



**HAL**  
open science

# Développement d'une évaluation génomique pour l'analyse de données longitudinales : application aux contrôles élémentaires chez les caprins laitiers

Mathieu Arnal

► **To cite this version:**

Mathieu Arnal. Développement d'une évaluation génomique pour l'analyse de données longitudinales : application aux contrôles élémentaires chez les caprins laitiers. Sciences agricoles. Institut National Polytechnique de Toulouse - INPT, 2019. Français. NNT : 2019INPT0124 . tel-04169801

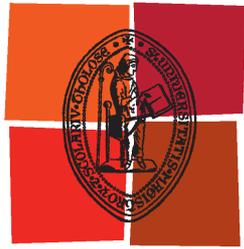
**HAL Id: tel-04169801**

**<https://theses.hal.science/tel-04169801>**

Submitted on 24 Jul 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Université  
de Toulouse

# THÈSE

En vue de l'obtention du

## DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

**Délivré par :**

Institut National Polytechnique de Toulouse (Toulouse INP)

**Discipline ou spécialité :**

Pathologie, Toxicologie, Génétique et Nutrition

---

**Présentée et soutenue par :**

M. MATHIEU ARNAL

le mercredi 11 décembre 2019

**Titre :**

Développement d'une évaluation génomique pour l'analyse de données longitudinales : application aux contrôles élémentaires chez les caprins laitiers

---

**Ecole doctorale :**

Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries (SEVAB)

**Unité de recherche :**

Génétique, Physiologie et Systèmes d'Elevage ( GENPHYSE)

**Directeur(s) de Thèse :**

MME CHRISTELE ROBERT-GRANIE

MME HELENE LARROQUE

**Rapporteurs :**

Mme EVA UGARTE SAGASTIZABAL, NEIKER TECNALIA

M. NICOLAS GENGLER, UNIVERSITE DE LIEGE

**Membre(s) du jury :**

M. ANDRES LEGARRA, INRA TOULOUSE, Président

M. DENIS LALOE, AGROPARISTECH, Membre

M. LUIZ BRITO, PURDUE UNIVERSITY EU, Membre

Mme CHRISTELE ROBERT-GRANIE, INRA TOULOUSE, Membre

Mme HELENE LARROQUE, INRA TOULOUSE, Membre



## Résumé

L'amélioration génétique des caprins laitiers est basée, à ce jour, sur la mesure de la quantité et de la qualité de la production laitière des femelles en ferme, à intervalles de 4 à 5 semaines au cours de la lactation, selon des protocoles stricts. L'évaluation génétique de la quantité de lait repose sur l'estimation de la quantité de lait totale produite par lactation. Cette sélection à partir de la quantité totale de lait à la lactation a tendance à sélectionner des animaux avec une production au pic de lactation de plus en plus élevée. Une production importante en début de lactation peut être à l'origine, pour les femelles, de problèmes métaboliques. De plus, en caprins laitiers, dans un contexte de production saisonnée, une production laitière qui se maintient après le pic, i.e. persistante, permettrait une production laitière plus étalée, en lien avec les attentes du marché. Il y a donc un intérêt zootechnique et économique de vouloir sélectionner des chèvres laitières plus persistantes.

Dans notre étude, l'approche consiste à modéliser la forme de la courbe de lactation à partir des informations recueillies lors de chaque contrôle en ferme. Les modèles permettant l'analyse de telles données longitudinales sont généralement appelés modèles de contrôles élémentaires. L'un des principaux intérêts, est une meilleure prise en compte d'effets d'environnement, affectant la production le jour du contrôle, avec un effet troupeau-jour de contrôle ne dépendant que des animaux présents lors du contrôle. Le deuxième avantage de ce type de modèle réside dans la modélisation de la plupart des effets génétiques et d'environnement sous forme de courbes ; il serait donc possible de sélectionner les animaux avec la meilleure valeur génétique pour la persistance. La mise au point de ces modèles nécessite l'étude préalable des effets d'environnement affectant la production laitière au cours du temps.

Suite à une analyse descriptive détaillée des courbes de lactation des deux principales races caprines françaises (Alpine et Saanen), nous avons montré qu'il existait une variabilité de la forme de courbes de lactation, et en particulier que le mois de mise bas était impliqué dans les différentes formes de courbe. Nous avons ensuite proposé un modèle de régression aléatoire, proche de celui développé chez les bovins laitiers français. La modélisation proposée permet d'obtenir directement deux index génétiques : un premier correspondant à la valeur génétique de l'animal pour la quantité totale de lait au cours de la lactation et un deuxième correspondant à une valeur génétique de la persistance laitière de l'animal, et cela sans corrélation entre les deux. Le modèle développé permet de prendre en compte de façon disjointe les chèvres en primipares et les chèvres en multipares. Nous avons également étudié les corrélations entre index de différents caractères, au cours de la lactation ainsi que les corrélations entre persistance et fertilité à l'IA ou entre persistance et longévité. Dans une dernière partie, nous avons étendu le modèle d'évaluation génétique des contrôles élémentaires à un modèle d'évaluation génomique (Single-step GBLUP) permettant d'exploiter l'ensemble des informations moléculaires disponibles (génotypes SNP 50K). Une validation de ce modèle et une comparaison au modèle actuel ont été réalisées. Les différences de moyennes des index estimés par année de naissance des boucs constituaient la principale différence entre le modèle Single-step GBLUP RRM et le Single-step GBLUP modèle à la lactation utilisé actuellement. Enfin, à partir du modèle Single-step GBLUP nous avons mis en évidence quelques régions du génome intéressantes liées à la persistance laitière.



## Abstract

Genetic improvement of dairy goats is based on the measurement of the quantity and quality of milk production of females on farms, at intervals of 4 to 5 weeks during lactation, according to strict protocols. The genetic evaluation of the quantity of milk is based on the estimation of the total quantity of milk produced per lactation. This selection based on the total quantity of milk in lactation tends to select animals with increasingly high peak lactation production. High production at the beginning of lactation can cause metabolic problems for females. In addition, in the case of dairy goats, in a context of seasonal production, a milk production that is maintained after the peak, i.e. persistent, would allow a more spread out milk production, in line with market expectations. There is therefore a zootechnical and economic interest in wanting to select more persistent dairy goats.

In our study, the approach consists in modelling the shape of the lactation curve based on the information collected during each farm test. Models allowing the analysis of such longitudinal data are generally called test-day models. One of the main interests is to take better account of environmental effects, affecting production on test-day, with a herd-test-day effect depending only on the animals present during the test. The second advantage of this type of model is that most genetic and environmental effects are modelled as curves, so it would be possible to select animals with the best genetic value for persistence. The development of these models requires the prior study of the environmental effects affecting milk production over time.

Following a detailed descriptive analysis of the lactation curves of the two main French goat breeds (Alpine and Saanen), we showed that there was a variability in the shape of the lactation curves, and in particular the month of calving was involved in the different curve shapes. Then we proposed a random regression model, similar to that developed in French dairy cattle. The proposed modeling makes it possible to obtain two genetic indexes directly: one corresponding to the genetic value of the animal for the total quantity of milk during lactation and a second corresponding to a genetic value of the animal's milk persistency, without correlation between the two. The model proposed is more relevant than the current one because it takes into account, in a disjointed way, the goats in primiparous and the goats in multiparous. We also studied correlations between indexes of different traits during lactation and correlations between persistence and AI fertility or between persistence and longevity. In the last part of the thesis, we extended the genetic evaluation of test-day to a genomic evaluation model (Single-step GBLUP) allowing to exploit all available molecular information (genotyping SNP 50K). A validation of this model and a comparison with the current model was carried out. The main difference between the Single-step GBLUP RRM and the Single-step GBLUP lactation model currently in use was the differences in the averages of the estimated indexes per year of birth of the bucks. Finally, from the Single-step GBLUP model we have identified some interested regions of the genome linked to milk persistence.



## Remerciements

On dit souvent qu'une thèse ne se fait pas seul et c'était évidemment vrai pour la mienne. Je tiens à remercier les personnes qui m'ont accompagné ces 3 dernières années et en particulier mes encadrantes de thèse Hélène Larroque et Christèle Robert-Granié qui ont su me faire confiance et me guider tout en me laissant la liberté d'explorer et d'entreprendre, en étant disponibles et bienveillantes, depuis mon stage de fin d'étude.

Merci à Vincent Ducrocq qui a su être présent tout au long de la thèse, de près ou de loin, qui m'a rassuré et aidé à progresser à des moments clefs.

Merci aux rapporteurs Nicolas Gengler et Eva Ugarte Sagastizabal ainsi qu'aux examinateurs Luiz Brito, Denis Laloé et Andrés Legarra d'avoir pris le temps de lire le manuscrit et de participer à la soutenance malgré leur emploi du temps très chargé. Merci à Andrés d'avoir présidé ce jury. L'après-midi de soutenance a été très agréable et j'en garde un très bon souvenir.

Je tiens à remercier les membres de mon comité de thèse pour leur regard extérieur qui a été précieux pour mener à bien cette thèse : Carine Capel, Virginie Clément, Sébastien Déjan, Vincent Ducrocq, Mohammed El Jabri, Hélène Leclerc, Sophie Mattalia, et Isabelle Pailhière.

Merci à l'Institut de l'élevage de m'avoir permis de réaliser cette thèse. Merci en particulier à Sophie Mattalia de m'avoir accompagné avec souplesse et de continuer à me faire confiance pour la suite. Merci également à l'antenne de l'Institut de l'élevage de Toulouse de m'avoir accueilli durant une partie de la thèse : Agnès, Annette, Aurélie, Catherine, Carole, Diane, Emmanuel, Emmanuelle, Florence, Gilles, Jacques, Jean, Jean-Christophe, Jean-Luc, Jean-Michel, Jérôme, Jocelyn, Mireille, Valérie, Yoann.

Merci à l'ensemble de l'unité GenPhySE et en particulier à l'équipe GeSPR : Carole, Charlotte, Christel, Christophe, Claire, Dominique F, Dominique H, Estelle, Flavie, Isabelle, Jean-Louis, Rachel, Sophie, Tiphaine, ainsi que l'équipe MG<sup>2</sup> : Andrés, Anne, Carolina, David, Fernando, Jean-Michel, Marc, Silvia, Zulma, qui ont su m'aider à améliorer ma thèse. Merci à Evelyne, Florence, Manuela, Nancy, et Valérie pour les tâches administratives. Merci à Alexandre, Davi et Frédéric pour nos belles discussions du midi et nos joyeux repas.

Merci à l'équipe Gaby de Jouy en Josas pour la qualité de leur accueil lors de mon séjour dans leur unité.

Merci à Karin Meyer, Ignacy Misztal et son équipe, Vincent Ducrocq, développeurs des logiciels d'estimation des paramètres génétiques et d'évaluation génétique utilisés durant cette thèse.

Je tiens à remercier Apis Gene, l'ANRT, l'INRA et l'institut de l'élevage d'avoir financé ma thèse.

Merci au GAEC Mordesson de m'avoir fait découvrir l'élevage caprin en stage d'école d'ingénieur ce qui m'a permis de mieux comprendre le contexte de ma thèse.

Merci à mes amis qui m'ont accompagné jusqu'à aujourd'hui et notamment mes colocataires toujours ressourçants.

Merci tout particulièrement à ma famille qui m'a toujours soutenu, qui rend mon quotidien très agréable et qui m'a énormément transmis notamment la passion de l'agriculture et de l'élevage. Les moments passés à la ferme m'ont permis, bien des fois, de souffler et de recharger les batteries.

Merci à Mélodie, de m'avoir toujours soutenu, écouté, et rassuré pendant ces années.



# Table des matières

Résumé.....	2
Abstract.....	4
Remerciements.....	6
Table des matières.....	8
Liste des abréviations.....	14
Introduction générale .....	16
I. Contexte de la filière caprine française.....	16
II. La sélection caprine française.....	18
III. Les données du contrôle laitier en race Alpine et Saanen.....	26
IV. L'évaluation génétique des caractères de production laitière et des cellules somatiques en races Alpine et Saanen .....	28
1. Les phénotypes à la lactation .....	28
2. Le modèle d'évaluation génomique à la lactation mis en œuvre pour les caractères de production laitière et les scores de cellules somatiques en race Alpine et Saanen depuis 2018.....	30
a. Description des effets fixes du modèle d'évaluation génomique.....	32
b. Les paramètres génétiques .....	33
c. Limites du modèle d'évaluation génétique mis en œuvre aujourd'hui .....	34
V. Modèles de contrôles élémentaires (Test day models) .....	34
1. Les modélisations pour la partie fixe .....	41
2. Modélisation pour les effets génétiques et environnements permanents .....	41
3. Limites des modèles de régression aléatoire.....	43
Chapitre 1: Etude de la diversité phénotypique des courbes de lactation des deux principales races caprines françaises .....	46
I. Diversité des formes de courbes de lactation des chèvres Alpine et Saanen françaises (Article1) .....	46
II. Analyses et résultats complémentaires .....	59
1. Analyse des profils moyens des courbes de lactation en fonction du nombre de clusters retenus pour la classification.....	59

2. Evolution des formes de courbes phénotypiques de lactation au cours des années .....	61
3. Comparaison des courbes de production au cours de la lactation pour les races Alpine et Saanen .....	64
Chapitre 2 : Estimation des paramètres génétiques pour les caractères de production laitière des chèvres en première lactation à l'aide d'un modèle de régression aléatoire .....	68
I. Paramètres génétiques des caractères de production laitière des chèvres Alpine et Saanen en utilisant un modèle de régression aléatoire (Article 2) .....	69
II. Analyses et résultats complémentaires .....	93
1. L'estimation de la production à la lactation par la méthode Fleischmann est-elle comparable à celle issue d'un modèle de régression aléatoire (RRM)? .....	93
2. Difficultés rencontrées lors de la réduction des matrices de variances-covariances génétique et d'environnement permanent .....	94
3. Définition de l'effet troupeau (HTD).....	96
Chapitre 3 : Estimation des paramètres génétiques des caractères de production laitière et des comptages de cellules somatiques des chèvres Alpine et Saanen françaises au cours des trois premières lactations à l'aide d'un modèle de régression aléatoire .....	104
I. Estimation des paramètres génétiques des caractères de production laitière et des comptages de cellules somatiques des trois premières lactations à l'aide d'un modèle de régression aléatoire chez les chèvres de race Alpine (Article 3).....	105
II. Estimation des paramètres génétiques des caractères de production laitière et des comptages de cellules somatiques des trois premières lactations à l'aide d'un modèle de régression aléatoire chez les chèvres de race Saanen.....	131
III. Etudes complémentaires réalisées lors de la mise en œuvre du modèle de régression aléatoire .....	140
1. L'effet Troupeau-Jour de contrôle (HTD) doit être en interaction avec le rang de lactation .....	141
2. Ajustement des covariables de Legendre pour combiner les rangs de lactation 2 et 3 .....	144
IV. Analyses complémentaires sur les variances génétiques et les effets de milieux issus des modèles LEG_FT et EGV_FT pour les caractères étudiés (Lait, MG, MP, TB, TP, SCS et ratio TB/TP) en race Alpine et Saanen.....	145
1. Etude complémentaire sur les variances génétiques .....	145

a. Pourcentage de variance génétique représenté par les trois premières valeurs propres des matrices de variances-covariances issues du modèle LEG_FT par rang de lactation en race Alpine et Saanen .....	145
b. Profils des fonctions propres issues du modèle EGV_FT et variabilité génétique du niveau et de la persistance pour les différents caractères étudiés en race Alpine et Saanen.....	147
c. Représentation de courbes phénotypiques pour des animaux extrêmes.....	147
2. Analyses complémentaires des effets fixes du modèle LEG_FT.....	148
V. Etude des relations entre caractères de durabilité (fertilité à l'IA, survie, longévité et durée de lactation) et les index, niveau et persistance génétique et d'environnement permanent des caractères de production et des comptages de cellules somatiques (SCS), issus du modèle EGV_PM en race Alpine et Saanen.....	156
1. Etude des index génétiques et d'environnement permanent des caractères de production (Lait, MG, MP, TB, TP, ratio TB/TP) et des comptages de cellules somatiques (SCS) selon les modalités de la durée de lactation en race Saanen.....	160
2. Etude des index génétiques et d'environnement permanent des caractères de production (Lait, MG, MP, TB, TP, ratio TB/TP) et des comptages de cellules somatiques (SCS) selon les modalités de la survie en race Saanen .....	162
3. Etude des index génétiques et d'environnement permanent des caractères de production (Lait, MG, MP, TB, TP, ratio TB/TP) et des comptages de cellules somatiques (SCS) selon les modalités de la réussite à l'insémination artificielle en race Saanen.....	166
4. Evolution de l'infertilité, de la survie, de la durée de lactation selon l'évolution des index niveau et persistance pour la production laitière.....	168
5. Etude des relations entre la longévité et les index génétiques de niveau et de persistance pour la production laitière.....	171
Chapitre 4 : Etude de la faisabilité d'une évaluation génomique basées sur les contrôles élémentaires de chèvres Alpine et Saanen à partir d'un modèle de régression aléatoire.....	176
I. Principes d'une validation d'un modèle d'évaluation génomique .....	177
II. Validation d'une évaluation génomique sur les contrôles élémentaires en caprin laitier à partir d'un modèle de régression aléatoire et analyse des effets des SNPs issus d'un modèle ssGBLUP .....	178
III. Résultats complémentaires sur l'analyse des effets des SNPs issus des modèles ssGBLUP RRM et ssGBLUP LM pour l'analyse des caractères de production laitière et scores de cellules somatiques en race Alpine et Saanen.....	225

1. Analyse des effets des allèles des SNPs significativement associés à la persistance pour la MG et la MP en Saanen et les SCS en Alpine .....	229
2. Evolution des effets des allèles de SNPs associées à plusieurs caractères pour le niveau et la persistance sur les chromosomes 6 et 19.....	230
3. Analyse des temps de calcul et de la mémoire utilisés pour l'exécution d'une évaluation génomique ssGBLUP basée sur les contrôles élémentaires et d'une évaluation ssGBLUP basée sur des phénotypes à la lactation.....	235
Discussion générale et perspectives.....	240
I. Le modèle RRM est bien adapté à l'évaluation génétique de la production globale à la lactation .....	240
II. L'évaluation génétique de la persistance par le modèle RRM.....	243
III. Utilisation des valeurs génétiques produites par le RRM en sélection .....	247
1. Corrélations entre index de production et de SCS .....	247
2. Corrélations entre index de production, SCS et fertilité .....	249
3. Corrélations entre index de production, SCS et longévité .....	253
4. Index issus du RRM et objectifs de sélection .....	254
IV. Utilisation des solutions des effets fixes produites par le RRM en appui technique .	257
V. Perspectives.....	258
Liste des tableaux.....	262
Liste des figures .....	266
Références.....	274





## Liste des abréviations

AOP : Appellation d'origine contrôlée

BLUP: Meilleur prédicteur linéaire non biaisé (Best Linear Unbiased Predictor)

CCS: Comptages de cellules somatiques

DGAT1: Diacylglycerol O-Acyltransferase 1

DYD: Daughter Yield Deviation

EBV: Valeur génétique estimée (Estimated Breeding Value)

FPR : Ratio Taux Butyreux/Taux Protéique (Fat-Protein ratio)

GBLUP: Meilleur prédicteur linéaire génomique non biaisé (Genomic Best Linear Unbiased Predictor)

GEBV: Valeur génomique estimée (Genomic Estimated Breeding Value)

GWAS: Etude association pangénomique (Genome Wide Association Study)

HTD : Effet troupeau jour de contrôle (Herd Test-day)

IA: Insémination artificielle

ICC: Index combiné caprin

IMC: Index morphologique caprin

IPC: Index de production caprine

LM : Modèle à la lactation (Lactation model)

LSCS: Performance à la lactation du score de cellules somatiques corrigé

MG: Quantité de matière grasse

MP: Quantité de matière protéique

NE : Nord-Est

NO : Nord-Ouest

QTL: Locus de caractères quantitatifs (Quantitative Trait Locus)

RRM : Modèle de régression aléatoire (Random regression model)

SCS: Score de cellules somatiques (Somatic cell score)

SE : Sud-Est

SO : Sud-Ouest

SNP: Polymorphisme d'une seule base nucléotidique (Single Nucleotide Polymorphism)

ssGBLUP: BLUP génomique en une étape (Single-step GBLUP)

TB : Taux butyreux

TDM : Modèle de contrôles élémentaires (Test day model)

TP : Taux protéique



# Introduction générale

## I. Contexte de la filière caprine française

La chèvre a été domestiquée il y a 10 000 ans, principalement pour la consommation de son lait (Vigne and Helmer, 2007). Aujourd'hui, les fromages de chèvres constituent l'essentiel des débouchés de la production laitière de la filière caprine française (120 000 tonnes). Les fromages affinés représentent 79% de la production laitière et les bûches et bûchettes représentent 75% de ces fromages.

En 2016, 4900 élevages avaient plus de 10 chèvres avec une activité laitière :

- 48% des élevages sont des livreurs, plutôt en plaine, avec 237 chèvres en moyenne (ce qui représente 72% de la population des chèvres et 79% du volume) et seulement 34 % d'entre eux sont spécialisés (une majorité est donc en polyculture-élevage) ;
- 47% des élevages sont des fromagers, ils ont 70 chèvres en moyenne et 50% d'entre eux sont spécialisés avec des contextes pédoclimatiques plus difficiles (soit 225 000 chèvres, 21% des chèvres et 15% du volume)

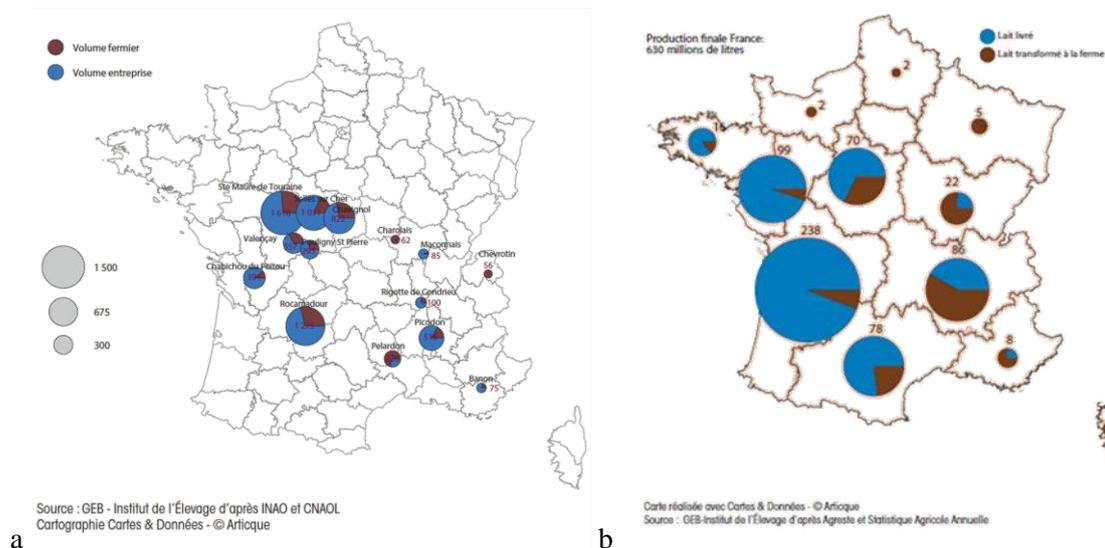
Il existe 14 Appellations d'Origine Contrôlée (AOP) fromagères en France qui représentent 6% du volume de lait. La Figure 1.a. présente la répartition et les volumes de lait des différentes AOP sur le territoire français. Les fromages en agriculture biologique représentent 1398 tonnes en 2018 contre 372 tonnes en 2010.

Le nombre d'élevages est passé de 200 000 en 1960, avec 4.5 chèvres en moyenne, à 4 900 élevages en 2016 avec en moyenne 150 chèvres. Après une forte croissance de la production entre 2000 et 2010, la filière a subi une crise importante de 2011 à 2014 en raison de la flambée du coût des matières premières qui a touché en premier les livreurs. Entre 2000 et 2016, la moitié des exploitations ont disparu (soit 25% des exploitations livreurs de Poitou Charentes), mais l'effectif de chèvres est resté constant et la production a augmenté de 27%. Malgré l'intensification de la production, la filière importe encore entre 9% (entre 2010-2012) et 21% (en 2017) de l'approvisionnement industriel en provenance principalement d'Espagne et des Pays-Bas (Prache et al., 2018). La tendance à l'agrandissement des élevages (jusqu'à 3000 chèvres) est une source de fragilité de la filière avec des risques sanitaires possibles par la concentration d'animaux, un montant de l'exploitation agricole très élevé limitant la reprise de l'exploitation et une mauvaise image auprès des consommateurs qui préfèrent des petites structures (Prache et al., 2018).

En Europe, la France est le premier producteur de lait mais possède le 4<sup>ème</sup> cheptel avec 770 000 chèvres, loin derrière la Grèce (3.9 millions), l'Espagne (3 millions) et la Roumanie (1.4 millions) (Maigret, 2018a).

En France, il existe 3 principaux bassins de production (Prache et al., 2018) comme le montre la Figure 1.b:

- Le bassin du grand-Ouest et de l’Aveyron (60% du volume national), 95% du lait est livré et 77% des livreurs utilisent des fourrages conservés (foin (41%), enrubannage (4%), ensilage de maïs (12%)). Les chèvres sont en majorité alimentées toute l’année en bâtiment. Les quantités de concentré ingérées sont en moyenne de 629g de concentrés/litre de lait (Idele, 2019a).
- Le bassin Rhône-Alpes-Bourgogne (17% du volume national), où 52% de la production est transformée à la ferme et 41% des fromagers sont pastorants ou pâturants (25% de pâturant). Les quantités de concentré ingérées pour les fromagers sont en moyenne de 400g de concentrés/litre de lait (Idele, 2019a).
- Le bassin Centre-Val de Loire (11% du volume national) possède 5 AOP représentant 60% du volume des AOP, 65% du volume produit est livré à des laiteries et le pâturage est plus présent que dans l’Ouest. Les quantités de concentré ingérées sont en moyenne de 607g de concentrés/litre de lait (Idele, 2019a).



**Figure 1 : Zones de production de fromages AOP, et volumes de lait associés (a), et répartition des volumes de lait produit par régions administratives et type de destination (livré à des laiteries ou transformé à la ferme) (b) (Maigret, 2018b)**

La production est très fortement saisonnée avec un pic de production au printemps et une production très faible en hiver en raison du saisonnement de la reproduction des chèvres qui a lieu naturellement en septembre. Le prix du lait aux mille litres est inversement proportionnel à l’abondance de la production de lait (Maigret, 2018a). La Figure 2 présente l’évolution des quantités produites et du prix du lait au fil de l’année. Afin d’étaler la période de production les éleveurs ont recours au désaisonnement avec l’utilisation de traitements lumineux et/ou d’hormones ou ont recours aux lactations longues afin d’étendre la période de production.

Depuis le début des années 1960, l’accent était mis sur l’augmentation de la production par animal et exploitation. La filière caprine fait face actuellement à de nouveaux défis. En effet, la filière reste fragile

par sa faible autonomie alimentaire qui n'est que de 55% en raison de l'utilisation importante de concentrés et la part très faible de pâturage. L'image de la production de lait de chèvre perçue par le consommateur est en décalage avec la réalité. En effet en 2013, 98% des consommateurs pensaient que les chèvres étaient élevées tout ou une partie de l'année en extérieur et 42% pensaient que l'importation de lait était mauvaise pour la filière (Prache et al., 2018). Aujourd'hui, comme toutes les filières animales, poussée par des associations environnementalistes et welfaristes, la filière caprine s'intéresse à son impact sur l'environnement (émission de gaz à effets de serre en lien avec le changement climatique, pollution de l'eau et des sols (effluents d'élevage)), à la concurrence entre l'alimentation humaine et l'alimentation animale (consommation de céréales, de soja cultivé sur des terres déforestées puis importé), au bien-être animal (associations welfaristes qui demandent plus de pâturage), à la santé du consommateur (diminution ou arrêt de l'utilisation d'antibiotiques et d'hormones), aux qualités organoleptiques des produits (ANICAP, 2015; ANICAP and Interbev, 2018).

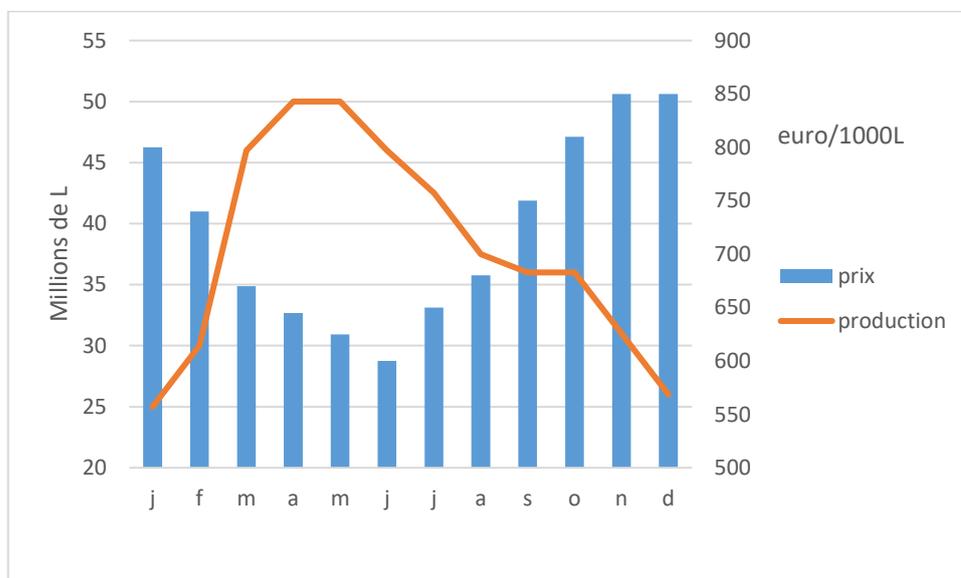


Figure 2: Evolution des quantités totales de lait produit et du prix du lait par mois en 2018 (Idele, 2019b)

## II. La sélection caprine française

La sélection caprine française a débuté avec la loi sur l'élevage de 1966 qui a mis en place l'UPRA (unité nationale de sélection et de promotion des races). Cette dernière regroupe les races Alpine chamoisée, Saanen et Poitevine (DANCHIN-BURGE and DUCLOS, 2009). L'Alpine et la Saanen, à poils court, ont été préférées par les éleveurs à la Poitevine qui a un poil plus long, pour des raisons sanitaires et de traite en bâtiment (Roche, 2008). Le nombre de Saanen a augmenté dans les années 80, car elle avait la réputation de valoriser de grande quantité de fourrage et d'être moins difficile dans des systèmes spécialisés (ensilage ou paille, déshydratés et concentrés) (Minot, 2002). Aujourd'hui, en France, l'Alpine et la Saanen sont les races laitières majoritaires (59,1% et 37,6% des données du contrôle laitier) à côté des races à plus faible effectif : la Chèvre des Fossés, la Chèvre de Savoie, la Chèvre du

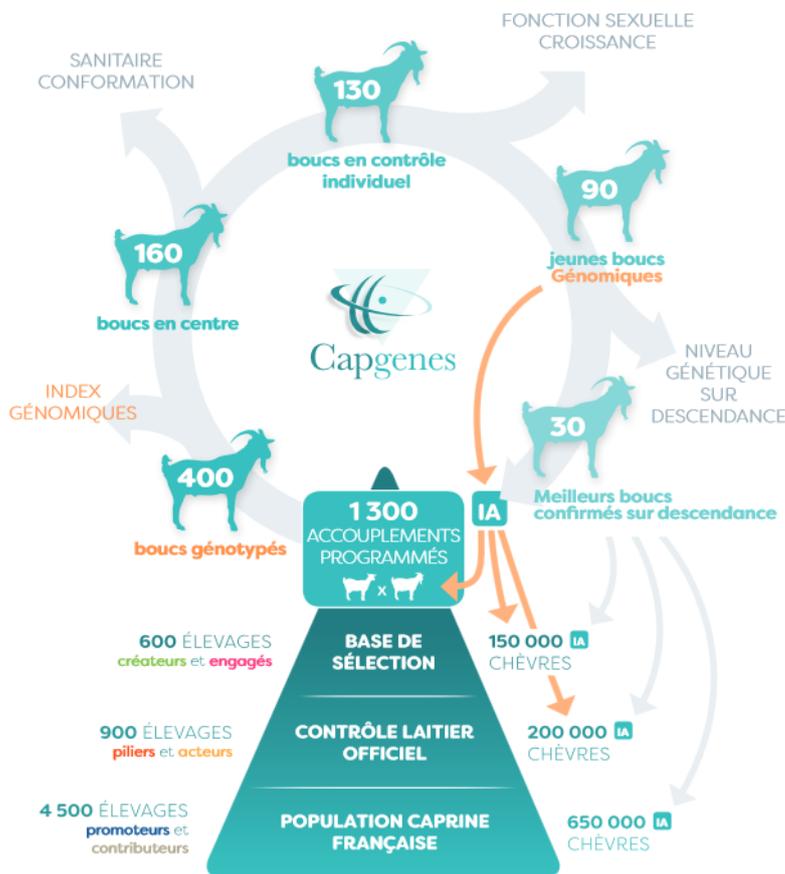
Massif Central, la Poitevine, la Provençale, la Pyrénéenne et la Rove, dont la production laitière est bien moins importante (Thomas, 2019). Le programme de sélection porté par Capgenes (seul organisme et entreprise de sélection française) consiste à l'amélioration génétique de ces deux races majoritaires en utilisant comme reproducteurs, les animaux avec les valeurs génétiques les plus élevées pour les caractères que l'on souhaite améliorer.

Initialement, l'objectif de sélection était basé principalement sur la quantité de MP par lactation puis le TP a pris de l'importance et a été inclus dans l'objectif de sélection. Un index de synthèse (ICC= index combiné caprin) combinant ces deux critères a été créé en 1996 (ICC = Index Matière Protéique + 0.4 Index Taux Protéique). A partir de 1999, la MG et le TB ont été ajoutés dans l'objectif de sélection afin de favoriser un rapport TB/TP optimum pour la qualité fromagère (ICC = Index Matière Protéique + 0.4 Index Taux Protéique + 0.2 Index Matière Grasse + 0.1 Index Taux Butyreux). Cet index est aujourd'hui appelé l'IPC (index production caprin).

Depuis 2006, les 2 races ont également un index synthétique combinant les caractères de morphologie de la mamelle (IMC : index morphologique caprin) (IMC= Profil de la mamelle + hauteur du plancher + Largeur de l'attache Arrière + Orientation des Trayons + Avant pis). En 2010, l'IMC et l'IPC ont été combinés pour former un nouveau ICC. En Saanen, il est égal à  $IPC + 0.6 IMC$  et en Alpine à  $IPC + 0.5 IMC$ . Les comptages de cellules somatiques ont aussi été prises en compte dans la sélection des boucs mais n'est toujours pas pris en compte dans l'objectif de sélection (Clément et al., 2015a) afin d'y intégrer la résistance aux mammites. Ces ajouts ont été faits sans pondération économique mais ont été confirmés par une approche économique réalisée en 2015 (Palhiere et al., 2015). L'objectif de préservation de la variabilité génétique au sein des races a conduit à prendre en compte le degré d'apparement des pères et des mères des futurs boucs d'insémination à partir de 2002 (Palhiere et al., 2014).

Avec l'arrivée de nouvelles technologies d'analyse du génome et de typage à haut débit, un consortium international (International Goat Genome Consortium (IGGC)), a été créé en 2010 afin de développer une puce à SNP pour l'espèce caprine. Une puce caprine 50k, l'Illumina goat SNP50k BeadChip (Tosser-Klopp et al., 2014) est désormais disponible et l'essentiel des boucs de testage français nés depuis la fin des années 90 ont été génotypés avec cette puce depuis 2012 (Carillier et al., 2013). Les travaux de thèse de Carillier-Jacquin (2015) et Teissier (2018) ont permis de tester et développer différents modèles d'évaluation génomique. Les résultats de ces travaux ont montré la supériorité du ssGBLUP (modèle génomique basé sur l'information moléculaire SNP couplée aux informations pedigree) sur le BLUP (modèle génétique basé uniquement sur l'information pedigree) pour prédire les valeurs génétiques des animaux. L'information moléculaire dense des marqueurs SNP permet donc une estimation plus précise de la valeur génétique d'un individu dès la naissance. Le schéma de sélection a alors été modifié en 2018. 1200 accouplements sont programmés à partir de mères à boucs sélectionnés au sein de 210 élevages adhérents à Capgenes (600 élevages soit 12% des élevages, et 160 000 chèvres soit 21% des

chèvres) et 400 jeunes boucs sont génotypés. A partir de ces index génomiques, 170 boucs sont sélectionnés. Environ 50 sont éliminés pour des problèmes sanitaires et 40 pour des problèmes de fonctions sexuelles et de croissance. Au final 90 boucs sur index génomiques sont utilisés en insémination. Une trentaine des boucs confirmés sur descendance sont utilisés pour les accouplements programmés. La Figure 3 présente ce schéma de sélection.



**Figure 3 : Schéma de sélection de l'organisme Capgenes (Capgenes, 2019)**

L'efficacité du schéma de sélection est évaluée grâce au progrès génétique annuel, i.e. la progression de la valeur génétique des reproducteurs retenus au fil des années, calculée de la façon suivante :  $\Delta G = \frac{iR\sigma_a}{\bar{T}}$  avec  $i$  l'intensité de sélection,  $R$  la précision de l'évaluation génétique et  $\sigma_a$  l'écart-type génétique et  $\bar{T}$  l'intervalle entre génération. L'intensité de sélection est définie comme la supériorité génétique exprimée en écart type génétique des animaux sélectionnés par rapport aux candidats. Pour augmenter l'intensité de sélection on peut augmenter le nombre de candidats, ce qui est couteux s'ils doivent être génotypés ou phénotypés, ou réduire le nombre d'animaux sélectionnés, ce qui est difficile car les boucs ne peuvent produire plus de 10 000 doses par an (Carillier-Jacquin, 2015). La précision de l'évaluation correspond à la corrélation entre les valeurs génétiques calculées par l'évaluation et la valeur génétique « vraie ». Plus la précision est élevée, plus le progrès génétique est important. La précision peut être améliorée en utilisant des modèles d'évaluation génétique qui prennent bien en compte les effets de

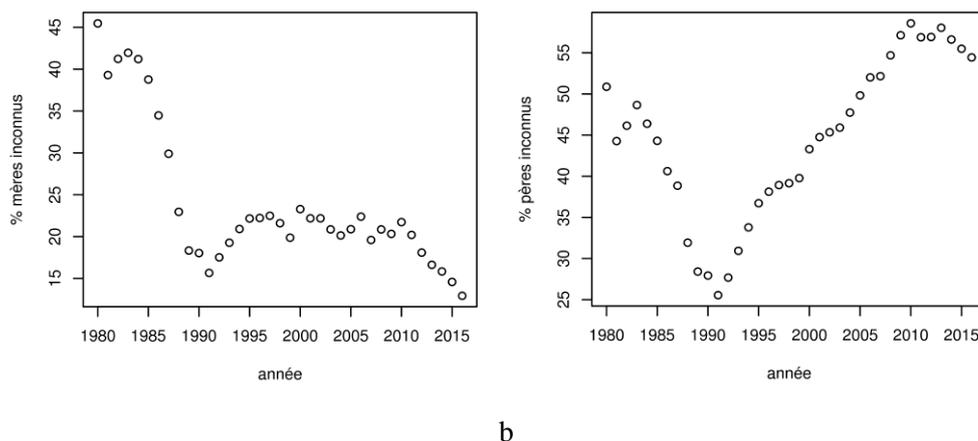
l'environnement et de la génétique, avec un pedigree précis et sur plusieurs générations, en mesurant de manière précise les phénotypes en ferme. L'écart type génétique correspond à la variabilité génétique du caractère. Plus la variance génétique est grande, plus il est facile de sélectionner des animaux extrêmes. L'intervalle entre génération correspond à l'âge moyen des parents au moment de la naissance de leurs descendants. Le testage sur descendance est une étape chronophage (en moyenne 3 ans), car il faut attendre les performances des filles de testage mais il aide à la précision de l'évaluation génétique. L'intervalle de génération dans le schéma caprin est en moyenne de 4.5 ans. L'emploi de l'information génomique aide à réduire l'intervalle de génération lorsqu'il se substitue au testage comme ce fut le cas en bovins laitiers (Schaeffer, 2006). En revanche, en caprin, cet intervalle est plus faible car les chèvres, dans la grande majorité, mettent bas à un an contre 2 ans, au plus tôt, pour les vaches laitières.

Plus un caractère est héritable, plus il sera facile de l'améliorer car la précision de l'évaluation sera élevée. L'héritabilité est le rapport entre la variance génétique et la variance phénotypique (i.e. la part génétiquement transmissible des différences entre animaux). L'amélioration génétique de l'ensemble des caractères peut s'avérer difficile car les caractères sont corrélés génétiquement entre eux positivement ou négativement comme nous pouvons le voir avec le Tableau 1. Les quantités (lait, MG et MP) sont très corrélées entre elles (au-dessus de 0.76), les taux également (au-dessus de 0.51), la MG et avec le TB (au-dessus de +0.49), les corrélations entre SCS et les autres caractères sont modérées (corrélation maximale de - 0.2 entre le TB et les SCS en Saanen), le TP et le lait ont une corrélation négative assez forte (-0.28).

**Tableau 1 : Corrélations génétiques entre les différents caractères laitiers pour les chèvres Alpine (triangulaire inférieure) et pour les chèvres Saanen (triangulaire supérieure) (Bélichon et al., 1999; Rupp et al., 2011)**

	Lait	MG	MP	TB	TP	SCS
Lait	.	0.76	0.92	-0.09	-0.28	0.12
MG	0.76	.	0.83	0.57	0.07	-0.02
MP	0.88	0.85	.	0.1	0.09	0.06
TB	-0.17	0.49	0.11	.	0.51	-0.2
TP	-0.28	0.14	0.18	0.61	.	-0.13
SCS	0	-0.13	-0.04	-0.2	-0.06	.

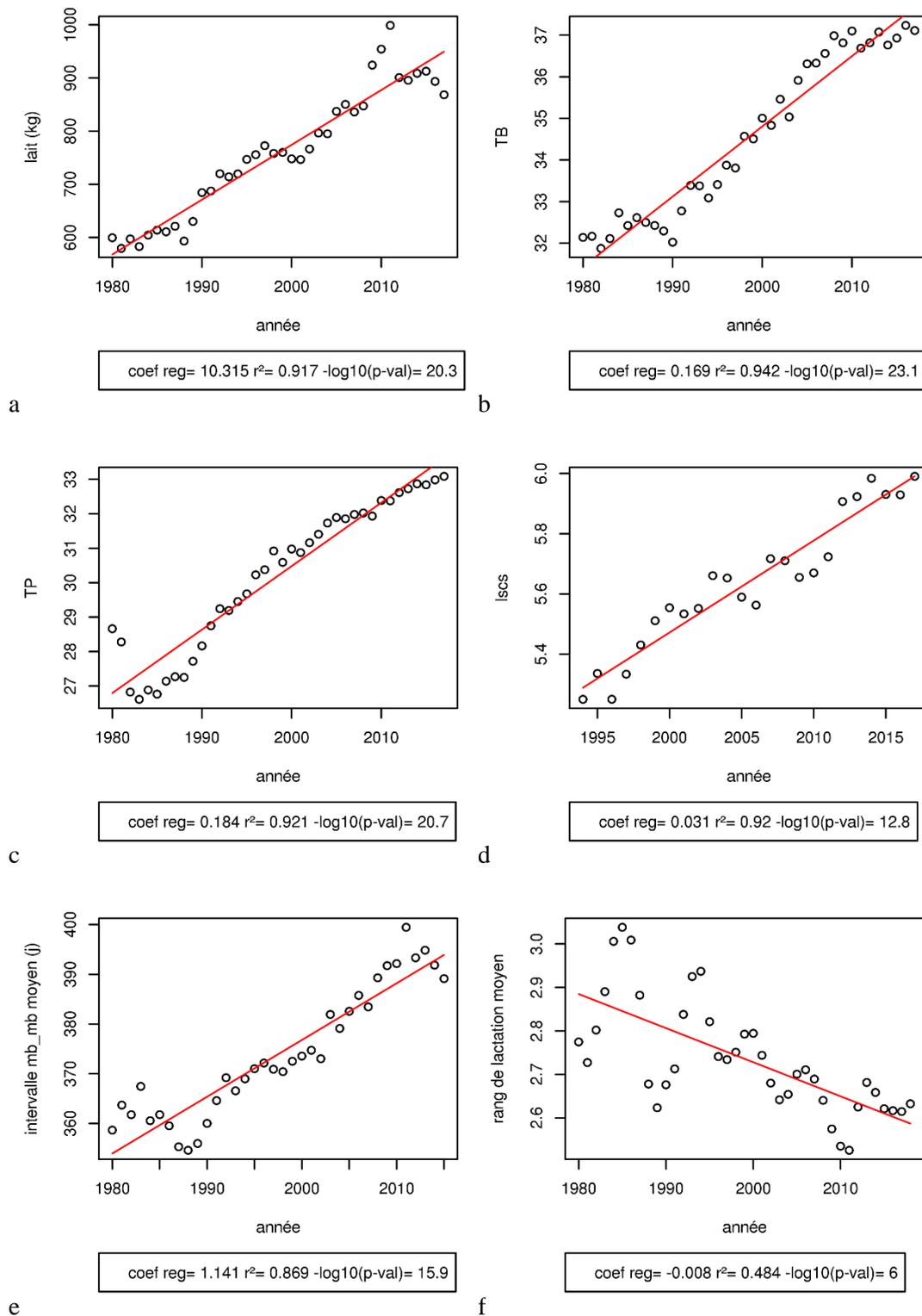
Un des enjeux de la filière caprine est la bonne connaissance des filiations car sans un pedigree complet il n'est pas possible de dissocier de la performance phénotypique l'effet de l'environnement et de la génétique. La Figure 4 présente l'évolution du pourcentage de mères inconnues et celui des pères inconnus au fil des années de naissance des animaux. Nous pouvons voir que le pourcentage de mères inconnues est passé de 45% en 1980 à 15 % en 1990 puis est remonté à 20%, il est aujourd'hui autour de 15%. En revanche pour le pourcentage de pères inconnus, il est passé de 50% dans les années 80 à 25% en 1990 pour remonter autour de 60% aujourd'hui.



**Figure 4: Evolution du pourcentage de mères inconnues (a) et de celui des pères inconnus (b) au fil des années de naissance (Saanen et Alpines confondues)**

En 2016, il y a eu 73 600 inséminations (soit 9% des chèvres) dans 844 élevages (soit 17% des élevages) (Maigret, 2018b) pour un taux de réussite de 62% en moyenne. L'insémination artificielle en caprin n'est pas aussi répandue qu'en bovin lait (80% d'insémination) ou en ovin lait (32.6% d'insémination) (FGE, 2010) mais le taux de réussite est plus élevé qu'en bovin lait.

Les estimations du progrès génétique montrent une progression de 13kg de lait par an et 0.1g/kg de TP et de TB (Piacère, 2008). La Figure 5 présente l'évolution phénotypique sur près de 40 ans de la quantité de lait moyenne à la lactation (a), du TB moyen à la lactation (b), du TP moyen à la lactation (c), du comptage cellulaire à la lactation (d), de l'intervalle mise bas-mise bas moyen (e) et du rang de lactation moyen (f). Les gains de productivité sont importants (+10kg de lait par an, +0.18g/l de TP par an), mais en même temps les intervalles entre mise bas ont augmenté (+1.14 jours par an). Ces phénomènes peuvent être une résultante des compromis entre fonctions de production et fonctions de reproduction comme en bovins (Lucy, 2001) ou à des changements de pratiques, i.e. lactations longues. Le rang moyen de lactation diminue également en élevage (-0.008 lactations par an) exprimant peut être là aussi un compromis entre production et longévité comme en bovins (Oltenacu and Algers, 2005) qui peut être associé à une dégradation du bien-être animal (Knaus, 2009).

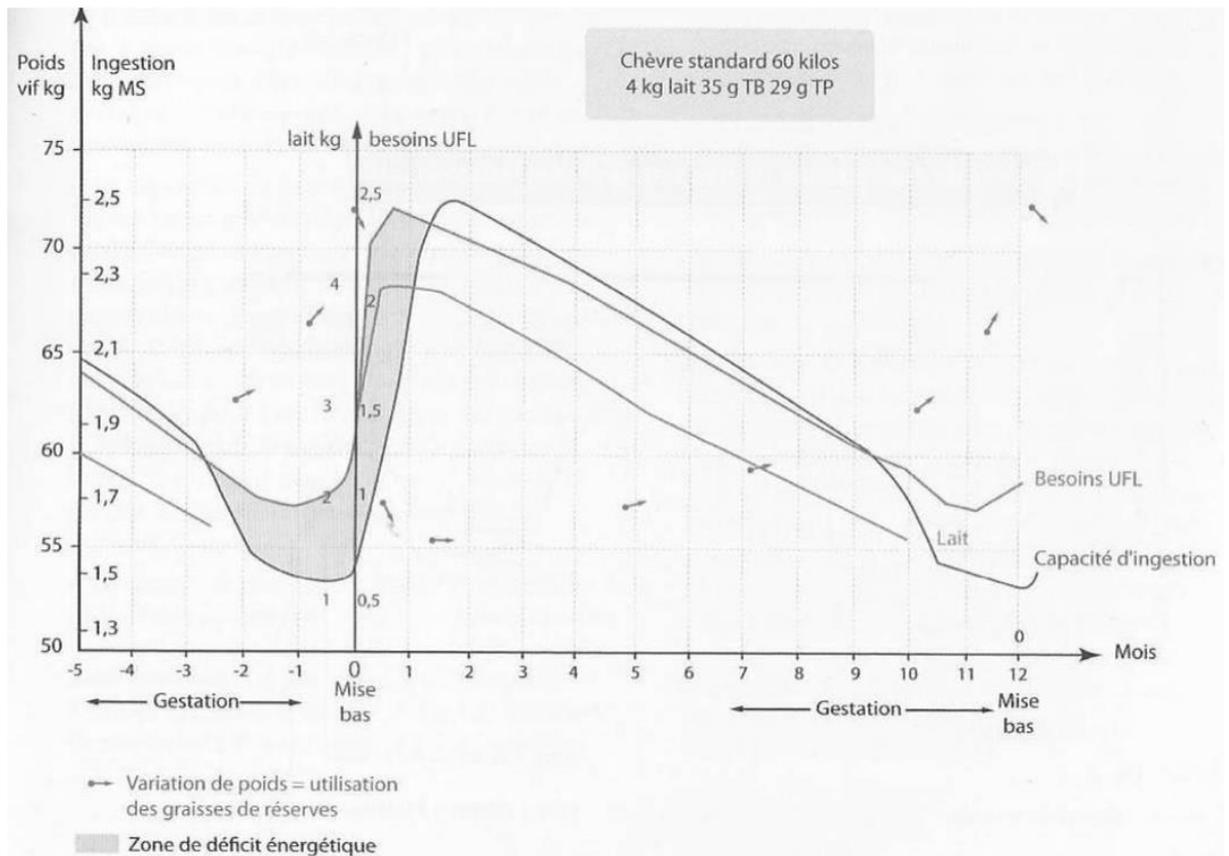


**Figure 5: Evolution de la quantité de lait moyenne à la lactation (a), du TB moyen à la lactation (b), du TP moyen à la lactation (c), du comptage cellulaire à la lactation (d), de l'intervalle mise-bas-mise bas moyen (e) et du rang de lactation moyen (f)**

(coef reg : coefficient de régression de la droite d'ajustement en rouge,  $r^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$ , p-val : p-value du coefficient de régression)

Face à la demande sociétale en matière d'économie des ressources, de bien-être animal et de qualité des produits, les nouveaux caractères d'intérêt pour la filière sont : la fertilité, la longévité, l'efficacité alimentaire, la composition fine des laits et la persistance laitière (Martin, 2014).

La persistance laitière est intéressante à plusieurs titres. Après la mise bas, une chèvre produit du lait pour nourrir son petit, sa production laitière augmente par la prolifération des cellules épithéliales sécrétrices de lait de la mamelle pour atteindre un pic ou un plateau autour de 40 jours puis cette production diminue par l'apoptose des cellules sécrétrices de lait (Strucken et al., 2015) jusqu'à environ 300 jours, la chèvre est alors tarie et remettra bas 1 an après la fois précédente. Un animal persistant, c'est-à-dire qui produit moins en début de lactation et plus à la fin de la lactation, comparé à un animal moyen produisant au total la même quantité de lait, est souhaitable car ses besoins alimentaires seront réduits au moment du pic de lactation (Dekkers et al., 1998) et par conséquent sa balance énergétique négative (balance énergétique = Énergie fournie par l'alimentation – énergie demandée pour la production laitière et pour le maintien des fonctions vitales) en début de lactation sera réduite (Muir et al., 2004). Une illustration de l'évolution des réserves corporelles, de l'énergie provenant de l'alimentation, et de l'énergie pour la production laitière et pour le maintien des fonctions vitales au cours de la lactation est donnée en Figure 6. Comme nous pouvons le voir, l'énergie demandée au pic de lactation est supérieure à l'énergie apportée par l'alimentation car les capacités d'ingestion sont limitées par la place prise par le fœtus dans le ventre de sa mère (Allen, 1996). La chèvre puise donc dans ses réserves corporelles. En bovins, un animal persistant est particulièrement recherché car si la balance énergétique négative est réduite l'animal peut avoir moins de problèmes métaboliques et être plus fertile (Jakobsen et al., 2003) d'autant plus que la mise à la reproduction dans un système de mise bas tous les ans intervient au bout de 3 mois de lactation. En caprins, la situation est différente puisque la reproduction, dans un système à une mise bas par an intervient au bout de 7 mois de lactation, c'est-à-dire en fin de lactation. Il peut donc y avoir un antagonisme entre persistance, i.e. production de lait importante en fin de lactation, et fertilité. Il est aussi possible d'étudier les variations des persistances d'autres caractères comme le TB ou le TP ou également le rapport TB/TP. En effet, il a été montré que l'évolution de ce rapport est intéressante en début de lactation car il est lié à la balance énergétique (Buttchereit et al., 2011, 2012) et que l'évolution du TB est intéressante par rapport à la fertilité (Albarrán-Portillo and Pollott, 2013).



**Figure 6: Evolution des réserves corporelles, de l'énergie provenant de l'alimentation, et de l'énergie nécessaire à la production laitière et aux fonctions vitales au cours de la lactation (Fragné, 2014; Chartier and Blain, 2018)**

Un autre avantage d'une persistance élevée à l'échelle d'un troupeau caprin ou d'une collecte laitière est que la production est mieux répartie au cours de la campagne de traite. Cela est particulièrement avantageux en caprins car la reproduction est saisonnée (en septembre) et induit donc une production de lait très saisonnée. Dans le cadre du projet Casdar Rustic (2016-2019), l'impact économique de la persistance a été évalué. Les bénéfices évalués étaient liés à une production différée dans le temps, induisant des prix plus élevés en fin d'année. Si toute la population est sélectionnée pour améliorer la persistance, on peut s'attendre à ce que le prix en fin d'année baisse.

Gengler (1996) a rapporté un grand nombre de paramètres permettant de mesurer la persistance qui sont plus ou moins corrélés à la quantité de lait à la lactation. Afin de sélectionner génétiquement de manière efficace la persistance laitière, il serait intéressant de déterminer des critères de persistance, non corrélés à la quantité totale de lait produit au cours de la lactation afin de pouvoir l'incorporer dans un index synthétique global de sélection.

L'objectif de mon travail de thèse est donc de s'intéresser à la modélisation de la forme de courbe de lactation à partir des informations recueillies lors de chaque contrôle en ferme et de pouvoir en déduire les composantes génétiques qui permettraient de sélectionner des chèvres plus persistantes.

### III. Les données du contrôle laitier en race Alpine et Saanen

1467 élevages (soit 30% des élevages caprins) étaient adhérents au contrôle laitier en 2017, cela représente une population de 386 836 chèvres (soit 50% des chèvres) pour une production moyenne de 953 kg de lait, 33 g/kg de TP et 37g/kg de TB en 312 jours (Idele, 2018). Le Tableau 2 présente un récapitulatif des chiffres clés de l'élevage caprin français actuel en 2018.

**Tableau 2: La filière caprine française : quelques chiffres clés**

	Livreurs	Fromagers	Adhérents contrôle laitier	Insémination	Adhérents Capgenes	Agriculture biologique	Total
Nombre d'élevages	2,352	2,303	1,467	844	600	845	4,900
Nombre de chèvres	554,400	161,700	386,836	73,000	160,000	53,284	770,000
% Elevage	48	47	30	17	12	17	100
% Chèvres	72	21	50	9	21	7	100

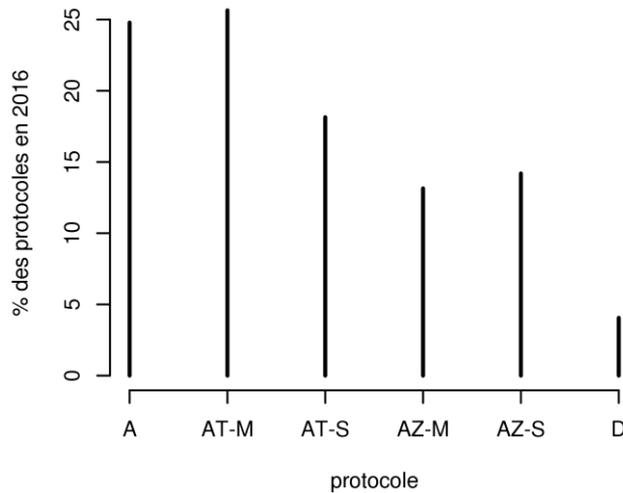
Les protocoles de contrôles de performances officiels sont agréés pour l'espèce caprine par l'organisme international ICAR (Kompan and Klopčic, 2004). Les chèvres sont considérées comme contrôlables si elles ont atteint leur 8ème jour de lactation (7 jours après la mise bas), jusqu'à leur tarissement.

Les caractères de production laitière et le comptage des cellules somatiques sont enregistrés dans le cadre du Contrôle Laitier Officiel. Le suivi des performances est réalisé grâce à 5 protocoles (A, D, AT, AZ, CZ). Pour les 4 protocoles, les enregistrements des performances se font toutes les 4 à 5 semaines entre le début de la lactation et le tarissement :

- Les protocoles A et D (30% des données de notre étude) comprend le contrôle de toutes les traites de 24 heures. Les teneurs en matières grasses et protéiques et la numération des cellules sont estimées à l'aide d'autant de prélèvements qu'il y a de traites contrôlées : 1, 2, ou 3 traites. Dans le cas général, l'opérateur de collecte reconstitue un échantillon représentatif du lait produit sur 24 heures (protocole A), mais il est également possible de soumettre à l'analyse un échantillon représentatif de chaque traite (protocole D).
- Le protocole AT (42% des données de notre étude) implique le contrôle (enregistrement du poids de lait et échantillonnage) d'une seule des 2 traites quotidiennes, alternativement celle du soir et celle du matin.
- Le protocole AZ (28% des données de notre étude) implique un enregistrement du poids de lait à partir des 2 traites quotidiennes et un enregistrement des taux et numérations cellulaires à partir d'un échantillon fait sur une seule des 2 traites, alternativement celle du soir et celle du matin.

- Le protocole CZ représente 6% de la base de données brutes implique que ce soit l'éleveur qui face lui-même la pesée. Ce type de protocole n'a pas été pris en compte dans nos études pour ne pas prendre le risque d'avoir un effet opérateur trop important et différents des autres protocoles.

La Figure 7 présente la répartition (en pourcentage) des 4 types de contrôles (A, AT, AZ et D) en 2016.



**Figure 7 : Proportion représentée par les différents types de protocoles de contrôle laitier en 2016 pris en compte dans cette thèse**

Lors de ces contrôles, la date du contrôle est enregistrée, et les quantités de lait, les taux (TB et TP) et les CCS sont mesurés. En moyenne, les chèvres sont contrôlées toutes les 4 à 5 semaines, soit en moyenne 7 fois au cours d'une lactation. Le Tableau 3 présente le nombre de contrôles élémentaires et le nombre de lactation enregistrés entre 2000 et 2018 en races Alpine et Saanen. On observe environ 7 fois plus de contrôles enregistrés que de lactations pour environ 1 million d'individus dans le pedigree pour les deux races.

**Tableau 3 : Nombre d'animaux dans le pedigree, nombre de contrôles élémentaires (entre 7 et 270 jours de lactation), et nombre de lactations, pour les rangs de lactation inférieur à 4, entre 2000 et 2018**

	Saanen		Alpine	
<b>Pedigree</b>	818 702		1 021 665	
	<b>Nombre contrôles</b>	<b>Nombre lactations</b>	<b>Nombre contrôles</b>	<b>Nombre lactations</b>
<b>Total</b>	7 908 192	1 308 307	10 483 205	701 229

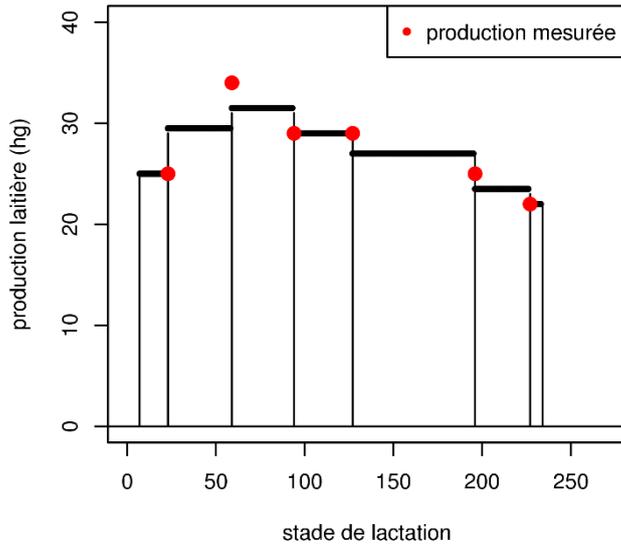
#### **IV. L'évaluation génétique des caractères de production laitière et des cellules somatiques en races Alpine et Saanen**

##### **1. Les phénotypes à la lactation**

Les quantités de lait du jour de contrôle sont mesurées en ferme puis interpolées afin d'obtenir les quantités à la lactation ( $Q_{lact}$ ) selon la formule :

$$Q_{lact} = I_0 \times Q_1 + \sum_{i=1}^{n-1} \left( I_i \times \frac{Q_i + Q_{i+1}}{2} \right) + I_n \times Q_n$$

où  $Q_i$  est la quantité mesurée lors du  $i^{\text{ème}}$  contrôle laitier,  $I_i$  est l'intervalle entre le contrôle laitier  $i$  et  $i+1$ ,  $I_0$  est l'intervalle entre le vêlage et le 1<sup>er</sup> contrôle et  $I_n$  l'intervalle entre le dernier contrôle et le tarissement. Une illustration est donnée en Figure 8. Les taux butyreux et protéique mesurés lors des contrôles sont transformés en quantités de matières journalières, puis en appliquant la formule de Fleischmann, les quantités de matières sont calculées à la lactation. Les taux à la lactation sont calculés à partir des quantités de lait et de matières ainsi calculées à la lactation.



**Figure 8 : Illustration de la méthode de Fleischmann sur une lactation (point rouge : mesure de la production réelle, barre noire : production moyenne des 2 contrôles encadrant la période)**

Les lactations de durée inférieure à 250 jours sont soit extrapolées pour les lactations en cours (Bouloc, 1991), soit corrigées pour la durée pour les lactations terminées. Les lactations qui dépassent 250 jours sont tronquées à 250 jours.

A partir de la production partielle et du dernier contrôle, la quantité de lait à la lactation extrapolée ( $Q_{\text{lact ext}}$ ) se calcule selon la formule :

$$Q_{\text{lact ext}} = \text{production partielle} + X * \text{production dernier contrôle}$$

X dépend de la variable (lait, MG, MP), de la durée de lactation en cours, et du numéro de lactation, il se calcule ainsi :  $X = Y_0 + Y_1 * \text{durée} + Y_2 * \text{durée}^2 + Y_3 * \text{durée}^3$  avec durée = durée de la lactation en cours.

Par exemple, pour la variable lait et, pour les primipares, X se calcule comme suit :  $X = 198.5 - 0.9475 * \text{durée} + 0.00243 * \text{durée}^2 - 0.0000074 * \text{durée}^3$

Pour les comptages de cellules somatiques, le calcul intègre tous les comptages cellulaires (CCS = comptages de cellules somatiques) entre 5 et 250 jours de lactation pour les lactations des rangs 1 à 3. La performance par lactation est basée sur une transformation des CCS en un score de cellules somatiques (SCS) pour avoir une distribution normale des performances :

$$SCS = \log_2 \left( \frac{CCS}{100000} \right) + 3$$

Une correction additive du SCS pour le stade et le rang de lactation est réalisée pour prendre en compte l'utilisation des lactations anciennes avec quelques comptages et des lactations en cours.

L'ajustement est estimé à partir des lactations saines (SCS moyen < 4). La moyenne pondérée des SCS

$$\text{SCSL} = \frac{\sum \frac{R}{\sigma} [\text{SCS corrigé}]}{\sum \frac{R}{\sigma}}$$

où R est la corrélation relative entre une mesure SCS corrigée et la moyenne des autres SCS corrigés (une corrélation de 1 a été attribuée au stade 90-120 jours), et  $\sigma$  écart-type du SCS corrigé fonction du stade et de la parité. La pondération est fonction de la race, du stade de lactation et de la parité pour  $\sigma$ .

Les SCSL sont transformés pour avoir une variabilité similaire entre rangs de lactation (multiplication de SCSL par 1.15 en L1 ; 1.08 en L2 et 1 en L3). Chaque lactation a un poids qui est fonction du nombre de contrôles avec CCS connu. Un minimum de 4 CCS est nécessaire pour obtenir un poids de 1 (Clément et al., 2015b).

## **2. Le modèle d'évaluation génomique à la lactation mis en œuvre pour les caractères de production laitière et les scores de cellules somatiques en race Alpine et Saanen depuis 2018**

Les 5 caractères laitiers (lait, TB, TP, MP et MG) sont aujourd'hui indexés séparément par un ssGBLUP modèle animal unicaractère mais les 2 races (Alpine et Saanen) sont indexées ensemble. Pour les cellules, les races Alpine et Saanen sont indexées séparément selon un modèle SSGBLUP. Toutes les parités sont traitées ensemble, il n'y a qu'un seul index génétique pour toutes les lactations, et qu'un seul effet d'environnement permanent. Le modèle utilisé est un modèle linéaire mixte c'est-à-dire qu'il comporte à la fois des effets fixes et des effets aléatoires.

Sous forme matricielle, le modèle ssGBLUP à la lactation s'écrit :  $y = Xb + Z_1a + Z_2p + e$

Avec  $y$  est le vecteur des phénotypes à la lactation

$b$  est le vecteur des effets fixes,

$a$  est le vecteur aléatoire des valeurs génomiques additives de dimension égale au nombre animaux dans le pedigree

$p$  est le vecteur aléatoire des effets de l'environnement permanent. L'effet d'environnement permanent permet de tenir compte de l'environnement commun (hors génétique) entre les différentes parités d'un animal.

$e$  est le vecteur aléatoire des résidus, distribués suivant une loi normale de moyenne nulle, et de variance hétérogène.

$X$  et  $Z_1$  et  $Z_2$  sont les matrices d'incidence des effets fixes, des effets aléatoires génétiques et des effets aléatoires de l'environnement permanent respectivement.

Les effets aléatoires du modèle suivent des lois normales telles que :  $E(a) = 0$  ;  $\text{Var}(a) = H \sigma_a^2$  où  $H$  est la matrice de parenté génomique et  $\sigma_a^2$  la variance génétique

$E(p) = 0$  ;  $\text{Var}(p) = I \sigma_p^2$  où  $I$  est la matrice identité et  $\sigma_p^2$  la variance de l'environnement permanent

$E(e) = 0$  ;  $\text{Var}(e) = R\sigma_e^2$  où  $R$  est la matrice de variances-covariances résiduelle, et  $\sigma_e^2$  la variance résiduelle. En plus des effets fixes qui corrigent pour la moyenne, une correction est appliquée dans la résiduelle pour les différences de variabilité associées au troupeau, à la campagne et au rang de lactation de la performance en effet fixe et pour l'élevage, campagne de mise bas, rang de lactation (1,2,3 et plus) en effet aléatoire.

Afin d'homogénéiser la variance pour tous les rangs de lactation, les premières lactations ont un poids de 1, les 2ièmes un poids de 0.9, et les 3ièmes et plus un poids de 0.85, pour prendre en compte la différence de caractère entre une première lactation et les lactations de rang supérieur. En effet, les chèvres en première lactation produisent moins qu'en deuxième lactation, elles-mêmes produisant moins qu'en troisième lactation et plus.

Les équations du modèle mixte s'écrivent :

$$\begin{pmatrix} X'X & X'Z_1 & X'Z_2 \\ Z_1'X & Z_1'Z_1 + \alpha H^{-1} & Z_1'Z_2 \\ Z_2'X & Z_2'Z_1 & Z_2'Z_2 + \gamma I \end{pmatrix} \begin{pmatrix} \hat{b} \\ \hat{a} \\ \hat{p} \end{pmatrix} = \begin{pmatrix} X'y \\ Z_1'y \\ Z_2'y \end{pmatrix}$$

Où  $\alpha = \frac{\sigma_e^2}{\sigma_a^2} = \frac{1-r}{h^2}$ ,  $\gamma = \frac{\sigma_e^2}{\sigma_p^2} = \frac{1-r}{r-h^2}$  et  $r = \frac{\sigma_a^2 + \sigma_p^2}{\sigma_y^2}$  avec  $r$  la répétabilité du caractère,  $h^2$  l'héritabilité du caractère,  $H^{-1} = A^{-1} + \begin{pmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{pmatrix}$  où  $A^{-1}$  est l'inverse de la matrice de parenté,  $A_{22}^{-1}$  est l'inverse de la matrice de parenté pour le bloc des animaux génotypés et  $G^{-1}$  est l'inverse de la matrice de parenté génomique

$G = 0.95 * \frac{ZZ'}{\sum_{i=1}^m p_i * (1 - p_i)} + 0.05 * A_{22}$ , avec  $Z = M - P$ , avec  $m$  est le nombre total de marqueurs SNP,  $p_i$  est la fréquence de l'allèle  $i$ ,  $M$  est la matrice centrée des génotypes,  $P$  est une matrice où chaque colonne contient :  $2 * p_i (1 - p_i)$ . La matrice  $G$  mesure la ressemblance entre les individus d'un point de vue génotypique, elle estime la proportion moyenne d'allèles partagés par deux individus pondérée par leurs fréquences alléliques. Les individus génotypés ont des relations de parentés calculées à partir des données génomiques ce qui permet de tenir compte de l'aléa de méiose. En effet, en l'absence de données génomiques, deux pleins frères ont en espérance une parenté de 0.5.

Les différentes variances ( $\sigma_a^2$ ,  $\sigma_p^2$ ,  $\sigma_e^2$ ) du modèle sont estimées à l'aide de la méthode AI-REML (Average Information Restricted Maximum Likelihood). Dans cette thèse, nous les avons réestimés à l'aide du logiciel Wombat (Meyer, 2007) qui est basé sur la méthode Expectation Maximization (EM) couplée à la méthode AIREML. Une fois les paramètres génétiques calculés, les effets génétiques et d'environnement permanent sont estimés en résolvant les équations du modèle mixte à l'aide du logiciel blupiod2 (Misztal et al., 2002).

### **a. Description des effets fixes du modèle d'évaluation génomique**

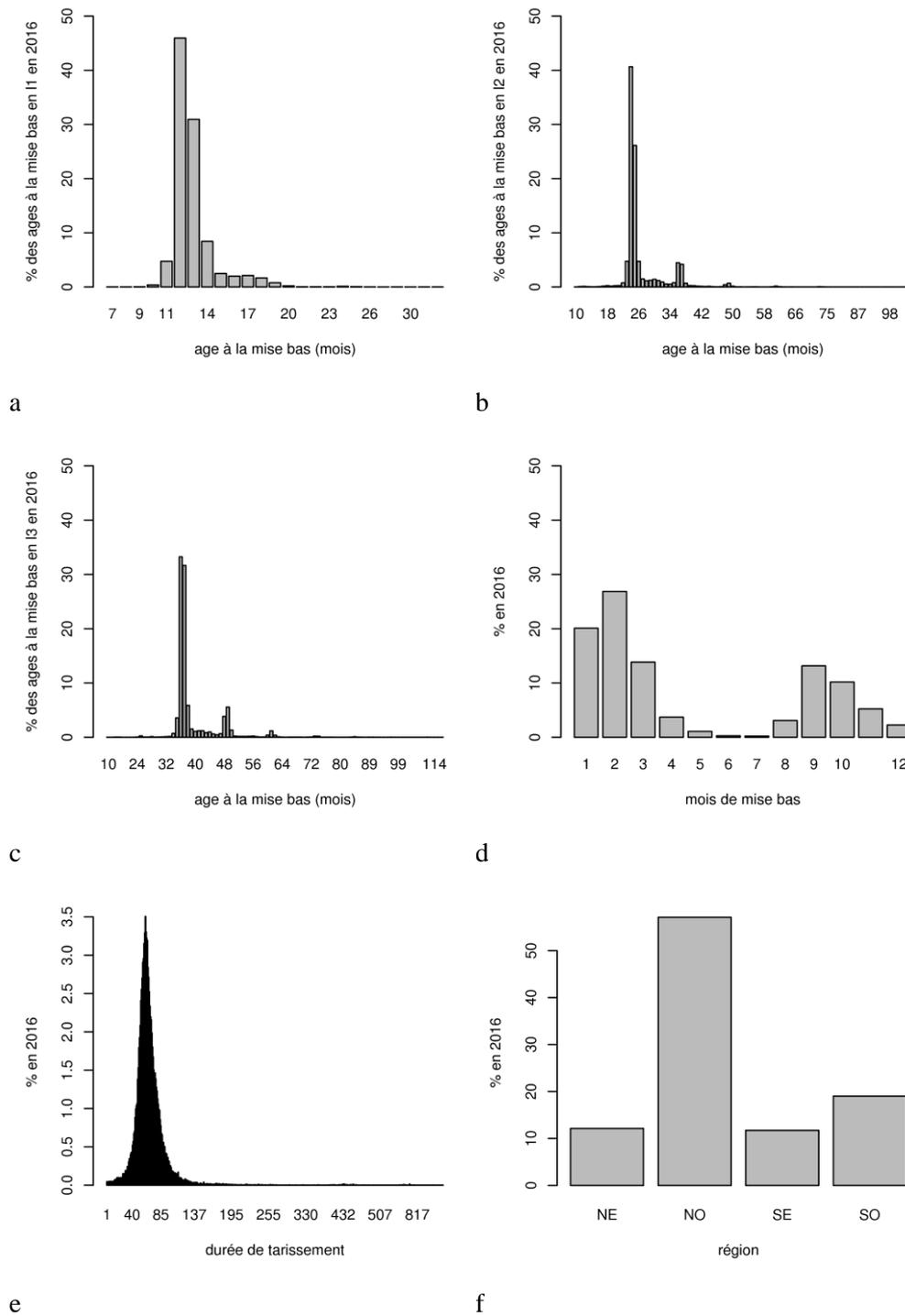
Les effets de milieu inclus dans le modèle d'évaluation génomique sont:

- Le troupeau, défini en fonction de la campagne de mise bas et du rang de lactation (1, 2, 3 et plus);
- Le mois de mise bas, défini en fonction de la région, la campagne de mise bas et le rang de lactation (1, 2 et plus);
- L'âge à la mise bas, défini en fonction de la région, la campagne de mise bas et le rang de lactation (1,2,3,4,5,6+);
- La durée de tarissement précédent, défini en fonction de la région, le rang de lactation (1,2,3+) et la campagne de mise bas.

La Figure 9 présente la répartition, pour les deux races confondues, des données de lactation en fonction de l'âge à la mise bas pour les chèvres en première (a), deuxième (b) et troisième lactation (c), du mois de mise bas (d), de la durée de tarissement (e), et de la région (f). Les chèvres mettent bas en majorité à 12 mois en première lactation, puis à 24 mois en deuxième lactation et à 36 mois en troisième lactation. Elles mettent bas majoritairement en janvier-février, sont taries environ 300 jours après leur mise bas durant une durée de 60 jours. La grande majorité des données du contrôle laitier provient du quart NO (55%). La répartition des 2 races est homogène en France.

Ces effets ont été regroupés en différentes classes :

- Âge à la mise bas (28 classes): pour les première lactation (en mois) : 0-9, 10, 11, 12, 13, 14, 15-18, 19-30, autre ; pour les deuxième lactations : 16-21, 22-23, 24-25, 26-27, 28-29, 30-33, 34-56, autres ; pour les troisième lactations : 22-31, 32-34, 35-38, 39-42, 43-54, 55-77, autres ; une classe unique pour les quatrième lactations, une classe unique pour les cinquième lactations et une dernière classe unique pour les sixième lactations et plus.
- Mois de mise bas (12 classes) : en première lactation : Septembre – Décembre, Janvier, Février, Mars, Avril, Mai-Aout, en deuxième lactation et plus : Septembre–Novembre, Décembre, Janvier, Février, Mars-Avril, Mai-Aout.
- Durée de tarissement (en jours) (9 classes): en première lactation (1 classe unique); en deuxième lactation : <30, entre 30 et 60, entre 60 et 365, autres ; en troisième lactation et plus : <30, entre 30 et 60, entre 60 et 365, autres



**Figure 9: Répartition de l'âge à la mise bas pour les chèvres en première (a), deuxième (b) et troisième lactation (c), du mois de mise bas (d), de la durée de tarissement (e), et de la région (f) en 2016, Alpine et Saanen confondues.**

### **b. Les paramètres génétiques**

L'héritabilité et la répétabilité ont été estimées respectivement à 30 % et 50 % pour les quantités de matières et le lait, et à 50 % et 70% pour les taux (Clément et al., 2015b) pour les deux races.

Pour les SCS, l'héritabilité du caractère estimée est de 19% en race Alpine et 21% en race Saanen, la répétabilité de 25% en race Alpine et 29% en race Saanen.

### **c. Limites du modèle d'évaluation génétique mis en œuvre aujourd'hui**

Comme nous l'avons déjà vu, le modèle utilisé pour l'évaluation génétique à la lactation a permis d'augmenter la quantité de lait totale à la lactation sans s'intéresser à la forme de la courbe de lactation. Le modèle d'évaluation actuel comporte ainsi plusieurs limites pour une analyse fine de la production laitière et de la persistance des chèvres. Il ne prend pas en compte les corrélations génétiques entre stade de lactation. En effet, le modèle à la lactation considère des corrélations de 1 entre tous les stades de lactation. Le modèle, basé sur la production à la lactation, ne permet donc pas d'évaluer génétiquement la persistance laitière des animaux.

Le phénotype extrapolé lorsque la lactation n'est pas terminée, n'est pas optimal (Bouloc, 1991). En effet, la formule d'extrapolation est générique pour tous les animaux, il existe une seule différenciation pour les primipares et les multipares. Aucun effet de milieu n'est considéré dans la formule d'extrapolation.

Une autre limite du modèle utilisé est la prise en compte des rangs de lactation. En effet, les lactations sont considérées ensemble quel que soit le rang de lactation supposant une corrélation génétique et de l'environnement permanent de 1 entre les rangs de lactation successifs. Ce résultat est probablement surestimé au vu des études caprines allemande (Zumbach et al., 2008) ou anglaise (Mucha et al., 2014).

L'effet troupeau est un effet défini en fonction de la campagne. Un animal appartient à une classe d'effet troupeau-campagne-numéro de lactation selon la date de sa mise bas. Deux animaux, l'un mettant bas en début de campagne et l'autre mettant bas en fin de campagne, peuvent avoir le même effet troupeau-campagne sans avoir produit réellement en même temps. Les autres effets fixes comme le mois de mise bas, l'âge à la mise bas, la durée de tarissement sont des effets qui ne prennent pas en compte le stade de lactation. Ces effets sont constants tout au long de la lactation. Enfin l'effet de la gestation n'est pas pris en compte dans le modèle. Ces résultats peuvent induire des biais dans le calcul d'extrapolation de la production à la lactation.

## **V. Modèles de contrôles élémentaires (Test day models)**

Contrairement au modèle utilisé actuellement en France pour les caprins, les modèles, appelés modèles de contrôles élémentaires ou Test day models, utilisent directement les données issues des mesures en ferme (les contrôles élémentaires). La quantité de données à traiter est alors plus importante (environ 7 fois plus de données à analyser).

Plusieurs modèles permettent de traiter de telles données : les modèles à répétabilité (Ptak and Schaeffer, 1993), les modèles multi-caractères par classes de stades de lactation (Misztal et al., 2000), les « character process models » (Pletcher and Geyer, 1999), les « structured antedependance models »

(Jaffrézic et al., 2004), les « dynamic linear models » (Forni et al., 2009) ou encore les modèles de régression aléatoire (Jamrozik et al., 1997). Les modèles à répétabilité considèrent une corrélation de 1 entre tous les stades de lactation et ne permettent donc pas une analyse et une sélection de la persistance. Les modèles multi-caractères par classes de stades de lactation ont un nombre de paramètres importants si nous voulons bien décrire la variation des corrélations au cours de la lactation, ce qui rend leur utilisation compliquée. De plus, lorsqu'un grand nombre de classes est utilisé pour ce type de modèle, peu de répétition par périodes sont disponibles ce qui diminue la précision des estimations. Il existe également des modèles avec des fonctions de covariances mais ils sont équivalents aux modèles de régression aléatoire (Meyer and Hill, 1997).

Dans la suite du document, nous ne détaillerons que les modèles de régression aléatoire, qui ont fait leur preuve dans les travaux développés en bovins laitiers (Druet et al., 2003, 2005; Leclerc, 2008).

En génétique quantitative, les modèles de régression aléatoire sont les plus utilisés pour traiter les données de production laitière (Jamrozik et al., 1997; Druet et al., 2005; Lidauer et al., 2015), ils sont implémentés dans les logiciels d'estimation de paramètres génétiques comme Wombat (Meyer, 2007) ou AIREMLF90 (Misztal et al., 2002).

Ils permettent de modéliser les effets aléatoires génétiques et d'environnement permanent au cours du temps sans restriction sur le nombre minimal de données par individu. Ils permettent d'estimer les valeurs génétiques des animaux pour des phénomènes biologiques qui se déroulent au cours du temps, comme la production laitière, la croissance d'animaux d'élevage (Meyer, 2005), la courbe de ponte d'œufs de poules (Anang et al., 2002). Ils produisent une valeur génétique par unité de temps (jour par exemple). Ces valeurs génétiques peuvent ensuite être exploitées finement dans un schéma de sélection. La forme de la courbe de lactation peut ainsi être exploitée à l'aide des modèles de régression aléatoire. Les modèles de régression aléatoire peuvent également être utiles pour étudier les interactions entre la performance et la qualité du milieu avec des modèles appelés « norme de réaction » (Huquet et al., 2012; Mota et al., 2018).

Les modèles de régression aléatoire sont actuellement utilisés en routine dans les évaluations génétiques des vaches laitières de nombreux pays : Allemagne-Autriche-Luxembourg, Australie, Belgique, Canada, Croatie, Danemark-Finlande-Suède, Estonie, Irlande, Italie, Japon, Lettonie, Lituanie, Nouvelle-Zélande, Pays-Bas, Pologne, République-Tchèque, Royaume-Uni, Slovaquie et Suisse (Oliveira et al., 2019a). Plusieurs études ont été menées en caprins en Allemagne (Zumbach et al., 2008), au Royaume-Uni (Mucha et al., 2014), en Norvège (Andonov et al., 2013), en Espagne (Menéndez-Buxadera et al., 2010), et au Brésil (Brito et al., 2017). La plupart de ces pays utilisent ces modèles afin d'obtenir des valeurs génétiques plus précises pour la quantité à la lactation car les effets environnementaux sont mieux modélisés et pris en compte (Schaeffer and Jamrozik, 2008). D'autres sont

intéressés par ces modèles pour mieux comprendre la relation entre les caractères au cours de la lactation (Buttchereit et al., 2011; Jamrozik and Schaeffer, 2012; Mucha et al., 2014).

Plusieurs pays diffusent un index persistance issu des modèles de régression aléatoire aux éleveurs et ces index font partie de l'index de synthèse de sélection global notamment pour les éleveurs en agriculture biologique car la persistance est associée à un caractère de santé et de robustesse en Suisse (Bapst, 2001), en Autriche (Baumung et al., 2001), en Allemagne (Krogmeier, 2003) et au Canada (Rozzi et al., 2007). Contrairement à la valeur génétique à la lactation qui est transmise à Interbull (organisme maintenant un système de calcul de routine des évaluations internationales génétiques et génomiques des principaux groupes de races et de caractères laitiers en bovin), la valeur de la persistance laitière ne l'est pas actuellement mais devrait l'être prochainement (Oliveira et al., 2019a). En caprins laitiers, il est plus difficile de savoir quels sont les modèles utilisés en routine car l'équivalent d'Interbull en caprin n'existe pas aujourd'hui.

De manière générale un modèle de régression aléatoire, pour modéliser la production laitière d'une chèvre au cours de la lactation, s'écrit sous la forme suivante:

$$y_{ikt n} = HTD_i + A_k + \sum_{o=1}^q \theta_{ko} N_{(o,t)} + \sum_{o=1}^q a_{no} \varphi_{(o,t)} + \sum_{o=1}^q p_{no} \varphi_{(o,t)} + e_{ikt n}$$

où  $y_{ikt n}$  est la production mesurée au stade de lactation  $t$  de la chèvre  $n$  appartenant à la classe d'âge à la mise bas-mois de mise bas  $k$ , au troupeau-jour de contrôle  $i$ .  $HTD_i$  est l'effet fixe troupeau-jour de contrôle;  $A_k$  est l'effet fixe âge à la mise bas-mois de mise bas;  $\theta_{ko}$  sont les coefficients de régression pour l'âge à la mise bas-mois de mise bas;  $N_{(o,t)}$  est la  $o^{i\text{ème}}$  covariable au stade  $t$ ;  $a_{no}$  et  $p_{no}$  sont respectivement les coefficients de régression aléatoires génétiques et de l'environnement permanent pour la  $o^{i\text{ème}}$  covariable,  $\varphi_{(o,t)}$  est la valeur de la  $o^{i\text{ème}}$  covariable au stade  $d$  et  $e_{ikt n}$  est le résidu. Dans cet exemple, la fonction utilisée est la même pour l'effet fixe et les deux effets aléatoires, elle est composée de  $q$  covariables.

Sous forme matricielle, le modèle ci-dessus s'écrit comme suit :

$$y = \mathbf{Xb} + \mathbf{Z}_1 \mathbf{a} + \mathbf{Z}_2 \mathbf{p} + \mathbf{e}$$

Où  $y$  est le vecteur des contrôles élémentaires,

$\mathbf{b}$  est le vecteur des effets fixes

$\mathbf{a}$  est le vecteur des coefficients des régressions aléatoires des effets génétiques additives (taille :  $n \times$  nombre animaux du pedigree, avec  $n$  le nombre de covariables)

$\mathbf{p}$  est le vecteur des coefficients des régressions aléatoires des effets d'environnement permanent (taille :  $n \times$  nombre animaux avec performance)

$\mathbf{e}$  est le vecteur des résidus

$\mathbf{X}$ ,  $\mathbf{Z}_1$  et  $\mathbf{Z}_2$  sont les matrices d'incidence qui relient les observations aux effets fixes et aléatoires

La matrice de variances-covariances s'écrit sous la forme suivante :

$$\text{var} \begin{pmatrix} a \\ p \\ e \end{pmatrix} = \begin{pmatrix} \mathbf{H} \otimes \mathbf{C} & 0 & 0 \\ 0 & \mathbf{I} \otimes \mathbf{P} & 0 \\ 0 & 0 & \mathbf{R} \end{pmatrix} \quad \text{avec} \quad \mathbf{C} = \begin{bmatrix} \sigma_{a_i}^2 & \cdot & \cdot & \cdot & \sigma_{a_i a_q}^2 \\ \vdots & \cdot & \cdot & \cdot & \vdots \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma_{a_q a_i}^2 & \cdot & \cdot & \cdot & \sigma_{a_q}^2 \end{bmatrix} \quad \text{et} \quad \text{avec}$$

$$\mathbf{P} = \begin{bmatrix} \sigma_{p_i}^2 & \cdot & \cdot & \cdot & \sigma_{p_i p_q}^2 \\ \vdots & \cdot & \cdot & \cdot & \vdots \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma_{p_q p_i}^2 & \cdot & \cdot & \cdot & \sigma_{p_q}^2 \end{bmatrix}$$

où  $\mathbf{C}$  et  $\mathbf{P}$  sont des matrices de dimension égale au nombre de covariables  $q$  pour les effets génétiques et d'environnement permanent respectivement,  $\mathbf{H}$  est la matrice de parenté génomique issue des informations du pedigree et génomiques,  $\mathbf{R}$  est la matrice diagonale de la variance résiduelle qui dépend du stade de lactation, et  $\otimes$  est le produit de Kronecker.

Les équations du modèle mixte s'écrivent alors :

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}_1 & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}_2 \\ \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{Z}_1 + \mathbf{H}^{-1} \otimes \mathbf{C}^{-1} & \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{Z}_2 \\ \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{Z}_1 & \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{Z}_2 + \mathbf{I} \otimes \mathbf{P}^{-1} \end{pmatrix} \begin{pmatrix} \hat{b} \\ \hat{a} \\ \hat{p} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix}$$

$$\text{avec } \mathbf{H} \otimes \mathbf{C} = \text{Var} \begin{bmatrix} a_i \\ \vdots \\ a_q \end{bmatrix} = \begin{bmatrix} \mathbf{H} \sigma_{a_i}^2 & \cdot & \cdot & \cdot & \mathbf{H} \sigma_{a_i a_q}^2 \\ \vdots & \cdot & \cdot & \cdot & \vdots \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \mathbf{H} \sigma_{a_q a_i}^2 & \cdot & \cdot & \cdot & \mathbf{H} \sigma_{a_q}^2 \end{bmatrix}$$

Les matrices de variances-covariances génétiques ( $\Sigma$ ) et de l'environnement permanent ( $\Phi$ ) s'obtiennent ainsi :

$$\Sigma = \mathbf{Q}\mathbf{C}\mathbf{Q}' \quad \text{et} \quad \Phi = \mathbf{Q}\mathbf{P}\mathbf{Q}'$$

avec  $\mathbf{Q}$  la matrice des covariables utilisées.

L'évolution des valeurs génétiques d'un individu  $n$  ( $\text{EBV}_n$ ) au cours de la lactation s'obtient comme suit :

$$\text{EBV}_n = \mathbf{Q}\hat{a}_n \quad \text{où } \hat{a}_n \text{ est le vecteur des coefficients de régression de l'animal } n.$$

Les paramètres génétiques, ici les matrices  $\mathbf{C}$ ,  $\mathbf{P}$  et  $\mathbf{R}$  sont estimées à l'aide de la méthode AI-REML (Average Information Restricted Maximum Likelihood) et la méthode EM à l'aide du logiciel Wombat (Meyer, 2007). Les valeurs génétiques et les effets d'environnement permanent sont estimés ensuite en résolvant les équations du modèle mixte à l'aide du logiciel blupiod2 (Misztal et al., 2002).

Dans le modèle de régression aléatoire, les covariables  $N_{(o,t)}$  (ou  $\varphi_{(o,t)}$  ( $o^{\text{ième}}$  covariable au stade  $t$ ) sont issues des modélisations retenues pour modéliser l'évolution des différents effets fixes du modèle. Ils peuvent être classés en trois catégories : les modèles paramétriques, les modèles semi-paramétriques et les modèles non paramétriques. On note  $y_t$  la production laitière moyenne du jour au temps  $t$ .

Les modèles paramétriques permettent de résumer les caractéristiques de l'évolution de la production laitière au cours du temps par des fonctions linéaires ou non linéaires à travers un faible nombre de paramètres. A partir de ces fonctions, la date du pic de production peut être déduite facilement : il s'agit de la valeur de  $t$  pour laquelle la dérivée première de la fonction s'annule. Le niveau de production au pic correspondra à la valeur de la fonction pour  $t = t_{\text{pic}}$ . Enfin la production entre 2 dates sera obtenue en calculant l'intégrale de la fonction sur cet intervalle de temps. Les modèles paramétriques classiquement utilisés pour modéliser une courbe de production laitière dans la littérature sont les suivants:

- Le modèle de Wilmink (1987) s'exprime comme  $E(y_t) = a + bt - de^{-ct}$ . Ce modèle, composé d'une fonction exponentielle, est représenté par une courbe non linéaire et comporte quatre paramètres à estimer ( $a$ ,  $b$ ,  $c$  et  $d$ ). Le paramètre  $a$  est associé au niveau de production,  $b$  à la phase décroissante de la production,  $c$  à la période du pic de lactation et  $d$  à l'augmentation de la production au cours de la phase ascendante.
- Le modèle d'Ali et Schaeffer (1987) se met sous la forme  $E(y_t) = a + bu + cu^2 + dv + ev^2$  où  $u = \frac{t}{270}$  et  $v = \ln\left(\frac{270}{t}\right)$ . Il a été largement utilisé dans de nombreuses études décrivant l'évolution d'une production ou d'une croissance. Il est composé de 5 paramètres ( $a$ ,  $b$ ,  $c$ ,  $d$  et  $e$ ) :  $a$  est associé au pic de production,  $b$  et  $c$  à la phase descendante de la courbe de production,  $d$  et  $e$  à la phase ascendante. L'inconvénient de ce modèle est l'existence, par construction, de fortes corrélations entre ses paramètres, ce qui limite fortement la convergence des algorithmes d'estimation des paramètres.
- Les polynômes de Legendre (Kirkpatrick et al., 1990), se définissent comme suit :

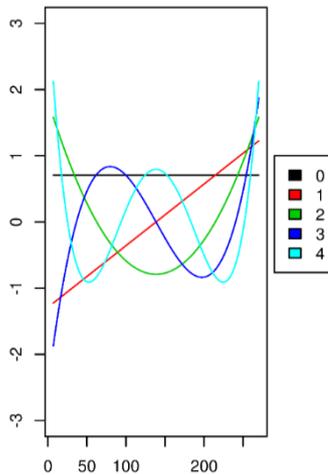
$$E(y_t) = \sum_{i=0}^q \alpha_i \sqrt{\frac{2q+1}{2}} \times P_i(\omega)$$

où  $\omega = 2 \times \left( \frac{t - t_{\text{min}}}{t_{\text{max}} - t_{\text{min}}} \right) - 1$ ,  $t_{\text{min}} = 7$  (premier jour de contrôle),  $t_{\text{max}} = 270$  (stade maximum de lactation),  $P_{q+1}(\omega) = \frac{1}{q+1} ((2q+1)\omega P_q(\omega) - n P_{q-1}(\omega))$ , avec  $P_0(\omega) = 1$  et  $P_1(\omega) = \omega$ , et  $\alpha_i$  coefficient de régression. Une des propriétés intéressantes des polynômes de Legendre est l'orthogonalité des covariables, elle permet de minimiser les corrélations entre les coefficients de régression et donc de faciliter la convergence de l'algorithme d'estimation des paramètres.

Pour  $q=4$ , le modèle des polynômes de Legendre s'écrit sous la forme suivante :

$$E(y_t) = \sqrt{0.5} \times a + \sqrt{\frac{3}{2}} \times b \times \omega + \sqrt{\frac{5}{2}} \times \frac{1}{2} \times c \times (3\omega^2 - 1) + \sqrt{\frac{7}{2}} \times \frac{1}{2} \times d \times (5\omega^3 - 3\omega) + \sqrt{\frac{9}{2}} \times \frac{1}{8} \times e \times (35\omega^4 - 30\omega^2 + 3).$$

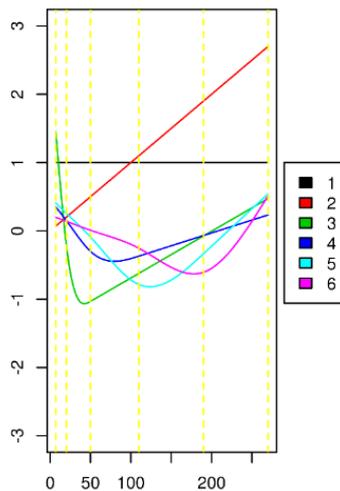
La Figure 10 représente les 5 premières covariables de Legendre. La première covariable ( $\varphi_{(0,t)}$ ) correspond à une droite constante au cours de la lactation, la deuxième covariable ( $\varphi_{(1,t)}$ ) correspond à une fonction linéaire (équation d'une droite), la troisième covariable ( $\varphi_{(2,t)}$ ) à une parabole, etc.



**Figure 10 : Représentation des 5 premières covariables de Legendre**

Les modèles semi-paramétriques que l'on retrouve le plus souvent dans la littérature pour modéliser ce type de données sont les splines. Il s'agit de fonctions polynomiales par morceaux dont la jonction se fait au niveau de nœuds. L'ajustement dépend du nombre et de la position des nœuds utilisés, que l'on place aux endroits où il y a le plus de variation (en début de lactation dans notre cas) ainsi que du degré du polynôme (degré 1 : spline linéaire, degré 2 : spline quadratique, degré 3 : spline cubique). Les splines cubiques, par exemple, sont définies par l'équation suivante :

$y_t = a_i + b_i(t - t_i) + c_i(t - t_i)^2 + d_i(t - t_i)^3$  pour  $t_i < t < t_{i+1}$ ,  $t_i$  étant le stade de lactation au  $i^{\text{ème}}$  nœud. Il existe une grande diversité de types de splines comme les splines naturelles qui ont leur dérivée seconde à chaque nœud extrême nulle, il existe aussi les  $\beta$  splines utilisées en génétique quantitative (Meyer, 2005). Les 6 covariables de la spline utilisée par la suite, qui est une spline naturelle cubique, sont présentées sur la Figure 11.



**Figure 11: Illustration des six covariables d'une spline naturelle cubique dont les nœuds (représentés par les traits pointillés jaunes) ont été positionnés aux stades de lactation suivants ( $t = 7, 20, 50, 110, 190, 270$ ) pour modéliser la forme de la courbe de lactation**

La modélisation de la production laitière à l'aide de modèles paramétriques ou semi-paramétriques permet généralement de lisser les variations aléatoires observées et de résumer la courbe de production laitière à l'aide de quelques paramètres. En revanche, des biais peuvent apparaître si l'ajustement aux données est mauvais. C'est pourquoi d'autres modèles ont été développés pour remédier à ce défaut.

Les modèles non paramétriques permettent de modéliser l'évolution de la production laitière au cours de la lactation sans a priori sur la forme de la courbe. Une façon de le faire simplement est de définir des classes de stade de lactations et d'estimer un effet par classe. L'inconvénient d'une telle fonction constante par palier est la définition d'un nombre important de paramètres à estimer. La formation des classes et leur nombre est l'enjeu de ce type de modélisation. Ces courbes sont donc relativement flexibles mais très coûteuses en nombre de paramètres à estimer.

Une étude (réalisée dans le cadre d'un stage de fin d'étude par Arnal (2016)) a permis de comparer différents modèles pour modéliser l'évolution phénotypique de la production laitière au cours de la lactation. Le modèle de Wilmink, des polynômes de Legendre de degré 5, un modèle incluant une variable dont les niveaux ont été définis comme des classes de 5 jours de lactation, le modèle d'Ali et Schaeffer et des splines ont été mis en œuvre et comparés en terme de qualité d'ajustement. Il a été montré que le modèle qui modélisait la courbe par classes de 5 jours, ajustait le mieux la courbe, devant les splines, le modèle d'Ali et Schaeffer et les polynômes de Legendre. Les différences étaient toutefois infimes. Cette étude a permis de valider sur les données caprines l'étude de Druet et al. (2003), réalisée sur des données bovines laitières françaises et qui avait choisi les splines pour modéliser la partie fixe de la courbe. La courbe par classes de 5 jours n'a pas été retenue en raison de son nombre important de paramètres.

Les polynômes de Legendre ou les splines sont des fonctions non spécifiques à l'étude de la production laitière. Différentes études ont été menées comparant les polynômes de Legendre aux splines (Bohmanova et al., 2008; Bignardi et al., 2011; Bouallegue et al., 2015). Les polynômes de Legendre modélisent mal, en général, les extrémités générant des effets de bord (Misztal, 2006). Les splines peuvent palier à ce problème en densifiant le nombre de nœuds aux extrémités (Druet et al., 2003). Les résultats sont dépendants des caractères étudiés mais en général, bien que l'ajustement aux données est légèrement différent, le classement des animaux sur leur valeur génétique est peu modifié et les mêmes animaux sont sélectionnés (Pereira et al., 2013).

## 1. Les modélisations pour la partie fixe

L'effet troupeau-jour de contrôle, aussi appelé Herd Test-day en anglais (HTD) permet de définir un effet commun pour un ensemble d'animaux partageant des conditions environnementales et alimentaires similaires. Cet effet est particulièrement contesté par Schaeffer (2016, 2018) qui considère qu'il ne représente pas suffisamment l'effet des contemporaines. En effet, cet effet peut regrouper des animaux d'un même troupeau à différents stades de lactation. Schaeffer (2016) propose de modéliser l'effet des contemporaines à l'aide d'un effet aléatoire troupeau-année-rang de lactation en fonction du stade de lactation. Lidauer et al. (2015) utilise l'effet HTD mais en effet aléatoire. Au vu de la littérature, nous avons choisi d'étudier dans une partie de la thèse l'influence de la modélisation de l'effet HTD en effet fixe par rapport à une modélisation de l'effet aléatoire troupeau-année-rang de lactation en fonction du stade de lactation.

Les modèles de contrôles élémentaires permettent de modéliser les effets environnementaux sous forme de courbe ( $\sum_{o=1}^q \theta_{ko} N_{(o,t)}$  dans notre modèle). Nous obtenons donc une courbe par modalité du facteur environnemental (ici  $k$ ). Dans la littérature, l'effet de la durée de tarissement, de l'âge à la mise bas, du mois de mise bas sont souvent modélisés en fonction du stade de lactation sous forme de courbe. Ces effets sont souvent combinés à l'effet année ou/et combinés entre eux (Lidauer et al., 2015; Schaeffer, 2016). Il est également possible de ne pas utiliser de courbes, effet ne dépendant pas du stade de lactation ( $A_k$  dans notre modèle) comme c'est le cas avec le modèle à la lactation actuel pour des effets comme la durée de tarissement, l'âge à la mise bas, le mois de mise bas.

## 2. Modélisation pour les effets génétiques et environnements permanents

Les effets modélisés dans la partie aléatoire sont généralement l'effet génétique ( $\sum_{o=1}^q a_{no} \varphi_{(o,t)}$  dans notre modèle) et l'effet de l'environnement permanent ( $\sum_{o=1}^q p_{no} \varphi_{(o,t)}$  dans notre modèle).

Dans la littérature certains auteurs utilisent des modèles de contrôles élémentaires avec une constante pour modéliser la partie génétique et la partie aléatoire, comme par exemple l'utilisation d'un polynôme de Legendre de degré 0 ( $q=1, \sum_{o=1}^1 a_{no} \varphi_{(o,t)} + \sum_{o=1}^1 p_{no} \varphi_{(o,t)}$  dans notre modèle), que l'on appelle modèle à répétabilité. Le principal avantage est la simplicité du modèle. Ce modèle bénéficie

d'une modélisation fine de la partie fixe mais les parties génétique et de l'environnement permanent ne se résument qu'à une seule valeur. Le principal inconvénient de ce type de modèle est qu'il ne valorise pas la possibilité de modéliser la valeur génétique et l'environnement permanent des animaux en fonction du stade de lactation.

Le plus souvent le degré du polynôme utilisé pour la partie génétique et de l'environnement permanent est identique afin d'avoir la même complexité pour les deux parties.

Dans les équations du modèle mixte, le produit de Kronecker entre la matrice de variances-covariances  $\mathbf{C}$  (de taille  $q \times q$ ) et la matrice de parenté  $\mathbf{H}$  (de taille  $n \times n$ ) génère une matrice de très grande dimension si  $q$  est grand, augmentant considérablement les temps de calcul. Afin de limiter la taille de cette matrice, il est possible de réduire la taille de la matrice  $\mathbf{C}$  en utilisant un degré plus faible pour le polynôme de Legendre, mais il est aussi possible de décomposer la matrice de variances-covariances  $\mathbf{C}$  en composantes principales. La décomposition canonique de la matrice  $\mathbf{C}$  s'écrit:

$$\mathbf{C} = \mathbf{U}\mathbf{D}\mathbf{U}' \text{ avec } \mathbf{U}\mathbf{U}' = \mathbf{I}_q \text{ et } \mathbf{U}'\mathbf{U} = \mathbf{I}_q$$

où  $\mathbf{U}$  est la matrice des vecteurs propres,  $\mathbf{D}$  la matrice diagonale des valeurs propres, et  $\mathbf{I}_q$  la matrice identité de dimension  $q \times q$

Si on ignore les  $p$  plus petites valeurs propres, on peut opérer à une réduction de la matrice  $\mathbf{C}$  en ne conservant que les  $m$  premières valeurs propres ( $m < q$ ). La matrice des fonctions propres s'obtient alors comme suit :  $\mathbf{F} = \mathbf{\Phi}\mathbf{U}^*$  où  $\mathbf{U}^*$  est la matrice des vecteurs propres conservés ( $q \times m$ ),  $\mathbf{\Phi}$  la matrice des coefficients de Legendre (nombre de stade de lactation  $\times q$ ), et  $\mathbf{F}$  la matrice des fonctions propres (nombre de stade de lactation  $\times m$ ). Il est possible de réaliser la même réduction sur la matrice de variances-covariances de l'environnement permanent  $\mathbf{P}$  en utilisant ou non le même  $m$ . En utilisant les fonctions propres comme nouvelles covariables, nous obtenons une matrice de variances-covariances  $\mathbf{C}'$  de dimension  $m \times m$ . Le produit Kronecker avec la matrice de parenté devient donc  $\mathbf{H} \otimes \mathbf{C}'$  et la matrice résultante de ce produit est de dimension réduite.

Chez les bovins laitiers français, Leclerc et al. (2009) ont étudié l'impact de l'utilisation de la réduction d'une matrice de variances-covariances sur les valeurs génétiques estimées et ont montré que les index obtenus étaient les mêmes que ceux obtenus sans la réduction (corrélation  $> 0.99$ ). Cette réduction de la matrice de variances-covariances a été réalisée dans plusieurs études (Van der Werf et al., 1998; Druet et al., 2003) et les formes des vecteurs propres ont été étudiées. Druet et al. (2005) a utilisé directement les fonctions propres comme nouvelles covariables (issues de la matrice  $\mathbf{F}$  dans notre modèle) afin de modéliser les parties génétique et d'environnement permanent pour réduire la taille de la matrice résultant du produit Kronecker  $\mathbf{A} \otimes \mathbf{C}$  (avec  $\mathbf{A}$  matrice de parenté obtenue sans information génomique). L'orthogonalité des fonctions propres est un avantage en terme de convergence du modèle. De plus, la réduction permet de lisser la variance estimée et donc d'éviter les effets de bords aux extrémités de la

lactation souvent rencontrés avec les polynômes de Legendre. Un autre avantage de l'utilisation des fonctions propres est que les coefficients de régression obtenus peuvent avoir une interprétation biologique. Il a été montré que le premier coefficient de régression est associé à la production à la lactation et le deuxième à la différence de production entre le début et la fin de la lactation, pouvant être interprété comme une mesure de la persistance laitière (Druet et al., 2005).

