



Analyse et modélisation des formes de courbes de lactation des caprins en France

Mathieu Arnal

► To cite this version:

Mathieu Arnal. Analyse et modélisation des formes de courbes de lactation des caprins en France. Sciences agricoles. 2016. dumas-01400860

HAL Id: dumas-01400860

<https://dumas.ccsd.cnrs.fr/dumas-01400860>

Submitted on 22 Nov 2016

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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



AGROCAMPUS OUEST

☐ CFR Angers ☒ CFR Rennes

Année universitaire : 2015.-2016

Spécialité :

Sciences de l'Animal pour l'élevage de demain.....

.....

Spécialisation (et option éventuelle) :

.....

...

Mémoire de Fin d'Études

- ☒ d'Ingénieur de l'Institut Supérieur des Sciences agronomiques, agroalimentaires, horticoles et du paysage
- ☐ de Master de l'Institut Supérieur des Sciences agronomiques, agroalimentaires, horticoles et du paysage
- ☐ d'un autre établissement (étudiant arrivé en M2)

Analyse et modélisation des formes de courbes de lactation des caprins en France

Par : Mathieu ARNAL

Soutenu à Rennes le 13 septembre 2016

Devant le jury composé de :

Maître de stage : Hélène LARROQUE et Christèle ROBERT-GRANIE

Enseignant référent : Jocelyne FLAMENT

Autres membres du jury :

Rapporteur : Philippe FAVERDIN

Présidente : Catherine DISENHAUS

Les analyses et les conclusions de ce travail d'étudiant n'engagent que la responsabilité de son auteur et non celle d'AGROCAMPUS OUEST

Remerciements

Je tiens à remercier mes deux encadrantes, Hélène Larroque et Christèle Robert-Granié, de m'avoir fait confiance en me permettant de réaliser mon stage de fin d'étude dans leur équipe. Je les remercie pour leur encadrement, leur soutien et leur très grande disponibilité. Ce fut un plaisir d'apprendre à leurs côtés.

Je tiens à remercier l'ensemble de l'équipe GenPhySe et plus particulièrement l'équipe des petits ruminants pour leur accueil et leur bonne humeur. J'ai grandement apprécié les réunions hebdomadaires qui m'ont beaucoup appris du contexte des petits ruminants.

Je remercie ma tutrice Jocelyne Flament pour sa disponibilité et ses conseils.

Je remercie l'ensemble des stagiaires, des thésards, l'équipe du flag rugby pour leur accueil et leurs conseils, et particulièrement Marc Teissier pour son aide et sa bonne humeur quotidienne.

Sommaire

Remerciements	I
Sommaire.....	II
Liste des abréviations	III
Liste des figures.....	IV
Liste des tableaux.....	VI
Liste des annexes.....	VII
Introduction	1
Matériels et méthodes	4
1. Sélection de deux jeux de données.....	4
2. Description des données issues de notre sélection et création des classes de facteurs.....	5
3. Classification des profils individuels de courbes de lactation.....	6
4. Modélisation des courbes de lactation moyenne.....	8
4. a. Description des modèles.....	8
4. b. Tests des modèles.....	10
4. c. Outils de comparaison des modèles	11
Résultats	11
1. Analyse descriptive des courbes de lactation en fonction des effets de milieu (premier jeu de données)	11
2. Classification des courbes de production laitière	12
2. a. Choix du nombre de clusters :	12
2. b. A partir de cette classification en 7 clusters :	13
2. c. Evolution de la forme de la courbe de lactation au cours de la carrière des chèvres :	14
3. Modélisation des courbes de production laitière	15
3. a. Sélection des types de modèles:.....	15
3. b. Modélisation des effets d'environnement :	15
Discussion	17
Conclusion et perspectives.....	20
Bibliographie.....	21
Annexes	24

Liste des abréviations

Protocole **A** ou **D** ou **AZ** : protocole laitier dont la pesée du lait est réalisée lors des 2 traites soir

AIC : Critère d'Information d'Akaike

protocole **AT** : protocole laitier dont la pesée du lait est réalisée lors d'une seule

HTD Herd Test Day

BIC Critère d'Information Bayésien

Fix de 1 à 5 : différents modèles des classes fixes

L3+ : les troisièmes lactations et plus du premier jeu de données de 2011-2012

L3++ : les troisièmes lactations et plus du premier jeu de données de 2013-2014

Leg 4 : modèle de Legendre de degré 4

Leg 5 : modèle de Legendre de degré 5

NE : Nord Est

NO : Nord-Ouest

R² : coefficient de détermination

SE : Sud Est

SO : Sud-Ouest

Sigma : écart type de l'erreur

TDM : Test Day Model

Liste des figures

Figure 1 : Répartition de la production de lait de chèvres en France en 2014 (Institut de l'élevage, 2015)	1
Figure 2 : Evolution de l'index lait entre 1990 et 2009 pour les chèvres de race Alpine (Larroque et al., 2011).....	2
Figure 3 : Quantité de lait moyenne produite par jour en hg en fonction du stade de lactation en jours (les trois principales phases de la lactation sont indiquées). La droite $y = 36.5$ hg identifie le niveau de production au plateau de la lactation.	2
Figure 4 : Nombre de données par stade de lactation selon le rang de lactation (première lactation (L1), deuxième lactation (L2) et troisième lactation et plus, (L3+) dans le deuxième jeu de données.....	4
Figure 5 : Quantité de lait (en hg) journalière moyenne lissée en fonction du stade de lactation (en jours) par des chèvres en première (L1) et deuxième lactation (L2) pour les campagnes de mise bas 2009 à 2011 (premier jeu de données)	4
Figure 6 : Quantité de lait (en hg) journalière moyenne lissée en fonction du stade de lactation (en jours) par des chèvres en troisième lactation et plus (L3 et +) pour les campagnes de mise bas 2011 à 2014 (premier jeu de données).....	4
Figure 7 : Quantité de lait (en hg) journalière moyenne lissée produite en fonction du stade de lactation (en jours) selon la classe de durée de tarissement (premier jeu de données)	4
Figure 8 : Quantité de lait moyenne lissée (fonction R « smooth.spline ») produite par jour en hg en fonction du stade de lactation en jours et du rang de lactation pour le jeu de données 1.	6
Figure 9 : Quantité de lait moyenne lissée (fonction R « smooth.spline ») produite par jour en hg en fonction du stade de lactation en jours par des chèvres en deuxième lactation issues de la région Nord-ouest de race Saanen ou Alpine pour le jeu de données 1.	6
Figure 10 : Quantité de lait moyenne lissée (fonction R « smooth.spline ») produite par jour en hg en fonction du stade de lactation en jours par des chèvres en deuxième lactation de race Alpine selon la région d'élevage (NE : Nord-Est, NO : Nord-Ouest, SE : Sud-Est, SO : Sud-Ouest) pour le jeu de données 1.	6
Figure 11 : Quantité de lait moyenne lissée (fonction R « smooth.spline ») produite par jour en hg en fonction du stade de lactation en jours par des chèvres en deuxième lactation de race Alpine selon l'âge à la mise bas en mois (jeu de données 1)	6
Figure 12 : Quantité de lait moyenne lissée (fonction R « smooth.spline ») produite par jour en hg en fonction du stade de lactation en jours par des chèvres en deuxième lactation de race Alpine selon le mois de mise bas (1 : Janvier, 2 : Février, 3 : Mars, 4-8 : Avril-Août, 9-12 : Septembre-Décembre) pour le jeu de données 1.....	7
Figure 13 : Quantité de lait moyenne lissée (fonction R « smooth.spline ») produite par jour en hg en fonction du stade de lactation en jours par des chèvres en deuxième lactation selon la durée de tarissement en jours (jeu de données 1).....	7
Figure 14 : Quantité de lait moyenne lissée (fonction R « smooth.spline ») produite par jour en hg en fonction du stade de lactation en jours par des chèvres en deuxième lactation selon la durée de gestation en jours (jeu de données 1).....	7
Figure 15 : Evolution de la valeur relative (échelle de 0 à 1) de différents critères de qualité en fonction du nombre de clusters. Droite verticale tracée pour 7 clusters. Résultats sur un échantillon aléatoire de 30 000 lactations du premier jeu de données.	8
Figure 16 : Courbes de lactation moyenne issues des 7 clusters retenus. Chaque point représente la moyenne de la production pour un stade donné des animaux d'un cluster. L'ensemble des lactations issues du jeu de donnée 1 a été pris en compte.....	8

Figure 17 : Représentation des courbes selon le modèle de Wilmink pour les différents rangs de lactation. Les points représentent la moyenne des contrôles par stade et par rang de lactation (modèle 0)	12
Figure 18 : Représentation des courbes selon le modèle d'Ali et Schaeffer pour les différents rangs de lactation. Les points représentent la moyenne des contrôles par stade et par rang de lactation (modèle 0)	12
Figure 19 : Représentation des courbes selon le modèle des polynômes de Legendre « leg5 » pour les différents rangs de lactation. Les points représentent la moyenne des contrôles par stade et par rang de lactation (modèle 0)	12
Figure 20 : Représentation des courbes selon le modèle des polynômes fractionnaires pour les différents rangs de lactation. Les points représentent la moyenne des contrôles par stade et par rang de lactation (modèle 0)	12
Figure 21 : Représentation des courbes selon le modèle à 236 classes fixes (« fix1 ») selon le rang de lactation. Les points représentent la moyenne des contrôles par stade et par rang de lactation (modèle 0)	13
Figure 22 : Représentation des courbes selon le modèle à 67 classes fixes (« fix5 ») selon le rang de lactation. Les points représentent la moyenne des contrôles par stade et par rang de lactation (modèle 0)	13
Figure 23 : Représentation des courbes selon le modèle splines cubiques pour les différents rangs de lactation. Les points représentent la moyenne des contrôles par stade et par rang de lactation (modèle 0)	13
Figure 24 : Représentation des résidus moyens des différents modèles (modèle 0) en fonction du stade de lactation	13
Figure 25 : Résidus moyens lissés par classes de 5 jours pour le modèle 4 selon la modélisation utilisée. Le deuxième jeu de données a été utilisé.	16
Figure 26 : Courbe de lactation modélisée avec le modèle 4 utilisant Ali et Schaeffer en fonction du stade de la gestation pour les chèvres Alpines en troisième lactation et plus. Le deuxième jeu de données a été utilisé.	17
Figure 27 : Courbe de lactation modélisée avec le modèle 4 utilisant Ali et Schaeffer en fonction du mois de mise bas pour les chèvres Alpines en troisième lactation et plus. Le deuxième jeu de données a été utilisé.	17
Figure 28 : Courbe de lactation modélisée avec le modèle 4 utilisant Ali et Schaeffer en fonction de l'âge à la mise bas pour les chèvres Alpines en troisième lactation et plus. Le deuxième jeu de données a été utilisé.	17
Figure 29 : Courbe de lactation modélisée avec le modèle 4 utilisant Ali et Schaeffer en fonction de la durée de tarissement pour les chèvres Alpines en troisième lactation et plus. Le deuxième jeu de données a été utilisé.	17
Figure 30 : Production mesurée (), production estimée (avec le modèle 4 utilisant Ali et Schaeffer) () d'une chèvre Alpine en deuxième lactation du NO ayant mis en troisième lactation en janvier, âgée de 39 mois, qui a été tarie 53 jours entre la deuxième et la troisième lactation et qui a fini sa troisième lactation gestante de 91 jours. L'effet de l'HTD au moment de ses contrôles (), les fonctions des différents facteurs (traits pleins) et les interactions des différents facteurs (traits pointillés) sont également représentés.	18

Liste des tableaux

Tableau 1 : Caractéristiques des données sélectionnées en fonction des classes d'effets de milieu	5
Tableau 2 : Répartition, en pourcentages, des lactations selon l'âge à la mise bas en fonction du mois de mise bas pour les chèvres en première lactation.....	5
Tableau 3 : Critères de choix (Logarithme de la fonction de vraisemblance, AIC et BIC) des différents modèles testés	9
Tableau 4 : Composition des clusters dits groupe "plate" ou groupe "pic" en pourcentage des modalités. Dans la colonne significativité, deux modalités sont différentes si elles n'ont pas de lettre en commun	10
Tableau 5 : Nombre de paramètres, logarithme de la fonction de vraisemblance ($\log(L)$), critère d'information d'Akaike (AIC), critère d'information bayésien (BIC), écart type de l'erreur (σ) pour les différents modèles retenus. Ils sont de la forme $y=f(\text{lact})+e$. Les données utilisées sont issues du jeu de données 1.....	11
Tableau 6 : Nombre de paramètres, logarithme de la fonction de vraisemblance ($\log(L)$), critère AIC, critère BIC, écart type de la variance résiduelle (σ), et qualité d'ajustement du modèle (R^2) pour différents modèles utilisant différentes fonctions. Les données sont issues du jeu de données 2.....	14
Tableau 7 : Nombre de paramètres, logarithme de la fonction de vraisemblance ($\log(L)$), critère AIC, critère BIC, écart type de la variance résiduelle (σ), et qualité d'ajustement du modèle (R^2) pour différents modèles. Les données sont issues du jeu de données 2 pour la race Alpine.....	15

Liste des annexes

Annexe 1 : Liste des numéros des départements par région.....	24
Annexe 2 : Délimitation des 4 régions françaises.....	24
Annexe 3 : Nombre de paramètres, logarithme de la fonction de vraisemblance ($\log(L)$), critère d'information d'Akaike (AIC), critère d'information bayésien (BIC), écart type de l'erreur (σ) pour les différents modèles testés. Ils sont de la forme $y_i = f(lact_i) + e_i$. Les données utilisées sont issues du jeu de données 1.....	25
Annexe 4 : Choix de l'emplacement des nœuds des splines à partir du modèle 4 et des données du deuxième jeu de données pour la race Alpine.....	26

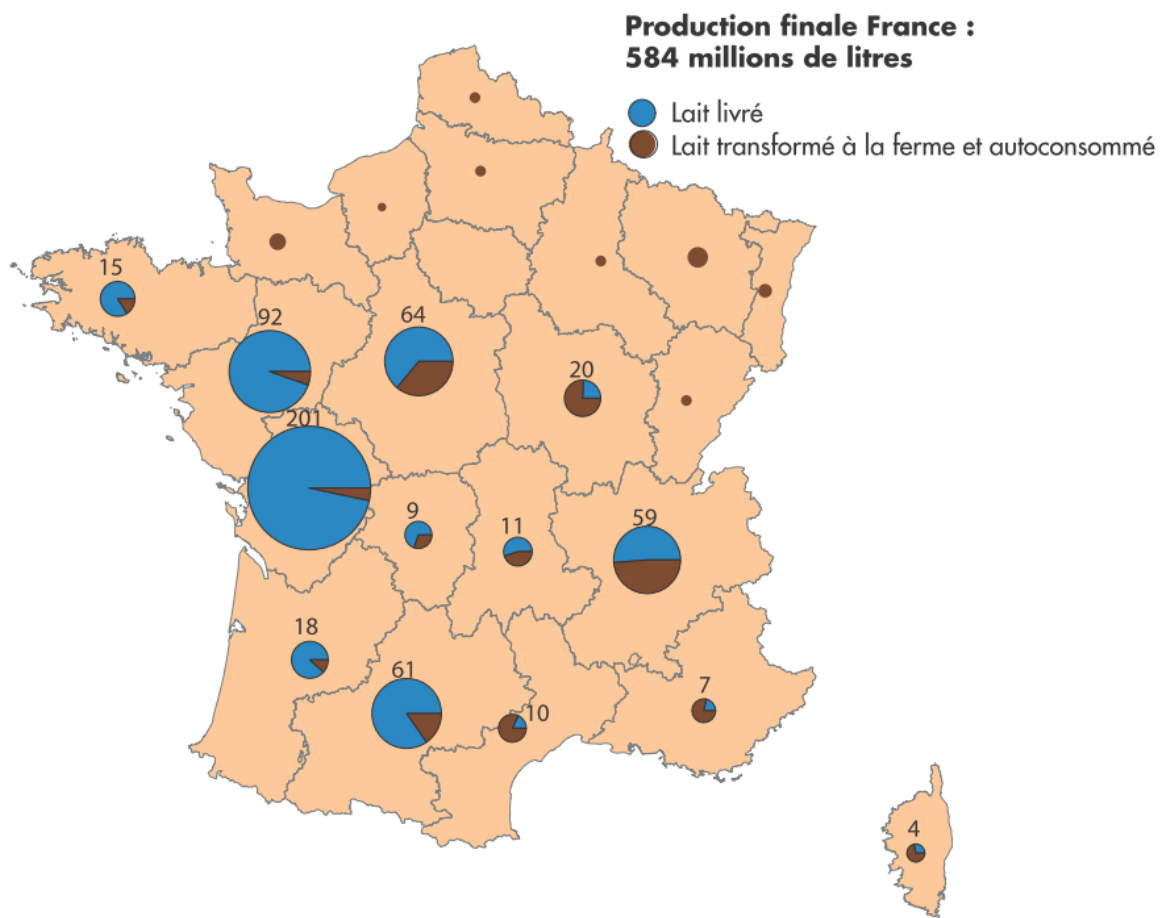


Figure 1 : Répartition de la production de lait de chèvres en France en 2014 (Institut de l'élevage, 2015)

Introduction

La France est le premier producteur de lait de chèvre en Europe avec 598 000 tonnes devant l'Espagne et la Grèce (Institut de l'élevage, 2015). Cette production est réalisée par 866 000 chèvres réparties en : (i) 3000 exploitations livrant directement leur lait à des laiteries ; elles détiennent 70% du cheptel avec en moyenne 225 chèvres par troupeau et produisent 80% du volume français et (ii) 2900 exploitations transformant une partie ou la totalité de leur production, elles possèdent en moyenne 79 chèvres (Institut de l'élevage, 2015). Un tiers des chèvres sont élevées en Poitou-Charentes. Les régions Pays de la Loire, Rhône-Alpes, Centre et Midi-Pyrénées concentrent chacune environ 12% de l'effectif national. L'élevage caprin en France est contrasté d'une région à l'autre : (i) en Poitou-Charentes, Pays de Loire et Centre les éleveurs produisent 70 % de la production nationale , ils sont livreurs, spécialisés ou associent à leur élevage de chèvres, un élevage de bovins ou des cultures, ils ont de gros troupeaux à l'exception des éleveurs fromagers de la région Centre ; (ii) dans les zones sèches du Sud-Est, les éleveurs ont des troupeaux de petites tailles, transforment leur lait en fromages sous signe de qualité et valorisent des parcours ; (iii) dans les systèmes des zones herbagères (Massif- Central, Bourgogne, Midi-Pyrénées, Rhône-Alpes) les éleveurs sont spécialisés (Sud-ouest) ou non, un tiers transforme leur lait en fromage sous signe de qualité (Peyraud *et al.*, 2013) (Figure 1). En France, dans des systèmes qui visent une production laitière importante par chèvre, la production moyenne de lait se situe entre 800 et 1100 kg/chèvre/an (Lefrileux *et al.*, 2012) et le pâturage est peu utilisé (seul 8% du cheptel du Poitou-Charentes pâture (Agreste, 2012)).

Un peu plus de 1600 élevages adhèrent au contrôle de performances officiel, ce qui représente 274 541 lactations par an (Institut de l'élevage, 2014). Toutes les 4 à 5 semaines, un technicien mesure la production laitière des chèvres lors d'un contrôle élémentaire (Ministère de l'agriculture, de l'agroalimentaire et de la forêt, 2014). Lors du contrôle laitier, la pesée du lait peut être réalisée soit lors des 2 traites matin et soir (protocoles laitiers : notés A, D, et AZ) soit lors d'une seule traite (protocole noté AT : en alternance le soir et le matin). En 2014, 69 000 inséminations artificielles ont été réalisées dans 863 élevages. La sélection est organisée par Capgènes, seul organisme et entreprise de sélection caprin en France. Le taux d'insémination artificielle est de l'ordre de 20% dans la population au contrôle laitier et de 40% dans le noyau de sélection (éleveurs adhérents à Capgènes) (Larroque *et al.*, 2014). Les deux principales races sont l'Alpine (59%) et la Saanen (38%). Les autres races sont des races locales. Grâce aux contrôles de performances et à une grande pression de sélection : 1000 chèvres sont sélectionnées dans la base de sélection (qui en comprend 170 000) pour produire des boucs candidats qui après plusieurs étapes de sélection sont seulement 40 à être agréés en tant que mâles d'insémination artificielle.

Dans le schéma de sélection actuel, la sélection pour l'amélioration de la quantité de lait par chèvre est basée sur la quantité totale de lait produite durant la lactation, période entre la mise bas et le tarissement de l'animal. Cette quantité est calculée selon la formule de Fleischmann (1932)

$$Q_{lact} = I_0 \times Q_1 + \sum_{i=1}^{n-1} (I_i \times \frac{Q_i + Q_{i+1}}{2}) + I_n \times Q_n$$
 avec I_0 l'intervalle entre la mise bas et le premier contrôle, I_i l'intervalle entre les contrôles i et $i+1$, I_n l'intervalle entre le dernier contrôle et le tarissement, Q_i la quantité de lait produite au $i^{ième}$ contrôle.

