



**HAL**  
open science

# Development of the French dairy cattle test-day model genetic evaluation and prospects of using results for herd management.

Hélène Leclerc

► **To cite this version:**

Hélène Leclerc. Development of the French dairy cattle test-day model genetic evaluation and prospects of using results for herd management.. Life Sciences [q-bio]. AgroParisTech, 2008. English. NNT : 2008AGPT0100 . pastel-00004976

**HAL Id: pastel-00004976**

**<https://pastel.hal.science/pastel-00004976v1>**

Submitted on 27 May 2009

**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.



Institut des sciences et  
industries du vivant  
et de l'environnement



Ecole doctorale  
ABIES



Institut national de  
la recherche agronomique  
Station de Génétique  
Quantitative et Appliquée



Institut de l'Élevage  
Département de  
Génétique

N° / / / / / / / / / / / / / / / /

## THÈSE

pour obtenir le grade de

**Docteur**  
de

**l'Institut des Sciences et Industries du Vivant et de l'Environnement  
(Agro Paris Tech)**

Spécialité : Génétique animale

*présentée et soutenue publiquement par*

**Hélène LECLERC**

**le 19 décembre 2008**

**Mise en place de l'évaluation génétique  
sur les contrôles élémentaires en bovins laitiers  
et perspectives d'utilisation des résultats en appui technique**

*Directeur de thèse : Vincent DUCROCQ*

*Travail réalisé à l'INRA, UR337 Station de Génétique Quantitative et Appliquée, F-78352 Jouy-en-Josas*

*Avec le soutien financier du ministère de l'Agriculture (action innovante UTILEG)*

Devant le jury :

**Etienne VERRIER**

**Leo DEMPFLÉ**

**Christèle ROBERT-GRANIE**

**Vincent DUCROCQ**

**Philippe FAVERDIN**

**Christophe LECOMTE**

**Professeur, AgroParisTech**

**Professeur, Univ. Technique de Munich**

**Chargé de Recherche, INRA SAGA**

**Directeur de Recherche, INRA SGQA**

**Directeur de Recherche, INRA UMR-PL**

**Ingénieur, France Contrôle Laitier**

**Président**

**Rapporteur**

**Rapporteur**

**Examinateur**

**Examinateur**

**Examinateur**

*L'Institut des Sciences et Industries du Vivant et de l'Environnement (Agro Paris Tech) est un Grand Etablissement dépendant du  
Ministère de l'Agriculture et de la Pêche, composé de l'INA PG, de l'ENGREF et de l'ENSIA*

*(décret n° 2006-1592 du 13 décembre 2006)*



*A Méline et Solène,  
A mon frère qui je l'espère bénéficiera un jour  
des applications de ce travail,  
A mes racines...*



# Remerciements

*A l'heure où une page se tourne, je tenais à adresser mes plus sincères remerciements à ceux qui ont contribué à l'écriture de ce manuscrit.*

*Tout d'abord, je pense à toi Vincent. Merci mille fois. Pour les conseils, la confiance que tu as pu me porter depuis mon arrivée en mai 2002... Merci d'avoir cru en moi (parfois plus que moi), de m'avoir ouvert des portes notamment celles du labo de génétique moléculaire (Merci à Sandrine pour son encadrement) et d'Interbull (Merci à Freddy et toutes les personnes rencontrées à Uppsala mais également à l'UNCEIA pour le financement de ces 6 mois). Merci de m'avoir concocté un sujet sur-mesure (même si par la force des choses celui-ci a été amené à évoluer) mais aussi de t'être préoccupé de son financement. A ce titre, je tenais à remercier le Ministère de l'Agriculture qui, à travers l'action innovante UTILEG, a financé trois ans de mon travail.*

*Ensuite, mes remerciements s'adressent à Sophie. Merci pour le temps et l'énergie consacrés à m'aider, à monter un groupe de travail (Merci aux participants enthousiastes) et à trouver des financements pour poursuivre les recherches entamées et les concrétiser afin que les principaux intéressés (les éleveurs) bénéficient le plus rapidement possible de ces travaux. Je souhaitais également remercier Pierre-Louis Gastinel et Claude Allo d'avoir accepté de financer ma fin de thèse.*

*Mes remerciements s'adressent également à Jean-Pierre. Merci de m'avoir permis de réaliser ce travail au sein de l'ex-SGQA, désormais GABI dans des conditions matérielles idéales pour tout thésard. Merci également à Didier, chef du département Génétique Animale de l'INRA... Je travaille sur ton ex-bureau, reçois encore certains de tes coups de téléphone... mais je n'ai pas l'espoir de t'arriver un jour à la cheville !*

*Merci à tous les membres du Jury : Monsieur Etienne Verrier pour en avoir assuré la présidence, l'organisation matérielle de la soutenance et avoir pris en charge l'ensemble des relations avec l'école doctorale ABIES au cours de cette thèse, Pr. Leo Dempfle et Madame Christèle Robert-Granié pour leur implication en tant que rapporteur de ce manuscrit à une période où tous deux avaient un emploi du temps surchargé. Merci également à Messieurs Philippe Faverdin et Christophe Lecomte pour leur participation en tant qu'examineur. Je tenais également à remercier les membres de mon comité de thèse qui ont su m'orienter, m'encourager et parfois définir des priorités dans mon travail, notamment Tom Druet (bonne continuation en Belgique), Christophe Vuillet, Agnès Hervé, Philippe Faverdin et Etienne Verrier.*

*Un grand Merci à celles qui m'ont supporté pendant ces années passées parmi vous, notamment Hélène, Sophie puis Anne et enfin Armelle. Un merci tout particulier pour toi Armelle, pour ta patience et ta tolérance du « bordel » (désolé il n'y a pas d'ordre mot pour qualifier l'état perpétuel de mon bureau), les moments de détente que tu m'as accordé lors de nos discussions, tes trucs et astuces pour SAS et les scripts... Merci Anne d'avoir toléré mes interruptions soudaines dans ton bureau afin d'avoir des précisions sur le modèle lactation, un nouveau fichier de données... Merci Hélène pour ta disponibilité, tu resteras celle qui m'a initié à Unix et SAS et je t'en suis à jamais reconnaissante. Enfin, merci à l'ensemble des membres du groupe bovins laitiers INRA et Institut, désormais absorbés*

au sein de G<sup>2</sup>B. Je n'ai pas forcément eu l'occasion de travailler directement avec vous, mais vous avez toujours été disponibles pour répondre à mes questions diverses et variées. Un merci à Eric pour le temps consacré à résoudre mes problèmes « toujours urgents » liés à la manipulation de mes gros fichiers ! Merci à l'équipe informatique Fabien, Hervé et Sylvie... je manque parfois de patience vis-à-vis des machines... mais grâce à votre aide, je suis toujours parvenue à mes fins. Merci aux sauvegardes informatiques qui m'ont évité quelques heures de découragements. Merci également au CTIG... désolé pour les quelques plantages de DGA2 dont je suis responsable et mes incessantes demandes d'espace disque... Je souhaitais aussi remercier Wendy pour sa disponibilité et les corrections qu'elle a pu apporter à mon anglais ; nos secrétaires Manuëla, Sylvie et Christelle pour les multiples services rendus ; Annie pour me m'avoir permis de retrouver l'accès à certaines revues électroniques et Serge pour sa maîtrise des outils reprographiques.

Je voulais également remercier deux personnes qui m'ont permis de prendre du recul en m'obligeant à expliquer mon sujet de façon claire... enfin à y essayer : Delphine avec qui j'ai pu collaborer quelques mois. Merci pour le travail accompli, un peu répétitif certes, mais qui m'a bien aidé et Laure pour le mois passé à Jouy à t'initier au modèle sur les contrôles élémentaires. Je voulais aussi remercier Pilar et Istvan, deux post-docs avec qui j'ai eu l'occasion de partager un peu de mon travail.

Je voulais également souhaiter bon courage et plein de réussite à celles et ceux qui se sont lancés dans cette folle course qu'est la thèse. Il vous faudra une bonne dose de persévérance pour rejoindre la ligne d'arrivée. Mes encouragements s'adressent à tous, qu'ils viennent de prendre le départ (Clotilde), soient à mi-course (Btissam, Laure et Sophie), entrent dans la dernière ligne droite (Guillemette et Alban) ou franchissent prochainement la ligne d'arrivée (François et Félicie). Une pensée spéciale à Flo avec qui j'ai partagé quelques casse-tête pendant le DEA, mais aussi beaucoup de péripéties thésardiennes pendant 3 ans. Profites bien de Göttingen et reviens nous vite !

Je voulais aussi remercier le Dr Miléa. J'ai souvent pensé à votre petite histoire dont la conclusion était qu'il ne fallait pas laisser les aléas de la vie nous détourner de nos projets sous peine de le regretter un jour. J'ai été au bout de celui-ci et c'est un peu grâce à vous.

Merci à ma famille et plus particulièrement à mes parents, Lucie et Maxime, et mes nièces Solène et Méline. Merci de votre soutien et des moments de détente partagés avec vous. Quel bonheur de voir Solène crier « Tata » devant la web-cam ! Un merci particulier à toi papa pour avoir su me communiquer ta passion des vaches. Merci à Orfée, Qualité et sa descendance, Marcie et les autres d'avoir contribué à mon épanouissement. A 6 ans, c'est envoutant de découvrir un embryon sous la lame d'un microscope. Enfin, Merci à toi Pascal... je n'aurai pas eu à écrire ces pages si tu n'avais pas été là. Merci de m'avoir soutenue au quotidien, de près comme de loin... Merci du chemin parcouru depuis 5 ans... Le 20 octobre sera désormais une double fête ! Une page se tourne, tu es de retour en France et j'espère que nous allons désormais pouvoir avancer ensemble vers les projets qui nous tiennent à cœur...

Il est maintenant l'heure de poursuivre... la page est blanche. Il me faut désormais esquisser un nouveau décor... dans le prolongement du précédent (Merci Sophie). J'espère que les discussions et rencontres à venir seront comme des touches de couleurs et retouches successives d'un peintre sur sa toile. A tous, merci d'avance pour vos contributions permettant de parfaire le dessin initial.

Bonne lecture...

# SOMMAIRE

<b>INTRODUCTION GENERALE</b>	<b>9</b>
<b>CHAPITRE 1 LA SELECTION CHEZ LES BOVINS LAITIERS</b>	<b>13</b>
Partie 1.1 L'organisation de la sélection	15
Partie 1.2 Les évaluations génétiques	23
Partie 1.3 L'appui technique	31
<b>CHAPITRE 2 D'UNE EVALUATION GENETIQUE BASEE SUR LES DONNEES « LACTATION » AUX « CONTROLES ELEMENTAIRES »</b>	<b>39</b>
Partie 2.1 Modélisation de la courbe de lactation	45
Partie 2.2 Les évaluations génétiques sur les données lactation	63
Partie 2.3 Les évaluations génétiques de données longitudinales	77
Partie 2.4 Le modèle « contrôles élémentaires » pressenti en France	103
<b>CHAPITRE 3 LA MISE EN ŒUVRE OPERATIONNELLE</b>	<b>123</b>
Partie 3.1 Validation du logiciel d'évaluation génétique	125
Partie 3.2 Synthétiser l'information, une nécessité	135
Partie 3.3 Validation des effets fixes du modèle et réduction des besoins informatiques	155
<b>CHAPITRE 4 ANALYSE DES EFFETS FIXES DU MODELE</b>	<b>177</b>
Partie 4.1 Impact des effets d'environnement sur la production	179
Partie 4.2 L'effet troupeau-jour de contrôle	207
<b>CHAPITRE 5 DISCUSSION GENERALE</b>	<b>229</b>
<b>CONCLUSION GENERALE</b>	<b>243</b>
<b>ANNEXES</b>	<b>245</b>



## TABLE DES ABBREVIATIONS

ADN	Acide DésoxyriboNucléique
AI	Average Information
ANOVA	ANalysis Of VAriance
ARIMA	AutoRegressive Integrated Moving Average
ARSOE	Association Régionale de Services aux Organismes d'Elevage
BLUP	Best Linear Unbiased Prediction
CCS	Comptage de Cellules Somatiques
CD	Coefficient de Détermination
CF	Covariance Function
CIA	Centre d'Insémination Animale
CNAG	Commission Nationale d'Amélioration Génétique
CNIEL	Centre National Interprofessionnel de l'Economie Laitière
CP	Character Process
CRI	Centre Régional Informatique
CTIG	Centre de Traitement de l'Information Génétique
DCC	Days Carried Calf (Stade de gestation)
DF	Derivative Free
DIM	Days In Milk (Stade de lactation)
EM	Expectation Maximisation
FAO	Food and Agriculture Organization
FCL	France Contrôle Laitier
FEZ	Fédération Européenne de Zootechnie
FGE	France Génétique Elevage
FNCL	Fédération Nationale des Coopératives Laitières
FNIL	Fédération Nationale des Industries Laitières
FNPL	Fédération Nationale des Producteurs de Lait
FR	Fixed Regression
GLM	General Linear Model
HTD	Herd – Test Day
HY	Effet Troupeau – année (Herd-Year)
IA	Insémination Artificielle
ICAR	International Committee for Animal Recording
IDF	International Dairy Federation
INEL	INdex Economique Laitier
INRA	Institut National de la Recherche Agronomique
INSEE	Institut National de la Statistique et des Etudes Economiques
ISO	Organisation Internationale de Standardisation
ISU	Index Synthétique UPRA
ITEB	Institut Technique de l'Elevage Bovin
LOA	Loi d'Orientation Agricole
MACE	Multiple-trait Across Country Evaluation
MG	Matière Grasse
ML	Maximum Likelihood
MP	Matière Protéique
MT	Multiple Trait
OCL	Organisme de Contrôle Laitier
ONILAIT	Office National Interprofessionnel du LAIT et des produits laitiers
OS	Organisme de Sélection
PDIN	Protéines Digestibles dans l'Intestin permise par l'Azote
PDIE	Protéines Digestibles dans l'Intestin permise par l'Energie
QL	Quantité de Lait
REML	REstricted Maximum Likelihood
RR	Random Regression
SAS	Statistical Analysis System
SCS	Score de Cellules Somatiques
SIEL	Système d'Information des Elevages Laitiers
SIG	Système d'Information Génétique
SGQA	Station de Génétique Quantitative et Appliquée
TB	Taux Butyreux
TDM	Test Day Model
TIM	Test Interval Method
TP	Taux Protéique
UE	Union Européenne
UFL	Unité Fourragère Laitière
UMT	Unité Mixte Technologique
UPRa	Unité de Promotion de la Race

# Introduction générale

Utilisé dans toutes les civilisations et à toutes les époques, le lait est une matière première universelle qui bénéficie d'une image très positive auprès des consommateurs. Source essentielle de protéines et de calcium, il est considéré comme un aliment de base de l'alimentation humaine occidentale comme en témoigne la consommation moyenne française, équivalente pour l'année 2005 à 398 kg de lait entier (CNIEL (Centre National Interprofessionnel de l'Economie Laitière), 2007). Sa composition, exceptionnellement riche, explique les multiples usages du lait découverts par les hommes au fil du temps. Il est classiquement transformé sous forme de produits laitiers tels que les yaourts, le fromage, la crème, le beurre, les crèmes glacées... mais peut aussi contribuer à des usages moins connus tels que la fabrication de colles, de peintures, de matière plastiques, de produits cosmétiques...

La France est un pays où la filière laitière contribue de façon importante à l'activité économique nationale. En 2005, plus de 23.5 milliards de litres de lait ont été produits dans près de 110 000 fermes laitières en France soit 4.6% de la production mondiale de lait de vaches et 17.0% de la production européenne à 25 états (CNIEL, 2007). En 2002, le chiffre d'affaire de l'industrie de la transformation laitière française s'est élevé à 17.2 milliards d'euros, contribuant à la balance commerciale avec un solde positif de 2.6 milliards. Pour cette même année, l'industrie laitière a employé 60 000 personnes sur 700 sites. Plusieurs des entreprises françaises sont de dimension internationale, notamment Lactalis n°2, Danone n°3 et Bongrain n°13 mondial.

En 2005, l'Europe avec ses 25 Etats membres comptait plus de 1.7 millions d'exploitations agricoles, auxquelles il faut ajouter les 1.3 millions d'exploitations de la Roumanie et de la Bulgarie. Le nombre de vaches laitières pour ces 27 pays s'élève à 26.2 millions, cachant cependant de grandes disparités dans la taille des exploitations. Avec 75 à 80 vaches en moyenne, les exploitations danoises et du Royaume-Uni sont les plus conséquentes. La France et l'Allemagne sont dans une position médiane avec en moyenne 35 vaches par exploitation. Les derniers pays entrés dans l'Union Européenne (UE) (Pologne, Roumanie et Bulgarie) avec une économie encore très centrée sur le secteur agricole, comptent en moyenne moins de 5 vaches par exploitation.

En profonde évolution, la filière laitière européenne et mondiale doit faire face à de nouveaux enjeux. Dans les régions où la production laitière est historiquement bien implantée, les critères sanitaires et de qualité des produits ne suffisent plus à répondre aux attentes de l'opinion publique. Cette dernière attend désormais que les produits consommés aient été produits de façon respectueuse en tenant compte de l'environnement et du bien-être des animaux. Au niveau mondial, la demande en lait et produits laitiers est en forte croissance. D'après l'Organisation des Nations Unies pour l'Alimentation et l'Agriculture (FAO pour Food and Agriculture Organization) (Speedy, disponible sur <http://www.fao.org/AG/aGA/workshop/feed/papers/4andy.doc>) l'Asie et plus particulièrement la Chine qui consommaient jusqu'à maintenant peu de produits laitiers, montrent un engouement de plus en plus

important pour le lait, ce qui laisse présager une demande en croissance soutenue dans ces pays (15 à 20% par an).

Ces quelques données permettent de justifier l'intérêt porté par les différents acteurs de la filière pour les activités de recherche ayant trait à la maîtrise zootechnique des populations exploitées. Elles ont pour finalité la production des connaissances nécessaires à la création et au développement de nouveaux outils répondant au mieux aux contraintes imposées par le contexte socio-économique. Dans le cadre des activités de production laitière, une situation excédentaire née de la faible progression des débouchés est apparue il y a une vingtaine d'années. Elle a quelque peu modifié les objectifs de la filière en substituant à une démarche essentiellement productiviste, une volonté de maîtrise de la qualité du lait et de diminution des coûts de production, en même temps que s'opérait une lente remise en question de certains modes d'élevage. Ces préoccupations relativement récentes font écho en aval à une demande croissante de la part des industriels transformateurs d'un lait mieux adapté par sa composition en matières utiles (matières protéiques et matières grasses) aux procédés de fabrication et de stockage des fromages et autres produits laitiers. Ces produits de transformation (environ 25% de la consommation de lait en France) sont en augmentation continue aux dépens de la consommation de lait brut (Combris, 1997).

Pour l'éleveur qui cherche à augmenter le revenu de son activité ou du moins à le maintenir, deux leviers d'action s'offrent à lui : augmenter ses « produits » en fournissant une matière première – le lait – répondant par sa composition aux attentes des industriels ; et/ou diminuer ses « charges » en réduisant ses coûts de production. Cette 2<sup>ème</sup> alternative passe par une optimisation de la conduite du troupeau et notamment par une maîtrise de l'alimentation, principal poste de dépense, et la maîtrise des aspects sanitaires et de reproduction. L'appui technique dans ce domaine est principalement assuré par l'organisme de contrôle laitier dont l'objectif est d'aider l'éleveur à maximiser le bénéfice de son atelier.

Cette thèse, financée à travers une « action innovante » du ministère de l'Agriculture intitulée UTILEG (pour UTILisation des Evaluations Génétiques), déposée par l'Institut National de la Recherche Agronomique (INRA) et l'Institut de l'Élevage, avait initialement l'objectif d'étudier les possibles valorisations des résultats des évaluations génétiques des vaches laitières sur la production à chaque contrôle à des fins d'appui technique. Pour développer des outils de pilotage de l'atelier laitier, il nous semblait intéressant de s'appuyer sur les résultats de la nouvelle évaluation génétique des caractères laitiers basés sur les performances journalières, dont l'étude avait débuté. En effet, certains pays plus avancés sur cette thématique mettaient en avant des perspectives intéressantes (Mayeres *et al.*, 2004). Ce travail permettait également d'accroître l'intérêt de la base de données nationale en y adjoignant une nouvelle finalité grâce à la diffusion d'indicateurs utiles pour le conseil en élevage. A travers le développement d'une nouvelle génération d'outils d'appui technique permettant un pilotage plus précis, les éleveurs, principaux contributeurs de ce système, peuvent entrevoir une valorisation des informations dont ils ont permis l'acquisition.

Suite au retard pris dans le développement de l'évaluation génétique sur les données de production à chaque contrôle, l'objectif de cette thèse a quelque peu été modifié en y incluant une participation aux travaux nécessaires à la mise en place de cette évaluation génétique. L'étude des principales perspectives d'utilisation des résultats en appui technique a donc été traitée dans un second temps.

Cette thèse s'articule autour de cinq chapitres. Afin de bien resituer le contexte, le premier chapitre est consacré à de brefs rappels sur l'organisation de la sélection bovine en France, les évaluations génétiques qui sont réalisées et enfin les principales caractéristiques des outils techniques existant actuellement.

Le deuxième chapitre consiste en une revue bibliographique des éléments nécessaires à la compréhension des évaluations génétiques sur les contrôles élémentaires. Pour mieux en percevoir les implications, quatre parties ont été distinguées. La première présente les nombreux travaux qui depuis près d'un siècle ont contribué à la modélisation des courbes de lactation. Ces travaux n'avaient toutefois pas été intégrés dans les évaluations génétiques jusqu'à récemment. Les limites des modèles d'évaluation génétique sur les données lactation, exposées dans une deuxième partie, ont amené à mettre en œuvre une nouvelle méthodologie présentée dans la troisième partie, basée sur les données de production du jour de contrôle et permettant de valoriser au mieux l'ensemble des données recueillies. Enfin, une quatrième partie est consacrée à la présentation des travaux portant sur le modèle contrôles élémentaires français réalisés préalablement à cette thèse et à la description du modèle d'évaluation génétique tel qu'il était pressenti fin 2004.

Le troisième chapitre aborde les travaux préalables nécessaires à la mise en œuvre opérationnelle de l'évaluation génétique sur les données de production à chaque contrôle. Les modèles d'analyse de ces données s'appuient sur une méthodologie éprouvée, celle du BLUP (Best Linear Unbiased Prediction ou meilleur prédiction linéaire non biaisée). Toutefois, compte tenu des spécificités du modèle d'évaluation génétique français et du nombre de performances à traiter, il a été nécessaire de développer un logiciel adapté aux besoins identifiés. La première problématique à laquelle cette thèse s'est attachée, a consisté à développer une méthode permettant de s'assurer de l'exactitude des développements informatiques effectués. Ainsi, ce troisième chapitre est tout d'abord consacré à la présentation d'une approche basée sur l'utilisation de données simulées permettant de valider formellement le logiciel d'évaluation génétique utilisé. Les tests effectués à partir du logiciel ainsi validé ont mis en évidence des besoins en mémoire et de temps de calcul incompatibles avec les exigences des évaluations génétiques qui ont lieu périodiquement (3 fois par an) et avec des contraintes temporelles (1 semaine à 10 jours maximum pour l'ensemble des races et des caractères laitiers). Pour répondre à ces besoins, les deuxième et troisième parties ont porté respectivement sur l'étude des effets aléatoires et des effets fixes du modèle avec pour objectif la simplification du modèle d'évaluation initialement pressenti. L'enjeu principal de ces deux parties était de combiner le maintien de la description fine des effets sous forme de courbes (spécificité du modèle d'évaluation génétique français) avec la limitation des besoins informatiques.

Les contraintes techniques ayant été levées grâce au développement et à l'utilisation de méthodologies appropriées, nous nous sommes intéressés aux possibles valorisations des effets non génétiques du modèle dans le quatrième chapitre. Les effets d'environnement (rang de lactation, âge au vêlage, mois de vêlage, durée de tarissement précédente, gestation...) ont été étudiés. Ceci visait notamment à quantifier leur impact sur les courbes de lactation. Dans un second temps, les estimations des effets troupeau-jour de contrôle, considérées comme de bons indicateurs de la conduite du troupeau, ont été analysées. Différentes méthodes de prédiction de ces effets ont été proposées et comparées, et les applications potentielles des effets troupeau-jour de contrôle à des fins d'appui technique ont été mises en avant.

Enfin, le dernier chapitre consiste en une discussion générale. L'état d'avancement de l'évaluation génétique sur les contrôles élémentaires et les perspectives de valorisation à travers la prédiction de la production à l'échelle de l'animal y sont présentés.

## **Références**

CNIEL. 2007. L'économie laitière en chiffre. 217p.

Combris P. 1997. La consommation des produits animaux en France : tendances et perspectives d'évolution. *INRA Productions Animales*, 10 : 267-274.

Mayeres P., Stoll J., Bormann J., Reents R. and Gengler N. 2004. Prediction of daily milk, fat and protein production by a random regression test-day model. *Journal of Dairy Science*. 87: 1925-1933.

# **Chapitre 1**

## **La sélection chez les bovins laitiers**

## SOMMAIRE du Chapitre 1

<b>Partie 1.1</b>	<b>L'organisation de la sélection</b>	<b>15</b>
1.	De la loi sur l'élevage à la Loi d'Orientation Agricole	16
2.	Les acteurs et leurs contributions	17
2.1.	Les Etablissements Départementaux d'Elevage	17
2.2.	L'organisme de sélection	19
2.2.1.	Le contrôle de performance	19
2.2.2.	La filière Insémination Animale	21
2.3.	Le réseau national de collecte et de traitement de l'information	21
2.4.	La Station de Génétique Quantitative et Appliquée	21
2.5.	L'institut de l'Elevage	22
<b>Partie 1.2</b>	<b>Les évaluations génétiques</b>	<b>23</b>
1.	Les caractères d'intérêt	24
1.1.	Les caractères laitiers	24
1.2.	Les caractères fonctionnels	25
1.3.	Les caractères de synthèse	27
1.3.1.	L'Index Economique Laitier	27
1.3.2.	L'Index de Synthèse UPRa	27
2.	Les évaluations internationales	28
2.1.	Le développement des évaluations internationales	28
2.2.	Les bases méthodologiques	29
2.3.	Les enjeux	30
<b>Partie 1.3</b>	<b>L'appui technique</b>	<b>31</b>
1.	Outils d'appui technique au niveau individuel	33
1.1.	L'analyse <i>a posteriori</i>	33
1.2.	Aspects prospectifs	34
2.	Outils d'appui technique au niveau du troupeau	35
2.1.	L'analyse <i>a posteriori</i>	35
2.2.	Aspects prospectifs	36

## **Partie 1.1**

# **L'organisation de la sélection**



L'importance accordée à la sélection des bovins depuis leur domestication il y a près de 10 000 ans est largement expliquée par leur grande diversité de valorisation. Ils constituent non seulement une source importante de protéines animales à travers la production de viande et de lait mais également une source non négligeable d'énergie dans certaines régions du globe à travers l'utilisation comme animal de trait pour la réalisation de travaux agricoles. Longtemps abordée à travers les aspects phénotypiques (sélection massale), la sélection animale a profondément évolué depuis le milieu du XX<sup>ème</sup> siècle grâce au développement de la génétique quantitative. Les informations phénotypiques recueillies y sont désormais interprétées en tenant compte des relations de parentés existantes entre l'ensemble des animaux.

En pratique, cette révolution méthodologique a été permise grâce à la centralisation des informations de filiations et des performances recueillies, voulue par la loi sur l'Élevage de 1966. Auparavant, nombre d'outils utiles à la sélection animale tels que les livres généalogiques, le contrôle de performance, l'insémination artificielle existaient mais n'étaient pas structurés autour d'un dispositif commun. Le nombre limité d'éleveurs participants, les disparités géographiques et le manque de coordination entre les différents partenaires expliquait le faible impact de ces organismes sur la sélection animale jusqu'aux années 1960. Après avoir présenté la loi sur l'élevage de 1966 et son évolution en loi d'orientation agricole en 2006, cette partie sera consacrée à la présentation des différents acteurs du dispositif français d'amélioration génétique sur lesquels reposent la sélection génétique chez les bovins laitiers.

## **1. De la loi sur l'élevage à la Loi d'Orientation Agricole**

La loi sur l'élevage du 28 décembre 1966 a permis de structurer la sélection en adaptant ou créant des organismes répondant aux besoins collectifs. Cette politique nationale a notamment permis de faire progresser le niveau génétique du cheptel national en organisant la création du progrès génétique et en assurant sa diffusion sur tout le territoire ; d'améliorer le niveau sanitaire grâce à l'identification des animaux et à une garantie sur la qualité et la traçabilité de la semence ; et enfin, de guider les choix techniques des éleveurs par la mise en place de structures de développement et d'animation aux échelons national, racial et local.

Afin de faciliter ce développement, les organismes d'élevage ont bénéficié d'un environnement protégé leur permettant d'investir en toute sécurité dans des schémas à longue échéance à travers l'exclusivité territoriale offerte aux centres de mise en place de la semence ainsi qu'aux Etablissements Départementaux de l'Élevage (EDE) pour le contrôle des performances et l'identification des animaux et à travers l'unicité de représentation d'une race pour chaque Unité de Promotion de la Race (UPRa). Au cours des 40 années d'application de cette loi, le bilan qui s'en dégage est très positif. Toutefois, confrontée à une évolution du contexte législatif avec la primauté du droit communautaire sur le droit national, la France s'est vue dans l'obligation de modifier le cadre législatif national afin de se mettre en conformité avec la réglementation européenne. Les points d'achoppement portaient principalement sur l'exclusivité territoriale consentie aux centres de mise en place de la semence, contraire au principe de libre concurrence. Cette remise en cause de la loi de 1966 a

toutefois suscité de multiples interrogations de la part de l'ensemble des acteurs du dispositif génétique. Les risques de déclin de l'élevage dans certaines zones du territoire français pouvaient être redoutés en cas d'un désengagement des acteurs présents et l'absence de volontaires pour poursuivre ces activités économiquement moins rentables. Hors « le maintien de l'élevage est la condition *sine qua non* de la préservation de l'activité économique de nombreuses communes rurales et est donc devenu un enjeu territorial déterminant pour la cohésion spatiale du pays » (Extrait de l'intervention du représentant du ministère de l'Agriculture et de la Pêche à la journée de formation sur la Loi d'Orientation Agricole, 18 octobre 2006).

Dans un environnement plus libéral avec un encadrement étatique limité, la Loi d'Orientation Agricole (LOA), votée en début d'année 2006, a introduit d'importantes modifications dans le dispositif génétique français hérité de la Loi sur l'Élevage de 1966. L'objectif principal de la réforme était de confier le dispositif génétique à ses principaux acteurs, les éleveurs à travers la création d'une interprofession génétique (France Génétique Elevage ou FGE), tout en garantissant les intérêts de la société dans son ensemble sur le long terme, notamment en terme de maintien de la variabilité génétique. D'un point de vue pratique, cette réforme a permis de se mettre en conformité avec la réglementation européenne en supprimant le monopole de mise en place de la semence, mais également l'encadrement de l'activité d'insémination, permettant ainsi aux vétérinaires d'exercer cette activité. Elle a offert également une plus grande souplesse vis-à-vis de la demande d'autonomie des organisations professionnelles. Toutefois pour assurer un service d'une qualité équivalente à l'ensemble des acteurs, l'ouverture à la concurrence est contrôlée afin de préserver l'équilibre des territoires. Un service universel pour la distribution et la mise en place de la semence, pour l'enregistrement et la certification des parentés mais également pour le contrôle de performance a été créé et est actuellement en cours de mise en place. Cette réforme doit également permettre de faire évoluer les organismes de gestion génétique des races (UPRa, entreprise de sélection) sous la forme d'un interlocuteur unique rassemblant tous les acteurs de la race sur un plan national : l'Organisme de Sélection (OS). Le pilotage de ce dispositif reste assuré par l'Etat à travers la CNAG (Commission Nationale d'Amélioration Génétique), qui reste également garant de l'évaluation génétique, mission déléguée à l'INRA.

## **2. Les acteurs et leurs contributions**

Comme l'illustre la figure 1.1, les acteurs du dispositif génétique français sont nombreux. Chacun à son échelon contribue à l'amélioration génétique, véritable « œuvre collective » (Bougler, 1992).

### **2.1. Les Etablissements Départementaux d'Élevage**

Les EDE en rassemblant les divers interlocuteurs à l'échelon départemental ou régional, assure la coordination des opérations de base du dispositif génétique que sont l'identification, l'établissement des filiations et le contrôle de performances. Pour des raisons sanitaires, la mission d'identification concerne également les élevages dits hors base de sélection, c'est-à-

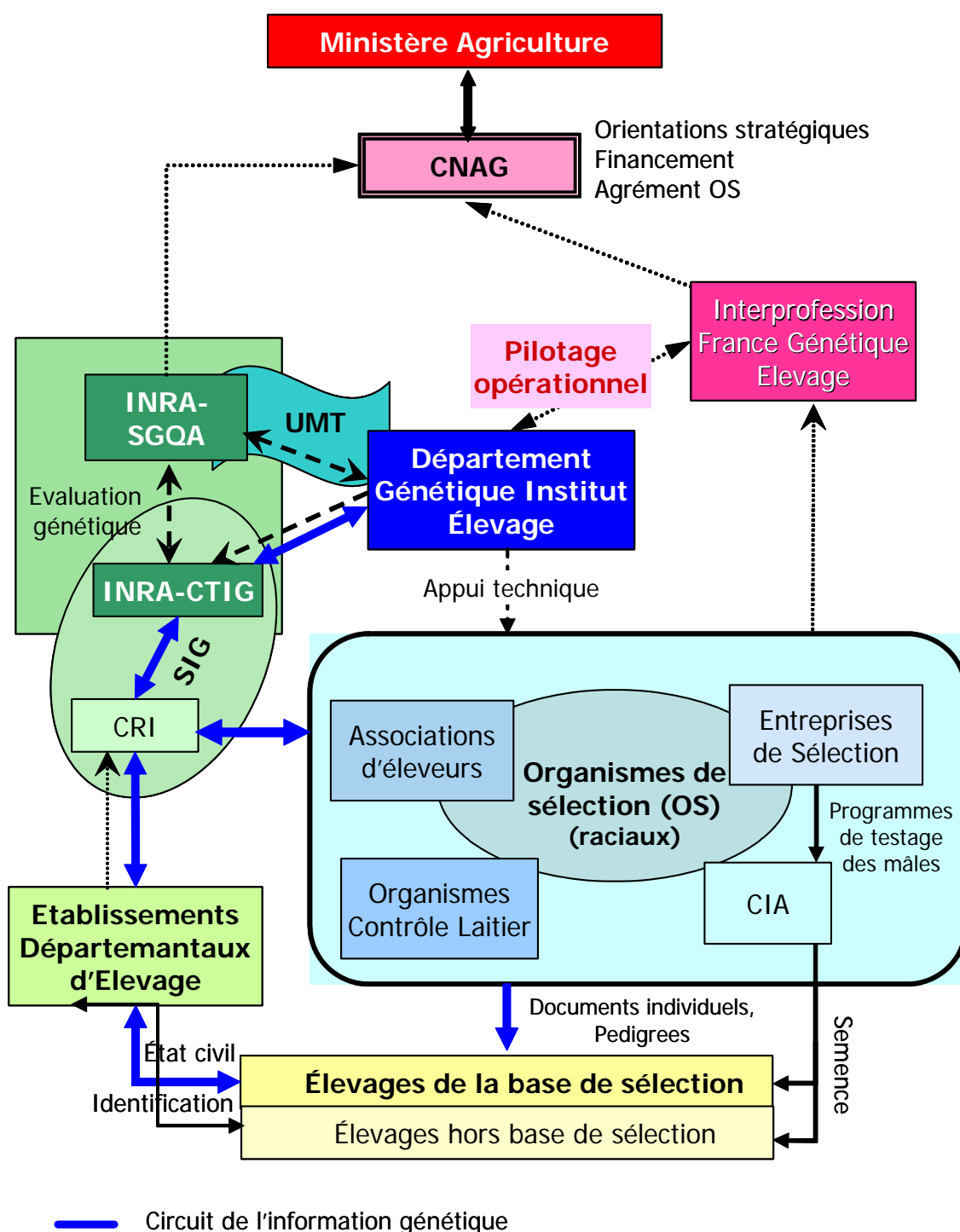


Figure 1.1. Schéma du dispositif français d'amélioration génétique suite à la loi d'orientation agricole du 5 janvier 2006. (CNAG : Commission Nationale d'Amélioration Génétique, ISO : Organisation Internationale de Standardisation, INRA : Institut National de la Recherche Agronomique, SGQA, Station de Génétique Quantitative et Appliquée, UMT : Unité Mixte Technologique, CTIG : Centre de Traitement de l'Information Génétique, SIG : Système d'Information Génétique, CRI : Centre Régional Informatique aussi appelé ARSOE (Association Régionale de Services aux Organismes d'Élevage), CIA : Centre d'Insémination Animale, OS : Organisme de Sélection)

dire ne contribuant pas au dispositif génétique. Concernant le contrôle de performance, cette activité est généralement confiée à un organisme spécialisé qui dans le cadre des bovins laitiers est appelé OCL pour Organisme de Contrôle Laitier. Le rôle des EDE ne se limite toutefois pas aux missions liées au dispositif génétique. Généralement associée aux Chambres d'Agriculture, elles contribuent au développement agricole en apportant du conseil aux éleveurs.

## **2.2. L'organisme de sélection**

L'organisme de sélection rassemble tous les acteurs de l'amélioration génétique d'une race sur le plan national, en vue de définir les caractéristiques de la race et certifier l'appartenance ou non d'un animal à cette race. Ses missions sont de définir les orientations de la race et d'en préciser les objectifs de sélection, de définir le programme d'amélioration génétique (création et diffusion du progrès génétique). L'organisme de sélection est également responsable de la tenue du livre généalogique de la race et délivre les documents officiels certifiant les informations relatives à chaque reproducteur.

L'OS regroupe donc différents organismes, notamment ceux responsables du contrôle de performance, qu'ils soient en lien avec les caractères laitiers ou de croissance dans les races allaitantes ainsi que les caractères de morphologie, mais également les entreprises conduisant les programmes de sélection des mâles reproducteurs et les Coopératives d'Insémination Animale (CIA) assurant la diffusion des semences sélectionnées.

### **2.2.1. Le contrôle de performance**

Le premier syndicat de contrôle laitier français a été mis en place dans le pays de Caux (Seine-Maritime) en 1907. Après avoir végété jusqu'au milieu des années soixante, période au cours de laquelle moins de 10% des 7 millions de vaches laitières étaient contrôlées, le contrôle laitier s'est rapidement étendu entre les années 1960 et 1990 (Figure 1.1), suite à la mise en place d'un cadre de développement propice avec la loi sur l'élevage. Aujourd'hui, le contrôle laitier français mesure mensuellement les performances de 2.5 millions de vaches dans un peu plus de 60 000 troupeaux, soit 66% des vaches laitières (cheptel de 3.8 millions de vaches laitières) et 86% du lait produit. Les organismes de contrôle laitier sont des entreprises autonomes, sous statut associatif, syndical, coopératif ou service de Chambre d'Agriculture, et sont directement gérés par des éleveurs élus par l'ensemble des adhérents. Les 81 organismes départementaux ou interdépartementaux sont structurés au niveau national par une Fédération, France Contrôle Laitier (FCL). Ces organismes mobilisent 5000 agents pour réaliser les opérations de collecte des informations nécessaires au dispositif génétique français et aux acteurs de la filière économique, et pour apporter aux éleveurs adhérents les conseils et services adaptés à leurs besoins.

Le cœur du métier du contrôle de performance chez les bovins laitiers est d'estimer la production des animaux à partir d'un sondage mensuel. Le jour du contrôle, la quantité de lait produite est mesurée pour chacune des traites contrôlées à partir de Tru-tests ou de compteurs à lait électronique homologué pour les exploitations qui en sont équipées. La richesse en matière grasse et protéique ainsi que les comptages de cellules somatiques sont obtenus à

partir d'un échantillon prélevé sur une fraction des laits prélevés au cours des traites contrôlées. Le laboratoire de référence réalise l'analyse de l'échantillon à partir d'appareils à dosage rapide par spectrophotométrie dans le moyen infra-rouge pour les teneurs en matière grasse et protéique, et une méthode fluoro-opto-électronique basée sur le comptage de noyaux cellulaires dont l'Acide DesoxyriboNucléique (ADN) est rendu fluorescent pour les cellules somatique (Leray, 1999). Pour estimer la production d'une vache, diverses méthodes d'échantillonnage (présentée dans la partie 2.2), variant principalement sur leur fréquence, ont été adoptées dans le monde entier. Le principe d'estimer la production à partir d'un sondage est bien accepté, car si nous souhaitions disposer de la production réelle, et non d'une production estimée, cela nécessiterait un relevé quotidien de la production, mais surtout l'analyse d'un échantillon de lait pour chacune des traites. De telles mesures seraient impraticables, extrêmement onéreuses et finalement d'une précision nettement supérieure aux besoins.

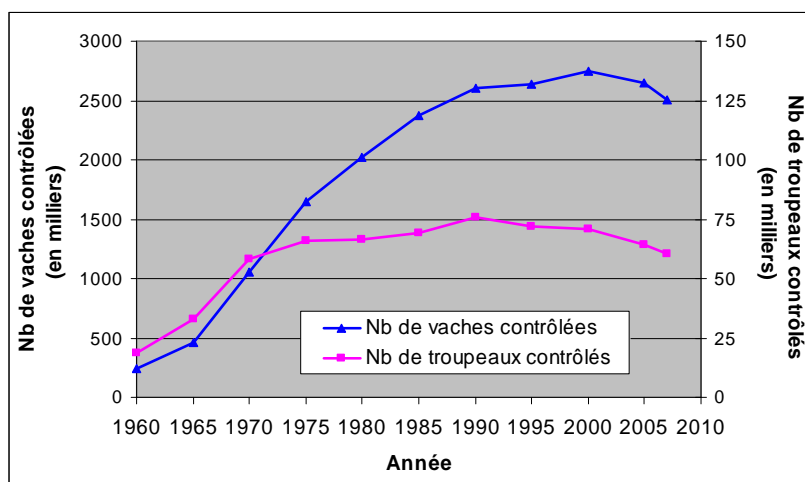


Figure 1.2. Evolution des effectifs de vaches contrôlées et de troupeaux adhérents au contrôle laitier entre 1960 et 2007

Outre la connaissance de la production laitière, le contrôleur laitier a un rôle d'appui technique auprès de l'éleveur. Dans ce cadre, les résultats du contrôle sont mis en relation avec l'alimentation du jour du contrôle. Il est alors possible d'effectuer un ajustement de la quantité et de la qualité de la ration afin de mieux s'adapter aux besoins et potentiels de production. Le rôle d'appui technique du contrôle laitier auprès des éleveurs sera développé ultérieurement (Partie 1.3).

Pour les caractères morphologiques, les animaux sont décrits à partir de notes définies d'après une table de pointage linéaire et des mensurations. Ces pointages, équivalents à des performances, sont recueillis d'une part par l'association raciale, qui hérite avec la LOA d'une partie de l'activité des anciennes UPRA, et d'autre part par les entreprises de sélection.

Le contrôle de performance est l'élément clé du dispositif d'amélioration génétique puisqu'il permet l'apport des données qui seront analysées. Son bon niveau de représentation de la population, avec 66% des vaches laitières françaises contrôlées, constitue une force pour l'amélioration génétique nationale.

### **2.2.2. La filière Insémination Animale**

Deux acteurs de la filière insémination oeuvrent dans le cadre du dispositif génétique. Tout d'abord les entreprises de sélection, qui contribue à la création du progrès génétique à travers la conduite de programme de sélection. Jusqu'à présent, de jeunes mâles sont testés et évalués par le biais d'un nombre limité de filles (dites de testage). Les meilleurs sont sélectionnés et agréés officiellement comme taureaux reproducteurs et sont diffusés à plus large échelle.

Les Centres d'Insémination Artificielle (CIA) sont les seconds contributeurs de cette filière. Ils assurent la diffusion du progrès génétique à travers l'insémination des femelles à partir de semences françaises ou étrangères de taureaux agréés au niveau national.

### **2.3. Le réseau national de collecte et de traitement de l'information**

Le dispositif génétique mis en place dans les années 1960 n'a pu atteindre l'envergure actuelle que grâce au développement d'un système de collecte et de traitement de l'information efficace, c'est-à-dire où les informations élaborées, utiles à l'éleveur pour la gestion quotidienne de son exploitation lui reviennent rapidement ; mais également où l'ensemble des informations de base convergent vers le Centre de Traitement de l'Information Génétique (CTIG) en vue de la réalisation de l'évaluation génétique.

Pour atteindre ce double objectif, le réseau national de collecte et de traitement de l'information a été structuré de la manière suivante :

1/ Au niveau départemental, les EDE sont chargés de collecter et de vérifier l'ensemble des informations recueillies dans les élevages et de les transmettre à l'échelon régional.

2/ Au niveau régional, des Centres Régionaux Informatiques (CRI) aussi appelés l'ARSOE (Association Régionale de Services aux Organismes d'Elevage) traitent les informations à des fins de gestions techniques pour permettre une valorisation rapide auprès des éleveurs et transmet les données de base vers la base de donnée nationale : le Système d'Information Génétique (SIG).

3/ Au niveau national, le CTIG rassemble toutes les données provenant des ARSOE (contrôle de performances et insémination artificielle) ainsi que des organismes en charge des pointages morphologiques et les gèrent à travers le SIG. Les résultats issus des évaluations génétiques suivent alors le chemin inverse.

### **2.4. La Station de Génétique Quantitative et Appliquée**

Outre le CTIG chargé de gérer la base de données génétique nationale, l'Etat a confié à l'INRA à travers la Station de Génétique Quantitative et Appliquée (SGQA) la responsabilité de réaliser les évaluations génétiques. Par ce contrôle, l'Etat s'assure de la fiabilité et de l'objectivité des informations mises à disposition des utilisateurs. Les travaux de recherche de l'équipe « bovins laitiers » de la SGQA sont largement conditionnés par cette responsabilité. En plus de la réalisation des évaluations génétiques trois fois par an selon un calendrier élaboré avec les professionnels, l'équipe a pour mission de contribuer au développement de méthodes et d'outils de gestion des populations notamment par ses travaux sur la définition des objectifs de sélection, la gestion de la variabilité génétique et l'optimisation des schémas

de sélection. Elle apporte également sa contribution à l'approfondissement des connaissances sur la variabilité génétique des caractères d'intérêt notamment à travers l'estimation de la variabilité polygénique, la détection de locus à effets quantitatifs et l'analyse des données génomiques à partir de l'exploitation des bases de données nationales ou expérimentales.

Afin de garantir la qualité de ses évaluations génétiques bovines auprès des utilisateurs, l'INRA s'est engagé dans une démarche qualité qui a abouti à la certification de l'ensemble de ses activités d'évaluation génétique depuis la gestion des bases de données par le CTIG jusqu'à la conception et la réalisation proprement dite des évaluations par la SGQA suivant une norme ISO 9001 en janvier 2006. Cette démarche a permis un fort rapprochement entre les équipes en charge des évaluations génétique au sein du CTIG et de la SGQA.

## **2.5. L'institut de l'Élevage**

L'Institut de l'Élevage est l'institut technique en charge des espèces herbivores. Il se situe à la charnière entre la recherche et le développement. Il conduit des travaux d'expérimentation, d'ingénierie, de coordination, de formation et de conseil afin d'améliorer la compétitivité des exploitations et les conditions d'exercice du métier d'éleveur, d'adapter la production et les systèmes d'élevage, de répondre aux demandes des filières sur la transformation des produits et les démarches « qualité ». L'Institut de l'Élevage est composé de 4 départements opérationnels : Identification animale et génétique ; Techniques d'élevage et qualité des produits ; Systèmes d'élevage et actions régionales ; Economie. L'action de l'Institut de l'Élevage s'inscrit dans une politique partenariale forte avec la recherche publique, en particulier l'INRA, les chambres d'agriculture, les organisations professionnelles et interprofessionnelles des filières herbivores. Avec la LOA, l'Institut de l'Élevage assure à travers son département génétique l'animation et la gestion administrative de l'interprofession FGE.

Suite à un appel d'offre du ministère de l'agriculture, une Unité Mixte Technologique (UMT) intitulée « évaluation génétique des bovins » regroupant une partie des équipes du département génétique de l'Institut de l'Élevage et des équipes de l'INRA-SGQA a été créée. Elle officialise et renforce la collaboration historique dans le domaine de l'évaluation génétique. L'UMT aide chacun des organismes à réaliser ses missions, confirmées par la LOA, dans les meilleures conditions de partenariat et de complémentarité. Elle répond à une complexification croissante des problématiques à traiter grâce à une planification concertée des travaux et en favorisant toutes les synergies entre les deux instituts.

A l'heure actuelle, l'amélioration génétique des bovins laitiers bénéficie donc d'un contexte favorable, largement hérité de la Loi sur l'Élevage de 1966, où l'intérêt collectif constituait le principal guide. L'investissement de l'Etat à travers le financement de ce dispositif a permis, même si ce soutien est désormais considérablement réduit, de combler le retard de l'élevage français des années 1960 et de lui permettre d'atteindre les premières places mondiales. Avec la LOA, le dispositif est désormais dans les mains de ses principaux acteurs : les éleveurs à travers les diverses organisation professionnelles.

## **Partie 1.2**

# **Les évaluations génétiques**



Le premier point de cette partie de ce chapitre sera consacré à une présentation générale des caractères sur lesquels porte la sélection chez les bovins laitiers. La sélection porte sur une combinaison de caractères dont la pondération varie suivant leur importance économique. La définition précise de l'objectif de sélection constitue le fondement de tout schéma de sélection.

L'étape suivante consiste à développer une évaluation génétique des reproducteurs permettant de choisir, sur la base d'un prédicteur de leur valeur génétique, les animaux les plus intéressants vis-à-vis de l'objectif précédemment défini. Les bases méthodologiques sur lesquelles repose toute évaluation génétique ainsi que les principes généraux et les propriétés des principales méthodes d'estimation de la variabilité génétique des caractères sont consultables en annexe.

Le second point développé dans cette partie sera consacré aux évaluations internationales, dont l'objectif premier est de permettre une comparaison équitable de l'ensemble des reproducteurs mâles des pays participants. Depuis le milieu des années 1990, ces dernières ont pris une place croissante dans les orientations et évolutions des évaluations génétiques nationales.

## **1. Les caractères d'intérêt**

Dans la filière bovine laitière, la principale source de revenus des éleveurs est la vente de lait ou de ses produits dérivés. Pour accroître son revenu, l'éleveur pouvait, jusqu'à la mise en place des quotas au milieu des années 1980, augmenter le nombre de vaches productives ou bien augmenter leur niveau de production. La maîtrise du niveau de production dépendait de l'alimentation sur le court terme et le niveau génétique sur le long terme. Avec l'évolution de ce contexte, l'éleveur a cherché à maximiser son revenu en optimisant les recettes liées à la commercialisation de ses produits tout en minimisant les charges liées à sa production. Dans ce contexte de production limité, l'éleveur recherche les animaux les plus rentables, c'est-à-dire des animaux ayant un niveau de production laitière élevé et présentant une qualité du lait, notamment des taux jugés suffisants, et ayant des coûts d'élevage limité. Pour répondre à ces nouvelles attentes, les caractères sur lesquels porte la sélection se sont multipliés depuis le milieu des années 1990, afin d'intégrer les caractères fonctionnels (Colleau et Régaldo, 2001).

### **1.1. Les caractères laitiers**

Les caractères laitiers, c'est-à-dire la quantité de lait et les taux butyreux (TB) et protéique (TP) (ou les matières grasses (MG) et protéique (MP)) sont la principale source de revenu de l'éleveur si les éventuels produits liés à la vente des animaux à des fins d'élevage ou de boucherie ne sont pas considérés. Le poids de lait est mesuré à travers le contrôle laitier généralement avec une périodicité mensuelle sur les traites d'une journée complète (24 heures). Les taux sont mesurés dans un laboratoire agréé grâce à l'échantillon prélevé lors du contrôle. L'évaluation génétique des caractères de production laitière est effectuée en France depuis 1964 (Colleau, 1996). Elle se base actuellement sur une production de lait estimée, exprimée suivant une durée de référence (les 305 premiers jours de lactation) en kilogramme. Les taux sont exprimés en ‰ soit des g/kg. Le taux protéique utilisé en France est le taux vrai,

inférieur de 5 % au taux azoté pourtant appelé ‘% protéine’ dans les publications étrangères (VanRaden, 2004). Les matières grasses et protéiques sont respectivement le produit de la quantité de lait et du taux butyreux et de la quantité de lait et du taux protéique, et sont elles aussi exprimées en unités de masse.

Les trois caractères de quantité (lait, MP et MG) sont génétiquement très fortement corrélés entre eux, avec une corrélation minimale de 0.85 entre le lait et la MG. Les taux sont aussi bien corrélés entre eux (corrélation de 0.60 entre le TB et le TP). Ils sont tous deux opposés à la quantité de lait (respectivement -0.30 et -0.45), et présentent donc des corrélations génétiques relativement faibles avec les quantités de matières. La MG et la MP étant obtenues grâce au produit de la quantité de lait avec le taux correspondant, seul 3 des 5 caractères laitiers évalués sont indépendants.

## 1.2. Les caractères fonctionnels

Les caractères fonctionnels (mammites, fertilité, longévité ...) influencent l'efficacité de la production non pas par un gain direct de productivité des animaux mais par une diminution des coûts d'élevage. Ces caractères sont généralement faiblement héréditaires et estimés avec un degré de précision faible. Ils présentent habituellement des corrélations génétiques négatives avec les caractères de production, ce qui complique fortement les possibilités d'amélioration génétique. Du fait de leur faible précision, les indices fonctionnels, à l'exception de l'évaluation sur la numération cellulaire et la morphologie, ne sont disponibles que sur les taureaux.

A travers les mesures de numérations cellulaires, l'objectif est d'améliorer la résistance aux mammites par une sélection indirecte sur ce critère, à défaut de disposer directement des informations sur les mammites cliniques. La mammite est une inflammation de la mamelle dont l'origine est infectieuse dans la majorité des cas. Elle est liée à la présence d'un pathogène, généralement une bactérie, dans la citerne de la mamelle qui entraîne une réponse immunitaire. La concentration du lait en cellules, principalement des leucocytes, est un témoin de l'état inflammatoire de la mamelle. Son incidence économique est double. D'une part, elle est prise en compte dans le système de paiement du lait et d'autre part elle est responsable de charges supplémentaires sur l'exploitation à travers le coût des traitements antibiotiques et les pertes financières liées aux réformes involontaires. Les données de numération cellulaire présentent l'avantage d'être disponibles pour la quasi-totalité des animaux puisque elles sont mesurées en routine à travers l'échantillon de lait réalisé par le contrôle laitier servant aussi à l'analyse des taux. Elles constituent pour le moment le seul critère de détection des mammites subcliniques disponible à grande échelle (Rupp et Boichard, 1997). Cette mesure constitue un outil indispensable à la maîtrise de l'état sanitaire des mamelles du troupeau. Les données, stockées dans le SIG, sont considérées être exhaustives en France depuis 1994 et disponibles partiellement depuis 1990. Même si les mammites cliniques et subcliniques constituent des pathologies distinctes, la corrélation génétique entre la numération cellulaire moyenne par lactation et l'incidence de mammites cliniques est élevée (corrélation de 0.7).

Parmi les autres caractères fonctionnels évalués en France, il y a la fertilité. L'objectif de cette évaluation est d'améliorer le niveau de fertilité des vaches laitières ou du moins de stopper la dégradation de la fertilité due à la sélection laitière, car ce caractère est génétiquement associé de façon défavorable à la production (Boichard, 1987). La fertilité représente un enjeu économique majeur pour les exploitations laitières de par les coûts directs comme les frais d'insémination, mais aussi indirects comme les coûts des réformes involontaires. La maximisation du taux de réussite à l'insémination constitue l'objectif principal (Boichard *et al.*, 1998). La réussite à l'IA est déterminée à partir des dates d'inséminations, disponibles dans le SIG pour les élevages adhérents au contrôle laitier depuis 1993 (Boichard *et al.*, 1998), de la date du vêlage et de la durée moyenne de gestation de la race. Malgré l'héritabilité très faible du caractère, la sélection de ce caractère reste intéressante car il existe une grande variabilité génétique. Sur des taureaux extrêmes, l'écart de réussite à l'insémination chez leurs filles peut atteindre 15%.

Autre caractère fonctionnel évalué, la longévité d'une vache laitière est généralement définie comme son aptitude à retarder la réforme involontaire c'est-à-dire pour des causes non liées à son niveau de production. Elle intègre l'influence des autres caractères fonctionnels (fertilité, résistances aux mammites et aux maladies, facilité de traite, qualité des aplombs...) et est mesurée comme le nombre de jours séparant le premier vêlage de la réforme (dernier contrôle connu). L'analyse statistique de la durée de vie productive est complexe (Ducrocq, 1997) car un animal en vie ne dispose que d'une borne inférieure (valeur actuelle) de la durée de vie productive vraie. Ce type de donnée, qualifiée de "censurée", nécessite le recours à des méthodes d'analyse de données de survie initialement développées dans le domaine biomédical. Le poids économique de la longévité est important, surtout en situation de quotas, car elle conditionne les coûts d'amortissement et de remplacement des animaux, ainsi que les possibilités de réforme volontaire sur la production laitière ou d'autres critères. Cette méthode fait maintenant référence et est utilisée dans de très nombreux pays (Interbull, 2008)

Les caractères liés à la mise bas tels que les facilités de naissance et de vêlage ainsi que la mortinatalité sont aussi des caractères fonctionnels. Les dystocies sont responsables chez les bovins de pertes économiques importantes du fait de l'augmentation de la mortalité, de la réduction de la fertilité et de surcoûts (vétérinaire). La stratégie consiste à maîtriser ces facteurs en exploitant les informations collectées par l'éleveur à travers les notes de conditions de naissance. L'objectif est d'identifier les taureaux extrêmes et ainsi d'éviter, surtout sur les génisses, les accouplements à risque à partir des conditions de naissance des veaux (effet direct) et des conditions de vêlage des filles des taureaux (effets direct et maternel). L'évaluation génétique est exprimée en pourcentage de vêlages attendus faciles. La mortinatalité dispose de nombreuses composantes communes avec les difficultés de naissance et de vêlage. En France, un veau est considéré mort-né dès lors qu'il ne survit pas au-delà de 2 jours. Cette mort peut être liée à la faible vitalité de l'animal, à un vêlage difficile ou bien à des tares génétiques associées à des déficiences physiologiques. La perte du veau a des conséquences économiques importantes, d'autant plus si elle touche un veau femelle car elle limite les possibilités de renouvellement de troupeau et entraîne une baisse du progrès génétique. L'objectif de cette évaluation est d'identifier les taureaux responsables d'une faible vitalité.

Enfin, les caractères de morphologie peuvent être classifiés parmi les caractères fonctionnels. Même si leur impact économique direct est limité au format de l'animal lors de la vente des vaches de réforme, ils ont une influence sur la facilité du travail, et sur des aspects liés aux risques sanitaires (mammites par exemple).

### **1.3. Les caractères de synthèse**

En France, il existe deux index de synthèse. L'index de synthèse laitière plus connu sous le nom de INEL (INdex Economique Laitier) et l'index de synthèse UPRa aussi appelé ISU. La dernière révision des pondérations économiques accordées à chaque caractère date de juin 2001 (Colleau et Régaldo, 2001).

#### **1.3.1. L'INdex Economique Laitier**

L'objectif de l'INEL était de maximiser la marge globale de l'exploitation, dans un contexte de quota laitier avec une référence sur le taux butyreux, tout en étant prospectif vis-à-vis du prix du lait et robuste face à la diversité des conditions technico-économiques des exploitations (Colleau et Régaldo, 2001). Après une période entre 1993 et juin 2001, où l'accent avait été mis sur la quantité et le taux de protéines suite à des promesses de la part des industriels laitiers de rémunération encourageant un fort taux protéique. En 2001, les conclusions des différents scénarios envisagés montraient que le paiement actuel du lait ne justifiait pas un coefficient aussi élevé du taux protéique. Faute de pouvoir prédire l'évolution du contexte des quotas, l'INEL a été redéfini de façon à maximiser la marge nette hors travail de l'atelier (coûts constants : contrainte de coûts opérationnels et de structure constante et en l'absence d'activités de substitution), après discussion avec les représentants de la filière 'lait' : FNPL (Fédération Nationale des Producteurs de Lait), FNCL (Fédération Nationale des Coopératives Laitières), FNIL (Fédération Nationale des Industries Laitières), ONILAIT (Office National Interprofessionnel du LAIT) comme  $INEL = MP + 0.2 MG + 1 TP + 0.5 TB$ , avec quelques variations sur les pondérations des taux suivant les races. Cette révision de L'INEL est plus favorable à la matière grasse et au taux butyreux. Nous observons une augmentation des réponses sur la matière grasse, une quasi-suppression des réponses négatives en TB et une réduction des réponses positives en TP (Colleau et Régaldo, 2001).

#### **1.3.2. L'Index de Synthèse UPRa**

L'index de Synthèse UPRa a pour objectif de synthétiser l'information donnée par les évaluations sur les différents caractères d'intérêt afin d'apprécier le plus justement possible la valeur génétique globale des animaux indexés et d'éviter une sélection par seuils, qui serait moins efficace. Lors de la mise en place d'un premier ISU défini autour de 1990, les pondérations économiques utilisées étaient basées sur la recherche d'un compromis entre les progrès génétiques attendus pour certains caractères (INEL, morphologie et vitesse de traite) et non sur un raisonnement économique. En 2000, pour favoriser l'utilisation des index fonctionnels disponibles depuis 1997-1998, l'inclusion explicite des index fonctionnels dans les objectifs de sélection et notamment à travers l'ISU a fait l'objet d'une étude. Les index

morphologiques (ou leur synthèse) présents dans le précédent ISU, même s'ils sont des prédictors partiels des caractères fonctionnels ne retranscrivaient pas l'ensemble des besoins. Les caractères actuellement inclus dans l'ISU sont l'INEL avec 50% du poids économique, la numération cellulaire, la fertilité, la longévité ainsi que la synthèse morphologie qui se répartissent de façon équitable les autres 50% (avec là encore quelques variations suivant les races). L'ISU est calculé suivant une méthodologie appropriée (Ducrocq *et al.*, 2001) permettant de tenir compte des corrélations entre caractères et dont les propriétés statistiques ont été validées par simulation (Lassen *et al.*, 2007). Les caractères fonctionnels se caractérisant par une faible héritabilité et une expression souvent tardive handicapant la sélection, des caractères prédictors, notamment morphologiques, sont utilisés pour enrichir leur information. Un autre intérêt de cette méthodologie est de permettre le calcul de la synthèse sur l'ensemble de la population, mâles et femelles, même si pour certains caractères fonctionnels, les valeurs génétiques ne sont actuellement estimées que chez les mâles.

## **2. Les évaluations internationales**

Avec l'émergence de l'insémination artificielle, le développement des méthodes de cryoconservation de la semence et les progrès du transfert embryonnaire, les programmes de sélection des bovins laitiers ne se raisonnent plus à l'échelle nationale. Les échanges de matériels génétiques : animaux, embryons et semences sont importants. L'amélioration génétique des bovins laitiers doit donc se réfléchir sur le plan international.

De nombreux de pays estiment la valeur génétique des taureaux et même des vaches à l'échelle nationale. Toutefois, pendant de nombreuses années, il n'était pas possible de comparer de manière précise les différentes valeurs génétiques car chaque pays dispose de sa propre façon de prédire et d'exprimer la valeur génétique, la base de référence, l'échelle, l'unité d'expression et la modélisation pouvant être différentes (Interbull, 2000).