### **3. Limites des modèles de régression aléatoire**

Le principal inconvénient des modèles de régression aléatoire est le temps de calcul pour estimer les valeurs génétiques des animaux. En effet, comparé à un modèle à la lactation, les modèles de régression aléatoire en caprin utiliseront en moyenne 7 fois plus de données par lactation. La matrice de parenté ( $A$  ou  $H$ ) n'est plus multipliée par une constante  $\alpha$  (comme dans un modèle à la lactation) mais est multipliée par un produit Kronecker avec la matrice de variances-covariances  $C$ , ce qui augmente considérablement la taille et la complexité de la matrice résultante ( $H \otimes C$ ). La matrice de parenté génomique  $G$  constituante de  $H$  est particulièrement dense ce qui rend les calculs encore plus longs qu'avec l'utilisation de la matrice  $A$  sans information génomique. L'utilisation d'un modèle ssGBLUP RRM en routine constitue donc un défi, que nous avons étudié dans le cadre de ce travail de thèse.

Dans la littérature, très peu d'études sont publiées sur la mise en œuvre d'un modèle de régression aléatoire couplé à des données génomiques pour modéliser les contrôles élémentaires (Oliveira et al., 2019b).

Les quelques études (Koivula et al., 2015; Kang et al., 2018; Oliveira et al., 2019d) ont montré que les précisions des valeurs génétiques étaient améliorées en utilisant les informations génomiques, par rapport à l'utilisation de l'information pedigree seule. L'accès à l'information génomique permet en effet une sélection précoce des animaux génotypés sans phénotypes.

Des analyses pangénomiques ont aussi été réalisées à partir de modèles de régression aléatoire ssGBLUP. Des SNPs ont été mis en évidence comme ayant une expression évoluant au cours du temps (Oliveira et al., 2019c).

En conclusion, l'étude de la forme de la courbe de lactation chez les caprins laitiers est d'un intérêt majeur pour répondre aux enjeux agro-écologiques, et économiques de la filière. Dans la suite du manuscrit, nous avons choisi d'étudier, dans un premier temps, la variabilité des courbes de lactation au niveau phénotypique afin de faire un état des lieux des formes de courbes caprines et d'étudier les différents facteurs de variation de ces courbes. Cette étude a été faite à l'aide d'une analyse en composantes principales sur les données des deux principales races caprines françaises (races Alpine et Saanen). Dans une deuxième partie, nous développerons un modèle de régression aléatoire adapté aux données caprines françaises afin de pouvoir sélectionner les animaux sur leur valeur génétique pour la

forme de courbe de lactation. Nous définirons les effets fixes et aléatoires du modèle, nous estimerons les paramètres génétiques de ce modèle, et nous comparerons les valeurs génétiques de ce modèle au modèle à la lactation actuel. Une première étude a été réalisée sur les données de chèvres en première lactation. Nous avons ensuite étendu le modèle développé aux données des trois premières lactations. Dans une dernière partie, nous avons intégré dans notre modèle l'information génomique disponible (génotypages 50K disponibles en race Alpine et Saanen). Le modèle proposé, ssGBLUP RRM, a été comparé aux modèles sans information génomique et à la lactation (avec ou sans information génomique). Plusieurs critères de comparaison ont été étudiés : la corrélation entre les index prédits pour les boucs sans filles et les écarts de production des filles (DYD), la pente de régression des DYD sur les index prédits, la différence de moyenne entre les DYD et les index prédits. Les corrélations entre les différents caractères au fil de la lactation ont été étudiés ainsi que les relations entre les caractères de production et la longévité ou la fertilité afin de comprendre si l'amélioration génétique de ces caractères de production améliore ou dégrade la longévité et/ou la fertilité, caractères très importants pour la rentabilité de l'élevage et le bien-être des animaux.



# **Chapitre 1: Etude de la diversité phénotypique des courbes de lactation des deux principales races caprines françaises**

L'objectif de ce chapitre est l'étude de la diversité phénotypique des courbes de lactation dans les populations françaises de chèvres de race Alpine et Saanen, en préambule aux études du déterminisme génétique qui sous-tend la forme de ces courbes. Cette première étude a également permis d'explorer les facteurs d'environnement et génétiques qui impactent la forme des courbes. Ce chapitre se divise en deux parties : la première est constituée d'un article publié dans la revue « Journal of dairy science » en 2018, la deuxième présente des résultats complémentaires sur la classification des formes de courbes et sur les formes phénotypiques des courbes de production selon l'année de mise bas et la race.

L'article présente une étude de classification des formes de courbes de lactation. L'ensemble des données se composait de 2 231 720 contrôles élémentaires, de 213 534 chèvres Saanen et Alpine, mesurés entre septembre 2008 et juin 2012. Une première étude, réalisée à partir d'une analyse en composantes principales a permis de résumer la forme des courbes de lactation suivant 3 composantes principales: la première composante représentait le niveau de production laitière tout au long de la lactation, la deuxième composante représentait la persistance laitière et la troisième composante rendait compte de la production laitière en milieu de lactation. Puis, à partir des scores de chaque courbe de lactation issus de l'analyse en composante principale, les lactations ont été rassemblées dans 5 différents groupes ou clusters. La plupart des lactations avaient une forme similaire à la courbe moyenne, à l'exception de 30 % des lactations qui se répartissaient en 3 groupes : elles avaient un niveau de production élevé au pic, et une persistance différente selon le cluster. L'index génétique pour la production laitière et la région d'origine de l'élevage étaient les facteurs les plus liés au niveau de production (1<sup>ère</sup> composante). Le mois de mise bas, la race, et le stade de gestation avaient les plus grands effets sur la persistance (2<sup>ème</sup> composante). Le mois de mise bas était le facteur le plus lié à la production en milieu de lactation (3<sup>ème</sup> composante). L'effet troupeau avait un effet important sur les 3 composantes principales.

La première partie des résultats complémentaires présente une étude comparative des profils de courbes obtenus en faisant varier le nombre de clusters de 3 à 8 afin de déterminer le nombre optimal de clusters à retenir. La deuxième partie présente une étude de l'évolution de la forme de courbes de lactation selon l'année de mise bas et selon la race à partir d'un jeu de données plus complet que celui utilisé dans l'article.

## **I. Diversité des formes de courbes de lactation des chèvres Alpine et Saanen françaises (Article1)**



## Diversity of dairy goat lactation curves in France

M. Arnal,\*†<sup>1</sup> C. Robert-Granié,\* and H. Larroque\*

\*GenPhySE, Université de Toulouse, INRA, INPT, ENVT, 31326 Castanet-Tolosan, France

†Institut de l'élevage (Idele), 31326 Castanet-Tolosan Cedex, France

### ABSTRACT

A high level of production at the peak of lactation may be associated with animal health disorders, high feeding costs, and reduced milk supply throughout the year. The objective of this study was to typologize the lactation curves in French dairy goats and analyze the influence of environmental and genetic factors on these curves. The data set consisted of 2,231,720 monthly test-day records of 213,534 French Saanen and Alpine goats recorded between September 2008 and June 2012. First, principal component analysis classified the shape of the lactation curves into 3 principal components: the first component accounted for milk yield level throughout lactation, the second component accounted for lactation persistency, and the third component accounted for milk yield in mid-lactation. Then, from the principal component scores, the lactations were clustered into 5 different groups. Most lactations had a similar shape to the mean curve, except 30% of the lactations that fell into 3 clusters that had a high production level at the peak and then a different persistency according to cluster. Estimated breeding value for milk yield and home region of breeding were the factors most related to lactation production level. Month of kidding, breed, and gestation stage had the biggest effect on persistency. Month of kidding was the factor most strongly linked to mid-lactation production. A herd effect was observed on all 3 principal components.

**Key words:** French dairy goat, test-day milk record, classification, lactation curve

### INTRODUCTION

A goat's level of milk production is often evaluated by its daily average yield or by its total production throughout lactation. However, neither approach factors in the change in milk production according to DIM (i.e., lactation curve). Shape of the lactation curve can affect an animal's dietary needs, and consequently its

health, as well as the time-course scale of milk production. Indeed, a high level of production at the peak can lead to an energy deficit for the animal if it is unable to ingest enough food to compensate. Goats, like dairy cattle, will then begin drawing on their body reserves, which may lead to metabolic and reproductive disorders (Gipson and Grossman, 1990). A goat in energy deficit has to be fed expensive supplement concentrates and high nutrition value forage, whereas a persistent goat can be fed lower-quality forage and fewer concentrates (Sölkner and Fuchs, 1987), which is an asset to farm systems aiming to reduce inputs. The ability to maintain lactation in a partially season-cycled production context meets the processing industry demand for more evenly spread milk production, in line with the needs of the market (Kearney, 2010), while also avoiding the need to freeze cheeses, which can change their flavor (Van Hekken et al., 2005).

Research on shape of the lactation curve has mainly addressed milk persistency, trying to find indicator criteria of the slope of the curve after the peak of lactation (Gengler, 1996), or modeling of overall curve shape. Some authors have focused on finding mathematical models that describe the biological processes of milk production by mammary gland cells (Pollott, 2000; Elvira et al., 2013a). Others have proposed a statistical approach by modeling the shape of the curve from elementary controls such as random regression test-day models including genetic and nongenetic effects (Menéndez-Buxadera et al., 2010; Mucha et al., 2014; Brito et al., 2017) or by using methods such as principal component analysis (PCA) to sum up the information contained in the data (Carta et al., 2014). Principal component analysis is a dimension-reduction tool that reduces a large set of variables to a small set containing most of the original information. Macciotta et al. (2006) used a PCA based on rank of bovine test-days where the first principal component (PC) summarized the averaged milk yield level and the second component summarized persistency. van der Werf et al. (1998) and Druet et al. (2003) performed data dimension reduction on a covariance genetic matrix derived from a random regression test-day model and also obtained 2 PC with the same biological interpretation.

Received April 26, 2018.

Accepted July 23, 2018.

<sup>1</sup>Corresponding author: [mathieu.arnal@idele.fr](mailto:mathieu.arnal@idele.fr)

Bouloc (1991) showed in French dairy goats that lactation curve shapes may depend on breed and genetics of the animal. Menéndez-Buxadera et al. (2010) showed in Murciano-Granadina goats that persistency is heritable (0.20). Curve shape may also vary with breeding and physiological factors such as parity, age at kidding, kidding season, dry-period length, gestation stage, and herd environment factors (Gipson and Grossman, 1990; Bouloc, 1991).

To our knowledge, no recent study has been done on lactation curve shapes using test-day records in French dairy goats. The latest study, performed 27 yr ago, was by Bouloc (1991), who classified protein yield curves based on different input features: lactation duration, average daily protein yield, persistency of the protein yield calculated as the ratio of d 50–d 200-period protein yield to d 31–d 60-period protein yield.

Here we typologized lactation curve shapes based on PCA scores taking into account test-day lactation stage in the 2 main French dairy goat breeds (Alpine and Saanen), using a recent large data set of test-day milk production records from the French genetic national database. We went on to study the influence of different genetic and environmental factors on these curves.

## MATERIALS AND METHODS

### Data

The data were extracted from the genetic database used for the national French genetic evaluations. France has been running a goat milk yield recording scheme since 1989 that applies the standard A, D, AT, AZ, or CZ methods registered by ICAR (2017) with sampling at 4- or 5-wk intervals. The data retained here was collected from the A, D, and AZ protocols with precise measurement of daily production, taking into account amount of milk at the 2 daily milkings. About 60% of the overall database comes from these protocols. The weights of milk produced by a goat in the morning and evening milkings were added together to obtain the weight of daily milk.

Lactations had to have a conventional duration between 180 and 350 d to be conserved. Test-day records before 7 and beyond 300 DIM were removed to have a sufficient number of records per DIM. Within a lactation, a goat was required to have at least 4 test-day records, with the first test day recorded before d 81 of lactation to assess production at the beginning of lactation, the peak of lactation occurring around 40 DIM. Only goats from herds having at least 30 conserved lactations per year were kept in the analysis.

The final data set consisted of 2,231,720 test-day records collected from 319,975 lactations of 213,534

French Saanen and Alpine goats in 910 French herds spread throughout France.

As described by Pulina et al. (2018), in France goats production systems present a great diversity according to feeding (grazing or not, concentrate quantity), reproduction (out of season or not), transformation of milk (cheese maker/deliverer), herd size, or even breeding goal (milk yield/composition). However, the database used recording animal milk productions and did not contain detailed information on production system of herds. Almost 60% of the lactations retained were on Alpine goats, and mainly (77.5%) in the north-west of France. The final data set contained 77% of lactations in parity 1 to 3, with other lactations up to parity 8 (Table 1). Lactations lasted 288 d on average, for an averaged total production of 964.1 kg of milk (minimum: 132.5 kg, maximum: 2,615 kg, SD: 253).

Goats were measured on average 7 times per lactation (minimum: 4, maximum: 11, SD: 1), at an average measure-to-measure interval of 39 d (minimum: 10, maximum: 158, SD: 10.9). Goats were not all measured at the same DIM, and production varied from 0.5 to 11.4 kg per test-day record.

For each lactation, parturition date served to calculate age at kidding and define kidding month. For goats in their second lactation or more, “dry-period length” was defined as the length of the period between the end of the previous lactation and parturition date. Breeders give the end date of the previous lactation; otherwise, it is a deducted date by adding to the date of the last test day 14 d (or 28 d if the interval with the following lactation is greater than 56 d). Gestation stage at 300 DIM was calculated for goats having subsequent lactation.

### Analysis

All statistical analyses were performed using R software (RCore Team, R Foundation for Statistical Computing, 2014, R: A Language and Environment for Statistical Computing, Version 3.0.2, Vienna, Austria, <http://www.R-project.org>).

The general approach of the analyses was to perform (1) a PCA on test-days after smoothing the data, then (2) a classification of the curves, and finally, (3) a study of the relationships between environmental and genetic factors and curve shapes.

**Principal Component Analysis.** The PCA was performed on the test-day milk yield data, for each lactation, using the R package *fdapace* (Dai et al., 2016). *Fdapace* was well-adapted to the structure of our data set: sparse data, irregular measurement stage, irregular number of measurements per lactation, small number of repeated measurements available per lactation. First,

**Table 1.** Percentage of data for each level of different factors in French dairy goats

Variable	%
Parity	
1	32.9
2	26.5
3	17.8
4	10.9
5	6.1
6	3.3
7	1.7
8	0.8
Breed	
Alpine	59.1
Saanen	40.9
Milk EBV (kg)	
–349 to –79	25.1
–79 to –33	25.3
–33 to 13	24.7
13 to 301	24.9
Region <sup>1</sup>	
NE	17.7
NW	77.5
S	4.9
Kidding month	
1	24.51
2	27.89
3	14.24
4	3.50
5	1.13
6	0.19
7	0.03
8	1.06
9	9.02
10	8.21
11	6.83
12	3.40
Kidding age <sup>2</sup> (mo)	
Other	12
P4+	22.8
Normal	50.8
Young	2.9
Old	11.6
Dry-period length <sup>3</sup> (d)	
1–60	19.5
60–90	32.7
90–120	9.1
Other lengths	5.8
Parity 1	32.9
SCC EBV <sup>4</sup> (base 100)	
101–105	23.8
105–131	21.2
65–97	26.2
97–101	25
EBV missing	3.8
Gestation stage <sup>5</sup> (d)	
120–150	5.2
150–180	1.3
1–60	5.9
60–90	32.5
90–120	24
Other stages	1.3
No consecutive lactation	29.9

<sup>1</sup>French herd region (NW = north-west; S = south; NE = north-east).

<sup>2</sup>Age at kidding for the first 3 parturitions (young = 9–10 mo at first kidding, 16–22 mo at second kidding, or 22–34 mo at third kidding; normal = 11–13 mo at first kidding, 23–25 mo at second kidding, or 35–37 mo at third kidding; old = 14–30 mo at first kidding, 26–56 mo at second kidding, or 38–77 mo at third kidding; P4+ = parity 4 and more; other = other ages at kidding for the first 3 parturitions).

<sup>3</sup>Dry-period length before last kidding (d; for goats in second and more lactation).

<sup>4</sup>Somatic cell count EBV (base 100): <100: increase in SCC, >100: decrease in SCC.

<sup>5</sup>Gestation stage (d) at 300 DIM.

fdapace calculated the smoothed mean curve using local linear Gaussian kernel regression between 51 equidistant support points (default option) aggregating all the measurements together. Then, the raw covariance for each curve was calculated, and all covariances were aggregated to generate the sample raw covariance. The smoothed covariance was estimated from the off-diagonal of sample raw covariance. Eigen analyses performed on the smoothed covariance matrix gave the estimated eigenfunctions, eigenvalues, and PC scores (Yao et al., 2005).

To evaluate the accuracy of the smoothed curves using local linear regression, the true and predicted milk yields were compared by studying the coefficient of determination ( $R^2$ ). This analysis can be done according to the number of PC.

**Classification.** Using the PC scores (**SPC**) of the lactations, an unsupervised classification, also called cluster analysis, was performed using the R package *rmixmod* (Langrognon et al., 2016). *Rmixmod* is a tool to fit a mixture model of multivariate components to a data set (Biernacki et al., 2006). A new feature of the *Rmixmod* software is that it considers a parameterization of the variance matrix of a cluster through its eigenvalue decomposition, leading to many meaningful models for clustering (Celeux and Govaert, 1995). Each variance matrix is decomposed as the product of one parameter determining the volume of the cluster, a matrix of eigenvectors determining its orientation, and a diagonal matrix determining its shape (Lebret et al., 2015). By allowing some but not all of these parameters and matrices to vary between clusters, parsimonious and easily interpreted models can be obtained. As SPC are continuous variables, we tested 14 different Gaussian models with different variance-covariance structures as proposed by Celeux and Govaert (1995), and we selected the best one based on the Bayesian information criterion (**BIC**). The number of components of the mixture model is generally known or fixed beforehand. If this number is not known, it is possible to compare models with different numbers of components and to use criteria proposed in the *Rmixmod* package to make the choice. Here, the optimal number of components was researched between 2 and 5 clusters, and the BIC was used to determine it.

The mixture parameter was estimated through maximization of the log-likelihood. In a first step, a stochastic expectation maximization algorithm was used 100 times with 1,000 iterations each time. Then, an expectation maximization algorithm was performed 100 times with 1,000 iterations each time. The stochastic expectation maximization algorithm used random drawing at each iteration, which avoided converging to

stationary points of the log-likelihood (Biernacki et al., 2003).

**Relationships Between Environmental and Genetic Factors and PC.** The effects of environmental and genetic factors on the first 3 PC were studied according to the following model:

$$y_{ijklmnopqr} = \mu + B_i + R_j + N_k + A_l + M_m + D_n + G_o + I_p + SCI_q + \varepsilon_{ijklmnopqr}, \quad [\text{model 1}]$$

where  $y_{ijklmnopqr}$  is the dependent variable (SPC) and  $\mu$  is overall mean. The fixed effects tested were  $B_i$ , breed (Saanen or Alpine);  $R_j$ , French herd region (NW = north-west; S = south; NE = north-east);  $N_k$ , parity (1 to 8);  $A_l$ , age at kidding for the 3 first parturitions (5 levels: 1 = 9–10 mo at first kidding, 16–22 mo at second kidding, or 22–34 mo at third kidding; 2 = 11–13 mo at first kidding, 23–25 mo at second kidding, or 35–37 mo at third kidding; 3 = 14–30 mo at first kidding, 26–56 mo at second kidding, or 38–77 mo at third kidding; 4 = parity 4 and more; 5 = other ages at kidding for the first 3 parturitions);  $M_m$ , kidding month (12 levels: 1 = January, . . . , 12 = December);  $D_n$ , class of dry-period length before the last kidding (d) for goats in second or more lactations (5 levels: 1 = [1–60], 2 = [60–90], 3 = [90–120], 4 = other lengths, 5 = 1st parity);  $G_o$ , class of gestation stage (d) at 300 DIM (7 levels: 1 = [1–60], 2 = [60–90], 3 = [90–120], 4 = [120–150], 5 = [150–180 d], 6 = other stages, 7 = no consecutive lactation);  $I_p$ , class of genetic EBV based on total-lactation milk yield (kg) from official French genetic evaluation (4 levels: 1 = [–349 to –79], 2 = [–79 to –33], 3 = [–33 to 13], 4 = [13 to 301]);  $SCI_q$ , class of SCC EBV (base 100): <100: increased SCC, >100: decreased SCC) from official French genetic evaluation (5 levels: 1 = [65–97], 2 = [97–101], 3 = [101–105], 4 = [105–131], 5 = missing information); and  $\varepsilon_{ijklmnopqr}$  are the normally distributed residuals. Model 1 using  $SPC_x$  with  $x = \{1, \dots, X\}$  is noted  $M_{SPC_x}$ . Percentages of data for each class of the different factors are presented in Table 1. Despite the low percentages of data, particularly for the kidding month effect in summer, the lowest number of data in a class was 87.

Each effect of model 1 was successively removed to study its relative effect using a Fisher test and comparing the  $R^2$  of the model.

Least squares means were calculated using proportions of the population as weights to take into account the unbalanced number of data in each level of effects, using the R package lsmeans (Lenth, 2016).

The influence of 2-way interactions between herd and year of lactation (2,145 classes) on SPC was studied

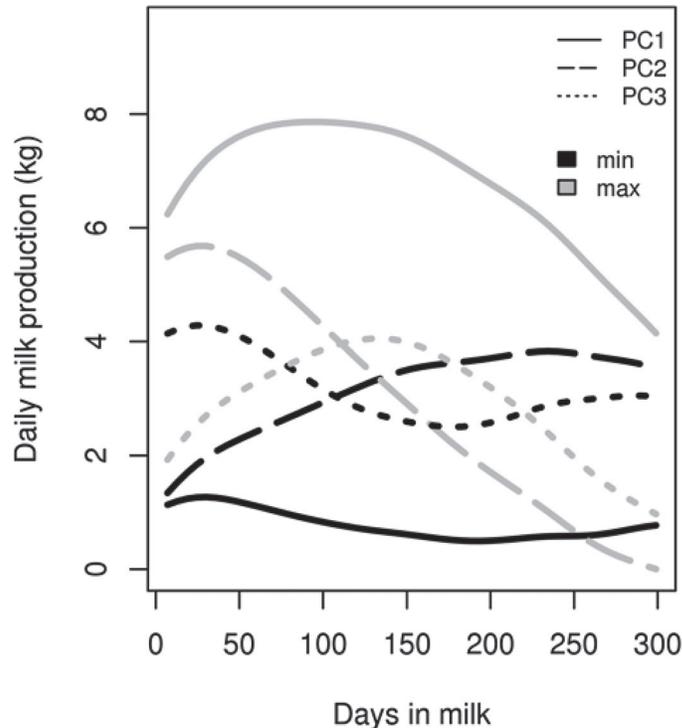
from the residuals of model 1, due to the total confusion between herd number and region.

## RESULTS AND DISCUSSION

### Principal Component Analysis

The first 3 components represented 99.5% of the original variance: 81% for PC1, 15%, for PC2, and 3.5% for PC3. It was not considered useful to further study PC due to the total variance already explained. The  $R^2$ , used to evaluate the precision of the fit by local linear regression, was 0.92 with 3 PC and fell to 0.88 with only the first 2 PC. Clearly, adjustment was better with all 3 components, especially considering the smoothed lactation curves of extreme goats after peak lactation (Figure 1). Therefore, 3 PC were retained.

Plotting extreme lactation curves, with SPC extremes for one PC and other SPC close to 0, showed that each PC characterized a component of curve shape (Figure 1). Principal component 1 was an indicator of level of milk production throughout lactation, where a low SPC1 corresponds to low milk production (mean  $\pm$  SD: SPC1:  $-1.4 \pm 144.4$ , total lactation production (kg):  $964.1 \pm 253.3$ ; correlation between SPC1 and lactation



**Figure 1.** Local linear regressions (curves) for lactations having a minimum (min) or a maximum (max) score for 1 principal component (PC; first, PC1; second, PC2; or third, PC3) and scores close to 0 for the 2 remaining principal components.

**Table 2.** Mean (SD) per cluster for the first, second, and third principal component (PC1, PC2, and PC3, respectively), and proportion of lactations per cluster

Cluster number	PC1	PC2	PC3	Proportion (%)
1	-76.5 (94.4)	107.1 (40.7)	-39.5 (17.7)	4.2
2	-121.7 (80.9)	7.2 (40.8)	9.7 (18.2)	39
3	221.9 (120.5)	-41 (69.3)	-23.8 (30.5)	8.9
4	101.1 (108.7)	64.2 (59.5)	29.4 (26.5)	16.2
5	41.7 (78)	-29 (42.1)	-4.9 (19)	31.7

production: 1); PC2 was an indicator of the difference in milk production between the peak and end of lactation, where a low SPC2 was equivalent to high milk production at the end of the lactation (mean  $\pm$  SD: SPC2:  $4.9 \pm 61.9$ , production difference (kg) between DIM 40 and 240:  $1.11 \pm 0.9$ ; correlation between SPC2 and production difference: 0.99); finally, PC3 was an indicator of milk production after the peak of lactation at around 120 DIM: a low SPC3 equated to low production around 120 DIM (mean  $\pm$  SD: SPC3:  $3.3 \pm 27.2$ , production difference between DIM 120 and the sum between the first and the last DIM (kg):  $-1.6 \pm 0.7$ ; correlation between SPC3 and production difference: 0.97).

On cow data, some authors have defined persistency as a degree of maintenance of peak milk yield (Grossman et al., 1999), others as the degree of decline in milk yield after the peak, using indicators of the speed of decline such as the ratio of milk yield in the first 100 DIM to milk yield in the last 100 DIM (Sölkner and Fuchs, 1987). Based on these definitions of persistency, PC2 and PC3 could both be considered persistency indicators, where “complete” persistency could be a minimal value for PC2 and a maximal value for PC3. For the rest of the study, we defined persistency as a difference in milk production between the peak and end of lactation. Consequently, PC2 was associated with persistency and PC3 was associated with milk production in mid-lactation.

Macciotta et al. (2006) performed a PCA on a correlation matrix from dairy cow test-day milk yields. They used not DIM but rank of the test-day during lactation and obtained 2 PC: the first representing total-lactation production, and the second representing persistency. Carta et al. (2014), with a similar approach, obtained the same results in dairy sheep.

### Classification of Curve Shapes

The best Gaussian mixture model for our data based on the BIC was defined by free volume, equal shape, and equal orientation between the different clusters. The number of clusters tested varied from 2 to 5. The

BIC criteria were 10,775,614, 10,767,801, 10,763,784, and 10,760,253 for 2, 3, 4, and 5 clusters, respectively. The best model was then obtained by considering 5 clusters. To obtain clusters with sufficient sizes, it was not desirable to have a greater number of clusters. On the whole data file, all iterative procedures used to build the best model were run several times, and the same clusters were found each time.

Table 2 reports mean and standard deviation per cluster of the lactation scores for each PC. Figures 2a and 2b represent each lactation according to their PC scores and their cluster membership. Each ellipse represents the internal covariance between 2 PC within a cluster.

Clusters 3 and 4, with the biggest ellipses in Figures 2a and 2b, had the largest SPC variances, with standard deviations of 120.5 and 108.7, respectively, for PC1. In contrast, clusters 2 and 5, although including proportionally more lactations (39 and 31.7%, respectively), had the lowest SPC variances (80.9 and 78, respectively, for PC1, for example).

The averages of daily milk production were plotted for all the data as well as according to cluster number (Figure 3). In each cluster, average daily production did not fluctuate between 2 consecutive DIM. Each shape of average daily production looked like a smooth function.

We estimated the probability of each lactation belonging to a cluster. To challenge the robustness of the clustering, lactations with a cluster probability higher than 0.7 were conserved ( $n = 161,676$  lactations; i.e., 51% of the data set). In this case, the clusters looked well separated (the ellipses did not overlap) and there was no substantial change in the means curve profiles (results not shown).

According to BIC, the repeatability of the classification results, the smoothness of the mean curve of each cluster and the similarity of the mean curves using a probability higher than 0.7, the choice of 5 clusters appeared robust.

Principal component 3 (representing only 3.5% of the original variance) contributed significantly to the diversity of the shape of curves highlighted, which is

easily understandable as PC3 was responsible for the flexibility in mid-lactation. Indeed, not taking into account PC3 in the classification would make cluster 1 disappear and change the shape of clusters 3 and 4 (results not shown).

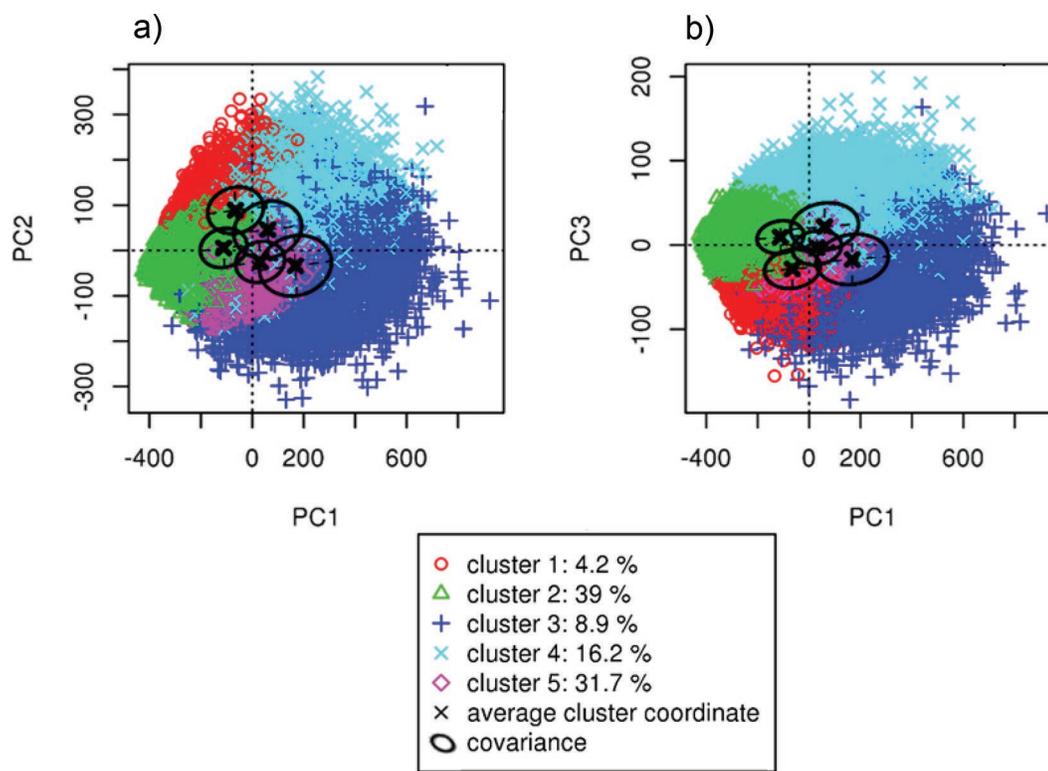
The average lactation curve could be decomposed into 3 phases according to daily milk yield: (1) an initial rapid increase until the 50th DIM, (2) a stabilization for 50 d, and finally, (3) a slow decrease through to the end of lactation.

Day of peak lactation yield differed according to shape of the lactation curve. The lactation peak, defined as the day of the maximal daily milk production, was before 40 DIM for clusters 1 and 5 but was 10 d later for clusters with the highest milk production level (clusters 3 and 4). Goats in cluster 4 maintained peak milk production for a far longer time than goats in cluster 1.

Cluster 2 (the green curve in Figure 3), representing 39% of lactations, was characterized by the lowest milk yield level, intermediate means for PC2 and PC3, and a curve shape similar to the mean curve (the black curve in Figure 3). Cluster 5 (the pink curve in Figure 3), representing 31.7% of lactations, was defined by intermediate means for PC1 and PC3 and the second-lowest

mean for PC2 corresponding to a “flat” curve (Figure 3). The mean curves of 3 others clusters (nearly 30% of the data) were very different from the all-data set mean curve. Indeed, cluster 3 (the dark blue curve), characterizing 8.9% of the data, had the highest mean milk production (PC1) and highest persistency (PC2) but the second-lowest value for PC3. Cluster 1 (the red curve), which ranked fourth for milk production level, had the lowest persistency and the highest decrease in mid-lactation, resulting in a very specific lactation peak (Figure 3), and constituted the smallest group with 4.2% of the lactations. Cluster 4 (the light blue curve in Figure 3), which was characterized by the second-highest total milk production level and ranked second for nonpersistency and first for maintaining milk in mid-lactation, giving an early-bulging convex curve, represented 16.2% of lactations.

On a sample of 146,280 French goats of Saanen breed (31.8%), Alpine breed (56.1%), and other breeds (12.1%) milked in 1987, Bouloc (1991) realized a classification based on 3 criteria: (1) persistency of protein yield calculated as the ratio of d 50–d 200-period protein yield to d 31–d 60-period protein yield, (2) average daily protein yield, and (3) duration of lactation. The author identified 8 distinct clusters and was able to



**Figure 2.** Plot of the lactations according to (a) the first 2 principal components (PC1 and PC2), and (b) the first and third principal components (PC1 and PC3).

**Table 3.** Coefficient of determination ( $R^2$ ) of model 1 where the lactation scores of each of the 3 principal components (PC) are explained by all the environmental factors, and of model 1 successively reduced by 1 environmental factor

Item	PC1	PC2	PC3
Model 1	0.402	0.225	0.135
Model 1 without parity	0.399	0.215	0.134
Model 1 without kidding age	0.399	0.222	0.133
Model 1 without kidding month	0.395	0.199	0.034
Model 1 without dry-period length	0.400	0.222	0.135
Model 1 without gestation stage	0.395	0.208	0.130
Model 1 without breed	0.394	0.201	0.135
Model 1 without milk EBV	0.161	0.224	0.135
Model 1 without SCS EBV	0.401	0.219	0.135
Model 1 without region	0.385	0.223	0.134

highlight: (1) lactations with atypical forms (i.e., low production during the first weeks, which returned to an average level thereafter), (2) persistent curves with a difference of 1.5 kg of milk yield between the peak and end of lactation, (3) curves with marked peaks and a difference of more than 2.0 kg in milk yield between the peak and end of lactation. In our data set, only the mean lactation curve of cluster 2 (low milk yield level and medium persistency) was similar to what Bouloc found. Bouloc's study did not find the profiles of clus-

ters 3 and 5 (high production and high persistency), probably due to lower levels of genetics and management 27 yr ago, nor the profiles of clusters 4 and 1, whereas the atypical curves found by Bouloc (1991) were not found here, probably due to the use of different phenotypes for the classification.

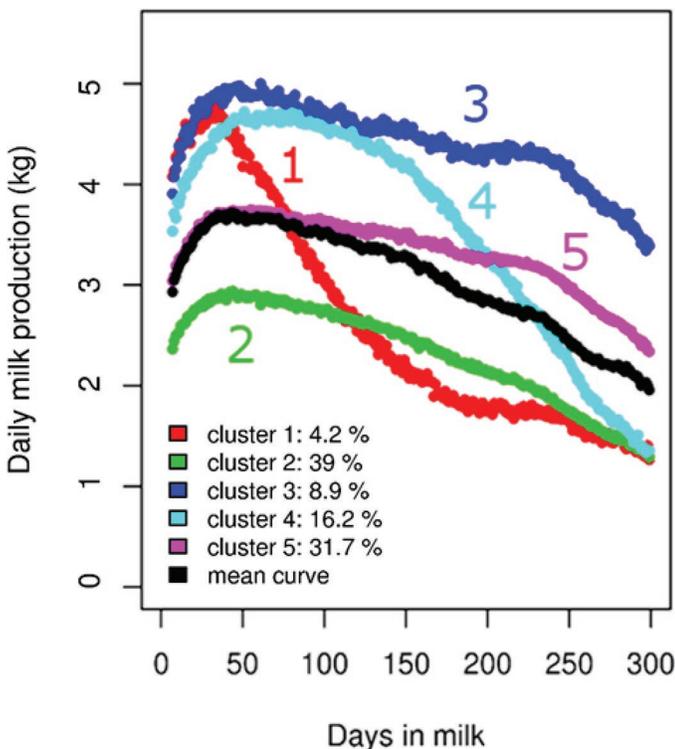
#### Relationship Between Environmental and Genetic Factors and PC

Model 1 was used to estimate the relative influence of each environmental and genetic factor on each PC. For each PC, Table 3 presents the  $R^2$  of model 1 and the  $R^2$  of model 1 iterative reduced by one of the factors. The Fisher's tests between model 1 and the different sub-models were all significant ( $P < 10^{-5}$ ) except for breed on PC3 ( $P = 0.08$ ).

The  $R^2$  of model 1 was 0.402 for PC1. The biggest decrease in  $R^2$  ( $R^2 = 0.161$ ) occurred when EBV of total milk yield was removed from model 1, meaning that it was the factor most related to PC1, far ahead of region ( $R^2 = 0.385$ ) and breed ( $R^2 = 0.394$ ). For PC2, the  $R^2$  of the model 1 was 0.225. Principal component 2 was not strongly related to environmental and genetic factors, as the biggest association was found for kidding month with  $R^2$  decreased to 0.199, followed by breed ( $R^2 = 0.201$ ), gestation stage ( $R^2 = 0.208$ ), parity ( $R^2 = 0.215$ ), and SCS EBV ( $R^2 = 0.219$ ). The  $R^2$  of model 1 for PC3 was 0.135. The only factor strongly linked to this PC was kidding month, with  $R^2$  reduced to 0.034.

Figures 4, 5, 6, 7, and 8 chart the effects on PC1 and PC2 of factors having the greatest effects.

**Milk EBV and Region.** Clearly, PC1 score increased with EBV milk yield classes (Figure 4). Furthermore, the correlation between milk EBV and PC1 was +0.51, meaning that animals with the highest genetic breeding value for total milk yield had the highest level of daily milk yield. Milk EBV was not linked with PC2, meaning that genetic selection on total-lactation



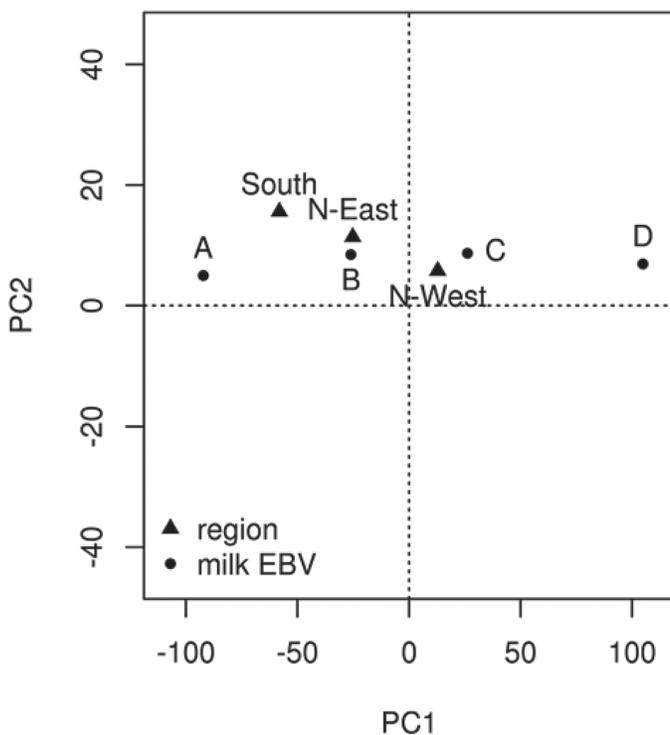
**Figure 3.** Average lactation curves for all data (in black) and according to cluster number (each point corresponds to the average milk output of the animals).

milk production was not linked to lactation persistency. The region factor showed a gradient of increasing production from south < north-east < north-west France.

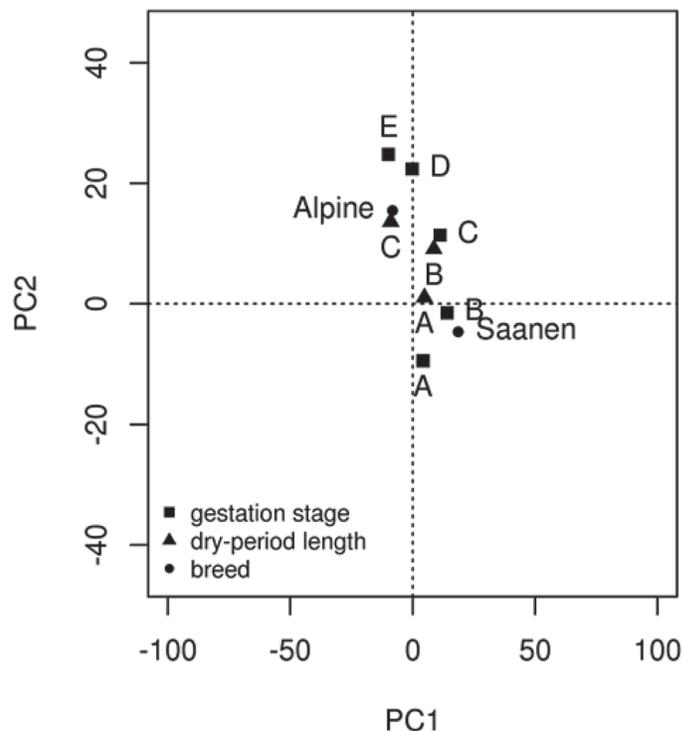
Bouloc (1991) observed differences in milk production between regions: goats from Pays-de-la-Loire (north-west in our study) were more numerous in high-producing groups whereas goats from Midi-Pyrenees (South in our study) were relatively more numerous in lower-producing groups.

**Gestation Stage, Dry-Period Length, and Breed.** Goats had more persistent lactations when their gestation was less advanced at 300 DIM and when they were dried for a short period lower than 90 d before their last kidding. Saanen goats were more productive and more persistent than Alpines (Figure 5).

Like the current study, Bouloc (1991) also observed that Saanen goats were more numerous in the groups of persistent lactations. Furthermore, Rupp et al. (2011) observed in the French goat population that Saanens were more productive than Alpines. Like the current study, Knight and Wilde (1988) reported that gestation causes a decrease in milk yield compared with nonges-

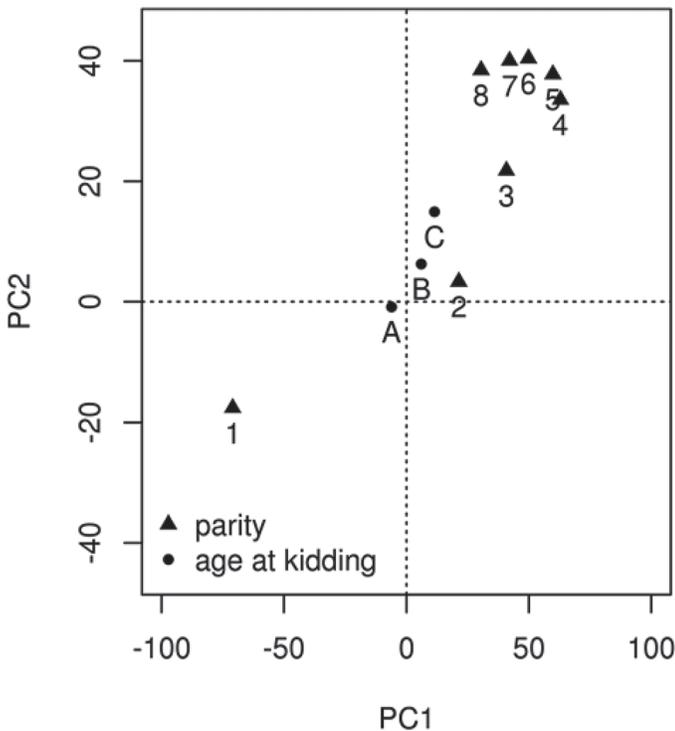


**Figure 4.** Least squares means on the first 2 principal components (PC1 and PC2) of the environmental factors most strongly linked to the first principal component: milk EBV (kg) from the French genetic evaluation model (A: -349 to -79, B: -79 to -33, C: -33 to 13, D: 13 to 301), and home region of the herd (N-West = north-west; S = south; N-East = north-east).



**Figure 5.** Least squares means on the first 2 principal components (PC1 and PC2) of the environmental factors strongly linked to the second principal component: gestation stage (d) at 300 DIM (A: 1-60, B: 60-90, C: 90-120, D: 120-150, E: 150-180), dry-period length (d) (A: 1-60, B: 60-90, C: 90-120), and breed (Saanen or Alpine).

tating does. Although their study counted a limited number of goats, they found that the influence of gestation was significant after 8 wk of gestation. Salama et al. (2005), on goats milked once a day, also found that the gestation caused a decrease of milk yield on gestating goats compared with goats in extended lactation. In the dairy cow, Druet et al. (2003) also showed a decrease of milk yield according to the progress of gestation stage. The effect of dry-period length on lactation curve has been understudied. Knight and Wilde (1988) specified that an absence of dry period cause a higher peak and lower persistency. Caja et al. (2006) studied the effect of dry-period length on 17 goats and found that no-dry-period goats produced less than dry-period goats. Atashi et al. (2013) studied the influence of the dry period length on lactation curve shape, in dairy cows. As in our study, they noticed the best persistency, the smallest peak, and the higher milk lactation production for the cow dried previously during a short time. In contrary, they observed also that animals previously dried for a long period produced higher milk yield at the peak than animals dried during an average period. Similarly, Elvira et al. (2013b) showed that the longer dry periods were not associated with high milk yield



**Figure 6.** Least squares means on the first 2 principal components (PC1 and PC2) of the environmental factors linked to the first 2 principal components: parity (1 to 8) and age at kidding (A: 9–10 mo at first kidding, 16–22 mo at second kidding or 22–34 mo at third kidding; B: 11–13 mo at first kidding, 23–25 mo at second kidding or 35–37 mo at third kidding; C: 14–30 mo at first kidding, 26–56 mo at second kidding or 38–77 mo at third kidding).

productions throughout the lactation, on their sheep data.

**SCS EBV.** Goats had a more persistent lactation when they had a high SCS EBV (less SCC). The relationship between high SCC index (indicator of genetically low SCC levels) and persistency in goats remains unknown. One hypothesis could be that goats with genetically low SCC levels are more resistant to mammary infections and thus better able to maintain high milk production. Indeed, Leitner et al. (2004) reported that infected goats produced less milk than the noninfected goats.

In bovines, Appuhamy et al. (2009) showed a negative genetic correlation ( $-0.24$ ) between persistency and a mastitis occurring in the second half of lactation. Cole and Null (2009) found a negative correlation between SCS persistency and milk persistency. Note that Nayeri et al. (2017) found a common QTL between persistency and mastitis.

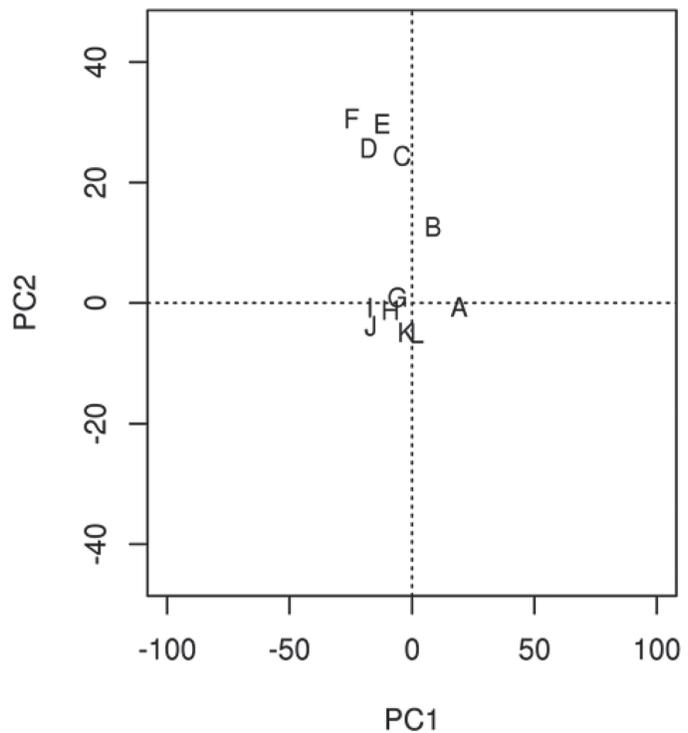
**Age at Kidding and Parity.** Goats that kidded youngest (9–10 mo at first kidding, 16–22 mo at second kidding, or 22–34 mo at third kidding) had a lower milk production levels but higher persistency (Figure

6). From first to fourth parity, the goats produced more milk on average but their lactations were less persistent. Particularly, this highlights the important difference in lactation production potential between goats in their first parity and others parities. After the fourth parity, level of milk production decreased more and more whereas lactations were a little less persistent than in the fourth lactation.

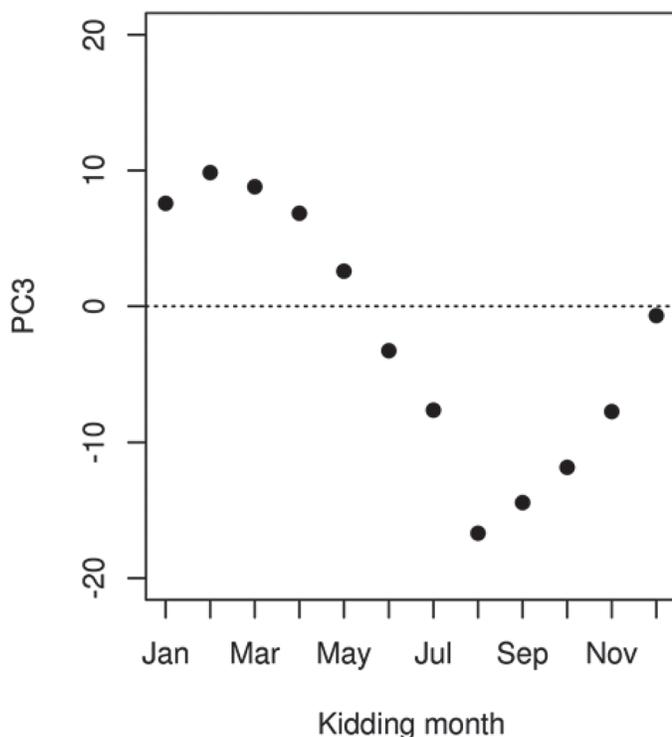
León et al. (2012) analyzed the production of 38,000 Murciano-Granadina breed goats following their first 5 lactations. Like in the current study, they showed that persistency was greater in first lactation than second lactation, which was in turn greater than third lactation and more.

In the French goat population, Boichard et al. (1989) showed that age at kidding had a positive near-linear relationship with milk yield. Bouloc (1991) found the same effect of age at kidding on shape of the lactation curve as the current study: goats kidding before 12 mo had a higher lactation persistency. French dairy cattle showed a similar pattern (Leclerc et al., 2008).

**Kidding Month.** Lactations were less persistent in goats kidding during the spring but more persistent in goats kidding in autumn (Figure 7). Note that the end



**Figure 7.** Least squares means on the first 2 principal components (PC1 and PC2) of the environmental factor the most linked to the second principal component: kidding month (A = January; B = February; C = March; D = April; E = May; F = June; G = July; H = August; I = September; J = October; K = November; L = December).



**Figure 8.** Least squares means on the third principal component (PC3) of kidding month, which is the environmental factor most linked to the PC3.

of the more persistent lactation occurred when days were long (at the beginning of summer), whereas the end of the less persistent lactation occurred when days were short (at the beginning of winter).

The effect of kidding month on PC3 showed broad variability, with a maximal value in February and a minimum value in August (Figure 8). Principal component 3 represented the production at approximately 120 DIM, and variations of the least squares means followed the evolution of the photoperiod duration with 120 d difference. For example, a goat that kids in February will be at its 120 DIM in June when photoperiod is maximal, whereas a goat that kids in August will be at its 120 DIM in December when photoperiod is minimal.

The goat producing more during the longer days will yield more milk from the middle to the end of the lactation according to kidding month. The influence of kidding month on persistency had already been highlighted in studies by Bouloc (1991), Montaldo et al. (1997), and León et al. (2012). In these 3 studies, the greater persistency in goats kidding between October and January was explained by the thriving spring grass, which occurs after the peak of lactation and maintains milk production. Here, 80% of the data were recorded in the north-west (Loire-Atlantique, Poitou-Charente)

where 84% of the goats did not pasture (Caillat et al., 2016). Therefore, the spring grass did not explain the persistency or the effect on PC3. However, abundant literature is available on the effect of duration of the day, which has a significant effect on goat lactation curves (Chemineau et al., 2007; Russo et al., 2013). These studies showed that long exposure to artificial light (16 h per day) has a big effect on milk yield throughout the lactation stage (up to +33% in France; Delouis and Mirman, 1984). Here, we found this same effect of photoperiod, independent of other environmental factors, on a large data set of goats raised in commercial herds.

A kidding in January seems interesting (high lactation yield, average persistency, and high production around DIM120) but has to be put in relation to the milk price, which fluctuates a lot seasonally, and to justify a kidding out of season.

**Herd Effect.** The influence of herd was studied on the residuals of model 1. The remaining variance explained by the herd effect was relatively significant, as illustrated by the  $R^2$  of 0.397, 0.178, and 0.162 for residuals of  $M_{SPC1}$ ,  $M_{SPC2}$ , and  $M_{SPC3}$ , respectively. The dispersion of herd effects on residuals was large, ranging from  $-243.8$  to  $314.5$  for PC1, from  $-90.43$  to  $101$  for PC2, and from  $-35.82$  to  $43.21$  for PC3. Bouloc (1991) demonstrated that herd effect had the biggest effect, in comparison to the other factors, on milk production and persistency.

When the herd effect was replaced by the estimated herd effect on total-lactation milk yield from the official genetic evaluation (4 classes),  $R^2$  was 0.329 for PC1 and 0 for PC2 and PC3. This showed that the herd effect from the official genetic evaluation well explained the average of all-herd milk production level. However, as expected, the herd effect did not take into account the effect of the herd on persistency and production at 120 DIM.

Lactation length was not studied as a factor influencing lactation curves because of its links with breeder decisions (according to the potential of the animal, the day of insemination, and breeder strategies). León et al. (2012) specified that number of kids affected the production and shape of the lactation curve. A goat with more kids had more milk in the beginning of the lactation and was less persistent. Zumbach et al. (2008), like Fernández et al. (2002), used the “number of born kids” factor in their genetic evaluation model. This factor was not tested here as the information was unavailable. Likewise, we did not have more precise data on the farm regarding, for example, feeding system, system of sales, breeding goal, or climatic conditions, to better understand the large influence of herd on the different SPC values.

## CONCLUSIONS

This study, based on a big recent data set, typologized lactation curve shapes in the current French goat population. It demonstrated a very real diversity of lactation curve shapes in France: 30% of the lactations with a shape very different from the mean. This great diversity of curves is mainly due to the global level of milk production and persistency of the lactation, but not only this. Indeed, we showed that the curve bends in mid-lactation, at 120 DIM, which causes particular lactation curve shapes tied to duration of the photo-period. Furthermore, this trait could be viewed as a complementary indicator of milk persistency. Lactation persistency plays a key role in the discrimination of lactation curve types, highlighting a certain variability. However, none of the environmental and genetic factors tested in our study had a strong effect on this trait. Some well-known factors could not be tested here, which may explain these results. This trait also has a genetic variability that we were unable to factor here. Other approaches will have to be tested to investigate the genetic variability of the lactation curves. The use of random regression models taking into account environmental and genetic effects that vary throughout lactation will have to be explored. The present study, showing the effect of some environmental factors on the lactation curve shape, will help us to model the environmental part of the random regression model. Once the genetic evaluation is carried out, the correlations between lactation shape curve and current or future selected characters (on milk quality, longevity, or health) will have to be evaluated to ensure the relevance of the selection. The effects of the environmental factors could also be used in management advice in the dairy goat husbandry.

## ACKNOWLEDGMENTS

This project receives financial support from Ministry of Agriculture (Compte d'affectation Spécial au Développement Agricole et Rural, CASDAR, Rustic). The first author received financial support from APIS-GENE (Paris, France) and Association nationale de la recherche et de la technologie (ANRT, Paris, France).

## REFERENCES

- Appuhamy, J. A., B. G. Cassell, and J. B. Cole. 2009. Phenotypic and genetic relationships of common health disorders with milk and fat yield persistencies from producer-recorded health data and test-day yields. *J. Dairy Sci.* 92:1785–1795.
- Atashi, H., M. J. Zamiri, and M. Dadpasand. 2013. Association between dry period length and lactation performance, lactation curve, calf birth weight, and dystocia in Holstein dairy cows in Iran. *J. Dairy Sci.* 96:3632–3638.
- Biernacki, C., G. Celeux, and G. Govaert. 2003. Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. *Comput. Stat. Data Anal.* 41:561–575.
- Biernacki, C., G. Celeux, G. Govaert, and F. Langrognet. 2006. Model-based cluster and discriminant analysis with the MIXMOD software. *Comput. Stat. Data Anal.* 51:587–600.
- Boichard, D., N. Bouloc, G. Ricordeau, A. Piacere, and F. Barillet. 1989. Genetic parameters for first lactation dairy traits in the Alpine and Saanen goat breeds. *Genet. Sel. Evol.* 21:205.
- Bouloc, N. 1991. Analyse de la forme de la courbe de lactation. Application à l'étude des modalités d'allègement du contrôle laitier et de prévision précoce de la production dans l'espèce caprine. PhD thesis, Institut national agronomique (INA) Paris Grignon, Paris, France.
- Brito, L. F., F. G. Silva, H. R. Oliveira, N. O. Souza, G. C. Caetano, E. V. Costa, G. R. O. Menezes, A. L. P. Melo, M. T. Rodrigues, and R. A. Torres. 2017. Modelling lactation curves of dairy goats by fitting random regression models using Legendre polynomials or B-splines. *Can. J. Anim. Sci.* 98:73–83.
- Caillat, H., B. Ranger, E. Bruneteau, C. Paraud, and R. Delargarde. 2016. Use of grazing in a dairy goat farm to design sustainable production systems in France. Page 794 in Book of Abstract of the 67th Annual Meeting of the European Federation of Animal 67. Annual Meeting of the European Association for Animal Production (EAAP). Wageningen Academic Publishers, Wageningen (Pays-Bas), Belfast, GBR.
- Caja, G., A. A. K. Salama, and X. Such. 2006. Omitting the dry-off period negatively affects colostrum and milk yield in dairy goats. *J. Dairy Sci.* 89:4220–4228.
- Carta, A., S. Casu, M. G. Usai, and S. Salaris. 2014. Heritability of persistency traits and their genetic correlations with milk yield and udder morphology in dairy sheep. Page 39th ICAR Session, Berlin (Germany). <https://www.icar.org/wp-content/uploads/2016/07/Berlin-2014-Carta-Casu-Usai-Heritability-of-persistency-traits-and-their-genetic-correlations.pdf>.
- Celeux, G., and G. Govaert. 1995. Gaussian parsimonious clustering models. *Pattern Recognit.* 28:781–793.
- Chemineau, P., B. Malpoux, J. P. Brillard, and A. Fostier. 2007. Seasonality of reproduction and production in farm fishes, birds and mammals. *Animal* 1:419–432.
- Cole, J. B., and D. J. Null. 2009. Genetic evaluation of lactation persistency for five breeds of dairy cattle. *J. Dairy Sci.* 92:2248–2258.
- Dai, X., P. Z. Hadjipantelis, H. Ji, H.-G. Mueller, and J.-L. Wang. 2016. Fdapace: Functional Data Analysis and Empirical Dynamics. R Package Version 0.2.5. Accessed Jan. 18, 2017. <https://cran.r-project.org/web/packages/fdapace/index.html>.
- Delouis, C., and B. Mirman. 1984. Influence de la durée quotidienne d'éclairage sur la production laitière de la chèvre. *Compte-Rendu 9e Journ. Rech. Ovine Caprine* 352–360.
- Druet, T., F. Jaffrézic, D. Boichard, and V. Ducrocq. 2003. Modeling lactation curves and estimation of genetic parameters for first lactation test-day records of French Holstein cows. *J. Dairy Sci.* 86:2480–2490.
- Elvira, L., F. Hernandez, P. Cuesta, S. Cano, J.-V. Gonzalez-Martin, and S. Astiz. 2013a. Accurate mathematical models to describe the lactation curve of Lacaune dairy sheep under intensive management. *Animal* 7:1044–1052.
- Elvira, L., F. Hernandez, P. Cuesta, S. Cano, J.-V. Gonzalez-Martin, and S. Astiz. 2013b. Factors affecting the lactation curves of intensively managed sheep based on a clustering approach. *J. Dairy Res.* 80:439–447.
- Fernández, C., A. Sánchez, and C. Garcés. 2002. Modeling the lactation curve for test-day milk yield in Murciano-Granadina goats. *Small Rumin. Res.* 46:29–41.
- Gengler, N. 1996. Persistency of lactation yields: A review. *Interbull Bull.* 12:87–96.
- Gipson, T. A., and M. Grossman. 1990. Lactation curves in dairy goats: A review. *Small Rumin. Res.* 3:383–396.
- Grossman, M., S. M. Hartz, and W. J. Koops. 1999. Persistency of lactation yield: A novel approach. *J. Dairy Sci.* 82:2192–2197.

- ICAR. 2017. ICAR guidelines. Accessed Mar. 8, 2018. <https://www.icar.org/index.php/icar-recording-guidelines/>.
- Kearney, J. 2010. Food consumption trends and drivers. *Philos. Trans. R. Soc. B Biol. Sci.* 365:2793–2807.
- Knight, C. H., and C. J. Wilde. 1988. Milk production in concurrently pregnant and lactating goats mated out of season. *J. Dairy Res.* 55:487–493.
- Langrognet, F., R. Lebre, C. Poli, S. Iovleff, B. Auder, C. Biernacki, G. Celeux, and G. Govaert. 2016. Rmixmod: Supervised, Unsupervised, Semi-Supervised Classification with MIXture MODELing (Interface of MIXMOD Software). R Package Version 2.1.1. Accessed Jan. 18, 2017. <https://cran.r-project.org/web/packages/Rmixmod/index.html>.
- Lebre, R., S. Iovleff, F. Langrognet, C. Biernacki, G. Celeux, and G. Govaert. 2015. Rmixmod: The r package of the model-based unsupervised, supervised and semi-supervised classification mixmod library. *J. Stat. Softw.* 67:241–270.
- Leclerc, H., D. Duclos, A. Barbat, T. Druet, and V. Ducrocq. 2008. Environmental effects on lactation curves included in a test-day model genetic evaluation. *Animal* 2:344–353.
- Leitner, G., U. Merin, and N. Silanikove. 2004. Changes in milk composition as affected by subclinical mastitis in goats. *J. Dairy Sci.* 87:1719–1726.
- Lenth, R. V. 2016. Least-Squares Means: The R Package lsmeans. *J. Stat. Softw.* 69:1–33.
- León, J. M., N. P. P. Macciotta, L. T. Gama, C. Barba, and J. V. Delgado. 2012. Characterization of the lactation curve in Murciano-Granadina dairy goats. *Small Rumin. Res.* 107:76–84.
- Macciotta, N. P. P., D. Vicario, and A. Cappio-Borlino. 2006. Use of multivariate analysis to extract latent variables related to level of production and lactation persistency in dairy cattle. *J. Dairy Sci.* 89:3188–3194.
- Menéndez-Buxadera, A., A. Molina, F. Arrebola, M. J. Gil, and J. M. Serradilla. 2010. Random regression analysis of milk yield and milk composition in the first and second lactations of Murciano-Granadina goats. *J. Dairy Sci.* 93:2718–2726.
- Montaldo, H., A. Almanza, and A. Juárez. 1997. Genetic group, age and season effects on lactation curve shape in goats. *Small Rumin. Res.* 24:195–202.
- Mucha, S., R. Mrode, M. Coffey, and J. Conington. 2014. Estimation of genetic parameters for milk yield across lactations in mixed-breed dairy goats. *J. Dairy Sci.* 97:2455–2461.
- Nayeri, S., M. Sargolzaei, M. K. Abo-Ismael, S. Miller, F. Schenkel, S. S. Moore, and P. Stothard. 2017. Genome-wide association study for lactation persistency, female fertility, longevity, and lifetime profit index traits in Holstein dairy cattle. *J. Dairy Sci.* 100:1246–1258.
- Pollott, G. E. 2000. A biological approach to lactation curve analysis for milk yield. *J. Dairy Sci.* 83:2448–2458.
- Pulina, G., M. J. Milán, M. P. Lavín, A. Theodoridis, E. Morin, J. Capote, D. L. Thomas, A. H. D. Francesconi, and G. Caja. 2018. Invited review: Current production trends, farm structures, and economics of the dairy sheep and goat sectors. *J. Dairy Sci.* 101:6715–6729.
- Rupp, R., V. Clément, A. Piacere, C. Robert-Granié, and E. Manfredi. 2011. Genetic parameters for milk somatic cell score and relationship with production and udder type traits in dairy Alpine and Saanen primiparous goats. *J. Dairy Sci.* 94:3629–3634.
- Russo, V. M., A. W. N. Cameron, F. R. Dunshea, A. J. Tilbrook, and B. J. Leury. 2013. Artificially extending photoperiod improves milk yield in dairy goats and is most effective in late lactation. *Small Rumin. Res.* 113:179–186.
- Salama, A. A., G. Caja, X. Such, R. Casals, and E. Albanell. 2005. Effect of pregnancy and extended lactation on milk production in dairy goats milked once daily. *J. Dairy Sci.* 88:3894–3904.
- Sölkner, J., and W. Fuchs. 1987. A comparison of different measures of persistency with special respect to variation of test-day milk yields. *Livest. Prod. Sci.* 16:305–319.
- van der Werf, J. H. J., M. E. Goddard, and K. Meyer. 1998. The use of covariance functions and random regressions for genetic evaluation of milk production based on test day records. *J. Dairy Sci.* 81:3300–3308.
- Van Hekken, D. L., M. H. Tunick, and Y. W. Park. 2005. Effect of frozen storage on the proteolytic and rheological properties of soft caprine milk cheese. *J. Dairy Sci.* 88:1966–1972. [https://doi.org/10.3168/jds.S0022-0302\(05\)72872-1](https://doi.org/10.3168/jds.S0022-0302(05)72872-1).
- Yao, F., H.-G. Müller, and J.-L. Wang. 2005. Functional data analysis for sparse longitudinal data. *J. Am. Stat. Assoc.* 100:577–590.
- Zumbach, B., S. Tsuruta, I. Misztal, and K. J. Peters. 2008. Use of a test day model for dairy goat milk yield across lactations in Germany. *J. Anim. Breed. Genet.* 125:160–167.

## **II. Analyses et résultats complémentaires**

### **1. Analyse des profils moyens des courbes de lactation en fonction du nombre de clusters retenus pour la classification**

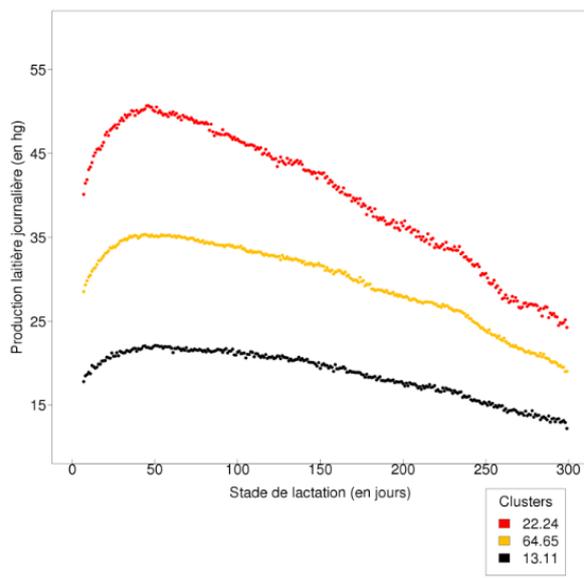
Dans l'article précédent nous avons privilégié une étude avec 5 clusters permettant au dernier cluster de représenter un effectif de courbes suffisamment important (supérieur à 4%) afin que les résultats puissent être représentatifs de l'ensemble de la population.

Pour arriver à ce nombre de clusters, nous avons réalisé de manière exploratoire une classification avec plusieurs nombres de clusters dont les résultats sont reportés à la Figure 12.

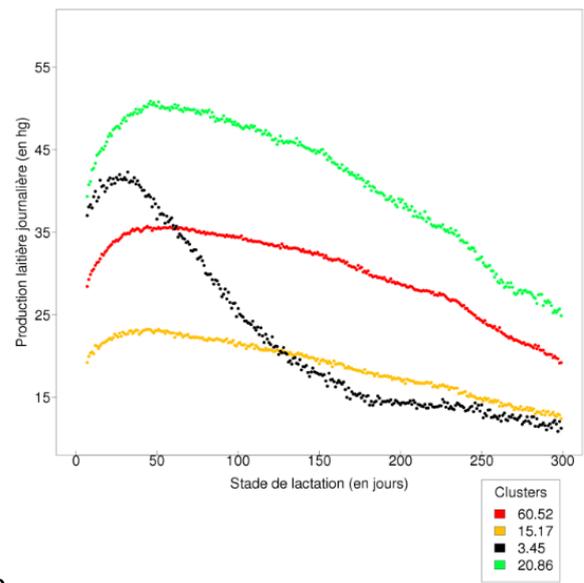
Lorsque seulement 3 clusters sont retenus, bien que présentant des niveaux de production différents, les formes de courbes de lactation étaient similaires (Figure 12.a). A partir de 4 clusters, nous avons observé qu'un cluster en particulier regroupait des courbes présentant un pic marqué et une très mauvaise persistance (courbe en noir Figure 12.b). Ce type de courbes était présent dans toutes les situations avec plus de 3 clusters. A partir de 5 clusters, un profil de courbes avec une grande quantité de lait et une grande persistance (8.69% des courbes individuelles) apparaît, ainsi qu'un profil avec une persistance importante à partir du pic de lactation jusqu'à 150 jours puis une diminution rapide de la production (16.09% des courbes individuelles) (en noir et vert respectivement Figure 12.c). Avec 6 clusters, le cluster qui comportait le moins d'individus représentait 2.55% des données (en bleu clair Figure 12.d). La courbe moyenne de ce cluster présentait une plus grande variabilité entre deux stade de lactation successifs que les autres clusters représentant une plus grande quantité de données. Ce cluster avait un profil moyen de courbe de lactation avec une production élevée et un pic à plus de 5.5kg. Avec 7 clusters, un profil de courbes à haut niveau de production et forte persistance apparaissait (en rouge Figure 12.e). Enfin avec 8 clusters, l'étude mettait en évidence un dernier cluster représentant moins de 2% des courbes de lactation, et dont le profil moyen de courbe comportait un pic très marqué et une production qui diminuait rapidement (en violet Figure 12.f).

En deçà de 4 clusters, les profils de courbes ne se distinguaient que par le niveau de production.

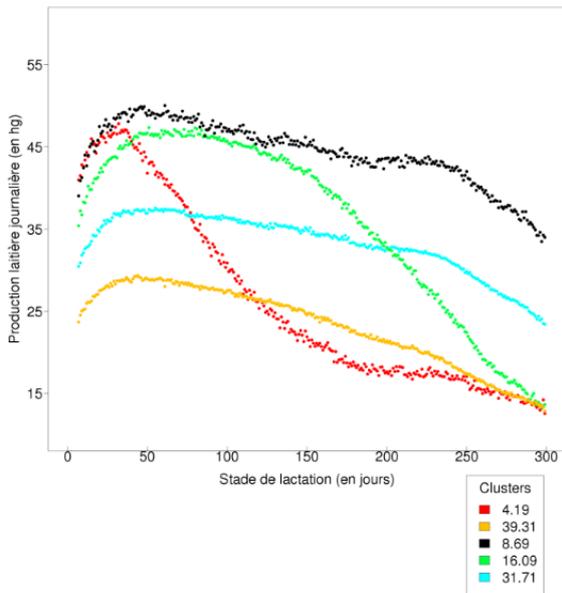
Les très faibles effectifs représentés par les clusters 6 et suivants ainsi que les formes de courbes moyennes plus erratiques de ces clusters nous ont conduits à privilégier l'étude avec 5 clusters comme étant le meilleur compromis entre distinction de différents profils de courbes et représentativité de ces courbes dans la population caprine.



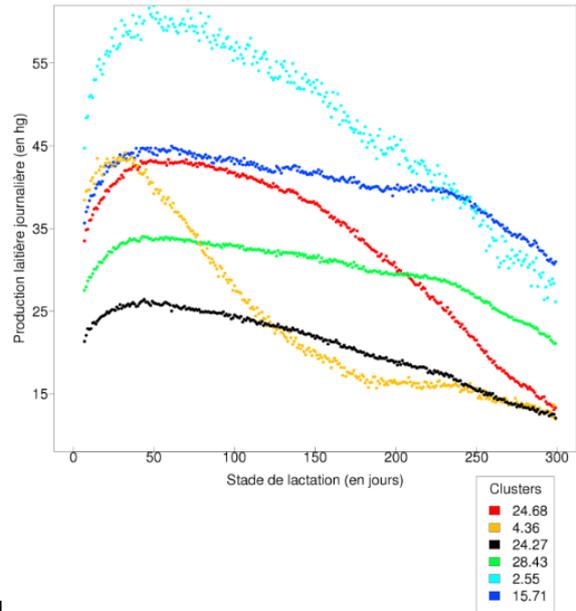
a



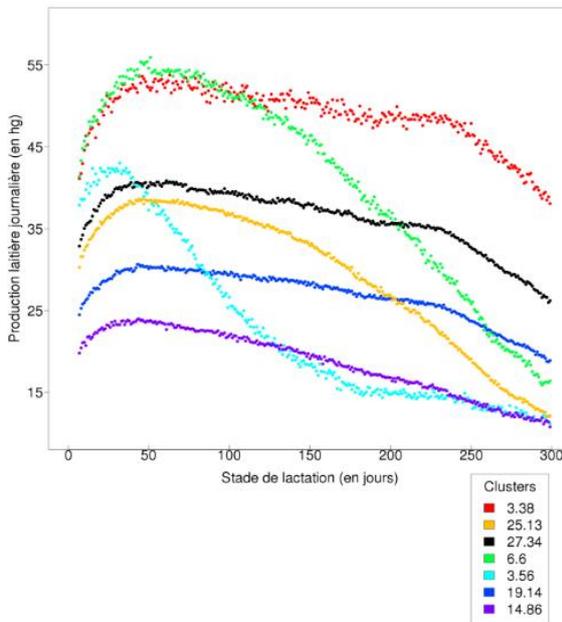
b



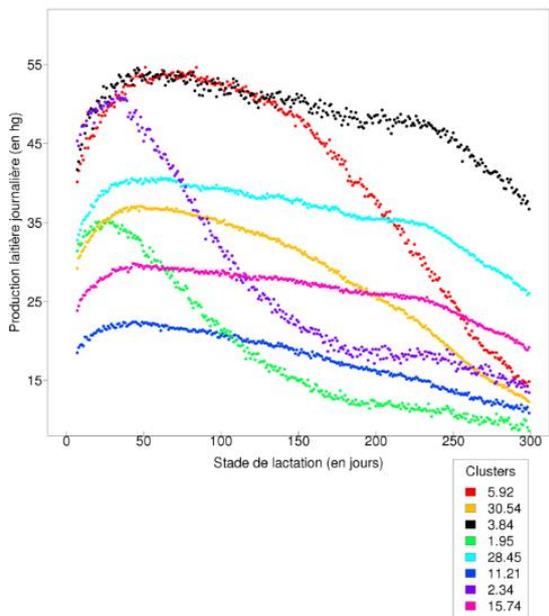
c



d



e



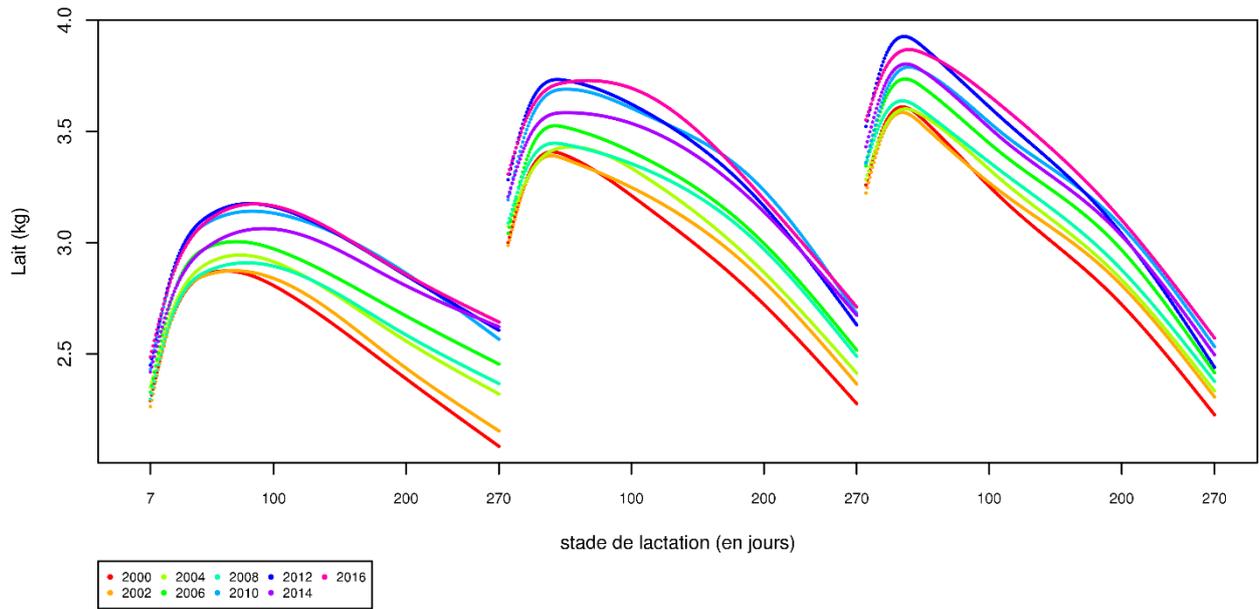
f

Figure 12: Courbes de lactation phénotypiques des différents clusters selon le nombre de clusters : a) 3 clusters, b) 4 clusters, c) 5 clusters, d) 6 clusters, e) 7 clusters, f) 8 clusters. Le pourcentage de lactations représenté par chaque cluster est précisé en légende à côté de la couleur du cluster.

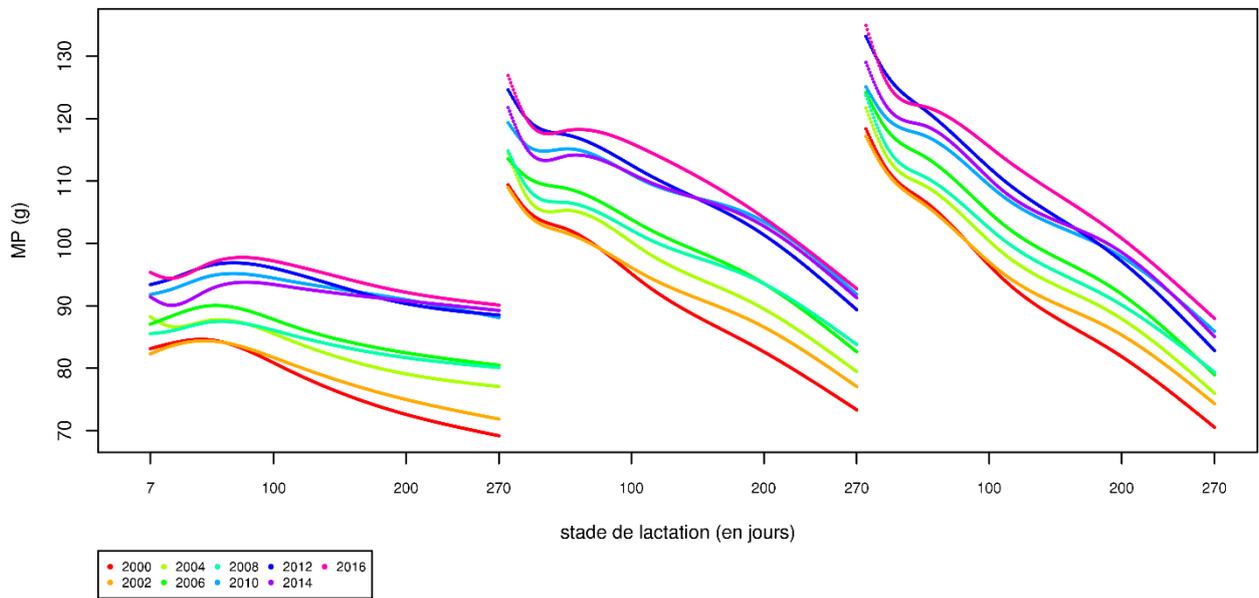
## 2. Evolution des formes de courbes phénotypiques de lactation au cours des années

L'étude précédente a été réalisée sur un jeu de données contenant peu de campagnes de lactation (de 2009 à 2012) ne nous permettant pas d'étudier l'évolution des formes de courbes de lactation au cours des ans. Les courbes phénotypiques annuelles pour tous les caractères pour les 3 premières lactations ont été tracées de l'année 2000 à 2016. Les Figure 13 à Figure 17 présentent ces courbes tous les deux ans pour les caractères production de lait, MP, TP, CCS, rapport TB/TP en race Saanen. Le jeu de données utilisé ici était différent du précédent. Il contenait tous les contrôles des chèvres contrôlées entre 2000 et 2016 de race Alpine et Saanen. Seules les performances des 3 premières lactations ont été conservées ce qui représentait un total de 7 908 192 contrôles élémentaires en Saanen pour 1 306 704 lactations et 10 483 205 contrôles élémentaires en Alpine pour 1 698 032 lactations.

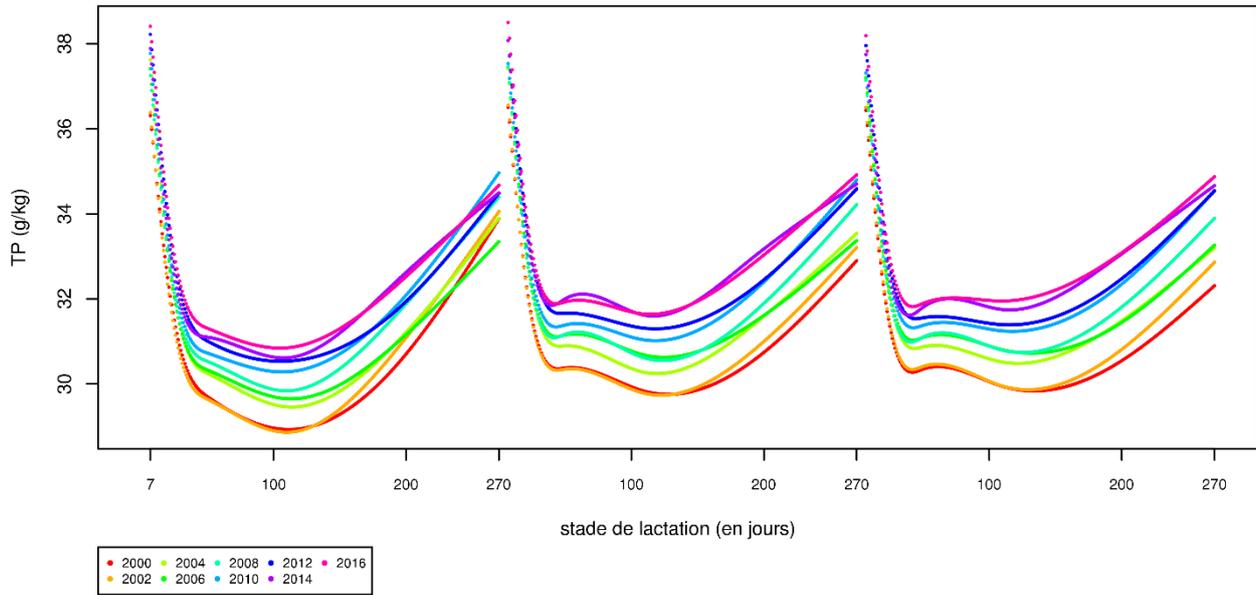
Nous constatons globalement une augmentation des quantités produites aux fils des ans sans réelles différences de formes de courbes (de 3.3 kg à 3.8 kg pour la production laitière, de 108 à 118 g pour la MP, au stade de lactation 40 en deuxième lactation) mis à part en Saanen, en primipares où la persistance phénotypique des quantités a été augmentée de 2000 à 2008 et ne semble pas progresser à partir de 2010 (de 2 kg à 2.7 kg pour le stade de lactation 270). Pour le TB, l'augmentation est plutôt régulière au fil des ans, elle est localisée en particulier en milieu de lactation autour de 150 jours, et est plus importante en race Saanen. Pour le TP, l'augmentation est plutôt régulière au fil des ans, le niveau a augmenté tout au long de la lactation, et cette augmentation est plus marquée en race Saanen (de 30 à 32 pour le TP au stade de lactation 40 en deuxième lactation). Les quantités de SCS ont, elles aussi, augmenté de manière régulière au fil des ans, et de façon plus importante en race Saanen (de 4.3 à 4.7 pour les SCS au stade de lactation 40 en deuxième lactation). Le ratio TB/TP a progressé jusqu'en 2008 et régressé les années suivantes tout le long de la lactation (de 1.17 à 1.25 pour les SCS et pas de différence pour le ratio TB/TP au stade de lactation 40 en deuxième lactation). Bouloc (1991) a étudié les courbes phénotypiques des chèvres ayant mis bas en 1987. Les productions étaient moindres en général par rapport à notre étude qui comportait des lactations réalisées entre 2000 et 2016 (au pic de lactation : 2.7 kg de lait en primipare, 3.5kg en multipares, 75g de MP en primipares et 90 g de MP en multipare et un TP de 26g/kg de TP en primipares et en multipares en 1987 contre au pic de lactation : 3.2 kg de lait en primipare, 3.7kg en multipares, 98g de MP en primipares et 120 g de MP en multipares et un TP de 32g/kg de TP en primipares et en multipares en 2016). Cette étude nous confirme qu'au fil des ans la production au pic est de plus en plus élevée créant possiblement des problèmes métaboliques à la chèvre si cette espèce réagit comme les bovins (Muir et al., 2004).



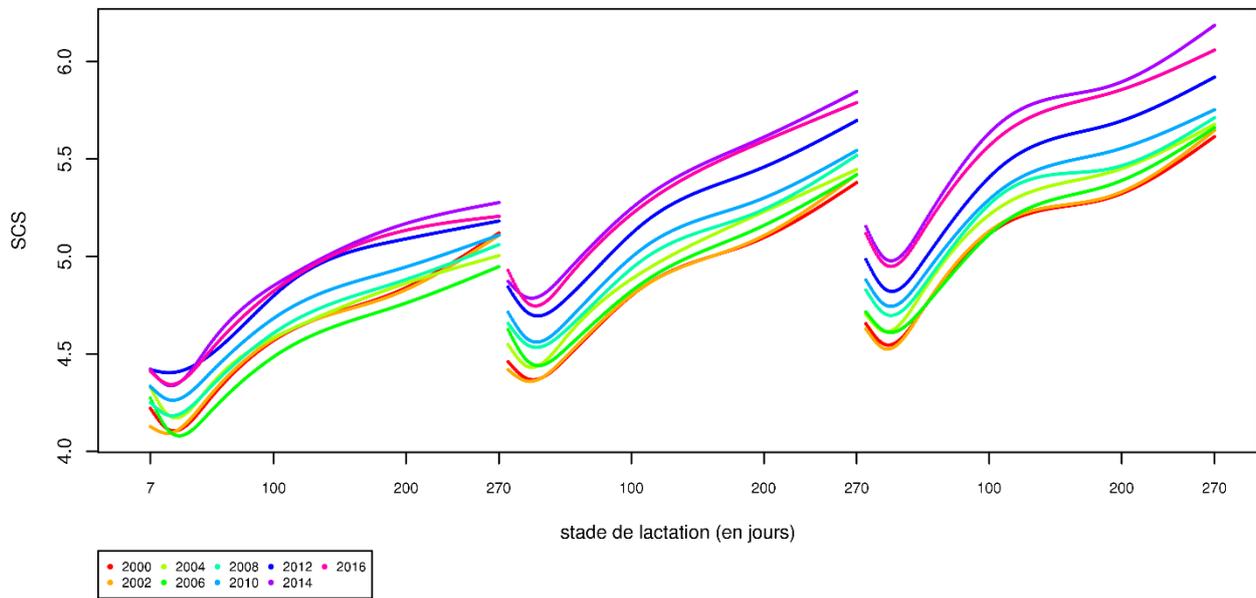
**Figure 13:** Evolution des courbes phénotypiques selon le rang de lactation et la campagne de mise bas pour la quantité de lait en Saanen



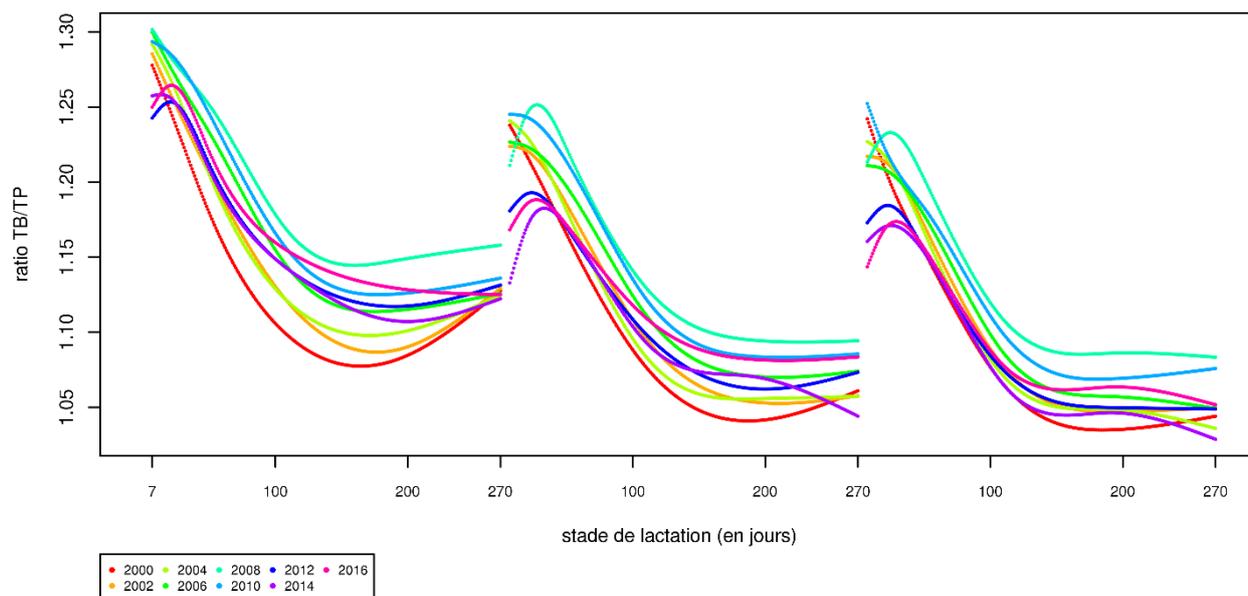
**Figure 14:** Evolution des courbes phénotypiques selon le rang de lactation et la campagne de mise bas pour la MP en Saanen



**Figure 15: Evolution des courbes phénotypiques selon le rang de lactation et la campagne de mise bas pour le TP en Saanen**



**Figure 16: Evolution des courbes phénotypiques selon le rang de lactation et la campagne de mise bas pour le comptage cellulaire en Saanen**

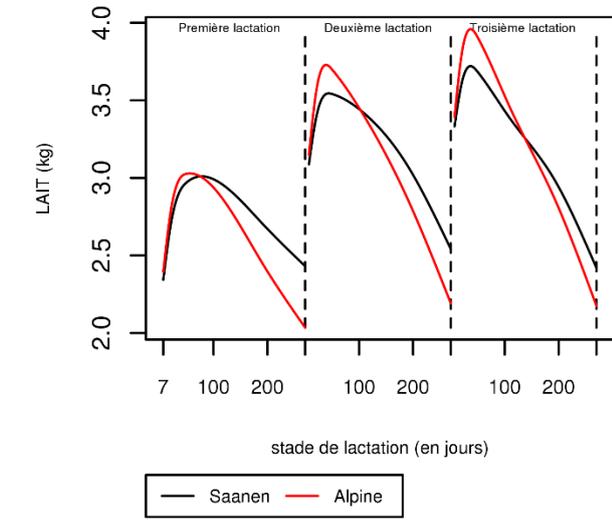


**Figure 17: Evolution des courbes phénotypiques selon le rang de lactation et la campagne de mise bas pour le ratio TB/TP en Saanen**

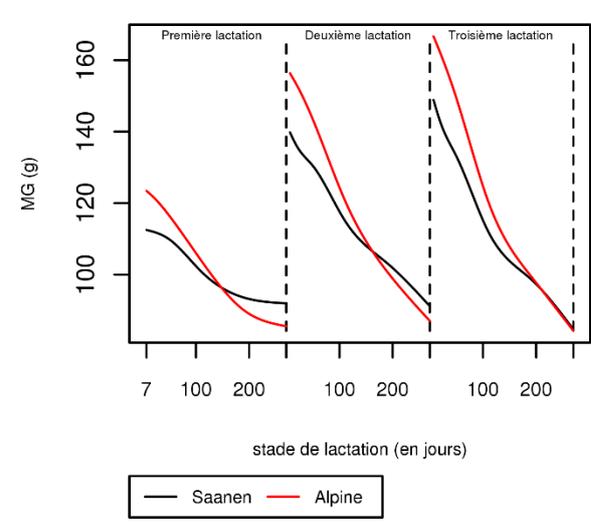
### 3. Comparaison des courbes de production au cours de la lactation pour les races Alpine et Saanen

Le même jeu de données que celui utilisé au paragraphe 2 a été utilisé pour étudier les différences de courbes entre les races Saanen et Alpine. La Figure 18 présente les évolutions moyennes phénotypiques des campagnes de mise bas de 2000 à 2016 des caractères de production et des comptages de cellules somatiques au cours des 3 premières lactations en race Saanen (en noir) et Alpine (en rouge). Pour les productions de lait et de matières (Figure 18.a, Figure 18.b, et Figure 18.c), les chèvres de race Saanen sont plus persistantes en accord avec les résultats de l'article 1 et ont un pic de lactation moins marqué avec une production au pic plus faible. Pour les taux (Figure 18.d et Figure 18.e), les Alpines produisent un lait plus riche. Passé le pic de lactation les courbes des taux ont sensiblement le même profil en Alpine et Saanen pour le TB, mais le TP des Saanen en fin de lactation ne progresse pas aussi vite que celui des Alpines. Pour les SCS, les chèvres de race Saanen produisent plus de SCS que les chèvres Alpines et en primipare, les chèvres Alpine ont une production en début de lactation plus faible (4 au lieu de 4.3) mais terminent au même SCS à 5. Pour le ratio TB/TP, il est plus élevé pour les chèvres de race Alpine et le profil des courbes est comparable entre les deux races, bien que le minimum soit atteint plus tôt en race Saanen. Bouloc (1991) a étudié les courbes phénotypiques des chèvres ayant mis bas en 1987. Les courbes de productions étaient différentes car les chèvres Alpine et Saanen avaient des pics de production laitière équivalents et les chèvres de race Saanen avaient en revanche une persistance légèrement supérieure aux chèvres de race Alpine comme dans notre étude. Pour le TP, les différences en fin de lactation étaient moins marquées par rapport à notre jeu de données

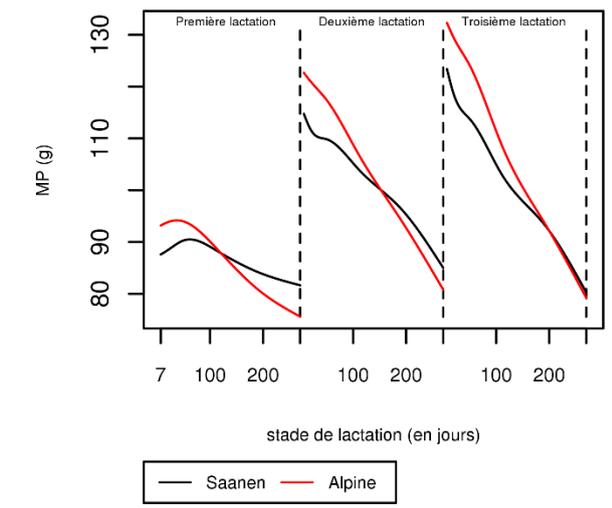
plus récent. La comparaison de nos résultats à ceux de Bouloc (1991), tend à montrer que les courbes de lactation n'ont pas évolué de la même manière selon la race Alpine et le Saanen.



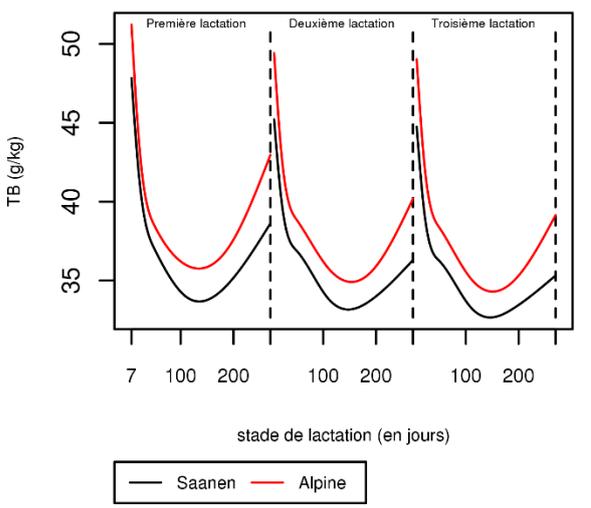
a



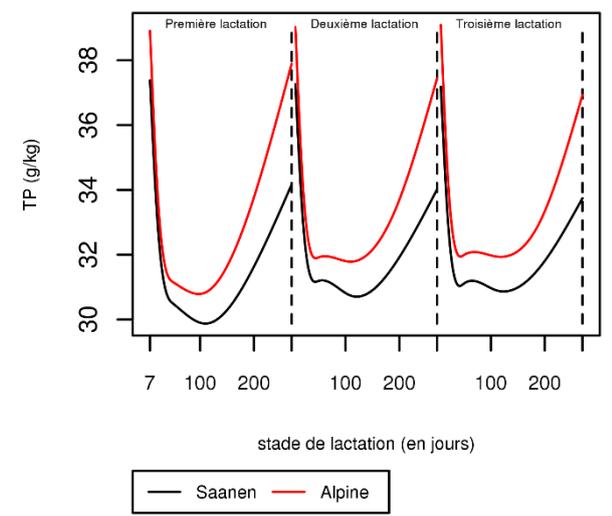
b



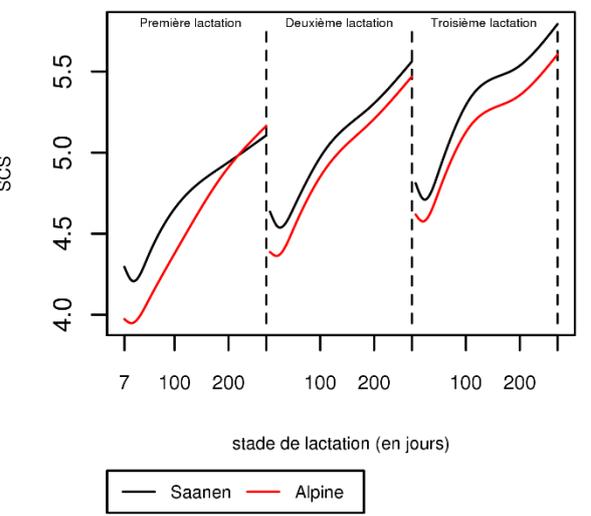
c



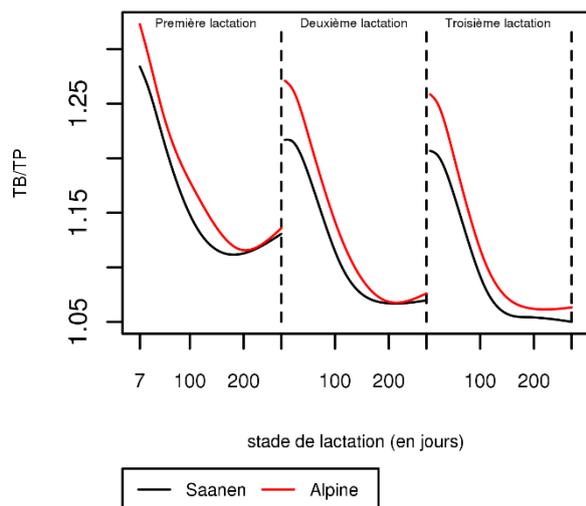
d



e



f



σ

**Figure 18 : Courbes d'évolution phénotypique moyenne en race Alpine et Saanen des caractères : a) lait, b) MG, c) MP, d) TB, e) TP, f) SCS, g) ratio TB/TP**

En conclusion, nous avons pu identifier 5 types de courbe de lactation avec des profils différents. L'analyse en composantes principales a permis de décomposer la forme de la courbe de lactation en 3 composantes : la production à la lactation, la persistance laitière et la production en milieu de lactation. Nous avons montré que la valeur génétique pour la production laitière d'un animal (calculée sur la base d'une quantité de lait totale) est le facteur le plus lié au niveau de production lors de la lactation. Alors que le mois de mise bas est le facteur environnemental le plus lié à la persistance laitière et à la production en milieu de lactation. L'analyse des courbes phénotypiques selon les années nous a permis d'apprécier l'augmentation de la production non seulement au niveau de la production totale mais aussi au niveau des formes de courbes. La comparaison des formes des courbes phénotypiques par race nous a aussi permis de voir que ces deux races ne produisent pas de la même façon au niveau des quantités mais aussi au niveau de la dynamique de production au cours de la lactation. Cette diversité de courbes est intéressante et à présent nous aimerions savoir quelle est la part de cette diversité qui est due à la génétique de l'animal. Les effets environnementaux liés à la forme de la courbe de lactation (troupeau, âge à la mise bas, mois de mise bas, durée de tarissement, stade de gestation) nous aideront à modéliser la partie liée à l'environnement de l'animal dans le futur modèle d'évaluation génétique.



## **Chapitre 2 : Estimation des paramètres génétiques pour les caractères de production laitière des chèvres en première lactation à l'aide d'un modèle de régression aléatoire**

Comme indiqué dans l'introduction générale, les modèles de régression aléatoire sont largement utilisés pour les évaluations génétiques des caractères de production laitière car ils modélisent l'évolution au cours du temps des effets génétiques et d'environnement. Ils permettent également de définir un effet cheptel en fonction du jour de chaque mesure.

L'objectif de ce chapitre est de mettre au point sur la base de données de contrôle élémentaires de la première lactation des chèvres, un modèle de régression aléatoire pour les caractères de production laitière permettant un bon ajustement aux données tout en étant simple à mettre en œuvre pour une évaluation génétique future, donc peu coûteux en temps de calcul, et dont les index sont facilement interprétables par le sélectionneur. L'estimation des paramètres génétiques en première lactation des chèvres a permis de tester et comparer différents modèles. Cette étude a été inspirée des travaux conduits en bovins laitiers en France (Druet et al., 2003).

La première partie de ce chapitre est constituée de l'article publié dans « Genetic Selection and Evolution » en 2019. Les données consistaient en l'enregistrement de 193 226 et 205 841 contrôles élémentaires laitiers de 30 186 et 32 256 premières lactations de chèvres Saanen et Alpine. Les productions de lait, de matière grasse, et de matière protéique, ainsi que les taux butyreux et taux protéique ont été analysés. Des splines ont été utilisées pour modéliser les facteurs environnementaux. Les effets génétiques et de l'environnement permanent ont été modélisés eux par les mêmes polynômes de Legendre. La qualité de l'ajustement et les paramètres génétiques dérivés des polynômes d'ordre 0 à 4 ont été testés. Les résultats ont également été comparés à ceux d'un modèle à la lactation avec un rendement laitier total calculé sur 250 jours (proche du modèle utilisé actuellement en routine pour les évaluations génétiques), ainsi qu'à ceux d'un modèle multi-caractères qui considère la performance sur six périodes de la lactation comme des caractères différents mais génétiquement corrélés. Les paramètres génétiques étaient cohérents entre les modèles. Les modèles avec des polynômes de Legendre de quatrième ordre ont permis d'obtenir le meilleur ajustement aux données. Afin de réduire la complexité, le temps de calcul et de faciliter l'interprétation et l'utilisation des valeurs génétiques, une réduction de rang des matrices de variances-covariances génétique et d'environnement permanent a été réalisée à l'aide d'une décomposition en valeurs propres. Avec une réduction au rang 2, les deux premières composantes principales de la matrice génétique résumaient correctement la variabilité génétique du niveau et de la persistance de la production laitière, avec une corrélation génétique proche de 0.

La deuxième partie de ce chapitre présente des résultats complémentaires sur l'évaluation de la précision du phénotype utilisé actuellement en évaluation génétique par le modèle à la lactation, ainsi que sur 2 postulats du modèle RRM finalement retenu. Le premier de ces postulats concerne l'utilisation de la même réduction des

matrices de variances-covariances pour l'effet génétique et l'effet d'environnement permanent, la deuxième porte sur la définition assez simple d'un effet troupeau-jour de contrôle. Nous avons voulu fournir des éléments d'appréciation de ces 2 postulats qui ont un impact majeur pour l'utilisation ultérieure de ce modèle en évaluation génétique.

## **I. Paramètres génétiques des caractères de production laitière des chèvres Alpine et Saanen en utilisant un modèle de régression aléatoire (Article 2)**

RESEARCH ARTICLE

Open Access



# Genetic parameters for first lactation dairy traits in the Alpine and Saanen goat breeds using a random regression test-day model

Mathieu Arnal<sup>1,2\*</sup> , H el ene Larroque<sup>1</sup>, H el ene Leclerc<sup>3</sup>, Vincent Ducrocq<sup>4</sup> and Christ ele Robert-Grani e<sup>1</sup>

## Abstract

**Background:** Random regression models (RRM) are widely used to analyze longitudinal data in genetic evaluation systems because they can better account for time-course changes in environmental effects and additive genetic values of animals by fitting the test-day (TD) specific effects. Our objective was to implement a random regression model for the evaluation of dairy production traits in French goats.

**Results:** The data consisted of milk TD records from 30,186 and 32,256 first lactations of Saanen and Alpine goats. Milk yield, fat yield, protein yield, fat content and protein content were considered. Splines were used to model the environmental factors. The genetic and permanent environmental effects were modeled by the same Legendre polynomials. The goodness-of-fit and the genetic parameters derived from functions of the polynomials of orders 0 to 4 were tested. Results were also compared to those from a lactation model with total milk yield calculated over 250 days and to those of a multiple-trait model that considers performance in six periods throughout lactation as different traits. Genetic parameters were consistent between models. Models with fourth-order Legendre polynomials led to the best fit of the data. In order to reduce complexity, computing time, and interpretation, a rank reduction of the variance covariance matrix was performed using eigenvalue decomposition. With a reduction to rank 2, the first two principal components correctly summarized the genetic variability of milk yield level and persistency, with a correlation close to 0 between them.

**Conclusions:** A random regression model was implemented in France to evaluate and select goats for yield traits and persistency, which are independent i.e. no genetic correlation between them, in first lactation.