Cette quantité de lait produite est expliquée par des facteurs génétiques et d'environnement de l'animal. A partir des performances et des généalogies, on estime simultanément les effets d'environnement et la valeur génétique additive des animaux, qui est transmissible en espérance pour moitié à la descendance. Actuellement, les effets environnementaux pris en compte dans l'évaluation

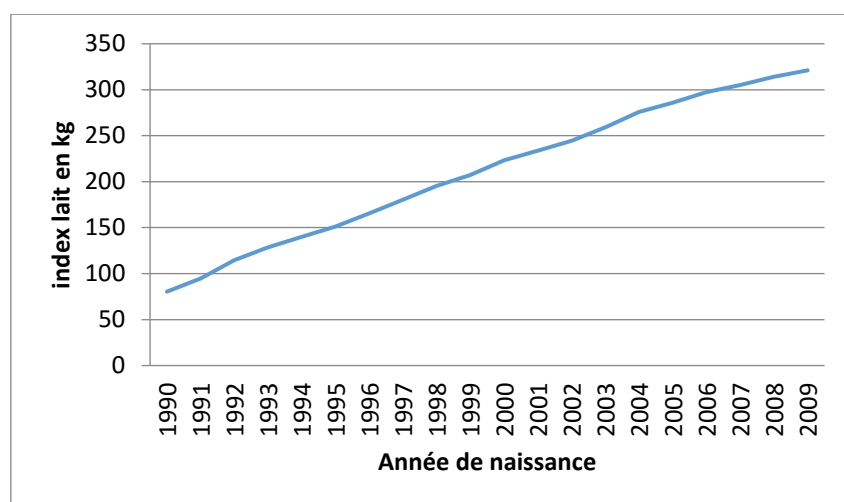


Figure 2 : Evolution de l'index lait entre 1990 et 2009 pour les chèvres de race Alpine (Larroque *et al.*, 2011)

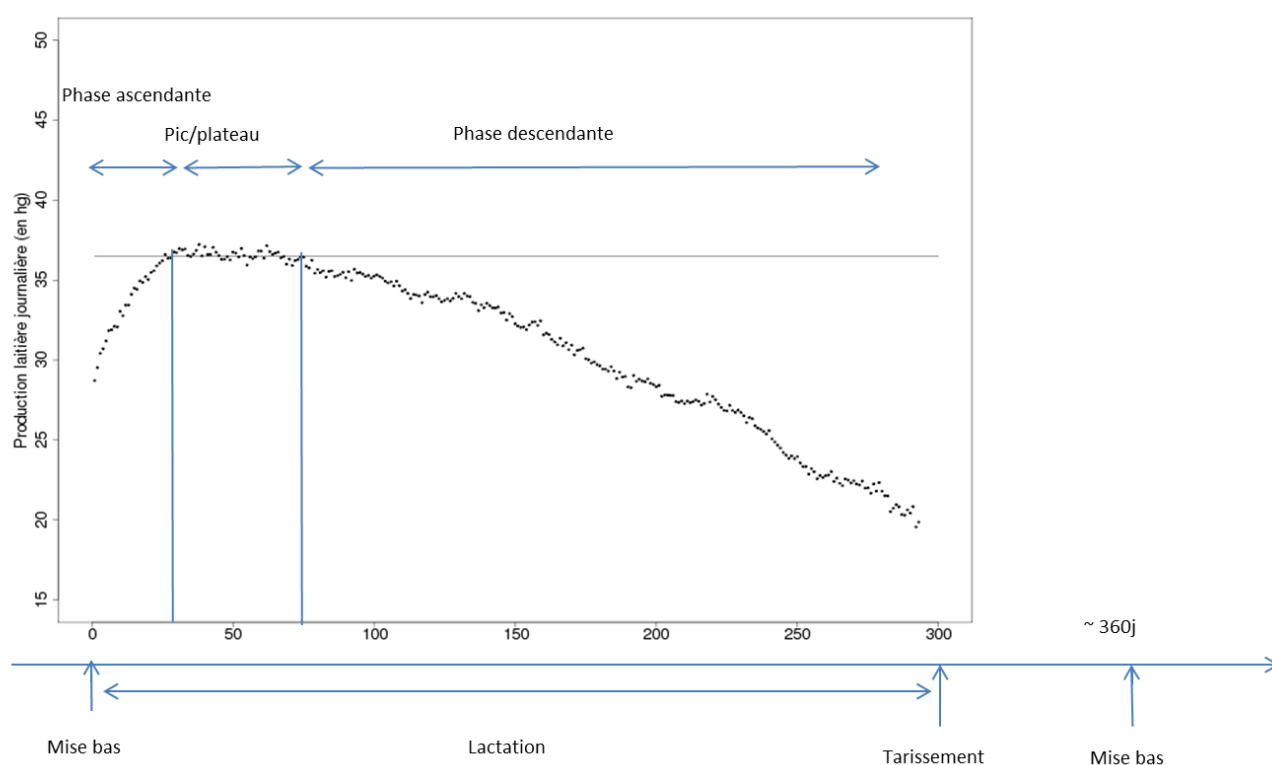


Figure 3 : Quantité de lait moyenne produite par jour en hg en fonction du stade de lactation en jours (les trois principales phases de la lactation sont indiquées). La droite $y = 36.5$ hg identifie le niveau de production au plateau de la lactation.

génétique officielle en France sont, en plus d'un effet d'environnement permanent (propre à l'animal mais non génétique), l'âge de l'animal à la mise bas, le mois de mise bas, la durée de tarissement et le numéro du cheptel. Ces effets de milieu sont en interaction avec la campagne, le rang de lactation et la région. La durée de tarissement est en interaction en plus avec la région d'élevage. Des modalités de ces facteurs ont été regroupées en différentes classes. La sélection génétique a permis de réaliser un progrès génétique de 13 kg de lait par an en moyenne, ce qui a permis d'augmenter la productivité laitière de manière conséquente (+270 kg de lait par lactation de progrès génétique en 22 ans pour les chèvres de race Saanen (Figure 2)).

Actuellement, la sélection génétique pour la production laitière est basée uniquement sur la quantité totale de lait produite et ne tient pas compte de l'évolution de la quantité de lait journalière au cours du stade de lactation (c'est-à-dire en fonction du nombre de jours écoulés depuis la mise bas). La courbe de lactation peut être décomposée en 3 phases (Figure 3): une phase ascendante où la production laitière augmente, en lien avec une activation rapide des cellules spécialisées épithéliales de la glande mammaire, jusqu'à atteindre un pic qui peut parfois se maintenir et former la phase de plateau, et enfin une phase descendante plus longue qui dure les deux tiers de la lactation et correspond à une phase de régression cellulaire (Macciotta *et al.*, 2008). La production totale au cours de la lactation, la date et le niveau de production atteint au début du pic de lactation, et le degré de maintien de la production du pic, appelé persistance permettent de décrire la courbe de lactation (Macciotta *et al.*, 2008).

Des facteurs environnementaux influencent la forme de la courbe de la lactation. En France, Bouloc (1991) les a étudiés sur la base des lactations de 146 000 chèvres en 1987. Elle a mis en évidence les effets: (i) du rang de lactation : les chèvres en première lactation (primipares) ont une persistance plus importante, un pic plus précoce et une production totale moins importante, par rapport aux chèvres en deuxième lactation ou plus (multipares), en accord avec d'autres études (Gipson et Grossman, 1990 ; Fernández *et al.*, 2002 ; Macciotta *et al.*, 2005 ; León *et al.*, 2012) ; (ii) du troupeau (regroupe différents facteurs comme l'alimentation, l'état sanitaire, ...) qui explique une part importante de la persistance ; (iii) de l'âge à la mise bas qui entraîne une meilleure persistance pour les animaux ayant mis-bas précocement et des productions plus élevées pour les animaux plus tardifs ; (iv) du mois de mise bas : les mises bas d'octobre à janvier entraînant une persistance plus importante, résultats en accord avec Montaldo *et al.* (1997) et León *et al.* (2012) ; (v) de la race avec une meilleure persistance pour les chèvres de race Saanen ; (vi) de la région avec un niveau production plus élevé au Nord-Ouest qu'au Sud-Est.

Bouloc (1991) a également étudié la diversité des types de courbes rencontrés en réalisant une classification sur la base de la persistance, du niveau de production et de la durée de lactation. Elle a ainsi mis en évidence des courbes de lactation atypiques caractérisées par un niveau de production médiocre en début de lactation mais qui ensuite augmente et compense le début de la lactation. Macciotta *et al.* (2008), en étudiant l'ajustement des courbes de lactations de 389 chèvres de race Frisa, ont aussi relevé des courbes atypiques (sans pic de lactation) dont l'ajustement au modèle était très faible ($R^2 < 0,2$).

La dernière phase de la courbe de la lactation est particulièrement intéressante pour plusieurs raisons. Premièrement, à même production totale sur une lactation, une chèvre persistante aura un niveau de production au pic de lactation moins élevé. Ceci est particulièrement intéressant car un niveau élevé de production au pic peut entraîner un déficit énergétique chez la chèvre dû à son incapacité à ingérer suffisamment d'aliment pour compenser sa grande production. La chèvre puise

alors dans ses réserves corporelles ce qui peut entraîner des problèmes métaboliques et reproductifs (Gipson et Grossman, 1990). Pour réduire ce déficit énergétique, les nutritionnistes ont préconisé la consommation de fourrages de haute valeur énergétique et de concentrés. A l'inverse, une chèvre persistante s'accommode de fourrages de moins bonne qualité et de moins de concentrés (Sölkner et Fuchs, 1987) ce qui, dans un contexte de réduction des intrants et d'accroissement des contraintes agro-environnementales, est un atout. Les chèvres dessaisonnées persistantes, qui mettent bas à l'automne (aux alentours du 1^{ier} novembre), peuvent mieux valoriser le pâturage du printemps car leur besoin est toujours important en fin de lactation. Une meilleure persistance entraîne également une production mieux répartie sur l'année ce qui intéresse les fromagers et les laiteries.

Il existe plusieurs critères définissant la persistance d'une lactation (Grossman *et al.*, 1999). Certains auteurs la définissent comme un rapport de productions entre différentes périodes de la lactation. Par exemple, Keown *et al* (1986) ont choisi la moyenne de la production par jour entre les jours 1 et 300 divisée par la production du jour 300. Roumeas *et al* (2014) ont utilisé le rapport entre le cumul de la production laitière observée entre 100 et 200 jours et entre 0 et 100 jours. D'autres ont cherché à modéliser la forme entière de la courbe de lactation et ont donc définie la persistance à partir des paramètres de leur modèle comme Wood (1967) ou Gipson et Grossman (1990).

De nombreux modèles mathématiques ont été proposés dans la littérature pour modéliser la courbe de lactation. Ils peuvent être classés en deux grandes catégories : (i) les modèles mécanistes qui s'intéressent au métabolisme tissulaire sous-jacent, (ii) les modèles empiriques qui ne prennent en compte que les performances enregistrées en ferme pour prédire la forme de la lactation. Les modèles mécanistes sont difficiles à mettre en œuvre car ils sont complexes et cette complexité n'est pas nécessaire lorsqu'on ne s'intéresse pas aux raisons biologiques qui infèrent sur la courbe de lactation (Leclerc 2008, Macciotta *et al.* 2008). Les modèles empiriques se subdivisent en trois groupes : (i) les modèles paramétriques (Ali et Schaeffer, 1987 ; Wilmink, 1987) nécessitant peu de paramètres pour ajuster la courbe de lactation mais dont l'interprétation biologique est parfois compliquée, et l'ajustement aux données, principalement aux extrémités de la courbe, difficile, (ii) les modèles semi-paramétriques représentés par les splines (fonctions polynômiales par morceaux reliées au niveau de nœuds), qui sont un bon compromis par rapport aux modèles précédents en terme de nombre de paramètres et d'ajustement aux données (White *et al.*, 1999) et enfin (iii) les modèles non paramétriques nécessitant un plus grand nombre de paramètres mais très flexibles (Druet *et al.*, 2003). Ces modèles peuvent avoir plusieurs applications d'appui technique: révéler des maladies lorsque l'animal s'écarte brutalement de la courbe moyenne prédite (ex : mammite sub-clinique) (Bouloc, 1993), ajuster l'alimentation selon les besoins de l'animal et aider les techniciens ou les économistes à prévoir la production de l'animal ou du troupeau (Leclerc, 2008).

L'objectif de mon étude est de modéliser la forme de la courbe de lactation à des fins d'évaluation génétique et notamment pour évaluer la part génétique de la persistance, à partir des informations recueillies lors de chaque contrôle en ferme. Les modèles sous-jacents permettant de traiter des données longitudinales (mesures répétées sur un même individu) sont généralement appelés « Test-Day Model » ou TDM, pour contrôles élémentaires (Schaeffer et Dekkers, 1994). Ils prennent en compte la notion de répétabilité des performances dans la mesure où ces observations, mesurées sur un même individu, ne sont pas indépendantes les unes des autres. L'un des principaux intérêts de ces modèles est une meilleure prise en compte d'effets d'environnement (comme l'alimentation ou la météorologie), affectant la production le jour du contrôle, avec un effet troupeau-jour de contrôle (HTD, Herd-Test Day) ne dépendant que des animaux présents à chaque contrôle (Meyer *et al.*, 1989 ; Ptak et Schaeffer, 1993). L'effet HTD est particulièrement intéressant lorsque les mises-bas ne sont pas

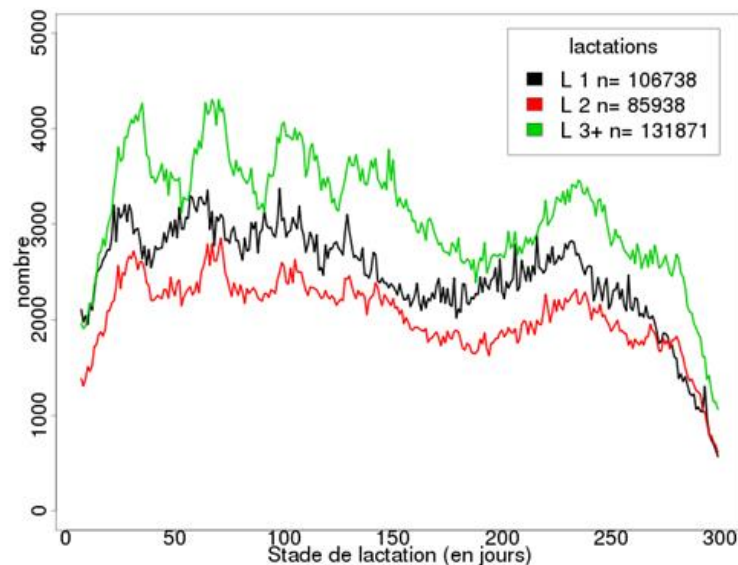


Figure 4 : Nombre de données par stade de lactation selon le rang de lactation (première lactation (L1), deuxième lactation (L2) et troisième lactation et plus, (L3+) dans le deuxième jeu de données.

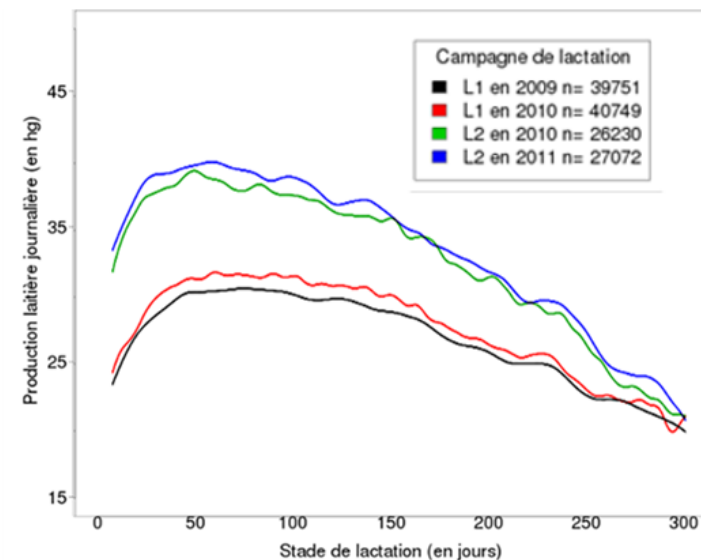


Figure 5 : Quantité de lait (en hg) journalière moyenne lissée en fonction du stade de lactation (en jours) par des chèvres en première (L1) et deuxième lactation (L2) pour les campagnes de mise bas 2009 à 2011 (premier jeu de données)

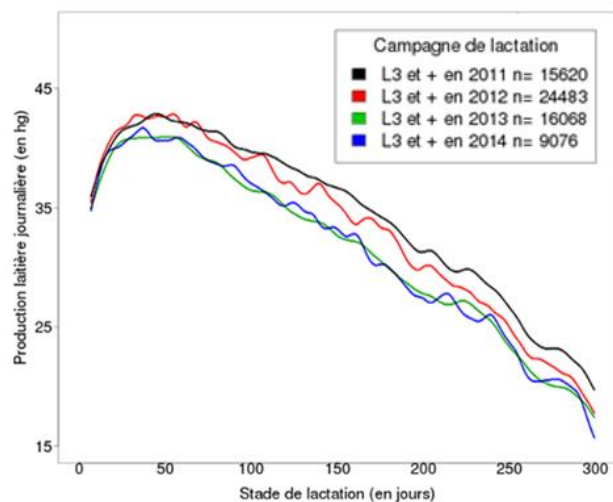


Figure 6 : Quantité de lait (en hg) journalière moyenne lissée en fonction du stade de lactation (en jours) par des chèvres en troisième lactation et plus (L3 et +) pour les campagnes de mise bas 2011 à 2014 (premier jeu de données)

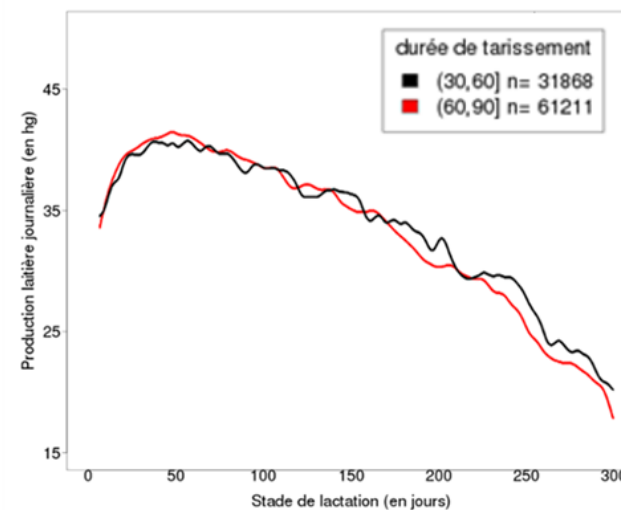


Figure 7 : Quantité de lait (en hg) journalière moyenne lissée produite en fonction du stade de lactation (en jours) selon la classe de durée de tarissement (premier jeu de données)

entièrement groupées. Dans le cadre d'un appui technique, il sera possible de le prédire et ainsi d'alermer l'éleveur si l'écart entre la prédiction et l'effet mesuré est trop grand (maladie dans le troupeau, alimentation, bâtiment, ...) (Leclerc, 2008). Le deuxième avantage de ce type de modèle réside dans la modélisation de la plupart des effets (génétiques ou d'environnement comme l'âge à la mise bas) sous forme de courbes c'est-à-dire dont l'influence sur le caractère varie au cours du stade de lactation (Leclerc, 2008). Ainsi il est possible d'obtenir la valeur génétique de la persistance d'un animal. La mise au point de ces modèles nécessite donc une première étape d'étude des effets d'environnement affectant la production laitière au cours du temps.

A partir de 2 échantillons de la base de données nationale caprine des contrôles laitiers élémentaires (plus de 62 millions de contrôles élémentaires actuellement stockés dans la base de données nationale caprine SIECL depuis 1988) mon stage a consisté à étudier l'effet de différents facteurs d'environnement sur les courbes de lactation; de proposer une classification des courbes individuelles afin de mettre en évidence différents profils types, et enfin de proposer des premiers modèles prenant en compte plusieurs effets d'environnements.

Matériels et méthodes

1. Sélection de deux jeux de données

Les données utilisées ont été extraites de la base de données nationale SIECL qui centralise les données du contrôle laitier caprin en France (plus de 62 millions de contrôles élémentaires actuellement stockés dans la base de données depuis 1988). En raison de la masse considérable de données disponibles et la nécessité d'optimiser les temps de calculs, deux jeux de données ont été utilisés au cours de mon stage en fonction des objectifs poursuivis. Le premier correspond à l'ensemble des données de lactation jusqu'en 2015, retenues pour l'évaluation génétique officielle pour la production laitière totale, de chèvres ayant débuté leur première lactation en 2009 ou 2010. Ce jeu de données permettra d'étudier l'évolution de la forme de la courbe de la lactation selon le rang de lactation des chèvres. Le second jeu contient les données de toutes les lactations débutées entre 2009 et 2011, quel que soit le rang de lactation des chèvres, retenues pour l'évaluation génétique officielle pour la production laitière totale. Il permettra de mimer l'hétérogénéité des données rencontrées lors des évaluations génétiques.

Pour ces deux jeux de données, les principaux critères de sélection des données ont été les mêmes. Seules les lactations des chèvres des deux principales races (Saanen et Alpine) ont été conservées pour l'étude. Afin d'avoir une estimation de la production par jour la plus précise possible, seules les lactations avec des productions laitières pesées en majorité lors des 2 traites quotidiennes ont été conservées (protocoles A, D, AZ). Dans l'objectif d'avoir des lactations de durée homogène et relativement classique, seules les lactations de plus de 180 jours et de moins de 350 jours ont été conservées (82% des lactations ont été conservées). Les contrôles laitiers, moins nombreux, réalisés après le 300^{ième} jour de lactation ont été éliminés afin de diminuer la variabilité en fin de lactation. Les contrôles réalisés moins de 7 jours après le début de la lactation ont aussi été supprimés en accord avec le cahier des charges du contrôle laitier (Ministère de l'agriculture, de l'agroalimentaire et de la forêt, 2014). Spécifiquement au premier jeu de données, et afin d'avoir des effectifs suffisants pour l'étude de l'effet des campagnes de mise bas, seules les campagnes 2010 et 2011 ont été gardées pour les deuxième lactations, et les années 2011, 2012, 2013, 2014 pour les troisième lactations.