### **2.1. Le développement des évaluations internationales**

Face à ces nombreuses difficultés, la Fédération Européenne de Zootechnie (FEZ) a créé un groupe de travail en 1975 chargé de réfléchir à la faisabilité de procédures de prédiction de valeurs génétiques homogénéisées entre les pays. A la fin des années 1970, la Fédération Internationale de Bovins Laitiers (IDF pour International Dairy Federation) exprimait les mêmes attentes. Au début des années 1980, la FAO a cherché à comparer le niveau de 10 populations de vaches Pie-Noire en les plaçant dans un même troupeau en Pologne (Stolzmann *et al.*, 1981). Cette méthode de comparaison était toutefois très onéreuse aussi bien du point de vue financier que du temps consacré et ne concernait qu'un nombre très limité de taureaux, pas forcément représentatifs des taureaux de chaque pays

Les besoins de comparaison internationale ont mené à la création de l'International Bull Service (Interbull) en 1983 par la FEZ (Fédération Européenne de Zootechnie), l'IDF et le Comité International pour le Contrôle des Performances en Élevage (ICAR pour International Committee for Animal Recording), sous la supervision de la FAO. Initialement, Interbull avait comme objectifs de permettre l'échange d'informations sur les méthodes d'évaluations génétiques des bovins laitiers dans les différents pays et de fournir des recommandations

quant à l'utilisation et l'interprétation des évaluations dans les autres pays (Philipsson, 1998). Chaque pays membre avait le devoir d'expliquer son modèle et les méthodes d'évaluation génétique retenues, Interbull retransmettant l'ensemble des informations obtenues à partir de formules de conversion entre chacun des pays pris 2 à 2 dès 1985. Ces indications ont permis de réduire les tensions politiques et commerciales dues à l'existence de différentes méthodes de comparaison. En 1988, Interbull est reconnu comme sous-comité permanent d'ICAR. L'importance et les enjeux de ce travail ont nécessité alors la création d'une unité opérationnelle qui a ouvert ses portes à Uppsala (Suède) en 1991.

En Août 1994 a lieu la première évaluation internationale entre les 4 pays nordiques en utilisant la méthode d'évaluation multicaractères entre pays (MACE pour Multiple-trait Across Country Evaluation). Dès février 1995, 10 pays d'Amérique du Nord et d'Europe participent à la seconde évaluation. A l'heure actuelle, ce sont 27 pays de tous les continents, à l'exception de l'Amérique du Sud, qui participent aux évaluations internationales, dans 6 races bovines et 6 groupes de caractères : production, morphologie, santé (numération cellulaire et mammites cliniques), longévité, naissance (facilités de vêlage et de naissance et mortinatalité) ainsi que fertilité. La mise en place en routine de l'évaluation des caractères associé à la facilité de travail (vitesse de traite et tempérament) est quant à elle prévue pour 2008/2009.

## 2.2. Les bases méthodologiques

Après l'utilisation de formules de conversions basées sur le principe de régression de 1985 à 1994, les comparaisons reposent aujourd'hui sur la méthode MACE (Schaeffer, 1994) dans laquelle tous les taureaux de l'ensemble des pays participants sont analysés simultanément. Le MACE est une évaluation multicaractères dans laquelle un caractère analysé est considéré être différent dans chacun des pays mais génétiquement corrélée.

Cette méthode présente l'avantage de prendre en compte les variances génétiques et résiduelles, et donc l'héritabilité, spécifiques de chacun des pays. De cette façon, les corrélations génétiques entre pays peuvent être inférieure à 1.00 et prendre en compte l'existence de différences entre pays du fait des interactions entre le génotype et l'environnement ou bien de disparités dans la définition du caractère.

Le modèle tient compte de l'ensemble des parentés entre mâles de tous les pays. Ces liens génétiques entre les pays via les parentés entre animaux contribuent à créer des connexions. Les importations et exportations de matériel génétique permettent des connexions directes entre pays alors que les liens de parenté des animaux importés et exportés permettent des connexions indirectes.

En sus des nombreuses vérifications imposées par Interbull avant l'intégration des données d'un pays dans les évaluations internationales, le progrès génétique de chaque caractère ou groupe de caractère doit faire l'objet d'une étape de validation à l'aide d'une des méthodes « agréées » (Boichard *et al.*, 1995). En effet, la valeur du progrès génétique a un impact important sur le résultat des évaluations internationales. Dans les évaluations nationales, en cas de surestimation du progrès génétique, le classement des animaux est correct mais les jeunes animaux sont surestimés (ou les anciens sous-estimés). Par contre, dans les évaluations

internationales, les taureaux anciens étant généralement les taureaux communs, c'est-à-dire ceux sur lesquels s'établissent les éléments de comparaison, les jeunes taureaux d'un pays surestimant son progrès génétique sont abusivement surclassés.

### **2.3. Les enjeux**

Les évaluations internationales constituent un enjeu important pour tous les pays, qu'ils soient importateurs ou exportateurs de semences, même si l'enjeu économique direct se concentre sur les pays exportateurs. En effet, Interbull, en permettant une comparaison équitable dans l'échelle de chacun des pays participants, permet à chaque entreprise de sélection de commercialiser le matériel génétique dont elle est propriétaire de manière objective. Au niveau national, l'importation de matériel génétique permet d'une part d'accroître le progrès génétique en utilisant de façon préférentielle les animaux présentant le meilleur potentiel et d'autre part de diversifier les origines génétiques et ainsi de limiter l'accroissement de la consanguinité dans les populations nationales. Les échanges de matériels génétiques sont donc primordiaux.

Cette concurrence entre pays est importante, notamment en race Holstein, mais elle est aussi stimulante. En effet, il existe une certaine émulation entre les différents pays, poussant chacun d'entre eux à s'afficher comme des leaders et forces de propositions au niveau international. Cela joue aussi bien sur les aspects méthodologiques des évaluations internationales que sur les caractères évalués. Ainsi les pays Nordiques, fortement en avance sur les aspects santé des animaux en évaluant de longue date les mammites cliniques, constituent une force motrice pour le développement d'une évaluation génétique sur ce caractère dans d'autres pays. Des pays comme le Canada et les Pays-Bas ont eux été leaders pour le développement d'évaluation génétique de la production laitière basée sur les contrôles élémentaires.

Pour maximiser les chances d'être bien placés dans les palmarès internationaux, les pays cherchent à homogénéiser la définition et la modélisation de leurs caractères. Ceci permet généralement de limiter les différences entre pays, qui se traduisent par une corrélation génétique d'autant plus faible que les différences sont importantes.

## **Partie 1.3**

### **L'appui technique**



Comme nous avons pu le voir dans la partie 1.1, le contrôle laitier a une mission d'appui technico-économique auprès des éleveurs, en plus de sa participation à la mission collective d'amélioration génétique du cheptel français à travers l'enregistrement de la quantité et de la qualité du lait produit par chacune des vaches des troupeaux adhérents. L'appui technique peut concerner de nombreux axes en lien avec les bovins laitiers (production laitière, alimentation, production fourragère, reproduction, accouplements, ...). Toutefois, notre travail étant axé sur les caractères de production laitière, seules les analyses découlant directement de ce domaine vont être présentées.

L'appui technique est une réponse de l'OCL à une demande de l'éleveur qui adhère à ce service en complément du service de base (mesure de la quantité et analyse de la qualité du lait produit par chacune des vaches). Le regard extérieur du technicien ou conseiller de l'OCL permet de réaliser une « photographie » technico-économique de l'élevage à partir des informations disponibles. Ce dernier point n'est généralement pas limitant car l'éleveur ayant fait la démarche de payer un service d'appui technique, est généralement disposé à donner le maximum d'informations susceptibles d'aider l'agent du contrôle laitier à affiner ses conseils. Ainsi, des informations telles que les livraisons et la rémunération du lait, les coûts des concentrés et autres aliments achetés sont utilisées dans les logiciels d'appui technique existants. Les réponses apportées par les techniciens et conseillers de l'OCL se doivent de tenir compte des objectifs de l'éleveur mais également de ses contraintes.

Plusieurs logiciels d'appui technique coexistent sur le territoire français, faute d'avoir réussi à construire un logiciel commun aux différentes régions. L'objectif de ces logiciels est d'aider l'agent du contrôle laitier dans son rôle d'analyse et de conseil en proposant différentes valorisations possibles des données enregistrées à partir de l'ensemble des informations recueillies dans l'élevage. A défaut de disposer d'éléments descriptifs pour l'ensemble de ces logiciels, nous ne présenterons dans cette partie que les principaux types de valorisation, appelées « fonctions » décrites dans les documents explicatifs du logiciel SIEL acronyme de Système d'Information des Elevages Laitiers (FCL, 2005a et 2005b), dont la mise en place est récente (2005-2007). Plusieurs des fonctions du logiciel SIEL sont basées sur l'utilisation de courbes théoriques de lactation dites « courbes-types » provenant du référentiel intitulé « ARCOLA » développé par FCL et l'Institut de l'Élevage au début des années 1990 (Bouloc, 1993). Les courbes-types intègrent quelques facteurs simples de variation de la forme de la courbe de lactation tels que le rang de lactation, le mois de vêlage et le niveau de production. Elles sont adaptées au contexte de chacune des régions utilisatrices grâce à différents jeux de coefficients. Deux applications principales du référentiel ARCOLA peuvent être mises en avant. La première (application ❶) permet de calculer, à partir de la production du jour du contrôle obtenue pour un stade de lactation donné, du niveau de production exprimé (production cumulée qu'aurait une vache au cours de sa lactation en 305 jours si sa production suivait strictement la courbe théorique passant par la production observée lors du dernier contrôle connu – voir Figure 1.3). La seconde (application ❷) prédit la production d'une vache pour un stade de lactation donné à partir de son niveau de lactation en cours (ou celui de sa lactation précédente).

Dans notre présentation, nous distinguerons deux niveaux d'appui technique : d'une part, un niveau individuel basé sur l'analyse de chacune des vaches de l'élevage, et d'autre part, un

niveau plus global basé sur l'analyse du troupeau, qui amène généralement à une comparaison par rapport à un réseau de référence (moyenne des troupeaux ayant un système de production similaire). Pour chacun des niveaux, nous distinguerons l'analyse des résultats passés (le « réalisé ») et la prédiction de la production attendue à court et moyen terme.

## 1. Outils d'appui technique au niveau individuel

### 1.1. L'analyse *a posteriori*

- ✓ La fonction intitulée « historique » dans SIEL permet d'analyser la production réalisée par l'animal aussi bien en termes de quantité que de qualité (TB, TP et cellules leucocytaires) à partir des données des 18 derniers contrôles. Cette fonction permet d'identifier des variations de production à court, moyen et long terme, notamment les conséquences de problèmes sanitaires tels que récemment ceux liés à la diffusion de la fièvre catarrhale ovine.
- ✓ La fonction « analyse des chutes » permet de détecter des variations anormales à court terme. SIEL calcule le pourcentage de variation de la production entre deux contrôles consécutifs pour tout contrôle réalisé avant 240 jours de lactation. Exprimée pour une durée standard de 1 mois, une chute de plus de 15% (seuil paramétrable) fait l'objet d'une alerte.
- ✓ La fonction « suivi de production laitière » est une autre approche pour détecter des variations anormales. Elle permet de comparer pour chacune des vaches la production réalisée par rapport à la production attendue, simulée *a posteriori* d'après leur niveau de production en cours et de leur stade de lactation (application ②).
- ✓ La fonction « vaches à problèmes » de SIEL permet de mettre en évidence dans le cas de vaches multipares, des animaux présentant un pic de lactation estimé à 30 jours de lactation jugé insuffisant, ainsi que les vaches dont la production a chuté. Le niveau de cette insuffisance est paramétrable. Le pic de production objectif est considéré comme atteint dès lors que le pic estimé de la lactation d'après les productions mesurées au cours de 70 premiers jours correspond au minimum à 90 % du pic attendu.

En ce qui concerne l'aspect sanitaire de la mamelle, l'agent de l'OCL effectue un suivi des statuts sanitaires à partir de la fonction « historique » ou la fonction « vaches à problèmes ». Pour qu'une vache soit considérée comme « saine », il faut que tous ses contrôles aient des comptages cellulaires inférieurs à 300 000 cellules par mL. Son statut sera « douteux » dès lors qu'un des comptages dépasse le seuil de 300 000 cellules et elle sera considérée comme « infectée » lorsque deux des contrôles sont supérieurs à 800 000 cellules (classification adaptée d'après Serieys, 1985). Le retour à un statut « sain » nécessite un minimum de 6 contrôles successifs avec moins de 300 000 cellules par mL.

Concernant la production cumulée depuis le vêlage (ou les taux moyens), ils sont calculés selon la méthode Fleischmann, considérée être la méthode de référence. Elle est basée sur l'interpolation de la production entre deux dates de contrôles (voir chapitre 2 pour plus de détails). Cette méthode simple donne une très bonne approximation de la production réelle et ne présente pas de biais (Schaeffer and Jamrozik, 1996).

## 1.2. Aspects prospectifs

✓ La fonction « simulation de production laitière » du logiciel SIEL permet de prédire la production laitière d'une vache sur la base des données de lactations antérieures lorsqu'il y en a, mais également sur les premiers contrôles de la lactation en cours si celle-ci a débuté. Dans ce dernier cas, SIEL prédit la production 305 jours probable d'une part à partir du cumul de production réalisé depuis le début de la lactation obtenu selon la méthode Fleischmann, et d'autre part à partir du cumul prévisible obtenu selon des courbes de lactation théoriques entre le dernier contrôle et le stade de 305 jours. En tenant compte de paramètres tels que le rang de lactation, le mois de vêlage, l'âge au vêlage (pour les primipares) ou l'intervalle entre vêlage, les courbes-types sont adaptées aux situations les plus communes.

La fonction « simulation de production laitière » s'appuie sur l'application ❶ du référentiel ARCOLA qui permet d'établir un niveau de production exprimée sur la durée de référence de la lactation (305 jours), corrigée pour des effets tels que le rang de vêlage, le mois de vêlage, l'âge ou l'intervalle entre vêlages, à partir de la production réalisée lors d'un contrôle et du stade de lactation correspondant.

✓ Cette application est également utilisée par la fonction « analyse des niveaux » qui permet de comparer le niveau de production de vaches à des stades de lactation différents. Par exemple, bien que la production laitière au contrôle de la vache B soit inférieure à celle de la vache A (Figure 1.3), le niveau exprimé de la vache B est supérieur (8000 kg contre 7000 kg) à cause d'un stade de lactation plus avancé. Cette fonction permet également de comparer le niveau de production obtenu pour une vache lors de contrôles successifs.

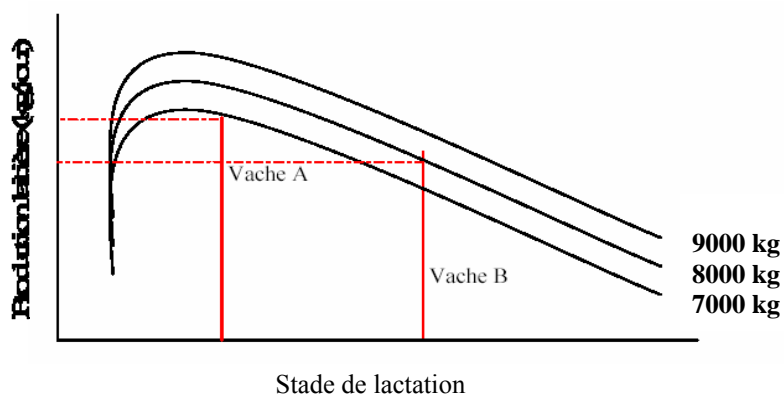


Figure 1.3. Comparaison du niveau de production exprimé en 305 jours en fonction de courbes-types de lactation et du stade de lactation (source : FCL, 2005a)

A partir des niveaux de production obtenus d'après les courbes-types, SIEL établit pour les multipares un pic objectif à partir du pic correspondant au 2<sup>ème</sup> meilleur niveau de production obtenu pour la lactation précédente sur la base des quantités de matières grasse et protéique produite, en considérant que le meilleur n'est pas forcément représentatif du potentiel de l'animal. Une moyenne des trois meilleurs contrôles aurait également pu être utilisée. Ce pic est standardisé, c'est-à-dire exprimé en équivalent adulte d'après des coefficients similaires à ceux de l'évaluation génétique sur les données lactation (cf partie 2.2), pour un vêlage en octobre et un intervalle vêlage-vêlage compris entre 390 et 419 jours. L'information relative au pic objectif est ensuite utilisée par d'autres fonctions pour déterminer si le pic est atteint ou non.

## 2. Outils d'appui technique au niveau du troupeau

### 2.1. L'analyse *a posteriori*

En regroupant les analyses effectuées au niveau individuel, il est possible de faire ressortir les principales tendances du troupeau.

✓ La fonction « vaches à problèmes » permet à SIEL d'éditer la liste des animaux à problèmes (chutes de production, pic faible, mammites, problème de fertilité).

✓ La fonction « analyse des chutes » permet d'établir des fréquences de chutes de production par rang de lactation et niveau de production afin de repérer plus facilement des conditions défavorables. Le pourcentage de chute, c'est-à-dire le nombre de vaches ayant chuté par rapport au nombre de vaches qui auraient pu chuter, est également calculé pour mettre en évidence ou relativiser leur importance dans le troupeau.

✓ La fonction « valorisation du contrôle » synthétise de nombreuses informations déterminées à l'aide d'autres fonctions. Elle détermine par exemple un niveau d'étable à partir de la moyenne des niveaux adultes obtenus pour le mois de contrôle considéré pour chacune des vaches présentes dans le troupeau. Son évolution par rapport au contrôle précédent est un bon indicateur de l'évolution de la conduite du troupeau. De façon similaire, le mois moyen de lactation est déterminé en divisant la somme des stades de lactation de chacune des vaches par le nombre de vaches du troupeau et la durée moyenne d'un mois de contrôle. La tendance qui s'en dégage permet de prévoir grossièrement l'évolution de la quantité de lait produite, de la numération cellulaire...

Cette fonction reprend également les principaux résultats d'un module de SIEL consacré à l'alimentation. En effet, ce domaine constitue un des axes majeurs d'intervention des agents du contrôle laitier. Si l'alimentation correspondant à la période de contrôle a fait l'objet d'un enregistrement, il est possible de calculer l'ensemble des critères décrivant la ration en terme quantitatif et qualitatif tels que les Unités Fourragères Laitière (UFL), les Protéines Digestibles dans l'Intestin permises par l'Azote (PDIN) et les Protéines Digestibles dans l'Intestin permises par l'Energie (PDIE). A partir de ces données, SIEL calcule le total des UFL concentrés distribués par vache. Sur la base d'une valorisation de l'énergie de 0.44 UFL sous forme de 1kg de lait, il calcule le lait par vache permis par les concentrés, qui une fois soustrait au lait total produit par vache permet d'obtenir le lait théoriquement produit par l'énergie des fourrages. Ces éléments, ainsi que les coûts relatifs à chacun des types d'aliment permettent d'en déduire 2 critères relatifs à l'efficacité du troupeau : d'une part la quantité moyenne de concentré utilisé pour produire 1 kg de lait et d'autre part le coût alimentaire pour produire 1000 kg de lait. Ces critères peuvent facilement être exploités pour jugés de l'efficacité économique de l'atelier laitier, par exemple sur le 1<sup>ère</sup> critère, une valeur proche de zéro est courante dans les troupeaux où l'alimentation est basée sur l'herbe alors que la valeur peut atteindre 0.5 kg dans des systèmes de type ration sèche où le concentré est primordial.

✓ La fonction « suivi de la production laitière » permet en comparant les productions simulées à partir d'un objectif défini en accord avec l'éleveur et les productions réalisées de réaliser un suivi technique de la production laitière du troupeau. Les différences observées peuvent également être interprétées comme des effets de la maîtrise (ou plus souvent d'une non maîtrise) de l'environnement de production et notamment des facteurs liés à

l'alimentation ou au suivi sanitaire et permettent aux techniciens de guider l'éleveur dans les orientations à prendre pour optimiser la conduite de l'élevage.

✓ La fonction « prévision quota » permet dans le contexte actuel de limitation de la production de réaliser un suivi quantitatif de la production réalisée par rapport aux références laitières (quotas) du troupeau afin d'ajuster le mieux possible les livraisons de la campagne laitière.

Comme au niveau individuel, le suivi de la qualité du lait du troupeau à travers le niveau leucocytaire est une tâche essentielle dans un élevage. Les fonctions « historique troupeau » ou « vaches à problèmes » facilitent cette analyse. Le niveau leucocytaire, qui constitue une image de la gestion du troupeau, du matériel de traite et de l'environnement de l'animal (bâtiment et salle de traite) a un impact économique important sur le paiement du lait. Réglementairement, le lait doit contenir moins de 300 000 cellules par mL, mais dès lors que les livraisons dépassent 400 000 cellules par mL, 15€ de pénalités sont en moyenne prélevés par les industries laitières collectrices pour 1000L de lait. Au contraire, un lait de très bonne qualité (moins de 200 000 cellules par mL) sera (en moyenne) payé 6€ de plus pour 1000L. Par conséquent pour une exploitation ayant un quota annuel de 300 000 L de lait, l'incidence des cellules sur le paiement du lait peut rapidement atteindre 6000€. Le taux cellulaire du troupeau doit donc être maîtrisé en ayant recours, si besoin est, à la réforme des animaux responsables des plus mauvais résultats.

## 2.2. Aspects prospectifs

✓ Comme dans le cas de prédiction individuelle, la fonction « simulation de la production laitière » permet de simuler la production laitière ou les taux mais cette fois-ci à l'échelle du troupeau pour la campagne en cours par exemple. Cette prédiction se base sur les événements réels enregistrés pour ce qui est déjà réalisé et sur les événements prévisibles compte tenu des dernières informations recueillies pour la fraction de la campagne restant à effectuer.

✓ La fonction « prévision quota » se base sur les résultats de la fonction « simulation de la production laitière » pour étudier la gestion du quota sur la campagne laitière. Par exemple, une simulation réalisée le 3 novembre se basera sur le lait livré à la laiterie entre le 1<sup>er</sup> avril et cette date et sur les prédictions de production entre le 4 novembre et le 31 mars pour les vaches actuellement en lactation, ainsi que les vaches qui débuteraient une lactation au cours de cette période. La qualité de la prédiction dépend étroitement de la qualité des connaissances des entrées et sorties du troupeau qui sont prévues jusqu'à la fin de la campagne laitière. Il est donc essentiel que le travail de prévision soit fait en collaboration avec l'éleveur de façon à identifier les animaux dont la réforme est prévue, mais également la date probable de cette réforme. Deux niveaux de production peuvent être utilisés pour cette prédiction : un niveau de production optimal ou bien un niveau de production probable. Il est ainsi possible de déterminer un niveau de production attendu mais également une fourchette haute qui peut être atteinte en cas de besoin, grâce à la maîtrise des conditions de production. Une autre variable d'ajustement possible de la production à la disposition de l'éleveur pour diminuer ou augmenter la quantité de lait produite est la durée de tarissement.

Pour le correctif de production lié à la matière grasse pour les quotas, SIEL propose trois alternatives aux éleveurs : des prédictions basées sur des données simulées compte tenu des TB actuellement mesurés ; l'utilisation des TB laiterie obtenus pour les mois correspondants l'année précédente ; le recours à un TB moyen pour le troupeau sur les mois à venir. SIEL permet également de tenir compte des quantités de lait qui seraient destinées à l'élevage des veaux, autoconsommées par l'éleveur et sa famille ou commercialisées en vente directe.

Le logiciel d'appui technique SIEL, récemment développé, dépend largement du référentiel de courbes de lactation ARCOLA, estimé au début des années 1990 (Bouloc, 1993) désormais vieillissant. Ce référentiel, qui présentait lors de sa mise en place l'intérêt de prendre en compte le niveau de production de l'animal, son âge au vêlage (primipare) ou l'intervalle entre vêlages, son rang de lactation et son mois de vêlage à l'aide de corrections multiplicatives ou additives, est aujourd'hui technologiquement dépassé. Seuls les facteurs de variation simples de la forme de la courbe de lactation sont pris en compte (rang de lactation, mois de vêlage et niveau de production). Les effets troupeaux ne sont donc pris que très partiellement en compte à travers le niveau de production de l'animal, et les effets race et région de production en sont complètement absents. Pour palier à certains de ces manques, les courbes-types sur lesquelles se basent SIEL ont fait l'objet d'ajustements pour tenir compte des particularités locales (pratiques d'alimentation, conditions pédo-climatiques, système fourrager...) à travers différents catalogues de jeux de coefficients. Ces ajustements ponctuels sont toutefois difficiles à maîtriser. La mise en place d'une évaluation génétique basée sur les contrôles élémentaires, présentée dans le chapitre suivant, ouvre toutefois de nouvelles perspectives, notamment pour le développement d'outils d'appui technique adaptés à l'ensemble des conditions françaises et constamment mis à jour, mais permettant également de valoriser les données de comptages leucocytaires, dont l'appui technique se limite pour l'instant à l'impact économique sur le paiement du lait.

### Références

- Boichard D. 1987. Analyse génétique des caractères de fertilité femelle chez les bovins laitiers. Thèse de Doctorat, Institut National Agronomique Paris-Grignon Paris.
- Boichard D., Barbat A. et Briend M. 1998. Evaluation génétique des caractères de fertilité femelle chez les bovins laitiers. *Rencontres Recherches Ruminants*. 5: 103-106.
- Boichard D., Bonaïti B., Barbat A. and Mattalia S. 1995. Three methods to validate the estimation of genetic trend for dairy cattle. *Journal of Dairy Science*. 78: 475-486.
- Bougler J. 1992. La loi sur l'élevage et l'organisation générale de la sélection en France. *INRA Productions Animales – Hors Série « Eléments de génétique quantitative et application aux populations animales »*. 219-221.
- Bouloc N. 1993. Guide d'utilisation du référentiel courbe de lactation. *Compte rendu Institut de L'Elevage* N°2086, 68p.
- Colleau J.J. 1996. Evaluation génétique des animaux d'élevages. *INRA Productions animales*. N°HS « 50 ans de recherches en production animales »: 27-40.
- Colleau J.J. et Régaldo D. 2001. Définition de l'objectif de sélection dans les races bovines laitières. *Rencontres Recherches Ruminants*. 8: 329-332.
- Ducrocq V. 1997. Mise en place et premiers résultats de l'évaluation génétique des taureaux laitiers sur la longévité des leur filles. *Rencontres Recherches Ruminants*. 4: 215-218.
- Ducrocq V., Boichard D., Barbat A. et Larroque H. 2001. Intégration des caractères fonctionnels dans un index de synthèse pour les races bovines laitières : de la théorie à la pratique. *Rencontres Recherches Ruminants*. 8: 333-336.
- France Contrôle Laitier. 2005a. SIEL : Fiche domaine valorisation du contrôle de performance. V1.7.0. 28p.
- France Contrôle Laitier. 2005b. SIEL : Fiche domaine production laitière. V1.7.0. 35p.
- Interbull. 2000. National genetic evaluation programmes for dairy production traits practiced in Interbull member countries 1999-2000. *Interbull Bulletin*. 24, 111p.
- Interbull. 2008. National Genetic Evaluation Systems Information. [http://www-interbull.slu.se/national\\_ges\\_info2/framesida-ges.htm](http://www-interbull.slu.se/national_ges_info2/framesida-ges.htm). Accessed July. 7, 2008.
- Lassen J., Sorensen M.K., Madsen P. and Ducrocq V. 2007. A stochastic simulation study on validation of an approximate multitrait model using preadjusted data for prediction of breeding values. *Journal of Dairy Science*. 90: 3002-3011.
- Leray O. 1999. Méthodes de comptages de cellules du lait et contrôle qualité. *Journée nationales GTV-INRA*. 7-13.
- Philipsson J. 1998. Global use of bulls and the Interbull system. *Acta Agriculturae Scandinavica, Section A, Animal Science*. Suppl 29: 98-107.
- Rupp R. et Boichard D. 1997. Evaluation génétique des bovins laitiers sur les comptages de cellules somatiques pour l'amélioration de la résistance aux mammites. *Rencontres Recherches Ruminants*. 4: 211-214.
- Schaeffer L.R. 1994. Multiple-Country Comparison of Dairy Sires. *Journal of Dairy Science*. 77: 2671-2678.
- Schaeffer L.R. and Jamrozik J. 1996. Multiple-trait prediction of lactation yields for dairy cattle. *Journal of Dairy Science*. 79: 2044-2055.
- Serieys F. 1985. Interprétation de concentrations cellulaires du lait individuel de la vache pour le diagnostic de l'état d'infection mammaire. *Annales de Recherches Vétérinaires*. 16: 263-269.
- Stolzmann M., Jasiorowski H., Reklewski Z. et al. 1981. Friesian cattle in Poland – Preliminary results of testing different strains. *World Animal Review*. 38: 9.
- VanRaden P.M. 2004. Choice of scales for delivery of genetic evaluations to the public. *Interbull Bulletin*. 32: 118-121.

## **Chapitre 2**

**D'une évaluation génétique basée sur  
les données « lactation » aux  
« contrôles élémentaires »**



## SOMMAIRE du Chapitre 2

<b>Partie 2.1</b>	<b>Modélisation de la courbe de lactation</b>	<b>45</b>
1.	Les modèles empiriques	47
1.1.	Courbes paramétriques	47
1.1.1.	Les fonctions exponentielles	48
1.1.2.	Les fonctions polynomiales	50
1.2.	Courbes avec prises en compte du phénomène d'autocorrélation	53
1.3.	Courbes non paramétriques	54
1.4.	Courbes semi-paramétriques : les splines	56
1.4.1.	Les splines linéaires	56
1.4.2.	Les splines quadratiques	57
1.4.3.	Les splines cubiques	58
1.4.4.	Les autres types de splines	59
2.	Les modèles mécanistes	60
<b>Partie 2.2</b>	<b>Les évaluations génétiques sur les données lactation</b>	<b>63</b>
1.	Etape préliminaire à l'évaluation génétique	64
1.1.	Les recommandations	64
1.1.1.	Les caractères laitiers	65
1.1.2.	La santé de la mamelle	68
1.2.	Le cas de la France	69
1.2.1.	Les caractères laitiers	69
1.2.2.	La santé de la mamelle	71
2.	L'évaluation génétique	71
2.1.	Les recommandations	72
2.2.	Le cas de la France	73
3.	Les étapes postérieures à l'évaluation génétique	74
3.1.	Les recommandations	74
3.2.	Le cas de la France	75
<b>Partie 2.3</b>	<b>Les évaluations génétiques de données longitudinales</b>	<b>77</b>
1.	Modélisation des données longitudinales	78
1.1.	Les différents modèles contrôles élémentaires (TDM)	78
1.1.1.	Les modèles mixtes à régression fixe (FR-TDM)	79
1.1.2.	Les modèles avec régression aléatoire (RR-TDM)	82
1.1.3.	Les modèles multicaractères (MT-TDM)	85
1.1.4.	Les modèles de régression aléatoire avec réduction du rang	86
1.1.5.	Les modèles « Character process » (CP)	87
1.2.	Intérêts des modèles longitudinaux : exemple du TDM sur les caractères laitiers	88
1.3.	Les modèles de régressions aléatoires utilisés à travers le monde	92
1.3.1.	Les autres caractères avec contrôles élémentaires	92
1.3.2.	Les caractères de croissance	92
1.3.3.	Les applications aux analyses de survie	93
1.3.4.	Les applications aux caractères de fertilité en bovins laitiers	93
1.3.5.	Les applications aux études d'interaction génotype-milieu	94

## Chapitre 2

2.	Description générale du modèle contrôles élémentaires	94
2.1.	Les effets dits « fixes »	95
2.1.1.	L'effet Troupeau – Jour de contrôle (HTD)	95
2.1.2.	Les autres effets fixes	95
2.2.	Les effets dits « aléatoires »	97
2.2.1.	L'effet génétique additif	97
2.2.2.	L'effet d'environnement permanent	99
2.2.3.	L'effet troupeau - année	99
2.3.	La résiduelle	100
<b>Partie 2.4</b>	<b>Le modèle « contrôles élémentaires » pressenti en France</b>	<b>103</b>
1.	Le modèle pressenti initialement	104
1.1.	Les effets « fixes »	104
1.2.	Les effets « aléatoires »	106
1.3.	La résiduelle	106
2.	L'estimation des paramètres génétiques	106
2.1.	Méthode d'estimation employée	106
2.2.	L'approche utilisée	107
2.3.	Les paramètres génétiques de la 1 <sup>ère</sup> lactation	108
2.4.	Les paramètres génétiques des 3 premières lactations	110
3.	Le modèle finalement retenu : le modèle de « référence »	111
3.1.	Les effets « fixes »	111
3.2.	Les effets « aléatoires »	111
3.3.	La résiduelle	112

Durant le siècle dernier, la sélection des caractères laitiers a principalement été basée sur la production totale au cours d'une lactation, souvent définie sur une durée de 305 jours. La méthode Fleischmann est généralement considérée comme la méthode de référence pour le calcul de la production au cours de la lactation. Elle est basée sur l'interpolation de la production entre deux dates de contrôles. Les quantités de matières  $Q_{lact}$  (Lait, MG et MP) sont calculées suivant la formule suivante :

$$Q_{lact} = I_0 \times Q_1 + \sum_{i=1}^{n-1} \left( I_i \times \frac{Q_i + Q_{i+1}}{2} \right) + I_n \times Q_n \quad (2.1)$$

où  $Q_i$  est la quantité produite lors du  $i^{\text{ème}}$  contrôle,  $I_i$  est l'intervalle entre le contrôle  $i$  et  $i+1$ ,  $I_0$  est l'intervalle entre le vêlage et le 1<sup>ère</sup> contrôle et  $I_n$  l'intervalle entre le dernier contrôle et le tarissement. Cette méthode ne reste toutefois qu'une approximation. Des scientifiques se sont donc intéressés à la modélisation de la courbe de lactation, avec l'objectif de proposer entre autres une approche plus précise pour estimer la production totale. La production laitière constitue un modèle intéressant pour les scientifiques, comme en témoignent les propositions de Brody *et al.* (1923) ou Gaines (1927) dans les années 1920 pour modéliser la courbe de lactation dont l'intérêt ne se limite pas au calcul de la production totale. Nous verrons dans une première partie, les principales modélisations proposées de la courbe de lactation au cours du siècle dernier.

Il s'est toutefois avéré que ces méthodes basées sur les courbes de lactation n'avaient pas un intérêt net pour le calcul de la production cumulée sur la lactation par rapport à la production estimée avec la méthode Fleischmann. L'évaluation génétique sur la production laitière reste pour le moment encore, dans certains pays dont la France, définie de manière très synthétique par quelques variables représentatives de la production totale par lactation généralement obtenues suivant la méthode Fleischmann pour les quantité de lait, de matières grasse et protéique, les taux butyreux et protéique, ainsi qu'un indicateur de l'état de santé de la mamelle : la numération cellulaire. Compte tenu de l'importance des échanges internationaux chez les bovins laitiers et de la prééminence des caractères laitiers dans la rentabilité des élevages, les évaluations génétiques internationales ont été initiées avec ces caractères. Elles ont permis d'établir des recommandations pour leur évaluation génétique. Nous ferons, dans une seconde partie, un parallèle entre les recommandations émises au niveau international et la situation actuelle de l'évaluation génétique française sur les données lactations.

L'évolution des outils informatiques permet désormais d'envisager la réalisation d'évaluations génétiques à partir de modèles basés sur l'ensemble des contrôles mensuels, valorisant au mieux la richesse des informations recueillies. Les données exploitées sont des productions (ou taux moyens) obtenus sur 24 heures, et sont couramment appelées performances ou contrôles « élémentaires » même si elles correspondent généralement à la somme (ou moyenne pondérée) des deux ou trois pesées quotidiennes. Les modèles valorisant ces données présentent deux principaux avantages : la prise en compte des effets d'environnement du jour du contrôle d'une part et d'autre part la possibilité de modéliser la plupart des effets (génétiques, effets fixes liées à l'âge au vêlage, à la gestation...) sous forme

## *Chapitre 2*

de courbes dont l'influence sur le caractère évalué varie suivant le stade de lactation. Les nombreux travaux sur la modélisation des courbes de lactation ont ainsi pu être valorisés. Les modèles d'évaluations génétiques sur les données longitudinales et plus particulièrement les modèles « contrôles élémentaires » utilisés dans le cadre des caractères de production laitière seront présentés dans une troisième partie.

Au passage d'une évaluation génétique sur les données « lactations » à une évaluation sur les contrôles élémentaires, chaque pays a développé un modèle spécifique, afin de répondre à ses besoins propres. La quatrième partie de ce chapitre sera donc consacrée à la présentation du modèle « contrôles élémentaires » français tel qu'il était pressenti au début de ce travail de thèse.



## **Partie 2.1**

# **Modélisation de la courbe de lactation**

La courbe de lactation traduit l'évolution de la production de lait et de ses constituants en fonction du stade de lactation. Son allure générale est relativement constante entre les diverses espèces domestiques. Au cours de la lactation, deux phases sont généralement distinguées. Suite à la mise bas, la production de lait augmente jusqu'à atteindre un maximum, le pic de lactation, quelques semaines (de 2 à 10 semaines) après la parturition. Knight et Wilde (1993) expliquent que ce phénomène est lié à l'augmentation exponentielle du volume de cellules sécrétrices, d'une part pendant la gestation grâce au phénomène d'hyperplasie (prolifération des cellules) et d'autre part entre le vêlage et le pic de lactation par hypertrophie (intensification de leur activité). Suit une phase plus longue durant laquelle la sécrétion lactée diminue progressivement jusqu'au tarissement, qui peut être provoqué ou physiologique, marqué par un arrêt brutal de la production. Cette seconde phase serait expliquée par l'involution des cellules sécrétrices mais surtout par la chute du nombre de cellules sécrétrices par apoptose.

La richesse du lait (teneur en matières grasses et protéiques) suit une courbe inverse à la sécrétion lactée. Elle diminue rapidement au cours des premières semaines, se stabilise à un niveau minimal puis s'enrichit de nouveau du fait d'une dilution moindre. La composition relative des constituants du lait se modifie profondément durant les premiers jours après la parturition. La concentration en immunoglobulines décroît rapidement après la mise bas au profit des caséines. De plus, la teneur relative en acides gras évolue, traduisant des changements d'ordre métabolique. L'évolution des profils d'acides gras au cours de la lactation obtenus par Soyeurt (2008) montre une baisse des acides gras à chaînes longues ( $C_{18:0}$  et  $C_{18:1}$ ) entre le vêlage et 120 jours de lactation au profit de chaînes plus courtes ( $C_{14:0}$  et  $C_{16:0}$ ). L'évolution des quantités de matières résulte de la combinaison du lait et des taux correspondants.

La courbe de lactation des femelles de ruminants a focalisé l'attention des chercheurs (généticiens, nutritionnistes, physiologistes, technologues...), des agents de développement, des économistes et industriels de la filière lait. L'intérêt considérable porté par des spécialistes de disciplines différentes à la courbe de lactation, tant comme moyen de prévision de la production que comme élément de diagnostic de troubles de la santé, s'explique d'une part par le rôle déterminant de la production laitière vis-à-vis du revenu de l'activité d'élevage correspondante, et d'autre part par les nombreux facteurs biologiques susceptibles d'en modifier le déroulement. La prédiction de l'évolution de la production laitière à l'échelle individuelle ou du troupeau est un puissant outil de gestion des performances du troupeau, d'autant plus important que le contingentement de la production nécessite une connaissance précise des productions envisageables à court et moyen terme (Coulon et Pérochon, 2000). En effet, la connaissance des paramètres des courbes de lactation peut permettre de prédire la production totale à partir d'un seul contrôle (Wood, 1974), et ce d'autant plus précisément que le nombre de contrôles augmente. Un des autres intérêts fréquemment avancé de la modélisation de la production sous forme de courbe est la mesure de ce qui est appelé la persistance. La sélection d'animaux présentant une persistance élevée (faible décroissance de la production au cours de la seconde phase de la lactation) est intéressante car les quantités de lait produites au cours de l'année sont à peu près stables, ce qui garantit des revenus réguliers

à l'éleveur ; le stress physiologique lié à un bilan énergétique négatif en début de lactation peut être réduit, ce qui peut entraîner une réduction du nombre et de la sévérité des troubles métaboliques et reproductifs (Gipson and Grossman, 1990). L'animal peut avoir une alimentation plus stable, favorisant notamment la proportion de fourrage dans la ration, et ainsi réduire les coûts alimentaires (Sölkner and Fuchs, 1987).

L'intérêt que présente l'étude de la courbe de lactation est reflété par une multiplicité des modèles mathématiques proposés pour la décrire ou la prévoir. Ces modèles sont d'autant plus appréciés et utilisés qu'ils ont une interprétation biologique ou économique simple. Les modèles déterministes (modèles dont la connaissance des lois d'évolution et l'état actuel permet de prévoir l'état futur), aussi qualifiés d'empiriques (modèles s'appuyant exclusivement sur l'expérience et l'observation et non sur une théorie), qu'ils soient linéaires ou non, représentent la grande majorité des études publiées dans la bibliographie. Ils se contentent de modéliser la production laitière à partir de données pré-existantes. Certains d'entre eux permettent d'obtenir de bons ajustements de toute ou partie de la courbe de lactation, ainsi que d'estimer certaines caractéristiques classiques de la courbe (date et niveau du maximum de production, persistance en phase décroissante, production totale). Cependant, l'interprétation des paramètres des modèles plus complexes est difficile et, dans ce cas, il peut être impossible de les relier de façon simple à ces caractéristiques classiques. Toutefois, ces modèles ne sont pas aptes à rendre compte de façon logique des variations à court terme du niveau de production autour d'une tendance générale. Pour traduire ces variations, quelques auteurs ont cherché à construire des modèles d'ajustement de la courbe de lactation, dits stochastiques (où le processus pour passage d'un stade à l'autre n'est soumis qu'au hasard), qui s'appuient sur un modèle déterministe de description de l'allure globale de la courbe, et une fonction traduisant le caractère autocorrélé des productions journalières successives. La majorité des modèles déterministes ne cherchent pas à rendre compte des phénomènes nutritionnels et physiologiques sous-jacents à la sécrétion lactée, contrairement aux modèles mécanistes. Ces derniers, désormais envisageables grâce au développement d'outils informatiques performants, cherchent à modéliser la réponse d'un animal par la représentation du métabolisme tissulaire sous-jacent. Ces modèles se différencient par leur qualité, qui dépend étroitement de leur aptitude à simuler le processus biologique d'évolution de la production laitière et de leur capacité d'ajustements aux facteurs de variations.

## **1. Les modèles empiriques**

### **1.1. Courbes paramétriques**

Pour décrire l'évolution temporelle de phénomènes biologiques (courbe de lactation, croissance...), les scientifiques ont généralement recours à des courbes paramétriques. La variation au cours du temps est modélisée à l'aide de fonctions linéaires comme celles d'Ali et Schaeffer (1987), Wilmink, (1987), Guo et Swalve (1995) ou non linéaires comme celles de Nelder (1966), Wood (1967), Grossman et Koops (1988). La production estimée  $y=f(t)$ , qui est souvent une production journalière obtenue le jour  $t$ , est une fonction plus ou moins complexe du stade de lactation.



Deux des avantages des courbes paramétriques sont premièrement de résumer les caractéristiques de distribution à travers un petit nombre de paramètres et deuxièmement de disposer algébriquement des caractéristiques de la courbe. En effet, la date du pic de production est obtenue pour la valeur de  $t$  où la dérivée première de  $f(t)$  s'annule et le niveau de ce pic  $t_{pic}$  correspond à la valeur de la fonction  $f(t)$  pour  $t$  égal à  $t_{pic}$ . La production entre 2 dates sera obtenue en calculant l'intégrale de  $f(t)$  sur cet intervalle de temps.

### 1.1.1. Les fonctions exponentielles

Parmi les modèles déterministes, les courbes de lactation modélisées par des fonctions exponentielles ont été largement utilisées. D'un point de vue historique, le premier modèle mathématique décrivant une courbe de lactation chez les vaches laitières, proposé par Brody *et al.* en 1923 était basé sur une fonction exponentielle. La production attendue  $E(y)$  du jour  $t$  y était décrite comme :

$$E(y_t) = a e^{-ct} \quad (2.2)$$

où  $a$  est le facteur d'échelle permettant d'ajuster la production au niveau de production du début de la lactation et  $c$  est associé à la phase descendante de la courbe de lactation, et peut être qualifié d'indicateur de persistance.

La principale limite de ce modèle était la non prise en compte de la partie ascendante de la courbe. Différentes propositions ont été faites pour améliorer cet aspect. La plus marquante est celle proposée par Wood (1967), aussi appelée fonction Gamma incomplète, qui a été l'une des plus communément employées dans les études portant sur la courbe de lactation. La production attendue  $E(y)$  du jour  $t$  est décrite comme :

$$E(y_t) = a t^b e^{-ct} \quad (2.3)$$

où  $a$  est le facteur d'échelle permettant d'ajuster la production au niveau de production moyenne en début de la lactation, les paramètres  $b$  et  $c$  étant associés respectivement à la phase ascendante et descendante de la courbe de lactation. Le succès de ce modèle non linéaire s'explique d'une part parce qu'il prend en compte la phase ascendante de la courbe, et d'autre part parce que les paramètres sont faciles à estimer après linéarisation de l'équation ( $\ln(y(t)) = a + b \ln(t) - ct$ ). Wood (1977 et 1979) ainsi que Masselin *et al.* (1987) ont tenté de donner une interprétation physiologique aux paramètres  $a$ ,  $b$  et  $c$  en lien avec les modalités d'utilisation de l'énergie durant la lactation. Le paramètre  $a$  traduirait le potentiel métabolique global à produire du lait. Les expressions  $t^b$  et  $e^{-ct}$  sont quant à elles à rapprocher de l'évolution métabolique liée à la production laitière (néoglucogénèse suivie de la diminution de la contribution des réserves corporelles) et de l'activité mammaire (mobilisation progressive des cellules sécrétrices suivie de la décroissance exponentielle de leur nombre).

Toutefois, ce modèle présente certaines limites. Par construction, la production du jour du vêlage est contrainte à être nulle. Cobby et Le Du (1978) mentionnent une surestimation de la production en début de lactation et sous-estimation du pic de lactation. Dhanoa (1981) ainsi que Goodall et Sprevak (1984) indiquent des problèmes de corrélations fortes entre estimés des paramètres pour le premier, et entre les résidus pour les seconds. Par ailleurs, les

tentatives d'application de cette équation à la description de l'évolution des taux ont été peu fructueuses, donnant des ajustements médiocres ce qui est toutefois logique car l'évolution est inverse.

Pour dépasser ces limites, de nombreux auteurs ont proposé des améliorations du modèle de Wood afin d'augmenter sa flexibilité (Schaeffer *et al.*, 1977 ; Cobby et Le Du, 1978 ; Dhanoa, 1981 ; Goodall, 1986 ; Grossman *et al.*, 1986 ; Grossman et Koops, 1988 ). Trois types d'améliorations ont été apportés par les modèles dérivés de celui de Wood (Beever, 1991) :

- ✓ Une plus grande souplesse dans la modélisation des formes de courbe (Cobby and Le du, 1978 ; Papajcsik and Bodero, 1988).
- ✓ Une amélioration des propriétés mathématiques du modèle (diminution de la corrélation entre les paramètres du modèle (Dhanoa, 1981) et de la méthode d'estimation des paramètres (Cobby et Le du, 1978)).
- ✓ La création de modèles prédictifs intégrant des séries temporelles.

Avec le souci de disposer de modèles qui s'ajustent à chacune des phases de production (phase ascendante – pic – phase descendante), plusieurs auteurs ont proposé des modèles dits à compartiments ou multiphasiques. Ces modèles sont exprimés sous la forme de sommes de plusieurs fonctions qu'ils espèrent pouvoir interpréter à partir des caractéristiques de production à différentes périodes. Ainsi, ces modèles de type additif présentent l'avantage de ne pas contraindre la production initiale à être nulle (Dematawewa *et al.*, 2007). Inspirés de l'équation de Wood (1967), Cobby et Le Du (1978) ont proposé par exemple, un modèle où la production attendue  $E(y)$  du jour  $t$  est décrite comme :

$$E(y_t) = a + b t - a e^{-ct} \quad (2.4)$$

Comparativement au modèle de Wood (2.3), le déclin de la production laitière  $y$  est modélisé de façon linéaire plutôt que de façon exponentielle (Sherchand *et al.*, 1995). Ce modèle permet notamment un meilleur ajustement aux performances du début de lactation, là où le modèle de Wood (2.3) sous-estimait le pic de lactation (Rowlands *et al.*, 1982).

D'autres modèles à compartiments existent, comme celui de Grossman et Koops (1988). Ce dernier comprend 2 à 3 phases. La production  $y$  du jour  $t$  est définie par une somme de fonctions logistiques faisant intervenir la fonction tangente hyperbolique. Toutefois, ce type de modèle présente l'inconvénient de multiplier le nombre de paramètres à estimer par le nombre de phases décrivant la courbe de lactation et donc tend à accroître les problèmes d'estimation (Scott *et al.*, 1996). En réponse aux critiques faites à ce modèle sur l'absence de lien avec les connaissances biologiques, Grossman et Koops (2003) ont proposé un nouveau modèle multiphasique (extension et reparamétrisation du modèle de 1988), avec 12 paramètres, qui se rapproche des modèles mécanistes de par l'interprétation biologique possible des paramètres de la courbe. Le modèle est construit autour des 3 phases de la lactation que sont la prolifération, l'apoptose et la gestation. Dematawewa *et al.* (2007) confirment le très bon ajustement obtenu avec ce modèle, mais nuancent les possibilités d'utilisation d'un tel modèle du fait de problèmes de convergence et de difficultés dans l'interprétation des paramètres obtenus.

Wilmink (1987) a proposé une fonction linéaire avec 3 (ou 4) paramètres dérivée du modèle de Cobby et Le Du (1978) pour décrire la courbe de lactation. La production attendue au jour  $t$  :  $E(y)$  est décrite comme :

$$E(y_i) = a + b t + d e^{-ct} \quad (2.5)$$

Selon l'auteur, le paramètre  $a$  est associé au niveau de production,  $b$  à la phase décroissante de la production et  $d$  à l'augmentation de la production au cours de la phase ascendante. Ce modèle offre une flexibilité accrue par rapport au modèle de Cobby et Le Du (1978) en distinguant le coefficient associé au niveau de production initial de celui de la phase ascendante. Le coefficient  $c$  a été estimé à 0.05. Il correspond à la période du pic de lactation, soit environ 50 jours après la mise bas. Ce dernier coefficient peut ou non faire l'objet d'estimation spécifique, en fonction des résultats d'analyses préliminaires. Ainsi Olori *et al.* (1999) ont estimé ce coefficient à 0.061, Brotherstone *et al.* (2000) à 0.100 et 0.068 suivant le fichier de données et Silvestre *et al.* (2006) à 0.065. Brotherstone *et al.* (2000) soulignent l'importance de paramètre  $d$  dans la capacité du modèle de Wilmink à modéliser la production journalière en début de lactation.

### 1.1.2. Les fonctions polynomiales

La production laitière peut être considérée comme une fonction polynomiale du temps écoulé depuis la mise bas ou d'une fonction simple de celui-ci (puissance, logarithme...). Masselin *et al.* (1987) font une revue complète de ces modèles. Historiquement, nous pouvons citer le modèle de Gaines (1927), un modèle linéaire du 1<sup>er</sup> degré où la production attendue au jour  $t$  est décrite comme :

$$E(y_i) = a + c t \quad (2.6)$$

le paramètre  $a$  est associé au niveau de production initial et  $c$  décrit la décroissance de la production au cours de la lactation et reflète le niveau de persistance. Le principal intérêt de ce modèle est sa simplicité et sa facilité d'interprétation des paramètres. Toutefois, l'incapacité du modèle à s'ajuster aux deux phases d'une courbe de lactation limite considérablement son utilisation.

Des polynômes de degré supérieur ont été utilisés pour modéliser la production laitière. Avec ces modèles, les paramètres restent simples à estimer. Toutefois, la fonction quadratique (degré 2) a l'inconvénient de présenter une symétrie au niveau du pic de production, les polynômes de degré 3 permettant quant à eux d'introduire une dissymétrie dans la courbe. Les ajustements obtenus par certains auteurs sont satisfaisants mais comme l'indiquent Masselin *et al.* (1987), ne concernent généralement qu'une portion de la courbe de lactation.

Nelder (1966) a proposé une famille de rapport de fonctions polynomiales (appelées fonctions inverses) pour ajuster les courbes de réponse dans les expériences plurifactorielles, notamment dans le cadre de la modélisation de la courbe de lactation. La production attendue au jour  $t$   $y$  est décrite comme :

$$E(y_i) = \frac{t}{a + b t + c t^2} \quad (2.7)$$

D'après Batra (1986), le coefficient  $a$  correspond à l'augmentation de la production au cours de la phase ascendante,  $b$  à la pente générale de la courbe et  $c$  au taux de déclin de la production après le pic de lactation. Dans certaines situations, ce modèle peut donner de bons résultats. C'est le cas lorsque l'étude porte plus spécifiquement sur le début de lactation. L'utilisation de moyennes de production hebdomadaires semble donner de très bons ajustements (valeur de  $R^2$  pouvant atteindre 0.99) comme le rapporte Batra (1986). Toutefois, le modèle de Nelder présente quelques limites. Comme pour le modèle de Wood, il contraint le niveau de production à être nul le jour du vêlage.

Ali et Schaeffer (1987) ont proposé l'utilisation d'une fonction linéaire à 5 variables où la production attendue au jour  $t$  est décrite comme :

$$E(y_t) = a + bu + cu^2 + dv + ev^2 \quad \text{où } u = \frac{t}{305} \text{ et } v = \ln\left(\frac{305}{t}\right) \quad (2.8)$$

où  $a$  est un paramètre associé au pic de production,  $d$  et  $e$  sont associés à la partie ascendante de la courbe de production et  $b$  et  $c$  à la partie descendante. D'après les auteurs, ce modèle a l'avantage de présenter des corrélations génétiques nulles entre les paramètres du modèle et la production totale de la lactation. Ceci permet donc de modifier la forme de la courbe de lactation sans modifier le niveau global de production. Un problème potentiel est toutefois soulevé par les auteurs : les paramètres du modèle de régression sont fortement corrélés les uns avec les autres (entre -0.73 et 0.46), ce qui peut limiter fortement son utilisation.

Toutefois, les résultats obtenus avec cette modélisation sont contradictoires. D'après l'étude faite par Jamrozik *et al.*, (1997a), ce modèle donne des résultats très proches de ceux obtenus avec le modèle de Wilmink (1987), malgré le fait que la modélisation de Ali et Schaeffer comporte un paramètre supplémentaire. Guo et Swalve (1995 et 1997) trouvent au contraire ce modèle inférieur aux autres, notamment à celui de Wilmink.

Concernant les limites de ce modèle, Macciotta *et al.* (2005) notent de très fortes corrélations en valeur absolue (entre 0.85 et 0.99) entre les coefficients des courbes de lactation ayant une forme standard, ce qui constitue d'après Kettunen *et al.* (2000) une limite bien connue du modèle de Ali et Schaeffer. Les corrélations obtenues par ces auteurs sont bien supérieures à celles évoquées par Ali et Schaeffer (1987). Pour cette raison, cette fonction a généralement été abandonnée au profit des polynômes de Legendre (Schaeffer, 2004).

Contrairement à la fonction polynomiale d'Ali et Schaeffer, développée uniquement à des fins de modélisation de la courbe de lactation, il existe des fonctions mathématiques plus générales pouvant être utilisées pour modéliser des fonctions biologiques de type longitudinal. Ainsi, les polynômes de Legendre, publiés autour de 1797, sont des fonctions polynomiales de degré  $n$  où la production attendue au jour  $t$  peut être décrite sous la forme :

$$E(y_t) = \sum_{i=0}^n \alpha_i \phi_i(\omega) \quad \text{avec } \omega = 2 \left( \frac{t - t_{\min}}{t_{\max} - t_{\min}} \right) - 1 \text{ et } \phi_n(\omega) = \sqrt{\frac{2n+1}{2}} P_n(\omega) \quad (2.9)$$

où  $\alpha_i$  est le coefficient de régression,  $\omega$  est une unité de temps standardisée variant entre -1 et +1 (Kirkpatrick *et al.*, 1990) avec  $t_{\min}$ , le premier jour de contrôle possible, généralement 5

ou 7 jours après la mise bas et  $t_{\max}$  le stade maximum pris en compte, soit 305, 335 voire 365 jours suivant les pays ; et  $\phi_n(\omega)$  est la version normalisée du polynôme  $P_n(\omega)$  de degré  $n$ .

Le polynôme de degré  $n+1$  peut être écrit sous forme de l'équation récursive suivante :

$$P_{n+1}(\omega) = \frac{1}{n+1}((2n+1)\omega P_n(\omega) - n P_{n-1}(\omega)) \text{ et } P_0(\omega) = 1 \text{ soit } P_1(\omega) = \omega, P_2(\omega) = \frac{3}{2}\omega^2 - \frac{1}{2} \dots \quad (2.10)$$

$\phi_n(\omega)$  est généralement multiplié par  $\sqrt{2}$  de sorte que le coefficient de régression du terme constant soit égal à 1. L'ordre du polynôme de Legendre utilisé a un impact important sur la flexibilité de la modélisation de la courbe de lactation. Une propriété mathématique intéressante des polynômes de Legendre est leur orthogonalité<sup>1</sup> sur l'intervalle  $[-1, 1]$  qui minimise les corrélations entre les coefficients de régression. Combinée à l'utilisation d'unité de temps standardisée, ces propriétés permettent d'avoir une meilleure convergence et une meilleure précision qu'avec des polynômes classiques. Toutefois, comme dans le cas du modèle d'Ali et Schaeffer, les modèles comportant un nombre élevé de paramètres (à partir de 5) et donc très flexibles sont aussi plus sensibles aux variations locales, notamment en début et fin de lactation (effets de bord). Meyer (1998) observe que ces variations ont un impact non négligeable sur les coefficients de régression.

De nombreuses autres fonctions polynomiales ont été proposées pour décrire des courbes telles que celle de la lactation. Parmi eux, nous pouvons citer les polynômes fractionnaires (Royston and Altman, 1994) dont l'application à des scores de cellules somatiques a permis d'illustrer la flexibilité (Robert-Granié *et al.*, 2002). Le polynôme fractionnaire de degré  $m$ , fonction du temps  $t$  peut être écrit sous la forme :

$$E(y_i) = b_0 + b_1 t^{p_1} + b_2 t^{p_2} + \dots + b_m t^{p_m} = \sum_{i=0}^m \beta_i t^{p_i} = \phi_m(t; \mathbf{p}) \quad (2.11)$$

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

soit d'après la transformation de Box-Tidwell  $t^{p_i} = \begin{cases} t^{p_i} & \text{si } p_i \neq 0 \\ \ln(t) & \text{si } p_i = 0 \end{cases} \quad (2.12)$

Le polynôme fractionnaire dépend donc du degré  $m$  choisi ainsi que du vecteur  $\mathbf{p}$  des puissances. Royston et Altman (1994) ont suggéré l'utilisation d'un ensemble prédéfini  $\mathcal{P}$  de puissances, avec  $\mathcal{P} = \{-2, -1, -\frac{1}{2}, -\frac{1}{3}, 0, \frac{1}{3}, \frac{1}{2}, 1, 2, \dots, \max(3, m)\}$  couvrant de façon appropriée un grand nombre de cas pratiques. Ainsi, le modèle d'Ali et Schaeffer (2.8) peut être réécrit sous la forme  $E(y_i) = a_0 + a_1 t + a_2 t^2 + a_3 \ln(t) + a_4 (\ln(t))^2 = b_0 + b_1 \ln(t) + b_2 (\ln(t))^2 + b_3 t + b_4 t^2 = \phi_4(t; \mathbf{p} = (0, 0, 1, 2))$ .

Les résultats obtenus par Robert-Granié *et al.* (2002) montre qu'un polynôme fractionnaire de degré 2 permet d'obtenir des résultats équivalents à un modèle d'Ali et Schaeffer. Ils constituent donc une alternative parcimonieuse et flexible pour modéliser la courbe de lactation.

<sup>1</sup> Deux fonctions  $f(x)$  et  $g(x)$  sont orthogonales sur l'intervalle  $a \leq x \leq b$  si l'intégrale de leur produit  $\int_a^b f(x) g(x) dx = 0$ .

## 1.2. Courbes avec prises en compte du phénomène d'autocorrélation

Les modèles déterministes reflètent assez fidèlement la courbe. Par conséquent, ils peuvent être d'un usage intéressant pour l'interprétation de résultats d'expérimentation. Mais ils ne sont pas aptes à rendre compte logiquement des variations à court terme du niveau de production autour de la tendance générale. Pour appréhender ces variations, quelques auteurs ont cherché à construire des modèles d'ajustement de la courbe de lactation dits stochastiques. Ces modèles s'appuient sur un modèle déterministe de description de l'allure globale de la courbe, et une fonction utilisant les autocorrélations existant entre les mesures successives pour modéliser les variations à court terme autour de la trajectoire attendue. De tels modèles ouvrent des perspectives nouvelles, notamment pour l'analyse des résultats du contrôle laitier, mais aussi en termes de prédiction de la production (Beever *et al.*, 1991).

Ce type de modèle repose sur un constat simple mentionné par Turner (1926). Sous des conditions normales, la production laitière postérieure au pic de lactation est une proportion stable de la production du mois précédent. A partir de cette observation, Dhanoa et Le Du (1982) ont proposé une méthode de calcul de la production laitière par récurrence, en faisant l'hypothèse que l'écart de production observé entre deux semaines consécutives est proportionnel à la variation de production espérée, qu'ils écrivent comme :

$$E(y_t) = k y_t^{est} + (1-k)y_{t-1} \quad (2.13)$$

où  $y_t$  et  $y_{t-1}$  sont les productions observées à  $t$  et  $t-1$ ,  $k$  est un paramètre compris entre 0 et 1 et  $y_t^{est}$  est la production estimée à la date  $t$ . Ce modèle permet de bien s'ajuster aux données, que le pic de lactation soit prononcé ou non. En outre, il est simple et sensible aux variations dues à des facteurs environnementaux. Il peut donc être utilisé pour l'interprétation des courbes de lactation (Dhanoa, 1981). Un tel modèle permet de rendre compte des variations à court terme de la production laitière, liées à des variations de conditions environnementales et/ou expérimentales, ce que ne font pas les modèles déterministes.

Une autre méthode récursive d'estimation de la production laitière a été proposée par Goodall et Sprevak (1985), le but étant d'obtenir une meilleure estimation de la production de lait en n'ayant que des données de début de lactation et ainsi d'améliorer la qualité prédictive du modèle. Cette proposition est basée sur une procédure d'estimation Bayésienne dans laquelle est utilisée l'information antérieure du troupeau et les observations de la lactation en cours afin d'obtenir des prédictions précoces et précises de la production totale sur la lactation. Ce modèle utilise la méthode dite du filtre de Kalman (1960), un estimateur récursif, généralement utilisé dans les problèmes de prédiction et d'estimation de systèmes linéaires dynamiques tels que les séries temporelles. Les paramètres nécessaires à l'obtention de la courbe de lactation avec la méthode de Goodall et Sprevak (1985) sont stables pour la plupart des troupeaux dès lors que 10 semaines de données sont disponibles. Cinq semaines suffisent dans de nombreux cas. Les courbes obtenues avec la méthode de Goodall et Sprevak (1985) montrent un très bon ajustement aux données réelles à partir de 10 semaines de données, contrairement à celles obtenues avec la méthode des moindres carrés qui nécessitent un minimum de 15 semaines de données pour obtenir un ajustement équivalent.

Deluyker *et al.* (1990) ont proposé d'utiliser une méthode basée sur l'analyse des séries temporelles des données laitières collectées quotidiennement à partir de modèle dit « Moyenne Mobile Intégrée AutoRégressive » (ARIMA pour AutoRegressive Integrated Moving Average). Cette approche exploite le niveau élevé d'autocorrélation entre les observations obtenues à intervalle régulier pour prédire la production à court terme. Elle est basée sur 3 étapes successives que sont : la phase d'identification du modèle sous-jacent (notamment des autocorrélations), la phase d'estimation des paramètres du modèle retenu et enfin une phase de prévision de la série à partir du modèle estimé dans l'étape précédente. D'après les auteurs, ce type de modèle peut s'avérer très intéressant en confrontant la production prédite à la production réalisée pour la détection de problèmes de santé tels que les mammites ou les cétozes. Il n'est cependant exploitable que dans les exploitations disposant de relevé journalier de la production laitière.

Comme l'affirment Masselin *et al.* (1987), de tels modèles ouvrent donc des perspectives nouvelles pour l'analyse des résultats du contrôle laitier, notamment en terme de détection et d'interprétation des variations de production. Toutefois, ces méthodes nécessitent des ressources informatiques importantes et ne se justifient que dans le cas de données relevées fréquemment (au minimum hebdomadaire) et pour un objectif de prévision à court terme notamment pour le suivi technique des performances (Dhanoa and Le Du, 1982). En effet, les corrélations entre mesures successives diminuant lorsque l'intervalle entre celles-ci augmente, la pertinence de telles démarches est limitée dans le cas de contrôles mensuels.