## Background

Random regression test-day (TD) models (RRM) are widely used in genetic evaluations of TD milk production in dairy cows but also in other species such as dairy goats [1–6]. RRM increase the accuracy of breeding value predictions and fit the variability of environmental effects throughout lactation more accurately [7].

Besides describing the variability of genetic parameters along the lactation curve, RRM serve to predict estimated breeding values (EBV) for lactation persistency, based on the variation of EBV throughout lactation. A persistent

animal is defined as producing on average less milk at the beginning but more at the end of the lactation period than animals with a similar overall production [8]. Lactation persistency is of interest for dairy producers because the shape of the lactation curve can affect an animal's nutritional needs, and consequently its health, as well as the distribution of the farm's milk output during the year [9, 10]. In a previous study, Arnal et al. [11] demonstrated the phenotypic variability of the shape of the dairy goat lactation curves in France, and showed that the main environmental factors influencing curve shapes are breed, kidding month, age at kidding, gestation stage, and length of the dry period.

Various functions have been used to model fixed, genetic and permanent environment effects on TD records, including the Wilmink function [12], Legendre polynomials [13, 14], and splines [15, 16]. Given that

\*Correspondence: Mathieu.arnal@idele.fr

<sup>1</sup> GenPhySE, Universit e de Toulouse, INRA, ENVT, Toulouse INP, 31326 Castanet-Tolosan, France

Full list of author information is available at the end of the article



RRM are computationally expensive, Legendre polynomial functions have the computational advantage of reducing the correlations between estimated regression coefficients, which impact the convergence [7] of the iterative algorithms used for variance component estimation and genetic evaluation. They are also sufficiently flexible to fit differently shaped curves.

With eigen decomposition, it becomes possible to reduce the rank of the resulting variance-covariance matrix by ignoring the contribution of the smallest eigenvalues and eigenvectors [14, 17]. This decreases computing time by reducing the number of genetic and permanent environment regression coefficients that are estimated.

To date, the implementation of RRM in dairy goats has not been studied in France. In this paper, we estimate genetic parameters for milk yield, fat yield, protein yield, fat content, and protein content using RRM with Legendre polynomial functions of different orders, with or without rank reductions, to obtain TD EBV for French dairy goats in first lactation.

## Methods

### Data

The data consisted of 193,226 milk TD records from 30,186 first-lactation Saanen goats (234 herds) and 205,841 milk TD records from 32,256 first-lactation Alpine goats (198 herds) from northwestern France, collected between 1995 and 2015. The pedigree consisted of 66,716 and 67,159 Saanen and Alpine goats, respectively. Each lactation included at least four TD between the 7th and 270th day in milk (DIM). Lactation had to last between 180 and 350 days. Goats were milked twice a day and their records were summed to obtain their daily production. More than four animals per herd × test date combination were required. The sires of these goats were artificial-insemination bucks with at least 20 progeny in the dataset, with 379 and 324 sires from the Alpine and Saanen goat breeds, respectively. The dams of the goats had to be known. The traits analyzed were milk, fat and protein yields and fat and protein contents.

### Random regression models (RRM)

Five RRM were tested separately for each breed and for all five traits and all had the form:

$$\begin{aligned}
 y_{ijkldgn} = & HTD_{ti} + A_{ijk} + M_{ijl} + \sum_{r=1}^6 \theta_{tkr} N_{(r,d)} \\
 & + \sum_{r=1}^6 \tau_{tlr} N_{(r,d)} + \sum_{s=1}^4 \gamma_{ts} M_{(s,g)} \\
 & + \sum_{o=0}^q a_{tno} \varphi_{(o,d)} + \sum_{o=0}^q p_{tno} \varphi_{(o,d)} + e_{ijkldgn},
 \end{aligned}$$

where  $y_{ijkldgn}$  is the observation for trait  $t$  (milk yield, fat yield, protein yield, fat content, protein content) on day in milk (DIM)  $d$  (from 7 to 270 day) of goat  $n$  that was pregnant for the past  $g$  days ( $g = 0$  if not pregnant), was in production year  $j$  (from 1995 to 2015), belonged to kidding age class  $k$  (7 classes [in months]: 9–11, 12, 13, 14, 15, 16, 17 or more), to kidding period class  $l$  (8 classes: January, February, March, April–May–June, July–August–September, October, November, December), and to herd × test-date class  $i$ ;  $HTD_{ti}$  is a herd-test-date fixed effect;  $A_{ijk}$  is the fixed effect of year-age at kidding;  $M_{ijl}$  is the fixed effect of year-kidding period;  $\theta_{tkr}$ ,  $\tau_{tlr}$  and  $\gamma_{ts}$  are the fixed regression coefficients for age at kidding, kidding period and gestation stage, respectively;  $N_{(r,d)}$  is covariate  $r$  at time  $d$  of a cubic natural spline function with six knots at  $d = 7, 20, 50, 110, 190, 270$ ;  $M_{(s,g)}$  is covariate  $s$  at time  $d$  of a cubic natural spline function with four knots at  $d = 31, 53, 76, 100$  (at  $g$  between 0 and 30 day, the coefficients were assumed to be equal to 0, implying that there was no effect of gestation on production; due to scarcity of records, the effect of gestation was considered to be constant after  $g = 100$ ).  $a_{tno}$  is the set of random additive genetic regression coefficients that follow a normal distribution ( $Var(a_{to}) \sim N(0, \sigma_{a_{to}}^2)$ ) for the  $o$ th term of a Legendre polynomial of order  $q$ ,  $p_{tno}$  is the set of random permanent environmental regression coefficients that follow a normal distribution ( $Var(p_{to}) \sim N(0, \sigma_{p_{to}}^2)$ ) for the  $o$ th term of a Legendre polynomial of order  $q$ ,  $\varphi_{(o,d)}$  is the value of the  $o$ th term of the Legendre polynomial at time  $d$  and  $e_{ijkldgn}$  is the normally-distributed residual term. The model assumed a heterogeneous residual variance along lactation that was considered to be constant within nine classes of DIM (7–36, 37–66, 67–96, 97–126, 127–156, 157–186, 187–216, 217–246, 247–270), without between-class correlations.

Legendre polynomials of order 0 to 4 (leg0 to leg4) were tested assuming the same order for both the random genetic and environmental effects. The EBV of a goat for a complete lactation milk yield was obtained by summing the EBV for each DIM (called SUM\_leg $q$ ) [18]. Given that the RRM can give an estimation of the shape of the lactation curve for each goat, EBV for persistency (denoted PERS\_leg $q$ ) was computed as the cumulative deviation in genetic contribution to yield from the DIM 40 to DIM 240 relative to an average animal having the same yield at DIM 40 [18].

The next step was to test various rank reductions of the genetic variance-covariance matrices based on an eigen decomposition of these matrices. To keep a consistent definition of persistency, the second eigenvector was set, by multiplying by  $-1$ , to be systematically negative at the beginning of lactation, and then positive. For a reduction to

rank  $z$ , the first  $z$  eigenvectors obtained with the complete model were multiplied by the terms of Legendre polynomial as in Druet et al. [14] to obtain the  $z$  first eigenfunctions. These eigenfunctions were used to model the genetic and permanent environment effects. The Legendre polynomial function of order  $q$  with a covariance matrix reduced (RRM\*) to rank  $z$  is written  $\text{leg}qRz$ . The reduced models were of the form:

$$y_{ijklgdn} = HTD_{ti} + A_{tjk} + M_{tjl} + \sum_{r=1}^6 \theta_{tkr} N_{(r,d)} + \sum_{r=1}^6 \tau_{tjr} N_{(r,d)} + \sum_{s=1}^4 \gamma_{ts} M_{(s,g)} + \sum_{o=1}^z b_{tno} \chi_{(o,d)} + \sum_{o=1}^z c_{tno} \chi_{(o,d)} + e_{ijklgdn}$$

where  $b_{tno}$  is the random additive genetic regression coefficient that follows a normal distribution ( $\text{Var}(b_{tno}) \sim N(0, \sigma_{b_{tno}}^2)$ ) for the  $o$ th eigenfunction of the genetic (co)variances matrix obtained with a  $q$ th-order Legendre polynomial model reduced to rank  $z$ ,  $c_{tno}$  is the random permanent environmental regression coefficient that follows a normal distribution ( $\text{Var}(c_{tno}) \sim N(0, \sigma_{c_{tno}}^2)$ ) for the  $o$ th eigenfunction of the genetic (co)variances matrix obtained with a  $q$ th-order Legendre polynomial model reduced to rank  $z$ ,  $\chi_{(o,d)}$  is the value of the  $o$ th eigenfunction of the genetic (co)variances matrix obtained with a  $q$ th-order Legendre polynomial model reduced to rank  $z$  ( $z = \{2, 3\}$ ) at DIM  $d$ .

The corresponding EBV for whole lactation performance of goat  $n$  from the RRM was obtained by summing its EBV throughout the DIM period (from DIM 7 to DIM 270) (SUM\_legxRz).

**Multiple-trait model (MT)**

The evolution of the estimated heritabilities throughout the DIM period and the genetic correlations between DIM that were obtained with RRM were compared to those obtained with a MT model, in which lactation was subdivided into six periods (1 = [7, 45], 2 = [46, 90], 3 = [91, 135], 4 = [136, 180], 5 = [181, 225], 6 = [226, 270]). TD yields and contents for each period  $r$  ( $r=1$  to 6) were considered as different traits. The frequency of TD recording (on average every 28 or 35 days) implied that only one TD record was performed normally per period but if two TD were carried out, the measures were averaged. The MT model was:

$$y_{ijklnro} = \pi_{tir} + \delta_{tjr} + \gamma_{tkr} + HY_{tjr} + a_{tnr} + e_{ijklnro}$$

where  $y_{ijklnro}$  is the observation for trait  $t$  (milk yield, fat yield, protein yield, fat content, protein content) of goat  $n$  for period  $r$  ( $r=1$  to 6), in subclass  $i$  for  $\pi$ , the age at

kidding effect, subclass  $j$  for  $\delta$ , the month at kidding effect, subclass  $k$  for  $\gamma$ , the gestation stage effect at DIM 270 (the last DIM studied), and class  $l$  for the herd-year (HY) effect. The fixed effects classes are the same as in the previous RRM.  $a_{nr}$  is the additive genetic breeding value of animal  $n$  for period  $r$  and followed a normal distribution ( $\text{Var}(a_r) \sim N(0, \sigma_{MT_r}^2)$ ).

**Lactation model (LACT)**

We implemented a lactation (LACT) model that was close to that routinely used for official genetic evaluations in France. The total milk, fat and protein yields throughout lactation were calculated as in routine genetic evaluations using the Fleischmann method [19]. Fat and protein contents were derived from the total lactation yields. The genetic evaluation model based on these phenotypes was:

$$y_{ijkln} = \pi_i + \delta_j + \gamma_k + HY_l + a_n + e_{ijkln}$$

where  $y_{ijkln}$  was the lactation observation for trait  $t$  (milk yield, fat yield, protein yield, fat content, protein content) of goat  $n$ , in subclass  $i$  for  $\pi$ , the age at kidding effect, subclass  $j$  for  $\delta$ , the month at kidding effect, subclass  $k$  for  $\gamma$ , the gestation stage effect at DIM 270 (the last DIM studied), and class  $l$  for herd-year effect (HY). Fixed effects were defined as in the previous RRM.  $a_n$  was the additive genetic breeding value of animal  $n$  and followed a normal distribution (mean:  $\mu_{LACT} = 0$ , variance is  $\sigma_{LACT}^2$ ).

All the genetic parameters were estimated using the WOMBAT software [20].

**Estimation of genetic correlations and heritabilities with the RRM**

For each trait, the heritability of the  $o$ th regression coefficients ( $h_b$ ) was calculated as in Schaeffer [21], by dividing the genetic variance of the  $o$ th regression coefficient by the sum of the genetic variance of the  $o$ th regression coefficient, the permanent environment variance of the  $o$ th regression coefficient, and the mean square error.

For each trait, the genetic variance–covariance matrix between all DIM was obtained following Druet et al. [14] as  $G_{264} = QK_gQ'$ , where  $G_{264}$  is a 264-by-264 genetic variance–covariance matrix,  $Q$  is a 264-by- $q$  matrix with the (daily) values of the  $q$  terms of the Legendre polynomial, and  $K_g$  is the  $q$ -by- $q$  genetic variance–covariance matrix. The same approach was used to obtain the permanent environmental variance–covariance matrix  $W_{264}$ . The phenotypic variance–covariance matrix between all DIM,  $P_{264}$ , was obtained by summing  $G_{264}$ ,  $W_{264}$  and the residual variance for the relevant DIM.

Heritabilities for each test day were obtained by dividing the diagonal elements of  $G_{264}$  by the corresponding diagonal elements of  $P_{264}$ .

The genetic correlations between DIM  $d$  and the other DIM were derived from  $\mathbf{G}_{264}$ .

Genetic variances for each trait throughout the whole lactation ( $g_{wl}$ ) were obtained following Hammami et al. [22] as  $g_{wl} = \mathbf{sG}_{264}\mathbf{s}'$ , where  $\mathbf{s}$  is a summation vector (vector of 1  $s$ ) of length 264. The same approach was used to obtain the permanent environmental ( $w_{wl}$ ) and phenotypic variances for the whole lactation  $p_{wl}$ . The heritability of each trait on the lactation scale ( $h_{wl}$ ) was obtained by dividing  $g_{wl}$  by  $p_{wl}$ .

### Criteria for comparing the models

The goodness-of-fit of the models for each trait was assessed by comparing the Bayesian information criterion (BIC) and the Pearson correlation coefficients ( $\rho$ ) between observed and predicted phenotypes for each model. Pearson correlation coefficients were also used to compare the EBV of the bucks obtained from different models.

## Results

### Rank reduction of the variance–covariance matrix of the most complex model

In the Eigen decomposition of the genetic matrix from the leg4 model, the first two principal components (PC) represented on average more than 97% of the total genetic variance (88 and 9%, respectively), and Additional file 1: Table S1 shows the percentage of variance represented by each PC for the five traits and the two breeds. The proportion of variance accounted for by the first eigenvalue was higher for yield traits and fat content in Saanen than in Alpine goats (from +2% for fat content to +4.5% for fat yield). In contrast, the proportion of variance accounted for by the third eigenvalue was higher in Alpine than in Saanen goats, although for all the traits it represented a very small fraction of the total variance (less than 4.2%).

The first eigenfunction was almost constant throughout the DIM period, which suggests that the first PC can be regarded as linked with the average production level throughout lactation, whereas the second PC varied almost linearly, which indicates extreme production levels at the beginning and end of lactation independently of the average production level, and thus it is associated with persistency. The correlations between these measures of average production level throughout lactation and persistency are equal to 0 by construction. The third eigenfunction showed contrasted production characteristics between those measured in the middle of lactation and those measured at the beginning and end of lactation. We found no significant between-trait differences in the shapes of eigenfunctions, see Additional file 2: Figure

S1 that shows the eigenfunctions of each PC for the five traits in the Saanen breed.

### Model fitting

Table 1 shows the BIC values for each trait and both breeds with the five complete models (leg0 to leg4) and the five reduced models (leg $x$ R $z$  with  $x = \{2, 3, 4\}$  and  $z = \{2, 3\}$ ).

Regardless of trait and breed, BIC decreased rapidly as the order of the Legendre polynomials increased from 0 to 2. The decay was smaller between orders 2 and 4 although 18 additional parameters had to be estimated. However, the differences between all models were significant (difference in BIC > 10), which indicates that, for all traits and both breeds, the fit to the data improved as the order of the Legendre polynomial used increased. As rank decreased, BIC decreased when the first three PC were kept instead of just the first two, without significant differences between leg3R3 and leg4R3. Regardless of the order of the Legendre polynomial, the BIC obtained by using two eigenfunctions of the genetic (co)variances matrix with model leg $x$ R2 were close and often better (i.e. smaller) than with model leg1, with the same number of estimated parameters. The BIC obtained by using three eigenfunctions (leg $x$ R3) was similar and better than that with leg2, again with the same number of estimated parameters.

Pearson correlation coefficients ( $\rho$ ) between observed data and predicted values were calculated to compare adjustment to data between traits and breeds and between RRM and MT models; Additional file 3: Table S2 presents the evolution of these correlations under the different models, for each trait in both breeds. For RRM, the conclusions drawn were similar to those for BIC. For the fourth-order Legendre polynomial,  $\rho$  were high for most traits ( $\sim 0.96$ ), but slightly lower for fat content in both breeds (0.93), which highlights a less satisfactory modeling of this trait than for the others. The MT model was the worst model for all traits, with  $\rho$  values ranging from 0.80 to 0.90, which can be explained by the genetic effect and certain fixed effects being constant throughout each period. With random regression models other than leg0, these effects gradually change with DIM.

The evolution of the residual variance with DIM is another criterion for characterizing the effect of this variation on the quality of the adjustments throughout lactation. In the French dairy goat breeding program, the most important trait for cheese production is protein yield, which can also be expressed as protein content with respect to the level of milk yield. The residual variances for milk yield and protein content for different lactation periods are shown in Fig. 1 and Additional file 4: Figure S2, respectively. To keep the figures easy

**Table 1 Bayesian information criterion<sup>a</sup> (BIC) for the five complete and five reduced models**

	Saanen					Alpine				
	Milk yield	Fat yield	Protein yield	Fat content	Protein content	Milk yield	Fat yield	Protein yield	Fat content	Protein content
Complete model										
leg0	19165	11543	13089	4229	14490	21314	14748	13014	4921	18671
leg1	7008	2631	4253	1749	4555	6539	2578	3392	1648	4913
leg2	1472	525	1023	747	1773	1739	835	975	826	2113
leg3	293	81	201	231	788	293	197	220	285	947
leg4	0	0	0	0	0	0	0	0	0	0
Reduced model										
leg2r2	3962	1967	3382	1737	4852	4575	2617	3259	1647	4769
leg3r2	3715	2258	3439	1825	4889	3827	2843	3352	1668	4858
leg4r2	3699	2307	3442	1765	4931	3794	2822	3292	1632	4849
leg3r3	1285	499	936	438	1521	952	612	906	570	1748
leg4r3	1040	500	915	251	1426	873	619	895	481	1898

<sup>a</sup> Values are expressed as deviations from the best (smallest) value (leg4)

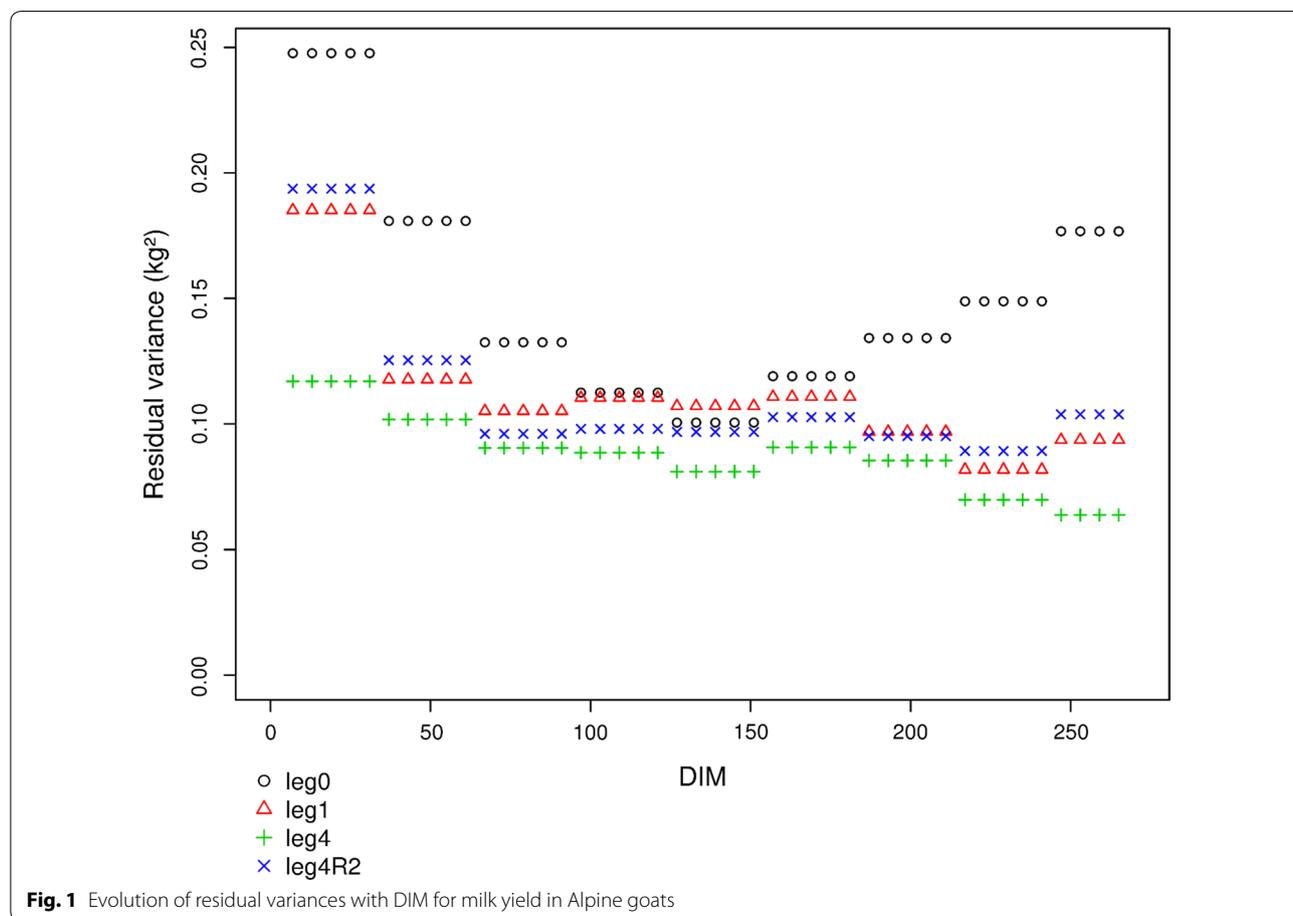
to interpret, the results for the leg2 and leg3 models are not presented, but these were very close to those of the leg4 model. The between-model differences in residual variance are larger at the beginning and end of lactation. As for BIC and  $\rho$ , the higher is the order of the polynomial, the better is the adjustment, as indicated by the smaller residual variances. Figure 1 and Additional file 4: Figure S2 illustrate the better adjustment obtained with leg4 compared to the other models, at both ends of lactation. legxR3 resulted in smaller residual variances than legxR2, particularly at the beginning (DIM 7–45) and end of lactation (DIM 225–270) (results not shown). For clarity, the residuals of the reduced models derived from leg2 and leg3 were not included in the figures but were close to those from leg4. Overall, the RRM\* adjustments were systematically much better than for the model that assumes a constant genetic value throughout lactation (leg0). For model fit, similar results were found for the other traits in both breeds.

**Choice of final models**

Model leg4 was chosen as the reference model for RRM because it resulted in a better fit to the data, but it is also the most complicated model in terms of number of genetic parameters to estimate. However, leg0 and leg1 appear too simplistic. In order to choose a robust model that is sufficiently manageable for large-scale routine evaluations (after extension to all lactations), we chose the leg4R2 model as an attractive compromise, since it derived from the best RRM and its interpretation was straightforward, i.e. it summarized two important lactation characteristics (average level and persistency). The third PC was considered less relevant because it accounted for much less variance. These different RRM (leg0, leg1, leg4R2 and leg4) were compared to a lactation model (LACT) and a MT model for the estimation of genetic parameters.

**Heritability estimates**

Heritabilities for each trait on the whole lactation ( $h_{wl}$ ) scale were estimated as in Hammami et al. [22] with the LACT and different RRM models and were close to 0.3



for milk yield in both breeds and for fat and protein yields in Alpine goats, and slightly higher i.e. 0.34 for protein yield and 0.37 for fat yield in Saanen goats (see Additional file 5: Table S3). Heritability estimates reached 0.68 for fat content and 0.65 for protein content in Saanen, and 0.75 for fat content and 0.70 for protein content in Alpine goats. We observed no differences in heritability between the RRM for any of the traits. Furthermore, heritabilities estimated with the RRM\* were close to those estimated with RRM. For all the traits, heritabilities estimated with the RRM were slightly higher, especially for fat content, than those estimated with the LACT model

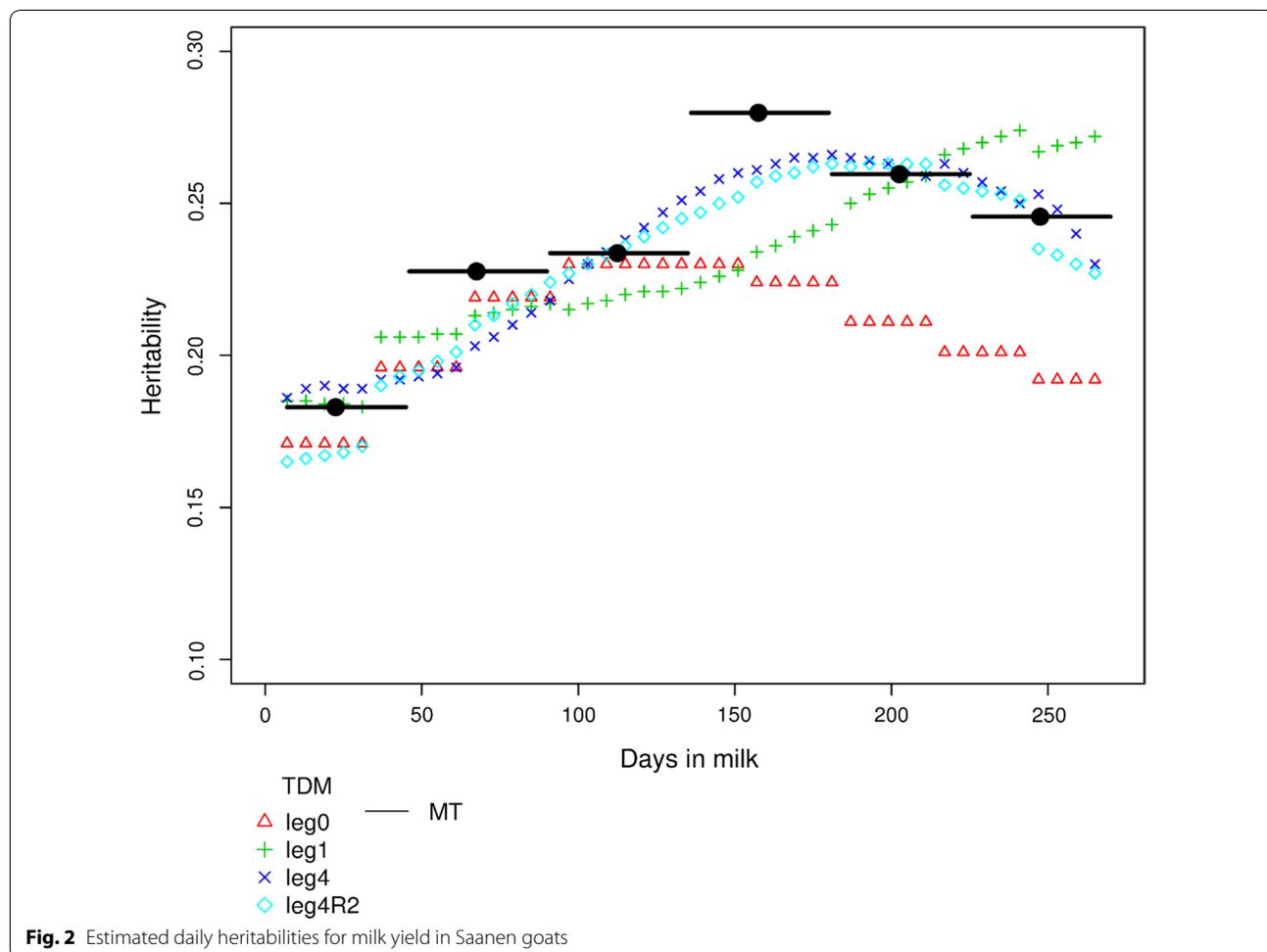
(0.65 for LACT and 0.75 with leg4 for fat content in the Alpine breed).

The heritabilities for the regression coefficients ( $h_b$ ) estimated with the leg4R2 model as in Schaeffer [21] are in Table 2. These can be interpreted as heritability of average yield (PC1) and persistency of yield (PC2), respectively. The heritabilities for persistency (PC2) were close to 0.18 for protein content and ~0.10 for the other traits.

Figure 2 shows the evolution of the estimated heritabilities of test day yield throughout the lactation period obtained with the MT model and different RRM models

**Table 2 Heritabilities of the regression coefficients ( $h_b$ ) according to the PC derived from the leg4R2 model**

	Saanen					Alpine				
	Milk yield	Fat yield	Protein yield	Fat content	Protein content	Milk yield	Fat yield	Protein yield	Fat content	Protein content
PC1	0.28	0.31	0.29	0.50	0.56	0.27	0.24	0.25	0.54	0.62
PC2	0.13	0.12	0.11	0.07	0.18	0.14	0.10	0.09	0.10	0.19

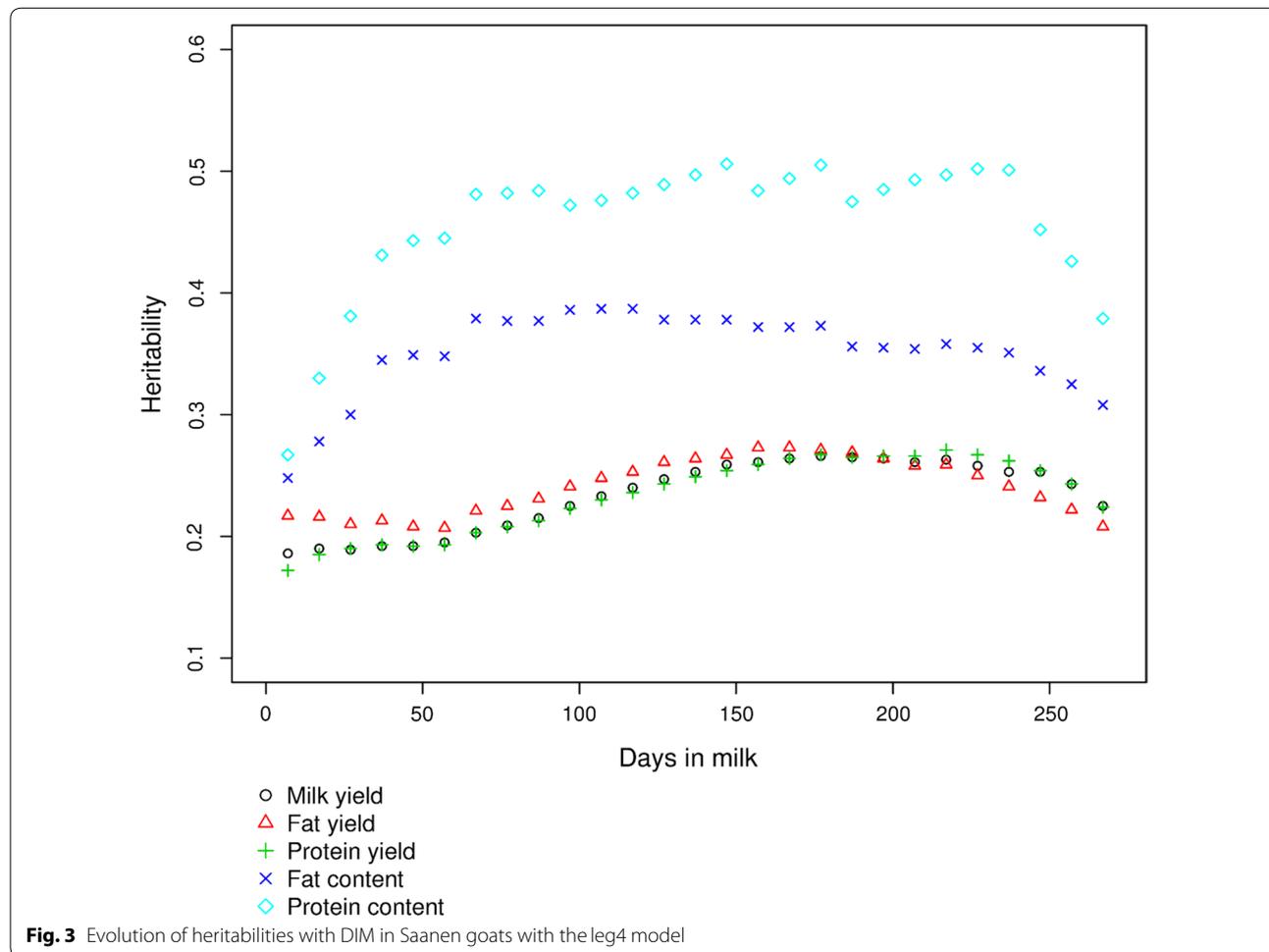


**Fig. 2** Estimated daily heritabilities for milk yield in Saanen goats

for milk yield in the Saanen breed, i.e. leg4 and leg4R2, which were the models without and with reduction, leg1, which was a model that had the same number of parameters as leg4R2, and finally leg0. Models leg4 and leg4R2 yielded the highest heritabilities. The discontinuous curve of the estimated heritability throughout the DIM period is due to the residual variance that was different for each of six 30-day periods in the different RRM. The heritability estimated with the MT model first increased up to a peak in mid-lactation (0.26) and then decreased. However, leg0 yielded a lower heritability after DIM 150 for milk yield (due to a larger residual variance), whereas leg1 resulted in a different pattern with low heritabilities at the beginning but high heritabilities at the end of lactation. Additional file 6: Figure S3 shows the evolution of the estimated heritabilities throughout the DIM period for protein content in Saanen goats. The comparison between models resulted in similar results for milk yield and protein content, except with leg1 for protein content, which showed a profile closer to that of leg4. Similar results were also observed for the Alpine breed and the

other traits. Estimated heritabilities obtained with RRM\* were close to those calculated with MT and leg4 models, because their eigenfunctions are neither constant nor linear throughout lactation, unlike those of the leg0 and leg1 models.

Figure 3 for the Saanen breed and Additional file 7: Figure S4 for the Alpine breed present the heritabilities estimates for test days with the leg4 model throughout the lactation period for all five traits and show the same within-breed patterns for the three yield traits. The highest heritability estimated for milk yield was 0.26 in both breeds, and was reached on ~DIM 175 and ~DIM 125 for the Saanen and Alpine breeds, respectively. For component yield traits in the Alpine breed, the highest heritability of 0.23 was reached on ~DIM 125, similar to that for milk yield. For component yield traits in the Saanen breed, the estimated heritabilities were highest, i.e. 0.27 on ~DIM 157 for fat yield and ~DIM 211 for protein yield. For protein and fat contents, the evolution of the heritabilities throughout the DIM period followed the same pattern in both breeds but values were higher in the



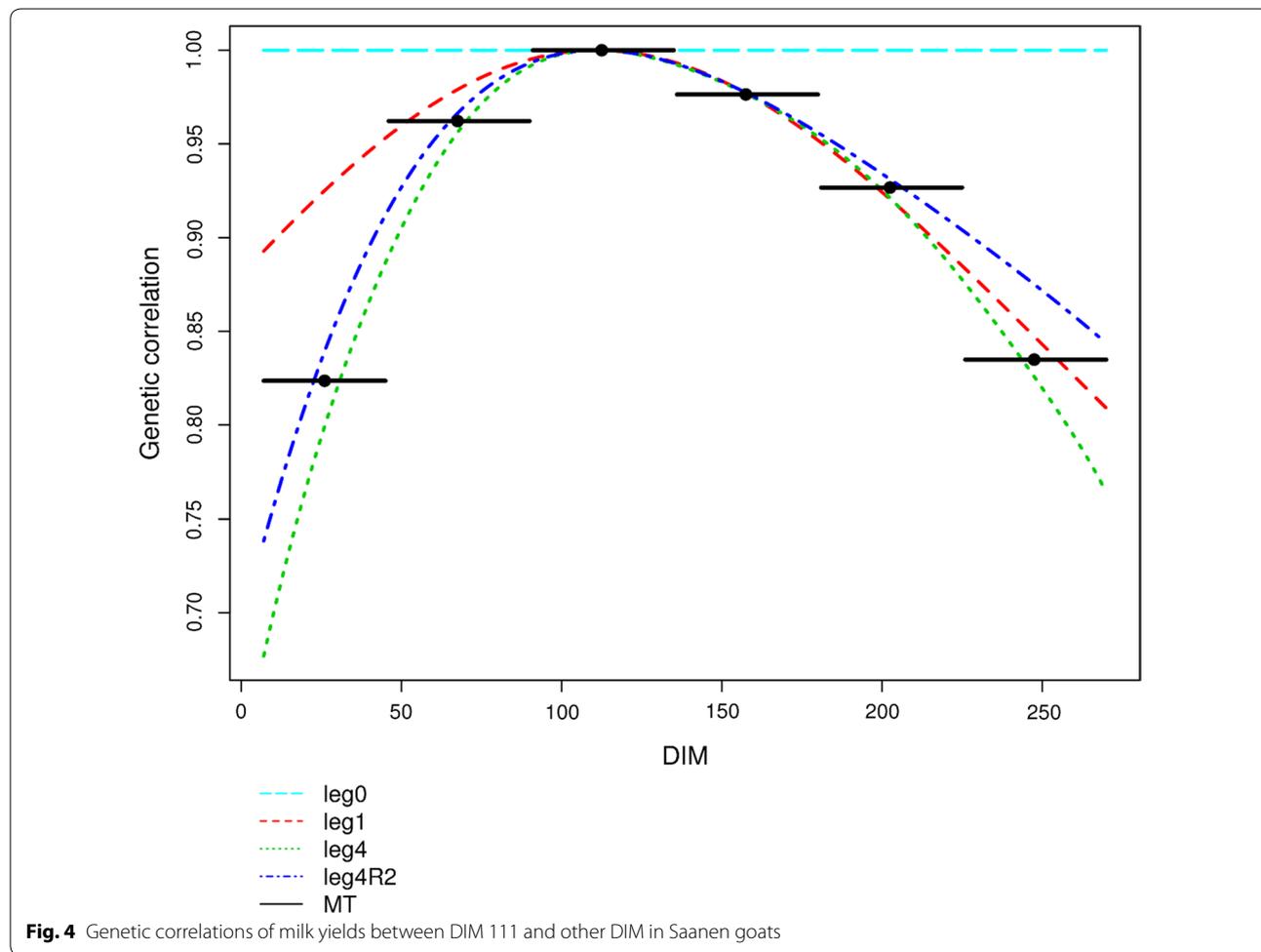
Alpine breed i.e. 0.45 and 0.57, respectively for fat and protein content on ~DIM 150 versus 0.38 for fat content on ~DIM 106 and 0.51 for protein content on ~DIM 150 in the Saanen breed.

**Genetic correlations between DIM from different models**

In order to compare the genetic correlations between periods obtained with the MT model and the genetic correlations between DIM obtained with the different RRM models, we selected the median day for each period of the MT model. Figure 4 and Additional file 8: Figure S5 show the genetic correlations of milk yield and protein content, respectively, at DIM 111 (i.e. the middle of the third period) with those at other DIM in the Saanen breed with the same models as those used for Fig. 2. The results are similar for both breeds. With model leg0, which assumes a constant genetic effect throughout lactation, the genetic correlation is equal to 1 between all DIM. This is also implicitly the case for the LACT model, which gives a global EBV for the whole lactation assuming a correlation of 1 between all DIM. The most similar

genetic correlations are those estimated with the leg4 and MT models. As observed earlier with BIC, leg4 was the best model among all RRM. For all traits, the genetic correlations between any two periods estimated with the MT and RRM models were positive. The lowest correlation i.e. 0.44 was found with the MT model for milk yield in the Alpine breed between periods 1 and 6. The genetic correlations between DIM from the RRM\* were fairly aligned to the pattern from the best full-rank model (leg4) for all traits and both breeds.

Additional file 9: Figure S6 and Additional file 10: Figure S7 show the genetic correlations, estimated with the leg4 model, for all production traits between TD at DIM 40, which corresponds to the DIM at lactation peak, and at the other DIM, in both the Saanen and Alpine breeds. Genetic correlations for milk yield between the peak and the end of lactation were positive, i.e. ~0.5 in the Saanen and ~0.4 in the Alpine breed. In the Alpine breed, daily genetic correlations for fat and protein contents were higher (e.g. a correlation of ~0.7 for protein content between DIM 40 and at the end of lactation) than those



for yield traits (e.g. a correlation of ~0.4 for milk yield between DIM 40 and at the end of lactation). For each trait, the genetic correlations between DIM were higher in the Saanen than the Alpine breed (e.g. correlations of ~0.4 (Alpine) and ~0.5 (Saanen) for milk yield between DIM 40 and end of lactation).

**Correlations between EBV for level of production and persistency**

**Within-trait**

Table 3 presents the correlations between EBV for milk yield of bucks that were sires of the phenotyped goats (379 in Alpine and 324 in Saanen) in Saanen under the leg1, leg4 and leg4R2 models.

We were able to evaluate the level of milk produced throughout lactation using the LACT model, or by considering the SUM\_legx and SUM\_leg4R2 values, or the first coefficient of RRM (a<sub>0</sub>) or RRM\* (b<sub>1</sub>). Correlations between these values reached a value of at least 0.99 for milk yield (Table 3) and all other traits, confirming that they characterize the same trait.

For the study of persistency, the leg0 and LACT models were unsuitable because the resulting EBV were not DIM-dependent. Persistency could be evaluated by considering the PERS\_leg4 value, or the second coefficient of RRM (a<sub>1</sub>) and RRM\* (b<sub>2</sub>). A correlation close to 1 was found between the a<sub>1</sub> coefficient of leg4 and PERS\_leg4 for milk yield in the Saanen breed. The correlation between b<sub>2</sub> and PERS\_leg4 was 0.98 for milk yield and 0.88 for protein yield.

The correlation between SUM\_leg4 and PERS\_leg4 was low for milk yield (i.e. 0.15, see Table 3) in Saanen and close to 0 in Alpine, but higher for protein content in both breeds (0.45 in Saanen and 0.58 in Alpine) (results not shown). The correlations between a<sub>0</sub> and

a<sub>1</sub> from leg1 and leg4 were equal to 0.44, whereas those between b<sub>1</sub>\_leg4R2 and b<sub>2</sub>\_leg4R2 were equal to 0.01. The aim is to have a low genetic correlation between the level and the persistency because it allows selection of animals with a desired persistency throughout the lactation without changing the net-total lactation output.

**Correlations between milk yield and other production traits**

Table 4 reports the correlations between EBV of SUM\_leg4 (i.e. milk production level) and PERS\_leg4 (i.e. persistency) for milk yield and of SUM\_leg4 and PERS\_leg4 for the other traits for Saanen and Alpine bucks.

**Table 4 Correlations between EBV for milk yield and the other traits for bucks**

	Milk yield			
	Saanen		Alpine	
	SUM_leg4	PERS_leg4	SUM_leg4	PERS_leg4
Fat yield				
SUM_leg4	0.75	0.15	0.73	-0.13
PERS_leg4	-0.07	0.86	-0.17	0.85
Protein yield				
SUM_leg4	0.90	0.18	0.88	-0.04
PERS_leg4	0.16	0.94	0.02	0.93
Fat content				
SUM_leg4	-0.20	0.06	-0.33	-0.14
PERS_leg4	-0.36	-0.10	-0.32	-0.17
Protein content				
SUM_leg4	-0.33	0.01	-0.44	-0.10
PERS_leg4	-0.29	-0.45	-0.28	-0.58

**Table 3 Correlations between EBV for milk yield of buck sires of the recorded goats in Saanen**

	LACT	SUM_leg4	PERS_leg4	a <sub>0</sub> _leg4	a <sub>1</sub> _leg4	SUM_leg1	a <sub>0</sub> _leg1	a <sub>1</sub> _leg1	SUM_leg4R2	b <sub>1</sub> _leg4R2
SUM_leg4	0.99									
PERS_leg4	0.06	0.15								
a <sub>0</sub> _leg4	0.99	1.00	0.15							
a <sub>1</sub> _leg4	0.11	0.20	1.00	0.20						
SUM_leg1	0.99	1.00	0.15	1.00	0.20					
a <sub>0</sub> _leg1	0.99	1.00	0.15	1.00	0.20	1.00				
a <sub>1</sub> _leg1	0.10	0.19	1.00	0.19	0.99	0.19	0.19			
SUM_leg4R2	0.99	1.00	0.15	1.00	0.20	1.00	1.00	0.19		
b <sub>1</sub> _leg4R2	0.99	1.00	0.18	1.00	0.23	1.00	1.00	0.22	1.00	
b <sub>2</sub> _leg4R2	-0.10	-0.01	0.98	-0.01	0.97	-0.01	-0.01	0.97	-0.01	0.02

LACT, lactation model; SUM\_legx, sum of the daily EBV of legx; PERS\_leg4, persistence calculated as in [18] from leg4; a<sub>0</sub>\_legx, regression coefficient for the 0th term of Legendre polynomial; a<sub>1</sub>\_legx, regression coefficient for the 1st term of the Legendre polynomial; b<sub>1</sub>\_leg4R2, regression coefficient for the first eigenfunction; b<sub>2</sub>\_leg4R2, regression coefficient for the second eigenfunction

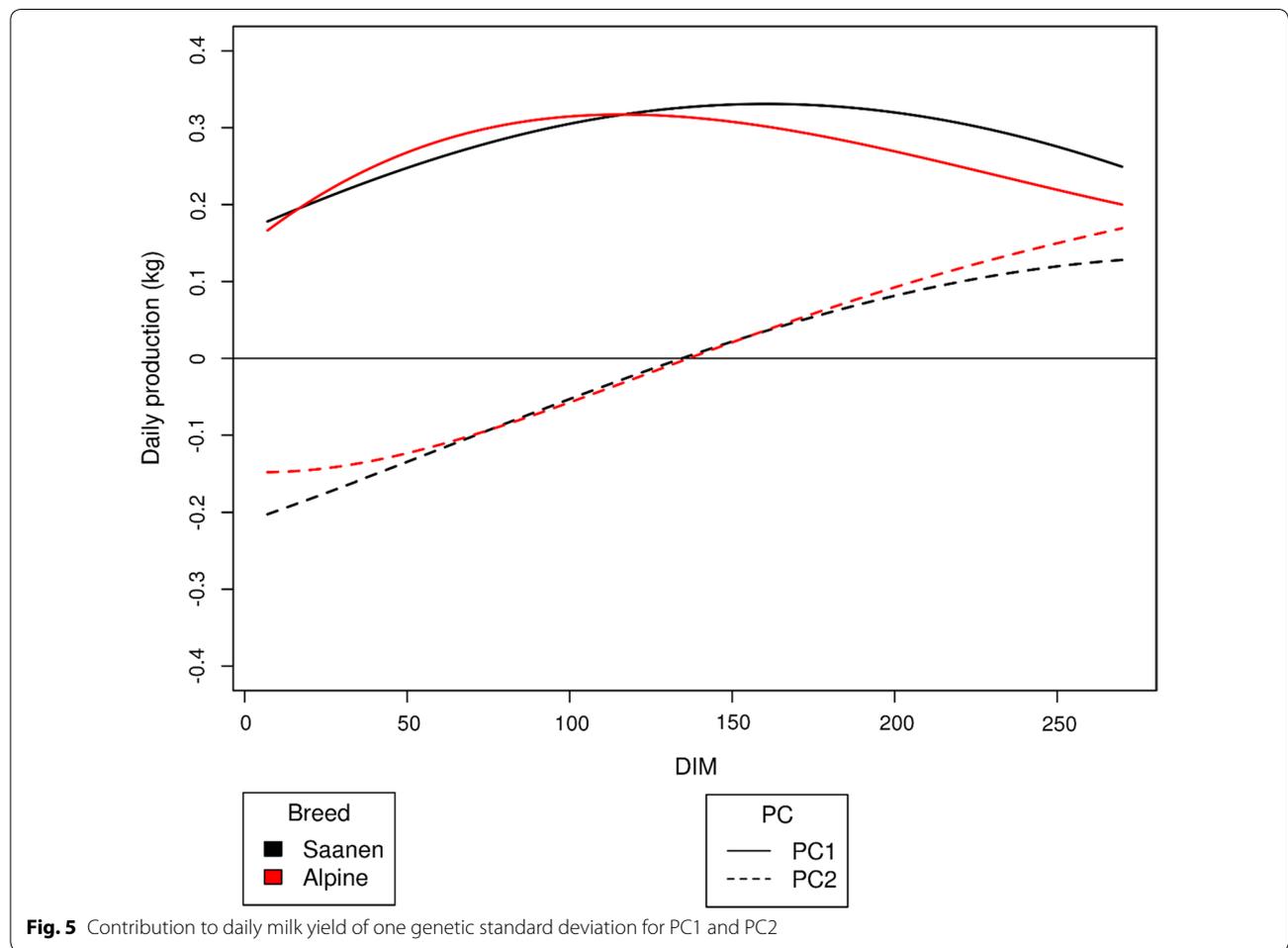
For levels of production, the correlations between SUM\_leg4 for milk yield and SUM\_leg4 for the other traits averaged 0.75 for fat yield, 0.89 for protein yield, -0.25 for fat content and around -0.4 for protein content. The correlations between SUM\_leg4 for milk yield and persistencies of the other traits were moderate (all < 0.36 in absolute value). The correlations between milk yield persistency and SUM\_leg4 for the other traits were close to 0 (between -0.14 and 0.18) for both breeds. The correlations between milk persistency and the other persistencies were very high (> 0.85) for protein and fat yields but very low for fat content (-0.13, on average). However, the correlation with persistency of protein content was negative and relatively high (-0.45 in Saanen and -0.58 in Alpine). These quite high correlations indicate that a higher milk persistency is positively correlated with a higher-than-average protein content at the beginning of lactation. Because the correlation between fat content persistency and milk yield persistency was low (-0.10 in Saanen and -0.17 in Alpine) and because protein content at the beginning of lactation was higher

than average, persistency is correlated with a reduced fat:protein ratio at the beginning of lactation.

As expected, the correlations of EBV between  $b_1\_leg4R2$  and  $b_2\_leg4R2$  for milk yield and  $b_1\_leg4R2$  and  $b_2\_leg4R2$  for the other traits (not shown) were very close to those obtained in Table 4. This result confirms that  $b_1\_leg4R2$  is a measure of lactation production and  $b_2\_leg4R2$  is a measure of persistency.

**Relative contribution of  $b_1$  and  $b_2$  coefficients to lactation production**

Figure 5 and Additional file 11: Figure S8 show the relative contribution to lactation production of a change of one standard deviation in the  $b_1$  (level) and  $b_2$  (slope) coefficients for milk yield and protein content, respectively, with leg4R2. An increase of one genetic standard deviation for  $b_1$  leads to an increased production level throughout lactation, with a higher increase at the end of lactation for Saanen compared to Alpine goats. This study on the one-standard-deviation increase of  $b_1$  and  $b_2$  allows to analyze the potential result of a selection



**Fig. 5** Contribution to daily milk yield of one genetic standard deviation for PC1 and PC2

on these two traits. As a consequence of the one-standard-deviation increase of  $b_1$ , the sum of the added daily milk productions represents 76 kg of milk for Saanen and 71.1 kg for Alpine goats, which is relatively consistent with the LACT model (genetic standard deviation equal to 92 kg in Saanen and 85 kg in Alpine goats). At the beginning of lactation, increasing the persistency ( $b_2$ ) by one standard deviation reduces production by 12.9 kg until DIM 129 and then increases it by 10.5 kg at the end of the lactation for Saanen. For protein content, a gain of one standard deviation of  $b_1$  increases average production more in Alpine than in Saanen goats. A loss of one standard deviation in protein content persistency leads to an increase of 1 g/kg at the beginning of lactation and a loss of 1 g/kg at the end of lactation.

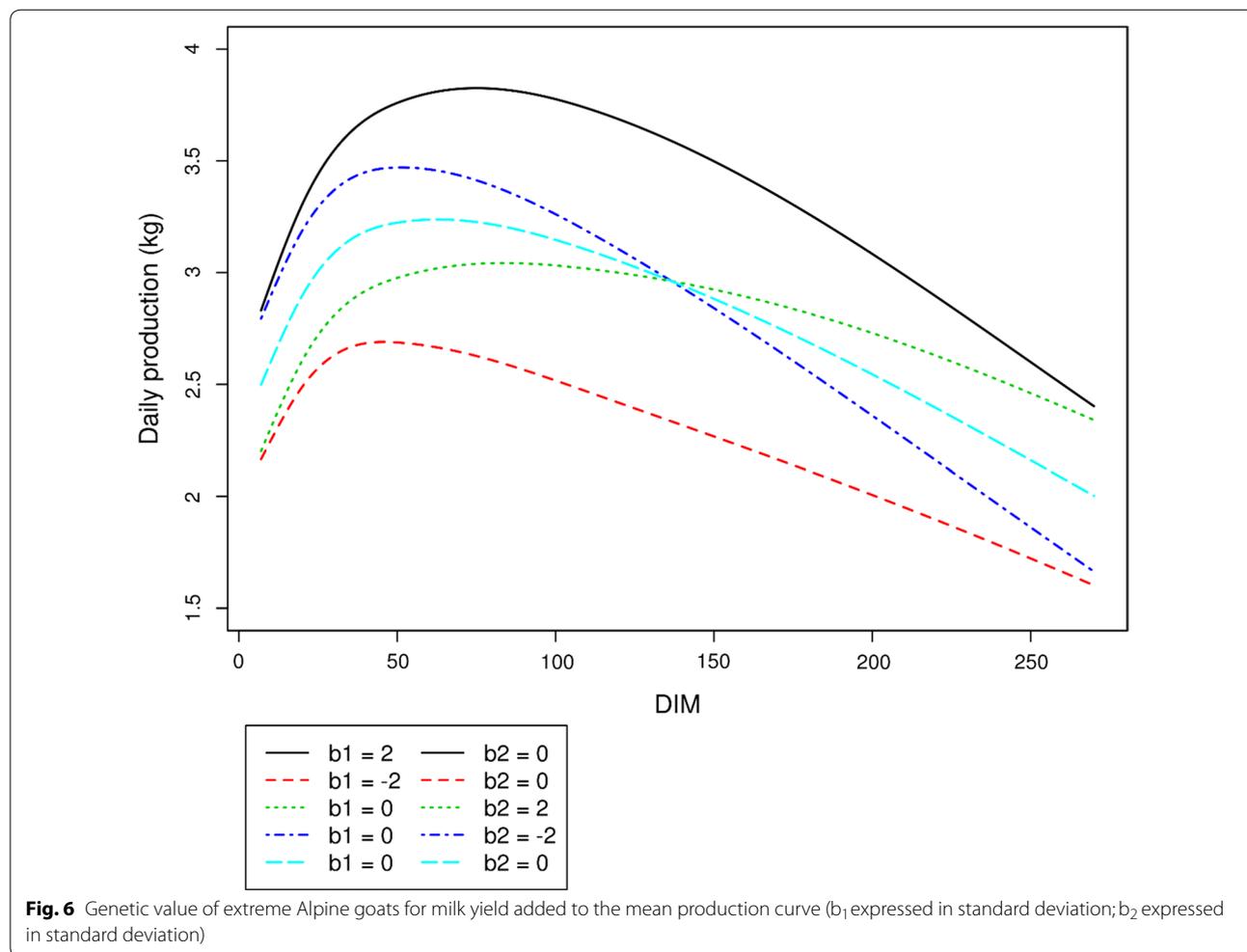
As an illustration, Fig. 6 displays the genetic daily EBV added to the average standard production curve, for two extreme-value goats for global production level and two extreme-value goats for milk yield persistency.

Additional file 12: Figure S9 shows the same results but for protein yield.

### Discussion

We estimated various genetic parameters in a large population of goats that had been regularly measured for five major dairy traits throughout lactation since 1995. This estimation was done by using a random regression model for the first time in France for the two main goat breeds Alpine and Saanen, which represent 97% of the goats recorded in France. Connectedness between the large number of herds was ensured by using phenotypes from daughters of sires from artificial insemination (379 and 324 in the Alpine and Saanen breed, respectively).

The fixed effects, i.e. age at kidding, month of kidding, dry-period length and gestation stage, were modeled with splines in the RRM model. In a previous study [11], we showed that these factors had an impact on the shape of the lactation curve, and that there was no interaction between these effects and year of lactation. These



results were consistent with other studies in dairy cows [14, 23]. Unlike other studies in dairy goats [4, 5], we did not use an effect on litter size since this information was not available in our dataset, but Mucha et al. [1] found no impact of this effect on EBV.

We applied a principal component analysis of the variance–covariance matrix of the most complex model, which showed that it was possible to associate biological significance with each PC. As in our study, van der Werf et al. [17], Olori et al. [24], Druet et al. [14] and Togashi and Lin [25] highlighted that the first eigenfunction was linked with the average production level throughout lactation, the second eigenfunction was associated with persistency, and the third eigenfunction opposed production around the middle of lactation against production at the beginning and end of lactation. The high percentages of variances explained by the first three PC (and even the two first PC), and the desire to reduce the overall dimension of the model in routine evaluations, pointed at the need to test a rank reduction of the variance–covariance matrix.

Implementing a RRM for genetic evaluation requires cumbersome testing of the best tradeoffs between model fit and complexity. Pool et al. [26] compared the residual variances of different RRM on cow data and found, as we did here, that the most suitable model for TD milk production traits was leg4. The use of other criteria (BIC and Pearson correlation coefficients) also confirmed that leg4 was the most suitable model for TD milk production traits. The leg $\times$ R3 reductions (for  $x = \{3,4\}$ ) had a better fit to the data than leg $\times$ R2 although leg $\times$ R2 reductions led to a better adjustment to the data compared to other more concise models such as leg0 and leg1, especially for milk yield. This rank reduction facilitates the extension of the model to subsequent lactations 2 and 3 in order to construct a genetic evaluation. Indeed, this model extension has to take into account how genetic correlations between first and following parities will probably differ from 1 (around 0.7 for milk yield) as found by several studies in dairy goats [1, 4].

For all traits, heritability estimates from the LACT model and for the regression coefficients from RRM and RRM\* were close to those reported by Rupp et al. [27] for the same five traits in French Alpine and Saanen goats in first lactation. The evolution of the heritability throughout lactation for each trait was similar between RRM and RRM\*. In Norwegian goats, Andonov et al. [6] showed that the heritability of milk yield estimated with the MT model increased up to a maximum at mid-lactation (0.26) and then decreased, which agrees with our observations. The evolution of the heritability throughout the DIM period found here for milk yield differed from that reported in other studies using RRM

in goats. Menéndez-Buxadera et al. [4] found a maximum heritability of 0.24 at the beginning of lactation, then a decrease with DIM in Murciano-Granadina goats whereas Zumbach et al. [5] found a maximum heritability of 0.4 and then a decrease with DIM in six German breeds. Andonov et al. [6] found a maximum heritability of 0.33 at  $\sim$  DIM 155, in a Norwegian goat breed and Mucha et al. [1] found a maximum heritability of 0.45 at  $\sim$  DIM 220, in a crossbred population including three goat breeds: Alpine, Saanen, and Toggenburg. The mean heritability estimates for protein yield found here ( $\sim$ 0.2) was in agreement with that reported by Muños-Mejías et al. [3] for goats of the Florida breed. This was not the case for fat yield, for which we found a higher heritability ( $\sim$ 0.2) than that for the Florida breed ( $\sim$ 0.15). Moreover, Muños-Mejías et al. [12] showed that the estimated heritabilities for fat and protein yield for Florida goats were higher at the end of lactation, which was not the case in our study. For protein and fat contents, the estimated heritabilities reported in Muños-Mejías et al. [3] and Andonov et al. [6] were lower than those found here, and followed a different shape, i.e. they increased with DIM. The RRM\* made it possible to calculate the heritability of persistency for each trait [21]. The heritabilities of persistency for milk yield calculated from RRM\* were close to those reported by Cole and VanRaden [8] in cattle, who also calculated milk yield persistency ensuring that it was not correlated with milk yield level. Menéndez-Buxadera et al. [4] reported a heritability of milk yield persistency 0.208, but the correlation with lactation yield level is unknown.

As in our study, Mucha et al. [1] found that all the genetic correlations for milk yield between two periods in the trajectory fitted in the RRM models were always positive. This indicates that selection of animals based on any daily EBV will yield positive responses for all the other days in the lactation curve. Also in agreement with our results, Muñoz-Mejías et al. [3], Andonov et al. [6] and Menéndez-Buxadera et al. [4] found that genetic correlations between days were higher for fat and protein contents than for yield traits. The moderate genetic correlations for milk yield between the peak and end of lactation indicated genetic variability in the level of milk production between the peak and end of lactation, and therefore a genetic variability for milk persistency. For all traits, the between-day genetic correlations were similar with RRM or RRM\*.

We showed that correlations between EBV from RRM and RRM\* were close to 1 for both average and persistency of yield as reported in Leclerc et al. [28] who compared EBV from a complete and a reduced model. The advantages of the reduced model are the smaller size of the variance–covariance matrices and the zero

correlation between EBV for lactation yield ( $b_1$ ) and persistency ( $b_2$ ) by construction. This allows selection of animals with a desired persistency throughout the lactation without changing the net-total lactation output. Model leg1 does not have this advantage as the correlation between mean level and persistency is high in that model. The large difference in lactation trajectory between extreme animals suggests that it can be valuable to consider persistency in selection. The form of the first eigenfunction (Fig. 5) is interesting because it represents the pattern of how a goat produces milk throughout lactation from its genetic makeup. The eigenfunction coordinates were higher at the end of the lactation period for the Saanen breed than for Alpine breed, confirming the observations of Arnal et al. [11] who showed a better persistency for Saanen goats than Alpine goats. This is also evidenced by the higher correlation between SUM\_leg4 and PERS\_leg4 for the Saanen compared to Alpine goats (0.15 in Saanen vs. 0 in Alpine) as well as the higher genetic correlations between DIM in Saanen compared to Alpine goats.

The correlations of the production level for milk yield with the production level for the other traits were very close to those reported by Bélichon et al. [29] on total lactation traits (0.90 for protein yield and 0.76 for fat yield). For protein content and fat content, Bélichon et al. [29] found slightly lower correlations, i.e.  $-0.28$  for protein content and  $-0.13$  for fat content while we found 0.39 and  $-0.27$ , respectively, in our study. The correlations between milk yield persistency and the production level for the other traits were weak or close to 0 for both breeds, indicating that milk yield persistency can be selected for with no impact on content-related traits. Furthermore, the difference in the correlations observed here between milk persistency and protein content persistency on the one hand, and between milk persistency and fat content persistency on the other hand, indicates that a goat with a high milk persistency will tend to have a lower fat:protein ratio in early lactation. Several studies in dairy cattle [13, 30, 31] showed that a high fat:protein ratio was associated with a negative energy balance, sub-clinical mastitis, and poor fertility. These results point out to the potential value of lactation persistency in breeding schemes.

Finally, the development of a test-day model for milk production opens up new perspectives. For example, the study of the genetic relationship between persistency and other traits such as longevity or fertility could help to explain the negative correlation between high production and fitness. The estimated fixed effects of such test-day models, especially the herd-test-day effect, also offer producers important clues on the impact of herd

management on these traits. For example, herd test-day estimates can be compared between farms under similar systems in terms of mean and variability throughout the year [23].

## Conclusions

In this paper, we show that the genetic parameters obtained with a test-day model using a fourth-order Legendre polynomial (leg4) for the genetic and permanent environmental components and a multi-trait model are consistent. However, this kind of model is complex and computationally demanding. Given that the aim was to develop TD genetic evaluations for traits in selection schemes (milk production traits and somatic cell score) over several parities and possibly including genomic information, a simpler model is necessary. We found that reducing the genetic and permanent environment (co) variance matrices of leg4 to its first two PC (leg4R2) was a satisfactory compromise, which accurately approximates genetic parameters and EBV under the complete model. This reduced model can give EBV for total lactation milk yield and persistency that are nearly independent (correlation close to 0). This negligible correlation is appealing because it allows to select animals with a desired lactation shape independently from selection for total lactation production. Therefore, for the extension of the TD models to several lactations, we consider that leg4R2 is more appropriate for implementation than the complete model leg4.

## Additional files

**Additional file 1: Table S1.** Proportion (%) of genetic variance explained by the first three principal components (PC) using leg4.

**Additional file 2: Figure S1.** Coordinates of the first three eigenfunctions of the test-day genetic covariance matrix computed with leg4 in Saanen goats.

**Additional file 3: Table S2.** Pearson correlation coefficients ( $\rho$ ) between observed and predicted values with the complete and reduced models. MT: multitrait model; legx: Legendre polynomial of order x; legxRz: Legendre polynomial of order x reduced to order z.

**Additional file 4: Figure S2.** Evolution of residual variances with DIM for protein content in Alpine goats (full-rank models: leg0, leg1, leg2, leg4; reduced model: leg4R2).

**Additional file 5: Table S3.** Whole-lactation heritabilities from the LACT model and heritabilities ( $h_{wl}$ ) calculated as in [22] in Saanen and Alpine goats.

**Additional file 6: Figure S3.** Daily estimated heritabilities for protein content in Saanen goats (full-rank models: leg0, leg1, leg4; best reduced model: leg4R2 and the MT model).

**Additional file 7: Figure S4.** Evolution of heritabilities with DIM in Alpine goats with the leg4 model.

**Additional file 8: Figure S5.** Genetic correlations of protein contents between DIM 111 and other DIM in Saanen goats.

**Additional file 9: Figure S6.** Genetic correlations between DIM 40 and other DIM in Alpine goats from the leg4 model.

**Additional file 10: Figure S7.** Genetic correlations between DIM 40 and other DIM in Saanen goats from the leg4 model.

**Additional file 11: Figure S8.** Contribution to daily protein content of one genetic standard deviation for PC1 and PC2.

**Additional file 12: Figure S9.** EBV of Alpine goats added to the mean production of the population for protein content ( $b_1$  expressed in standard deviation;  $b_2$  expressed in standard deviation).

**Acknowledgements**

The authors thank Karin Meyer for the Wombat program.

**Authors' contributions**

CRG, HL and VD proposed the models for study and test. MA performed the analysis and wrote the paper. MA, CRG, HL, VD and HL interpreted the results. CRG, HL, VD and HL revised and improved the manuscript. All authors read and approved the final manuscript.

**Funding**

The first author received financial support from APIS-GENE (Paris, France) and the French National Association for Research and Technology (ANRT, Paris, France).

**Availability of data and materials**

The datasets analyzed in this study are not publicly available since they were partially produced by private professional partnerships.

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Author details**

<sup>1</sup> GenPhySE, Université de Toulouse, INRA, ENVT, Toulouse INP, 31326 Castanet-Tolosan, France. <sup>2</sup> Institut de l'Élevage, Chemin de Borde Rouge, 31326 Castanet-Tolosan Cedex, France. <sup>3</sup> INRA UMR1313 GABI, Institut de l'Élevage, 78350 Jouy-en-Josas, France. <sup>4</sup> UMR1313 GABI, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France.

Received: 6 February 2019 Accepted: 31 July 2019

Published online: 13 August 2019

**References**

1. Mucha S, Mrode R, Coffey M, Conington J. Estimation of genetic parameters for milk yield across lactations in mixed-breed dairy goats. *J Dairy Sci.* 2014;97:2455–61.
2. Brito LF, Silva FG, Oliveira HR, Souza NO, Caetano GC, Costa EV, et al. Modelling lactation curves of dairy goats by fitting random regression models using Legendre polynomials or B-splines. *Can J Anim Sci.* 2017;98:73–83.
3. Muñoz-Mejías ME, Menéndez-Buxadera A, Sánchez-Rodríguez M, Serradilla JM. Genetic progress attained in the selection program of Florida breed of goats in Spain. *Option Méditerranéennes Ser A.* 2013;108:134–9.
4. Menéndez-Buxadera A, Molina A, Arrebola F, Gil MJ, Serradilla JM. Random regression analysis of milk yield and milk composition in the first and second lactations of Murciano-Granadina goats. *J Dairy Sci.* 2010;93:2718–26.
5. Zumbach B, Tsuruta S, Misztal I, Peters KJ. Use of a test day model for dairy goat milk yield across lactations in Germany. *J Anim Breed Genet.* 2008;125:160–7.
6. Andonov S, Ødegård J, Svendsen M, Ådnøy T, Vegara M, Klemetsdal G. Comparison of random regression and repeatability models to predict

- breeding values from test-day records of Norwegian goats. *J Dairy Sci.* 2013;96:1834–43.
7. Schaeffer LR, Jamrozik J. Random regression models: a longitudinal perspective. *J Anim Breed Genet.* 2008;125:145–6.
8. Cole JB, VanRaden PM. Genetic evaluation and best prediction of lactation persistency. *J Dairy Sci.* 2006;89:2722–8.
9. Sölkner J, Fuchs W. A comparison of different measures of persistency with special respect to variation of test-day milk yields. *Livest Prod Sci.* 1987;16:305–19.
10. Gipson TA, Grossman M. Lactation curves in dairy goats: a review. *Small Ruminant Res.* 1990;3:383–96.
11. Arnal M, Robert-Granié C, Larroque H. Diversity of dairy goat lactation curves in France. *J Dairy Sci.* 2018;101:11040–51.
12. Schaeffer LR, Jamrozik J, Kistemaker GJ, Van Doormaal J. Experience with a test-day model. *J Dairy Sci.* 2000;83:1135–44.
13. Jamrozik J, Schaeffer LR. Test-day somatic cell score, fat-to-protein ratio and milk yield as indicator traits for sub-clinical mastitis in dairy cattle. *J Anim Breed Genet.* 2012;129:11–9.
14. Druet T, Jaffrézic F, Boichard D, Ducrocq V. Modeling lactation curves and estimation of genetic parameters for first lactation test-day records of French Holstein cows. *J Dairy Sci.* 2003;86:2480–90.
15. White IMS, Thompson R, Brotherstone S. Genetic and environmental smoothing of lactation curves with cubic splines. *J Dairy Sci.* 1999;82:632–8.
16. Misztal I. Properties of random regression models using linear splines. *J Anim Breed Genet.* 2006;123:74–80.
17. van der Werf JHJ, Goddard ME, Meyer K. The use of covariance functions and random regressions for genetic evaluation of milk production based on test day records. *J Dairy Sci.* 1998;81:3300–8.
18. Jamrozik J, Schaeffer LR, Dekkers JCM. Genetic evaluation of dairy cattle using test day yields and random regression model. *J Dairy Sci.* 1997;80:1217–26.
19. Sargent FD, Lytton VH, Wall OG. Test interval method of calculating dairy herd improvement association records. *J Dairy Sci.* 1968;51:170–9.
20. Meyer K. WOMBAT—A tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). *J Zhejiang Univ Sci B.* 2007;8:815–21.
21. Schaeffer LR. Random regression models; 2016. <http://animalbioscienc.es.uoguelph.ca/~lrs/BOOKS/rrmbook.pdf> Accessed 3 July 2019.
22. Hammami H, Reik B, Soyeurt H, Gara AB, Gengler N. Genetic parameters for Tunisian Holsteins using a test-day random regression model. *J Dairy Sci.* 2008;91:2118–26.
23. Leclerc H. Development of the French dairy cattle test-day model genetic evaluation and prospects of using results for herd management. Ph.D. thesis, AgroParisTech; 2008.
24. Olori VE, Hill WG, McGuirk BJ, Brotherstone S. Estimating variance components for test day milk records by restricted maximum likelihood with a random regression animal model. *Livest Prod Sci.* 1999;61:53–63.
25. Togashi K, Lin CY. Selection for milk production and persistency using eigenvectors of the random regression coefficient matrix. *J Dairy Sci.* 2006;89:4866–73.
26. Pool MH, Janss LLG, Meuwissen THE. Genetic parameters of Legendre polynomials for first parity lactation curves. *J Dairy Sci.* 2000;83:2640–9.
27. Rupp R, Clément V, Piacere A, Robert-Granié C, Manfredi E. Genetic parameters for milk somatic cell score and relationship with production and udder type traits in dairy Alpine and Saanen primiparous goats. *J Dairy Sci.* 2011;94:3629–34.
28. Leclerc H, Nagy I, Ducrocq V. Impact of using reduced rank random regression test-day model on genetic evaluation. *Interbull Bull.* 2009;40:42–6.
29. Bélichon S, Manfredi E, Piacère A. Genetic parameters of dairy traits in the Alpine and Saanen goat breeds. *Genet Sel Evol.* 1999;31:52934.
30. Negussie E, Strandén I, Mäntysaari EA. Genetic associations of test-day fat:protein ratio with milk yield, fertility, and udder health traits in Nordic Red cattle. *J Dairy Sci.* 2013;96:1237–50.
31. Buttchererit N, Stamer E, Junge W, Thaller G. Genetic relationships among daily energy balance, feed intake, body condition score, and fat to protein ratio of milk in dairy cows. *J Dairy Sci.* 2011;94:1586–91.

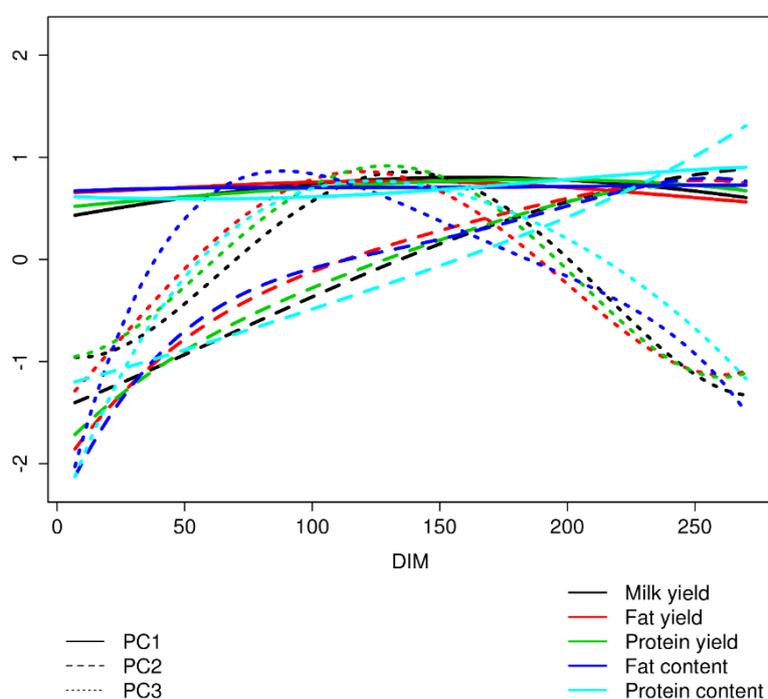
**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Additional files

Additional file 1: Table S1: Proportion (%) of genetic variance explained by the first three principal components (PC) using leg4

	Milk yield		Fat yield		Protein yield		Fat content		Protein content	
	Saanen	Alpine	Saanen	Alpine	Saanen	Alpine	Saanen	Alpine	Saanen	Alpine
<b>PC1</b>	87.5	84.4	87.6	83.1	89.2	85.4	92.7	90.7	90.3	91.7
<b>PC2</b>	10.6	12.1	11	12.3	9.1	9.9	4.8	6.6	7.5	6.9
<b>PC3</b>	1.5	2.9	1.1	4.2	1.2	4	1.8	2.2	1.8	1
<b>sum of first 2 PC</b>	98.1	96.5	98.6	95.4	98.3	95.3	97.5	97.3	97.8	98.6
<b>sum of first 3 PC</b>	99.6	99.4	99.7	99.6	99.5	99.3	99.3	99.5	99.6	99.6

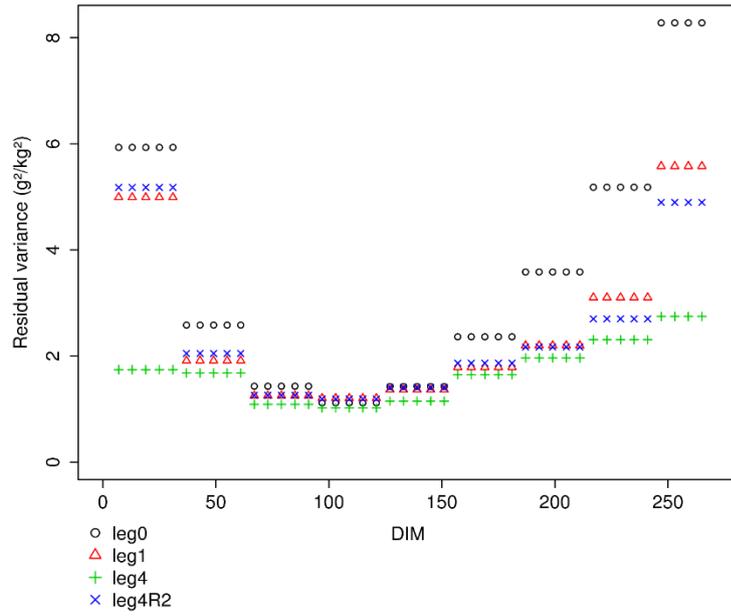


Additional file 2: Figure S1 Coordinates of the first three eigenvectors of the test-day genetic covariance matrix computed with leg4 in Saanen goats

Additional file 3: Table S2: Pearson correlation coefficients ( $\rho$ ) between observed and predicted values with complete and reduced models

		Saanen					Alpine				
		Milk yield	Fat yield	Protein yield	Fat content	Protein content	Milk yield	Fat yield	Protein yield	Fat content	Protein content
Complete model	MT	0.81	0.83	0.82	0.85	0.88	0.81	0.81	0.80	0.85	0.90
	leg0	0.91	0.89	0.90	0.87	0.90	0.91	0.88	0.90	0.87	0.91
	leg1	0.94	0.92	0.93	0.90	0.93	0.94	0.92	0.93	0.89	0.94
	leg2	0.96	0.94	0.95	0.91	0.95	0.95	0.93	0.94	0.90	0.96
	leg3	0.96	0.94	0.96	0.92	0.96	0.96	0.94	0.95	0.92	0.97
	leg4	0.96	0.95	0.96	0.93	0.97	0.96	0.95	0.96	0.93	0.97
Reduced model	leg2r2	0.94	0.93	0.94	0.90	0.93	0.94	0.92	0.93	0.89	0.94
	leg3r2	0.94	0.93	0.94	0.90	0.93	0.94	0.92	0.93	0.89	0.95
	leg4r2	0.94	0.93	0.94	0.90	0.93	0.94	0.92	0.93	0.89	0.95
	leg3r3	0.96	0.94	0.95	0.92	0.95	0.96	0.94	0.94	0.91	0.96
	leg4r3	0.96	0.94	0.95	0.92	0.95	0.96	0.94	0.94	0.91	0.96

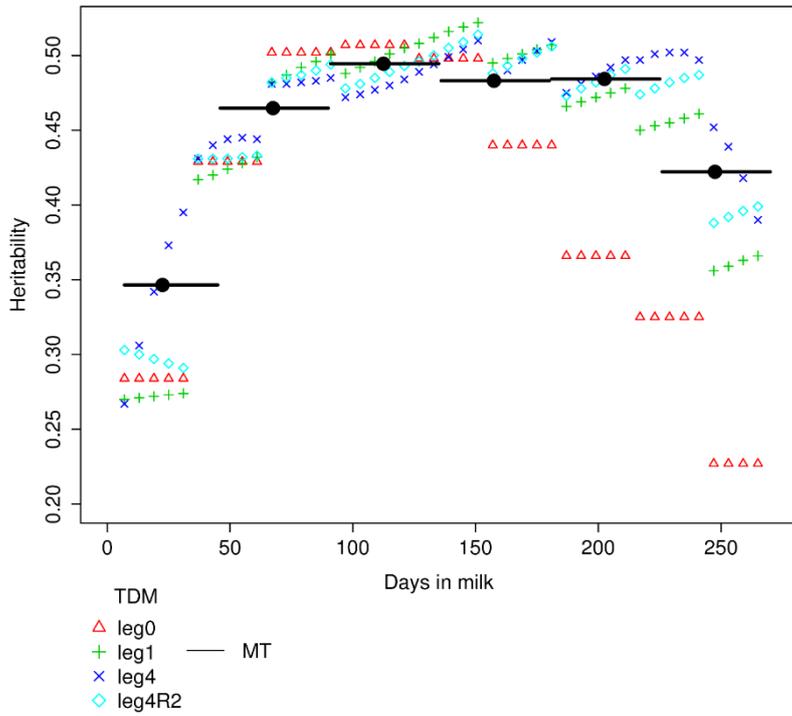
(MT: multitrait model, leg $x$ : Legendre polynomial of order  $x$ , leg $x$ R $z$ : Legendre polynomial of order  $x$  reduced to order  $z$ )



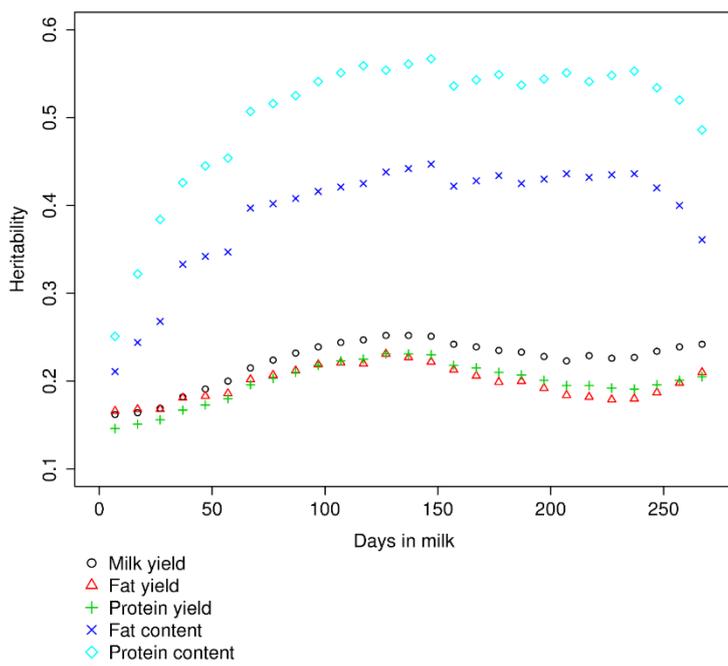
Additional file 4: Figure S2: Evolution of residual variances according to DIM for protein content in Alpine goats (full-rank models: leg0, leg1, leg2, leg4; reduced model: leg4R2)

Additional file 5: Table S3: Whole-lactation heritabilities from LACT and heritabilities ( $h_{wl}$ ) calculated as in [22] in both Saanen-breed and Alpine-breed goats.

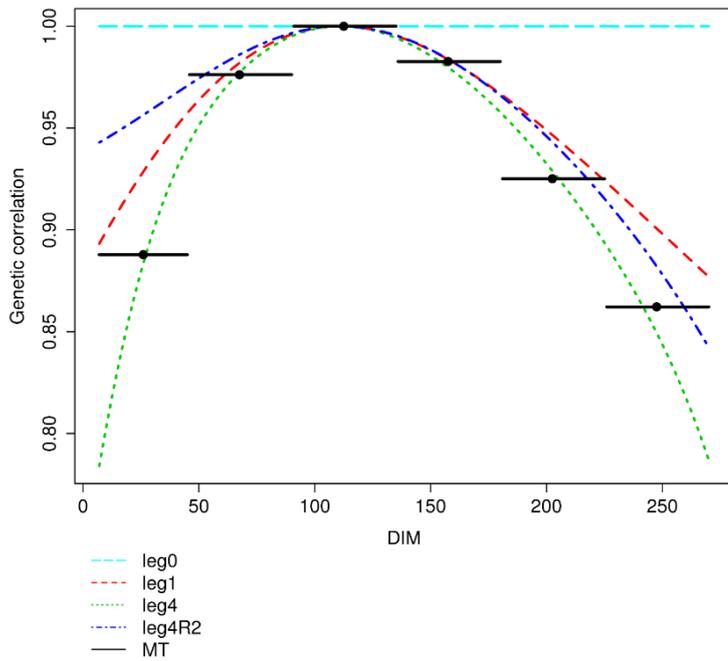
		Saanen					Alpine				
		Milk yield	Fat yield	Protein yield	Fat content	Protein content	Milk yield	Fat yield	Protein yield	Fat content	Protein content
	<b>Lact</b>	0.29	0.33	0.30	0.6	0.62	0.26	0.26	0.25	0.65	0.66
<b>Reduced model</b>	<b>Complete model</b>										
	<b>Leg0</b>	0.32	0.39	0.35	0.69	0.64	0.3	0.32	0.3	0.77	0.7
	<b>Leg1</b>	0.32	0.38	0.35	0.68	0.65	0.29	0.31	0.29	0.77	0.7
	<b>Leg2</b>	0.32	0.38	0.34	0.68	0.66	0.29	0.3	0.29	0.76	0.71
	<b>Leg3</b>	0.31	0.37	0.34	0.68	0.65	0.29	0.3	0.29	0.76	0.71
	<b>Leg4</b>	0.31	0.37	0.34	0.68	0.65	0.29	0.3	0.29	0.75	0.7
	<b>Leg2r2</b>	0.32	0.38	0.34	0.68	0.65	0.31	0.31	0.3	0.77	0.71
	<b>Leg3r2</b>	0.32	0.38	0.34	0.69	0.65	0.31	0.32	0.3	0.77	0.71
	<b>Leg4r2</b>	0.32	0.38	0.34	0.69	0.65	0.31	0.32	0.3	0.77	0.71
	<b>Leg3r3</b>	0.32	0.38	0.34	0.68	0.66	0.3	0.3	0.29	0.76	0.71
<b>Leg4r3</b>	0.32	0.38	0.34	0.68	0.66	0.3	0.31	0.29	0.76	0.71	



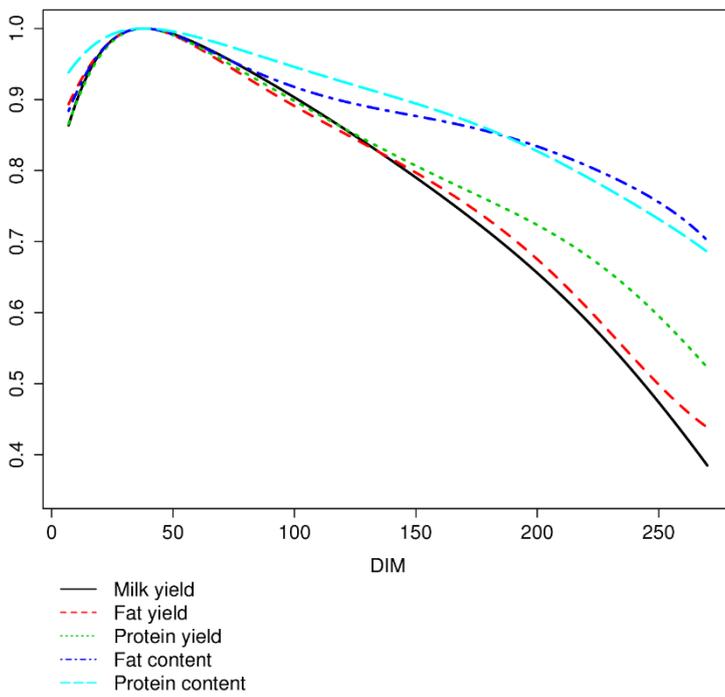
Additional file 6: Figure S3: Daily estimated heritabilities for protein content in Saanen goats (full-rank models: leg0, leg1, leg4; best reduced model: leg4R2 and the MT model)



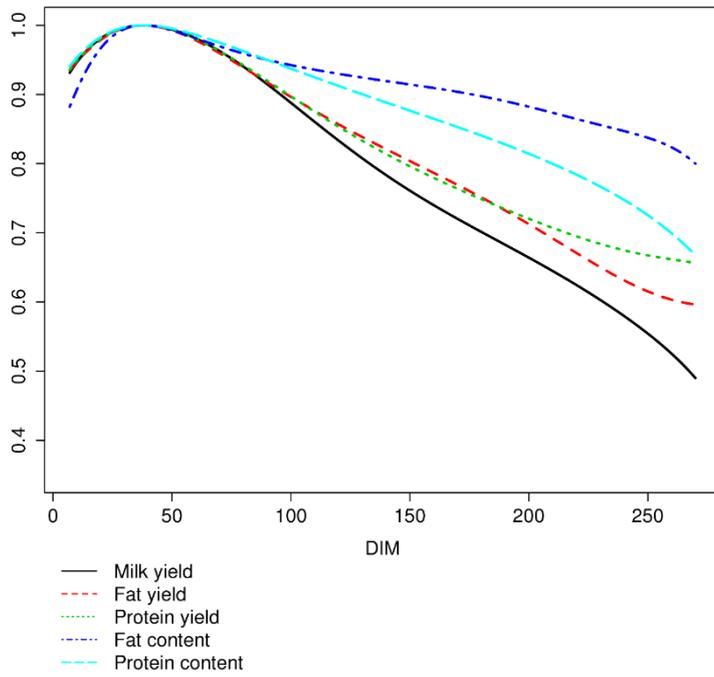
Additional file 7: Figure S4: Evolution of heritability according to DIM in Alpine-breed goats with the leg4 model



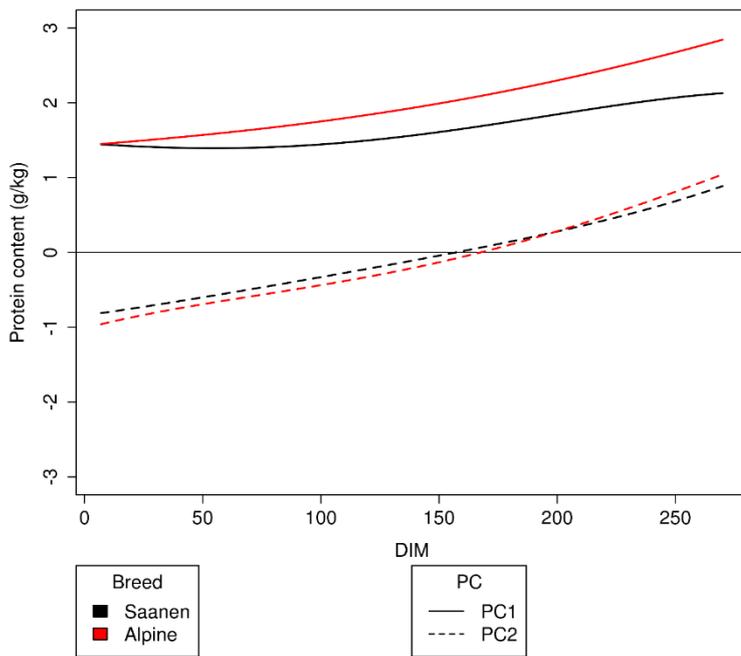
Additional file 8: Figure S5: Genetic correlations between DIM 111 and other DIMs for protein content in Saanen-breed goats.



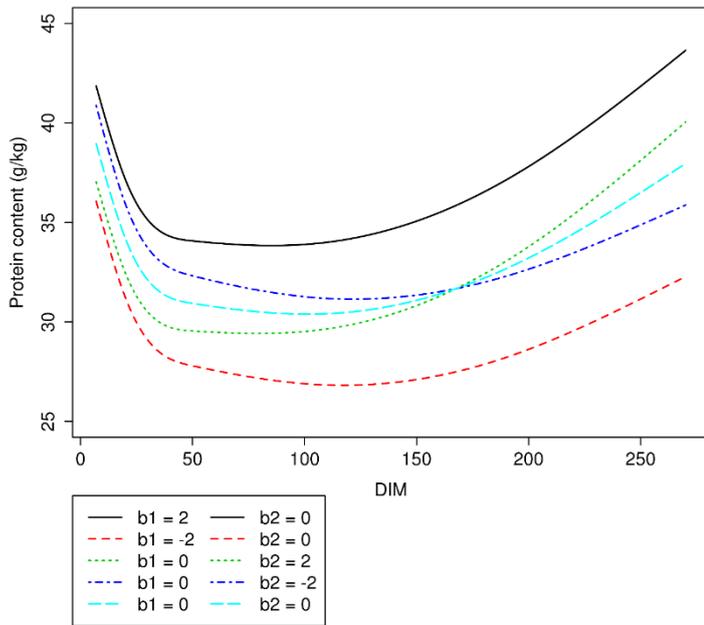
Additional file 9: Figure S6: Genetic correlations between DIM 40 and other DIMs in Alpine-breed goats from the leg4 model.



Additional file 10: Figure S7: Genetic correlations between DIM 40 and other DIMs in Saanen-breed goats from the leg4 model.



Additional file 11: Figure S8: Contribution to daily protein content of 1 genetic standard deviation for PC1 and PC2



Additional file 12: Figure S9: EBV of Alpine goats added to the mean production of the population for protein content (sigma  $b_1$ :  $b_1$  expressed in standard deviation, sigma  $b_2$ :  $b_2$  expressed in standard deviation).

## II. Analyses et résultats complémentaires

### 1. L'estimation de la production à la lactation par la méthode Fleischmann est-elle comparable à celle issue d'un modèle de régression aléatoire (RRM)?

Nous avons comparé des modèles RRM au modèle d'évaluation à la lactation qui utilise une production laitière totale phénotypique calculée par la méthode de Fleischmann. En effet jusqu'à aujourd'hui, nous manquions d'éléments pour apprécier la précision d'estimation de la méthode de Fleischmann sur le calcul de la production totale. Pour cela, nous avons comparé la production de lait calculée par la méthode de Fleischmann à une production calculée et modélisée selon un RRM basé sur une spline naturelle à 6 nœuds. Nous avons utilisé les données de chèvres Saanen en première lactation entre 2000 et 2018 ce qui représentait 3 363 167 contrôles élémentaires pour 526 410 chèvres. Nous avons sélectionné les lactations avec plus de 3 contrôles élémentaires afin d'être dans une situation comparable à celle de l'utilisation de la méthode de Fleischmann. En effet, la méthode de Fleischmann est appliquée uniquement lorsque la lactation comporte suffisamment de contrôles.

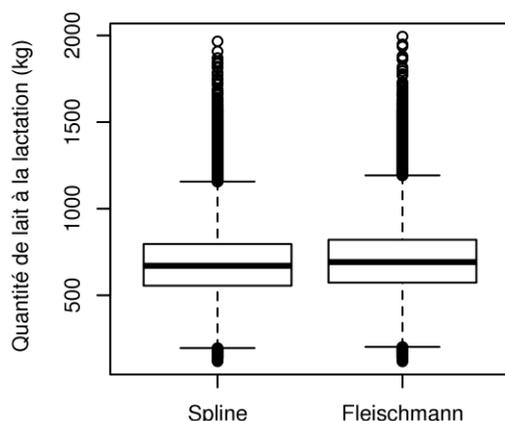
Le modèle RRM était le suivant :

$$y_{dk} = \sum_{o=1}^6 \theta_{ko} N_{(o,d)} + e_{kd} \text{ (Modèle 1)}$$

avec  $y_{dk}$  la production laitière de la chèvre  $k$  pour le stade de lactation  $d$  (7, ... , 270),  $N_{(o,d)}$  est la  $o^{ième}$  covariable pour le stade de lactation  $d$  pour la spline cubique naturelle avec 6 nœuds à  $d = 7, 20, 50, 110, 190, 270$ ,  $\theta_{ko}$  est le  $o^{ième}$  coefficient de régression aléatoire et  $e_{kd}$  le résidu supposé normalement distribué selon une loi normale centrée et de variance homogène.

A partir des coefficients de régression obtenus avec le modèle 1, nous avons calculé la production de lait totale de chaque lactation. La corrélation entre la production de lait obtenue avec le modèle 1 et celle calculée par la méthode de Fleischmann était de 0.99.

Nous avons comparé la distribution des productions à la lactation calculées selon la méthode de Fleischmann et celles selon le modèle 1 (Figure 19). Les moyennes sont très proches, même si celle avec la méthode de Fleischmann est significativement supérieure (703.3kg pour Fleischmann contre 681.9kg pour le modèle 1). L'interpolation de Fleischmann est donc précise et estime un phénotype proche de celui que l'on peut calculer à l'aide de splines. Ces résultats sont proches de ceux d'Anderson et al. (1989) en bovins laitiers qui trouvaient des résidus pour l'estimation de la production à la lactation proches entre la méthode d'interpolation de Fleischmann et l'utilisation de courbes paramétriques (courbes de Wood (1967)). La méthode d'interpolation prend bien en compte la forme de la courbe de lactation de l'animal.



**Figure 19 : Distribution des productions de lait totales à la lactation calculées selon la méthode Fleischmann et avec l'utilisation de splines (modèle 1)**

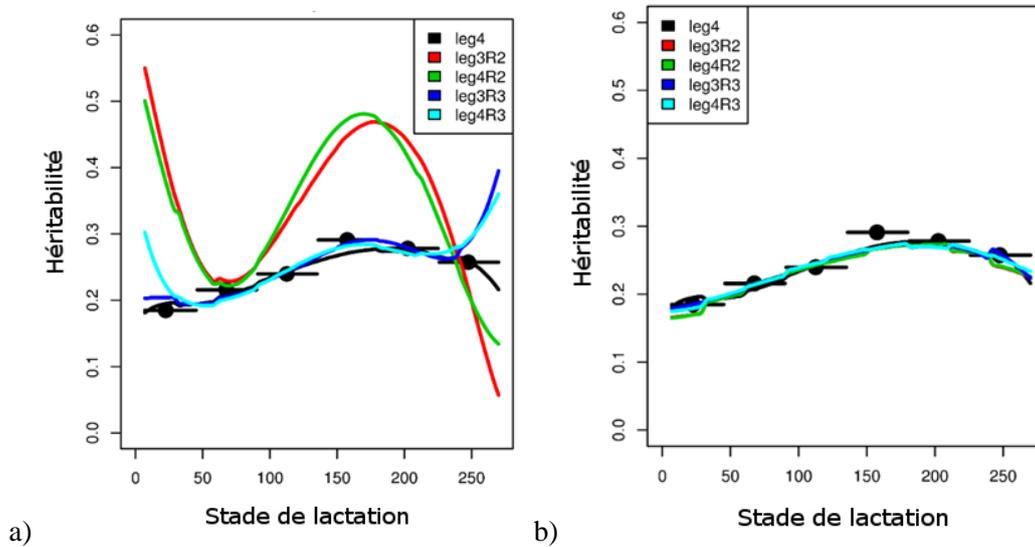
## **2. Difficultés rencontrées lors de la réduction des matrices de variances-covariances génétique et d'environnement permanent**

Lorsque nous avons voulu réduire les matrices de variances-covariances génétique et de l'environnement permanent, nous avons été confronté à quelques difficultés liées à l'utilisation du logiciel Wombat (Meyer, 2007). En premier lieu, nous avons eu des problèmes en utilisant la réduction de ces matrices directement fournies par le logiciel, et dans un deuxième temps voulant utiliser des vecteurs propres spécifiques à chacun des deux effets aléatoires.

Le logiciel Wombat propose une réduction de la matrice de variances-covariances. Malheureusement, lorsque nous l'avons utilisée, l'option n'a pas fourni les résultats escomptés par rapport au modèle non réduit. La Figure 20 présente les héritabilités obtenues lorsque la réduction était réalisée par le logiciel Wombat (Figure 20.a), ou par une étape extérieure au logiciel Wombat avec le logiciel R (Figure 20.b) en utilisant simplement la fonction « eigen ». Les héritabilités estimées lorsque la réduction des matrices de variances-covariances est réalisée par logiciel Wombat sont très souvent éloignées de celles estimées avec le modèle complet de référence leg4 (Figure 20.a) ou encore avec le modèle multi caractère (Figure 20.a). En revanche, lorsque la réduction est réalisée hors du logiciel Wombat, l'héritabilité suit bien l'évolution tout au long de la lactation de l'héritabilité estimée par le modèle complet (Figure 20.b) et par le modèle multi caractère.

Les problèmes rencontrés avec le logiciel Wombat ont fait l'objet de plusieurs échanges avec sa conceptrice K. Meyer (Meyer and Kirkpatrick, 2008). K. Meyer nous a expliqué que ces problèmes provenaient de la tentative de réduction des deux matrices de variances-covariances en même temps. Nous avons donc décidé de réaliser la réduction des deux matrices en dehors du logiciel Wombat et de donner ensuite les vecteurs propres calculés comme nouvelles fonctions (dites fonctions utilisateur) en paramètres

d'entrée au logiciel Wombat, étant donné les résultats beaucoup plus satisfaisant obtenus ainsi avec notre modèle.



**Figure 20 : Evolution de l'héritabilité selon le stade de lactation pour la production laitière chez les chèvres de race Saanen en première lactation avec une réduction des de variances- covariances génétique et de l'environnement permanent : a) par le logiciel Wombat, b) hors du logiciel Wombat avec R.**

Le logiciel Wombat ne permettait pas d'utiliser deux fonctions utilisateur différentes pour les effets aléatoires (génétique d'une part et environnement permanent d'autre part). Nous avons décidé d'utiliser les vecteurs propres issu de la matrice génétique comme fonction pour modéliser aussi l'effet d'environnement permanent. Au préalable, nous avons cherché à estimer l'impact de cette pratique sur les paramètres génétiques estimés par le modèle.

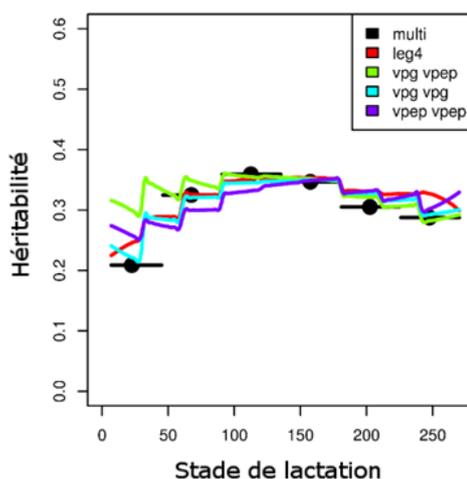
Pour cela, nous avons utilisé le logiciel AIREMLF90 (Miszta et al., 2002), avec lequel il est possible d'utiliser des vecteurs propres spécifiques à chacun des effets aléatoires génétique ou d'environnement permanent. Le TB a été choisi comme caractère de test car les vecteurs propres génétiques et de l'environnement permanent sont des vecteurs avec des différences relativement importantes comparés aux autres caractères. La Figure 21 présente l'évolution de l'héritabilité selon le stade de lactation pour le TB en Saanen en primipares selon différents modèles : le multi caractère : multi (appelé MT dans l'article 2), le modèle de régression aléatoire de référence leg4 (appelé aussi leg4 dans l'article 2 : modèle de référence), ainsi que les modèles avec réduction des matrices de variances-covariances :

- vpg vpep : en utilisant pour chaque effet une réduction spécifique et leurs vecteurs propres associés: vpg pour la valeur génétique et vpep pour l'effet d'environnement permanent ;

- vpg vpg : en utilisant pour les deux effets les vecteurs propres obtenus par réduction de la matrice de variances-covariances génétique (correspondant au modèle leg4R2 de l'article 2),

- vpep vpep : en utilisant pour les deux effets les vecteurs propres obtenus par réduction de la matrice de variances-covariances de l'environnement permanent

Les héritabilités estimées avec le modèle « vpg vpg » (leg4R2 dans l'article 2) sont très proches de celles estimées à l'aide du modèle complet « leg4 ». Ces estimées sont plus proches de celles obtenues avec le modèle de référence que ne le sont celles calculées avec les modèles « vpg vpep » et « vpep vpep ». Les mêmes observations ont été faites en comparant les variances génétiques. Nous avons conclu de ces résultats que l'utilisation des vecteurs propres génétiques pour modéliser aussi bien l'effet génétique que l'effet d'environnement permanent était satisfaisante afin d'avoir une bonne estimation des valeurs génétiques des animaux, ce qui nous intéresse le plus dans un contexte de sélection animale.



**Figure 21 : Evolution de l'héritabilité selon le stade de lactation pour le TB chez les primipares de race Saanen selon différents modèles**

*multi (multi caractère appelé MT dans l'article 2), leg4 (appelé leg4 dans l'article 2, modèle de référence), vpg vpep (modèle utilisant les vecteurs propres de la matrice de variances-covariances génétique pour modéliser la génétique et les vecteurs propres de la matrice de variances-covariances de l'environnement permanent pour modéliser l'environnement permanent), vpg vpg (modèle utilisant les vecteurs propres de la matrice de variances-covariances génétique pour modéliser la génétique et l'environnement permanent correspondant au modèle leg4R2 de l'article 2), vpep vpep (modèle utilisant les vecteurs propres de la matrice de variances-covariances de l'environnement permanent pour modéliser la génétique et l'environnement permanent)*

### 3. Définition de l'effet troupeau (HTD)

Dans les modèles de régression aléatoire pour l'analyse de la production laitière, la définition du principal effet d'environnement du modèle, l'effet troupeau-jour de contrôle, soulève certaines questions. Dans son livre de 2016, L. Schaeffer explique que l'effet troupeau-jour de contrôle (HTD) ne devrait pas être utilisé car il ne prend pas en compte le fait que toutes les chèvres lors du contrôle ne sont pas au même

stade de lactation (car elles n'ont pas mis bas toutes le même jour), ce qui entache l'effet HTD d'imprécision. Il est aussi opposé à l'utilisation d'un effet fixe pour l'effet des contemporaines et préconise un effet troupeau-saison de mise bas-année aléatoire fonction du stade de lactation. En effet, il considère que les effets fixes doivent être réservés à des effets avec peu de niveaux et des effets dont on veut connaître la valeur de l'effet spécifique. En revanche, un effet aléatoire doit être utilisé quand on s'intéresse à la part de la variance phénotypique représentée par un effet, que nous avons une grande quantité de niveaux pour l'effet étudié et que nous ne sommes pas intéressés par l'effet d'un niveau précis (Schaeffer, 2018). Afin de prendre en compte ces remarques, nous avons testé 2 modèles dits de Schaeffer (Sch\_or et Sch\_HY) que nous avons comparés au modèle de référence de notre étude « leg4 ».

Pour ces 2 modèles, nous avons redéfini l'effet âge à la mise bas en âge à la mise bas\*saison de mise bas\*années regroupées par 5 ans. De plus, un effet troupeau\*année\*saison a été modélisé en fonction du stade de lactation par un polynôme de Legendre de degré 4.