Afin d'apprécier correctement la forme de la courbe de lactation, 3 autres critères de sélection ont été appliqués. Lorsque le premier contrôle de la lactation a eu lieu 81 jours après le début de la lactation, la lactation entière a été éliminée afin de bien caractériser le pic de la lactation. Les lactations

Tableau 1 : Caractéristiques des données sélectionnées en fonction des classes d'effets de milieu

Facteurs	Répartition jeu 1	Répartition jeu 2
Rang de lactation (1ère, 2ième, 3ième, 4ième, 5ième, 6ième, 7ième, 8ième)	40.7%, 26.9%, 17.0%, 9.5%, 4.7%, 1.2%, 0%, 0%	32.9%, 26.5%, 17.8%, 10.9%, 6.1%, 3.3%, 1.7%, 0.8%
Groupe de lactation (1 ^{ère} , 2ième, 3 +, 3++)	40.7%, 26.9%, 20.0%, 12.3%	32.9%, 26.5%, 40.6%, 0%
Carrière (1, 2, 3, 4, 5, 6, 7 lactations)	33.3%, 26.2%, 18.7%, 11.3%, 7.9%, 2.6%, 0.0%	
Région (NO, NE, SE, SO)	80.2%, 15.4%, 2.4%, 2%	76.8%, 17.9%, 3.2%, 2.1%
Race (Alpine, Saanen)	60.1%, 39.9%	59.3%, 40.7%
Age à la mise bas en première lactation (10-11 mois, 12 mois, 13 mois, 14-18 mois, autres mois)	21.6%, 39%, 13.7%, 10.4%, 15.4%	21.7%, 38.6%, 13.7%, 10.1%, 15.9%
Age à la mise bas en deuxième lactation (16-23 mois, 24-25 mois, 26-33 mois, 34-56 mois, autres mois)	34.3%, 41.1%, 8.4%, 4.2%, 12%	30.7%, 36.3%, 8.6%, 8.2%, 16.2%
Age à la mise bas en troisième lactation (22-34 mois, 35-38 mois, 39-77 mois, autres mois)	5.4%, 67.7%, 15.8%, 11.1%	6.2%, 61.1%, 15.7%, 17.1%
Mois de mise bas (janvier, février, mars, avril-mai, septembre-décembre)	25.9%, 29.4%, 14.9%, 5.9%, 23.9%	24.3%, 27.9%, 14.4%, 6%, 27.4%
Durée de tarissement (en jours) ([30,90], [90,120], autre durée)	46.8%, 8.3%, 45%	50.3%, 8.9%, 40.8%
Stade de gestation (en jours) ([1,30], [30,60], [60,90], [90,150], non gestante)	1.9%, 11.8%, 38.8%, 10.2%, 37.2%	2%, 11.6%, 37.8%, 9.9%, 38.8%
Production en kg en première lactation ([0,800], [800,1000], [1000,1200], [1200,2500].)	29%, 35%, 25%, 12% (moyenne 932 ±230kg)	28%, 34%, 25%, 13% (moyenne 935 ±233 kg)
Production en kg en deuxième lactation ([0,800], [800,1000], [1000,1200], [1200,2500].)	17%, 28%, 30%, 25% (moyenne 1035 ±255 kg)	22%, 30%, 28%, 21% (moyenne 1000 ±257 kg)
Production en kg en 3+ ([0,800], [800,1000], [1000,1200], [1200,2500].)	22%, 28%, 27%, 23% (moyenne 1010 ±264 kg)	28%, 29%, 25%, 18% (moyenne 959 ±265 kg)
Production en kg en 3++ ([0,800], [800,1000], [1000,1200], [1200,2500].)	29%, 31%, 24%, 17% (moyenne 951 ±260 kg)	

Tableau 2 : Répartition, en pourcentages, des lactations selon l'âge à la mise bas en fonction du mois de mise bas pour les chèvres en première lactation

	10-11 mois	12 mois	13 mois	14-18 mois	autres mois
Janvier	31.9%	22.1%	5.7%	11.2%	14.9%
Février	24.5%	36.1%	28.1%	21.2%	28.9%
Mars	7.2%	13.8%	40.1%	36.8%	27.1%
Avril-Aout	5.3%	2.9%	10.1%	25.8%	11.7%
Septembre-Décembre	31.1%	25.2%	16.1%	5%	17.4%

avec moins de 4 contrôles élémentaires, trop peu nombreux pour apprécier correctement toute la forme de la courbe, ont été supprimées. Enfin, afin de bien estimer l'effet troupeau d'un jour de contrôle, si le nombre de chèvres contrôlées au sein d'un élevage, pour une race donnée, un jour de contrôle donné, était inférieur à 5, tous les contrôles de cet élevage pour ce jour de contrôle ont été supprimés. Les trois derniers critères de sélection étant interdépendants, la sélection sur ces critères a été faite de manière itérative afin de tous les satisfaire.

Cette sélection a réduit considérablement les données d'origine puisque seulement 40% des données initiales ont été conservées. Le critère de sélection le plus sévère a été le nombre de pesées par jour car 46% des contrôles ont été écartés sur ce seul critère. A la fin de la sélection, le premier jeu se compose de 1 366 565 contrôles provenant 198 507 lactations, de 89 729 animaux. Le deuxième jeu comprend 2 245 628 contrôles issus 324 547 lactations, de 213 534 chèvres.

2. Description des données issues de notre sélection et création des classes de facteurs

Au cours d'une lactation, en moyenne 6.9 contrôles sont réalisés (entre 4 et 11), ce qui représente un écart moyen entre contrôles de 39 jours. La figure 4 présente le nombre de contrôles par stade de lactation. La baisse du nombre de données entre les stades de lactation 150 et 210 jour correspond à l'absence de contrôle au mois d'août.

La répartition des données selon les différents critères de milieu est présentée dans le tableau 1. 60% des lactations ont été produites par des chèvres de race Alpine et 40% par des chèvres de race Saanen.

Dans l'évaluation génétique officielle sur la production laitière, la France est divisée en 4 grandes régions (cf. annexe 1 pour la composition en départements de chaque région, et annexe2 pour la carte de la France divisée en 4 régions). Les régions Sud-Ouest (SO) et Sud-Est (SE) ont été très pénalisées par nos critères de sélection car dans ces régions 94% et 86% des contrôles élémentaires, respectivement, sont issus de protocole de pesée AT. Avant sélection, dans nos deux jeux de données, 19% des lactations étaient réalisées en région SO, et 12% en région SE (13% dans le Nord-Est (NE) et 57% dans le Nord-Ouest (NO)), alors qu'après sélection la région SO ne représente plus que 2% des lactations et la région SE 3%. Après sélection, le Nord-Est et le Nord-Ouest représentent 16% et 78% des lactations respectivement.

Les rangs de lactation s'échelonnent de la première à la huitième lactation pour le deuxième jeu de données. Les chèvres du premier jeu de données sont en moyenne plus jeunes que celles du deuxième jeu car elles n'ont été suivies que jusqu'en 2014, tronquant ainsi leur carrière. Elles sont suivies sur 2.42 lactations en moyenne, et sont issues de 997 cheptels différents.

Le modèle d'évaluation génétique caprine comprend notamment comme effet d'environnement : le rang de lactation, l'âge à la mise bas, le mois de mise bas et la durée de tarissement. Ces facteurs ont été découpés en classes, en fonction de leurs effets sur la performance et de leurs effectifs. En partant du principe que ces facteurs ont un effet sur la production laitière globale, une étude préliminaire a consisté à visualiser les courbes de lactation moyennes, pour chacune de ces modalités de facteurs, à l'aide du logiciel R et de la fonction « smooth.spline » qui lisse la courbe moyenne des données par une fonction spline cubique.

Pour le premier jeu de données, après avoir tracé les courbes de lactation selon le rang de lactation et la campagne de mise bas, il n'a pas paru nécessaire de différencier les premières (L1) et les deuxièmes (L2) lactations selon les campagnes de mise bas () ; en revanche, les troisièmes lactations (L3) ont été séparées entre les campagnes 2011-2012 d'une part et les campagnes 2013-2014 d'autre

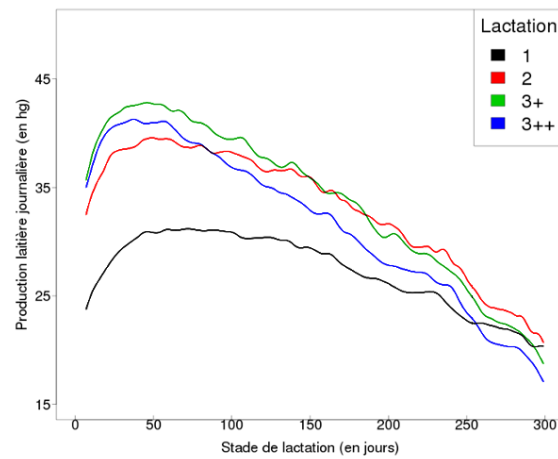


Figure 8 : Quantité de lait moyenne lissée (fonction R « smooth.spline ») produite par jour en hg en fonction du stade de lactation en jours et du rang de lactation pour le jeu de données 1.

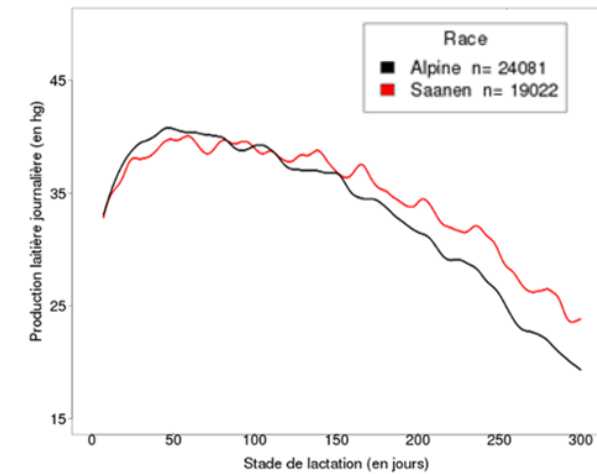


Figure 9 : Quantité de lait moyenne lissée (fonction R « smooth.spline ») produite par jour en hg en fonction du stade de lactation en jours par des chèvres en deuxième lactation issues de la région Nord-ouest de race Saanen ou Alpine pour le jeu de données 1.

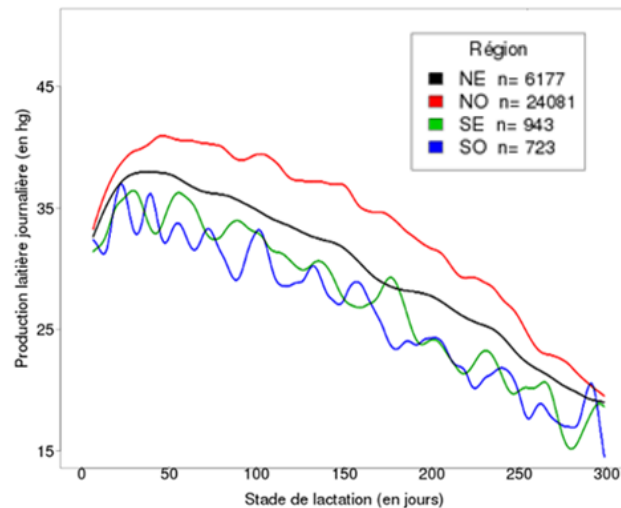


Figure 10 : Quantité de lait moyenne lissée (fonction R « smooth.spline ») produite par jour en hg en fonction du stade de lactation en jours par des chèvres en deuxième lactation de race Alpine selon la région d'élevage (NE :Nord-Est, NO :Nord-Ouest, SE :Sud-Est, SO :Sud-Ouest) pour le jeu de données 1.

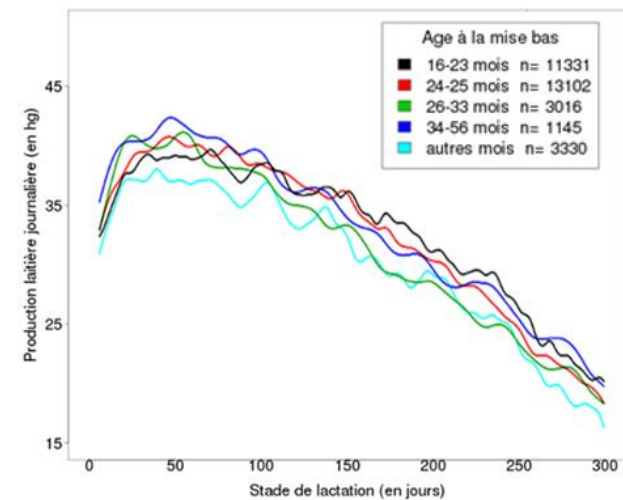


Figure 11 : Quantité de lait moyenne lissée (fonction R « smooth.spline ») produite par jour en hg en fonction du stade de lactation en jours par des chèvres en deuxième lactation de race Alpine selon l'âge à la mise bas en mois (jeu de données 1)

part (Figure 6) car elles avaient des profils différents. Les troisièmes lactations et plus de 2011-2012 ont été appelées les « L3+ » et celles de 2013-2014 les « L3++ ». Dans le deuxième jeu de données, la campagne ne semblait pas avoir d'effet sur les courbes de lactations.

La médiane de l'âge à la mise bas est de 12, 24, 36 mois respectivement en première, deuxième et troisième lactation. Le premier et le troisième quartile est égal à plus ou moins un mois de la médiane. La différenciation selon l'âge à la mise bas n'est réalisée que pour les trois premières lactations en accord avec l'indexation officielle.

L'étude de la durée de tarissement n'est possible que pour les chèvres en deuxième lactation et plus. Le nombre important (40%) de chèvres dans la catégorie « autre » est dû à la présence des premières lactations. Après avoir étudié les courbes de lactation (cf. Figure 7), il est apparu que les chèvres taries entre 30 et 60 jours avaient les mêmes courbes que les chèvres taries entre 60 et 90 jours, ces deux catégories ont donc été regroupées.

Contrairement aux facteurs précédents, l'effet de la gestation n'est pas intégré dans le modèle d'indexation officielle. Les courbes de lactation ont été tronquées au 300^{ième} jour et il aurait été difficile d'apprécier l'effet de la gestation pour les lactations se terminant après. Le stade de gestation a donc été défini au moment du tarissement, pour les chèvres ayant eu une durée de lactation inférieure à 300 jours, et au 300^{ième} jour de lactation pour les autres. 5 classes de stade de gestation ont été définies. Il a été calculé uniquement chez les chèvres ayant eu une lactation supplémentaire en considérant une durée de gestation moyenne de 150 jours. Les chèvres dites « non gestante » n'ont pas remis bas l'année d'après (ou cette information n'était pas disponible) ou bien étaient toujours vides, d'après les calculs, au 300^{ième} jour de lactation ou au tarissement. Lorsqu'une chèvre est gestante, la médiane du stade de gestation au tarissement ou au 300^{ième} jour de lactation est de 76 jours (1^{ier} quartile : 62 jours, 3^{ième} quartile : 86 jours).

La production totale produite lors d'une lactation est connue et a été classifiée en trois catégories (en kg) : [0,800],]800,1000],]1000,1200],]1200,2500].

Les différents facteurs ne sont pas indépendants les uns des autres. Le tableau 2 illustre la relation entre l'âge à la mise bas et le mois de mise bas. En effet, pour les premières lactations, par exemple, les chèvres chevrotant en janvier-février sont en moyenne plus jeunes que celles qui mettent bas entre mars et août. Les chèvres mettant bas entre mars et août représentent une part plus importante des lactations dans la région NE en comparaison avec les autres régions françaises. Il en est de même pour les deuxièmes lactations, où l'on constate aussi que les chèvres chevrotant précocement ont en moyenne un tarissement plus court.

3. Classification des profils individuels de courbes de lactation

L'objectif de la classification est d'explorer la variabilité des formes des courbes individuelles et de les caractériser. Il sera ainsi possible de catégoriser les formes de courbes et de repérer des courbes différentes des profils moyens de la partie précédente. Le premier jeu de données a été utilisé afin de pouvoir suivre l'évolution de la forme de la courbe de lactation des chèvres tout au long de leur carrière.

Pour déterminer la classification, nous avons utilisé le package R « kml » (Genolini *et al.*, 2015). Ce package requiert des données longitudinales. Dans notre cas, nous disposons pour chaque individu de 6.9 mesures en moyenne pour 294 stades possibles (entre le stade jour 7 et 300). Le package offre plusieurs méthodes d'estimation de ces données manquantes (ou imputation) et nous avons choisie celle recommandée par l'auteur. Cette méthode impute les données manquantes en prenant en

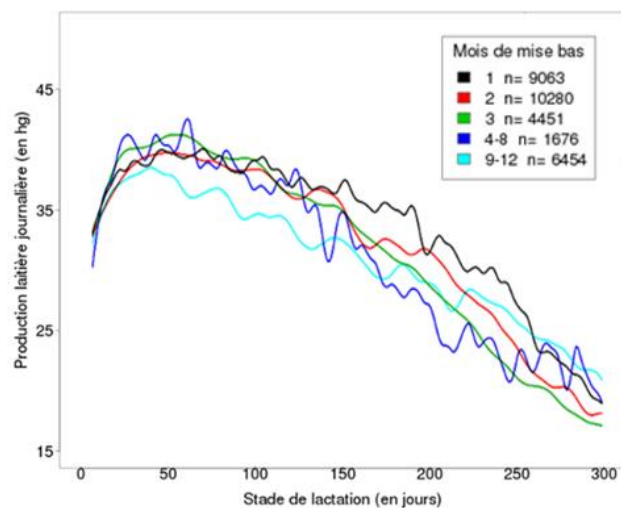


Figure 12 : Quantité de lait moyenne lissée (fonction R « smooth.spline ») produite par jour en hg en fonction du stade de lactation en jours par des chèvres en deuxième lactation de race Alpine selon le mois de mise bas (1 : Janvier, 2 : Février, 3 : Mars, 4-8 : Avril-Août, 9-12 : Septembre-Décembre) pour le jeu de données 1.

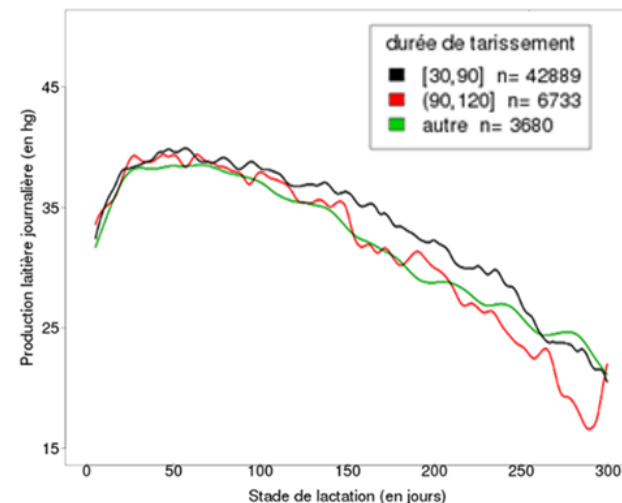


Figure 13 : Quantité de lait moyenne lissée (fonction R « smooth.spline ») produite par jour en hg en fonction du stade de lactation en jours par des chèvres en deuxième lactation selon la durée de tarissement en jours (jeu de données 1)

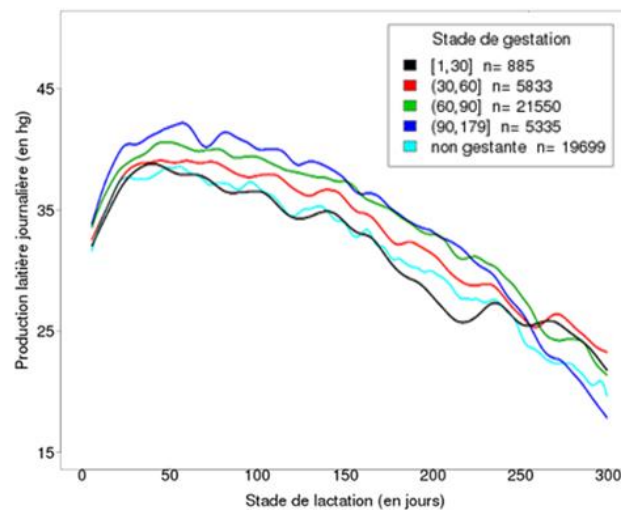


Figure 14 : Quantité de lait moyenne lissée (fonction R « smooth.spline ») produite par jour en hg en fonction du stade de lactation en jours par des chèvres en deuxième lactation selon la durée de gestation en jours (jeu de données 1)

compte la courbe moyenne de la population.