### 1.3. Courbes non paramétriques

La modélisation de la lactation sous forme d'une courbe permet d'une part de lisser les variations aléatoires observées, d'autre part de résumer la courbe en quelques paramètres. Cependant, ces variables synthétiques peuvent engendrer des biais importants si l'ajustement avec les données est médiocre. La flexibilité du modèle doit donc être suffisante pour permettre de bien s'ajuster à n'importe quelle fluctuation. Les courbes non paramétriques permettent de modéliser la courbe de lactation sans a priori sur la forme de cette courbe. La façon la plus simple d'estimer des courbes de lactation de type non paramétriques est d'avoir recours aux équations du modèle mixte. En effet, chaque classe aura comme valeur celle qui minimise l'erreur quadratique moyenne. Les résiduelles de chacune des classes de la courbe ont donc, par construction, une moyenne nulle (Druet *et al.*, 2003). La flexibilité des courbes non paramétriques dépend étroitement du nombre de classes comme l'illustre la figure 2.1.

Suivant les auteurs, la formation des classes et leur nombre varient de façon importante. Stanton *et al.* (1992) utilisent des classes de 5 jours jusqu'à 200 jours de lactation et de 10 jours entre 201 et 369 jours de lactation pour définir des courbes des lactations par rang de lactation, soit 56 paramètres par courbes. Keown *et al.* (1986) définissent des courbes de lactation par rang de lactation (3 classes), période de vêlage (6 classes de 2 mois) et niveau de production du troupeau (5 classes) soit 90 courbes par caractère, en utilisant des classes de 1 jour en début et fin de lactation (avant 100 jours et entre 231 et 305 jours de lactations) et des classes de 10 jours entre ces 2 périodes soit 182 classes par courbe. Ces modèles se caractérisent donc par leur flexibilité mais sont coûteux en terme de nombre de paramètres à estimer et donc en degrés de liberté perdus.

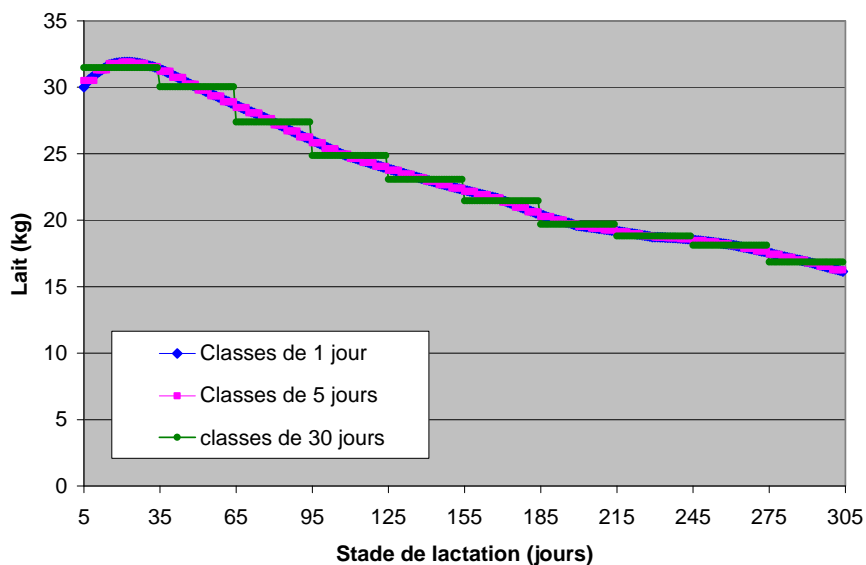


Figure 2.1. Effet de l'amplitude des classes de stade de lactation sur la modélisation de la courbe de lactation à partir de courbes non paramétriques

Les tables développées dans le cadre de la construction des référentiels de courbes de lactation telles que celles dites « ARCOLA » (Institut de l'Élevage, 1993) ou dans l'étude de Baraton et Masselin de l'ITEB (Institut Technique de l'Élevage Bovin, 1990) sont typiquement des courbes non paramétriques. Les valeurs de la courbe de base sont estimées pour chaque décade à partir de quelques dizaines de milliers de lactation représentant les situations les plus fréquentes. Des coefficients additifs et multiplicatifs permettent ensuite d'adapter la courbe de base aux différentes situations de rang de lactation, de mois de vêlage, d'âge au vêlage. Outre le grand nombre de paramètres générés par ces courbes, leur mise à jour est souvent un facteur crucial. Les paramètres utilisés actuellement ont été estimés au début des années 1990. L'évolution des niveaux de production et les changements de conduite, notamment de la périodicité des vêlages, les rendent certainement inappropriés.

En 1989, Elston *et al.* ont proposé une méthode non paramétrique, qui répond toutefois à certaines hypothèses générales sur la courbe de lactation (un pic de production, existence ou non d'un point d'inflexion...). Ces hypothèses se traduisent par des contraintes sur les signes des dérivées premières, secondes et troisièmes de la courbe de telles sortes qu'elles soient de signes équivalents aux dérivées obtenues avec une courbe de Wood. Associée à une méthode d'interpolation entre les points ajustés préservant la forme de la courbe, elle aboutit à une représentation graphique. Le recours à des dérivées troisièmes complique toutefois la détermination de la méthode d'interpolation satisfaisant l'ensemble des restrictions établies. Cette modélisation se révèle être un outil descriptif intéressant qui permet de donner des estimations des caractéristiques de la courbe. Toutefois, les auteurs ne recommandent pas l'utilisation de cette méthode dans le cadre de la prédiction de la production future, car les contraintes sur la forme de la courbe de lactation sont jugées insuffisantes.



## 1.4. Courbes semi-paramétriques : les splines

Pour répondre aux besoins de flexibilité d'une part et s'affranchir du choix d'un modèle de courbe paramétrique d'autre part, des courbes dites semi-paramétriques ont été proposées pour la modélisation de données longitudinales. Les splines sont des fonctions polynomiales par morceaux, c'est-à-dire des courbes construites à partir de segments de polynômes de faible degré dont la jonction se fait au niveau de nœuds. Le terme anglophone de spline désigne de longue et fine bande de bois utilisé pour les lattes de la literie. Quand elle sont fixés à certains points donnés : les nœuds, le bois prend une forme qui minimise l'énergie nécessaire pour l'incurver, ce qui permet d'obtenir la forme la plus régulière qui soit. L'approche mathématique des splines est particulièrement utilisée dans les analyses non-paramétriques nécessitant le lissage de courbes (Green and Silvermann, 1994 ; Ruppert *et al.*, 2003). Comme dans le cas des polynômes de Legendre, ces courbes peuvent modéliser nombre de fonctions biologiques. Verbyla *et al.* (1999) les ont utilisées pour modéliser des courbes de croissance, White *et al.* (1999) pour des courbes de lactation bovines.

Avec les splines, les observations n'influencent pas l'ensemble de la courbe, mais seulement les segments adjacents. La qualité d'ajustement des splines dépend étroitement du nombre de nœuds pris en compte dans la fonction. D'après White *et al.* (1999), limiter le nombre de nœuds du modèle contribue directement à lisser la courbe. Inversement, augmenter le nombre de nœuds contribue à un meilleur ajustement de la courbe aux données, mais aussi à une plus grande sensibilité aux données. Le nombre de nœuds est donc modulable en fonction de la complexité de la trajectoire. Par cet aspect, les splines constituent une alternative très souple pour modéliser la courbe de lactation. De plus, la position des nœuds peut être ajustée de sorte qu'ils soient denses en début de lactation, lorsque la trajectoire de la courbe change rapidement et au contraire plus épars sur la seconde moitié de la lactation lorsque la trajectoire est plus régulière. Les splines peuvent être classées selon leur complexité. Nous n'évoquerons dans cette présentation que les splines linéaires jusqu'aux cubiques.

### 1.4.1. Les splines linéaires

Dans le cas le plus simple, la fonction spline est définie par un segment linéaire, équivalente à l'interpolation linéaire entre les 2 nœuds consécutifs. La courbe ainsi obtenue est une ligne brisée et correspond à l'extension directe de la régression linéaire paramétrique. Dans le contexte des courbes de lactation, si  $y_i$  correspond à la performance enregistrée au temps  $t_i$ ,  $T_k$  représente la position du  $k^{\text{ème}}$  nœud de la courbe de lactation formée par  $k - 1$  périodes. La spline linéaire de l'observation  $y_i$  prend la forme :

$$y_i = g_k(t_i) + e_i = \beta_0 + \sum_k \beta_{1k} (t_i - T_k)_+ + e_i \quad (2.14)$$

où  $\beta_0$  et  $\beta_{1k}$  sont respectivement la constante et le coefficient de régression linéaire,  $e_i$  est l'erreur relative à l'observation  $y_i$  et  $(x)_+ = \max(0, x)$ , c'est-à-dire égal à  $x$  si  $x$  est positif (soit dans notre situation, égal à l'intervalle entre le nœud précédent et le stade de lactation  $t_i$ ) ou  $x$

est nul autrement (soit, nul si le stade de lactation  $t_i \notin [T_k, T_{k+1}]$ ). La fonction  $g_k(t_i)$  peut être décrite comme :

$$g_k(t_i) = y_k \frac{T_{k+1} - t_i}{T_{k+1} - T_k} + y_{k+1} \frac{t_i - T_k}{T_{k+1} - T_k} = y_k + \frac{y_{k+1} - y_k}{T_{k+1} - T_k} (t_i - T_k), t_i \in [T_k, T_{k+1}] \quad (2.15)$$

Selon cette formulation,  $y_k$  correspond à la constante et  $\frac{y_{k+1} - y_k}{T_{k+1} - T_k}$  au coefficient de régression

linéaire ou pente de la courbe. Les splines linéaires sont donc mathématiquement très simples. Bohmanova *et al.* (2008) les utilisent pour modéliser les effets fixes et aléatoires d'un modèle de contrôles élémentaires sur les caractères de production laitière en raison de leur simplicité. Les changements de pente au niveau des nœuds peuvent être brutaux avec des splines linéaires, car la continuité de la dérivée première n'est pas garantie. Pour remplir cette condition et améliorer la modélisation, il faut recourir aux splines quadratiques.

### 1.4.2. Les splines quadratiques

Dans ce type de splines, une fonction polynomiale quadratique est utilisée pour s'ajuster aux données. La formulation équivalente à l'équation 2.14 est dans ce cas :

$$y_i = g_k(t_i) + e_i = \beta_0 + \sum_k \beta_{1k} (t_i - T_k)_+ + \sum_k \beta_{2k} (t_i - T_k)_+^2 + e_i \quad (2.16)$$

$(x)_+^p$  est une fonction puissance tronquée, la spline quadratique étant donc basée sur une puissance tronquée de degré 2. Afin d'obtenir une courbe lissée, les pentes des deux segments contigus doivent être équivalentes au nœud de jonction, ce qui se traduit par les contraintes suivantes sur les dérivées premières :

$$g'_k(T_{k+1}) = g'_{k+1}(T_{k+1}) \quad (2.17)$$

où  $g'_k(T_{k+1})$  est la dérivée première de la fonction  $g_k$  au stade de lactation du  $k+1^{\text{ème}}$  nœud.

La fonction  $g_k(t_i)$  répondant à ces contraintes peut être décrite comme :

$$g_k(t_i) = y_k + z_k (t_i - T_k) + \frac{z_{k+1} - z_k}{2(T_{k+1} - T_k)} (t_i - T_k)^2, t_i \in [T_k, T_{k+1}] \quad (2.18)$$

Les coefficients  $z$  peuvent être déterminés en choisissant un  $z_0$  (généralement égal à 0) et en utilisant la relation de récurrence  $z_{k+1} = -z_k + 2 \frac{y_{k+1} - y_k}{T_{k+1} - T_k}$  pour définir les coefficients suivants.

Parmi les différents types de splines, Misztal (2006) recommande l'utilisation des splines quadratiques qu'il juge être un bon compromis entre complexité et qualité d'ajustement. Toutefois, ce type de splines présente parfois des oscillations importantes, comme l'illustre la figure 2.2. L'utilisation de splines cubiques permet d'améliorer la trajectoire des courbes.

### 1.4.3. Les splines cubiques

Une spline cubique est une courbe lissée sur un intervalle donné obtenue par la jonction, au niveau de nœuds, de segments de polynômes cubiques de telle sorte que la courbe ainsi que ses dérivées premières et secondes sont continues sur les deux segments consécutifs. Avec des variables similaires aux équations 2.14 et 2.16, une spline cubique sera écrit sous la forme :

$$y_i = g_k(t_i) + e_i = \beta_0 + \sum_k \beta_{1k} (t_i - T_k)_+ + \sum_k \beta_{2k} (t_i - T_k)_+^2 + \sum_k \beta_{3k} (t_i - T_k)_+^3 + e_i \quad (2.19)$$

Le lissage de la courbe est obtenu par les contraintes sur les premières et secondes qui doivent être continues entre chaque nœuds. Ces contraintes peuvent être écrites sous la forme :

$$g'_k(T_{k+1}) = g'_{k+1}(T_{k+1}) \text{ et } g''_k(T_{k+1}) = g''_{k+1}(T_{k+1}) \quad (2.20)$$

où  $g'_k(T_{k+1})$  et  $g''_k(T_{k+1})$  sont respectivement la dérivée première et seconde de la fonction  $g_k$  au stade de lactation du  $k+1$ ème nœud.

La fonction  $g_k(t_i)$  répondant à ces contraintes est écrite sous la forme :

$$g_k(t_i) = \frac{(t_i - T_k)y_{k+1} + (T_{k+1} - t_i)y_k}{T_{k+1} - T_k} - \frac{1}{6}(t_i - T_k)(T_{k+1} - t_i) \left[ \left( 1 + \frac{t_i - T_k}{T_{k+1} - T_k} \right) v_{i+1} + \left( 1 + \frac{T_{k+1} - t_i}{T_{k+1} - T_k} \right) v_k \right] \quad (2.21)$$

avec le stade de lactation  $t_i \in [T_k, T_{k+1}]$ . La fonction et la dérivée seconde ont respectivement pour valeur  $y_k$  et  $v_k$  au nœud  $k$ .

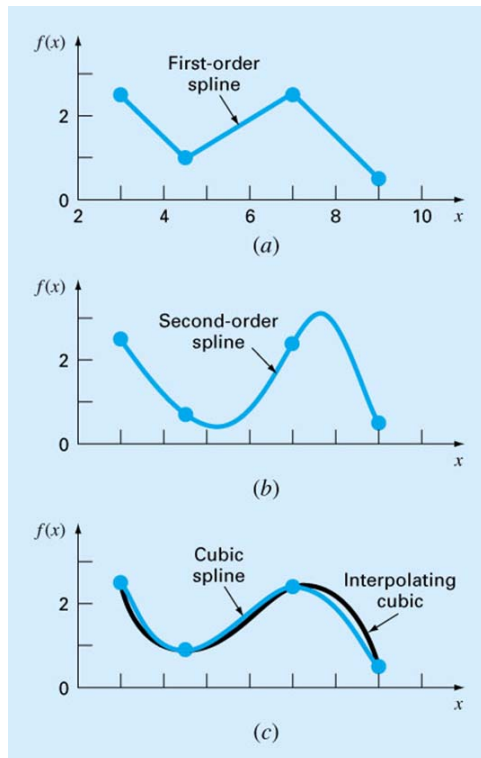


Figure 2.2. Représentation schématique des 3 principaux types de splines : spline linéaire (first-order spline), spline quadratique (second-order spline) et spline cubique (Cubic spline) Source : <http://ceprofs.tamu.edu/jzhang/cven-302/9> consulté le 15 septembre 2008

#### 1.4.4. Les autres types de splines

Différents types de splines ont été développés à partir des splines cubiques. Parmi ces modèles, nous pouvons citer les splines cubiques naturelles qui présentent l'intérêt d'être linéaire en dehors de l'intervalle formé par le premier et le dernier nœud. Elles sont particulièrement utiles lorsque les modalités extrêmes comportent peu d'observations.

La spline de régression est une généralisation de l'équation 2.19, qui peut s'écrire sous la forme :

$$y_i = g_k(t_i) + e_i = \beta_0 + \beta_1 t_i + \dots + \beta_p t_i^p + \sum_k \beta_{pk} (t_i - T_k)_+^p + e_i \quad (2.22)$$

La spline de lissage (smoothing spline) considère généralement un nœud pour chaque observations. En tenant compte des données  $(t_1, y_1), \dots, (t_n, y_n)$ , la spline lissée est celle qui minimise :

$$\sum (y_i - g(t_i))^2 + \alpha \int_{x_1}^{x_q} \left( \frac{d^2 g}{dt^2} \right)^2 dt \quad (2.23)$$

où l'intégrale pénalise les fluctuations et  $\alpha$  est un paramètre (pénalité) qui détermine le compromis entre la fidélité de la courbe par rapport aux données et l'allure lissée de cette courbe. L'objectif du lissage est de réduire les fluctuations ponctuelles associées à certaines données extrêmes. Le paramètre de lissage pénalise les variations brutales et contribue avec l'accroissement de la taille des segments (diminution du nombre de nœuds) à obtenir une courbe lisse. L'estimation du paramètre  $\alpha$  complexifie toutefois nettement sa mise en œuvre.

Les B-splines permettent d'obtenir des ajustements équivalents à des splines basées sur une fonction puissance tronquée, mais possède des propriétés numériques très intéressantes. Torres et Quaas (2001) les ont utilisés dans des modèles de régression aléatoire pour modéliser les performances laitières élémentaires, Meyer (2005) pour modéliser des courbes de croissance.

L'univers des splines n'ayant pour limites que la fertile créativité des chercheurs, il existe également des splines pénalisées (P-splines), les M-splines qui sont toujours positives (utilisées en analyse de survie), les I-splines qui sont basées sur l'intégrale de M-splines, des pseudosplines...

Pour résumer, les splines sont des fonctions présentant de nombreuses propriétés intéressantes. Les comparaisons de modélisation de la courbe de lactation réalisées par Druet *et al.* (2003) montrent que les splines cubiques présentent une flexibilité accrue par rapport aux approches paramétriques classiques. Comparativement aux modèles non paramétriques, les splines permettent d'obtenir des résiduelles de tendances similaires, c'est-à-dire

globalement nulles quel que soit le stade de lactation, à l'exception de très légères déviations en début et fin de lactation. En revanche, elles montrent peu de variations locales contrairement aux courbes non paramétriques et surtout nécessitent un nombre de paramètres réduit. La qualité d'ajustement de la courbe aux données dépend cependant étroitement du nombre de nœuds, dont l'augmentation induit un accroissement des besoins en terme calculatoire. D'après Druet *et al.* (2003), les splines constituent toutefois un bon compromis en terme de qualité d'ajustement, de sensibilité aux données, de comportement local, de lissage et de nombre de paramètres.

## 2. Les modèles mécanistes

De manière complémentaire aux modèles empiriques, plusieurs modèles mécanistes ont été développés pour simuler la réponse métabolique de l'animal, chaque terme du modèle ayant une interprétation biologique (Beever *et al.*, 1991). Bien que certains auteurs de modèles descriptifs aient tenté de donner une interprétation physiologique de leur modèle, comme l'a fait Wood en 1977 en introduisant les notions de stockage/déstockage de réserves et de développement de la glande mammaire, l'un des premiers modèles sur la lactation qui peut être qualifié de mécaniste est celui de Neal et Thornley (1983). Il est basé sur la différenciation des cellules mammaire et leur mort cellulaire programmée (apoptose) en intégrant trois facteurs : un flux d'apport de nutriments, une sécrétion d'hormone et le nombre de cellules sécrétrices. La lactation  $y(t)$  est décrite en fonction de l'activité des cellules sécrétrices. Ce modèle a toutefois une utilisation pratique très limitée du fait de l'indisponibilité des paramètres nécessaires au modèle, notamment sur la connaissance des apports nutritionnels.

Dijkstra *et al.* (1997) ont proposé un modèle très simple basé sur 4 ou 5 paramètres pour décrire le développement et la mort des cellules de la glande mammaire au cours de la gestation et de la lactation des mammifères. D'après les auteurs, il est difficile de donner une interprétation précise des paramètres compte tenu de la grande variabilité observée entre les animaux. Toutefois, Val-Arreola *et al.* (2004) interprètent les paramètres de Dijkstra *et al.* comme le niveau initial théorique de production laitière, le taux de différenciation des cellules sécrétrices (en séparant la phase ayant lieu pendant la gestation et celle du début de la lactation dans le cas d'un modèle à 5 paramètres), le déclin de la production et le taux de mort cellulaire. Ce modèle, développé par certains auteurs à partir de différentes espèces, principalement des rongeurs, présente toutefois un bon ajustement aux données de vaches laitières comme en témoigne Val-Arreola *et al.* (2004) et Dematawewa *et al.* (2007).

Pollott (2000) a proposé un modèle basé sur 6 paramètres qui décrit les 3 processus : la différenciation, l'apoptose et la sécrétion lactée de la glande mammaire. Les paramètres du modèle sont interprétés comme le potentiel de sécrétion maximum de la glande mammaire, le taux de croissance cellulaire, la proportion de cellules sécrétrices (différenciées) présentes au début de la lactation, le taux de différenciation cellulaire, la proportion de cellules disparaissant lors de la parturition et le taux de mort cellulaire. Ce modèle a été étendu aux composants du lait : matière grasse, matière protéique, lactose (Pollott, 2004). Malgré l'intérêt de ce modèle en terme d'interprétation des paramètres, les résultats sont controversés puisque

d'un point de vue pratique, il présente une mauvaise convergence d'après Val-Arreola *et al.* (2004) et Dematawewa *et al.* (2007). Malgré l'utilisation de différents algorithmes proposés dans la procédure du logiciel SAS (Statistical Analysis System) dédiée au modèle mixte non linéaire (NLMIXED) et notamment de l'algorithme de quasi-newton proposé par défaut, ces auteurs mentionnent des problèmes de convergence pour les modèles avec plus de 5 paramètres comme celui de Pollott. Ces problèmes seraient liés à une grande sensibilité du modèle aux valeurs initiales des paramètres. De plus, il existe très peu de méthodes exactes pour l'estimation des paramètres des modèles mixtes non linéaire ce qui réduit considérablement le niveau de confiance accordé à ces approches. Toutefois, l'émergence de méthode exacte d'estimation des paramètres basée sur l'algorithme SAEM-MCMC ouvre de nouvelles voies (Duval, 2008) d'utilisation de ces modèles.

Les modèles mécanistes les plus sophistiqués permettent donc de décrire précisément la courbe de lactation, mais sont souvent très difficiles à mettre en œuvre du fait de problèmes de convergence liés à une sur-paramétrisation et de difficulté dans l'interprétation des paramètres obtenus. Les modèles comportant un nombre limité de paramètres sont donc généralement favorisés pour leur facilité d'interprétation biologique.

Dans le cadre d'une analyse génétique avec un modèle mécaniste simplifié dépendant uniquement du potentiel de sécrétion maximum de la glande mammaire et du taux de mort cellulaire, Albarrán-Portillo et Pollott (2008) ont estimé l'héritabilité de ces deux caractères à 0.27 et 0.08 respectivement. Les corrélations élevées obtenues entre le pic de production et le potentiel de sécrétion maximum de la glande mammaire (0.99) et entre la persistance et le taux de mort cellulaire (0.84) en font un modèle très intéressant.

Dans la plupart des modèles vus précédemment, deux caractéristiques sous-jacentes, que sont le pic de lactation et la persistance, sont prédominantes. Toutefois, cette dernière notion a une définition et un mode d'expression qui varient largement suivant les modèles. La courbe de lactation est la résultante d'un nombre important de facteurs. Les principaux sont ceux relatifs à la conduite du troupeau (alimentation, état sanitaire, période de mise bas...), à l'état physiologique de l'animal (âge, gestation) et à son niveau génétique. L'étude des courbes de lactation permet de quantifier l'effet de ces différents facteurs, dans la mesure où ils peuvent être distingués les uns des autres. L'impact de ces facteurs sera abordé ultérieurement.

La modélisation des courbes de lactation a souvent été utilisée comme outil descriptif des formes de courbes observées. Désormais, les recherches sont consacrées à la modélisation de la variabilité entre courbes de lactation individuelles plutôt qu'à la modélisation d'une courbe commune à l'ensemble des vaches (Stanton *et al.*, 1992 ; Pérochon *et al.*, 1996 ; White *et al.*, 1999). Avec le développement des moyens informatiques, il s'avère que le principal intérêt des courbes de lactation réside dans son aspect prédictif (Coulon et Pérochon, 1996 ; Olori *et al.*, 1999). Prédire l'évolution de la production permet, en confrontant la prédiction à la production réelle, d'établir un diagnostic dynamique. Ce type de modélisation peut éventuellement être utilisé à des fins de management pour moduler l'alimentation en ajustant au mieux la production réelle à la production souhaitée. De plus, en cas de variation atypique, la forme de la déviation de la production peut constituer un élément de diagnostic pour identifier le ou les facteurs causaux. Ces variations anormales devraient être encore plus

facilement détectables et interprétables lorsque les données de production laitière sont enregistrées quotidiennement comme c'est le cas dans les troupeaux équipés de compteurs à lait. D'autre part, à partir du moment où, sous l'influence d'un facteur exogène, la production s'est éloignée significativement de sa trajectoire la plus probable, il importerait d'étudier la ou les lois du retour à une trajectoire normale. Dans ce domaine, il semble indispensable de développer la voie tracée par Dhanoa et Le Du (1982) et Goodall et Sprevak (1984) pour aboutir à des modèles qui soient capables d'intégrer l'existence des variations de la production autour d'une trajectoire moyenne. De tels modèles pourraient également être utilement appliqués dans le cadre de la valorisation des résultats du contrôle laitier. En effet, la règle généralement adoptée consiste à souligner une chute mensuelle de production supérieure à 10%, en ne considérant que deux enregistrements consécutifs (Masselin *et al.*, 1987). Dans les outils d'appui technique actuellement disponibles, tels que dans SIEL, cette valeur est paramétrable. Or l'application de modèles stochastiques devrait permettre d'aboutir à une meilleure interprétation des variations de production mensuelles en intégrant plus largement le passé de chaque courbe de lactation. Le développement des outils informatiques permet désormais de faire évoluer cette situation. Toutefois, le risque d'aboutir ainsi à un modèle lourd et complexe n'est pas négligeable, ce qui signifie qu'il ne faut pas délaissier la mise au point d'une gamme de modèles simples et opérationnels adaptés aux objectifs précis des différents utilisateurs (alimentation, génétique...).

Le choix d'un modèle ainsi que la quantité et la qualité des informations nécessaires à son estimation doivent donc être raisonnés en fonction de l'utilisation souhaitée. Ainsi, le calcul de la production totale sur la lactation à l'aide de 10 contrôles mensuels apparaît moins satisfaisante à partir de l'ajustement à une courbe de Wood qu'avec la méthode classique de Fleischmann : la précision n'est pas améliorée et le biais reste légèrement plus élevé (Schaeffer and Burnside, 1976 ; Congleton and Everett, 1980 ; Rowlands *et al.*, 1982 ; Anderson *et al.*, 1989). Le choix du modèle doit se baser dans ce cas d'une part sur la facilité d'estimation des paramètres, sa polyvalence (modélisation possible des différents constituants du lait et pas uniquement de la quantité de lait) et sur la qualité d'ajustement jugée non seulement sur les corrélations entre production prédite et production réalisée et la moyenne des résiduelles mais aussi sur leur variance et leur distribution. A niveau d'ajustement comparable, Guo et Swalve (1997) recommandent l'utilisation du modèle présentant le plus faible nombre de paramètres possible.

## **Partie 2.2**

# **Les évaluations génétiques sur les données lactation**



Les évaluations génétiques nationales sur les caractères de production sont basées (ou l'ont été) sur un modèle dit « lactation » qui se base sur les données cumulées au cours de la lactation. En effet, certains pays ont évolué vers un modèle dit « contrôles élémentaires » qui sera présenté dans la partie suivante. Malgré une apparente homogénéité dans la définition des caractères laitiers, de nombreuses différences subsistent comme en témoignent les corrélations entre les pays participant aux évaluations internationales qui restent inférieures à 1 (Interbull, 2008). Selon Wiggans (2000), ces corrélations reflètent les différences dans la collecte des données de production laitière, les tailles de populations, les méthodes de calcul et outils statistiques disponibles lors du développement du modèle d'évaluation, les compétences et préférences du personnel ainsi que le temps disponible pour effectuer les améliorations souhaitées. Améliorer ces corrélations en homogénéisant les pratiques nationales permettrait de favoriser les échanges mondiaux et profiterait à l'ensemble des pays. Après avoir présenté les recommandations faites par le Comité International pour le Contrôle des Performances en Élevage (ICAR) sur l'enregistrement des données de production laitière (ICAR, 2007) et par Interbull (sous-comité d'ICAR) sur les évaluations génétiques nationales sur les caractères laitiers (extraite de l'Interbull Bulletin N°28, 2001), les indexations actuellement réalisées en France sur les caractères qui nous intéressent seront décrites. Une évaluation génétique se décompose classiquement en trois étapes : une étape de préparation des données, l'étape d'évaluation génétique sensu stricto et l'étape post-évaluation.

## **1. Etape préliminaire à l'évaluation génétique**

Cette étape couvre la période allant de la collecte des données jusqu'à leur utilisation dans les évaluations génétiques. Des vérifications sont généralement faites sur les données afin de vérifier la cohérence de celles-ci avant leur utilisation pour les évaluations génétiques. Les précorrections éventuelles des données pour différents facteurs sont faites lors de cette étape.

### **1.1. Les recommandations**

Sur les aspects liés à l'enregistrement des performances laitières, ICAR recommande aux organismes des différents pays de s'engager dans des protocoles agréés au niveau international afin de garantir la qualité des données utilisées pour leurs évaluations génétiques. Les informations relatives à l'identification de l'animal, sa généalogie sur un minimum de 3 générations, ses dates de naissance et de vêlage, son troupeau de production sont jugés indispensables à la prise en compte des données dans l'évaluation génétique.

Le contrôle de performances laitières est réalisé suivant des modalités répondant aux normes du protocole auquel l'éleveur a adhéré. Ces modalités sont reconnues au niveau international par le Comité International pour le Contrôle des Performances en Élevage (ICAR) pour le calcul de performances laitières individuelles utilisées pour l'évaluation génétique des reproducteurs et les échanges commerciaux. Pour répondre aux attentes des éleveurs en terme de réduction des contraintes et de coût lié au contrôle de performances plus faible, la diversité des protocoles s'est accrue. Initialement basées sur un contrôle mensuel (toutes les 4 semaines) réalisé par un agent agréé sur une période de 24 heures (dit protocole A4, « A »

pour Agent et « 4 » pour la fréquence), un premier assouplissement dans les années 1970 a permis d'une part d'espacer les contrôles et d'autre part de ne contrôler qu'une fraction de la production journalière (protocole AT), en alternant les traites contrôlées (matin et soir). Dans les années 1990, de nouvelles demandes d'allègement provenant du monde entier afin de réduire les coûts et de s'adapter à l'automatisation de la traite ont abouti à la mise en place de nouveaux protocoles agréés internationalement. Le contrôle de performances laitières peut désormais être réalisé par l'agent, l'éleveur ou bien alternativement l'agent et l'éleveur en fonction des traites du matin et du soir (respectivement codés A, B et C) ; concerner la totalité des traites sur 24h (aucune codification), une seule des traites (T), correspondre à l'utilisation d'un robot (R) ou bien décrire une quantité de lait collectée sur 24h mais un échantillonnage des taux provenant d'une seule de traite alternativement (Z) ; et enfin être réalisé suivant une fréquence variable de 4 à 6 semaines. Ainsi, le nombre de protocoles agréés s'est accru depuis les années 2000. En regroupant les fréquences des protocoles de types B et C ainsi que les enregistrements provenant de robot, nous obtenons en France la distribution figurant à la figure 2.3 pour l'année 2007 (ICAR, 2008). Comparativement à d'autres pays laitiers, la France avec 78% de contrôle de type A montre un suivi très présent puisque ce protocole représente pour les Pays-Bas autour de 60% des contrôles, 45% pour l'Allemagne, 20% pour les Etats-Unis et 10% pour le Danemark. Les protocoles alternés de type AT avec près de 17% des contrôles sont eux relativement peu présent en France alors qu'ils représentent jusqu'à 65% des contrôles aux Etats-Unis. Les protocoles de contrôles par l'éleveur (B) sont eux en phase de démarrage en France, où ils représentent seulement 2.4% des contrôles, alors qu'ils représentent 16% aux Etats-Unis, 33% en Allemagne et 90% au Danemark.

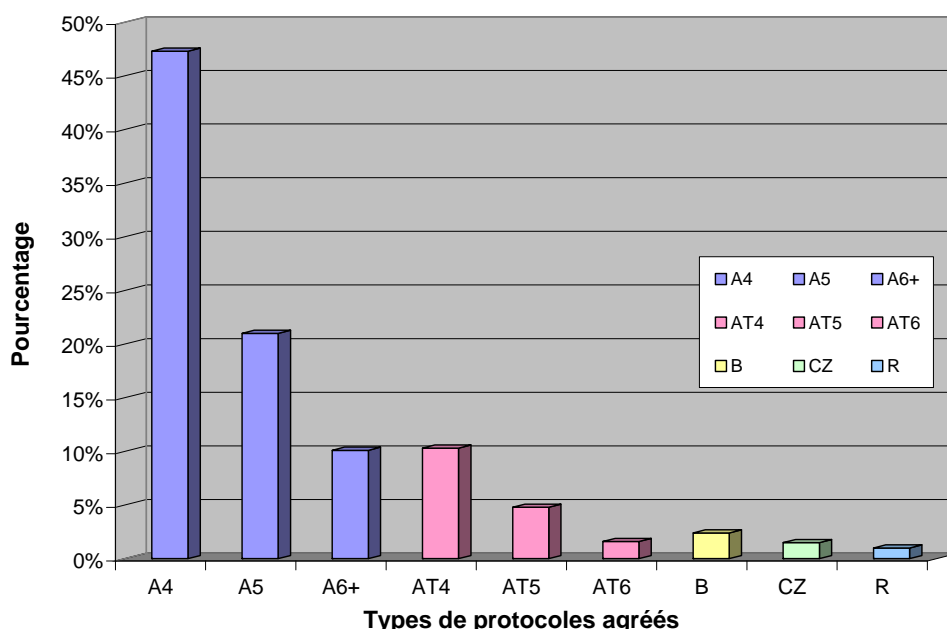


Figure 2.3. Fréquence des différents protocoles de contrôles laitiers agréés en France pour l'année 2007

### 1.1.1. Les caractères laitiers

Le premier point de cette étape consiste à disposer d'une production mesurée sur 24 heures. Avec certains protocoles, notamment ceux dits A et B, les mesures sont faites sur 24 heures et

ne posent donc aucun problème. Par contre, pour les protocoles de type AT, la production n'est recueillie que sur l'une des deux traites. Différentes méthodes ont donc été proposées pour reconstituer la production 24 heures, consistant notamment à corriger les données pour les horaires de traites (DeLorenzo et Wiggans, 1986 ; Liu *et al.*, 2000a). Pour les contrôles continus réalisés par des automates de traite, des méthodes spécifiques ont été développées (Bouloc *et al.*, 2003).

Le deuxième point consiste à vérifier les données élémentaires. En effet, les données doivent répondre à des normes précises pour être prises en compte dans le calcul de la production de référence. ICAR recommande d'intégrer dans le calcul de la lactation les performances mesurées à partir du 5<sup>ème</sup> jour de lactation, présentant un poids de lait minimum de 3kg sur 24 heures en deçà duquel la vache est considérée comme tarie et des taux compris entre 15 et 90 g/kg pour le TB et entre 10 et 70 g/kg pour le TP (à l'exception des races ayant un TB moyen supérieur à 50g/kg, soit la race Jersiaise pour les populations présentes en France). La cohérence des données est aussi vérifiée. Elles doivent être biologiquement plausibles, c'est-à-dire correspondre à des plages de valeurs « homologuées » en terme de niveau de production, d'âge, de rang de lactation.

Le troisième point concerne le calcul de la production de référence ou « standard » (i.e. d'une durée de 305 jours) qui constitue certainement l'aspect le plus important de cette étape. Malheureusement, c'est celui qui diffère le plus entre les pays et qui est donc sujet à débat. Interbull recommande la prise en compte dans les évaluations génétiques des trois premières lactations de l'animal au minimum en intégrant les nouvelles lactations dès lors que leur dernier contrôle connu dépasse le stade de 45 jours de lactation ou qu'elles comptent ont moins 2 contrôles (Interbull, 2001). Interbull préconise une mise à jour régulière des méthodes et facteurs d'ajustement utilisés dans le cas de précorrection des données (au moins tous les 5 ans). Compte tenu de la durée de référence, les lactations en cours ou celles de vaches réformées inférieures à 305 jours doivent donc être étendues. Celles de vaches tarées avant 305 jours ou naturellement terminées avant 305 jours doivent être étendues en tenant compte de l'effet de la gestation. Les lactations supérieures à 305 jours doivent être tronquées à 305 jours. Théoriquement, les règles et la méthode d'extension utilisées doivent être les mêmes pour toutes les données au sein d'une évaluation raciale, seuls les facteurs d'ajustement peuvent varier notamment en fonction du rang de lactation.

Cependant, chaque pays a développé une méthode de calcul de la lactation de référence qui lui est propre, généralement fondée sur des aspects historiques, parfois assez éloignée des recommandations internationales. Parmi les méthodes recommandées par ICAR, il y a la méthode Fleischmann basée sur l'interpolation de la production entre deux dates de contrôle. Cette méthode, malgré sa simplicité, donne une très bonne approximation de la production réelle, et ne présente pas de biais (Schaeffer et Jamrozik, 1996).

Des méthodes d'interpolation basées sur l'utilisation de courbes standard sont aussi utilisées. C'est dans cette optique que certaines des approches, présentées dans la première partie de ce chapitre, ont été développées. Les facteurs de variations tels que le rang de lactation, l'âge au vêlage et la période de vêlage sont généralement pris en compte dans ces courbes. Schaeffer

et Jamrozik (1996) affirment que ces méthodes de prédiction ont une bonne précision dès lors que 4 ou 5 contrôles sont disponibles. Leur qualité prédictive n'est toutefois pas supérieure à celle de la méthode Fleischmann lorsque toutes les données caractérisant la production laitière sont mesurées à chaque contrôle et que la périodicité mensuelle des contrôles est respectée (Schaeffer and Burnside, 1976 ; Anderson *et al.*, 1989). Dans le contexte de lactation terminée, répondant à un protocole de type A4 ou B4, la méthode Fleischmann constitue un bon compromis, notamment grâce à sa facilité de mise en œuvre.

Plus récemment, d'autres méthodes ont été développées pour améliorer la prédiction de la production de référence. Face aux pressions visant à alléger les contraintes et les coûts des protocoles de collecte des performances laitières et les risques de possible déconnexion dans le temps des mesures de lait et de taux, des méthodes exploitant les corrélations entre ces caractères se sont développées. Ainsi, VanRaden (1997) a proposé d'exploiter les propriétés de la méthode dite de « meilleure prédiction ». Les covariances entre les différents jours de contrôle et la moyenne de production du troupeau sont utilisées pour prédire la production de chacun des jours de la lactation ainsi que la production en 305 jours. Cette méthode présente l'avantage de régresser la performance prédite vers la moyenne, qui a donc une variance inférieure à celle de la performance réelle, contrairement à la performance prédite avec la méthode Fleischmann. Norman *et al.* (1999) ont montré l'intérêt de cette méthode comparativement à la méthode Fleischmann dans le contexte des protocoles AT ou d'un protocole A trimestriel, qui permet respectivement une diminution de l'erreur d'estimation de 6 et 10%. Lorsque les taux ne sont pas mesurés sur l'ensemble des contrôles, l'utilisation d'une approche multicaractères permet de compenser partiellement la perte d'information. Schaeffer et Jamrozik (1996) ont quant à eux proposé l'emploi d'une prédiction multicaractères basée sur les approches de régression aléatoire (cf 3<sup>ème</sup> partie de ce chapitre). Comme l'approche précédente, l'intérêt de cette méthode est d'exploiter non seulement les corrélations élevées entre les contrôles successifs, mais aussi les corrélations importantes entre les différents caractères laitiers. Dans l'article de Schaeffer et Jamrozik (1996), le modèle de Wood (1967) est utilisé pour modéliser la courbe de lactation en tenant compte de la race, de la région, du rang de lactation, de l'âge et de la période de vêlage. Outre le calcul de la production de référence en tenant compte des intervalles entre contrôles et de la prise en compte des contrôles incomplets, ce modèle permet d'estimer les caractéristiques de la courbe de lactation telles que le pic ou la persistance pouvant être utiles à des fins d'appui technique. Différentes approches sont envisageables pour calculer la production de référence à partir des nombreuses modélisations de courbes de lactation vues précédemment. Guo et Swalve (1997) conseillent la méthode des splines cubiques pour interpoler les données de production en cas de valeurs manquantes.

Pour tenir compte d'éventuelles différences de précision dans le calcul de la lactation de référence, Interbull recommande de faire varier le poids accordé à chacune des lactations en fonction de leur qualification (protocole de collecte des données).

Le dernier point de cette étape est l'éventuelle précorrection des données faite par certains pays pour différents facteurs environnementaux, à partir de facteurs additifs ou multiplicatifs. Le choix entre facteurs additifs et multiplicatifs dépend de l'impact de l'effet sur le caractère.

L'utilisation de facteurs additifs suppose des effets d'importance équivalente intra classe quel que soit le niveau de production de l'animal alors que les facteurs multiplicatifs présument que l'effet est proportionnel au niveau de production. Ce choix entre les deux méthodes est nécessaire pour les effets ayant un aspect temporel tels que l'âge au vêlage, l'intervalle entre vêlage, le mois de vêlage... mais aussi le numéro de lactation pour obtenir des lactations standardisées sur la production adulte. Interbull rappelle toutefois qu'il s'avère préférable d'inclure ces effets dans le modèle, ou si cela n'est pas envisageable d'avoir recours à des facteurs multiplicatifs qui permettent d'ajuster la variance ainsi que la moyenne (Wiggans, 2000). L'inclusion des effets dans le modèle permet de limiter les risques de biais sur le progrès génétique (Bonaïti *et al.*, 1993 ; Boichard *et al.*, 1995), sinon en cas de précorrection, les facteurs d'ajustement doivent être réestimés régulièrement (*cf* 2.1 pour plus d'explication).

### 1.1.2. La santé de la mamelle

La santé de la mamelle est évaluée à travers la mesure du taux cellulaire (nombre de cellules par ml de lait), qui lors d'une infection augmente du fait de la réponse immunitaire. Les cellules épithéliales étant toujours présentes dans le lait de par le remplacement du tissu sécréteur, le taux cellulaire d'une vache saine est généralement compris entre 30 000 et 100 000 cellules / ml suivant le stade de lactation (Serieys, 1985).

Pour le calcul d'un taux cellulaire sur 24 heures, la problématique est similaire à celle des caractères laitiers lorsque le lait analysé ne provient pas d'un mélange entre la traite du matin et celle du soir, ce qui est le cas dans les protocoles de type AT. Les données cellulaires étant collectées via le même réseau d'information que les caractères laitiers, elles répondent aux mêmes normes de prise en compte et de cohérence que ces derniers.

En ce qui concerne le calcul d'une variable résumant les observations sur une lactation, différentes méthodes ont été proposées telles que l'utilisation d'une moyenne arithmétique, d'une moyenne géométrique, ou d'une procédure combinant les différentes mesures en fonction de leur covariance avec la moyenne des autres contrôles (Wiggans et Shook, 1987). Toutefois, préalablement à ce calcul d'une variable sur la lactation, les données subissent généralement une transformation logarithmique de sorte que la variable obtenue appelée Score de Cellules Somatiques (SCS) suive une distribution à peu près normale. Il existe cependant une grande diversité dans le calcul de la variable représentant le niveau de cellules sur la lactation (Interbull, 2008). La transformation logarithmique utilisée est variable suivant le pays (VanRaden, 2004). La majorité utilise un logarithme de base 2, mais d'autres emploient un logarithme de base 10 (cas de l'Australie) ou de base  $e$  (Royaume-Uni, Afrique du Sud) (Interbull, 2008). Les données élémentaires utilisées pour le calcul de la variable par lactation peuvent aussi avoir fait l'objet de correction pour le stade de lactation et le rang de lactation, ainsi que pour le mois de contrôle, voire l'âge au vêlage (Schutz *et al.*, 1995) avant leur combinaison.

Comme dans le cas des caractères laitiers, la dernière étape concerne un éventuel préajustement additif ou multiplicatif de la variable résumant les données de la lactation ( $SCS_L$ ) pour certains facteurs environnementaux de façon à ajuster la variance de cette variable. Toutefois la recommandation d'Interbull d'éviter ces pratiques reste valable quel que soit le caractère.

## 1.2. Le cas de la France

Le système de management de la qualité du dispositif génétique français suivant une norme ISO 9001 est actuellement en cours de déploiement. Il permettra de répondre pleinement aux recommandations faites par ICAR. En effet, son périmètre comprend l'identification des animaux, les activités relatives à l'insémination, la certification des parentés, l'ensemble des processus tel que le contrôle de performance concourant à l'évaluation sur la production laitière. La disponibilité et la qualité de l'ensemble des informations nécessaires à l'évaluation génétique sont donc garanties.

Concernant les normes relatives aux protocoles de contrôle de performance, la France présente la spécificité par rapport aux autres pays de vérifier pour chaque lactation terminée que le protocole déclaré par contrat entre l'éleveur et l'OCL a bien été appliqué. Ainsi à partir de l'ensemble des dates de contrôles enregistrées dans le SIG, il est possible de vérifier que l'intervalle entre les contrôles et le nombre de contrôles réalisés au cours d'une lactation sont compatibles avec le protocole déclaré. Un intervalle exceptionnellement long est autorisé pendant la lactation, généralement sur un des contrôles de la période estivale. En cas de non-conformité, la qualification est dégradée jusqu'à un protocole de niveau compatible en termes d'intervalle et de nombre de contrôles. Si aucun protocole n'est compatible, la lactation est disqualifiée et n'est alors plus incluse dans les évaluations. Chaque lactation est prise en compte dans l'évaluation génétique avec un poids relatif à la perte de précision de la mesure par rapport à une lactation enregistrée suivant un protocole A4 (Institut de l'Élevage-INRA, 2005).

### 1.2.1. Les caractères laitiers

Pour le calcul de la production en 24 heures dans le cas de protocole AT, la production est multipliée par 2 sans tenir compte de l'intervalle horaire entre traites qui, d'après le règlement technique, ne doit pas dépasser les limites 10 heures – 14 heures. Cette méthode, certes très simplifiée, est basée sur une stricte alternance entre les traites du matin et du soir, ce qui permet de compenser les biais potentiels de chacun des contrôles. Sur une lactation en 305 jours, elle présente ainsi un biais relativement faible : 1.1% soit + ou -69 kg pour une lactation de référence de 6195 kg, avec un signe qui dépend de la 1<sup>ère</sup> traite contrôlée (matin vs soir) (Leclerc *et al.*, 2005). Pour les robots de traite, la production 24 heures est calculée suivant la méthode décrite par Bouloc *et al.* (2003).

Les données doivent par ailleurs satisfaire les normes du règlement technique du contrôle laitier français pour être prises en compte dans les évaluations génétiques, notamment en terme d'intervalle entre deux contrôles successifs. En France, seuls les contrôles réalisés à partir de 7 jours de lactation sont pris en compte dans le calcul de la performance laitière de l'animal. Le seuil minimum de prise en compte de la production est, quant à lui, de 2kg de lait sur 24 heures ; les critères sur les taux sont ceux d'ICAR.

L'évaluation génétique sur les caractères laitiers a été décrite par Bonaïti *et al.* (1990). Pour le calcul de la production de référence, la production totale de la lactation (en cours ou terminée) est d'abord estimée selon la méthode Fleischmann pour le lait et les quantités de matières. Les

taux butyreux et protéique sont des taux moyens obtenus à partir du calcul des matières. Pour les premières lactations, et les lactations de rang supérieur ou égal à 2 de plus de 305 jours, la production brute totale est corrigée pour la durée et ramenée à un niveau adulte pour obtenir la production standard selon la formule (Poutous *et al.*, 1981) :

$$QL_{réf} = QL \times \frac{385}{D+80} \times C \quad (2.24)$$

où  $QL_{réf}$  est la quantité de lait précorrigée (en kg), QL est la quantité de lait brute produite (en kg), D est la durée de lactation et C est un coefficient dépendant du rang de lactation (1.30 en 1<sup>ère</sup> lactation, 1.17 en 2<sup>ème</sup> lactation et 1.10 pour les 3<sup>èmes</sup> lactations et suivantes). La constante 80 du dénominateur est une valeur ad-hoc qui, à l'époque, rendait  $QL_{réf}$  à peu près indépendant de la durée de la lactation au-delà de 250 jours, afin de ne pas défavoriser les femelles ayant une bonne fertilité, tout en pénalisant les lactations plus courtes (Poutous et Mocquot, 1975). Le coefficient multiplicatif C vise à exprimer les lactations en équivalent adulte. Il ne corrige que partiellement pour les différences de moyenne entre rangs de lactation. En revanche, il gomme grossièrement les différences de variance d'un rang de lactation à l'autre (Bonaïti *et al.*, 1990). Cette méthode est contraire avec les recommandations internationales qui conseillent de tronquer la lactation au 305<sup>ème</sup> jour de production. D'après Bonaïti *et al.* (1990), cette correction se justifie afin de ne pas favoriser les animaux peu fertiles qui ont généralement des lactations longues.

Les quantités de matière sont obtenues simplement à partir du TB et TP moyens :

$$MG_{réf} = QL_{réf} \times TB \text{ et } MP_{réf} = QL_{réf} \times TP \quad (2.25)$$

Quand une lactation n'est pas terminée, mais a démarré plus de 90 jours avant l'indexation et compte au moins 60 jours, elle est extrapolée en une durée de 305 jours. L'extrapolation porte sur les quantités de lait, de MG et de MP ; les taux sont obtenus par le rapport des quantités extrapolées. La formule d'extrapolation (Auran et Mocquot, 1974) est de la forme :

$$Q_{extrapolée} = (a \times Q_{partielle}) + (b \times P_{dernier\ controle}) \quad (2.26)$$

où les coefficients  $a$  et  $b$  dépendent de la race, du stade de lactation et du rang de lactation. Dès qu'elles atteignent 180 jours, les lactations extrapolées interviennent avec la même pondération que les lactations terminées. Dans le cas contraire, leur pondération est moindre. Suivant les cas décrits dans le tableau 2.1, les lactations sont donc soit extrapolées soit corrigées pour la durée.

Tableau 2.1 : Synthèse des corrections appliquées au calcul de quantité de lait de référence suivant le rang de la lactation et son statut (en cours ou terminée)

Rang de lactation	En cours	Terminée	
		< durée de référence	> durée de référence
1 <sup>ère</sup> lactation	Extrapolation	Correction durée	Correction durée
2 <sup>ème</sup> lactation et +	Extrapolation	Extrapolation	Correction durée

L'utilisation d'une correction pour la durée dans le cas de lactation inférieure à la durée de référence est en désaccord avec les recommandations internationales. Toutefois, Barbat *et al.* (1995) ont montré que l'utilisation des formules d'extrapolation entraînait une surestimation des taureaux de valeur génétique médiocre sur la production laitière et qu'il était préférable de

corriger les données pour la durée. Ainsi, la production corrigée pour les lactations de plus de 270 jours est indépendante de la durée, alors que pour des durées inférieures elle reste positivement corrélée à la durée de la lactation (Bonaïti *et al.*, 1990). Aucune autre précorrection n'est faite en France. Les effets liés à l'âge, au mois de vêlage... sont directement pris en compte dans le modèle, comme le recommande Interbull.

### 1.2.2. La santé de la mamelle

Pour le calcul de la variable  $SCS_L$  c'est-à-dire le SCS moyen sur la lactation, l'ensemble des contrôles réalisés sur les vaches au cours de leurs 3 premières lactations et ayant un stade d'au plus 450 jours sont pris en compte. Le taux cellulaire sur 24 heures est celui mesuré par le laboratoire d'analyse quel que soit la durée effective d'échantillonnage correspondant au prélèvement de lait. Dans le cadre d'un protocole dit alterné, il n'y a donc pas de correction de la mesure pour les horaires de traites ou l'effet de dilution dans le lait.

Concernant le calcul de la  $SCS_L$ , la France (Rupp et Boichard, 1997) a fait le choix de suivre le cadre méthodologique présenté par Wiggans et Shook (1987). Le premier point consiste à transformer l'ensemble des données de comptages cellulaires (CCS) en un score de cellules somatiques afin de normaliser leur distribution à partir de la formule suivante :

$$SCS = \log_2 \left( \frac{CCS}{100\,000} \right) + 3 \quad (2.27)$$

Ainsi, un CCS de 12 500 équivaut à un score de 0, celui de 25 000 à un score de 1... toute augmentation du SCS de 1 correspondant au doublement de la valeur de CCS précédente. Le second point consiste à appliquer au SCS une correction additive pour le stade et rang de lactation estimée à partir de lactations saines (SCS inférieur à 4) lors de la mise en place de l'évaluation sur les comptages cellulaires en 1997. Ces données de SCS sont ensuite combinées pour obtenir une moyenne pondérée sur la lactation selon la formule :

$$SCS_L = \frac{\sum \frac{R}{\sigma} [SCS_{corrigé}]}{\sum \frac{R}{\sigma}} \quad (2.28)$$

où  $SCS_L$  la variable synthétisant les SCS sur l'ensemble de la lactation, R est la corrélation entre les SCS du stade et rang de lactation considérés et la moyenne des autres SCS du même rang de lactation et  $\sigma$  est l'écart-type des SCS à ce stade de lactation. Enfin, le dernier point consiste à préajuster les données  $SCS_L$  pour la variabilité liée au rang de lactation ( $SCS_L$  multiplié par 1.15 en 1<sup>ère</sup> lactation, par 1.08 en 2<sup>ème</sup> lactation et 1.00 en 3<sup>ème</sup> lactation).

Le poids accordé à chaque  $SCS_L$  est fonction du nombre de comptages cellulaires pris en compte dans son calcul. Un minimum de 5 contrôles est nécessaire pour obtenir un poids de 1.

## 2. L'évaluation génétique

Cette étape concerne l'évaluation génétique *sensu stricto* ainsi que certains points en lien direct avec celle-ci tels que la définition du modèle d'évaluation ou l'estimation des paramètres génétiques. Cette étape est jugée être la plus sensible de l'ensemble du processus d'évaluation génétique. Les choix portant sur la méthode d'évaluation, le modèle d'évaluation et les paramètres génétiques utilisés ont des implications fortes sur les classements nationaux



et internationaux. Les recommandations émises par Interbull pour les caractères laitiers peuvent être étendues à l'indicateur de l'état sanitaire de la mamelle que sont les comptages cellulaires.

## 2.1. Les recommandations

Le premier point sur lequel des divergences peuvent exister au niveau mondial concerne les effets qui doivent être pris en compte dans le modèle. Le modèle retenu doit être un compromis entre la simplicité d'une part et le niveau d'ajustement d'autre part. Toutefois, la simplification consistant à précorriger les données pour certains des effets dans l'étape précédente doit, dans la mesure du possible, être évitée, car elle est source de biais dans l'évaluation génétique. En effet, l'une des propriétés majeures du BLUP est l'obtention de prédictions non biaisées grâce à l'estimation simultanée des effets fixes et aléatoires. De plus, cette prise en compte des effets directement dans le modèle permet aussi d'éviter les problèmes liés à la mise à jour de ces facteurs. De nombreux effets sont jugés avoir un impact sur la production. Parmi eux, nous comptons le troupeau, l'année, la saison, le rang de lactation, l'âge au vêlage, le mois de vêlage, l'intervalle entre vêlage, l'intervalle vêlage – IA fécondante... Une des difficultés consiste à savoir s'il s'agit d'effets « fixes » dont l'impact sera le même sur l'ensemble des animaux, ou au contraire d'effets « aléatoires ». Classiquement, les effets cités précédemment sont pris en compte en tant qu'effets fixes dans le modèle, mais souvent sous forme d'interaction telle que la race – région – année – âge au vêlage afin d'en obtenir une bonne caractérisation. Dans de pareils cas, une des recommandations d'Interbull est de s'assurer que l'effectif de chacune des classes ainsi formées est suffisant pour que son effet soit estimé avec une précision satisfaisante, sans toutefois donner plus d'indication (Interbull, 2001) compte tenu de la grande variabilité des effectifs suivant les pays et les races analysées.

Le second point responsable de différences entre les pays est la description de la partie génétique dans le modèle. L'ensemble des pays participant aux évaluations internationales utilise un modèle animal en remplacement du modèle père utilisé avant les années 90. Interbull recommande d'utiliser un modèle basé sur des lactations multiples plutôt que sur la seule première lactation. Les pays ayant un modèle « lactations multiples » employaient majoritairement (2/3 des pays dans l'enquête datant de 2000 (Interbull, 2001)) une approche prenant en compte la répétabilité du caractère entre les différentes lactations.

Parmi les aspects ayant un impact sur les évaluations internationales, nous pouvons aussi ajouter la prise en compte de groupes de parents inconnus et d'hétérogénéités de variance dans le modèle. Ainsi, Interbull juge essentiel la prise en compte des différences entre les populations d'origine à travers la formation de groupes génétiques homogènes et ayant un effectif suffisant. Par ailleurs, de nombreuses études (Bonaïti, 1982 ; Hill *et al.*, 1983 ; Brotherstone and Hill, 1986 ; Meuwissen and Van der Werf, 1993 ; Robert, 1996...) ont montré l'existence de variances hétérogènes des effets génétique, d'environnement permanent et résiduel pour les caractères de production laitière. Auparavant, la variabilité des performances était considérée être la même à l'intérieur de toutes les modalités de milieu, par

exemple dans l'ensemble des troupeaux quelle que soit leur région. Supposer les variances homogènes avait donc des conséquences sur le classement des animaux, et plus particulièrement les femelles en faveur de celles issues de troupeaux présentant une forte variabilité. Des méthodes permettant la prise en compte des variances hétérogènes ont donc été développées. Par exemple, Wiggans et VanRaden (1991) ont proposé de précorriger les données préalablement à l'évaluation génétique pour tenir compte de l'hétérogénéité de la variance phénotypique intra-troupeau. Mais, comme dans le cas de la précorrection pour les effets fixes, ces approches consistant à précorriger les données ne sont pas jugées optimales. Des méthodes, telles que celle de Meuwissen *et al.* (1996) ou Robert-Granié *et al.* (1999), intégrant la correction pour les variances hétérogènes directement dans le modèle d'évaluation sont plus satisfaisantes.

Afin de garantir l'absence de biais dans les évaluations génétiques, la méthode employée doit reposer sur des estimateurs non biaisés et le modèle statistique décrivant les données doit être adapté. La condition portant sur la méthode ne pose pas de problème puisqu'il s'agit d'une des propriétés essentielles du BLUP. Par contre, pour contrôler les possibles biais du modèle, des méthodes de validation tels que celles de Boichard *et al.* (1995) ou basées sur l'analyse des aléas de méiose (Miglior *et al.*, 2002) ont été proposés. L'absence de biais détectés par les trois méthodes de Boichard *et al.* (1995) est un pré requis pour la participation d'un pays à une évaluation internationale pour un groupe de caractère donné. La première méthode se base sur la comparaison des évaluations réalisées à partir de l'ensemble des lactations *vs* les premières lactations. La seconde méthode repose sur l'analyse des déviations moyennes des filles par taureau et par année de vêlage et enfin la troisième méthode compare les variations entre évaluations successives.

L'estimation des paramètres génétiques constitue un autre point essentiel de cette étape. Ceux-ci doivent théoriquement être estimés de façon régulière, i.e. d'après Interbull au moins une fois par génération suivant un modèle similaire à celui de l'évaluation génétique (Interbull, 2001).

## 2.2. Le cas de la France

Afin de garantir la qualité de ses évaluations génétiques bovines auprès des utilisateurs, l'INRA s'est engagé dans une démarche qualité qui a abouti à la certification de l'ensemble de ses activités d'évaluation génétique suivant une norme ISO 9001 en janvier 2006.

Actuellement, le modèle d'évaluation français des taureaux et des vaches à partir des performances laitières est un modèle animal unicaractère basé sur les 3 premières lactations de l'animal calculées selon la méthode décrite précédemment avec prise en compte de la répétabilité entre les lactations. Ce même modèle est appliqué aux données sur les cellules somatiques. La production de référence de l'animal  $i$  est décrit comme :

$$y_i = \sum f_i + g_i + p_i + e_i \quad (2.29)$$

Où  $y_i$  est la performance de l'animal  $i$ ,

- $\sum f_i$  est la somme des effets de milieu identifiés et enregistrés (effet fixes) de l'animal  $i$ . Cet effet inclut un effet du troupeau intra année ; le rang de lactation intra région et année ; le mois de vêlage intra rang de lactation (première *vs* autres lactations), région et année ; l'âge au vêlage intra rang de lactation, région et année ; et pour les multipares, la durée de tarissement intra région et année,
- $g_i$  est l'effet génétique additif de l'animal  $i$  où  $g \sim N(0, \sigma_g^2)$ ,
- $p_i$  est l'effet d'environnement permanent, c'est-à-dire l'effet propre de l'animal  $i$  hors effet génétique additif où  $p \sim N(0, \sigma_p^2)$ ,
- $e_{ij}$  est la valeur résiduelle de la performance de l'animal  $i$  qui correspond à l'ensemble des effets non identifiés ou non enregistrables du modèle  $e \sim N(0, \sigma_e^2)$ .

Depuis 1999, le modèle (à l'exception des cellules somatiques) suppose des variances résiduelles hétérogènes et les rapports de variances (héritabilité et répétabilité) constants. Ces derniers ont respectivement pour valeur 0.30 et 0.50 pour les quantités de lait, matière grasse et matière protéique, 0.50 et 0.70 pour les taux butyreux et protéique et 0.15 et 0.35 pour la numération cellulaire. Robert-Granié *et al.* (1999) ont décrit la méthodologie employée en France suivant les travaux de Foulley *et al.* (1992) et San Cristobal *et al.* (1993) (modèle linéaire mixte structural des logarithmes des variances résiduelles avec un effet fixe région – année et un effet aléatoire troupeau – année où l'autocorrélation intra-troupeau entre les différentes années est modélisée). Dans cette étude, les auteurs ont montré l'impact de la prise en compte des variances hétérogènes sur le classement des animaux. En effet, même si cette inclusion dans le modèle n'a engendré aucun changement notable dans l'indexation des mâles compte tenu de la répartition de leurs filles dans un grand nombre de troupeaux, de régions et d'années, la situation était différente chez les femelles. Les meilleures vaches des troupeaux présentant une forte variabilité dans leur conduite ont vu leur valeur génétique régressée vers la moyenne de la population. Au contraire, les troupeaux ayant une conduite très homogène de leurs animaux ont vu leurs meilleures vaches progresser dans le classement national. Cette méthode d'évaluation a largement permis de faire progresser la fiabilité des index.

### 3. Les étapes postérieures à l'évaluation génétique

Cette étape concerne les activités relatives à la mise en forme des valeurs génétiques estimées et à leurs diffusions auprès des acteurs de la filière professionnelle : entreprise de sélection, éleveurs... Contrairement aux précédentes étapes, cette dernière n'a aucun impact sur les comparaisons internationales. Elle joue seulement un rôle au niveau de la compréhension et de la transparence pour les utilisateurs.

#### 3.1. Les recommandations

Les règles de publications des index font parties des éléments importants de cette étape. Suivant les pays et les caractères évalués, le niveau de précision requis, aussi appelé coefficient de détermination (CD), peut varier fortement (généralement entre 0.50 et 0.75).

Celui-ci est défini comme le carré du coefficient de corrélation entre la valeur génétique additive vraie et son estimation. Pour les caractères laitiers, le nombre de filles requises pour voir l'index du père diffusé peut ainsi varier entre 5 et 100 suivant le pays (Interbull, 2008). De même, le nombre minimum de troupeaux varie de 1 à 20. En pratique, Interbull accepte dans les évaluations internationales les informations de taureaux de testage testés dans un minimum de 10 troupeaux et de taureaux de service (import) ayant au minimum 75 filles dans 50 troupeaux en race Holstein (moins dans les autres races).