Pour le modèle (Sch\_HY), un effet troupeau\*année (HY<sub>ji</sub>), en effet fixe, a été conservé alors que le modèle (Sch\_or) ne comporte pas d'autres effets liés au troupeau. Le modèle le plus complet (Sch\_HY) s'écrit donc :

$$y_{ijklm d g n t} = HY_{ji} + AMY_{tkm} + MY_{lj} + \sum_{r=1}^6 \theta_{tkm} N_{(r,d)} + \sum_{r=1}^6 \tau_{lj} N_{(r,d)} + \sum_{s=1}^4 \gamma_s M_{(s,g)} + \sum_{o=0}^4 a_{no} \varphi_{(o,d)} + \sum_{o=0}^4 p_{no} \varphi_{(o,d)} + \sum_{o=0}^4 h_{jio} \varphi_{(o,d)} + e_{ijklm d g n t}$$

où  $y_{ijklm d g n t}$  est l'observation pour le stade de lactation  $d$  (du 7 au 270<sup>ième</sup> jour de contrôle) de la chèvre  $n$  gestante de  $g$  jours ( $g = 0$  si pas gestante), produisant l'année  $j$  (de 1995 à 2015), durant la période de 5 ans  $m$  (4 classes: 1995-2000, 2000-2005, 2005-2010, 2010-2015), appartenant à la classe d'âge à la mise bas  $k$  (7 classes [en mois]: 9–11, 12, 13, 14, 15, 16, 17 ou plus), qui a mis bas durant la classe  $l$  (8 classes: janvier, février, mars, avril – mai – juin, juillet – août – septembre, octobre, novembre, décembre), durant la saison  $t$  (2 classes: janvier-juin, juillet-décembre), et dans le troupeau  $i$ .  $HY_{ji}$  est l'effet troupeau année;  $AMY_{tkm}$  est l'effet âge à la mise bas\*saison\*période de 5 ans ;  $MY_{lj}$  est l'effet mois de mise bas\*période de 5 ans;  $\theta_{tkr}$  et  $\gamma_s$  sont les coefficient de régression pour la période\*mois de mise bas\*âge à la mise bas, et le stade de gestation, respectivement.  $N_{(r,d)}$  est la covariable  $r$  pour le stade de lactation  $d$  de la spline cubique naturelle avec 6 nœuds à  $d = 7, 20, 50, 110, 190, 270$ ;  $M_{(s,g)}$  est la covariable  $s$  au stade de lactation  $d$  de la spline cubique naturelle avec 4 nœuds à  $d = 31, 53, 76, 100$  (et  $g$  entre 0 et 30 jours, les coefficient ont été mis à 0, ce qui implique qu'il n'y avait pas d'effet de gestation sur la production. Enfin, dû au manque de données, l'effet de la gestation était considéré comme constant après  $g = 100$ ).  $a_{no}$  est le coefficient de régression aléatoire de la valeur génétique additive qui suit une loi normale ( $\text{Var}(a_o) \sim N(0, \sigma_{a_o}^2)$ ) pour la  $o^{\text{ième}}$  covariable de Legendre de degré  $q$ ,  $p_{no}$  est le coefficient de régression aléatoire de l'environnement permanent qui suit une loi normale ( $\text{Var}(p_o) \sim N(0, \sigma_{p_o}^2)$ ) pour la  $o^{\text{ième}}$  covariable de Legendre de degré  $q$ ,  $h_{jio}$  est le coefficient de régression aléatoire pour l'effet troupeau-année

saison (1769 modalités) qui suit une loi normale ( $\text{Var}(p_o) \sim N(0, \sigma_{p_o}^2)$ ) pour la  $o^{\text{ième}}$  covariable de Legendre de degré  $q$ ,  $\varphi_{(o,d)}$  est la valeur de la  $o^{\text{ième}}$  covariable de Legendre pour le stade de lactation  $d$  et  $e_{ijklmdgnt}$  est le résidu qui suit une loi normale. La matrice de variances-covariances résiduelle est supposée diagonale avec des variances hétérogènes tout au long de la lactation et en étant constante au cours de 9 classes de jour de lactations (7–36, 37–66, 67–96, 97–126, 127–156, 157–186, 187–216, 217–246, 247–270).

Le Tableau 4 présente les matrices de variances-covariances des effets aléatoires des modèles leg4, Sch\_or et Sch\_HY. Les valeurs des variances et des covariances sont très proches entre les 3 différents modèles pour la génétique et l'environnement permanent. En revanche, pour l'effet troupeau\*année\*saison, lorsque l'effet troupeau-année est ajouté en effet fixe (HY), la variance estimée pour son premier coefficient de Legendre est proche de 0 dans le modèle Sch\_HY, en accord avec Leclerc (2008), alors qu'elle est égale à 37 pour le modèle Sch\_or. Pour les quatre autres coefficients les variances sont très proches entre les deux modèles.

Le Tableau 5 présente les corrélations entre les coefficients des polynômes de Legendre pour les effets génétiques, de l'environnement permanent, et l'effet troupeau-année-saison obtenus avec les différents modèles. Les coefficients sont très proches entre modèles avec des corrélations supérieures à 0.98 pour le premier coefficient du polynôme ( $a_0$ ) pour la génétique et l'environnement permanent. Pour les coefficients des ordres supérieurs, représentant peu de part de variance, les corrélations sont plus faibles mais supérieures à 0.86. En revanche, la corrélation entre les premiers coefficients des polynômes de Legendre de l'effet troupeau-année-saison entre les deux modèles de Schaeffer est faible (0.28), mais elle est égale à 0.99 pour les autres coefficients de cet effet.

Comme le montre le Tableau 6, le modèle (leg4) utilisé dans notre étude estime les variances résiduelles les plus faibles et a un BIC plus faible que les modèles dits de Schaeffer. Ce modèle utilise moins de paramètres ce qui est un avantage pour la suite de nos études, lorsqu'il faudra incorporer les performances des chèvres de rang de lactation 2 et 3. La proximité des paramètres génétiques et des index estimés avec ce modèle et avec les modèles dits de Schaeffer, nous incite à penser que la modélisation choisie pour l'effet troupeau par un effet HTD dans le modèle leg4 est relativement robuste et n'est pas à remettre en cause par la suite.

**Tableau 4 : Matrices de variance –covariance génétique, de l’environnement permanent, et de l’effet troupeau-année pour les différents modèles : Sch\_or (modèle de Schaeffer), Sch\_HY (modèle de Schaeffer comprenant l’effet troupeau-année), leg4 (modèle de l’article 2, avec un effet troupeau-jour de contrôle)**

	Degré du Polynôme	Sch_or					Sch_HY					leg4				
		a_0	a_1	a_2	a_3	a_4	a_0	a_1	a_2	a_3	a_4	a_0	a_1	a_2	a_3	a_4
<b>Matrice de variance – covariance génétique</b>	<b>a_0</b>	18.76					16.86					16.6				
	<b>a_1</b>	1.15	2.12				1.2	2.14				1.08	2.1			
	<b>a_2</b>	-2.19	-0.41	0.58			-2.06	-0.43	0.58			-1.92	-0.37	0.54		
	<b>a_3</b>	-0.03	-0.12	0.03	0.04		-0.01	-0.12	0.03	0.04		-0.04	-0.07	0.02	0.05	
	<b>a_4</b>	-0.09	0.01	-0.05	0	0.06	-0.06	0.01	-0.05	0	0.06	0.03	0	-0.06	-0.01	0.04
<b>Matrice de variance – covariance de l’environnement permanent</b>	<b>a_0</b>	34.45					35.69					36.17				
	<b>a_1</b>	-0.28	5.34				-0.29	5.33				-0.27	5.5			
	<b>a_2</b>	-2.87	-0.03	2.05			-2.97	-0.03	2.05			-3.15	-0.15	2.33		
	<b>a_3</b>	0.86	-0.68	-0.15	0.83		0.84	-0.68	-0.15	0.83		0.8	-0.7	-0.17	1.08	
	<b>a_4</b>	-0.29	0.11	-0.34	0.01	0.19	-0.32	0.11	-0.34	0.01	0.19	-0.33	0.08	-0.39	-0.03	0.41
<b>Matrice de variance-covariance de l’effet troupeau-année</b>	<b>a_0</b>	37.4					3.36									
	<b>a_1</b>	2.51	3.95				-0.5	3.97								
	<b>a_2</b>	-2.96	0.17	1.73			-0.99	0.15	1.77							
	<b>a_3</b>	0.8	-0.08	0.08	1.13		-0.28	-0.07	0.06	1.12						
	<b>a_4</b>	-0.74	-0.02	-0.01	0.08	0.68	-0.47	-0.03	0.01	0.06	0.68					

**Tableau 5 : Corrélations entre les coefficients de régression des covariables de Legendre pour les effets génétiques, de l'environnement permanent et du troupeau-année-saison des différents modèles : Sch\_or (modèle de Schaeffer), Sch\_HY (modèle de Schaeffer comprenant l'effet troupeau-année), leg4 (modèle de l'article 2, avec un effet troupeau-jour de contrôle)**

	Modèles	a_0	a_1	a_2	a_3	a_4
Corrélations pour les effets génétiques	leg4/ Sch_or	0.98	0.96	0.98	0.86	0.86
	leg4/ Sch_HY	0.99	0.96	0.98	0.86	0.87
	Sch_or/Sch_HY	0.99	1	1	1	1
Corrélations pour les effets de l'environnement permanent	leg4/ Sch_or	0.99	0.98	0.97	0.95	0.91
	leg4/ Sch_HY	0.99	0.97	0.97	0.95	0.91
	Sch_or/Sch_HY	1	1	1	1	1
Corrélations pour les effets du troupeau-année-saison	Sch_or/Sch_HY	0.28	0.99	0.99	0.99	0.99

**Tableau 6: Variances résiduelles et BIC pour les différents modèles : : Sch\_or (modèle de Schaeffer), Sch\_HY (modèle de Schaeffer comprenant l'effet troupeau-année), leg4 (modèle de l'article 2, avec un effet troupeau-jour de contrôle)**

	Périodes de stade de lactation	Sch_or	Sch_HY	leg4
Variances résiduelles par classes	7–36	12.22	12.17	10.78
	37–66	11.37	11.35	9.98
	67–96	10.58	10.57	9.19
	97–126	10.04	10.04	8.75
	127–156	10.05	10.05	8.6
	157–186	10.49	10.5	8.82
	187–216	10.31	10.3	8.85
	217–246	9.09	9.09	7.78
	247–270	7.4	7.4	6.67
<b>BIC</b>		793,349.29	787,498.21	755,966.61

Dans ce chapitre, nous avons développé un modèle de régression aléatoire adapté aux données caprines en première lactation pour le lait, la MG, la MP, le TB, et le TP. Nous avons vu comment définir de manière robuste l'effet troupeau à inclure dans le modèle, comment réduire les temps de calcul par une réduction des matrices de variances-covariances et comment obtenir des index facilement interprétables. Nous avons également comparé les index obtenus avec le modèle RRM au modèle utilisé actuellement en routine pour l'évaluation génétique et nous avons obtenu des index très proches. Le modèle retenu permet un bon ajustement aux données, avec des temps de calcul acceptables, des paramètres génétiques similaires (héritabilités, variances génétiques, corrélations génétiques entre jours de lactation, corrélations entre index) à ceux obtenus avec un modèle plus complet. Les index génétiques produits par ce modèle sont facilement interprétables, résumant le niveau de production à la lactation et la persistance laitière de façon

décorrélée. L'index représentant le niveau de production à la lactation obtenu par notre modèle a une corrélation proche de 1 avec l'index calculé avec le modèle actuel de l'évaluation génétique. Après avoir développé un modèle adapté aux primipares, avec des temps de calcul réduit et une interprétation aisée des index calculés, nous voulons étendre ce modèle aux parités 2 et 3 afin d'obtenir des index qui représenteront la valeur génétique des animaux pour les trois premières lactations de la vie de l'animal.





## **Chapitre 3 : Estimation des paramètres génétiques des caractères de production laitière et des comptages de cellules somatiques des chèvres Alpine et Saanen françaises au cours des trois premières lactations à l'aide d'un modèle de régression aléatoire**

L'objectif de ce chapitre est d'adapter aux trois premiers rangs de lactation, le modèle de régression aléatoire développé pour les chèvres en première lactation dans le chapitre 2, pour les caractères de production laitière (quantité de lait, de matière grasse (MG), de matière protéique (MP), taux butyreux (TB) et protéique (TP)), les comptages de cellules somatiques (SCS) et le rapport TB/TP.

La première partie de ce chapitre est consacrée au développement du modèle de régression aléatoire pour les 3 premières lactations en race Alpine. Les résultats de cette étude font l'objet d'un article scientifique, en cours de révision, pour la revue Journal of Dairy Science. Les contrôles élémentaires de 49 849 chèvres de race Alpine suivies de leur première lactation à leur troisième lactation constituent la base des données de cette étude. Quatre modèles ont été comparés selon différents critères : la qualité de l'ajustement aux données, l'évolution des héritabilités et de la variance génétique au cours de la lactation et entre rangs de lactation, les corrélations génétiques entre stade de lactation des différentes parités ainsi que des corrélations entre index génétiques. Les modèles de régression aléatoire considérés utilisaient des polynômes de Legendre pour modéliser les effets génétiques et d'environnement permanent avec ou sans réduction des matrices de variances-covariances (EGV vs LEG). Nous avons également étudié l'impact de combiner ou non les parités 2 et 3 (PM vs FT) dans un même niveau de parité dans les modèles étudiés. 4 modèles ont ainsi été définis : LEG\_FT (sans réduction de la matrice de variances-covariances et avec les 3 rangs de lactation), EGV\_FT (avec réduction de la matrice de variances-covariances et les 3 rangs de lactation), LEG\_PM (sans réduction de la matrice de variances-covariances et en combinant dans un même niveau les rangs de lactation 2 et 3), et EGV\_PM (avec réduction de la matrice de variances-covariances et en combinant dans un même niveau les rangs de lactation 2 et 3). En considérant une réduction des matrices de variances-covariances de rang 2 (EGV) et en combinant les deuxièmes et troisièmes parités (PM), deux premières fonctions propres permettaient de capter correctement la variabilité génétique du niveau et de la persistance du rendement laitier, avec une corrélation proche de zéro entre leurs coefficients de régression. De plus, les temps de calcul avec ce modèle étaient relativement réduits par rapport au modèle de référence (EGV\_PM).

La deuxième partie de ce chapitre présente les résultats obtenus en race Saanen en suivant la même démarche que celle réalisée en race Alpine. Les données consistaient en des contrôles élémentaires de 60 071 chèvres de race Saanen suivies de leur première lactation à leur troisième lactation. La troisième partie regroupe les analyses et résultats supplémentaires menés lors de la mise au point des modèles de régression aléatoire pour les 3 parités, en particulier sur la définition de l'effet troupeau (résultats présentés en Saanen pour la quantité de lait) et sur l'ajustement des variances entre rang de lactation (résultats présentés pour les deux races et l'ensemble des caractères). La quatrième partie de ce chapitre présente les résultats complémentaires obtenus sur l'analyse de la partie génétique des modèles étudiés (pourcentage de variance représenté par les valeurs propres des matrices de variances-covariances génétiques, représentation des courbes phénotypiques pour des individus ayant des valeurs génétiques extrêmes) et sur l'analyse des effets d'environnement (évolution de la moyenne des HTD par mois, année, effets de l'âge à la mise bas, du mois de mise bas, de la durée de tarissement et de la gestation). Enfin, une étude des relations entre index (niveau, persistance des caractères de production et de comptage de cellules somatiques) et caractères de durabilité (i.e., durée de lactation, survie en fin de lactation, réussite à l'insémination animale) est présentée dans la dernière partie de ce chapitre.

## **I. Estimation des paramètres génétiques des caractères de production laitière et des comptages de cellules somatiques des trois premières lactations à l'aide d'un modèle de régression aléatoire chez les chèvres de race Alpine (Article 3)**

Cet article a été soumis à « Journal of Dairy Science » en Août 2019 et a été accepté début janvier 2020.

1 **Interpretive summary**

2 **Estimation of genetic parameters for dairy traits and somatic cell score in the first three parities**  
3 **using a random regression test-day model in French Alpine goats**

4 M. Arnal

5 Lactation curve shape is linked to an animal's health, feed requirements, and milk production  
6 throughout the year. We implemented a random regression model for the genetic evaluation of lactation  
7 curve shapes of dairy traits in French Alpine goats for their first three parities. A rank reduction of the  
8 variance-covariance matrix was performed using an eigenvalue decomposition for each parity.  
9 Combining the second and third parities, the first two principal components correctly summarized the  
10 genetic variability of milk yield level and persistency, with a near-nil correlation between the two.

11 **Running head: RANDOM REGRESSION IN FRENCH DAIRY GOATS**

12 **Estimation of genetic parameters for dairy traits and somatic cell score in the first three parities**  
13 **using a random regression test-day model in French Alpine goats**

14 M. Arnal<sup>1,2\*</sup>, H. Larroque<sup>1</sup>, H. Leclerc<sup>3</sup>, V. Ducrocq<sup>4</sup>, C. Robert-Granié<sup>1</sup>

15 <sup>1</sup>GenPhySE, Université de Toulouse, INRAE, ENVT, F-31326, Castanet Tolosan, France

16 <sup>2</sup>Institut de l'Élevage, Chemin de Borde Rouge, 31326 Castanet-Tolosan cedex, France

17 <sup>3</sup>Institut de l'Élevage, INRAE UMR1313 GABI, 78350 Jouy-en-Josas, France

18 <sup>4</sup>UMR1313 GABI, INRAE, AgroParisTech, Univ. Paris-Saclay, 78350 Jouy-en-Josas, France

19 \*Corresponding author: Mathieu Arnal, [mathieu.arnal@idele.fr](mailto:mathieu.arnal@idele.fr)

20

## 21           **ABSTRACT**

22           Lactation curve shape can affect an animal's health, feed requirements, and milk production  
23 throughout the year. We implemented a random regression model (**RRM**) for the genetic evaluation of  
24 lactation curve shapes of dairy traits in French Alpine goats for their first three parities. Milk, fat and  
25 protein yields, fat and protein contents, somatic cell score and fat/protein ratio (**FPR**) were considered.  
26 The data consisted of test-day (**TD**) records from 49,849 first lactation Alpine goats during their first  
27 three lactations. The reference model used a Legendre polynomial of order 2 for each parity to describe  
28 the genetic and permanent environmental effects, and was compared to a model that combined the  
29 second and third parities. A rank reduction of the variance-covariance matrix was also performed using  
30 an eigenvalue decomposition for each parity from the two models. Genetic parameters were consistent  
31 between the models tested. With a reduction to rank 2 and combining the second and third parities, the  
32 first two principal components correctly summarized the genetic variability of milk yield level and  
33 persistency, with a near-nil correlation between the two, and with a much shorter computation time than  
34 the reference model. A favorable correlation of +0.43 between milk yield persistency and FPR  
35 persistency at the beginning of the lactation was found from buck EBVs.

36           **Keywords:** random regression model, reduced rank matrix, genetic parameter estimation,  
37 persistency, dairy goat

## 38           **INTRODUCTION**

39           Random regression models (**RRMs**) are used for genetic evaluations of test-day (**TD**) milk  
40 production in dairy goats worldwide (Zumbach et al., 2008; Menéndez-Buxadera et al., 2010; Mucha et  
41 al., 2014). RRM allow more accurate breeding value predictions and fit the variability of environmental  
42 effects on lactation more closely (Schaeffer and Jamrozik, 2008).

43           RRMs allow the prediction of estimated breeding values (**EBVs**) for lactation persistency, based  
44 on the variation of genetic effects on lactation. A persistent animal is defined as one producing on  
45 average less milk at the beginning of lactation than animals with similar overall production, but more at  
46 the end (Cole and VanRaden, 2006). Lactation persistency interests dairy producers because the shape  
47 of the lactation curve affects the distribution of the farm's milk production during the year. In dairy  
48 cattle, several authors (Buttchereit et al., 2011; Jamrozik and Schaeffer, 2012; Koeck et al., 2013;  
49 Negussie et al., 2013) studied the fat/protein ratio (**FPR**) throughout lactation with an RRM. They  
50 showed that a high FPR at the beginning of the lactation was associated with a negative energy balance,  
51 subclinical mastitis and impaired fertility, and suggested that selecting on this criterion could help avoid  
52 these adverse traits.

53           In France, the implementation of an RRM in dairy goats has been studied in first parity in Alpine  
54 and in Saanen breeds (Arnal et al., 2019). However, in Germany, Zumbach et al. (2008) reported an  
55 average genetic correlation of 0.64 for milk yield between the first and second lactations, of 0.46

56 between the first and third lactations and of 0.75 between second and third lactations in German breeds  
57 dairy goats. From first to fourth parity, the goats produced more milk on average, but their lactations  
58 were less persistent in Alpine and in Saanen breeds (Arnal et al., 2018). This highlights in particular the  
59 marked difference in lactation production potential between goats in successive parities. An RRM for  
60 the first three parities therefore needs to be developed to correctly evaluate French goats throughout  
61 their lives.

62 One disadvantage of an RRM is its long computing time. However, the relatively high genetic  
63 correlation between second and third parity reported by Zumbach et al. (2008), together with those found  
64 by Leclerc (2008) in French dairy cattle, means that these two parities could be modeled together. With  
65 eigendecomposition, the rank of the variance-covariance matrix can be reduced by ignoring the  
66 contribution of the smallest eigenvalues (Van der Werf et al., 1998; Druet et al., 2003).  
67 Eigendecomposition and combining second and third parities would save computing time by reducing  
68 the number of genetic and permanent environment regression coefficients to be estimated.

69 This study set out to estimate genetic parameters for milk, fat and protein yields, and fat and  
70 protein contents, somatic cell score (SCS) and FPR using an RRM with Legendre polynomial functions,  
71 with and without rank reductions, and with and without combining second and third parities, to obtain  
72 EBVs for French Alpine dairy goats in their first three parities.

## 73 MATERIALS AND METHODS

### 74 *Data*

75 The data consisted of 692,743 TD records from 49,849 Alpine goats from the French North-  
76 West region collected between 1995 and 2015. The pedigree consisted of 98,075 animals. When  
77 possible, goats were followed during their first three parities. The details of the distribution of TD  
78 records and lactations per parity are given in Table 1. Each lactation included at least four TDs between  
79 day in milk (DIM) 7 and DIM 270. The lactation had to last for at least 180 days. Goats were milked  
80 twice a day and their records were summed to obtain their daily production. More than five animals per  
81 herd per parity and per test date were required. The sires had to be artificial insemination bucks (**AI-**  
82 **bucks**) and to have at least 20 progeny each in the dataset. There were 486. The dams of the goats had  
83 to be known. The first lactation had to be available to study the subsequent parities. The traits analyzed  
84 were milk yield, fat yield, protein yield, fat content, protein content, SCS and FPR. SCS were calculated  
85 from somatic cell count after Wiggans and Shook (1987).

### 86 *First to third parity (FT)*

87 *Legendre polynomial (LEG)*. A random regression model used in the first parity, described in  
88 Arnal et al. (2019), was extended to the first three parities. The three parities were analyzed

89 simultaneously in a sort of multiple-lactations model. The genetic and permanent environmental effects  
 90 were modeled by a Legendre polynomial of order 2 for each parity (**LEG\_FT**):

$$\begin{aligned}
 91 \quad y_{trsijklm d g n} = & HTD_{tri} + A_{tjk} + M_{trjl} + D_{trjm} + \sum_{o=1}^6 \theta_{tko} N_{(o,d)} + \sum_{o=1}^6 \tau_{trlo} N_{(o,d)} + \sum_{o=1}^6 \pi_{trmo} N_{(o,d)} \\
 92 \quad & + \sum_{o=1}^4 \gamma_{tso} M_{(o,g)} + \sum_{o=0}^2 a_{trno} \varphi_{(o,d)} + \sum_{o=0}^2 p_{trno} \varphi_{(o,d)} + e_{trsijklm d g n}
 \end{aligned}$$

93 where  $y_{trsijklm d g n}$  is the observation for trait  $t$  (milk yield, fat yield, protein yield, fat content,  
 94 protein content, SCS, FPR) of day in milk (DIM)  $d$  (7, ..., 270) of goat  $n$  carrying a kid for  $g$  day(s) (0,  
 95 ..., 100), in production year  $j$  (1995, ..., 2015), in parity  $r$  (1, 2, 3), in parity class  $s$  (primiparous (P) or  
 96 multiparous (M)), belonging to kidding age class  $k$  (7 classes for first lactation [in months]: 9-11, 12,  
 97 13, 14, 15, 16, +17; 6 classes for second lactation :21-23, 24, 25, 26, 27, +28; 6 classes for third lactation  
 98 : 31-35, 36, 37, 38, 39, +40), kidding month class  $l$  (7 classes: January, February, March-April-May,  
 99 June-July-August-September, October, November, December), class  $m$  of dry period length (6 classes  
 100 (days): [First parity], [0,50], [50,75], [75,100], [100,125], [125,+), herd  $\times$  test-date class  $i$ .  $HTD_{tri}$  is the  
 101 fixed effect for the herd test-date;  $A_{tjk}$  is the fixed effect for the age at kidding;  $M_{trjl}$  is the fixed effect  
 102 for the kidding period;  $D_{trjm}$  is the fixed effect for length of the dry period;  
 103  $\theta_{tko}$ ,  $\tau_{trlo}$ ,  $\pi_{trmo}$  and  $\gamma_{tso}$  are fixed regression coefficients for age at kidding, the kidding month, dry  
 104 period length and gestation stage respectively;  $N_{(o,d)}$  is the  $o^{\text{th}}$  covariate at time  $d$  of a cubic natural  
 105 spline function with 6 knots at  $d = 7, 20, 50, 110, 190, 270$ ;  $M_{(o,g)}$  is the  $o^{\text{th}}$  covariate at time  $d$  of a cubic  
 106 natural spline function with 4 knots at  $d = 31, 53, 76, 100$  (between  $g = 0$  and  $g = 30$ , the coefficients  
 107 were taken as equal to 0; if the gestation stage was greater than 100, it was taken as equal to 100);  
 108  $a_{trno}$  and  $p_{trno}$  are respectively the random additive genetic and permanent environmental regression  
 109 coefficients for the  $o^{\text{th}}$  term of a Legendre polynomial of order  $q$ ,  $\varphi_{(o,d)}$  is the value of the  $o^{\text{th}}$  term of  
 110 the Legendre polynomial at time  $d$  and  $e_{trsijklm d g n}$  is the residual term. The model assumed  
 111 heterogeneous residual variances modeled with nine classes within each parity: [7, 35], [36, 64], [65,  
 112 93], [94, 122], [123, 151], [152, 180], [181, 209], [210, 238], [239, 270].

113 The genetic variance-covariance matrix (C) for LEG\_FT was :

$$114 \quad C_{\text{LEG\_FT}} = \begin{bmatrix}
 \sigma_{a_{F\_P0}}^2 & \cdot & \sigma_{a_{F\_P0}a_{T\_P2}} \\
 \cdot & \sigma_{a_{F\_P1}}^2 & \cdot \\
 \cdot & \cdot & \sigma_{a_{F\_P2}}^2 & \cdot \\
 \cdot & \cdot & \cdot & \sigma_{a_{S\_P0}}^2 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
 \cdot & \cdot & \cdot & \cdot & \sigma_{a_{S\_P1}}^2 & \cdot & \cdot & \cdot & \cdot & \cdot \\
 \cdot & \cdot & \cdot & \cdot & \cdot & \sigma_{a_{S\_P2}}^2 & \cdot & \cdot & \cdot & \cdot \\
 \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \sigma_{a_{T\_P0}}^2 & \cdot & \cdot & \cdot \\
 \cdot & \sigma_{a_{T\_P1}}^2 & \cdot & \cdot \\
 \sigma_{a_{F\_P0}a_{T\_P2}} & \cdot & \sigma_{a_{T\_P2}}^2 & \cdot
 \end{bmatrix}$$

115 with  $\sigma_{a_{x_{P_o}}}^2$  the variance of the  $o^{\text{th}}$  regression coefficient of Legendre polynomial (3 classes: 0,  
 116 1, 2) for the  $x^{\text{th}}$  parity (3 classes: first (F), second (S), third (T)), and  $\sigma_{a_{x_{P_o}}a_{x_{P_o}}}$  the covariance between  
 117 the  $o^{\text{th}}$  regression coefficient of Legendre polynomial of the  $x^{\text{th}}$  parity and the  $o^{\text{th}}$  regression coefficient  
 118 of Legendre polynomial of the  $x^{\text{th}}$  parity. The permanent environment variance-covariance matrix was  
 119 defined in the same manner.

120 The EBV of goat  $n$  for its three parities was obtained by summing the EBV for each DIM (**3-**  
 121 **lact-EBV**).

122 **RRM using eigenfunction (EGV).** In a second step, the first two eigenvectors for each parity  
 123 of the genetic variance-covariance matrix from LEG\_FT, which can be associated with level of  
 124 production and persistency, were multiplied by the Legendre polynomial coefficients to obtain  
 125 eigenfunctions as shown in Druet et al. (2005). This makes it possible to obtain a reduced model that  
 126 requires less computation capacity because fewer parameters have to be estimated ( $2 * 3$  instead of  $3 * 3$ ).  
 127 It also allows the information to be summarized in terms of level and persistence as shown in Arnal  
 128 et al. (2019). These eigenfunctions were used for the genetic and the permanent environment parts of  
 129 the model. The second eigenfunction was set to be negative at the beginning of the lactation, then  
 130 positive as in Arnal et al. (2019) to model persistency. The reduced model (**EGV\_FT**) was:

$$131 \quad y_{trsiijklmdgn} = HTD_{tri} + A_{tjk} + M_{trjl} + D_{trjm} + \sum_{o=1}^6 \theta_{tko} N_{(o,d)} + \sum_{o=1}^6 \tau_{trlo} N_{(o,d)}$$

$$132 \quad + \sum_{o=1}^6 \pi_{trmo} N_{(o,d)} + \sum_{o=1}^4 \gamma_{tso} M_{(o,g)} + \sum_{o=1}^2 b_{trno} \chi_{(or,d)} + \sum_{o=1}^2 c_{trno} \chi_{(or,d)}$$

$$133 \quad + e_{trsiijklmdgn}$$

134 where  $b_{trno}$  and  $c_{trno}$  are the random additive genetic and permanent environmental regression  
 135 coefficients for the  $o^{\text{th}}$  eigenfunctions, and  $\chi_{(or,d)}$  the value of the  $o^{\text{th}}$  eigenfunction at DIM  $d$  for parity  
 136  $r$ . The other effects were the same as in LEG\_FT. The random additive genetic regression coefficient  
 137 for the first eigenfunction,  $b_1$ , is noted **LEV** (related to production level) and for the second  
 138 eigenfunction,  $b_2$ , is noted **PERS** (persistency). The three-lactation LEV EBV (**3-lact-LEV-EBV**) of  
 139 goat  $n$  was obtained by summing the LEV EBV of each parity. The three-lactation PERS EBV (**3-lact-**  
 140 **PERS-EBV**) of goat  $n$  was obtained by summing the PERS EBV of each parity.

141 The genetic variance-covariance matrix (C) for EGV\_FT was :

$$142 \quad C_{EGV\_FT} = \begin{bmatrix} \sigma_{a_{F\_EGV1}}^2 & \cdot & \cdot & \cdot & \cdot & \sigma_{a_{F\_EGV1}a_{T\_EGV2}} \\ \cdot & \sigma_{a_{F\_EGV2}}^2 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \sigma_{a_{S\_EGV1}}^2 & \sigma_{a_{S\_EGV2}}^2 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \sigma_{a_{T\_EGV1}}^2 & \cdot & \cdot \\ \sigma_{a_{F\_EGV1}a_{T\_EGV2}} & \cdot & \cdot & \cdot & \cdot & \sigma_{a_{T\_EGV2}}^2 \end{bmatrix}$$

143 with  $\sigma_{a_{x\_EGV_o}}^2$  the variance of the  $o^{\text{th}}$  regression coefficient of the eigenfunctions (2 classes: 1  
 144 (level of production throughout the lactation), 2 (persistence)) for the  $x^{\text{th}}$  parity (3 classes: first (F),  
 145 second (S), third (T)), and  $\sigma_{a_{x\_P_0}a_{x\_P_0}}$  the covariance between the  $o^{\text{th}}$  regression coefficient of Legendre  
 146 polynomial of the  $x^{\text{th}}$  parity and the  $o^{\text{th}}$  regression coefficient of Legendre polynomial of the  $x^{\text{th}}$  parity.  
 147 The permanent environment variance-covariance matrix was defined in the same manner.

148 ***Primiparous and multiparous (PM)***

149 To further simplify the model, the second and third parities (multiparous) were modeled together  
 150 for genetic and permanent environment effects. For the fixed effects and the residual variance, the  
 151 second and third parities were kept separate. The genetic and permanent environment effects were first  
 152 modeled using a Legendre polynomial of degree 2 (model LEG\_PM). The first two eigenvectors of the  
 153 genetic variance-covariance matrix for primiparous and multiparous from LEG\_PM were then  
 154 multiplied by the terms of the Legendre polynomial to obtain eigenfunctions as above. These  
 155 eigenfunctions were used for the genetic and the permanent environment parts of the model in  
 156 (EGV\_PM). As above, the second eigenfunction was set to be negative at the beginning of the lactation  
 157 to model persistency. The random additive genetic regression coefficients for these two eigenfunctions  
 158 are  $b_1$ , noted LEV (production level), and  $b_2$ , noted PERS (persistence).

159 The genetic variance-covariance matrix (C) for LEG\_PM was defined as:

160 
$$C_{\text{LEG\_PM}} = \begin{bmatrix} \sigma_{a_{P\_P_0}}^2 & \cdot & \cdot & \cdot & \cdot & \sigma_{a_{P\_P_0}a_{M\_P_2}} \\ \cdot & \sigma_{a_{P\_P_1}}^2 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \sigma_{a_{P\_P_2}}^2 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \sigma_{a_{M\_P_0}}^2 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \sigma_{a_{M\_P_1}}^2 & \cdot \\ \sigma_{a_{P\_P_0}a_{M\_P_2}} & \cdot & \cdot & \cdot & \cdot & \sigma_{a_{M\_P_2}}^2 \end{bmatrix}$$

161 with  $\sigma_{a_{x\_P_0}}^2$  the variance of the  $o^{\text{th}}$  regression coefficient of Legendre polynomial (3 classes: 0,  
 162 1, 2) for the  $x^{\text{th}}$  parity (2 classes: Primiparous (P), multiparous (M)), and  $\sigma_{a_{x\_P_0}a_{x\_P_0}}$  the covariance  
 163 between the  $o^{\text{th}}$  regression coefficient of Legendre polynomial of the  $x^{\text{th}}$  parity and the  $o^{\text{th}}$  regression  
 164 coefficient of Legendre polynomial of the  $x^{\text{th}}$  parity. The permanent environment variance-covariance  
 165 matrix was defined in the same manner.

166 The genetic variance-covariance matrix (C) for EGV\_PM was :

167 
$$C_{\text{EGV\_PM}} = \begin{bmatrix} \sigma_{a_{P\_EGV1}}^2 & \cdot & \cdot & \sigma_{a_{P\_EGV1}a_{M\_EGV2}} \\ \cdot & \sigma_{a_{P\_EGV2}}^2 & \cdot & \cdot \\ \cdot & \cdot & \sigma_{a_{M\_EGV1}}^2 & \cdot \\ \sigma_{a_{P\_EGV1}a_{M\_EGV2}} & \cdot & \cdot & \sigma_{a_{M\_EGV2}}^2 \end{bmatrix}$$

168 with  $\sigma_{a_{x\_EGV_o}}^2$  the variance of the  $o^{\text{th}}$  regression coefficient of the eigenfunctions (2 classes: 1  
 169 (level of production throughout the lactation), 2 (persistence)) for the  $x^{\text{th}}$  parity (2 classes: Primiparous  
 170 (P), multiparous (M)), and  $\sigma_{a_{x\_EGV_o}a_{x\_EGV_o}}$  the covariance between the  $o^{\text{th}}$  regression coefficient of

171 Legendre polynomial of the  $x^{\text{th}}$  parity and the  $o^{\text{th}}$  regression coefficient of Legendre polynomial of the  
172  $x^{\text{th}}$  parity. The permanent environment variance-covariance matrix was defined in the same manner.

173 The EBV of goat  $n$  for its three parities (3-lact-EBV) was obtained as for LEG\_FT and EGV\_FT.  
174 The three-lactation LEV EBV (3-lact-LEV-EBV) of goat  $n$  was obtained by multiplying the LEV EBV  
175 for primiparous by 0.33 and for multiparous by 0.66. The same calculation was carried out for  
176 persistency to obtain 3-lact-PERS-EBV.

### 177 ***Genetic correlations and heritability***

178 The genetic variance-covariance matrix among all DIM was derived following Druet et al.  
179 (2003) as  $G_{792} = QK_aQ'$  where  $G$  is a 792 (264 DIM  $\times$  3 parities = 792 DIM) by 792 genetic  
180 (co)variance matrix,  $K_a$  the  $(o \times r)$  by  $(o \times r)$  variance covariance matrix and  $Q$  a 792 by an  $(o \times$   
181  $r)$  matrix with the values of the  $o$  terms of Legendre polynomial repeated for each parity for LEG  
182 models or the  $o$  eigenfunctions for each parity for EGV models.

183 The same approach was used to obtain the permanent environmental variance-covariance matrix  
184  $W_{792}$ . The phenotypic variance-covariance matrix between all DIM,  $P_{792}$ , was obtained by summing  
185  $G_{792}$ ,  $W_{792}$  and the residual variance for the relevant DIM. The daily heritabilities were obtained by  
186 dividing the diagonal elements of  $G_x$  by the diagonal elements of  $P_x$ .

187 The genetic correlation between the  $d^{\text{th}}$  DIM and the other DIM is the  $d^{\text{th}}$  column of  $G_{792}$ .

188 The heritability of the  $o^{\text{th}}$  regression coefficients were calculated by dividing the genetic  
189 variance of the  $o^{\text{th}}$  regression coefficient by the sum of the genetic variance of the  $o^{\text{th}}$  regression  
190 coefficient, the permanent environment variance of the  $o^{\text{th}}$  regression coefficient and the mean square  
191 error as in Schaeffer (2016). This calculation was used to calculate the heritabilities of LEV and PERS.

192 All the genetic parameters were estimated using the WOMBAT software (Meyer, 2007).

### 193 **Model selection criteria**

194 The Pearson correlations between the EBVs of bucks obtained from the different models  
195 were studied to compare and assess the closeness of the models.

## 196 **RESULTS**

197 The genetic correlations between the second and third parity regression coefficients with the  
198 LEG\_FT and EGV\_FT models are shown in Table 2. Genetic correlations were very high (values  
199 between 0.92 and 1) for both models and all traits, except for SCS, where genetic correlations were  
200 equal to 0.73, 0.80 and 0.83 for PERS,  $a_2$  and  $a_1$  Legendre coefficients respectively. These results  
201 prompted us to combine the second and third lactations into a single category, i.e. multiparous, when  
202 modeling the genetic and the permanent environment components of each trait. This was done through  
203 the LEG\_PM model, which used a Legendre polynomial of order 2 for the genetic and permanent

204 environment components for each parity, and the EGV\_PM model, which used the first two eigenvectors  
205 of the LEG\_PM variance-covariance matrix for each parity.

### 206 ***Model fitting***

207 The different models were compared based on their estimated residual variances for milk yield  
208 according to DIM in Figure 1. The residual variance was smallest with LEG\_FT for all DIM, parities  
209 and traits. For the first lactation, residual variance according to days for LEG\_PM and LEG\_FT were  
210 totally superposed. And similarly the evolution of residual variance for EGV\_PM and EGV\_FT were  
211 also totally superposed in first lactation (Figure 1). When the multiparous lactations were modeled  
212 together, the residuals were higher, especially for the third lactation. Residual variance estimates with  
213 the LEG\_FT model were very close to those with the EGV\_FT and to those of the LEG\_PM and  
214 EGV\_PM models, except at the beginning and end of lactation, where estimates with LEG models were  
215 smaller (i.e. better) than with EGV models. Changing the Legendre polynomials by eigenfunctions  
216 (LEG\_FT to EGV\_FT or LEG\_PM to EGV\_PM) generate less residuals than putting together  
217 multiparous parities (LEG\_FT to LEG\_PM or EGV\_FT to EGV\_PM). The higher residual variances  
218 were due to permanent environment effects being not well modeled with EGV\_FT and combining 2 and  
219 3 parities (results not shown).

### 220 ***Estimated genetic variance***

221 The evolution of estimated test-day genetic variances according to DIM for milk yield is  
222 presented in Figure 2. The variances increased from the first to the third parity following the same pattern  
223 (a bell curve). The evolution of genetic variance in first lactation for LEG\_PM and LEG\_FT were totally  
224 superposed. And similarly genetic variances according to DIM for EGV\_PM and EGV\_FT were also  
225 totally superposed in first lactation. Smaller differences were found between FT models and between  
226 PM models compared to residual variances. For yield traits, the genetic variances were higher in third  
227 than in second parity. To combine the second and the third parities together, we multiplied, for the third  
228 parity, the Legendre covariables for LEG\_PM and the eigenfunction for EGV\_PM by 1.1 for MY, and  
229 by 1.2 for FY and PY. These coefficients were chosen in order to have the best third lactation genetic  
230 variance (results not shown). When the second and third parities were combined, the genetic variance  
231 for yield traits and SCS with PM models was higher than with FT models in the second parity, but  
232 smaller in the third parity. However, for contents, the phenotypic means and the genetic variances were  
233 similar between parities (results not shown).

### 234 ***Heritabilities***

235 The evolution of test-day heritabilities for the different traits is shown in Figure 3. Few  
236 variations were found between parities. The highest heritabilities were found for protein content (around  
237 0.6), then fat content (around 0.45), FPR (around 0.35), yield traits (around 0.25) and SCS (around 0.15).  
238 Table 3 compares the heritabilities for the regression coefficients of EGV\_FT and EGV\_PM. Again,

239 few differences were observed between the two models (FT and PM). The heritabilities of LEV were  
240 quite similar between parities and equal to the mean of the test-day heritabilities shown in Figure 3. The  
241 heritabilities of PERS ranged between 0.04 (for SCS) and 0.21 (for protein content) in the first parity.  
242 Heritabilities of PERS for the second and third parities for yield traits were higher than those found in  
243 the first parity (e.g. 0.10 for milk yield in the first parity and 0.17 for multiparous); this was not the case  
244 for contents, FPR and SCS, for which heritabilities were equal to those found in the first lactation.

#### 245 *Estimated genetic correlations*

246 The genetic correlations between DIM 40 (close to the lactation peak) in the second parity and  
247 all other DIM for milk yield were compared according to the models used in Figure 4. Few differences  
248 were observed between models. Correlations decreased down to about +0.5 in the second parity. As seen  
249 in Table 2, the evolution of the genetic correlations between DIM 40 in the second lactation and all DIM  
250 in the third lactation was similar to the evolution in the second parity. Combining second and third  
251 lactations in the LEG\_PM and EGV\_PM models is thus sound, giving a genetic correlation estimated  
252 of 1 between DIM 40 in the two lactations compared with 0.98. By contrast, the correlation between  
253 DIM 40 in first and second lactations was only 0.85, confirming that yield traits in first and later  
254 lactations are genetically different. However, the genetic correlation curve shapes were the same  
255 between parities.

256 The same correlations between DIM 40 in the second lactation and all DIM from LEG\_FT for  
257 all the traits are plotted in Figure 5. The highest correlations between the beginning and end of lactation  
258 in the second parity were observed for SCS (higher than 0.85), then contents (around 0.80), FPR (around  
259 0.70) and finally yield traits. The correlations between DIM 40 in the second and third parities were  
260 always close to 1, and the shapes of the correlations for the different traits were again very similar, as  
261 seen in Table 2. Smaller correlations between DIM 40 in the second parity and end of lactation were  
262 observed for yields in the third compared to the second parity. However, clear trait differences were  
263 observed between DIM 40 in the first and second lactations: correlations were much higher for contents  
264 (0.95) than FPR (0.92), yields (0.85) and SCS (<0.75). For all traits and parities, the within-parity  
265 correlations decreased with distance from DIM 40, except for SCS in the first parity. Correlations  
266 between DIM 40 in the second parity and DIM of the first lactation for SCS were higher in mid-lactation  
267 (between DIM 150 and DIM 200).

#### 268 *EBV correlations between models*

269 The Pearson correlations between the 3-lact-EBVs from the LEG models, 3-lact-LEV-EBVs  
270 and 3-lact-PERS-EBVs from EGV models for AI-bucks, sires of the recorded goats ( $n = 486$ ), were  
271 calculated to evaluate the closeness of the models in Table 4 for milk yield. Sires of the recorded goats  
272 were all from an artificial insemination center, with at least 20 daughters in the dataset. The EBVs of  
273 these animals were considered reliable given daughter number. The correlations between the 3-lact-

274 EBVs derived from the two LEG models (LEG-FT and LEG-PM), between the 3-lact-LEV-EBVs and  
275 between the 3-lact-PERS-EBVs of the two EGV models were all greater than or equal to 0.99,  
276 confirming that modeling second and third parities together generated little differences. The correlations  
277 between the 3-lact-LEV-EBVs from EGV and the 3-lact-EBVs from LEG were also equal to 1, showing  
278 that the reduced model using the two main eigenvalues fitted well, and that its LEV component truly  
279 represented the level of production throughout the lactations. The correlations between 3-lact-LEV-  
280 EBVs and 3-lact-PERS-EBVs were close to 0 ( $-0.04$  for EGV\_FT and  $-0.02$  for the EGV\_PM) showing  
281 that the reduced model uncorrelated level of production and persistency. Similar results were obtained  
282 for the other traits (results not shown).

### 283 ***EBV correlations between traits***

284 For residual variances, genetic variances, genetic correlations between DIM and EBV  
285 correlations, the four models gave similar results. EGV\_PM, which took the shortest computational time  
286 (EGV\_FT was 1.15 times longer, LEG\_PM 1.43 times longer and LEG\_FT 2.25 times longer for milk  
287 yield) and whose interpretation was easy was chosen.

288 The correlations between milk yield EBVs of Alpine AI-bucks and yields, contents and SCS  
289 EBVs from the EGV\_PM model are presented in Table 5. High positive correlations were observed  
290 between 3-lact-LEV-EBVs of milk yield and 3-lact-LEV-EBVs of other yield traits (0.73 for fat yield  
291 and 0.88 for protein yield). Negative correlations were found between 3-lact-LEV-EBVs of milk yield  
292 and 3-lact-LEV-EBVs of contents ( $-0.3$  with fat content,  $-0.46$  with protein content). Correlations were  
293 nil between milk 3-lact-LEV\_EBVs and FPR 3-lact-LEV-EBVs and slightly negative with SCS 3-lact-  
294 LEV-EBVs ( $-0.14$ ). Correlation between 3-lact-LEV-EBVs of milk yield and 3-lact-PERS-EBVs of  
295 other traits were close to 0 (min: 0 for protein yield, max:  $-0.19$  for SCS) together with the correlations  
296 between 3-lact-PERS-EBVs of milk yield and for the other trait 3-lact-LEV-EBVs. The correlation  
297 between 3-lact-PERS-EBVs of milk yield and fat content 3-lact-PERS-EBVs was close to 0 and  
298 moderately negative with SCS 3-lact-PERS-EBVs ( $-0.24$ ), while this correlation was strongly negative  
299 with protein content 3-lact-PERS-EBVs ( $-0.64$ ). Consequently, the correlation between 3-lact-PERS-  
300 EBVs of milk yield and FPR 3-lact-PERS-EBVs was positive (0.43). Like for production level, the  
301 correlations between persistencies of yield traits were very high (0.89 for fat yield and 0.94 for protein  
302 yield). The same correlations between traits were obtained when studying only measured goat EBVs.

### 303 **DISCUSSION**

304 The results presented pertain the Alpine breed (almost 60% of goats in France). The same  
305 analyses were performed in French Saanen goats and similar results were obtained for all traits (results  
306 not shown). The estimation of genetic parameters was based on a large number of goats measured  
307 regularly throughout their lactation. The connectedness between the large number of herds was ensured  
308 by the use of phenotypes of daughters of artificial insemination bucks.

309 LEG\_FT was considered as the reference (full) model. Genetic and permanent environment  
310 effects were modeled with a Legendre polynomial of order 2 as proposed by Arnal et al.(2019), which  
311 studied first lactation traits in dairy goats in France. They showed that a Legendre polynomial of order  
312 2 was close to a more complex Legendre polynomial of order 4 in goodness-of-fit and genetic  
313 parameters. A model using a Legendre polynomial of order 4, for each parity, would thus provide little  
314 additional precision and be too time-consuming in routine use for genetic evaluation on a complete data  
315 set.

316 EGV\_FT was compared to LEG\_FT. The EGV\_FT model was the extension of the reduced  
317 model proposed by Arnal et al. (2019), which was satisfactory in terms of goodness-of-fit and genetic  
318 parameters, summarizing the relevant information for production level (LEV) and persistency (PERS).

319 Goodness-of-fit was better at the beginning and end of lactation with a complete model than  
320 with a reduced model using eigenfunctions, as shown by Arnal et al. (2019) with data on first parity  
321 goats only. However, the genetic variances and between-DIM genetic correlations were very similar  
322 between complete models and reduced models using eigenfunctions. The reduced models therefore  
323 represent attractive alternatives to the complete analysis for routine genetic evaluations.

324 We showed that EBV correlations for production level were close to 1 between LEG\_FT,  
325 EGV\_FT, LEG\_PM and EGV\_PM, in good agreement with Leclerc et al. (2009) and Arnal et al. (2019),  
326 who compared EBVs from a complete and a reduced model. The advantages of the reduced model are  
327 first the smaller size of the variance-covariance matrices, a shorter computing time, and also the zero  
328 correlation between the EBV for production level and persistency by construction. This enables us to  
329 select animals with a desired persistency throughout lactation without changing the total lactation  
330 production.

331 Modeling the second and third parities together led to larger residuals, especially for the third  
332 parity. A way to avoid this increase could be to not model together parities 2 and 3 for the permanent  
333 environment effect but only for genetic part. This was not implemented in order the more reduced  
334 models could be fitted. For the genetic variances, some differences were also found between models  
335 considering all parities separately or those combining the second and third. However, regarding the  
336 heritabilities and the buck EBVs (production throughout three parities and persistency) when the  
337 reduced model was applied, few differences were found from the models considering all parities  
338 separately. The model combining second and third lactations and using eigenfunctions was therefore  
339 preferred for routine calculations because of its shorter computing time and its ease of interpretation.

340 The heritabilities estimated for production level and persistency for primiparous goats  
341 confirmed those reported by Arnal et al. (2019), with partly the same animals and also a reduced model.  
342 The production level heritabilities were in line with Rupp et al. (2011), who found a heritability of 0.2  
343 for SCS in primiparous Alpine goats in France with a lactation model, and similar heritabilities for yields

344 and contents. We found heritabilities for total milk production yield similar to other studies on different  
345 breeds of goats using RRM (Zumbach et al., 2008; Menéndez-Buxadera et al., 2010; Mucha et al.,  
346 2014). For yield traits, we also found heritabilities similar to Menéndez-Buxadera et al. (2010). Using  
347 an RRM, Apodaca-Sarabia et al. (2009) found, like us, a heritability between 0.12 and 0.25 for SCS in  
348 a goat population of New Zealand, mixing parities. We found a heritability close to 0.3 for FPR for the  
349 first three parities, in good agreement with Nishiura et al. (2015) who studied FPR in Holstein dairy  
350 cows (2014), using an RRM.

351 We report a heritability of 0.1 for 3-lact milk yield persistency, but Menéndez-Buxadera et al.  
352 (2010) reported a higher heritability (0.20) considering the second eigenfunction of their variance-  
353 covariance matrix. These heritabilities are of the same order, and the small difference may be due to  
354 breed differences. In dairy cows, Haile-Mariam et al. (2003) found like us that fat persistency heritability  
355 (0.06) was smaller than milk yield persistency. Cole and VanRaden (2006) reported a milk yield  
356 persistency heritability similar to ours in dairy cows, mixing parities, with no correlation with milk yield.  
357 Like us, they found smaller heritabilities for fat yield persistency (0.07), protein yield persistency (0.09)  
358 and especially SCS persistency (0.03) than for milk yield persistency.

359 In line with our study, the other RRM goat studies (Zumbach et al., 2008; Menéndez-Buxadera  
360 et al., 2010; Mucha et al., 2014) reported similar genetic correlations between DIM and between parities  
361 for milk yield. For other yield traits, Menéndez-Buxadera et al. (2010) reported correlations similar to  
362 ours. For SCS in dairy cows, Zavadilova et al. (2011) investigated the genetic correlation between DIM  
363 40 and all DIM of all parities with a random regression model, and reported, like us, that the highest  
364 correlation between DIM 40 in the second lactation and DIM in the first lactation was not at DIM 40  
365 but at DIM 100. As in Druet et al. (2005) and Arnal et al. (2019), higher genetic correlations were found  
366 between the beginning and end of lactation for content traits than for yield traits.

367 The correlations between production level for milk yield and for the other traits were very close  
368 to those estimated in the first parity by Arnal et al. (2019) and by Bélichon et al. (1999) on total lactation  
369 traits (0.90 for protein yield and 0.76 for fat yield in Bélichon et al. (1999) and 0.88 and 0.73 for protein  
370 and fat yields respectively in our study). They were slightly weaker in Bélichon's analysis for contents:  
371  $-0.28$  for protein content and  $-0.13$  for fat content vs.  $-0.46$  and  $-0.3$  respectively in the present study.  
372 In terms of selection, these negative correlations between contents and milk yield are a brake on  
373 increasing protein yield. Rupp et al. (2011) found a zero correlation between milk yield and SCS in the  
374 first parity, close to our result ( $-0.14$ ). In Holstein dairy cows, Cole and VanRaden (2006) found, like  
375 us, small correlations between milk production level and milk persistency and a negative one between  
376 milk yield level and SCS persistency ( $-0.23$ ). They reported correlations close to 1 between milk yield  
377 persistency and fat and protein yield persistencies, high correlation between fat and protein yield  
378 persistencies as did we, and a negative correlation between milk persistency and SCS persistency as in  
379 our application ( $-0.24$ ). The correlation between milk yield persistency and LEV of contents was close

380 to zero, so that milk yield persistency can be selected with no impact on contents. However, the negative  
381 high genetic correlation between protein content persistency and milk yield persistency indicated that a  
382 goat with a more persistent milk production will often have a higher protein content at the beginning of  
383 lactation than an average animal with the same protein content throughout lactation. Similar conclusions  
384 were drawn when the permanent environment of the animal was added to its EBV. Buttchereit et al.  
385 (2011) and Negussie et al. (2013), in dairy cows, reported that a decrease in FPR at the beginning of the  
386 lactation was genetically correlated with a decrease in milk yield. We confirmed this result in goats.  
387 Milk yield persistency (less milk at the beginning of lactation than a goat producing the same amount  
388 of milk throughout lactation) was positively correlated (+0.43) with FPR persistency (i.e. smaller FPR  
389 at the beginning of the lactation), which could be used to select for better fitness, because metabolic  
390 diseases were found to be correlated with high FPR at the beginning of lactation (Buttchereit et al., 2011;  
391 Jamrozik and Schaeffer, 2012; Koeck et al., 2013; Negussie et al., 2013). A negative genetic correlation  
392 was found in dairy ewes (Riggio et al., 2009) and in dairy cows (Caraviello et al., 2005; Sewalem et al.,  
393 2006) between high SCS and longevity. The negative correlation between PERS milk yield and LEV  
394 SCS suggests that a more persistent goat could live longer. This emphasizes the potential benefit of  
395 selecting for lactation persistency in breeding programs. The correlation between milk yield persistency  
396 and fertility has to be studied to evaluate a potential antagonism between the two (Pryce et al., 2004;  
397 Dobson et al., 2007). In dairy cows, reproduction, in a scheme with a calving each year, occurs around  
398 DIM 90, and so no antagonism between fertility and persistency was found (Harder et al., 2006;  
399 Albarrán-Portillo and Pollott, 2013). However, in dairy goats, the reproduction is at the end of lactation  
400 (around 210 DIM), in herds with a kidding each year, and so an antagonism could arise.

## 401 **CONCLUSIONS**

402 We show that the genetic parameters and EBVs obtained with a reduced rank test-day model  
403 grouping together multiparous goats are similar to those derived from a complete model using a  
404 traditional Legendre polynomial for each parity. This reduced model yields production level and  
405 persistency EBVs per lactation. These are nearly genetically independent, so that animals with a desired  
406 lactation shape can be selected independently of their total milk yield production. The positive genetic  
407 relationship between milk persistency and FPR persistency may offer a way to genetically offset  
408 negative energy balance at the beginning of lactation.

## 409 **ACKNOWLEDGMENTS**

410 The first author received financial support from APIS-GENE (Paris, France) and the French  
411 National Association for Research and Technology (ANRT, Paris, France). The authors thank Karin  
412 Meyer for the Wombat program.

## 413 **REFERENCES**

414           Albarrán-Portillo, B., and G.E. Pollott. 2013. The relationship between fertility and lactation  
415 characteristics in Holstein cows on United Kingdom commercial dairy farms. *J. Dairy Sci.* 96:635–646.

416           Apodaca-Sarabia, C.A., N. Lopez-Villalobos, H.T. Blair, and G.G. Prosser. 2009. Genetic  
417 parameters for somatic cell score in dairy goats estimated by random regression. Pages 206–209 in  
418 Proceedings of the New Zealand Society of Animal Production. New Zealand Society of Animal  
419 Production.

420           Arnal, M., H. Larroque, H. Leclerc, V. Ducrocq, and C. Robert-Granié. 2019a. Genetic  
421 parameters for first lactation dairy traits in the Alpine and Saanen goat breeds using a random regression  
422 test-day model. *Genet. Sel. Evol.* 51:43.

423           Arnal, M., C. Robert-Granié, and H. Larroque. 2018. Diversity of dairy goat lactation curves in  
424 France. *J. Dairy Sci.* 101:11040–11051.

425           Bélichon, S., E. Manfredi, and A. Piacère. 1999. Genetic parameters of dairy traits in the Alpine  
426 and Saanen goat breeds. *Genet. Sel. Evol.* 31:529–534.

427           Buttchereit, N., E. Stamer, W. Junge, and G. Thaller. 2011. Genetic relationships among daily  
428 energy balance, feed intake, body condition score, and fat to protein ratio of milk in dairy cows. *J. Dairy  
429 Sci.* 94:1586–1591.

430           Caraviello, D.Z., K.A. Weigel, G.E. Shook, and P.L. Ruegg. 2005. Assessment of the impact of  
431 somatic cell count on functional longevity in Holstein and Jersey cattle using survival analysis  
432 methodology. *J. Dairy Sci.* 88:804–811.

433           Cole, J.B., and P.M. VanRaden. 2006. Genetic evaluation and best prediction of lactation  
434 persistency. *J. Dairy Sci.* 89:2722–2728.

435           Dobson, H., R.F. Smith, M.D. Royal, C.H. Knight, and I.M. Sheldon. 2007. The high-producing  
436 dairy cow and its reproductive performance. *Reprod. Domest. Anim.* 42:17–23.

437           Druet, T., F. Jaffrézic, D. Boichard, and V. Ducrocq. 2003. Modeling lactation curves and  
438 estimation of genetic parameters for first lactation test-day records of French Holstein cows. *J. Dairy  
439 Sci.* 86:2480–2490.

440           Druet, T., F. Jaffrézic, and V. Ducrocq. 2005. Estimation of genetic parameters for test day  
441 records of dairy traits in the first three lactations. *Genet. Sel. Evol.* 37:257.

442           Haile-Mariam, M., P.J. Bowman, and M.E. Goddard. 2003. Genetic and environmental  
443 relationship among calving interval, survival, persistency of milk yield and somatic cell count in dairy  
444 cattle. *Livest. Prod. Sci.* 80:189–200.

445 Harder, B., J. Bennewitz, D. Hinrichs, and E. Kalm. 2006. Genetic parameters for health traits  
446 and their relationship to different persistency traits in German Holstein dairy cattle. *J. Dairy Sci.*  
447 89:3202–3212.

448 Jamrozik, J., and L.R. Schaeffer. 2012. Test-day somatic cell score, fat-to-protein ratio and milk  
449 yield as indicator traits for sub-clinical mastitis in dairy cattle. *J. Anim. Breed. Genet.* 129:11–19.

450 Koeck, A., F. Miglior, J. Jamrozik, D.F. Kelton, and F.S. Schenkel. 2013. Genetic associations  
451 of ketosis and displaced abomasum with milk production traits in early first lactation of Canadian  
452 Holsteins. *J. Dairy Sci.* 96:4688–4696.

453 Leclerc, H. 2008. Development of the French dairy cattle test-day model genetic evaluation and  
454 prospects of using results for herd management. PhD Thesis Thesis. AgroParisTech,.

455 Leclerc, H., I. Nagy, and V. Ducrocq. 2009. Impact of using reduced rank random regression  
456 test-day model on genetic evaluation. *Interbull Bull.* 40:42–46.

457 Menéndez-Buxadera, A., A. Molina, F. Arrebola, M.J. Gil, and J.M. Serradilla. 2010. Random  
458 regression analysis of milk yield and milk composition in the first and second lactations of Murciano-  
459 Granadina goats. *J. Dairy Sci.* 93:2718–2726.

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

462 Mucha, S., R. Mrode, M. Coffey, and J. Conington. 2014. Estimation of genetic parameters for  
463 milk yield across lactations in mixed-breed dairy goats. *J. Dairy Sci.* 97:2455–2461.

464 Negussie, E., I. Strandén, and E.A. Mäntysaari. 2013. Genetic associations of test-day fat:  
465 protein ratio with milk yield, fertility, and udder health traits in Nordic Red cattle. *J. Dairy Sci.* 96:1237–  
466 1250.

467 Nishiura, A., O. Sasaki, M. Aihara, H. Takeda, and M. Satoh. 2015. Genetic analysis of fat-to-  
468 protein ratio, milk yield and somatic cell score of H olstein cows in J apan in the first three lactations by  
469 using a random regression model. *Anim. Sci. J.* 86:961–969.

470 Pryce, J.E., M.D. Royal, P.C. Garnsworthy, and I.L. Mao. 2004. Fertility in the high-producing  
471 dairy cow. *Livest. Prod. Sci.* 86:125–135.

472 Riggio, V., D.O. Maizon, B. Portolano, H. Bovenhuis, and J.A.M. Van Arendonk. 2009. Effect  
473 of somatic cell count level on functional longevity in Valle del Belice dairy sheep assessed using survival  
474 analysis. *J. Dairy Sci.* 92:6160–6166.

475 Rupp, R., V. Clément, A. Piacere, C. Robert-Granié, and E. Manfredi. 2011. Genetic parameters  
476 for milk somatic cell score and relationship with production and udder type traits in dairy Alpine and  
477 Saanen primiparous goats. *J. Dairy Sci.* 94:3629–3634.

478 Schaeffer, L.R. 2016. Random Regression Models. Accessed July 3, 2019.  
479 <http://animalbiosciences.uoguelph.ca/~lrs/BOOKS/rrmbook.pdf>.

480 Schaeffer, L.R., and J. Jamrozik. 2008. Random regression models: a longitudinal perspective.  
481 *J. Anim. Breed. Genet.* 125:145–146.

482 Sewalem, A., F. Miglior, G.J. Kistemaker, and B.J. Van Doormaal. 2006. Analysis of the  
483 relationship between somatic cell score and functional longevity in Canadian dairy cattle. *J. Dairy Sci.*  
484 89:3609–3614.

485 Van der Werf, J.H.J., M.E. Goddard, and K. Meyer. 1998. The use of covariance functions and  
486 random regressions for genetic evaluation of milk production based on test day records. *J. Dairy Sci.*  
487 81:3300–3308.

488 Wiggans, G.R., and G.E. Shook. 1987. A lactation measure of somatic cell count. *J. Dairy Sci.*  
489 70:2666–2672.

490 Zavadilová, L., J. Wolf, M. Štípková, E. Němcová, and J. Jamrozik. 2011. Genetic parameters  
491 for somatic cell score in the first three lactations of Czech Holstein and Fleckvieh breeds using a random  
492 regression model. *Czech J Anim Sci* 56:251–260.

493 Zumbach, B., S. Tsuruta, I. Misztal, and K.J. Peters. 2008. Use of a test day model for dairy  
494 goat milk yield across lactations in Germany. *J. Anim. Breed. Genet.* 125:160–167.

495

496

497

**TABLES**

498

Table 1. Number of test-day records and lactations by parity in Alpine breed

<b>Parity</b>	<b>1</b>	<b>2</b>	<b>3</b>
<b>Test-day records</b>	318,612	225,658	148,473
<b>Lactations</b>	49,849	34,859	23,207

499

500 Table 2. Genetic correlation estimates of the three terms of the Legendre polynomial coefficients  
 501 ( $a_0$ ,  $a_1$ ,  $a_2$  in LEG\_FT\*) and of the regression coefficients of the first two eigenfunctions (LEV, PERS  
 502 in EGV\_FT\*\*) in second and third lactations in Alpine goats

	Milk yield	Fat yield	Protein yield	Fat content	Protein content	SCS	FPR
$a_0$	0.95	0.95	0.95	1.00	1.00	0.99	0.99
$a_1$	0.95	0.95	0.95	0.97	0.99	0.83	0.96
$a_2$	0.94	0.92	0.93	0.97	0.97	0.80	0.96
<b>LEV</b>	0.95	0.95	0.94	1	0.99	0.99	0.99
<b>PERS</b>	0.96	0.92	0.95	0.99	0.98	0.73	0.97

503 \* *LEG\_FT*: Legendre polynomial of order 2 for each parity to describe the genetic and permanent  
 504 environment effects,

505 \*\* *EGV\_FT*: eigenfunctions used for each parity to describe the genetic and permanent environment  
 506 effects

507

508 Table 3. Heritability of each component from EGV\_FT and EGV\_PM for each trait in Alpine  
 509 goats

Model	Parity		Milk yield	Fat yield	Protein yield	Fat content	Protein content	SCS	FPR
<b>EGV_FT</b>	<b>First</b>	<b>LEV<sub>F</sub></b>	0.28	0.25	0.26	0.56	0.66	0.19	0.43
<b>EGV_PM</b>	<b>Primiparous</b>	<b>LEV<sub>P</sub></b>	0.26	0.23	0.24	0.55	0.65	0.19	0.42
<b>EGV_FT</b>	<b>First</b>	<b>PERS<sub>F</sub></b>	0.12	0.09	0.09	0.1	0.21	0.05	0.1
<b>EGV_PM</b>	<b>Primiparous</b>	<b>PERS<sub>P</sub></b>	0.1	0.08	0.07	0.1	0.2	0.04	0.09
<b>EGV_FT</b>	<b>Second</b>	<b>LEV<sub>S</sub></b>	0.27	0.26	0.24	0.55	0.66	0.22	0.45
<b>EGV_FT</b>	<b>Third</b>	<b>LEV<sub>T</sub></b>	0.28	0.29	0.24	0.53	0.65	0.21	0.46
<b>EGV_PM</b>	<b>Multiparous</b>	<b>LEV<sub>M</sub></b>	0.28	0.27	0.25	0.54	0.67	0.23	0.46
<b>EGV_FT</b>	<b>Second</b>	<b>PERS<sub>S</sub></b>	0.17	0.15	0.14	0.1	0.2	0.02	0.11
<b>EGV_FT</b>	<b>Third</b>	<b>PERS<sub>T</sub></b>	0.17	0.16	0.16	0.1	0.17	0.03	0.1
<b>EGV_PM</b>	<b>Multiparous</b>	<b>PERS<sub>M</sub></b>	0.17	0.14	0.14	0.1	0.19	0.02	0.11

510 *EGV\_FT*: eigenfunctions used for each parity to describe the genetic and permanent environment  
 511 effects, *EGV\_PM*: eigenfunctions used for primiparous and multiparous to describe the genetic and  
 512 permanent environment effects, *LEV*: random additive genetic regression coefficient for the first  
 513 eigenfunction, *PERS*: random additive genetic regression coefficient for the second eigenfunction, *SCS*:  
 514 somatic cell score, *FPR*: Fat/protein ratio

515

516 Table 4. EBV correlations of bucks ( $n = 486$ ) according to the model for milk yield in Alpine  
 517 goats

		<b>LEG_FT</b>	<b>LEG_PM</b>	<b>EGV_FT</b>	<b>EGV_PM</b>	<b>EGV_FT</b>
		<b>3-lact-EBV</b>	<b>3-lact-EBV</b>	<b>3-lact-LEV-EBV</b>	<b>3-lact-LEV-EBV</b>	<b>3-lact-PERS-EBV</b>
<b>LEG_PM</b>	<b>3-lact-EBV</b>	1				
<b>EGV_FT</b>	<b>3-lact-LEV-EBV</b>	1	0.99			
<b>EGV_P M</b>	<b>3-lact-LEV-EBV</b>	1	1	1		
<b>EGV_FT</b>	<b>3-lact-PERS-EBV</b>	-0.06	-0.07	-0.04	-0.06	
<b>EGV_P M</b>	<b>3-lact-PERS-EBV</b>	-0.01	-0.02	0	-0.02	0.99

518 *LEG\_FT*: Legendre polynomial of order 2 for each parity to describe the genetic and permanent  
 519 environment effects, *LEG\_PM*: Legendre polynomial of order 2 for primiparous and multiparous to  
 520 describe the genetic and permanent environment effects, *EGV\_FT*: eigenfunctions used for each parity  
 521 to describe the genetic and permanent environment effects, *EGV\_PM*: eigenfunctions used for  
 522 primiparous and multiparous to describe the genetic and permanent environment effects, *3-lact-EBV*:  
 523 sum the EBV for each DIM throughout all parities, *3-lact-LEV-EBV*: sum of the random additive  
 524 genetic regression coefficient for the first eigenfunction throughout all parities (for *LEG\_PM* and  
 525 *EGV\_PM*, the random additive genetic regression coefficient for the first eigenfunction in primiparous  
 526 was multiplied by 0.33 and the one of multiparous by 0.66), *3-lact-PERS-EBV*: sum of the random  
 527 additive genetic regression coefficients for the second eigenfunction throughout all parities (for  
 528 *LEG\_PM* and *EGV\_PM*, the random additive genetic regression coefficient for the second eigenfunction  
 529 in primiparous was multiplied by 0.33 and that of multiparous by 0.66)

530

531 Table 5. EBV correlations of Alpine bucks ( $n = 486$ ) from EGV\_PM between milk yield and  
 532 other traits (fat yield, protein yield, fat content, protein content, SCS)

		Milk yield	
		3-lact-LEV-EBV	3-lact-PERS-EBV
Fat yield	3-lact-LEV-EBV	0.73	-0.06
	3-lact-PERS-EBV	0.13	0.89
Protein yield	3-lact-LEV-EBV	0.88	-0.02
	3-lact-PERS-EBV	0	0.94
Fat content	3-lact-LEV-EBV	-0.3	0.03
	3-lact-PERS-EBV	-0.05	-0.01
Protein content	3-lact-LEV-EBV	-0.46	-0.07
	3-lact-PERS-EBV	-0.04	-0.64
SCS	3-lact-LEV-EBV	-0.14	-0.16
	3-lact-PERS-EBV	-0.19	-0.24
FPR	3-lact-LEV-EBV	0.01	0.05
	3-lact-PERS-EBV	0.02	0.43

533

534 **FIGURES**

535 **Arnal, Figure 1: Evolution of the estimated residual variance according to days in milk (DIM) for**  
536 **the different models for milk yield in Alpine goats.**

537 *LEG\_PM and LEG\_FT were totally superposed in first lactation. EGV\_PM and EGV\_FT were also*  
538 *totally superposed in first lactation.*

539 *LEG\_FT: Legendre polynomial used for the modeling of genetic and permanent environment from first*  
540 *to third parity, LEG\_PM: Legendre polynomial used for the modeling of genetic and permanent*  
541 *environment for primiparous and multiparous, EGV\_FT: eigenfunctions used for the modeling of*  
542 *genetic and permanent environment from first to third parity, EGV\_PM: eigenfunctions used for the*  
543 *modeling of genetic and permanent environment for primiparous and multiparous*

544 **Arnal, Figure 2. Test-day genetic variances for four models for milk yield in Alpine goats.**

545 *Curves of LEG\_PM and LEG\_FT were totally superposed in first lactation. Curve of EGV\_PM and*  
546 *EGV\_FT were also totally superposed in first lactation.*

547 *LEG\_FT: Legendre polynomial of order 2 for each parity to describe the genetic and permanent*  
548 *environment effects, LEG\_PM: Legendre polynomial of order 2 for primiparous and multiparous to*  
549 *describe the genetic and permanent environment effects, EGV\_FT: eigenfunctions used for each parity*  
550 *to describe the genetic and permanent environment effects, EGV\_PM: eigenfunctions used for*  
551 *primiparous and multiparous to describe the genetic and permanent environment effects*

552 **Arnal, Figure 3. Test-day heritabilities for the different traits with the LEG\_FT model in Alpine**  
553 **goats.**

554 *SCS: somatic cell score, FPR: fat/protein ratio*

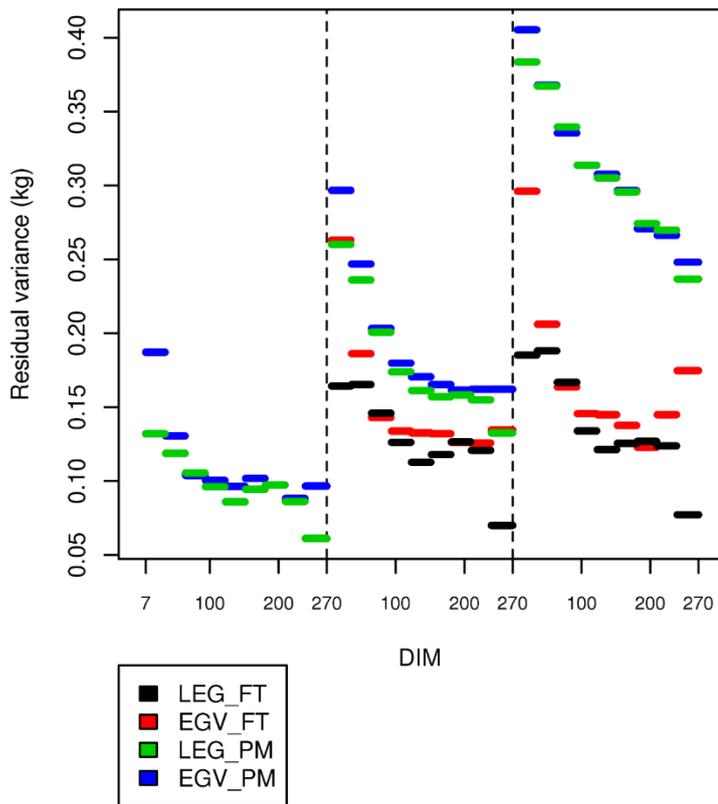
555 **Arnal, Figure 4. Genetic correlations between DIM 40 in second parity and the other DIM with**  
556 **the different models for milk yield in Alpine goats**

557 *LEG\_FT: Legendre polynomial of order 2 for each parity to describe the genetic and permanent*  
558 *environment effects, LEG\_PM: Legendre polynomial of order 2 for primiparous and multiparous to*  
559 *describe the genetic and permanent environment effects, EGV\_FT: eigenfunctions used for each parity*  
560 *to describe the genetic and permanent environment effects, EGV\_PM: eigenfunctions used for*  
561 *primiparous and multiparous to describe the genetic and permanent environment effects*

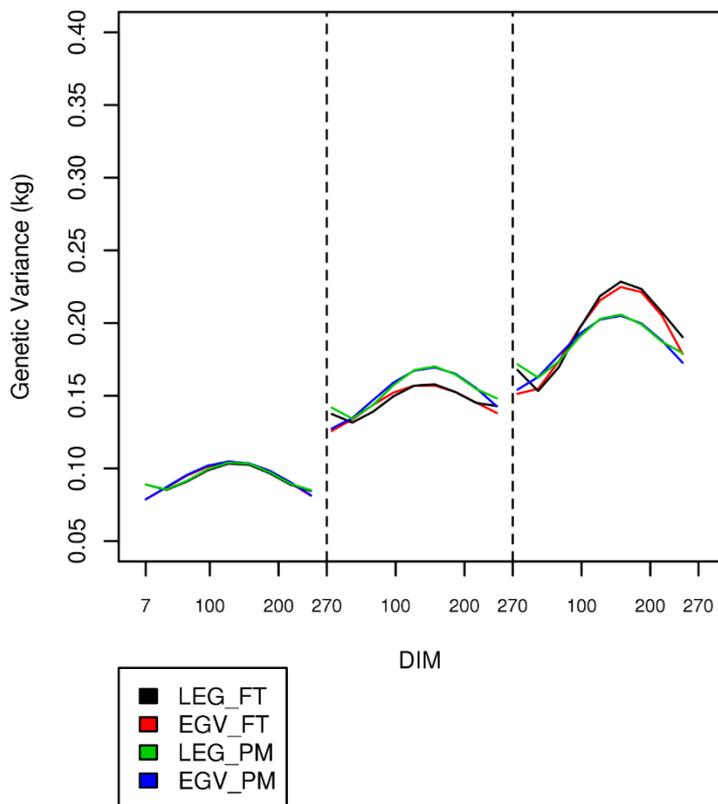
562 **Arnal, Figure 5. Genetic correlations with model LEG\_FT between DIM 40 in second lactation**  
563 **and all DIM for all traits in Alpine goats**

564 *SCS: somatic cell score, FPR: fat/protein ratio*

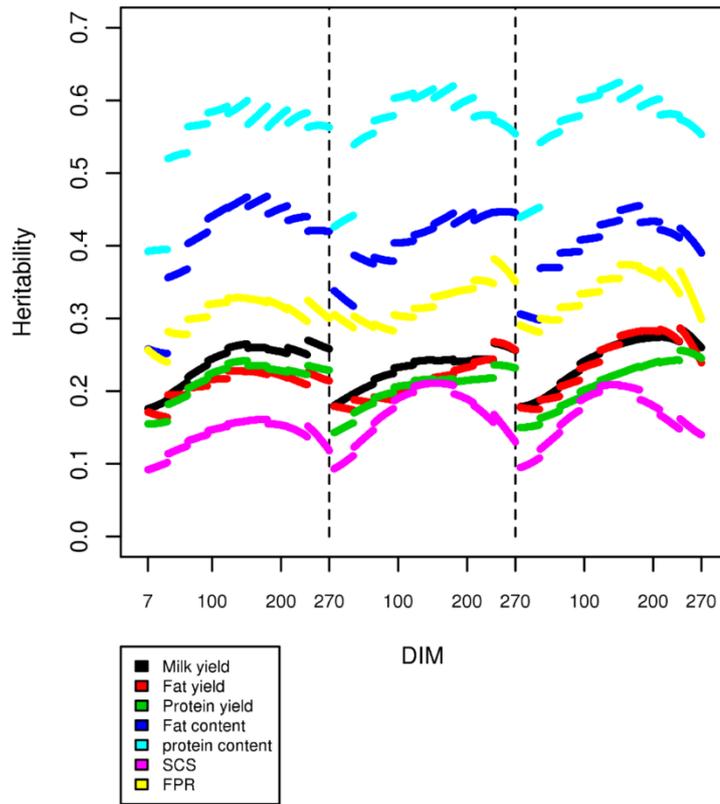
Arnal, Figure 1.



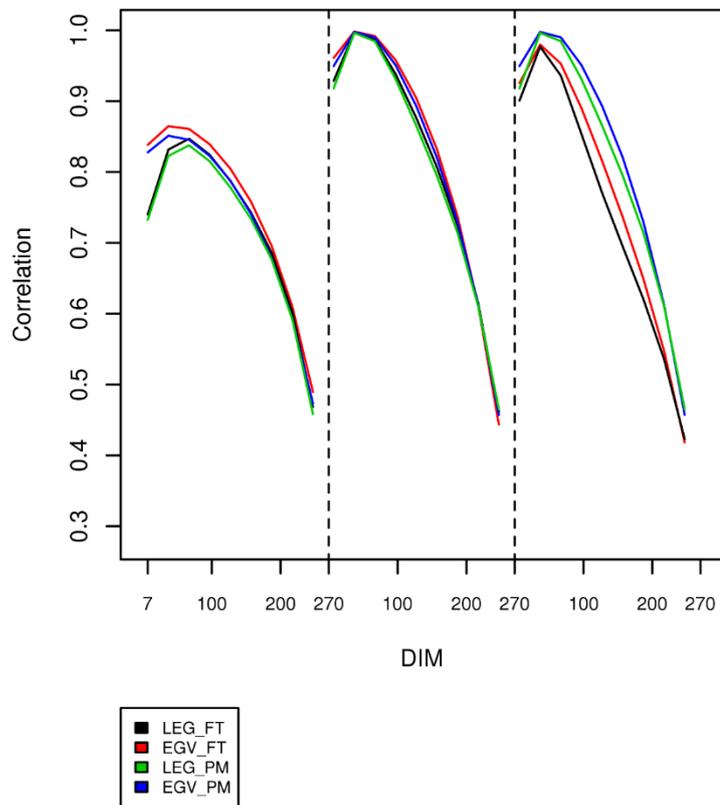
Arnal, Figure 2.



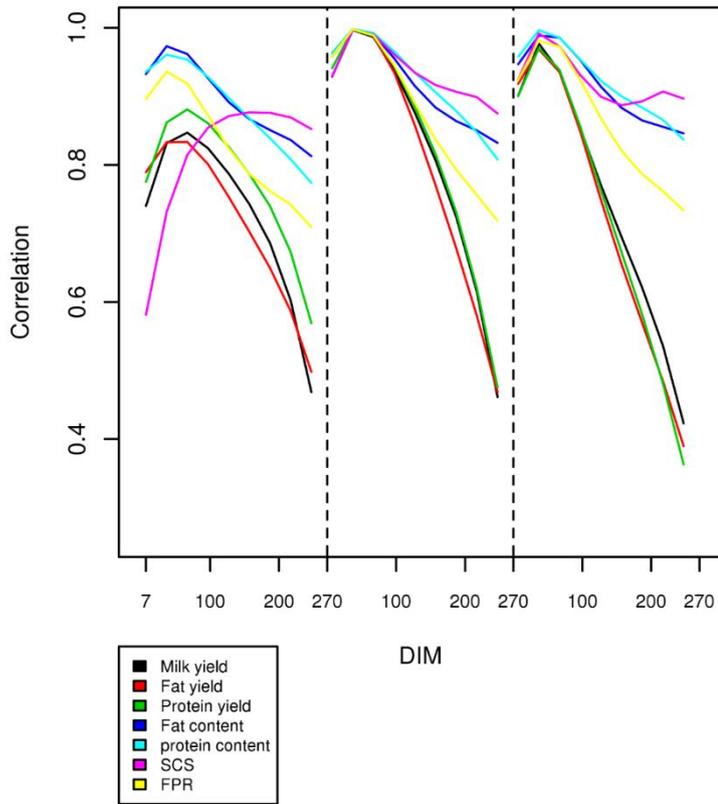
Arnal, Figure 3.



Arnal, Figure 4.



Arnal, Figure 5.



## II. Estimation des paramètres génétiques des caractères de production laitière et des comptages de cellules somatiques des trois premières lactations à l'aide d'un modèle de régression aléatoire chez les chèvres de race Saanen

Nous avons réalisé une étude similaire sur les données de race Saanen. Le nombre de contrôles élémentaires et de lactation étaient légèrement plus importants en Saanen qu'en Alpine (Tableau 7).

*Tableau 7. Nombre de contrôles élémentaires et de lactations par parité pour la race Saanen*

Parité	1	2	3
Nombre de contrôles	388 964	271 551	179 585
Nombre de Lactations	60 071	41 605	27 770

Nous présentons dans cette partie l'ensemble des résultats obtenus en race Saanen selon la même présentation que dans l'article précédent. Les mêmes modèles, ont été comparés.

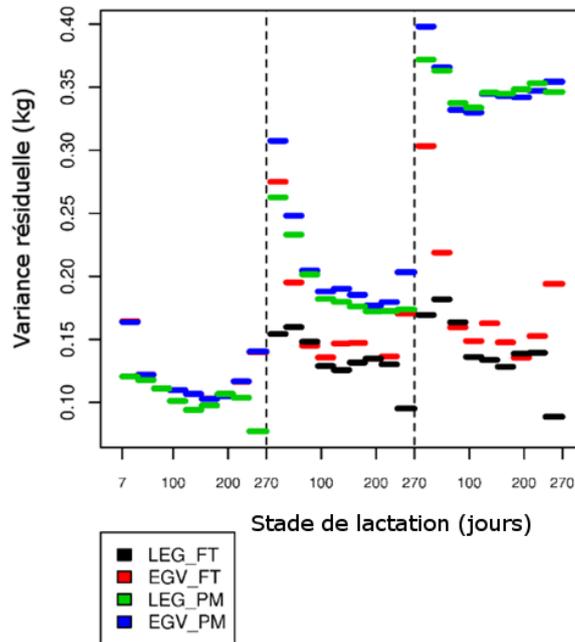
Les corrélations génétiques entre les coefficients de régression de la deuxième et de la troisième lactation sont présentées dans le Tableau 8. Les corrélations génétiques étaient supérieures à 0.95 pour le premier coefficient de LEG\_FT( $a_0$ ), et la première fonction propre relative au Niveau (EGV\_FT), qui représente la majorité de la variance génétique comme en race Alpine (Tableau 2 de l'article du chapitre 3). Pour le deuxième coefficient de LEG\_FT ( $a_1$ ), et la fonction propre relative à la persistance (EGV\_FT), la corrélation minimale était de 0.91 pour la MP et maximale (supérieure à 0.96) pour le TB, TP et le ratio TB/TP. Les corrélations génétiques entre le deuxième coefficient de la deuxième lactation et le deuxième coefficient de la troisième lactation étaient plus importantes en Saanen qu'en Alpine pour LEG\_FT (0.96 vs 0.83) et pour EGV\_FT (0.95 vs 0.73) pour les SCS. Comme pour la race Alpine, la deuxième et la troisième lactation peuvent être considérées comme un même niveau de facteur au vu des corrélations très importantes obtenues.

**Tableau 8. Estimations des corrélations génétiques entre les coefficients de régression des polynômes de Legendre ( $a_0$ ,  $a_1$ ,  $a_2$  dans LEG\_FT\*) de la deuxième parité avec ceux de la troisième parité et des corrélations génétiques entre les coefficients de régression des deux premières fonctions propres (Niveau, Persistance dans EGV\_FT\*\*) de la deuxième parité avec ceux de la troisième parité pour la race Saanen**

	Lait	MG	MP	TB	TP	SCS	Ratio TB/TP
<b>a<sub>0</sub></b>	0.96	0.95	0.95	0.99	0.99	0.97	0.99
<b>a<sub>1</sub></b>	0.92	0.94	0.91	0.98	0.97	0.96	0.98
<b>a<sub>2</sub></b>	0.97	0.93	0.99	0.93	0.92	0.80	0.92
<b>Niveau</b>	0.96	0.95	0.95	0.99	0.99	0.97	0.98
<b>Persistance</b>	0.92	0.94	0.91	0.97	0.97	0.95	0.98

\* LEG\_FT : Polynôme de Legendre d'ordre 2 pour chaque parité pour décrire les effets génétiques et d'environnement permanent, \*\* EGV\_FT : fonctions propres utilisées pour chaque parité pour décrire les effets génétiques et d'environnement permanent.

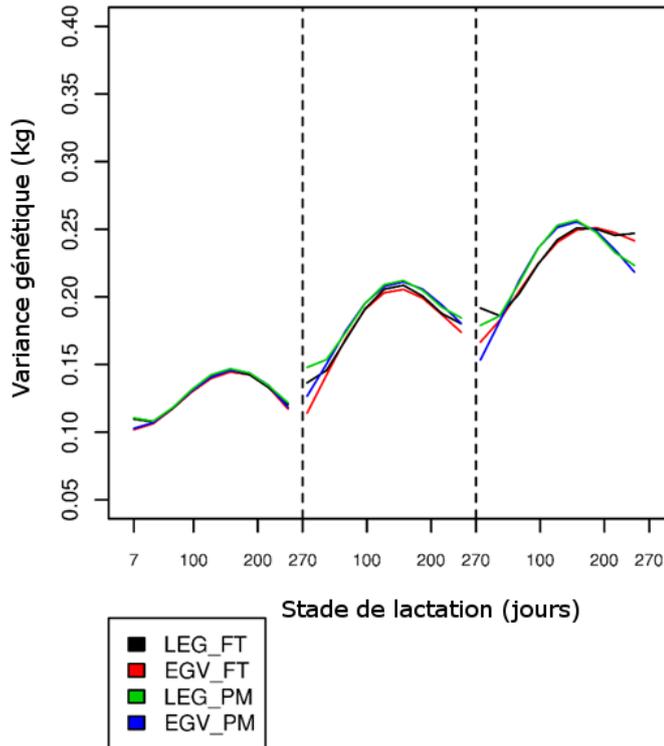
L'évolution de la variance résiduelle de la production laitière en Saanen est représentée sur la Figure 22. Elle est très proche des résultats obtenus en race Alpine. On note cependant une évolution différente de la variance résiduelle en troisième lactation, elle est stable autour de 0.35 kg<sup>2</sup> en Saanen alors qu'elle décroît jusqu'à 0.25 kg<sup>2</sup> en Alpine. Comme en race Alpine, le regroupement des lactations 2 et 3 dans un même niveau conduit à accroître la variance résiduelle, de manière modérée en lactation 2 et plus marquée en lactation 3.



**Figure 22: Evolution de la variance résiduelle estimée en fonction du stade de lactation pour les différents modèles pour la quantité de lait pour les chèvres Saanen.**

*LEG\_FT : Polynôme de Legendre utilisé pour la modélisation des effets génétique et de l'environnement permanent de la première à la troisième parité, LEG\_PM : Polynôme de Legendre utilisé pour la modélisation des effets génétique et de l'environnement permanent pour les primipares et multipares, EGV\_FT : fonctions propres utilisées pour la modélisation des effets génétique et de l'environnement permanent de la première à la troisième parité, EGV\_PM : fonctions propres utilisées pour la modélisation des effets génétique et de l'environnement permanent pour les primipares et les multipares*

Les évolutions des variances génétiques selon le modèle utilisé et présentées à la Figure 23, montrent relativement peu de différences en fonction des modèles. Contrairement à la race Alpine, les modèles regroupant les deuxièmes et les troisièmes lactations ensemble (LEG\_PM, EGV\_PM) présentent des évolutions similaires de la variance génétique que les modèles séparant toutes les lactations (LEG\_FT, EGV\_FT). De même, les évolutions des variances génétiques entre les modèles avec ou sans réduction des matrices de variances-covariances (EGV vs LEG) sont très proches. Au vu de ces résultats, regrouper dans un même niveau les lactations 2 et 3 et utiliser les fonctions propres à la place des polynômes de Legendre semblent pertinents.



**Figure 23. Evolution des variances génétiques en fonction du stade de lactation pour les quatre modèles pour la quantité de lait chez les chèvres Saanen.**

*LEG\_FT* : Polynôme de Legendre utilisé pour la modélisation des effets génétique et de l'environnement permanent de la première à la troisième parité, *LEG\_PM* : Polynôme de Legendre utilisé pour la modélisation des effets génétique et de l'environnement permanent pour les primipares et multipares, *EGV\_FT* : fonctions propres utilisées pour la modélisation des effets génétique et de l'environnement permanent de la première à la troisième parité, *EGV\_PM* : fonctions propres utilisées pour la modélisation des effets génétique et de l'environnement permanent pour les primipares et les multipares

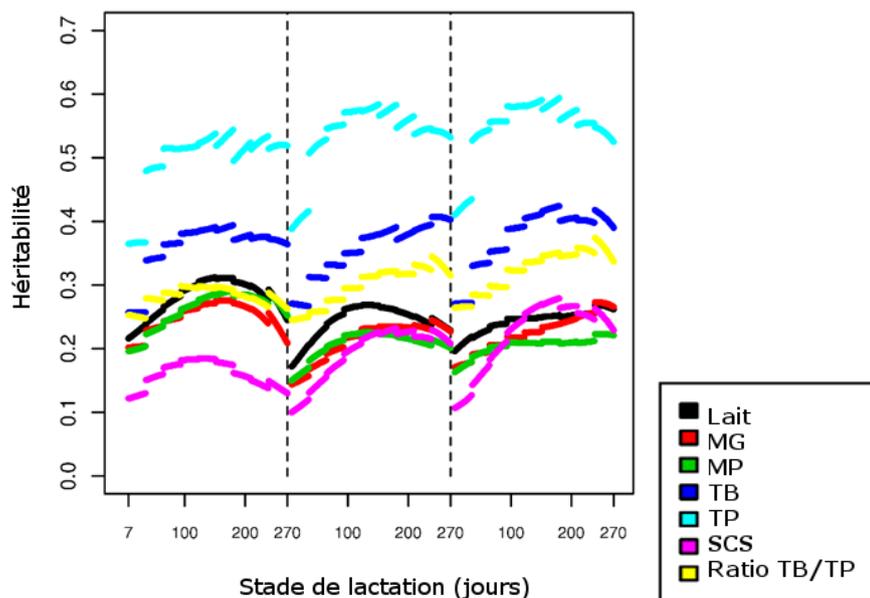
Le Tableau 9 présente les héritabilités de chaque composante des modèles *EGV\_FT* et *EGV\_PM* en race Saanen. Les héritabilités pour les quantités à la lactation (Niveau) sont proches de 0.3, 0.5 pour le TB et 0.6 pour le TP, 0.24 pour les SCS et autour de 0.4 pour le ratio TB/TP. Les héritabilités des quantités, des taux et des SCS sont proches de celles de la littérature (Rupp et al., 2011). Les héritabilités pour les persistances sont proches de 0.1 pour les quantités en primipares et autour de 0.15 en multipares, pour le TB et le ratio autour de 0.07, pour les SCS autour de 0.04 et enfin le TP a les héritabilités les plus élevées autour de 0.2. Les héritabilités de l'ensemble de ces composantes sont très proches de celles obtenues en Alpine pour tous les caractères et tous les rangs de lactation.

**Tableau 9. Héritabilité de chaque composante pour les modèles EGV\_FT et EGV\_PM pour chaque caractère chez les chèvres Saanen**

Modèle	Parité		Lait	MG	MP	TB	TP	SCS	Ratio TB/TP
EGV_FT	Première	NIV <sub>F</sub>	0.33	0.3	0.31	0.5	0.6	0.22	0.41
EGV_PM	Primipare	NIV <sub>P</sub>	0.31	0.29	0.29	0.49	0.59	0.22	0.4
EGV_FT	Première	PERS <sub>F</sub>	0.13	0.1	0.1	0.07	0.21	0.05	0.06
EGV_PM	Primipare	PERS <sub>P</sub>	0.1	0.08	0.08	0.07	0.2	0.05	0.06
EGV_FT	Deuxième	NIV <sub>S</sub>	0.29	0.27	0.25	0.49	0.63	0.25	0.43
EGV_FT	Troisième	NIV <sub>T</sub>	0.28	0.27	0.24	0.5	0.63	0.28	0.44
EGV_PM	Multipares	NIV <sub>M</sub>	0.3	0.28	0.26	0.49	0.64	0.27	0.43
EGV_FT	Deuxième	PERS <sub>S</sub>	0.15	0.11	0.11	0.07	0.2	0.04	0.06
EGV_FT	Troisième	PERS <sub>T</sub>	0.18	0.14	0.15	0.08	0.2	0.05	0.08
EGV_PM	Multipares	PERS <sub>M</sub>	0.16	0.11	0.11	0.07	0.21	0.05	0.07

*EGV\_FT : fonctions propres utilisées pour chaque parité pour décrire les effets génétiques et de l'environnement permanent, EGV\_PM : fonctions propres utilisées pour décrire les effets génétiques et de l'environnement permanent, NIV : coefficient de régression génétique additive aléatoire pour la première fonction propre, PERS : coefficient de régression génétique additive aléatoire pour la deuxième fonction propre, SCS : score de cellules somatiques*

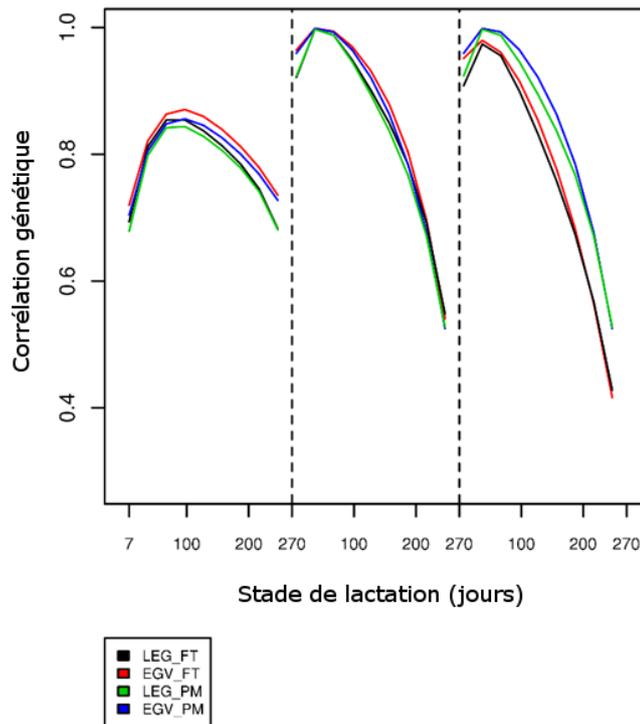
Les évolutions des héritabilités au cours des 3 lactations en race Saanen pour l'ensemble des caractères étudiés sont présentées à la Figure 24. Les évolutions sont très proches de celles obtenues en race Alpine. Cependant, on peut noter en multipares une héritabilité des comptages de cellules somatiques plus élevée en fin de lactation en race Saanen comparée à la race Alpine (autour de 0.2 contre 0.1 en Alpine).



**Figure 24. Evolution des hérabilités au cours du temps pour les caractères de production laitière (Lait, MG, MP, TB et TP), les scores de cellules somatiques (SCS) et le ratio TB/TP avec le modèle LEG\_FT\* chez les chèvres Saanen.**

*LEG\_FT\* : Polynôme de Legendre utilisé pour la modélisation des effets génétique et de l'environnement permanent de la première à la troisième parité*

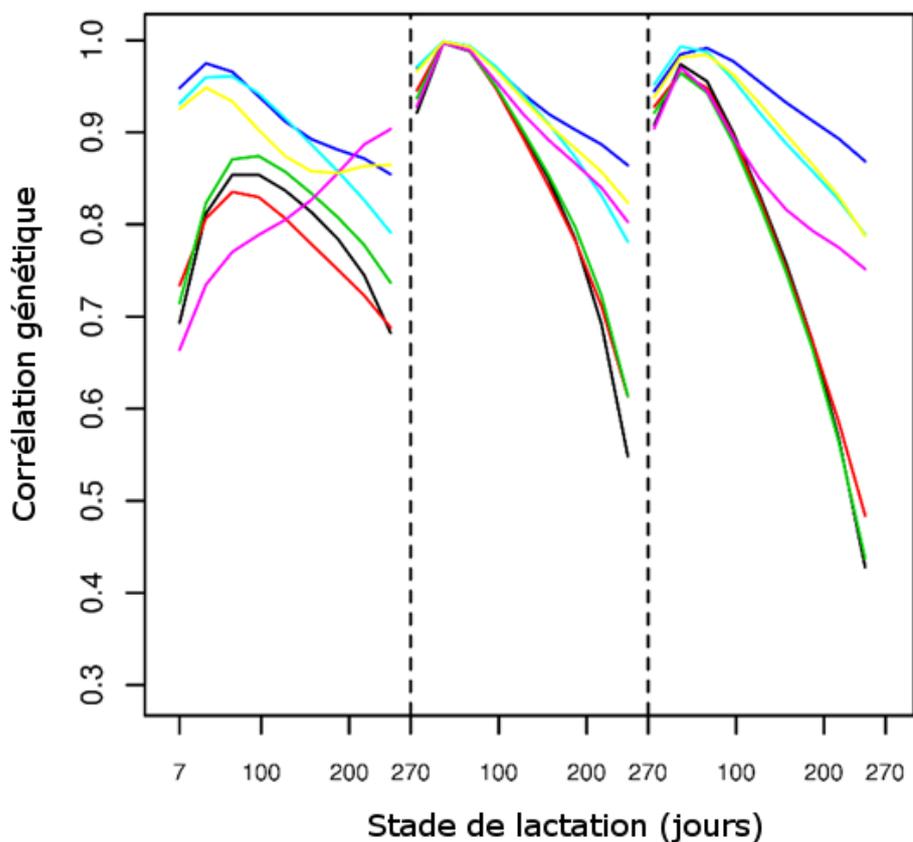
Les corrélations génétiques pour la quantité de lait entre le 40<sup>ième</sup> jour de lactation en deuxième lactation et les autres jours de lactation étaient similaires quel que soit le modèle utilisé en race Saanen (Figure 25). Ces résultats étaient similaires et de même ordre de grandeur qu'en race Alpine. La seule différence observée concerne la corrélation entre le jour 40 de la deuxième lactation et le jour 270 de la première lactation, qui était de 0.45 en Alpine et 0.7 en Saanen. Enfin, les résultats entre les 4 modèles étant très peu différents, un regroupement des lactations 2 et 3 dans un même niveau de facteur (PM) et une réduction de la matrice de variance-covariance des effets génétiques et d'environnement permanent (EGV) peuvent être envisagés pour modéliser les contrôles élémentaires de la quantité de lait à l'aide d'un modèle de régression aléatoire.



**Figure 25. Corrélations génétiques entre le stade de lactation 40 de la seconde parité et les autres stades de lactation selon les différents modèles (LEG\_FT, EGV\_FT, LEG\_PM et EGV\_PM) pour la quantité de lait des chèvres de race Saanen**

*LEG\_FT : Polynôme de Legendre d'ordre 2 pour chaque parité pour décrire les effets génétiques et d'environnement permanent, LEG\_PM : Polynôme de Legendre d'ordre 2 pour les primipares et les multipares pour décrire les effets génétiques et d'environnement permanent, EGV\_FT : fonctions propres utilisées pour chaque parité pour décrire les effets génétiques et d'environnement permanent, EGV\_PM : fonctions propres utilisées pour les primipares et les multipares pour décrire les effets génétiques et d'environnement permanent*

Les corrélations génétiques entre le stade 40 de la deuxième lactation et les autres stades de lactation pour les autres caractères (MG, MP, TB, TP, SCS et ratio TB/TP) et avec le modèle LEG\_FT sont présentées à la Figure 26 en Saanen. Les évolutions des corrélations sont très proches entre les deux races (en valeurs et forme d'évolution). Quelques petites différences peuvent être malgré tout notées. Pour les scores de cellules somatiques (SCS), la corrélation des différents stades de première lactation avec le 40<sup>ième</sup> jour de la deuxième lactation augmente de façon régulière passant de 0.7 à 0.9 alors qu'en Alpine la corrélation passe de 0.6 à 0.85 au stade 110 puis se maintient à ce niveau. En multipares, toujours pour les SCS, la corrélation entre le jour 40 et le jour 270 est de 0.9 en deuxième comme en troisième lactation en Alpine contre 0.8 en Saanen en deuxième lactation et 0.75 en troisième lactation. Les niveaux de corrélation entre le jour 40 en deuxième lactation et le jour 270 en première lactation des matières MG et MP sont du même ordre de grandeur que le niveau de corrélation pour la quantité de lait (Figure 25), de 0.7 en Saanen, contre 0.45 en race Alpine.



**Figure 26.** *Corrélations génétiques avec le modèle LEG\_FT\* entre le stade de lactation 40 en deuxième lactation et tous les stades de lactation pour les caractères étudiés (Lait, MG, MP, TB, TP, SCS et ratio TB/TP) pour les chèvres de race Saanen*

*SCS : score des cellules somatiques*

*LEG\_FT\* : Polynôme de Legendre d'ordre 2 pour chaque parité pour décrire les effets génétiques et d'environnement permanent*

Le Tableau 10 reporte pour la race Saanen les corrélations pour la quantité de lait entre les index génétiques pour la quantité totale selon les 4 modèles étudiées (LEG\_FT, EGV\_FT, LEG\_PM et EGV\_PM). Plusieurs index génétiques ont été calculés : 3-lact-EBV représente la somme des index de chaque stade de lactation pour toutes les parités, 3-lact-NIV-EBV représente la somme du coefficient de régression génétique additive aléatoire pour la première fonction propre pour toutes les parités (pour LEG\_PM et EGV\_PM, le coefficient de régression génétique additive aléatoire pour la première

fonction propre chez les primipares était multiplié par 0.33 et celle des multipares par 0.66), 3-lact-PERS-EBV représente la somme des coefficients de régression génétique additive aléatoire pour la deuxième fonction propre pour toutes les parités (pour LEG\_PM et EGV\_PM, le coefficient de régression génétique additive aléatoire pour la deuxième fonction propre des primipares était multiplié par 0.33 et celui des multipares par 0.66). Les corrélations entre les index équivalents (cités ci-dessus) pour les chèvres de race Saanen issues des 4 modèles étudiés sont très proches de 1 (entre 0.99 et 1) comme celles obtenues en race Alpine. Si comme en race Alpine le niveau laitier est très fortement corrélé à la quantité totale produite, les corrélations entre niveaux et persistances ou entre 3-lact-EBV et persistances étaient en revanche plus fortes en Saanen qu'en race Alpine (proche de -0.2 en Saanen contre 0 en Alpine).

**Tableau 10. Corrélations entre index des boucs (n = 486) selon les 4 modèles étudiés pour la quantité de lait des chèvres de race Saanen**

	LEG_FT	LEG_PM	EGV_FT	EGV_PM	EGV_FT
	3-lact-EBV	3-lact-EBV	3-lact-NIV-EBV	3-lact-NIV-EBV	3-lact-PERS-EBV
LEG_PM	3-lact-EBV	0.99			
EGV_FT	3-lact-NIV-EBV	1	0.99		
EGV_PM	3-lact-NIV-EBV	0.99	0.99	1	
EGV_FT	3-lact-PERS-EBV	-0.19	-0.25	-0.17	-0.19
EGV_PM	3-lact-PERS-EBV	-0.17	-0.24	-0.16	-0.17
					1

*LEG\_FT* : Polynôme de Legendre d'ordre 2 pour chaque parité pour décrire les effets génétiques et d'environnement permanent, *LEG\_PM* : Polynôme de Legendre d'ordre 2 pour les primipares et les multipares pour décrire les effets génétiques et d'environnement permanent, *EGV\_FT* : fonctions propres utilisées pour chaque parité pour décrire les effets génétiques et d'environnement permanent, *EGV\_PM* : fonctions propres utilisées pour les primipares et les multipares pour décrire les effets génétiques et d'environnement permanent, *3-lact-EBV* : somme des index de chaque stade de lactation pour toutes les parités, *3-lact-NIV-EBV* : somme du coefficient de régression génétique additive aléatoire pour la première fonction propre pour toutes les parités (pour LEG\_PM et EGV\_PM, le coefficient de régression génétique additive aléatoire pour la première fonction propre chez les primipares était multiplié par 0.33 et celle des multipares par 0.66), *3-lact-PERS-EBV* : somme des coefficients de régression génétique additive aléatoire pour la deuxième fonction propre pour toutes les parités (pour LEG\_PM et EGV\_PM, le coefficient de régression génétique additive aléatoire pour la deuxième fonction propre des primipares était multiplié par 0.33 et celui des multipares par 0.66)

Les corrélations entre les valeurs génétiques pour la quantité de lait et les autres caractères sont reportées au Tableau 11. Les corrélations entre niveau de lait et les niveaux des matières sont importantes (au-dessus de 0.76), les corrélations entre niveau de lait et taux sont inférieures à -0.17, les corrélations avec les SCS et le ratio TB/TP sont inférieures à 0.1 et proches de 0. Les corrélations entre la persistance laitière et les persistances des niveaux sont très fortes (au-dessus de 0.89), la corrélation entre la

persistance du TB et la persistance du lait est nulle. En revanche, la corrélation entre la persistance du lait et du TP est proche de -0.58, avec celle du ratio elle est de 0.4 et avec les SCS elle est de -0.34. Les corrélations entre niveau et persistances sont faibles, comprises entre -0.21 (entre persistance du TB et niveau de lait) et 0.18 (entre niveau du TB et persistance du lait) à une exception près celle du niveau de SCS et la persistance du lait (-0.37). Les corrélations montrent donc peu de différences avec celles estimées en race Alpine à l'exception toujours de la corrélation entre niveau de SCS et persistance du lait. Enfin, les corrélations entre les valeurs génétiques relatives au niveau de production laitière sont cohérentes avec celles rapportées dans la littérature (Bélichon et al., 1999).