Le package partitionne des données longitudinales en utilisant la méthode des k-moyennes. Cette méthode est itérative. En premier lieu, tous les individus (ici lactations) sont assignés à un groupe (ou cluster). Ensuite deux phases se succèdent : une première où le centre de gravité du groupe est calculé et une deuxième où les individus sont à nouveau répartis au sein des nouveaux groupes. Ces deux étapes sont réalisées jusqu'à ce qu'il n'y ait plus de changements dans les clusters. L'objectif de la classification est donc de maximiser la variance inter groupe et minimiser la variance intra groupe. Cette méthode a pour prérequis le nombre de clusters. Ce nombre qu'il est possible de faire varier entre 2 et 26, doit être choisi par l'utilisateur en s'appuyant sur des critères de qualité proposés, dont:

- le critère de Calinsky et Harabasz (1974) modifié par Genolini (2015) appelé « Calinsky.Harabasz3 »: $C_G(k) = \frac{Trace(B)}{Trace(W)} \times \frac{n-k}{\sqrt{k-1}}$ où B est la matrice de covariance entre les clusters et W est la matrice de covariance intra-cluster, k le nombre de groupes et n le nombre de lactations. L'objectif est de maximiser ce critère.
- le critère de Ray et Turi (1999) : $R(k) = \frac{V_{intra}}{V_{inter}}$ avec $V_{intra} = \sum_x (dist(x, center(x)))$ représente la variance intra-cluster et $V_{inter} = \min_{i,j} (dist(center_i, center_j))^2$ représente la variance inter-clusters. L'objectif est de minimiser ce critère.
- le BIC (Schwarz, 1978) : $BIC = 2 \times \log(L) - h \times \log(n)$ avec L la fonction de vraisemblance et h le nombre de paramètres du modèle (cf. 4.c Matériels et Méthodes)
- le BIC2 : $BIC2 = 2 \times \log(L) - h \times \log(N)$ avec N le nombre d'observations = $n \times t$ avec t le nombre de stades de lactation
- l'AIC (Akaike, 1974) $AIC = 2 \times \log(L) - 2 \times h$ (cf. 4.c Matériels et Méthodes)

Les trois derniers critères sont pertinents pour des variables normales ce qui est le cas de la production de lait journalière. Afin de pouvoir comparer ces critères, ils ont été transformés sur une échelle de 0 à 1. Le nombre de clusters optimal est celui pour lequel le critère se rapproche de 1.

Les temps de calcul nécessaires pour la classification des courbes de lactation à partir de l'ensemble du premier jeu de données étant très longs, il n'était pas possible de réaliser l'ensemble des tests pour un panel de 2 à 26 clusters. Quatre échantillons de 30 000 lactations tirées au sort aléatoirement ont donc été étudiés afin de choisir le nombre optimal de cluster.

Une fois le nombre de clusters déterminé, nous avons caractérisé chacun d'eux en fonction des effets environnementaux dont nous disposons. L'appartenance à un cluster a été analysé suivant un modèle linéaire généralisé tel que:

$$(I. a) y = région + lact + race + age + mois + dtar + gest + prod_lact$$

avec y le numéro de cluster, *région* l'effet de la région, *lact* l'effet du rang de lactation, *race* l'effet race de l'animal, *âge* l'effet de l'âge à la mise bas, *mois* l'effet mois de mise bas, *dtar* l'effet de la durée de tarissement, *gest* l'effet du stade de gestation au tarissement ou au 300^{ème} jour de lactation, et *prod_lact* la production laitière totale de la lactation. Les effets étant définis comme au paragraphe 2 du Matériels et Méthodes. La significativité des différences de répartition au sein des clusters a été étudiée grâce au package R « lsmmeans » (Lenth, 2016). Dans les résultats, deux modalités seront interprétées comme différentes si elles n'ont pas de lettre en commun. Nous avons ensuite testé la significativité des différents effets introduits dans le modèle à l'aide du test du rapport de vraisemblance et des critères de choix de modèles tels que AIC et BIC (cf. paragraphe 4.c du Matériels

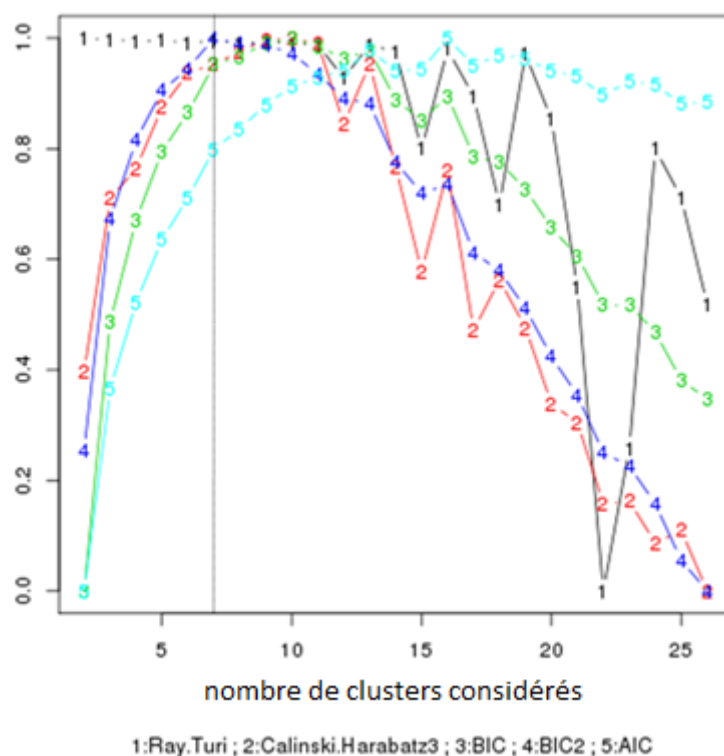


Figure 15 : Evolution de la valeur relative (échelle de 0 à 1) de différents critères de qualité en fonction du nombre de clusters. Droite verticale tracée pour 7 clusters. Résultats sur un échantillon aléatoire de 30 000 lactations du premier jeu de données.

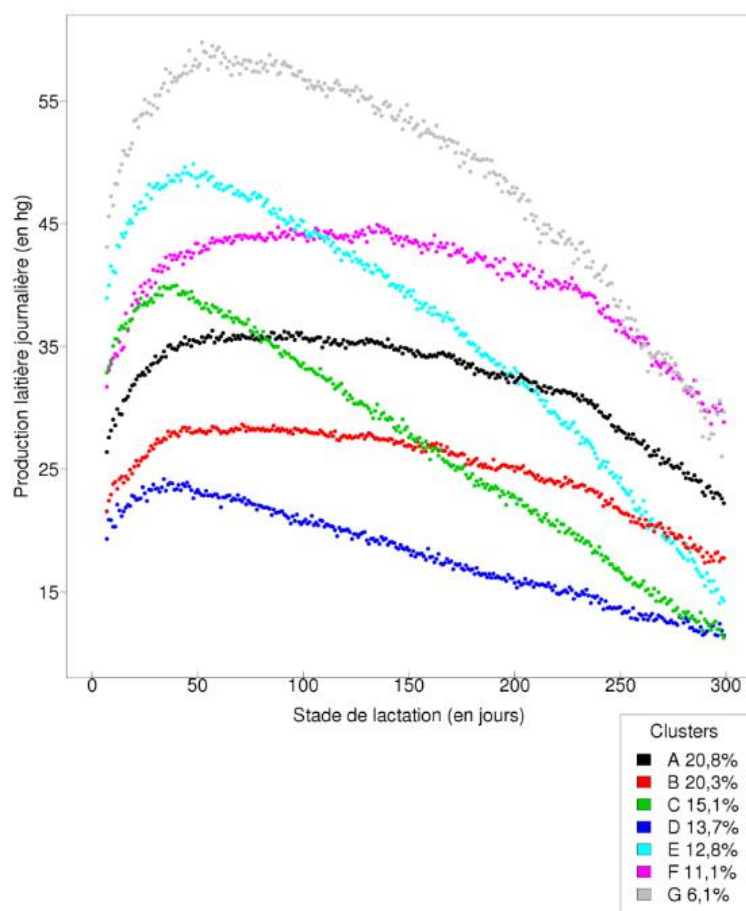


Figure 16 : Courbes de lactation moyenne issues des 7 clusters retenus. Chaque point représente la moyenne de la production pour un stade donné des animaux d'un cluster. L'ensemble des lactations issues du jeu de donnée 1 a été pris en compte.

et Méthodes).

Un modèle linéaire a été utilisé afin d'étudier l'influence de la forme de la courbe sur la production de lait sur les chèvres ayant eu au moins 5 lactations complètes (n=9487) :

(I. b) $y = \text{alternance des formes} + \text{région} + \text{race}$ avec y la production cumulée de lait sur les 5 premières lactations, *alternance des formes* la succession de clusters d'un animal durant ses 5 premières lactations (par exemple : cluster 1-cluster 1-cluster 3- cluster 7-cluster 7), *région* l'effet de la région, *race* l'effet race de l'animal.

4. Modélisation des courbes de lactation moyenne

Notre objectif est de modéliser la courbe de lactation des chèvres en prenant en compte plusieurs effets d'environnements. Les modèles devront être le plus précis possible tout en ayant un nombre de paramètres limités afin de réduire les temps de calculs.

4. a. Description des modèles

Les modèles utilisés afin de modéliser la courbe de lactation moyenne en fonction du stade de lactation 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 et $E(y_t)$ son espérance.

Les modèles paramétriques :

Ils permettent de résumer les caractéristiques de l'évolution de la production laitière au cours du temps par des fonctions linéaires ou non linéaires avec un petit nombre de paramètres. A partir de ces fonctions, la date du pic de production peut être obtenue. La production entre 2 dates est obtenue en calculant l'intégrale de la fonction sur cet intervalle de temps.

Le modèle de Wilmink (Wilmink, 1987) : $E(y_t) = a + bt - de^{-ct}$. Ce modèle, composé d'une fonction exponentielle, est un modèle non linéaire et a 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 (Ali et Schaeffer, 1987): $E(y_t) = a + bu + cu^2 + dv + ev^2$ où $u = \frac{t}{300}$ et $v = \ln(\frac{300}{t})$. Ce modèle 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). Le paramètre 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.

Le modèle des polynômes de Legendre (Kirkpatrick *et al.*, 1990),

$$E(y_i) = \sum_{i=0}^n \alpha_i \sqrt{\frac{2n+1}{2}} \times P_i(\omega) \text{ où } \omega = 2 \times \left(\frac{t-t_{\min}}{t_{\max}-t_{\min}} \right) - 1, t_{\min}=7, t_{\max} = 300,$$

$P_{n+1}(\omega) = \frac{1}{n+1} ((2n+1)\omega P_n(\omega) - nP_{n-1}(\omega))$, avec $P_0(\omega)=1$, n le degré du polynôme et α_i coefficient de régression. Une propriété intéressante des polynômes de Legendre est leur orthogonalité qui permet de minimiser les corrélations entre les coefficients de régression. Les polynômes de degrés 4 et 5 ont été testés car ils ont un nombre de paramètres équivalents aux autres modèles et seront appelés respectivement « leg4 » et « leg5 ».

Le modèle des polynômes fractionnaires (Royston et Altman, 1994):

$$E(y_t) = \sum_{i=0}^n \beta_i t^{p_i} \text{ où } p_0 = 0 \text{ et } t^{p_i} = \begin{cases} t^{p_i} \text{ si } p_i \neq 0 \\ t^{p_{i-1}} \ln(t) \text{ si } p_i = p_{i-1} \end{cases}$$

Tableau 3 : Critères de choix (logarithme de la fonction de vraisemblance (log(L)), critère d'information d'Akaike (AIC), critère d'information bayésien (BIC)) des différents modèles testés. Les données utilisées sont issues du jeu de données 1

	Log(L)	AIC	BIC	Classement ⁽¹⁾
Modèle complet^(a)	-98 356.2	196 776.5	197 102.9	
-région^(b)	-98 414.3	196 888.7	197 194.7	8
- âge à la mise bas^(b)	-98 812.3	197 664.6	197 868.6	5
-rang de lactation^(b)	-99 117.6	198 293.2	198 589	4
-mois de mise bas^(b)	-99 781.9	199 619.8	199 905.4	2
-stade de gestation^(b)	-98 575.0	197 206.1	197 491.6	6
-durée de tarissement^(b)	-98 424.8	196 909.7	197 215.6	7
-race^(b)	-99 497.3	199 056.8	199 372.9	3
-production_lactation^(b)	-108 823.3	217 704.7	218 000.4	1

(a) Modèle complet correspondant au modèle I a

(b) Modèle complet sans le facteur x (le facteur soustrait est indiqué sur la ligne du tableau.

(1) Importance du facteur enlevé (du plus important 1 au moins important 8)

Les polynômes fractionnaires ont pour propriété de pouvoir être de degré non entier, ce qui permet de définir des fonctions relativement flexibles pour modéliser la courbe de lactation moyenne. Par définition, pour un même degré n , plusieurs polynômes existent.

Les modèles semi-paramétriques : les splines

Les splines sont des fonctions polynomiales par morceaux dont la jonction se fait au niveau de nœuds. L'ajustement dépend du nombre, de la position des nœuds utilisés et du degré du polynôme.

Dans notre étude, nous avons choisi d'utiliser les splines cubiques, le degré des polynômes est 3. Les splines cubiques ont la propriété d'avoir leurs deux premières dérivées continues entre deux segments consécutifs. Elles 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. Après l'étude de ces modèles sur nos données, nous avons choisi de fixer le nombre de nœuds à 4 et plusieurs positions de nœuds ont été testées : (i) 60, 120, 180, 240, (ii) 80, 180, 200, 220, (iii) 80, 160, 180, 260, (iv) 30, 60, 100, 230, (v) 30, 60, 100, 200. Dans les modèles 4 et 5 (cf. 4.b du Matériels et Méthodes), pour la modélisation du stade de gestation, les nœuds ont été placés à la fin de la lactation, aux stades 200, 240, 260, 280, lorsque l'effet de la gestation sur la production laitière est le plus significatif.

Les modèles non paramétriques:

La modélisation de la production laitière sous forme de courbe par des modèles paramétriques ou semi-paramétriques permet de lisser les variations aléatoires observées et de résumer la courbe à l'aide de quelques paramètres. En revanche des biais peuvent être créés si l'ajustement aux données est mauvais. Les modèles non paramétriques permettent de modéliser la courbe de lactation sans a priori sur la forme de la courbe. Une façon de le faire simplement est de définir des classes de stade de lactation et d'estimer un effet par classe. L'inconvénient d'un tel modèle 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 modèles sont donc relativement flexibles mais très coûteux en nombre de paramètres à estimer et donc en degrés de liberté perdus.

Plusieurs classes de stades de lactation ont été créées afin d'estimer le moins de paramètres possibles tout en ayant un bon ajustement aux données :

- Modèle noté « **Fix1** » (236 classes) : définition de classes de 5 jours sur toutes les lactations,
- Modèle noté « **Fix2** » (124 classes) : définition de classes de 5 jours jusqu'au 30^{ème} jour de lactation, puis une classe de 10 jours ; une classe de 25 jours et enfin 24 classes de 10 jours jusqu'à la fin de la lactation,
- Modèle noté « **Fix3** » (92 classes): définition de classes de 5 jours jusqu'au 30^{ème} jour de lactation, puis une classe de 10 jours ; une classe de 25 jours et enfin 16 classes de 15 jours jusqu'à la fin de la lactation,
- Modèle noté « **Fix4** » (120 classes): définition de classes de 10 jours sur toutes les lactations,
- Modèle noté « **Fix5** » (67 classes): les classes sont définies intra lactation

1^{ère} lactation : classes de 5 jours jusqu'à 40^{ème} jour de lactation, puis [40-125], [125-140], [140-165], [165-180], [180-200], [200-230], [230-245], [245-280], [280-300]

2^{ème} lactation : classes de 5 jours jusqu'au 20^{ème} jour de lactation, puis [20-40], [40-80], [80-100], [100-115], [115-150], [150-165], [165-180], [180-200], [200-210], [210-240], [240-250], [250-260], [260-280], [280-300]

3+ (lactations 3 et plus de 2011 et 2012): classes de 5 jours jusqu'au 20^{ème} jour de lactation, puis

Tableau 4 : Composition des clusters dits groupe "plate" ou groupe "pic" en pourcentage des modalités. Dans la colonne significativité, deux modalités sont différentes si elles n'ont pas de lettre en commun

facteurs	modalités	pic	plat	significativité
Race	Alpine	38.40%	61.60%	a
	Saanen	27.60%	72.40%	b
Rang de lactation	1	13.90%	86.10%	a
	2	39.70%	60.30%	b
	3+	55%	45%	c
	3++	54.70%	45.30%	c
Région	NE	33.20%	66.80%	a
	NO	34.40%	65.60%	b
	S	31.70%	68.30%	c
Age à la mise bas	10-11 mois	8.30%	91.70%	g
	12mois	13%	87%	f
	13mois	19.30%	80.70%	bcd
	14-18mois	21.40%	78.60%	abc
	16-23mois	33%	67%	f
	24-25mois	42.90%	57.10%	de
	26-33mois	51.90%	48.10%	ab
	34-56mois	45.30%	54.70%	ab
	22-34mois	46.40%	53.60%	ef
	35-38mois	54.70%	45.30%	cde
	39-77mois	54.60%	45.40%	ab
	autres	26.50%	73.50%	e
Mois de mise bas	14et+	56.40%	43.60%	a
	1	34.10%	65.90%	c
	2	36.80%	63.20%	b
	3	38.10%	61.90%	a
	4-Aug	36.90%	63.10%	a
	9-Dec	27.50%	72.50%	c
Durée de tarissement	[30,90]	46.80%	53.20%	a
	(90,120]	54.90%	45.10%	b
	autre	17%	83%	a
Durée de gestation	[1,30]	31.90%	68.10%	a
	(30,60]	32%	68%	a
	(60,90]	32.60%	67.40%	a
	(90,179]	33.40%	66.60%	b
	non gestante	36.60%	63.40%	b
Production par lactation	(0,800]	11.90%	88.10%	a
	(800,1000]	34.40%	65.60%	b
	(1000,1200]	38.10%	61.90%	c
	(1200,2600]	56.90%	43.10%	d

[20-30], [30-70], [70-85], [85-110], [110-140], [140-155], [155-180], [180-195], [195-215], [215-230], [230-245], [245-255], [255-275], [275-290],[290-300]

3++ (lactations 3 et plus de 2013 et 2014): classes de 5 jours jusqu'au 20^{ième} jour de lactation, puis [20-65], [65-85], [85-100], [100-120], [120-135], [135-150], [150-165], [165-185], [185-210], [210-240], [240-255], [255-285], [285-300]

Dans une étude préliminaire, l'ensemble des modèles décrits ci-dessus ont été testés sur le premier jeu de données à l'aide du logiciel R en différenciant les courbes selon le rang de lactation des chèvres (modèle 0 : $y_i = f(lact_i) + e_i$). Pour les polynômes de Legendre, les modèles « leg4 » et « leg5 » ont été testés. Pour les polynômes fractionnaires, les degrés 2 et 3 ont été testés avec toutes les combinaisons possibles des puissances $\{-3, -2, -1, -0.5, -\frac{1}{3}, \frac{1}{3}, 0.5, 1, 2, 3\}$. Il en a été de même pour les différentes positions des nœuds des splines. A l'issue de ce test, le meilleur modèle de chaque catégorie a été conservé pour la suite de l'étude.

4. b. Tests des modèles

A la suite de cette première analyse, les deux races ont été séparées et seules les Alpines ont été étudiées. Des modèles associant les différents facteurs de variation (effets environnementaux) ont été testés grâce au logiciel Genokit développé par V. Ducrocq (INRA-UMR GABI, Jouy-en-Josas).

Le premier modèle testé est : (1) $y_{ik} = HTD_i + e_{ik}$

avec y_{ik} est la production laitière de l'individu k le jour de contrôle i, HTD_i le $i^{ième}$ effet HTD (Herd Test Day : date du contrôle*nombre de cheptel), et e_{ik} est l'erreur résiduelle du modèle. Le modèle (1) a été utilisé avec le premier et le deuxième jeu de données. Par la suite, seul le deuxième jeu de données a été utilisé.

Le deuxième modèle consiste à ajouter au modèle précédent une fonction f (paramétrique, semi-paramétrique ou non paramétrique), dépendant du rang de lactation appelé « lact »:

(2) $y_{ijk} = HTD_i + f(lact_j) + e_{ijk}$.

avec y_{ijk} est la production laitière de l'individu k le jour de contrôle i pour le $j^{ième}$ rang de lactation, f un des modèles étudiés (modèle paramétrique, semi-paramétrique ou non paramétrique) et $lact_j$ le $j^{ième}$ rang de lactation.

Le troisième modèle consiste à ajouter les différents facteurs environnementaux âge à la mise-bas, mois de mise-bas, durée de tarissement et stade de gestation selon la région et le rang de lactation sur le niveau moyen des courbes de lactation. Le modèle (3) complet s'écrit:

(3) $y_{ijklmn} = HTD_i + f(lact_j) + (région * lact * age)_k + (région * lact * mois)_l + (région * lact * dtar)_m + (région * lact * gest)_n + e_{ijklmn}$

avec $(région * lact * age)_k$, $(région * lact * mois)_l$, $(région * lact * dtar)_m$, $(région * lact * gest)_n$ correspondant à la $k^{ième}$, $l^{ième}$, $m^{ième}$, $n^{ième}$ interactions « région*rang de lactation*âge à la mise bas », « région*rang de lactation*mois de mise bas », « région*rang de lactation*durée de tarissement », « région*rang de lactation*stade de gestation » respectivement. Ces interactions sont constantes au cours de la lactation.

Le modèle 3a correspond au modèle 3 soustrait du terme d'interaction $(région * lact * age)_k$. Il en est de même pour les modèles 3b, 3c, 3d qui eux sont soustraits des interactions $(région * lact * mois)_l$, $(région * lact * dtar)_m$, $(région * lact * gest)_n$ respectivement.

Le quatrième modèle consiste à moduler la fonction f d'Ali et Schaeffer et des splines en fonction des facteurs environnementaux sans la région en interaction:

Tableau 5 : Nombre de paramètres, logarithme de la fonction de vraisemblance ($\log(L)$), critère d'information d'Akaike (AIC), critère d'information bayésien (BIC), écart type de l'erreur (σ) pour les différents modèles retenus. Ils sont de la forme $y_i = f(lact_i) + e_i$. Les données utilisées sont issues du jeu de données 1.