Le mode d'expression des index est aussi un sujet sur lequel les pays ne parviennent pas à s'homogénéiser, chacun souhaitant conserver son mode d'expression pour garder ses repères historiques. Ainsi, VanRaden (2004) énumère 4 modes d'expression : valeur génétique estimée, valeur génétique relative avec une moyenne fixée et un écart-type variable, valeur génétique standardisée avec une moyenne généralement nulle et un écart-type de 1 ainsi que la valeur génétique transmise (soit la moitié de la valeur génétique de l'animal). Interbull recommande l'utilisation des valeurs génétiques estimées exprimées dans le système métrique international.

L'expression des index à travers une base génétique est un autre point sujet à des différences entre pays. La base peut être constituée d'une population mâle ou femelle de référence ; elle peut changer à toutes les évaluations, tous les ans ou tous les 5 ans. Afin d'harmoniser ces pratiques, Interbull recommande d'établir la base à partir de la population femelle qui a le mérite d'être numériquement plus conséquente et donc plus stable, de faire des changements de bases tous les cinq ans, les années se terminant en 0 et en 5, avec comme base la population née 5 ans auparavant. Selon ces recommandations, la base 2005 utilisée depuis 2005 est constituée des vaches nées en 2000.

### **3.2. Le cas de la France**

Pour la diffusion des évaluations génétiques des mâles, une précision minimale de 0.50 est exigée réglementairement au niveau de l'Europe quel que soit le caractère évalué. Toutes valeurs génétiques d'un animal franchissant ce seuil peuvent donc faire l'objet de diffusion. Cependant, pour mettre en avant l'importance de la fiabilité des index, un critère plus contraignant, le « seuil FGE », a été mis en place pour les caractères laitiers. Seuls les animaux franchissant ce seuil bénéficient d'une large diffusion à travers les palmarès et ont un index diffusable dans le cadre des évaluations internationales. Ce seuil est de 0.70 et de 40 filles dans 10 étables en Holstein ; 0.70 et 40 filles dans 30 étables en Montbéliarde et Normande ; 0.60 et de 20 filles pour les autres races. Pour la numération cellulaire, la précision nécessaire à la diffusion de l'index est de 0.50 quelle que soit la race (Institut de l'Elevage-INRA, 2007). Chez les femelles, la précision minimale exigée est de 0.30 pour l'ensemble des caractères.

En ce qui concerne le mode d'expression des index, la France a choisi d'exprimer ses valeurs génétiques en kilogrammes de Lait et de matières, en % pour les taux et en unité d'écart-type génétique pour la numération cellulaire. Ces index sont exprimés relativement à une base mobile fonction du sexe de l'animal, changeant tous les ans (lors de l'indexation de juin jusqu'en 2007 et en février à partir de 2009). Pour les mâles des 3 principales races, la base

prend en compte l'ensemble des taureaux nés entre n-10 et n-7 inclus, alors que cette période est élargie de n-12 à n-7 pour les races dites régionales. Pour les femelles, les années considérées sont n-6 pour les principales races et n-5 à n-7 pour les autres. Pour la numération cellulaire, le signe de l'index est inversé de sorte qu'un taureau ayant un faible comptage cellulaire soit indexé positivement.

Dans les modèles d'évaluation « lactation », généralement basés sur les modèles dits à répétabilité, les mesures répétées sont considérées être l'expression d'un même caractère dans le temps. En d'autres termes, une corrélation génétique de 1 est supposée entre les différentes lactations d'un même animal. Ce type de modèle a largement été employé pour l'évaluation des caractères laitiers comme en témoigne le bulletin d'Interbull N°24 (2000). Ses principaux avantages sont sa simplicité, ses faibles besoins de calcul et le faible nombre de paramètres requis en comparaison d'un modèle multivarié. Cependant, ce modèle présente de nombreuses limites. D'un point de vue méthodologique, la principale est l'hypothèse de complète dépendance entre observations. En effet, plusieurs études ont compromis cette hypothèse : Intra lactation, l'héritabilité de la production journalière varie suivant le stade de lactation (Danell, 1982 ; Pander *et al.*, 1992 ; ...) et les corrélations génétiques entre les mesures répétées tendent généralement à décroître avec l'augmentation de l'intervalle entre celles-ci (Pander *et al.*, 1992 ; Rekaya *et al.*, 1995). De plus, entre lactations, les corrélations obtenues (entre 0.88 et 1.00 (Meyer, 1985)) sont loin d'appuyer l'hypothèse que la production 305 jours est la mesure d'un seul et même caractère quel que soit le rang de lactation. Ce type de données nécessite donc le recours à une méthodologie adaptée, permettant d'analyser les modifications individuelles au cours du temps et ainsi d'exploiter pleinement les observations recueillies.

## **Partie 2.3**

# **Les évaluations génétiques de données longitudinales**

Les données de production laitière sont des données dites longitudinales, c'est-à-dire provenant de mesures sur un même individu répétées dans le temps. L'étude de ces données peut être très puissante, car elle permet d'analyser les modifications des performances d'un individu au cours du temps. L'analyse portant sur des observations répétées sur un même animal, il est nécessaire de prendre en compte la notion de répétabilité des performances, dans la mesure où ces observations ne sont pas indépendantes les unes des autres.

Dans un contexte similaire, chez les bovins allaitants, les mesures répétées de la croissance ont été analysées différemment, avec l'hypothèse que les mesures à des âges différents étaient des caractères différents mais génétiquement corrélés, et donc analysées avec un modèle multicaractères. Bien que le modèle multicaractères soit une sophistication par rapport au modèle à répétabilité grâce à la prise en compte des corrélations génétiques entre les différentes mesures, il peut être grandement surparamétrisé si les mesures sont faites de façon fréquente. De plus, pour un modèle multicaractères sur une mesure fréquente, il serait difficile d'obtenir des estimations précises des paramètres génétiques.

Un modèle approprié pour l'étude de données répétées dans le temps ou avec l'âge doit prendre en compte la moyenne et les changements au cours du temps de la structure des covariances tout en permettant l'estimation des paramètres génétiques requis. En 1994, Schaeffer et Dekkers ont introduit le concept de régression aléatoire (RR pour « Random Regression ») pour l'analyse des données de contrôles journaliers chez les bovins laitiers comme un moyen de prendre en compte la structure de covariance entre les mesures répétées dans le temps. A peu près à la même époque, Kirkpatrick *et al.* (1990, 1994) introduisaient le concept de fonction de covariance (CF pour « Covariance Function ») pour l'analyse de données longitudinales, illustrant cette méthodologie à partir de données de croissance. Après avoir présenté les différents modèles utilisés pour l'analyse de ces données, leurs potentielles applications seront développées. Une seconde partie sera consacrée à la présentation, dans le contexte de l'évaluation laitière, des divers effets inclus dans le modèle contrôles élémentaires.

## **1. Modélisation des données longitudinales**

Pour le moment, les modèles traitant les données longitudinales ont principalement concerné les données de production laitière, classiquement mesurées mensuellement. De ce fait, les appellations des modèles comportent généralement le terme « contrôles élémentaires » (TDM pour « Test-Day Model ») et l'ensemble des modèles décrits ultérieurement se réfère aux caractères de production laitière. Toutefois, comme nous pourrions le voir dans la troisième partie de cette section 1., les applications de ces modèles peuvent être beaucoup plus larges.

### **1.1. Les différents modèles contrôles élémentaires (TDM)**

Il existe de nombreuses variantes des modèles contrôles élémentaires, témoignant de l'évolution des recherches au cours de 15 dernières années. Tout d'abord, suivant la nomenclature de Swalve (2000), nous pouvons distinguer deux stratégies d'évaluation TDM : les modèles en deux étapes *vs* ceux en une étape. Les modèles en deux étapes sont en fait des variantes du modèle lactation. Une 1<sup>ère</sup> étape consiste à évaluer les effets environnementaux

sur la base des contrôles journaliers. Les productions journalières sont ensuite corrigées pour ces effets environnementaux afin d'obtenir des performances ajustés et sont combinées pour déterminer une performance résumant la lactation, en tenant compte dans certains pays (Nouvelle-Zélande) de pondérations pour les différents contrôles en fonction des corrélations entre ceux-ci. Si la lactation a une durée inférieure à 305 jours, des facteurs d'extension sont nécessaires. Dans une 2<sup>nd</sup>e étape, ces données sont utilisées dans un modèle lactation. La différence par rapport au modèle lactation traditionnel réside dans le préajustement préalable des données avant les éventuelles corrections pour la durée. Cette méthodologie a principalement été utilisée (et l'est parfois encore) en Australie (Jones and Goddard, 1990), en Nouvelle-Zélande (Johnson, 1996) et dans le Nord-Est des Etats-Unis (Université Cornell, Everett *et al.*, 1994).

Deux principaux contextes d'utilisation de cette méthode sont recensés (Swalve, 2000) :

- Les pays en transition qui n'ont pas conservé les performances journalières depuis suffisamment longtemps et ne peuvent donc pas se lancer directement dans une évaluation sur la base de contrôles élémentaires. Ces pays ont recours à ce type de modèle pour combiner les données « lactations » anciennes aux variables résumant les lactations établies à partir des contrôles élémentaires.
- Les pays dont la population de vaches contrôlées est trop importante pour inclure les données élémentaires dans un TDM. Ces pays utilisent cette approche pour optimiser leur évaluation génétique et valoriser au mieux les informations disponibles à travers les performances journalières.

L'évaluation américaine réalisée par l'université Cornell ajuste les enregistrements journaliers pour les effets de conduite intra-troupeau, d'âge, de stade de lactation, de mois de vêlage, et de gestation avant de les combiner en une production en 320 jours. Cette approche et, par une extension douteuse, l'utilisation des modèles « jours de contrôles » a été brevetés par cette université pour certains pays. Ceci a jusqu'à présent empêché les Etats-Unis d'opter pour un tel modèle, préférant attendre l'expiration du brevet en 2014.

Ces modèles restent très similaires aux modèles dits « à répétabilité » utilisés dans le modèle d'évaluation génétique lactation, et permettent seulement une meilleure prise en compte de l'environnement spécifique du jour du contrôle. Les limites concernant les corrélations entre lactations et entre contrôles évoquées précédemment ne sont généralement pas prises en compte. D'après Jensen (2001), les modèles en deux étapes devraient théoriquement tenir compte, lors du calcul de la donnée résumant la lactation, des différences de précision entre les données élémentaires, comme c'est le cas en Nouvelle-Zélande.

La manière la plus aisée de tenir compte des corrélations existant entre contrôles est de recourir à un modèle en une étape estimant simultanément l'ensemble des effets du modèle.

### **1.1.1. Les modèles mixtes à régression fixe (FR-TDM)**

Le cadre théorique du modèle mixte à régression fixe et son application à l'analyse des données longitudinales telles que les contrôles journaliers de production laitière a été présenté par Meyer *et al.* (1989) ainsi que Ptak et Schaeffer (1993).



De façon similaire au modèle avec répétabilité (les principes des modèles de bases de l'évaluation génétique sont consultables en annexe), le modèle de régression fixe suppose pour l'effet génétique que les performances journalières intra-lactation sont des mesures répétées du même caractère. Ainsi, si  $g_i(t)$  est la valeur génétique de l'animal  $i$  au temps  $t$ , la corrélation  $\rho(g_i(t), g_i(t'))$  est égale à 1 pour tout  $t, t'$ . Généralement, l'effet d'environnement permanent inclus dans le modèle prend en compte les facteurs environnementaux ayant un effet permanent sur l'ensemble des contrôles de la lactation ( $\rho(p_i(t), p_i(t'))=1$ ). Avec l'extension à un modèle multi-lactation, les contrôles élémentaires sont considérés comme des mesures répétées intra-lactation mais comme des caractères différents entre lactations.

Dans le modèle de régression fixe (FR pour Fixed Regression), la performance est décrite comme :

$$y_{ijkt} = HTD_j + \sum_{n=0}^{nf-1} \phi_{knt} \beta_{kn} + g_i + p_i + e_{ijkt} \quad (2.30)$$

- où  $y_{ijkt}$  est le contrôle journalier de l'animal  $i$ , à un stade de lactation  $t$ , avec un effet troupeau  $j$  et appartenant au groupe de référence  $k$ ,  
 $HTD_j$  est l'effet fixe troupeau – jour de contrôle  $j$   
 $\phi_{knt}$  est le  $n^{ième}$  des  $nf$  paramètres de la courbe de lactation (par exemple, du polynôme de Legendre) correspondant au groupe de référence  $k$  de l'animal  $i$  pour un stade de lactation  $t$ ,  
 $\beta_{kn}$  est le  $n^{ième}$  coefficient de la régression fixe du groupe de référence  $k$ ,  
 $g_i$  est l'effet génétique transmissible de l'animal  $i$ ,  
 $p_i$  est l'effet d'environnement permanent de l'animal  $i$ ,  
 $e_{ijkt}$  est la valeur résiduelle de la performance de l'animal  $i$ , obtenue au stade de lactation  $t$ , avec un effet troupeau  $j$  et appartenant au groupe de référence  $k$ .

Sous forme matricielle, la performance peut être écrite :

$$\mathbf{y} = \mathbf{X} \mathbf{b} + \mathbf{Z} \mathbf{a} + \mathbf{W} \mathbf{p} + \mathbf{e} \quad (2.31)$$

- où :  $\mathbf{y}$  est le vecteur des observations élémentaires,  
 $\mathbf{b}$  est le vecteur des effets fixes comprenant l'effet Troupeau – Jour de contrôle et les coefficients de régression fixe,  
 $\mathbf{a}$  est le vecteur des valeurs génétiques additives de tous les animaux avec ou sans performances,  
 $\mathbf{p}$  est le vecteur des valeurs d'effet permanent de tous les animaux avec performances,  
 $\mathbf{e}$  est le vecteur des résiduelles, distribuées suivant une loi normale de moyenne nulle,  
 $\mathbf{X}, \mathbf{Z}$  et  $\mathbf{W}$  sont les matrices d'incidence qui relient les observations aux effets fixes et aléatoires qui les ont influencées.

- où :  $E(\mathbf{a}) = \mathbf{0}$        $\text{Var}(\mathbf{a}) = \mathbf{A} \sigma_a^2$       où  $\mathbf{A}$  est la matrice de parenté,  
 $E(\mathbf{p}) = \mathbf{0}$        $\text{Var}(\mathbf{p}) = \mathbf{I} \sigma_p^2$       où  $\mathbf{I}$  est la matrice identité,  
 $E(\mathbf{e}) = \mathbf{0}$        $\text{Var}(\mathbf{e}) = \mathbf{I} \sigma_e^2$ .

Les équations du modèle mixte s'écrivent alors :

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} & \mathbf{X}'\mathbf{W} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \alpha\mathbf{A}^{-1} & \mathbf{Z}'\mathbf{W} \\ \mathbf{W}'\mathbf{X} & \mathbf{W}'\mathbf{Z} & \mathbf{W}'\mathbf{W} + \gamma\mathbf{I} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{p}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \\ \mathbf{W}'\mathbf{y} \end{bmatrix} \quad (2.32)$$

Où  $\alpha = \sigma_e^2 / \sigma_a^2$  et  $\gamma = \sigma_e^2 / \sigma_p^2$ .

Le modèle de régression fixe inclut une fonction du stade de lactation pour modéliser la courbe de lactation, élément essentiel du modèle contrôles élémentaires. Un compromis doit être trouvé pour la constitution des groupes définissant les courbes de lactation, de sorte que ces derniers soient suffisamment fins pour éviter des biais mais aussi suffisamment conséquents pour permettre une estimation précise des coefficients de la courbe de régression (Ptak and Schaeffer, 1993). Ces auteurs recommandent au minimum la prise en compte de l'âge et de la saison de vêlage dans la définition des groupes. Dans le modèle allemand utilisé pour les évaluations génétiques nationales des caractères de production et de santé de la mamelle entre 1995 et 2002, les coefficients de régression de 7560 courbes étaient ainsi estimés (2 classes de rang de lactation x 9 régions – race x 5 classes d'âge au vêlage x 3 saisons de vêlage x 7 classes d'intervalle entre vêlage x 4 groupes d'années de vêlage).

Selon Jamrozik et Schaeffer (1997b) ainsi que Jensen (2001), une extension intéressante du modèle de régression fixe serait d'inclure l'effet du troupeau dans la modélisation de la courbe de lactation afin de prendre en compte des différences liées au système de conduite d'élevage et notamment à l'alimentation.

Ptak et Schaeffer (1993) et Jamrozik *et al.* (1997b) soulignent l'importance de la prise en compte de l'hétérogénéité de la variance résiduelle dans le modèle. Dans les modèles proposés, la variance résiduelle varie suivant le stade de lactation. Toutefois, dans ces études, elle n'intègre pas les sources de variations de la variance résiduelle telles que le troupeau ou l'année de production dont l'importance a notamment été montrée par Wiggans et VanRaden (1991), Meuwissen *et al.* (1996) ainsi que Robert-Granié *et al.* (1999) et qui sont désormais largement intégrées dans le modèle lactation.

L'inclusion d'une régression fixe sur le stade de lactation dans le modèle permet la prise en compte de la forme de la courbe de lactation liée au groupe de référence de l'animal. Bien qu'il soit possible de modéliser la variance résiduelle en fonction du stade de lactation dans le cas de régression fixe, le modèle ne permet pas de modéliser l'effet génétique suivant une structure de covariance donnée. Les valeurs génétiques influent donc uniquement sur le niveau des courbes, et en aucun cas sur la forme. En cela, l'effet génétique dans le modèle de régression fixe est modélisé de façon similaire à un modèle dit « avec répétabilité » comme le rappelle Jensen (2001). La persistance, caractère d'intérêt pour les éleveurs laitiers, ne peut donc pas être estimée à partir d'un tel modèle. Pour répondre à ce besoin, il est nécessaire d'inclure dans le modèle une fonction du stade de lactation pour la valeur génétique de chaque animal, c'est-à-dire une régression aléatoire.

### 1.1.2. Les modèles avec régression aléatoire (RR-TDM)

Schaeffer et Dekkers (1994) ont étendu le modèle FR-TDM pour permettre d'ajuster spécifiquement la trajectoire de la courbe de lactation à l'animal en incluant une régression aléatoire sur l'effet génétique. La courbe de lactation de l'animal est ainsi définie à partir de deux séries de régressions sur le stade de lactation : une régression fixe décrivant la forme générale de la courbe de lactation en fonction du groupe de référence (par exemple région, âge au vêlage, saison de vêlage) et une régression aléatoire décrivant la déviation par rapport à la courbe de référence liée à l'effet génétique. Ceci permet d'avoir une modulation du niveau et de la forme de la courbe de lactation dépendant de l'animal, ce qui offre la possibilité d'évaluer génétiquement la persistance d'une lactation. La régression aléatoire permet aussi de structurer la variance entre observations.

L'utilisation des modèles de régression aléatoire (RR pour Random Regression) avait été proposée par Henderson Jr (1982). Les premiers modèles RR utilisaient des covariables fortement corrélées entre elles ce qui ralentissait la convergence vers les solutions du modèle mixte. Malgré cela, les résultats obtenus étaient prometteurs (Schaeffer and Dekkers, 1994). Pour remédier à cette difficulté, de nombreux travaux ont été consacrés à l'étude de la modélisation des courbes et aux covariables utilisées, éléments essentiels des modèles contrôles élémentaires. Ainsi, les effets modélisés sous forme de courbes, qu'ils soient fixes ou aléatoires, peuvent l'être à partir de courbes paramétriques (Guo and Swalve, 1997), de polynômes orthogonaux tels que les polynômes de Legendre (Brotherstone *et al.*, 2000), de courbes non paramétriques, semi-paramétriques telles que les splines cubiques (White *et al.*, 1999) ou toutes autres fonctions telles que celles décrites dans la partie 2.1. Même s'il n'est nullement nécessaire de modéliser les différents effets fixes et aléatoires avec des fonctions similaires, il s'avère que la modélisation des formes de courbes de lactation à partir des fonctions les plus simples, dépendant directement du stade de lactation, a quasiment été abandonnée au profit des polynômes orthogonaux. Outre l'avantage de ne faire aucune hypothèse sur la forme des courbes, ces derniers présentent en effet des corrélations entre coefficients de régressions plus faibles et les matrices de covariance estimées tendent à être plus stables vis-à-vis des différents fichiers utilisés pour les estimer.

Schaeffer et Dekkers (1994) recommandent l'inclusion d'un effet d'environnement permanent sous forme de régression aléatoire pour permettre de tenir compte des corrélations des effets non génétiques propres à l'animal entre les différents contrôles d'une même lactation, voire entre les différentes lactations. Jamrozik *et al.* (1997b) vont encore plus loin en mentionnant la possibilité d'inclure une régression fixe ou aléatoire sur l'effet troupeau – année pour tenir compte des effets liés à la conduite du troupeau sur la forme de la courbe de lactation des vaches du troupeau.

Concernant l'ordre des polynômes nécessaire à l'ajustement du modèle de régression aléatoire, Pool *et al.* (2000) ont montré que l'effet génétique additif nécessitait un ordre inférieur à celui de l'environnement permanent. Cependant des études sont nécessaires pour déterminer l'ordre optimum des régressions expliquant la plus grande part de la variation entre les observations. En 2004, Schaeffer expliquait que les travaux en cours sur le modèle Canadien concernaient cet aspect.

Pour les contrôles journaliers chez les bovins laitiers, la performance est de plus en plus classiquement décrite comme :

$$y_{ijkt} = HTD_j + \sum_{n=0}^{nf-1} \phi_{knt} \beta_{kn} + \sum_{n=0}^{nr_g-1} \varphi_{int} g_{in} + \sum_{n=0}^{nr_p-1} \psi_{int} p_{in} + e_{ijkt} \quad (2.33)$$

où  $y_{ijkt}$  est le contrôle journalier de l'animal  $i$ , à un stade de lactation  $t$ , avec un effet troupeau  $j$  et appartenant au groupe de référence  $k$ ,

$HTD_j$  est l'effet fixe troupeau – jour de contrôle  $j$

$\phi_{knt}$  est le  $n^{ième}$  des  $nf$  paramètres de la courbe des effets fixes correspondant au groupe de référence  $k$  de l'animal  $i$  pour un stade de lactation  $t$ ,

$\beta_{kn}$  sont le  $n^{ième}$  coefficient de la régression fixe du groupe de référence  $k$ ,

$\varphi_{int}$  est le  $n^{ième}$  des  $nr_g$  paramètres de la courbe des effets génétiques de l'animal  $i$  pour un stade de lactation  $t$ ,

$g_{in}$  est le  $n^{ième}$  effet génétique transmissible de l'animal  $i$ ,

$\psi_{int}$  est le  $n^{ième}$  des  $nr_p$  paramètres de la courbe des effets d'environnement permanent de l'animal  $i$  pour un stade de lactation  $t$ ,

$p_{in}$  est le  $n^{ième}$  effet d'environnement permanent de l'animal  $i$ ,

$e_{ijkt}$  est la valeur résiduelle de la performance de l'animal  $i$ , obtenue au stade de lactation  $t$ , avec un effet troupeau  $j$  et appartenant au groupe de référence  $k$ .

Sous forme matricielle, la performance est écrite comme précédemment :

$$\mathbf{y} = \mathbf{X} \mathbf{b} + \mathbf{Z} \mathbf{a} + \mathbf{W} \mathbf{p} + \mathbf{e} \quad (2.34)$$

où :  $\mathbf{y}$  est le vecteur des observations élémentaires,

$\mathbf{b}$  est le vecteur des effets fixes comprenant l'effet Troupeau – Jour de contrôle et les coefficients de régression fixes,

$\mathbf{a}$  est désormais le vecteur des coefficients des régressions aléatoires des effets génétiques additives de l'ensemble des animaux,

$\mathbf{p}$  est le vecteur des coefficients des régressions aléatoires des effets d'environnement permanent des animaux avec performance,

$\mathbf{e}$  est le vecteur des résiduelles, distribuées suivant une loi normale de moyenne nulle,

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

$$\text{avec } \begin{pmatrix} \mathbf{a} \\ \mathbf{p} \\ \mathbf{e} \end{pmatrix} \sim N(\mathbf{0}, \mathbf{V}) \quad \text{où } \mathbf{V} = \begin{pmatrix} \mathbf{A} \otimes \mathbf{G} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \otimes \mathbf{P} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R} \end{pmatrix}$$

$\mathbf{G}$  est une matrice ayant une dimension égale à  $nr_g$  et  $\mathbf{P}$  égale à  $nr_p$ .

Les équations du modèle mixte s'écrivent alors :

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{W} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{A}^{-1} \otimes \mathbf{G}^{-1} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{W} \\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{W}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \mathbf{I} \otimes \mathbf{P}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{p}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \quad (2.35)$$

La modélisation de l'effet génétique et potentiellement d'autres effets aléatoires tels que l'environnement permanent dans un modèle RR entraîne des besoins calculatoires en forte hausse par rapport à un modèle FR-TDM. Certains chercheurs se sont interrogés sur la pertinence de modéliser la valeur génétique sous forme d'une courbe plutôt que par une constante, comme c'est le cas dans un modèle FR-TDM. Olori (1997) cité par Swalve (2000) estime qu'il n'y a pas (ou peu) de variation individuelle, un terme constant expliquant à lui seul 93% de la variabilité génétique. Au contraire, Van der Werf *et al.* (1998) estiment que la part de variation de la seule persistance est de 11%, et nécessite donc une prise en compte à travers un modèle RR-TDM. La question sous-jacente à cette interrogation concerne le nombre de paramètres à estimer, qui dépend largement de la complexité du modèle.

Kirkpatrick *et al.* (1990) ont présenté une méthode générale pour l'analyse génétique de données répétées, mesurées sur une échelle continue telle que le temps ou l'âge. Cette méthode repose sur la notion de fonction de covariance et est particulièrement adaptée aux caractères ayant des changements continus comme les données de croissance, ou la production laitière journalière (Kirkpatrick *et al.*, 1994). Le nombre de mesures décrivant ces caractères peut être conçu comme étant de dimension infinie. La structure de covariance d'un caractère ayant une dimension infinie est, dans ce type de modèle, définie comme une fonction du temps. Par conséquent, la fonction de covariance est un équivalent de dimension infinie d'une matrice de covariance pour un nombre donné d'enregistrements recueillis au cours du temps à différents âges. Cette méthode s'appuie sur l'utilisation de polynômes orthogonaux tels que les polynômes de Legendre pour estimer les fonctions de covariance à partir des matrices de covariance observées. A travers le travail d'extension de cette méthode à l'estimation directe des fonctions de covariance à partir des données avec une méthode REML (Restricted Maximum Likelihood : maximum de vraisemblance restreinte), Meyer et Hill (1997) ainsi que Van der Werf *et al.* (1998) ont montré que cette méthode était équivalente aux modèles RR-TDM lorsque que ces méthodes utilisaient des modélisations similaires (polynôme de Legendre du même ordre).

Un élément ennuyeux dans les modèles RR utilisant les polynômes de Legendre est leur tendance en début et fin de lactation à donner lieu à des estimations de variances génétiques très supérieures (ou très inférieures) à celles du milieu de la lactation, aussi appelé « effets de bord ». Les héritabilités résultantes sont donc plus élevées sur ces périodes. Pour répondre à cette limite, une des propositions consiste à utiliser des splines. Certains ont proposé d'utiliser un modèle multicaractères, où chaque période de 30 jours est considéré comme un caractère différent, à la place de polynômes de Legendre (Rekaya *et al.*, 1999). Cependant, dans un modèle multicaractères, le nombre de paramètre est vite très important.

Un autre aspect portant préjudice aux modèles contrôles élémentaires est l'instabilité des valeurs génétiques estimées dans le temps. En effet, l'importance des différences obtenues entre les valeurs génétiques estimées par lactation accrédite l'hypothèse selon laquelle les différentes lactations constituent des caractères distincts (Schaeffer *et al.*, 2000). Cependant la prise en compte de cette information dans le modèle, même si elle est le reflet de la réalité

biologique, entraîne des variations d'index importantes lors de la prise en compte de nouvelles données dans le modèle, notamment à l'arrivée des 2<sup>èmes</sup> puis des 3<sup>èmes</sup> lactations.

Dans le modèle RR, les corrélations génétiques entre contrôles sont inférieures à 1 grâce à la modélisation de(s) l'effet(s) aléatoires sous forme de courbes. Comme nous l'avons évoqué précédemment, les différentes observations peuvent être considérées comme des caractères distincts où les corrélations entre les différents contrôles sont prises en compte, et sont donc similaires à des modèles multicaractères. D'ailleurs, un modèle RR où la fonction modélisant la production dans le temps aurait un nombre de paramètres strictement équivalent au nombre de contrôles serait équivalent à un modèle multicaractères. Meyer *et al.* (1991), Jensen *et al.* (1997), Rekaya *et al.*, (1999) ont employé la méthodologie multicaractères pour analyser les corrélations entre les différents contrôles d'une lactation, et entre lactations. Ainsi, 3 lactations de 305 jours chacune peuvent donner lieu à 915 caractères distincts soit plus de 400 000 corrélations à estimer. L'intérêt de cette méthode, qui est de ne pas imposer de structure de covariance entre les contrôles, est largement annihilé par les besoins liés à l'estimation des nombreux paramètres. Même en réduisant la définition des caractères à une donnée mensuelle, les paramètres estimés ne sont pas toujours très stables. Elle est donc peu adaptée à la situation et par conséquent pas utilisée en tant que telle. Dans les TDM, le terme multicaractères, lorsqu'il est employé, est généralement réservé aux modèles permettant l'analyse simultanée de plusieurs caractères tels que le lait, les matières grasse et protéique et les comptages leucocytaires.

### 1.1.3. Les modèles multicaractères (MT-TDM)

Selon cette sémantique, seuls deux modèles multicaractères sont utilisés à notre connaissance, à travers le monde : d'une part, le modèle Canadien et ses dérivés Italien (race Holstein uniquement) et Suisse (race Brune uniquement) et d'autre part, le modèle des pays nordiques (Danemark, Finlande et Suède). Ces modèles se caractérisent par leur nombre élevé de paramètres et par la stratégie employée pour la mise en œuvre opérationnelle de cette évaluation. Le modèle Canadien sera utilisé à des fins d'illustration.

Depuis février 1999, le Canada dispose en routine d'une évaluation génétique basée sur un modèle RR-TDM multicaractères (MT-TDM pour Multiple-Trait TDM). Ce modèle intègre quatre caractères (lait, MG, MP et SCS) évalués chacun sur les 3 premières lactations, soit 12 caractères (Schaeffer *et al.*, 2000). Le modèle est en constante évolution pour apporter des améliorations telles que la prise en compte du stade de gestation dans le modèle ou l'extension du stade de lactation de 305 jours à 365 jours (Schaeffer and Jamrozik, 2008). Chaque caractère est modélisé par un polynôme de Legendre d'ordre 4 (*i.e.* 5 covariables), soit 60 covariables pour l'effet génétique additif et 60 covariables pour l'effet d'environnement permanent de chaque animal. Les coefficients de régression estimés pour chaque individu sont convertis en 12 valeurs génétiques : une pour chacun des caractères de production 305 jours et chacune des lactations. Ces valeurs sont ensuite synthétisées avec des pondérations entre les différentes lactations pour obtenir une seule valeur génétique par caractère. Les informations sur les formes de courbes de lactation et les changements de

niveau de production d'une lactation sur l'autre sont malheureusement occultées par cette pratique. Seule une valeur génétique caractérisant la persistance est calculée, mais son usage et son importance sont méconnus.

L'ensemble des limites décrites dans les modèles RR-TDM reste valable pour les modèles MT-TDM. Les besoins informatiques sont encore accentués avec ces modèles multi-lactation multicaractères. Selon Schaeffer et Jamrozik (2008), la question principale avec les modèles RR est de savoir si ce type de modèle est intéressant compte tenu de leur lourdeur calculatoire.

#### 1.1.4. Les modèles de régression aléatoire avec réduction du rang

Les modèles RR-TDM et MT-TDM dépendent d'un nombre élevé de paramètres et nécessitent donc de résoudre un nombre important d'équations du modèle mixte. La plupart des paramètres de ces modèles sont utilisés pour décrire l'évolution des variances et covariances au cours de la lactation. Toutefois, les évolutions observées entre les différentes variables (comme les différentes lactations) sont quasiment identiques. Or, comme nous l'avons vu dans la partie 2.2, les recommandations internationales prônent la prise en compte des données des 3 premières lactations. Pour répondre à cette demande, les modèles généralement utilisés considèrent les 3 lactations séparément ou bien la 1<sup>ère</sup> lactation *vs* les lactations suivantes. L'estimation de paramètres génétiques spécifiques à chaque lactation n'a cependant que peu d'intérêt pour la prédiction des valeurs génétiques. Au contraire, elle peut, du fait de matrices proches de la singularité, engendrer des problèmes numériques. Afin de réduire le nombre de paramètres à estimer, Jensen (2001) a proposé de considérer non pas les différentes lactations (ou groupes de lactation) comme des caractères distincts mais comme une dimension supplémentaire dans le temps. Dans cette situation, le stade de lactation modélise la trajectoire de la lactation intra lactation et le rang de lactation modélise les changements entre lactations. Selon Misztal *et al.* (2000), une solution appropriée est de recourir aux méthodes visant à faciliter les calculs grâce aux approches basées sur une décomposition canonique. D'une part, la décomposition canonique permet d'orthogonaliser les différents éléments de la matrice de sorte qu'ils ne sont plus corrélés les uns aux autres et d'autre part, elle permet de réduire la dimension des matrices en supprimant les contributions des vecteurs propres associés aux valeurs propres les plus proches de 0. Dans le cadre des approches RR, ces méthodes ont été étudiées par Meyer (1997) ainsi que Van der Werf *et al.* (1998).

La décomposition canonique de la matrice  $\mathbf{A}$  s'écrit sous la forme :

$$\mathbf{A} = \mathbf{U} \mathbf{D} \mathbf{U}' \text{ avec } \mathbf{U} \mathbf{U}' = \mathbf{I}_m \text{ et } \mathbf{U} \mathbf{U}' = \mathbf{I}_m \quad (2.36)$$

où  $\mathbf{U}$  est la matrice orthogonale des vecteurs propres,  $\mathbf{D}$  la matrice diagonale des valeurs propres et  $\mathbf{I}_i$  la matrice identité de dimension  $i$ . Le rang de la matrice  $\mathbf{A}$  peut être réduit en supposant nulle les plus petites valeurs propres de  $\mathbf{D}$  et en supprimant les vecteurs propres correspondant de  $\mathbf{U}$ . Ainsi :

$$\mathbf{A}^* = \mathbf{U}^* \mathbf{D}^* \mathbf{U}^{*'} \text{ avec } \mathbf{U}^{*'} \mathbf{U}^* = \mathbf{I}_n \quad (2.37)$$

où  $\mathbf{A}^*$  est la matrice de rang  $n < m$ ,  $\mathbf{D}^*$  est la matrice diagonale de rang réduit,  $\mathbf{U}^*$  est la matrice orthogonale des vecteurs propres conservés. L'information est ainsi synthétisée sous forme d'un nombre limité de composantes qui sont directement interprétables.

Misztal *et al.* (2000) proposent de répéter ce processus d'élimination des covariables les moins représentatives plusieurs fois au cours des étapes d'estimation des paramètres génétiques jusqu'à la convergence. Selon ces auteurs, des changements modérés dans les paramètres utilisés (de l'ordre de 20% par rapport à leurs valeurs optimales) ont des conséquences négligeables sur la précision des valeurs génétiques prédites. Les modèles recourant à la réduction du rang de la matrice présentent une erreur résiduelle légèrement supérieure aux modèles RR classique selon Lidauer *et al.* (2003), mais qui reste très largement inférieure à celle d'un modèle FR-TDM.

Van der Werf *et al.* (1998) soulignent un autre intérêt de la transformation canonique pour les évaluations génétiques. Elle réduit avantageusement les temps de calcul et les ressources mémoire nécessaires du fait d'un nombre plus faible d'éléments hors diagonaux à estimer dans les équations du modèle mixte et de corrélations modérées entre les coefficients de la régression aléatoire.

Par exemple, dans un contexte d'évaluation multicaractères, Wiggans and Goddard, (1997) ont proposé d'utiliser la méthode de transformation canonique adaptée au contexte de données manquantes (Ducrocq and Besbes, 1993) pour synthétiser l'information des 60 caractères initiaux (10 contrôles pour le lait, la MG et la MP, en 1<sup>ère</sup> lactation vs lactations suivantes). Six pseudo-caractères ont ainsi été conservés, correspondant aux valeurs propres les plus importantes, et représentant la majeure partie des variations observées sur les données initiales.

### 1.1.5. Les modèles « Character process » (CP)

Contrairement aux modèles RR, les modèles « Character Process » (CP) ne cherchent pas à modéliser la partie aléatoire de la fonction décrivant la trajectoire de la lactation, mais directement la fonction de covariance (Pletcher and Geyer, 1999). Par exemple, la corrélation entre deux contrôles est supposée être une fonction dépendant seulement de l'intervalle entre ces contrôles. Même si cette hypothèse n'est pas strictement vraie, Pletcher et Geyer (1999) ont montré qu'elle fournissait une approximation convenable dans la plupart des cas. Lorsque l'intervalle entre les différents contrôles est strictement équivalent, il est possible de trouver un modèle CP avec une variance constante et une fonction de corrélation équivalente à un modèle autorégressif d'ordre 1. Comparativement aux modèles RR et CF, Jaffrézic et Pletcher (2000) ont montré que le modèle CP donnait la meilleure description de la structure de covariance sous-jacente. D'après eux, ce modèle présente un meilleur ajustement, des paramètres directement interprétables et en nombre réduit.



## 1.2. Intérêts des modèles longitudinaux : exemple du TDM sur les caractères laitiers

Les modèles contrôles élémentaires permettent l'utilisation des données élémentaires de production collectées par le contrôle laitier, plutôt que la donnée synthétique constituée de la production sur la durée de référence, généralement calculée sur 305 jours. De nombreux facteurs sont inclus dans le modèle comme l'effet de la conduite du troupeau ou de différents lots d'animaux, l'effet du jour du contrôle et notamment de la météorologie, l'effet du stade de lactation et de gestation de chaque animal... soit la majorité des facteurs spécifiques au troupeau et à l'animal pour ce jour de contrôle donné. Certains des effets, notamment le stade de lactation, y sont modélisés à travers une fonction du temps, ce qui offre une certaine souplesse notamment sur la description de la forme de la courbe de lactation.

Le principal intérêt lié à l'utilisation des TDM est la meilleure prise en compte des effets d'environnement affectant la production. Tout d'abord, sur l'effet environnemental lié au troupeau : il est dans le modèle lactation modélisé par l'interaction entre le troupeau et l'année de production et selon les pays, la saison de mise bas. Cet effet affecte donc l'ensemble des vaches vêlées la même année soit pour l'exemple de la figure 2.4 : l'année 1 pour les vaches 1, 2 et 3, l'année 2 pour les vaches 4 et 5 et l'année 3 pour la vache 6), même si elles ne partagent pas (vaches 1 et 3) ou peu (vaches 2 et 3) de contrôles communs, comme l'illustre la figure 2.4. Les animaux soumis aux mêmes effets d'environnement sur une période longue peuvent ainsi se retrouver dans des effets troupeau – année différents (vaches 3 et 4). De nombreuses études ont montré qu'une grande part des effets environnementaux n'était pas pris en compte dans ce modèle (Meyer *et al.*, 1989 ; Ptak et Schaeffer, 1993).

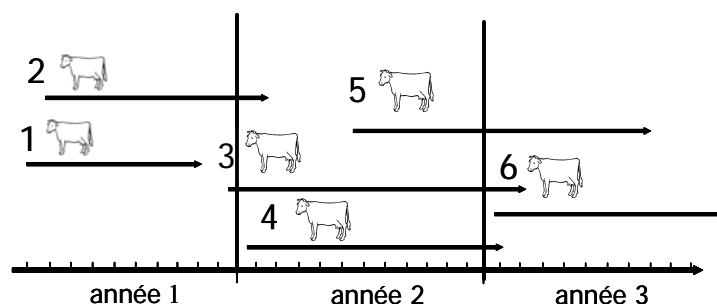


Figure 2.4. Représentation de la définition de l'effet troupeau – année (– saison) utilisé dans le modèle « lactation »

Au contraire, dans le modèle TDM, il est possible de définir un effet troupeau – jour de contrôle (HTD pour Herd – Test Day) dépendant des vaches présentes lors du contrôle de production (Figure 2.5). La prise en compte dans le modèle d'un effet de l'environnement beaucoup plus précis permet de réduire de façon importante la variance résiduelle, comme l'ont montré Meyer *et al.* (1989), Ptak et Schaeffer (1993), Pösö *et al.* (1996) et ainsi d'accroître légèrement l'héritabilité du caractère. Dans un modèle TDM, Ptak et Schaeffer (1993) ont montré que la prise en compte de l'effet troupeau à travers l'effet HTD plutôt qu'un effet troupeau – année – saison permettait de gommer les différences dans l'estimation des coefficients de régression de la courbe de lactation liées à la saison de mise bas. Le

modèle contrôles élémentaires permet donc une correction automatique pour les effets dont l'impact varie au cours de la lactation.

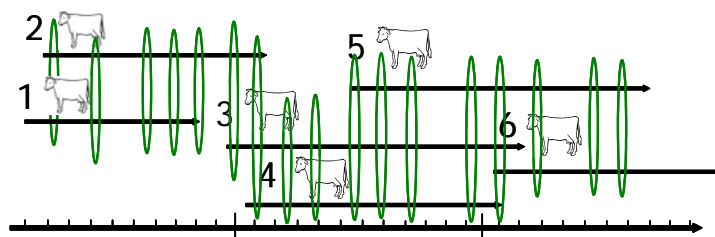


Figure 2.5. Représentation de la définition de l'effet troupeau – jour de contrôle (HTD) utilisé dans le modèle « contrôles élémentaires » (TDM)

Le stade de lactation est le second effet, en plus de l'effet troupeau – jour de contrôle qui impacte de façon importante la production laitière d'un jour donné. La modélisation de la production à travers une ou plusieurs courbes permet de tenir compte d'un niveau de production variable en fonction du stade de lactation. Les corrélations phénotypiques entre les différents contrôles, mais aussi de l'intervalle entre les contrôles sont automatiquement prises en compte. En effet, la production d'un contrôle est statistiquement corrélée avec les contrôles précédents et suivants et ce d'autant plus fortement que l'intervalle entre les contrôles est réduit. Danell (1982) a estimé que les corrélations entre contrôles adjacents étaient supérieures à 0.90, tout en diminuant lorsque les contrôles s'espacent. Selon Pander *et al.* (1992), les corrélations génétiques entre contrôles extrêmes sont au minimum de 0.70, Rekaya *et al.*, (1995) obtenant une valeur de 0.52.

Parmi les autres avantages du modèle TDM, nous avons vu dans la partie 2.2 qu'il existait de nombreux points de divergence entre les pays concernant le calcul de la lactation de référence. L'utilisation de TDM répond à ces problèmes : d'une part, parce qu'il permet de s'affranchir des aspects méthodologiques liés au calcul de la production de référence et d'autre part parce qu'il permet de se libérer des problèmes liés à l'extrapolation de la production sur la durée de référence et des corrections appliquées aux lactations courtes.

Un autre intérêt du modèle contrôles élémentaires est sa rapidité d'exploitation. En effet, d'un point de vue théorique, les vaches peuvent faire l'objet d'évaluation génétique dès lors qu'elles ont au moins un contrôle enregistré. Les informations des contrôles élémentaires étant potentiellement utilisées plus vite qu'avec un modèle lactation, les décisions de sélection peuvent intervenir plus rapidement et ainsi contribuer à la réduction de l'intervalle entre générations (Reents and Dopp, 1996 ; Swalve 1998). Jamrozik *et al.* (1997b) suggèrent même d'accroître la fréquence des évaluations afin de minimiser le temps nécessaire pour l'exploitation des données. Nous noterons toutefois que les premiers pays ayant publiés des index contrôles élémentaires (Allemagne, Pays-Bas) ont accru la sévérité des règles de publication par rapport à la situation initiale suite à d'importantes variations d'index liées à un grand nombre de lactations débutantes ou les passages successifs d'une lactation à l'autre.

La précision des évaluations génétiques constitue aussi un aspect attractif de ces modèles. En effet, Ptak et Schaeffer (1993) et Meyer *et al.*, (1989) estiment respectivement à 4 et 5, le

nombre de contrôles élémentaires pour obtenir une précision équivalente à la production par lactation. Les lactations complètes comptant généralement 9 ou 10 contrôles, elles permettent d'obtenir une évaluation génétique de précision supérieure, notamment sur la population femelle (Lidauer *et al.*, 2003). De plus, l'héritabilité des données de productions élémentaires, étudiée par de nombreux auteurs (Wilmink, 1987 ; Pander *et al.*, 1992 ; Swalve, 1995a) peut atteindre au milieu de la lactation un ordre de grandeur similaire à celle des données lactations (supérieure à 0.30). Toutefois, en début et fin de lactation, l'héritabilité est plus faible notamment à cause de variances résiduelles supérieures (Ptak and Schaeffer, 1993).

Comparativement à l'évaluation génétique basée sur un modèle lactation, le modèle contrôles élémentaires n'entraîne pas de profonds changements dans les index. Les corrélations entre valeurs génétiques obtenues avec un modèle lactation et celles du modèle contrôles élémentaires sont supérieures à 0.90 et 0.95 pour les taureaux disposant respectivement de plus de 20 et 50 filles (Swalve, 1995a). Lidauer *et al.* (2000) obtiennent eux des corrélations entre 0.94 et 0.97 pour les mâles sur les 3 caractères de production (Lait, MG et MP) et entre 0.88 et 0.89 pour les femelles.

Du fait d'une meilleure précision à nombre de contrôles équivalent, il semble envisageable, comme le mentionnent Ptak et Schaeffer (1993) d'assouplir les protocoles de collecte de données de production laitière, notamment en allégeant la fréquence des contrôles et donc les contraintes en terme d'organisation ou par la pratique du contrôle par l'éleveur. Swalve (1998) avance l'intérêt économique lié à l'extensification des protocoles de collectes des données, très coûteux dans la version A classique. Jamrozik *et al.* (1997b) ou encore Wiggans et Goddard, (1997) vont encore plus loin en exploitant les propriétés du modèle multicaractères. En effet, ces auteurs envisagent des protocoles où les taux ne seraient pas mesurés de façon systématique mais seraient évalués en partie grâce aux corrélations entre caractères. C'est une réalité en Finlande, où le modèle multicaractères est utilisé pour compenser le faible nombre de mesures des taux avec en moyenne 4.5 observations par lactation à comparer aux 9.4 observations pour la quantité de lait (Lidauer *et al.*, 2003). A l'opposé, les troupeaux équipés de compteurs à lait pourraient souhaiter le recueil de l'ensemble des données collectées par leur système informatique individuel.

L'utilisation d'un modèle d'évaluation basé sur les contrôles élémentaires ouvre de nouvelles perspectives en terme de caractère d'intérêt. La sélection sur la persistance de la production laitière ou son utilisation comme prédicteur précoce par exemple de la fertilité sont ainsi envisageables, dès lors que l'effet génétique est modélisé sous forme de courbe. Comme nous l'avons vu dans la partie 2.1, la persistance constitue depuis longtemps un caractère intéressant pour les éleveurs, aussi bien pour des raisons économiques dues au fonctionnement de l'exploitation agricole que pour des raisons physiologiques liées au stress des vaches productrices. Jamrozik *et al.* (1997b) estiment qu'il est possible de modifier la trajectoire de la courbe de lactation et donc la persistance grâce à la sélection. Toutefois, ces auteurs s'interrogent sur les recommandations liées à la sélection sur ce caractère, ne connaissant pas les conséquences à long terme sur des caractères tels que la reproduction ou la longévité. Outre les informations sur la persistance, les évaluations génétiques sur les

contrôles élémentaires offrent des perspectives de développement d'outils d'appui technique. Cet élément a rarement été mis en avant, du moins jusque récemment, à l'exception de Mayeres *et al.* (2004) (en Belgique) et Koivula *et al.* (2007) (en Finlande). Les néerlandais ont également travaillé sur la mise au point d'outils d'appui technique, mais sans divulguer d'informations sur leur mise en oeuvre. Au Canada, ce volet n'a pas pu être développé pour le moment du fait du brevet sur les contrôles élémentaires détenu par l'université Cornell. La constitution de bases de données plus larges que les simples données de production, incluant notamment les événements relatifs à la santé de l'animal, la conduite du troupeau... devrait permettre d'offrir de nouvelles perspectives de valorisation.

Concernant les limites des modèles contrôles élémentaires, la principale est de nature informatique (Ptak and Schaeffer, 1993 ; Jamrozik *et al.*, 1997b). Même si les capacités de stockage et de mémoire se sont considérablement accrues au cours des dernières années, la gestion et le traitement de millions de données suivant des modèles toujours plus complexes n'est pas sans poser de problèmes. Ainsi, en Janvier 2008, la race Holstein comptait plus de 34 millions de lactations enregistrées en France et environ 10 fois plus de contrôles élémentaires. La complexité du modèle étant elle aussi fortement accrue, les temps de calcul et besoins en mémoire augmentent selon un rapport plus important que celui lié au nombre de données. Jamrozik *et al.* (1997b) soulignent donc l'importance du compromis à trouver entre la complexité de la modélisation des différents effets, le nombre de paramètres à estimer et donc les besoins calculatoires.

Comme le signalent Ptak et Schaeffer (1993), malgré tous les avantages et intérêts que le TDM comporte, il ne peut prendre en compte dans les évaluations que les effets connus. Ainsi, si dans certains troupeaux les animaux sont conduits différenciellement en fonction de leur niveau de production (traitement préférentiel), le TDM ne pourra en tenir compte à moins que la constitution de ces groupes ne soit renseignée à travers la variable « Unité Laitière » comme le permettent les tables du SIG en France. A défaut d'avoir cette information, la prise en compte dans le modèle de variances hétérogènes amoindrit l'impact du traitement préférentiel mais reste une correction partielle.

Un autre point sensible du modèle contrôles élémentaires concerne la définition fine du groupe de contemporaines à travers l'effet HTD. Dans certains cas, il peut constituer un facteur limitant son développement. En effet, dans certains pays, la taille des troupeaux est réduite (moins de 20 vaches en Finlande) et ceux-ci peuvent être constitués de plusieurs races. Dans le cas d'évaluation génétique raciale ne considérant généralement que les 3 premières lactations et les contrôles réalisés au cours de la période de référence (entre 5 à 305 jours), la définition des effets HTD peut être cruciale pour obtenir des estimés précis comme le souligne Swalve (1995a, 1995b).

Le dernier aspect pouvant poser quelques problèmes touche les pays souhaitant conserver un certain niveau d'historique. Pour cela, Swalve (1995b) ainsi que Jamrozik *et al.* (1997b) soulignent le besoin de développer une méthode composite permettant de combiner les données de contrôles élémentaires aux données lactations. Jamrozik *et al.* (1997b) et Mäntysaari (2002) suggéraient l'utilisation d'un modèle multicaractères similaire au MACE utilisé pour les évaluations internationales afin de combiner les différentes sources

d'informations. L'intérêt de cette méthode dépasse la simple conservation des données « historiques » puisque qu'elle peut être envisagée comme un moyen de réduire le nombre de données à analyser.

### **1.3. Les modèles de régressions aléatoires utilisés à travers le monde**

Le champ d'application des modèles de régressions aléatoires est bien plus large que les seuls caractères de production chez les bovins laitiers comme le montre Schaeffer (2004) qui passe en revue les différentes applications possibles. Les modèles RR peuvent avoir un intérêt dès lors qu'il y a une relation entre la variable analysée et une variable continue telle que la date de la mesure ou l'âge. Selon cet auteur, un modèle RR doit inclure une ou des courbes d'effets fixes permettant de décrire la relation globale entre les variables, pour les différentes classes d'individus, et des régressions aléatoires pour l'effet génétique et l'effet d'environnement permanent de l'animal. D'autres régressions aléatoires sur l'effet du troupeau ou l'effet génétique maternel peuvent aussi être incluses.

#### **1.3.1. Les autres caractères avec contrôles élémentaires**

De façon similaire aux bovins laitiers, les modèles contrôles élémentaires ont été utilisés chez les ovins et les caprins laitiers. La modélisation de la trajectoire phénotypique est toutefois difficile chez ces deux espèces car les agneaux et chevreaux sont généralement allaités par la mère au cours des premières semaines. Un modèle RR peut être appliqué à la production hebdomadaire ou mensuelle d'œufs chez les volailles qui suit une courbe similaire à la lactation (Anang *et al.*, 2000 ; Luo *et al.*, 2007), à la production annuelle de laine chez les moutons, la taille des portées des truies par mise bas...

Une autre application peut concerner la fertilité de la semence des mâles reproducteurs. Les observations peuvent être le volume des éjaculats par semaine (Carabaño *et al.*, 2007), une mesure de la motilité du sperme (Serrano *et al.*, 2006), ou le pourcentage de spermatozoïdes mort analysés par éjaculat. L'âge du mâle lors de la récolte, l'agent assurant le prélèvement, la période de l'année, l'intervalle entre deux prélèvements et le régime alimentaire pourraient être selon Schaeffer (2004) les facteurs pris en compte dans le modèle.

#### **1.3.2. Les caractères de croissance**

Les données de croissance peuvent aussi être modélisées avec des modèles RR. Dans la plupart des espèces d'élevage, la croissance constitue un caractère économique majeur. Quelques caractères tels que la consommation alimentaire, l'efficacité alimentaire, le dépôt adipeux, le développement musculaire, la croissance osseuse, la maturité et l'état corporel peuvent y être associés. La différence majeure, par rapport aux modèles contrôles élémentaires sur la production laitière, est l'aspect cumulatif des données de croissance. Une des premières applications des modèles RR aux données de croissance concernait les porcs (Andersen and Pedersen, 1996). Le poids et la croissance étaient modélisés d'une part en fonction du temps et d'autre part en fonction de la consommation alimentaire. L'idéal aurait été d'utiliser un modèle RR multicaractères, qui aurait permis la prise en compte simultanée

du poids corporel et de la consommation alimentaire et ainsi de modéliser l'évolution de la relation entre ces deux caractères au cours du temps. Albuquerque et Meyer (2001) ont étudié un modèle RR modélisant les effets maternels (génétique et environnement permanent), prépondérants pour la croissance des bovins allaitants, à partir de régressions aléatoires. La corrélation entre les effets directs et maternels est supposée nulle afin de simplifier le modèle.

Les modèles RR ont été utilisés dans de nombreux travaux de recherche sur les caractères de croissance de différentes espèces : bovins allaitants pour Magnobosco *et al.* (2000), Meyer (1999, 2000, 2001, 2004, 2005) et Robert-Granié *et al.* (2002), truite arc-en-ciel pour McKay *et al.* (2002), consommation alimentaire chez le porc pour Schnyder *et al.* (2001), caractères de conformation chez les bovins laitiers pour Uribe *et al.* (2000). Dans ces applications, les modèles RR permettent une plus grande souplesse vis-à-vis des questions concernant le nombre de mesures et la période ou l'âge auxquels elles doivent être réalisées. L'utilisation de modèles RR pour modéliser la croissance des plantes ou de populations bactériennes peut aussi être envisagée.

### 1.3.3. Les applications aux analyses de survie

La survie est un caractère qui peut être observé tous les jours de la vie de l'animal, mais seule la date à laquelle l'animal décède ou est réformé est généralement enregistrée. Classiquement, les analyses de survie sont réalisées avec une fonction modélisant le risque de réforme (Ducrocq, 1987). Cependant, Veerkamp *et al.* (1999) ont proposé l'utilisation de modèle RR pour ce type d'analyse, en modélisant chaque mois de présence d'un animal par des valeurs 1, et les mois où il est réformé par des valeurs 0. Toutefois, la méthodologie RR faisant appel à une analyse BLUP semble moins adaptée aux données de survie qui sont de type binaire. Dans le cas de l'analyse de survie s'ajoute un problème majeur : la série d'observations ne peut être composée que de 1 suivi de 0. L'ajout d'un 1 entre deux 1 n'apporte aucune information supplémentaire. Une seule donnée résume toute l'information : le mois du passage d'un 1 à un 0.

### 1.3.4. Les applications aux caractères de fertilité en bovins laitiers

Les performances reproductives sont des caractères économiquement importants chez les bovins laitiers. Ces performances incluent l'âge au 1<sup>er</sup> oestrus, l'âge à la 1<sup>ère</sup> insémination, le nombre d'IA nécessaires pour obtenir une gestation, l'âge à la fécondation, l'âge au vêlage, la facilité de vêlage... Ces caractères sont observés à chaque vêlage et leur valeur génétique peut donc évoluer au cours du temps. Le nombre de vêlages peut être utilisé comme unité de temps, de façon équivalente au stade de lactation. Les différentes inséminations d'une même lactation peuvent également être analysées avec des modèles de régression aléatoire comme l'ont fait Averill *et al.* (2006) pour des vaches en première lactation. Les caractères de fertilité ayant une très faible héritabilité, il est intéressant d'utiliser un modèle multicaractères de façon à analyser l'ensemble des caractères simultanément.

La principale difficulté de ces modèles reste l'estimation des paramètres génétiques, les données se raréfiant avec l'accroissement du rang de lactation. Cela nécessiterait la sélection d'un groupe d'animaux ayant des performances supérieures à la moyenne de la population,

source de biais dans l'estimation des paramètres génétiques. La définition d'un index économique sur la fertilité serait nécessaire pour combiner les valeurs génétiques de chacune des lactations et de l'ensemble des caractères. D'un point de vue méthodologique, l'utilisation d'une méthode basée sur le BLUP pour des caractères non continus pose des problèmes. Malgré ces difficultés, Schaeffer (2004) souligne les nombreuses perspectives des modèles RR sur les caractères de fertilité, notamment pour l'analyse des effets d'une lactation à la suivante.

### 1.3.5. Les applications aux études d'interaction génotype-milieu

L'existence d'interactions génotype-milieu suppose que le classement des animaux dépend étroitement de leur environnement. Plutôt qu'une régression sur le stade de lactation ou l'âge de l'animal, les modèles RR utilisent une variable environnementale telle qu'un index basée sur la température et l'humidité, l'altitude ou le niveau de production du troupeau. Ravagnolo et Misztal (2000) ont utilisé l'index température-humidité pour estimer des paramètres génétiques prenant en compte la variabilité de l'environnement de production afin d'établir un classement fonction du niveau de stress lié à la chaleur. Lawlor *et al.* (2002) ont utilisé un modèle RR pour modéliser l'évolution de la variance génétique des caractères de production laitière, de durée de vie productive et quatre caractères morphologiques entre 1981 et 1996. Toutefois, les données enregistrées sont spécifiques à un environnement donné (localisation ou période), d'après Schaeffer (2004), il est difficile de conclure sur l'intérêt des modèles RR pour la modélisation des interactions génotype-milieu.

Pour le moment, ces modèles ont principalement été utilisés en génétique animale. Toutefois, leurs propriétés ouvrent de larges perspectives pour des applications telles que les productions végétales, les sciences sociales, la santé humaine et beaucoup d'autres situations biologiques.

## 2. Description générale du modèle contrôles élémentaires

Dans les évaluations sur les contrôles élémentaires, un certain nombre d'effets sont pris en compte dans le modèle quel que soit le pays. De façon synthétique, la production journalière  $y$  est décrite comme :

$$production = HTD + courbes\ fixes + effets\ fixes + génét. + env.\ perm. + erreur \quad (2.38)$$

Un élément important du modèle est la période d'analyse des données. Avec les modèles lactation, les performances étaient calculées à partir des contrôles effectués entre 5 et 305 jours de lactation, à l'exception de la Nouvelle-Zélande qui se limite à 270 jours compte tenu de son système de production. Les modèles contrôles élémentaires offrent une plus grande souplesse vis-à-vis de la durée de référence d'une lactation. Même si la plupart des pays ont conservé un modèle d'analyse incluant les performances enregistrées entre 5 et 305 jours de lactation, cette durée est facilement extensible. En effet, compte tenu de l'allongement observé des lactations, certains pays ont choisi d'inclure les données jusque 330 (en Allemagne), 335 (aux Pays-Bas), 365 (au Danemark, en Finlande, et en cours d'étude au Canada) ou 400 jours. Ces variations compliquent toutefois la comparaison des matrices de

covariance estimées dans les différents pays. La tendance est d'allonger la durée de prise en compte des données même si les valeurs génétiques restent exprimées en fonction de la durée de référence du pays.

## **2.1. Les effets dits « fixes »**

### **2.1.1. L'effet Troupeau – Jour de contrôle (HTD)**

L'effet HTD définissant le groupe de contemporaines en remplacement d'un effet troupeau – année – saison a unanimement été intégré dans les modèles contrôles élémentaires et en constitue un des intérêts majeurs. Son impact sur la qualité du modèle a largement été démontré, notamment par Ptak et Schaeffer (1993). Suivant les pays, lorsque la taille du troupeau est suffisante, l'effet troupeau – jour de contrôle peut être défini par rang de lactation pour tenir compte d'une conduite de troupeau différenciée suivant le rang de lactation de l'animal. Cette définition encore plus fine de l'effet HTD classique est utilisée par l'Allemagne, le Canada, l'Italie. L'analyse de l'effet troupeau – jour de contrôle offre de nouvelles perspectives en terme d'outils d'aide au management pour les éleveurs (Pösö *et al.*, 1996 ; Mayeres *et al.*, 2004 ; Koivula *et al.*, 2007).

### **2.1.2. Les autres effets fixes**

Les modèles contrôles élémentaires incluent généralement une fonction dépendante du temps  $t$  qui permet de tenir compte des relations phénotypiques entre les performances mesurées et les covariables temporelles. Cette fonction, qui prend en compte la forme de la courbe de lactation pour un caractère et un groupe d'animaux donnés appartenant généralement à la même race, est définie de façon très variable suivant les pays aussi bien en terme de type de courbes (polynomiales, non-paramétriques, semi-paramétriques...) qu'en terme de nombre de courbes dans le modèle (une courbe définie avec de nombreuses interactions *vs* plusieurs courbes additives). Ainsi, dans son modèle actuel, l'Allemagne décrit la courbe de lactation d'un animal en fonction de sa combinaison région – rang de lactation – classe d'année de vêlage – classe d'âge au vêlage – classe de saison de vêlage – classe d'intervalle de vêlage soit l'une des 8547 courbes possibles. Au Canada, une courbe de base est définie par rang de lactation à laquelle s'ajoute une courbe définie par région – année – rang de lactation – classe d'âge au vêlage – classe de saison de vêlage. Quant aux Pays-Bas, une courbe non-paramétrique est utilisée pour définir, par rang de lactation, l'effet du stade de lactation à laquelle s'ajoutent deux courbes dépendantes du stade de lactation et une courbe fonction du stade de gestation. Les deux courbes affinant l'effet du stade de lactation sont définies pour l'une par l'année – rang de lactation – classe d'âge au vêlage – classe de saison de vêlage – classe de stade de lactation et pour l'autre par classe de durée de tarissement – classe de stade de lactation. Pour la courbe portant sur l'effet de la gestation, elle est définie par rang de lactation – classe de stade de gestation. Le soin apporté au choix de la méthode de modélisation de la courbe phénotypique de lactation est essentiel dans le cadre des modèles contrôles élémentaires. L'importance des effets environnementaux pris en compte dans les courbes d'effets fixes nécessite une bonne flexibilité de ces dernières. Quel que soit le type de courbe utilisé, le nombre de paramètres de la régression fixe est globalement élevé.



Brotherstone *et al.* (2000) propose de modéliser la partie fixe de la courbe de lactation avec une courbe paramétrique telle que celle d'Ali et Schaeffer (1987) ou à l'aide de splines cubiques lissées (White *et al.*, 1999) ayant un nombre de nœuds suffisant. En règle générale, les régressions fixes comptent un minimum de 5 paramètres. En plus de ces courbes, certains pays incluent des termes constants, plus simples à estimer, définis comme des interactions entre effets, interactions qui ne pouvaient généralement pas être prises en compte dans les courbes précédemment décrites.