**Tableau 11. Corrélations d'index pour les boucs Saanen (n = 486) obtenus avec le modèle EGV\_PM entre la production laitière et les autres caractères (MG, MP, TB, TP, SCS, ratio TB/TP)**

		Lait	
		3-lact-LEV-EBV	3-lact-PERS-EBV
MG	3-lact-LEV-EBV	0.76	0.06
	3-lact-PERS-EBV	-0.08	0.89
MP	3-lact-LEV-EBV	0.89	-0.02
	3-lact-PERS-EBV	-0.08	0.94
TB	3-lact-LEV-EBV	-0.17	0.19
	3-lact-PERS-EBV	-0.21	-0.01
TP	3-lact-LEV-EBV	-0.33	0.02
	3-lact-PERS-EBV	-0.08	-0.58
SCS	3-lact-LEV-EBV	0.1	-0.37
	3-lact-PERS-EBV	-0.17	-0.34
ratio TB/TP	3-lact-LEV-EBV	0.04	0.18
	3-lact-PERS-EBV	-0.08	0.4

En conclusion, nous pouvons noter que les quatre modèles étudiés se comportent de la même façon en races Saanen et Alpine. Les paramètres génétiques sont globalement similaires pour l'ensemble des caractères: héritabilités, corrélations génétiques, corrélations entre caractères, variances génétiques. En revanche, de légères différences de paramètres génétiques entre Alpine et Saanen ont été observées: corrélations entre caractères, héritabilités, corrélations entre jours de lactation.

### **III. Études complémentaires réalisées lors de la mise en œuvre du modèle de régression aléatoire**

L'objet des résultats présentés dans cette partie est de compléter l'étude de la mise en œuvre du modèle de régression aléatoire pour les données de contrôles élémentaires sur les lactations 1 à 3 des chèvres Alpine et Saanen, présentée dans les parties 1 et 2, par une étude sur d'une part la définition de

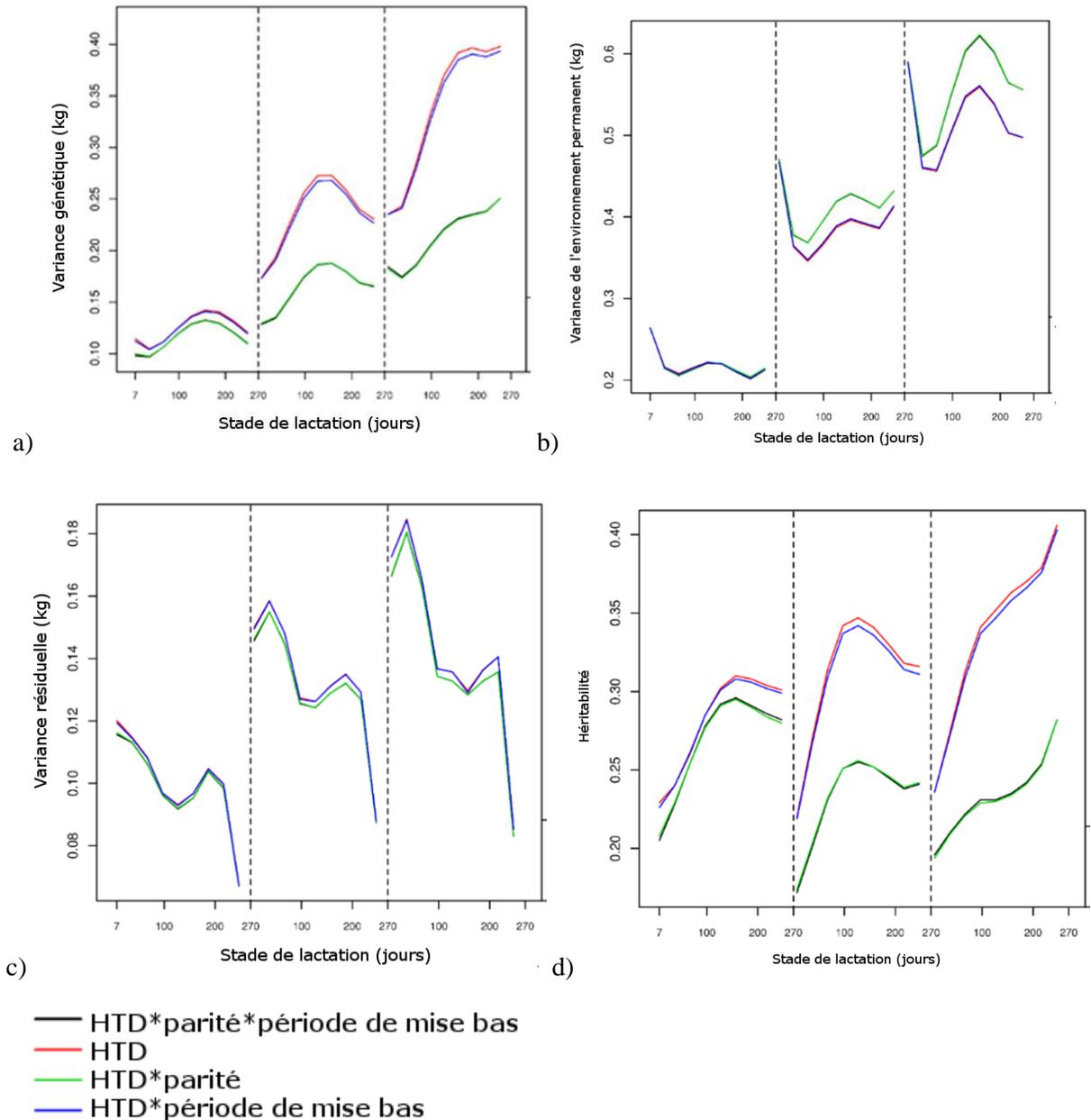
l'effet troupeau inclus dans le modèle et une étude sur les variances lorsque les lactations 2 et 3 sont considérées comme un seul et même niveau (modèle PM) d'autre part.

### **1. L'effet Troupeau-Jour de contrôle (HTD) doit être en interaction avec le rang de lactation**

Comme évoqué au chapitre 2, la définition de l'effet troupeau est un facteur important à prendre en compte dans le modèle. L'extension du modèle à 3 lactations a nécessité une redéfinition de cet effet (effet HTD en interaction avec le rang de lactation, parties I et II de ce chapitre) et une analyse fine que nous détaillons ci-dessous.

Les résultats sont présentés pour la production laitière en race Saanen à partir du modèle de référence (LEG\_FT). 4 définitions de l'effet troupeau ont été testés et comparés : un effet HTD simple (identique quel que soit le rang de lactation), un effet HTD en interaction avec l'effet de la période de mise bas (année civile divisée en deux périodes de six mois pour prendre en compte des lots différents au sein des troupeaux s'il y a deux saison de mise bas par an), un effet HTD en interaction avec le rang de lactation et un effet HTD en interaction avec le rang de lactation et la période de mise bas.

La Figure 27 présente les évolutions de la variance génétique (a), de la variance de l'environnement permanent (b), de la variance résiduelle (c), et de l'héritabilité (d) selon le stade de lactation et la modélisation de l'effet troupeau dans le cadre du modèle LEG\_FT. L'interaction entre HTD et le rang de lactation (courbes vertes et noire) modifie de façon considérable, en particulier en multipares, les évolutions des variances génétiques, des variances de l'environnement permanent et de l'héritabilité comparé à la non interaction entre le rang de lactation et HTD (courbes rouge et bleue). Lorsque l'effet HTD est en interaction avec le rang de lactation (courbe noire et verte), la variance génétique est diminuée, la variance de l'environnement permanent est augmentée et l'héritabilité est diminuée par rapport à la non interaction entre HTD et le rang de lactation (courbes rouge et bleue). La variance résiduelle est peu affectée par la modélisation de l'effet troupeau mais l'interaction avec le rang de lactation la diminue légèrement ce qui signifie que l'ajustement aux données est meilleur. L'interaction entre HTD et la période de mise bas (courbes bleue et noire) n'induit pas de changement important dans les évolutions des variances et d'héritabilité par rapport à un modèle avec un effet HTD simple (courbes rouge et verte). L'interaction entre HTD et le rang de lactation semble donner les meilleurs résultats.

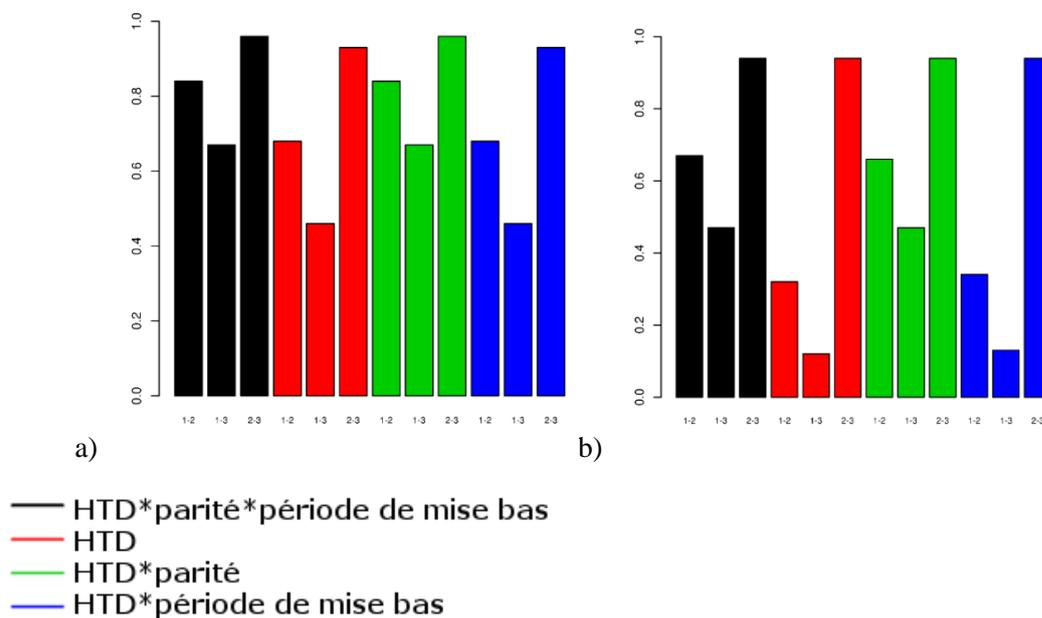


**Figure 27: Evolution de la variance génétique (a), de la variance de l'environnement permanent (b), de la variance résiduelle (c), et de l'héritabilité (d) de la production laitière en race Saanen selon le stade de lactation et la définition de l'effet troupeau dans un modèle LEG\_FT.**

En noir: HTD est défini en interaction avec la période de mise bas et le rang de lactation, en rouge: HTD, en vert: HTD est en interaction avec le rang de lactation, en bleu: HTD est en interaction avec la période de mise bas. La courbe verte et noire sont quasiment confondues dans tous les graphes

La Figure 28 présente les corrélations génétiques pour la production laitière en Saanen entre les coefficients de régression entre deux rangs de lactation pour la première covariable de Legendre  $\varphi_{(0,t)}$ , et pour la deuxième covariable  $\varphi_{(1,t)}$ , selon la définition de l'effet troupeau dans un modèle LEG\_FT. Lorsque l'interaction entre HTD et le rang de lactation est prise en compte (en noir et vert), les

corrélations génétiques entre parités des coefficients de régression pour la première covariable de Legendre sont plus élevées que lorsque l'interaction n'est pas prise en compte (en rouge et bleu) (corrélation entre première et troisième lactation de 0.65 avec l'interaction et de 0.45 sans l'interaction). Selon la prise en compte de l'interaction entre HTD et la parité, les différences pour les corrélations génétiques entre parités des coefficients de régression pour la deuxième covariable de Legendre sont encore plus marquées que pour la première covariable. En effet, lorsque l'interaction entre HTD et le rang de lactation est prise en compte (en noir et vert), les corrélations génétiques entre parités des coefficients de régression pour la deuxième covariable de Legendre sont plus élevées que lorsque l'interaction n'est pas prise en compte (en rouge et bleu) (corrélacion entre première et troisième lactation de 0.45 avec l'interaction et de 0.1 sans l'interaction).



**Figure 28 : Corrélations génétiques pour la production laitière en race Saanen entre les coefficients de régression entre deux rangs de lactation : pour la première covariable de Legendre  $\varphi_{(0,t)}$  (a), et pour la deuxième covariable  $\varphi_{(1,t)}$  (b), en fonction de la définition de l'effet troupeau dans un modèle LEG\_FT.**

En noir: HTD est défini en interaction avec la période de mise bas et le rang de lactation, en rouge: HTD, en vert: HTD est en interaction avec le rang de lactation et en bleu: HTD est en interaction avec la période de mise bas.

L'augmentation de la corrélation entre les coefficients de régression de la première parité et ceux des suivantes est la raison qui nous a amené à choisir de conserver dans le modèle LEG\_FT l'interaction entre l'effet HTD et la parité. En effet, les classements des boucs ne devraient pas être fortement impactés avec l'arrivée des données de multipares. La prise en compte dans le modèle d'une interaction de l'effet troupeau avec le rang de lactation est nécessaire afin de bien prendre en compte l'effet des contemporaines. Ce choix se justifie par la conduite des animaux dans un élevage. En effet,

les troupeaux en caprins sont souvent de grande taille, ils sont gérés par lots d'animaux homogènes avec souvent un lot de primipares et un lot de multipares.

Dans le modèle développé par H. Leclerc (2008) pour les bovins laitiers, l'effet HTD n'est pas défini en interaction avec le rang de lactation. En revanche, leur modèle possède un effet troupeau dépendant du stade de lactation en interaction avec le rang de lactation. Le niveau du troupeau en interaction avec le rang de lactation était ainsi pris en compte. Nous n'avons pas choisi d'opter pour un tel modèle par soucis de simplification du modèle avec moins d'effets aléatoires comme vu au chapitre 2 pour l'analyse des données primipares.

## **2. Ajustement des covariables de Legendre pour combiner les rangs de lactation 2 et 3**

Dans ce paragraphe nous allons étudier les développements des modèles regroupant les deuxièmes et les troisièmes lactations ensemble (EGV\_PM et LEG\_PM). Nous utiliserons toujours les mêmes corpus de données en Alpine et en Saanen. Comme nous l'avons vu dans l'article 3 (partie I de ce chapitre) en race Alpine, les corrélations génétiques entre les polynômes de Legendre entre la deuxième et la troisième lactation sont très élevées, au-dessus de 0.9 (Table 2, article 3). Ce résultat était le même en Saanen (II. Tableau 8). Ces corrélations très élevées permettent de modéliser ensemble les deux rangs de lactation comme nous l'avons fait avec LEG\_PM et EGV\_PM. En revanche, les variances génétiques et de l'environnement permanent sont plus importantes en troisième lactation qu'en deuxième lactation pour les quantités (lait, MG et MP) comme montré dans les parties précédentes (Figure 2, article 3).

En regroupant les deuxièmes et les troisièmes lactations ensemble dans la modélisation, la variance génétique de la troisième lactation se trouvait réduite conséquemment. Afin d'estomper ces différences de variance pour la troisième lactation, nous avons proposé d'utiliser un coefficient multiplicateur pour les coefficients des polynômes de Legendre en troisième lactation. Le Tableau 12 présente les valeurs du critère BIC pour les différentes valeurs du coefficient multiplicateur testées sur l'ensemble des caractères étudiés en race Alpine et Saanen.

Nous avons testé des coefficients multiplicateurs de 1 à 1.2 avec un pas de 0.05. Le meilleur modèle est le modèle qui a le BIC le plus faible. Les BIC étaient les plus faibles pour les coefficients 1.1 pour le lait, 1.2 pour les matières et 1 pour les autres caractères en Alpine. Ces coefficients ont donné des BIC faibles également en Saanen. Aux vues de ces résultats et afin d'avoir des coefficients multiplicateurs identiques pour les 2 races, nous avons choisi de retenir un coefficient multiplicateur de 1.1 pour le lait, 1.2 pour les matières (MG et MP) et 1 pour les autres caractères (TB, TP, SCS, ratio TB/TP).

La prise en compte de ces coefficients multiplicateurs dans les modèles (PM), permet de mieux modéliser la variance génétique en troisième lactation et induit des variances différentes en deuxième et

troisième lactation malgré une modélisation commune des deuxièmes et troisièmes lactations (Figure 2 de l'article 3 et Figure 23 du II.).

**Tableau 12 : BIC\* du modèle LEG\_PM selon différentes valeurs du coefficient multiplicateur testé et appliqué aux coefficients de Legendre de troisième lactation pour les caractères (LAIT, MG, MP, TB, TP, SCS, ratio TB/TP) en race Saanen et Alpine.**

Coefficient Multiplicateur	Saanen							Alpine						
	LAIT	MG	MP	TB	TP	SCS	ratio TB/TP	LAIT	MG	MP	TB	TP	SCS	ratio TB/TP
<b>1</b>	126	1023	462	0	214	0	0	316	749	720	0	0	0	0
<b>1.05</b>	0	561	191	168	0	493	14	98	388	385	146	143	532	48
<b>1.1</b>	15	244	45	569	370	1136	218	0	148	162	498	794	1151	247
<b>1.15</b>	138	61	0	1172	1255	1904	587	1	22	38	1029	1883	1840	579
<b>1.2</b>	338	0	40	1946	2584	2777	1100	83	0	0	1713	3342	2589	1026

\*pour chaque caractère étudié, la meilleure valeur du critère BIC a été mise à 0 et les valeurs présentées dans le tableau sont en écart par rapport à cette valeur du BIC.

Une autre réduction était possible en décomposant en valeurs propres la matrice de variances-covariances génétique du modèle EGV\_FT (de dimension 6×6) et en ne gardant que les 4 premières valeurs propres, comme cela a été fait chez les bovins laitiers (Leclerc et al., 2009). A partir des 4 vecteurs propres nous pouvons calculer des fonctions propres et les utiliser comme nouvelles covariables. Cependant, les coefficients de régression de ces fonctions propres ne sont pas facilement interprétables biologiquement comme c'était le cas pour le niveau et la persistance. Cette réduction nécessite de transformer les coefficients de régression afin qu'ils soient de nouveau interprétables en terme de niveau et de persistance. Afin d'éviter cette étape, nous avons donc préféré multiplier les coefficients de Legendre par les coefficients multiplicateurs comme précisé ci-dessus aux vues des paramètres génétiques inchangés entre EGV\_FT et EGV\_PM.

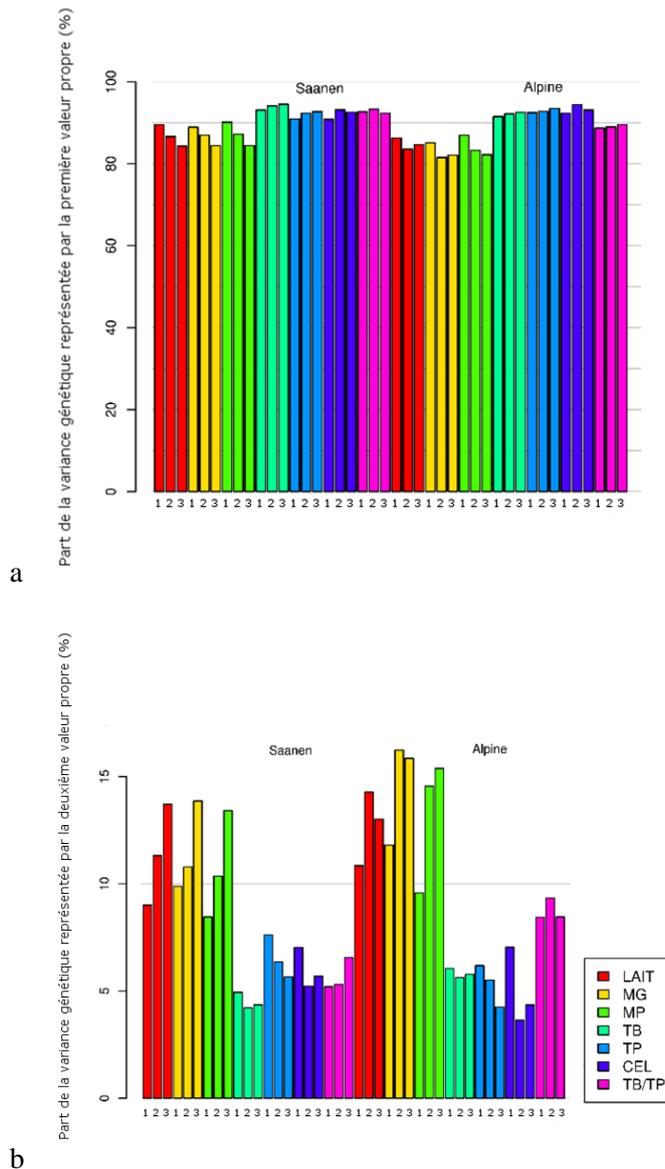
#### **IV. Analyses complémentaires sur les variances génétiques et les effets de milieux issus des modèles LEG\_FT et EGV\_FT pour les caractères étudiés (Lait, MG, MP, TB, TP, SCS et ratio TB/TP) en race Alpine et Saanen**

##### **1. Etude complémentaire sur les variances génétiques**

###### **a. Pourcentage de variance génétique représenté par les trois premières valeurs propres des matrices de variances-covariances issues du modèle LEG\_FT par rang de lactation en race Alpine et Saanen**

Le pourcentage de variance génétique représentée par chacune des valeurs propres a été étudié par caractère et par rang de lactation pour les deux races à partir du modèle de référence LEG\_FT. Les Figure 29.a et Figure 29.b présentent les parts de variances génétiques pour les 2 premières valeurs

propres respectivement. Le pourcentage de variance représenté par la première valeur propre (Figure 29.a) est maximal pour les taux (TB, TP) et les SCS, quel que soit le rang de lactation, dans les deux races (supérieur à 90%). Ce pourcentage est minimal pour les quantités (MG, MP) en troisième lactation (autour de 80%). Le pourcentage de variance génétique représenté par la deuxième valeur propre (Figure 29.b) est supérieur à 8% pour les quantités (Lait, MG, MP) et atteint 16% pour la MG en race Alpine. Il est intéressant de noter que ce pourcentage est minimal en première lactation. En revanche, pour les taux et les SCS, ce pourcentage est proche de 5%. La troisième valeur propre (résultats non montrés) représente un très faible pourcentage de la part de variance génétique (au maximum 3.5%) et peu de différences sont notables entre caractères et rang de lactation pour les deux races.



**Figure 29: Part de variance génétique représentée par la première valeur propre (a) et la deuxième valeur propre (b) selon le rang de lactation et le caractère en race Alpine et Saanen**

### **b. Profils des fonctions propres issues du modèle EGV\_FT et variabilité génétique du niveau et de la persistance pour les différents caractères étudiés en race Alpine et Saanen**

Les fonctions propres, issues du modèle de régression EGV\_FT, ont été multipliées par plus ou moins un écart type génétique du premier (niveau) ou du deuxième (persistance) coefficient de régression du modèle pour chaque lactation, chaque caractère et chacune des deux races. Les profils des fonctions sont représentés sur la Figure 30. Des différences de formes, d'évolution des fonctions et d'écart types génétiques sont observables entre les différentes lactations et entre les deux races.

Pour la première fonction propre (représentant le niveau de production), les Figures 30.a, b et c montrent qu'il existe une plus grande variabilité génétique pour les quantités de lait, de matière grasse et de matière protéique en race Saanen comparé à la race Alpine. A l'inverse pour les taux, le TB (Figure 30.d) et le TP (Figure 30.e), les chèvres Alpines présentent une variabilité génétique supérieure du niveau de production par rapport aux chèvres de race Saanen.

Pour la deuxième fonction propre (représentant la persistance) et pour les quantités de lait (Figures 30 a, b, c), les chèvres Saanen présentent une plus grande différence de production dans les tout premiers jours de lactation qu'en toute fin de lactation. Cette différence est moins marquée pour les chèvres de race Alpine. On retrouve également pour le ratio TB/TP un différentiel de ratio en tout début de lactation plus important qu'en fin de lactation. Il y a une différence marquée entre les quantités et les taux pour la deuxième composante. En effet, pour les quantités, le différentiel de production en début et en fin de lactation est égal à un écart type de la première composante alors que pour le TB, le TP, et les SCS en Saanen, le différentiel de production en début et en fin de lactation est égal à la moitié de l'écart type de la première composante. Ces différences sont en accord avec les résultats observés sur les pourcentages représentés par les différentes valeurs propres. La persistance représentait une part supérieure pour les quantités de lait (autour de 10%) que pour les taux (TB, TP) (autour de 5%). Et elles peuvent également être reliées aux plus faibles corrélations génétiques entre stade de lactation pour les quantités (corrélation génétique autour de 0.6 entre le jour 40 et le jour 270 en deuxième lactation) comparées aux taux (corrélation génétique de 0.8 entre le jour 40 et le jour 270 en deuxième lactation).

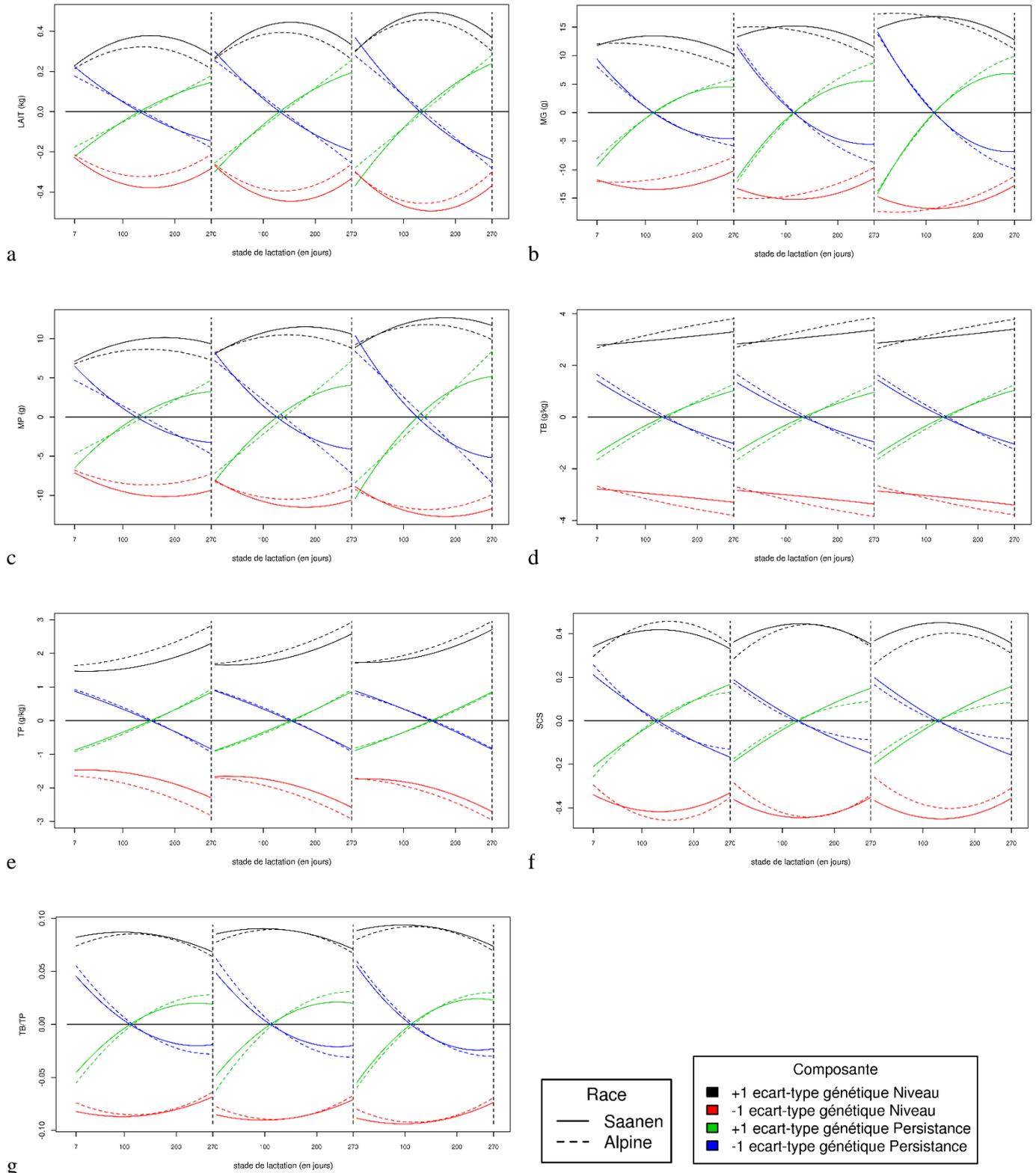
### **c. Représentation de courbes phénotypiques pour des animaux extrêmes**

Nous avons ensuite ajouté la courbe de production phénotypique correspondante (obtenue par une modélisation des données phénotypiques à l'aide d'une spline naturelle) aux index extrêmes issus du modèle LEG\_FT ( $\pm 2$  écarts types génétiques du coefficient de régression de la première et de la deuxième fonction propre) afin de visualiser la courbe phénotypique d'animaux génétiquement extrêmes pour la forme de leur courbe de lactation. Ce type de courbes peut être utile pour communiquer auprès des éleveurs. Les courbes moyennes d'animaux extrêmes sont représentées sur la Figure 31. Nous

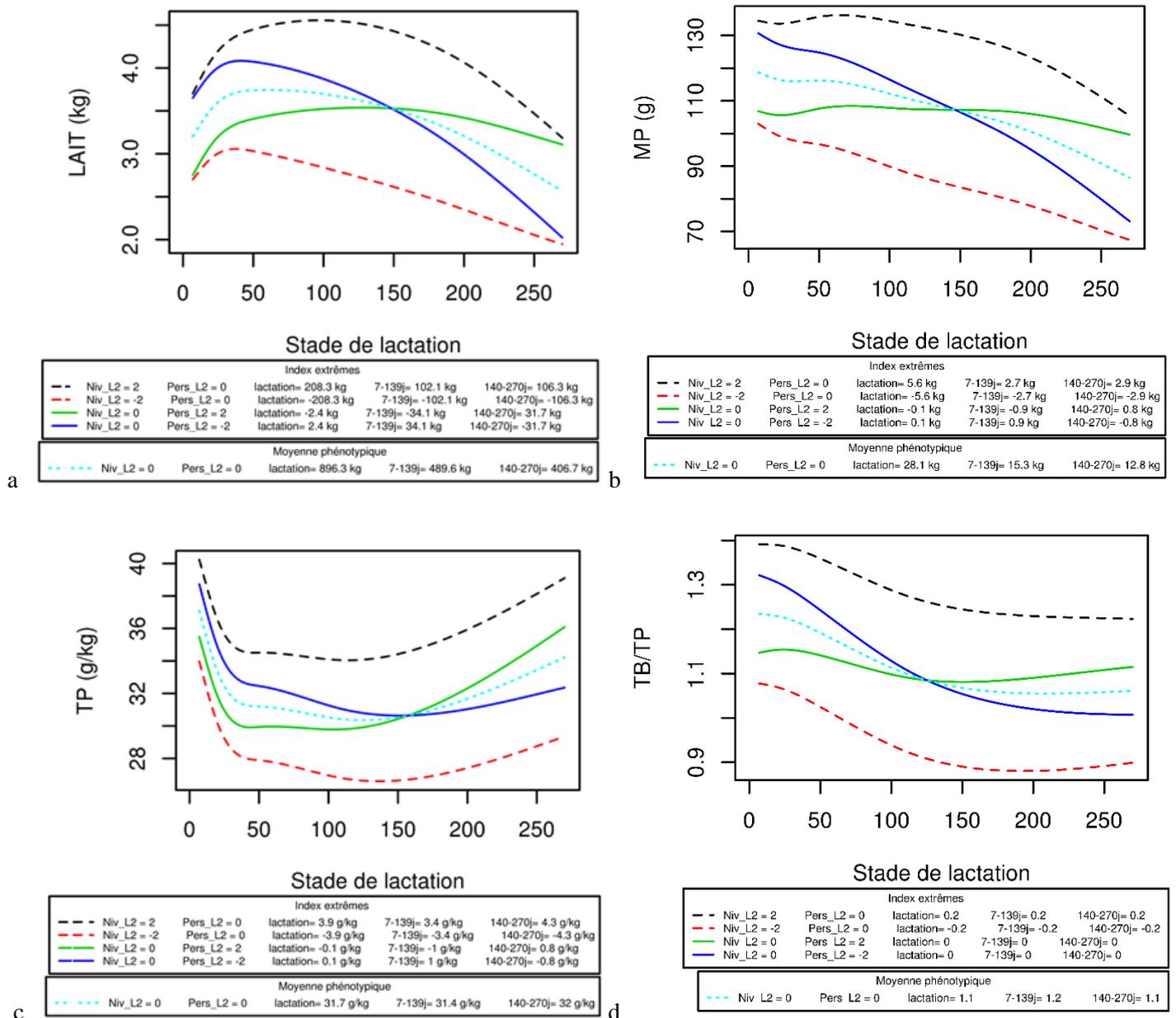
observons une grande variabilité des courbes de lactation pour des animaux extrêmes de 2 écarts-types génétiques pour le niveau ou la persistance quel que soit le caractère. Elles illustrent les gains potentiels d'une sélection de la persistance laitière pour différents caractères laitiers. Pour le caractère lait, les sommes d'index par période nous montrent l'importance du lait perdu en début de lactation par une chèvre avec une bonne persistance (-34kg, courbe verte) mais qui est bien récupéré en fin de lactation (+31kg). Ces quantités, gagnées en fin de lactation ou perdues en début de lactation, représentent environ le tiers de ce qui est gagné par le niveau (+102kg, courbe noire) sur les mêmes périodes. Au moment du pic, la différence de production entre les deux courbes extrêmes (4 écart types génétiques de persistance de différence, courbe bleue et verte) pour la persistance est de l'ordre de 0.7 kg par jour soit 20% de la production moyenne au pic sachant que la production moyenne au pic est de 3.6 kg. Cette différence de production est donc importante. Pour la MP, la courbe extrême pour la persistance (courbe verte) montre une courbe relativement très plate et une production quasiment constante. Ces courbes extrêmes pour la persistance montrent des productions en fin de lactation très importantes qui semblent peut-être trop importantes pour un tarissement réussi dans un schéma classique avec une mise bas par an. Pour le TP et le ratio TB/TP, des différences importantes de taux au moment du pic de lactation sont observées. Pour le TP, il semble souhaitable de contre sélectionner pour la persistance pour obtenir un TP élevé en début de lactation. Un ratio TB/TP peu élevé en début de lactation est souhaitable avec une stabilisation rapide. En bovin lait, un ratio élevé en début de lactation est généralement associé à un déficit énergétique car le TB est important. Ce phénomène traduit une augmentation de la lipolyse et de l'absorption d'acides gras des réserves corporelles. En même temps un apport insuffisant en glucides fermentescibles peut provoquer une synthèse protéique insuffisante chez les bactéries du rumen et entraîne donc une baisse du TP.

## **2. Analyses complémentaires des effets fixes du modèle LEG\_FT**

Nous avons étudié les solutions des effets fixes du modèle LEG\_FT pour l'ensemble des caractères chez les 2 races Alpine et Saanen. Seuls les résultats pour les caractères lait, TP, SCS et le ratio TB/TP en race en Saanen sont présentés dans cette partie. Les résultats obtenus pour la MG et la MP sont très proches de ceux obtenus pour le lait, ainsi que ceux sur le TB et le TP. Enfin les résultats en Alpine sont proches de ceux de la Saanen. Nous nous sommes intéressés à l'évolution des différents effets environnementaux suivant : effet du troupeau (HTD) moyen selon le mois de production (Figure 32), effet moyen de la campagne de mise bas (Figure 33), effet de l'âge à la mise bas (Figure 34), effet du mois de mise bas (Figure 35), effet de la durée de tarissement (Figure 36), et effet du stade de gestation (Figure 37). Ces effets environnementaux sont d'intérêt car ils peuvent être valorisés dans le cadre de conseils à l'élevage.



**Figure 30: Profils des première et deuxième fonctions propres (+1 ou -1 écart type génétique du premier ou du deuxième coefficient de régression du modèle EGV\_FT) selon le rang de lactation et la race (Alpine et Saanen) : a) lait, b) MG, c) MP, d) TB, e) TP, f) SCS, g) ratio TB/TP**



**Figure 31: Courbes de lactation phénotypiques moyennes d'animaux ayant des valeurs génétiques extrêmes ( $\pm 2$  écarts types génétiques pour le niveau et la persistance) en deuxième lactation avec le modèle EGV\_FT en race Saanen pour le lait (a), la MP (b), le TP (c), et le ratio TB/TP (d).**

*Niv\_L2* : index niveau exprimé en écarts types génétiques (*LEV\_L2* en anglais), *Pers\_L2* : index persistance exprimé en écart type génétique, *lactation* : somme des index quotidiens sur l'ensemble de la lactation, *7-139j* : somme des index entre les stades de lactation 7 et 139, *140-270j* : somme des index entre les stades 140 et 270. Les valeurs de production à la lactation et entre les jours 7 et 139 ainsi qu'entre les jours 140 et 270 sont précisées dans l'encart « Moyenne phénotypique ».

Les effets HTD moyen par mois de production atteignent un maximum au mois de Juillet et un minimum au mois d'Octobre pour la quantité de lait. Ce résultat est cohérent avec les résultats obtenus au premier chapitre (classification des courbes de lactation). L'évolution des effets HTD pour les taux est contraire à celle de la quantité de lait avec un maximum en janvier/décembre et un minimum en Juillet. Ces résultats sont conformes à ceux de la littérature (Chemineau et al. (2007)). L'effet HTD pour les SCS atteint un maximum au mois d'Octobre et un minimum au mois de Février. Enfin pour le ratio TB/TP, l'effet HTD évolue de la même façon que celle du TP. Par rapport aux autres effets environnementaux du modèle, on constate que les différentiels de production sont importants d'un mois à l'autre en particulier pour le lait et les taux.

Les effets de la campagne de mise bas ont été calculés en additionnant les effets d'une même campagne : moyenne du mois de mise bas par campagne, moyenne de l'âge à la mise bas par campagne, moyenne de la durée de tarissement par campagne et la moyenne des effets HTD par campagne. L'effet de la campagne permet d'avoir un aperçu de l'évolution des performances au cours des ans (1995 à 2015) en s'affranchissant de l'évolution des index génétiques des animaux au cours des ans. Les effets campagne de mise bas pour le lait oscillent d'une année à l'autre sans tendance nette. Sur les 20 années de l'étude il n'y a pas d'augmentation ou de diminution des performances (Figure 33). On peut noter toutefois une cassure entre les années 2011 et 2013 avec une chute de production de 0.3 kg. Cette cassure coïncide avec la crise de l'élevage caprin. Les éleveurs, subissant un prix du lait peu rémunérateur, ont probablement diminué les quantités de concentré distribuées. Pour les taux et les SCS, on observe une augmentation des effets campagne de mise bas constante au cours des ans. Pour le ratio TB/TP, on observe une augmentation de 1995 à 2009, puis une légère diminution. Le différentiel de production entre années extrêmes (2000 et 2011) est inférieur au différentiel de production entre mois extrêmes (juillet et octobre) pour le lait et le TP traduisant peu de différences au fil des campagnes de mise bas.

Sur la Figure 34, seuls les résultats en 1<sup>ière</sup> lactation sont représentés car l'effet de l'âge à la mise bas est plus important en 1<sup>ière</sup> lactation lorsque les chèvres n'ont pas atteint leur poids adulte. Pour la production laitière, nous pouvons voir que plus la chèvre est âgée à la mise bas, plus son pic de lactation est élevé et plus elle produit. On peut observer une interaction entre l'âge à la mise bas et le stade de lactation car l'écart de production est plus important au pic de lactation qu'en fin de lactation entre les plus jeunes et les plus âgées à la mise bas. Pour les taux, peu de différences sont visibles comparées aux autres effets. Pour les SCS, les chèvres les plus âgées produisent plus de SCS que les plus jeunes à la mise bas et le différentiel est important comparé aux autres effets. Pour le ratio, les chèvres âgées de moins de 12 mois ainsi que celles de plus de 14 mois ont des ratios TB/TP inférieurs à celles qui étaient âgées de 12 à 14 mois.

Pour l'analyse de l'effet du mois de mise bas, seuls les résultats en première lactation sont présentés en Figure 35. Les résultats obtenus pour les autres lactations sont similaires. Les chèvres qui mettent bas en début d'année ont des pics de lactation qui durent plus longtemps (Figure 35). Les pics

de lactations sont plus hauts en novembre et en octobre. La majorité des différences sont situées autour de 150 jours de lactation ce qui est en accord avec les résultats du chapitre où la production à cette période était représentée par la troisième composante. En effet, lorsque les jours s'allongent la production est plus importante. Pour les taux, les différences entre les effets des mois de mise bas sont importantes comparées aux autres effets. Les différences sont en accord avec les changements de HTD observés au cours des mois de lactation. Enfin, les comptages cellulaires (SCS) sont plus importants en fin de lactation pour les chèvres qui mettent bas en début d'année. Les chèvres qui mettent bas de Juin à novembre ont des ratios TB/TP supérieurs aux ratios des chèvres mettant bas entre décembre et mai.

Les effets de la durée de tarissement (Figure 36) en 3<sup>ème</sup> lactation sont similaires à ceux de la 2<sup>ème</sup> lactation. Les chèvres tariées sur une courte période, ont une production très différente des autres classes d'effets. Au pic, elles ont une production de lait moins élevée que pour les autres durées de tarissement (Figure 36). Elles produisent plus de TP que les autres chèvres, ont un ratio TB/TP plus faible que les autres et ont plus de SCS dans les 20 premiers jours. La différence de ratio est importante ce qui est cohérent puisque le tarissement est important pour la phase de mobilisation corporelle de début de lactation.

L'effet de la gestation est négatif sur la quantité de lait et est positif sur les taux, les SCS et le ratio TB/TP (Figure 37). Lorsque nous comparons les effets des différents facteurs environnementaux, nous pouvons voir que l'effet de la gestation est le plus important pour les stades les plus avancés pour les taux et la quantité de lait.

Nous voyons que les effets entre caractères respectent les corrélations phénotypiques entre ces mêmes caractères, notamment pour la corrélation la plus marquée entre les caractères étudiés, l'opposition entre lait et TP.

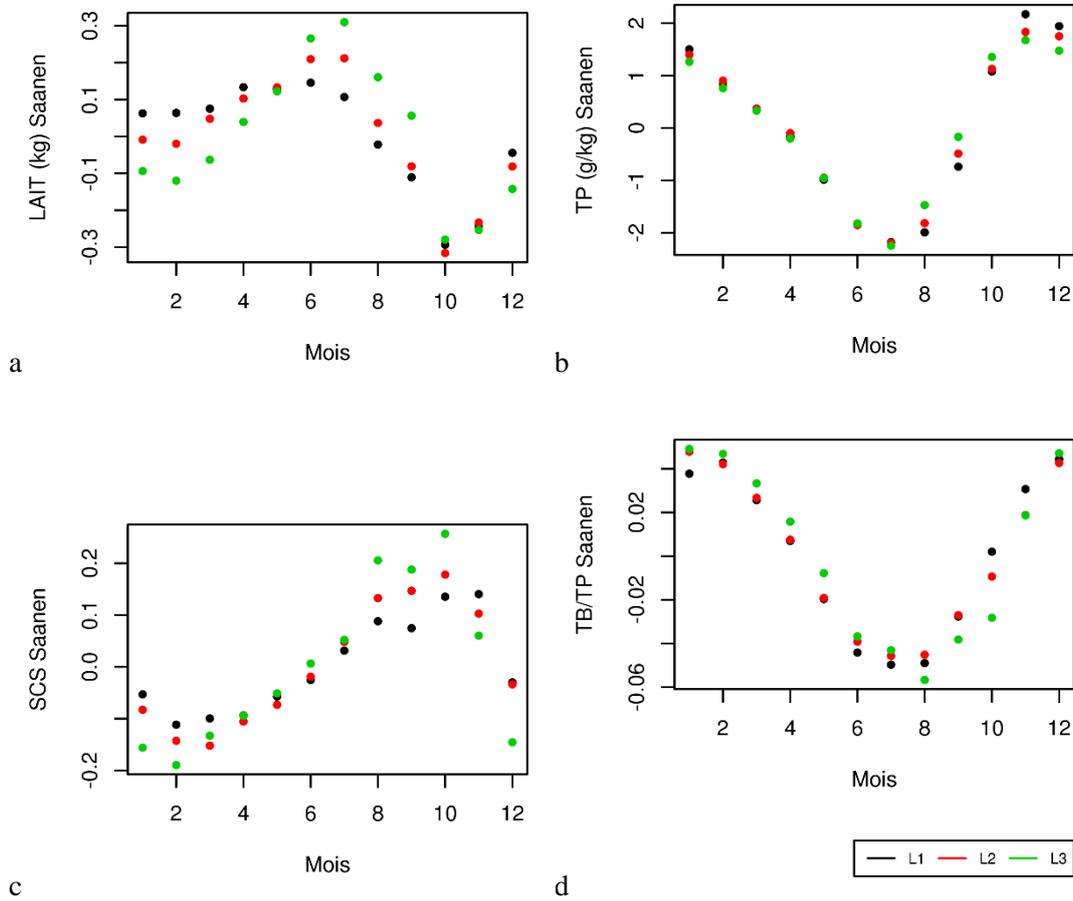


Figure 32: Evolution de la moyenne des HTD par mois de lactation selon le caractère étudié ((a) Lait, (b) TP, (c) SCS et (d) ratio TB/TP) et le rang de lactation en race Saanen

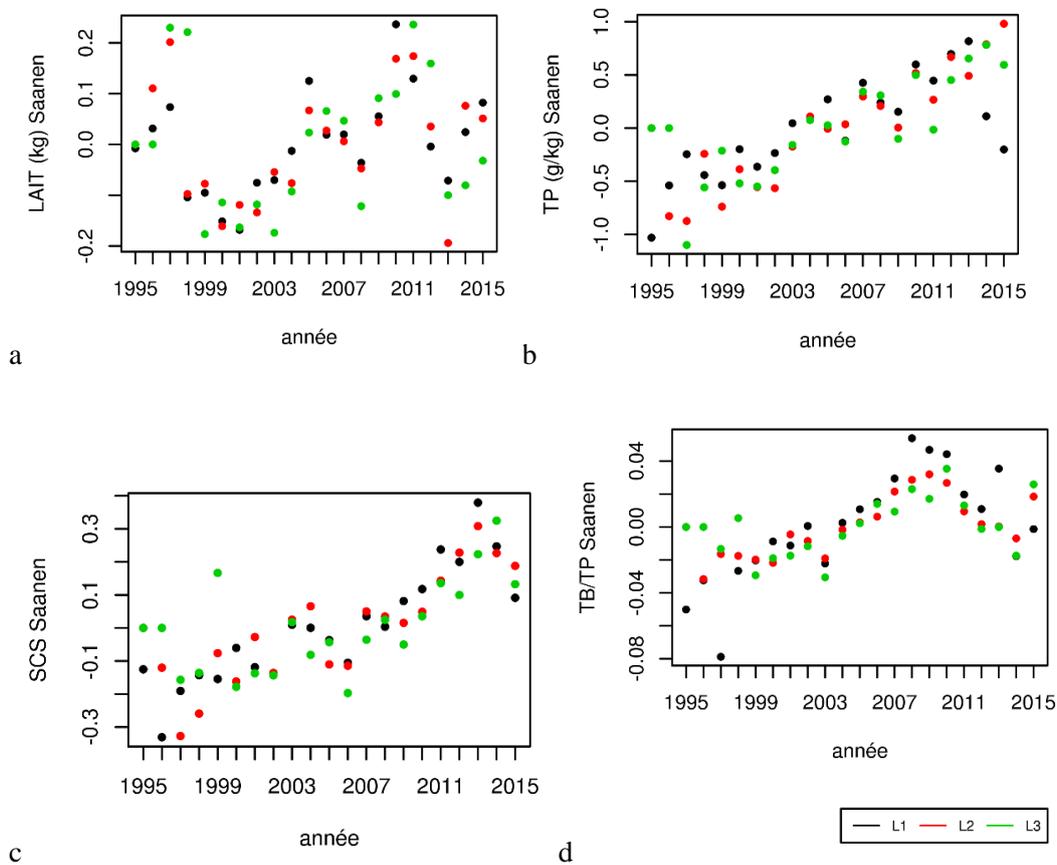
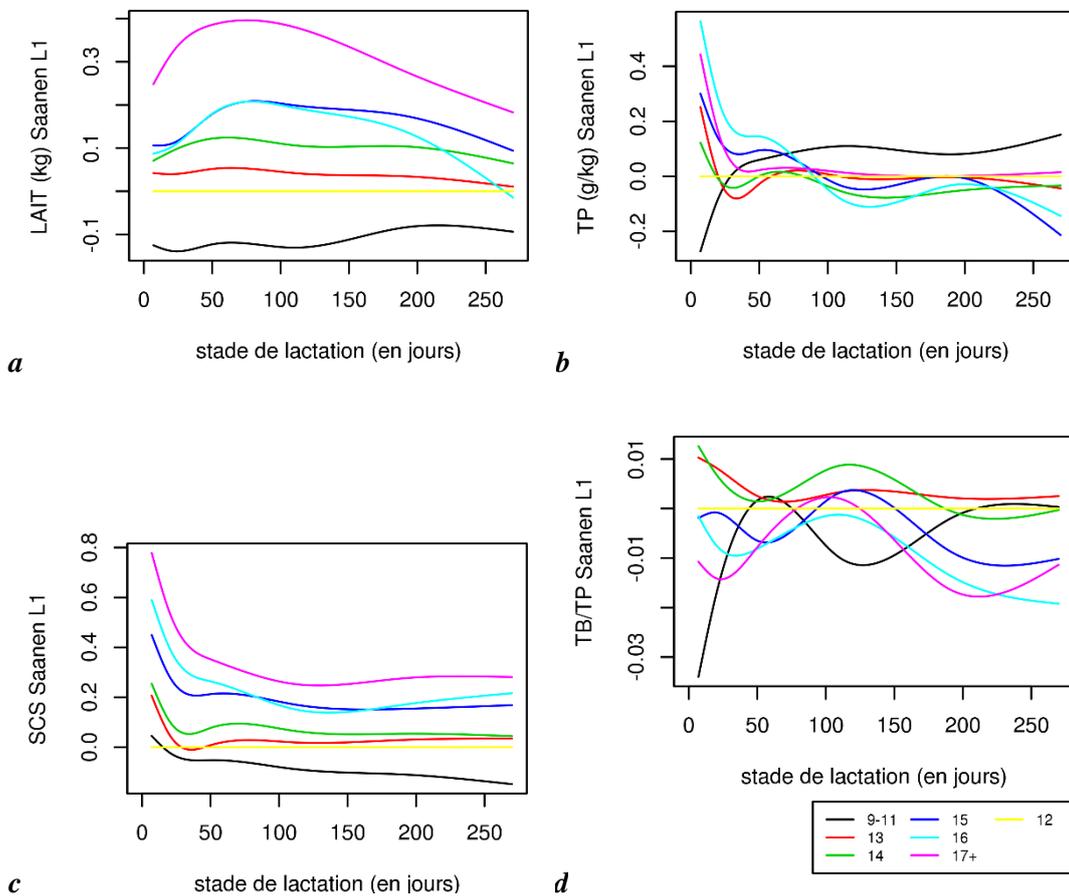
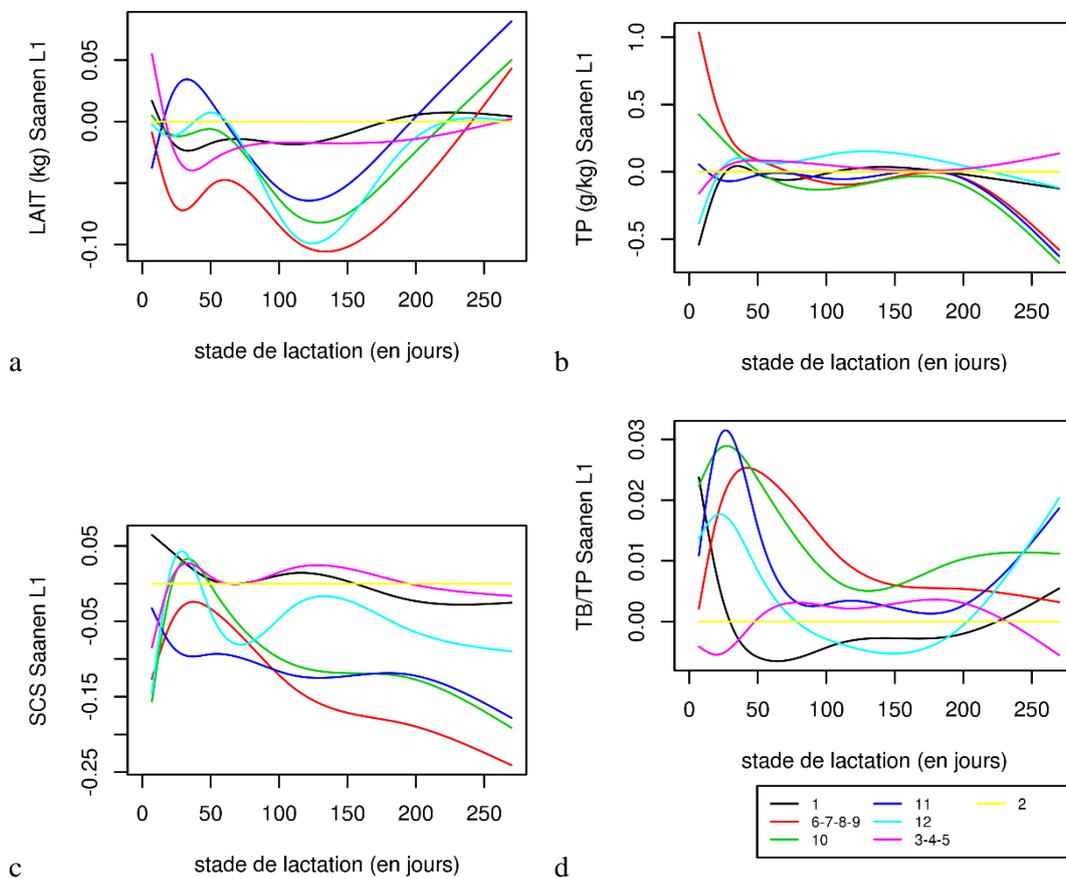


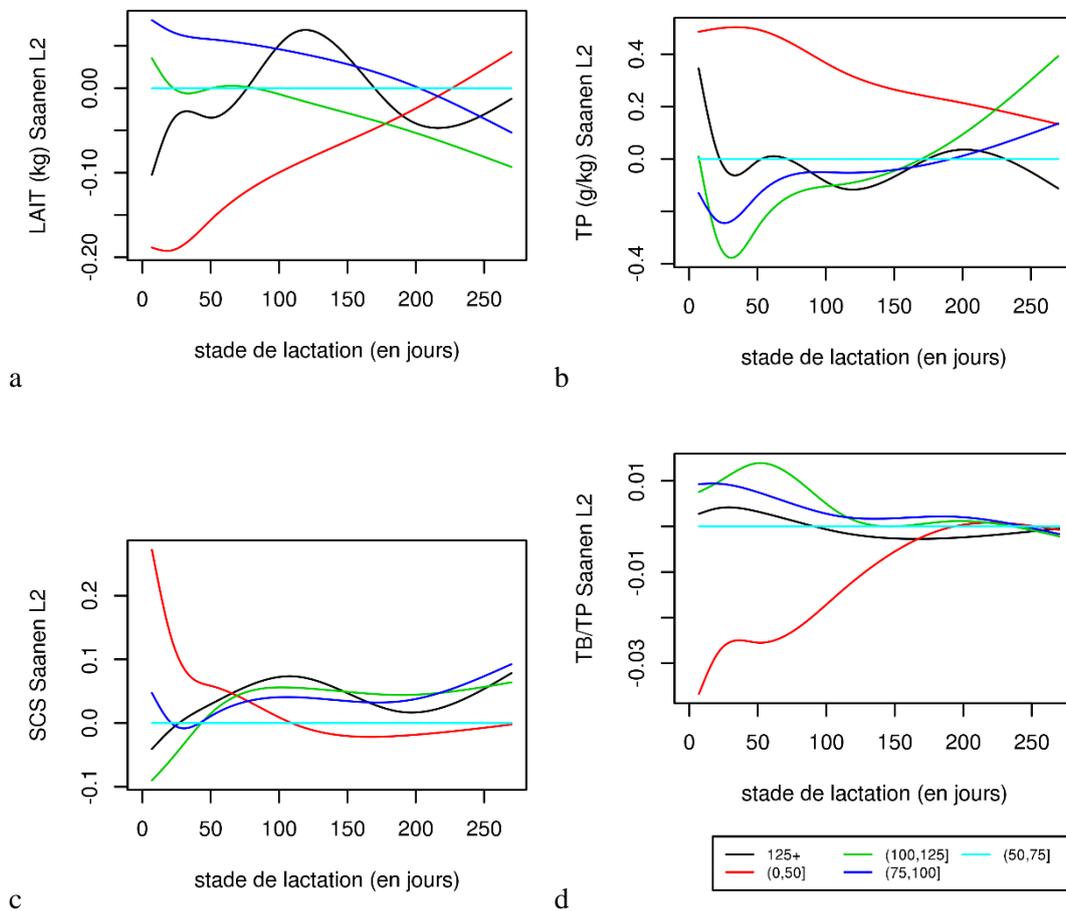
Figure 33: Evolution des effets campagne de mise bas selon le caractère étudié ((a) Lait, (b) TP, (c) SCS et (d) ratio TB/TP) et le rang de lactation en race Saanen



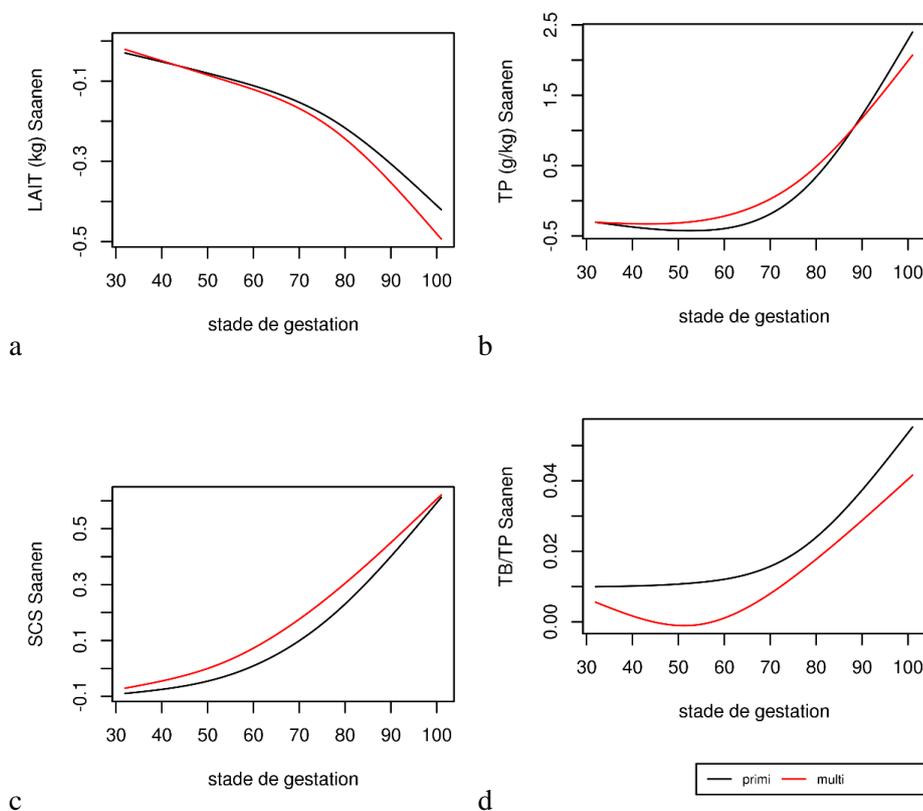
**Figure 34: Evolution des effets de l'âge à la mise bas pour les différents caractères étudiés ((a) Lait, (b) TP, (c) SCS et (d) ratio TB/TP) en L1 en race Saanen**



**Figure 35: Evolution des effets du mois de mise bas pour les différents caractères étudiés ((a) Lait, (b) TP, (c) SCS et (d) ratio TB/TP) en L1 en race Saanen**



**Figure 36: Evolution des effets de la durée de tarissement pour les différents caractères étudiés ((a) Lait, (b) TP, (c) SCS et (d) ratio TB/TP) en L2 en race Saanen**



**Figure 37: Courbes de production en fonction du stade de gestation pour les différents caractères en primipares et en multipares en Saanen : a) lait, b) TP, c) SCS, d) ratio TB/TP**

## **V. Etude des relations entre caractères de durabilité (fertilité à l'IA, survie, longévité et durée de lactation) et les index, niveau et persistance génétique et d'environnement permanent des caractères de production et des comptages de cellules somatiques (SCS), issus du modèle EGV\_PM en race Alpine et Saanen**

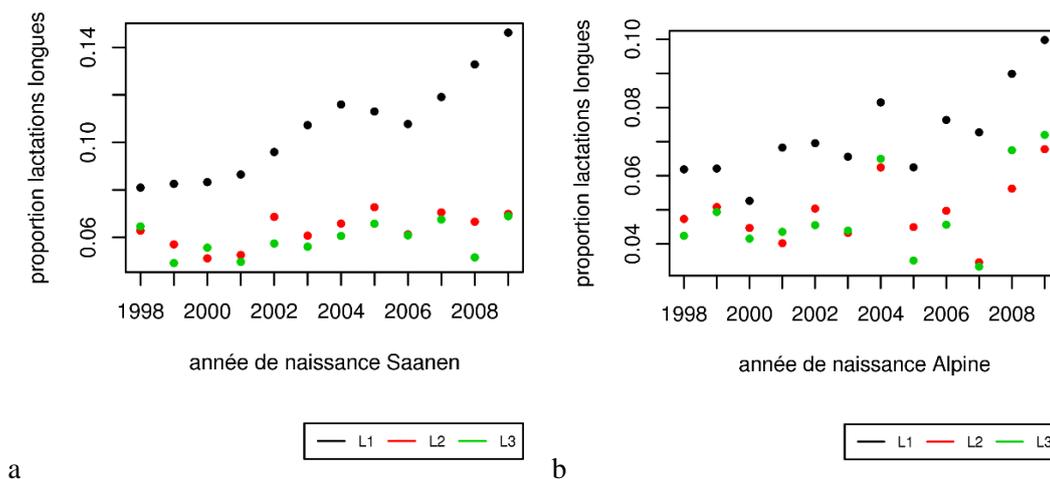
L'objectif de cette partie est d'étudier les liens entre les index niveau et persistance génétique et niveau et persistance d'environnement permanent issus du modèle EGV\_PM, et quelques caractères de durabilité (fertilité à l'IA et survie) d'une part et avec la durée de la lactation d'autre part. Pour cette étude, nous avons utilisé le jeu de données étudié précédemment en Alpine et en Saanen, soit 49849 chèvres de race Alpine et 60071 chèvres de race Saanen suivies de leur première lactation à leur troisième lactation.

Nous constatons un allongement régulier de la durée de la lactation d'années en années dans de nombreux élevages caprins. Une lactation est considérée comme longue si elle dure plus de 485 jours. Dans les élevages à deux périodes de reproduction, la chèvre peut subir deux périodes de reproduction : une à 215 jours de lactation et l'autre à 395 jours. Si la chèvre est toujours productive à 485 jours elle est considérée comme ayant une lactation longue. La répartition de ces deux modalités dans notre jeu de données est présentée dans le Tableau 13. On observe que les chèvres ont plus de lactations longues en première lactation que dans les suivantes (10% en L1 et autour de 5% pour les multipares en Saanen, contre 6% en L1 et 4% en multipares en Alpine respectivement).

La Figure 38 présente l'évolution de la proportion de lactations longues dans les deux races en fonction de l'année de naissance. Le pourcentage de lactations longues en primipares augmente au cours du temps passant de 8% pour les animaux nés en 1998 à 14% pour les animaux nés en 2009 en Saanen. Le pourcentage en multipares est plus stable (autour de 6%). En Alpine, nous constatons une augmentation de la proportion de lactations longues, de 6% à 10% en primipares et une légère augmentation de 4 à 7% en multipares. Les augmentations observées en L1 peuvent être liées à des difficultés de reproduction en fin de lactation.

**Tableau 13 : Répartition des pourcentages des chèvres Alpine et Saanen selon la durée de lactation, la survie et la fertilité à l'IA en fonction du rang de lactation**

			L1	L2	L3
Durée de lactation	Saanen	C (durée lactation <486j)	90%	94%	95%
		L (durée lactation >486j)	10%	6%	5%
	Alpine	C (durée lactation <486j)	94%	96%	96%
		L (durée lactation >486j)	6%	4%	4%
Survie	Saanen	S (sortie)	13%	18%	22%
		C (conservée)	87%	82%	78%
	Alpine	S (sortie)	12%	16%	22%
		C (conservée)	88%	84%	78%
Fertilité à l'IA	Saanen	F (Fertile)	50%	51%	48%
		I (infertile)	24%	22%	21%
		pas info	25%	27%	31%
	Alpine	F (Fertile)	55%	57%	53%
		I (infertile)	21%	20%	20%
		pas information	24%	23%	27%

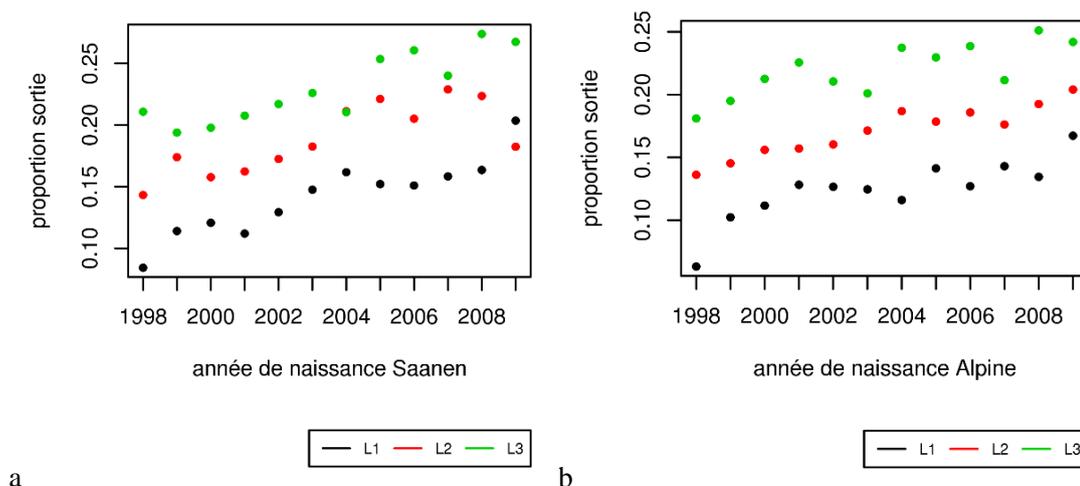


**Figure 38: Evolution de la proportion de lactations longues selon le rang de lactation et l'année de naissance en race Saanen (a), et Alpine (b)**

Les chèvres ont été suivies tout au long de leur carrière et nous avons étudié les moyennes d'index génétiques et d'environnement permanent selon le devenir de la chèvre en fin de lactation (gardée ou sortie de l'élevage). Les chèvres qui ont quitté l'élevage (sorties) représentent 13%, 18% et 22% en L1, L2 et L3 en race Saanen contre respectivement 12%, 16% et 22% en race Alpine (Tableau 13).

La Figure 39 présente l'évolution de la proportion de chèvres sorties par lactation et par race selon l'année de naissance. La proportion de chèvres sorties augmente par année de naissance pour les trois rangs de lactation, avec une augmentation de 5% pour tous les rangs de lactation de 1998 à 2009.

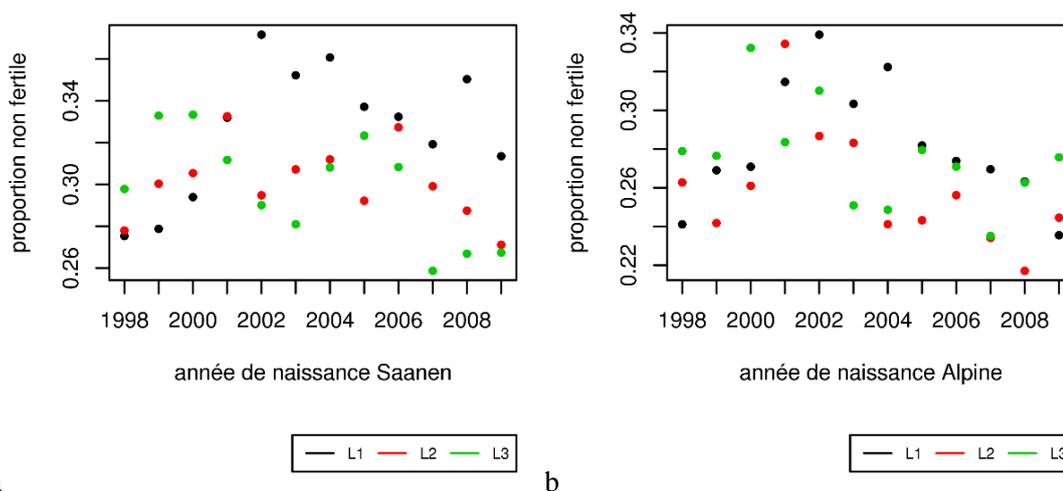
Cette augmentation mérite une attention particulière car la longévité est un caractère d'importance pour les éleveurs en raison du coût d'élevage des jeunes de remplacement.



**Figure 39: Evolution de la proportion d'animaux sortis selon le rang de lactation et l'année de naissance en race Saanen (a), et Alpine (b)**

Concernant la réussite ou non à l'insémination artificielle, nous avons choisi de coder cette variable selon trois modalités: fertile à l'insémination artificielle, infertile à l'insémination artificielle ou « pas d'information ». Cette dernière modalité signifie que nous ne savons pas si la chèvre a été inséminée ou pas, et nous n'avons pas le résultat de l'insémination éventuelle. Le taux de réussite se situe autour de 50% en Saanen et autour de 55% en Alpine, il varie peu d'une lactation à l'autre. Le taux d'infertilité et « pas d'information » sont similaires d'une race à l'autre et d'une lactation à l'autre, autour de 25% (Tableau 13).

La Figure 40 présente l'évolution de la proportion de non réussite à l'IA par rapport à la réussite à l'IA selon le rang de lactation et pour les 2 races. Cette proportion est variable d'une année à l'autre et ce quel que soit le rang de lactation mais elle oscille autour d'une valeur moyenne de 30% pour les deux races. En revanche le nombre de femelles sans information augmente.



**Figure 40: Evolution de la proportion de chèvres non fertiles selon le rang de lactation et l'année de naissance en race Saanen (a), et Alpine (b)**

Le Tableau 14 présente les relations entre la durée de lactation (lactation courte ou lactation longue) par rang de lactation et la survie (chèvres sorties ou non de l'élevage) en race Alpine et Saanen. 23% des chèvres sorties ont fait une lactation longue contre 9% des conservées en Saanen et 14% des sorties ont fait des lactations longues contre 6% des conservées en Alpine en primipare. En multipares, nous observons la même tendance mais les différences sont moins marquées 6% des sorties ont fait une lactation longue contre 4% des conservées. Les chèvres sorties de l'élevage ont donc fait généralement des lactations longues plus souvent que les chèvres conservées et en particulier en primipare. Ce résultat est identique pour les 2 races et pour tous les rangs de lactation.

**Tableau 14 : Répartition des pourcentages de chèvres selon la durée lactation (lactation courte ou lactation longue) par rang de lactation (L1, L2 et L3) et la longévité (chèvres sorties ou non de l'élevage) en race Alpine et Saanen**

		L1		L2		L3	
		C (durée lactation <486j)	L (durée lactation >486j)	C (durée lactation <486j)	L (durée lactation >486j)	C (durée lactation <486j)	L (durée lactation >486j)
<b>Saanen</b>	<b>S (sortie)</b>	77%	23%	91%	9%	92%	8%
	<b>C (conservée)</b>	91%	9%	95%	5%	95%	5%
<b>Alpine</b>	<b>S (sortie)</b>	86%	14%	94%	6%	94%	6%
	<b>C (conservée)</b>	94%	6%	96%	4%	96%	4%

Le Tableau 15 présente le pourcentage de chèvres dans les élevages selon la durée de lactation (lactation courte ou lactation longue) ou la survie (chèvres sorties ou non de l'élevage) et la fertilité par rang de lactation en race Alpine et Saanen. La plupart des chèvres sorties n'ont pas d'information pour la réussite à l'IA (autour de 55% des sorties contre 20% des conservées pour les deux races pour tous

les rangs de lactation). Les chèvres sorties qui ont une information sont moins fertiles (40% d'infertiles pour les sorties contre 20% pour les conservées en primipares). Les chèvres fertiles et sorties sont très peu nombreuses (autour de 8%). 60% des chèvres en lactation longue en première lactation sont des chèvres qui ont échoué à l'IA contre 20% pour les chèvres qui n'ont pas fait de lactation longue.

Les chèvres sorties ont moins bien réussi à la reproduction et en plus, il semblerait qu'elles n'aient pas été choisies pour la reproduction par insémination artificielle car le nombre de « pas d'information » est plus important.

**Tableau 15 : Répartition des pourcentages de chèvres selon la survie (chèvres sorties ou non de l'élevage) ou la durée de lactation (lactation courte ou lactation longue) et la fertilité par rang de lactation en race Alpine et Saanen**

		L1			L2			L3		
		F (Fertile)	I (infertile)	pas information	F (Fertile)	I (infertile)	pas information	F (Fertile)	I (infertile)	pas information
Saanen	S (sortie)	5%	40%	55%	6%	31%	62%	8%	27%	65%
	C (conservée)	58%	22%	20%	62%	20%	19%	61%	18%	21%
Alpine	S (sortie)	7%	39%	54%	8%	34%	58%	8%	30%	62%
	C (conservée)	61%	19%	20%	66%	17%	16%	66%	17%	17%
Saanen	C (durée lactation <486j)	55%	20%	25%	52%	21%	26%	49%	20%	31%
	L (durée lactation >486j)	17%	61%	22%	42%	28%	31%	42%	22%	37%
Alpine	C (durée lactation <486j)	57%	18%	24%	57%	20%	23%	54%	20%	26%
	L (durée lactation >486j)	21%	60%	19%	51%	27%	23%	45%	23%	33%

Au vu de ces premiers résultats, nous avons poursuivi notre analyse par l'étude des relations entre les index génétiques et de l'environnement permanent et les variables citées précédemment (la survie, la fertilité et la durée de lactation). A l'aide d'un test de Student, nous avons étudié si les moyennes des index étaient différentes selon les modalités des variables étudiées : la survie, la fertilité ou non à l'IA et la durée de la lactation.

Dans la suite de ce chapitre, nous ne présenterons que les résultats obtenus en Saanen car les résultats en Alpine étaient très similaires.

### **1. Etude des index génétiques et d'environnement permanent des caractères de production (Lait, MG, MP, TB, TP, ratio TB/TP) et des comptages de cellules somatiques (SCS) selon les modalités de la durée de lactation en race Saanen**

La Figure 41 présente les valeurs de  $-\log_{10}(p\text{-value})$  des tests de Student de comparaison de moyenne entre les modalités de la durée de lactation des index génétiques et de l'environnement

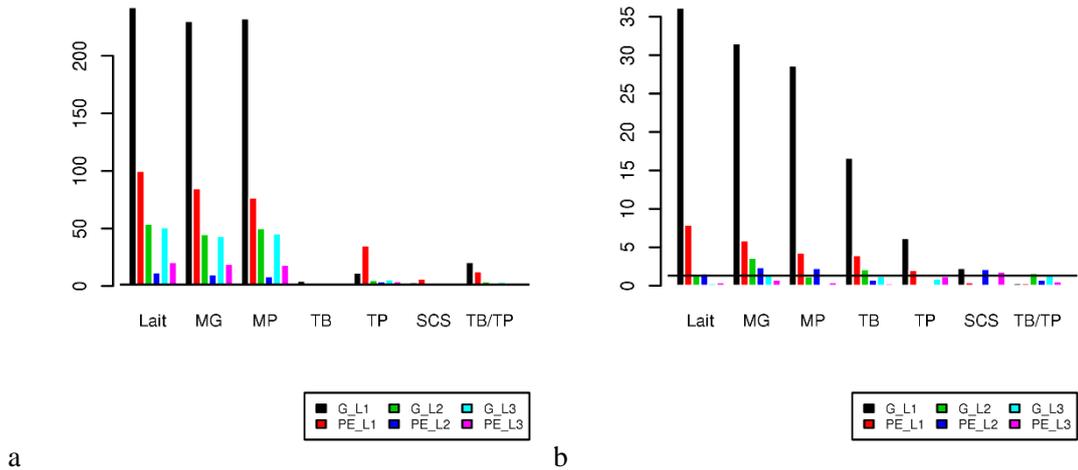
permanent selon le rang de lactation, pour le niveau (a) et la persistance (b), et pour l'ensemble des caractères étudiés en race Saanen.

Pour le coefficient de régression représentant le niveau, les index les plus liés à la durée de la lactation pour les trois rangs de lactation sont les index génétiques pour le lait, la MG et la MP puis on trouve les index de l'environnement permanent pour les mêmes caractères. Les index des autres caractères sont globalement moins fortement liés à la durée de lactation en primipares ; comme par exemple l'index de l'environnement permanent pour le TP, l'index génétique pour le TP et l'index génétique pour le ratio TB/TP. Pour les multipares, l'index génétique des matières (Lait, MG, MP) est le plus lié à la durée de lactation devant les index de l'environnement permanent pour les matières également. Les autres caractères sont peu liés en multipares avec la durée de tarissement.

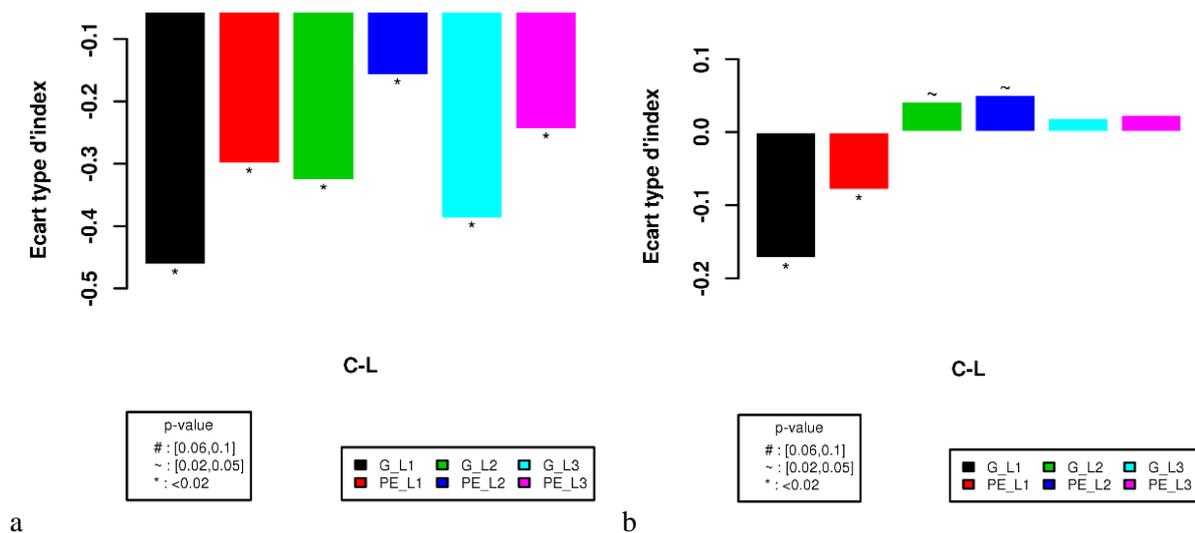
Pour le coefficient de régression représentant la persistance, les index les plus liés à la durée de la lactation sont les index génétiques pour les quantités (Lait, MG, MP) puis les index de l'environnement permanent pour les mêmes caractères en primipares. En multipares, les index génétiques de la persistance laitière sont peu liés à la durée de la lactation. Les valeurs de  $-\log_{10}(p\text{-values})$  sont beaucoup plus faibles pour les persistances que pour les niveaux, quel que soit le rang de lactation. Le niveau est donc majoritairement plus lié à la durée de lactation que la persistance.

Les chèvres ayant une lactation longue ont des index génétiques et de l'environnement permanent plus élevés pour les trois rangs de lactation (+0.45 écart type génétique pour les L1 à +0.3 pour les L2 ; et pour l'environnement permanent : +0.3 en L1 à +0.15 en L2) ( Figure 42). Ces chèvres à lactation plus longues, produisent moins de TP, et ont un ratio TB/TP plus élevé. Ce résultat est en accord avec l'étude (Idele, 2013a), montrant que les éleveurs pratiquent des lactations longues afin de conserver les meilleures chèvres ayant des difficultés à se reproduire. En mettant les chèvres en lactation longue pour les remettre à la reproduction plus tard, ils souhaitent ainsi amortir la génétique investie dans leur troupeau. Ces résultats vont dans le sens de l'étude qui révèle que les éleveurs gardent les chèvres de très haut potentiel en lactations longues (Idele, 2013b).

Les chèvres à lactation longue ont aussi une meilleure persistance génétique et de l'environnement permanent en première lactation en race Saanen (0.15 écart type génétique et 0.07 écart type de l'environnement permanent).



**Figure 41 :**  $-\log_{10}(p\text{-value})$  des tests de Student de comparaison de moyenne des index génétiques (G) et des index de l'environnement permanent (PE) entre les modalités de la durée de lactation, selon le rangs de lactation (L1, L2, L3) pour le niveau (a), et la persistance (b) pour les différents caractères (Lait, MG, MP, TB, TP, SCS, ratio TB/TP) en race Saanen



**Figure 42:** Moyenne de la différence entre les index, génétique ou d'environnement permanent pour le niveau (a) et la persistance (b), des animaux ayant une durée de lactation normale et ceux qui ont une lactation longue, pour les différents rang de lactation pour le lait en race Saanen

## 2. Etude des index génétiques et d'environnement permanent des caractères de production (Lait, MG, MP, TB, TP, ratio TB/TP) et des comptages de cellules somatiques (SCS) selon les modalités de la survie en race Saanen

Nous avons calculé des index moyens à partir des index quotidiens par stade de lactation. Les stades choisis pour l'étude étaient : 1, 30, 60, 90, 120, 150, 180, 210, 240, 270. Les moyennes de ces index, pour le niveau et la persistance, par stade de lactation, ont été testées pour chacune des variables d'étude (survie, fertilité, durée de lactation). Il s'est avéré que les index quotidiens sont nettement plus liés que le niveau ou la persistance uniquement pour la survie. Nous n'avons pas présenté ces résultats

pour la durée de la lactation et nous ne les présenterons pas pour la fertilité à l'IA. Le stade de lactation qui était le plus lié à la survie est l'index pour le stade de lactation 240.

Comme dans le paragraphe précédent, nous avons étudié les moyennes des index génétique et d'environnement permanent, pour le niveau, la persistance et le stade de lactation 240 jours, et pour chaque rang de lactation selon les modalités de la survie (chèvres sorties ou conservées dans l'élevage) à l'aide d'un test de Student. Les valeurs des  $-\log_{10}(\text{p-values})$  de chaque test sont reportées sur la Figure 43.

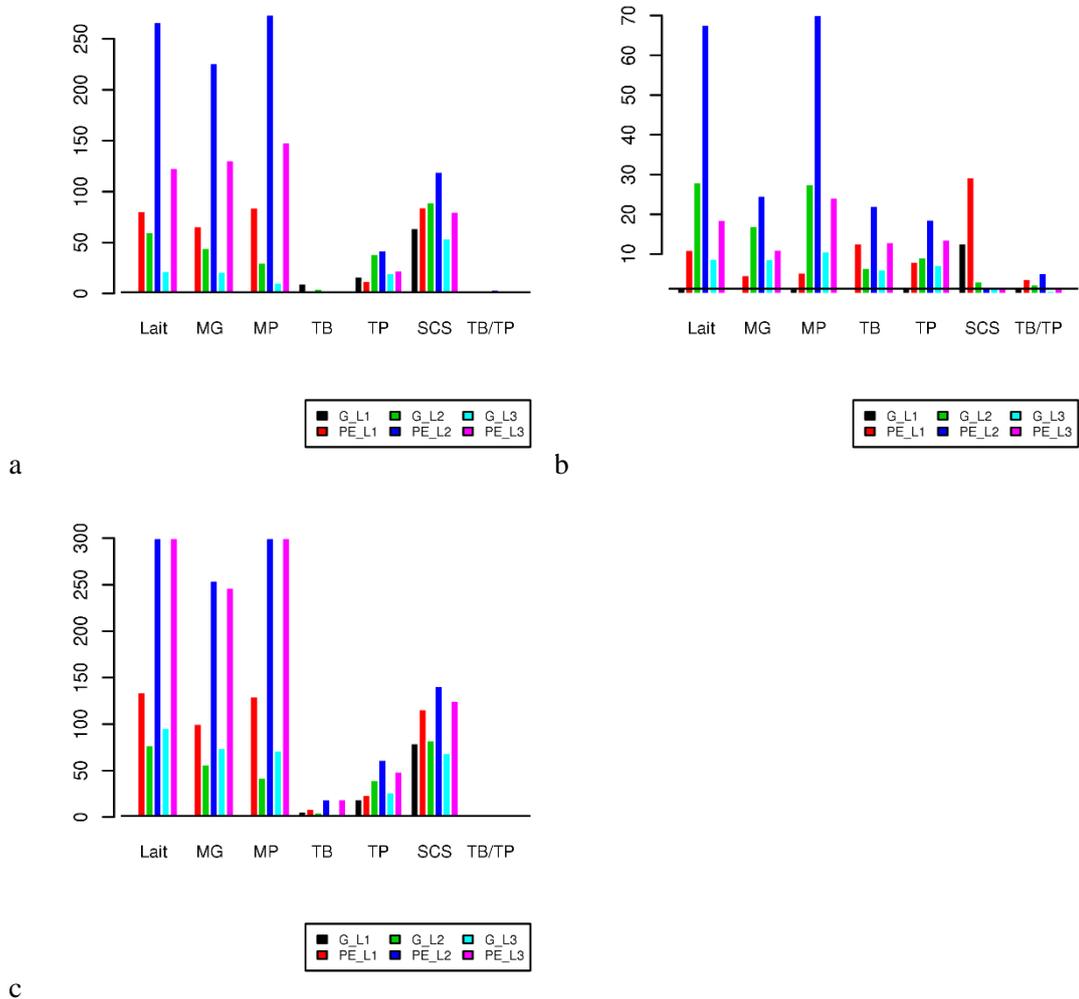
Pour le niveau, les valeurs de  $-\log_{10}(\text{p-values})$  les plus grandes sont celles des index de l'environnement permanent pour les quantités (Lait, MG, MP) pour les trois rangs de lactation. Les index génétiques sont les plus significatives pour les SCS et pour tous les rangs de lactation. Pour les quantités, les index génétiques sont différents selon la survie, uniquement en multipares. Les index génétiques et de l'environnement permanent pour le TP sont également différents selon la survie mais dans une moindre mesure, comparés aux index des quantités de lait et des SCS.

Pour la persistance (Figure 43.b), les liaisons les plus importantes entre survie et index ont été observées en deuxième lactation pour les index de l'environnement permanent pour les quantités (Lait, MG, MP). En primipares, les liaisons étaient surtout importantes pour l'index persistance de l'environnement permanent des SCS et de manière un peu moins marqué pour l'index génétique persistance des SCS. En multipares, les index de l'environnement permanent des quantités étaient les plus liés à la survie, on trouve ensuite les index génétiques et enfin ceux de l'environnement permanent du TB et TP.

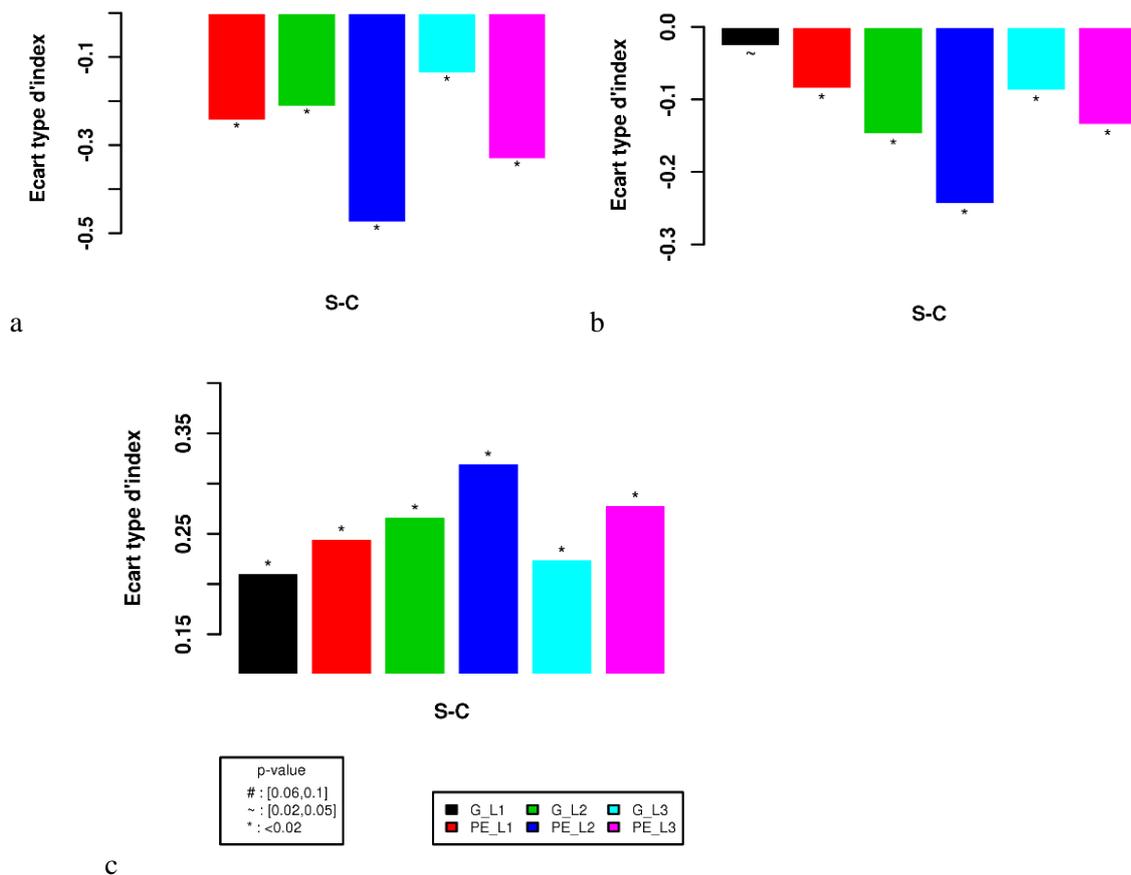
Les valeurs de  $-\log_{10}(\text{p-values})$  des moyennes des index selon les modalités de la variable survie pour le stade 240 jours sont plus importants que celles observées pour le niveau et la persistance pour tous les caractères étudiés (Figure 43.c). Ce résultat s'explique par la forte liaison entre la persistance et survie ainsi qu'entre le niveau et la survie donc en sommant niveau et persistance, nous obtenons l'index 240, qui est plus lié à la survie que les deux index séparés. Pour la fertilité et la durée de lactation, le niveau étant beaucoup plus lié que la persistance à la fertilité et à la durée de lactation que la somme du niveau et de la persistance, i.e. avec l'index 240, apporte peu de différence par rapport au niveau seul.

Les chèvres conservées en fin de lactation dans l'élevage ont des valeurs génétiques et d'environnement permanent pour les SCS significativement inférieures aux chèvres sorties (Figure 44). Les chèvres sorties de l'élevage ont des valeurs génétiques et d'environnement permanent pour le lait plus faibles que les chèvres conservées au sein de l'élevage à l'exception des primipares pour lesquelles les index génétiques sont similaires. Les chèvres sorties de l'élevage ont des index génétiques et de l'environnement permanent pour le TP significativement supérieurs aux chèvres conservées.

Les chèvres Saanen non conservées ont une persistance pour le TP, le TB et le ratio TB/TP plus importante que les chèvres conservées, elles ont donc moins de TB, de TP et de ratio TB/TP en début de lactation comparé aux chèvres moyennes avec des niveaux similaires. Par contre les chèvres sorties ont des persistances laitières moins élevées que les chèvres conservées.



**Figure 43:**  $-\log_{10}(p\text{-value})$  des tests de Student de comparaison de moyennes des index génétiques (G) et des index de l'environnement permanent (PE) entre les modalités de la survie, selon le rang de lactation (L1, L2, L3) pour le niveau (a), et la persistance (b) et le stade de lactation 240 pour les différents caractères (Lait, MG, MP, TB, TP, SCS, ratio TB/TP) en race Saanen



**Figure 44: Moyenne de la différence d'index génétique ou de l'environnement permanent entre les animaux ayant été sortis et ceux ayant été conservés pour les différents caractères et rangs de lactation en Saanen : niveau du lait (a), persistance du lait (b), niveau des SCS (c)**

Pour la survie, des différences marquées de liaisons entre les primipares et les multipares sont observées. En effet, le niveau de l'environnement permanent et de la génétique sont très liés à la survie en multipares alors que le niveau génétique pour la quantité de lait en primipares n'est pas lié. Une hypothèse est que l'index génétique en primipares n'est pas lié de manière linéaire à la survie. Nous avons montré que les index les plus liés à la survie étaient les index des niveaux de production de l'environnement permanent. Les chèvres conservées avaient des index environnement permanent pour la production plus importants. Il est intéressant de voir qu'ils sont plus liés que les index génétiques pour la survie. Ces résultats sont complémentaires de l'étude « améliorer la survie des troupeaux caprins » (Inosys, 2014) qui montre des différences en matière d'effet troupeau, et d'effet de durée de tarissement. Nous avons mis en évidence des liaisons importantes avec les SCS, pour des chèvres sorties qui ont plus de SCS d'un point de vue génétique et de l'environnement permanent. Ces comptages cellulaires élevés pour les chèvres sorties de l'élevage peuvent être expliqués par une infection par un pathogène tel que l'arthrite encéphalite caprine (CAEV) ou un autre pathogène (Idele, 2012).

### 3. Etude des index génétiques et d'environnement permanent des caractères de production (Lait, MG, MP, TB, TP, ratio TB/TP) et des comptages de cellules somatiques (SCS) selon les modalités de la réussite à l'insémination artificielle en race Saanen

La relation entre la fertilité à l'IA et les index génétique et d'environnement permanent pour le niveau et pour la persistance et pour chaque rang de lactation a été étudiée à l'aide d'un test de Student. La valeurs de  $-\log_{10}(p\text{-values})$  de chaque test sont reportées sur la Figure 45. Seules les chèvres qui ont réussi ou échoué à la reproduction à l'IA ont été prises en compte, i.e. nous avons exclu les chèvres appartenant à la classe « sans information ».

Pour le coefficient de régression relatif au niveau et pour les primipares, les index les plus liés à la fertilité sont par ordre d'importance les index de l'environnement permanent pour les quantités, les index génétiques des quantités, l'index de l'environnement permanent pour le TP, les index génétiques pour les SCS et le TP et l'index de l'environnement permanent pour le ratio TB/TP. Pour les multipares, les index génétiques des quantités sont les plus liés à la fertilité, viennent ensuite les index de l'environnement permanent pour les quantités.

Pour le coefficient de régression relatif à la persistance, les liaisons sont globalement moins importantes que pour les niveaux. En primipare, les liaisons les plus importantes entre index et fertilité à l'IA sont l'index des cellules (SCS) pour l'environnement permanent, les index génétiques du TB et des SCS, puis les index de l'environnement permanent pour le TP et pour le lait. En multipares les liaisons sont faiblement significatives.

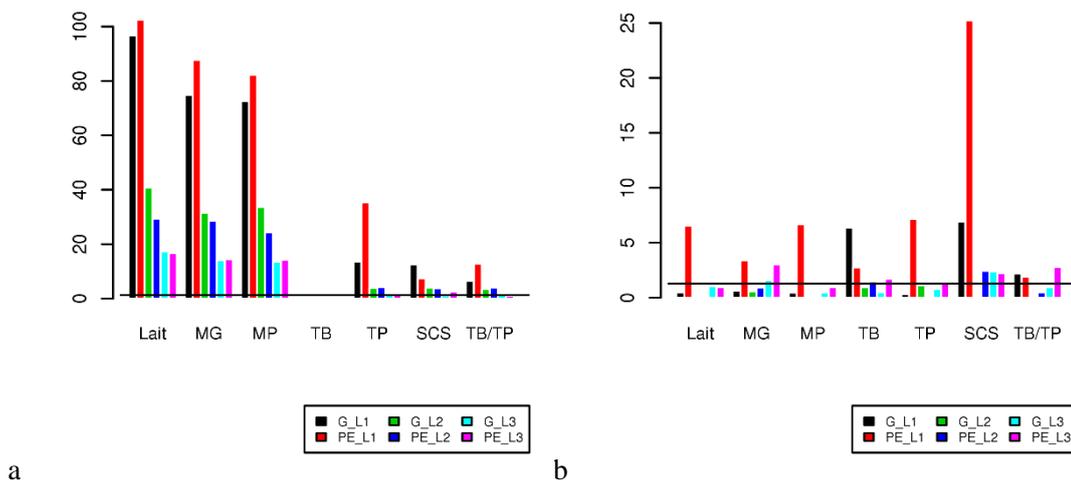
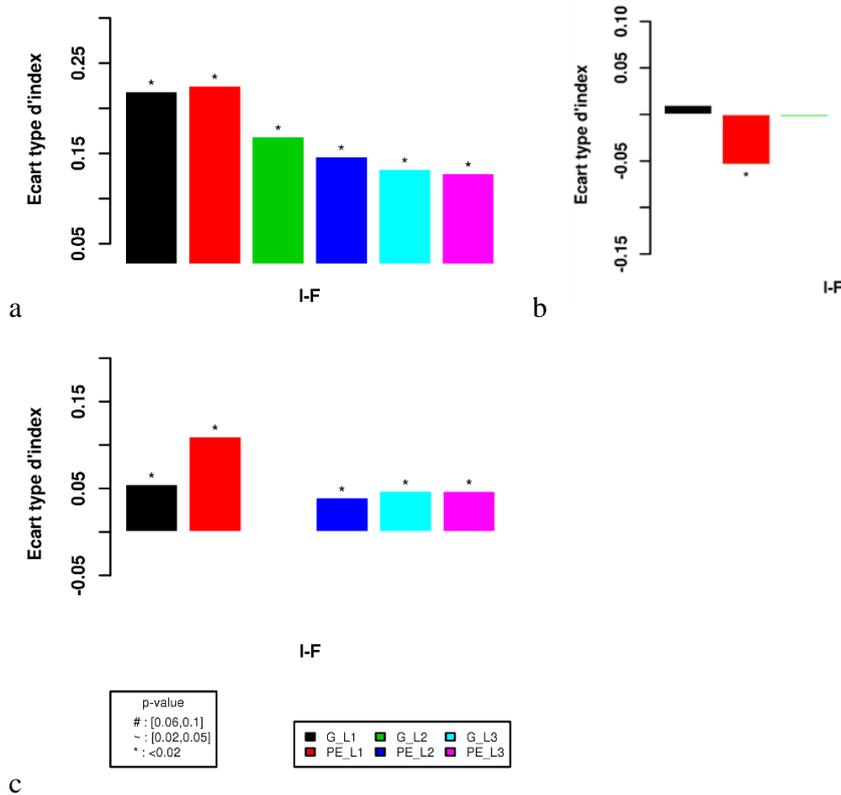


Figure 45 :  $-\log_{10}(p\text{-value})$  des tests de Student de comparaison de moyennes des index génétiques (G) et des index de l'environnement permanent (PE) entre les modalités de la fertilité à l'IA, selon le rang de lactation (L1, L2, L3) pour le niveau (a), et la persistance (b) pour les différents caractères (Lait, MG, MP, TB, TP, SCS, ratio TB/TP) en race Saanen

Les chèvres fertiles ont des index lait plus faibles pour les quantités (-0.20 écart-types génétiques en primipare, -0.15 en multipares), et des index TP et du ratio TB/TP plus élevés (Figure 46). A contrario, les chèvres infertiles ont un index persistance pour les SCS plus élevé, et une persistance pour le TB moins élevée. Enfin, les chèvres infertiles pour l'environnement permanent avaient un index persistance laitière plus faible que les fertiles.



**Figure 46: Moyenne de la différence d'index génétique ou de l'environnement permanent entre les animaux ayant été infertiles et ceux ayant été fertiles pour les différents caractères et rangs de lactation en Saanen : niveau du lait (a), persistance du lait (b), persistance des SCS (c)**

En conclusion, les liaisons entre les caractères de production et la fertilité sont moins fortes que celles entre les caractères de production et la survie ou la durée de lactation pour le niveau comme pour la persistance. Ces résultats sont peut-être liés à l'IA. En effet, toutes les chèvres n'ont pas été mises à la reproduction par insémination artificielle ou certaines ont été mises mais sorties ensuite, entrant dans la catégorie « sans information ». Ce phénomène peut sous-estimer les liaisons observées avec les caractères de production, car dans l'étude pour la durée de lactation ou la survie, nous avons l'ensemble des animaux.

Nous avons étudié la fertilité à l'IA mais il aurait été aussi intéressant d'étudier la fertilité plus généralement avec notamment la réussite à la monte naturelle qui suit en général l'échec à l'insémination artificielle (appelé « IA+retours ») comme dans l'étude de l'« évolution de la fertilité à l'insémination artificielle chez les chèvres de 2004 à 2010 » (Idele, 2013b). Cette étude montre que le taux de fertilité

à l'IA est stable alors que le taux de fertilité IA+retours se dégrade avec une diminution de 5% en 6 ans. Dans cette étude, le niveau de réussite en primipares est inférieur à celui en multipares alors que dans notre étude nous avons montré de plus grandes liaisons entre niveau de production et fertilité en primipares qu'en multipares.

En conclusion, les index génétiques de la quantité étaient les plus liés pour la durée de la lactation, les index génétiques et d'environnement permanent étaient liés avec la même intensité pour la fertilité. En revanche pour la survie, les index de l'environnement permanent étaient les plus liés.

#### **4. Evolution de l'infertilité, de la survie, de la durée de lactation selon l'évolution des index niveau et persistance pour la production laitière**

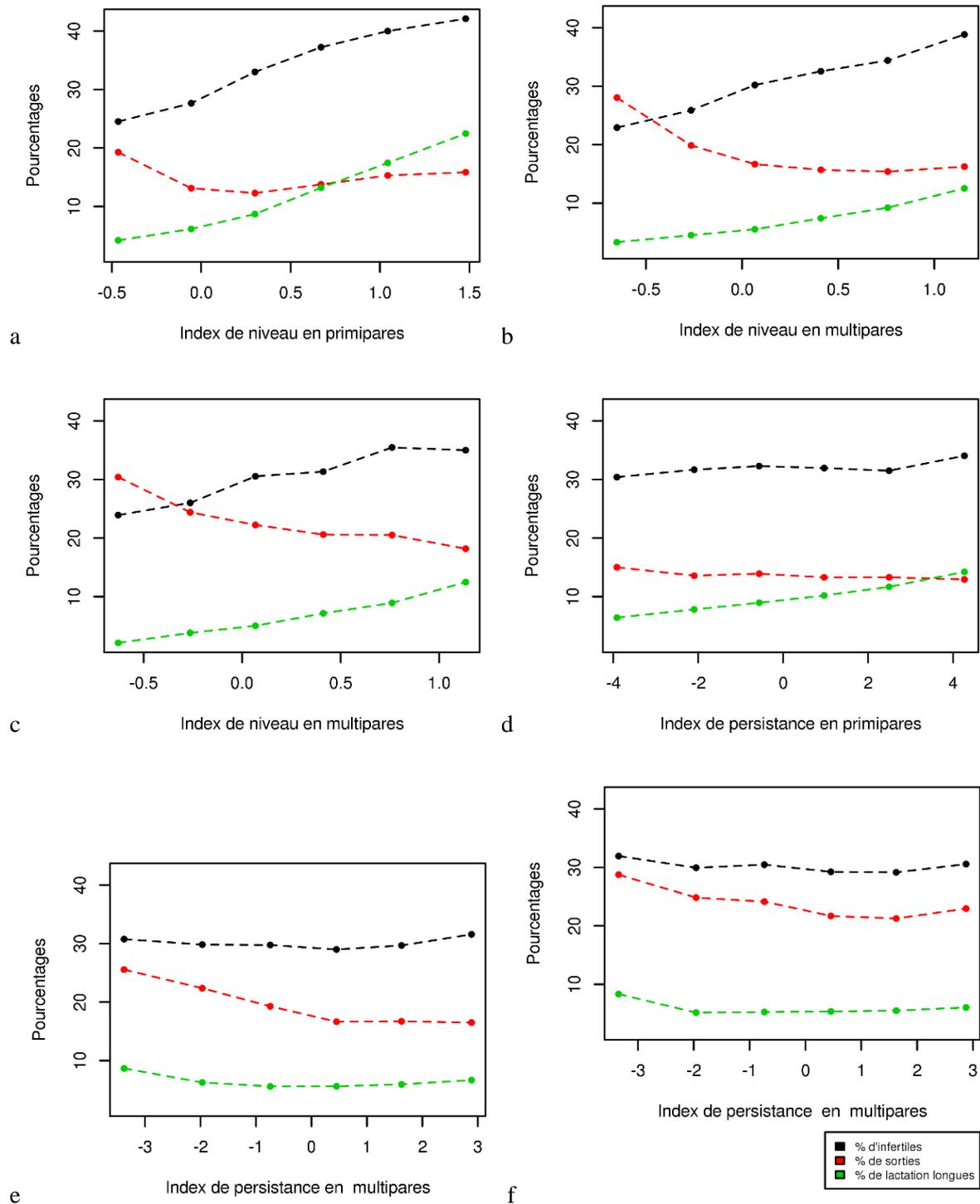
Afin d'avoir une vision d'ensemble des liens entre fertilité, survie et durée de lactation, nous avons calculé les moyennes, par classe d'index génétiques de la quantité de lait pour le niveau et la persistance pour chaque rang de lactation, du pourcentage de chèvres infertiles, du pourcentage de chèvres sorties et du pourcentage de lactations longues et nous les avons présentées à la Figure 47.

6 classes d'index génétiques ont été définies de la façon suivante : nous avons découpé l'intervalle entre l'index minimal et maximal en 10 intervalles égaux et avons regroupé les 3 intervalles de chaque extrémité.

L'évolution des moyennes de pourcentage sont représentées dans la Figure 47. Lorsque les index des coefficients de régression du niveau augmentent en primipare, le pourcentage d'infertilité augmente linéairement (de 25% à 42%), le pourcentage de lactations longues augmente également avec la même pente que l'infertilité (de 4% à 20%), le pourcentage de sortie est passé de 20% à 15% pour la troisième classe puis remonte à 17%. En multipares, lorsque l'index de niveau augmente, l'infertilité augmente également de manière linéaire de 22% à 40% en deuxième lactation et de 24% à 35% en troisième lactation. Le pourcentage de lactation longues augmente également de manière linéaire de 4 à 11%. Le pourcentage de sortie passe de 30% à 20% avec l'augmentation des index. Des antagonismes entre fertilité et survie en fonction des index sont observés. En effet, plus les index niveaux augmentent, plus l'infertilité est grande, mais le taux de sortie diminue (surtout en multipares). En primipares et en Saanen, le taux de sortie n'est pas minimal pour les plus gros index niveaux. Il est intéressant de noter que les courbes d'infertilité et de taux de lactation longues augmentent avec la même pente.