Modèles	Nombre de paramètres	Log(L)	AIC	BIC	Sigma
Wilmink	16	-5 055 310	10 110 653	10 110 859	9.7801
Ali et Schaeffer	20	-5 054 591	10 109 224	10 109 478	9.7750
Legendre (leg5)	24	-5 054 774	10 109 598	10 109 901	9.7763
Polynômes fractionnaires de degré 3 (f3)	16	-5 054 698	10 109 429	10 109 636	9.7757
fix1	236	-5 053 732	10 107 937	10 110 812	9.7696
fix5	67	-5 054 634	10 109 404	10 110 228	9.7755
Spline	32	-5 054 284	10 108 633	10 109 033	9.7728

$$(4) \quad y_{ijklmn} = HTD_i + \text{région} * \text{lact} * \text{age}_k + \text{région} * \text{lact} * \text{mois}_l + \text{région} * \text{lact} * \text{dtar}_m + \text{région} * \text{lact} * \text{gest}_n + f(\text{lact} * \text{age}_k) + f(\text{lact} * \text{mois}_l) + f(\text{lact} * \text{dtar}_m) + f(\text{lact} * \text{gest}_n) + e_{ijklmn}$$

Le modèle 4a correspond au modèle 4 soustrait du terme $f(\text{lact} * \text{age}_k)$. Il en est de même pour les modèles 4b, 4c, 4d qui eux sont soustraits des termes : $f(\text{lact} * \text{mois}_l)$, $f(\text{lact} * \text{dtar}_m)$, $f(\text{lact} * \text{gest}_n)$ respectivement.

Le cinquième modèle consiste à moduler la fonction f en fonction des facteurs environnementaux avec la région en interaction:

$$(5) \quad y_{ijklmn} = HTD_i + \text{région} * \text{lact} * \text{age}_k + \text{région} * \text{lact} * \text{mois}_l + \text{région} * \text{lact} * \text{dtar}_m + \text{région} * \text{lact} * \text{gest}_n + f(\text{région} * \text{lact} * \text{age}_k) + f(\text{région} * \text{lact} * \text{mois}_l) + f(\text{région} * \text{lact} * \text{dtar}_m) + f(\text{région} * \text{lact} * \text{gest}_n) + e_{ijklmn}$$

4. c. Outils de comparaison des modèles

Plusieurs critères peuvent être utilisés afin de comparer les différents modèles que nous avons mis en œuvre.

- Le critère d'information d'Akaike (**AIC**) (Akaike, 1974) $AIC = 2k - 2 \times \ln(L)$
où k est le nombre de paramètres à estimer et L est le maximum de la fonction de vraisemblance du modèle.
- le critère d'information bayésien (**BIC**) (Schwarz, 1978) $BIC = \ln(n) \times k - 2 \times \ln(L)$
avec n le nombre d'observations dans l'échantillon étudié et k le nombre de paramètres. Le meilleur modèle est celui dont la valeur du critère est la plus faible.
- L'écart type de l'erreur, appelé « **sigma** » a lui aussi été comparé pour chaque modèle.

$$\text{Sigma} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-1}} \quad \text{où } \hat{y}_i \text{ est l'estimation donnée par le modèle et } n \text{ est le nombre de données.}$$

- Le coefficient de détermination notée **R²**, mesure la qualité d'ajustement. Il est défini comme suit: $R^2 = \frac{\text{Var}(\hat{y})}{\text{Var}(y)} = 1 - \frac{\text{Var}(\hat{e})}{\text{Var}(y)}$

Ce coefficient est compris entre 0 et 1. Le meilleur modèle sera celui ayant le R^2 le plus élevé.

Le premier terme du critère d'AIC et de BIC représente une fonction de pénalité mesurant la complexité du modèle à travers son nombre de paramètres. Le second terme peut être interprété comme une mesure de l'inadéquation du modèle aux données. Le meilleur modèle est celui dont la valeur du critère est la plus faible. L'AIC et le BIC sont basés sur une estimation de la distance entre chacun des modèles candidats et le vrai modèle (inconnu). Ils ne disent rien de la qualité d'ajustement absolue du modèle : ils ne rendraient pas compte d'un mauvais ajustement de l'ensemble des modèles.

L'évolution de la moyenne des résidus de chaque modèle en fonction du stade de lactation a également été étudiée.

Résultats

1. Analyse descriptive des courbes de lactation en fonction des effets de milieu (premier jeu de données)

La figure 8 présente les courbes moyennes de production laitière des chèvres en fonction de leur rang de lactation. Les premières lactations ont le pic de lactation le moins élevé, (inférieur de 10 hg environ par rapport aux autres) et il intervient à 50 jours de stade de lactation soit 10 jours environ après celui des multipares. Les troisièmes lactations et plus ont les pics les plus élevés, celui des « L3+ » étant plus élevé que celui des « L3++ » (+2 hg). La persistance laitière, définie comme élevée quand

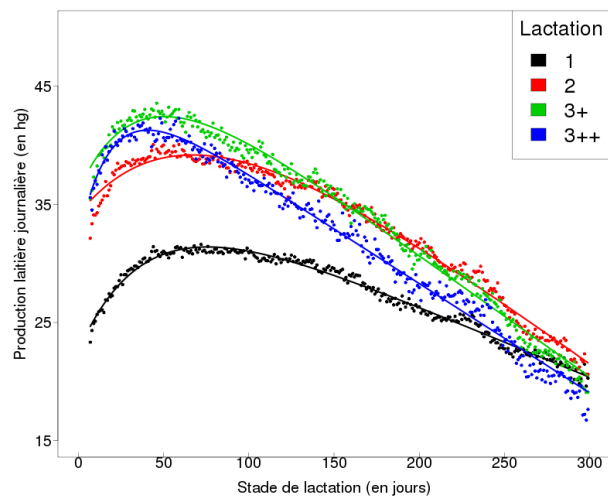


Figure 17 : Représentation des courbes selon le modèle de Wilink pour les différents rangs de lactation. Les points représentent la moyenne des contrôles par stade et par rang de lactation (modèle 0)

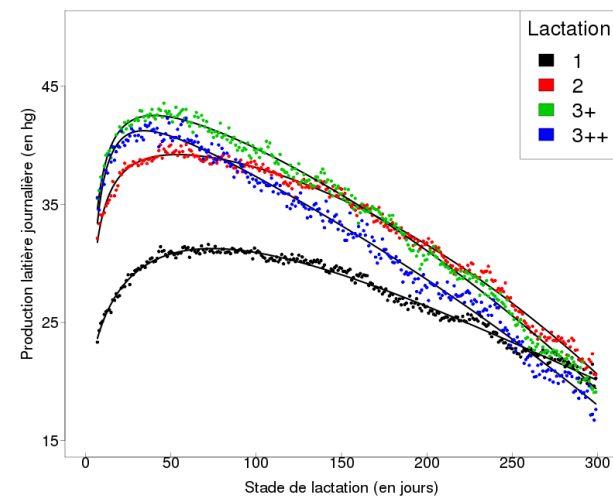


Figure 18 : Représentation des courbes selon le modèle d'Ali et Schaeffer pour les différents rangs de lactation. Les points représentent la moyenne des contrôles par stade et par rang de lactation (modèle 0)

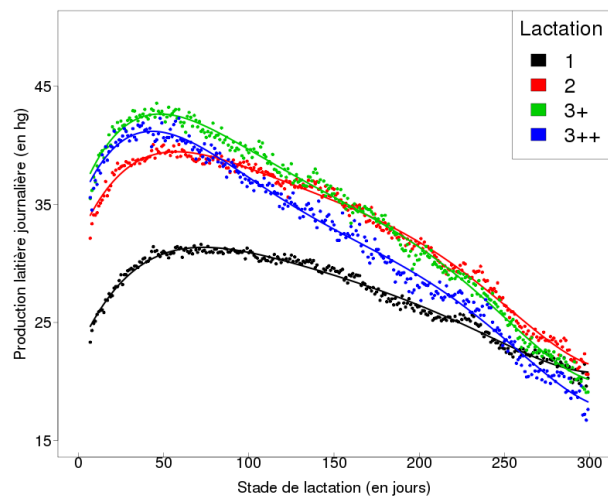


Figure 19 : Représentation des courbes selon le modèle des polynômes de Legendre « leg5 » pour les différents rangs de lactation. Les points représentent la moyenne des contrôles par stade et par rang de lactation (modèle 0)

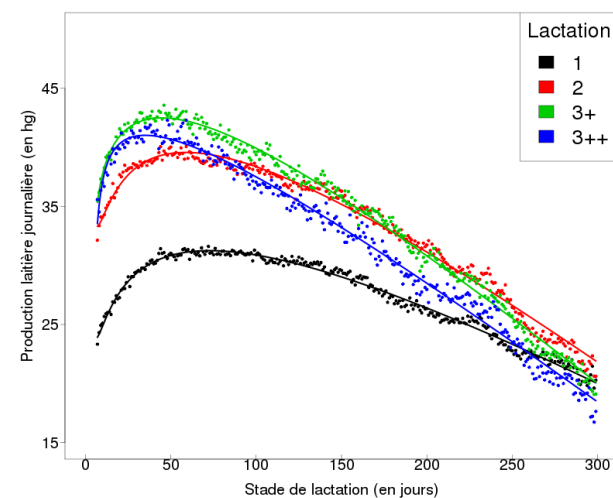


Figure 20 : Représentation des courbes selon le modèle des polynômes fractionnaires pour les différents rangs de lactation. Les points représentent la moyenne des contrôles par stade et par rang de lactation (modèle 0)

l'écart de production au pic et à 300 jours est faible, est maximale pour les premières lactations (avec un écart de production de 10 hg), intermédiaire pour les deuxièmes lactations (écart de 18 hg) et minimale pour les troisièmes lactations et plus (écart de 22 hg).

Quel que soit la région et le rang de lactation, les Alpines ont un pic de lactation plus précoce et une persistance laitière moins importante que les Saanens. En deuxième lactation et plus, les Alpines produisent plus au pic de lactation quel que soit la région comme l'illustre la figure 9 pour les deuxièmes lactations.

Tout au long de la lactation, le niveau de production est différent d'une région à l'autre. La production journalière moyenne est plus élevée dans le Nord-Ouest, elle est intermédiaire dans le Nord-Est, et est plus basse dans le Sud-Est et le Sud-Ouest (cf. Figure 10). Les écarts entre régions semblent plus marqués en race Alpine qu'en race Saanen quel que soit le rang de lactation. Cependant, ce constat est à nuancer puisque dans notre jeu de données 80% des chèvres étaient dans le NO et seulement 2% et 3% dans le SO et le SE respectivement. Les faibles effectifs sont plus variables et expliquent les variations de productions « parasites ».

Les chèvres qui mettent bas précocement (âge à la mise bas < 12 mois) ont un pic de lactation moins haut mais une plus grande persistance comme l'illustre la figure 11. Ces différences semblent plus marquées en race Alpine et il est à noter que les chèvres Saanen sont en moyenne plus précoces.

Quel que soit le mois de mise-bas, les niveaux de production au pic de lactation sont similaires mais la persistance est différente (cf figure 12 pour les Alpine en L2). En effet les chèvres chevrotant en janvier ou entre septembre et décembre ont des lactations plus persistantes que celles mettant bas en février-mars. Ceci est vérifié quel que soit la race ou le rang de lactation.

Les chèvres restées taries plus longtemps ont, à la lactation suivante, une courbe de lactation moins persistante (cf. figure 13 pour les deuxièmes lactations). La durée de tarissement est en lien avec le niveau de production à la fin de la lactation. En effet, il est probable que les éleveurs tarissent les chèvres à partir d'un certain niveau de production. Une chèvre plus persistante produit plus à la fin de sa lactation, elle est tarie plus tard et donc moins longtemps qu'une chèvre moins persistante. Cette hypothèse laisse penser qu'une chèvre persistante l'est d'une lactation à l'autre.

Plus la gestation est avancée au moment du tarissement ou à 300 jours, plus la chute de la production laitière en fin de lactation est importante (cf. Figure 14 pour les deuxièmes lactations), quel que soit le rang de lactation ou la race. Comme nous venons de le voir, les chèvres les plus persistantes sont taries plus tard au cours de leur lactation et ont par conséquent un stade de gestation au tarissement plus avancé.

Le tarissement intervient lorsque l'effet de la gestation abaisse la production laitière à un certain niveau qui entraîne le tarissement de l'animal.

Une des principales difficultés rencontrées pour évaluer les effets des modalités des différents facteurs est le lien entre ces différentes modalités.

2. Classification des courbes de production laitière

2. a. Choix du nombre de clusters :

Les résultats à partir de ces différents échantillons de 30 000 lactations étant très proches, nous ne montrons à la figure 15 que les résultats pour l'un d'entre eux. Le choix du nombre de clusters à retenir a été défini à l'aide de différents critères présentés dans le package. Le critère de Ray et Turi est très proche de 1 pour un nombre de clusters compris entre 2 et 11, puis il oscille et se rapproche de 1 avec

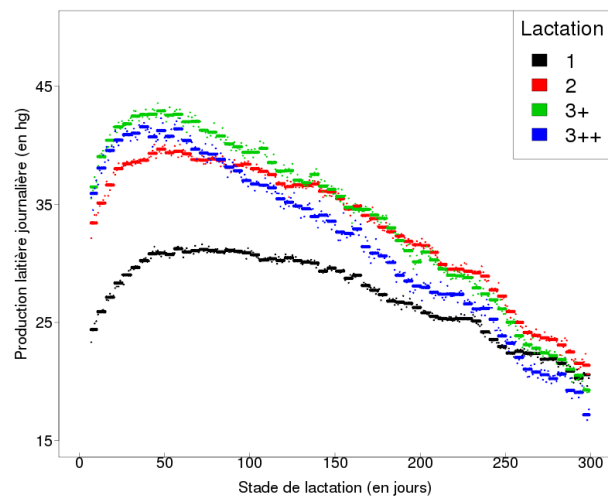


Figure 21 : Représentation des courbes selon le modèle à 236 classes fixes (« fix1 ») selon le rang de lactation. Les points représentent la moyenne des contrôles par stade et par rang de lactation (modèle 0)

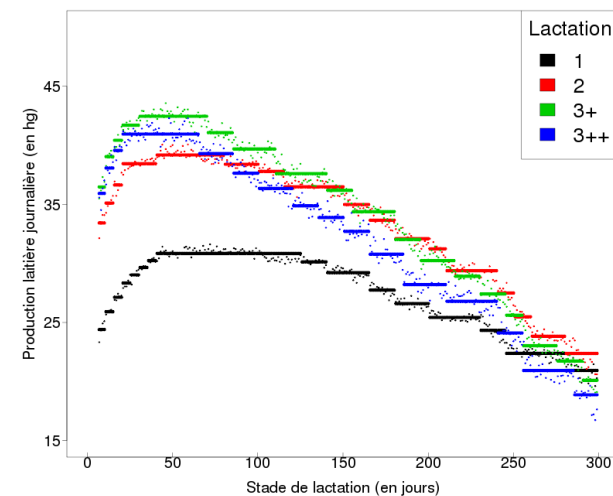


Figure 22 : Représentation des courbes selon le modèle à 67 classes fixes (« fix5 ») selon le rang de lactation. Les points représentent la moyenne des contrôles par stade et par rang de lactation (modèle 0)

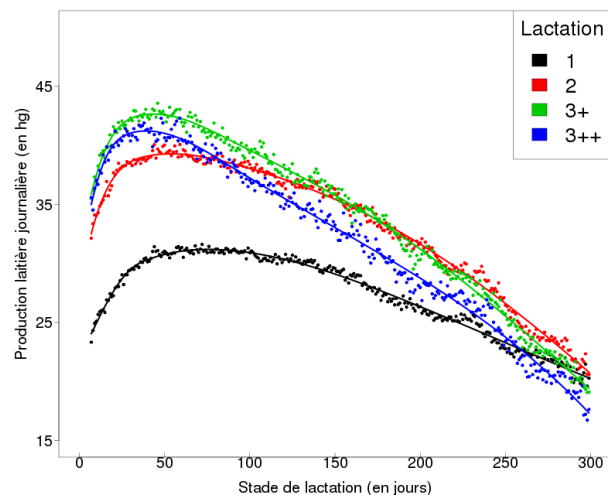


Figure 23 : Représentation des courbes selon le modèle splines cubiques pour les différents rangs de lactation. Les points représentent la moyenne des contrôles par stade et par rang de lactation (modèle 0)

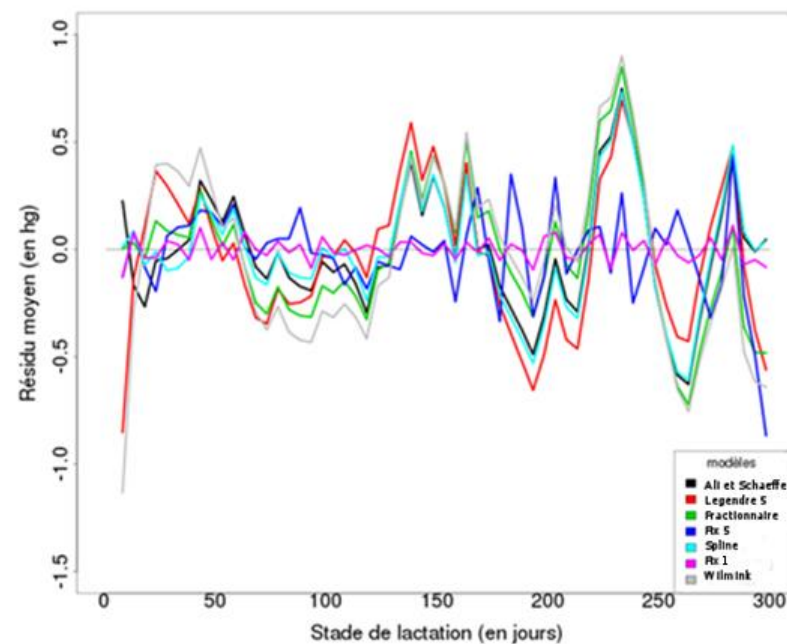


Figure 24 : Représentation des résidus moyens des différents modèles (modèle 0) en fonction du stade de lactation

13, 14, 16, ou 19 clusters. Les autres critères (Calinski.Harabatz3, BIC, BIC2 et AIC) ont des valeurs faibles pour 2 clusters puis augmentent de manière importante jusqu'à un nombre de clusters maximal avant de rediminuer. Pour 7 clusters, le BIC2 est à son maximum, les critères BIC et Calinsky.Harabatz3 sont à 95% de leur valeur maximale. Le BIC et l'AIC ont une forte progression entre 2 et 7 clusters puis cette progression ralentit. Sur la base de l'évolution de ces critères, 7 clusters ont donc été choisis ce qui permet également d'avoir un nombre réduit de classes pour dégager les grandes caractéristiques des profils de courbes de lactation de chèvres.

2. b. A partir de cette classification en 7 clusters :

Les calculs ont ensuite été réalisés afin de classer toutes les lactations du premier jeu de données en sept clusters. Les clusters comprennent de 6.1 à 20.8% des lactations. La valeur brute (et non modifiée pour la figure 15) du coefficient de Ray et Turi est de 0.00213. Cette valeur proche de 0 signifie donc que la variance inter cluster est 470 fois plus grande que la variance intra-cluster.

La figure 16 présente la courbe de lactation moyenne de chaque cluster. Des profils de courbes différents peuvent être remarqués: (1) les clusters C, E, G présentent des profils avec un pic de lactation prononcé (groupe « pic ») avec une différence de production entre le pic et le 300^{ième} jour de lactation supérieure à 25 hg, ils représentent 34% des lactations; (2) les clusters A, B, D, F, présentent des profils relativement plats (groupe « plate »), avec une différence entre le pic de lactation et le 300^{ième} jour de lactation inférieure à 15hg.

Afin de caractériser chaque cluster, nous avons eu recours au modèle linéaire généralisé (l.a) : chaque individu est attribué à un cluster selon son profil (il s'agit de notre variable réponse) et nous avons testé la significativité des effets d'environnement dont nous disposons sur cette variable réponse (Tableau 3). Le modèle complet est un modèle additif composé des effets suivants : région, âge à la mise bas, rang de lactation, mois de mise bas, stade de gestation, durée de tarissement, race et niveau de production totale à la lactation. Tous les effets se sont avérés significatifs. Ils peuvent donc expliquer le partitionnement des lactations dans les différents clusters mais certains comme la région ou la durée de tarissement n'apportent pas beaucoup au modèle. En revanche, le niveau de production a grandement contribué à la répartition des clusters, viennent ensuite le mois de mise bas, la race, le rang de lactation et l'âge à la mise bas.

Les 2 groupes (pic vs plate) de courbes de lactation ont été caractérisés en fonction de leur composition pour les différents critères d'environnement (Tableau 4). Ainsi on observe : (1) les lactations de chèvres Alpine ont plus souvent une courbe avec un pic que les lactations de chèvres Saanen (38.4 vs 27.6%) ; (2) les chèvres en première lactation ont des courbes plates dans leur très grande majorité (86.1%), c'est un peu moins vrai en deuxième lactation (60.3%), à la différence des animaux en troisième lactation et plus qui ont majoritairement un pic (55%) ; (3) le Nord-ouest a légèrement plus de chèvres en pic que le Sud et le Nord-Est (+2%) ; (4) les chèvres ayant mis bas précocement ont des courbes plus plates que les chèvres plus tardives pour les 3 premiers rangs de lactation ; (5) les chèvres mettant bas entre septembre et janvier ont des courbes de lactation plus plates que celles mettant bas entre février et août ; (6) les chèvres tarées moins longtemps ont des courbes plus plates (+8%) ce qui est peut-être dû au fait que des chèvres ayant un tarissement plus long ont plus de réserves corporelles qu'elles peuvent utiliser plus facilement durant la lactation suivante ; (7) les chèvres les plus avancées dans leur gestation à 300 jours ou non gestantes ont légèrement plus de courbes avec un pic que les autres ; (8) les chèvres avec un faible niveau total de production ont des courbes plus plates. On observe également que les chèvres produisant plus de 1200 kg ont plus de courbes avec un pic mais que néanmoins dans 43.1% le même niveau de production est

Tableau 6 : Nombre de paramètres, logarithme de la fonction de vraisemblance (Log(L)), critère AIC, critère BIC, écart type de la variance résiduelle (sigma), et qualité d'ajustement du modèle (R^2) pour différents modèles utilisant différentes fonctions. Les données sont issues du jeu de données 2.