Chaque pays modélise donc les effets d'environnement suivant un modèle qui lui est propre, mais dans l'ensemble, les effets pris en compte dans les modèles sont assez similaires. Ces effets sont la race, le rang de lactation, le stade de lactation, le stade de gestation ou l'intervalle entre vêlage, la région de production, l'année, l'âge au vêlage, la saison de vêlage, la durée de tarissement, l'effet d'hétérosis dans le cas d'évaluation multi- raciale... La plupart de ces effets ont largement été étudiés entre 1970 et 1990 dans le contexte de performances sur la lactation notamment par Auran (1976), Congleton et Everett (1980), Danell (1982), Keown *et al.* (1986), Schutz *et al.* (1990)... mais restent tout à fait valables dans le cadre de la production journalière.

Le stade de lactation est certainement l'effet qui a été le plus largement étudié, du moins pour la quantité de lait. Nombre de courbes de lactation ont été proposées pour modéliser cette variation dont les principales sont celles de Wood (1967), Ali et Schaeffer (1987), Wilmlink (1987). La prise en compte du stade de gestation dans le modèle est essentielle, même si elle a souvent été négligée comme le rappelle Swalve (1995b). Sa non prise en compte pénalise les lactations courtes et favorise au contraire les lactations longues, car la gestation interfère de façon importante avec le niveau de production journalier, surtout dans le dernier tiers de la gestation. Suivant les pays, l'effet gestation peut être pris en compte à travers l'effet de l'intervalle entre vêlages, notamment lorsque la date d'insémination fécondante ne peut être déterminée. Cependant, l'intégration de l'effet du stade de gestation dans le modèle contrôle élémentaires constitue une approche plus directe. Le stade de lactation est généralement défini en interaction avec le rang de lactation. Il permet ainsi de tenir compte de la variabilité des formes de courbes de lactation, notamment la plus faible amplitude mais aussi la plus forte persistance des courbes chez les primipares. Pour expliquer ce phénomène, Stanton *et al.* (1992) avance l'effet de compétition entre les besoins de croissance de l'animal et les besoins liés à la production laitière. Concernant l'âge au vêlage, son effet décroît au cours de la lactation. Même si certains, notamment Meyer *et al.*, 1989 ont proposé d'utiliser l'âge au contrôle à la place de l'âge au vêlage pour l'analyse des contrôles élémentaires, Stanton *et al.* (1992) recommandent de conserver ce dernier qui présente l'avantage d'être commun à l'ensemble des contrôles d'une lactation et limite le nombre de classes à constituer. En ce qui concerne l'effet de la saison sur la production, des généticiens comme Wood (1976) considéraient que l'effet du groupe de contemporaines (HTD) était suffisant pour en quantifier l'impact. Même si cet effet prédomine largement celui de la saison de mise bas, ce dernier n'est pas négligeable, comme l'ont montré Stanton *et al.* (1992). La durée de tarissement est un effet qui entraîne un biais important s'il n'est pas pris en compte dans le modèle d'après De Roos *et al.* (2002). Cependant, cet effet est délicat à intégrer dans le

modèle car les pratiques de tarissement varient grandement d'un élevage à l'autre, comme l'indique Swalve (1995c). Les autres facteurs décrivant le contexte de production comme la région ou l'année de production peuvent faire varier les performances de façon importante. Les variations climatiques selon les années sont en grande part responsables de la quantité et de la qualité des ressources fourragères. Au sein d'un pays, notamment pour les pays couvrant une grande superficie comme les Etats-Unis, le Canada, l'Allemagne, la France, les différentes régions témoignent d'un environnement de production plus ou moins favorable, notamment à travers le contexte pédo-climatique.

Compte tenu du nombre important de facteurs ayant un impact sur la production, il est donc nécessaire de trouver des compromis pour la définition des courbes de lactations et des interactions considérées. L'objectif est d'une part de bien caractériser les différents groupes d'animaux avec un découpage suffisamment fin et d'autre part d'avoir des effectifs suffisants dans chacun des groupes pour que les paramètres estimés le soient avec une précision suffisante.

## **2.2. Les effets dits « aléatoires »**

Les régressions aléatoires sont utilisées pour modéliser les déviations d'individus ou groupes d'individus par rapport à la trajectoire de la courbe définie par les effets fixes. Parmi les effets modélisés comme « aléatoires », il y a bien sûr l'effet génétique additif, l'effet d'environnement permanent et dans certains pays un effet troupeau – année. L'utilisation de polynômes orthogonaux exprimés en fonction d'unités de temps standardisées a été recommandée (Kirkpatrick *et al.*, 1990). L'ordre des polynômes utilisés pour modéliser les différentes régressions aléatoires (génétique, environnement permanent, troupeau) n'est pas contraint à être équivalent, mais l'est souvent par mesure d'équilibre et souci de simplicité.

### **2.2.1. L'effet génétique additif**

Outre la mesure du niveau de production de l'animal, les modèles contrôles élémentaires avec régressions aléatoires permettent, grâce à la variation de l'effet génétique en fonction du stade de lactation, de décrire des déviations génétiques individuelles par rapport à un profil moyen et en particulier de caractériser la persistance. Les animaux peuvent présenter des profils très différents (Figure 2.6). Par exemple, des taureaux peuvent avoir une valeur génétique similaire en début de lactation, puis évoluer de façon opposée. Le potentiel de production du taureau A augmente au cours de la lactation, montrant une bonne persistance, alors que le taureau E chute de façon importante. La différence entre ces taureaux s'amplifie avec le rang de lactation pour passer de 6 kg en fin de 1<sup>ère</sup> lactation à 12 kg en fin de 3<sup>ème</sup> lactation. Certains taureaux, comme le B sont supérieurs aux autres sur la quasi-totalité des lactations, d'autres présentent un profil opposé (Taureau F).

La persistance, caractère économiquement important dans un élevage, joue un rôle primordial sur la santé de l'animal, notamment en reflétant le stress métabolique au cours de la lactation. Les animaux présentant une bonne persistance sont généralement des vaches avec un faible pic de lactation : Les résultats de Rekaya *et al.* (1999) ont mis en évidence un antagonisme entre les coefficients de régression aléatoire liés au pic de production et ceux liés à la

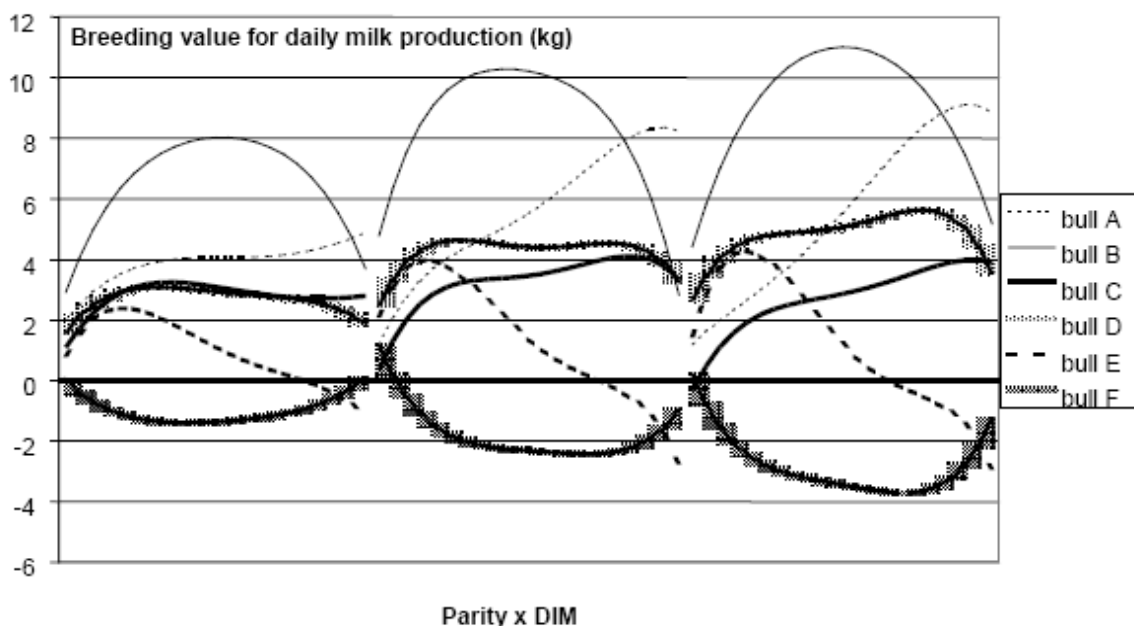


Figure 2.6. Evolution des valeurs génétiques sur la production laitière des 3 premières lactations sur 6 taureaux Néerlandais (extrait du document explicatif du modèle contrôles élémentaires de ce pays ; source : NRS (Royal Dutch Cattle Syndicate), 2008)

persistance. Ces animaux ont des lactations généralement qualifiées de « plates ». La sélection sur la persistance permettrait selon Zimmerman and Sommer, (1973) de réduire le stress métabolique lié au pic de lactation à production équivalente. Les animaux les plus persistants semblent également être plus résistants aux maladies, d'après les résultats obtenus par Jakobsen (2000), qui a montré sur 8000 primipares danoises, l'existence d'une corrélation de l'ordre de 0.2 à 0.5 entre la persistance et la résistance aux maladies. La sélection sur la persistance semble prometteuse d'autant plus que ce caractère est, d'après Jamrozik *et al.* (1998), génétiquement indépendant du niveau de production laitière et se caractérise par une héritabilité modérée.

Différentes mesures de persistance ont été proposées, notamment par Jamrozik *et al.* (1998), mais aucune ne fait l'unanimité. Au final, le Canada utilise une mesure basée sur le déclin de la production entre 60 et 280 jours de lactation pour apprécier la persistance (Schaeffer *et al.*, 2000). Une autre alternative serait de représenter graphiquement la régression aléatoire sur l'effet génétique. Ce mode de communication permettrait à la fois de modéliser le niveau de production, mais aussi de percevoir la persistance de l'animal. D'après nos informations, cette alternative n'est utilisée dans aucun pays pour l'instant. Un autre intérêt d'un mode de représentation graphique est de disposer de l'évolution des profils de valeurs génétiques au cours des lactations, ce qui permet d'aborder la notion de maturité de l'animal. Ainsi, à la figure 2.6, le taureau C progresse légèrement entre la 1<sup>ère</sup> et la 2<sup>ème</sup> lactation, mais sa valeur génétique chute sur la 1<sup>ère</sup> moitié de sa 3<sup>ème</sup> lactation, alors que le taureau D, qui avait un profil similaire au taureau C sur la 1<sup>ère</sup> lactation, a un potentiel qui continue à progresser au cours des lactations suivantes. A partir de cette figure, nous pouvons supposer que le taureau C était plus précoce que le taureau D, dont la maturité est au plus tôt atteinte au cours de la 3<sup>ème</sup> lactation. Dans leurs évaluations, les Pays-Bas expriment le taux de maturité comme une différence entre la valeur génétique en 1<sup>ère</sup> lactation (corrigée pour être exprimée en équivalent adulte : facteur de 1.3) et la valeur génétique en 3<sup>ème</sup> lactation (NRS, 2008).

### 2.2.2. L'effet d'environnement permanent

L'effet d'environnement permanent modélise la part de la performance de l'animal qui ne peut être expliquée par les effets environnementaux dits fixes tels que l'effet troupeau – jour de contrôle, le stade de lactation, l'âge au vêlage... ni par l'effet génétique additif et qui explique la ressemblance entre les contrôles successifs d'un animal. Cet effet, dont l'impact varie suivant le stade de lactation, permet par exemple de tenir compte d'éventuelles maladies ou du retard de croissance que l'animal aurait eu pendant sa phase de développement et qui peuvent avoir un impact sur l'ensemble de sa carrière.

L'effet d'environnement permanent est généralement modélisé de façon similaire à l'effet génétique additif. Ces deux effets ont ainsi une flexibilité équivalente, ce qui évite des problèmes de transfert de variance de l'effet génétique vers l'effet d'environnement permanent et vice versa (Pool and Meuwissen, 2000). Pool *et al.* (2000) ont toutefois montré que l'effet génétique additif pouvait être modélisé avec une matrice de variance – covariance d'ordre inférieur à celui de l'environnement permanent.

### 2.2.3. L'effet troupeau - année

De Roos *et al.* (2002) indiquent la présence d'un biais important sur la résiduelle quand l'interaction troupeau – année – stade de lactation n'est pas pris en compte dans le modèle. Pour tenir compte de cette interaction et prendre en compte des différences liées au système de conduite d'élevage et notamment à l'alimentation, Jamrozik et Schaeffer (1997b) ainsi que Jensen (2001) ont proposé d'inclure l'effet du troupeau dans la modélisation de la courbe de lactation. Mais à défaut d'avoir des troupeaux de taille suffisante pour inclure cet effet dans la définition de la trajectoire de la lactation, celui-ci peut être modélisé à partir d'une régression aléatoire, c'est-à-dire avec une courbe spécifique au troupeau pour une année donnée.

La courbe de l'effet troupeau – année permet de tenir compte de l'impact de la conduite du troupeau sur la production. Par exemple, certains troupeaux avec des systèmes de ration complète peuvent avoir une alimentation relativement pauvre pour les vaches fraîchement vêlées et au contraire supérieure aux besoins pour les vaches en fin de lactation, alors que les troupeaux en ration semi-complète, avec complémentation individuelle en fonction du niveau de production de la vache, tendent à maximiser la production au moment du pic et au contraire à la limiter en fin de lactation. Contrairement à l'effet fixe troupeau – jour de contrôle qui modélise le niveau moyen de production du troupeau, l'effet aléatoire troupeau – année modélise la déviation de la trajectoire de la lactation spécifique au troupeau, par rapport à la courbe de production moyenne. Les différences de persistance liées à la conduite du troupeau peuvent ainsi être prises en compte.

Gengler *et al.* (2001) et De Roos *et al.* (2002) ont montré, en utilisant un effet troupeau – année défini bisannuellement, que celui-ci avait un impact important sur la structure de la covariance génétique. Cet effet présente une variance non négligeable en début et fin de lactation, mais très faible (5 à 15 fois moins) en milieu de lactation. De Roos *et al.* (2004) l'ont estimé à 7 – 12% de la variance phénotypique en début et fin de lactation, soit 2 à 4% en moyenne sur la lactation. Ils ont montré que son inclusion dans le modèle contrôles élémentaires permettait de corriger les effets de bord observés lors de l'estimation des

variances génétiques (Misztal *et al.*, 2000), accroissait la stabilité des paramètres estimés et réduisait l'héritabilité du caractère.

Pour expliquer la diminution de la variance génétique liée à l'inclusion de l'effet aléatoire troupeau – année dans le modèle observée par Jamrozik *et al.* (2001), Gengler *et al.* (2001) et De Roos *et al.* (2004), ces derniers avancent comme possible explication une confusion dans l'origine de la variance, les vaches et leur mère ayant généralement effectué leur carrière dans le même troupeau.

### 2.3. La résiduelle

La principale source d'hétérogénéité de variance dans les modèles contrôles élémentaires est celle liée au stade de lactation. Cependant, d'après une étude faite par Misztal *et al.* (2000), la plupart des modèles supposent que la résiduelle est normalement et indépendamment distribuée avec une moyenne nulle et une variance homogène sur la lactation ou au sein de classes. Le nombre de classes est toutefois très variable suivant les pays. Dans le modèle canadien (Schaeffer *et al.*, 2000), quatre périodes sont définies : 5 – 45, 46 – 115, 116 – 265 et 266 – 305 jours. Au Royaume-Uni, la variance résiduelle est modélisée en fonction de 12 classes ayant des amplitudes variables de une semaine en début de lactation à plus de 100 jours lors de la seconde moitié de la lactation (Brotherstone *et al.*, 2000). En Espagne, Rekaya *et al.* (1999) ont défini un modèle où la variance résiduelle est définie à travers 30 classes de 10 jours chacune. La distribution des résidus au cours de la lactation tend toutefois à remettre en cause l'hypothèse d'homogénéité de la variance résiduelle au sein des classes (Jamrozik *et al.*, 1997b). Pour répondre à cette limite, il est possible d'exprimer la variance résiduelle comme une fonction du temps  $t$  (Robert-Granié *et al.*, 2004 ; Schaeffer, 2004) ou de la modéliser en supposant une structure autorégressive (Schaeffer, 2004). La modélisation de la variance résiduelle à travers une fonction continue du stade de lactation reste la solution optimale, car elle permet de s'affranchir de l'hypothèse d'homogénéité de la variance à l'intérieur d'une classe et d'avoir un nombre faible de paramètres.

Ces études, généralement préliminaires à la mise en place de l'évaluation génétique, n'intègrent pas (à l'exception de Schaeffer *et al.*, 2000) les sources d'hétérogénéité des variances souvent incluses dans les modèles lactation, telles que le troupeau ou l'année de production, dont l'importance a été montrée par Wiggans et VanRaden (1991) et Meuwissen *et al.* (1996). Ces sources de variation existent pourtant aussi dans les modèles contrôles élémentaires. Par exemple, Veerkamp et Goddard (1998) ont montré que les fonctions de covariance dépendaient du niveau de production du troupeau et plus spécifiquement que la variance génétique augmentait avec celui-ci. Ces résultats traduisent l'existence d'une hétérogénéité de variances liée au troupeau. Cependant, pour des raisons calculatoires, la prise en compte du niveau de production du troupeau est limitée à quelques classes (4 dans l'étude en question). Selon Jensen (2001), les méthodes se contentant de définir le niveau de production du troupeau comme covariable ne sont toutefois pas optimales car le niveau de production du troupeau dépend étroitement des données analysées. Il est donc nécessaire de corriger pour les variances hétérogènes simultanément à l'évaluation génétique, ou à défaut d'effectuer une correction des données préalablement à l'évaluation génétique (Kistemaker and Schaeffer, 1998).

Lorsque nous nous intéressons à la prise en compte de l'hétérogénéité des variances résiduelles dans les modèles contrôles élémentaires actuellement utilisés, nous constatons qu'il subsiste toujours les 2 approches : précorrection *vs* intégration dans le modèle. Le Canada précorrigé les performances pour l'hétérogénéité des variances entre troupeaux – jour de contrôle – rang de lactation préalablement à l'évaluation génétique selon la procédure décrite par Schaeffer *et al.*, (2000). Une approche de précorrection des données est également utilisée en Allemagne et en Italie. Au contraire, les Pays-Bas intègre un ajustement pour les variances hétérogènes directement au modèle d'évaluation en ajustant à l'aide d'un facteur de correction la partie « résiduelle » du modèle définie comme la différence entre la performance et la somme des effets fixes et des effets génétiques.

Parmi les autres sources d'hétérogénéité de variance connue, le Canada a intégré à son évaluation génétique la prise en compte de la précision des protocoles de collectes des performances. La variance résiduelle  $y$  est modulée en fonction de la précision des données journalières, en la divisant par la précision du protocole en question, *e.g.* 0.89 dans le cas d'un protocole AT *vs* 1.00 avec un protocole A. Ainsi, les données recueillies suivant des protocoles allégés ont des variances résiduelles plus importantes.

Le développement d'évaluation génétique basée sur les contrôles élémentaires est donc délicat à mettre en œuvre compte tenu de la complexité du modèle statistique requis, du nombre d'effets à inclure dans le modèle, et surtout du nombre de données à traiter. Outre les aspects techniques, les pays précurseurs dans le développement de ce type d'évaluation ont eu maille à partir avec l'université Cornell qui a déposé un brevet sur les méthodes de management du troupeau laitier en février 1993. Ce brevet déposé en Amérique du Nord et en Europe porte sur l'analyse et la valorisation des données à l'échelle du troupeau. L'intégration de l'effet HTD dans les modèles contrôles élémentaires empiète selon cette université sur leur droit de propriété. En juin 2007, le bureau des brevets européen a rejeté la demande en appel faite par cette université et confirmé la décision de janvier 2005 de rejet du brevet, pour manque de caractère « inventif » par rapport à un document de Nichols datant de 1977. Toutefois, le brevet reste valable pour l'Amérique du Nord et impose au Canada de payer des droits afin d'utiliser cette méthodologie. Les risques financiers encourus sont une des explications du faible investissement de la France sur ce domaine pendant de nombreuses années.



## **Partie 2.4**

# **Le modèle « contrôles élémentaires » présenti en France**



Outre l'attrait d'une évaluation basée sur une méthodologie plus performante incluant une prise en compte plus fine des effets d'environnement et un impact variable des effets aléatoires au cours de la lactation, l'évaluation génétique sur les contrôles élémentaires présente un intérêt important en terme d'image des évaluations nationales dans un but commercial. En effet, la majorité des pays jouant un rôle majeur dans les échanges internationaux de matériel génétique sont passés à une évaluation génétique des caractères laitiers basée sur les données élémentaires. Au niveau des évaluations internationales, comme expliqué dans la partie 1.2, ceci joue un rôle important sur le classement des animaux à travers les corrélations entre pays. Ainsi, Schaeffer *et al.* (2000) mentionnaient que les corrélations entre le Canada et l'Allemagne avaient augmenté suite au changement de modèle canadien pour se rapprocher du modèle allemand. Le développement d'une évaluation génétique basée sur les contrôles élémentaires pour les caractères laitiers est une étape nécessaire pour que la France conserve son rang parmi les pays leaders de la génétique bovine laitière.

L'intérêt de la France pour les modèles contrôles élémentaires, notamment sur les aspects méthodologiques, s'est traduit par une thèse soutenue par Florence Jaffrézic en 2001. Les premiers travaux sur le développement d'un modèle contrôles élémentaires en France n'ont toutefois démarré qu'en 2002. Ceux-ci ont principalement consisté à définir le modèle d'évaluation et plus particulièrement la modélisation des différentes régressions fixes et aléatoires. Une méthodologie permettant de combiner les paramètres génétiques estimés sur différents fichiers a aussi été mise au point. Les paramètres génétiques des trois principales races laitières françaises ont ainsi été estimés.

Comme l'ensemble des pays, la France a procédé par étapes pour définir son modèle d'évaluation et estimer les paramètres génétiques. Nous retracerons cette démarche d'un point de vue chronologique en présentant tout d'abord les grandes lignes du modèle pressenti, puis la phase d'estimation des paramètres génétiques en 1<sup>ère</sup> lactation (Druet *et al.*, 2003), son extension aux 3 premières lactations, en incluant les contrôles enregistrés jusqu'à 335 jours de lactation (Druet *et al.*, 2005) et enfin le modèle finalement retenu.

## 1. Le modèle pressenti initialement

Ce modèle unicaractère est défini par race. D'une façon générale, la performance journalière peut être décrite suivant l'équation 2.38.

(pour rappel :  $production = HTD + courbes\ fixes + effets\ fixes + génét. + env.\ perm. + erreur$  )

### 1.1. Les effets « fixes »

Parmi les effets fixes, le modèle prend en compte un effet HTD défini indépendamment du rang de lactation. Cet effet est constitué à partir des contrôles de vaches en 1<sup>ère</sup>, 2<sup>ème</sup> ou 3<sup>ème</sup> lactation et ayant un stade de lactation entre 7 et 335 jours. Trois vaches sont nécessaires au minimum pour définir l'effet HTD afin de ne pas écarter des évaluations génétiques les animaux provenant de troupeaux ayant un très faible effectif ou composés de plusieurs races. Toutefois, un effectif minimum de cinq aurait été plus adéquat pour garantir la fiabilité des effets HTD estimés. Un regroupement des données de deux contrôles successifs n'est toutefois pas envisageable car l'effet HTD représente l'environnement à court-terme (notamment celui de l'alimentation).

Afin de définir le type de courbe utilisé pour modéliser la partie fixe de la courbe de lactation, Druet *et al.* (2003) ont étudié à partir de données de premières lactations différentes fonctions : une courbe de Wilmink (3 paramètres), une courbe d'Ali et Schaeffer (5 paramètres), un polynôme de Legendre d'ordre 4 (5 paramètres), une spline de régression avec 6 nœuds (6 paramètres) et une courbe non paramétrique avec des classes de 5 jours (60 paramètres). Chacune de ces fonctions ( $f_1$ ) a été utilisée pour modéliser deux effets variants suivant le stade de lactation : l'effet de l'âge et du mois de vêlage définis tous deux en interaction avec le rang de lactation. De façon similaire, l'effet de la gestation au-delà de 100 jours après l'insémination fécondante et jusqu'à 265 jours est modélisé par une fonction  $f_2$  définie à partir d'un polynôme de Legendre d'ordre 4, une spline de régression avec 4, 5 ou 6 nœuds ou bien une courbe non paramétrique avec des classes de 5 jours. Le modèle d'analyse pour les effets fixes pour les premières lactations était donc :

$$y_{ijkltr} = HTD_j + f_1(age_k, DIM_t) + f_1(mois_l, DIM_t) + f_2(DCC_r) + e_{ijkltr} \quad (2.39)$$

où  $y_{ijkltr}$  est le contrôle journalier de l'animal  $i$ , à un stade de lactation  $t$  et de gestation  $t'$ ,  
 avec un effet troupeau  $j$ , un âge au vêlage  $k$  et un mois de vêlage  $l$ ,  
 $HTD_j$  est l'effet fixe troupeau – jour de contrôle  $j$ ,  
 $f_1$  est la fonction testée pour modéliser l'effet du stade de lactation,  
 $age_k$  est la classe d'âge au vêlage  $k$ ,  
 $DIM_t$  est le stade de lactation (Days In Milk)  $t$ ,  
 $mois_l$  est la classe de mois de vêlage  $l$ ,  
 $f_2$  est la fonction testée pour modéliser l'effet du stade de gestation,  
 $DCC_r$  est le stade de gestation (Days Carried Calf)  $t'$ ,  
 $e_{ijkltr}$  est la valeur résiduelle de la performance de l'animal  $i$ , obtenue au stade de lactation  $t$  et de gestation  $t'$ , avec un effet troupeau  $j$ , un âge au vêlage  $k$  et un mois de vêlage  $l$ .

Les courbes non-paramétriques présentaient, comme attendu, le meilleur ajustement sur l'effet du stade de lactation d'après l'analyse des résiduelles puisque par construction, la moyenne des résiduelles de la classe est nulle. La grande flexibilité de ce type de courbe procurée par le nombre élevé de paramètres entraîne cependant un coût informatique important et une sensibilité accrue aux données. Druet *et al.* (2003) observent des variations locales importantes sur la fin de la courbe de gestation ne reflétant pas la réalité biologique, certainement dues à un manque de données. Lorsque les courbes sont modélisées à partir de splines, les résiduelles sont en moyenne quasi nulles au cours de la lactation avec de très légères déviations en début et fin de lactation. Avec ce type de modélisation, les données influencent la courbe sur le segment en question, mais aussi sur les segments adjacents, ce qui évite l'inconvénient du comportement local parfois surprenant des courbes non-paramétriques. Quant aux trois autres types de courbes, elles présentent toutes des biais importants, avec des effets de bord caractéristiques pour les polynômes de Legendre.

A partir de ces résultats, il a été choisi de modéliser la forme de la courbe de la lactation pour chaque race et chaque caractère comme une somme de courbes semi-paramétriques. Dans le modèle pressenti pour les trois premières lactations, les courbes modélisant l'effet du stade de

lactation pour la classe d'âge au vêlage, le mois de mise bas ainsi que la durée de tarissement sont définies par région, par année et par rang de lactation. L'effet de la durée de tarissement a été ajouté avec la prise en compte des données de productions des multipares. D'après les résultats obtenus sur les tests préliminaires, le choix s'est porté sur des splines avec 6 nœuds positionnés respectivement à 7, 20, 50, 135, 245 et 335 jours de lactation. La courbe modélisant l'effet de la gestation sur la lactation est définie en fonction du stade de gestation pour une région, une année et un rang de lactation donnés. L'effet de la gestation étant supposé nul sur la production avant 100 jours, une spline avec 4 nœuds, positionnés respectivement à 100, 150, 200, et 265 jours de gestation a été utilisée. L'effet de la gestation est déterminé à partir de la date de vêlage et la durée de gestation moyenne de la race pour l'ensemble des lactations achevées, et à partir de la date de la dernière insémination, supposée fécondante, pour les lactations en cours.

## 1.2. Les effets « aléatoires »

Compte tenu du nombre d'animaux qui doivent être pris en compte dans les évaluations génétiques de la race Holstein, l'idée principale était de limiter au maximum le nombre d'effets aléatoires nécessaires à la modélisation des effets génétique, d'environnement permanent et troupeau – année. Un nombre raisonnable d'effets génétiques et d'environnement permanent est paru souhaitable. Ce nombre a été fixé à 6 (2 paramètres x 3 lactation) lors de l'étape d'estimation des paramètres génétiques décrite ci-dessous. Afin de conserver une flexibilité équivalente des deux effets et éviter des problèmes de transfert de variance entre ceux-ci, ces deux effets sont modélisés avec des modèles strictement identiques comme le recommandent Pool et Meuwissen (2000).

## 1.3. La résiduelle

Dans le modèle lactation, les hétérogénéités des performances entre région – année et troupeau – année sont prises en compte grâce à la modélisation de variances résiduelles hétérogènes. Afin de conserver cet avantage, un modèle similaire est souhaité pour le modèle contrôles élémentaires. Ce dernier doit cependant intégrer une nouvelle source de variance évidente : le stade de lactation.

# 2. L'estimation des paramètres génétiques

## 2.1. Méthode d'estimation employée

Les paramètres génétiques ont été estimés à partir du programme de Misztal *et al.* (2002) basé sur un algorithme AI-REML (Average Information-REML) modifié afin de combiner les estimations de l'AI-REML avec ceux d'un algorithme EM-REML (Expectation Maximisation-REML) en cas de matrices non-définies positives et permettre la prise en compte des modèles de régression aléatoire. Outre le faible nombre d'itérations nécessaire avec l'algorithme AI, celui-ci permet également d'obtenir les écarts-types des paramètres estimés. Le logiciel a été adapté pour pouvoir combiner l'information de plusieurs fichiers distincts afin de pouvoir baser l'estimation des paramètres génétiques sur un plus grand nombre de données suivant les approches proposées par Babb (1986) et Yerex (1988). En combinant les

matrices de coefficients obtenues pour chacun des fichiers ainsi que le membre de droite des équations du modèle mixte à chaque itération, il est possible d'estimer un seul jeu de paramètres génétiques en cumulant l'ensemble des informations en ignorant le fait que les données issues de fichiers différents proviennent d'animaux apparentés. Cette approche, développée spécifiquement pour l'application aux données de contrôles élémentaires, permet d'accroître la précision des paramètres estimés. Huit à dix fichiers pour les 3 principales races laitières françaises, contenant chacun près de 100 000 contrôles élémentaires ont ainsi été traités. Dans leur étude, Druet *et al.* (2003) ont quantifié le gain de précision lié à l'utilisation de cette approche par rapport à un traitement indépendant des échantillons. Cette méthode a permis de réduire l'écart-type par un facteur variant entre 1.16 et 3.25 selon l'échantillon.

## 2.2. L'approche utilisée

La phase d'estimation des paramètres génétiques est la partie la plus coûteuse du développement d'une nouvelle évaluation génétique, aussi bien en terme de temps de calcul qu'au niveau des ressources mémoire. Pour aboutir à des paramètres génétiques sur les 3 premières lactations dans l'ensemble des races et des caractères analysés, la démarche a été scindée pour les effets génétique et d'environnement permanent en trois étapes (Figure 2.7).

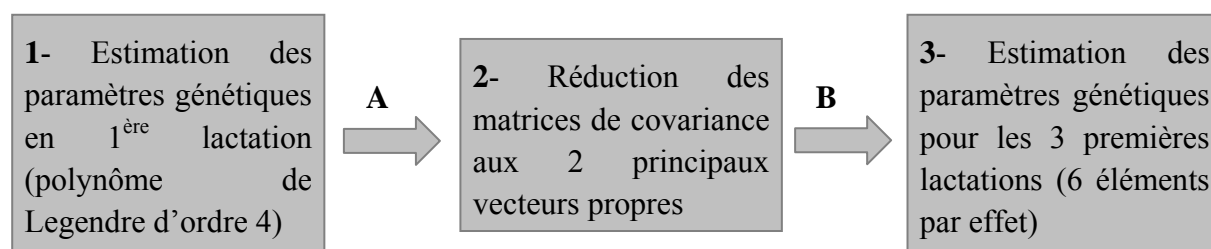


Figure 2.7. Schéma de l'approche utilisée pour estimer les paramètres pour les effets génétique et d'environnement permanent du modèle contrôles élémentaires.

Cette approche est permise par différents éléments. Tout d'abord, l'étape A est basée sur l'importance relative de la part de variance contrôlée par les principaux vecteurs propres des matrices de (co)variances génétique et d'environnement permanent. Les résultats d'une analyse préliminaire avaient permis de montrer que les 3 premiers vecteurs propres obtenus pour l'effet génétique et celui d'environnement permanent avec un modèle non structuré de rang 10 (équivalent à un modèle multi-caractère) étaient similaires aux 3 premiers vecteurs propres obtenus avec une régression aléatoire utilisant des polynômes de Legendre d'ordre 4. D'après les résultats obtenus par Druet *et al.* (2005), l'impact attendu de la réduction du rang de la matrice de covariance aux 2 principaux vecteurs propres est faible sur l'effet génétique. En effet, le pourcentage de variance expliquée par ces vecteurs varie suivant les caractères entre 96.5 et 99.6%. Le 1<sup>er</sup> vecteur propre, qui lorsque nous le traçons en fonction du stade de lactation est globalement constant au cours de la lactation, peut être interprété comme le potentiel de production moyen sur la lactation. Le second vecteur, de trajectoire oblique, est interprété comme une mesure de la persistance opposant le début et la fin de la lactation. La synthèse de l'information sous forme d'un nombre réduit de composantes permet de limiter les besoins calculatoires mais également d'accroître la robustesse du modèle et d'en simplifier

l'interprétation puisque les composantes ainsi conservées ont une signification biologique claire. La situation est toutefois différente avec l'effet d'environnement permanent. En terme de part de la variance expliquée par les 2 principaux vecteurs propres, Druet *et al.* (2005) obtenaient des valeurs comprises entre 83.0 et 90.8% suivant les caractères. Ces résultats sont cependant similaires à ceux de Pool *et al.* (2000). Malgré la perte d'information entraînée par la sélection des 2 premiers vecteurs propres, Druet *et al.* (2005) estiment que la part de variation totale prise en compte dans le modèle est largement suffisante et n'entraîne que des différences mineures. L'impact des vecteurs propres écartés se concentrait sur de petites périodes de la lactation, au début et à la fin de la courbe de lactation et étaient notamment responsables des effets de bord classiquement observés avec les polynômes de Legendre. La part de variance non conservée à travers les vecteurs propres est alors ajoutée à la résiduelle du modèle avant réduction de rang.

L'étape B repose sur l'utilisation des vecteurs propres de la 1<sup>ère</sup> lactation comme covariables pour les 3 lactations dans le modèle de régression aléatoire à la place de covariables classiques telles que les polynômes de Legendre. Elle s'appuie sur une étude de De Roos (*communication personnelle*) qui montrait que les 2 premiers vecteurs propres obtenus pour chacune des 3 premières lactations étaient similaires. L'utilisation de vecteurs propres permet plus de souplesse dans la modélisation des effets aléatoires comparativement à des fonctions ayant un nombre de paramètres raisonnables d'un point de vue calculatoire telles que les courbes de Wilmlink ou les polynômes de Legendre d'ordre 2.

Cette approche porte sur l'estimation des paramètres de l'effet génétique et celui d'environnement permanent. Pour l'effet troupeau – année, il a été choisi de le modéliser avec un polynôme de Legendre d'ordre 3 pour chacune des lactations.

### **2.3. Les paramètres génétiques de la 1<sup>ère</sup> lactation**

Les courbes de variances de la quantité de lait en race Holstein obtenues à partir des données de 1<sup>ère</sup> lactation sont selon Druet *et al.* (2003) similaires à de nombreux autres auteurs utilisant des polynômes de Legendre. L'héritabilité suit une courbe en cloche et est située dans les valeurs moyennes d'études équivalentes. Elle est minimale en début de lactation et maximale vers 200 jours. Outre l'héritabilité, les variances obtenues sont représentées sur la figure 2.8. Nous pouvons également y voir l'impact de la réduction du rang des matrices de covariance. La variance de l'effet génétique est alors légèrement réduite en début et fin de lactation. Les changements sont plus conséquents sur l'effet d'environnement permanent comme laissait présager la plus faible part de variance captée par les 2 principaux vecteurs propres. Pour maintenir la variance totale constante, la part de la variance « perdue » est transférée à la résiduelle.

En 1<sup>ère</sup> lactation, les corrélations obtenues par Druet *et al.* (2003) sont élevées au milieu de la lactation. Par exemple, la corrélation génétique entre le 150<sup>ème</sup> jour de lactation et les autres stades de lactation est supérieure à 0.90 entre le 45<sup>ème</sup> jour et le 280<sup>ème</sup>. Ces auteurs jugent les corrélations obtenues cohérentes avec celles d'analyses multicaractères (Meyer *et al.*, 1989 ; Pander *et al.*, 1992 ; Kettunen *et al.*, 1998) et de modèles de régression aléatoire (Brotherstone *et al.*, 2000 ; Olori *et al.*, 1999).

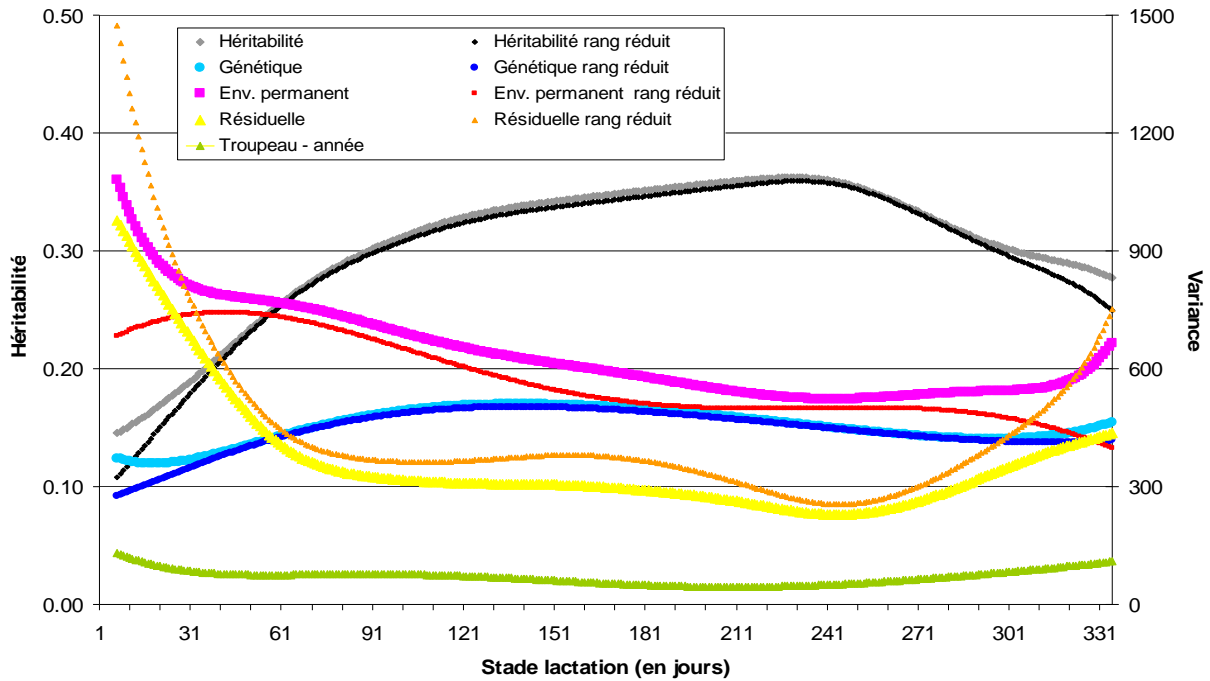


Figure 2.8. Héritabilité et Variance du caractère 'quantité de lait' en 1<sup>ère</sup> lactation pour la race Holstein avant et après réduction du rang de la matrice de covariance<sub>335x335</sub> aux 2 principaux vecteurs propres.

Si nous comparons les corrélations génétiques obtenues entre les différents stades de lactation avec la matrice de plein rang d'une part et celles obtenues avec la matrice de rang réduit d'autre part, nous notons peu de différences liées à la réduction du rang (Figure 2.9 A). Au contraire, pour l'effet d'environnement permanent (Figure 2.9 B), nous observons des différences importantes entre les corrélations obtenues avec une matrice de plein rang vs une matrice de rang réduit. Les valeurs des corrélations augmentent en moyenne de 0.08 à cause de la réduction du rang de la matrice de covariance.

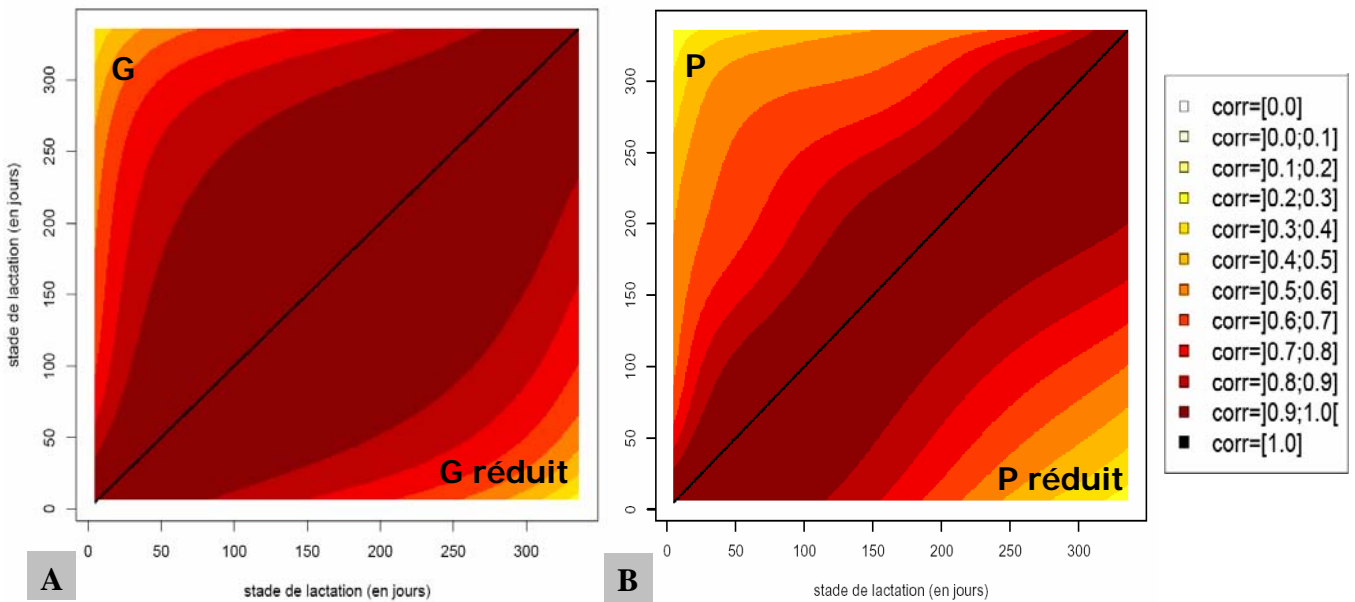


Figure 2.9. Corrélations génétiques **G** (partie A) et de l'effet d'environnement permanent **P** (partie B) entre les différents stades de lactation du caractère 'quantité de lait' pour la 1<sup>ère</sup> lactation de la race Holstein avant (triangle supérieur) et après (triangle inférieur) réduction du rang de la matrice de covariance<sub>335x335</sub> aux 2 principaux vecteurs propres.

## 2.4. Les paramètres génétiques des 3 premières lactations

Les profils des courbes de variances obtenues par Druet *et al.* (2005) pour les 3 premières lactations sont semblables entre lactations, mais montrent cependant une augmentation des variances avec l'augmentation du rang de lactation. Les héritabilités obtenues avec les modèles contrôles élémentaires sont stables d'une lactation à l'autre (Figure 2.10) et cohérentes avec les valeurs des modèles lactations. Pour les quantités de lait, l'héritabilité est généralement comprise entre 0.25 et 0.40. Pour les quantités de matières, elle est, comme d'autres études (Liu *et al.*, 2000b ; Jakobsen *et al.*, 2002) légèrement inférieure mais reste la plupart du temps comprise entre 0.20 et 0.30. Comme attendu, les héritabilités des taux butyreux et protéique sont plus élevées avec globalement des valeurs comprises entre 0.30 et 0.60 ; l'héritabilité du taux cellulaires est plus faible et augmente avec le stade de lactation (entre 0.10 et 0.20), de manière semblable à ce que Rupp (2000) avait obtenu.

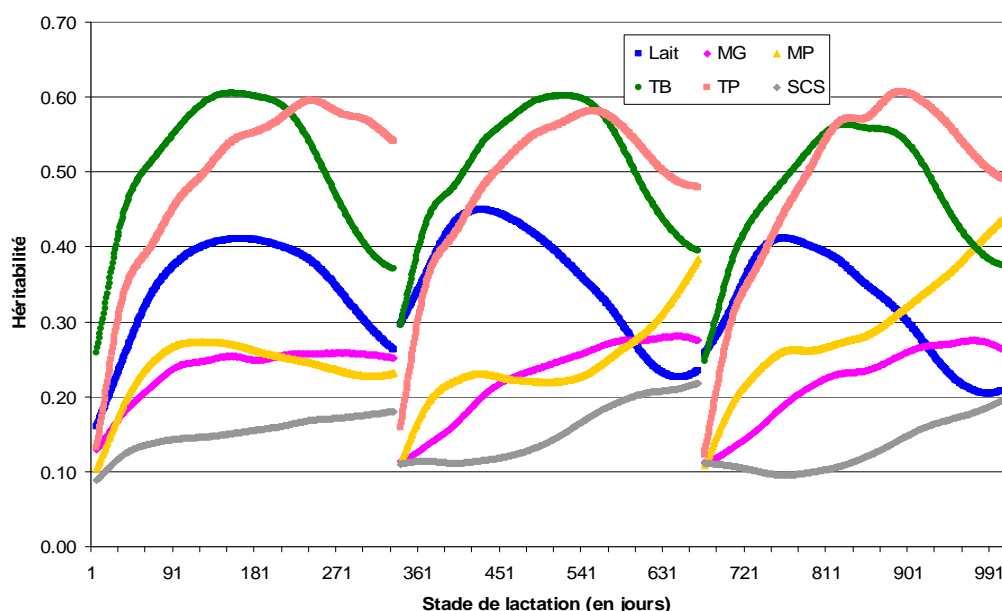


Figure 2.10. Héritabilités des caractères Lait, matières grasses et protéiques (MG et MP), taux butyreux et protéique (TB et TP) et scores de cellules somatiques (SCS) en race Holstein pour les 3 premières lactations.

Les corrélations entre les différents stades de lactation des 3 premières lactations pour la quantité de lait en race Holstein (Figure 2.11) montrent un niveau élevé pour l'effet génétique (triangle supérieur). Ainsi, l'effet génétique du 150<sup>ème</sup> jour de la 2<sup>ème</sup> lactation sera corrélé à plus de 0.90 avec les effets génétiques de la 1<sup>ère</sup> lactation entre 127 et 248 jours ; de la 2<sup>ème</sup> lactation entre 26 et 252 jours ; de la 3<sup>ème</sup> lactation entre 28 et 247 jours. Dans l'ensemble, les corrélations génétiques entre la 2<sup>ème</sup> et la 3<sup>ème</sup> lactation (carré central du haut) sont très élevées, et inférieures avec la 1<sup>ère</sup> lactation. Pour l'effet d'environnement permanent (triangle inférieur), les corrélations entre les différents stades des 3 premières lactations sont nettement inférieures (autour de 0.38 entre lactations différentes).

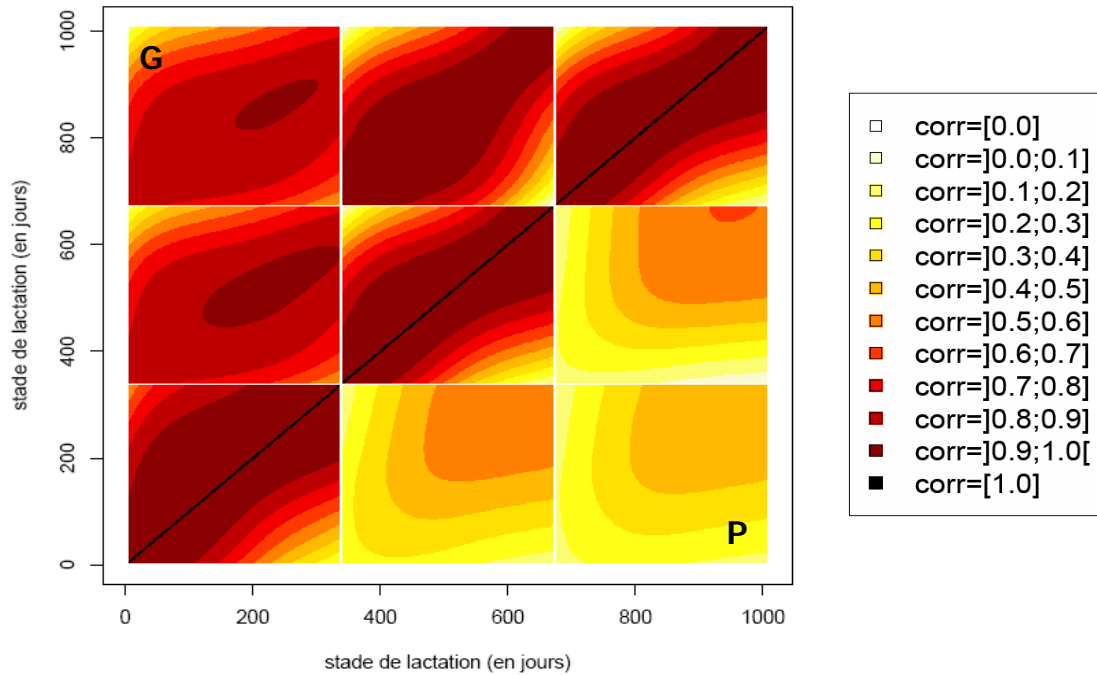


Figure 2.11. Corrélations entre les différents stades de lactation du caractère 'quantité de lait' pour les 3 premières lactations (de 335 jours chacune) de la race Holstein sur l'effet génétique **G** (triangle supérieur) et l'effet d'environnement permanent **P** (triangle inférieur).

### 3. Le modèle finalement retenu : le modèle de « référence »

#### 3.1. Les effets « fixes »

Par rapport au modèle pressenti initialement, le modèle retenu contient, outre les courbes modélisant l'effet du mois et de l'âge au vêlage sur la performance, ainsi que l'effet de la gestation, une courbe prenant en compte l'effet de la durée du tarissement. Cette dernière est déterminée à partir de l'intervalle entre la date du dernier contrôle enregistré de la lactation précédente à laquelle est ajoutée une durée forfaitaire (14 ou 28 jours suivant l'intervalle avec le contrôle suivant dans l'élevage) et la date de vêlage de la lactation considérée. Les 1<sup>ères</sup> lactations sont regroupées en une classe « forfaitaire » puisqu'il n'y a pas de tarissement précédent.

#### 3.2. Les effets « aléatoires »

Comme indiqué précédemment, les effets génétiques et d'environnement permanent du modèle d'évaluation retenu sont décrits de façon identique, afin de conserver une flexibilité équivalente de ces deux effets. Ils sont modélisés à partir de 2 vecteurs propres pour chacune des lactations conduisant ainsi à six valeurs génétiques pouvant être interprétées comme des mesures du niveau de production d'une part et du niveau de persistance d'autre part pour les différentes lactations. La matrice de covariance de ces deux effets est donc de dimension (6, 6). Pour la composante génétique, de fortes corrélations ont été obtenues entre les coefficients aléatoires du premier vecteur propre (niveau de production) de la 2<sup>ème</sup> et 3<sup>ème</sup> lactation (mais également entre de deuxième vecteur propre (persistance) de la 2<sup>ème</sup> et 3<sup>ème</sup> lactation) : entre 0.95 et 0.99 pour presque l'ensemble des caractères. Le déterminisme génétique des caractères laitiers pour les 2<sup>ème</sup> et 3<sup>ème</sup> lactations semble donc très proche. Ces résultats



laissent présager d'une possible réduction ultérieure du rang de la matrice de covariance. Pour cette raison, les nombres d'effets aléatoires décrits dans l'équation 2.41 correspondant au modèle d'évaluation prévu au début de mon travail est mentionné comme étant potentiellement révisables à la baisse. Il sera toutefois nécessaire d'étudier l'impact des réductions supplémentaires de rang des matrices de covariance pour chacun des effets.

Une courbe prenant en compte la modification de la trajectoire de la lactation du fait de l'effet troupeau – année est également prise en compte dans le modèle français. Son importance est assez modeste comme en témoigne le faible pourcentage de variance totale au cours de la lactation pour le caractère 'quantité de lait' (5%) dû à cet effet. Des résultats similaires ont été obtenus dans les autres pays incluant un effet aléatoire troupeau – année dans leur modèle. Cet effet est modélisé différemment des effets génétique et environnement permanent. Un polynôme de Legendre d'ordre 3 était initialement pressenti pour modéliser l'effet troupeau – année pour chacune des lactations. Ce nombre a finalement été limité à 2 dans le modèle de référence pour aboutir à une matrice de covariance de taille 9 x 9. Là encore, nous chercherons à réduire le rang de cette matrice (cf partie 3.2).

### 3.3. La résiduelle

Dans l'étude de Druet *et al.* (2005), la résiduelle était modélisée par une fonction du stade de lactation à l'aide de l'exponentielle d'une spline de régression de 12 nœuds positionnés à intervalle régulier tous les 30 jours de lactation pour chacune des lactations. Cette modélisation procure une grande flexibilité sans imposer de forme et garantit que la variance résiduelle soit positive. Toutefois, l'hétérogénéité de variance liée au stade de lactation n'est pas le seul élément dont la prise en compte est prévue le modèle contrôles élémentaires français. Il semble souhaitable de conserver dans l'évaluation génétique les sources d'hétérogénéité de variance actuellement incluses dans le modèle lactation (Robert-Granié *et al.*, 1999), telles que l'effet troupeau – année ainsi que l'effet région – année – rang de lactation en ajoutant une interaction avec le type de protocole de contrôle laitier car la précision des données décroît avec l'augmentation de la flexibilité du protocole. La correction pour les hétérogénéités de variances qui se fera simultanément à l'évaluation génétique peut être écrite sous la forme :

$$\sigma_{e,jklqrt}^2 = \sigma_{e,t}^2 \cdot \sigma_{e,jklqr}^2 \quad (2.40)$$

où  $\sigma_{e,jklqrt}^2$  est la variance résiduelle dépendant du stade de lactation  $t$ , de la région  $j$ , de l'année  $k$ , du rang de lactation  $l$ , du troupeau  $q$  et du type de protocole de contrôle de performance  $r$ ,

$\sigma_{e,t}^2 = \exp\left(\sum_{i=1}^{12} c_{il} \psi_{it}\right)$  est la variance résiduelle qui dépend exclusivement du stade de lactation  $t$  et du rang de lactation  $l$  de l'animal déterminée à partir d'une spline à 12 nœuds ( $c_{il}$  étant le coefficient de la spline pour le  $i^{\text{ème}}$  nœud et  $\psi_{it}$  est la valeur de la covariable pour ce nœud au stade de lactation  $t$ ),

$\sigma_{e,jklqr}^2$  est la variance résiduelle modélisée en fonction de l'effet troupeau – année  $kq$ , de l'effet région – année – rang de lactation – type de protocole contrôle de performance  $ijklr$ ,

Au début de cette thèse, il était donc prévu que dans le modèle d'évaluation, la performance soit décrite comme :

$$Perf_{ijklmnopqrtt'} = HTD_i + \left. \begin{aligned} & spl_{DIM}(\text{région}_j \times \text{année}_k \times nlac_l \times \text{durée\_tarissement}_m, t) + \\ & spl_{DIM}(\text{région}_j \times \text{année}_k \times nlac_l \times \text{mois}_n, t) + \\ & spl_{DIM}(\text{région}_j \times \text{année}_k \times nlac_l \times \text{age}_o, t) + \\ & spl_{DCC}(\text{région}_j \times \text{année}_k \times nlac_l, t') + \\ & \sum_{a=1}^6 (\text{génétique}_{pa} \cdot v_{at}) + \sum_{b=1}^6 (\text{environ}'\_permanent_{pb} \cdot \zeta_{bt}) + \\ & \sum_{c=1}^9 ((\text{troupeau}_q \times \text{année}_k)_c \cdot \tau_{ct}) + e_{ijklmnopqrtt'} \end{aligned} \right\} \begin{array}{l} \text{Effets} \\ \text{fixes} \\ \text{Effets} \\ \text{aléatoires} \end{array} \quad (2.41)$$

où  $Perf_{ijklmnopqrtt'}$  est le contrôle journalier de l'animal  $p$ , à un stade de lactation  $t$  et de gestation  $t'$ , avec un effet troupeau – jour de contrôle  $i$ , dans la région  $j$ , l'année  $k$ , un rang de lactation  $l$ , une durée de tarissement  $m$ , un mois de vêlage  $n$ , un âge au vêlage  $o$ , un troupeau  $q$  et un protocole de contrôle de performance  $r$ ,

$HTD_i$  est l'effet fixe troupeau – jour de contrôle  $i$ ,

$spl_{DIM}$  est une spline avec 6 nœuds utilisée pour modéliser l'effet du stade de lactation,

$spl_{DCC}$  est une spline avec 4 nœuds utilisée pour modéliser l'effet du stade de gestation,  $génétique_{pa}$  est le  $a^{\text{ème}}$  effet génétique de l'animal  $p$ , le nombre d'effet génétique maximum étant de 6,

$v_{at}$  est la valeur du  $a^{\text{ème}}$  vecteur propre de l'effet génétique au stade de lactation  $t$ ,

$environ'_permanent_{pb}$  est le  $b^{\text{ème}}$  effet d'environnement permanent de l'animal  $p$ , avec au maximum 6 effets,

$\zeta_{bt}$  est la valeur du  $b^{\text{ème}}$  vecteur propre de l'effet environnement permanent pour un stade de lactation  $t$ ,

$(\text{troupeau}_q \times \text{année}_k)_c$  est le  $c^{\text{ème}}$  effet troupeau – année du troupeau  $q$  pour l'année  $k$ , avec au maximum 9 effets,

$\tau_{ct}$  est la valeur du  $c^{\text{ème}}$  vecteur propre de l'effet troupeau – année pour un stade de lactation  $t$ ,

$e_{ijklmnopqrtt'}$  est la valeur résiduelle de la performance et  $\log(\sigma_e)$  dépend du stade de lactation  $t$ , de l'interaction entre région  $j \times$  année  $k \times$  nlac  $l \times$  protocole CL  $r$ , et de l'interaction troupeau  $q \times$  année  $k$ .

Nous verrons dans le chapitre suivant qu'il est possible de simplifier assez nettement ce modèle.

## Références

- Albarrán-Portillo B. and Pollott G.E. 2008. Genetic Parameters derived from using a biological model of lactation on records of commercial dairy cows. *Journal of Dairy Science*. 91: 3639-3648.
- Albuquerque L.G. and Meyer K. 2001. Estimated of covariance functions for growth from birth to 630 days of age in Nelore cattle. *Journal of Animal Science*. 79: 2776-2789.
- Ali T.E. and Schaeffer L.R. 1987. Accounting for covariances among test day milk yields in dairy cows. *Canadian Journal of Animal Science*. 67: 637-644.
- Anang A., Mielenz N. and Schuler L. 2000. Genetic and phenotypic parameters for monthly egg production in White Leghorn hens. *Journal of Animal Breeding and Genetics*. 117: 407-415.
- Andersen S. and Pedersen B. 1996. Growth and food intake curves for group-housed gilts and castrated male pigs. *Animal Science*. 63: 457-464.
- Anderson S.M. Mao I.L. and Gill J.L. 1989. Effect of frequency and spacing of sampling on accuracy and precision of estimating total lactation milk yield and characteristics of the lactation curve. *Journal of Dairy Science*. 72: 2387-2394.
- Auran T. 1986. Studies on monthly and cumulative monthly milk yield records. III. Estimates of genetic parameters. *Acta Agriculturae Scandinavica* 26: 3-9.
- Auran T. et Mocquot J.C. 1974. Etudes sur la production laitière des bovins. II – Intérêt des lactations partielles pour la sélection. a) étude critique de différentes méthodes d'estimation des productions totales à partir des résultats de lactation partielle. *Annales de génétique et de sélection animale*. 6: 429-444.
- Averill T., Rekaya R. and Weigel K. 2006. Random regression models for male and female fertility evaluation using longitudinal binary data. *Journal of Dairy Science*. 89: 3681-3689.
- Babb J.S. 1986. Pooling maximum likelihood estimates of variance components obtained from subsets of unbalanced data. M.Sc. Thesis, Cornell University, Ithaca, United-States.
- Batra T.R. 1986. Comparison of two mathematical models in fitting lactation curves for pureline and crossline dairy cows. *Canadian Journal of Animal Science*. 66: 405-414.
- Barbat A., Bonaïti B. et Boichard D. 1995. Comparaison de 2 méthodes de précorrection des lactations courtes pour l'évaluation des reproducteurs laitiers. *Annales de Zootechnie*. 44: 161-172.
- Beever D.E., Rook A.J., France J., Dhanoa M.S. and Gill M. 1991. A review of empirical and mechanistic models of lactational performance by the dairy cow. *Livestock Production Science*. 29: 115-130.
- Bohmanova J., Miglior F., Jamrozik J., Misztal I. and Sullivan P.G. 2008. Comparison of random regression models with Legendre polynomials and linear splines for production traits and somatic cell score of Canadian Holstein cows. *Journal of Dairy Science*. 91: 3627-3638.
- Boichard D., Bonaïti B., Barbat A. and Mattalia S. 1995. Three methods to validate the estimation of genetic trend for dairy cattle. *Journal of Dairy Science*. 78: 475-486.
- Bonaïti B. 1982. Interaction génotype milieu sur la production laitière chez les bovins. II. Corrélation génétique entre niveaux de production et variation de l'héritabilité. *Annales de génétique et de sélection animale*. 14 441-452.
- Bonaïti B., Boichard D., Barbat A. and Mattalia S. 1993. Problems arising with genetic trend estimation in dairy cattle. *Interbull Bulletin*. 8.
- Bonaïti B., Boichard D., Verrier E., Ducrocq V., Barbat A. and Briend M. 1990. La méthode française d'évaluation génétique des reproducteurs laitiers. *INRA Productions Animales*. 3(2): 83-92.
- Bouloc, N. Delacroix J. and Dervishi V. 2003. Milk recording and automatic milking systems: features and simplification possibilities of recording procedures. In: Proceedings of the 33<sup>th</sup> ICAR session, Interlaken, Switzerland, May 26-31, 2002. *Performance recording of animals, state of the art 2002*. 2003. EAAP 107:57-63.
- Brody S.A., Ragsdale A.C. and Turner C.W. 1923. The rate of decline of milk secretion with the advance of the period of lactation. *Journal of General Physiology*. 5: 441-444.
- Brotherstone S. and Hill W.G. 1986. Heterogeneity of variance amongst herds for milk production. *Animal Production*. 42: 297-303.
- Brotherstone S., White I.M.S and Meyer K. 2000. Genetic modelling of daily milk yield using orthogonal polynomials and parametric curves. *Animal Science*. 70: 407-415.