Pour la persistance, lorsque l'index persistance augmente, l'infertilité est plutôt stable, tout comme le pourcentage de sortie en primipare. Le pourcentage de lactation longue lui a tendance à augmenter. En multipares, le pourcentage de lactations longues et d'infertilité est plutôt stable alors que le pourcentage de sortie a tendance à diminuer linéairement avec l'augmentation des index persistance de 25 à 18% en seconde lactation et de 30 à 25% en troisième lactation.

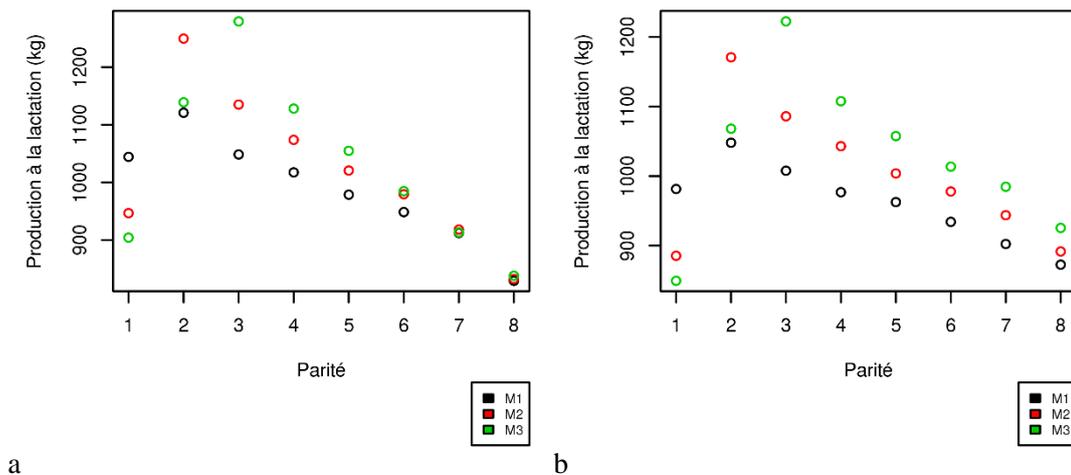
Les chèvres qui ont des lactations avec des hauts niveaux de production ont des troubles de la fertilité mais sont gardées dans le troupeau notamment en allongeant la durée de la lactation. Ces résultats permettent de relier des paramètres d'élevage concrets exprimés en pourcentage et des index niveau et persistance génétique pour le lait. Ces résultats pourraient être utiles aux éleveurs.



**Figure 47 : Evolution des pourcentages d'infertiles, de sorties et de lactation longues en fonction des index génétique de la quantité de lait en Saanen : niveau en primipares pour la première lactation (a), en multipares pour la deuxième lactation (b), en multipares pour la troisième lactation (c), persistance pour la première lactation (d), en multipares pour la deuxième lactation (e), en multipares pour la troisième lactation (f)**

## 5. Etude des relations entre la longévité et les index génétiques de niveau et de persistance pour la production laitière

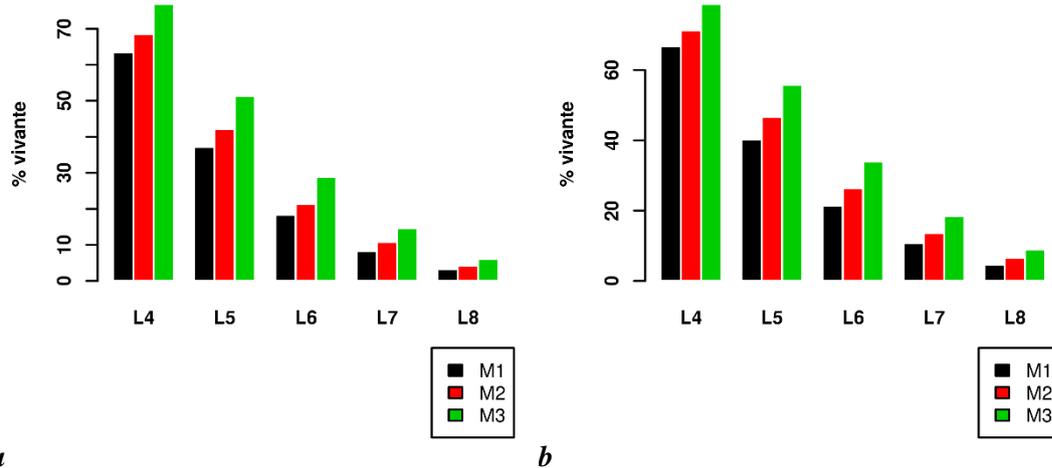
Nous avons étudié la survie au-delà de la troisième lactation et avons mis en perspective cette longévité au regard de la production et des index niveau et persistance par parité. Le jeu de données étudié est identique au précédent en sélectionnant dans ce dernier les chèvres ayant au moins 3 lactations et nées avant 2008 afin d’avoir l’enregistrement de leurs 8 lactations potentielles, ce qui représente 25 828 chèvres de race Saanen et 21 673 chèvres de race Alpine. Nous avons regroupé les 3000 chèvres ayant produit le plus de lait en première lactation dans une classe notée M1. Nous avons fait de même avec les chèvres en deuxième lactation (notée M2) et en troisième lactation (notée M3). Les productions à la lactation calculées par la méthode Fleischmann de ces trois groupes de chèvres sont présentées dans la Figure 48 en fonction de la parité. Les M1 sont les plus productives en L1 mais les moins productives de la L2 à la L8. Les M2 sont les deuxièmes plus productives en L1, les premières en L2 et les deuxièmes de la L3 à la L8. Les M3 sont les moins productives en L1, les deuxièmes en L2 et les premières de la L3 à la L8 en accord avec l’étude de Haiger (2018) en bovin lait. Les différences entre classes s’amenuisent à partir de la lactation 6 en Saanen.



**Figure 48 : Evolution de la production à la lactation calculée par la méthode de Fleischmann pour les 3000 chèvres qui ont produit le plus en L1 (M1), en L2 (M2) et en L3 (M3) en Saanen (a) et en Alpine (b).**

*M1 : chèvres faisant partie de la classe des 3000 meilleures en L1, M2 : chèvres faisant partie de la classe des 3000 meilleures en L2, M3 : chèvres faisant partie de la classe des 3000 meilleures en L3*

L’évolution du pourcentage de chèvres conservées au cours des lactations 4 à 8 (L4 à L8) est présenté dans la Figure 49. Les M1 sont les chèvres ayant la moins bonne longévité, suivies des M2. Les M3 ont la meilleure longévité en accord avec l’étude de Haiger (2018). Lorsque nous regardons la production à la carrière : les M3 ont le plus produit suivies des M2. La longévité est meilleure en Alpine qu’en Saanen (plus de chèvres non sorties de la L4 à la L8 pour toutes les classes).



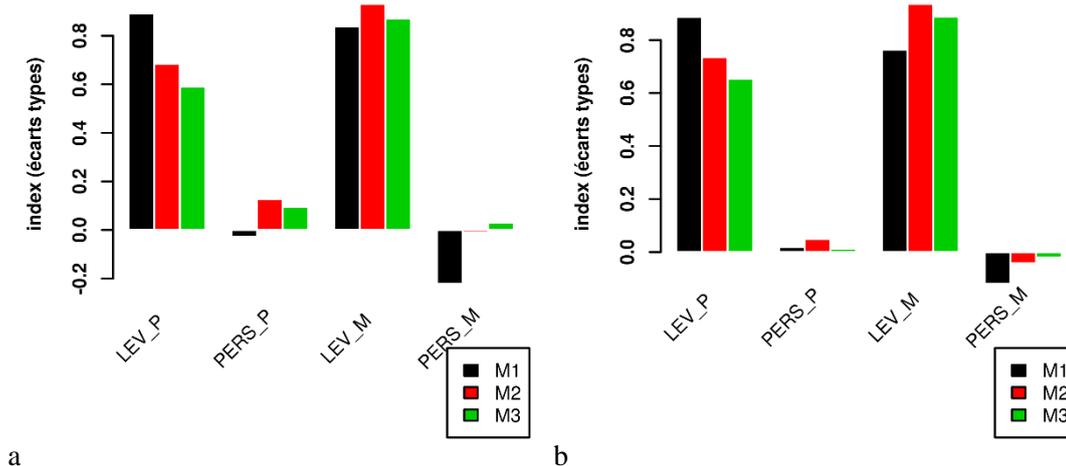
**Figure 49 : Evolution du pourcentage de chèvres toujours présentes dans les parités 4 à 8 (L4, ..., L8) en Saanen (a) et en Alpine (b) pour les 3000 chèvres qui ont produit le plus en L1 (M1), en L2 (M2) et en L3 (M3)**

*M1 : chèvres faisant partie de la classe des 3000 meilleures en L1, M2 : chèvres faisant partie de la classe des 3000 meilleures en L2, M3 : chèvres faisant partie de la classe des 3000 meilleures en L3*

Les index moyens génétiques pour la production laitière du modèle EGV\_PM des chèvres selon leur classe (M1, M2, M3) sont présentés dans la Figure 50. Ces index ont été standardisés pour être comparables. Les chèvres de la classe M1 ont les index en primipares les plus élevés suivies des chèvres du groupe M2 (différence de +0.3 écart types génétiques entre les M1 et les M3 en Saanen et en Alpine). A noter toutefois que les M3 font parties des meilleures chèvres de la population avec des index à +0.6 d'écart type. Pour le niveau en multipares les chèvres du groupe M2 ont les index les plus élevés devant les groupe M3 et M1, bien que les index des trois catégories soient proches.

Pour les persistances, pour les index en primipares, en Saanen, les chèvres des groupes M2 et M3 ont des index plus élevés que celles du groupe M1 (+0.1 écart-types) alors qu'il n'y a pas de différence en Alpine. En multipares, les chèvres M3 ont des index persistance proche de 0 (moyenne de la population) mais plus importants que les chèvres M1 (+0.2 écart types en Saanen, +0.1 écart types en Alpine).

L'ensemble des résultats reste inchangé si on sélectionne 10 000 chèvres pour chaque catégorie (M1, M2, M3).



**Figure 50 : Index moyens des chèvres issus du modèle EGV\_LM (LEV\_P, PERS\_P, LEV\_M, PERS\_M) pour les 3000 chèvres qui ont produit le plus en L1 (M1), en L2 (M2) et en L3 (M3) en Saanen (a) et en Alpine (b)**

M1 : chèvres faisant partie de la classe des 3000 meilleures en L1, M2 : chèvres faisant partie de la classe des 3000 meilleures en L2, M3 : chèvres faisant partie de la classe des 3000 meilleures en L3, LEV\_P: index pour le niveau en primipare, PERS\_P : index pour la persistance en primipare, LEV\_M : index pour le niveau en multipares, PERS\_M : index pour la persistance en multipares,

Ces résultats sont particulièrement intéressants car ils mettent en évidence que les chèvres les plus productives avec les meilleurs index en primipares ne sont pas les chèvres les plus productives dans les lactations suivantes. Les chèvres qui produisent beaucoup en L1 ont une moins bonne longévité que les chèvres qui produisent le plus en L3. Ces résultats montrent tout l'intérêt de séparer les primipares et les multipares dans un modèle d'indexation afin d'obtenir des index primipares et multipares distincts. Pour la création d'un index combiné il semble donc important de privilégier la L3 au détriment de la L1. Ces résultats montrent aussi l'intérêt d'une sélection de la persistance afin d'améliorer la longévité et la production de lait sur la carrière de l'animal.

La corrélation négative entre grande production et fertilité a été approchée précédemment (V.4.) avec une liaison plus importante en primipares qu'en multipares. Nous avons également vu une liaison positive entre production et longévité ou survie à partir de la deuxième lactation. La liaison positive entre persistance et survie a, là aussi, été réaffirmée.

En conclusion, un modèle de régression aléatoire pour la modélisation des contrôles élémentaires des trois premières lactations des chèvres Alpine et Saanen est proposé avec un temps de calcul réduit par rapport au modèle le plus complet. Pour ce faire, nous avons utilisé des fonctions propres pour la modélisation de la partie génétique et de l'environnement permanent et avons combiné dans un même niveau les deuxièmes et troisièmes lactations ensemble. Les index obtenus avec ce modèle RRM

(EGV\_PM) sont cohérents avec ceux du modèle à la lactation. Le modèle EGV\_PM permet d'obtenir de manière directe des valeurs génétiques pour le niveau d'une part et la persistance d'autre part en différenciant les primipares des multipares, ces 2 vecteurs de valeurs génétiques étant non corrélés. Ces index semblent intéressants pour sélectionner des animaux productifs et durables au vu des relations positives entre index niveau et persistance en multipares et la longévité. Les effets de milieux issus du modèle RRM peuvent être des outils intéressants à exploiter d'un point de vue zootechnique pour un conseil auprès des éleveurs. Cependant, le modèle d'évaluation génétique utilisé actuellement en routine intègre les informations génomiques (puce SNP 50K) disponibles. Il est donc nécessaire de développer et d'étudier la faisabilité d'un modèle RRM génomique chez les caprins laitiers.



## **Chapitre 4 : Etude de la faisabilité d'une évaluation génomique basées sur les contrôles élémentaires de chèvres Alpine et Saanen à partir d'un modèle de régression aléatoire**

En caprins laitiers, la disponibilité d'une puce de 50 000 polymorphismes d'un seul nucléotide (single-nucleotide polymorphism : SNP) en 2011 a poussé la filière caprine à s'intéresser à la possible mise en place d'une sélection génomique dans cette espèce pour les caractères évalués en routine dans un premier temps. Suite à de nombreuses études depuis 2012 (Carillier-Jacquin, 2015; Marc Teissier, 2018), la sélection génomique est déployée pour l'espèce caprine en France depuis janvier 2018 à l'aide d'un modèle Single-step GBLUP (ssGBLUP). Elle concerne les 2 principales races laitières (Alpine et Saanen) et est appliquée à l'ensemble des caractères de production actuellement en sélection (Lait, MG, MP, TB, TP, SCS). Le gain de progrès génétique attendu a ainsi été quantifié entre 20% et 30% à coût constant pour le schéma de sélection.

Dans la continuité des travaux précédents, nous avons souhaité exploiter l'information génomique pour étendre les modèles génomiques ssGBLUP, à l'analyse des contrôles élémentaires de la production laitière.

Dans une première partie, nous introduirons les principes de la validation d'une évaluation génomique. Dans une seconde partie rédigée sous forme d'un article scientifique (article 4 : soumission prévue dans la revue *Journal of Dairy Science*), nous présentons la comparaison de 2 modèles d'évaluation génétique ou génomique, l'un basé sur les contrôles élémentaires et l'autre basé sur les performances à la lactation. Ces évaluations ont été testées et comparées par une validation qui consistait à comparer les index génétiques de jeunes boucs sans descendance à leur index une fois leur descendance produite et suivie sur plusieurs lactations. Pour la validation, nous avons calculé et analysé plusieurs paramètres : le biais, les rapports de variances, les corrélations et les pentes entre les index des jeunes boucs sans descendance et leurs index une fois la descendance produite et mesurée, pour évaluer les qualités prédictives des différents modèles testés.

La deuxième partie présente des analyses complémentaires non présentées dans l'article 4 par soucis de clarté. Ces analyses portent, tout d'abord, sur l'association significative des SNPs aux caractères de production pour les chromosomes non présentés dans l'article 4. Et dans un second temps, nous avons étudié les effets des allèles des SNPs significativement associés aux caractères, tout au long de la lactation, en primipares et en multipares dans les deux races.

Enfin une étude sur les temps de calcul a été réalisée afin d'évaluer la faisabilité de déployer ce modèle pour des évaluations génomiques en routine. Les données utilisées pour cette étude s'étendaient

de 2000 à 2017, pour les trois premières lactations, soit 7 908 192 contrôles élémentaires pour les chèvres de race Saanen (1 308 307 lactations) et 10 483 205 contrôles élémentaires pour les chèvres de race Alpine (1 701 229 lactations).

## I. Principes d'une validation d'un modèle d'évaluation génomique

Le critère le plus couramment utilisé dans la littérature pour comparer les différentes méthodes d'évaluations est la précision des évaluations. Il est calculé comme le coefficient de corrélation de Pearson entre les valeurs génétiques « vraies » et les valeurs génétiques prédites. La valeur génétique « vraie » d'un animal est en général inconnue, on utilise alors une approximation de cette valeur à l'aide du calcul des écarts des productions des filles (Daughter Yield Deviation (DYD)) (VanRaden and Wiggans, 1991).

Les DYD sont calculés à partir des Yield deviation (YD) qui sont des performances brutes corrigées des effets fixes, des effets aléatoires autres que génétiques. Ils s'expriment de la façon suivante :

$$YD = y - (\hat{\beta} + \hat{p})$$

où  $y$  est le vecteur des phénotypes,  $\hat{\beta}$  un vecteur représentant l'estimation des effets fixes, et  $\hat{p}$  l'estimation des effets d'environnement permanent. Pour un mâle, le DYD est calculé comme:

$$DYD = \frac{\sum_{i=1}^n (\widehat{YD}_i - 0.5 * g_{mère}) * w_i}{\sum_{i=1}^n w_i} \text{ où } \widehat{YD}_i \text{ est le YD de la femelle } i, g_{mère} \text{ est l'effet polygénique}$$

de la mère de la femelle  $i$  et  $w_i$  un poids associé au nombre de lactations de la femelle et  $n$  le nombre de filles du mâle. La précision du DYD dépend de son nombre de filles.

La méthode d'évaluation génomique la plus intéressante sera celle pour laquelle la corrélation entre les valeurs génétiques « vraies » et les valeurs génétiques prédites sera la plus proche de 1. La capacité prédictive d'un modèle est particulièrement intéressante pour le choix des jeunes animaux candidats à être utilisés comme producteurs de semence destinée à l'insémination artificielle. En effet, le choix de ces individus largement diffusés dans la population est primordial.

La précision des évaluations génomiques est calculée à l'aide d'une méthode de validation croisée appelée « forward validation ». Elle consiste à diviser la population de référence en 2 sous-populations en fonction de l'année de naissance des animaux. La première sous-population, appelée population d'apprentissage, est constituée des animaux les plus anciens pour lesquels les phénotypes (performances des filles), le pedigree et les génotypes sont connus et utilisés. La seconde sous population, appelée population de validation, contient les animaux les plus jeunes pour lesquels seuls les génotypes et l'ascendance (pedigree) sont connus. La population de validation simule la population actuelle, nous excluons donc toutes les données qui concernent la descendance de ces animaux.

**II. Validation d'une évaluation génomique sur les contrôles élémentaires en caprins laitiers à partir d'un modèle de régression aléatoire et analyse des effets des SNPs issus d'un modèle ssGBLUP**

1 **Interpretive summary**

2 **Validation of a Single-step GBLUP random regression evaluation and analysis of SNP effects on**  
3 **dairy production traits and somatic cell score in French dairy goat**

4 M. Arnal

5 A random regression model (**RRM**) was implemented for the genetic evaluation of goat  
6 lactation curve shapes for milk production and somatic cell scores. Quality of genetic predictions was  
7 assessed on candidate bucks comparing RRM daughter yield deviations (**DYD**) and estimated breeding  
8 values (**EBV**) obtained without progeny, from RRM and a model using accumulated production (**LM**)  
9 with and without genomic information. LM and RRM had quite the same performances for the  
10 regression slopes and for correlations but RRM reduced the bias. Analysis of SNP effects was performed  
11 following genomic evaluations, same genomic regions were found with LM and RRM and specific  
12 regions associated with persistency were found.

13 **Running head:** GENOMIC RANDOM REGRESSION MODEL IN GOATS

14 **Validation of a Single-step GBLUP random regression evaluation and SNP effects analysis on**  
15 **dairy production traits and somatic cell score in French dairy goat**

16 M. Arnal<sup>1,2\*</sup>, C. Robert-Granié<sup>1</sup>, V. Ducrocq<sup>3</sup>, H. Larroque<sup>1</sup>

17 <sup>1</sup>GenPhySE, Université de Toulouse, INRA, INPT, ENVT, 31326 Castanet-Tolosan, France

18 <sup>2</sup>Institut de l'Elevage, Chemin de Borde Rouge, 31326 Castanet-Tolosan cedex, France

19 <sup>3</sup>UMR1313 GABI, INRA, AgroParisTech, Univ. Paris-Saclay, 78350 Jouy-en-Josas, France

20 \*Corresponding author: Mathieu Arnal, [mathieu.arnal@idele.fr](mailto:mathieu.arnal@idele.fr)

21

22           **ABSTRACT**

23           Lactation curve shape is linked to an animal's health, feed requirements, and milk production  
24 throughout the year. Random regression model (**RRM**) are widely used for genetic evaluation of total  
25 milk production throughout the lactation and for milk yield persistency. Genomic information improved  
26 quite dramatically the selection in main dairy cow breeds using Single-step method (**SSGBLUP**). The  
27 aim of this study was to implement a RRM using SSGBLUP in dairy goat in France. A validation  
28 evaluation was realized to evaluate the performance of this type of evaluation with our data targeting  
29 candidate bucks. The model was compared to a non-genomic evaluation and to traditional evaluation  
30 that use accumulated performance throughout the lactation (**LM**). The genomic information increased  
31 correlations between daughter yield deviation (**DYD**) and estimated breeding values (**EBV**) obtained  
32 with partial dataset for candidate bucks. The lactation model and the RRM had quite the same  
33 performance for the slope and correlation. However, the RRM reduced the bias. This study shows that  
34 a ssGBLUP RRM evaluation is possible in dairy goat in France and that its performance is comparable  
35 to LM but with, in addition, a genomic evaluation of persistency. SNP effects study was realized  
36 following the SSGBLUP, the  $-\log_{10}(p\text{-value})$  were analyzed and the same regions were found between  
37 LM and RRM. Several regions associated with persistency were also found.

38           **Keywords** random regression model, dairy goat persistency, genomic, SNP p-values

39

40           **INTRODUCTION**

41           Traditional genetic evaluation for milk yield was based on the accumulated production on 250  
42 days in dairy goats in France (i.e. **LM**). Interpolation between test day records (**TD**) was realized using  
43 the so-called Fleischmann method (Sargent et al., 1968).

44           RRM are used in a lot of country to evaluate genetic values of dairy animals based on their TD  
45 (Oliveira et al., 2019a). RRM permits to better take into account environmental effects and to compute  
46 EBV for milk yield persistency (Schaeffer and Jamrozik, 2008; Arnal et al., 2019a). Persistency is  
47 defined as the ability for an animal to produce less at the beginning of the lactation and more at the end  
48 of the lactation than an average animal that produce the same amount of milk throughout the lactation.  
49 Selection on persistency is useful to reduce health disorders by limiting milk production at the peak of  
50 the lactation. Moreover, it permits a spread of milk production throughout the year, what is an advantage  
51 in dairy goat whose production is very seasoned with a majority of kidding in the middle of winter. The  
52 main disadvantage of this model is its time consuming compared to LM.

53           Genomic information from single-nucleotide polymorphism (**SNP**) is worldly used in  
54 genetic evaluation to improve accuracy of EBVs (Boichard et al., 2016). Single-step GBLUP  
55 (**SSGBLUP**) is the reference method to evaluate genetic values in population where not all animals are

56 genotyped (Legarra et al., 2009; Aguilar et al., 2010). It use directly the phenotypes, the pedigree and  
57 the SNP genotype information building a relationship matrix (**H**) associating genotyped animals with  
58 non-genotyped animals (Legarra et al., 2009). Building H is quite time consuming, so using H matrix in  
59 RRM are not very used despite all the possible advantages of this type of model (Oliveira et al., 2019a).

60 For genetic evaluation validation, in most of the case a cohort of newborn male  
61 candidates for Artificial insemination scheme is the target population. Indeed, it is important to select  
62 the animals that have the best genetic values, when there are newborn because if they are not selected  
63 they are slaughtered. In a genetic evaluation validation, the EBV obtained by the newborn is compared  
64 to its genetic value obtained when it has several recorded progenies because its genetic value can be  
65 considered as the real one if the number of recorded progeny is sufficiently large. Its real genetic value  
66 can be defined as its deregressed EBV (VanRaden et al., 2009), its yield deviation, its DYD (VanRaden  
67 and Wiggans, 1991), or its EBV obtained with its recorded progeny in the dataset (Legarra and Reverter,  
68 2018). Several criteria could be used to compare candidates EBV at birth and its real genetic value after  
69 daughters' lactations. Accuracy of prediction must be evaluated by the correlation between both,  
70 difference in dispersion must be measured by regression coefficient of the real genetic value on the EBV  
71 at birth (slope, better closest to 1), and the bias must be evaluated by the difference between the two  
72 means (zero is expected).

73 After a SSGBLUP evaluation, it is possible to perform an association analysis between  
74 SNP and traits studied (Aguilar et al., 2014). It can be calculated the variance explained by  $n$  adjacent  
75 SNP (Wang et al., 2014) or the p-values of the association (Aguilar et al., 2019). With a random  
76 regression model, it is possible to study the regions associated with persistency as it has been done in  
77 several studies in bovine (Strucken et al., 2012; Oliveira et al., 2019c). In goat, Cardona et al. (2016)  
78 also showed that genes can be expressed differently throughout the lactation.

79 The aim of this study was to implement a genomic SSGBLUP RRM evaluation in France for  
80 dairy goats (Alpine and Saanen breeds) and to evaluate the advantages of this type of model in our  
81 population. To do so, we followed a cohort of young bucks, and assessed the power of prediction at birth  
82 from RRM and LM, with or without genomic information, of the genetic value with progeny (measured  
83 by the accuracy, the dispersion and the bias). Then, we compared analysis of SNP effects with LM and  
84 RRM, and study genomic regions associated with persistency generated by RRM.

## 85 MATERIALS AND METHODS

### 86 *Data*

87 Only the first three parities of Alpine and Saanen goats were analyzed. Each lactation included  
88 TDs between day in milk (**DIM**) 7 and DIM 270. The traits analyzed were milk yield (**MY**), fat yield  
89 (**FY**), protein yield (**PY**), fat content (**FC**), protein content (**PC**), somatic cell score (**SCS**) and fat protein  
90 ratio (**FPR**) which is the ratio of FC to PC. SCS were calculated from somatic cell count as Wiggans et

91 Shook (1987). To be in a situation close to a routine evaluation, we kept all the lactations (from first to  
92 third parity) retained for the current official evaluation (Larroque et al., 2011). Goats were milked twice  
93 a day, and one or two milking were measured. In the case of one single measured milking (32% of our  
94 data), the production was multiplied by a coefficient estimated by a technician from the milk cooling  
95 tank. The total milk, fat and protein yields throughout the lactation were calculated as for routine genetic  
96 evaluation using the so-called Fleischmann method (Sargent et al., 1968). Fat and protein contents were  
97 derived from the total lactation yields.

98 To test the qualities of prediction of the genetic evaluation, we assumed a situation close to the  
99 current breeding scheme: the future bucks are selected as newborn so we divided our complete dataset  
100 (c) and we kept the part from 2000 to 2011, called partial dataset (p). The assumed newborn bucks,  
101 called candidate bucks, were the bucks born between 2010 and 2013. The complete dataset (c) consisted  
102 in 7,908,192 TD records from 1,308,307 lactations of Saanen goats and 10,483,205 test-day records  
103 from 1,701,229 lactations of Alpine goats, collected in France between 2000 and 2017. With the  
104 complete dataset, it will be possible to calculate the true genetic value of the candidate bucks based on  
105 the performance of their progeny. For the analysis of SNP effects, due to software limitation, we had to  
106 reduce the size of our complete dataset so we kept only lactations of goats with two milking a day  
107 recorded, being part of a HTD (**Herd Test Day**) with more than 4 goats, with more than 4 TD, and  
108 whose sire had at least 25 recorded daughters. The details of the number of TD and lactations by dataset  
109 are given in Table 1. The pedigree consisted of 818,702 animals in Saanen and 1,021,665 in Alpine.

110 French dairy goats were genotyped with the Illumina goat SNP50 BeadChip (Tosser-Klopp et  
111 al., 2014). Applied rules for SNP quality control were the same as in Teissier et al. (2019). At the end  
112 of quality control, we had 47,206 SNPs. Table 2 presented the number of genotyped animals and the  
113 number of animals in the pedigree: 1,242 Saanen genotyped animals (430 Saanen genotyped bucks) and  
114 1,925 Alpine genotyped animals (555 Saanen genotyped bucks) were involved in the evaluation. The  
115 assumed newborn genotyped candidate bucks were 133 in Saanen and 172 in Alpine (Table 2). The  
116 number of candidates was not large, but we could not have more candidates without further reducing  
117 the reference population.

### 118 *Models of evaluation*

119 Four different evaluations were implemented and compared. The evaluations were run with the  
120 software blup90iod2 (Misztal et al., 2002). Unknown parent groups were defined according to year of  
121 birth (classes of 4 years) and the sex of the animal. A single group was created for animals born before  
122 1975.

123 *Classical lactation model (LM)*. A LM close to the one routinely used for official genetic  
124 evaluations in France (Larroque et al., 2011) was implemented. The genetic evaluation model based on  
125 these phenotypes was:

$$y_{rtnjlkmpi} = H_{trji} + A_{tjkp} + M_{trjlp} + D_{trjmp} + a_{tn} + p_{tn} + e_{rtnjlkmpi}$$

where  $y_{rtnjlkmpi}$  was the lactation observation for trait  $t$  (MY, FY, PY, FC, PC, SCS and FPR) of goat  $n$ , in production year  $j$  (2000, ..., 2017), in parity  $r$  (1, 2, 3), belonging to kidding age class  $k$  (7 classes for first lactation (in months): 9-11, 12, 13, 14, 15, 16, +17; 6 classes for second lactation :21-23, 24, 25, 26, 27, +28; 6 classes for third lactation : 31-35, 36, 37, 38, 39, +40), kidding month class  $l$  (7 classes: January, February, March-April-May, June-July-August-September, October, November, December), class  $m$  of dry period length (6 classes (days): [First parity], [0,50], ]50,75], ]75,100], ]100,125], 125+), in region  $p$  (4 classes: North-West, North-East, South-West, South-East) and herd  $i$ .  $H_{trji}$  is the fixed effect of herd;  $A_{tjkp}$  is the fixed effect of age at kidding;  $M_{trjlp}$  was the fixed effect for the kidding period;  $D_{trjmp}$  the fixed effect of length of the dry period;  $a_{tn}$  was the additive genetic breeding value that followed a normal distribution (mean:  $\mu_{LACT}=0$ , variance= $\sigma_{LACT}^2A$ ) with  $A$  the additive genetic relation matrix based on pedigree information, or (mean:  $\mu_{LACT}=0$ , variance= $\sigma_{LACT}^2H$ ) with  $H$  the additive genetic relation matrix based on pedigree and genomic information as in Legarra et al. (2009) with  $H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ \tau((1-w)G + wA_{22})^{-1} & -\omega A_{22}^{-1} \end{bmatrix}$ . We used  $\tau=1$ ,  $w=0.05$  and  $\omega=1$  which are the default values used in blupiod2 (Misztal et al., 2002). For the scaling of  $G$ , the mean of the diagonal of  $G$  was set equal to the mean of the diagonal of  $A$ . The mean of the out of diagonal of  $G$  was set to be equal of the out of diagonal of  $A$ .  $p_{tn}$  was the permanent environment value that followed a normal distribution (mean:  $\mu_{PE\_LACT}=0$ , variance:  $\sigma_{PE\_LACT}^2$ ).

Using the partial dataset, the EBVs obtained from the model that used pedigree information for building the relationship matrix were called **A\_LM**. Using the complete dataset, the EBVs obtained from the model that used pedigree information for building the relationship matrix were called **A\_LM\_COMP**. Using the partial dataset, the EBVs obtained from the model that used the genomic information and the pedigree information to build the relationship matrix was called **H\_LM**.

**Random regression model (RRM).** The model used was the “EGV\_PM” described in Arnal et al. (2019b). The second and the third parities (multiparous) were modelled together for genetic and permanent environment. For the fix effects, the second and third parities were kept separated.

$$y_{trsijklm d g n p q} = HTD_{tri} + A_{tjkp} + M_{trjlp} + D_{trjmp} + \sum_{o=1}^6 \theta_{tko} N_{(o,d)} + \sum_{o=1}^6 \tau_{trlo} N_{(o,d)} + \sum_{o=1}^6 \pi_{trmo} N_{(o,d)} + \sum_{o=1}^4 \gamma_{tso} M_{(o,g)} + \sum_{o=1}^2 b_{tson} \chi_{(to,d)} + \sum_{o=1}^2 c_{tson} \chi_{(to,d)} + e_{trsijklm d g n p q}$$

where  $y_{trsijklm d g n p q}$  = observation for trait  $t$  (MY, FY, PY, FC, PC, SCS, FPR) of day in milk (DIM)  $d$  (7, ..., 270) of goat  $n$  carrying a kid for  $g$  day(s) (0, ..., 100), in production year  $j$  (1995, ...,

157 2015), in parity  $r$  (1, 2, 3), in parity class  $s$  (primiparous (P) or multiparous (M)), belonging to kidding  
 158 age class  $k$  (7 classes for first lactation (in months): 9-11, 12, 13, 14, 15, 16, +17; 6 classes for second  
 159 lactation :21-23, 24, 25, 26, 27, +28; 6 classes for third lactation : 31-35, 36, 37, 38, 39, +40), kidding  
 160 month class  $l$  (7 classes: January, February, March-April-May, June-July-August-September, October,  
 161 November, December), class  $m$  of dry period length (6 classes (days): [First parity], [0,50], ]50,75],  
 162 ]75,100], ]100,125], 125+), in region  $p$  (4 classes: North-West, North-East, South-West, South-East),  
 163 herd  $\times$  test-date class  $i$ .  $HTD_{tri}$  is the fixed effect for the herd-test-date;  $A_{tj_kp}$  is the fixed effect for the  
 164 age at kidding;  $M_{trjlp}$  is the fixed effect for the kidding period;  $D_{trjmp}$  is the fixed effect for length of  
 165 the dry period;  $\theta_{tko}$ ,  $\tau_{trlo}$ ,  $\pi_{trmo}$  and  $\gamma_{tso}$  are fixed regression coefficients for age at kidding, the  
 166 kidding month, dry period length and gestation stage respectively;  $N_{(o,d)}$  is the  $o^{\text{th}}$  covariate at time  $d$  of  
 167 a cubic natural spline function with 6 knots at  $d= 7, 20, 50, 110, 190, 270$ ;  $M_{(o,g)}$  is the  $o^{\text{th}}$  covariate at  
 168 time  $d$  of a cubic natural spline function with 4 knots at  $d= 31, 53, 76, 100$  (between  $g=0$  and  $g=30$ , the  
 169 coefficient were assumed to be equal to 0; if the gestation stage was greater than 100, it was supposed  
 170 to be equal to 100);  $b_{trno}$  and  $c_{trno}$  are the random additive genetic and permanent environmental  
 171 regression coefficient for the  $o^{\text{th}}$  eigenvectors of the genetic (co)variances matrix obtained with a 2nd-  
 172 order Legendre polynomial model reduced to rank 2,  $\chi_{(to,d)}$  the value of the  $o^{\text{th}}$  eigenvector of the  
 173 genetic (co)variances matrix obtained with a 2nd-order Legendre polynomial model reduced to rank 2  
 174 at DIM  $d$  for parity class  $s$   $e_{trsjklmdgnqp}$  is the residual term. A homogenous residual variance was  
 175 considered due to software limitations. The random additive genetic regression coefficient for the first  
 176 eigenvector,  $b_1$ , is noted **LEV** (related to production level) and for the second eigenvector,  $b_2$ , is noted  
 177 **PERS** (persistence). LEV is noted **LEV\_P** for primiparous, and is noted **LEV\_M** for multiparous. PERS  
 178 is noted **PERS\_P** for primiparous, and is noted **PERS\_M** for multiparous.

179 EBVs for total production were computed separately for primiparous (**primi**) and multiparous  
 180 (**multi**). To do that, each daily EBV was computed as:

$$181 \quad EBV_{primi\ d} = LEV\_P * \chi_{(1,d)} + PERS\_P * \chi_{(2,d)}$$

$$182 \quad EBV_{multi\ d} = LEV\_M * \chi_{(3,d)} + PERS\_M * \chi_{(4,d)}$$

183 Daily EBVs were then added to obtain a total EBV for primiparous ( $EBV_{primi}$ ) and for  
 184 multiparous ( $EBV_{multi}$ ), as:

$$185 \quad EBV_{primi} = \sum_{d=1}^{264} EBV_{primi\ d}$$

$$186 \quad EBV_{multi} = \sum_{d=1}^{264} EBV_{multi\ d}$$

187 Finally, EBVs from primiparous and multiparous were combined as follow:

188 
$$EBV_{tot} = 0.33 * EBV_{primi} + 0.66 * EBV_{multi}$$

189 The two coefficients were chosen to give same weight for the first three parities.

190 The same procedure was applied to obtain an EBV for the level of production and for  
191 persistency.

192 
$$lev_{tot} = 0.33 * lev_{primi} + 0.66 * lev_{multi}$$

193 
$$pers_{tot} = 0.33 * pers_{primi} + 0.66 * pers_{multi}$$

194 Using partial dataset,  $EBV_{tot}$  from the model that used only pedigree information for building  
195 the relationship matrix was called **A\_RRM**. Using complete dataset,  $EBV_{tot}$  from the model that used  
196 only pedigree information for building the relationship matrix was called **A\_RRM\_COMP**. Using  
197 partial dataset,  $EBV_{tot}$  from the model that used genomic information in addition to the pedigree for  
198 building the relationship matrix as in Legarra et al. (2009) was called **H\_RRM**. Using partial dataset,  
199  $lev_{tot}$  from the model that used only pedigree information for building the relationship matrix was  
200 called **A\_LEV**. Using partial dataset,  $lev_{tot}$  from the model that used genomic information in addition  
201 to the pedigree for building the relationship matrix as in Legarra et al. (2009) was called **H\_LEV**. Using  
202 partial dataset,  $pers_{tot}$  from the model that used only pedigree information for building the relationship  
203 matrix was called **A\_PERS**. Using partial dataset,  $pers_{tot}$  from the model that used genomic  
204 information in addition to the pedigree for building the relationship matrix as in Legarra et al. (2009)  
205 was called **H\_PERS**.

206 ***Daughter Yield deviation (DYD)***

207 DYD were calculated from RRM using the complete dataset but without genomic information.  
208 We note **DYD\_RRM** the DYD for total lactation production obtained as  $EBV_{tot}$ . We note **DYD\_LEV**  
209 and **DYD\_PERS** the DYD for LEV and PERS calculated as  $lev_{tot}$  and  $pers_{tot}$ .

210 DYD were calculated also from LM using the complete dataset and was noted **DYD\_LM**.

211 DYD were calculated using genekit software (Ducrocq, 1998) as in Täubert et al (2010).

212 ***Criteria of evaluation comparison***

213 We looked at the number of progeny of candidate bucks in 2018 and analyzed the reliability  
214 obtained in 2018 as in Ducrocq and Schneider (2007) with genekit software. These elements are  
215 presented in Table 3. Because the number of multiparous progeny was heterogeneous between bucks  
216 (from 6 to 152, median 40 in Saanen, and from 6 to 179 median 61 in Alpine) calculations of bias,  
217 slopes, variance ratios and correlations were weighted by the number of multiparous progeny.

218 ***Bias*** The mean of DYD and EBVs were set to zero for the bucks (not necessarily AI bucks) with  
219 more than 25 progeny in multiparous born in 1997. The bias was calculated as:

220 
$$\mu_{DYD,p} = \frac{\hat{u}_p - DYD}{\sigma_{DYD}}$$

221 with  $\mu$  the bias,  $\hat{u}_p$  EBV obtained with partial data, and  $\sigma_{DYD}$  a standard deviation of DYD.  
 222 Significance of difference between biases from two evaluations was tested by using a Student t-test and  
 223 a difference was considered significant if the p-value was smaller than 0.1.

224 **Slope.** The slope was calculated as:

225 
$$b_{DYD,p} = \frac{cov(\hat{u}_p, DYD)}{var(\hat{u}_p)}$$

226 with  $b_{DYD,p}$  the slope of the regression of DYD on EBV obtained with partial dataset. A  
 227 generalized linear model was used to test the difference of slope between the evaluations. The model  
 228 was:

229 
$$DYD = \hat{u}_p + M + \hat{u}_p: M$$

230 with M the evaluation and  $\hat{u}_p: M$  the interaction between the index from partial dataset and the  
 231 evaluation. The p-value of the interaction was studied and considered significant if the p-value was  
 232 smaller than 0.1.

233 **Correlation.** The correlation was:

234 
$$\rho_{DYD,p} = \frac{cov(\hat{u}_p, DYD)}{\sqrt{var(DYD)var(\hat{u}_p)}}$$

235 with  $\rho_{p,DYD}$  the coefficient of correlation between EBVs obtained with partial dataset and the  
 236 DYD. The Hotelling-Williams test (Steiger, 1980) was used to compare the correlation obtained with  
 237 the different evaluations.

238 **Variance ratio.** The variance ratio was:

239 
$$R = \frac{Var(\hat{u}_p)}{Var(DYD)}$$

240 with R the variance ratio between the variance of EBVs obtained with partial dataset ( $Var(\hat{u}_p)$ ),  
 241 and the variance of DYD ( $Var(DYD)$ ). A Fisher test was used to test differences of variances and  
 242 considered significant if the p-value was smaller than 0.1.

243 **Analysis of SNP effects**

244 Markers effects were calculated by SSGBLUP after estimation of breeding values of H\_LM and  
 245 H\_RRM models. P-values of SNPs were calculated with PostGSf90 (Aguilar et al., 2014) according to  
 246 Aguilar et al. (2019). The variance explained by 20 adjacent SNPs was also calculated.

247 Each calculated SNP effect was subjected to a *t*-test to calculate the  $-\log_{10}(P\text{-value})$ . A  
248 Bonferroni correction of  $\alpha = 5\%$  was applied (Significance threshold =  $\alpha/\text{number of SNPs}$ ). Any SNP  
249 with  $P < 1.06 \times 10^{-6}$  were considered significantly associated with the trait studied which is equivalent  
250 to a  $-\log_{10}(p\text{-value})$  superior to 6.

251 For the percentage of variance explained by 20 adjacent SNPs, we kept variance explained for  
252 the biggest non-overlapping windows segments; it means that the sum of variance percentages is equal  
253 to 100% of the total variance.

## 254 **RESULTS**

### 255 *Correlations between evaluations*

256 Correlations between EBVs for candidate bucks with partial dataset (without performances of  
257 progeny) were presented in Figure 1.

258 The correlation between A\_LM and H\_LM were high: close to 0.9 for contents and SCS in both  
259 breeds; close to 0.8 for MY, in both breeds and FY in Saanen; and close to 0.72 for PY in Saanen and  
260 FY in Alpine. The smallest correlation (0.62) was found for PY in Alpine. The correlation between  
261 A\_RRM and H\_RRM were very close to the correlation between A\_LM and H\_LM except for SCS in  
262 Alpine where the correlation was close to 0.7, therefore, the genomic information brings the same changes  
263 with both models (LM or RRM).

264 The correlation between A\_LM and A\_RRM were higher than 0.9 for all traits in both breeds.  
265 The correlations between H\_LM and H\_RRM were close to those between A\_RRM and H\_RRM but a  
266 bit smaller and particularly for SCS in Alpine where the correlation was close to 0.8.

### 267 **Choice of the reference genetic value of bucks after daughters' lactations**

268 D<sub>YD</sub>\_RRM were compared to, D<sub>YD</sub>\_LM and EBVs of A\_RRM from complete dataset  
269 (A\_RRM\_COMP) in terms of correlations, means and variances in Supplemental Figure S1, S2 and S3  
270 respectively. Correlations with A\_RRM\_COMP were high: between 0.9, for PY in Saanen, and 0.98,  
271 for FC in both breeds. Correlations with D<sub>YD</sub>\_LM were also high: between 0.92, for PY in Saanen, and  
272 0.98 for contents in both breeds (Figure S1).

273 D<sub>YD</sub>\_RRM and A\_RRM\_COMP means were very similar (Figure S2): the differences between  
274 them ranged from 0 D<sub>YD</sub>\_RRM standard deviations for FPR in Saanen, to 0.12 D<sub>YD</sub>\_RRM standard  
275 deviations for PY in Saanen. The means were significantly different only for PY in Saanen ( $p\text{-value}$   
276 = 0.03). However, difference of means between D<sub>YD</sub>\_RRM and D<sub>YD</sub>\_LM were higher: between 0.03  
277 D<sub>YD</sub>\_RRM standard deviations for SCS in Alpine and 0.84 D<sub>YD</sub>\_RRM standard deviations for PY in  
278 Alpine. The means were significantly different for yields in both breeds and for FPR in Alpine.

279 The ratios of variances between DYD\_LM or A\_RRM\_COMP and DYD\_RRM were studied  
280 to compare variances (Figure S3). Variances of A\_RRM\_COMP were always smaller than those of  
281 DYD\_RRM, with a ratio between 0.75 for FY in Saanen and around 0.9 for contents in both breeds.  
282 These variances were significantly different for FY and PY in Saanen and for SCS in Alpine. For  
283 DYD\_LM, the ratio range from 0.8 for PC in Alpine to 1.25 for MY in both breeds for all traits except  
284 SCS. For SCS, a big difference was found with a ratio of 1.9 in Saanen and 2.2 in Alpine. The variances  
285 were significantly different only for this trait in both breeds.

286 These results showed that DYD\_RRM and A\_RRM\_COMP were relatively close in terms of  
287 correlations, means and variances. Therefore, DYD can be kept as 'reference parameter' because it  
288 based on a raw phenotype corrected for fix effects. DYD\_LM and DYD\_RRM were highly correlated,  
289 but shown important differences of means for yield traits and of variances for SCS. DYD from RRM  
290 were finally chosen because Arnal et al. (2019a) showed that modeling the genetic and permanent  
291 environment effects according to DIM is more accurate than to consider these effects constant along  
292 lactation. More over in a RRM, other environmental parameters in the fix part are defined more precisely  
293 (gestation stage, age at kidding, month of kidding and dry period length depending on the lactation stage,  
294 herd-test day effect instead of herd-year effect). For the genetic part, the separation of parities improved  
295 also the r-squared in a preliminary study (results not shown). Finally, DYD\_RRM was used as the  
296 reference genetic value of bucks for the validation study.

### 297 *Validation of the different genetic evaluations*

298 We used different criteria to test the prediction abilities of the different models,  
299 comparing the EBVs of genotyped bucks obtained as newborn to their DYD\_RRM obtained from the  
300 performances of their progeny. A model is valuable if the EBVs of the newborns are close to their  
301 DYD\_RRM for the mean (studied by the bias), for the variance (studied by the ratio of variances), with  
302 a correlation close to one (assessed by the accuracy criterion), a slope of the regression of the  
303 DYD\_RRM on the EBVs close to one, this criteria measure of closeness and of the EBVs dispersion.

304 The results for LEV were the same as observed for RRM ( $EBV_{tot}$ ), this result was observed in  
305 a previous study (Arnal et al., 2019a) with the correlation close to 1 between A\_RRM and A\_LEV.

306 **Correlations.** Correlations between the DYD\_RRM from complete dataset and A\_LM,  
307 A\_RRM, H\_LM and H\_RRM obtained with partial dataset were presented in Figure 2. EBVs for traits  
308 with the highest heritabilities, i.e. contents (Arnal et al., 2019b), had the highest correlations with  
309 DYD\_RRM in all models (around 0.65). Correlations between DYD\_RRM and H\_LM or H\_RRM were  
310 around 0.4 for yields and SCS in Saanen, as well as for MY and SCS in Alpine. The correlations between  
311 EBVs and DYD\_RRM were the lowest ones (around 0.25) for FY and PY in Alpine.

312 The use of genomic information improves the correlations around +0.07 for all traits in Saanen  
313 and until +0.19 for PY with the LM for Saanen. The gain of correlation was the same with both models

314 LM and RRM and was significant for all traits in Saanen. For Alpine, the same tendency was observed  
315 except for FC and SCS, but this gain was significant only for PC with the RRM. However, the  
316 improvement of correlation using genomic information (around +0.04) was smaller than in Saanen. In  
317 Alpine breed, the highest gain (+0.1) was observed for PY for LM and RRM, like in Saanen.

318 Correlations between DYD\_RRM and LM or RRM were very close for both breeds and with or  
319 without genomic information. For FY, FC and FPR in Saanen, the RRM have significant lowest  
320 correlations than LM with genomic. It was also the case for SCS in Alpine. For production traits in  
321 Alpine, the RRM gave higher correlations than LM and it was significant for FY without genomic  
322 information and for MY and FY with genomic information.

323 Correlation between DYD for LEV and PERS and the EBVs for LEV and PERS obtained with  
324 partial dataset with or without genomic information are presented it in Supplemental Figure S4. For  
325 yield traits and SCS in Alpine, correlations for persistency were higher than correlations for the level,  
326 particularly without genomic information. In Alpine, the maximal difference was found for PY, where  
327 the correlation for A\_LEV was equal to 0.19 whereas the correlation for A\_PERS was equal to 0.48.  
328 For contents in both breeds and SCS in Saanen, correlation of DYD\_LEV with A\_LEV was higher than  
329 with A\_PERS. In Saanen, the highest difference was for FC where the A\_LEV correlation was equal to  
330 0.55 and the A\_PERS equal to 0.32. For most of the traits, the genomic information improve more the  
331 LEV (around +0.07 in Saanen) than the PERS (around +0.04 in Saanen), which was particularly true for  
332 yields. As for total EBVs and LEV, the correlations for PERS were more improved in Saanen than in  
333 Alpine. The genomic information improved significantly only MY and PC persistency in Saanen.

334 **Slope.** Slopes of the regression of DYD\_RRM on the A\_LM, A\_RRM, H\_LM and H\_RRM  
335 obtained with partial dataset are compared in Figure 3. As for correlations, the best slopes (closest to 1)  
336 were obtained for traits with the highest heritabilities, i.e contents (between 0.9 and 1.3 for both breed),  
337 and the worst for FY and PY (between 0.3 and 0.5) in Alpine. For yields and SCS in both breeds contrary  
338 to contents, slopes were inferior to 1, indicating an over-dispersion of bucks EBVs at birth.

339 The RRM seemed to give slopes closest to 1 compared to LM for almost all the traits and  
340 especially for MY in Alpine, whose difference between the two models was significant when using  
341 genomic information (H\_LM vs H\_RRM).

342 For LM, for most of the traits, the use of genomic information improve slopes and this difference  
343 of slopes (A\_LM vs H\_LM) was significant for PY in Saanen and FC in Alpine. With RRM, the use of  
344 genomic information improved also the slopes except for SCS in Alpine where the slope was  
345 significantly worse.

346 Regression slopes of DYD\_LEV on A\_LEV and H\_LEV estimated with partial dataset and  
347 regression slope of DYD\_PERS on A\_PERS and H\_PERS obtained with partial dataset are presented  
348 in Supplemental Figure S5. For yields traits, the slopes obtained for persistency were similar to those

349 obtained for LEV. For contents and SCS the slope of PERS were smaller than those obtained for LEV.  
350 All the slopes for PERS were relatively close, around 0.6. The genomic information improved the slope  
351 of PERS for MY and PY in Saanen and decreased the slope for the other traits, significantly for SCS in  
352 Alpine.

353 **Variations.** Variance ratios between candidate bucks' EBVs from the partial dataset and from  
354 the 4 models and the DYD\_RRM from the complete dataset, for all traits in Saanen and Alpine were  
355 given in Figure 4. For all models, the ratio were inferior of 1, meaning that the variance of candidate  
356 bucks EBVs were smaller than DYD\_RRM. The variances were all significantly different from  
357 DYD\_RRM except for SCS in Saanen with H\_LM (p-value=0.08). LM gave a significantly higher  
358 variance of EBVs than RRM for SCS in Saanen, MY, FY (without genomic), PY, and SCS in Alpine.  
359 For PC in Alpine, RRM gave significantly a higher variance than LM.

360 The genomic information increased the variance of EBVs for PY with RRM and FC with LM  
361 and RRM in Saanen. The genomic information increased also the variance of EBVs for FY and SCS  
362 with RRM, and PC and FPR with LM and RRM in Alpine. For the other traits, the variances were not  
363 significantly different.

364 Variance ratios between candidate bucks' EBVs from the partial dataset for LEV and PERS and  
365 the DYD\_LEV and DYD\_PERS from the complete dataset for the different traits in Saanen and Alpine  
366 were given in Supplemental Figure S6. For PERS, the ratio was always smaller than 1 for all traits in  
367 both breeds and was significantly different except for FPR in Alpine using genomic information. The  
368 use of genomic information increased significantly the variance of PERS for FC, SCS and FPR in  
369 Alpine.

370 **Bias.** Bias of the different EBVs at birth was compared in Figure 5. The LM gave important  
371 bias, between 0.16 and 1.1 DYD\_RRM standard deviation for yields traits and SCS in both breeds and  
372 for FC in Saanen. The means of EBVs were significantly different to DYD\_RRM for these traits. The  
373 bias was all positive except for PC and FPR in both breeds, which means that the  $EBV_{tot}$  were mainly  
374 overestimated.

375 Smallest bias was found with RRM compared to LM, significantly for yields in both breeds and  
376 for SCS in Saanen. Indeed, the bias ranges between 0.19 DYD\_RRM standard deviation and 0.45 for  
377 both breeds. For SCS, the bias was higher in Saanen, around 0.75 DYD\_RRM standard deviation using  
378 LM and around 0.5 using RRM whereas it was around 0.25 in Alpine using LM and RRM. For contents,  
379 the bias was smaller than for yields and SCS (smaller than 0.25).

380 The use of genomic information did not impact the bias except for FY and PY in both breeds  
381 for which the bias was significantly reduced using LM. The use of genomic information with RRM did  
382 not change significantly the bias.

383 The bias was reported for LEV and PERS in Supplemental Figure S7. The means of A\_PERS  
384 and H\_PERS were significantly smaller than DYD\_PERS for MY, FC, and FRR in Saanen, and MY,  
385 FY, PY, FC and FPR in Alpine. The bias for PERS was higher than those for LEV for yield traits, FC  
386 and FPR in Alpine, comprised between -0.38 and -0.6 DYD\_PERS standard deviation. It was even  
387 higher, between -0.6 and -1.15, in Saanen for PERS for MY, FC and FPR. The bias was around -0.2 for  
388 PERS for FY and PY in Saanen. The bias for PERS was negative for yields and FC and FPR, which  
389 means that the EBVs of PERS were underestimated compared to DYD. For PERS, the bias was positive  
390 and smaller than 0.3 for SCS and PC in both breeds. The genomic information has no significant effect  
391 on bias for any traits in any breed.

### 392 *Evolution of mean and variance EBV throughout the years*

393 The mean of EBVs with partial and complete dataset and the DYD\_RRM and DYD\_LM of  
394 genotyped bucks were plotted in Supplemental Figure S8 for MY in Alpine. DYD\_LM and DYD\_RRM  
395 were parallel confirming the correlations close to one between DYD\_LM and DYD\_RRM of  
396 Supplemental Figure S1. However, the means were not the same as it could see for the bias. The  
397 difference increased from 1997, with a difference of 10 kg, which was close to 50 kg in 2013. It was  
398 also the case for the EBVs but with an inflammation from 2000 compared to DYD. A\_LM and H\_LM  
399 were very close each other as A\_RRM and H\_RRM. A\_LM and H\_LM were close to DYD\_LM until  
400 2010 (first year of candidate bucks) and then we could observe an inflation of EBVs as it could be  
401 observed for candidate bucks (2010-2013) for bias (Figure 5). The same results were observed with  
402 RRM. The EBVs obtained with the complete dataset without genomic were very close to the DYD for  
403 both models.

404 The variance of EBVs with partial and complete dataset and the DYD\_RRM and DYD\_LM of  
405 genotyped bucks were plotted in Supplemental Figure S9 for MY in Alpine. All the variances were quite  
406 the same until 2008 and then the variances of EBVs obtained with partial dataset dropped significantly:  
407 EBVs with partial dataset were around 2000kg<sup>2</sup> whereas DYD\_RRM around 6200kg<sup>2</sup>.

### 408 *Analysis of SNP effects*

409 The maximal  $-\log_{10}(\text{p-value})$  of SNP by chromosome with a  $-\log_{10}(\text{p-value})$  superior  
410 to 6 are presented in Figure 6. Positions of SNP of the maximal  $-\log_{10}(\text{p-value})$  by chromosome  
411 superior to 6 are presented in Table 4. Regions significantly associated were found for all the traits in  
412 Saanen. However, no region was found for yield traits in Alpine. The positions of the 1<sup>st</sup> SNP of the  
413 segment of 20 SNPs that were maximal by chromosome and that represented more than 1 % of the total  
414 variance are given in Supplemental Table S1. The percentage of variance by chromosome, which were  
415 superior to 1% of the total variance, was given in Supplemental Figure S10.

416 **LM** For LM, same SNP positions were found for FC and FPR in Alpine on chromosome 8 and  
417 14; for FC and PC in Saanen on chromosome 6; for MY, FY and PY in chromosome 19 for Saanen.

418 Close SNP positions were found on chromosome 6: less than 200 kb for FC and PC in both breeds; on  
419 chromosome 14: less than 600 kb for FY in Saanen, FC and FPR in both breeds, on chromosome 19:  
420 less than 2 Mb for MY, FY, PY and SCS in Saanen. Few regions with  $-\log_{10}(\text{p-value})$  very close to 6  
421 (Figure 6) were found in chromosome 2 and 3 in Alpine for PC and SCS respectively. The regions with  
422 highest  $-\log_{10}(\text{p-value})$  were found for PC in chromosome 6 in both breed (around 40). For the other  
423 traits,  $-\log_{10}(\text{p-value})$  was between 6 and 20.

424 **LM/RRM** Same positions as LM were found for LEV\_P and LEV\_M for MY, PY, FC, PC, SCS  
425 and FPR in Saanen and for FC, PC, FPR in Alpine. In a close region, around 1 Mb, same region as LM  
426 and LEV\_P were found for LEV\_M, in Saanen for FY. No region on chromosome 2 and 3 were  
427 significant with RRM. The region on chromosome 14 for FY in Saanen, that had a  $-\log_{10}(\text{p-value})$  close  
428 to 6 was not found with RRM. One region with a  $-\log_{10}(\text{p-value})$  close to 6 was observed on  
429 chromosome 23 for SCS in Alpine at 24.8 Mb for LEV\_M.  $-\log_{10}(\text{p-value})$  were higher for LEV\_P  
430 than LM which was higher than LEV\_M for MY, FY, PY in Saanen.  $-\log_{10}(\text{p-value})$  were higher for  
431 RRM than LM also for FC on chromosome 6 and 8 and FPR on chromosome 8 in Alpine. However, the  
432 differences were small. The biggest  $-\log_{10}(\text{p-value})$  difference was for PC on chromosome 6:  $-\log_{10}(\text{p-}$   
433  $\text{value})$  was equal to 40 for LM and around 26 for RRM. The biggest  $-\log_{10}(\text{p-value})$  difference between  
434 primiparous and multiparous was for MY and FY in Saanen on chromosome 19 with a difference of  $-\log_{10}(\text{p-}$   
435  $\text{value})$  of 5. Three regions were found associated with persistency: for FY for primiparous in  
436 Saanen on chromosome 19 ( $-\log_{10}(\text{p-value}) = 8$ ), PY for multiparous in Saanen on chromosome 19 ( $-\log_{10}(\text{p-}$   
437  $\text{value})$  around 6) and finally for SCS for primiparous on chromosome 23 in Alpine 19 ( $-\log_{10}(\text{p-}$   
438  $\text{value})$  around 6).

439 Figure 7 presents the  $-\log_{10}(\text{p-value})$  for each SNP on chromosome 19, in Saanen breed, for  
440 MY, PY, FY and SCS, and for LM, LEV\_P, PERS\_P, LEV\_M and PERS\_M. Majority of SNPs  
441 associated with MY, FY, PY and SCS were all in the same region: between 21 and 30 Mb. SNPs  
442 associated with FY primiparous persistency were also in this area. Primiparous persistency for MY and  
443 PY in Saanen were also in this area but their  $-\log_{10}(\text{p-values})$  were below 6 (5.2 for PY and 4.5 for  
444 MY). The persistency of PY for multiparous was not in the same area but for persistency of MY in  
445 multiparous the same region was found with a  $-\log_{10}(\text{p-value}) > 5$ .

446 **Percentage of variance.** Some genomic regions were not exactly found when we studied the  
447 variance explained by 20 adjacent SNPs. Indeed, only regions representing more than 1% of the total  
448 variance on chromosomes 6 (around  $85.8 \pm 0.1$  Mb) and 19 (around  $26 \pm 0.2$  Mb) in Saanen and on  
449 chromosome 6 (around  $85 \pm 0.1$  Mb) in Alpine were the same as the ones found with p-values for the  
450 same traits and models. Areas with high  $-\log_{10}(\text{p-values})$  on chromosome 14 in both breeds and on  
451 chromosome 8 in Alpine represented less than 1% of the variance. Regions more anecdotic as on  
452 chromosome 19 associated with PY multiparous persistency, on chromosome 2, 3 and 23 represented  
453 also less than 1%. However, some regions represented more than 1% of the total variance even though

454 no SNP in the region had a  $-\log_{10}(\text{p-value}) > 6$ . It was the case on chromosome 6 (around  $85.8 \pm 0.1$   
455 Mb) with MY (LM, LEV\_P, LEV\_M), FC (PERS\_M), SCS (LM, LEV\_P, PERS\_P, PERS\_M), FPR  
456 (PERS\_P, PERS\_M) in Saanen and with MY (LEV\_M) in Alpine, and on chromosome 19 (around  $26 \pm$   
457 0.2 Mb) for MY (PERS\_P) and PY (PERS\_P) in Saanen.

458 Maximal percentage of total variance represented by segment of 20 consecutive SNPs, by  
459 chromosome and higher than 1% were presented in Supplemental Figure S10. The maximal percentage  
460 of variance represented by a segment was equal to 15% for PC in Saanen on chromosome 6 with LM  
461 and LEV\_M. For the other traits, it ranges between 1 and 5.2% for total production throughout the  
462 lactation and between 1 and 2.7 % for persistencies.

463 Supplemental Figure S11 presents the explained percentage of variance for each window of 20  
464 SNPs on chromosome 19, in Saanen breed, for MY, PY, FY and SCS, and for LM, LEV\_P, PERS\_P,  
465 LEV\_M and PERS\_M. The majority of SNPs associated with MY, FY, PY and SCS were all in the  
466 same region as with p-values (Figure 7): between 21 and 30 Mb. SNPs associated with MY, FY and PY  
467 primiparous persistency were in this area. Persistency of SCS in primiparous was also in this region  
468 representing 0.8% of the total variance. Percentages of total variance for multiparous persistency of MY,  
469 FY and PY were higher in the same area as for p-values i.e. around 37.6 Mb.

## 470 **DISCUSSION**

471 For the construction of the H matrix, we did not want to change the parameters  $\omega$  and  $\tau$ , we used  
472 default parameters. However,  $\omega = 1$ ,  $\tau = 1$  and  $w = 0.05$  gave satisfactory results compared to other values  
473 in other studies (Kang et al., 2018; Oliveira et al., 2019d).

### 474 ***Choice of the reference genetic value of bucks after daughters' lactations***

475 We saw that the correlation between A\_RRM and DYD\_RRM for candidate bucks using the  
476 complete dataset was close to one for all traits. It was also the case for the means of DYD and EBVs by  
477 year of birth for genomic bucks with LM and RRM. These results confirm that using DYD is relevant.  
478 The method proposed by Legarra et Reverter (2018) that used EBVs instead of DYD to do a genomic  
479 validation seems also relevant.

480 We chose DYD from RRM instead from LM as reference values because Arnal et al. (2019a)  
481 showed that modeling the genetic and permanent environment effects according to DIM is more accurate  
482 than to consider these effects constant along lactation. More over with our RRM, other environmental  
483 parameters in the fix part are more precisely defined (gestation stage, age at kidding, month of kidding  
484 and dry period length depending on the lactation stage, herd-test day effect instead of herd-year effect,  
485 and separation of parities). The main difference between LM and RRM was on the mean of EBVs or  
486 DYDs for yields in both breeds and for FPR in Alpine.

### 487 ***Validation of the different genetic evaluations***

488 For total production throughout the lactation, with A/H and LM/RRM, for yields and SCS in  
489 both breeds, we overestimated EBVs, with a variance of EBVs of candidate bucks smaller than DYD,  
490 with an over-dispersion of EBVs and with a poor accuracy of EBVs (validation correlations <0.46).  
491 More heritable traits, as contents compared to yields and SCS (Arnal et al., 2019b) had better  
492 correlations, slopes and small bias. These results are consistent with those obtained by Kang et al.  
493 (2017) in their simulation study using a RRM with A and H matrix.

494 For the comparisons of LM with RRM or the comparison of the use or not of genomic  
495 information, no general conclusion could be made: results were largely depending to the trait and to the  
496 breed. This was probably due to our small number of candidate bucks.

497 **LM/RRM comparison.** The bias with RRM was reduced compared to LM for yields in both  
498 breeds and SCS in Saanen. The lower bias with RRM was probably a consequence of a better adjustment  
499 of the model to the data as seen previously (Schaeffer and Jamrozik, 2008; Arnal et al., 2019a). EBVs  
500 variance of candidate bucks was closer to the reference (DYD\_RRM) with LM than RRM for yields in  
501 Alpine, SCS in both breeds but EBVs variance of candidate bucks was closer to the reference  
502 (DYD\_RRM) with RRM than LM for PC in Alpine. However, it is a property of BLUP that variances  
503 of EBVs' candidate bucks estimated with partial dataset due to small information is reduced compared  
504 to the DYD calculated with complete dataset.

505 Correlations between DYD and EBVs from LM were higher than with EBVs from RRM for  
506 candidate bucks for SCS in both breeds, FC and FPR in Saanen and for FY when using genomic  
507 information in Saanen. However, RRM showed higher correlations with DYD for candidate bucks for  
508 FY and MY with genomic information in Alpine. Slope of the regression of DYD on the EBVs of  
509 candidate bucks was better (closest to 1) with RRM with genomic information for MY in Alpine. For  
510 slopes and correlations, the differences between LM and RRM were not significant. For yields and SCS  
511 in both breeds contrary to contents, slopes were inferior to 1 so EBVs were over-dispersed as in the  
512 majority of genomic studies (Legarra and Reverter, 2018). In dairy cow, Oliveira et al. (2019b) found,  
513 as in our study, that RRM do not improve the validation correlations or the slope compared to LM using  
514 A or H.

515 LM does not model the shape of the lactation curve but relies on a phenotype that takes it into  
516 account, giving this model an accuracy close to that of the RRM for lactation production traits. This was  
517 certainly one of the reasons why LM and RRM are close in terms of correlation and slope.

518 **A/H comparison.** When we compared the results with or without genomic information, means  
519 of EBVs of candidate bucks were closer to the means of DYD when using genomic information with  
520 LM for FY and PY in Saanen reducing the bias. With RRM, using genomic information did not improve  
521 bias. Some bias could come from the use of unknown parent groups as explained in Legarra and Reverter  
522 (2018). Variances of EBVs of candidate bucks were closer to the DYD variances when using genomic

523 information for FC in both breeds and for FPR in Alpine using RRM and LM, and using RRM for PY  
524 in Saanen, FY and SCS in Alpine. For the persistency, variances of EBVs were higher with genomic  
525 information for FC, SCS and FPR in Alpine. This inflation of the variance must be related to the use of  
526 H matrix in the models. Indeed, H takes also into account the Mendelian sampling term, i.e. two full-  
527 sibs that have the same EBVs based on A would have two different with H.

528         Correlations between EBVs of genotyped bucks and DYD were higher using genomic  
529 information for all traits in Saanen using LM and RRM and for PC in Alpine with RRM. For the  
530 persistency, the correlations were higher for MY and PC in Saanen with genomic information for  
531 persistency. The slope was better (closest to 1) with genomic information for PY using LM and for FC  
532 in Alpine using RRM and LM. For the persistency, the slope was better for SCS in Alpine without  
533 genomic for persistency. The slope was better without genomic information with LM for SCS in Alpine.  
534 Several studies (Koivula et al., 2015; Mucha et al., 2015; Baba et al., 2017; Kang et al., 2017, 2018;  
535 Oliveira et al., 2019b) which compared A\_RRM with H\_RRM found an increase of validation  
536 correlations and a slope closer to 1, as in our study for the majority of traits. This was also due to H  
537 taking into account the Mendelian sampling term. Oliveira et al. (2019b) obtained similar results when  
538 they compare A\_LM with H\_LM.

539         In our study, the reference population was very small, which limited the performance of genomic  
540 evaluation. With more animals genotyped, correlations and slope should be improved. However, the  
541 reference population is of very good quality because artificial insemination bucks constitute the large  
542 part of this reference population. Indeed, these bucks have been genotyped since at least 1998, so they  
543 have very accurate genetic values and have been derived from goats of the entire population, whatever  
544 the breeding system. We could not have a larger population of candidate bucks because we needed  
545 enough bucks in the reference population as well as candidate bucks newborn at least 5 years before the  
546 end of our performance dataset for these males to have multiparous offsprings.

547         The study of the correlations between models showed that with our population, the gain from a  
548 LM\_H to a RRM\_H would be smaller than the one from a LM\_A to a LM\_H.

#### 549         *Analysis of SNP effects*

550         The percentages of variance were calculated with 20 adjacent SNPs, whereas the p-value was  
551 considered with only the specific SNP and not the adjacent ones was taken into account. The difference  
552 observed on results with these two criteria can explained the different genomic regions found. A  
553 difference was particularly important for the region of DGAT1 on chromosome 14 that was not found  
554 with the percentages of variance.

555         The genomic regions highlighted by the  $-\log_{10}(\text{p-value})$  with LM were the same as those found  
556 in other French studies (Martin et al., 2017, 2018) on the same breeds but using other methodologies.  
557 Same regions were found to have an impact on FC in Alpine on chromosome 8 (Martin et al., 2017), on

558 chromosome 6 for FC/PC in both breeds in the region of caseins (Martin et al., 2017); on chromosome  
559 19 for MY, FY and PY in Saanen (Martin et al., 2018), on chromosome 14 in the region of DGAT1 for  
560 FC in both breeds. Between our study and the one of Martin et al. (2017), the map of the SNPs on  
561 chromosome 14 was inverted, it is why they found the QTL at the beginning of the chromosome and us  
562 at the end. The region around 26.1 Mb on chromosome 19 was also found by Mucha et al. (2018) for  
563 milk yield in crossbred goats (Saanen-Toggenburg-Alpine)

564           However, we did not find for SCS the same region on chromosome 19. Indeed, Martin et al.  
565 (2018) found it around 41 Mb whereas we found it around 26 Mb as for yields traits.

566           We found regions for FPR on chromosome 14 at the same position as FC and with a higher p-  
567 value than FC for Saanen, and on chromosome 8 in Alpine as for FC. This trait was not studied before  
568 in the literature. It was interesting to note that the genomic region highlighted on chromosome 6 and  
569 common for PC and FC, was not associated with the ratio FPR.

570           The other more anecdotic regions on chromosome 2, 3 and 23 were not found in the goat  
571 genomic literature.

572           ***LM/RRM comparison.*** Same regions were found for LM and LEV with RRM for all the traits  
573 in both breeds except for more anecdotic regions. These results confirm the equivalence between the  
574 two models. The p-values, as well as the percentages of variance were equivalent between both models.  
575 Differences between primiparous and multiparous were observed but it was known that EBVs in  
576 primiparous and in multiparous are not exactly the same trait: there is a genetic correlation between  
577 them of 0.7 (Arnal et al., 2019b).

578           ***Persistency and genomic*** Three regions were found associated with persistency when using p-  
579 values: for FY for primiparous in Saanen on chromosome 19, PY for multiparous in Saanen on  
580 chromosome 19 and for SCS for primiparous on chromosome 23 in Alpine. With the percentages of  
581 variance, only the persistency of FY for primiparous in Saanen on chromosome 19 were found again.  
582 Persistency of MY and PY in Saanen for primiparous were found associated in the same region as FY  
583 persistency in primiparous. A SNP of this region, for the two traits, had also a p-value close to 6. Other  
584 regions associated with FC and FPR persistency on chromosome 6 for primiparous and multiparous in  
585 Saanen and with SCS for primiparous and multiparous persistency in Saanen were also found.

586           Regions associated with persistency were often the same as the regions associated with LEV:  
587 MY, FY, PY, FC, SCS in primiparous; but it was not the case for SCS in multiparous, and for FPR. On  
588 chromosome 6, no region was found for LEV but one for PERS was detected for FPR in Saanen.  
589 Cardona et al. (2016) found, as in our study, in Creole goats, that there are region associated with caseins  
590 on chromosome 6 that are also associated with FY persistency. We did not observe a region associated  
591 with FY persistency on chromosome 6 but for FC and FPR we found one and they are fat traits as FY.

592 In dairy cows, Pryce et al. (2010), Strucken et al. (2012), and Koelbehdari et al. (2009) also  
593 found regions associated with persistency. They did not use persistency decorrelated to LEV as in the  
594 present study but did by classes of DIM throughout the lactation. In dairy cows, Macciotta et al. (2015)  
595 used as in our study the second eigenvector to perform a GWAS but they did it at a phenotypic level.  
596 They found no region associated with the second eigenvector. It could be because they did not evaluate  
597 the persistency at a genetic level.

598 In dairy species, other studies looked at the effect of genotypes on the lactation curve shape  
599 (Pauciullo et al., 2012; Szyda et al., 2014) using the different allele of genes to explain the shape of the  
600 lactation curve. They found phenotypic differences on lactation curve shape according to the genotypes.

601 The analysis of the effects of SNPs from SSGBLUP for LEV and PERS shows that there are  
602 SNPs more associated than other are with some traits that they would be useful to weight in a genomic  
603 evaluation model. A ssWeightedGBLUP random regression model, as proposed in Teissier et al.  
604 (2019), could improve accuracy of EBVs of young candidates at birth.

## 605 **CONCLUSIONS**

606 A random regression model using Single-step genomic evaluation in dairy goat in France was  
607 implemented in the present study. The presented study was based on the French Alpine and Saanen  
608 goats, the two main French breeds, which represents 96% of goats in France. We showed that a genomic  
609 RRM evaluation is possible on the total number of goats measured regularly throughout their lactation  
610 nowadays. A validation evaluation have compared the EBVs of candidate bucks obtained at birth using  
611 LM or RRM models with or without genomic information with the DYD obtained later based on progeny  
612 performances. The genomic information increased correlations between DYD and EBVs for candidate  
613 bucks in Saanen breed. The lactation model and the RRM had quite the same performances for the slopes  
614 of the regressions of the DYD on EBVs of the candidate bucks and for correlations between DYD and  
615 EBVs of the candidates. However, the ssGBLUP RRM reduced the bias. This study shows that a Single-  
616 step RRM is feasible in dairy goat in France and that its performance is comparable to those of LM but  
617 with a genomic evaluation of persistency. The analysis of SNP effects highlight the same genomic  
618 regions founded with LM or RRM. Interested genomic regions associated with persistency were also  
619 found.

## 620 **ACKNOWLEDGEMENTS:**

621 We thank Ignacy Misztal (University of Georgia, Georgia, USA) and Andres Legarra (INRA,  
622 UMR GenPhySE, France) for access to executable files of blupiod2, M.Teissier (INRA, UMR  
623 GenPhySE, France) for the filtering of genomic data and A. Legarra (INRA, UMR GenPhySE, France)  
624 for constructive and valuable discussions.

## 625 **REFERENCES:**

626           Aguilar, I., A. Legarra, F. Cardoso, Y. Masuda, D. Lourenco, and I. Misztal. 2019. Frequentist  
627 p-values for large-scale-single step genome-wide association, with an application to birth weight in  
628 American Angus cattle. *Genet. Sel. Evol.* 51:28.

629           Aguilar, I., I. Misztal, D.L. Johnson, A. Legarra, S. Tsuruta, and T.J. Lawlor. 2010. Hot topic:  
630 A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation  
631 of Holstein final score. *J. Dairy Sci.* 93:743–752.

632           Aguilar, I., I. Misztal, S. Tsuruta, A. Legarra, and H. Wang. 2014. PREGSF90–POSTGSF90:  
633 computational tools for the implementation of Single-step genomic selection and genome-wide  
634 association with ungenotyped individuals in BLUPF90 programs. Page in Proceedings of the 10th world  
635 congress of genetics applied to livestock production.

636           Arnal, M., H. Larroque, H. Leclerc, V. Ducrocq, and C. Robert-Granié. 2019a. Genetic  
637 parameters for first lactation dairy traits in the Alpine and Saanen goat breeds using a random regression  
638 test-day model. *Genet. Sel. Evol.* 51:43.

639           Arnal, M., H. Larroque, H. Leclerc, V. Ducrocq, and C. Robert-Granié. 2019b. Estimation of  
640 genetic parameters for dairy traits and somatic cell score in the first three parities using a random  
641 regression test-day model in French Alpine goats (submitted).

642           Baba, T., Y. Gotoh, S. Yamaguchi, S. Nakagawa, H. Abe, Y. Masuda, and T. Kawahara. 2017.  
643 Application of Single-step genomic best linear unbiased prediction with a multiple-lactation random  
644 regression test-day model for Japanese Holsteins. *Anim. Sci. J.* 88:1226–1231.

645           Boichard, D., V. Ducrocq, P. Croiseau, and S. Fritz. 2016. Genomic selection in domestic  
646 animals: Principles, application and perspectives. *C R Biol* 339:274–277.

647           Cardona, S.J.C., H.C. Cadavid, J.D. Corrales, S. Munilla, R.J. Cantet, and A. Rogberg-Muñoz.  
648 2016. Longitudinal data analysis of polymorphisms in the  $\kappa$ -casein and  $\beta$ -lactoglobulin genes shows  
649 differential effects along the trajectory of the lactation curve in tropical dairy goats. *J. Dairy Sci.*  
650 99:7299–7307.

651           Ducrocq, V. 1998. Genokit, BLUP software; version October 30, 2017. INRA SGQA Jouy-En-  
652 Josas Fr.

653           Kang, H., C. Ning, L. Zhou, S. Zhang, Q. Yan, and J.-F. Liu. 2018. Single-step genomic  
654 evaluation of milk production traits using multiple-trait random regression model in Chinese Holsteins.  
655 *J. Dairy Sci.* 101:11143–11149.

656           Kang, H., L. Zhou, R. Mrode, Q. Zhang, and J.F. Liu. 2017. Incorporating the Single-step  
657 strategy into a random regression model to enhance genomic prediction of longitudinal traits. *Heredity*  
658 119:459.

659 Koivula, M., I. Strandén, J. Pösö, G.P. Aamand, and E.A. Mäntysaari. 2015. Single-step  
660 genomic evaluation using multitrait random regression model and test-day data. *J. Dairy Sci.* 98:2775–  
661 2784.

662 Kolbehdari, D., Z. Wang, J.R. Grant, B. Murdoch, A. Prasad, Z. Xiu, E. Marques, P. Stothard,  
663 and S.S. Moore. 2009. A whole genome scan to map QTL for milk production traits and somatic cell  
664 score in Canadian Holstein bulls. *J. Anim. Breed. Genet.* 126:216–227.

665 Larroque, H., J.-M. Astruc, A. Barbat, F. Barillet, D. Boichard, B. Bonaiti, V. Clément, I. David,  
666 G. Lagriffoul, I. Palhiere, A. Piacère, C. Robert-Granie, and R. Rupp. 2011. National Genetic  
667 Evaluations in Dairy Sheep and Goats in France. Wageningen Academic, Stavanger, Norway.

668 Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and  
669 genomic information. *J. Dairy Sci.* 92:4656–4663.

670 Legarra, A., and A. Reverter. 2018. Semi-parametric estimates of population accuracy and bias  
671 of predictions of breeding values and future phenotypes using the LR method. *Genet. Sel. Evol.* 50:53.

672 Macciotta, N.P.P., G. Gaspa, L. Bomba, D. Vicario, C. Dimauro, M. Cellesi, and P. Ajmone-  
673 Marsan. 2015. Genome-wide association analysis in Italian Simmental cows for lactation curve traits  
674 using a low-density (7K) SNP panel. *J. Dairy Sci.* 98:8175–8185.

675 Martin, P., I. Palhière, C. Maroteau, P. Bardou, K. Canale-Tabet, J. Sarry, F. Woloszyn, J.  
676 Bertrand-Michel, I. Racke, and H. Besir. 2017. A genome scan for milk production traits in dairy goats  
677 reveals two new mutations in *Dgat1* reducing milk fat content. *Sci. Rep.* 7:1872.

678 Martin, P., I. Palhière, C. Maroteau, V. Clément, I. David, G.T. Klopp, and R. Rupp. 2018.  
679 Genome-wide association mapping for type and mammary health traits in French dairy goats identifies  
680 a pleiotropic region on chromosome 19 in the Saanen breed. *J. Dairy Sci.* 101:5214–5226.

681 Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, and D.H. Lee. 2002. BLUPF90 and  
682 related programs (BGF90). Pages 743–744 in *Proceedings of the 7th world congress on genetics applied  
683 to livestock production.*

684 Mucha, S., R. Mrode, M. Coffey, M. Kizilaslan, S. Desire, and J. Conington. 2018. Genome-  
685 wide association study of conformation and milk yield in mixed-breed dairy goats. *J. Dairy Sci.*  
686 101:2213–2225.

687 Mucha, S., R. Mrode, I. MacLaren-Lee, M. Coffey, and J. Conington. 2015. Estimation of  
688 genomic breeding values for milk yield in UK dairy goats. *J. Dairy Sci.* 98:8201–8208.

689 Oliveira, H.R., L.F. Brito, D.A.L. Lourenco, F.F. Silva, J. Jamrozik, L.R. Schaeffer, and F.S.  
690 Schenkel. 2019a. Invited review: Advances and applications of random regression models: From  
691 quantitative genetics to genomics. *J. Dairy Sci.*

692 Oliveira, H.R., L.F. Brito, F.F. Silva, D.A.L. Lourenco, J. Jamrozik, and F.S. Schenkel. 2019b.  
693 Genomic prediction of lactation curves for milk, fat, protein, and somatic cell score in Holstein cattle.  
694 *J. Dairy Sci.* 102:452–463.

695 Oliveira, H.R., D.A.L. Lourenco, Y. Masuda, I. Misztal, S. Tsuruta, J. Jamrozik, L.F. Brito, F.F.  
696 Silva, J.P. Cant, and F.S. Schenkel. 2019c. Single-step genome-wide association for longitudinal traits  
697 of Canadian Ayrshire, Holstein, and Jersey dairy cattle. *J. Dairy Sci.*

698 Oliveira, H.R., D.A.L. Lourenco, Y. Masuda, I. Misztal, S. Tsuruta, J. Jamrozik, L.F. Brito, F.F.  
699 Silva, and F.S. Schenkel. 2019d. Application of Single-step genomic evaluation using multiple-trait  
700 random regression test-day models in dairy cattle. *J. Dairy Sci.* 102:2365–2377.

701 Paucullo, A., G. Cosenza, R. Steri, A. Coletta, L. Jemma, M. Feligini, D. Di Bernardino, N.P.  
702 Macciotta, and L. Ramunno. 2012. An association analysis between OXT genotype and milk yield and  
703 flow in Italian Mediterranean river buffalo. *J. Dairy Res.* 79:150–156.

704 Pryce, J.E., M. Haile-Mariam, K. Verbyla, P.J. Bowman, M.E. Goddard, and B.J. Hayes. 2010.  
705 Genetic markers for lactation persistency in primiparous Australian dairy cows. *J. Dairy Sci.* 93:2202–  
706 2214.

707 Sargent, F.D., V.H. Lytton, and O.G. Wall. 1968. Test interval method of calculating dairy herd  
708 improvement association records. *J. Dairy Sci.* 51:170–179.

709 Schaeffer, L.R., and J. Jamrozik. 2008. Random regression models: a longitudinal perspective.  
710 *J. Anim. Breed. Genet.* 125:145–146.

711 Steiger, J.H. 1980. Tests for comparing elements of a correlation matrix. *Psychol. Bull.* 87:247–  
712 251.

713 Strucken, E.M., R.H. Bortfeldt, D.J. De Koning, and G.A. Brockmann. 2012. Genome-wide  
714 associations for investigating time-dependent genetic effects for milk production traits in dairy cattle.  
715 *Anim. Genet.* 43:375–382.

716 Szyda, J., J. Komisarek, and I. Antkowiak. 2014. Modelling effects of candidate genes on  
717 complex traits as variables over time. *Anim. Genet.* 45:322–328.

718 Täubert, H., Z. Liu, J. Tarrès, and V. Ducrocq. 2010. An approach to compute EDC and DYD  
719 for test-day models. Page in Proceedings of the 9th WCGALP. 2010; 9. World Congress on Genetics  
720 Applied to Livestock Production, Leipzig, DEU, 2010-08-01-2010-08-06, 1-4.

721 Teissier, M., H. Larroque, and C. Robert-Granie. 2019. Accuracy of genomic evaluation with  
722 weighted Single-step genomic best linear unbiased prediction for milk production traits, udder type  
723 traits, and somatic cell scores in French dairy goats. *J. Dairy Sci.* 102:3142–3154.

724 Tossler-Klopp, G., P. Bardou, O. Bouchez, C. Cabau, R. Crooijmans, Y. Dong, C. Donnadiou-  
725 Tonon, A. Eggen, H.C. Heuven, and S. Jamli. 2014. Design and characterization of a 52K SNP chip for  
726 goats. *PloS One* 9:e86227.

727 VanRaden, P.M., C.P. Van Tassell, G.R. Wiggans, T.S. Sonstegard, R.D. Schnabel, J.F. Taylor,  
728 and F.S. Schenkel. 2009. Invited review: Reliability of genomic predictions for North American  
729 Holstein bulls. *J. Dairy Sci.* 92:16–24.

730 VanRaden, P.M., and G.R. Wiggans. 1991. Derivation, calculation, and use of national animal  
731 model information. *J. Dairy Sci.* 74:2737–2746.

732 Wang, H., I. Misztal, I. Aguilar, A. Legarra, R.L. Fernando, Z. Vitezica, R. Okimoto, T. Wing,  
733 R. Hawken, and W.M. Muir. 2014. Genome-wide association mapping including phenotypes from  
734 relatives without genotypes in a Single-step (ssGWAS) for 6-week body weight in broiler chickens.  
735 *Front. Genet.* 5:134.

736 Wiggans, G.R., and G.E. Shook. 1987. A lactation measure of somatic cell count. *J. Dairy Sci.*  
737 70:2666–2672.

738

739

## TABLES

740 Table 1: TD and lactation number by parity in Saanen and in Alpine for the different datasets

		Saanen		Alpine	
		TD	Lactation	TD	Lactation
<b>Complete dataset (2000-2017)</b>	<b>1<sup>st</sup> parity</b>	3,466,430	570,641	4,532,245	734,790
	<b>2<sup>nd</sup> parity</b>	2,628,369	431,407	3,495,849	563,710
	<b>3<sup>rd</sup> parity</b>	1,809,447	304,656	2,443,457	402,729
	<b>Total</b>	7,908,192	1,306,704	10,483,205	1,698,032
<b>Partial dataset (2000-2011)</b>	<b>1<sup>st</sup> parity</b>	2,566,171	413,817	3,265,177	515,612
	<b>2<sup>nd</sup> parity</b>	1,942,222	311,255	2,498,949	390,772
	<b>3<sup>rd</sup> parity</b>	1,346,841	221,250	1,748,618	279,398
	<b>Total</b>	5,855,234	946,322	7,512,744	1,185,782
<b>Dataset for SNP effects analysis (2000-2017)</b>	<b>1<sup>st</sup> parity</b>	642,673	96,628	1,049,090	156,794
	<b>2<sup>nd</sup> parity</b>	504,037	75,573	837,620	125,011
	<b>3<sup>rd</sup> parity</b>	338,551	51,423	578,530	87,591
	<b>Total</b>	1,485,261	223,624	2,465,240	369,396

741

742

743 Table 2: Number of reference and candidate bucks genotyped by year of birth and number of individuals  
 744 in the pedigree file in Alpine and Saanen

		<b>Saanen</b>		<b>Alpine</b>			
<b>Pedigree</b>		818,702		1,021,665			
<b>Females</b>		812		1370			
		<b>Year of birth</b>	<b>Number of bucks by year of birth</b>	<b>Total</b>	<b>Number of bucks by year of birth</b>	<b>Total</b>	
<b>Genotyped animals</b>	<b>Reference bucks</b>	<b>1993</b>	0		1		
		<b>1996</b>	0		1		
		<b>1997</b>	0		2		
		<b>1998</b>	19		19		
		<b>1999</b>	21		22		
		<b>2000</b>	18		18		
		<b>2001</b>	25		23		
		<b>2002</b>	28	297	27	383	
		<b>2003</b>	29		36		
		<b>2004</b>	25		38		
		<b>2005</b>	29		40		
		<b>2006</b>	25		40		
		<b>2007</b>	29		40		
	<b>2008</b>	21		35			
	<b>2009</b>	28		41			
		<b>candidate bucks</b>	<b>2010</b>	30		43	
			<b>2011</b>	32		43	
	<b>2012</b>		35	133	43	172	
	<b>2013</b>		36		43		

745

746 Table 3: Number of multiparous progeny and reliabilities with the complete dataset for the candidate  
 747 bucks

			<b>Minimum</b>	<b>Median</b>	<b>Maximum</b>
<b>Number of multiparous progeny</b>	<b>Saanen</b>		6	40	152
	<b>Alpine</b>		6	61	179
<b>Reliability</b>	<b>A_LM</b>	<b>Saanen</b>	0.67	0.88	0.98
		<b>Alpine</b>	0.78	0.91	0.98
	<b>A_LEV</b>	<b>Saanen</b>	0.65	0.86	0.96
		<b>Alpine</b>	0.71	0.89	0.96
	<b>A_PERS</b>	<b>Saanen</b>	0.55	0.78	0.91
		<b>Alpine</b>	0.62	0.86	0.94

748 *A\_LM: model based on lactation phenotypes with the complete dataset without genomic, A\_LEV:*  
 749 *regression coefficient for the first eigenfunction with the complete dataset without genomic (level of milk*  
 750 *throughout the lactation), A\_PERS: regression coefficient for the second eigenfunction with the*  
 751 *complete dataset without genomic (persistency)*

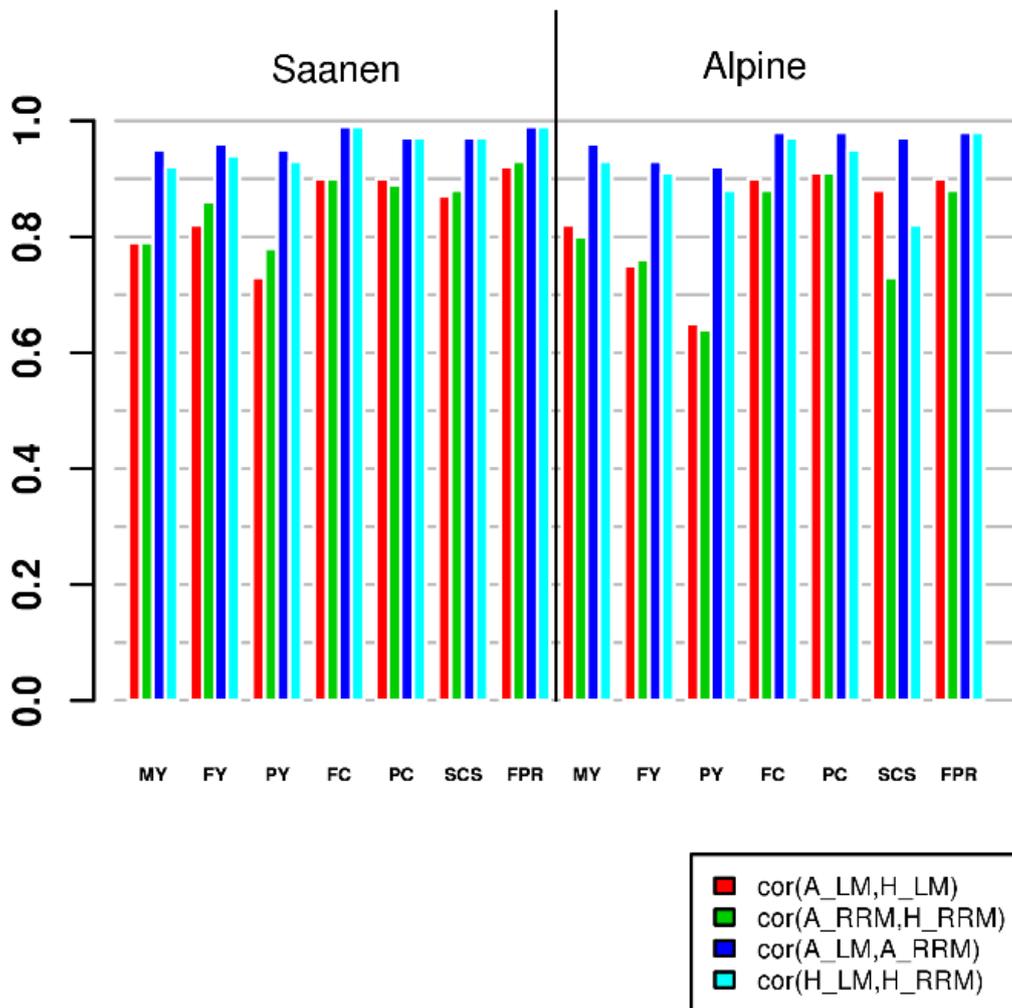
752

753

754 Table 4: Regions with a  $-\log_{10}(\text{p-value})$  superior to 6 detected by trait and model studied and that are maximal on the chromosome in Alpine and Saanen breeds

<b>Chromosome</b>	<b>Breed</b>	<b>Trait (Model)</b>	<b>Position</b>
2	Alpine	PC (LM)	130,868,973
3	Alpine	SCS (LM)	5,784,974
6	Alpine	FC (LM, LEV_P, LEV_M)	85,861,248
6	Both	PC (LM, LEV_P, LEV_M)	85,979,715
6	Saanen	FC (LM, LEV_P, LEV_M)	85,979,715
8	Alpine	FC (LM, LEV_P, LEV_M), FPR (LM, LEV_P, LEV_M)	23,810,468
14	Saanen	FY (LM)	81,112,809
14	Both	FC (LM, LEV_P, LEV_M), FPR (LM, LEV_P, LEV_M)	81,658,443
19	Saanen	FY (LEV_M)	25,147,677
19	Saanen	MY (LM, LEV_P, LEV_M), FY (LM, LEV_P), PY (LM, LEV_P, LEV_M)	26,192,128
19	Saanen	FY (PERS_P), SCS (LM, LEV_P, LEV_M)	26,610,610
19	Saanen	PY (PERS_M)	37,634,643
23	Alpine	SCS (LEV_M)	24,811,860
23	Alpine	SCS (PERS_P)	27,664,633

755 *LM*: model based on lactation phenotypes, *LEV\_P*: regression coefficient for the first eigenfunction for primiparous (level of milk throughout the lactation),  
756 *PERS\_P*: regression coefficient for the second eigenfunction for primiparous (persistence), *LEV\_M*: regression coefficient for the first eigenfunction for  
757 multiparous (level of milk throughout the lactation), *PERS\_M*: regression coefficient for the second eigenfunction for multiparous (persistence)

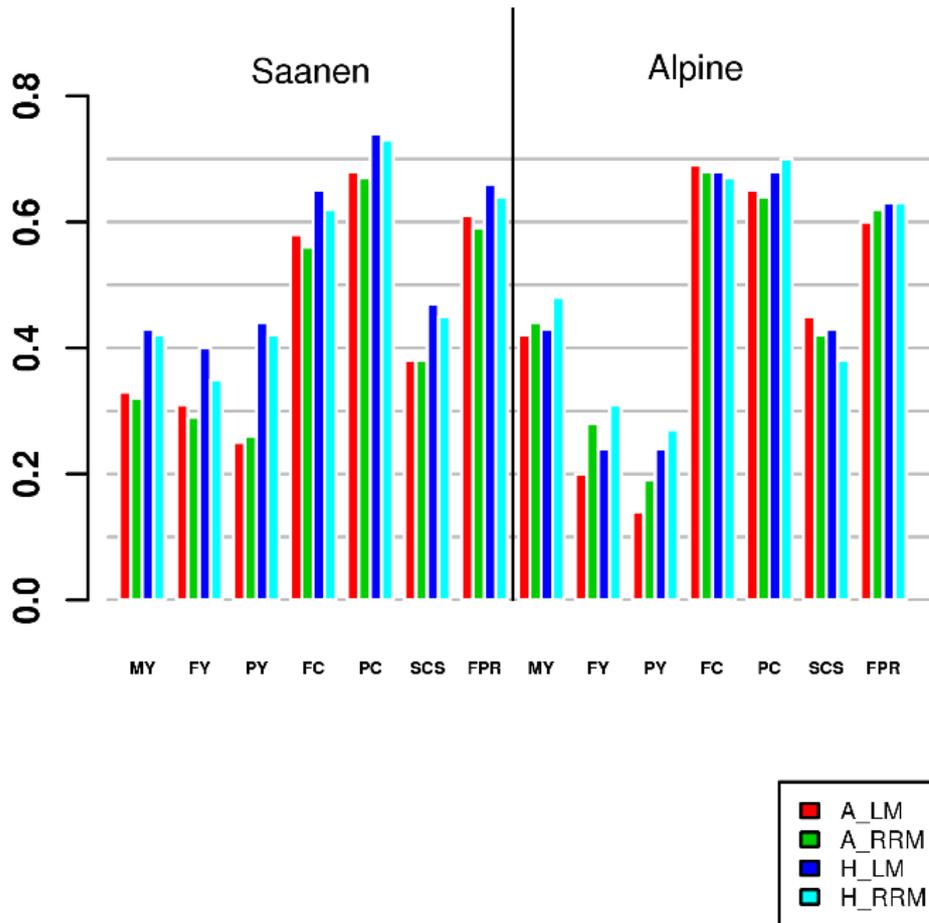


759

760 Figure 1: Correlations weighted by the number of progeny between the EBVs from the different models  
 761 of the candidate bucks calculated with the partial dataset for the different traits in Saanen and Alpine

762 (*MY: milk, FY: Fat yield, PY: protein yield, FC: fat content, PC: protein content, SCS: somatic cell*  
 763 *score, FPR: fat protein ratio*)

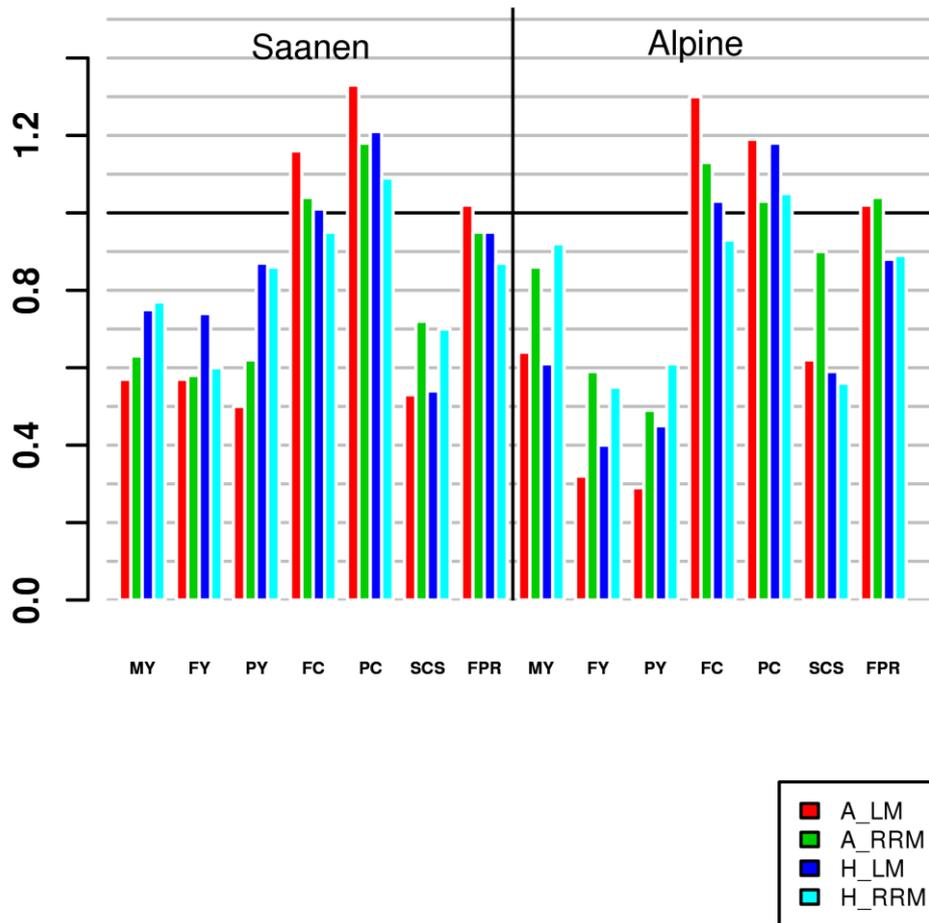
764 *A\_LM: lactation model without genomic information for the relationship matrix building, H\_LM:*  
 765 *lactation model with genomic information for the relationship matrix building, A\_RRM: random*  
 766 *regression model without genomic information for the relationship matrix building, H\_RRM: random*  
 767 *regression model with genomic information for the relationship matrix building*



768

769 Figure 2: Correlations between the candidate bucks EBVs from the partial dataset from 4 models studied  
 770 and their DYD\_RRM for the different traits in Saanen and Alpine weighted by the number of progeny  
 771 (MY: milk, FY: Fat yield, PY: protein yield, FC: fat content, PC: protein content, SCS: somatic cell  
 772 score, FPR: fat protein ratio)

773 *A\_LM*: lactation model without genomic information for the relationship matrix building, *H\_LM*:  
 774 lactation model with genomic information for the relationship matrix building, *A\_RRM*: random  
 775 regression model without genomic information for the relationship matrix building, *H\_RRM*: random  
 776 regression model with genomic information for the relationship matrix building

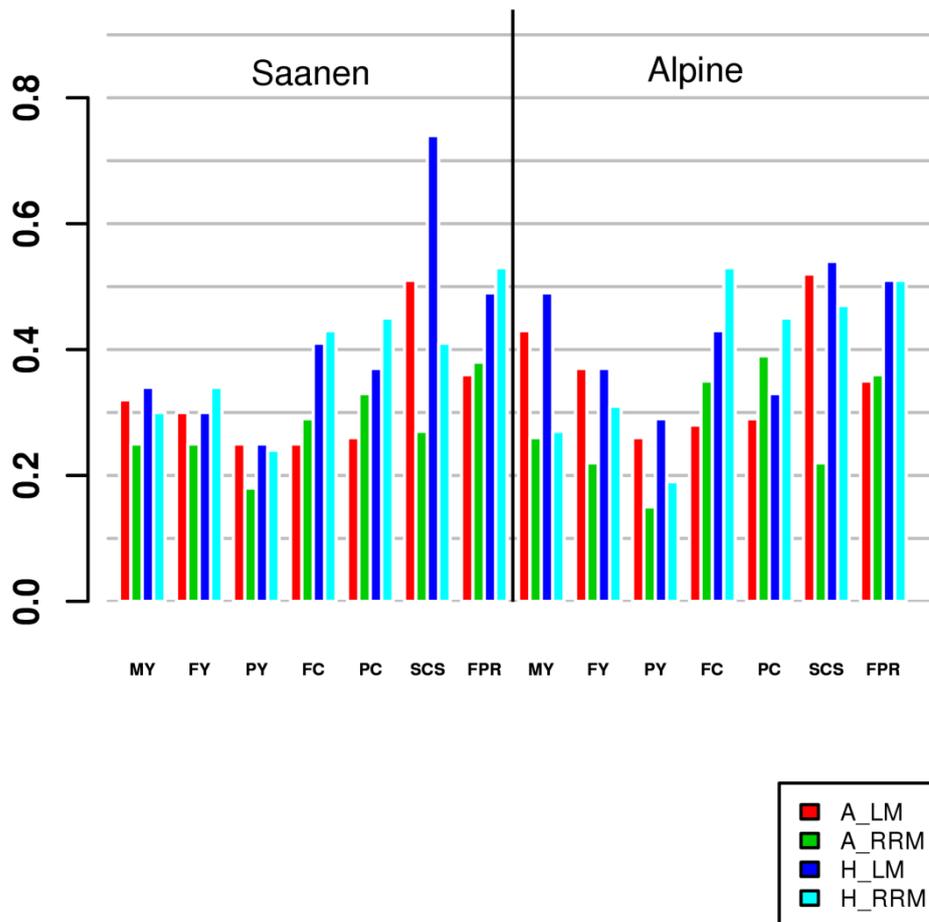


777

778 Figure 3: Regression slopes between the candidate bucks EBVs from the partial dataset from 4 models  
 779 studied and their DYD\_RRM for the different traits in Saanen and Alpine weighted by the number of  
 780 progeny

781 (*MY: milk, FY: Fat yield, PY: protein yield, FC: fat content, PC: protein content, SCS: somatic cell*  
 782 *score, FPR: fat protein ratio*)

783 *A\_LM: lactation model without genomic information for the relationship matrix building, H\_LM:*  
 784 *lactation model with genomic information for the relationship matrix building, A\_RRM: random*  
 785 *regression model without genomic information for the relationship matrix building, H\_RRM: random*  
 786 *regression model with genomic information for the relationship matrix building*

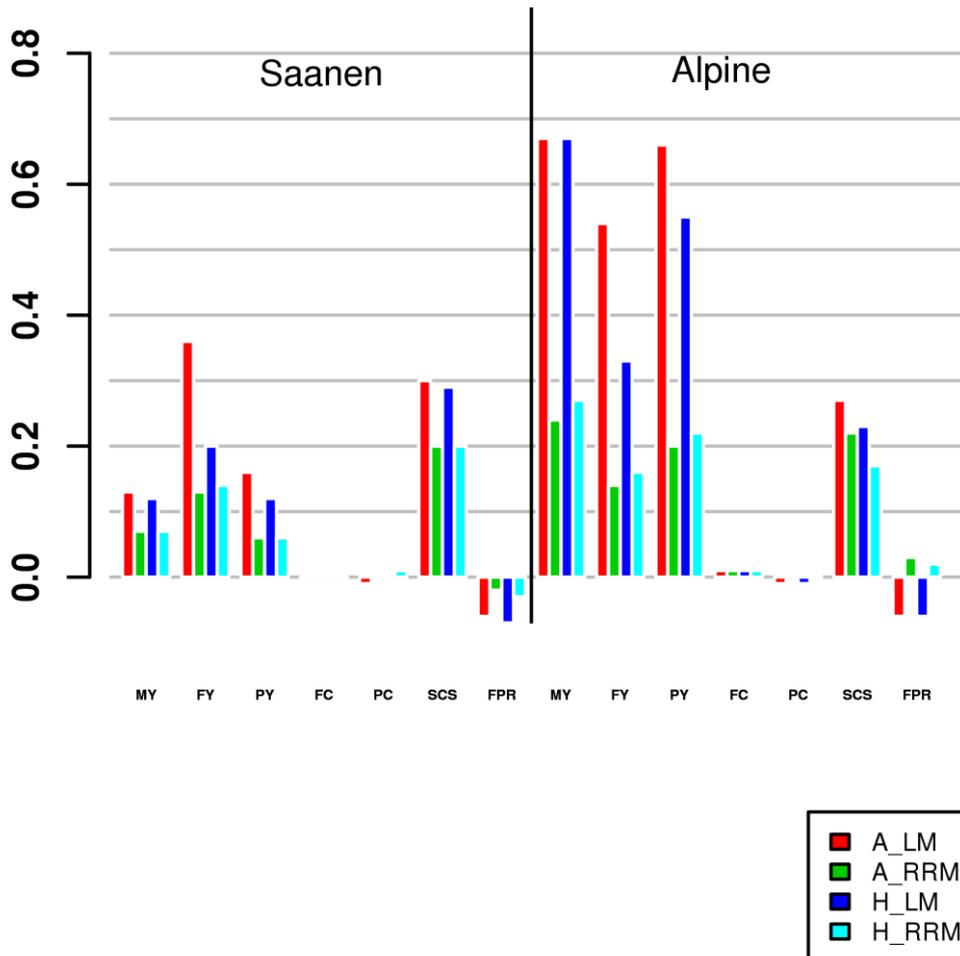


787

788 Figure 4: Variance ratios between the candidate bucks EBVs from the partial dataset from the different  
 789 models and their DYD\_RRM for the different traits in Saanen and Alpine weighted by the number of  
 790 progeny