	Modèle	Fonctions	Nombre de paramètres	Log(L)	AIC	BIC	Sigma	R²
1	$y_i = HTD_i + e_i$		13626	-4 754 247	9 535 747	9 700 539	8.5525	0.6732
2	$y_{ij} = HTD_i + f(lact_j) + e_{ij}$	Ali et Schaeffer	13641	- 4 646 904	9 321 092	9 486 065	7.8910	0.7311
		Splines	13644	-4 646 857	9 321 004	9 486 014	7.8907	0.7311
		Classes fixes	13677	-4 650 020	9 327 397	9 492 805	7.9095	0.7296
3	$y_{ijklmn} = HTD_i + f(lact_j) + (région * lact * age)_k + (région * lact * mois)_l + (région * lact * dtar)_m + (région * lact * gest)_n + e_{ijklmn}$	Ali et Schaeffer	13794	-4 617 493	9 262 576	9 429 398	7.7189	0.7447
		Splines	13800	-4 617 395	9 262 393	9 429 287	7.7183	0.7447
		Classes fixes	13833	-4 620 671	9 269 011	9 436 304	7.7373	0.7432
4	$y_{ijklmn} = HTD_i + (région * lact * age)_k + (région * lact * mois)_l + (région * lact * dtar)_m + (région * lact * gest)_n + f(lact * age_k) + f(lact * mois_l) + f(lact * dtar_m) + f(lact * gest_n) + e_{ijklmn}$	Ali et Schaeffer	13990	-4 613 784	9 255 551	9 424 741	7.6974	0.7463
		Splines	14094	-4 613 489	9 255 168	9 425 615	7.6957	0.7465
5	$y_{ijklmn} = HTD_i + (région * lact * age)_k + (région * lact * mois)_l + (région * lact * dtar)_m + (région * lact * gest)_n + f(région * lact * age_k) + f(région * lact * mois_l) + f(région * lact * dtar_m) + f(région * lact * gest_n) + e_{ijklmn}$	Ali et Schaeffer	14406	-4 613 315	9 255 444	9 429 660	7.6947	0.7465
		Splines	14718	-4 613 024	9 255 487	9 433 472	7.6930	0.7467
Classes fixes	16284	-4 614 852	9 262 274	9 459 177	7.7036	0.7459		

atteint avec un profil plat.

Lorsqu'on étudie les clusters individuellement, on peut remarquer que : (1) le cluster G est composé de chèvres de tous les rangs de lactation (5% de premières lactations, 36% de deuxième lactation) provenant en majorité du NO, ayant mis bas tardivement, il regroupe 33.5% des chèvres qui produisent plus de 1200kg de lait ; (2) le cluster F regroupe principalement des chèvres Saanen (14% contre 8% d'Alpine), les chèvres sont majoritairement en deuxième lactation (40%), précoces, et représentent 39% des chèvres produisant plus de 1200kg ; (3) dans le cluster D on a 20.7% des premières lactations qui représentent 61% des chèvres du cluster, avec 30% des lactations du sud, et représentent 56.5% des chèvres produisant moins de 800L.

Certains clusters ont des niveaux de production particulièrement élevés par rapport aux courbes moyennes de la partie 1 des Résultats. En effet, dans cette dernière partie, le pic des chèvres en 3^{ème} lactation et plus atteint 44 hg en moyenne alors que pour les clusters F, E, et G ils atteignent 45 hg ou plus, jusqu'à 60 hg pour le cluster G.

2. c. Evolution de la forme de la courbe de lactation au cours de la carrière des chèvres :

Nous pouvons noter que :

(1) pour les chèvres ayant réalisé les deux premières lactations (i) 54% de chèvres sont dans un cluster « groupe plate » en première lactation et y restent en 2^{ème} lactation (cela correspond au passage majoritaire du cluster B vers le cluster A (11%)), (ii) 31 % passant d'un cluster « groupe plate » à un cluster « groupe pic » avec 31% des Alpines et 25% des Saanens (passage du cluster B vers le cluster C (6%)),

(2) entre la 2^{ème} et la 3^{ème} lactation, (i) 33% des animaux sont passées d'un cluster « groupe plate » à un cluster « groupe plate » (41% des Saanens contre 28% des Alpines), (ii) 26% d'un cluster « groupe plate » à un cluster « groupe pic », (iii) 28% d'un cluster « groupe pic » à un cluster « groupe pic » (31% des Alpines contre 21% des Saanens).

En ce qui concerne les carrières des chèvres, ayant accompli 5 lactations (n=6801) : (i) 12% ont eu leurs 5 courbes de lactation dans un cluster du « groupe plate », (ii) 14% n'ont eu que la première lactation dans un cluster « groupe plate », (iii) 10% les deux premières lactations dans un cluster « groupe plate », (iv) 6% les 3 premières lactations dans un cluster « groupe plate », (v) 6% les 4 premières lactations dans un cluster « groupe plate », (vi) 5% n'ont eu que la quatrième lactation dans un cluster « groupe pic » et (vii) 5 % n'ont eu que des lactations dans un cluster « groupe pic ». Lorsque nous avons appliqué un modèle linéaire généralisé sur cet ensemble de données (modèle I.b), la production moyenne cumulée des 5 lactations, d'une chèvre Alpine du Nord Est, ayant eu que des profils de production dans un cluster « groupe pic » dans sa carrière, est égale à 5490kg de lait. Les chèvres de race Alpine du NE ayant eu que des courbes dans un cluster « groupe plate » ont produit 1514 kg de moins en moyenne, celles ayant eu leur courbe L1 dans un cluster « groupe plate » ont produit 155kg de moins, celles ayant eu les deux premières lactations dans un cluster « groupe plate » 495kg de moins et, enfin, celles ayant eu les 3 premières lactations dans un cluster « groupe plate » 940kg de moins. Les chèvres de race Saanen du NE ont produit 238 kg de plus. Les chèvres de race Alpine du NO 500 kg de plus, et celles du Sud 122kg de moins.

Logiquement, nous retrouvons les mêmes effets que précédemment pour les facteurs région et race sur les formes des courbes. Les profits « pics » sont associés à des lactations plus productives bien que les Saanens qui ont des profils plus « plats » produisent légèrement plus que les chèvres de race Alpine après 5 lactations.

Tableau 7 : Nombre de paramètres, logarithme de la fonction de vraisemblance (Log(L)), critère AIC, critère BIC, écart type de la variance résiduelle (sigma), et qualité d'ajustement du modèle(R²) pour différents modèles. Les données sont issues du jeu de données 2 pour la race Alpine.

Modèle	Fonctions	Nombre de paramètres	Log(L)	AIC	BIC	Sigma	R ²
(3) : $y_{ijklmn} = HTD_i + f(lact_j)$ + (région * lact * age) _k + (région * lact * mois) _l + (région * lact * dtar) _m + (région * lact * gest) _n + e_{ijklmn}	(3) base Ali et Schaeffer	13 794	- 4 617 493	9 262 576	9 429 398	7.719	0.745
	Splines	13 800	-4 617 395	9 262 393	9 429 287	7.718	0.745
	Classes fixes	13 833	-4 620 671	9 269 011	9 436 304	7.737	0.743
	(3a) –(région * lact * age) _k Ali et Schaeffer	13 749	-4 619 534	9 266 568	9 432 846	7.731	0.744
	Splines	13 755	-4 619 432	9 266 376	9 432 727	7.730	0.744
	Classes fixes	13 788	-4 622 795	9 273 168	9 439 918	7.750	0.742
	(3b) –(région * lact * mois) _l Ali et Schaeffer	13 749	-4 622 723	9 272 947	9 439 225	7.749	0.742
	Splines	13 755	-4 622 669	9 272 850	9 439 201	7.749	0.742
	Classes fixes	13 788	-4 625 950	9 279 478	9 446 228	7.768	0.741
	(3c) –(région * lact * dtar) _m Ali et Schaeffer	13 773	-4 618 430	9 264 408	9 430 976	7.724	0.744
	Splines	13 779	-4 618 342	9 264 245	9 430 886	7.724	0.744
	Classes fixes	13 812	-4 621 529	9 270 684	9 437 723	7.742	0.743
	(3d) –(région * lact * gest) _n Ali et Schaeffer	13 749	-4 637 487	9 302 474	9 468 753	7.836	0.736
	Splines	13 755	-4 637 381	9 302 275	9 468 626	7.835	0.736
	Classes fixes	13 788	-4 640 578	9 308 734	9 475 483	7.854	0.734
(4) : $y_{ijklmn} = HTD_i$ + (région * lact * age) _k + (région * lact * mois) _l + (région * lact * dtar) _m + (région * lact * gest) _n + $f(lact * age_k)$ + $f(lact * mois_l)$ + $f(lact * dtar_m)$ + $f(lact * gest_n) + e_{ijklmn}$	(4) base Ali et Schaeffer	13 990	-4 613 784	9 255 551	9 424 741	7.697	0.746
	Splines	14 094	-4 613 489	9 255 168	9 425 615	7.696	0.747
	(4a)– $f(lact * age_k)$ Ali et Schaeffer	13 930	-4 614 678	9 257 219	9 425 684	7.703	0.746
	Splines	14 004	-4 614 396	9 256 803	9 426 162	7.701	0.746
	(4b)– $f(lact * mois_l)$ Ali et Schaeffer	13 930	-4 614 771	9 257 405	9 425 870	7.703	0.746
	Splines	14 004	-4 614 468	9 256 946	9 426 306	7.701	0.746
	(4c)– $f(lact * dtar_m)$ Ali et Schaeffer	13 962	-4 614 050	9 256 027	9 424 879	7.699	0.746
	Splines	14 052	-4 613 765	9 255 636	9 425 575	7.697	0.746
	(4d)– $f(lact * gest_n)$ Ali et Schaeffer	13 930	-4 615 295	9 258 453	9 426 918	7.706	0.746
	Splines	14 004	-4 615 156	9 258 322	9 427 681	7.705	0.746

3. Modélisation des courbes de production laitière

3. a. Sélection des types de modèles:

Des études préliminaires, basées sur l'ajustement des données (issues du jeu de données 1) par le modèle 0 de la forme $y=f(\text{rang de lact})$ avec f fonctions définies dans la partie 4a du Matériel & Méthodes, ont permis : (1) de montrer qu'un polynôme de Legendre de degré 5 était statistiquement meilleur qu'un polynôme de Legendre de degré 4 (meilleur BIC) ; (2) de retenir deux modèles utilisant des classes fixes : un modèle à 236 classes (noté « fix1 ») car il est le meilleur en termes d'ajustement (selon le test du rapport de vraisemblance) et un modèle à 67 classes (noté « fix5 ») car c'est le modèle ayant le moins de paramètres, tout en ayant un bon ajustement aux données et avec le meilleur BIC des modèles à classes fixes ; (3) de conserver le meilleur polynôme fractionnaire, il est de degré 3 et s'écrit sous la forme suivante $f_3(t) = a + bt^{b_1} + ct^{b_1}\ln(t) + dt^{b_2}$ avec $b_1 = \frac{1}{3}$ et $b_2=3$; (4) de sélectionner les nœuds aux positions (7, 80, 180, 200, 220, 299) pour les modèles splines (cf. Annexe 3). Les résultats sur l'ensemble des modèles testés sont présentés dans le tableau en annexe 3.

Les modèles conservés sont comparés dans le tableau 5. Les valeurs des différents critères varient peu d'un modèle à l'autre, bien que le nombre de paramètres puisse être très différent d'un modèle à l'autre. Pour l'écart-type de l'erreur, l'écart maximal est de 0.01 entre le meilleur modèle et le moins bon. Le modèle « fix1 » basé sur 236 classes de stades de lactation présente l'AIC et l'écart type résiduel le plus faible. Mais son nombre de paramètres élevés à estimer le pénalise d'où son classement en avant dernier sur le critère BIC. Le modèle des splines cubiques se trouve en première position sur le critère BIC et en deuxième position sur les critères AIC et écart type de l'erreur. Le modèle d'Ali et Schaeffer est 3^{ième} sur AIC et sur l'écart type de l'erreur, et deuxième sur le BIC. Le modèle des polynômes de Legendre et le modèle « fix5 » avec 67 classes de stade de lactation sont proches du modèle d'Ali et Schaeffer.

Les figures 17 à 23, illustrent l'ajustement des modèles aux données moyennes selon le rang de lactation et pour chacune des fonctions utilisées. Les courbes sont très similaires quel que soit selon la fonction. Seuls les ajustements aux extrémités des courbes de productions diffèrent selon la fonction choisie, le modèle d'Ali et Schaeffer présente le meilleur ajustement (Figure 18).

La figure 24 présente les résidus moyens en fonction du stade de lactation selon les différentes modélisations. Les 2 modèles à classes fixes (fix1 et fix5) ont les résidus qui oscillent le plus près de la droite $y=0$. Les modèles de Wilmink et des polynômes de Legendre semblent plus éloignés et avoir les plus grands résidus. Pour la suite de l'étude, nous retiendrons les 3 modèles suivants : (1) le modèle d'Ali et Schaeffer comme étant le meilleur modèle paramétrique, (2) le modèle à 67 classes « fix5 », (3) les splines cubiques comme le meilleur modèle semi paramétrique.

3. b. Modélisation des effets d'environnement :

Nous avons souhaité ensuite inclure dans le modèle l'ensemble des facteurs environnementaux dont nous disposons afin de les tester conjointement. Nous avons alors utilisé le jeu de données 2 car le jeu de données 1 surevaluait l'effet troupeau x jour de contrôle (HTD) et nous avons des niveaux pour ce facteur avec peu d'effectif. En effet, dans le jeu de données 1, en 2009, les troupeaux étaient composés que de chèvres en première lactation, de 2012 à 2014 seulement de chèvres de plus de 3 lactations. Il y a donc une confusion entre l'effet HTD et le rang de lactation pour ces années-là. Le jeu de données 2 a ensuite été divisé selon le facteur race pour ne considérer que les Alpines afin de simplifier l'analyse. Pour le modèle 1, le R^2 pour le premier jeu de données était de 0.73 (avec uniquement les Alpines) contre 0.67 pour le deuxième jeu.

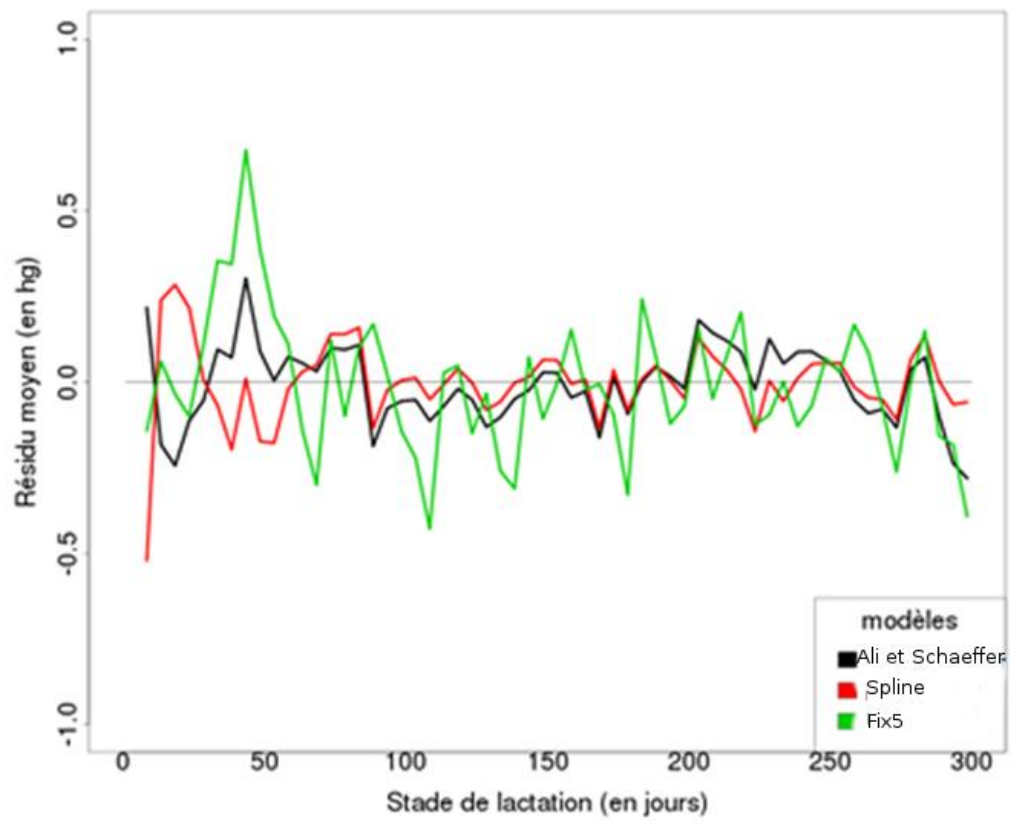


Figure 25 : Résidus moyens lissés par classes de 5 jours pour le modèle 4 selon la modélisation utilisée. Le deuxième jeu de données a été utilisé.

Pour ce jeu de données et le modèle splines cubiques, plusieurs positions de nœuds ont été testées (cf. Annexe 4). Le meilleur modèle selon les critères testés (AIC, BIC, Sigma et R^2) est un modèle dont les 6 nœuds sont positionnés aux stades de lactation 7, 30, 60, 100, 230, 299.

Le tableau 6 regroupe les niveaux des différents critères d'ajustement aux données en fonction des modèles. Globalement, la modélisation des formes de courbes de lactation selon le modèle d'Ali et Schaeffer est très proche de celle par les splines quel que soit le modèle, avec un R^2 identique à 10^{-3} . Le modèle d'Ali et Schaeffer est légèrement meilleur que le modèle splines pour le modèle 4 selon le BIC, car il a 5 paramètres à estimer au lieu de 6 pour les splines. Quel que soit le modèle considéré la modélisation de la courbe de lactation par le modèle d'Ali et Schaeffer ou les splines est meilleure que celle avec les classes fixes

Le modèle 5 avec la région en interaction avec le rang de lactation et le facteur d'environnement, n'est pas le modèle le plus intéressant par son BIC, légèrement par son AIC et par sa fonction de vraisemblance.

Le tableau 7 présente les niveaux des différents critères d'ajustement aux données des modèles 3 et 4 ainsi que les modèles qui en sont dérivés (3a, 3b, 3c, 3d, 4a, 4b, 4c, 4d). Dans le modèle 3, chaque facteur influe sur le niveau de production car lorsqu'on enlève une interaction du modèle, le critère BIC augmente quelle que soit la fonction utilisée. L'importance des différents facteurs n'est pas la même : le facteur le plus important est l'effet de la gestation puis vient l'effet du mois de mise bas, ensuite l'âge à la mise bas et enfin la durée de tarissement.

Dans le modèle 4, lorsqu'on enlève une fonction qui dépend d'un facteur, le critère BIC augmente quel que soit le facteur à l'exception de la durée de tarissement pour le modèle splines. L'ordre d'importance des facteurs est le même que pour le modèle 4. Il peut donc être conclu que chaque facteur influence la forme de la courbe de lactation.

Suite aux analyses réalisées, le meilleur modèle selon le critère BIC pour modéliser la courbe de production laitière est le modèle 4 utilisant Ali et Schaeffer (Tableau 6).

La figure 25 représente les résidus moyens du modèle 4 en fonction du stade de lactation pour chacune des fonctions utilisées. L'ensemble des modèles modélisent bien les données car les résidus moyens oscillent autour de l'axe des abscisses. Le résidu maximal égal à 0.7 hg est obtenu avec les classes fixes. Il peut être remarqué que le modèle avec les classes fixes a le meilleur ajustement avant le pic de lactation. Cela résulte de la taille réduite (5 jours) des classes en début de lactation. Les résidus moyens avant le pic de lactation sont symétriques entre les splines et Ali et Schaeffer. Les splines sous estiment puis surestiment la production alors que la fonction d'Ali et Schaeffer procède inversement tout en ayant des résidus moyens plus faibles. Lors du pic de lactation, les classes fixes ont les résidus les plus élevés. Ce mauvais ajustement aux données est probablement dû aux regroupements de classes qui ont été faits et qui ne permettent pas de bien apprécier les particularités de chèvres extrêmes. En effet, les classes ont été réalisées à partir de courbes moyennes et nous avons vu que certaines catégories de chèvres peuvent s'en éloigner de façon importante. Entre le stade 60 jours et le stade 280 jours, les fonctions splines et Ali et Schaeffer sont équivalentes et meilleures que les classes fixes. Entre 280 jours et 300 jours les splines sont meilleures que les deux autres modèles qui sous estiment les données. Comme espéré, les résidus moyens ont été bien réduits par rapport au modèle 0.