## Chapitre 2

- Carabaño M.J., Diaz C., Ugarte C. and Serrano M. 2007. Exploring the use of random regression models with Legendre polynomials to analyze measures of volume of ejaculate in Holstein bulls *Journal of Dairy Science*. 90: 1044-1057.
- Cobby J.M. and Le Du Y.L.P. 1978. On fitting curves to lactation data. *Animal Production*. 26: 127-133.
- Congleton Jr. W.R. and Everett R.W. 1980. Error and bias in using the incomplete gamma function to describe lactation curves. *Journal of Dairy Science*. 73: 539-548.
- Coulon J.B. et Pérochon L. 2000. Evolution de la production laitière au cours de la lactation : modèle de prédiction chez la vache laitière. *INRA Productions Animales*. 13(5): 349-360.
- Danell B. 1982. Studies on lactation yield and individual test-day yields of Swedish dairy cows. I. Environmental influence and development of adjustment factors. *Acta Agriculturae Scandinavica* 32: 65-81.
- DeLorenzo M.A. and Wiggans G.R. 1986. Factors for Estimating Daily Yield of Milk, Fat, and Protein from a Single Milking for Herds Milked Twice a Day. *Journal of Dairy Science*. 69: 2386-2394.
- Deluyker H.A., Shumway R.H., Wecker W.E., Azari A.S. and Weaver L.D. 1990. Modeling daily milk yield in Holstein cows using time series analysis. *Journal of Dairy Science*. 73: 539-548.
- Dematawewa C.M.B., Pearson R.E. and VanRaden P.M. 2007. Modeling extended lactations of Holstein. *Journal of Dairy Science*. 90: 3924-3936.
- De Roos A.P.W., Harbers A.G.F and De Jong G. 2004. Random Herd Curves in a Test-Day Model for Milk, Fat, and Protein Production of Dairy Cattle in The Netherlands. *Journal of Dairy Science*. 87: 2693-2701.
- De Roos A.P.W., Koenen E.P.C., Harbers A.G.F and De Jong G. 2002. Model validation and rank reduction of covariance matrices in the random regression test-day model in The Netherlands. *Interbull Bulletin*. 29: 91-94.
- Dhanoa M.S. 1981. A note on an alternative form of the lactation model of Wood. *Animal Production*. 32: 349-351.
- Dhanoa M.S. and Le Du Y.L.P 1982. A partial adjustment model to describe the lactation curve of a dairy cow. *Animal Production*. 34: 243--247.
- Dijkstra J., France J., Dhanoa M.S., Maas J.A., Hanigan M.D. Rook A.J. and Beever D. 1997. A model to describe growth patterns of the mammary gland during pregnancy and lactation. *Journal of Dairy Science*. 80: 2340-2354.
- Druet T., Jaffrézic F., Boichard D. and 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*. 86: 2480-2490.
- Druet T., Jaffrézic F. and Ducrocq V. 2005. Estimation of genetic parameters for test day record of dairy traits in the first three lactations. *Genetics, Selection and Evolution*. 37: 257-271.
- Ducrocq V. 1987. An analysis of productive life in dairy cattle. Ph.D. Thesis, Cornell University, Ithaca, United-States.
- Ducrocq V. and Besbes B. 1993. Solution of multiple trait animal models with missing values on some traits. *Journal of Animal Breeding and Genetics*. 110: 81-92.
- Duval M. 2008. Modélisation et estimation de variances hétérogènes dans les modèles non linéaires mixtes. Thèse de Doctorat, AgroParisTech. Paris. France.
- Elston D.A., Glasbey C.A. and Neilson D.R. 1989. Non-parametric lactation curves. *Animal Production*. 48: 331-339.
- Everett R.W., Schmitz F., Wadell L.H. 1994. A test-day model for monitoring management and genetics in dairy cattle. *Journal of Dairy Science*. 77: Suppl.1, 267 Abstract..
- Foulley J.L., San Cristobal M., Gianola D. and Im S. 1992. Marginal likelihood and Bayesian approaches to the analysis of heterogeneous residual variances in mixed linear Gaussian models. *Computational Statistics and Data Analysis*. 13: 291-305.
- Gaines W.L. 1927. Persistency of lactation in dairy cows. Univ. Illinois Agric. Exp. Stn. Bull. 288: 355-424.
- Gengler N., Tijani A., Wiggans G.R. and Philpot J.C. 2001. Indirect estimation of (co)variance functions for test-day yields during first and second lactations in the United States. *Journal of Dairy Science*. 84: suppl. [<http://www.adsa.org/jds/jds9296E/jds9296.html>, consulté le 15 septembre 2008].
- Gipson T.A. and Grossman M. 1990. Lactation curves in dairy goats: a review. *Small Ruminant Research*. 3: 383-396.

## Chapitre 2

- Goodall E.A. 1986. A note on the use of a categorical variable to explain seasonality deviation from the lactation curve. *Animal Production*. 42: 153-156.
- Goodall E.A. and Sprevak D. 1984. A note on a stochastic model to describe the milk yield of a dairy cow. *Animal Production*. 38: 133-136.
- Goodall E.A. and Sprevak D. 1985. A Bayesian estimation of the lactation curve of a dairy cow. *Animal Production*. 40: 189-193.
- Green P.J. and Silverman B.W. 1994. Nonparametric regression and generalized linear models. A roughness penalty approach. Chapman & Hall, London.
- Grossman M. and Koops J.W. 1988. Multiphasic analysis of lactation curves in dairy cattle. *Journal of Dairy Science*. 71: 1598-1608.
- Grossman M. and Koops J.W. 2003. Modeling extended lactation curves of dairy cattle: A biological basis for the multiphasic approach. *Journal of Dairy Science*. 86: 988-998.
- Grossman M., Kuck A.L and Norton H.W. 1986. Lactation curve of purebred and crossbred dairy cattle. *Journal of Dairy Science*. 69: 195-203.
- Guo Z. and Swalve H.H. 1995. Modelling of the lactation curve as a sub-model in the evaluation of test day records. *Interbull Bulletin*. 11.
- Guo Z. and Swalve H.H. 1997. Comparison of different lactation curve sub-models in test day models. *Interbull Bulletin*. 16:75-79.
- Henderson C.R. Jr. 1982. Analysis of covariance in the mixed model: higher level, non homogeneous, and random regressions. *Biometrics*. 38: 623-640.
- Hill W.G., Edwards M.R. Ahmed M.K.A and Thompson R. 1983. Heritability of milk yield and composition at different levels of variability of production. *Animal Production*. 36: 56-68.
- ICAR (International Committee for Animal Recording). 2007. International Agreement of Recording Practices. <http://www.icar.org/> Accessed June. 4, 2008.
- ICAR (International Committee for Animal Recording). 2008. Yearly Milk Enquiry – On-line Database. <http://www.waap.it/enquiry/> Accessed September. 1, 2008.
- Interbull. 2001. Interbull guidelines for national and international genetic evaluation systems in dairy cattle with focus on production traits. *Interbull Bulletin*. 28, 30 p.
- Interbull. 2008. National Genetic Evaluation Systems Information. [http://www-interbull.slu.se/national\\_ges\\_info2/framesida-ges.htm](http://www-interbull.slu.se/national_ges_info2/framesida-ges.htm). Accessed July. 7, 2008.
- Institut de l'Élevage. 1993. Guide d'utilisation du référentiel courbe de lactation. *Compte rendu N°2086*. 69 p.
- Institut de l'Élevage-INRA. 2005. Gestion des protocoles de Contrôle Laitier dans le SIG. *Notes Indexation Bovine Laitière*. 2005-9.
- Institut de l'Élevage-INRA. 2007. Règles de publication des index laitiers des taureaux. *Notes Indexation Bovine Laitière*. 2007-3.
- ITEB (Institut Technique de l'Élevage Bovin). 1990. Etude des facteurs de variation des courbes de lactation et d'évolution des taux. *Compte rendu ITEB*. 122 p.
- Jaffrézic F. and Pletcher S.D. 2000. Statistical models for estimating the genetic basis of repeated measures and other function-valued traits. *Genetics*. 156: 913-922.
- Jakobsen J.H. 2000. Genetic correlations between the shape of the lactation curve and disease resistance in dairy cattle. Ph.D. Thesis, Danish Institute of Agricultural Science, Foulum, Denmark.
- Jakobsen J.H., Madsen P, Jensen J., Pedersen J., Christensen L.G. and Sorensen D.A. 2002. Genetic parameters for milk production and persistency for Danish Holstein estimated in random regression models using REML. *Journal of Dairy Science*. 85: 1607-1616.
- Jamrozik J., Gianola D. and Schaeffer L.R. 2001. Bayesian estimation of genetic parameters for test day records in dairy cattle using linear hierarchical models. *Livestock Production Science*. 71: 223-240.
- Jamrozik J., Jansen G., Schaeffer L.R. and Liu Z. 1998. Analysis of persistency of lactation calculated from a random regression test day model. *Interbull Bulletin*. 17: 64-69.
- Jamrozik J., Kistemaker G.J., Dekkers J.C.M. and Schaeffer L.R. 1997a. Comparison of possible covariates for use in a random regression model for analyses of test day yields. *Journal of Dairy Science*. 80: 2550-2556.
- Jamrozik J., Schaeffer L.R. and Dekkers J.C.M. 1997b. Genetic evaluation of dairy cattle using test day yields and random regression models. *Journal of Dairy Science*. 80: 1217-1226.

## Chapitre 2

- Jensen J. 2001. Genetic evaluation of dairy cattle using test-day models. *Journal of Dairy Science*. 84: 2803-2812.
- Jensen J., Mäntysaari E., Madsen P. and Thompson R. 1997. Residual maximum likelihood estimation of (co)variance components in multivariate mixed linear models using average information. *Journal of the Indian Society of Agricultural Statistics*. (Golden Jub. No.) 1997: 215-236.
- Johnson, D. L. 1996. Estimation of lactation yield from repeated measures of test day yields. *Proc. New-Zealand Society of Animal Production* 56:16-18.
- Jones L.P and Goddard M.E. 1990. Combining test day and full lactation records in prediction of breeding values: Five years experience with animal model for dairy evaluation in Australia. In: Proceedings of the 4<sup>th</sup> World Congress on Genetic Applied to Livestock Production, Edinburgh, Scotland. XIII: 382-385.
- Kalman, R. E. 1960. A New Approach to Linear Filtering and Prediction Problems. Transactions of the ASME. *Journal of Basic Engineering*. 82:35-45.
- Keown J.F., Everett R.W., Empet N.B. and Wadell L.H. 1986. Lactation curves. *Journal of Dairy Science*. 69: 769-781.
- Kettunen A., Mäntysaari E.A. and Pösö J. 2000. Estimation of genetic parameters for daily milk yield of primiparous Ayrshire cows by random regression test-day models. *Livestock Production Science*. 66: 251-261.
- Kettunen A., Mäntysaari E.A., Strandén I., Pösö J. and Lidauer M. 1998. Estimation of genetic parameters for first lactation test day milk production using random regression models. In: Proceedings of the 6<sup>th</sup> World Congress on Genetic Applied to Livestock Production, Armidale, Australia. XXIII: 307-310.
- Kirkpatrick M., Hill W.G. and Thompson R. 1994. Estimating the covariance structure of traits during growth and ageing illustrated with lactations in dairy cattle. *Genetics Research*. 64: 57-69.
- Kirkpatrick M., Lofsvold D. and Bulmer M. 1990. Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics*. 124: 979-993.
- Kistemaker G.J. and Schaeffer L.R., 1998. Adjustment for heterogeneous Herd-Test-Day variances. *Interbull Bulletin*. 17: 55-59.
- Knight C.H. and Wilde C.J. 1993. Mammary cell changes during pregnancy and lactation. *Livestock Production Science*. 35: 3-19.
- Koivula M., Nousiainen J.I., Nousiainen J. and Mäntysaari E.A. 2007. Use of herd solutions from a random regression test-day model for diagnostic dairy herd management. *Journal of Dairy Science*. 90: 2563-2568.
- Lawlor T.J., Tsuruta S., Klei L. and Misztal, I. 2002. Use of a random regression model to investigate changes in genetic parameters over time. In: Proceedings of the 7<sup>th</sup> World Congress on Genetic Applied to Livestock Production, Montpellier, France. XXXII: 235-238.
- Leclerc H, Delacroix J, Larroque H, Gallard Y, Mattalia S. 2005. Milk Recording: a comparison of the T, Z and standard methods (Z= Milk yield recorded on 2 milkings and the Contents on one alternate milking). In: Proceedings of the 34th ICAR session, Sousse, Tunisia, June 1-3, 2004. *Performance recording of animals, state of the art 2004*. 2005. EAAP 113:237-242.
- Lidauer M., Mäntysaari E.A. and Strandén I. 2003. Comparison of test-day models for genetic evaluation of production traits in dairy cattle. *Livestock Production Science*. 79: 73-86.
- Lidauer M., Mäntysaari E.A., Strandén I. and Pösö J. 2000. Multiple-trait random regression test-day model for all lactations. *Interbull Bulletin*. 25: 81-84.
- Liu Z., Reents R., Reinhardt F. and Kuwan K. 2000a. Approaches to estimating daily yield from single milk testing schemes and use AM-PM records in test-day model genetic evaluation in dairy cattle. *Journal of Dairy Science*. 83: 2672-2682.
- Liu Z., Reinhardt F. and Reents R. 2000b. Estimating parameters of a random regression test day model for first three lactation milk production traits using the covariance function approach. *Interbull Bulletin*. 25:74-80.
- Luo P.T., Yang R.Q. and Yang N. 2007. Estimation of genetic parameters for cumulative egg numbers in a broiler dam line by using a random regression model. *Poultry Science*. 86: 30-36.
- Macciotta N.P.P., Vicario D. and Cappio-Borlino A. 2005. Detection of different shapes of lactation curve for milk yield in dairy cattle by empirical mathematical models. *Journal of Dairy Science*. 88: 1178-1191.
- Mäntysaari E. 2002. Combining test day and full lactation records in prediction of breeding values. In: Proceedings of the 7<sup>th</sup> World Congress on Genetic Applied to Livestock Production, Montpellier, France. XXIX: 35-42.

## Chapitre 2

- Masselin S., Sauvant D., Chapoutot P. et Milan D. 1987. Les modèles d'ajustement des courbes de lactation. *Annales de Zootechnie*. 36: 171-206.
- Mayeres P., Stoll J., Bormann J., Reents R. and Gengler N. 2004. Prediction of daily milk, fat and protein production by a random regression test-day model. *Journal of Dairy Science*. 87: 1925-1933.
- Magnabosco C., Lobo R.B. de U. and Famula T.R. 2000. Bayesian inference for genetic parameter estimation on growth traits for Nelore cattle in Brazil, using the Gibbs sampler *Journal of Animal Breeding and Genetics*. 117: 169-188.
- McKay L.R., Schaeffer L.R. and McMillan I. 2002. Analysis of growth curves in rainbow trout using random regression. In: Proceedings of the 7<sup>th</sup> World Congress on Genetic Applied to Livestock Production, Montpellier, France. XXX: 505-508.
- Meuwissen T.H.E., De Jong G. and Engel B. 1996. Joint estimation of breeding values and heterogeneous variances of large data files. *Journal of Dairy Science*. 79: 310-316.
- Meuwissen T.H.E. and Van der Werf J.H.J. 1993. Impact of heterogeneous within herd variances on dairy cattle breeding schemes. *Livestock Production Science*. 33: 31-41.
- Meyer K. 1985. Genetic parameters for dairy production of Australian black and white cows. *Livestock Production Science*. 12: 205-219.
- Meyer K. 1998. Estimating covariance functions for longitudinal data using a random regression model. *Genetics, Selection and Evolution*. 29: 97-116.
- Meyer K. 1991. Estimating variances and covariances for multivariate animal models by restricted maximum likelihood. *Genetics, Selection and Evolution*. 23: 67-83.
- Meyer K. 1997. An 'average information' restricted maximum likelihood algorithm for estimating reduced rank genetic covariance matrices or covariance functions for animal models with equal design matrices. *Genetics, Selection and Evolution*. 29: 97-116.
- Meyer K. 1999. Estimates of genetic and phenotypic covariance function for postweaning growth and mature weight of beef cows. *Journal of Animal Breeding and Genetics*. 116: 181-205.
- Meyer K. 2000. Random regressions to model phenotypic variation in monthly weights of Australian beef cows. *Livestock Production Science*. 65: 19-38.
- Meyer K. 2001. Estimates of direct and maternal covariance functions for growth of Australian beef calves from birth to weaning. *Genetics, Selection and Evolution*. 33: 487-514.
- Meyer K. 2004. Scope for a random regression model in genetic evaluation of beef cattle for growth. *Livestock Production Science*. 86: 69-83.
- Meyer K. 2005. Random regression analyses using B-splines to model growth of Australian Angus cattle. *Genetics, Selection and Evolution*. 37: 473-500.
- Meyer K., Graser H.-U. and Hammond K. 1989. Estimates of genetic parameters for first lactation test day production of Australian black and white cows. *Livestock Production Science*. 21: 177-199.
- Meyer K., Hill W.G. 1997. Estimation of genetic and phenotypic covariance functions for longitudinal or 'repeated' records by restricted maximum likelihood. *Livestock Production Science*. 47: 185-200.
- Miglior F., Sullivan P.G. and Van Doormaal B.J. 2002. Preliminary analysis of Mendelian sampling terms for genetic evaluation validation. *Interbull Bulletin*. 28: 183-187.
- Misztal I. 2006. Properties of random regression models using linear splines. *Journal of Animal Breeding and Genetics*. 123: 74-80.
- Misztal I., Strabel T., Jamrozik J., Mäntysaari E.A. and Meuwissen T.H.E. 2000. Strategies for estimating the parameters needed for different test-day models. *Journal of Dairy Science*. 83: 1125-1134.
- Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T. and Lee D.H. 2002. BLUPF90 and related programs (BFG90). In: Proceedings of the 7<sup>th</sup> World Congress on Genetic Applied to Livestock Production, Montpellier, France. XXXIII: 743-744.
- Neal H.D.St.C. and Thornley J.H.M. 1983. The lactation curve in cattle: a mathematical model of the mammary gland. *Journal of Agricultural Science (Cambridge)*. 101: 389-400.
- Nedler J.A. 1966. Inverse polynomials, a useful group of multi-factor response functions. *Biometrics*. 22: 128-141.

## Chapitre 2

- Norman H.D., VanRaden P.M., Wright J.R. and Clay J.S. 1999. Comparison of test interval and best prediction methods for estimation of lactation yield from monthly, AM-PM and trimonthly testing. *Journal of Dairy Science*. 82: 438-444.
- NRS. 2008. Breeding value estimation of milk production traits with test-day model E7. <http://www.nrs.nl/index-eng.htm>. Accessed July. 15, 2008.
- Olori V.E. 1997. Utilisation of daily milk records in genetic evaluation of dairy cattle. Ph.D thesis, University of Edinburgh, Edinburgh, Scotland.
- Olori V.E., Brotherstone S., Hill W.G. and McGuirk B.J. 1999. Fit of standard models of the lactation curve to weekly records of milk production of cows in a single herd. *Livestock Production Science*. 58: 55-63.
- Pander B.L., Hill W.G. and Thompson R. 1992. Genetic parameters of test-day records of British Holstein-Friesian heifers. *Animal Production*. 55: 11-21.
- Papajcsik I.A. and Bodero J. 1988. Modelling lactation curves of Friesian cows in a sub-tropical climate. *Animal Production*. 47: 201-208.
- Pérochon L., Coulon J.B. and Lescourret F. 1996. Modelling lactation curves of dairy cows with emphasis on individual variability. *Animal Science*. 63: 189-200.
- Pletcher S.D. and Geyer C.J. 1999. The genetic analysis of age dependent traits: Modeling a character process. *Genetics*. 153: 825-833.
- Pollott G.E. 2000. A biological approach to lactation curve analysis for milk yield. *Journal of Dairy Science*. 83: 2448-2458.
- Pollott G.E. 2004. Deconstructing milk yield and composition during lactation using biologically based lactation models. *Journal of Dairy Science*. 87: 2375-2387.
- Pool M.H., Janss L.L.G. and Meuwissen T.H.E. 2000. Genetic parameters of Legendre polynomials for first parity lactation curves. *Journal of Dairy Science*. 83: 2640-2649.
- Pool M.H. and Meuwissen T.H.E. 2000. Reduction of the number of parameters needed for a polynomial random regression test-day model. *Livestock Production Science*. 64: 133-145.
- Pösö J., Mäntysaari E.A. and Kettunen A. 1996. Estimation of genetic parameters of test day production in Finnish Ayrshire cows. *Interbull Bulletin*. 14: 45-48.
- Poutous M., Briend M., Calomiti S., Doan D., Felgines C et Steier G. 1981. Méthode de calcul des index laitiers. *Bulletin technique d'information*. 361: 433-446.
- Poutous M. et Mocquot J.C. 1975. Etudes sur le production laitière des bovins. III. Relations entre critères de production, durée de lactation et intervalle entre le 1<sup>er</sup> et le 2<sup>ème</sup> vêlage. *Annales de génétique et de sélection animale*. 7: 181-189.
- Ptak E. and Schaeffer L.R. 1993. Use of test day yields for genetic evaluation of dairy sires and cows. *Livestock Production Science*. 34: 23-34.
- Ravagnolo O. and Misztal I. 2000. Genetic component of heat stress in dairy cattle, parameter estimation. *Journal of Dairy Science*. 83: 2126-2130.
- Reents R. and Dopp L. 1996. Genetic evaluation for dairy production traits with a test day model for multiple lactations. *Interbull Bulletin*. 14: 113-117.
- Rekaya R., Carabaño M.J and Toro M.A. 1999. Use of test day yields for the genetic evaluation of production traits in Holstein-Friesian cattle. *Livestock Production Science*. 57: 203-217.
- Rekaya R., Béjar F., Carabaño M.J. and Alenda R. 1995. Genetic parameters for test day measurements in Spanish Holstein-Friesian. *Interbull Bulletin*. 11.
- Robert C. 1996. Etude de quelques problèmes liés à la mise en oeuvre du REML en génétique quantitative. Thèse de Doctorat, Université Paul Sabatier. Toulouse. France.
- Robert-Granié C., Bonaïti B., Boichard D. and Barbat A. 1999. Accounting for variance heterogeneity in French dairy cattle genetic evaluation. *Livestock Production Science*. 60: 343-357.
- Robert-Granié C., Foulley J.L., Maza E. and Rupp R. 2004. Statistical analysis of somatic cell scores via mixed model methodology for longitudinal data. *Animal Research*. 53: 259-273.
- Robert-Granié C., Heude B. and Foulley J.L. 2002. Modelling the growth curve of Maine-Anjou beef cattle using heteroskedastic random coefficients models. *Genetics, Selection and Evolution*. 34: 423-445.



## Chapitre 2

- Robert-Granié C., Maza E., Rupp R. and Foulley J.L. 2002. Use of fractional polynomial for modelling somatic cell scores in dairy cattle. In: Proceedings of the 7<sup>th</sup> World Congress on Genetic Applied to Livestock Production, Montpellier, France. XXXII: 153-156.
- Rowlands G.J., Lucey S. and Russell A.M. 1982. A comparison of different models of the lactation curve in dairy cattle. *Animal Production*. 35: 135-144.
- Royston P. and Altman D.G. 1994. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Applied Statistics*. 43: 429-467.
- Rupp R. 2000. Analyse génétique de la résistance aux mammites chez les ruminants laitiers. Thèse de Doctorat, Institut National Agronomique Paris-Grignon. Paris. France.
- Rupp R. et Boichard D. 1997. Evaluation génétique des bovins laitiers sur les comptages de cellules somatiques pour l'amélioration de la résistance aux mammites. *Rencontres Recherches Ruminants*. 4: 211-214.
- Ruppert D., Wand M.P. and Carroll R.J. 2003. Semiparametric Regression. Cambridge University Press, New-York.
- San Cristobal M., Foulley J.L. and Manfredi E. 1993. Inference about multiplicative heteroskedastic components of variance in a mixed linear Gaussian model with an application to beef cattle breeding. *Genetics, Selection and Evolution*. 25: 3-30.
- Schaeffer L.R. 2004. Application of random regression models in animal breeding. *Livestock Production Science*. 86: 35-45.
- Schaeffer L.R. and Burnside E.B. 1976. AM-PM testing for estimating lactation yield. *Canadian Journal of Animal Science*. 56: 157-170.
- 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 5<sup>th</sup> World Congress on Genetic Applied to Livestock Production, Guelph, Canada. XVIII: 443-446.
- Schaeffer L.R. and Jamrozik J. 1996. Multiple-trait prediction of lactation yields for dairy cattle. *Journal of Dairy Science*. 79: 2044-2055.
- Schaeffer L.R. and Jamrozik J. 2008. Random regression models: a longitudinal perspective. *Journal of Animal Breeding and Genetics*. 125: 145-146.
- Schaeffer L.R., Jamrozik J., Kistemaker G.J. and van Doormaal B.J. 2000. Experience with a test-day model. *Journal of Dairy Science*. 83: 1135-1144.
- Schaeffer L.R., Minder C.E., McMillan I. and Burnside E.B. 1977. Nonlinear techniques for predicting 305-day lactation production of Holstein and Jerseys. *Journal of Dairy Science*. 60: 1636-1644.
- Schnyder U., Hofer A. and Kunzi N. 2001. Impact of variation in length of individual testing periods on estimation of covariance components of a random regression model for feed intake of growing pigs. *Journal of Animal Breeding and Genetics*. 118: 235-246.
- Schutz M.M., Hansen L.B., Steuernagel G.R. and Kuck A.L. 1990. Variation of Milk, Fat, Protein, and Somatic Cells for Dairy Cattle. *Journal of Dairy Science*. 73: 484-493.
- Schutz M.M., VanRaden P.M., Wiggans G.R. and Norman H.D. 1995. Standardization of lactation means of somatic cell scores for calculation of genetic evaluations. *Journal of Dairy Science*. 78: 1843-1854.
- Scott T.A., Yandell B., Zepeda L., Shaver R.D. and Smith T.R. 1996. Use of lactation curves for analysis of milk production data. *Journal of Dairy Science*. 79: 1881-1894.
- Serrano M., Diaz C. and Carabano M.J. 2006. Modelling sperm motility of Holstein bulls using random regression methodology. In: Proceedings of the 8<sup>th</sup> World Congress on Genetic Applied to Livestock Production, Belo Horizonte, Brazil. Communication 25-14.
- Serieys F. 1985. Interprétation de concentrations cellulaires du lait individuel de la vache pour le diagnostic de l'état d'infection mammaire. *Annales de Recherches Vétérinaires*. 16: 263-269.
- Sherchand L., McNew R.W., Kellogg D.W. and Johnson Z.B. 1995. Selection of a mathematical model to generate lactation curves using daily milk yields of Holstein cows. *Journal of Dairy Science*. 78: 2507-2513.
- Silvestre A.M., Petim-Batista F. and Colaço 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*. 89: 1813-1821.
- Sölkner J. and Fuchs W. 1987. A comparison of different measures of persistency with special respect to variation of test-day milk yields. *Livestock Production Science*. 16: 305-319.

## Chapitre 2

- Soyeurt H. 2008. Study of genetic variability of fatty acid profile in bovine milk and fat using mid-infrared spectrometry. Thèse de Doctorat, Faculté Universitaire des Sciences Agronomiques de Gembloux [[http://bictel-fusagx.ulg.ac.be/ETD-db/collection/available/FUSAGxetd-04222008-134044/unrestricted/Soyeurt\\_Bictel.pdf](http://bictel-fusagx.ulg.ac.be/ETD-db/collection/available/FUSAGxetd-04222008-134044/unrestricted/Soyeurt_Bictel.pdf), consulté le 6 mai 2008].
- Swalve H.H. 1995a. The Effect of Test Day Models on the Estimation of Genetic Parameters and Breeding Values for Dairy Yield Traits. *Journal of Dairy Science*. 74: 1690-1699.
- Swalve H.H. 1995b. Test day models in the analysis of dairy production data – a review. *Archiv Tierzucht*. 38: 591-612.
- Swalve H.H. 1995c. Genetic relationship between dairy lactation persistency and yield. *Journal of Animal Breeding and Genetics*. 112: 303-311.
- Swalve H.H. 1998. Use of test day records in genetic evaluation. In: Proceedings of the 6<sup>th</sup> World Congress on Genetic Applied to Livestock Production, Armidale, Australia. XXIII: 295-302.
- Swalve H.H. 2000. Theoretical basis and computational methods for different test-day genetic evaluation methods. *Journal of Dairy Science*. 83: 1115-1124.
- Stanton T.L., Jones L.R., Everett R.W. and Kachman S.D. 1992. Estimating milk, fat and protein lactation curves with a test day model. *Journal of Dairy Science*. 75: 1691-1700.
- Torres R.A. and Quaas R.L. 2001. Determination of covariance functions for lactation traits on dairy cattle using random coefficient regressions on B-splines. *Journal of Animal Science*. 79: Suppl.1, 112 Abstract.
- Turner C.W. 1926. A quantitative form of expressing persistency of milk or fat secretion. *Journal of Dairy Science*. 9: 203-214.
- Uribe H., Schaeffer L.R., Jamrozik J. and Lawlor T.J. 2000. Genetic evaluation of dairy cattle for conformation traits using random regression models. *Journal of Animal Breeding and Genetics*. 117: 247-259
- Val-Arreola D., Kebreab E., Dijkstra J. and France J. 2004. Study of the lactation curve in dairy cattle on farms in central Mexico. *Journal of Dairy Science*. 87: 3789-3799.
- Van der Werf J.H.J., Goddard M.E. and Meyer K. 1998. The use of covariance functions and random regressions for genetic evaluation of milk production based on test day records. *Journal of Dairy Science*. 81: 3300-3308.
- VanRaden P.M. 1997. Lactation yields and accuracies computed from test day yields and (Co)Variances by best prediction. *Journal of Dairy Science*. 80: 3015-3022.
- VanRaden P.M. 2004. Choice of scales for delivery of genetic evaluations to the public. *Interbull Bulletin*. 32: 118-121.
- Veerkamp R.F., Brotherstone S., and Meuwissen, T.H.E. 1999. Survival analysis using random regression models. *Interbull Bulletin*. 21: 36-40.
- Veerkamp R.F. and Goddard M.E. 1998. Covariance function across herd production levels for test day records on milk, fat and protein yields. *Journal of Dairy Science*. 81: 1690-1701.
- Verbyla, A.P., Cullis, B.R., Kenward, M.G. and Welham, S.J. 1999. The analysis of designed experiments and longitudinal data by using smoothing splines. *Applied Statistics*. 48: 269-311.
- White I.M.S., Thompson R. and Brotherstone S. 1999. Genetic and environmental smoothing of lactation curves with cubic splines. *Journal of Dairy Science*. 82: 632-638.
- Wiggans G.R. 2000. Issues in defining a genetic evaluation model. *Interbull Bulletin*. 26: 8-12.
- Wiggans G.R. and Goddard M.E. 1997. A computationally feasible test day model for genetic evaluation of yields traits in the United States. *Journal of Dairy Science*. 80: 1785-1800.
- Wiggans G.R. and Shook G.E. 1987. A lactation measure of somatic cell count. *Journal of Dairy Science*. 70: 2666-2672.
- Wiggans G.R. and VanRaden P.M. 1991. Method and effect of adjustment for heterogeneous variances. *Journal of Dairy Science*. 74: 4350-4357.
- Wilmink J.B.M. 1987. Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. *Livestock Production Science*. 16: 335-348.
- Wood P.D.P. 1967. Algebraic model of the lactation curve in cattle. *Nature*. 216: 164-165.
- Wood P.D.P. 1974. A note on the estimation of total lactation yield from production of a single day. *Animal Production*. 19: 393-396.
- Wood P.D.P. 1976. Algebraic models of the lactation curves for milk, fat and protein production, with the estimates of seasonal variation. *Animal Production*. 22: 35-40.

## Chapitre 2

- Wood P.D.P. 1977. The biometry of the lactation. *Journal of Agricultural Science*. 88: 333-339.
- Wood P.D.P. 1979. A simple model of lactation curves for milk yields, food requirement and body weight. *Animal Production*. 28: 55-63.
- Yerex R.P. 1988. Pooling restricted maximum likelihood estimates from data subsets under the animal model. Ph.D. Thesis, Cornell University, Ithaca, United-States.
- Zimmermann E. and Sommer H.. 1973. Zum Laktationsverlauf von Kühen in Hochleistungsherden und dessen Beeinflussung durch nichterbliche Faktoren. *Züchtungskunde*. 45:75-88.

# **Chapitre 3**

## **La mise en œuvre opérationnelle**

La mise en œuvre d'une nouvelle évaluation génétique dépend d'une part de la disponibilité des données mais également du logiciel permettant sa réalisation. Dans le cas des caractères de production laitière, le premier élément n'est pas le facteur limitant en France puisque les données de production par jour de contrôle sont sauvegardées exhaustivement depuis 1987 dans la base nationale du SIG (1994 pour les comptages leucocytaires). Par contre, le logiciel permettant de réaliser une évaluation génétique suivant le modèle retenu par la France n'existait pas. Il a donc été nécessaire de développer un logiciel d'évaluation génétique pour répondre à nos besoins spécifiques d'une part en terme de modélisation (utilisation de splines, plusieurs effets modélisés par des régressions aléatoires, variances résiduelles hétérogènes incluses dans le modèle) et d'autre part qui soit capable de traiter plusieurs centaines de millions de données simultanément. Le développement d'un logiciel reste toutefois une étape généralement longue, fastidieuse et semée d'embûches. Pour garantir l'exactitude du logiciel, une étape de validation a été entreprise au début de ma thèse. Une approche novatrice, basée sur la simulation, a été développée et est présentée dans la 1<sup>ère</sup> partie de ce chapitre.

Une fois l'exactitude du logiciel d'évaluation génétique vérifiée, nous nous sommes attachés à valider la modélisation des effets aléatoires inclus dans le modèle pressenti décrit dans la partie 2.4. Comme expliqué précédemment, Druet *et al.* (2005) avaient montré l'existence de fortes corrélations entre les coefficients aléatoires des vecteurs propres des 2<sup>ème</sup> et 3<sup>ème</sup> lactation, laissant supposer une possible réduction ultérieure du rang des matrices. Dans un contexte où de nombreux facteurs de variation sont pris en compte dans le modèle et où les données à analyser sont extrêmement nombreuses, il nous paraissait essentiel de limiter le nombre d'équations du modèle mixte. Dans un premier temps, une réduction d'un tiers de la dimension des trois effets aléatoires (génétique, environnement permanent et troupeau-année) a été testée et les éventuelles conséquences sur l'évaluation génétique ont été mesurées. Ce travail sera présenté dans la 2<sup>ème</sup> partie de ce chapitre.

Enfin, la dernière partie de ce chapitre est consacrée, dans une démarche similaire à la 2<sup>ème</sup> partie, à valider la modélisation des effets fixes inclus dans le modèle pressenti présenté précédemment. Toujours dans l'objectif de limiter le coût informatique d'une telle analyse, nous avons également testé une alternative technique, permise par la mise en place d'un modèle d'évaluation génétique plus simple que celui qui était pressenti.

## **Partie 3.1**

# **Validation du logiciel d'évaluation génétique**

## INTRODUCTION

Comme nous l'avons vu dans la partie 2.2, la qualité d'une évaluation génétique dépend classiquement de l'étape de préparation des données, de l'évaluation génétique *sensu stricto* et de l'étape postérieure à l'évaluation. Dans la seconde étape, la qualité est liée principalement à l'estimation des paramètres génétiques nécessaires à l'évaluation et la définition du modèle. Or, nous oublions souvent un des éléments majeurs de cette étape qui est le logiciel d'évaluation génétique utilisé. Dans la mesure du possible, les centres d'évaluation utilisent des logiciels partagés au sein de la communauté scientifique, afin de limiter le temps consacré à son développement et surtout à sa validation. Les équipes se lançant toutefois dans un projet de développement d'un nouveau logiciel sont systématiquement confrontées au besoin de valider leur logiciel. La plupart du temps, quand la complexité du modèle le permet, cette étape est réalisée en comparant les résultats exacts obtenus par l'inversion directe des équations du modèle mixte avec les résultats obtenus par le logiciel en phase de développement, ou bien en comparant ces derniers résultats avec ceux obtenus avec un logiciel de référence. Cependant, ces approches de validation ne sont pas toujours forcément envisageables. C'est le cas pour certains modèles complexes ou s'appuyant sur des approches novatrices comme l'évaluation sur les contrôles élémentaires que la France va mettre en place, où sont combinées l'utilisation de splines pour décrire la forme des courbes de lactation, celle de vecteurs propres pour décrire les effets aléatoires et une variance résiduelle hétérogène en fonction du stade de lactation et du troupeau. Afin de s'affranchir des méthodes précédemment citées, Robin Thompson (*communication personnelle*) a proposé une approche de validation générale basée sur la simulation de données avec des valeurs génétiques et des phénotypes déterminés. Ainsi, lorsque le logiciel d'évaluation génétique testé est exact, les valeurs génétiques calculées et conduisant à des solutions des équations du modèle mixte connues à l'avance par celui-ci sont égales à celles simulées et utilisées pour déterminer les phénotypes. Malgré la volonté d'être générale, cette approche présente des contraintes importantes qui limitent fortement sa portée. Par exemple, les résiduelles du modèle sont supposées nulles impliquant des contraintes sur les effets fixes et aléatoires. Elles se traduisent ainsi par la nécessité de déterminer une performance pour tous les animaux inclus dans le modèle, y compris les mâles. De ce fait, cette méthode s'éloigne parfois de façon importante de la réalité des fichiers dont nous disposons en génétique animale. De plus, elle nous a paru difficilement extensible aux modèles de régression aléatoire utilisés pour les contrôles élémentaires.

Compte tenu du contexte et de la spécificité de nos besoins, notre objectif était de développer une méthode originale de validation certaine du logiciel d'évaluation génétique plus flexible que l'approche présentée par Robin Thompson. Afin de contourner les limites de cette dernière, les contraintes de construction des performances simulées de l'approche proposée portent sur la résiduelle et non sur les effets fixes et aléatoires inclus dans le modèle. L'objectif est de s'assurer de l'exactitude des calculs mis en œuvre dans *Genekit*, le logiciel BLUP développé par Vincent Ducrocq pour répondre aux spécificités du modèle français.

## MATERIEL ET METHODES

Le programme de simulation proposé est très généraliste. Il se base sur un fichier de données réelles pour lequel nous disposons de l'ensemble des effets pris en compte dans le modèle et dont les performances enregistrées sont remplacées par des performances simulées. Nous disposons également d'un fichier réel de généalogie correspondant aux performances. Ceci permet d'éviter d'avoir à simuler la structure d'une population sélectionnée sur plusieurs générations. Nous supposons également connus les paramètres génétiques ainsi que les autres éléments nécessaires aux calculs des coefficients de régression et coefficient des splines. D'un point de vue méthodologique, l'approche proposée repose sur la simulation d'une résiduelle répondant à des conditions spécifiques détaillées dans l'article suivant. En ce point, elle se distingue nettement de l'approche proposée par Robin Thompson qui supposait une résiduelle nulle.

La procédure de validation peut être décomposée en 3 étapes :

- ✓ A partir d'un programme de simulation, une valeur est simulée pour chacun des effets du modèle, la résiduelle étant simulée selon une procédure particulière. Ces valeurs sont combinées selon le modèle choisi afin de reconstituer une performance simulée pour chacun des enregistrements du fichier réel,
- ✓ A partir du logiciel d'évaluation génétique à valider, *Genekit* dans notre cas, les données simulées sont utilisées comme fichier d'apport. L'ensemble des effets du modèle sont alors estimés,
- ✓ Les effets fixes et aléatoires estimés à partir du logiciel d'évaluation génétique sont comparés à ceux qui avaient été simulés lors de la 1<sup>ère</sup> étape et qui sont considérés comme les effets « vrais ». Le logiciel est validé lorsque les effets obtenus sont identiques.

Le programme de simulation a été développé pour répondre aux besoins de notre équipe en terme de validation du logiciel destiné au modèle contrôles élémentaires unicaractère multi-lactation. Il permet donc la prise en compte des effets fixes [effet troupeau – jour de contrôle, des constantes, des courbes modélisant la trajectoire phénotypique de la lactation], des effets aléatoires [génétique additif, environnement permanent, troupeau – année] ainsi qu'une résiduelle dont la variance varie suivant le stade de lactation, sans toutefois permettre la prise en compte de l'hétérogénéité liée aux troupeaux. Le traitement des groupes de parents inconnus a aussi été intégré au programme afin de valider le plus largement possible tous les éléments du modèle d'évaluation français. Dans un souci de flexibilité, le programme est modulable selon de nombreux paramètres portant aussi bien sur la modélisation de la trajectoire des courbes de lactation (courbes non paramétriques, polynômes de Legendre, splines) que sur les effets aléatoires pour lesquels il est possible de réduire le rang de la matrice de covariance et d'utiliser les vecteurs propres correspondants comme covariables.

Deux fichiers de données ont principalement été utilisés pour illustrer cette approche. D'une part, un fichier de taille modeste (377 080 contrôles élémentaires) correspondant à des vaches Montbéliardes en 1<sup>ère</sup> lactation dans le département du Jura enregistrés entre septembre 1995 et



Août 2005 a servi de support à nos analyses. Des précisions supplémentaires sur la constitution du fichier de données et le modèle employé pour décrire la performance sont disponibles dans l'article qui suit. D'autre part, un fichier contenant les contrôles élémentaires des 3 premières lactations des vaches du Jura pour la même race enregistrées entre septembre 1988 et août 2005, soit 2 533 137 données et 203 327 animaux dans le fichier de généalogie, a été analysé selon un modèle très proche de celui qui sera utilisé en pratique.

## RESULTATS ET DISCUSSION

Quel que soit le fichier de données utilisé et le modèle employé pour décrire les performances, les effets estimés par le logiciel d'évaluation génétique et leur valeur simulée suivant des modèles similaires sont mathématiquement strictement identiques. La corrélation entre ces effets est donc égale à 1 pour l'ensemble des effets. Le logiciel d'évaluation génétique *Genedit* est donc considéré être exact pour les modèles testés.

L'approche proposée est extrêmement flexible. L'extension à des modèles multicaractères, des modèles prenant en compte des effets maternels, des modèles gamétiques... est envisageable, ce qui permettrait de valider n'importe quel logiciel d'évaluation génétique.

J. Dairy Sci. 91:3179–3183

doi:10.3168/jds.2007-0972

© American Dairy Science Association, 2008.

## A General Method to Validate Breeding Value Prediction Software

H. Leclerc,<sup>\*1,2</sup> M. Wensch-Dorendorf,<sup>†2</sup> J. Wensch,<sup>‡</sup> V. Ducrocq,<sup>\*</sup> and H. H. Swalve<sup>†</sup>

<sup>\*</sup>Institut National de la Recherche Agronomique, UR337, Station de Génétique Quantitative et Appliquée, F-78352 Jouy-en-Josas, France

<sup>†</sup>Institute of Agricultural and Nutritional Sciences, Martin-Luther University, Halle-Wittenberg, 06099 Halle, Germany

<sup>‡</sup>Institute of Scientific Computing, Technical University Dresden, 01062 Dresden, Germany

### ABSTRACT

The validity of national genetic evaluations depends on the quality of input data, on the model of analysis, and on the correctness of genetic evaluation software. A general strategy was developed to validate national breeding value prediction software: performances from a real data file were replaced with simulated ones, created from simulated fixed and random effects and residuals in such a way that BLUP estimates from the evaluation software must be equal to the simulated effects. This approach was implemented for a multiple-trait model and a random regression test-day model. An example was presented on test-day observations analyzed with a random regression animal model including a lactation curve described as a sum of fixed polynomial regression and fixed spline regression on days in milk, and with genetic and permanent environmental effects modeled by using Legendre polynomials of order 2. Residuals had heterogeneous variances, and phantom parent groups were included. This method can be easily extended to other linear models. The comparison of genetic evaluation results with simulated true effects is used to demonstrate the great efficiency and usefulness of the proposed method.

**Key words:** genetic evaluation, validation, best linear unbiased predictor, random regression

### INTRODUCTION

At the national level, the reliability of genetic evaluation depends on a large number of factors, such as 1) the quality of raw data, 2) the data edits, 3) the correctness of the evaluation model and of the software used, and 4) the postprocessing steps. A large number of tests are done at the national level by genetic evaluation centers to ensure the quality of their EBV. For instance, the results from 2 consecutive evaluations are system-

atically compared. More and more countries, such as France, Germany, or the Netherlands, have developed or are developing a quality management system based on ISO 9001 standards. At the worldwide level, Interbull has been providing international predicted breeding values of dairy bulls since 1995 on the scale of each participating country by using national genetic evaluation results. Input data quality is a crucial issue in international genetic evaluations, because the validity of results from complex genetic and statistical analyses depends on it. Therefore, monitoring and validation of input data are essential for Interbull (Fikse, 2004): data included in international evaluations have to pass a series of stringent tests before acceptance. The consistency of evaluations is assessed by the comparison of breeding values from consecutive evaluations to identify changes larger than expected based on statistical properties of the breeding values (Klei et al., 2002). Genetic trends are estimated to check that national breeding values are unbiased (Boichard et al., 1995). These checks are also used at the domestic level to guarantee the quality of national genetic evaluations and to keep customers satisfied. However, these tests do not guarantee correctness of computation. The diversity and complexity of models used in the various countries to analyze different traits have led to a situation in which new methods of validation of national data, models, or both are increasingly needed. One research project identified by Interbull in 2002 was the development of a general simulation tool to validate national genetic evaluation systems, especially the development of a simulation environment to test breeding value prediction software. With this aim, a program was developed from a strategy described by R. Thompson (Rothamsted Research, Harpenden, UK; personal communication) to simulate data with known breeding values and phenotypes for a single-trait animal model (Täubert et al., 2002) in such a way that BLUP solutions for breeding values should be equal to the simulated ones. This strategy assumes that residuals are zero but with constrained fixed and random effects. This method was later extended to a multiple-trait animal model (Wensch-Dorendorf et al., 2005).

Received December 21, 2007.

Accepted March 26, 2008.

<sup>1</sup>Corresponding author: helene.leclerc@jouy.inra.fr

<sup>2</sup>Equal contribution.

Making use of a simulation tool is not the only option to validate new genetic evaluation software. A simpler alternative may be preferred. For instance, the direct inversion of mixed-model equations (MME) for a small data example could verify that the iterative EBV match those obtained from direct inversion. Nevertheless, when the genetic evaluation model is more sophisticated, this does not guarantee that the MME are properly set up, and numerical problems are often detected only on larger data sets. Another option is to compare results from the new software with the ones obtained from reference software. This method is widely used when it is technically possible.

In actuality, a large number of BLUP software programs have been developed worldwide, fulfilling specific needs. For instance, in France, BLUP software (GeneKit, V. Ducrocq, personal communication) was developed to deal with test-day models (TDM). Indeed, national genetic evaluation models for dairy traits are increasingly based on TDM instead of 305-d lactation models. A large variety of models have been proposed, differing in 1) how the lactation curve is modeled as a function of DIM [with fixed classes, parametric curves, or semiparametric (spline) curves; White et al., 1999], 2) how the genetic and permanent environmental components are described (fixed or random regression using Legendre or other polynomials), and 3) how heterogeneous residual variances are accounted for. Unfortunately, no general evaluation software including all possible models is available for TDM with very large data sets. Therefore, countries have developed custom software to perform routine genetic evaluations for their own population using TDM. The lack of reference software makes the software validation step complex. Furthermore, the routine genetic evaluation software usually relies on iterative solving algorithms, which makes it even more difficult to debug them and makes the results complicated to verify. For TDM needs, extension from R. Thompson's strategy to random regression situations is not straightforward.

The objective of this paper was to present a general and flexible strategy that could be used to validate the correctness of newly developed genetic evaluation software. Consistent phenotypic data were generated in such a way that BLUP estimates from correct evaluation software were mathematically equal to the simulated effects. This methodology can be considered a helpful tool in the development and further refinement of BLUP software.

## MATERIALS AND METHODS

### Outline of the Procedure

The starting point is a pedigree file and a data file containing, for each record, the relevant levels, vari-

ables, or a combination of both for all effects, the animal's recoded number and permanent environmental effect level, and all other pertinent pieces of information [elements required to compute random regression coefficients; the weight of records; the genetic, permanent environment, and residual (co)variance matrices, etc.]. These files can be real data sets. The procedure to check genetic evaluation software can then be divided into 3 steps:

1. For each effect as well as for one residual per observation, simulate values following the approach described below, leading to a simulated performance record for each record in the data file.
2. Include as input data these simulated performance records in the national genetic evaluation software. Estimates are obtained for all effects included in the model.
3. Compare estimates of fixed effects and predicted random effects from the national genetic evaluation software with the true (simulated) ones. If the resulting breeding values, permanent environmental effects, and all estimable contrasts of fixed effects are identical to the true ones, then the genetic evaluation software can be considered as correct.

### Estimation Method

The following multiple-trait linear model  $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{a} + \mathbf{e}$  is considered to describe the derivation of constraints enforcing the simulated performance to fulfill the required properties.  $\mathbf{y}$  is the vector of observations;  $\mathbf{b}$  is a vector of fixed effects;  $\mathbf{a}$  is the vector of breeding values following a normal distribution, with  $E[\mathbf{a}] = 0$  and  $\text{Var}[\mathbf{a}] = \mathbf{G} = \mathbf{G}_0 \otimes \mathbf{A}$ ; and  $\mathbf{e}$  is the vector of random residuals following a normal distribution, with  $E[\mathbf{e}] = 0$  and  $\text{var}[\mathbf{e}] = \mathbf{R} = \mathbf{R}_0 \otimes \mathbf{I}$ .  $\mathbf{X}$  and  $\mathbf{Z}$  are matrices relating  $\mathbf{y}$  to the appropriate fixed and genetic effects.  $\mathbf{G}_0$  is the covariance matrix for the genetic effects, and  $\mathbf{A}$  is the additive genetic relationship matrix.

The MME corresponding to this model are

$$\begin{bmatrix} \mathbf{X}\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{X}\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}. \quad [1]$$

The MME are analogous to the normal equation in the standard linear model. Indeed, MME were initially obtained by maximizing the posterior distribution of a given  $\mathbf{y}$  (Henderson et al., 1959). System [1] can also be rewritten as  $\mathbf{F}'\mathbf{F}\mathbf{x} = \mathbf{F}'\mathbf{c}$ , with  $\mathbf{F} = \begin{bmatrix} \mathbf{R}^{-1/2}\mathbf{X} & \mathbf{R}^{-1/2}\mathbf{Z} \\ \mathbf{0} & \mathbf{G}^{-1/2} \end{bmatrix}$ ,  $\mathbf{x} = \begin{bmatrix} \mathbf{b} \\ \mathbf{a} \end{bmatrix}$ , and  $\mathbf{c} = \begin{bmatrix} \mathbf{R}^{-1/2}\mathbf{y} \\ \mathbf{0} \end{bmatrix}$ , where  $\mathbf{M}^{1/2}$  is, for a symmetric general positive definite matrix  $\mathbf{M}$ , the unique positive

definite matrix  $\mathbf{B}$ , with  $\mathbf{B}\mathbf{B} = \mathbf{M}$ . System [1] can then be viewed as “normal equations” (Searle, 1971), giving the least squares solutions of the overdetermined system  $\mathbf{F}\mathbf{x} = \mathbf{c}$ . The least squares solutions are obtained by minimizing the squared Euclidean norm of the residual  $\|\mathbf{F}\mathbf{x} - \mathbf{c}\|_2^2$ :

$$\|\mathbf{F}\mathbf{x} - \mathbf{c}\|_2^2 = \mathbf{x}'\mathbf{F}'\mathbf{F}\mathbf{x} - 2\mathbf{x}'\mathbf{F}'\mathbf{c} + 2\mathbf{c}'\mathbf{c}. \quad [2]$$

To minimize [2], its derivative with respect to  $\mathbf{x}$  is set to zero, that is,  $2\mathbf{F}'\mathbf{F}\mathbf{x} - 2\mathbf{F}'\mathbf{c} = \mathbf{0}$ . After simplification, the minimizing vector  $\mathbf{x}$  is a solution of the normal equation  $\mathbf{F}'\mathbf{F}\mathbf{x} = \mathbf{F}'\mathbf{c}$ , which can be written as  $\mathbf{F}'(\mathbf{c} - \mathbf{F}\mathbf{x}) = \mathbf{0}$ .

Therefore, to simulate the data  $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{a} + \mathbf{e}$  such that  $\mathbf{F}'(\mathbf{c} - \mathbf{F}\mathbf{x}) = \mathbf{0}$  whatever the values of  $\mathbf{b}$ , we must have  $\mathbf{F}'\begin{bmatrix} \mathbf{R}^{-1/2}\mathbf{e} \\ -\mathbf{G}^{-1/2}\mathbf{a} \end{bmatrix} = \mathbf{0}$ , which leads to the 2 following conditions:  $\mathbf{X}\mathbf{R}^{-1}\mathbf{e} = \mathbf{0}$  and  $\mathbf{Z}\mathbf{R}^{-1}\mathbf{e} = \mathbf{G}^{-1}\mathbf{a}$ .

#### Application to a Random Regression TDM with Unknown Parent Groups

The previous model can be extended to

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{Q}\mathbf{g} + \mathbf{Z}\mathbf{a}^* + \mathbf{W}\mathbf{p} + \mathbf{e},$$

where  $\mathbf{y}$  is the vector of observations;  $\mathbf{b}$  is a vector of fixed effects;  $\mathbf{g}$  is the vector of genetic group effects (e.g., phantom parents groups);  $\mathbf{a}^*$  is the vector of breeding values corrected for genetic group effects, that is, with  $E[\mathbf{a}^*] = \mathbf{0}$  or, alternatively,  $E[\mathbf{a}] = \mathbf{Q}\mathbf{g}$ , with  $\mathbf{a} = \mathbf{Q}\mathbf{g} + \mathbf{a}^*$  and  $\text{Var}[\mathbf{a}^*] = \mathbf{G} = \mathbf{G}_0 \otimes \mathbf{A}$ ;  $\mathbf{p}$  is the vector of permanent environmental effects, with  $E[\mathbf{p}] = \mathbf{0}$  and  $\text{Var}[\mathbf{p}] = \mathbf{P} = \mathbf{P}_0 \otimes \mathbf{I}$ ;  $\mathbf{e}$  is the vector of random residuals, with  $E[\mathbf{e}] = \mathbf{0}$  and  $\text{var}(\mathbf{e}) = \mathbf{R}$ ;  $\mathbf{X}$ ,  $\mathbf{Z}$ , and  $\mathbf{W}$  are matrices relating  $\mathbf{y}$  to the appropriate fixed, genetic, and permanent environmental effects, possibly through continuous covariates; and  $\mathbf{Q}$  is the matrix assigning animals in  $\mathbf{a}^*$  to groups in  $\mathbf{g}$ . In the case of random regressions,  $\mathbf{G}_0$  and  $\mathbf{P}_0$  are the covariance matrices for the genetic and permanent environmental effects, respectively, and  $\mathbf{A}$  is the additive genetic relationship matrix.  $\mathbf{R}$  is a diagonal matrix of residual variances. In the case of heterogeneous variances, for example, as a function of DIM (e.g., Druet et al., 2003), the diagonal terms of  $\mathbf{R}$  vary from one record to the next. If heterogeneous variances are estimated simultaneously with the other effects during iterations, the validation test is no longer applicable because solutions are then nonlinear functions of the data. However, the validation test can still be used to validate part of the software if the updating of heterogeneous variance estimates can be bypassed.

This model is an extension of the previous one, with  $\mathbf{X}$  and  $\mathbf{b}$  replaced by  $[\mathbf{X} \ \mathbf{Z}\mathbf{Q}]$  and  $\begin{bmatrix} \mathbf{b} \\ \mathbf{g} \end{bmatrix}$ , respectively, and  $\mathbf{Z}$  and  $\mathbf{a}$  replaced by  $[\mathbf{Z} \ \mathbf{W}]$  and  $\begin{bmatrix} \mathbf{a}^* \\ \mathbf{p} \end{bmatrix}$ , respectively. In this case, the MME are

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\mathbf{Q} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{W} \\ \mathbf{Q}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Q}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\mathbf{Q} & \mathbf{Q}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{Q}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{W} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\mathbf{Q} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{W} \\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{W}'\mathbf{R}^{-1}\mathbf{Z}\mathbf{Q} & \mathbf{W}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \mathbf{P}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{g} \\ \mathbf{a}^* \\ \mathbf{p} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Q}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}. \quad [3]$$

and the residual conditions derived above become  $\mathbf{X}'\mathbf{R}^{-1}\mathbf{e} = \mathbf{0}$ ,  $\mathbf{Q}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{e} = \mathbf{0}$ ,  $\mathbf{G}^{-1}\mathbf{a}^* = \mathbf{Z}'\mathbf{R}^{-1}\mathbf{e}$ , and  $\mathbf{P}^{-1}\mathbf{p} = \mathbf{W}'\mathbf{R}^{-1}\mathbf{e}$ . Most models used in animal breeding, such as the maternal model, dominance, and the epistatic model or gametic model, can be easily implemented.

#### Fulfilling of Residual Conditions

To fulfill the first 2 residual conditions, a working variable  $\varepsilon$  is first simulated for all observations with any underlying distribution, for example,  $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$  with variance  $\sigma_\varepsilon^2$ . Two vectors,  $\beta$  and  $\gamma$ , are then computed by using any least squares solver, for example, the one included in the software package FSPAK90 (Misztal and Perez-Enciso, 1998), as solutions for

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\mathbf{Q} \\ \mathbf{Q}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Q}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\mathbf{Q} \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\varepsilon \\ \mathbf{Q}'\mathbf{Z}'\mathbf{R}^{-1}\varepsilon \end{bmatrix}.$$

If we choose  $\mathbf{e} = \delta(\varepsilon - \mathbf{X}\mathbf{b} - \mathbf{Z}\mathbf{Q}\gamma)$ , one can check that  $\mathbf{X}'\mathbf{R}^{-1}\mathbf{e} = \mathbf{0}$  and  $\mathbf{Q}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{e} = \mathbf{0}$ . Here,  $\delta$  is a scaling factor that can take any value. Then  $\mathbf{b}$  and  $\mathbf{g}$  can be simulated by using any underlying distribution because they do not influence  $\mathbf{e}$ .

The next step is to derive  $\mathbf{a}^*$  such that  $\mathbf{G}^{-1}\mathbf{a}^* = (\mathbf{G}_0^{-1} \otimes \mathbf{A}^{-1})\mathbf{a}^* = (\mathbf{I} \otimes \mathbf{A}^{-1})(\mathbf{G}_0^{-1} \otimes \mathbf{I})\mathbf{a}^* = \mathbf{Z}'\mathbf{R}^{-1}\mathbf{e}$ . Define  $\mathbf{u} = (\mathbf{G}_0^{-1} \otimes \mathbf{I})\mathbf{a}^*$  so we need  $\mathbf{a}^*$  in such a way that  $(\mathbf{I} \otimes \mathbf{A}^{-1})\mathbf{u} = \mathbf{Z}'\mathbf{R}^{-1}\mathbf{e}$ . For example, if the random regression model includes 3 genetic terms (3 genetic values  $j$  per animal  $i$ ,  $\mathbf{a}_{i1}^*$ ,  $\mathbf{a}_{i2}^*$ , and  $\mathbf{a}_{i3}^*$ ),  $\mathbf{u} = \{\mathbf{u}_j\}$  also includes 3 terms,  $\mathbf{u}_{i1}$ ,  $\mathbf{u}_{i2}$ , and  $\mathbf{u}_{i3}$ , for each animal  $i$ , which are linear combinations of the terms  $\mathbf{a}_{i1}^*$ ,  $\mathbf{a}_{i2}^*$ , and  $\mathbf{a}_{i3}^*$ . For each  $j$ , we need to solve  $\mathbf{A}^{-1}\mathbf{u}_j = \mathbf{Z}'\mathbf{R}^{-1}\mathbf{e}_j$ . This is easily done by using the decomposition  $\mathbf{A}^{-1} = \mathbf{T}^{-1}\mathbf{D}^{-1}\mathbf{T}^{-\text{T}}$  (Quaas, 1976).  $\mathbf{T}^{-1}$  is a lower triangular matrix with at most 2 nonzero off-diagonal elements: the  $i$ th diagonal element is equal to 1 and the columns corresponding to the sire and dam of animal  $i$  are equal to  $-1/2$ .  $\mathbf{D}$  is the diagonal matrix

of the proportion of the genetic variance corresponding to the Mendelian sampling terms. First, we must solve  $(\mathbf{T}^{-1}\mathbf{D}^{-1})\mathbf{v}_j = \mathbf{Z}'\mathbf{R}^{-1}\mathbf{e}_j$  for  $\mathbf{v}_j$  and then  $\mathbf{T}^{-1}\mathbf{u}_j = \mathbf{v}_j$ . These 2 triangular systems are very simple. Finally, we can compute  $\mathbf{a}^* = \mathbf{G}_0 \otimes \mathbf{u}$  [or for animal  $i$  ( $[\mathbf{a}_{i,1}^* \ \mathbf{a}_{i,2}^* \ \mathbf{a}_{i,3}^*]'$  =  $\mathbf{G}_0[\mathbf{u}_{i,1} \ \mathbf{u}_{i,2} \ \mathbf{u}_{i,3}]'$ ].

We obtain  $\mathbf{p}$  enforcing  $\mathbf{P}^{-1}\mathbf{p} = (\mathbf{P}_0^{-1} \otimes \mathbf{I})\mathbf{p} = \mathbf{W}\mathbf{R}^{-1}\mathbf{e}$  by choosing  $\mathbf{p} = (\mathbf{P}_0 \otimes \mathbf{I})\mathbf{W}\mathbf{R}^{-1}\mathbf{e}$ . This reduces to the product of  $\mathbf{P}_0$  by the terms in  $\mathbf{W}\mathbf{R}^{-1}\mathbf{e}$  corresponding to animal  $i$ . Finally,  $\mathbf{y}$  is constructed as  $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{Q}\mathbf{g} + \mathbf{Z}\mathbf{a}^* + \mathbf{W}\mathbf{p} + \mathbf{e}$ .

**Application to a TDM**

As an illustration, this strategy was applied to a real data set with 47,492 first lactations of Montbéliarde cows from a sample of herds in Jura (a region of France), that is, 377,080 test-day observations recorded between September 1995 and August 2005 in 28,020 herd test-day (HTD) combinations. The average number of test-day records per lactation was 8, with a minimum of 5. On average, 115 test-day records were known per herd and per year, with a minimum of 50 and a maximum of 431. A total of 118 classes of calving month and 96 classes of calving age were considered. Days in milk had to be between 5 and 335. Information about gestation (days carried calf, DCC) was also available. The pedigree file included 102,339 animals with 8 groups of unknown parents. The approach described above was tested on the following model. For animal  $i$ , test-day  $j$ , and DCC  $k$ , the performance recorded after  $t$  days of lactation is

$$y_{ijkrs(t)} = \text{HTD}_j + \beta \cdot \text{DCC}_k + \sum_{l=1}^{Nb} \omega_{l(t)} \cdot b_{rl} + \sum_{p=1}^{Nc} \xi_{p(t)} \cdot c_{sp} + \sum_{n=1}^{Na} \alpha_{m(t)} \cdot a_{im} + \sum_{n=1}^{Np} \psi_{n(t)} \cdot p_{in} + e_{ijkrs(t)},$$

where  $y_{ijkrs(t)}$  is the performance recorded,  $\text{HTD}_j$  is the  $j$ th fixed herd by test-day effect applied to all animals in the herd on test day  $j$ ,  $\beta$  is a regression coefficient related to the effect of the calf after  $k$  days of gestation on performance,  $b_{rl}$  and  $c_{sp}$  are fixed regression coefficients specific to calving month  $r$  and calving age class  $s$ ,  $a_{im}$  are the genetic values for cow  $i$ , and  $p_{in}$  are the permanent environmental effects for cow  $i$ .  $\omega_{l(t)}$ ,  $\xi_{p(t)}$ ,  $\alpha_{m(t)}$ , and  $\psi_{n(t)}$  are continuous covariates depending on DIM. For this model, random regression curves and fixed regression curves for calving age were modeled by using Legendre polynomials of order 2 ( $Nc = Na = Np = 3$ ). Fixed regression curves for calving month were modeled by using 6 knot splines (White et al., 1999) at DIM 5, 20, 50, 135, 245, and 335.  $e_{ijkrs(t)}$  is the residual

effect for each observation, with a heterogeneous variance continuously changing with DIM.

**RESULTS AND DISCUSSION**

Correlations between simulated effects and estimates obtained with the national genetic evaluation software from the 377,080 test-day records were all equal to 1. Standard deviations for all effects included in the model were identical. For contrast analysis, the largest relative difference observed between “estimates” (estimable contrasts or predicted random effects) and true values (i.e., 478,468 estimated effects) was  $4 \cdot 10^{-4}$  when the convergence criterion used for BLUP evaluation (i.e., average solution change between 2 iterations) was  $10^{-6}$ . The software implemented for the national evaluation was therefore considered as correct for the tested model.

In contrast with other simulation approaches (R. Thompson, Rothamsted Research, Harpenden, UK; personal communication; Täubert et al., 2002; Wensch-Dorendorf et al., 2005), the proposed strategy to generate breeding values and performances can be applied to actual data sets by simply replacing performance values with simulated ones. Its main specificity is that appropriate residuals are created when the other approaches do not simulate any residual. This is more complex but also more realistic: for example, R. Thompson’s (Rothamsted Research, Harpenden, UK; personal communication) approach requires animals from the base generation to have “virtual” records. Such a requirement is not imposed here.