791 *A\_LM*: lactation model without genomic information for the relationship matrix building, *H\_LM*:  
 792 lactation model with genomic information for the relationship matrix building, *A\_RRM*: random  
 793 regression model without genomic information for the relationship matrix building, *H\_RRM*: random  
 794 regression model with genomic information for the relationship matrix building

795

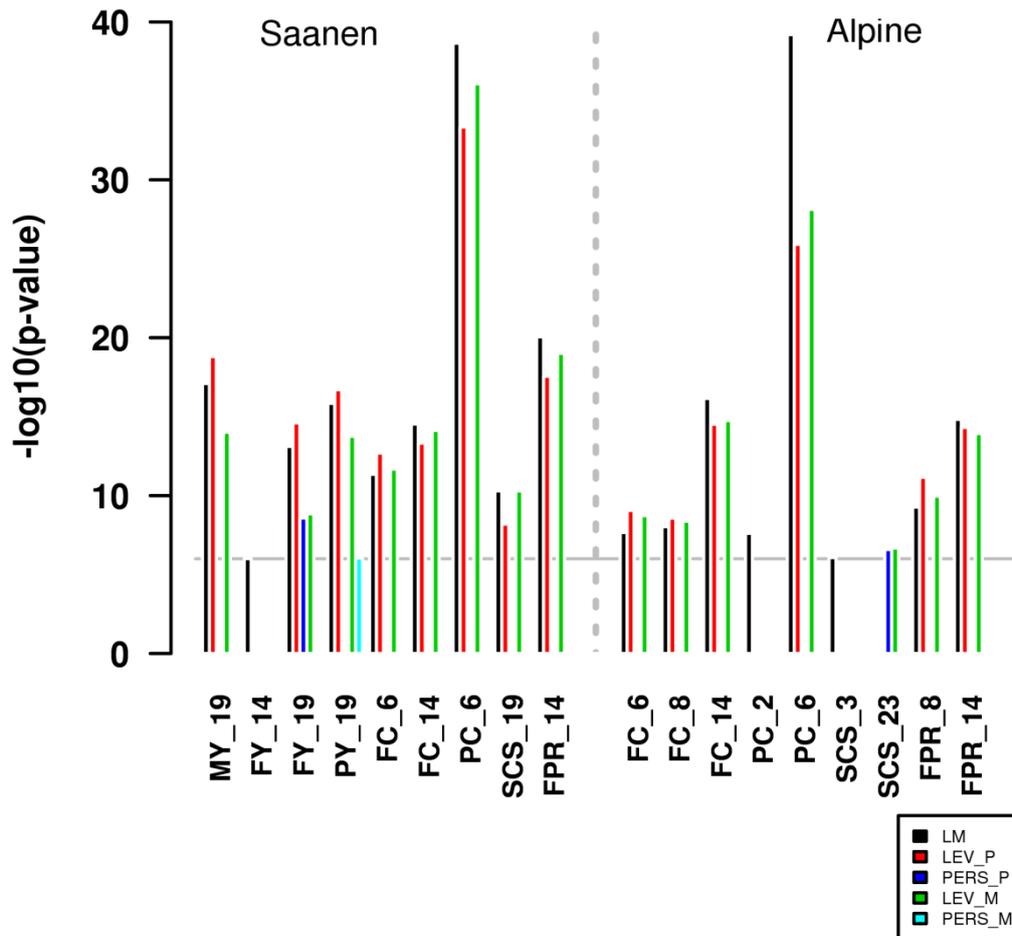


796

797 Figure 5: Bias calculated with DYD\_RRM for 4 models studied and traits in Saanen and Alpine for the  
 798 candidate bucks weighted by the number of progeny

799 (*MY: milk, FY: Fat yield, PY: protein yield, FC: fat content, PC: protein content, SCS: somatic cell*  
 800 *score, FPR: fat protein ratio*)

801 *A\_LM: lactation model without genomic information for the relationship matrix building, H\_LM:*  
 802 *lactation model with genomic information for the relationship matrix building, A\_RRM: random*  
 803 *regression model without genomic information for the relationship matrix building, H\_RRM: random*  
 804 *regression model with genomic information for the relationship matrix building*



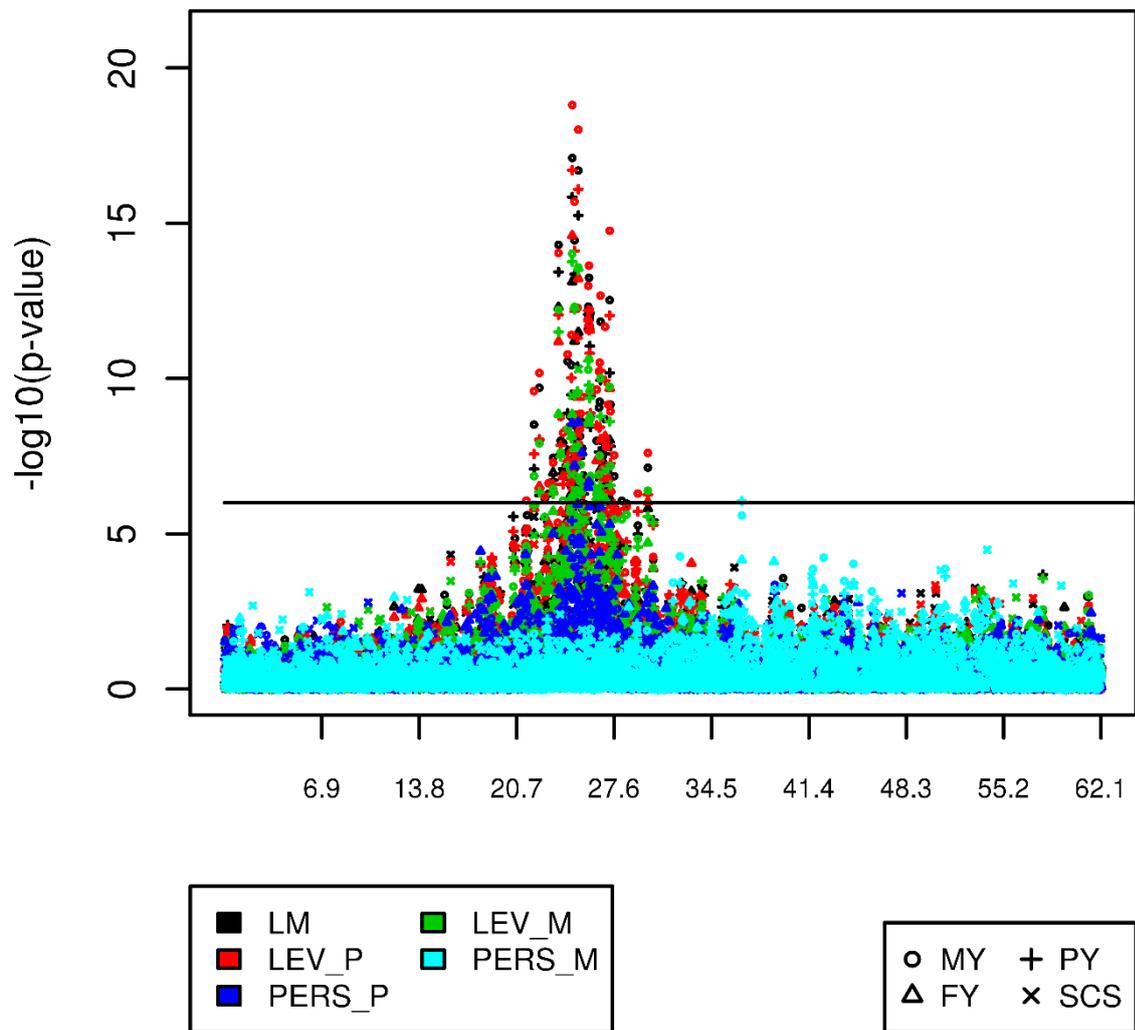
805

806 Figure 6: Maximal  $-\log_{10}(\text{p-value})$  of SNP by chromosome which were superior to 6 for the different  
 807 traits studied in Alpine and Saanen

808 *LM*: model based on lactation phenotypes, *LEV\_P*: regression coefficient for the first eigenfunction for  
 809 primiparous (level of milk throughout the lactation), *PERS\_P*: regression coefficient for the second  
 810 eigenfunction for primiparous (persistence), *LEV\_M*: regression coefficient for the first eigenfunction  
 811 for multiparous (level of milk throughout the lactation), *PERS\_M*: regression coefficient for the second  
 812 eigenfunction for multiparous (persistence)

813

814



815

816 Figure 7:  $-\log_{10}(\text{p-value})$  of SNP on chromosome 19 in Saanen for MY, FY, PY and SCS

817 *LM*: model based on lactation phenotypes, *LEV\_P*: regression coefficient for the first eigenfunction for  
 818 primiparous (level of milk throughout the lactation), *PERS\_P*: regression coefficient for the second  
 819 eigenfunction for primiparous (persistence), *LEV\_M*: regression coefficient for the first eigenfunction  
 820 for multiparous (level of milk throughout the lactation), *PERS\_M*: regression coefficient for the second  
 821 eigenfunction for multiparous (persistence)

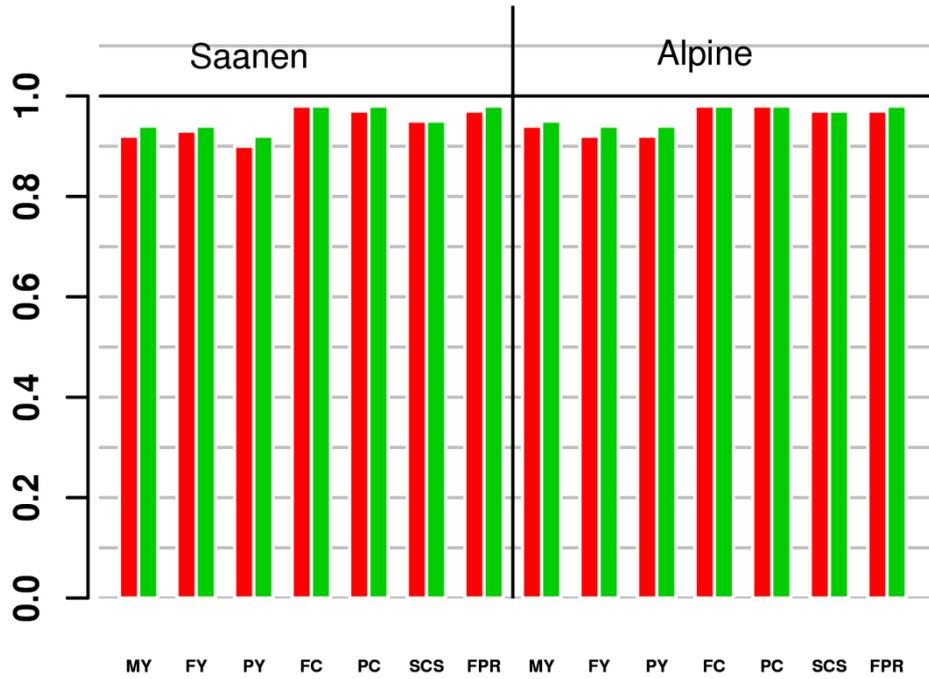
822 **SUPPLEMENTARY MATERIALS**

823 Supplemental Table S1: Positions of the maximal  $-\log_{10}(\text{p-value})$  superior to 6 by chromosome in Alpine and Saanen

Chromosome	Breed	Trait (Model)	Position of the 1 <sup>st</sup> SNP of the segment
6	Saanen	FC (PERS_M), FPR (PERS_M)	85,723,685
6	Alpine	MY( LEV_M), FC (LM, LEV_P, LEV_M), PC (LM, LEV_P, LEV_M)	85,723,685
6	Saanen	MY (LM, LEV_P, LEV_M), FC (LM, LEV_P, LEV_M), PC (LM, LEV_P, LEV_M), SCS (LM, LEV_P, PERS_P, PERS_M), FPR (PERS_P)	85,978,529
19	Saanen	FY (LM, LEV_M)	25,743,935
19	Saanen	MY(LM, LEV_P, LEV_M), FY (LEV_P, PERS_P), PY (LM, LEV_P, LEV_M), SCS (LM, LEV_P)	26,115,456
19	Saanen	MY (PERS_P), PY(PERS_P), SCS (LEV_M)	26,148,755

824 *LM: model based on lactation phenotypes, LEV\_P: regression coefficient for the first eigenfunction for primiparous (level of milk throughout the lactation),*  
 825 *PERS\_P: regression coefficient for the second eigenfunction for primiparous (persistency), LEV\_M: regression coefficient for the first eigenfunction for*  
 826 *multiparous (level of milk throughout the lactation), PERS\_M: regression coefficient for the second eigenfunction for multiparous (persistency), MY: milk yield,*  
 827 *FY: Fat yield, PY: Protein yield, FC: Fat content, PC: Protein content, SCS: Somatic cell score, FPR: Fat protein content*

828

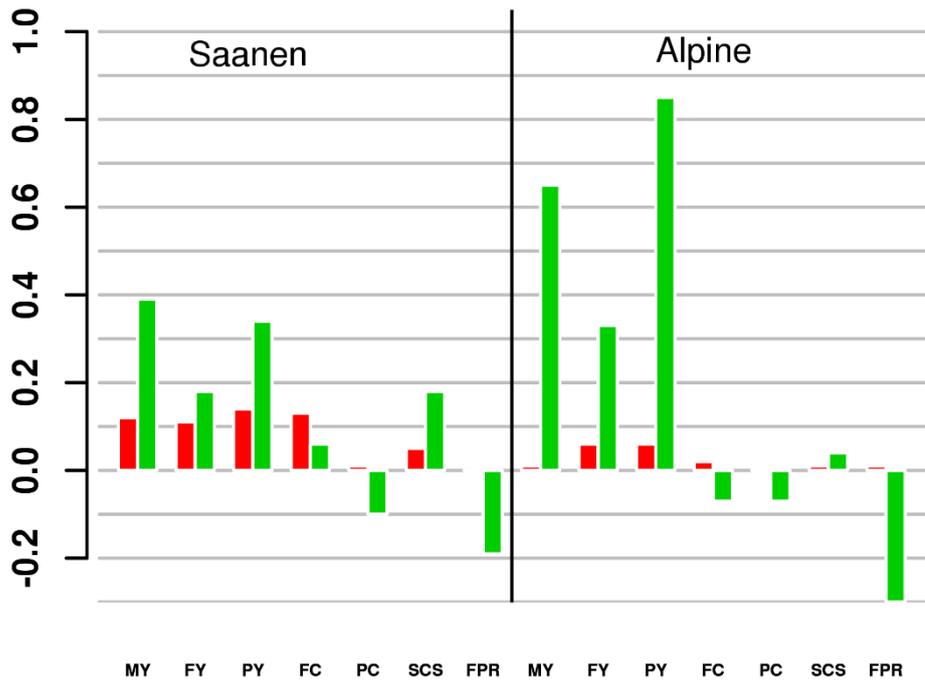


829

830 Supplemental Figure S1: Correlations between DYD\_RRM and EBVs from A\_RRM with complete  
 831 dataset (A\_RRM\_COMP) and between DYD\_RRM and DYD\_LM for candidate bucks in Alpine and  
 832 Saanen.

833 *A\_RRM\_COMP: EBVs from RRM with complete dataset without genomic data, DYD\_LM: DYD from*  
 834 *LM with complete dataset without genomic data*

835



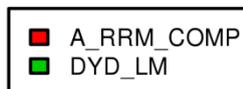
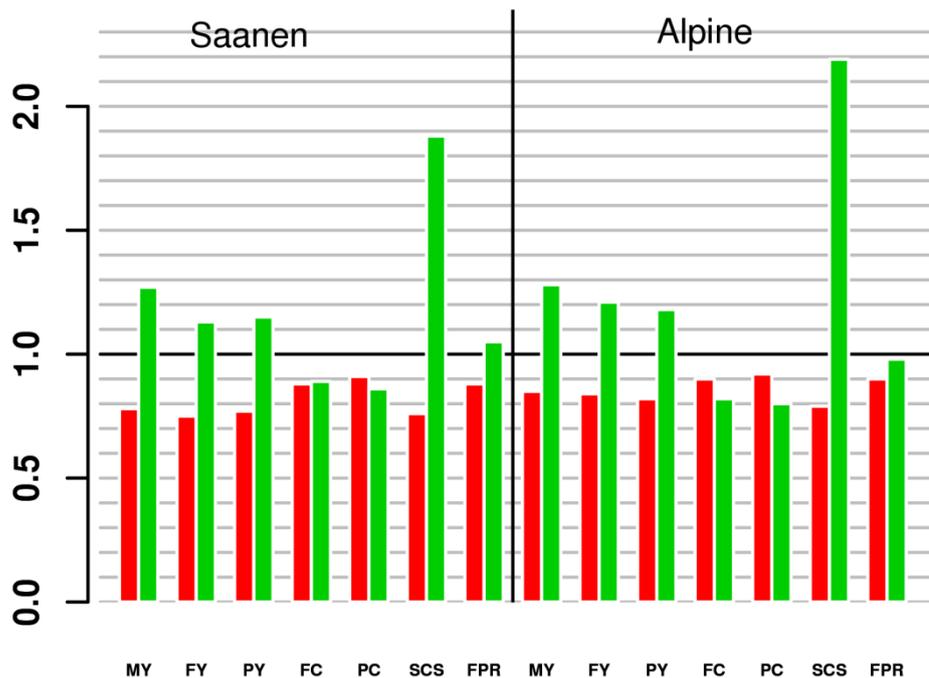
836

837 Supplemental Figure S2: Differences of means of EBVs between A\_RRM and DYD\_RRM and  
 838 differences of means between DYD\_LACT and DYD\_RRM for candidate bucks with complete dataset  
 839 I Alpine and Saanen

840 *A\_RRM\_COMP: EBVs from RRM with complete dataset without genomic data, DYD\_LM: DYD from*  
 841 *LM with complete dataset without genomic data*

842 *Differences of means of EBVs between A\_RRM and DYD\_RRM:*  
 843  *$\frac{\text{weighted mean by number of progeny of EBVs from A\_RRM} - \text{weighted mean by number of progeny of DYD\_RRM}}{\text{variance of DYD\_RRM}}$*

844 *Differences of means between DYD\_LACT and DYD\_RRM:*  
 845  *$\frac{\text{weighted mean by number of progeny of DYD\_LACT} - \text{weighted mean by number of progeny of DYD\_RRM}}{\text{variance of DYD\_RRM}}$*



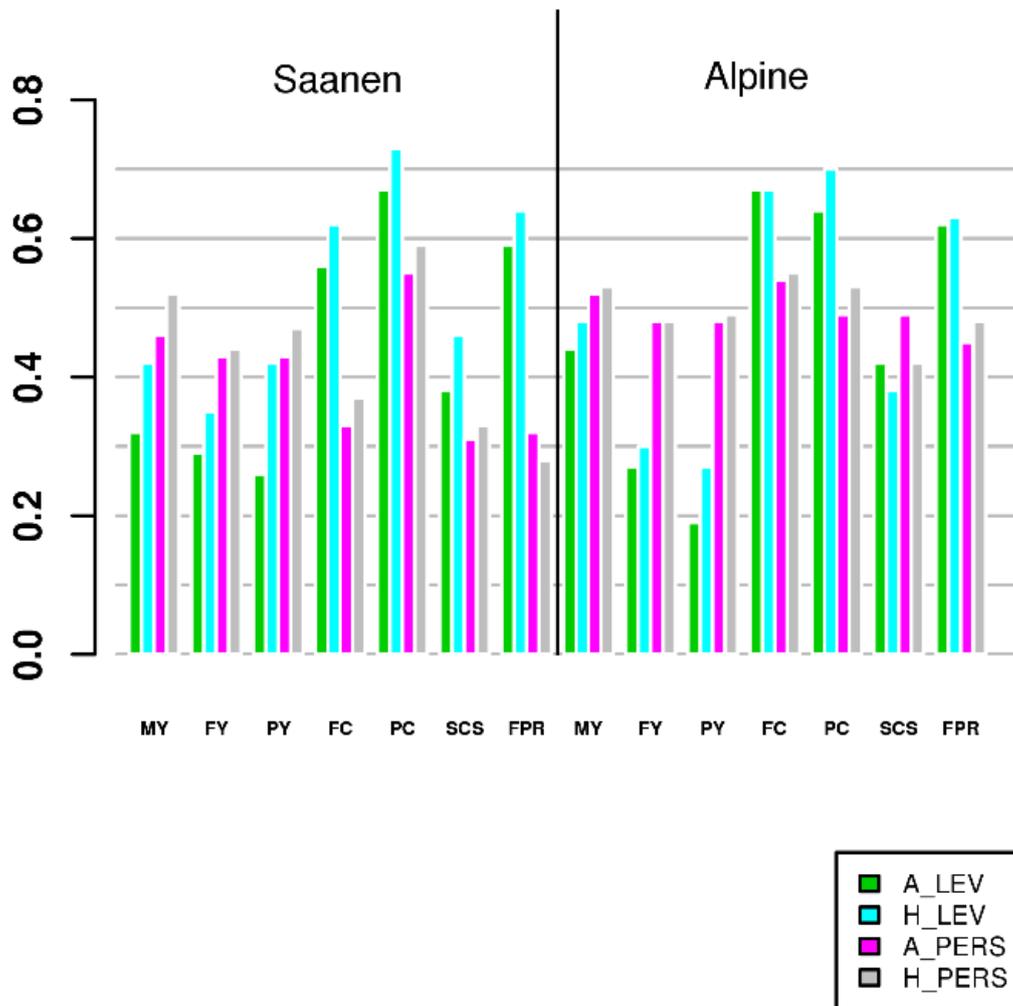
846

847 Supplemental Figure S3: Ratios of variances of A\_RRM\_COMP on DYD\_RRM and ratios of variances  
 848 of DYD\_LACT on DYD\_RRM for candidate bucks with the complete dataset in Alpine et Saanen

849 *A\_RRM\_COMP: EBVs from RRM with complete dataset without genomic data, DYD\_LM: DYD from*  
 850 *LM with complete dataset without genomic data*

851

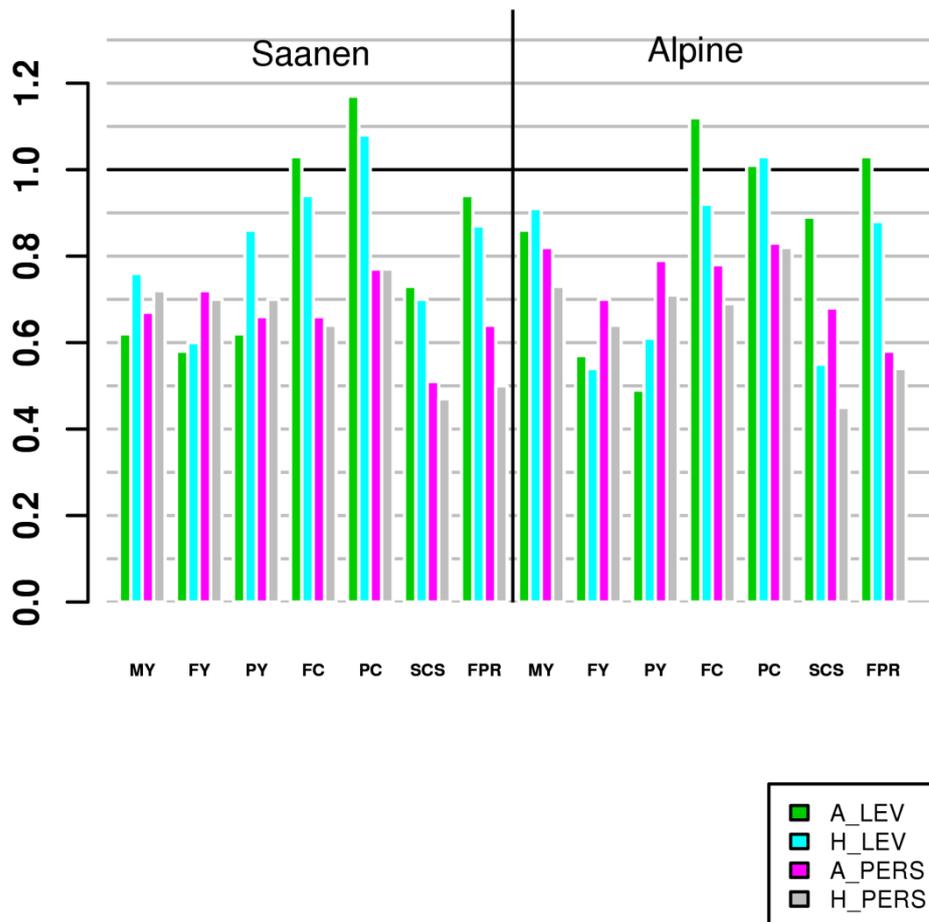
852



853

854 Supplemental Figure S4: Correlation with or without genomic information for the different models and  
 855 traits for the candidate bucks weighted by the number of progeny in Alpine and Saanen

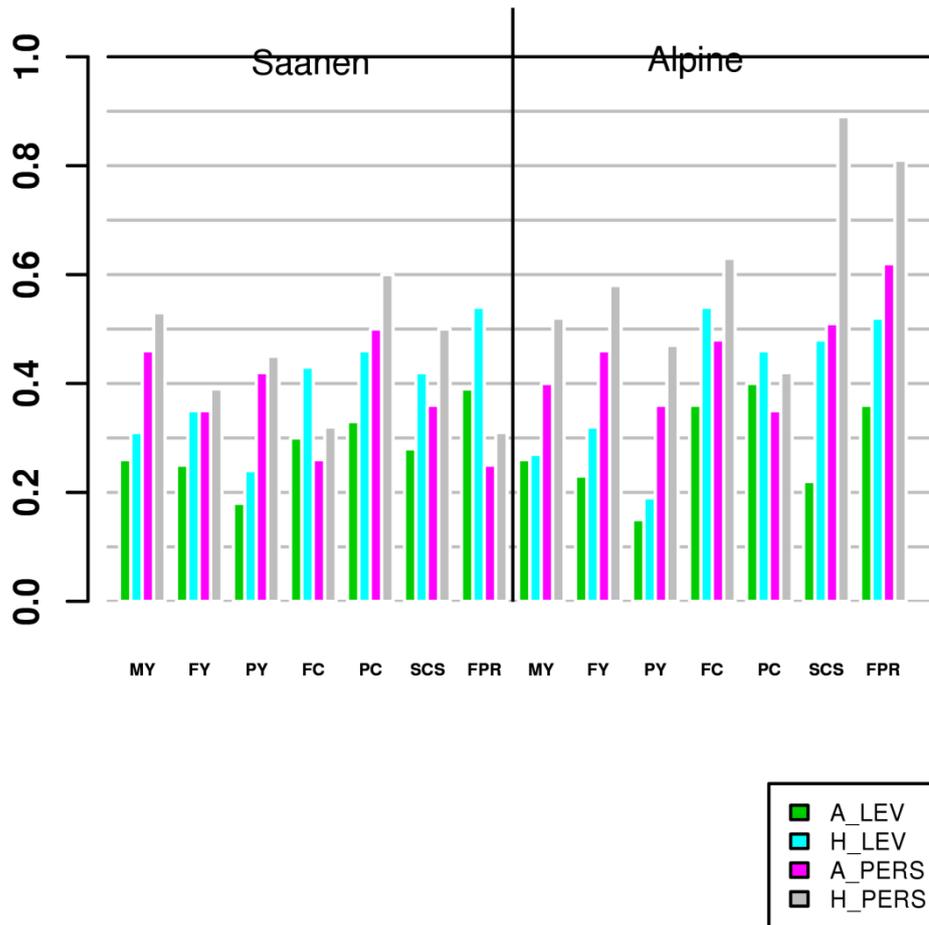
856 (*MY: milk, FY: Fat yield, PY: protein yield, FC: fat content, PC: protein content, SCS: somatic cell*  
 857 *score, FPR: fat protein ratio*) *A\_LEV: regression coefficient for the first eigenfunction estimated without*  
 858 *genomic information (level of milk throughout the lactation), A\_PERS: regression coefficient for the*  
 859 *second eigenfunction estimated without genomic information (persistence), H\_LEV: regression*  
 860 *coefficient for the first eigenfunction estimated with genomic information (level of milk throughout the*  
 861 *lactation), H\_PERS: regression coefficient for the second eigenfunction estimated with genomic*  
 862 *information (persistence)*



863

864 Supplemental Figure S5: Regression slope of DYD\_LEV on A\_LEV and H\_LEV estimated with partial  
 865 dataset and the slope of the regression of DYD\_PERS on A\_PERS and H\_PERS obtained with partial  
 866 dataset weighted by the number of progeny in Alpine and Saanen

867 *A\_LEV: regression coefficient for the first eigenfunction estimated without genomic information (level*  
 868 *of milk throughout the lactation), A\_PERS: regression coefficient for the second eigenfunction estimated*  
 869 *without genomic information (persistence), H\_LEV: regression coefficient for the first eigenfunction*  
 870 *estimated with genomic information (level of milk throughout the lactation), H\_PERS: regression*  
 871 *coefficient for the second eigenfunction estimated with genomic information (persistence)*

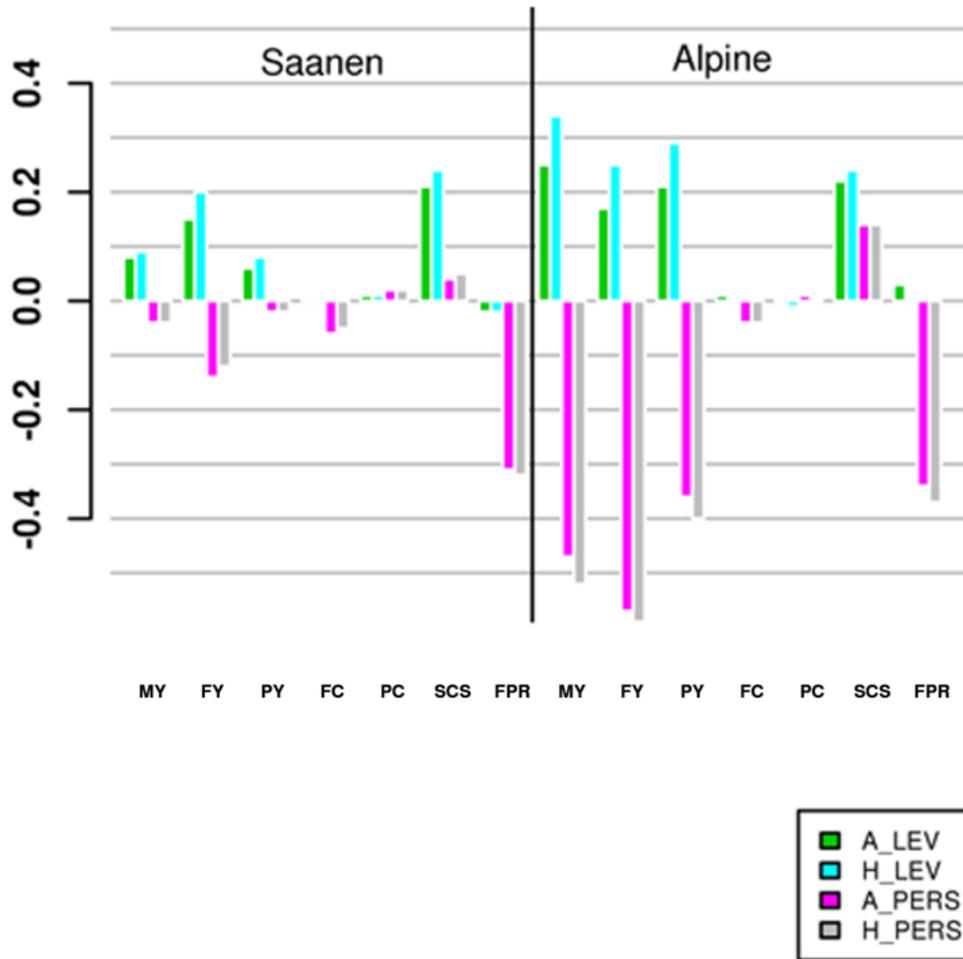


872

873 Supplemental Figure S6: Variance ratios of DYD\_LEV on A\_LEV and H\_LEV estimated with partial  
 874 dataset and variance ratios of DYD\_PERS on A\_PERS and H\_PERS obtained with partial dataset  
 875 weighted by the number of progeny in Alpine and Saanen

876 *A\_LEV*: regression coefficient for the first eigenfunction estimated without genomic information (level  
 877 of milk throughout the lactation), *A\_PERS*: regression coefficient for the second eigenfunction estimated  
 878 without genomic information (persistence), *H\_LEV*: regression coefficient for the first eigenfunction  
 879 estimated with genomic information (level of milk throughout the lactation), *H\_PERS*: regression  
 880 coefficient for the second eigenfunction estimated with genomic information (persistence)

881



882

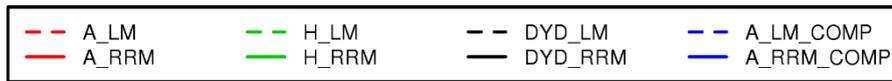
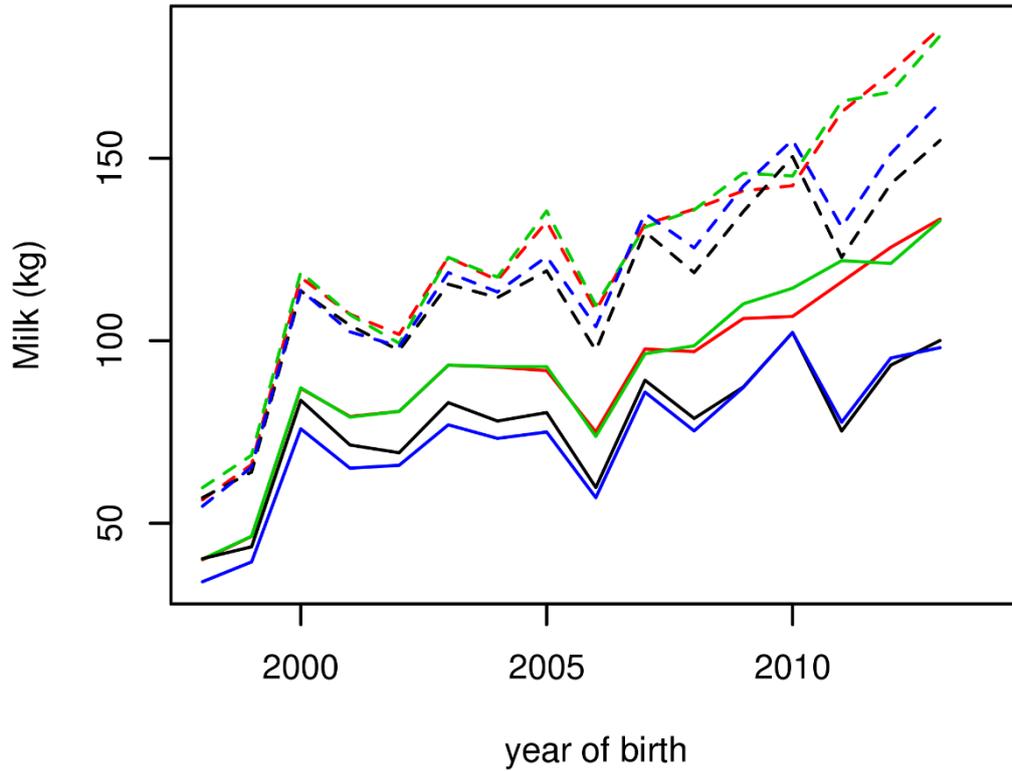
883 Supplemental Figure S7: Bias for the different models and traits for candidate bucks in Saanen and  
 884 Alpine for LEV and PERS weighted by the number of progeny in Alpine and Saanen

885 (*MY: milk, FY: Fat yield, PY: protein yield, FC: fat content, PC: protein content, SCS: somatic cell*  
 886 *score, FPR: fat protein ratio*)

887 
$$\text{Bias: } \mu_{D_{YD,p}} = \frac{\overline{\hat{u}_p - D_{YD}}}{\sigma_{D_{YD}}}$$

888 with  $\mu_{D_{YD,p}}$  the bias,  $\hat{u}_p$  EBV obtained with partial data, and  $\sigma_{D_{YD}}$  a standard deviation of  $D_{YD}$ .

889 *A\_LEV: regression coefficient for the first eigenfunction estimated without genomic information (level*  
 890 *of milk throughout the lactation), A\_PERS: regression coefficient for the second eigenfunction estimated*  
 891 *without genomic information (persistence), H\_LEV: regression coefficient for the first eigenfunction*  
 892 *estimated with genomic information (level of milk throughout the lactation), H\_PERS: regression*  
 893 *coefficient for the second eigenfunction estimated with genomic information (persistence)*

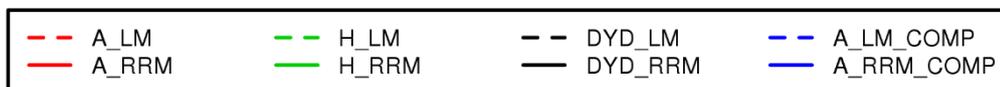
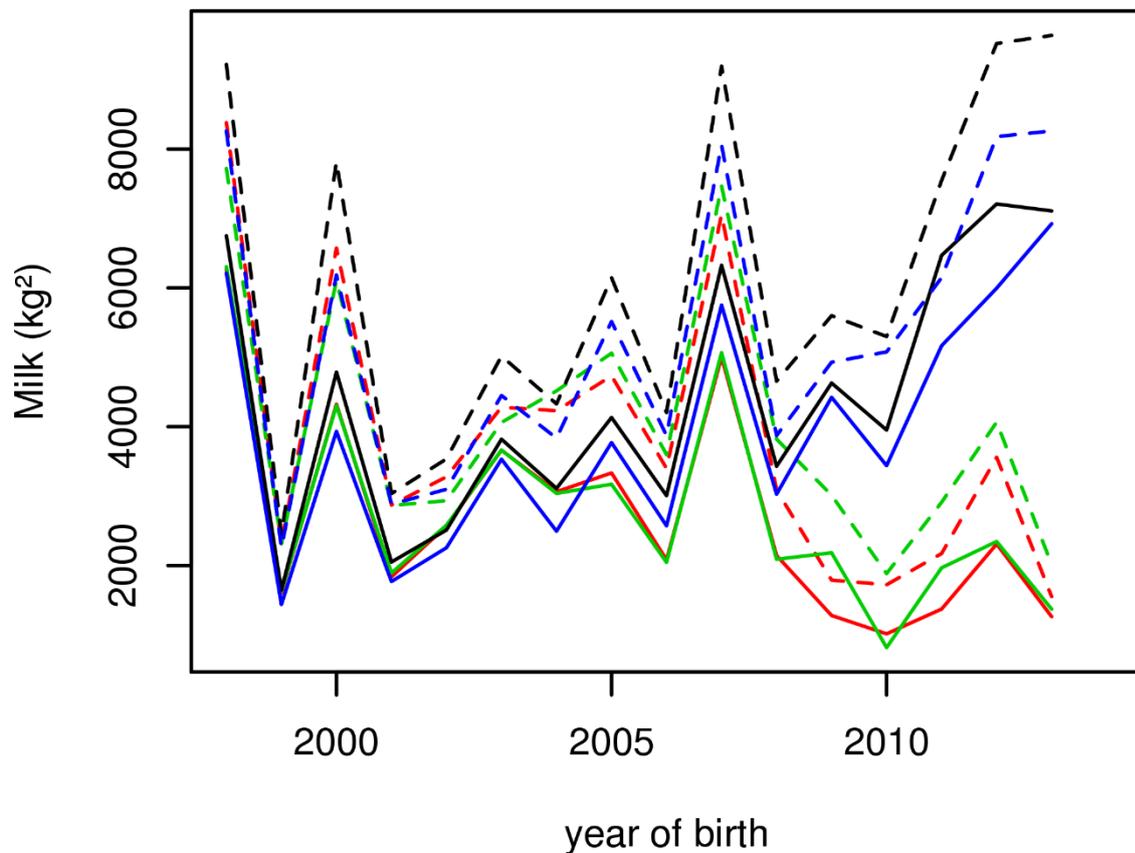


894

895 Supplemental Figure S8: Mean of MY EBV<sub>tot</sub> by year of birth of genotyped bucks weighted by the  
 896 number of progeny in Alpine

897 *A\_LM*: lactation model without genomic information for the relationship matrix building with  
 898 incomplete dataset, *H\_LM*: lactation model with genomic information for the relationship matrix  
 899 building with incomplete dataset, *A\_RRM*: random regression model without genomic information for  
 900 the relationship matrix building with incomplete dataset, *H\_RRM*: random regression model with  
 901 genomic information for the relationship matrix building with incomplete dataset, *A\_LM\_COMP*:  
 902 lactation model without genomic information for the relationship matrix building with complete dataset,  
 903 *A\_RRM\_COMP*: random regression model without genomic information for the relationship matrix  
 904 building with complete dataset

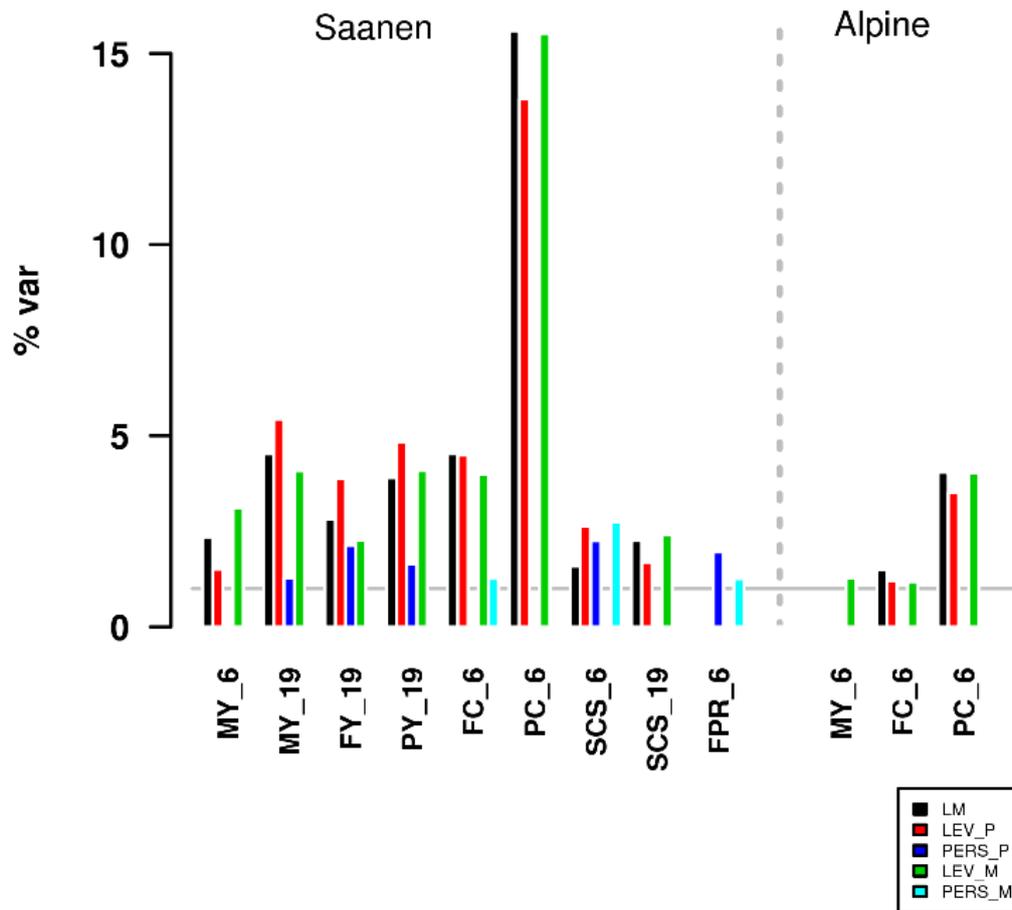
905



906

907 Supplemental Figure S9: Evolution of the variance by year of birth of genotyped bucks weighted by the  
 908 number of progeny for the different models in Alpine

909 *A\_LM*: lactation model without genomic information for the relationship matrix building with  
 910 incomplete dataset, *H\_LM*: lactation model with genomic information for the relationship matrix  
 911 building with incomplete dataset, *A\_RRM*: random regression model without genomic information for  
 912 the relationship matrix building with incomplete dataset, *H\_RRM*: random regression model with  
 913 genomic information for the relationship matrix building with incomplete dataset, *A\_LM\_COMP*:  
 914 lactation model without genomic information for the relationship matrix building with complete dataset,  
 915 *A\_RRM\_COMP*: random regression model without genomic information for the relationship matrix  
 916 building with complete dataset



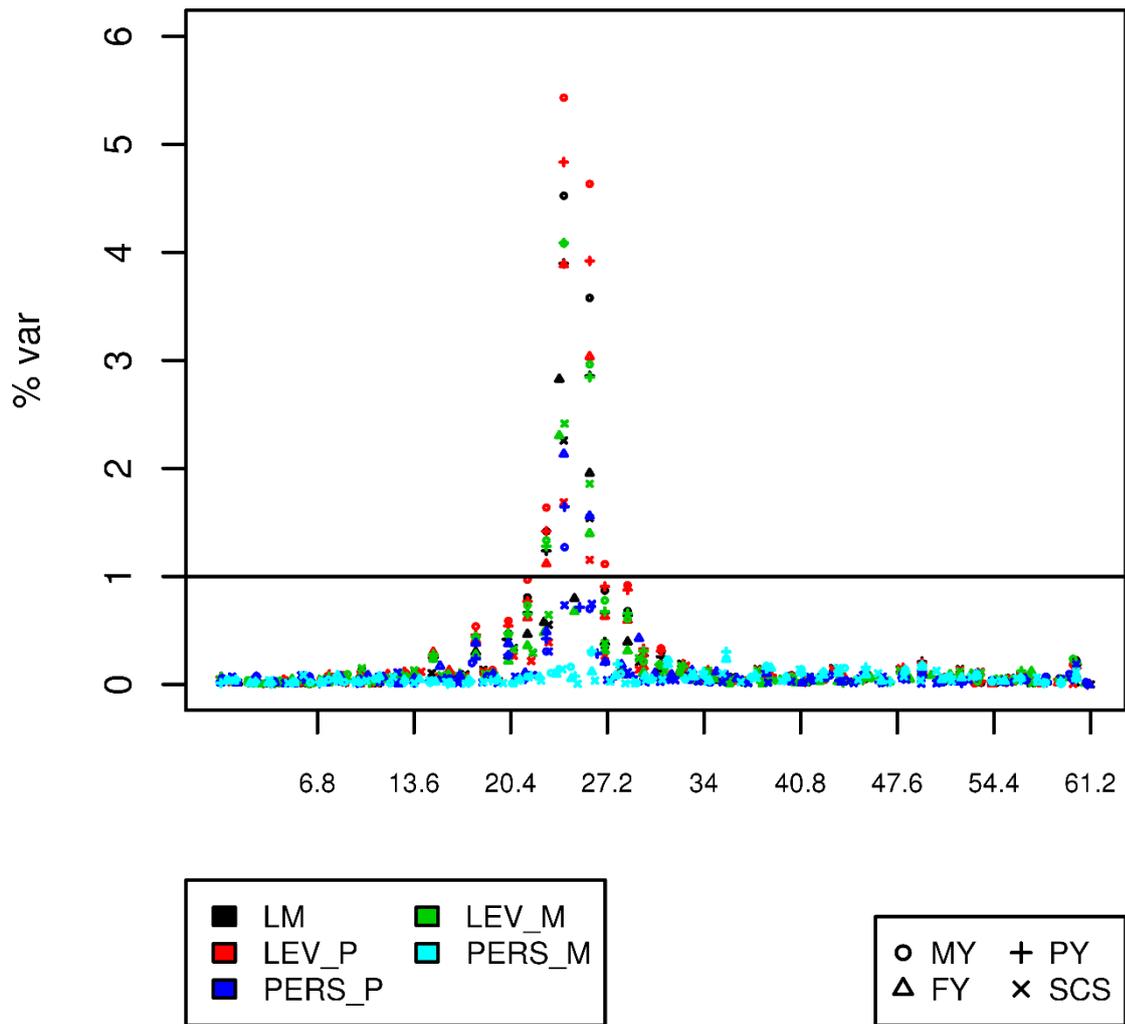
917

918 Supplemental Figure S10: Percentage of total variance (>1%), represented by segment of 20 consecutive  
 919 SNP, by chromosome and trait in Alpine and Saanen

920 *LM*: model based on lactation phenotypes, *LEV\_P*: regression coefficient for the first eigenfunction for  
 921 primiparous (level of milk throughout the lactation), *PERS\_P*: regression coefficient for the second  
 922 eigenfunction for primiparous (persistence), *LEV\_M*: regression coefficient for the first eigenfunction  
 923 for multiparous (level of milk throughout the lactation), *PERS\_M*: regression coefficient for the second  
 924 eigenfunction for multiparous (persistence)

925

926



927

928 Supplemental Figure S11: Percentages of total variance represented by non overlapping segments of 20  
 929 adjacent SNP on chromosome 19 in Saanen for MY, FY, PY and SCS

930 *LM*: model based on lactation phenotypes, *LEV\_P*: regression coefficient for the first eigenfunction for  
 931 primiparous (level of milk throughout the lactation), *PERS\_P*: regression coefficient for the second  
 932 eigenfunction for primiparous (persistence), *LEV\_M*: regression coefficient for the first eigenfunction  
 933 for multiparous (level of milk throughout the lactation), *PERS\_M*: regression coefficient for the second  
 934 eigenfunction for multiparous (persistence)

935

936

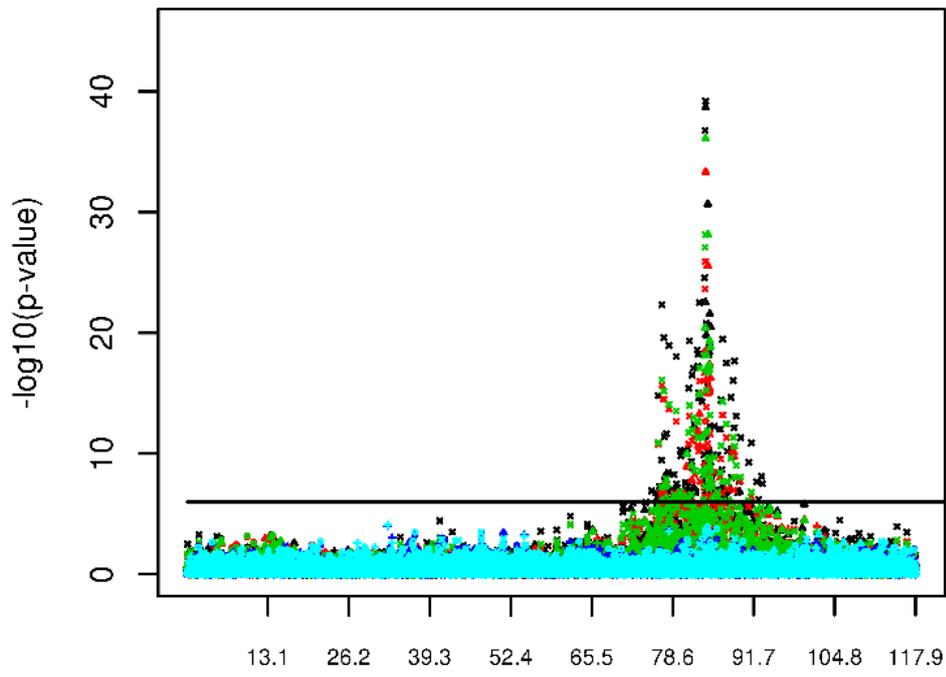
### **III. Résultats complémentaires sur l'analyse des effets des SNPs issus des modèles ssGBLUP RRM et ssGBLUP LM pour l'analyse des caractères de production laitière et scores de cellules somatiques en race Alpine et Saanen**

Dans la partie précédente, nous avons fait le choix de ne présenter que les  $-\log_{10}(\text{p-value})$  et les pourcentages de variances représentés par 20 SNPs consécutifs pour le chromosome 19 sur les caractères pour lesquels la  $-\log_{10}(\text{p-value})$  est supérieure à 6 ou le pourcentage de variance supérieur à 1%, c'est-à-dire pour le Lait, la MP, la MG et les SCS. La Figure 51 présente les  $-\log_{10}(\text{p-value})$  des effets des SNPs pour les 4 chromosomes (6, 8, 14 et 23) pour lesquels des SNPs ont été détectés avec une  $-\log_{10}(\text{p-value})$  supérieure à 6.

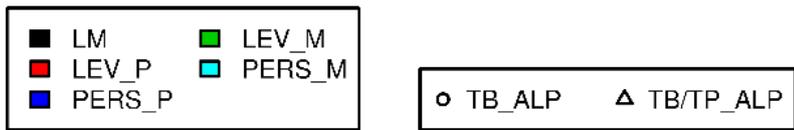
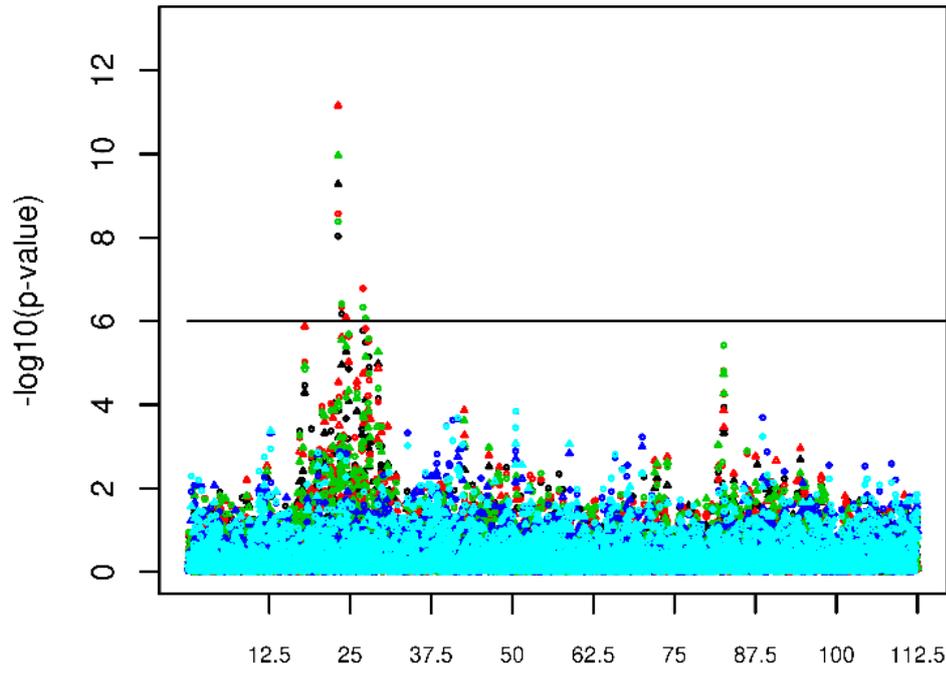
Comme nous l'avons vu pour le chromosome 19, les Manhattan plots du chromosome 6, 8, 14 et 23 nous permettent de visualiser des régions génomiques de plusieurs mégabases liées à plusieurs caractères : le TB et le TP pour le chromosome 6 pour les deux races (Figure 51.a), le TB et le ratio TB/TP pour le chromosome 8 en Alpine uniquement (Figure 51.b) et ces mêmes caractères pour le chromosome 23 pour les deux races (Figure 51.c). Pour le chromosome 8, un deuxième pic autour de 83 mégabases est identifié pour le TB et le ratio TB/TP en Alpine. Pour le chromosome 23, les résultats sont différents, nous n'observons que 2 SNPs significatifs (Figure 51.d), la persistance en primipares des SCS et les SNPs 24 811 860 et 27 664 633 du chromosome 23. Des résultats similaires ont été observés sur les chromosomes 2 et 3 (résultats non montrés): un SNP (130,868,973) est détecté significatif pour le TP en race Alpine sur le chromosome 2 et un autre SNP (5,784,974) détecté significatif pour les SCS en Alpine également.

La Figure 52 présente les pourcentages de variance représentés par 20 SNPs consécutifs du chromosome 6, pour les caractères pour lesquels le pourcentage de variance est supérieur à 1%, c'est-à-dire le lait, la MP, le TB, le TP, les SCS, le ratio TB/TP en Saanen, et le lait, le TB, et le TP en Alpine. Une même région génomique semble être associée à plusieurs caractères. Contrairement au  $-\log_{10}(\text{p-value})$ , d'autres caractères sont associés probablement dû au fait que en regroupant plusieurs SNP, qui ont une  $-\log_{10}(\text{p-value})$  légèrement inférieure à 6, ils représentent tout de même plus de 1% de la variance du caractère.

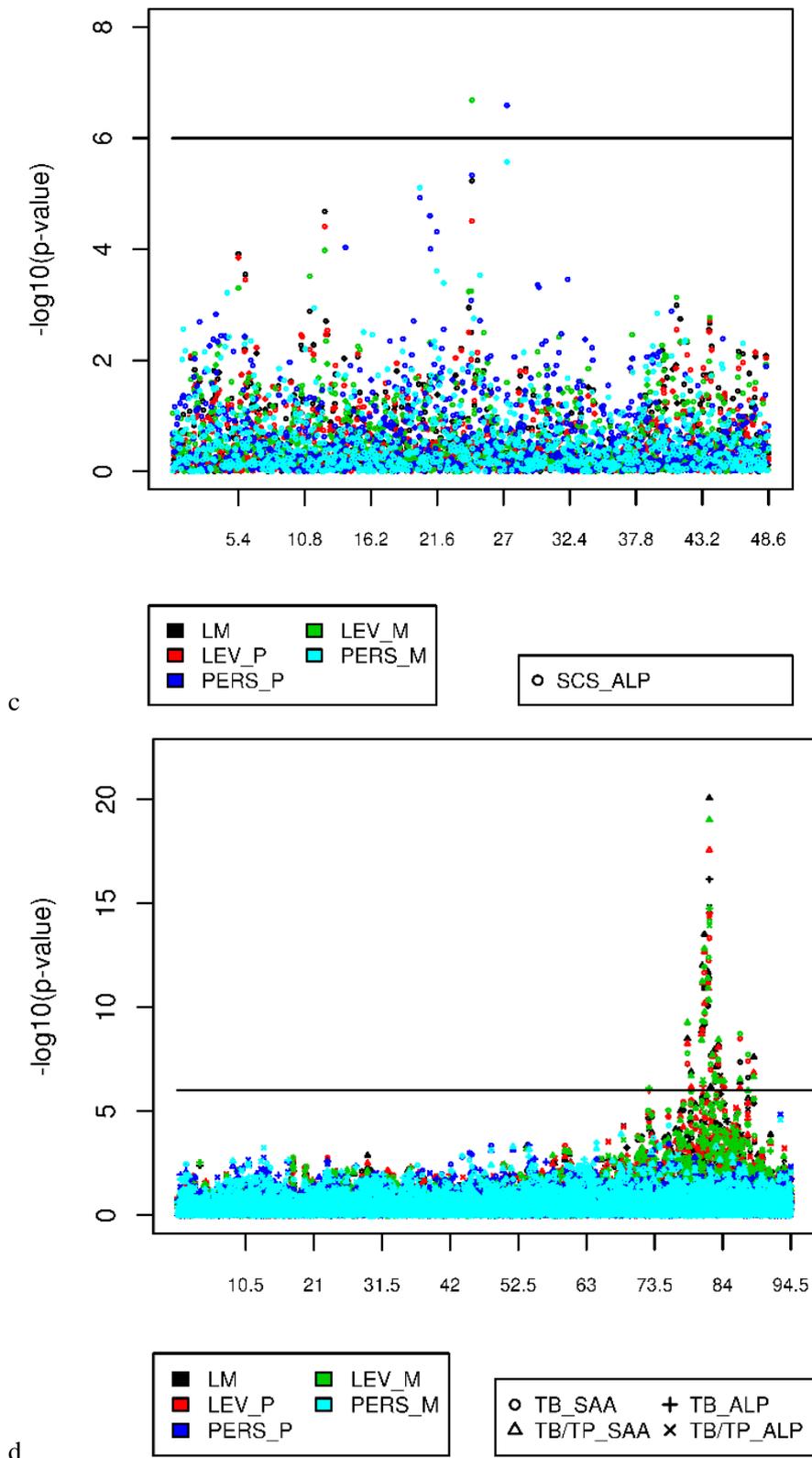
Les régions génomiques, les plus significatives, pourraient être valorisées dans un modèle d'évaluation génomique, en utilisant des approches de type Weighted ssGBLUP, approche développées dans Teissier et al. (2018). Ces approches permettent d'attribuer un poids supérieur aux SNPs les plus associés au caractère évalué.



a

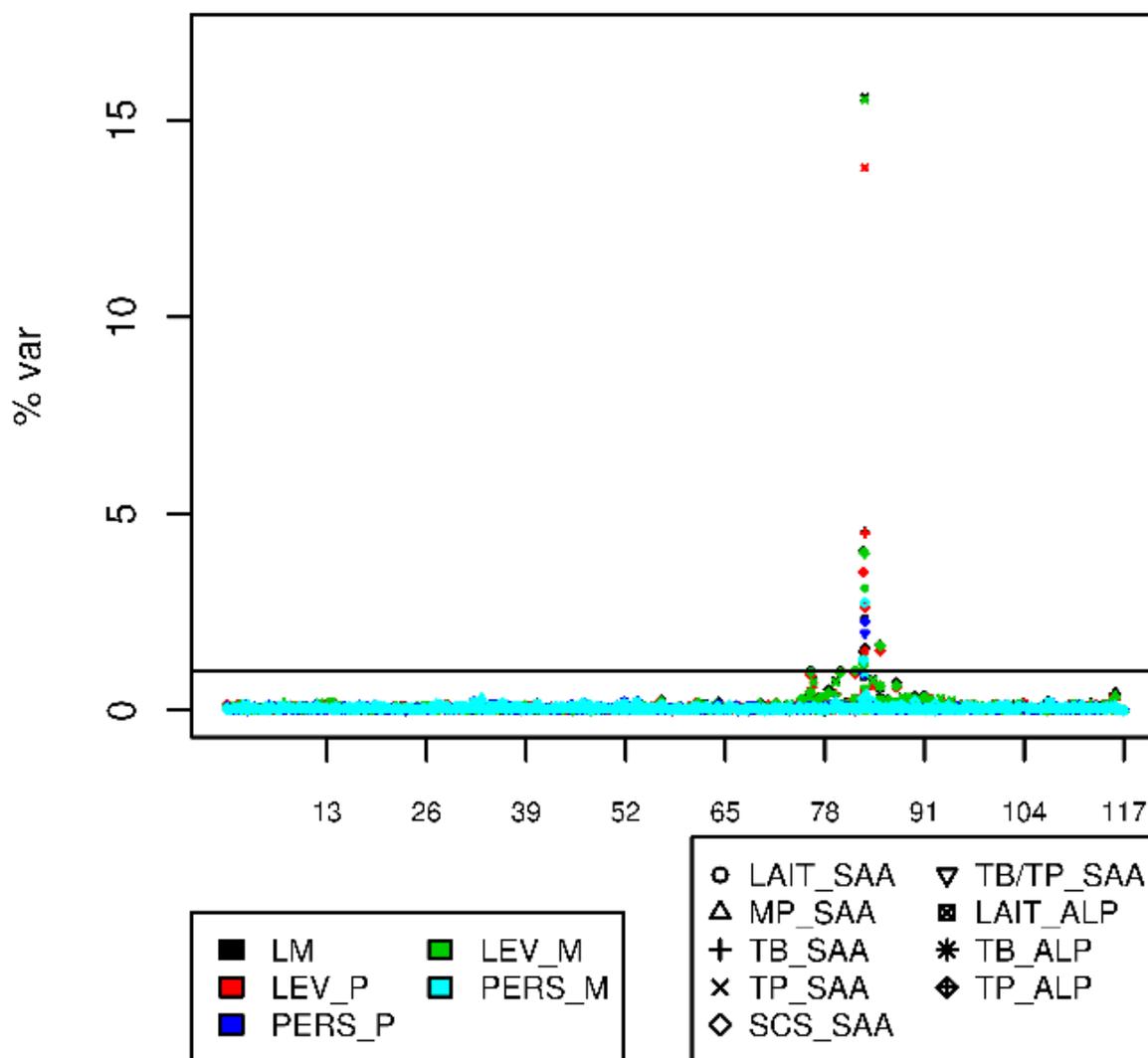


b



**Figure 51: Manhattan plot des effets des SNPs pour le: a) chromosome 6, b) chromosome 8, c) chromosome 14, d) chromosome 23, issus des modèles ssGBLUP<sub>LM</sub>, ou ssGBLUP<sub>RRM</sub> pour les coefficients de régression (LEV<sub>P</sub>, PERS<sub>P</sub>, LEV<sub>M</sub>, PERS<sub>M</sub>) pour les caractères pour lesquels des SNPs étaient détectés significatifs**

*LM* : modèle basé sur les phénotypes à la lactation, *LEV<sub>P</sub>* : coefficient de régression pour la première fonction propre des primipares (niveau de lait à la lactation), *PERS<sub>P</sub>* : coefficient de régression pour la deuxième fonction propre des primipares (persistance), *LEV<sub>M</sub>* : coefficient de régression pour la première fonction propre des multipares (niveau de lait à la lactation), *PERS<sub>M</sub>* : coefficient de régression pour la seconde fonction propre des multipares (persistances)



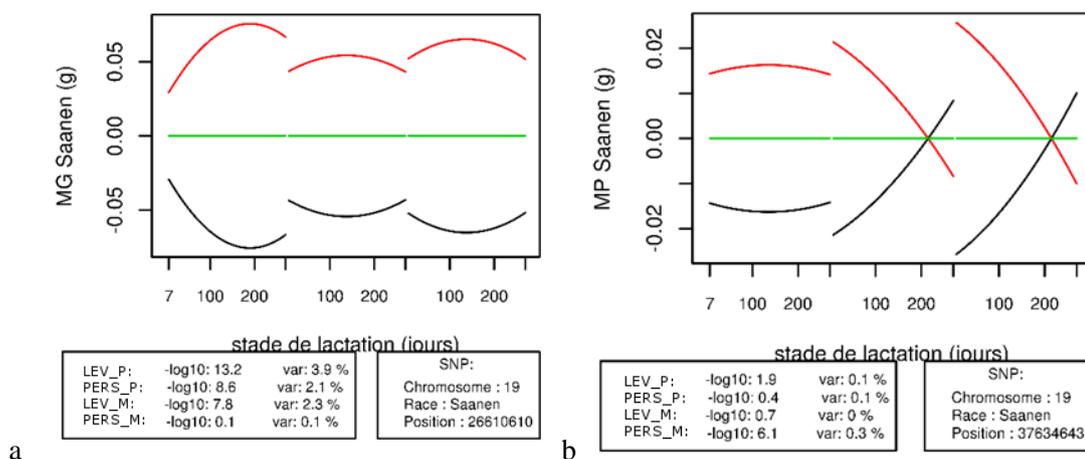
**Figure 52: Pourcentage de variance génétique représenté par 20 SNPs consécutifs sur le chromosome 6. Les résultats sont issus des modèles *ssGBLUP\_LM*, ou *ssGBLUP\_RRM* pour les coefficients de régression (*LEV\_P*, *PERS\_P*, *LEV\_M*, *PERS\_M*) pour les caractères pour lesquels des SNPs étaient détectés significatifs**

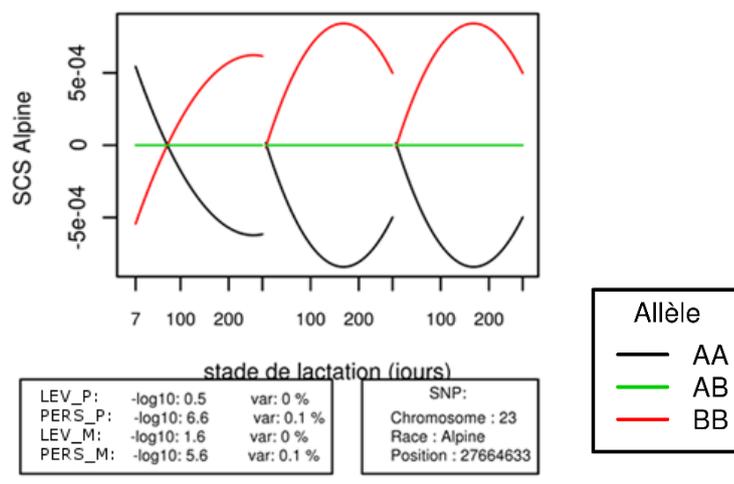
*LM* : modèle basé sur les phénotypes à la lactation, *LEV\_P* : coefficient de régression pour la première fonction propre des primipares (niveau de lait à la lactation), *PERS\_P* : coefficient de régression pour la deuxième fonction propre des primipares (persistance), *LEV\_M* : coefficient de régression pour la première fonction propre des multipares (niveau de lait à la lactation), *PERS\_M* : coefficient de régression pour la seconde fonction propre des multipares (persistances)

## 1. Analyse des effets des allèles des SNPs significativement associés à la persistance pour la MG et la MP en Saanen et les SCS en Alpine

Les effets des différents allèles (AA, AB et BB) des SNPs, ayant une  $-\log_{10}(\text{p-value})$  supérieure à 6 pour la persistance en primipares (LEV\_P) ou en multipares (PERS\_M) sont représentés au cours de la lactation pour les 3 lactations, sur la Figure 53.a pour la persistance de la MG en primipares en Saanen, sur la Figure 53.b pour la persistance de la MP en multipares en Saanen, sur la Figure 53.c pour la persistance des SCS en primipares en Alpine. Un allèle est représenté en noir (AA) et son opposé en rouge (BB) et l'hétérozygote (AB) en vert. Pour chacun des coefficients de régression (LEV\_P, PERS\_P, LEV\_M, PERS\_M), les valeurs des  $-\log_{10}(\text{p-value})$  et la part de variance représentée par un des segments de 20 SNPs dont ce SNP fait partie sont notées dans l'encart de la légende de la figure.

Pour la MG en Saanen, sur le chromosome 19 (Figure 53.a), une même région est significative à la fois pour le niveau et la persistance. En revanche, pour la persistance de la MP (Figure 53.b) et les SCS (Figure 53.c) une région est significative uniquement pour la persistance. Il est intéressant de noter que pour la persistance, il y a un seul SNP qui est significatif par sa  $-\log_{10}(\text{p-value})$  et appartient à un segment de 20 SNPs représentant plus de 1% de variance. Il s'agit du SNP significatif pour la persistance de la MG en Saanen. En effet, pour la MP, le SNP significatif pour la persistance appartient à des segments de 20 SNPs qui représentent au maximum seulement 0.3% de la variance génétique pour la persistance de la MP en multipares et pour les SCS seulement 0.1%. La différence de production entre l'allèle AA et BB est très faible (0.14 g pour la MG entre l'allèle AA et BB à 200 jours de lactation en Saanen en primipares, 0.04 g pour la MP entre l'allèle AA et BB à 7 jours de lactation en Saanen en multipares,  $1.4 \times 10^{-3}$  pour les SCS entre l'allèle AA et BB à 200 jours de lactation en Saanen en multipares) pour les trois caractères probablement car ce sont des effets estimés au sein d'un ssGBLUP qui est un modèle mixte donc les solutions des SNPs sont proches de 0. Les SNPs situés à proximité du SNP le plus significatif et en déséquilibre de liaison ont probablement également un effet non nul. Pour aller plus loin dans cette étude, il serait intéressant de comparer nos résultats, à ceux qui sont obtenus suite à des analyses pangénomiques de type GWAS où les SNPs sont pris un après l'autre.





c

**Figure 53: Evolution des effets des allèles des SNPs significatifs au cours de la lactation et pour les 3 lactations pour le coefficient de régression persistance : a) en MG en Saanen sur le chromosome 19, b) en MP sur le chromosome 19, c) en SCS pour le chromosome 23**

*LM : modèle basé sur les phénotypes à la lactation, LEV\_P : coefficient de régression pour la première fonction propre des primipares (niveau de lait à la lactation), PERS\_P : coefficient de régression pour la deuxième fonction propre des primipares (persistance), LEV\_M : coefficient de régression pour la première fonction propre des multipares (niveau de lait à la lactation), PERS\_M : coefficient de régression pour la seconde fonction propre des multipares (persistances), -log10 : valeur de la -log10(p-value) pour le SNP étudié, var : pourcentage de variance représenté par un des segments de 20 SNPs contenant le SNP étudié, chromosome : numéro du chromosome contenant le SNP étudié, Race : race pour laquelle l'effet du SNP est étudié, position : position du SNP étudié exprimé en base*

## **2. Evolution des effets des allèles de SNPs associées à plusieurs caractères pour le niveau et la persistance sur les chromosomes 6 et 19**

Comme illustré par la Figure 52, plusieurs segments de 20 SNPs, localisés à la fin du chromosome 6, représentent plus de 1% de variance pour plusieurs caractères : le lait, le TB, le TP, les SCS, le ratio TB/TP en Saanen ; le lait, le TB et le TP en Alpine.

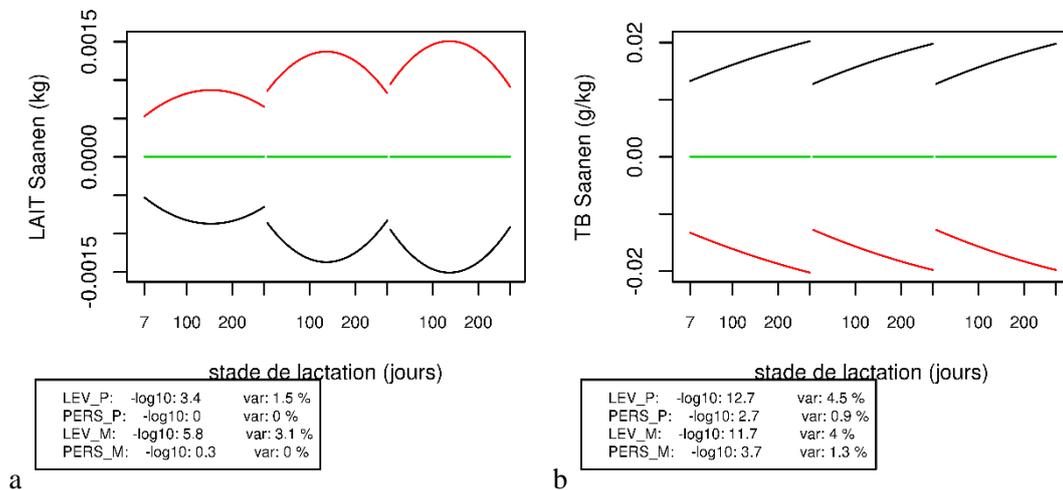
Pour la Saanen, les effets des allèles du SNP le plus significatif pour le TP ont été étudiés. La -log10(p-value) de ce SNP est la plus grande de tous les caractères étudiés et ce SNP se situe à la position 85 979 715 sur le chromosome 6. Nous avons ensuite analysé les effets de ce SNP pour les autres caractères. Les résultats sont présentés sur la Figure 54. Un allèle est représenté en noir (AA) et son opposé en rouge (BB). Nous pouvons voir que l'allèle AA de ce SNP est lié à moins de lait (part de variance associée au segment de 20 SNPs contenant ce SNP : 1.5% en primipares et 3.5% en multipares), plus de TB (part de variance associée au segment de 20 SNPs contenant ce SNP : 4.5% en primipares et 4% en multipares), plus de TP (part de variance associée : 13.8% en primipares et 15.5% en multipares) et moins de SCS pour les niveaux en primipares (part de variance associée : 2.6%). Par contre, on

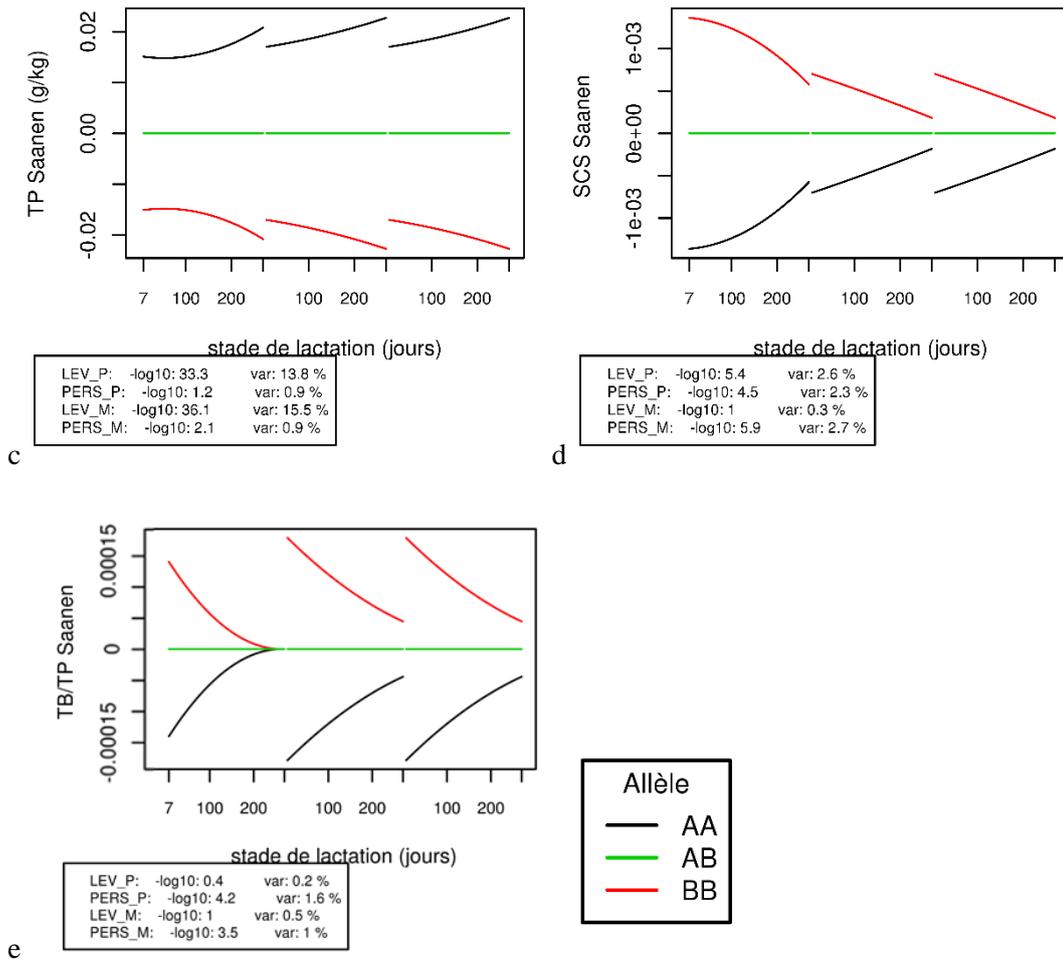
n'observe aucun effet sur le niveau du ratio TB/TP (Figure 54.e), les parts de variance associées au segment de 20 SNPs contenant ce SNP sont de 0.2% en primipares et 0.5% en multipares).

Ce SNP est aussi lié à la persistance: le TB, le TP, le ratio TB/TP et les SCS augmentent au cours du temps. Les parts de variance associées au segment de 20 SNPs contenant le SNP (en position 85979715) sont de 0.9%, 0.9%, 1.6% et 2.3% pour le TB, le TP, le ratio TB/TP et les SCS respectivement en primipares et de 1.3%, 0.9%, 1% et 2.7% respectivement en multipares. Le TB augmente plus vite que le TP. Concernant la persistance sur la production laitière elle est stable au cours du temps avec des parts de variance associées au segment de 20 SNPs contenant le SNP 85 979 715 nulles en primipares et multipares.

En race Alpine, les résultats sont légèrement différents. Nous observons que l'allèle AA est lié à moins de lait, plus de TB, plus de TP pour le niveau, comme en race Saanen mais il n'existe pas de variation au cours du temps (résultats non montrés).

Comme nous venons de l'étudier, le SNP 85 979 715 a un effet sur le niveau et la persistance pour le TP, le TB et les SCS en primipares. Pour le lait, il a un effet uniquement sur le niveau, et un effet uniquement sur la persistance du ratio TB/TP. Ces associations sont détectables lorsque nous nous intéressons aux variances représentées par des segments de 20 SNPs consécutifs. Si on ne s'intéresse qu'aux p-values supérieures à un seuil de 6, le SNP 85 979 715 n'est significatif que pour le niveau du TP et du TB.





**Figure 54: Etude des effets des allèles (AA, AB et BB) du SNP en position 85 979 715 sur le chromosome 6 en race Saanen pour les caractères pour lesquels ce SNP représente une part de variance supérieure à 1% pour l'une des 4 composantes (LEV\_P, PERS\_P, LEV\_M, PERS\_M) du modèle ssGBLUP\_RRM : a) lait, b) TB, c) TP, d) SCS, e) ratio TB/TP**

*LM* : modèle basé sur les phénotypes à la lactation, *LEV\_P* : coefficient de régression pour la première fonction propre des primipares (niveau de lait à la lactation), *PERS\_P* : coefficient de régression pour la deuxième fonction propre des primipares (persistance), *LEV\_M* : coefficient de régression pour la première fonction propre des multipares (niveau de lait à la lactation), *PERS\_M* : coefficient de régression pour la seconde fonction propre des multipares (persistances), *-log10* : valeur de la  $-\log_{10}(p\text{-value})$  pour le SNP étudié, *var* : pourcentage de variance représenté par un des segments de 20 SNPs contenant le SNP étudié, *chromosome* : numéro du chromosome contenant le SNP étudié, *Race* : race pour laquelle l'effet du SNP est étudié, *position* : position du SNP étudié exprimé en base

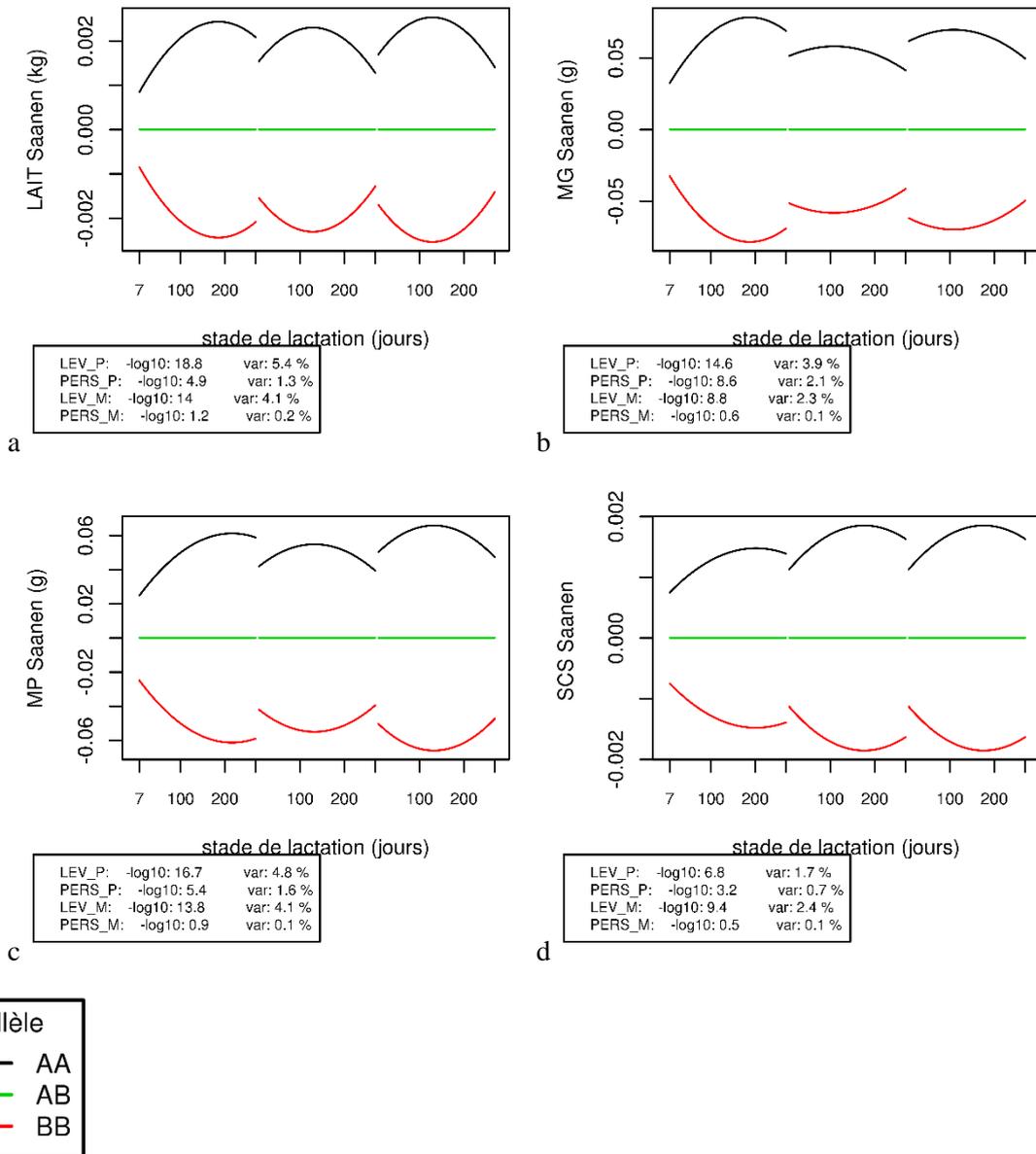
Comme illustré dans l'article 4, plusieurs segments de 20 SNPs, localisés à la fin du chromosome 19, représentent plus de 1% de variance pour plusieurs caractères : le lait, la MG, la MP et les SCS en Saanen.

Pour la Saanen, les effets des allèles du SNP le plus significatif pour le lait ont été étudiés. La  $-\log_{10}(p\text{-value})$  de ce SNP est la plus grande de tous les caractères étudiés et ce SNP se situe à la position 26 192 128 sur le chromosome 19. Nous avons ensuite analysé les effets de ce SNP pour les autres caractères associés au début du chromosome 19 dans la Figure 55. Un allèle est représenté en noir (AA) et son opposé en rouge (BB). Nous pouvons voir que l'allèle AA de ce SNP est lié à plus de lait (part de variance associée à un segment de 20 SNPs contenant ce SNP : 5.4% en primipares et 4.1% en multipares), plus de MG (part de variance associée à un segment de 20 SNPs contenant ce SNP : 3.9% en primipares et 2.3% en multipares), plus de MP (part de variance associée à un segment de 20 SNPs contenant ce SNP : 4.8% en primipares et 4.1% en multipares), plus de SCS (part de variance associée à un segment de 20 SNPs contenant ce SNP : 1.7% en primipares et 2.4% en multipares).

Ce SNP est aussi lié à la persistance en primipares avec une augmentation de la persistance du lait, de la MG, de la MP et des SCS. Les parts de variance associées à un segment de 20 SNPs contenant ce SNP 26 192 128 sont de 1.3%, 2.1%, 1.6%, et 0.7% en primipares. Par contre, on n'observe aucun effet sur la persistance en multipares.

Ces associations sont visibles lorsque nous nous intéressons aux variances représentées par 20 SNPs en revanche les p-values ne sont supérieures à 6 uniquement pour le niveau du lait, de la MP, de la MG et des SCS pour les primipares et les multipares, et de la persistance de la MG en primipares.

Ces résultats illustrent le fait qu'un même SNP peut avoir des effets importants sur plusieurs caractères d'une part et d'autre part des effets à la fois sur le niveau et sur la persistance. Il est intéressant de noter qu'un même SNP a des effets qui peuvent être positifs pour plusieurs caractères (TB et TP pour le SNP 85 979 715 sur le chromosome 6 en Saanen ; lait, MG et MP pour le SNP 26 192 128 sur le chromosome 19) ou des effets opposés (lait et TP pour le SNP 85 979 715 sur le chromosome 6 en Saanen). Ces résultats seront retrouvés par l'analyse des corrélations entre index génétiques des boucs (voir discussion) notamment les fortes corrélations entre TB-TP, lait-MG-MP, les corrélations positives entre persistance en primipares et niveau en multipares pour le lait-MG-MP-SCS, et les corrélations négatives entre Lait et TP.



**Figure 55: Etude des effets des allèles (AA, AB et BB) du SNP 85 979 715 sur le chromosome 19 en race Saanen pour les caractères pour lesquels ce SNP représente une part de variance supérieure à 1% pour l'une des 4 composantes (LEV\_P, PERS\_P, LEV\_M, PERS\_M) du ssGBLUP RRM : a) lait, b) MG, c) MP, d) SCS**

*LM : modèle basé sur les phénotypes à la lactation, LEV\_P : coefficient de régression pour la première fonction propre des primipares (niveau de lait à la lactation), PERS\_P : coefficient de régression pour la deuxième fonction propre des primipares (persistance), LEV\_M : coefficient de régression pour la première fonction propre des multipares (niveau de lait à la lactation), PERS\_M : coefficient de régression pour la seconde fonction propre des multipares (persistances), -log10 : valeur de la -log10(p-value) pour le SNP étudié, var : pourcentage de variance représenté par un des segments de 20 SNPs contenant le SNP étudié, chromosome : numéro du chromosome contenant le SNP étudié, Race : race pour laquelle l'effet du SNP est étudié, position : position du SNP étudié exprimé en base*

### **3. Analyse des temps de calcul et de la mémoire utilisés pour l'exécution d'une évaluation génomique ssGBLUP basée sur les contrôles élémentaires et d'une évaluation ssGBLUP basée sur des phénotypes à la lactation**

Une des principales limites des modèles basés sur les contrôles élémentaires est le temps de calcul. En effet, nous utilisons en moyenne 7 fois plus de performances et un nombre plus important d'effets aléatoires génétique et d'environnement permanent est inclus. Par exemple, dans le modèle réduit EGV\_PM, il y a 4 effets génétiques et 4 effets d'environnement permanent, nécessitant la construction de matrice de variances-covariances plus conséquentes que dans le modèle à la lactation.

Nous avons étudié le temps de calcul d'un modèle en utilisant un jeu de données qui pourrait être celui utilisé en routine afin d'étudier la faisabilité d'une telle évaluation en routine. Les jeux de données correspondent à ceux utilisés dans l'article 4 avec 7 908 192 contrôles élémentaires pour les chèvres de race Saanen (1 308 307 lactations) et 10 483 205 contrôles élémentaires pour les chèvres de race Alpine (1 701 229 lactations) sélectionnés à partir de l'année 2000 pour les chèvres qui sont entre leur première et troisième parité. Le Tableau 16 présente les temps de calcul et la mémoire utilisée par le modèle à la lactation (H\_LM) et par le modèle contrôles élémentaires EGV\_PM (appelé H\_RRM dans l'article 4). Ces deux modèles ont été testés en utilisant l'information génomique disponible.

Le modèle H\_RRM nécessite entre 1.1 et 5.3 fois plus d'itérations que le modèle H\_LM et les itérations durent moins d'une seconde pour le modèle H\_LM et entre 14 et 25 secondes par itération pour le modèle H\_RRM. Le modèle H\_RRM a été entre 25 et 100 fois plus long en temps de calcul que le modèle H\_LM et a nécessité 27h de calculs, ce qui reste convenable. Concernant le besoin en capacité de mémoire, les deux modèles sont similaires avec une consommation autour de 1Go. L'utilisation d'un modèle de régression aléatoire pour l'évaluation génomique des chèvres Françaises reste donc possible. De plus, les temps de calcul pour H\_LM sont faibles et non comparables à ceux de l'évaluation de routine actuelle qui utilise un historique bien antérieur à l'année 2000, qui est en multiracial, et qui considère, en plus des 3 premières parités, les parités 4 et suivantes.

En conclusion, nous avons montré dans ce chapitre, que les performances prédictives des modèles sont améliorées par l'utilisation des données génomiques pour le modèle à la lactation comme pour le modèle de contrôles élémentaires pour la quantité à la lactation comme pour la persistance pour les corrélations entre index et DYD pour les boucs candidats. Le modèle de contrôles élémentaires permet une estimation de la valeur génétique des animaux à la lactation avec la même qualité de prédiction que le modèle à la lactation. En revanche, la principale différence entre les deux modèles réside dans le biais plus important avec le modèle LM qui se répercute sur les courbes de progrès génétique avec des moyennes d'index LM supérieures. Les régions du génome connues comme étant associées à la production à la lactation ont été retrouvées avec le modèle de contrôles élémentaires pour

le chromosome 6 (TP, TB), le chromosome 8 (TB) en Alpine, le chromosome 14 (TB), le chromosome 19 (Lait, MG, MP, SCS) en Saanen. Des zones du génome impliquées dans la persistance ont été mises en évidence par le modèle contrôles élémentaires pour le chromosome 19 (MG, MP) en Saanen et chromosome 23 pour les SCS et nous avons également mis en évidence des régions du génome pouvant être liée à la fois au niveau et à la persistance et ce pour plusieurs caractères pour le chromosome 6 (TP, TB, Ratio TB/TP, SCS, Lait) et pour le chromosome 19 (Lait, MG, MP, SCS). L'analyse des temps de calcul des modèles de contrôles élémentaires a permis de valider la faisabilité d'une telle évaluation en routine.

Tableau 16 : Temps de calcul et mémoire utilisée pour le ssGBLUP

Modèle	Race	Caractère	Temps réel en h	Maximum Mémoire Utilisée	Nombre Itérations	Comparaison LM pour le temps
H_LM	SAANEN	LAIT	0.28	567.60MB	1136	1.0
		MG	0.26	573.33MB	1116	1.0
		MP	0.26	575.77MB	1110	1.0
		TB	0.30	573.08MB	1278	1.0
		TP	0.27	598.98MB	1098	1.0
		SCS	0.24	575.77MB	1081	1.0
		ratio TB/TP	0.28	571.84MB	1176	1.0
	ALPINE	LAIT	0.42	944.86MB	1323	1.0
		MG	0.41	876.59MB	1330	1.0
		MP	0.41	944.82MB	1278	1.0
		TB	0.44	914.74MB	1326	1.0
		TP	0.39	907.17MB	1173	1.0
		SCS	0.35	880.83MB	1157	1.0
		ratio TB/TP	0.36	878.24MB	1175	1.0
H_RRM	SAANEN	LAIT	26.53	736.61MB	6038	94.9
		MG	22.59	736.61MB	4764	85.3
		MP	21.55	763.21MB	5228	83.4
		TB	7.82	736.62MB	1410	25.8
		TP	8.34	736.31MB	1735	31.2
		SCS	25.01	736.46MB	5003	102.9
		ratio TB/TP	8.05	736.61MB	1594	28.5
	ALPINE	LAIT	20.83	1.05GB	4679	49.8
		MG	19.92	1.12GB	3660	48.2
		MP	24.42	1.08GB	3780	59.2
		TB	18.87	1.11GB	3286	42.9
		TP	14.70	1.14GB	2043	37.9
		SCS	25.13	1.14GB	4487	72.8
		ratio TB/TP	26.57	1.05GB	3938	73.7





## **Discussion générale et perspectives**

### **I. Le modèle RRM est bien adapté à l'évaluation génétique de la production globale à la lactation**

Comme mentionné en introduction, la modulation de la forme des courbes de lactation des chèvres laitières relève d'un réel intérêt zootechnique, économique, et aussi agro-écologique (Sölkner and Fuchs, 1987; Dekkers et al., 1998). Cependant, la sélection de ces populations a été basée depuis de nombreuses années sur des phénotypes de caractères de production mesurés à la lactation (Bouloc, 1991). Ces caractères constituent une part importante de l'objectif de sélection et leur importance économique a été démontrée pour les chèvres en France par le modèle bioéconomique développé dans le cadre du projet CASDAR OSIRIS (Palhiere et al., 2015). L'évaluation génétique des caractères de production à la lactation constitue donc un incontournable afin de garantir la possibilité de poursuivre les objectifs de sélection actuels. Le modèle RRM que nous avons développé devait donc être aussi performant que le modèle LM actuel pour l'évaluation génétique de la production laitière totale et apporter des innovations pour la sélection des formes de courbes.

Nous avons vu au chapitre 2 que l'interpolation par la méthode de Fleischmann (Sargent et al., 1968), qui calcule la production à la lactation à partir des contrôles élémentaires, est précise car son calcul équivaut à celui par une modélisation de la courbe par des splines. On pouvait alors considérer que la précision des phénotypes n'interférait pas dans la comparaison des modèles LM et RRM pour l'évaluation génétique de la production laitière totale.

Nous avons vu, que l'évaluation génétique de la quantité de lait totale produite à la lactation avec le modèle LM équivalait à celle estimée par le modèle RRM en sommant les valeurs génétiques journalières. En effet, aux chapitres 2-3-4 les corrélations estimées entre valeurs génétiques issues de LM et de RRM étaient très élevées, supérieures à 0.9 pour tous les caractères. Enfin, dans le premier chapitre, nous avons observé que les valeurs génétiques officielles issues de LM expliquaient la plus grande part de variance de la première composante de l'analyse en composantes principales des formes de courbes de lactation, qui représentait la quantité totale de lait à la lactation.

L'étude pangénomique réalisée pour tous les caractères à partir du modèle génomique a montré que beaucoup de régions du génome mises en évidence après une évaluation génomique LM étaient aussi mises en évidence après une évaluation génomique RRM pour les coefficients de régression associés au niveau de production (voir Chapitre 4) avec des p-values des SNP et des pourcentages de variance expliqués similaires entre modèles.

Cependant, les corrélations entre les valeurs génétiques estimées par les 2 modèles ne sont pas de 1, ce qui indique un écart d'estimation entre les 2 modèles qui pourrait être lié à la relative simplicité du modèle LM. C'est par exemple le cas de la différence génétique entre primipares et multipares. Au

chapitre 3, l'utilisation du modèle RRM a montré qu'il existait des différences de corrélations génétiques entre primipares et multipares. Pour les taux, ces corrélations sont élevées, supérieures à 0.9, mais pour les quantités, ces corrélations chutent à 0.7. Dans le modèle LM, ces corrélations ne sont pas directement prises en compte, des coefficients multiplicateurs sont appliqués aux phénotypes en fonction de la parité. Les modèles RRM que nous avons développés modélisent l'effet d'environnement permanent avec des variances différentes entre les 3 parités (modèle EGV\_FT), ou entre les primipares et les multipares (modèle EGV\_PM), et prennent en compte des corrélations entre parités plus faibles que pour l'effet génétique (Chapitre 3).

Afin de poursuivre la comparaison des valeurs génétiques issues des modèles LM et RRM, nous avons comparé le progrès génétique estimé à partir des 2 évaluations. Pour cela, nous avons utilisé les valeurs génétiques des boucs génotypés utilisés en insémination artificielle, nés après 1997 et avant 2014, soit 430 boucs en Saanen et 551 boucs en Alpine, obtenus à partir du jeu de données complet (lactations de 2000 à 2017). Le modèle RRM choisi pour cette étude est l'A\_RRM du chapitre 4. Un modèle linéaire (modèle 1) a été utilisé pour évaluer le progrès génétique annuel et les différences entre modèles.

$$y_{ij} = ax_i + m_j + e_{ij} \text{ (Modèle 1)}$$

avec  $y_{ij}$  l'index pour la production à la lactation calculé avec le modèle  $j$  (2 classes : LM, RRM) pour le bouc  $i$ ,  $a$  l'effet de l'année de naissance,  $x_i$  l'année de naissance du bouc  $i$ ,  $m_j$  l'effet du modèle, et  $e_{ij}$  les résidus du modèle. L'interaction entre l'année de naissance et la méthode a été testée, elle n'était pas significative sauf pour la quantité de lait et de MP en Alpine, avec des index qui augmentent plus vite pour le modèle LM.

Le Tableau 17 présente les coefficients de régression du progrès génétique obtenus pour les quantités à la lactation. Les valeurs génétiques pour les quantités de lait et de matières ont augmenté (+6.3 kg de lait par an, +0.29 kg de MP par an en Saanen, +5.3 kg de lait et +0.27 kg de MP pour la race Alpine) ainsi que pour les taux (+0.09 g/kg de TP en Saanen, +0.1 g/kg pour la race Alpine) pour les deux races (coefficient de régression positifs et p-values inférieures à 0). Depuis la mise au point de leur indexation, le niveau génétique des SCS a diminué, de façon significative en Alpine (-0.01 SCS par an), ainsi qu'en Saanen (p-value=0.07). Pour le ratio qui n'est pas sélectionné, il n'y a pas eu d'évolution significative des valeurs génétiques.

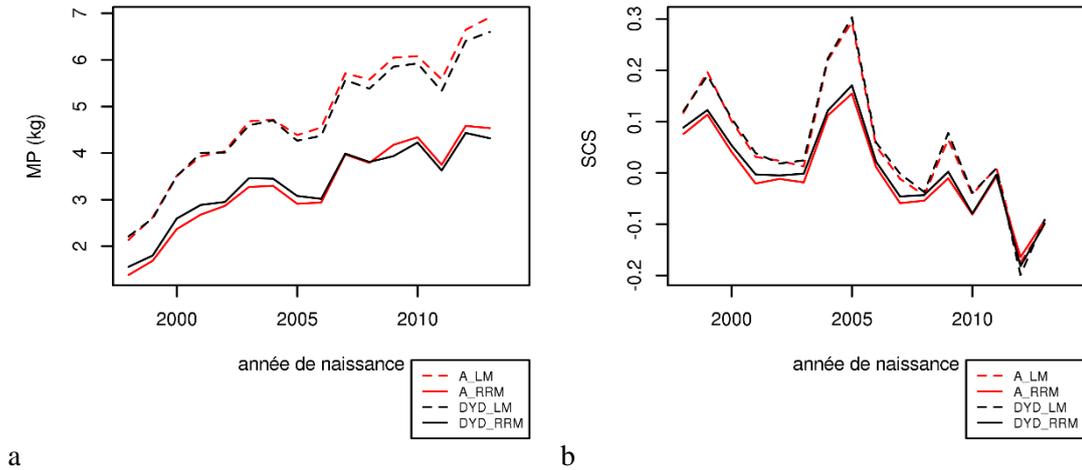
Le modèle 1 permet de montrer que le modèle LM surévalue les index, comme cela a déjà été observé au chapitre 4, pour les quantités de lait et de matières dans les deux races. C'est également le cas pour les SCS. Par contre pour le ratio TB/TP, le modèle RRM présente des index supérieurs à ceux du modèle LM.

**Tableau 17 : Coefficients de régression et estimés de l'effet « modèle » du modèle 1 pour les races Saanen et Alpine**

		<b>Coefficient régression (an)</b>	<b>Estimé de l'effet modèle RRM (LM en référence à 0)</b>
<b>Saanen</b>	<b>Lait</b>	6.3	-26.17
	<b>MG</b>	0.31	-0.47
	<b>MP</b>	0.29	-0.62
	<b>TB</b>	0.07	0.11
	<b>TP</b>	0.09	0.17
	<b>SCS</b>	-0.01	-0.07
	<b>Ratio TB/TP</b>	-0.06	0.9
<b>Alpine</b>	<b>Lait</b>	5.3	-42.44
	<b>MG</b>	0.32	-0.78
	<b>MP</b>	0.27	-1.49
	<b>TB</b>	0.11	0.16
	<b>TP</b>	0.1	0.12
	<b>SCS</b>	-0.01	-0.04
	<b>Ratio TB/TP</b>	0.03	1.56

*En gras les coefficients significativement différents de 0*

La Figure 56 présente les moyennes des index et des DYD pour la quantité à la lactation de MP et de SCS par année de naissance des boucs Alpains génotypés, obtenus selon les modèles A\_RRM et A\_LM. Afin de rendre comparable les index des boucs d'une année à l'autre, nous les avons exprimés par rapport à une sous population de référence constituées des boucs nés en 1997 avec plus de 25 filles en multipares. Nous retrouvons des résultats comparables à ceux obtenus en Saanen (« Supplemental Figure S8 », chapitre 4). Les DYD et les index sont très proches intra type de modèle. Les niveaux génétiques sont systématiquement supérieurs avec le modèle LM et l'écart entre les 2 modèle se creuse, ce qui surévalue le progrès génétique avec la modèle LM par rapport à celui observé avec le modèle RRM pour le caractère MP en Alpine. Pour les SCS, les différences entre modèles sont beaucoup plus faibles comme l'indiquait au Tableau 18 l'estimation de l'effet « modèle » pour ce caractère de -0.04 (p-value de la différence de 0.1).



**Figure 56: Evolution des valeurs génétiques (A) et des DYD selon le modèle utilisé LM et RRM pour la MP (a) et les SCS (b) à la lactation selon l'année de naissance des boucs en race Alpine**

*A\_LM* : index du modèle à la lactation, *A\_RRM* : index du modèle RRM, *DYD\_LM* : DYD obtenus avec le modèle à la lactation, *DYD\_RRM* : DYD obtenus avec le modèle à la lactation.

Nous retenons qu'il y a bien un progrès génétique quel que soit le modèle utilisé, et que les différences entre les modèles sont visibles lorsqu'on trace les courbes de progrès génétique. La principale différence réside dans les différences de moyennes entre LM et RRM.

## II. L'évaluation génétique de la persistance par le modèle RRM

Le principal intérêt du modèle RRM est d'estimer une valeur génétique pour la persistance laitière. Il permet de décorrélérer le niveau de production global à la lactation de la persistance laitière grâce à une décomposition en valeurs propres de la matrice de variances-covariances génétique (Van der Werf et al., 1998; Druet et al., 2003). En effet, la première valeur propre représente entre 83% et 92% de la variance originelle (chapitre 2 : Additional file 1) et représente le niveau global à la lactation, tandis que la deuxième valeur propre, qui représente entre 4 et 12% de la variance originelle, représente la persistance. La persistance a été définie comme la capacité à produire moins de lait au début de la lactation et plus à la fin par rapport à un animal moyen qui produit la même quantité totale de lait lors de sa lactation. La troisième valeur propre qui représente entre 1 et 4% de la variance originelle oppose la production du début et de la fin de lactation avec celle du milieu de lactation. Nous verrons par la suite que cette troisième composante aurait été intéressante à sélectionner si elle avait représenté plus de part de variance, car la production laitière d'une chèvre en milieu de lactation n'est alors pas en concurrence avec la reproduction ou la longévité.