Les figures 26 à 29 représentent les différentes fonctions de chaque facteur modélisées selon Ali et Schaeffer pour le modèle 4. Les écarts entre les courbes ne peuvent pas être commentés car ce sont les interactions avec la région qui vont fixer les niveaux de ces courbes. La figure 26 montre les

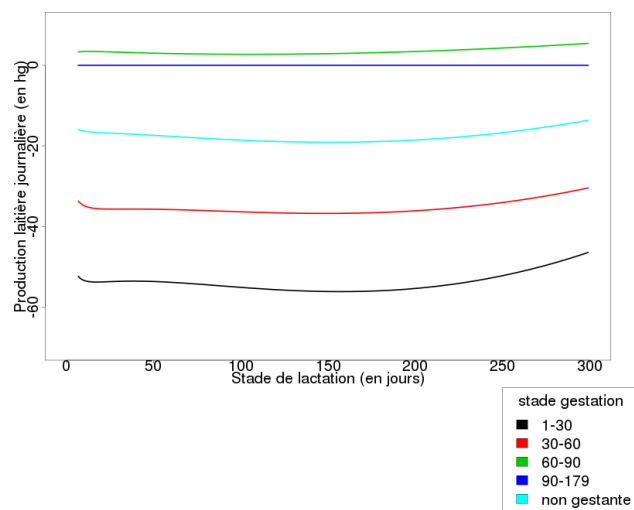


Figure 26 : Courbe de lactation modélisée avec le modèle 4 utilisant Ali et Schaeffer en fonction du stade de la gestation pour les chèvres Alpines en troisième lactation et plus. Le deuxième jeu de données a été utilisé.

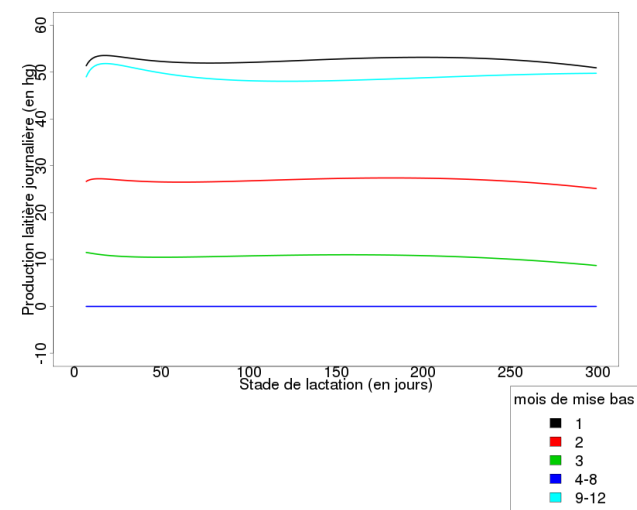


Figure 27 : Courbe de lactation modélisée avec le modèle 4 utilisant Ali et Schaeffer en fonction du mois de mise bas pour les chèvres Alpines en troisième lactation et plus. Le deuxième jeu de données a été utilisé.

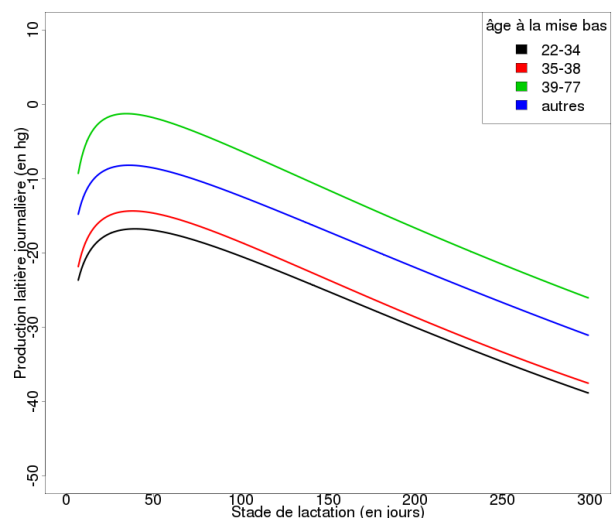


Figure 28 : Courbe de lactation modélisée avec le modèle 4 utilisant Ali et Schaeffer en fonction de l'âge à la mise bas pour les chèvres Alpines en troisième lactation et plus. Le deuxième jeu de données a été utilisé.

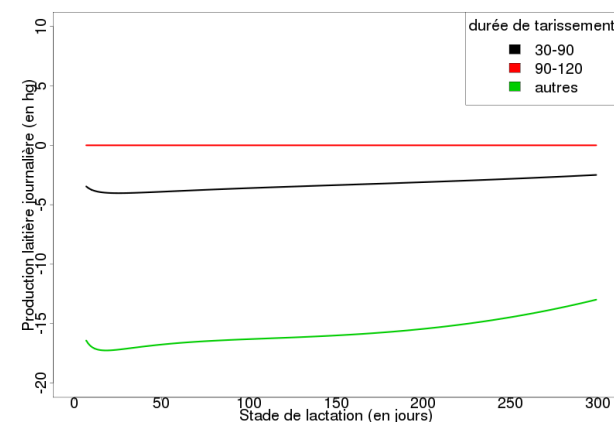


Figure 29 : Courbe de lactation modélisée avec le modèle 4 utilisant Ali et Schaeffer en fonction de la durée de tarissement pour les chèvres Alpines en troisième lactation et plus. Le deuxième jeu de données a été utilisé.

différences entre les différents stades de gestation. Il peut être observé une gradation entre les chèvres peu gestantes (moins de 60 jours) qui ont une courbe dédiée à la gestation qui croît à partir de 200 jours et celle des plus gestantes (plus de 90 jours) qui est parfaitement plate. On peut voir que l'accroissement de la production commence d'autant plus tôt que la chèvre sera moins gestante.

La figure 27 souligne la plus grande persistance en fin de lactation des chèvres ayant mis bas entre septembre et décembre alors que la production a tendance à baisser pour les chèvres mettant bas entre janvier et mars. Le pic est plus marqué pour les chèvres qui mettent bas entre septembre et janvier. Les courbes selon l'âge à la mise ne semblent pas très différentes les unes des autres même si le pic de lactation semble plus précoce pour les chèvres les plus âgées lors de la mise bas (Figure 28). La figure 29 montre une plus grande persistance des chèvres qui sont tarées moins longtemps. Ces courbes illustrent les différences de courbes que nous avons déjà mises en avant. Contrairement aux autres parties, les facteurs sont ici indépendants les uns des autres et l'effet troupeau est pris en compte.

La figure 30 montre les différents termes du modèle d'une chèvre Alpine en deuxième lactation du NO ayant mis bas en janvier, âgée de 39 mois, qui a été tarée 53 jours entre la deuxième et la troisième lactation et qui a fini sa lactation gestante de 91 jours. Les quantités de lait relevées lors des contrôles de cet animal et la valeur de l'HTD correspondant à ces contrôles ont été représentées. Si l'on somme les interactions, les fonctions et l'HTD, pour un stade donné, nous obtiendront la valeur prédite par le modèle. Une courbe appelée « production prédite simplifiée » a été obtenue en moyennant les HTD des contrôles de cet animal et en sommant pour chaque stade les interactions et les fonctions. Les estimations sont plutôt bonnes pour cet animal. Pour le deuxième contrôle, l'animal est en dessous de la valeur prédite. Un outil technique, qui aurait dans sa base de données la courbe prédite simplifiée, aurait prévenu l'éleveur d'un décrochage de l'animal éventuellement induit par une maladie. Le 5^{ème} contrôle illustre l'amélioration de la prédiction par HTD. En effet, le HTD était particulièrement haut lors de ce contrôle et notre modèle a pu estimer la production de l'animal avec précision alors que si l'effet troupeau avait une moyenne sur l'année, la prédiction aurait été plus proche de la courbe de la production prédite simplifiée et aurait donc été moins bonne (résidu plus grand).

Discussion

Dans notre étude, les chèvres sont plus productives que dans l'étude réalisée par Bouloc (1991) : en moyenne 423 kg de lait de plus pour les primipares et 315kg de plus pour les multipares. Par contre, la forme des courbes de production laitière est similaire (date du pic, persistance) selon le rang de lactation. León *et al.* (2012), qui a analysé la production de 38 000 chèvres de race Murciano-Granadina en suivant leurs 5 premières lactations, a montré, comme dans notre étude, qu'en première lactation la persistance est plus importante qu'en deuxième lactation qui est elle-même plus importante qu'en troisième lactation et plus.

A l'aide de la classification, nous avons pu caractériser les courbes de production laitière les plus couramment rencontrées dans la population de chèvres françaises de race Alpine et Saanen. Le niveau de production de certains clusters était particulièrement élevé, à un niveau supérieur à la moyenne. Nous avons pu mettre en évidence l'effet de différents facteurs environnementaux sur les profils de production laitière tels que la race, l'âge à la mise bas, le mois de mise bas ou encore la région. Nous avons pu également identifier des chèvres ayant des profils atypiques. Par exemple, certaines chèvres en première lactation, très minoritaires (0.7% des premières lactations) appartiennent aux clusters ayant une très haute production laitière et présentant un pic de lactation prononcé et très haut. D'autres chèvres produisent beaucoup en ayant, malgré tout, des courbes de lactation

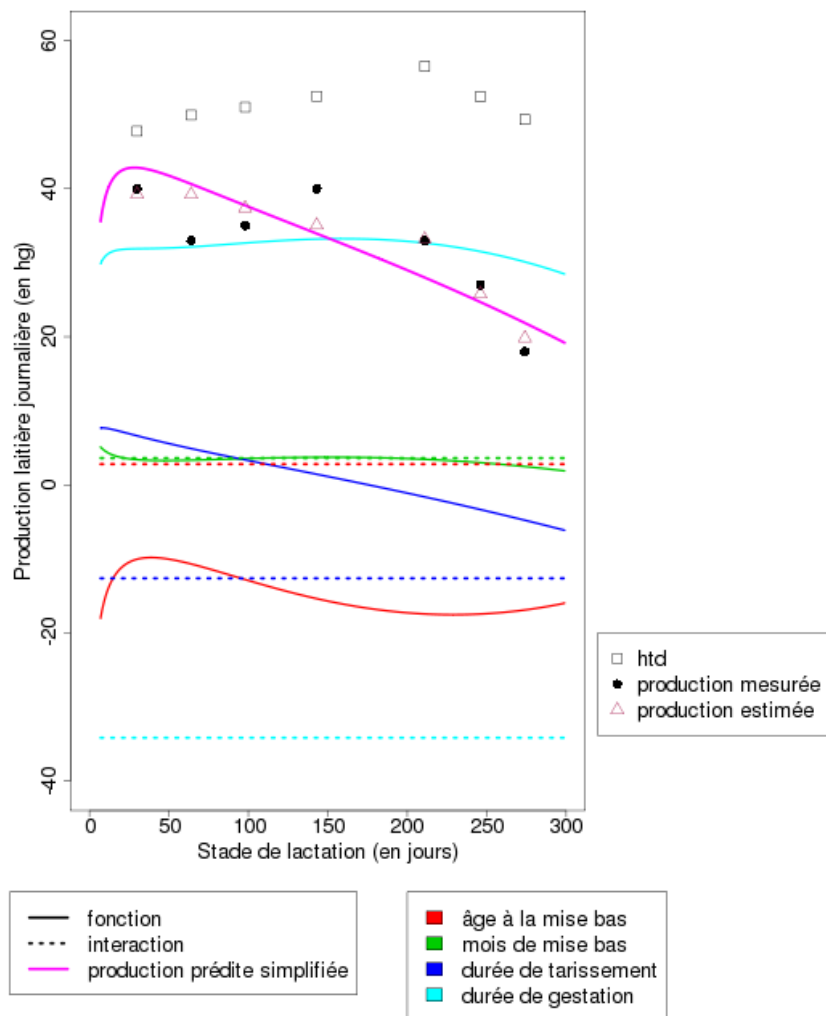


Figure 30 : Production mesurée (●), production estimée (avec le modèle 4 utilisant Ali et Schaeffer) (▲) d'une chèvre Alpine en deuxième lactation du NO ayant mis en troisième lactation en janvier, âgée de 39 mois, qui a été tarie 53 jours entre la deuxième et la troisième lactation et qui a fini sa troisième lactation gestante de 91 jours. L'effet de l'HTD au moment de ses contrôles (◻), les fonctions des différents facteurs (traits pleins) et les interactions des différents facteurs (traits pointillés) sont également représentés.

plutôt « plates », même si en général la succession de courbes au profil « plat » conduit à une production globale moindre. A l'inverse, on note qu'il existe des lactations de chèvres de rang supérieur à 1, avec des faibles niveaux de production et des courbes de lactation proches de celles rencontrées en première lactation.

Bouloc (1991) avait réalisé une classification basée sur 3 critères : le rapport entre la production à 305 jour et la production au pic afin d'estimer la persistance de la chèvre, le niveau de production et la durée de la lactation. Elle avait ainsi identifié 8 clusters et pu mettre en évidence : (1) des lactations aux formes atypiques avec une production médiocre durant les premières semaines puis la production remontait à un niveau normal par la suite, (2) des courbes persistantes avec une différence de production entre le pic et la fin de la lactation de 15 hg, (3) des courbes avec des pics marqués et une différence de production de plus de 20hg. Nous n'avons pas pu mettre en évidence dans notre classification les mêmes courbes atypiques que celles trouvées par Bouloc (1991). Une hypothèse pour expliquer cette absence est peut-être la manière dont les données manquantes sont imputées par le logiciel en prenant en compte la trajectoire moyenne de la population. Bouloc (1991) avait remarqué, comme dans notre étude, que les chèvres de race Saanen étaient plus présentes dans les groupes des lactations persistantes alors que les chèvres de race Alpine étaient plus présentes dans les groupes des lactations moins persistantes. Une différence de répartition des animaux les plus productifs selon la région a été observée dans l'étude de Bouloc (1991) comme dans la nôtre : les chèvres des Pays de la Loire (Nord-ouest dans notre étude) étaient plus présentes dans les groupes à forte production alors que ceux de Midi-Pyrénées (Sud dans notre étude) étaient plus présentes dans les groupes produisant moins. Contrairement à notre étude, elle n'a pas remarqué de différences de répartition des chèvres selon leur rang de lactation ou selon leurs différents facteurs environnementaux, entre les différents clusters

De nombreuses études utilisent des modèles relativement simples, avec 3 paramètres uniquement (Cappio-Borlino *et al.*, 1995 ; Cobby et Le Du, 1978 ; Wood, 1967), pour modéliser les courbes de lactation des chèvres (Fernández *et al.*, 2002 ; Gipson et Grossman, 1990 ; Macciotta *et al.*, 2008 ; Montaldo *et al.*, 1997 ; Nagel-Alne *et al.*, 2014). De plus, ces études s'appuient le plus souvent sur des effectifs réduits (entre 1000 et 38 000 chèvres) et répartis dans peu de troupeaux (maximum 130).

En vaches laitières des modèles plus complexes ont été utilisés. Silvestre *et al.* (2006), à partir d'une étude sur 144 lactations de vaches laitières, a comparé le modèle d'Ali et Schaeffer, les splines cubiques, les polynômes de Legendre et le modèle de Wilmink. Comme dans notre étude, il a trouvé peu de différences entre les différents modèles bien que les splines avaient le meilleur niveau d'ajustement devant le modèle d'Ali et Schaeffer et le polynôme de Legendre de degré 4. Druet *et al.* (2003) a étudié des données de vaches laitières en France (95 000 vaches, 776 000 contrôles) et a comparé les modèles de Legendre de degré 4, de Wilmink, d'Ali et Schaeffer, avec des classes fixes de 5 jours et des splines cubiques composées de 6 nœuds. Comme dans notre étude, il a obtenu le meilleur AIC et fonction de vraisemblance pour le modèle avec des classes fixes, il a eu le meilleur BIC pour Ali et Schaeffer, les splines arrivant en deuxième position. Les nœuds pour les splines ont été choisis aux positions 20, 50, 130, 230 pour les facteurs âge à la mise bas, mois de mise bas et durée de tarissement ; et aux positions 173, 206, 239, 272 pour la gestation. Comme dans notre étude, les meilleures positions des nœuds étaient celles avec une concentration plus élevée au début de la courbe pour les premiers facteurs. Les splines avaient été choisies dans leur étude car les résidus moyens étaient moins importants au début et à la fin de la lactation.

Le modèle que nous avons retenu (le modèle 4 utilisant Ali et Schaeffer), fourni de très bons

résultats d'ajustement avec des résidus moyens inférieurs à 0.4 hg tout au long de la lactation. Il a été choisi selon son BIC afin de limiter le nombre de paramètres à estimer. Ce modèle confirme l'influence des différents facteurs environnementaux sur la forme de la courbe de lactation mis en évidence dans l'observation des courbes et la classification. Leclerc (2008), dans son étude sur des vaches laitières en France, a étudié les effets environnementaux influant la forme de la courbe de lactation. Contrairement à notre étude, elle a inclus l'effet de la région pour modéliser la forme de la courbe de lactation chez les bovins.

Dans son étude, l'effet environnemental le plus important était l'effet du stade de gestation comme dans notre étude. Tous les autres effets étaient significatifs avec, par ordre d'importance, le mois de mise bas puis la durée de tarissement et enfin l'âge à la mise bas. Dans notre étude, la durée de tarissement est l'effet le moins important. Le stade de gestation qui n'est pas utilisé aujourd'hui semble avoir un effet important sur la forme de la courbe en fin de lactation. Il pourrait être pris en compte dans le cadre d'une évaluation génétique.

L'effet du mois de mise bas sur la persistance avait déjà été mis en avant dans les études de Montaldo *et al* (1997), de León *et al* (2012) et de Bouloc (1991). Dans ces études, la plus grande persistance des chèvres ayant mis bas entre octobre et janvier est expliquée par la poussée de l'herbe printanière, riche, qui intervient après le pic de lactation et qui maintient la production laitière.

Bouloc (1991) a montré que l'effet de l'âge à la mise bas était moins important sur la production de lait totale que la période de mise bas, et qu'il engendre les mêmes conséquences sur la forme de la courbe de lactation que dans notre étude. En effet, les chèvres mettant bas avant 12 mois ont une meilleure persistance. Cet effet a été observé également chez les vaches laitières (Leclerc, 2008).

En ce qui concerne l'effet race, Bouloc (1991) n'avait pas mis en évidence de différence de forme de courbe de lactation entre les deux races Alpine et Saanen pour les premières lactations alors que dans notre étude, ces différences existent. Pour les multipares, Bouloc (1991) avait identifié, comme nous, une meilleure persistance pour les chèvres de race Saanen.

Pour l'effet de la région, les écarts entre régions que nous observons semblent s'être accentués avec des courbes qui n'étaient pas différentes pour les multipares dans l'étude de Bouloc (1991) alors qu'aujourd'hui la différence entre le NO et les autres régions est importante.

Leclerc (2008) a utilisé un modèle semblable au nôtre, bien qu'elle ait pris en compte l'effet « année » dans ses interactions. En effet, elle a montré que cet effet influençait le niveau de production sans modifier la forme de la courbe de lactation. Dans notre étude, cet effet ne semblait pas important car nos jeux de données sont, au plus, étalés sur 3 années consécutives. Cependant dans le cadre d'une mise en place de ce modèle pour l'évaluation génétique nationale (qui comprend des données sur plus de 30 ans), il apparaît nécessaire de tester ce facteur.

León (2012) a montré que le nombre de chevreaux influençait le niveau production et la forme de la courbe de lactation. Dans son étude, plus la chèvre avait de chevreaux, plus elle produisait du lait et plus son pic de lactation était prononcé. Zumbach *et al.* (2008) tout comme Fernández *et al.* (2002) ont d'ailleurs utilisé le facteur « nombre de chevreaux nés » dans leur modèle d'évaluation génétique. C'est un facteur qu'il serait peut-être intéressant de relever en ferme afin de l'étudier.

Afin de quantifier l'effet de l'alimentation sur la forme des courbes de lactation, des données sur l'alimentation des chèvres auraient été intéressantes. Elles permettraient de comprendre les effets HTD. Cependant, cet effet est déjà intégré dans notre effet HTD puisqu'il représente l'ensemble des effets environnementaux d'un troupeau un jour de contrôle donné. Cet effet dans notre modèle représente une très grande partie du R^2 et son effet a d'ailleurs pu être illustré dans un exemple (Figure

30). L'effet troupeau avait d'ailleurs été évalué comme étant le facteur le plus important par Bouloc (1991).

Conclusion et perspectives

Dans cette étude, à partir de deux jeux de données de grande taille de mesures en ferme de la quantité de lait produite lors des 2 traites quotidiennes, il a été possible d'étudier l'influence de facteurs environnementaux sur la forme des courbes de lactation des chèvres en France. Une analyse par classification de la variabilité des formes de courbes a permis de mettre en évidence 7 profils de courbes qui ont ensuite été caractérisés en fonction de facteurs environnementaux et du niveau global de production. Des chèvres produisent plus que la moyenne et des formes de courbe « plates » ont pu être mis en évidence. Ces courbes ont ensuite été modélisées en fonction du rang de lactation par des modèles relativement complexes de différentes catégories: les paramétriques, les semi paramétriques et les non paramétriques. Ces modèles ont ensuite été complétés en intégrant les différents facteurs environnementaux connus ainsi qu'un effet « cheptel-jour de contrôle » (HTD). Le modèle utilisant une modélisation d'Ali et Schaeffer des facteurs tout au long de la courbe de lactation a été préféré pour son très bon ajustement et son faible nombre de paramètres à estimer.

Par la suite, afin de compléter l'exploration de la variabilité des courbes, nous pourrions étudier les résidus de chaque contrôle et refaire une classification en utilisant les résidus à la place des observations. Cette classification prendra en compte l'effet HTD. Il sera alors intéressant de repérer et ensuite étudier le profil des chèvres atypiques, les plus éloignées des courbes moyennes.