Moreover, the method could help to investigate the relevance of different iterative algorithms and convergence criteria as long as the simulated effects are of a magnitude similar to the real effects. This can be adjusted by choosing an appropriate scaling factor  $\delta$ . Miztal et al. (1988) showed that the real accuracy of solutions could be far from the one suggested by some convergence criteria. Here, true solutions of the system are directly available. The relative average difference between current and true solutions, as proposed by Miztal et al. (1988), can be computed at each iteration for a particular data set. Usually, the convergence of genetic evaluation software is checked by comparing solutions obtained from a limited number of iterations with the ones resulting from an increased number of iterations. This method is easy to implement because it does not require modification of the original software, but it does not allow one to study the relevance of the convergence criteria nor the correctness of the converged solutions.

The main target for the proposed approach may be researchers implementing new genetic evaluation soft-

ware to be used for sophisticated models for which no reference software exists. The development of a specific simulation tool to generate genetic and nongenetic effects as well as data under the most frequent models used in countries participating in international evaluation can be envisioned but should be done once for all countries. Covering a large number of potential models is a difficult task but is probably easier to consider than creating an all-purpose reference software. Such a tool should present 2 main advantages: it would enable any Interbull participating country to check the correctness of its national genetic evaluation software and to verify that the number of iteration rounds or the convergence criteria are adequate.

#### ACKNOWLEDGMENTS

This research was partly funded by a grant from the French Ministry of Agriculture (Action Innovante UTILEG, Paris, France) and from the German Research Foundation (Deutsche Forschungsgemeinschaft, Bonn, Germany). One part of this study was realized within the framework of the Unité Mixte Technologique (Joint Technology Unit) INRA-Institut de l'Élevage (Jouy-en-Josas, France) dealing with cattle genetic evaluation.

#### REFERENCES

- Boichard, D., B. Bonaiti, A. Barbat, and S. Mattalia. 1995. Three methods to validate the estimation of genetic trend for dairy cattle. *J. Dairy Sci.* 78:431–437.
- Druet, T., F. Jaffrézic, D. Boichard, and V. Ducrocq. 2003. Modeling lactation curves and estimations of genetic parameters for first lactation test-day records of French Holstein cows. *J. Dairy Sci.* 86:2480–2490.
- Fikse, W. F. 2004. Interbull guides through the labyrinth of national genetic evaluations. Page 326 in Proc. of the 55th Mtg. Eur. Assoc. Anim. Prod., Interlaken, Switzerland. [http://www-interbull.slu.se/w-agera/getfile.php/interbullforum\\_edumaterial/216/EAAP2004.pdf](http://www-interbull.slu.se/w-agera/getfile.php/interbullforum_edumaterial/216/EAAP2004.pdf) Accessed Mar. 25, 2008.
- Henderson, C. R., O. Kempthorne, S. R. Searle, and C. N. Von Krosigk. 1959. Estimation of environmental and genetic trends from records subject to culling. *Biometrics* 15:192–218.
- Klei, B., T. Mark, W. F. Fikse, and T. Lawlor. 2002. A method for verifying genetic evaluation results. *Interbull Bull.* 29:178–182.
- Misztal, I., D. Gianola, and L. R. Schaeffer. 1988. Convergence rates in animal model solutions. *J. Dairy Sci.* 70:2577–2584.
- Misztal, I., and M. Perez-Enciso. 1998. FSPAK90: A Fortran 90 interface to sparse-matrix package FSPAK with dynamic memory allocation and sparse matrix structure. Proc. 6th World Congr. Genet. Appl. Livest. Prod. 27:467–468.
- Quaas, R. L. 1976. Computing the diagonal elements and inverse of a large numerator relationship matrix. *Biometrics* 32:949–953.
- Searle, S. R. 1971. *Linear Models*. John Wiley and Sons Inc., New York, NY.
- Täubert, H., H. H. Swalve, and H. Simianer. 2002. The Interbull Audit Project Part II: Development of a program for auditing breeding value estimation programs. *Interbull Bull.* 29:165–167.
- Wensch-Dorendorf, M., H. H. Swalve, and J. Wensch. 2005. Simulation of multiple trait data for testing breeding value estimation programs. Page 207 in Proc. 56th Mtg. Eur. Assoc. Anim. Prod., Uppsala, Sweden Wageningen Acad. Publ., Wageningen, the Netherlands.
- White, I. M. S., R. Thompson, and S. Brotherstone. 1999. Genetic and environmental smoothing of lactation curves with cubic splines. *J. Dairy Sci.* 82:632–638.



## **Partie 3.2**

# **Synthétiser l'information, une nécessité**



# Réduction du rang d'un modèle de contrôles élémentaires : quelles conséquences sur l'évaluation génétique ?

H. Leclerc<sup>1,2†</sup> et V. Ducrocq<sup>1</sup>

<sup>1</sup>INRA, UR337, Station de Génétique Quantitative et Appliquée, 78352 Jouy-en-Josas

<sup>2</sup>Institut de l'Elevage, Département de Génétique, INRA-SGQA, 78352 Jouy-en-Josas  
Courriel : helene.leclerc@jouy.inra.fr

## Résumé

Le développement des évaluations génétiques des caractères de production laitière basées sur les données de contrôles élémentaires constitue un défi en termes informatiques. En effet, d'une part le nombre de paramètres des modèles employés est généralement élevé afin de modéliser finement la production sous forme de courbes fonction du stade de lactation et d'autre part le nombre de données à traiter peut atteindre dans certains pays plusieurs centaines de millions d'enregistrements. Pour répondre à ces contraintes, une des approches proposées consiste à utiliser des modèles de contrôles élémentaires de rang réduit. A travers cette étude, nous avons cherché à quantifier l'impact de tels modèles sur les valeurs estimées des différents effets aléatoires du modèle. Pour réduire le rang des matrices des effets aléatoires, deux approches ont été testées : l'une basée sur la décomposition de la matrice de corrélation et l'autre basée sur la matrice de covariance. Les corrélations obtenues entre les valeurs estimées avec un modèle de référence et les valeurs estimées avec un modèle de rang réduit étant proches ou supérieures à 0.99, elles permettent d'envisager sereinement l'utilisation d'un modèle de contrôles élémentaires de rang réduit pour la réalisation de l'évaluation génétique. Quant au choix de l'approche de réduction du rang, les résultats obtenus mettent clairement en avant les avantages de l'approche basée sur la matrice de covariance, surtout dans le cas de l'effet d'environnement permanent. Les corrélations obtenues pour cet effet varient de façon plus ou moins importante suivant l'approche employée.

---

† Cet article n'a pas vocation d'être soumis dans l'état actuel mais constitue une base de travail pour des études complémentaires.

## Introduction

Les caractères de production laitière occupent une place primordiale dans le choix des animaux reproducteurs. L'évaluation génétique sur les caractères laitiers reposait jusqu'il y a peu sur la production cumulée en 305 jours. Depuis le début des années 2000, la majeure partie des grands pays laitiers se sont orientés vers les modèles dits de contrôles élémentaires (TDM pour Test-Day Model) qui valorisent pleinement l'ensemble des informations recueillies au cours de la lactation de l'animal (Schaeffer *et al.*, 2000 ; De Roos *et al.*, 2001 ; Reinhardt *et al.*, 2002 ; Lidauer *et al.*, 2006). Ces modèles présentent de nombreux intérêts. Ils permettent notamment de prendre en compte les effets d'environnement spécifiques du jour du contrôle, de modéliser la plupart des effets fixes et aléatoires sous forme de courbes dont la trajectoire varie suivant le stade de lactation. Cependant, ces modèles dépendent d'une part d'un nombre élevé de paramètres et d'autre part nécessitent d'analyser un nombre de données environ dix fois supérieur aux modèles basés sur les données de lactation. Les aspects calculatoires tels que les besoins en mémoire et les temps de calcul constituent donc un enjeu majeur du développement des modèles « contrôles élémentaires ».

Dans les modèles de régression aléatoire (RR-TDM pour Random Regression-TDM), les effets aléatoires tels que l'effet génétique et l'effet d'environnement sont modélisés sous forme de courbes. L'impact de ces effets sur la performance peut ainsi varier suivant le stade de lactation. Même si les modèles « contrôles élémentaires » diffèrent suivant les pays, les effets aléatoires sont généralement modélisés à l'aide de 3 paramètres par lactation, permettant de tenir compte de l'évolution de la structure des variances et covariances au cours de la lactation. C'est notamment le cas du modèle unicaractère allemand (Reinhardt *et al.*, 2002), où l'effet génétique et l'effet d'environnement permanent dépendent de 9 paramètres chacun. C'est également le cas au Canada, où les quatre caractères (quantité de lait, quantité de matière grasse (MG), quantité de matière protéique (MP) et score de cellules somatiques) pour chacune des lactations sont évalués conjointement ce qui nécessite l'estimation de 36 effets génétiques par animal et 36 effets d'environnement permanent par vache ayant des contrôles enregistrés. Face au défi calculatoire que représentent de tels modèles, certains chercheurs ont proposé de recourir à des modèles RR-TDM de rang réduit (Wiggans and Goddard, 1997 ; Van der Werf *et al.*, 1998, Misztal *et al.*, 2000). Ces modèles découlent des approches dites de composantes principales (PC), techniques largement utilisées pour réduire la dimension par extraction et sélection des caractéristiques principales des données. Les analyses en composantes principales reposent sur le calcul des valeurs propres et vecteurs propres qui conditionnent les matrices de variance-covariance. Les vecteurs propres, aussi

appelés composantes principales, sont des fonctions linéaires ne dépendant plus directement des variables initiales. Les valeurs propres correspondantes mesurent la part de variation expliquée par chacun des effets associés à ces vecteurs propres. Ainsi, la composante  $n$  explique le maximum de la variation restée sans explication dans les composantes 1 à  $n-1$ . Souvent, les dernières composantes contribuent peu à la variabilité des données et peuvent être exclues de l'analyse sans réel impact sur les résultats. Le principal intérêt est de réduire ainsi les besoins en mémoire et les temps de calcul. Par exemple, dans une étude sur une évaluation TDM multicaractères où chaque mois de lactation est considéré comme un caractère distinct, Wiggans et Goddard (1997) proposent de réduire le rang d'une matrice de (co)variance génétique de 60 caractères (3 caractères laitiers : (Lait, MG et MP) ; 1<sup>ère</sup> lactation vs lactations suivantes ; 10 périodes mensuelles de lactation) à 6 caractères combinés. Dans le modèle d'évaluation génétique sur les contrôles élémentaires testés en Finlande, Lidauer *et al.* (2003) conservent parmi les 12 coefficients de l'effet génétique et les 9 coefficients de l'effet d'environnement permanent les 6 valeurs propres les plus importantes pour chacun des effets, expliquant respectivement 96% et 97% de la variance. Les corrélations entre valeurs génétiques estimées avec un modèle de plein rang et un modèle de rang réduit sont supérieures à 0.995 aussi bien pour les taureaux que pour les jeunes vaches. Ces résultats prometteurs ouvrent des perspectives pour contourner les limites calculatoires liées au développement d'une évaluation génétique sur les contrôles élémentaires dans les pays où la population compte par exemple plus de 10 millions d'animaux à évaluer comme c'est le cas en France. Ces approches calculatoires ont également été proposées dans le cadre des évaluations internationales afin de rendre possible la mise en œuvre d'évaluations multicaractères (Sullivan and Wilton, 2001 ; Liu *et al.*, 2004), notamment pour permettre la prise en compte des différentes informations issues des modèles « contrôles élémentaires ».

Dans un contexte où les contraintes calculatoires limitent actuellement la mise en œuvre de l'évaluation génétique sur les contrôles élémentaire en France, le premier objectif de cette étude était de quantifier l'impact d'une réduction du rang des matrices des différents effets aléatoires sur les solutions estimées. Pour cela, nous avons évalué les conséquences d'une réduction d'un tiers de la dimension des matrices sur l'évaluation génétique, en l'occurrence sur d'une part les index élémentaires des régressions aléatoires et d'autre part leur combinaison en une valeur globale. Deux approches de réduction du rang des matrices ont été testées, l'une portant sur les matrices de corrélation (COR) et l'autre sur les matrices de covariance (COV). Le second objectif de cette étude était donc de déterminer l'approche la plus adaptée pour effectuer cette réduction.

## Matériels et Méthodes

### Données

Cette étude a été réalisée à partir des performances journalières sur le lait, la matière grasse (MG) et protéique (MP), le taux butyreux (TB) et protéique (TP) de vaches Montbéliardes ayant vêlé entre septembre 1988 et août 2004 dans le département du Jura. Seules les données ayant un stade de lactation compris entre 7 et 335 jours ont été prises en compte. Au total, de l'ordre de 2,4 millions de données élémentaires, issues de 137 158 contrôles et provenant de 135 743 vaches ont été utilisées. Le fichier de généalogie relatif à ces données comptait 195 214 animaux dont 7 groupes de parents inconnus formés à partir de l'année de naissance et le sexe de l'animal.

### Le modèle d'évaluation génétique

Dans le modèle d'évaluation génétique unicaractère pressenti en France, la performance est décrite comme :

$$Perf_{ijklmnopqt'} = HTD_i + \text{effets fixes}_{klmnott'} + \sum_{a=1}^A (\text{génétique}_{pa} \cdot v_{at}) + \sum_{b=1}^B (\text{environ}^t\_permanent_{pb} \cdot \zeta_{bt}) + \sum_{c=1}^C ((\text{troupeau}_q \times \text{année}_k)_c \cdot \tau_{ct}) + e_{ijklmnopqt'} \quad (1)$$

où  $Perf_{ijklmnopqt'}$  est le contrôle journalier de l'animal  $p$ , à un stade de lactation  $t$  et de gestation  $t'$ , avec un effet troupeau – jour de contrôle  $i$ , l'année  $k$ , un rang de lactation  $l$ , une durée de tarissement  $m$ , un mois de vêlage  $n$ , un âge au vêlage  $o$  et un troupeau  $q$ ,

$HTD_i$  est l'effet fixe troupeau – jour de contrôle  $i$ ,

effets fixes $_{klmnott'}$  est la somme des effets fixes obtenu pour un stade de lactation  $t$ , un stade de gestation  $t'$ , l'année  $k$ , un rang de lactation  $l$ , une durée de tarissement  $m$ , un mois de vêlage  $n$ , un âge au vêlage  $o$ . Cette somme inclut notamment des courbes de lactation modélisées à l'aide de splines pour l'effet du mois de vêlage, de l'âge au vêlage, de la durée de tarissement et de la gestation,

génétique $_{pa}$  est le  $a^{\text{ème}}$  effet aléatoire d'origine génétique de l'animal  $p$  suivant une distribution normale, d'espérance nulle et de variance

$$\mathbf{A} \otimes \mathbf{G} = \text{Var} \begin{bmatrix} gen_a \\ \vdots \\ gen_{a'} \end{bmatrix} = \begin{bmatrix} \mathbf{A} \sigma_{gen_a}^2 & \cdots & \mathbf{A} \sigma_{gen_a gen_{a'}} \\ \vdots & \ddots & \vdots \\ \mathbf{A} \sigma_{gen_{a'} gen_a} & \cdots & \mathbf{A} \sigma_{gen_{a'}}^2 \end{bmatrix} \quad \text{où } \mathbf{A} \text{ est la matrice de parenté}$$

et  $\mathbf{G}$  est la matrice de covariance génétique,

$\nu_{at}$  est la valeur du  $a^{\text{ème}}$  vecteur propre de la matrice de variance-covariance de rang  $A$  de l'effet génétique au stade de lactation  $t$ ;  $A$  étant initialement de rang 6 et réduit à 4 dans le cadre de cette étude,

$\text{environ}'_{\text{permanent}}_{pb}$  est le  $b^{\text{ème}}$  effet aléatoire prenant en compte l'effet d'environnement permanent de l'animal  $p$  suivant une distribution normale,

d'espérance nulle et de variance  $\mathbf{I} \otimes \mathbf{P} = \text{Var} \begin{bmatrix} perm_b \\ \vdots \\ perm_b \end{bmatrix}$

$$= \begin{bmatrix} \sigma_{perm_b}^2 & \cdots & \sigma_{perm_b perm_b'} \\ \vdots & \ddots & \vdots \\ \sigma_{perm_b' perm_b} & \cdots & \sigma_{perm_b'}^2 \end{bmatrix} \text{ où } \mathbf{P} \text{ est la matrice de covariance de l'effet}$$

d'environnement permanent,

$\zeta_{bt}$  est la valeur du  $b^{\text{ème}}$  vecteur propre de la matrice de variance-covariance de rang  $B$  de l'effet environnement permanent pour un stade de lactation  $t$ ;  $B$  étant initialement de rang 6 et réduit à 4 dans le cadre de cette étude,

$(\text{troupeau}_q \times \text{année}_k)_c$  est le  $c^{\text{ème}}$  effet aléatoire de l'effet troupeau – année du troupeau  $q$  pour l'année  $k$ , suivant une distribution normale, d'espérance nulle et de

$$\text{variance } \mathbf{I} \otimes \mathbf{H} = \text{Var} \begin{bmatrix} trx h_c \\ \vdots \\ trx h_{c'} \end{bmatrix} = \begin{bmatrix} \sigma_{trx h_c}^2 & \cdots & \sigma_{trx h_c trx h_{c'}} \\ \vdots & \ddots & \vdots \\ \sigma_{trx h_{c'} trx h_c} & \cdots & \sigma_{trx h_{c'}}^2 \end{bmatrix} \text{ où } \mathbf{H} \text{ est la matrice de}$$

covariance de l'effet troupeau – année,

$\tau_{ct}$  est la valeur du  $c^{\text{ème}}$  vecteur propre de la matrice de variance-covariance de rang  $C$  de l'effet troupeau – année pour un stade de lactation  $t$ ;  $C$  étant initialement de rang 9 et réduit à 6 dans le cadre de cette étude,

$e_{ijklmnopqt}$  est la valeur résiduelle de la performance, dont la variance est une fonction du stade de lactation  $t$ .

L'approche par étapes utilisée pour estimer les paramètres génétiques (Druet *et al.*, 2003 et 2005) a permis de retenir les deux principaux vecteurs propres des effets génétique et d'environnement permanent obtenus sur la 1<sup>ère</sup> lactation comme covariables pour l'estimation des paramètres pour les trois premières lactations, soit 6 effets pour chacun. Les effets génétique ou d'environnement permanent associés à ces vecteurs propres peuvent être interprétés respectivement comme des mesures du potentiel de production moyen sur la lactation et de la persistance pour chacune des lactations. Cette approche par étapes a permis d'estimer directement les six caractéristiques sous-jacentes qui nous intéressent pour le choix des animaux reproducteurs. En race Holstein, Druet *et al.* (2005) avaient toutefois mis en

avant de fortes corrélations (généralement entre 0.95 et 0.99) entre les coefficients aléatoires représentant le niveau de production (et la mesure de la persistance) de la 2<sup>ème</sup> lactation et ceux de la 3<sup>ème</sup> lactation, laissant présager une possible réduction supplémentaire du rang des matrices de variance-covariance génétique et d'environnement permanent. Dans le cadre de cette étude, nous analyserons l'impact d'une réduction du rang de ces matrices de 6 à 4. Pour l'effet aléatoire troupeau – année, son effet pour chaque lactation est modélisé à travers un polynôme de Legendre d'ordre 2, soit une matrice de variance-covariance de rang 9. Pour des raisons calculatoires, il apparaît indispensable d'étudier la possibilité d'une réduction du rang de la matrice pour synthétiser l'information sous forme d'un nombre plus réduit de composantes. Dans le cadre de cette étude, nous testerons l'impact d'une réduction de la dimension de la matrice de covariance de l'effet troupeau – année de 9 à 6.

### Les approches de réduction du rang des matrices : covariance vs corrélation

La première approche est basée sur la décomposition de la matrice de covariance (COV) de chacun des effets génétique (G), d'environnement permanent (P) et troupeau – année (HY pour Herd-Year). Afin de simplifier la présentation, nous exposerons la démarche des deux approches pour la réduction du rang de matrice de l'effet génétique G. La matrice des (co)variances **G** entre les effets génétiques pour des jours de lactation différents est obtenue en pré- et post-multipliant la matrice **H** des (co)variances entre les six caractéristiques sous-jacentes que sont la production et la persistance pour chacune des lactations par la matrice **U** constituée des covariables correspondant aux vecteurs propres estimés en 1<sup>ère</sup> lactation (équation 2). Compte tenu du nombre de lactations (3) et de la plage de prise en compte des contrôles sur une lactation (335 jours), cette matrice **G** a une dimension de 1005 x 1005, mais un rang de 6.

$$\mathbf{G}_{1005, 1005} = \mathbf{U}_{1005, 6} \mathbf{H}_{6, 6} \mathbf{U}'_{6, 1005} \quad (2)$$

D'après la décomposition canonique de la matrice **H** :

$$\mathbf{H}_{6, 6} = \mathbf{V}_{6, 6} \mathbf{D}_{6, 6} \mathbf{V}'_{6, 6} \text{ avec } \mathbf{V}_{6, 6} \mathbf{V}'_{6, 6} = \mathbf{I}_6 \text{ et } \mathbf{V}'_{6, 6} \mathbf{V}_{6, 6} = \mathbf{I}_6 \quad (3)$$

où **V** est la matrice orthogonale des vecteurs propres, **D** la matrice diagonale des valeurs propres et **I<sub>n</sub>** la matrice identité de dimension *n*. Le rang de la matrice **H** peut être réduit en ignorant les 2 plus petites valeurs propres de **D** et en supprimant les vecteurs propres correspondants de **V**. Ainsi :

$$\mathbf{H}_{6, 6}^* \simeq \mathbf{V}_{6, 4}^* \mathbf{D}_{4, 4}^* \mathbf{V}'_{4, 6} \text{ avec } \mathbf{V}'_{4, 6} \mathbf{V}_{6, 4}^* = \mathbf{I}_4 \quad (4)$$

où  $\mathbf{H}^*$  est une matrice de rang 4,  $\mathbf{D}^*$  est la matrice diagonale de rang réduit incluant les 4 plus grandes valeurs propres de  $\mathbf{D}$ ,  $\mathbf{V}^*$  est la matrice des vecteurs propres conservés. L'information est ainsi résumée sous forme de 4 composantes. Pour l'évaluation génétique, afin de limiter les ressources informatiques nécessaires à sa réalisation, les 6 covariables issues de la matrices  $\mathbf{U}_{1005,6}$  sont remplacées par 4 covariables issues de  $\mathbf{U}_{1005,4}^* = \mathbf{U}_{1005,6} \mathbf{V}_{6,4}^*$  et la matrice  $\mathbf{H}_{6,6}$  est remplacée par  $\mathbf{D}_{4,4}^*$ . Afin de maintenir la variance totale constante, la part de la variance « perdue » lors de la réduction de rang de chacun des effets est ajoutée à la résiduelle.

Dans la seconde approche, basée sur la matrice de corrélation (COR), la matrice  $\mathbf{H}$  est décomposée sous la forme :

$$\mathbf{H}_{6,6} = \mathbf{S}_{6,6} \mathbf{G}_{r,6,6} \mathbf{S}_{6,6} = \mathbf{S}_{6,6} \mathbf{V}_{r,6,6} \mathbf{D}_{r,6,6} \mathbf{V}_{r,6,6}' \mathbf{S}_{6,6} \quad (5)$$

où  $\mathbf{S}$  est la matrice diagonale des écarts-types génétique et  $\mathbf{G}_r$  est la matrice de corrélation génétique. La matrice  $\mathbf{G}_r$  peut ensuite être décomposée sous forme de vecteurs et valeurs propres comme pour l'approche COV. Le rang de la matrice  $\mathbf{G}_r$  peut être réduit comme pour la matrice  $\mathbf{H}$ .

$$\mathbf{G}_{r,6,6}^* \simeq \mathbf{T}_{6,6}^{-1} \mathbf{V}_{r,6,4}^* \mathbf{D}_{r,4,4}^* \mathbf{V}_{r,4,6}^{*'} \mathbf{T}_{6,6}^{-1} \quad (6)$$

où  $\mathbf{G}_r^*$  est un matrice de rang 4, qui a nécessité l'utilisation d'une matrice  $\mathbf{T}$  (matrice diagonale des écarts-types de  $\mathbf{V}_{r,6,4}^*$ ,  $\mathbf{D}_{r,4,4}^*$ ,  $\mathbf{V}_{r,4,6}^{*}$ ) dans l'équation 6 afin de conserver une structure de matrice de corrélation avec des éléments diagonaux égaux à 1. Pour la réalisation de l'évaluation génétique, la matrice des covariables  $\mathbf{U}_{1005,6}$  est remplacée par  $\mathbf{U}_{r,1005,4}^* = \mathbf{U}_{1005,6} \mathbf{S}_{6,6} \mathbf{T}_{6,6}^{-1} \mathbf{V}_{r,6,4}^*$  et la matrice  $\mathbf{H}_{6,6}$  est remplacée par  $\mathbf{D}_{r,4,4}^*$ . Comme dans le cas de l'approche COV, la variance totale est maintenue constante.

### Les comparaisons effectuées

Les deux approches de réduction du rang (COV vs COR) ont donc été comparées sur la base d'un modèle réduit, où les effets aléatoires sont modélisés avec un nombre limité de coefficients, soit 4 pour chacun des effets génétique et d'environnement permanent et 6 pour l'effet troupeau – année par rapport à une référence commune : le modèle presenté initialement. Ce modèle correspond à une évaluation génétique où les effets aléatoires sont

modélisés de façon identique au modèle utilisé pour l'estimation des composantes de la variance (Druet *et al.*, 2005), c'est à dire un effet génétique et un effet d'environnement permanent décrits chacun par 6 coefficients et un effet troupeau – année décrit par 9 coefficients. Le terme « référence » sera utilisé pour désigner ce modèle dans la suite de cette étude.

Plusieurs éléments ont été pris en compte dans la comparaison des deux approches. Tout d'abord, nous avons analysé les valeurs génétiques ou d'environnement permanent « globales » obtenues en sommant le produit entre les coefficients estimés au cours de l'évaluation génétique pour chacun des effets (génétique, d'environnement permanent ou troupeau – année) et les covariables préalablement définies pour les stades de lactation considérés. Nous avons ainsi calculé la valeur « globale » des différents effets pour la 1<sup>ère</sup>, la 2<sup>ème</sup>, et la 3<sup>ème</sup> lactation ainsi qu'une valeur « globale » moyenne sur ces 3 lactations, avec une pondération équivalente entre les différentes lactations. Ces effets que nous qualifierons dans la suite de cette étude de « globaux » sont semblables dans leur interprétation à ceux dont nous disposons avec l'évaluation génétique sur les données lactations. Cependant, l'évaluation contrôles élémentaires offre l'opportunité de disposer de nouvelles informations telles qu'une mesure de la persistance pour chacune des lactations. Comme nous l'avons vu précédemment, Druet *et al.* (2005) ont mis en évidence l'interprétation possible des six coefficients estimés dans le modèle de référence comme potentiel de production et mesure de persistance pour chacune des lactations. Ces informations novatrices constituent un des intérêts majeurs du modèle contrôles élémentaires. Elles peuvent permettre notamment la sélection d'animaux plus persistants, ce qui présente un intérêt économique pour le fonctionnement de l'exploitation agricole (régularité des livraisons de lait et donc du revenu) et un intérêt physiologique pour les vaches productives grâce à la réduction du stress métaboliques en début de lactation (Gipson and Grossman, 1990). Ces effets, que nous qualifierons d'« élémentaires » dans la suite de cette étude, constituent donc le deuxième élément de comparaison des deux approches, la valeur « globale » correspondant à une combinaison des index élémentaires. Toutefois, la réorganisation de la structure de la matrice liée de l'utilisation d'approche de rang réduit ne nous permet plus de disposer directement de ces informations. Afin de dépasser cet obstacle, Tarrés *et al.* (2008) ont proposé une méthode pour ré-exprimer les valeurs génétiques sous forme des caractères élémentaires initiaux, avant réduction de rang, basée sur les propriétés des vecteurs orthogonaux ( $\mathbf{E}^{-1} \mathbf{E} = \mathbf{E}' \mathbf{E} = \mathbf{I}$ ), en utilisant une rétro-transformation. Il est possible d'obtenir un vecteur de solutions génétique



ou d'environnement permanent ( $\mathbf{sol}$ ) pour chaque jour de lactation de taille identique au modèle de référence où  $\mathbf{U}_{1005,4}^* \mathbf{sol}_4^* = \mathbf{U}_{1005,6} \mathbf{sol}_6^*$  en prenant  $\mathbf{sol}_6^* = \mathbf{V}_{6,4}^* \mathbf{sol}_4^*$ . Nous disposons ainsi de solutions directement interprétables. Une expression similaire s'applique à l'approche COR avec  $\mathbf{sol}_{r,6}^* = \mathbf{S}_{6,6} \mathbf{T}_{6,6}^{-1} \mathbf{V}_{r,6,4}^* \mathbf{sol}_{r,4}^*$ .

## Résultats

Pour évaluer les conséquences de la réduction du rang des matrices pour les différents effets aléatoires, le pourcentage de variance captée par les 4 valeurs propres et vecteurs propres conservés (6 dans le cas de l'effet troupeau – année) est un bon indicateur des changements attendus (Tableau 1). Ainsi, sur l'effet génétique, ce pourcentage varie entre 98.4% et 99.8% suivant l'approche employée et le caractère. Ce niveau très élevé laisse supposer que la réduction du rang entraîne des modifications mineures sur les valeurs des effets estimés. Les deux approches pour la réduction de rang sont globalement semblables. Pour l'effet d'environnement permanent, la situation est différente. L'approche COR donne des pourcentages de variance expliquée compris entre 87.6 et 91.3% alors que l'approche COV fluctue entre 91.3 et 94.4% suivant les caractères. Les changements attendus sur les valeurs estimées sont plus importants. Comme pour l'effet génétique, les 6 valeurs propres et vecteurs propres conservés pour l'effet troupeau – année captent entre 96.6 et 99.7% de la variance, sans présenter de différences suivant l'approche utilisée.

L'analyse de l'évolution des corrélations entre jours de lactation liée à la réduction du rang des matrices peut également contribuer à déterminer l'approche la plus adéquate. La figure 1 représente les corrélations entre jours de lactation estimées avec les matrices de référence pour l'effet génétique (triangle supérieur) et l'effet d'environnement permanent (triangle inférieur). L'effet de la méthode de réduction du rang (COV vs COR) est clairement visible sur la figure 2 qui représente les corrélations obtenues pour l'effet génétique avec chacune des 2 approches. La figure 3 correspond quant à elle à l'effet d'environnement permanent. Comparativement aux corrélations de référence (figure 1), l'approche COR (triangle supérieur) montre une augmentation du niveau moyen des corrélations de l'effet génétique de 0.05 entre la 1<sup>ère</sup> lactation et les lactations ultérieures. L'approche COV (triangle inférieur) présente des augmentations d'amplitude plus faible : en moyenne 0.01. Pour l'effet d'environnement permanent, ce sont les corrélations entre la 1<sup>ère</sup> lactation et la 2<sup>ème</sup> lactation, ainsi qu'entre la 2<sup>ème</sup> et la 3<sup>ème</sup> lactation qui augmentent respectivement de 0.15 et 0.26 en moyenne avec l'approche COR (triangle supérieur). Les augmentations sont en moyenne

nettement plus modérées (entre 0.02 et 0.04) avec l'approche COV (triangle inférieur), à l'exception des corrélations au sein de la 1<sup>ère</sup> lactation qui augmentent en moyenne de plus de 0.07. La réduction de rang basée sur les matrices de (co)variance (COV) perturbe donc nettement moins les corrélations entre effets mesurés à des stades de lactation différents qu'une réduction de rang basée sur les matrices de corrélation (COR).

L'analyse des corrélations entre valeurs « globales » estimées à partir des matrices de référence et les valeurs « globales » obtenues avec les approches de rang réduit pour les 3 premières lactations (Tableau 2) montre un niveau de corrélation élevé pour l'effet génétique. Les corrélations sont comprises entre 0.997 et 1.000 quelle que soit l'approche de réduction du rang utilisée, à l'exception des caractères quantité de lait et matière protéique où l'approche COR présente des corrélations légèrement inférieures (respectivement 0.974 et 0.985). Pour l'effet d'environnement permanent, les corrélations obtenues avec l'approche COV sont comprises entre 0.992 et 0.999 ce qui laissent supposer des changements mineurs. Par contre, avec l'approche COR, les corrélations avec les valeurs de référence sont plus faibles (entre 0.948 et 0.993). Pour l'effet troupeau – année, la situation se rapproche de celle de l'environnement permanent avec respectivement des corrélations comprises entre 0.901 et 0.994 pour l'approche COR et 0.986 et 0.996 pour l'approche COV. Ces résultats, globalement satisfaisants pour les deux approches sont toutefois à nuancer. En effet, les valeurs « globales » obtenues sur l'ensemble des lactations cachent des disparités importantes suivant les lactations. Ainsi, la 3<sup>ème</sup> lactation présente des corrélations nettement plus faibles pour l'effet d'environnement permanent et plus particulièrement avec l'approche COR. Par exemple pour la quantité de lait, la corrélation entre les valeurs d'environnement permanent obtenues avec la matrice de référence et les valeurs d'environnement permanent obtenues avec la matrice de rang réduit n'est que de 0.760.

Lorsque nous nous intéressons aux corrélations obtenues entre le modèle de référence et les approches de rang réduit pour les index « élémentaires » correspondants aux effets « interprétables » que sont le niveau de production et celui de persistance pour chacune des lactations, nous observons là encore des différences suivant l'effet analysé (génétique ou environnement permanent), le rang de lactation et l'approche employée. D'une façon générale, les différences entre les 2 types d'approche de réduction du rang sont peu marquées sur l'effet génétique et ne montrent pas de variation importante suivant le rang de lactation que ce soit pour le niveau de production (Tableau 4) ou le niveau de persistance (Tableau 5). Le niveau de corrélation est élevé (généralement entre 0.98 et 1.00). Par contre, pour l'effet

d'environnement permanent, nous obtenons de meilleures corrélations pour les index « élémentaires » correspondant au niveau de production (Tableau 4) avec l'approche COV qu'avec l'approche COR (corrélations respectivement comprises entre 0.972 et 0.999 vs 0.761 et 0.981), les différences étant particulièrement marquées en 3<sup>ème</sup> lactation. L'index « élémentaire » correspondant à la mesure du niveau de persistance de l'effet d'environnement permanent (Tableau 5) présente des corrélations entre le modèle de référence et les approches de rang réduit globalement plus faibles. Pour la 3<sup>ème</sup> lactation, l'approche COV montre des corrélations d'un niveau semblable à celle de l'approche COR contrairement au niveau de persistance de la 1<sup>ère</sup> lactation où les corrélations obtenues entre la méthode de référence et cette approche sont faibles (entre 0.700 et 0.953). Toutefois, lorsque nous nous intéressons à la moyenne sur les 3 premières lactations des index « élémentaires » correspondant au niveau de persistance (Tableau 6), les corrélations sont élevées (supérieures à 0.984) quels que soient l'approche utilisée et le caractère analysé (génétique ou environnement permanent).

## Discussion

Parmi les trois effets (génétique, environnement permanent et troupeau – année) analysés précédemment, l'effet d'environnement permanent est apparu être l'effet le plus sensible à la réduction de rang quelle que soit l'approche utilisée. Ce résultat est cohérent avec les résultats de Druet *et al.* (2005) qui indiquaient que la part de variance expliquée par les 2 principaux vecteurs propres de la 1<sup>ère</sup> lactation n'expliquaient pour l'effet d'environnement permanent que 83.0 à 90.8% suivant les caractères alors que pour l'effet génétique, les 2 principaux vecteurs propres expliquaient 96.5 à 99.6%. D'une façon générale, le choix de l'approche COV permet d'obtenir des corrélations globalement meilleures qu'avec l'approche COR, surtout pour l'effet d'environnement permanent. Cette conclusion n'est toutefois pas vérifiée pour l'effet d'environnement permanent lorsqu'il est exprimé à travers les effets interprétables que sont le niveau de production et de persistance pour chacune des lactations. Contrairement au niveau de production, les corrélations obtenues pour l'index élémentaires correspondant à la mesure du niveau de persistance de l'effet d'environnement permanent sont faibles, surtout pour la 1<sup>ère</sup> lactation (entre 0.70 et 0.80 pour la quantité de lait et les quantités de matières). Ces résultats médiocres sont liés à des changements majeurs lors de la réduction de rang de la matrice de covariance, qui contrairement à la réduction de rang sur la matrice de corrélation, peut entraîner des modifications importantes des variances (diagonale de la matrice). La

variance de la persistance de l'effet d'environnement permanent en 1<sup>ère</sup> lactation change ainsi de façon importante avec l'approche COV. Les corrélations entre la persistance en 1<sup>ère</sup> et 2<sup>ème</sup> lactations et entre 1<sup>ère</sup> et 3<sup>ème</sup> lactations sont profondément modifiées puisqu'elles passent respectivement de 0.51 et 0.45 avec la matrice de référence à 0.98 et 0.98 avec la matrice de rang réduit. Ces changements, même si l'index « élémentaire » de la mesure du niveau de persistance en 1<sup>ère</sup> lactation ne reflète plus la nature du caractère initial avec l'utilisation d'une approche COV, présentent l'avantage d'accroître la stabilité des index au passage de la 1<sup>ère</sup> lactation aux lactations suivantes. L'effet d'environnement permanent n'a toutefois pas d'impact sur la sélection des taureaux reproducteurs, puisque par définition, ce n'est pas un effet transmissible. De plus, l'approche de réduction de rang du modèle a été développée avec le souci de limiter l'impact sur l'estimation des effets aléatoires grâce au maintien de la variance totale constante.

Concernant la diffusion des index estimés, aucune décision n'est arrêtée au moment de la rédaction de ce document. Cependant, il est probable qu'elle se limite, dans un premier temps au moins, à l'index « global » d'origine génétique obtenu pour chaque caractère avec des pondérations entre lactations qui restent à déterminer, ainsi qu'à une mesure de la persistance (sûrement celle de la 1<sup>ère</sup> lactation). A notre connaissance, la limitation de la diffusion à un index génétique « global » est employée dans l'ensemble des pays dont l'évaluation génétique est basée sur un modèle « contrôles élémentaires ». Diffuser pour chacun des caractères de multiples index ne serait pas pertinent, la multiplication des informations nuisant généralement à leur utilisation. Des pays comme le Canada ou les Pays-Bas ont également développé un index quantifiant l'aspect génétique d'un caractère décrit comme la persistance. Un indicateur relatif à la maturité de l'animal pourrait aussi être envisageable, estimé à partir de la différence entre la valeur génétique en 1<sup>ère</sup> lactation et la valeur génétique en 3<sup>ème</sup> lactation, comme c'est actuellement le cas au Pays-Bas (NRS, 2008). Aucun pays ne transmet, à notre connaissance, les effets d'environnement permanent estimés (et encore moins l'effet troupeau-année) en vue d'une valorisation ultérieure.

## **Conclusion**

Compte tenu des fortes corrélations obtenues entre les valeurs estimées des effets avec les matrices de référence et les valeurs estimées avec des matrices de rang réduit, il est donc possible de recourir à des approches de réduction du rang des matrices de variance-covariance afin de limiter les besoins calculatoires. Avec des corrélations généralement proches ou

supérieures à 0.99, l'impact de la réduction de rang est très limité dès lors que nous nous intéressons aux valeurs « globales » de chacun des effets génétique, environnement permanent et troupeau-année sans distinguer le rang de lactation (valeur moyenne sur l'ensemble des lactations). L'approche COV donne des valeurs estimées plus proches des valeurs de référence, notamment pour l'effet d'environnement permanent que l'approche COR. C'est donc cette approche qui est utilisée pour l'établissement du modèle final d'évaluation génétique française sur les contrôles élémentaires.

### **Remerciements**

Cette étude a été réalisée dans le cadre de l'Unité Mixte Technologique (UMT) INRA - Institut de l'Élevage sur l'évaluation génétique des bovins. Le ministère de l'agriculture est remercié pour avoir contribué à son financement à travers l'action innovante UTILEG.

### **Références**

- De Roos, A.P.W., Harbers A.G.F. and De Jong G. 2001. Random regression test-day model in the Netherlands. *Interbull Bull.* 27: 155-158.
- Druet T., Jaffrézic F., Boichard D. and 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.* 86: 2480-2490.
- Druet T., Jaffrézic F. and Ducrocq V. 2005. Estimation of genetic parameters for test day record of dairy traits in the first three lactations. *Genetics, Selection and Evolution.* 37: 257-271.
- Gipson T.A. and Grossman M. 1990. Lactation curves in dairy goats: a review. *Small Ruminant Research.* 3: 383-396.
- Lidauer M., Mäntysaari E.A. and Strandén I. 2003. Comparison of test-day models for genetic evaluation of production traits in dairy cattle. *Livestock Production Science.* 79: 73-86.
- Lidauer M., Pedersen J., Pösö J., Mäntysaari E.A., Strandén I., Madsen P., Nielsen U.S., Eriksson J.-Å., Johansson K. and Aamand G.P. 2006. Joint Nordic test-day model: evaluation model. *Interbull Bull.* 35: 103-107.

Liu Z., Reinhardt F. and Reents R. 2004. A multiple-trait MACE model for international bull comparison based on daughter yield deviations. *Interbull Bull.* 32: 46-52.

Misztal I., Strabel T., Jamrozik J., Mäntysaari E.A. and Meuwissen T.H.E. 2000. Strategies for estimating the parameters needed for different test-day models. *Journal of Dairy Science.* 83: 1125-1134.

NRS. 2008. Breeding value estimation of milk production traits with test-day model E7. <http://www.nrs.nl/index-eng.htm>. Accessed July. 15, 2008.

Reinhardt, F., Liu Z., Bünger A., Dopp L., Bormann J. and Reents R. 2002. Impact of application of a random regression test-day model to production trait genetic evaluations in dairy cattle. *Interbull Bull.* 29: 103-107.

Schaeffer, L.R., Jamrozik J., Kistemaker G.J. and Van Doormaal B.J. 2000. Experience with a test-day model. *J. Dairy Sci.* 83: 1135-1144.

Sullivan P. and Wilton J.W.. 2001. Multiple-trait MACE with a variable number of traits per country. *Interbull Bull.* 27: 68-72.

Tarrés J., Liu Z., Ducrocq V., Reinhardt F. and Reents R. 2008. Data transformation for rank reduction in multiple-trait MACE model for international bull comparison. *Genetics, Selection and Evolution.* 40: 295-308.

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

Wiggans G.R. and Goddard M.E. 1997. A computationally feasible test day model for genetic evaluation of yields traits in the United States. *Journal of Dairy Science.* 80: 1785-1800.

Tableau 1. Pourcentage de variance captée par les 4 principaux vecteurs propres associés à l'effet génétique et à l'effet d'environnement permanent (6 pour l'effet troupeau – année) à partir d'une décomposition basée soit sur la matrice de corrélation (COR), soit sur la matrice de covariance (COV) de référence ayant un rang initial de 6 (9 pour l'effet troupeau – année).

Effet	Génétique		Environ <sup>t</sup> permanent		Troupeau – année	
	COR	COV	COR	COV	COR	COV
Lait	98.37	98.50	87.56	94.37	99.67	99.74
MG	99.67	99.80	87.98	93.98	96.93	97.98
MP	98.43	98.68	88.26	94.38	96.78	97.64
TB	99.69	99.85	91.29	91.27	98.14	97.98
TP	99.40	99.26	89.21	92.40	96.59	96.60

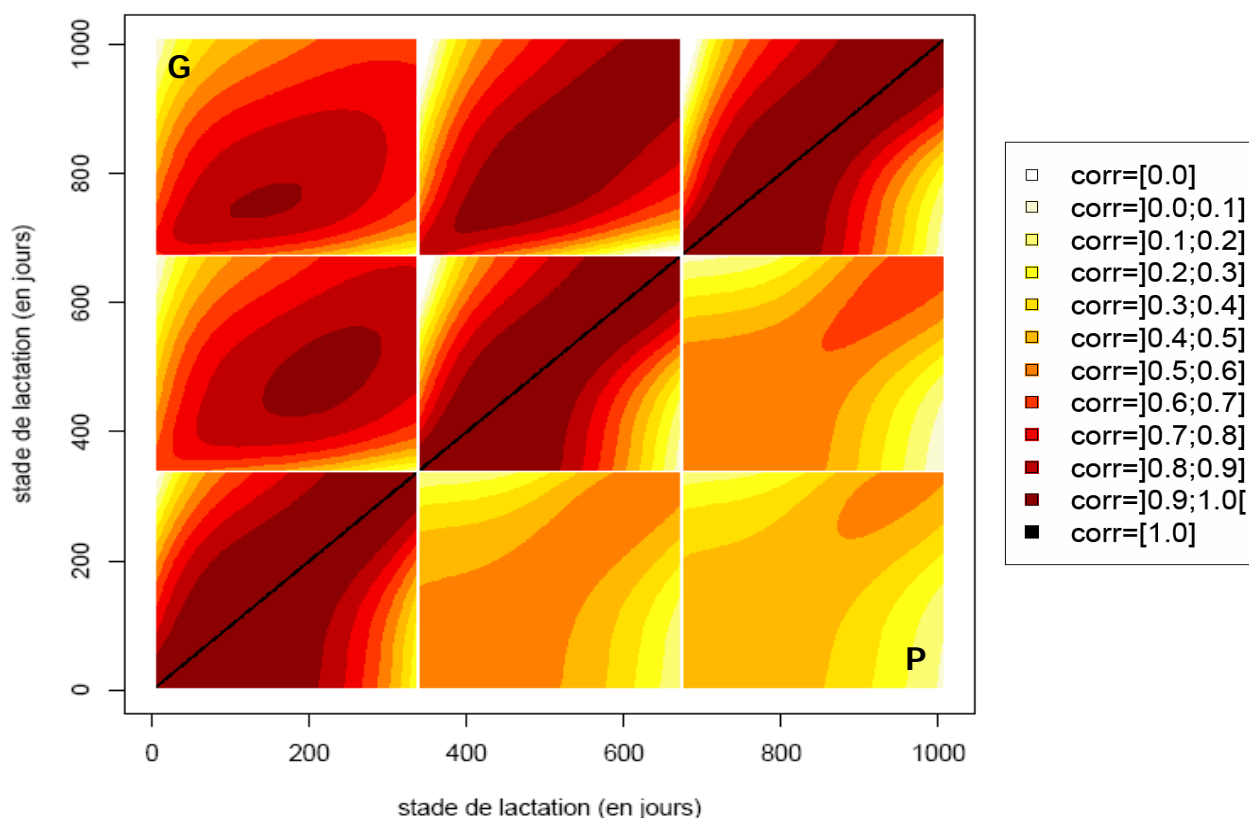


Figure 1. Corrélations entre les effets à différents stades de lactation estimées à partir de la **matrice de référence** pour l'effet génétique G (triangle supérieur) et pour l'effet d'environnement permanent P (triangle inférieur) du caractère 'quantité de lait' pour les 3 premières lactations (de 335 jours chacune) de la race Montbéliarde pour.

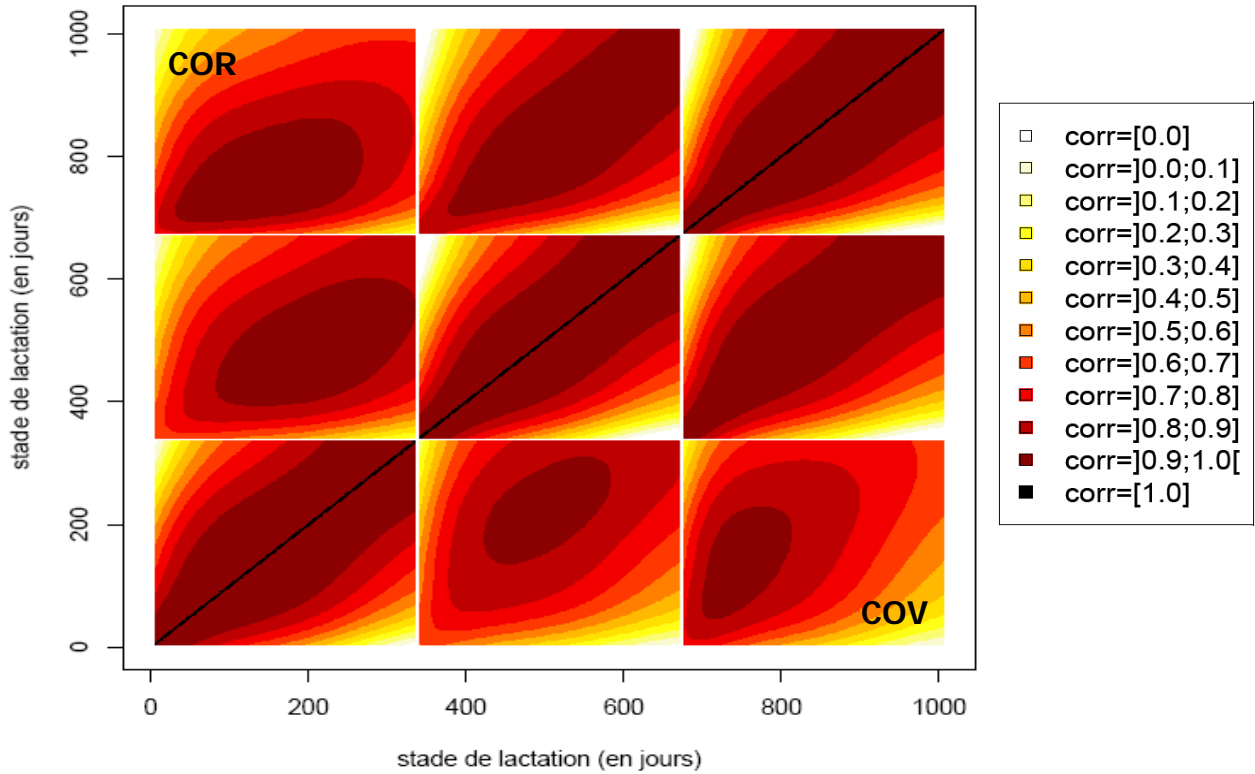


Figure 2. Corrélations entre les effets des différents stades de lactation pour l'effet génétique du caractère 'quantité de lait' pour les 3 premières lactations (de 335 jours chacune) de la race Montbéliarde obtenues sur une **matrice de rang réduit à 4 vecteurs propres** à partir de la matrice de corrélation **COR** (triangle supérieur) et à partir de la matrice de covariance **COV** (triangle inférieur).

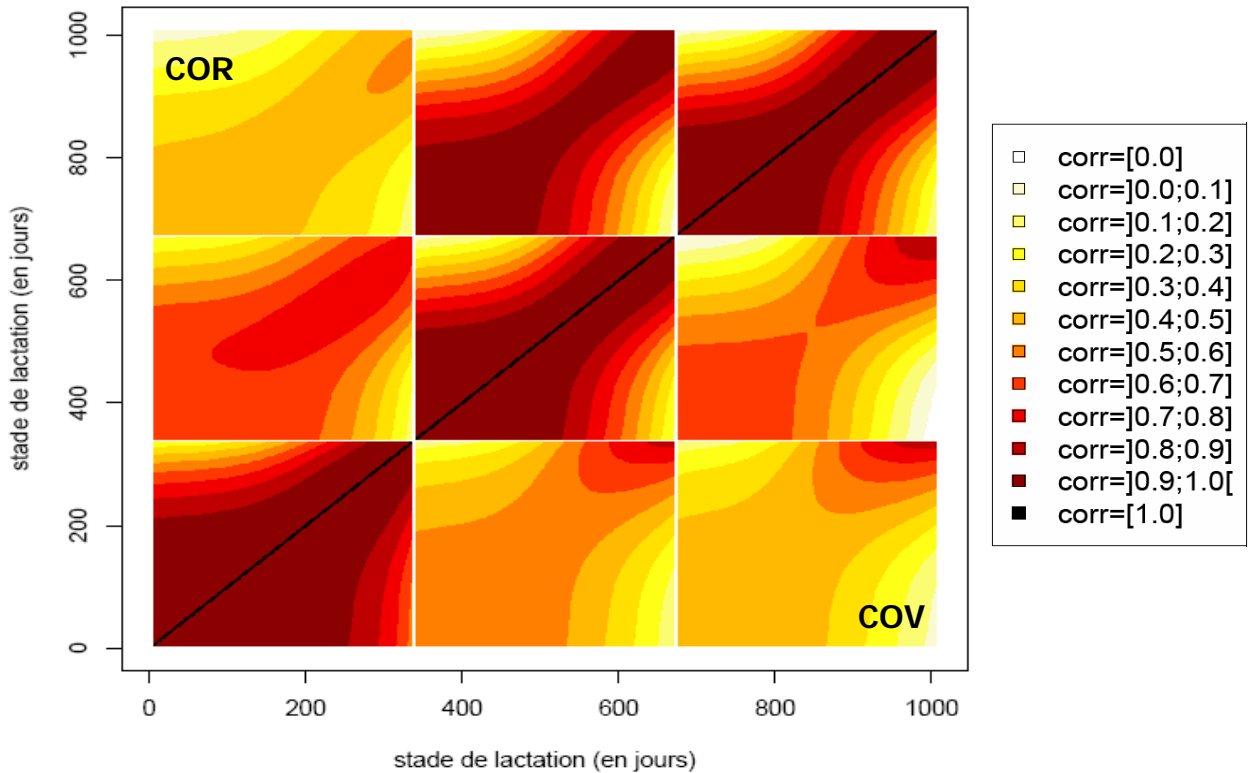


Figure 3. Corrélations entre les effets des différents stades de lactation pour l'effet d'environnement permanent du caractère 'quantité de lait' pour les 3 premières lactations (de 335 jours chacune) de la race Montbéliarde obtenues sur une **matrice de rang réduit à 4 vecteurs propres** à partir de la matrice de corrélation **COR** (triangle supérieur) et à partir de la matrice de covariance **COV** (triangle inférieur).



Tableau 2. Corrélations entre les **valeurs « globales »** des index des **3 premières lactations** (pondération de  $\frac{1}{3}$  pour chacune des lactations) obtenues d'une part avec une matrice de référence (rang = 6 pour l'effet génétique et l'effet d'environnement permanent et 9 pour l'effet troupeau – année) et d'autre part avec une matrice de rang réduit (rang = 4 pour l'effet génétique et l'effet d'environnement permanent et 6 pour l'effet troupeau – année) selon la matrice (corrélation COR ou covariance COV) sur laquelle est basée la décomposition canonique.

Effet	Génétique		Environ <sup>t</sup> permanent		Troupeau – année	
	COR	COV	COR	COV	COR	COV
Approche						
Lait	0.974	0.999	0.948	0.999	0.901	0.987
MG	0.997	0.999	0.989	0.999	0.946	0.995
MP	0.985	0.999	0.971	0.999	0.915	0.992
TB	0.999	1.000	0.993	0.999	0.994	0.996
TP	0.997	0.999	0.952	0.992	0.974	0.986

Tableau 3. Corrélations entre les **valeurs « globales »** des index pour la **3<sup>ème</sup> lactation** obtenues d'une part avec une matrice de référence (rang = 6 pour l'effet génétique et l'effet d'environnement permanent et 9 pour l'effet troupeau – année) et d'autre part avec une matrice de rang réduit (rang = 4 pour l'effet génétique et l'effet d'environnement permanent et 6 pour l'effet troupeau – année) selon la matrice (corrélation COR ou covariance COV) sur laquelle est basée la décomposition canonique.

Effet	Génétique		Environ <sup>t</sup> permanent		Troupeau – année	
	COR	COV	COR	COV	COR	COV
Approche						
Lait	0.957	0.996	0.760	0.997	0.897	0.987
MG	0.996	0.999	0.905	0.999	0.948	0.997
MP	0.978	0.997	0.855	0.997	0.918	0.992
TB	0.997	1.000	0.928	0.952	0.995	0.996
TP	0.994	0.995	0.893	0.976	0.974	0.985

Tableau 4. Corrélations entre les **valeurs « élémentaires »** de l'index **« niveau de production »** obtenues d'une part avec une matrice de plein rang (6 pour l'effet génétique et l'effet d'environnement permanent) et d'autre part avec une matrice de rang réduit (4 pour l'effet génétique et l'effet d'environnement permanent) dont les solutions ont été transformées pour être directement interprétables selon la matrice (corrélation COR ou covariance COV) sur laquelle est basée la décomposition canonique pour la 1<sup>ère</sup> lactation et la 3<sup>ème</sup> lactation.

Effet	Génétique				Environnement permanent			
	1 <sup>ère</sup> lactation		3 <sup>ème</sup> lactation		1 <sup>ère</sup> lactation		3 <sup>ème</sup> lactation	
	COR	COV	COR	COV	COR	COV	COR	COV
Approche								
Lait	0.977	0.997	0.960	0.997	0.948	0.997	0.761	0.997
MG	0.992	1.000	0.996	0.999	0.973	0.999	0.906	0.999
MP	0.982	0.996	0.978	0.996	0.957	0.996	0.862	0.998
TB	0.999	1.000	0.997	1.000	0.980	0.997	0.926	0.972
TP	0.992	0.996	0.994	0.996	0.981	0.985	0.978	0.991

Tableau 5. Corrélations entre les **valeurs « élémentaires »** de l'index « **niveau de persistance** » pour la **1<sup>ère</sup> lactation** et la **3<sup>ème</sup> lactation** obtenues d'une part avec une matrice de référence (rang = 6 pour l'effet génétique et l'effet d'environnement permanent) et d'autre part avec une matrice de rang réduit (rang = 4 pour l'effet génétique et l'effet d'environnement permanent) dont les solutions ont été transformées pour être directement interprétables selon la matrice (corrélation COR ou covariance COV) sur laquelle est basée la décomposition canonique.

Effet	Génétique				Environnement permanent			
	1 <sup>ère</sup> lactation		3 <sup>ème</sup> lactation		1 <sup>ère</sup> lactation		3 <sup>ème</sup> lactation	
Approche	COR	COV	COR	COV	COR	COV	COR	COV
Lait	0.937	0.982	0.977	0.976	0.972	0.795	0.924	0.933
MG	0.999	0.989	0.995	0.996	0.937	0.700	0.928	0.938
MP	0.992	0.978	0.990	0.992	0.974	0.709	0.918	0.916
TB	0.999	1.000	0.999	0.998	0.990	0.953	0.987	0.938
TP	0.994	0.996	0.995	0.994	0.896	0.874	0.827	0.862

Tableau 6. Corrélations entre les **valeurs « élémentaires »** de l'index « **niveau de persistance** » des **3 premières lactations** (pondération de  $1/3$  pour chacune des lactations) obtenues d'une part avec une matrice de référence (rang = 6 pour l'effet génétique et l'effet d'environnement permanent) et d'autre part avec une matrice de rang réduit (rang = 4 pour l'effet génétique et l'effet d'environnement permanent) dont les solutions ont été transformées pour être directement interprétables selon la matrice (corrélation COR ou covariance COV) sur laquelle est basée la décomposition canonique.

Effet	Génétique		Environnement permanent	
	COR	COV	COR	COV
Lait	0.989	0.993	0.984	0.988
MG	0.998	0.995	0.990	0.992
MP	0.992	0.992	0.988	0.987
TB	1.000	1.000	0.999	0.999
TP	0.998	0.998	0.991	0.995



## **Partie 3.3**

# **Validation des effets fixes du modèle et réduction des besoins informatiques**

## INTRODUCTION

Comme nous avons pu le décrire précédemment, de nombreux effets influencent la production laitière en termes de quantité (lait, matière grasse, matière protéique) et de qualité (taux butyreux et protéique, numération cellulaire). Outre les effets aléatoires qui dépendent généralement de l'animal (effet génétique et d'environnement permanent), nombre d'effets environnementaux fixes interviennent sur les performances enregistrées. Ces effets sont partagés par l'ensemble des vaches d'une même race, produisant la même année dans une même région, vêlées à la même saison, à un âge similaire... Dans le cadre de la mise en place d'un modèle d'évaluation génétique basé sur les contrôles élémentaires en France, un modèle, où chacune des races est analysée séparément, a été défini après des analyses préliminaires. Afin de valider ce modèle, présenté dans la partie 2.4, nous avons effectué quelques analyses complémentaires.

Compte tenu de la complexité du modèle pressenti et donc du coût informatique lié à une telle analyse, nous avons 2 objectifs :

- ✓ confirmer que chacun des effets fixes autres que l'effet troupeau – jour de contrôle (noté HTD pour Herd – Test day), inclus dans le modèle et notamment les nombreuses interactions était nécessaire pour modéliser correctement la performance,
- ✓ tester une alternative technique pour alléger les coûts informatiques d'un tel modèle afin qu'il soit envisageable de l'utiliser pour des évaluations génétiques officielles à l'échelle nationale (supérieur à 200 millions de données pour chacun des caractères évalués (Lait, MG, MP et SCS) en race Holstein, 3 fois par an dans un délai restreint d'environ 15 jours). Cette alternative, découlant des résultats obtenus sur le 1<sup>er</sup> aspect de cette étude, consiste à précorriger les données pour l'ensemble des effets fixes qui ne varient pas d'une année sur l'autre.

## MATERIEL ET METHODES

Les analyses ont été réalisées sur les 5 caractères de production laitière ainsi que la numération cellulaire, à partir de données de la race Montbéliarde collectées par le contrôle laitier entre septembre 1988 et décembre 2005 dans le département du Jura, soit de l'ordre de 2.4 millions de contrôles élémentaires. La base de données du SIG permet de disposer de l'ensemble des informations nécessaires à la détermination des effets environnementaux inclus dans les modèles.

Pour répondre à notre premier objectif, nous avons étudié la partie du modèle d'évaluation génétique correspondant aux effets fixes. Outre l'effet troupeau – jour de contrôle, le modèle pressenti (modèle A, appelé *Meff x year* dans l'article suivant) comprend 4 courbes modélisées par des splines décrivant la partie « fixée » de la trajectoire de la lactation. Trois des splines comprennent 6 nœuds (6 paramètres) et sont définies en fonction du stade de lactation pour tenir compte de l'effet de l'âge au vêlage, du mois de vêlage et de la durée de tarissement. La dernière spline ne comprend que 4 nœuds et permet de tenir compte de l'effet du stade de

gestation sur la lactation au-delà de 100 jours de gestation. La trajectoire globale de la courbe de lactation comprend donc 22 paramètres. L'ensemble de ces effets est défini en interaction avec l'année, la région et le rang de lactation. Afin de déterminer les effets fixes environnementaux réellement nécessaires pour la modélisation correcte des performances, ce modèle a été comparé à des modèles simplifiés. Dans un premier temps, pour évaluer l'importance de la prise en compte de l'interaction de l'effet de l'année sur la forme des courbes de lactation, un modèle similaire en terme de nombre de splines mais n'incluant pas d'interaction avec l'année a été testé (modèle B, appelé *Meff* dans l'article suivant). Pour prendre en compte l'effet global de l'année de production, 3 constantes ont été ajoutées au modèle de manière à tenir compte de l'interaction de l'année, la région et le rang de lactation avec les effets du mois de vêlage, de l'âge au vêlage et de la durée de tarissement. Le modèle B suppose donc que les courbes de lactation ne diffèrent entre les années que par une constante. Dans un second temps, nous nous sommes attachés à vérifier l'importance des différents effets fixes que sont le mois de vêlage, l'âge au vêlage, la durée de tarissement et la gestation sur la modélisation des performances. Pour ce faire, nous avons supprimé du modèle B (ou *Meff* dans l'article suivant) les splines modélisant chacun des effets une à une (modèle noté *Meff-effet*). Ces 2 aspects ont été étudiés sous SAS avec des procédures d'analyse de variance GLM (General Linear Model) des résiduelles en fonction du stade de lactation et de gestation ainsi qu'à travers des analyses graphiques des courbes obtenues. Tout écart systématique des résiduelles (par rapport à une moyenne de 0 attendue) à un moment quelconque de la lactation traduit une inadéquation du modèle.