Selon le caractère, la deuxième valeur propre était: plus importante pour les quantités que pour les taux (chapitre 3). Ces propriétés propres à chaque caractère peuvent aussi être illustrées par les corrélations génétiques entre DIM. En effet, pour les caractères de quantité, la corrélation était plus faible entre le pic de lactation et la fin de lactation que pour les taux.

La différence de production en début et fin de lactation due à la persistance n'est pas négligeable comparée au gain quotidien dû au niveau. En effet, nous pouvons voir à la Figure 5 de l'article du Chapitre 2 ainsi que dans le Chapitre 3 (Figure 31), que le déficit de production en début de lactation est aussi important que le gain de production par la première fonction propre. Il en est de même en fin de lactation. En d'autres termes, comme nous l'avons vu au chapitre 3, les chèvres les plus persistantes produisent en début de lactation autant que les chèvres ayant une persistance moyenne mais de faible niveau de production, et en fin de lactation elles produisent autant que les chèvres de persistance moyenne mais qui ont un très bon niveau de production laitière.

L'étude d'association pangénomique a permis de mettre en évidence des zones du génome, associées à la persistance de plusieurs caractères (voir chapitre 4). Ces régions étaient pour la plupart dans des zones elles-mêmes associées au niveau de production à la lactation. Il serait intéressant d'aller plus loin en étudiant plus précisément ces zones du génome qui influencent le déterminisme de cette persistance. De plus, il pourrait être intéressant de découper la lactation en plusieurs périodes et de regarder si certains gènes sont associés seulement à une seule période (Oliveira et al.; Lu and Bovenhuis, 2019).

Nous avons voulu également comparer le progrès génétique obtenu pour le niveau et la persistance en primipares et multipares. Pour cela, nous avons utilisé les valeurs génétiques des boucs génotypés utilisés en insémination artificielle, nés après 1997 et avant 2014, soit 430 boucs en Saanen et 551 boucs en Alpine, obtenus à partir du jeu de données complet (lactations de 2000 à 2017). Le modèle RRM utilisé est l'A\_RRM. Un modèle linéaire (modèle 2) a été utilisé afin d'évaluer le progrès génétique annuel pour le niveau et la persistance, et les différences entre primipares et multipares.

$$y_{ijk} = a_k x_i + p_{kj} + e_{ijk} \text{ (Modèle 2)}$$

avec  $y_{ijk}$  l'index pour le caractère  $k$  (niveau ou persistance) calculé pour la parité  $j$  (2 classes : primipare, multipares), pour le bouc  $i$ ,  $a_k$  l'effet de l'année de naissance,  $x_i$  l'année de naissance du bouc  $i$ ,  $p_{kj}$  l'effet de la parité, et  $e_{ijk}$  les résidus du modèle. L'interaction entre l'année de naissance et la parité a été testée mais n'était pas significative ni pour le niveau ni pour la persistance, il n'y a donc pas de différence de vitesse de progrès génétique entre primipares et multipares.

Le Tableau 18 présente les coefficients de régression du progrès génétique obtenu pour le niveau et la persistance. Le progrès génétique pour le niveau de production, est significatif (coefficient de régression positifs et p-values inférieures à 0) pour les mêmes caractères qu'à la lactation (quantités de lait et de matières, taux et SCS pour les deux races) (Tableau 17). Ce résultat semble logique étant donné les corrélations de 1 estimées entre valeurs génétiques pour la quantité totale à la lactation et pour la première composante du modèle RRM. Les valeurs génétiques primipares et multipares sont peu différentes, on note cependant en Saanen des valeurs génétiques en primipares inférieures pour le TP, et

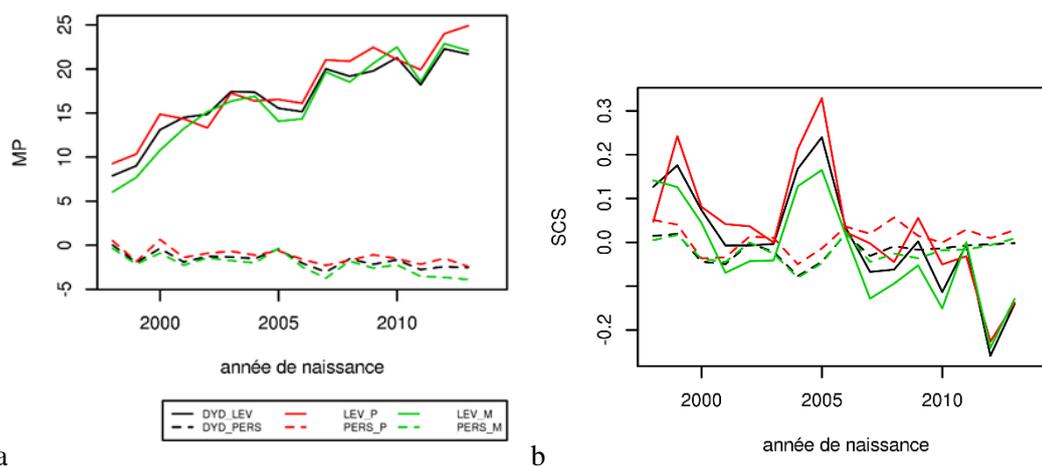
les SCS et supérieures pour le ratio, et en race Alpine des valeurs génétiques primipares supérieures pour le lait, la MP et les SCS.

Pour la persistance, il existe des différences de moyennes entre primipares et multipares pour certains caractères: les primipares ont des niveaux génétiques plus élevés que les multipares pour la MP et les SCS en Alpine et pour la MG pour les chèvres de race Saanen et pour d'autres caractères les multipares ont des niveaux génétiques plus élevés que les primipares TB en Saanen et en Alpine et le ratio TB/TP en Alpine. Les coefficients de régression sur l'année de naissance sont proches de 0 pour la plupart des caractères, ce qui montre qu'il n'y a pas eu d'évolution génétique de la persistance excepté pour le TB et le ratio qui a augmentée au cours du temps dans les deux races. De plus, le progrès génétique de la persistance du lait et de la MP en race Alpine a légèrement diminué comme nous pouvons le voir sur la Figure 57. Cette figure présente les moyennes des valeurs génétiques des boucs Alps selon leur année de naissance pour le niveau et la persistance en primipares et multipares, pour la MP (Figure 57 a) et les SCS (Figure 57 b).

**Tableau 18 : Coefficients de régression et estimés de l'effet parité du modèle 2 pour le niveau et la persistance en Saanen et en Alpine**

		Niveau		Persistance	
		coefficient de régression (année de naissance)	Coefficient des primipares (multipares en référence à 0)s	coefficient de régression (année de naissance)	coefficient des primipares (multipares en référence à 0)
<b>Saanen</b>	<b>Lait</b>	0.29	0.35	0	0.03
	<b>MG</b>	1.47	1.63	0.06	1.34
	<b>MP</b>	1.32	0.16	-0.03	0.15
	<b>TB</b>	1.21	3.63	0.17	-2.29
	<b>TP</b>	1.32	-4.13	-0.01	-0.52
	<b>SCS</b>	-0.01	-0.07	0	0
	<b>Ratio TB/TP</b>	-0.01	2.36	0.04	0.18
<b>Alpine</b>	<b>Lait</b>	0.21	0.43	-0.03	-0.02
	<b>MG</b>	1.48	0.11	0.02	0.44
	<b>MP</b>	1.1	1.21	-0.14	0.89
	<b>TB</b>	1.74	-0.51	0.3	-3.52
	<b>TP</b>	1.46	-2.45	-0.01	0.19
	<b>SCS</b>	-0.02	0.06	0	0.03
	<b>Ratio TB/TP</b>	0.12	0.44	0.11	-0.76

*En gras les coefficients significativement différents de 0*



**Figure 57: Evolution de la moyenne des valeurs génétiques et des DYD pour la MP (a) et les SCS selon l'année de naissance des boucs en race Alpine pour les primipares (P) et les multipares (M), pour le niveau (LEV) et la persistance (PERS) (index obtenus avec le jeu de données complet)**

Nous avons étudié les corrélations entre index officiels actuels et les index produits par RRM, malgré le fait que les index officiels actuels soient calculés à partir d'un historique de données plus important. Le Tableau 19 présente les corrélations entre index. Comme attendu les corrélations entre index officiels actuels à la lactation et les index produits par notre RRM pour le niveau sont très élevées. Ces corrélations sont plus importantes avec les index primipares qu'avec les index multipares pour les quantités de lait et de matières (+0.92 entre l'index officiel et le niveau en primipare, +0.87 entre l'index officiel et le niveau en multipares pour la quantité de lait en Saanen). De plus, les corrélations avec les persistances sont faibles comme attendues (décorrélation entre le niveau et la persistance). En revanche, la persistance pour le lait et la MP est légèrement corrélée négativement pour les primipares (-0.08 et -0.17 respectivement) comme pour les multipares (-0.17 et -0.23 respectivement) en Alpine. La sélection actuelle étant basée sur ces index officiels, ce résultat peut expliquer la baisse du niveau génétique de la persistance au fil des ans, observée pour la MP dans cette race à la Figure 57 a et quantifiée au Tableau 18. Les corrélations entre les index du RRM et l'index officiel pour les SCS sont négatives car le mode d'expression de l'index officiel est inversé : un index officiel élevé traduit un faible niveau de SCS. Pour ce caractère, l'index officiel actuel est fortement corrélé à l'index niveau du modèle RRM (de -0.92 à -0.96) légèrement plus en primipares qu'en multipares mais la différence est peu marquée. Par contre, on observe une corrélation modérée avec la persistance et en sens inverse entre primipares et multipares.

L'index actuel est plus proche d'un index niveau en primipare. Ce caractère a été largement amélioré au cours des dernières années. La persistance n'a pas beaucoup évolué, mis à part en Alpine pour la MP et le lait où elle a légèrement diminué. Le modèle de RRM offre la possibilité de sélectionner les animaux pour le niveau et la persistance en multipares de manière décorrélée, si ces valeurs génétiques sont des caractères intéressants à améliorer avec le futur objectif de sélection.

**Tableau 19 :Corrélations entre l'index officiel actuel et les index niveau et persistance primipares et multipares estimés avec le modèle RRM pour chaque caractère de production**

		Saanen						Alpine					
		LAIT	MG	MP	TB	TP	SCS	LAIT	MG	MP	TB	TP	SCS
Primipares	Niveau	0.92	0.93	0.93	0.98	0.98	-0.94	0.92	0.93	0.91	0.99	0.97	-0.96
	Persistance	-0.01	0.14	0.08	-0.15	-0.21	-0.24	-0.08	0.01	-0.17	0.01	-0.11	-0.17
Multipares	Niveau	0.87	0.90	0.89	0.99	0.96	-0.92	0.91	0.93	0.88	0.99	0.98	-0.94
	Persistance	-0.17	-0.08	-0.07	-0.09	-0.28	0.22	-0.16	-0.06	-0.23	0.08	-0.08	0.21

### III. Utilisation des valeurs génétiques produites par le RRM en sélection

Les corrélations génétiques entre caractères sont importantes en sélection animale car elles nous permettent d'apprécier la capacité d'évolution de plusieurs caractères sélectionnés ensemble. En effet, des caractères qui sont corrélés positivement ou décorrélés sont faciles à améliorer ensemble, en revanche des caractères corrélés négativement sont plus difficiles à sélectionner ensemble. Il est également important de connaître les corrélations génétiques entre les caractères sélectionnés et les caractères non sélectionnés, mais qui sont également importants en élevage, afin de comprendre qu'elles seront leur évolution au cours des ans. Si ces caractères non sélectionnés ont des corrélations négatives avec les index sélectionnés et sont dégradés, alors il est important de les intégrer à l'index synthétique de sélection global afin de stopper leur dégradation.

A partir du jeu de données utilisé au chapitre 4 qui comprend les lactations de 2000 à 2017, nous avons calculé les corrélations entre index des boucs d'insémination artificielle génotypés nés entre 1998 et 2013 et ayant plus de 25 filles en multipares, soit 389 boucs Saanen et 529 boucs Alpine. Ces index provenaient de l'évaluation officielle, et du modèle A\_RRM, ainsi que d'indexations expérimentales pour la réussite à l'insémination artificielle (Idele, V. Clément, action innovante FGE : CAPFER) dans chacune des 2 races et pour la longévité (INRA, I. Palhière, projet CASDAR Rustic) en race Alpine. Ces corrélations sont présentées à la Figure 58 pour les Saanen et à la Figure 59 pour les Alpines. Elles approchent les corrélations génétiques car les index des boucs sont précis avec un nombre important de filles multipares dans différents élevages en Alpine de 26 à 1459 filles en multipares et en Saanen de 26 à 1346 filles en multipares). Les caractères ont été regroupés en utilisant un regroupement hiérarchique à l'aide du package R corrplot (Wei et al., 2017).

#### 1. Corrélations entre index de production et de SCS

Pour les caractères de production et les SCS, nous avons mis en évidence des corrélations positives élevées entre :

- les niveaux des quantités de lait et de matières pour les primipares, les niveaux des quantités pour les multipares et entre niveaux des primipares et multipares en Saanen et en Alpine (entre 0.5 et 0.9),

- les persistances des quantités pour les primipares, les persistances des quantités pour les multipares et entre persistances des primipares et multipares en Saanen et en Alpine (entre 0.4 et 1),
- niveaux du TB et du ratio TB/TP en primipares, en multipares et entre primipares et multipares en Saanen et en Alpine (entre 0.7 et 0.8),
- persistance du TB et du ratio en primipares, en multipares et entre primipares et multipares en Saanen et en Alpine (entre 0.5 et 0.6),
- persistance des SCS et du TP en Saanen entre primipares, entre multipares et entre primipares et multipares (entre 0.2 et 0.4),
- TB et TP pour le niveau en primipares, en multipares et entre primipares et multipares en Saanen et en Alpine (0.5),
- MG et TB et le ratio pour les niveaux en primipares, en multipares et entre primipares et multipares en Saanen et en Alpine (entre 0.3 et 0.6),
- MG et TB et le ratio pour les persistances en primipares, en multipares et entre primipares et multipares en Saanen et en Alpine (entre 0.2 et 0.7),

Des corrélations négatives élevées ont été trouvées entre :

- les niveaux du TP et les niveaux des quantités en primipares, en multipares et entre primipares et multipares en Saanen et en Alpine (entre -0.6 et 0),
- les persistances du TP et les persistances des quantités pour les niveaux en primipares, en multipares et entre primipares et multipares en Saanen et en Alpine (entre -0.2 et -0.7),
- les niveaux du TP et les niveaux du ratio TB/TP en primipares, en multipares et entre primipares et multipares en Saanen et en Alpine (-0.2),
- les persistances du TP et les persistances du ratio TB/TP pour les niveaux en primipares, en multipares et entre primipares et multipares en Saanen et en Alpine (entre -0.4),
- le niveau du TB (et du ratio TB/TP) et le niveau du lait pour les niveaux, en multipares et entre primipares et multipares en Saanen et en Alpine (entre -0.5 et -0.4).
- les corrélations entre niveau de SCS et persistance des quantités en multipares en Saanen (-0.4 en Saanen)

Les principales différences entre Alpine et Saanen résident dans les corrélations entre :

- les niveaux des quantités et les niveaux des SCS (entre -0.3 et +0.3 en Saanen et de -0.4 à 0 en Alpine),
- les persistances des SCS et les persistances du TP (entre 0.2 et 0.4 en Saanen et entre 0.1 et 0.2 en Alpine),
- les index persistance pour les quantités entre les primipares et les multipares (moins élevées en Saanen entre 0.4 et 0.6, qu'en Alpine entre 0.6 et 0.8).

- les corrélations entre niveau de SCS et persistance des quantités en multipares (-0.4 en Saanen et entre -0.2 et -0.1 en Alpine)

Afin d'illustrer ces corrélations, nous avons calculé les corrélations entre index journaliers. La Figure 60 présente les corrélations entre valeurs génétiques quotidiennes des mêmes boucs en race Saanen (Figure 60 a) et Alpine (Figure 60 b). Les corrélations entre caractères sont en général relativement constantes au cours de la lactation comme celles entre le TB et le ratio TB/TP. Certaines ne le sont pas, et nous avons retrouvés la corrélation positive entre la persistance des SCS et la persistance du TP pour les multipares (+0.2 en Alpine, +0.4 en Saanen), la corrélation négative entre persistance laitière et la persistance des SCS (pour les multipares -0.3 en Alpine, -0.4 en Saanen) ainsi que persistance laitière et niveau des SCS en multipares (-0.4 en Saanen), et la corrélation entre la persistance du ratio et la persistance laitière (pour les multipares +0.5 dans les deux races).

Les corrélations entre niveaux sont proches des corrélations génétiques estimées avec un modèle à la lactation (Bélichon et al., 1999; Rupp et al., 2011). Nous retrouvons les corrélations élevées entre Lait-MG-MP, TB-TB, MG-TB et les corrélations négatives TP et quantités (Lait, MG, MP) et TB et lait. Avec le rapport TB/TP, nous observons des corrélations positives fortes avec le TB, fortes et négatives avec le TP. Le modèle RRM a pour avantage de montrer les évolutions de ces corrélations au cours de la lactation. Une des évolutions les plus importantes est celle entre les SCS et le lait qui est positive ou proche de 0 en primipares et qui décroît jusqu'à -0.5 en multipares pour la race Saanen. La raison de cette chute pourrait être que des animaux qui tombent malades en début de lactation et qui ne guérissent pas, gardent un SCS élevé et une production laitière faible due à la maladie. Mais cette évolution n'est pas la même en primipares et en multipares ce qui reste à comprendre.

## 2. Corrélations entre index de production, SCS et fertilité

Les corrélations entre index niveau et persistance des différents caractères de production et de SCS avec le modèle RRM et l'index de la réussite à l'insémination artificielle (ou fertilité) sont reportées aux Figure 58 et Figure 59. Les corrélations les plus élevées avec l'index fertilité sont celles avec les index des quantités (-0.3 avec l'index niveau du lait en primipares, et avec l'index persistance en primipares pour la MP en Saanen et +0.3 pour le niveau du TB en Alpine) ce qui signifie qu'une production laitière importante est associée génétiquement à de l'infertilité. Ces corrélations sont plus faibles en multipares (-0.1) qu'en primipares (-0.3) en Saanen. En Alpine, la corrélation est de -0.2 avec la quantité de lait pour toutes les parités. Pour le TB, il existe une corrélation d'environ +0.3 entre le TB et la fertilité pour les primipares et les multipares en Alpine et proche de 0 en Saanen. Pour la persistance des quantités, les corrélations sont négatives entre la persistance et la fertilité en particulier en primipares (entre -0.3 et -0.2 en Saanen et -0.1 en Alpine). Ces corrélations sont plus faibles mais restent négatives pour les multipares (entre -0.1 et 0). Les chèvres semblent avoir des difficultés à combiner production et fertilité en primipares quand elles sont encore en pleine croissance. La plus grande difficulté à se

reproduire en primipares a été observée dans une étude précédente avec des pourcentages de réussite à l'IA plus faibles pour les primipares comparés aux multipares (Idele, 2013).

La Figure 61 présente les corrélations entre index quotidiens et la fertilité à l'IA. Nous voyons pour les deux races que l'antagonisme génétique entre la production et fertilité est le plus marqué au moment de la reproduction (autour de 210 jours de lactation).

La persistance est plus problématique en primipares qu'en multipares pour les deux races car en primipares la croissance est en compétition énergétique avec les fonctions de production et reproduction. Ces corrélations négatives soulignent l'importance de suivre l'évolution de cet index persistance si l'objectif pour l'éleveur est d'avoir une mise bas par an. Ce type de corrélation n'est pas étudiée en bovins laitiers car la reproduction a lieu environ 3 mois après la mise bas, et donc en début de lactation (corrélation de -0.24 entre l'index fertilité et l'index production laitière pour le jour 60 en bovin lait) (Strucken et al., 2012).

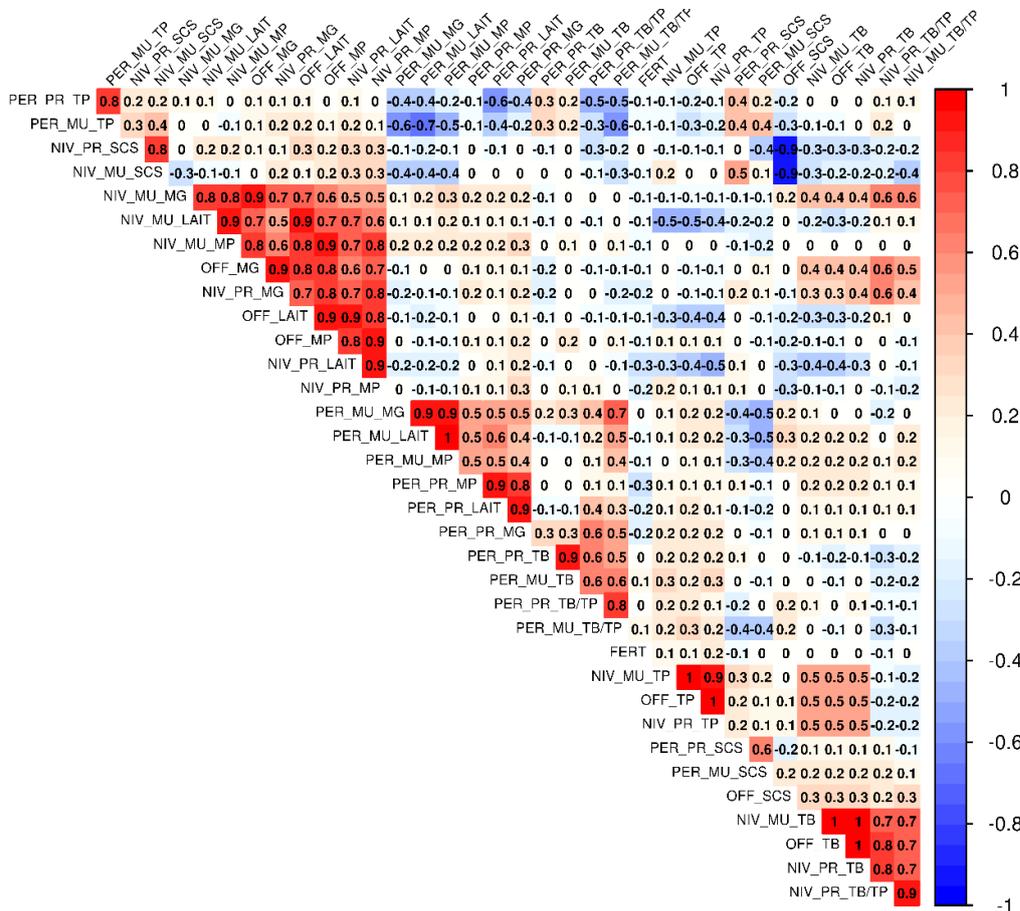


Figure 58: Corrélations entre index des 389 boucs Saanen génotypés nés entre 1998 et 2013 avec plus de 25 filles multipares (pour le niveau (NIV) ou la persistance (PER), en primipares (PR) ou multipares (MU), ou officiels (OFF), pour les caractères de production et de SCS, et pour la fertilité (FERT)).

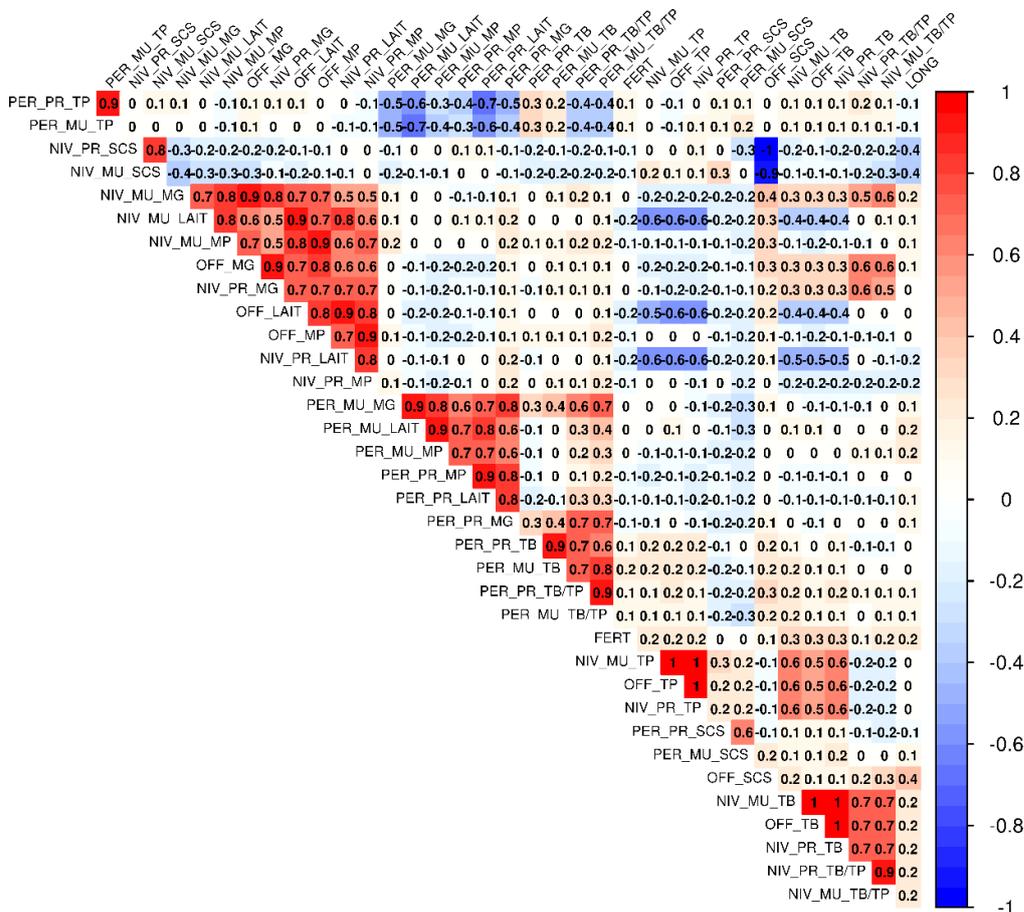
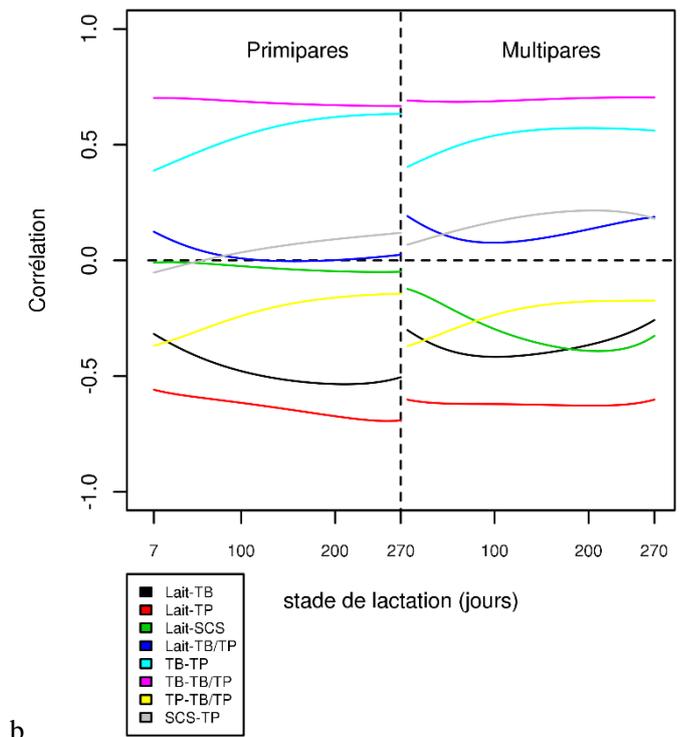
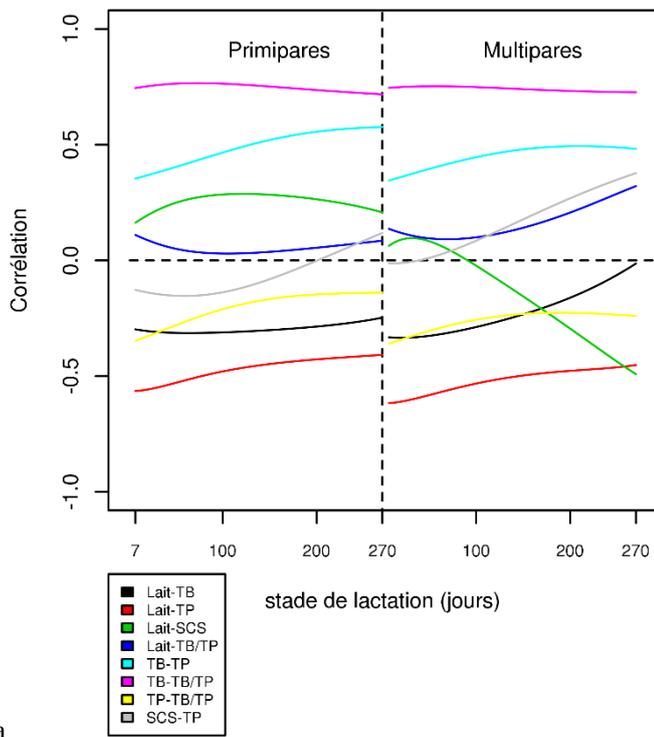


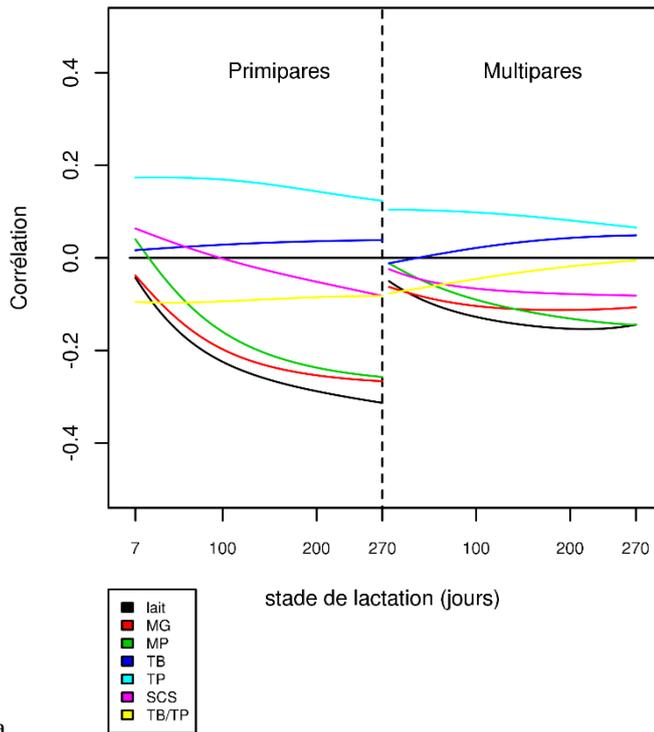
Figure 59: Corrélations entre index des 529 boucs Alpine nés entre 1998 et 2013 avec plus de 25 filles en multipares (pour le niveau (NIV) ou la persistance (PER), en primipares (PR) ou multipares (MU), ou officiels (OFF), pour les caractères de production et de SCS, pour la fertilité (FERT), et pour la longévité (LONG)).



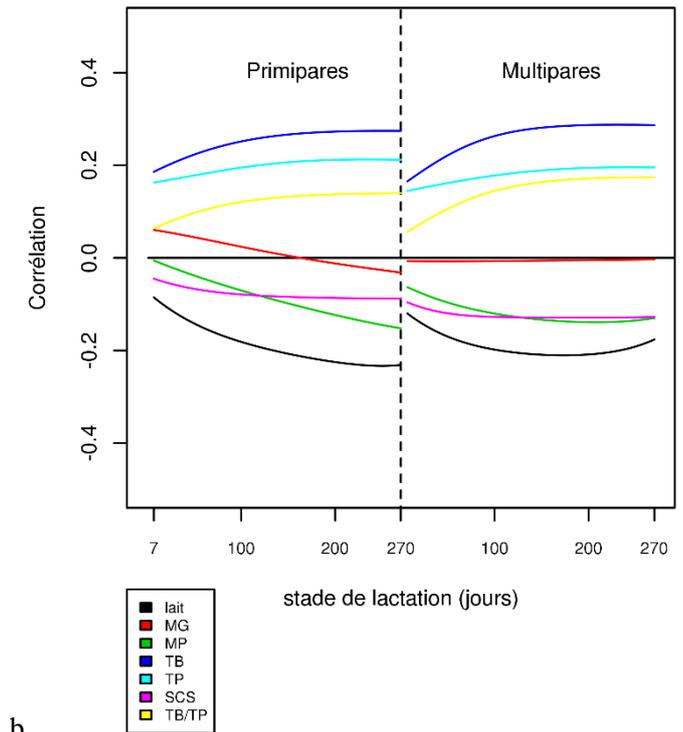
a

b

**Figure 60 : Evolution des corrélations en fonction du stade de lactation : (a) entre les index de 389 boucs Saanen génotypés nés entre 1998 et 2013 avec plus de 25 filles multipares et (b) entre les index de 529 boucs Alpine nés entre 1998 et 2013 avec plus de 25 filles en primipares**



a



b

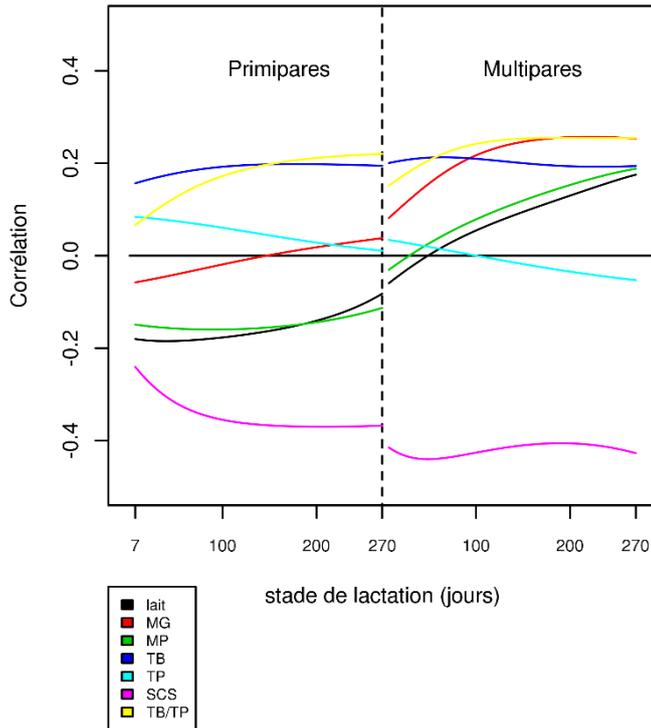
**Figure 61 Evolutions des corrélations entre index journaliers de production et de fertilité à l'IA en fonction du stade de lactation : (a) de 389 boucs Saanen génotypés nés entre 1998 et 2013 avec plus de 25 filles multipares et (b) de 529 boucs Alpine nés entre 1998 et 2013 avec plus de 25 filles en primipares**

### 3. Corrélations entre index de production, SCS et longévité

Les corrélations entre l'index longévité et les index de production sont reportées à la Figure 59 en race Alpine. La corrélation la plus forte avec l'index longévité a été observée avec les index officiels à la lactation, les index niveau et persistance du RRM pour les SCS (-0.4), i.e. plus une chèvre produit de SCS moins elle reste longtemps dans le troupeau. La deuxième corrélation la plus importante est celle avec les quantités de lait et de MP en primipares (-0.2), une chèvre qui produit beaucoup en première lactation restera moins longtemps dans le troupeau. En revanche, pour les multipares cette corrélation est positive autour de +0.1. On trouve également une corrélation de même ordre entre index longévité et ratio TB/TP (+0.2). Plus le ratio est important plus la chèvre restera dans le troupeau. Cette corrélation est peut-être le reflet de la corrélation entre SCS et le ratio TB/TP : lorsqu'une chèvre a beaucoup de SCS de façon récurrente, son ratio TB/TP est plus faible (corrélation de -0.3 entre niveau de SCS et niveau de TB/TP). Windig et al. (2005) explique que le ratio peut être soit élevé soit faible quand il y a un pic de SCS. En effet, les auteurs distinguent une infection contagieuse qui dure longtemps et qui est associée à un ratio TB/TP faible, d'une infection non contagieuse qui dure moins longtemps et qui est associée à un ratio TB/TP plus élevé. Pour les taux, il existe une corrélation de +0.2 entre le TB et la longévité pour les primipares et les multipares, et une corrélation de 0 pour le TP.

Les corrélations de la longévité avec la persistance sont relativement importantes : +0.2 pour la persistance du lait et de la MP en multipares et de +0.1 pour la persistance de la MG et du lait en primipares. La corrélation entre la fertilité et la longévité est de +0.23.

La Figure 62 présente l'évolution des corrélations entre index quotidiens de production et SCS avec l'index longévité. Nous voyons que contrairement à la fertilité, c'est le lait en début de lactation qui est le plus fortement corrélé défavorablement à la longévité. En revanche, une production élevée en multipares en fin de lactation est corrélée de façon positive avec la longévité ce qui est à l'opposé de la fertilité. La corrélation positive entre longévité et production laitière en fin de lactation illustre la corrélation positive entre longévité et persistance laitière (+0.2 Figure 59) et est un argument en faveur de l'incorporation de la persistance laitière dans l'objectif de sélection. Une quantité de lait importante au pic, entraîne un déficit énergétique qui entraîne une baisse du système immunitaire et donc l'apparition de problèmes de santé (Oltenucu and Algers, 2005). Il est intéressant de noter que la production de gras (TB, MG) est corrélée plus positivement à la longévité que la production de protéines. L'intégration des SCS dans l'objectif de sélection semble primordial afin d'augmenter la longévité.



**Figure 62 : Corrélations entre index de la longévité et les index quotidiens des autres caractères de production en fonction du stade de lactation de 529 boucs Alpine nés entre 1998 et 2013 avec plus de 25 filles en primipares**

#### 4. Index issus du RRM et objectifs de sélection

La dernière étude du chapitre 3 (V.5.) a montré que les chèvres qui produisent le plus en primipares ne sont pas les chèvres qui produisent le plus dans les lactations suivantes et qui ont la meilleure longévité. Elle nous a aussi montré que les chèvres qui produisent le plus en multipares ont une meilleure longévité, et une meilleure production totale au cours de leur carrière. Comme nous l'avons vu dès le chapitre 1, la production augmente d'une lactation à l'autre jusqu'en 5<sup>ème</sup> lactation. Les éleveurs ont donc intérêt à conserver leurs animaux. De plus, les jeunes animaux sont plus sujets à des problèmes de mise bas et les animaux plus âgés ont des capacités corporelles supérieures et donc peuvent valoriser plus de fourrages grossiers. Cette différence entre primipares et multipares semble indiquer un compromis entre croissance, production et longévité qui est bien documenté en bovins laitiers (Knaus, 2009). Ces résultats nous inciteraient à pondérer dans l'objectif de sélection plus faiblement l'index niveau des primipares par rapport à l'index des multipares. Il est probable que la profession préférera pour l'objectif de sélection une seule valeur génétique pour la quantité de lait. Dans ce cas, un index combinant les valeurs génétiques primipares et multipares devra être mis au point. Haiger (2018) rapporte une étude sur les bovins où les index de production ont été pondérés par parité : avec un poids de 0.25 en primipare, 0.30 en seconde lactation et 0.45 en troisième lactation parce que ces poids correspondent à l'augmentation de la performance des vaches qui ont de longues carrières, qui n'atteignent leur maximum qu'après la 5<sup>ème</sup> lactation. Ce type de poids appliqués aux index des chèvres

permettraient de réellement mettre en avant les troisièmes lactations, d'améliorer la longévité et donc aussi le bien-être animal (Oltenacu and Algers, 2005).

Enfin, la dernière étude du chapitre 3 (V.5.) a montré à nouveau le lien positif entre index persistance élevé et une meilleure longévité. Les éleveurs voulant limiter leur taux de renouvellement, qui est un objectif souhaitable pour limiter les coûts de production, pourraient sélectionner des chèvres avec une persistance laitière supérieure à la moyenne même si la sélection directe sur l'index longévité ou les SCS semble plus efficace.

Lorsque nous avons étudié les liens entre index génétiques et de l'environnement permanent avec la fertilité à l'IA, la survie et la durée de lactation (Chapitre 3 V.1.2.3.4), nous avons mis en évidence des liens entre caractères de production pour les index de niveau mais aussi pour les index de persistance. Pour les index niveau, cette étude nous a permis de mettre en évidence des relations entre la moins bonne survie et des index niveau SCS faibles, des index niveau élevés pour les quantités et une infertilité élevée, plus de lactations longues (surtout en primipare) et plus de survie (surtout en multipares). Pour les index persistance, elle a aussi permis de mettre en évidence des liens entre index de persistance laitière élevés et des lactations longues plus nombreuses, une survie améliorée et légèrement plus d'infertilité. Tous ces résultats ont été confirmés par les corrélations entre index de longévité et de production ainsi que par celles entre index de fertilité et ceux de production.

Ces liaisons entre fertilité, survie, durée de lactation qui sont trois paramètres importants de rentabilité d'un élevage et les persurances des caractères mettent, en valeur l'intérêt d'un modèle de régression aléatoire. Afin de caractériser plus finement les relations entre ces caractères, il aurait été intéressant d'utiliser le modèle EGV\_FT qui sépare L2 et L3 pour la partie génétique comme pour l'environnement permanent. Cette étude a montré que les index génétiques et d'environnement permanent issus d'un modèle RRM (niveau + persistance, lactations séparées) peuvent être des outils novateurs de caractérisation des animaux. Ces outils pourraient être utilisés notamment pour le suivi de chèvres particulières par exemple en les phénotypant précisément pour de nouveaux caractères comme l'efficacité alimentaire ou le bien être.

L'évaluation des persurances des différents caractères peut permettre la sélection d'individus ayant une forme désirée de courbe de production. Certains éleveurs peuvent souhaiter sélectionner un animal ayant une très grande persistance pour réaliser de très longues lactation jusqu'à l'extrême du modèle néerlandais où une chèvre peut ne mettre bas qu'une fois dans sa vie. Une étude récente sur les lactations longues des chèvres en France (Idele, R. de Crémoux, communication personnelle) montre que les chèvres en lactations longues n'ont pas de baisse de production moyenne comparée à des lactations classiques, et que des animaux sont capables de produire pendant 10 ans ce qui semble être une disposition propre de cette espèce. Sachant que la mise bas est à risque par rapport à la santé de l'animal (mise bas, délivrance, transition alimentaire, pic de lactation), qu'elle augmente la quantité de

travail (gestion des chevreaux et des mises bas), que les chevreaux ne représentent pas une grande valeur économique pour l'éleveur, qu'ils pourraient incarner une image négative pour la filière (devenir des chevreaux : conditions de vie, de transport...) et que la sélection actuelle continue à augmenter la quantité de lait qui est défavorable à la fertilité, l'utilisation des lactations longues peut devenir une pratique plébiscitée par les éleveurs. Dans des systèmes d'élevage basés sur la production herbagère du printemps, un animal ayant une persistance modérée peut être sélectionné afin de ne pas concurrencer la reproduction pour obtenir une mise bas par an, même si la relation entre persistance et infertilité n'est pas élevée. Dans ces systèmes des animaux avec des index intermédiaires pour le niveau devraient aussi être sélectionnés étant donné les corrélations négatives entre index lait élevés et fertilité. D'un autre point de vue, une production importante en fin de lactation impacterait aussi des éleveurs ayant la volonté de faire des lactations courtes et de ne jamais réaliser de lactations longues pour se libérer de la contrainte de la traite durant les mois de tarissement.

Dans le cadre du projet (CASDAR) « Rustic », l'impact économique d'un accroissement de la persistance laitière a été étudié et il est économiquement positif. Malheureusement, cette étude dans un premier temps n'a pris en compte qu'un changement du prix du lait dû à une production différée (par exemple plus de lait en fin d'année quand les prix sont au plus haut). Il n'est pas à exclure que le prix du lait diminuerait en fin d'année si la persistance de la population était améliorée. L'effet positif sur la longévité n'a pas été pris en compte dans cette étude.

Il serait intéressant de voir l'impact de l'augmentation de la persistance sur la réussite de la mise au tarissement. En effet, une production importante au moment du tarissement pourrait entraîner des difficultés (Strucken et al., 2015). L'éleveur par la conduite alimentaire peut également réduire de façon considérable la production laitière et pourrait peut-être avec une bonne préparation alimentaire réussir malgré tout le tarissement.

Il est important de rappeler que même si la persistance entraîne des différentiels de production importants en début et en fin de lactation et que ses corrélations avec la fertilité ou la longévité ne sont pas nulles, elle est moins liée à ces caractères que le niveau de production. C'est bien la somme des deux qui est négative ou positive car le niveau et la persistance peuvent s'annuler en début et en fin de lactation pour un animal donné.

Avant l'inclusion de la persistance dans l'index de synthèse, un travail plus approfondi des conséquences de la modification de la forme de la courbe de lactation doit être mené. En effet, la définition d'un objectif de sélection est lourde de conséquences et une analyse sur l'ensemble de la population doit être réalisée (Oltenucu and Algers, 2005).

Nous avons vu que les index officiels sont plus corrélés aux index niveau primipares alors que l'index niveau multipares est plus favorablement corrélé à la longévité. Nous avons également observé que la persistance associée à la longévité est corrélée négativement au niveau officiel et qu'elle est

dégradée au fil des ans en Alpine pour le lait et la MP. Ces deux observations laissent à penser qu'un modèle comme le modèle RRM que nous avons développé est pertinent car il évalue la persistance laitière de manière non corrélée au niveau et sépare l'index niveau pour les primipares et les multipares. La prise en compte dans un index global de sélection de l'index SCS, fertilité, longévité, persistance et niveau avec des pondérations en fonctions des rangs de lactation permettrait de sélectionner des animaux plus robustes, objectif recherché par la filière.

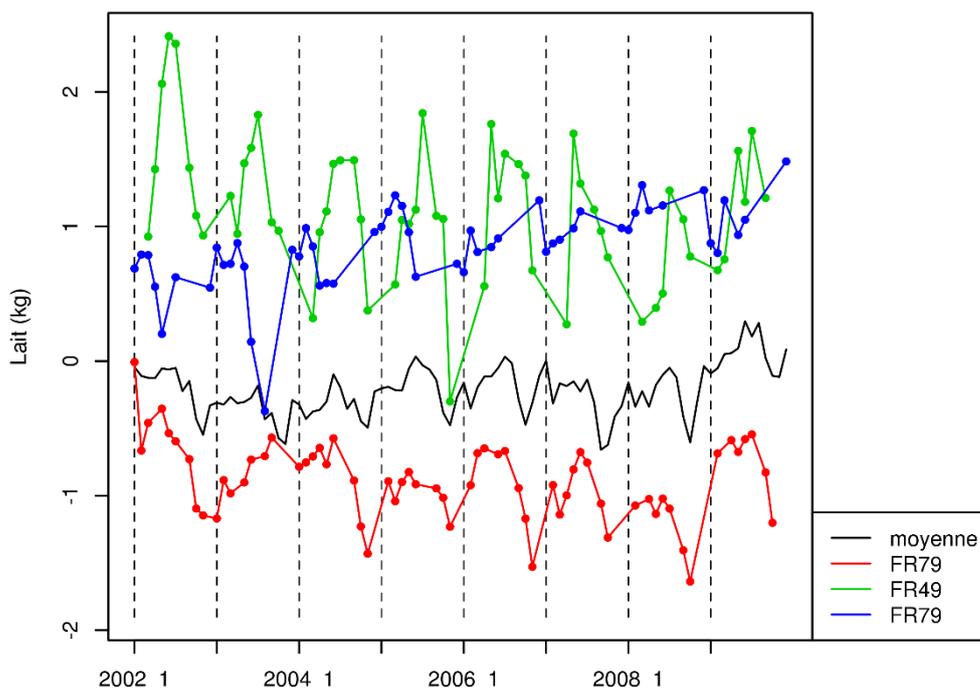
#### **IV. Utilisation des solutions des effets fixes produites par le RRM en appui technique**

L'effet troupeau-jour de contrôle (HTD) estimé par le modèle RRM est restitué aux éleveurs en bovins laitiers (Leclerc, 2008). Chaque éleveur peut ainsi suivre les variations de cet effet le concernant au cours de l'année. Il peut être intéressant de comparer cet effet année après année au sein de l'élevage mais aussi de comparer cet effet à d'autres élevages partageant le même système de production. Il peut être intéressant d'utiliser le modèle de Schaeffer du chapitre 2 qui utilise un effet aléatoire troupeau-année en fonction du stade de lactation afin d'obtenir l'effet troupeau année sur la forme de la courbe de lactation. L'éleveur pourrait ainsi voir l'effet de sa conduite sur la forme de la courbe de lactation. Là aussi une comparaison entre année et entre troupeaux partageant le même système peut être intéressante.

La Figure 63 présente une illustration de cet effet HTD pour 3 troupeaux suivis entre 2002 et 2009 ( Figure 63). Nous pouvons voir que le cheptel en rouge a globalement un effet troupeau faible par rapport à la moyenne, et qu'il a des variations interannuelles faibles et intersaison plus élevées que la moyenne. Le troupeau bleu a un effet troupeau élevé de faible variabilité interannuelle et intersaison. Le cheptel vert a un effet troupeau élevé, avec des variations interannuelles faibles mais des variations entre saisons très fortes.

Ces effets HTD peuvent aussi être utilisés afin de caractériser des élevages dans le cadre de travaux de recherche (Huquet et al., 2012).

Il a pu être noté au chapitre 3 des variations au cours de l'année de l'effet troupeau. Ces variations sont en accord avec celles observées au chapitre 1. Elles semblent liées à l'évolution de la photopériode au cours de l'année sous nos latitudes (Chemineau et al., 2007).



**Figure 63 : Evolution des effets HTD au cours des années, pour l'ensemble des troupeaux (moyenne) et pour 3 troupeaux différents**

Nous avons pu mettre en évidence des effets de l'âge à la mise bas, du mois de mise bas et de l'effet du tarissement et de la gestation sur l'ensemble des caractères étudiés. Ils peuvent être utilisés en zootechnie comme illustrations afin de conseiller les éleveurs sur les conduites à mettre en place.

Le modèle utilisé n'est pas le plus aisé pour tracer ces courbes car se mêlent des effets fixes non dépendant du stade de lactation en interaction avec l'année de lactation. L'effet du mois de mise bas, a une interaction avec le HTD qui peut « fausser » la forme de la courbe. Ces résultats sont donc à traiter avec une certaine prudence et peuvent nécessiter le développement d'un nouveau modèle de régression aléatoire adapté au conseil en élevage (alimentation, âge à la mise bas, durée de tarissement, ...).

Les effets fixes calculés peuvent être très utiles pour prédire avec précision la production à l'échelle d'un troupeau ou à celle d'une coopérative (Leclerc, 2008). La production caprine étant très saisonnée, la simulation d'un désaisonnement à l'échelle d'un troupeau ou d'une coopérative devrait être améliorée comparativement à un modèle à la lactation connaissant avec plus de précision les effets du mois à la mise bas sur la production laitière (quantité de lait mais aussi qualité (TP, TB)).

## V. Perspectives

Le modèle contrôles élémentaires doit être testé pour voir l'impact d'un allègement du contrôle laitier, qui permettrait de réduire les coûts de phénotypage, sur la prédiction des valeurs génétiques, en dégradant le nombre de contrôles. Aux vues des corrélations génétiques entre stades de lactation

(Chapitre 3), le nombre de contrôles pour les taux pourrait être plus facilement allégé que pour les quantités (Lait, MG, MP) (Schaeffer and Jamrozik, 2008). Il serait également intéressant d'étudier la qualité d'extrapolation du modèle RRM pour les lactations en cours comparativement au modèle à la lactation, utilisé actuellement.

De plus en plus d'animaux sont génotypés chaque année. En effet depuis 2018, tous les jeunes boucs candidats sont génotypés à la naissance et plus seulement les reproducteurs diffusés pour l'insémination artificielle. Il est possible que le génotypage femelle soit réalisé à l'avenir augmentant de façon conséquente le nombre d'animaux génotypés. Les précisions de l'évaluation génomique devraient être ainsi améliorées. En revanche, l'utilisation de modèles de régression aléatoire très coûteux en temps de calcul, comme nous l'avons vu précédemment, pourrait être freinée par le nombre croissant de génotypages. Le logiciel Blupf90 propose l'utilisation d'une méthode appelée APY (Misztal et al., 2014) qui permet de réduire par décomposition en valeurs propres la matrice génomique pour ne garder que les individus « cœur » du pedigree. Cette méthode permet de réduire considérablement la taille du système et elle pourra à terme être testée et envisagée sur les données caprines. Il est aussi possible d'utiliser un SNP-BLUP (Taskinen et al., 2017) dont la taille de la matrice de relation génomique a la taille du nombre de SNPs utilisés donc invariable par rapport au nombre d'animaux génotypés.

Le modèle qui a été développé en caprins laitiers dans cette thèse mais, peut sans difficultés, être adapté à d'autres espèces laitières comme les bovins ou les ovins. La principale difficulté chez les bovins est le nombre de contrôles élémentaires et le nombre de génotypages déjà présents dans la base de données mais des applications de SSGBLUP ont été faites au Canada (Oliveira et al., 2019) ou dans les pays Nordiques (Koivula et al., 2015) avec succès avec des jeux de données conséquents.

L'évolution au cours du temps de différents caractères peut être étudiée à l'aide de modèles de régression aléatoire comme la croissance des animaux d'élevage (Meyer, 2005), différentes concentrations de métabolites sanguins pour étudier l'efficacité alimentaire (Tran, 2018), la balance énergétique d'animaux (Buttchereit et al., 2011). Avec l'avènement de multiples capteurs (radars, distributeurs automatique de concentrés, distributeurs automatiques d'eau, peseurs automatiques, mesures d'émission de gaz à effet de serre...), de nouvelles données longitudinales seront disponibles pour des évaluations génétiques à l'avenir. La méthode de classification de courbes de lactation développée au premier chapitre a permis d'établir des profils d'évolution de note d'état corporel (Macé et al., 2019). Ces caractères peuvent être étudiés de manière simultanée dans un RRM multi caractères afin d'obtenir des corrélations génétiques au cours de la lactation. Si un tel modèle est trop compliqué à mettre en œuvre (temps de calcul) il est possible, comme dans le chapitre 3 de calculer des corrélations entre index pour différents stades de lactation.

Les modèles de régression aléatoire peuvent être étendus en tant que modèles de norme de réaction (Bohlouli et al., 2019). Pour les modèles de norme de réaction, la valeur génétique ne dépend

pas du temps comme pour les modèles de régression aléatoire mais peuvent dépendre de la température, l'altitude, le THI. Ainsi, les interactions entre ces variables et la génétique (interaction entre la génétique et l'environnement :  $G \times E$ ) peuvent être étudiées. Ces modèles permettent ainsi, en cas d'interactions, de sélectionner des individus qui ont de bonnes performances dans le milieu souhaité ou qui manifestent peu de changement selon les caractéristiques du milieu.



## Liste des tableaux

Tableau 1 : Corrélations génétiques entre les différents caractères laitiers pour les chèvres Alpine (triangulaire inférieure) et pour les chèvres Saanen (triangulaire supérieure) (Bélichon et al., 1999; Rupp et al., 2011).....	21
Tableau 2: La filière caprine française : quelques chiffres clés .....	26
Tableau 3 : Nombre d'animaux dans le pedigree, nombre de contrôles élémentaires (entre 7 et 270 jours de lactation), et nombre de lactations, pour les rangs de lactation inférieur à 4, entre 2000 et 2018....	28
Tableau 4 : Matrices de variance –covariance génétique, de l’environnement permanent, et de l’effet troupeau-année pour les différents modèles : Sch_or (modèle de Schaeffer), Sch_HY (modèle de Schaeffer comprenant l’effet troupeau-année), leg4 (modèle de l’article 2, avec un effet troupeau-jour de contrôle).....	99
Tableau 5 : Corrélations entre les coefficients de régression des covariables de Legendre pour les effets génétiques, de l’environnement permanent et du troupeau-année-saison des différents modèles : Sch_or (modèle de Schaeffer), Sch_HY (modèle de Schaeffer comprenant l’effet troupeau-année), leg4 (modèle de l’article 2, avec un effet troupeau-jour de contrôle) .....	100
Tableau 6: Variances résiduelles et BIC pour les différents modèles : : Sch_or (modèle de Schaeffer), Sch_HY (modèle de Schaeffer comprenant l’effet troupeau-année), leg4 (modèle de l’article 2, avec un effet troupeau-jour de contrôle).....	100
Tableau 7. Nombre de contrôles élémentaires et de lactations par parité pour la race Saanen .....	131
Tableau 8. Estimations des corrélations génétiques entre les coefficients de régression des polynômes de Legendre (a0, a1, a2 dans LEG_FT*) de la deuxième parité avec ceux de la troisième parité et des corrélations génétiques entre les coefficients de régression des deux premières fonctions propres (Niveau, Persistance dans EGV_FT**) de la deuxième parité avec ceux de la troisième parité pour la race Saanen.....	132
Tableau 9. Héritabilité de chaque composante pour les modèles EGV_FT et EGV_PM pour chaque caractère chez les chèvres Saanen .....	135
Tableau 10. Corrélations entre index des boucs (n = 486) selon les 4 modèles étudiés pour la quantité de lait des chèvres de race Saanen .....	139
Tableau 11. Corrélations d’index pour les boucs Saanen (n = 486) obtenus avec le modèle EGV_PM entre la production laitière et les autres caractères (MG, MP, TB, TP, SCS, ratio TB/TP) .....	140

Tableau 12 : BIC* du modèle LEG_PM selon différentes valeurs du coefficient multiplicateur testé et appliqué aux coefficients de Legendre de troisième lactation pour les caractères (LAIT, MG, MP, TB, TP, SCS, ratio TB/TP) en race Saanen et Alpine. ....	145
Tableau 13 : Répartition des pourcentages des chèvres Alpine et Saanen selon la durée de lactation, la survie et la fertilité à l'IA en fonction du rang de lactation.....	157
Tableau 14 : Répartition des pourcentages de chèvres selon la durée lactation (lactation courte ou lactation longue) par rang de lactation (L1, L2 et L3) et la longévité (chèvres sorties ou non de l'élevage) en race Alpine et Saanen .....	159
Tableau 15 : Répartition des pourcentages de chèvres selon la survie (chèvres sorties ou non de l'élevage) ou la durée de lactation (lactation courte ou lactation longue) et la fertilité par rang de lactation en race Alpine et Saanen.....	160
Tableau 16 : Temps de calcul et mémoire utilisée pour le ssGBLUP .....	237
Tableau 17 : Coefficients de régression et estimés de l'effet « modèle » du modèle 1 pour les races Saanen et Alpine.....	242
Tableau 18 : Coefficients de régression et estimés de l'effet parité du modèle 2 pour le niveau et la persistance en Saanen et en Alpine .....	245
Tableau 19 :Corrélations entre l'index officiel actuel et les index niveau et persistance primipares et multipares estimés avec le modèle RRM pour chaque caractère de production .....	247





## Liste des figures

Figure 1 : Zones de production de fromages AOP, et volumes de lait associés (a), et répartition des volumes de lait produit par régions administratives et type de destination (livré à des laiteries ou transformé à la ferme) (b) (Maigret, 2018b) .....	17
Figure 2: Evolution des quantités totales de lait produit et du prix du lait par mois en 2018 (Idele, 2019b) .....	18
Figure 3 : Schéma de sélection de l'organisme Capgenes (Capgenes, 2019) .....	20
Figure 4: Evolution du pourcentage de mères inconnues (a) et de celui des pères inconnus (b) au fil des années de naissance (Saanen et Alpines confondues) .....	22
Figure 5: Evolution de la quantité de lait moyenne à la lactation (a), du TB moyen à la lactation (b), du TP moyen à la lactation (c), du comptage cellulaire à la lactation (d), de l'intervalle mise-bas-mise bas moyen (e) et du rang de lactation moyen (f) .....	23
Figure 6: Evolution des réserves corporelles, de l'énergie provenant de l'alimentation, et de l'énergie nécessaire à la production laitière et aux fonctions vitales au cours de la lactation (Fagné, 2014; Chartier and Blain, 2018) .....	25
Figure 7 : Proportion représentée par les différents types de protocoles de contrôle laitier en 2016 pris en compte dans cette thèse .....	27
Figure 8 : Illustration de la méthode de Fleischmann sur une lactation (point rouge : mesure de la production réelle, barre noire : production moyenne des 2 contrôles encadrant la période).....	29
Figure 9: Répartition de l'âge à la mise bas pour les chèvres en première (a), deuxième (b) et troisième lactation (c), du mois de mise bas (d), de la durée de tarissement (e), et de la région (f) en 2016, Alpine et Saanen confondues. ....	33
Figure 10 : Représentation des 5 premières covariables de Legendre.....	39
Figure 11: Illustration des six covariables d'une spline naturelle cubique dont les nœuds (représentés par les traits pointillés jaunes) ont été positionnés aux stades de lactation suivants ( $t = 7, 20, 50, 110, 190, 270$ ) pour modéliser la forme de la courbe de lactation .....	40
Figure 12: Courbes de lactation phénotypiques des différents clusters selon le nombre de clusters : a) 3 clusters, b) 4 clusters, c) 5 clusters, d) 6 clusters, e) 7 clusters, f) 8 clusters. Le pourcentage de lactations représenté par chaque cluster est précisé en légende à côté de la couleur du cluster. ....	60
Figure 13: Evolution des courbes phénotypiques selon le rang de lactation et la campagne de mise bas pour la quantité de lait en Saanen.....	62

Figure 14: Evolution des courbes phénotypiques selon le rang de lactation et la campagne de mise bas pour la MP en Saanen.....	62
Figure 15: Evolution des courbes phénotypiques selon le rang de lactation et la campagne de mise bas pour le TP en Saanen.....	63
Figure 16: Evolution des courbes phénotypiques selon le rang de lactation et la campagne de mise bas pour le comptage cellulaire en Saanen .....	63
Figure 17: Evolution des courbes phénotypiques selon le rang de lactation et la campagne de mise bas pour le ratio TB/TP en Saanen .....	64
Figure 18 : Courbes d'évolution phénotypique moyenne en race Alpine et Saanen des caractères : a) lait, b) MG, c) MP, d) TB, e) TP, f) SCS, g) ratio TB/TP .....	66
Figure 19 : Distribution des productions de lait totales à la lactation calculées selon la méthode Fleischmann et avec l'utilisation de splines (modèle 1).....	94
Figure 20 : Evolution de l'héritabilité selon le stade de lactation pour la production laitière chez les chèvres de race Saanen en première lactation avec une réduction des de variances- covariances génétique et de l'environnement permanent : a) par le logiciel Wombat, b) hors du logiciel Wombat avec R. ...	95
Figure 21 : Evolution de l'héritabilité selon le stade de lactation pour le TB chez les primipares de race Saanen selon différents modèles .....	96
Figure 22: Evolution de la variance résiduelle estimée en fonction du stade de lactation pour les différents modèles pour la quantité de lait pour les chèvres Saanen. ....	133
Figure 23. Evolution des variances génétiques en fonction du stade de lactation pour les quatre modèles pour la quantité de lait chez les chèvres Saanen.....	134
Figure 24. Evolution des héritabilités au cours du temps pour les caractères de production laitière (Lait, MG, MP, TB et TP), les scores de cellules somatiques (SCS) et le ratio TB/TP avec le modèle LEG_FT* chez les chèvres Saanen. ....	136
Figure 25. Corrélations génétiques entre le stade de lactation 40 de la seconde parité et les autres stades de lactation selon les différents modèles (LEG_FT, EGV_FT, LEG_PM et EGV_PM) pour la quantité de lait des chèvres de race Saanen.....	137
Figure 26. Corrélations génétiques avec le modèle LEG_FT* entre le stade de lactation 40 en deuxième lactation et tous les stades de lactation pour les caractères étudiés (Lait, MG, MP, TB, TP, SCS et ratio TB/TP) pour les chèvres de race Saanen.....	138
Figure 27: Evolution de la variance génétique (a), de la variance de l'environnement permanent (b), de la variance résiduelle (c), et de l'héritabilité (d) de la production laitière en race Saanen selon le stade de lactation et la définition de l'effet troupeau dans un modèle LEG_FT. ....	142

Figure 28 : Corrélations génétiques pour la production laitière en race Saanen entre les coefficients de régression entre deux rangs de lactation : pour la première covariable de Legendre $\varphi(\mathbf{0}, \mathbf{t})(\mathbf{a})$ , et pour la deuxième covariable $\varphi(\mathbf{1}, \mathbf{t})(\mathbf{b})$ , en fonction de la définition de l'effet troupeau dans un modèle LEG_FT. ....	143
Figure 29: Part de variance génétique représentée par la première valeur propre (a) et la deuxième valeur propre (b) selon le rang de lactation et le caractère en race Alpine et Saanen .....	146
Figure 30: Profils des première et deuxième fonctions propres (+1 ou -1 écart type génétique du premier ou du deuxième coefficient de régression du modèle EGV_FT) selon le rang de lactation et la race (Alpine et Saanen) : a) lait, b) MG, c) MP, d) TB, e) TP, f) SCS, g) ratio TB/TP.....	149
Figure 31: Courbes de lactation phénotypiques moyennes d'animaux ayant des valeurs génétiques extrêmes ( $\pm 2$ écarts types génétiques pour le niveau et la persistance) en deuxième lactation avec le modèle EGV_FT en race Saanen pour le lait (a), la MP (b), le TP (c), et le ratio TB/TP (d). ....	150
Figure 32: Evolution de la moyenne des HTD par mois de lactation selon le caractère étudié ((a) Lait, (b) TP, (c) SCS et (d) ratio TB/TP) et le rang de lactation en race Saanen.....	153
Figure 33: Evolution des effets campagne de mise bas selon le caractère étudié ((a) Lait, (b) TP, (c) SCS et (d) ratio TB/TP) et le rang de lactation en race Saanen.....	153
Figure 34: Evolution des effet de l'âge à la mise bas pour les différents caractères étudiés ((a) Lait, (b) TP, (c) SCS et (d) ratio TB/TP) en L1 en race Saanen.....	154
Figure 35: Evolution des effets du mois de mise bas pour les différents caractères étudiés ((a) Lait, (b) TP, (c) SCS et (d) ratio TB/TP) en L1 en race Saanen.....	154
Figure 36: Evolution des effets de la durée de tarissement pour les différents caractères étudiés ((a) Lait, (b) TP, (c) SCS et (d) ratio TB/TP) en L2 en race Saanen .....	155
Figure 37: Courbes de production en fonction du stade de gestation pour les différents caractères en primipares et en multipares en Saanen : a) lait, b) TP, c) SCS, d) ratio TB/TP .....	155
Figure 38: Evolution de la proportion de lactations longues selon le rang de lactation et l'année de naissance en race Saanen (a), et Alpine (b).....	157
Figure 39: Evolution de la proportion d'animaux sortis selon le rang de lactation et l'année de naissance en race Saanen (a), et Alpine (b) .....	158
Figure 40: Evolution de la proportion de chèvres non fertiles selon le rang de lactation et l'année de naissance en race Saanen (a), et Alpine (b).....	159
Figure 41 : $-\log_{10}(\text{p-value})$ des tests de Student de comparaison de moyenne des index génétiques (G) et des index de l'environnement permanent (PE) entre les modalités de la durée de lactation, selon le	

rangs de lactation (L1, L2, L3) pour le niveau (a), et la persistance (b) pour les différents caractères (Lait, MG, MP, TB, TP, SCS, ratio TB/TP) en race Saanen.....	162
Figure 42: Moyenne de la différence entre les index, génétique ou d'environnement permanent pour le niveau (a) et la persistance (b), des animaux ayant une durée de lactation normale et ceux qui ont une lactation longue, pour les différents rang de lactation pour le lait en race Saanen.....	162
Figure 43: $-\log_{10}(\text{p-value})$ des tests de Student de comparaison de moyennes des index génétiques (G) et des index de l'environnement permanent (PE) entre les modalités de la survie, selon le rang de lactation (L1, L2, L3) pour le niveau (a), et la persistance (b) et le stade de lactation 240 pour les différents caractères (Lait, MG, MP, TB, TP, SCS, ratio TB/TP) en race Saanen.....	164
Figure 44: Moyenne de la différence d'index génétique ou de l'environnement permanent entre les animaux ayant été sortis et ceux ayant été conservés pour les différents caractères et rangs de lactation en Saanen : niveau du lait (a), persistance du lait (b), niveau des SCS (c).....	165
Figure 45 : $-\log_{10}(\text{p-value})$ des tests de Student de comparaison de moyennes des index génétiques (G) et des index de l'environnement permanent (PE) entre les modalités de la fertilité à l'IA, selon le rang de lactation (L1, L2, L3) pour le niveau (a), et la persistance (b) pour les différents caractères (Lait, MG, MP, TB, TP, SCS, ratio TB/TP) en race Saanen.....	166
Figure 46: Moyenne de la différence d'index génétique ou de l'environnement permanent entre les animaux ayant été infertiles et ceux ayant été fertiles pour les différents caractères et rangs de lactation en Saanen : niveau du lait (a), persistance du lait (b), persistance des SCS (c) .....	167
Figure 47 : Evolution des pourcentages d'infertiles, de sorties et de lactation longues en fonction des index génétique de la quantité de lait en Saanen : niveau en primipares pour la première lactation (a), en multipares pour la deuxième lactation (b), en multipares pour la troisième lactation (c), persistance pour la première lactation (d), en multipares pour la deuxième lactation (e), en multipares pour la troisième lactation (f) .....	170
Figure 48 : Evolution de la production à la lactation calculée par la méthode de Fleischmann pour les 3000 chèvres qui ont produit le plus en L1 (M1), en L2 (M2) et en L3 (M3) en Saanen (a) et en Alpine (b). .....	171
Figure 49 : Evolution du pourcentage de chèvres toujours présentes dans les parités 4 à 8 (L4, ..., L8) en Saanen (a) et en Alpine (b) pour les 3000 chèvres qui ont produit le plus en L1 (M1), en L2 (M2) et en L3 (M3) .....	172
Figure 50 : Index moyens des chèvres issus du modèle EGV_LM (LEV_P, PERS_P, LEV_M, PERS_M) et du modèle EGV_FT pour le jour de lactation 240 (L1_240, L2_240, L3_240) pour les 3000 chèvres qui ont produit le plus en L1 (M1), en L2 (M2) et en L3 (M3) en Saanen (a) et en Alpine (b) .....	173

Figure 51: Manhattan plot des effets des SNPs pour le: a) chromosome 6, b) chromosome 8, c) chromosome 14, d) chromosome 23, issus des modèles ssGBLUP_LM, ou ssGBLUP_RRM pour les coefficients de régression (LEV_P, PERS_P, LEV_M, PERS_M) pour les caractères pour lesquels des SNPs étaient détectés significatifs.....	227
Figure 52: Pourcentage de variance génétique représenté par 20 SNPs consécutifs sur le chromosome 6. Les résultats sont issus des modèles ssGBLUP_LM, ou ssGBLUP_RRM pour les coefficients de régression (LEV_P, PERS_P, LEV_M, PERS_M) pour les caractères pour lesquels des SNPs étaient détectés significatifs .....	228
Figure 53: Evolution des effets des allèles des SNPs significatifs au cours de la lactation et pour les 3 lactations pour le coefficient de régression persistance : a) en MG en Saanen sur le chromosome 19, b) en MP sur le chromosome 19, c) en SCS pour le chromosome 23 .....	230
Figure 54: Etude des effets des allèles (AA, AB et BB) du SNP en position 85 979 715 sur le chromosome 6 en race Saanen pour les caractères pour lesquels ce SNP représente une part de variance supérieure à 1% pour l'une des 4 composantes (LEV_P, PERS_P, LEV_M, PERS_M) du modèle ssGBLUP_RRM : a) lait, b) TB, c) TP, d) SCS, e) ratio TB/TP.....	232
Figure 55: Etude des effets des allèles (AA, AB et BB) du SNP 85 979 715 sur le chromosome 19 en race Saanen pour les caractères pour lesquels ce SNP représente une part de variance supérieure à 1% pour l'une des 4 composantes (LEV_P, PERS_P, LEV_M, PERS_M) du ssGBLUP RRM : a) lait, b) MG, c) MP, d) SCS .....	234
Figure 56: Evolution des valeurs génétiques (A) et des DYD selon le modèle utilisé LM et RRM pour la MP (a) et les SCS (b) à la lactation selon l'année de naissance des boucs en race Alpine .....	243
Figure 57: Evolution de la moyenne des valeurs génétiques et des DYD pour la MP (a) et les SCS selon l'année de naissance des boucs en race Alpine pour les primipares (P) et les multipares (M), pour le niveau (LEV) et la persistance (PERS) (index obtenus avec le jeu de données complet).....	246
Figure 58: Corrélations entre index des 389 boucs Saanen génotypés nés entre 1998 et 2013 avec plus de 25 filles multipares (pour le niveau (NIV) ou la persistance (PER), en primipares (PR) ou multipares (MU), ou officiels (OFF), pour les caractères de production et de SCS, et pour la fertilité (FERT)).	251
Figure 59: Corrélations entre index des 529 boucs Alpine nés entre 1998 et 2013 avec plus de 25 filles en multipares (pour le niveau (NIV) ou la persistance (PER), en primipares (PR) ou multipares (MU), ou officiels (OFF), pour les caractères de production et de SCS, pour la fertilité (FERT), et pour la longévité(LONG)).	251
Figure 60 : Evolution des corrélations en fonction du stade de lactation : (a) entre les index de 389 boucs Saanen génotypés nés entre 1998 et 2013 avec plus de 25 filles multipares et (b) entre les index de 529 boucs Alpine nés entre 1998 et 2013 avec plus de 25 filles en primipares .....	252

Figure 61 Evolutions des corrélations entre index journaliers de production et de fertilité à l'IA en fonction du stade de lactation : (a) de 389 boucs Saanen génotypés nés entre 1998 et 2013 avec plus, (b) de 25 filles multipares de 529 boucs Alpine nés entre 1998 et 2013 avec plus de 25 filles en primipares .....	252
Figure 62 :Corrélations entre index de la longévité et les index quotidiens des autres caractères de production en fonction du stade de lactation de 529 boucs Alpine nés entre 1998 et 2013 avec plus de 25 filles en primipares .....	254
Figure 63 : Evolution des effets HTD au cours des années, pour l'ensemble des troupeaux (moyenne) et pour 3 troupeaux différents .....	258





## Références

Aguilar, I., A. Legarra, F. Cardoso, Y. Masuda, D. Lourenco, and I. Misztal. 2019. Frequentist p-values for large-scale-single step genome-wide association, with an application to birth weight in American Angus cattle. *Genet. Sel. Evol.* 51:28.

Aguilar, I., I. Misztal, D.L. Johnson, A. Legarra, S. Tsuruta, and T.J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93:743–752.

Aguilar, I., I. Misztal, S. Tsuruta, A. Legarra, and H. Wang. 2014. PREGSF90–POSTGSF90: computational tools for the implementation of Single-step genomic selection and genome-wide association with ungenotyped individuals in BLUPF90 programs. Page in Proceedings of the 10th world congress of genetics applied to livestock production.

Albarrán-Portillo, B., and G.E. Pollott. 2013. The relationship between fertility and lactation characteristics in Holstein cows on United Kingdom commercial dairy farms. *J. Dairy Sci.* 96:635–646.

Ali, T.E., and L.R. Schaeffer. 1987. Accounting for covariances among test day milk yields in dairy cows. *Can. J. Anim. Sci.* 67:637–644.

Allen, M.S. 1996. Physical constraints on voluntary intake of forages by ruminants. *J. Anim. Sci.* 74:3063–3075.

Anang, A., N. Mielenz, and L. Schüler. 2002. Monthly model for genetic evaluation of laying hens II. Random regression. *Br. Poult. Sci.* 43:384–390.

Anderson, S.M., I.L. Mao, and J.L. Gill. 1989. Effect of Frequency and Spacing of Sampling on Accuracy and Precision of Estimating Total Lactation Milk Yield and Characteristics of the Lactation Curve I. *J. Dairy Sci.* 72:2387–2394.

Andonov, S., J. Ødegård, M. Svendsen, T. Adnøy, M. Vegara, and G. Klemetsdal. 2013. Comparison of random regression and repeatability models to predict breeding values from test-day records of Norwegian goats. *J. Dairy Sci.* 96:1834–1843.

ANICAP. 2015. Filière caprine-RSE.

ANICAP, and Interbev. 2018. Plan de la filière caprine française.

Apodaca-Sarabia, C.A., N. Lopez-Villalobos, H.T. Blair, and G.G. Prosser. 2009. Genetic parameters for somatic cell score in dairy goats estimated by random regression. Pages 206–209 in Proceedings of the New Zealand Society of Animal Production. New Zealand Society of Animal Production.

Arnal, M. 2016. Analyse et modélisation des formes de courbes de lactation des caprins en France.

Arnal, M., H. Larroque, H. Leclerc, V. Ducrocq, and C. Robert-Granié. 2019a. Genetic parameters for first lactation dairy traits in the Alpine and Saanen goat breeds using a random regression test-day model. *Genet. Sel. Evol.* 51:43.

Arnal, M., H. Larroque, H. Leclerc, V. Ducrocq, and C. Robert-Granié. 2019c. Estimation of genetic parameters for dairy traits and somatic cell score in the first three parities using a random regression test-day model in French Alpine goats (submitted).

Arnal, M., C. Robert-Granié, and H. Larroque. 2018. Diversity of dairy goat lactation curves in France. *J. Dairy Sci.* 101:11040–11051.

Baba, T., Y. Gotoh, S. Yamaguchi, S. Nakagawa, H. Abe, Y. Masuda, and T. Kawahara. 2017. Application of Single-step genomic best linear unbiased prediction with a multiple-lactation random regression test-day model for Japanese Holsteins. *Anim. Sci. J.* 88:1226–1231.

Bapst, B. 2001. Swiss experiences on practical cattle breeding strategies for organic dairy herds. Pages 44–50 in *Breeding and feeding for animal health and welfare in organic livestock systems. Proceedings of the 4th NAHWOA Workshop, Wagenningen.*

Baumung, R., J. Sölkner, E. Gierzinger, and A. Willam. 2001. Ecological total merit index for an Austrian dual purpose cattle breed. *Arch. Anim. Breed.* 44:5–14.

Bélichon, S., E. Manfredi, and A. Piacère. 1999. Genetic parameters of dairy traits in the Alpine and Saanen goat breeds. *Genet. Sel. Evol.* 31:529–534.

Bignardi, A.B., L. El Faro, R.A.A. Torres Júnior, V.L. Cardoso, P.F. Machado, and L.G. de Albuquerque. 2011. Random regression models using different functions to model test-day milk yield of Brazilian Holstein cows. *Genet. Mol. Res.* 3565–3575.

Bohlouli, M., S. Alijani, S. Naderi, T. Yin, and S. König. 2019. Prediction accuracies and genetic parameters for test-day traits from genomic and pedigree-based random regression models with or without heat stress interactions. *J. Dairy Sci.* 102:488–502.

Bohmanova, J., F. Miglior, J. Jamrozik, I. Misztal, and P.G. Sullivan. 2008. Comparison of random regression models with Legendre polynomials and linear splines for production traits and somatic cell score of Canadian Holstein cows. *J. Dairy Sci.* 91:3627–3638.

Boichard, D., V. Ducrocq, P. Croiseau, and S. Fritz. 2016. Genomic selection in domestic animals: Principals, application and perspectives. *C R Biol* 339:274–277.

Bouallegue, M., R. Steri, N. M'hamdi, and M.B. Hamouda. 2015. Modelling of individual lactation curves of Tunisian Holstein-Friesian cows for milk yield, fat, and protein contents using parametric, orthogonal and spline models. *J. Anim. Feed Sci.* 24.

Bouloc, N. 1991. Analyse de la forme de la courbe de lactation. Application à l'étude des modalités d'allègement du contrôle laitier et de prévision précoce de la production dans l'espèce caprine. PhD thesis, INA Paris Grignon, France.

Brito, L.F., F.G. Silva, H.R. Oliveira, N.O. Souza, G.C. Caetano, E.V. Costa, G.R.O. Menezes, A.L.P. Melo, M.T. Rodrigues, and R.A. Torres. 2017. Modelling lactation curves of dairy goats by fitting random regression models using Legendre polynomials or B-splines. *Can. J. Anim. Sci.* 98:73–83.

Buttchereit, N., E. Stamer, W. Junge, and G. Thaller. 2011. Genetic relationships among daily energy balance, feed intake, body condition score, and fat to protein ratio of milk in dairy cows. *J. Dairy Sci.* 94:1586–1591.

Buttchereit, N., E. Stamer, W. Junge, and G. Thaller. 2012. Genetic parameters for energy balance, fat /protein ratio, body condition score and disease traits in German Holstein cows. *J. Anim. Breed. Genet.* 129:280–288 <http://dx.doi.org/10.1111/j.1439-0388.2011.00976.x>.

Capgenes. 2019. La sélection caprine, à l'ère de la génomique.

Caraviello, D.Z., K.A. Weigel, G.E. Shook, and P.L. Ruegg. 2005. Assessment of the impact of somatic cell count on functional longevity in Holstein and Jersey cattle using survival analysis methodology. *J. Dairy Sci.* 88:804–811.

Cardona, S.J.C., H.C. Cadavid, J.D. Corrales, S. Munilla, R.J. Cantet, and A. Rogberg-Muñoz. 2016. Longitudinal data analysis of polymorphisms in the  $\kappa$ -casein and  $\beta$ -lactoglobulin genes shows differential effects along the trajectory of the lactation curve in tropical dairy goats. *J. Dairy Sci.* 99:7299–7307.

Carillier-Jacquin, C. 2015. Etude de la prédiction génomique chez les caprins: faisabilité et limites de la sélection génomique dans le cadre d'une population multiraciale et à faible effectif. PhD Thesis Thesis.

Chartier, C., and S. Blain. 2018. Pathologie Caprine: Du Diagnostic À La Prévention. Les éditions du point vétérinaire.

Chemineau, P., B. Malpoux, J.P. Brillard, and A. Fostier. 2007. Seasonality of reproduction and production in farm fishes, birds and mammals. *Animal* 1:419–432.

Clément, V., C. Ceglowski, R. de Cremoux, P. Martin, and R. Rupp. 2015a. Concentrations cellulaires dans les élevages caprins: état des lieux et mise en place d'un programme d'amélioration génétique. *Renc Rech Rumin.* 22:45–48.

Clément, V., I. Palhiere, and H. Larroque. 2015b. Évaluation génétique dans l'espèce caprine: Caractères de production laitière, de comptage de cellules somatiques et de morphologie. *Compte-Rendu* 10.

Cole, J.B., and P.M. VanRaden. 2006. Genetic evaluation and best prediction of lactation persistency. *J. Dairy Sci.* 89:2722–2728.

DANCHIN-BURGE, C., and D. DUCLOS. 2009. Situation et perspectives d’avenir des races caprines à petits effectifs. *Ethnozootechnie* 85:10–16.

Dekkers, J.C.M., J.H. Ten Hag, and A. Weersink. 1998. Economic aspects of persistency of lactation in dairy cattle. *Livest. Prod. Sci.* 53:237–252.

Dobson, H., R.F. Smith, M.D. Royal, C.H. Knight, and I.M. Sheldon. 2007. The high-producing dairy cow and its reproductive performance. *Reprod. Domest. Anim.* 42:17–23.

Druet, T., F. Jaffrézic, D. Boichard, and V. Ducrocq. 2003. Modeling lactation curves and estimation of genetic parameters for first lactation test-day records of French Holstein cows. *J. Dairy Sci.* 86:2480–2490.

Druet, T., F. Jaffrézic, and V. Ducrocq. 2005. Estimation of genetic parameters for test day records of dairy traits in the first three lactations. *Genet. Sel. Evol.* 37:257.

Ducrocq, V. 1998. Genedit, BLUP software; version October 30, 2017. INRA SGQA Jouy-En-Josas Fr.

FGE. 2010. La génétique française pour les filières bovines, ovines et caprines.

Forni, S., D. Gianola, G.J.M. Rosa, and G. de los Campos. 2009. A dynamic linear model for genetic analysis of longitudinal traits. *J. Anim. Sci.* 87:3845–3853 <http://dx.doi.org/10.2527/jas.2008-1514>.

Fragné, M. 2014. L’élevage caprin en France: situation actuelle et perspectives. PhD Thesis Thesis.

Gengler, N. 1996. Persistency of lactation yields: A review. *Interbull Bull.* 12:87–96.

Haiger, A. 2018. Life Performance-The “Natural Selection Index”. *Int J Biotech Bioeng* 4:5–102.

Haile-Mariam, M., P.J. Bowman, and M.E. Goddard. 2003. Genetic and environmental relationship among calving interval, survival, persistency of milk yield and somatic cell count in dairy cattle. *Livest. Prod. Sci.* 80:189–200.

Harder, B., J. Bennewitz, D. Hinrichs, and E. Kalm. 2006. Genetic parameters for health traits and their relationship to different persistency traits in German Holstein dairy cattle. *J. Dairy Sci.* 89:3202–3212.

Huquet, B., H. Leclerc, and V. Ducrocq. 2012a. Modelling and estimation of genotype by environment interactions for production traits in French dairy cattle. *Genet. Sel. Evol.* 44:35.

Huquet, B., H. Leclerc, and V. Ducrocq. 2012b. Characterization of French dairy farm environments from herd-test-day profiles. *J. Dairy Sci.* 95:4085–4098.

Idele. 2012. Que sont les cellules?

- Idele. 2013a. Les lactations longues.
- Idele. 2013b. Evolution de la fertilité à l'insémination artificielle chez les chèvres de 2004 à 2010.
- Idele. 2018. Résultats contrôle laitier caprin.
- Idele. 2019a. Résultats 2017 des ateliers caprins livreurs et fromagers.
- Idele. 2019b. Dossier annuel Caprin-Une production relancée en 2018. Une position à consolider en 2019.
- Inosys. 2014. Améliorer la longévité des troupeaux caprins.
- Jaffrézic, F., E. Venot, D. Laloë, A. Vinet, and G. Renand. 2004. Use of structured antedependence models for the genetic analysis of growth curves. *J. Anim. Sci.* 82:3465–3473.
- Jakobsen, J.H. er, R. Rekaya, J. Jensen, D.A. Sorensen, P. Madsen, D. Gianola, L.G. Christensen, and J. Pedersen. 2003. Bayesian estimates of covariance components between lactation curve parameters and disease liability in Danish Holstein cows. *J. Dairy Sci.* 86:3000–3007.
- Jamrozik, J., and L.R. Schaeffer. 2012. Test-day somatic cell score, fat-to-protein ratio and milk yield as indicator traits for sub-clinical mastitis in dairy cattle. *J. Anim. Breed. Genet.* 129:11–19.
- Jamrozik, J., L.R. Schaeffer, and J.C.M. Dekkers. 1997. Genetic evaluation of dairy cattle using test day yields and random regression model. *J. Dairy Sci.* 80:1217–1226.
- Kang, H., C. Ning, L. Zhou, S. Zhang, Q. Yan, and J.-F. Liu. 2018. Single-step genomic evaluation of milk production traits using multiple-trait random regression model in Chinese Holsteins. *J. Dairy Sci.* 101:11143–11149.
- Kang, H., L. Zhou, R. Mrode, Q. Zhang, and J.F. Liu. 2017. Incorporating the Single-step strategy into a random regression model to enhance genomic prediction of longitudinal traits. *Heredity* 119:459.
- Kirkpatrick, M., D. Lofsvold, and M. Bulmer. 1990. Analysis of the inheritance, selection and evolution of growth trajectories.. *Genetics* 124:979–993.
- Knaus, W. 2009. Dairy cows trapped between performance demands and adaptability. *J. Sci. Food Agric.* 89:1107–1114.
- Koeck, A., F. Miglior, J. Jamrozik, D.F. Kelton, and F.S. Schenkel. 2013. Genetic associations of ketosis and displaced abomasum with milk production traits in early first lactation of Canadian Holsteins. *J. Dairy Sci.* 96:4688–4696.
- Koivula, M., I. Strandén, J. Pösö, G.P. Aamand, and E.A. Mäntysaari. 2015. Single-step genomic evaluation using multitrait random regression model and test-day data. *J. Dairy Sci.* 98:2775–2784.

Kolbehdari, D., Z. Wang, J.R. Grant, B. Murdoch, A. Prasad, Z. Xiu, E. Marques, P. Stothard, and S.S. Moore. 2009. A whole genome scan to map QTL for milk production traits and somatic cell score in Canadian Holstein bulls. *J. Anim. Breed. Genet.* 126:216–227.

Kompan, D., and M. Klopčič. 2004. Report of WG Milk Recording in Goats. Univ. Ljublj. Biotech. Fac. Zoo Tech. Dep. Groblje 3.

Krogmeier, D. 2003. Der Ökologische Gesamtzuchtwert für Fleckvieh, Braunvieh und Gelbvieh.

Larroque, H., J.-M. Astruc, A. Barbat, F. Barillet, D. Boichard, B. Bonaiti, V. Clément, I. David, G. Lagriffoul, I. Palhière, A. Piacère, C. Robert-Granie, and R. Rupp. 2011. National Genetic Evaluations in Dairy Sheep and Goats in France. Wageningen Academic, Stavanger, Norway.

Leclerc, H. 2008a. Mise en place de l'évaluation génétique sur les contrôles élémentaires en bovins laitiers et perspectives d'utilisation des résultats en appui technique. Theses AgroParisTech Dec.

Leclerc, H. 2008b. Development of the French dairy cattle test-day model genetic evaluation and prospects of using results for herd management. PhD Thesis Thesis. AgroParisTech,.

Leclerc, H., D. Duclos, A. Barbat, T. Druet, and V. Ducrocq. 2008. Environmental effects on lactation curves included in a test-day model genetic evaluation. *Animal* 2:344–353.

Leclerc, H., I. Nagy, and V. Ducrocq. 2009. Impact of using reduced rank random regression test-day model on genetic evaluation. *Interbull Bull.* 40:42–46.

Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. *J. Dairy Sci.* 92:4656–4663.

Legarra, A., and A. Reverter. 2018. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. *Genet. Sel. Evol.* 50:53.

Lidauer, M.H., J. Pösö, J. Pedersen, J. Lassen, P. Madsen, E.A. Mäntysaari, U.S. Nielsen, J.-Å. Eriksson, K. Johansson, T. Pitkänen, I. Strandén, and G.P. Aamand. 2015. Across-country test-day model evaluations for Holstein, Nordic Red Cattle, and Jersey. *J. Dairy Sci.* 98:1296–1309 <http://dx.doi.org/10.3168/jds.2014-8307>.

Lu, H., and H. Bovenhuis. 2019. Genome-wide association studies for genetic effects that change during lactation in dairy cattle. *J. Dairy Sci.*

Lucy, M.C. 2001. Reproductive loss in high-producing dairy cattle: where will it end?. *J. Dairy Sci.* 84:1277–1293.

Macciotta, N.P.P., G. Gaspa, L. Bomba, D. Vicario, C. Dimauro, M. Cellesi, and P. Ajmone-Marsan. 2015. Genome-wide association analysis in Italian Simmental cows for lactation curve traits using a low-density (7K) SNP panel. *J. Dairy Sci.* 98:8175–8185.

Macé, T., E. González-García, F. Carrière, S. Douls, D. Foulquié, C. Robert-Granié, and D. Hazard. 2019. Intra-flock variability in the body reserve dynamics of meat sheep by analyzing BW and body condition score variations over multiple production cycles. *animal* 1–13.

Maigret, C. 2018a. Dossier annuel caprins : conjoncture 2017 et Perspectives 2018. Prod. Caprines Lait Viande Inst. L’Elevage Confédération Natl. L’Elevage Httpidele Fr.

Maigret, C. 2018b. Chiffres clés 2017. Prod. Caprines Lait Viande Inst. L’Elevage Confédération Natl. L’Elevage Httpidele Fr.

Marc Teissier. 2018. Intégration de données génomiques (mutations, gènes majeurs, marqueurs SNP, haplotypes) dans les modèles d’évaluations génétiques des chèvres laitières pour améliorer l’efficacité de la sélection.

Martin, P. 2014. Schéma de sélection caprin : bilan et perspectives. Page in Journées techniques, Selection animale en AB, Chateauroux.

Martin, P., I. Palhière, C. Maroteau, P. Bardou, K. Canale-Tabet, J. Sarry, F. Woloszyn, J. Bertrand-Michel, I. Racke, and H. Besir. 2017. A genome scan for milk production traits in dairy goats reveals two new mutations in *Dgat1* reducing milk fat content. *Sci. Rep.* 7:1872.

Martin, P., I. Palhière, C. Maroteau, V. Clément, I. David, G.T. Klopp, and R. Rupp. 2018. Genome-wide association mapping for type and mammary health traits in French dairy goats identifies a pleiotropic region on chromosome 19 in the Saanen breed. *J. Dairy Sci.* 101:5214–5226.

Menéndez-Buxadera, A., A. Molina, F. Arrebola, M.J. Gil, and J.M. Serradilla. 2010. Random regression analysis of milk yield and milk composition in the first and second lactations of Murciano-Granadina goats. *J. Dairy Sci.* 93:2718–2726.

Meyer, K. 2005. Random regression analyses using B-splines to model growth of Australian Angus cattle. *Genet. Sel. Evol.* 37:473.

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.

Meyer, K., and W.G. Hill. 1997. Estimation of genetic and phenotypic covariance functions for longitudinal or “repeated” records by restricted maximum likelihood. *Livest. Prod. Sci.* 47:185–200 [http://dx.doi.org/10.1016/S0301-6226\(96\)01414-5](http://dx.doi.org/10.1016/S0301-6226(96)01414-5).

Meyer, K., and M. Kirkpatrick. 2008. Perils of Parsimony: Properties of Reduced-Rank Estimates of Genetic Covariance Matrices. *Genetics* 180:1153–1166 <http://dx.doi.org/10.1534/genetics.108.090159>.

Minot, P. 2002. L’aventure de la chèvre en Poitou-Charentes au XXème siècle. *Ethnozootechnie* 11–16.

- Misztal, I. 2006. Properties of random regression models using linear splines. *J. Anim. Breed. Genet.* 123:74–80.
- Misztal, I., A. Legarra, and I. Aguilar. 2014. Using recursion to compute the inverse of the genomic relationship matrix. *J. Dairy Sci.* 97:3943–3952.
- Misztal, I., T. Strabel, J. Jamrozik, E.A. Mäntysaari, and T.H.E. Meuwissen. 2000. Strategies for estimating the parameters needed for different test-day models. *J. Dairy Sci.* 83:1125–1134.
- Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, and D.H. Lee. 2002. BLUPF90 and related programs (BGF90). Pages 743–744 in *Proceedings of the 7th world congress on genetics applied to livestock production*.
- Mota, R.R., F.F. Silva, P.S. Lopes, R.J. Tempelman, B.P. Sollero, I. Aguilar, and F.F. Cardoso. 2018. Analyses of reaction norms reveal new chromosome regions associated with tick resistance in cattle. *animal* 12:205–214 <http://dx.doi.org/10.1017/S1751731117001562>.
- Mucha, S., R. Mrode, M. Coffey, and J. Conington. 2014. Estimation of genetic parameters for milk yield across lactations in mixed-breed dairy goats. *J. Dairy Sci.* 97:2455–2461.
- Mucha, S., R. Mrode, M. Coffey, M. Kizilaslan, S. Desire, and J. Conington. 2018. Genome-wide association study of conformation and milk yield in mixed-breed dairy goats. *J. Dairy Sci.* 101:2213–2225.
- Mucha, S., R. Mrode, I. MacLaren-Lee, M. Coffey, and J. Conington. 2015. Estimation of genomic breeding values for milk yield in UK dairy goats. *J. Dairy Sci.* 98:8201–8208.
- Muir, B.L., J. Fatehi, and L.R. Schaeffer. 2004. Genetic relationships between persistency and reproductive performance in first-lactation Canadian Holsteins. *J. Dairy Sci.* 87:3029–3037.
- Negussie, E., I. Strandén, and E.A. Mäntysaari. 2013. Genetic associations of test-day fat: protein ratio with milk yield, fertility, and udder health traits in Nordic Red cattle. *J. Dairy Sci.* 96:1237–1250.
- Nishiura, A., O. Sasaki, M. Aihara, H. Takeda, and M. Satoh. 2015. Genetic analysis of fat-to-protein ratio, milk yield and somatic cell score of H olstein cows in J apan in the first three lactations by using a random regression model. *Anim. Sci. J.* 86:961–969.
- Oliveira, H.R., L.F. Brito, D.A.L. Lourenco, F.F. Silva, J. Jamrozik, L.R. Schaeffer, and F.S. Schenkel. 2019a. Invited review: Advances and applications of random regression models: From quantitative genetics to genomics. *J. Dairy Sci.*
- Oliveira, H.R., L.F. Brito, F.F. Silva, D.A.L. Lourenco, J. Jamrozik, and F.S. Schenkel. 2019b. Genomic prediction of lactation curves for milk, fat, protein, and somatic cell score in Holstein cattle. *J. Dairy Sci.* 102:452–463.

Oliveira, H.R., D.A.L. Lourenco, Y. Masuda, I. Misztal, S. Tsuruta, J. Jamrozik, L.F. Brito, F.F. Silva, J.P. Cant, and F.S. Schenkel. 2019c. Single-step genome-wide association for longitudinal traits of Canadian Ayrshire, Holstein, and Jersey dairy cattle. *J. Dairy Sci.*

Oliveira, H.R., D.A.L. Lourenco, Y. Masuda, I. Misztal, S. Tsuruta, J. Jamrozik, L.F. Brito, F.F. Silva, and F.S. Schenkel. 2019d. Application of Single-step genomic evaluation using multiple-trait random regression test-day models in dairy cattle. *J. Dairy Sci.* 102:2365–2377.

Oliveira, H.R., F.F. Silva, L.F. Brito, J. Jamrozik, D.A.L. Lourenco, and F.S. Schenkel. Genome-wide association study for milk, fat and protein yields in different lactation stages in Canadian Holstein and Jersey cattle.

Oltenucu, P.A., and B. Algers. 2005. Selection for increased production and the welfare of dairy cows: are new breeding goals needed?. *AMBIO J. Hum. Environ.* 34:311–316.

Palhiere, I., V. Clément, I. Croué, and D. Pinard. 2015. Un objectif de sélection qui augmente le profit des éleveurs caprins. Paris Fr.

Palhiere, I., V. Clément, P. Martin, and J.J. Colleau. 2014. Bilan de la méthode de sélection à parenté minimum après 6 ans d'application dans le schéma de sélection caprin. Proc. 21èmes Rencontres Autour Rech. Sur Rumin. 3–4.

Pauciuolo, A., G. Cosenza, R. Steri, A. Coletta, L. Jemma, M. Feligini, D. Di Berardino, N.P. Macciotta, and L. Ramunno. 2012. An association analysis between OXT genotype and milk yield and flow in Italian Mediterranean river buffalo. *J. Dairy Res.* 79:150–156.

Pereira, R.J., A.B. Bignardi, L. El Faro, R.S. Verneque, A.E. Vercesi Filho, and L.G. Albuquerque. 2013. Random regression models using Legendre polynomials or linear splines for test-day milk yield of dairy Gyr (*Bos indicus*) cattle. *J. Dairy Sci.* 96:565–574.

Piacére, A. 2008. Le schéma de sélection caprin laitier en France: son histoire et son évolution. *Ethnozootecnie* 41–46.

Pletcher, S.D., and C.J. Geyer. 1999. The genetic analysis of age-dependent traits: modeling the character process. *Genetics* 153:825–835.

Prache, S., H. Caillat, and G. Lagriffoul. 2018. Diversité dans la filière petits ruminants: une source de résilience. *Innov. Agron.* 68:171–191.

Pryce, J.E., M. Haile-Mariam, K. Verbyla, P.J. Bowman, M.E. Goddard, and B.J. Hayes. 2010. Genetic markers for lactation persistency in primiparous Australian dairy cows. *J. Dairy Sci.* 93:2202–2214.

Pryce, J.E., M.D. Royal, P.C. Garnsworthy, and I.L. Mao. 2004. Fertility in the high-producing dairy cow. *Livest. Prod. Sci.* 86:125–135.

- Ptak, E., and L.R. Schaeffer. 1993. Use of test day yields for genetic evaluation of dairy sires and cows. *Livest. Prod. Sci.* 34:23–34.
- Riggio, V., D.O. Maizon, B. Portolano, H. Bovenhuis, and J.A.M. Van Arendonk. 2009. Effect of somatic cell count level on functional longevity in Valle del Belice dairy sheep assessed using survival analysis. *J. Dairy Sci.* 92:6160–6166.
- Roche, C.V. 2008. La chèvre poitevine: son histoire et sa Renaissance. *Ethnozootechnie* 33–39.
- Rozzi, P., F. Miglior, and K.J. Hand. 2007. A total merit selection index for Ontario organic dairy farmers. *J. Dairy Sci.* 90:1584–1593.
- Rupp, R., V. Clément, A. Piacere, C. Robert-Granié, and E. Manfredi. 2011. Genetic parameters for milk somatic cell score and relationship with production and udder type traits in dairy Alpine and Saanen primiparous goats. *J. Dairy Sci.* 94:3629–3634.
- Sargent, F.D., V.H. Lytton, and O.G. Wall. 1968. Test interval method of calculating dairy herd improvement association records. *J. Dairy Sci.* 51:170–179.
- Schaeffer, L.R. 2006. Strategy for applying genome-wide selection in dairy cattle. *J. Anim. Breed. Genet.* 123:218–223.
- Schaeffer, L.R. 2016. *Random Regression Models*. Accessed July 3, 2019. <http://animalbiosciences.uoguelph.ca/~lrs/BOOKS/rrmbook.pdf>.
- Schaeffer, L.R. 2018. Necessary changes to improve animal models. *J. Anim. Breed. Genet.* 135:124–131 <http://dx.doi.org/10.1111/jbg.12321>.
- Schaeffer, L.R., and J. Jamrozik. 2008. Random regression models: a longitudinal perspective. *J. Anim. Breed. Genet.* 125:145–146.
- Sewalem, A., F. Miglior, G.J. Kistemaker, and B.J. Van Doormaal. 2006. Analysis of the relationship between somatic cell score and functional longevity in Canadian dairy cattle. *J. Dairy Sci.* 89:3609–3614.
- Sölkner, J., and W. Fuchs. 1987. A comparison of different measures of persistency with special respect to variation of test-day milk yields. *Livest. Prod. Sci.* 16:305–319.
- Steiger, J.H. 1980. Tests for comparing elements of a correlation matrix. *Psychol. Bull.* 87:247–251.
- Strucken, E.M., R.H. Bortfeldt, D.J. De Koning, and G.A. Brockmann. 2012a. Genome-wide associations for investigating time-dependent genetic effects for milk production traits in dairy cattle. *Anim. Genet.* 43:375–382.

Strucken, E.M., R.H. Bortfeldt, J. Tetens, G. Thaller, and G.A. Brockmann. 2012b. Genetic effects and correlations between production and fertility traits and their dependency on the lactation-stage in Holstein Friesians. *BMC Genet.* 13:108.

Strucken, E.M., Y.C. Laurenson, and G.A. Brockmann. 2015. Go with the flow—biology and genetics of the lactation cycle. *Front. Genet.* 6:118.

Szyda, J., J. Komisarek, and I. Antkowiak. 2014. Modelling effects of candidate genes on complex traits as variables over time. *Anim. Genet.* 45:322–328.

Taskinen, M., E.A. Mäntysaari, and I. Strandén. 2017. Single-step SNP-BLUP with on-the-fly imputed genotypes and residual polygenic effects. *Genet. Sel. Evol.* 49:36.

Täubert, H., Z. Liu, J. Tarrès, and V. Ducrocq. 2010. An approach to compute EDC and DYD for test-day models. Page in *Proceedings of the 9th WCGALP. 2010; 9. World Congress on Genetics Applied to Livestock Production, Leipzig, DEU, 2010-08-01-2010-08-06, 1-4.*

Teissier, M., H. Larroque, and C. Robert-Granié. 2018. Weighted Single-step genomic BLUP improves accuracy of genomic breeding values for protein content in French dairy goats: a quantitative trait influenced by a major gene. *Genet. Sel. Evol.* 50:31.

Teissier, M., H. Larroque, and C. Robert-Granie. 2019. Accuracy of genomic evaluation with weighted Single-step genomic best linear unbiased prediction for milk production traits, udder type traits, and somatic cell scores in French dairy goats. *J. Dairy Sci.* 102:3142–3154.

Thomas. 2019. RESULTATS DE CONTROLE LAITIER - ESPECE CAPRINE FRANCE 2018.

Tosser-Klopp, G., P. Bardou, O. Bouchez, C. Cabau, R. Crooijmans, Y. Dong, C. Donnadiu-Tonon, A. Eggen, H.C. Heuven, and S. Jamli. 2014. Design and characterization of a 52K SNP chip for goats. *PLoS One* 9:e86227.

Tran, V.H.H. 2018. NEW GENETIC LONGITUDINAL MODELS FOR FEED EFFICIENCY. PhD Thesis. Institut National Polytechnique de Toulouse,.

Van der Werf, J.H.J., M.E. Goddard, and K. Meyer. 1998. The use of covariance functions and random regressions for genetic evaluation of milk production based on test day records. *J. Dairy Sci.* 81:3300–3308.

VanRaden, P.M., C.P. Van Tassell, G.R. Wiggans, T.S. Sonstegard, R.D. Schnabel, J.F. Taylor, and F.S. Schenkel. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci.* 92:16–24.

VanRaden, P.M., and G.R. Wiggans. 1991. Derivation, calculation, and use of national animal model information. *J. Dairy Sci.* 74:2737–2746.

Vigne, J.-D., and D. Helmer. 2007. Was milk a “secondary product” in the Old World Neolithisation process? Its role in the domestication of cattle, sheep and goats. *Anthropozoologica* 42:9–40.

Wang, H., I. Misztal, I. Aguilar, A. Legarra, R.L. Fernando, Z. Vitezica, R. Okimoto, T. Wing, R. Hawken, and W.M. Muir. 2014. Genome-wide association mapping including phenotypes from relatives without genotypes in a Single-step (ssGWAS) for 6-week body weight in broiler chickens. *Front. Genet.* 5:134.

Wei, T., V. Simko, M. Levy, Y. Xie, Y. Jin, and J. Zemla. 2017. Package “corrplot”. *Statistician* 56:316–324.

Wiggans, G.R., and G.E. Shook. 1987. A lactation measure of somatic cell count. *J. Dairy Sci.* 70:2666–2672.

Wilmink, J.B.M. 1987. Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. *Livest. Prod. Sci.* 16:335–348.

Windig, J.J., M.P. Calus, G. de Jong, and R.F. Veerkamp. 2005. The association between somatic cell count patterns and milk production prior to mastitis. *Livest. Prod. Sci.* 96:291–299.

Wood, P.D.P. 1967. Algebraic model of the lactation curve in cattle.

Zavadilová, L., J. Wolf, M. Štípková, E. Němcová, and J. Jamrozik. 2011. Genetic parameters for somatic cell score in the first three lactations of Czech Holstein and Fleckvieh breeds using a random regression model. *Czech J Anim Sci* 56:251–260.

Zumbach, B., S. Tsuruta, I. Misztal, and K.J. Peters. 2008. Use of a test day model for dairy goat milk yield across lactations in Germany. *J. Anim. Breed. Genet.* 125:160–167.