Afin d'améliorer le modèle et l'adapter à la base de donnée tout en entière, l'effet « année » devra sans doute être pris en compte. De même, l'effet de la race devra aussi être analysé. Un modèle plus simple n'incluant que l'effet région pour fixer le niveau de production pourra être évalué.

Ce genre de modélisation ouvre des perspectives en appui technique notamment grâce au facteur HTD. Il peut permettre de comparer différents élevages de la même région à la même période d'une même année, ou de mettre en évidence des problèmes d'élevage.

Mais la meilleure estimation des effets de ces facteurs d'environnement doit prendre en compte simultanément les effets propres à l'animal dus à sa génétique ou à son histoire (l'effet d'environnement permanent). Pour cela un modèle TDM doit être développé en chèvres en France. Il sera alors possible d'identifier la partie génétique responsable de la persistance qu'un reproducteur pourra transmettre à sa descendance et ainsi sélectionner les chèvres les plus persistantes.

Il serait intéressant d'étudier les courbes de lactation (assez nombreuses) des chèvres qui ont une durée de lactation de plus de 350 jours. En effet, ce sont des animaux élevés dans des élevages recherchant des lactations très longues. La persistance dans de tels élevages est primordiale, s'agit-il d'une aptitude différente à moins ou plus de 350 jours de lactation ? Cette pratique d'élevage permet de limiter le temps de vie de la chèvre passé en début de lactation, période à risque pour les animaux.

Une autre amélioration de ces modèles serait de prendre en compte l'information de marqueurs ADN pour appréhender plus précisément la partie génétique et permettre ainsi une sélection génomique qui accélère le progrès génétique et réduit l'intervalle générationnel.

Bibliographie

- Agreste (2012). Recensement agricole 2010-Poitou Charentes-L' élevage caprin, *Direction Régionale de l'Alimentation, de l'Agriculture et de la Forêt*, n° 13, 4p.
- Akaike H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, Vol. 19, n°6, pp. 716–723.
- Ali T. E., Schaeffer L. R. (1987). Accounting for covariances among test day milk yields in dairy cows. *Canadian Journal of Animal Science*, Vol. 67, n°3, pp. 637–644.
- 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*. Thèse de doctorat INA-PG, Paris
- Bouloc N. 1993. Guide d'utilisation du référentiel courbe de lactation. *Compte rendu Institut de L'Elevage*, N°2086, 68p.
- Caliński T., Harabasz J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, Vol. 3, n°1, pp. 1–27.
- Cappio-Borlino A., Pulina G., Rossi G. (1995). A non-linear modification of Wood's equation fitted to lactation curves of Sardinian dairy ewes. *Small Ruminant Research*, Vol. 18, n°1, pp. 75–79.
- Cobby J. M., Le Du Y. L. P. (1978). On fitting curves to lactation data. *Animal Production*, Vol. 26, n°2, pp. 127–133.
- Druet T., Jaffrézic F., Boichard D., Ducrocq V. (2003). Modeling lactation curves and estimation of genetic parameters for first lactation test-day records of French Holstein cows. *Journal of dairy science*, Vol. 86, n°7, pp. 2480–2490.
- Fernández C., Sánchez A., Garcés C. (2002). Modeling the lactation curve for test-day milk yield in Murciano-Granadina goats. *Small Ruminant Research*, Vol. 46, n°1, pp. 29–41.
- Fleischmann W. (1932). *Lehrbuch der milchwirtschaft*. 7 éd., P. Parey, Berlin, p. 84
- Genolini C., Alacoque X., Sentenac M., Arnaud C. (2015). kml and kml3d: R Packages to Cluster Longitudinal Data. *Journal of Statistical Software*, Vol. 65, n°1, pp. 1–34.
- Gipson T. A., Grossman M. (1990). Lactation curves in dairy goats: a review. *Small Ruminant Research*, Vol. 3, n°4, pp. 383–396.
- Grossman M., Hartz S. M., Koops W. J. (1999). Persistency of lactation yield: A novel approach. *Journal of Dairy Science*, Vol. 82, n°10, pp. 2192–2197.
- Institut de l'élevage (2014) Résultats de contrôle laitier-espèce caprine, *Institut de l'élevage*, Paris, 32p
- Institut de l'élevage (2015) Chiffres clés 2015- Productions caprines- lait et viande, *GEB- Institut de l'élevage*, Paris, 10p
- Keown J. F., Everett R. W., Empet N. B., Wadell L. H. (1986). Lactation curves. *Journal of Dairy Science*, Vol. 69, n°3, pp. 769–781.
- Kirkpatrick M., Lofsvold D., Bulmer M. (1990). Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics*, Vol. 124, n°4, pp. 979–993.
- Larroque H., Astruc J.-M., Barbat A., Barillet F., Boichard D., Bonaiti B., Clément V., David I., Lagriffoul G., Palhière I., Piacère A., Robert-Granie C., Rupp R. (2011). National genetic evaluations in dairy sheep and goats in France . In: *Proceedings of the 62th European federation of animal science (EAAP): 29 August—2 September 2011; Stavanger, Norway*, 344 p.

- Larroque H., Barillet F., Baloché G., Astruc J., Buisson D., Shumbusho F., Clément V., Lagriffoul G. (2014). Toward genomic breeding programs in French dairy sheep and goats. In : 10th World Congress on Genetics Applied to Livestock Production, Vancouver, Canada, 17-22 août 2014
- Leclerc H. (2008). 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*, Paris, 258 p
- Lefrileux Y., Morand-Fehr P., Pommaret A. (2012). Aptitude des chèvres hautes productrices de lait à valoriser les prairies temporaires au pâturage. *Productions Animales*, Vol. 25, n°3, pp. 277-290
- Lenth R. V. (2016). Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software*, Vol. 69, n°1, pp. 1–33.
- León J. M., Macciotta N. P. P., Gama L. T., Barba C., Delgado J. V. (2012). Characterization of the lactation curve in Murciano-Granadina dairy goats. *Small Ruminant Research*, Vol. 107, n°2–3, pp. 76-84.
- Macciotta N. P., Fresi P., Usai G., Cappio-Borlino A. (2005). Lactation curves of Sarda breed goats estimated with test day models. *Journal of dairy research*, Vol. 72, n°4, pp. 470–475.
- Macciotta N. P., Dimauro C., Steri R., Cappio-Borlino A. (2008). Mathematical modelling of goat lactation curves. *Dairy goats feeding and nutrition*, pp. 31–46.
- Meyer K., Graser H.-U., Hammond K. (1989). Estimates of genetic parameters for first lactation test day production of Australian Black and White cows. *Livestock Production Science*, Vol. 21, n°3, pp. 177–199.
- Ministère de l'agriculture, de l'agroalimentaire et de la forêt (2014) Référentiel de contrôle des performances dans le cadre du service public d'enregistrement et de contrôle des performances des ruminants pour la filière « production de lait de chèvre »-Protocoles A, AT, AZ, CZ, *Bulletin officiel du 6 octobre 2014*, Paris, 32 p
- Montaldo H., Almanza A., Juárez A. (1997). Genetic group, age and season effects on lactation curve shape in goats. *Small Ruminant Research*, Vol. 24, n°3, pp. 195–202.
- Nagel-Alne G. E., Krontveit R., Bohlin J., Valle P. S., Skjerve E., Sølverød L. S. (2014). The Norwegian Healthier Goats program—Modeling lactation curves using a multilevel cubic spline regression model. *Journal of dairy science*, Vol. 97, n°7, pp. 4166–4173.
- Peyraud J.-L., Agabriel J., Benoit M., Duhem K., Lagriffoul G., Legarto J., Morin E. (2013). Vers des systèmes d'élevage de ruminants à hautes performances. In : 20. *Rencontres autour des Recherches sur les Ruminants*, Institut de l'Elevage-INRA, pp. 21-30
- Ptak E., Schaeffer L. R. (1993). Use of test day yields for genetic evaluation of dairy sires and cows. *Livestock Production Science*, Vol. 34, n°1, pp. 23–34.
- Ray S., Turi R. H. (1999). Determination of number of clusters in k-means clustering and application in colour image segmentation. In : *Proceedings of the 4th international conference on advances in pattern recognition and digital techniques*, Calcutta, India, 1999, pp. 137–143.
- Roumeas A., Gaudillière N., Dubief F., Adam H., BELOT P., Delaby L. (2014). Pic de lactation, persistance et lien avec les performances de reproduction de vaches Montbéliarde en Franche-Comté. In : 21. *Rencontres autour des Recherches sur les Ruminants*, Institut de l'Elevage-INRA, pp.277-280
- Royston P., Altman D. G. (1994). Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Applied statistics*, pp. 429–467.
- Schaeffer L.R. and Dekkers J.C.M. 1994. Random regressions in animal model for test-day production

in dairy cattle. In: *Proceedings of the 5th World Congress on Genetic Applied to Livestock Production*, Guelph, Canada, 7-12 août 1994, pp. 443-446.

Schwarz G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, Vol. 6, n°2, pp. 461–464.

Silvestre A. M., Petim-Batista F., Colaco J. (2006). The accuracy of seven mathematical functions in modeling dairy cattle lactation curves based on test-day records from varying sample schemes. *Journal of dairy science*, Vol. 89, n°5, pp. 1813–1821.

Sölkner J., Fuchs W. (1987). A comparison of different measures of persistency with special respect to variation of test-day milk yields. *Livestock Production Science*, Vol. 16, n°4, pp. 305–319.

White I. M. S., Thompson R., Brotherstone S. (1999). Genetic and environmental smoothing of lactation curves with cubic splines. *Journal of Dairy Science*, Vol. 82, n°3, pp. 632–638.

Wilmink J. B. M. (1987). Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. *Livestock Production Science*, Vol. 16, n°4, pp. 335–348.

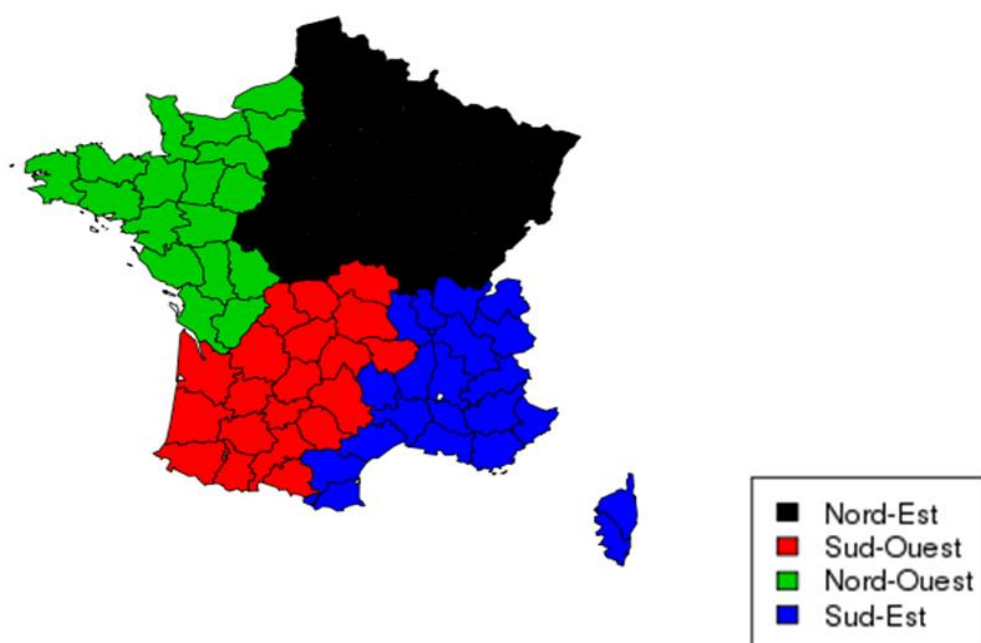
Wood P. D. P. (1967). Algebraic model of the lactation curve in cattle. *Nature*, Vol. 216, pp. 164-165.

Zumbach B., Tsuruta S., Misztal I., Peters K. J. (2008). Use of a test day model for dairy goat milk yield across lactations in Germany. *Journal of Animal Breeding and Genetics*, Vol. 125, n°3, pp. 160–167.

Annexe

Annexe 1 : Liste des numéros des départements par région

Régions	Départements
Nord-Est (NE)	2,8,10,18,21,25,28,36,37,39,41,45,51,52,54,55, 57,58,59,60,62,67,68,70,71,75,77,78,80,88,89, 90,91,92,93,94,95
Sud-Ouest (SO)	3,9,12,15,19,23,24,31,32,33,40,43,46,47,63,64, 65,81,82,87
Nord-Ouest (NO)	14,16,17,22,27,29,35,44,49,50,53,56,61,72,76, 79,85,86,
Sud-Est (SE)	1,4,5,6,7,11,13,20,26,30,34,38,42,48,66,69,73, 74,83,84



Annexe 2 : Délimitation des 4 régions françaises

Annexe 3 : Nombre de paramètres, logarithme de la fonction de vraisemblance ($\log(L)$), critère d'information d'Akaike (AIC), critère d'information bayésien (BIC), écart type de l'erreur (sigma) pour les différents modèles testés. Ils sont de la forme $y_i = f(lact_i) + e_i$. Les données utilisées sont issues du jeu de données 1.

Modèle		Nombre de paramètres	Log(L)	AIC	BIC	sigma
Wilmink		16	- 5 055 310	10 110 653	10 110 859	9.780
Ali et Schaeffer		20	- 5 054 591	10 109 224	10 109 478	9.775
Legendre	leg5	24	- 5 054 774	10 109 598	10 109 901	9.776
	leg4	20	- 5 055 384	10 110 811	10 111 065	9.781
Fractionnaire degré 2	(1)	12	- 5 062 251	10 124 528	10 124 686	9.830
	(2)	12	- 5 054 999	10 110 032	10 110 239	9.778
	(3)	12	- 5 055 039	10 110 112	10 110 318	9.778
	(4)	12	- 5 055 080	10 110 185	10 110 343	9.778
Fractionnaires degré 3	(1)	16	- 5 054 717	10 109 468	10 109 674	9.776
	(2)	16	- 5 054 787	10 109 608	10 109 814	9.776
	(3)	16	- 5 054 698	10 109 429	10 109 636	9.776
	(4)	16	- 5 054 930	10 109 894	10 110 100	9.777
	(5)	16	- 5 054 713	10 109 461	10 109 667	9.776
	(6)	16	- 5 054 825	10 109 683	10 109 890	9.777
	(7)	16	- 5 054 730	10 109 493	10 109 699	9.776
	(8)	16	- 5 055 580	10 111 194	10 111 400	9.782
classes fixes	fix1	236	- 5 053 732	10 107 937	10 110 812	9.770
	fix2	124	- 5 054 123	10 108 496	10 110 012	9.772
	fix3	92	- 5 054 678	10 109 542	10 110 670	9.776
	fix4	120	- 5 054 271	10 108 785	10 110 252	9.773
	fix5	67	- 5 054 634	10 109 404	10 110 228	9.776
splines	7, 60, 120, 180, 240,299	32	- 5 054 421	10 108 907	10 109 308	9.774
	7, 80, 180, 200, 220,299	32	- 5 054 284	10 108 633	10 109 033	9.773
	7, 80, 160, 180, 260,299	32	- 5 054 427	10 108 920	10 109 320	9.774
	7, 30, 60, 100, 230,299	32	- 5 054 485	10 109 035	10 109 436	9.774
	7, 30, 60, 100, 200,299	32	- 5 054 479	10 109 024	10 109 424	9.774

Polynômes fractionnaires de degré 2 : (1) : $f_1(t) = a + b \ln(t) + c(\ln(t))^2$, (2) : $f_2(t) = a + b \ln(t) + ct^d$ avec $d \neq 0$, (3) $f_3(t) = a + bt^d + c \ln(t) t^d$, (4) $f_4(t) = a + bt^{b_1} + ct^{b_2}$ avec $b_1 < b_2$; Polynômes fractionnaires de degré 3 : (1) $f_1(t) = a + bt^{b_1} + ct^{b_2} + dt^{b_3}$, (2) $f_2(t) = a + bt^{b_1} + ct^{b_2} + dt^{b_2} \ln(t)$, (3) $f_3(t) = a + bt^{b_1} + ct^{b_1} \ln(t) + dt^{b_2}$, (4) $f_4(t) = a + bt^{b_1} + ct^{b_1} \ln(t) + dt^{b_1} \ln^2 t$, (5) $f_5(t) = a + b \ln(t) + ct^{b_1} + dt^{b_2}$, (6) $f_6(t) = a + b \ln(t) + ct^{b_1} + dt^{b_1} (\ln(t))$, (7) $f_7(t) = a + b \ln(t) + c \ln(t)^2 + dt^{b_1}$, (8) $f_8(t) = a + b \ln(t) + c \ln(t)^2 + d \ln(t)^3$

Annexe 4 : Choix de l'emplacement des nœuds des splines à partir du modèle 4 et des données du deuxième jeu de données pour la race Alpine

Nœuds	Nombre de paramètres	Log(L)	AIC	BIC	Sigma	R ²
7, 60, 120, 180, 240,299	14094	-4 614 826.6	9 257 843	9 428 290.0	7.7035	0.7459
7, 80, 180, 200, 220,299	14094	-4 615 509.4	9 259 209	9 429 655.7	7.7074	0.7456
7, 80, 160, 180, 260,299	14094	-4 615 817.5	9 259 825	9 430 271.8	7.7092	0.7454
7, 30, 60, 100, 230,299	14094	-4 613 488.9	9 255 168	9 425 614.6	7.6957	0.7465
7, 30, 60, 100, 200,299	14094	-4 613 489.2	9 255 168	9 425 615.2	7.6957	0.7465



Diplôme : Diplôme d'Ingénieur de l'Institut Supérieur des Sciences Agronomiques, Agroalimentaires, Horticoles et du Paysage

Spécialité : **Sciences de l'Animal pour l'élevage de demain (SAED)**

Spécialisation / option :

Enseignant référent : Jocelyne FLAMENT

Auteur(s) : Mathieu ARNAL

Date de naissance* : 21/07/1992

Nb pages : 20 Annexe(s) : 4

Année de soutenance : 2016

Organisme d'accueil : INRA

Adresse : 24 Chemin de Borde Rouge,
31326 Castanet-Tolosan

Maître de stage : Hélène LARROQUE,
Christèle ROBERT-GRANIE

Titre français : Analyse et modélisation des formes de courbes de lactation des caprins en France

Titre anglais : Analysis and modeling of lactation curves forms in goats in France

Actuellement les chèvres sont sélectionnées sur la quantité totale de lait produite au cours de la lactation. La sélection ne tient pas compte de l'évolution de la quantité de lait journalière produite qui forme la courbe de la lactation. Celle-ci peut être décomposée en 3 phases : une phase ascendante, suivie d'un pic qui peut parfois se maintenir et former un plateau, et d'une phase descendante. Le degré de maintien de la production au pic est appelé persistance. La persistance est intéressante car à un même niveau de production au cours d'une lactation, une chèvre persistante aura son déficit énergétique moins marqué et des besoins en intrants inférieurs. L'objectif de cette étude est d'analyser les formes des courbes de lactation et de les modéliser à des fins d'évaluation génétique par la suite, notamment pour évaluer la part génétique de la persistance. L'étude est basée sur l'analyse des contrôles élémentaires de 213 000 chèvres, réalisés en ferme, par le contrôle laitier. Nous avons évalué l'influence de facteurs environnementaux (rang de la lactation, région, race, âge à la mise bas, mois de mise bas, durée de tarissement, stade de gestation) sur la courbe de lactation. Tous ont un effet sur le niveau de la courbe et/ou sa forme. Une classification a été réalisée afin d'étudier la diversité des formes de courbe de lactations. Nous avons pu mettre en évidence 7 profils de courbes différents dont certains à des niveaux de production élevés et/ou au profil « plat ». Différents modèles présents dans la littérature (modèles paramétriques, semi-paramétriques et non paramétriques) ont ensuite été testés en fonction du rang de lactation. En incluant les différents facteurs environnementaux (par une modélisation d'Ali et Schaeffer) et l'effet du cheptel pour le jour de contrôle (HTD), nous avons réussi à obtenir un modèle satisfaisant par son degré d'ajustement aux données et par son nombre réduits de paramètres à estimer.

Currently goats are selected on the total quantity of milk produced during lactation. The selection does not take into account changes in the quantity of milk produced daily that forms the curve of lactation. It can be divided into three phases: a rising phase, followed by a peak that can sometimes maintain, and a descending phase. The degree of continued production peak is called persistence. Persistence is interesting because at the same level of total lactation production, persistent goat will have its energy deficit less marked and needs lower inputs. The objective of this study is to analyze the shapes of lactation curves and model them for genetic evaluation, especially to assess the genetic part of persistency. The study is based on analysis of test-day from 213,000 goats, made by French test program. The influence of environmental factors (rank of lactation, region, race, age at calving, month of calving, drying time, stage of gestation) were tested on lactation curve. All have an effect on the level of the curve and / or its shape. A classification was used to study the diversity of lactation curve shape. Seven different profiles were observed and some at high production levels. Different models present in literature (parametric models, semi-parametric and non-parametric) were then tested according to the rank of lactation. Including the different environmental factors (modeled with Ali and Schaeffer model) and the impact of Herd Test Day (HTD), a satisfactory model were obtained with a good adjustment to the data and a reduced number of parameters to estimate.

Mots-clés : courbe de lactation, modélisation, facteurs environnementaux, persistance, chèvre

Key Words: lactation curve, modeling, environmental factors, persistency, goat

* Elément qui permet d'enregistrer les notices auteurs dans le catalogue des bibliothèques universitaires