Compte tenu des résultats obtenus dans la première partie de cette étude et pour répondre au second objectif, il est envisagé de précorriger les données pour les effets modélisant la forme de la courbe de lactation et ne variant pas d'une année sur l'autre. Comme nous allons le voir dans la synthèse des résultats, il s'agit des 4 splines du modèle B ou *Meff*, soit les 22 coefficients décrits précédemment. Cette alternative technique permettant de réduire les besoins de mémoire et de limiter les temps de calcul nous a conduits à étudier une stratégie consistant à scinder l'évaluation génétique en deux étapes. La première étape consisterait à estimer les facteurs de précorrection, c'est-à-dire les coefficients des splines, à partir d'un modèle n'incluant que les effets fixes du modèle, soit 26 paramètres (HTD + 3 constantes + 22 coefficients de splines). Cette étape serait par exemple réalisée une fois par an, afin d'avoir des facteurs régulièrement mis à jour. A travers ce re-calcul systématique, les reproches faits à certains modèles lactation précorrigent les données avec des facteurs obsolètes (Ptak *et al.*, 1993) ne pourront être adressés au modèle contrôles élémentaires français. La deuxième étape serait réalisée à partir de données ainsi précorrigées pour la forme des courbes de lactation. Le modèle utilisé contiendrait alors les effets fixes dépendant du temps tels que l'effet troupeau – jour de contrôle et les autres effets incluant une interaction avec l'année, ainsi que l'ensemble des effets aléatoires, soit 18 paramètres (HTD + 3 constantes + 4 effets génétiques + 4 effets d'environnement permanent + 6 effets troupeau – année). Deux pré-requis sont toutefois nécessaires pour mettre en place une telle approche. Il faut tout d'abord s'assurer que les formes des courbes de lactation sont stables au cours du temps afin de ne pas engendrer de biais dans les évaluations. De plus, il est nécessaire de quantifier l'impact de la précorrection sur l'estimation des effets aléatoires. Pour cela, nous avons comparé les résultats d'une

évaluation génétique en 2 étapes, c'est-à-dire avec précorrection des données pour la forme de la courbe de lactation, à ceux obtenus lors d'une évaluation en 1 seule étape.

Ces analyses détaillées sont décrites en détails dans l'article qui suit. Seuls les principaux résultats sont présentés dans cette synthèse.

## RESULTATS ET DISCUSSION

### Effet de l'année dans la définition de l'interaction définissant les courbes de lactation

La prise en compte du facteur *année* dans la définition de la forme des courbes de lactation contribue de façon importante à la complexité du modèle. Même si le nombre de termes à estimer augmente peu proportionnellement au nombre total d'équations à résoudre dans le modèle, estimer 15 courbes à la place d'une seule peut devenir très délicat notamment pour les classes basées sur un faible nombre de données (vêlage très précoce, durée de tarissement très courte...). Malgré la suppression des classes ayant de très faibles effectifs (moins de 250 contrôles élémentaires soit environ 25 lactations), les courbes estimées à partir d'effectifs restreints présentent une précision inférieure et augmentent considérablement le nombre d'itérations nécessaires pour atteindre la convergence de l'algorithme itératif de résolution. Regrouper les années permet dans une telle situation de disposer de coefficients plus robustes et de réduire nettement les temps de calcul. Ainsi pour le fichier de données présenté précédemment, nous sommes passés de 760 itérations pour un critère de convergence donné lorsque les courbes dépendaient de l'année (modèle A ou  $Meff \times year$  dans l'article suivant) à 161 itérations sans cette interaction (modèle B ou  $Meff$  dans l'article suivant).

Les analyses de variances effectuées sur les résiduelles des modèles A (ou  $Meff \times year$ ) et B (ou  $Meff$ ) ne permettent pas de mettre en évidence un impact important de l'année sur la forme

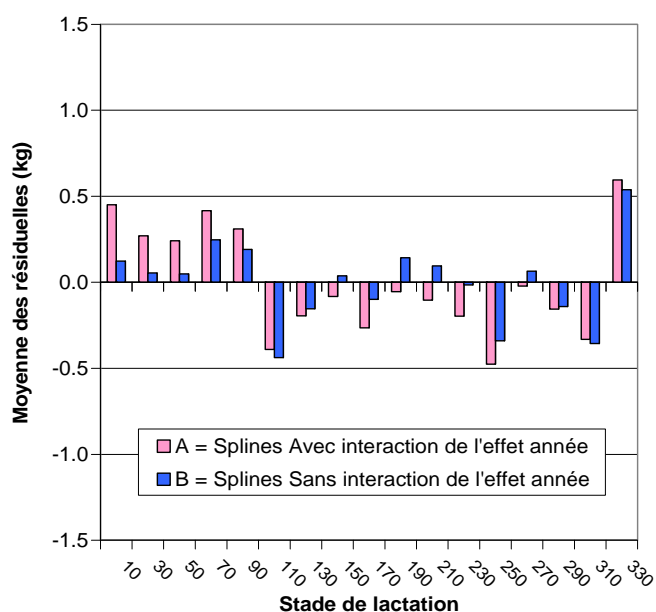


Figure 3.1. Comparaison des résiduelles obtenues avec les modèles A et B par classe de 20 jours de stade de lactation (modèle A : courbes définies en fonction de l'année ; modèle B : courbes ne tenant pas compte de l'année, effet modélisé sous forme d'une constante).

des courbes de lactation comme le montrent les résultats figurant dans l'article suivant. De plus, l'analyse des résidus des modèles A et B en fonction du stade de lactation ne montre pas de biais liés à l'utilisation d'un modèle simplifié tel que le modèle B (Figure 3.1).

D'après ces résultats, la forme des courbes de lactation est indépendante de l'année de production. Les modèles A (ou  $Meff \times year$ ) et B (ou  $Meff$ ) modélisent donc les performances avec une qualité quasi équivalente. Ces 2 modèles peuvent être représentés schématiquement selon la figure 3.2.

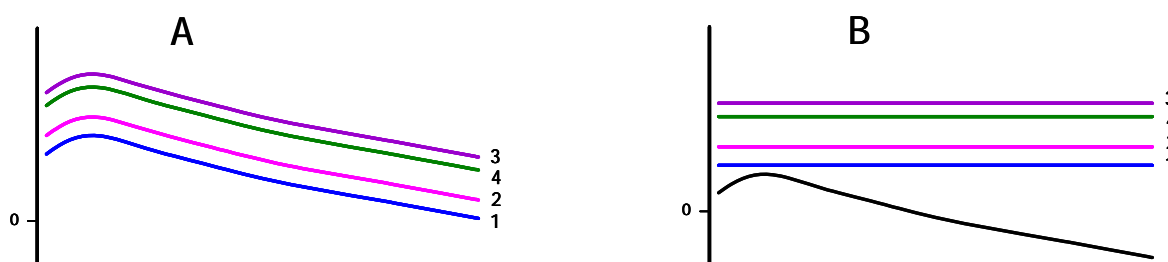


Figure 3.2. Représentation schématique de la prise en compte de l'effet de l'année, notée de 1 à 4, dans un modèle d'évaluation génétique en interaction avec les autres facteurs définissant la courbe de lactation (A) ou sous forme d'une constante s'ajoutant à une courbe de lactation commune à toutes les années (B).

### Confirmation des effets requis pour modéliser la forme de la courbe de lactation

Afin de confirmer l'intérêt de chacun des effets (âge au vêlage, mois de vêlage, durée de tarissement et gestation) pris en compte à travers une spline, les analyses des résiduelles obtenues avec les modèles où l'une des splines était supprimée (par exemple  $Meff-age$ ) ont été comparés à celles obtenues avec le modèle  $Meff$ . Les analyses en fonction du stade de lactation ou de gestation à l'aide de GLM ont systématiquement mis en évidence des biais importants sur l'évaluation génétique lorsque l'effet en question était supprimé. Ainsi la suppression de l'effet de gestation (Figure 3.3) entraîne un biais pouvant atteindre 3 kg de lait par jour pour

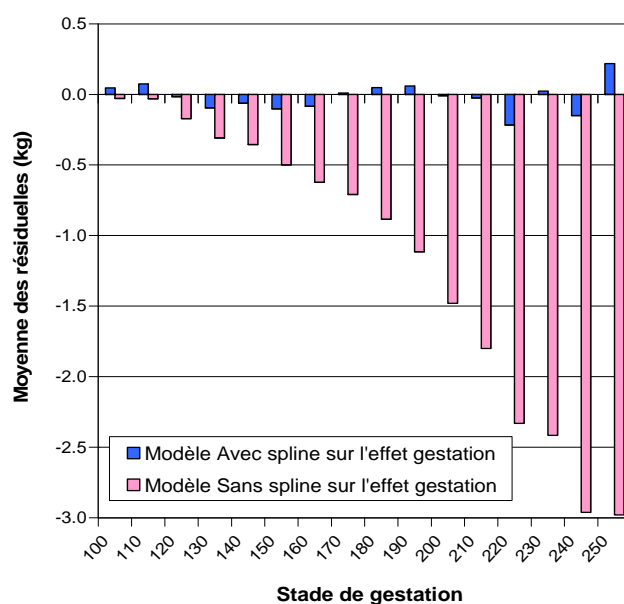


Figure 3.3. Comparaison des résiduelles obtenues avec un modèle incluant des splines sur les 4 effets (âge, mois de vêlage, durée de tarissement et gestation) et un modèle dans lequel la spline sur l'effet de la gestation a été supprimé par classe de 10 jours de stade de gestation.



un animal à 8 mois de gestation. L'impact sur la production des effets de l'âge, du mois de vêlage, de la durée de tarissement et de la gestation sont décrits et discutés comparativement aux données bibliographiques dans l'article qui suit.

### **Stabilité des courbes de lactation au cours du temps**

En pratique, il paraîtrait a priori raisonnable que les coefficients de correction des courbes de lactation soient estimés chaque année. Toutefois, pour se placer dans une situation plus extrême pour vérifier la stabilité des splines estimés pour chacun des effets, la comparaison a été effectuée à partir de coefficients estimés à 4 ans d'intervalle. Les coefficients ont ainsi été estimés avec l'ensemble des performances enregistrées jusqu'à fin 2001 d'une part et fin 2005 d'autre part. Pour le caractère « quantité de lait », les corrélations entre les coefficients obtenus dans chacun des fichiers sont en moyenne de 0.985. Quoique très élevée, cette valeur laisse présager de légères différences entre les courbes obtenues à 4 ans d'intervalle.

L'intérêt principal des évaluations génétiques reste cependant l'estimation des valeurs génétiques qui sont utilisées pour la sélection des animaux reproducteurs. Nous avons donc cherché à évaluer l'impact que pouvait avoir ces légères différences sur les effets aléatoires estimés (valeurs génétiques et d'environnement permanent). Deux fichiers de données ont été obtenus en précorrigant les performances recueillies jusqu'à fin 2005 pour chacun des jeux de facteurs de précorrection (2001 ou 2005) sur l'ensemble des caractères laitiers à l'exception de la numération cellulaire. Exprimées à l'échelle de la lactation, les valeurs génétiques obtenues avec ces 2 fichiers lors de la deuxième étape de l'évaluation génétique sont très fortement corrélées (supérieures à 0.99991 pour les 5 caractères) et comprises entre 0.99969 et 0.99994 pour les effets d'environnement permanent. La forme des courbes de lactation est donc très stable dans le temps en race Montbéliarde et les faibles variations observées ont un impact minime sur l'estimation des effets aléatoires. Une mise à jour annuelle des facteurs de précorrection des données apparaît donc tout à fait suffisante.

### **Comparaison d'une évaluation génétique en 1 et 2 étapes**

L'aspect essentiel de l'évaluation génétique étant l'estimation des effets aléatoires et plus particulièrement des effets génétiques, il est essentiel de vérifier que la réalisation de l'évaluation génétique sous forme de 2 étapes n'a pas d'impact sur l'estimation de ces derniers. Les corrélations entre les valeurs génétiques, exprimées à l'échelle de la lactation, obtenues avec une évaluation génétique en 1 étape d'une part et en 2 étapes d'autre part varient entre 0.9984 et 0.9992 selon les caractères (entre 0.9981 et 0.9991 pour les effets d'environnement permanent). L'impact est donc négligeable et permet de valider l'utilisation de cette approche en 2 étapes pour réaliser les évaluations génétiques nationales. Cette approche permet aussi de réduire par 4 selon nos tests, les temps de calcul en limitant de façon importante le nombre de facteurs dont dépend la performance à chacune des étapes. Ainsi les 40 paramètres (HTD + 3 constantes + 22 coefficients de splines + 4 effets génétiques + 4 effets d'environnement permanent + 6 effets troupeau – année) nécessaires à l'évaluation en 1 étape sont réduits dans le cadre d'une évaluation en 2 étapes à 26 paramètres pour la première étape et 18 pour la seconde.

## Environmental effects on lactation curves included in a test-day model genetic evaluation

H. Leclerc<sup>†</sup>, D. Duclos, A. Barbat, T. Druet and V. Ducrocq

UR337, Station de Génétique Quantitative et Appliquée, Institut National de la Recherche Agronomique, 78352 Jouy-en-Josas, France

(Received 15 June 2007; Accepted 7 November 2007)

*A large number of environmental factors affect the daily milk production of a cow. Lactation curves included in the French test-day model are modelled as a function of days in milk with semi-parametric curves (splines). The proper modelling of environmental effects in the test-day analysis was investigated using test-day records collected from the first three lactations of French Montbéliarde cows from 1988 to 2005. Four lactation-curve effects describing calving month, length of dry period, age at calving and gestation defined within parity-class were fitted. The shape of lactation curves did not depend on year of calving, which can be modelled as a constant over the whole lactation. To reduce computational requirements and time, data were pre-adjusted in a first step for fixed effects with no year interaction, and then used for genetic evaluation. Correlations for each lactation between 305-day estimates of genetic and permanent environment effects computed using pre-adjustment factors obtained at a 4-year interval were virtually one. The use of a two-step procedure had a very limited impact on the estimates of genetic and permanent environment effects. The minimum correlations with values estimated with a one-step procedure were 0.9984 and 0.9974, respectively. The knowledge of systematic environmental effects affecting the cow daily yield through lactation curves offers interesting perspectives to predict future daily milk production.*

**Keywords:** dairy cows, environmental factors, genetic evaluation, lactation curve, test-day model

### Introduction

In addition to genetic influences, a large number of factors, usually referred to as environmental effects, such as calving year, parity, calving age, and calving season (e.g. Wood, 1969; Danell, 1982; Wilmink, 1987b) are used in models as explanatory variables of recorded traits in dairy cattle. Yet, only calving season and year are environmental factors strictly speaking, the others being characteristics of the recorded animal. Some of them are related to physiological aspects, such as parity and pregnancy effects, others, such as length of dry period and age at calving, are more linked with management practices. The aggregated 305-day yields used for genetic evaluation based on lactation data are usually adjusted for all these environmental effects. Depending on country, production records are either pre-adjusted for the identified environmental effects with multiplicative or additive factors, or corrected by fitting environmental effects as fixed effects in the animal models for genetic evaluation.

Effects of calving age on milk production were generally studied to facilitate comparison of cows differing in

maturity. Wilmink (1987a) reported an evolution of calving-age effects through the years indicating a need regularly to update factors used to pre-adjust data before genetic evaluation. Mao *et al.* (1972), among others, reported the existence of parity effects and suggested that age effects be nested within parities. Moreover, it was shown that age at calving effects depend on lactation stage. The effect of calving month exerts a considerable influence on the cow's daily yield. Herd, year and region have been shown to be the main components responsible for variation of calving month effect, in particular through food supply. In addition to these two effects which have been considered in numerous studies, length of dry period was reported as a significant environmental effect on milk production of subsequent lactation (e.g. Schaeffer and Henderson, 1972). Pregnancy is also known to have a physiological effect on the daily milk production in the last part of the lactation, through hormonal changes and partition of nutrient supply. Knowledge of systematic environmental effects affecting cow daily yield makes it possible to standardise test-day (TD) production and unbiasedly compare cows within herd (Wilmink, 1987b, Pérochon *et al.*, 1996).

During the last decade, more and more countries have expressed interest in TD models and many of them have

<sup>†</sup> E-mail: helene.leclerc@jouy.inra.fr

implemented them in routine genetic evaluations. Genetic evaluation based on TD models presents a large number of advantages in comparison with 305-day lactation model as mentioned in several reviews (in particular Swalve (2000) and Jensen (2001)). In addition to the usual benefits of TD models from a genetic viewpoint, they are useful for prediction and management purposes such as feeding or cow culling decision support, but only a few published studies specifically refer to these aspects (Mayeres *et al.*, 2004). Each country that already moved from a lactation model to a TD model, developed its own model. These models differ in the way the lactation curve is modelled as a function of days in milk (DIM), with fixed classes, parametric or semi-parametric (splines) curves; in the way the genetic and permanent environment components are described (fixed or random regression using Legendre or other polynomials); and in the way heterogeneous residual variances are accounted for. In France, regression splines (White *et al.*, 1999) were chosen to model lactation curves (Druet *et al.*, 2003).

In contrast to lactation models, there have been limited published studies on how properly to include environmental effects in TD models. The aim of the present study was two-fold. First, to determine which environmental effects should be included in the national TD model for milk yield. The definition of effects in the genetic evaluation and their possible interactions with year of calving were studied. Second, to decrease computational requirements and computing time of such models. An alternative approach relying on data pre-adjustment for shape of lactation curves was evaluated.

## Material and methods

### Data

Data were TD yields for the first three lactations of French Montbéliarde (animals with at least 7/8 of Montbéliarde blood) collected after September 1988. DIM ranged from 7 to 335 days. Days carried calf (DCC) were considered and limited to a maximum of 265 days. Cows were required to have a first-lactation record and known sire and dam to be included in the data file. Daily test contents outside of the International Committee for Animal Recording (ICAR) limits, i.e. lower than 1.5% and 1.0% or greater than 9.0% and 7.0% for fat and protein percentages, respectively (ICAR, 2005), and milk productions lower than 2.0 kg (Institut de l'Élevage, 2005) were discarded. Only cows with at least three TD were considered. At least five records were needed to define a herd–test-date effect (HTD) and a minimum number of 20 records per herd, parity and year were required.

To reduce computational requirements, most analyses were made from a regional subset of the Montbéliarde data file limited to the Jura, a French administrative region. Because of the different modellings of fixed effects, two data files hereafter called 'Jura test' and 'Jura 2005', differing by their end date of TD recording were used (Table 1). 'Jura 2005' was created as the usual national genetic evaluation files, i.e. including all TD records until a known date, December 2005 in the present. In contrast, the 'Jura test' was created to assess

various models for fixed effects through curves, of which one included year of calving interaction. Because the year effect was defined from September of year  $n$  to August of year  $n + 1$ , the data file 'Jura test' included all TD records from cows calving from September 1988 to August 2004. Daily milk productions (mean  $\pm$  s.e.) of Jura data files were representative of the Montbéliarde breed, i.e.  $19.6 \pm 4.6$  kg for the first-lactation cows,  $22.4 \pm 6.8$  kg for the second-lactation cows and  $23.6 \pm 7.0$  kg for the third-lactation cows.

### Model

$$\begin{aligned}
 Y_{ijklnopqtt'vw} = & \text{HTD}_i + y\text{month}_j + y\text{dry}_k + y\text{age}_l \\
 & + \text{spl}_{\text{DIM}}(\text{month}_n, t) + \text{spl}_{\text{DIM}}(\text{dry}_o, t) + \text{spl}_{\text{DIM}}(\text{age}_p, t) \\
 & + \text{spl}_{\text{DCC}}(\text{parity}_q, t') + \sum_{r=1}^4 (v_{\text{tar}} a_{vr}) + \sum_{r=1}^4 (\zeta_{\text{tar}} p_{vr}) \\
 & + \sum_{r=1}^6 (\tau_{\text{tar}} h_{vr}) + e_{ijklnopqtt'vw}, \quad (1)
 \end{aligned}$$

where  $y_{ijklnopqtt'vw}$  is the record of animal  $v$  in lactation  $q$  on DIM  $t$  and days carried calf (DCC)  $t'$ .  $\text{HTD}_i$  is the  $i$ th HTD effect. The interaction terms  $y\text{month}$ ,  $y\text{dry}$  and  $y\text{age}$  correspond to the  $j$ th,  $k$ th and  $l$ th year–parity–calving month, year–parity–length of dry period and year–parity–calving age, respectively. These terms are constant over the lactation.  $\text{spl}_{\text{var}}(\text{eff})$  denotes a regression spline function of the variable  $\text{var}$  specific to the effect  $\text{eff}$ . Four spline functions were included to model lactation curves according to the effects of the parity–month of calving class  $n$  ( $\text{month}$ ), the parity–length of dry period class  $o$  ( $\text{dry}$ ) and the parity–age at calving class  $p$  ( $\text{age}$ ) as a function of DIM  $t$ ,  $\text{spl}_{\text{DCC}}$  is a regression spline modelling the effect of gestation of parity  $q$  as a function of DCC  $t'$  and  $e$  is the residual term. Interactions with region are not considered in model (1) since a regional subset was considered in this study, but will be included in the national model.

The random additive genetic, permanent environment and herd–year of calving random regression effects are  $a_{vr}$ ,  $p_{vr}$  and  $h_{vr}$ , respectively, for animal  $v$  and herd–year of calving  $w$  with associated coefficients  $v_{\text{tar}}$ ,  $\zeta_{\text{tar}}$  and  $\tau_{\text{tar}}$ , at DIM  $t$  and parity  $q$ .

*Environmental fixed effects.* Regression splines (White *et al.*, 1999) were used to model lactation curves. Six-knot regression splines, with knots positioned at days 7, 20, 50, 135, 245 and 335, were considered to be an accurate way to model lactation curves (Druet *et al.*, 2003). To model the effect of gestation, DCC defined as DIM minus days open (DO, the number of days from calving to successful insemination) was used. The conception date was approximated as the next calving date minus the average gestation length for the Montbéliarde breed, i.e. 288 days. When the next calving was not (yet) available, we assumed that the last artificial insemination led to a conception, except if the

**Table 1** Description of the data sets ('Jura test' and 'Jura 2005') and number of levels for each effect

	Jura test	Jura 2005
Available test-day (TD) records	3 574 488	3 574 488
Starting date	Calving September 1988	Calving September 1988
Ending date	Calving August 2004	TD December 2005
TD selected	2 406 345	2 563 040
Herd-test-date <sup>†</sup>	137 175	143 924
Year-parity-month <sup>†</sup>	541	578
Year-parity-age <sup>†</sup>	357	383
Year-parity-dry <sup>†</sup>	165	177
Parity-month <sup>†</sup>	36	36
Parity-age <sup>†</sup>	24	24
Parity-dry <sup>†</sup>	11	11
Genetic effect <sup>†</sup>	195 217	206 294
Permanent environment <sup>†</sup>	135 749	145 247
Herd-parity-year <sup>†</sup>	33 903	36 108

<sup>†</sup>Number of levels for each combination.

resulting gestation length was abnormally large. When there was no successful insemination, DCC was zero. When DCC was less than 100 days, the effect of gestation on the analysed traits was assumed to be nil. The DCC effect ranged from 100 to 265 days. It was assumed that the effect of DCC did not differ over DIM because large DCC values are only observed in late lactation stages while small DCC values were assumed to have small effects not varying much with DIM. Four-knot regression splines with knots at DCC 100, 150, 200 and 265 (Druet *et al.*, 2003) were used to model the effect of gestation. Curves for month of calving, age at calving, length of dry period and gestation were assumed to be independent and therefore the overall fixed part of the lactation curve of an animal can be described as the sum of elementary curves over the relevant effects.

The definition of the fixed effect classes was based on the current lactation model used for the national evaluation. In the national data file, three regions were defined based on seasonal patterns and feeding systems. However, data from the Jura subset considered in this study were from the same region. Twelve seasons of calving were defined based on calendar months. Most calvings occurred during the autumn, with a peak in September (26.4% of the first lactations in the 'Jura 2005' data file). Recently, the main calving season occurred earlier (August to November in 2004 *v.* September to December in the 1990s). For the calving age effect, 24 age classes were formed: 10 for first lactation by 2-month classes starting at 22 months, the last one including calving at 40 months and more, 7 for the second lactation by 3-month classes ( $\leq 36$ , 37 to 39, ...,  $\geq 52$ ) and 7 for the third lactation by 3-month classes ( $\leq 48$ , 49 to 51, ...,  $\geq 64$ ). The desired age at first calving for the Montbéliarde breed is usually between 32 and 36 months (61.3% of the data), corresponding to extensive management of heifers. In contrast, first-lactation cows having very young calving (22 to 23 months) represented only 0.09% of calving age class for first-lactation cows. For length of dry

period, five classes were created for second- and third-lactation cows: dry period  $\leq 25$ , 26 to 45, 46 to 60, 61 to 75 days, and  $\geq 76$  days. A specific class was created for heifer effects (first-lactation cows). In fact, dry period length  $\leq 25$  days represented around 2% of the data for second and third lactations each, whereas the dry period length between 61 and 75 days represented around 40% of the data for second and third lactations. For all interaction effects, at least 100 TD yields were required to take year  $\times$  parity  $\times$  effect interaction into account and 250 for the interaction parity  $\times$  effect.

*Random effects.* The covariance matrix of the random effects was

$$\text{var} \begin{pmatrix} a \\ p \\ h \\ e \end{pmatrix} = \begin{pmatrix} \mathbf{G} \otimes \mathbf{A} & 0 & 0 & 0 \\ 0 & \mathbf{P} \otimes \mathbf{I} & 0 & 0 \\ 0 & 0 & \mathbf{H} \otimes \mathbf{I} & 0 \\ 0 & 0 & 0 & \mathbf{R} \end{pmatrix},$$

where **G**, **P** and **H** were the random regression covariance matrices initially modelled from two coefficients (three for herd-year of calving) by parity for the genetic, permanent environment and herd-year of calving effects estimated previously (Druet *et al.*, 2005), of which ranks were reduced from six to four for **G** and **P** and from nine to six for **H**. The four selected eigenvectors explained 98.4% of the total genetic variation in the three lactations and 87.6% for the permanent environment variance. For the herd-year of calving variance, the first six eigenvectors explained 99.7% of the total variance. **A** was the additive genetic relationship matrix, and **R** was the residual variance, expressed as a function of 12-knot regression splines of DIM separately for the three lactations (Druet *et al.*, 2005). Other sources of variance heterogeneity, e.g. due to herd-year effect were not accounted for, although they will be in the national genetic evaluation. Phenotypic variance was ensured to

remain the same before and after rank reduction by adding the loss of variance in **G**, **P** and **H** to **R**. The heritability ranged from 0.23 to 0.44 depending on DIM.

*Computational approach*

*Determination of environmental effects affecting the lactation curves.* Six models were compared using the 'Jura test' data to assess different models for the fixed part of the lactation curve.

The Meff × year model (2) was defined as

$$y_{ijklmtt'} = \text{HTD}_i + \text{spl}_{\text{DIM}}(\text{ymonth}_j, t) + \text{spl}_{\text{DIM}}(\text{ydry}_k, t) + \text{spl}_{\text{DIM}}(\text{yage}_l, t) + \text{spl}_{\text{DCC}}(\text{yparity}_m, t') + e_{ijklmtt'}. \quad (2)$$

Unlike the fixed effects in model (1) used as reference, model (2) did not include any constant taking year interaction into account. Instead, year interaction was included in the lactation curve modelling. Hence, the four spline functions included in model (2) accounted for the effects of year–parity–class of calving month (ymonth), year–parity–class of dry period length (ydry) and year–parity–class of calving age (yage) on variable DIM  $t$ .  $\text{spl}_{\text{DCC}}$  was a regression spline modelling the effect of gestation of year–parity (yparity)  $m$  on DCC  $t'$ .

The Meff model (3) was equivalent to the fixed part of the model (1): the regression splines modelling the lactation curve did not include any interaction with year in the definition of classes (month = parity–class of calving month  $n$ , dry = parity–class of dry period length  $o$ , age = parity–class of calving age  $p$ ). The only interaction terms including the year effect were the fixed effects ymonth, ydry and yage. In other words, the Meff model assumed that the spline functions differed over years only by a constant.

$$y_{ijklnopqtt'} = \text{HTD}_i + \text{ymonth}_j + \text{ydry}_k + \text{yage}_l + \text{spl}_{\text{DIM}}(\text{month}_n, t) + \text{spl}_{\text{DIM}}(\text{dry}_o, t) + \text{spl}_{\text{DIM}}(\text{age}_p, t) + \text{spl}_{\text{DCC}}(\text{parity}_q, t') + e_{ijklnopqtt'}. \quad (3)$$

The last four models were derived from the Meff model by excluding one type of regression splines at a time: the Meff–age model was equivalent to the Meff model without regression splines on the calving age effect. Similarly, in the Meff–dry, Meff–month and Meff–gest models, regression splines for length of dry period, calving month and gestation effect were excluded, respectively.

The daily milk productions recorded in the 'Jura test' data file (2 406 345 TD) were used to compare these models through an analysis of the residuals obtained for each of these six models. Because of memory requirements for the Statistical Analysis Systems GLM procedure used, the analyses of variance (ANOVA) were applied only to the last 6 years of the data file (1 201 018 observations). The residuals

were on average equal to zero for the whole range of DIM or DCC values and for various effect combinations not included in the model. In a first step, the residuals of the Meff × year and Meff models were compared using an ANOVA with class of 20 days of milk (cl(DIM)), class of 10 days of carried calf (d(DCC)), ymonth, ydry, yage, interaction between ymonth and cl(DIM), interaction between ydry and cl(DIM), interaction between yage and cl(DIM), interaction between yparity and cl(DCC) as the class variable. This enabled to assess the need to include interaction with year in the elementary curve definition. In a second step, residuals of the Meff, Meff–age, Meff–dry, Meff–month and Meff–gest models were analysed using cl(DIM), cl(DCC), ymonth, ydry, yage, interaction between month and cl(DIM), interaction between dry and cl(DIM), interaction between age and cl(DIM), interaction between yparity and cl(DIM) and interaction between parity and cl(DCC) as the class variable. This allowed assessing whether each elementary curve, i.e. for calving month, calving age, length of dry period and gestation effects was needed in the model.

In addition to the residual analyses, the models were compared on the basis of graphical analysis of the lactation curves obtained with the 'Jura test' data (2 406 345 test-day).

*A two-step approach for the solution of the mixed model equations.* To reduce computational requirements and computing time when fitting the extended model, 'Jura 2005' data were first pre-adjusted for effects related to the shape of the lactation curve with no year interaction (parity–class of calving month, parity–class of dry period length and parity–class of calving age with 6-knot regression splines each and gestation effect with 4-knot regression splines, i.e. 22 spline coefficients) running a model including only fixed effects as in the Meff model (1). In a second step, pre-adjusted records were analysed using a model including the HTD effect, three fixed effects with year interaction (ymonth, ydry and yage) and four additive genetic effects, four permanent environment effects and six herd–year of calving effects, i.e. 18 effects, instead of 40, in a one-step approach.

For practical implementation of the pre-adjustment, one element of interest was the stability of regression splines over years and with data files including more and more observations. If spline coefficients really do not depend on year, lactation curves should be consistent whether they are estimated from the 'Jura 2005' data set or a subset truncated 4 years earlier (TD data collected between September 1988 and December 2001), hereafter referred to as the 'Jura 2001' data set. In addition to correlations between spline coefficients, both the genetic and permanent environment values estimated from the 'Jura 2005' data file pre-adjusted for lactation curves estimated from the 'Jura 2001' data file on the one hand and the 'Jura 2005' data file on the other hand were compared (Figure 1). Moreover, it was important to know whether the data pre-adjustment for the lactation curve itself did not induce variation of the random effect solutions of the model. Therefore, genetic and permanent environment values



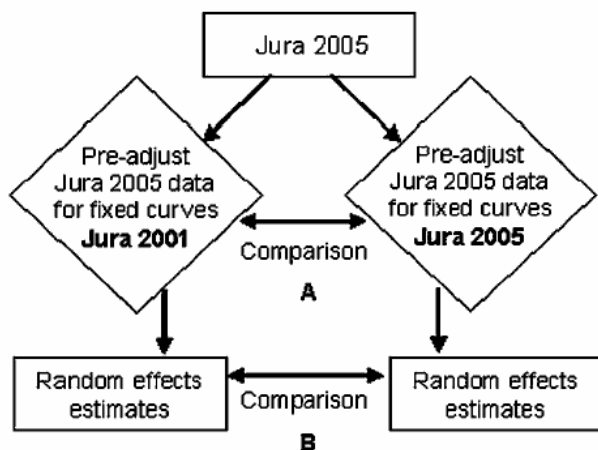


Figure 1 Approach implemented to check consistency of the shape of the lactation curves over years.

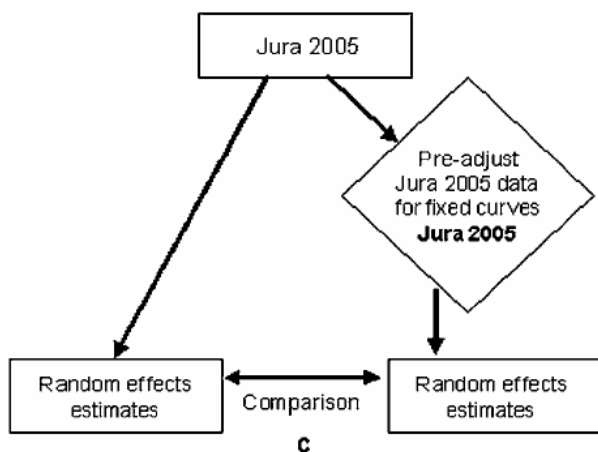


Figure 2 Approach implemented to assess the impact of a two-step v. a one-step genetic evaluation.

obtained for each 305-day lactation with a one-step procedure were compared with the ones obtained with the two-step procedure for the 'Jura 2005' data file (Figure 2).

## Results

### Comparison of six models to select environment fixed effects

The results of the comparison of the Meff × year and Meff models are given in detail in Table 2. The largest *F* values were for cl(DCC). The impact of gestation through DCC in the Meff × year model was not better modelled than with the Meff model, with *F* values of 2.15 and 2.68, respectively. In all models, the effect of gestation on milk production was only taken into account for DCC ≥ 100. In the GLM analysis, cl(DCC) analysed DCC by 10-day class, except for DCC < 100 they were merged into a single class. For cl(DIM), *F* values were also significant but limited in both models. None of the interactions including year had a significant *F* value, except the interaction between yparity and cl(DCC) with the Meff model. There was a small increase in *F* value between the Meff ×

Table 2 *F* statistics and *R*<sup>2</sup> of the GLM analysis of residual from two models (Meff × year: lactation curves (calving month, length of dry period, calving age and gestation) defined using splines including year interaction (Meff: lactation curves defined using splines without year interaction)

Effect <sup>†</sup>	d.f.	Meff × year	Meff
<i>R</i> <sup>2</sup>		0.0040	0.0060
cl(DIM)	48	1.50*	1.95***
cl(DCC)	17	2.15**	2.68***
ymonth	198	0.14	0.33
ydry	48	0.12	0.32
yage	122	0.08	0.27
ymonth × cl(DIM)	3168	0.55	0.85
ydry × cl(DIM)	768	0.50	0.75
yage × cl(DIM)	1952	0.51	0.78
yparity × cl(DCC)	323	0.76	1.30***

<sup>†</sup>Classes included in the analysis were the following: classes of 20 days of milk (cl(DIM)), classes of 10 days of carried calf (cl(DCC)), year–parity–calving month (ymonth), year–parity–length of dry period (ydry), year–parity–calving age (yage), interaction ymonth × cl(DIM), interaction ydry × cl(DIM), interaction yage × cl(DIM), interaction year–parity (yparity) × cl(DCC).

year and Meff models underlying that the definition of gestation effect in the Meff model (4-knot regression splines on DCC defined by parity) was slightly less efficient than the Meff × year model (4-knot regression splines on DCC defined by yparity). However, the large number of records involved in the analyses had the consequence of making some of the effects significant, although their influence in the overall variation may be very small. The *R*<sup>2</sup> obtained for the GLM analysis of the residuals of models Meff × year and Meff, with values of 0.0040 and 0.0060, respectively, showed the low impact of all effects included in the residual analyses.

The results obtained on the Meff model were used as the reference in the comparison of Meff, Meff–age, Meff–dry, Meff–month and Meff–gest models. The *R*<sup>2</sup> obtained for the GLM analysis of the residuals of the Meff, Meff–age, Meff–dry, Meff–month and Meff–gest models were 0.0016, 0.0024, 0.0023, 0.0049 and 0.0295, respectively. For all models, the impact of cl(DIM) and cl(DCC) were similar to the one described previously, i.e. significant (Table 3). The interaction yparity × cl(DIM) was significant but presented a small *F* value (i.e. 1.28 or 1.29) for all models. For all models, the deletion of the regression splines on the eff effects (month, dry and age) led to an increase of the *F* value for the interaction eff × cl(DIM) in comparison with the Meff model. The removal of the regression splines on gestation effect (Meff–gest) resulted in a jump of the *F* value from 1.22 to 25.26 for the interaction parity × cl(DCC), cl(DCC) soaring from 3.09 to 191.07. To sum up, all effects included as elementary curves in the model were significant, with in descending order of impact on TD production the gestation, calving month, length of dry period and calving age effects. Considering that the Meff model had 15 times fewer regression splines to estimate and that convergence was faster than the Meff × year model, the Meff model seemed to be a very good compromise.

**Table 3** F statistics and  $R^2$  of the GLM analysis of residual from five models (Meff: lactation curves (calving month, length of dry period, calving age and gestation) defined using splines including parity interaction; Meff-age: lactation curves (calving month, length of dry period and gestation); Meff-dry: lactation curves (calving month, calving age and gestation); Meff-month: lactation curves (length of dry period, calving age and gestation); Meff-gest: lactation curves (calving month, length of dry period and calving age))

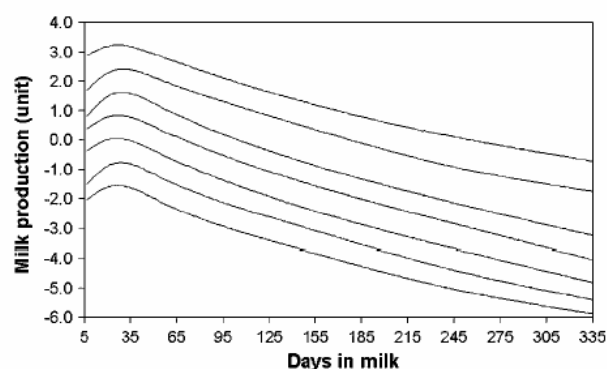
Effect <sup>†</sup>	d.f.	Meff	Meff-age	Meff-dry	Meff-month	Meff-gest
$R^2$		0.0016	0.0024	0.0023	0.0049	0.0295
cl(DIM)	48	2.00***	4.27***	6.78***	4.06***	73.82***
cl(DCC)	17	3.09***	3.25***	3.11***	3.52***	191.07***
y <sub>month</sub>	198	0.05	0.05	0.05	0.07	0.17
y <sub>dry</sub>	48	0.03	0.03	0.08	0.04	0.19
y <sub>age</sub>	122	0.03	0.04	0.04	0.03	0.15
y <sub>parity</sub> × cl(DIM)	288	1.29***	1.29***	1.29***	1.28**	1.28***
month × cl(DIM)	528	1.03	1.11*	1.12*	8.24***	1.74***
dry × cl(DIM)	128	1.02	1.05	6.97***	1.49***	1.47***
age × cl(DIM)	336	0.61	3.11***	0.60	1.46***	1.27***
parity × cl(DCC)	17	1.22	1.22	1.22	1.08	25.26***

<sup>†</sup>Classes included in the analysis were the following: classes of 20 days of milk (cl(DIM)), classes of 10 days of carried calf (cl(DCC)), year-parity-calving month (y<sub>month</sub>), year-parity-length of dry period (y<sub>dry</sub>), year-parity-calving age (y<sub>age</sub>), interaction year-parity (y<sub>parity</sub>) × cl(DIM), interaction parity-calving month (month) × cl(DIM), interaction parity-length of dry period (dry) × cl(DIM), interaction parity-calving age (age) × cl(DIM), interaction parity × cl(DCC).

*Comparison of lactation curves obtained from the Meff × year and Meff models*

The 541, 357 and 165 curves on y<sub>month</sub>, y<sub>age</sub> and y<sub>dry</sub> effects, respectively, and the 46 y<sub>parity</sub> curves on gestation effect of the Meff × year model were plotted by year, by parity and by effect (class of calving month, class of calving age, class of dry period length, respectively). Curves showing parity × class of effect according to year effect all presented similar profiles. For instance, curves obtained for the period 1990/1991 to 2002/2003 on the second lactation Montbéliarde cows calving between 43 and 45 months ('Jura test' data file) on TD milk production are displayed in Figure 3. For clarity of presentation, only curves obtained every 2 years were plotted, adding an arbitrary constant of one unit around peak yield days to each one (at DIM = 32 days). No large interaction between year and parity × class of calving age was noticed. Only small changes in shape at the beginning of the lactation and a slower decrease for curves of recent years (the top ones) were observed. Similar results were obtained for other effects: y<sub>month</sub>, y<sub>dry</sub> and gestation (results not shown). When curves estimated with the Meff × year model are regressed on a mean curve for each effect plus a constant year effect, the  $R^2$  of the regression was 0.984 for the calving month effect, 0.993 for the calving age effect, 0.994 for the length of dry period effect and 0.995 for the gestation effect. These results confirmed that the inclusion of year interaction in the spline definition was not necessary.

The impact of gestation on milk production modelled with 4-knot regression splines beyond 100 days of carried calf was plotted for the three parities (Figure 4). With a value around zero at 100 days, milk production in first lactation decreased more slowly than in the second and third lactations. The average loss of production at DCC 230 (i.e. a common time point when drying-off occurs 2 months before the next calving) in comparison with a non-pregnant cow was 3.2 kg for a first-lactation cow and 4.1 kg for a third-lactation cow (4.7 and



**Figure 3** Effects of years 1990/1991 to 2002/2003 for second lactation Montbéliarde cows calving between 43 and 45 months on test-day milk production modelled with 6-knot splines (only alternate yearly curves separated by a value of 1 at DIM 32 are shown).

5.7 kg at DCC 265, respectively). Curves for calving month classes showed large variations across month and parity. They were plotted for second lactation in Figure 5, without making any adjustment for HTD effect. The largest peak production on second parity was observed for cows calving in November, whereas the lowest was for cows calving in July. The analysis of the length of dry period class effect (Figure 6) showed that the first-lactation cows class (that could be used as reference: no effect of dry period length) had an intermediate impact on milk production. The 46 to 60 days and 61 to 75 days classes for both second and third lactation were not drawn in Figure 6. All corresponding curves would remain between the third lactation 26 to 45 days class and the third lactation 76 to 100 days class, with the curves not showing any specific impact of those classes of dry period length. Only a dry period ≤ 25 days for second and third lactations largely penalised milk production. Around peak production, cows dried during 25 days or less (i.e. 2% of the data) had a milk production loss between 2.5 and 3.5 kg in comparison with other classes of dry period length. Regarding the calving age effect (Figure 7), for first

Leclerc, Duclos, Barbat, Druet and Ducrocq

lactation Montbéliarde cows, calving between 30 and 39 months and more did not seem to have a large impact on the curve shape since all these curves overlapped each other.

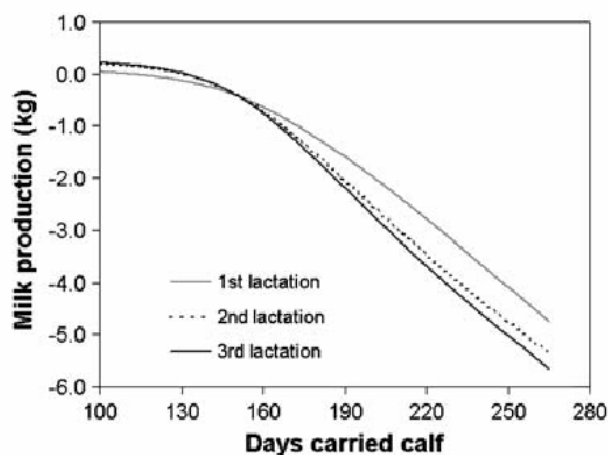


Figure 4 Days of carried calf effects on Montbéliarde test-day milk production modelled with 4-knot splines for first- to third-lactation cows.

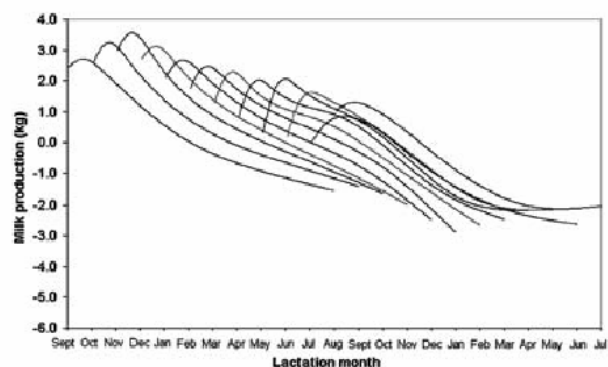


Figure 5 Month of calving effects on Montbéliarde test-day milk production modelled with 6-knot splines for second-lactation cows.

Similar characteristics were observed for second and third lactations. However, for first-lactation cows calving before 28 months of age, milk production was penalised, especially for cows calving before 24 months.

*Impact of spline coefficients estimated at different dates*

Two series of estimated regression spline coefficients were calculated: the first series with the 'Jura 2001' data file and the second series with the 'Jura 2005' data file. The correlations between the spline coefficients estimated with these two data sets were not as high as expected. For milk yield, the mean value was 0.985. When two corresponding splines were plotted, the shapes were not exactly the same, but it was difficult to determine if the differences were really significant over the whole lactation. However, the consistency of data pre-adjustment over years was verified through the stability of random estimates. The solutions for genetic and permanent environmental effects running the second step of the model on 'Jura 2005' data after pre-adjustment for lactation and gestation splines estimated from 'Jura 2001' and 'Jura 2005' data files were compared. Correlations for each of the four associated coefficients were between 0.99935 and 0.99987 for genetic effect and between 0.99941 and 0.99985 for the permanent environment effect (results not shown). When the solutions were brought back to a lactation value (one to three), all the correlations were between 0.99991 and 0.99997 for genetic effect and between 0.99969 and 0.99994 for the permanent environment effect. In other words, genetic and permanent environment values were not modified when the data were pre-adjusted with spline coefficients calculated 4 years before.

*Impact of two-step procedure v. one-step procedure*

Very high correlations were obtained for both genetic and permanent environment effects. When the solutions were

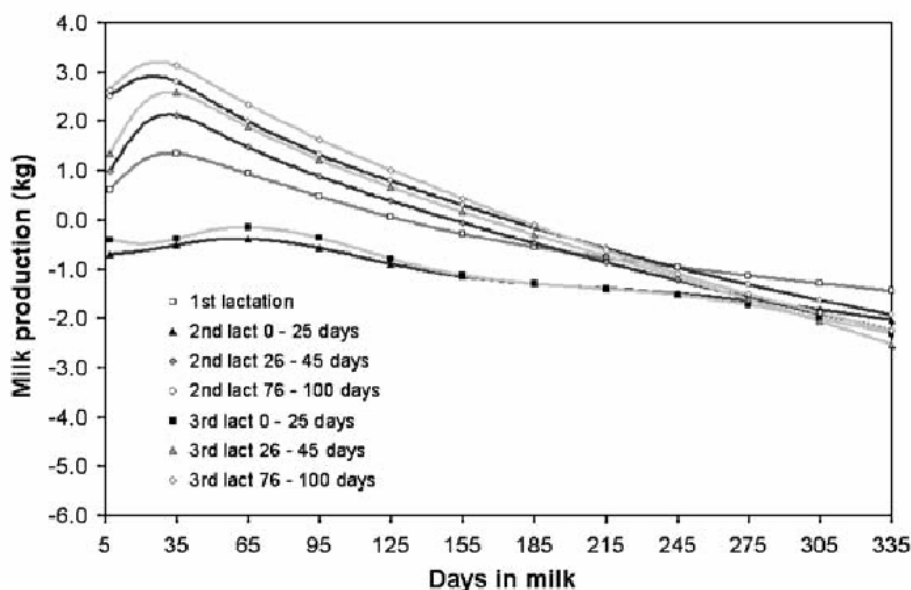


Figure 6 Effects of length of dry period on Montbéliarde test-day milk production modelled with 6-knot splines.



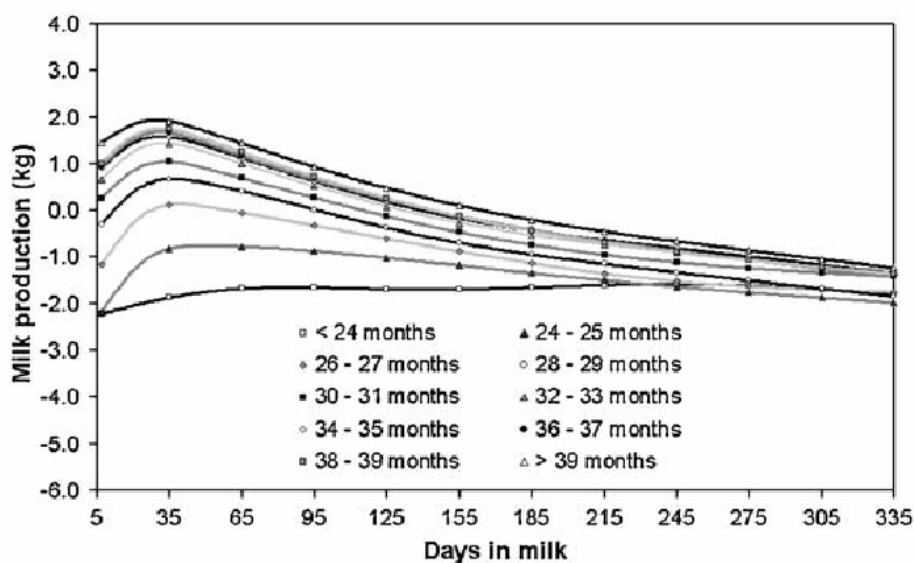


Figure 7 Calving age effects on Montbéliarde test-day milk production modelled with 6-knot splines for first-lactation cows.

expressed as lactation values (1 to 3), the correlations were between 0.9984 and 0.9992 for the genetic effect and between 0.9981 and 0.9991 for the permanent environment effect. Furthermore, similar correlations were obtained for milk trait analysed with the complete Montbéliarde data file (i.e. 21 128 460 TD) with a minimum value of 0.9996 for both genetic and permanent environment effects. Computing time was reduced by a factor of four when the two-step procedure was used, compared with the one-step procedure.

## Discussion

As reported by many authors (Wood, 1969; Danell, 1982; Wilmlink, 1987b; Pérochon *et al.*, 1996), this study showed that daily yield was strongly affected by environmental factors such as region, year, parity, month of calving, age at calving, length of dry period before calving and pregnancy. In the 1980s, the objectives of most studies were to develop additive or multiplicative adjustment factors for lactation yield. Most models did not take into account the changes in the shape of the lactation curves due to environmental factors, except in the case of the Wilmlink curve (Wilmlink, 1987b). Nowadays, with TD model genetic evaluation, these changes on the shape of the lactation curves can be modelled more easily. Some of the studied factors, like parity, month of calving, age at calving, length of dry period and pregnancy showed a large effect on the shape of the lactation curve, whereas others like interactions including year effect affected lactation curve only through a constant over the whole lactation. Parity had a large effect on variation in both the scale and shape of the lactation curve. Daily milk yield increased until about 25 to 50 days *post partum* depending on the environmental factors. Cows with unfavourable environmental factors had a delayed peak in their lactation, e.g. first-parity cows *v.* later parities, cows calving in the spring and summer *v.* cows calving in the autumn and winter, cows calving younger within parity *v.*

older ones. For instance, cows calving at 24 or 25 months had a peak production around 48 days after calving whereas cows calving at 34 or 35 months had peak production around 32 days for a first lactation. As mentioned by Wilmlink (1987b), milk yield differences between classes were found to be the largest at the beginning of lactation and were almost zero at the end of lactation. For instance, differences for second-parity cows between a length of dry period  $\leq 25$  days and a length of dry period of 76 to 99 days were 3.3 kg around yield peak and 0.8 kg in the last-third of lactation (at DIM 215).

## Pregnancy effects

Pregnancy had a significant effect on daily milk yields (Table 3). In our model, the penalising effect of pregnancy on milk yield was assumed to begin 100 days after conception. However, elementary curves representing the effect of gestation on milk production showed that in fact the negative effect of pregnancy began at DCC 112, 131 and 132 for first-, second- and third-lactation cows, respectively. The results observed were in agreement with the stage of 100 to 150 days after conception reported by many authors (Oltenucu *et al.*, 1980; Danell, 1982; Coulon *et al.*, 1995). Similar to many other authors (Louca and Legates, 1968; Oltenucu *et al.*, 1980; Coulon *et al.*, 1995), we also observed that the effect of pregnancy was lower in primiparous than in multiparous cows. The milk yield drop-off observed at DCC 230 (Figure 4) was consistent with the 3.0 to 5.0 kg/day reported by Danell (1982) for Swedish Red and White and Swedish Friesian. On a TD model considering only fixed effects, Bohmanova *et al.* (2006) observed a drop-off of 4.2 kg *v.* 3.5 kg for a first-lactation cow and 5.2 kg *v.* 4.1 kg for a third-lactation cow in the Holstein and Ayrshire breeds, respectively. The reduction of milk yield during pregnancy has been attributed to a modification in the hormonal balance linked to foetal growth. Changes in placental steroid hormones, increased oestrogens in

Leclerc, Duclos, Barbat, Druet and Ducrocq

particular, from the second-third of pregnancy were considered to be responsible (Bauman and Currie, 1980; Oltenacu *et al.*, 1980). Foetus growth might have an impact on rumen volume and distribution of nutrients between the foetus, milk synthesis and replenishment of energy reserves for the next lactation (Oltenacu *et al.*, 1980). However, foetus growth essentially occurred during the last 2 months of pregnancy (Bauman and Currie, 1980), when cows were usually dried off. Taking the effect of pregnancy on milk production into account requires the knowledge of conception date. Two limits can be mentioned in the establishment of pregnancy effect. First, when the date of next calving was available, we assumed the usual gestation length for the Montbéliarde breed. In case of abortion, gestation effect was not properly taken into account. Secondly, when the date of next calving was not (yet) available, we assumed that the last artificial insemination led to a conception, which was not always the case in reality. However, an effect of pregnancy starting at DCC 100 limits the risk to attribute this effect to a non-pregnant cow, except if this cow was intentionally not re-inseminated.

#### *Calving month effect*

Many authors mentioned calving month as an effect with considerable influence on the cow's milk production (Wood, 1969; Danell, 1982; Wilmink, 1987b; Ptak *et al.*, 1993). In this study, the effect of calving month was found to be significant on daily milk yield (Table 3). Time of peak yield changed with calving month. It occurred earlier in autumn and winter calvings (about 25 to 30 days for multiparous cows) and later in spring and summer calvings (about 30 to 45 days for multiparous cows). Cows calving in the autumn and winter had a higher 305-day production than cows calving in the spring and summer. Similar features were reported by Danell (1982), Wilmink (1987b) and Ptak *et al.* (1993). Cows calving in November produced 449, 475 and 456 kg more in the first, second and third lactations, respectively, in comparison with cows calving in July. Danell (1982) mentioned that the yield differences between the months with the highest *v.* lowest production were about 350 kg milk, i.e. 100 kg less than what was found in this study. Month of calving did not affect yield in the same way at the beginning and at the end of lactation. The shape of the lactation curve was found to be largely dependent upon month of calving (Figure 5) as reported by Danell (1982). Spring-calving cows produced most in early lactation taking advantage of better/higher fodder supply and quality, whereas autumn-calving cows produced more in late lactation. In this study, the interaction between calving month and year were assessed to be non-significant for the shape of the lactation curves based on residuals and lactation curve analysis. However, Danell (1982) and Ptak *et al.* (1993) mentioned the existence of an interaction between month of calving and year due to changes in herd management or in weather conditions, but stated that the general pattern seemed to be the same overall. In the Meff model, which will be used in the French genetic evaluation,

the interaction between calving month and year was included through a constant over years.

#### *Length of dry period effect*

As reported by Schaeffer and Henderson (1972) and Pérochon *et al.* (1996), the results obtained in this study showed a significant effect of length of dry period on lactation curve (Table 3). As we observed, the first classes showed that a dry period of  $\leq 1$  month had a negative effect on the subsequent lactation. For them, cows with a dry period of 50 to 59 days gave the highest average production, without great difference between classes 40 to 49 and 60 to 69 days, but milk production following a very long dry period (more than 120 days) was lower. In our study, the second and third lactations presented similar lactation curves, in contrast with Schaeffer and Henderson (1972) who found that parities later than the second one are less affected by length of dry period. Based on the large difference observed between curves corresponding to classes 0 to 25 and 26 to 45 days of dry period for the second and the third lactations (Figure 6), the class definition used for this study did not appear appealing. For the routine evaluation, new definition of traits with six classes instead of five will be used, i.e. 0 to 15, 16 to 30, 31 to 45, 46 to 60, 61 to 80 and  $\geq 81$  days. Schaeffer and Henderson (1972) mentioned that length of dry period influenced the survival of cows in the following lactation. Under normal management conditions, the dairy cows are overfed during the last-third of lactation to replenish body energy stores in preparation for the next lactation. One possible explanation of a lower production level after a short dry period was the competition for nutrients between the foetus, the current lactation and the cow replenishment. During the 1st month of lactation, the body reserves being utilised are energetically equivalent to about 23% of the milk produced (Bauman and Currie, 1980). Thus, without a sufficiently long dry period, cows might have a lower production in the next lactation or a lower survival rate. Another explanation could be the effect of a premature calving or abortion.

#### *Calving age effect*

A large number of authors such as Danell (1982), Wilmink (1987b) and Ptak *et al.* (1993) mentioned age at calving as a factor affecting lactation milk yield, and dependent on lactation stage. In this study, the model comparison step (Table 3) showed a significant effect of calving age on the lactation curve. Daily milk yield increased with increasing age at calving. Danell (1982) reported that a 1-month increase in age at first calving raised milk yield during lactation by about 50 kg. This corresponds to a higher value than the 590 kg milk yield difference between extreme classes of first-parity cows found here. However, differences between age classes decreased with parity, with 290 kg in second parity and 170 kg in third parity. As mentioned previously, the effect was the largest for TDs early in lactation e.g. 3.8 kg around peak production against 1.1 kg in the last-third of lactation (at DIM 215) between extreme classes of first parity (Figure 7).

### Comparison with national fixed effects model from other countries

National genetic evaluation models for dairy traits are nowadays more and more often based on TD models instead of 305-day lactation models as in Canada, Denmark–Finland, Germany–Austria, Italy and The Netherlands (Interbull, 2007). In these countries, lactation curves were mainly modelled using Legendre polynomials of order four, Wilmink function or a mix of both, except for The Netherlands which included only constants over the whole lactation and modelled lactation curves as an interaction between parity and DIM. For most countries, effects of age at calving and/or calving season were included in the lactation curve modelling. However, few of these countries considered the impact of pregnancy on TD production. The Netherlands and Finland (Lidauer *et al.*, 2003) included a constant defined by class of pregnancy stage, with an interaction with parity in the case of The Netherlands. In view of the large pregnancy effect affecting cows during the last-third of their lactation highlighted in this study, not taking this effect into account in the genetic evaluation model highly penalised cows with a short DO period. Recently, Canavesi *et al.* (2007) studied the pre-adjustment of TD records for the pregnancy stage. Some countries, like Canada, are studying the effect of pregnancy on the TD model (Bohmanova *et al.*, 2006). The length of dry period was included only in The Netherlands and Denmark–Finland models. As for the pregnancy effects, cows with short dry period were found to be penalised.

### Conclusion

In the Montbéliarde breed, the shape of lactation curves has not changed over the years. The changes of production curve shape over the years could only be related to changes of random effects, i.e. genetic, permanent environment and/or herd effects. This conclusion has a positive impact on the computation strategy to solve the mixed model equations. It was possible to pre-adjust TD records for time-independent fixed effects, i.e. lactation curves, in a first step. Random effects and year-dependent fixed effects were estimated in the second step. Our results validated this strategy since genetic and permanent environment effects obtained from a two-step procedure were virtually identical to those from a one-step procedure. Two main advantages can be put forward from this approach: it required less computing time (here reduced by a factor of four) and less memory (22 and 18 unknowns for the first and second steps, respectively, instead of 40 unknowns in a one-step procedure). Moreover, the first step (computation of the spline coefficients for fixed effects) does not necessarily have to be performed at each evaluation and could be done at broader intervals, for example once a year.

### Acknowledgements

The research was partly funded by a grant from the French Ministry of Agriculture (Action Innovante UTILEG). This study was realised within the framework of the Unité Mixte Technologique (Joint Technology Unit) INRA – Institut de l'Élevage dealing with cattle genetic evaluation.

### References

- Bauman DE and Currie WB 1980. Partitioning of nutrients during pregnancy and lactation: a review of mechanisms involving homeostasis and homeorhesis. *Journal of Dairy Science* 63, 1514–1529.
- Bohmanova J, Miglior F, Kelly M, Kistemaker G and Loker S 2006. Effect of pregnancy on milk yield of Canadian dairy cattle. Retrieved September 17, 2007, from [http://www.aps.uoguelph.ca/~jbohmano/publication/report\\_08.pdf](http://www.aps.uoguelph.ca/~jbohmano/publication/report_08.pdf)
- Canavesi F, Biffani S and Nicolazzi EL 2007. Three versus four traits random regression test-day model genetic evaluation for the Holstein breed in Italy. In Proceedings of the 58th Annual Meeting of the European Association for Animal Production, Dublin, 26–29 August, p. 64 (abstract).
- Coulon JB, Pérochon L and Lescouret F 1995. Modelling the effect of the stage of pregnancy on dairy cow's milk yield. *Animal Science* 60, 401–408.
- Danell B 1982. Studies on lactation yield and individual test-day yields of Swedish dairy cows. I. Environmental influence and development of adjustment factors. *Acta Agriculturae Scandinavica* 32, 65–81.
- Druet T, Jaffrézic F, Boichard D and 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* 86, 2480–2490.
- Druet T, Jaffrézic F and Ducrocq V 2005. Estimation of genetic parameters for test day records of dairy traits in the first three lactations. *Genetics Selection Evolution* 37, 257–271.
- Institut de l'Élevage and France Contrôle Laitier 2005. Le règlement technique du contrôle laitier zootechnique des espèces bovine et caprine et annexes. Retrieved April 17, 2007, from [http://www.inst-elevage.asso.fr/html1/rubrique.php?id\\_rubrique=648](http://www.inst-elevage.asso.fr/html1/rubrique.php?id_rubrique=648)
- Interbull 2007. Description of National Genetic Evaluation Systems for dairy cattle traits as applied in different Interbull member countries. Retrieved April 17, 2007, from [http://www-interbull.slu.se/national\\_ges\\_info2/framesida-ges.htm](http://www-interbull.slu.se/national_ges_info2/framesida-ges.htm)
- International Committee for Animal Recording 2005. International agreement of recording practices – guidelines approved by the general assembly held in Sousse, Tunisia, June 2004. ICAR, Rome, Italy.
- Jensen J 2001. Genetic evaluation of dairy cattle using test-day models. *Journal of Dairy Science* 84, 2803–2812.
- Lidauer M, Mäntysaari EA and Strandén I 2003. Comparison of test-day models for genetic evaluation of production traits in dairy cattle. *Livestock Production Science* 79, 73–86.
- Louca A and Legates JE 1968. Production losses in dairy cattle due to days open. *Journal of Dairy Science* 51, 573–583.
- Mao IL, Wilton JW and Freeman MG 1972. The effect of parity on age adjustment for milk and fat yield. *Journal of Dairy Science* 55 (Suppl. 1), 716.
- Mayeres P, Stoll J, Bormann J, Reents R and Gengler N 2004. Prediction of daily milk, fat and protein production by a random regression test-day model. *Journal of Dairy Science* 87, 1925–1933.
- Oltenu PA, Rounsaville TR, Milligan RA and Hintz RL 1980. Relationship between days open and cumulative milk yield at various intervals from parturition from high and low producing cows. *Journal of Dairy Science* 63, 1317–1327.
- Pérochon L, Coulon JB and Lescouret F 1996. Modelling lactation curves of dairy cows with emphasis on individual variability. *Animal Science* 63, 189–200.
- Ptak E, Horst HS and Schaeffer LR 1993. Interaction of age and month of calving with year of calving for production traits of Ontario Holsteins. *Journal of Dairy Science* 76, 3792–3798.
- Schaeffer LR and Henderson CR 1972. Effects of days dry and days open on Holstein milk production. *Journal of Dairy Science* 55, 107–112.
- Swalve HH 2000. Theoretical basis and computational methods for different test-day genetic evaluation methods. *Journal of Dairy Science* 83, 1115–1124.
- White IMS, Thompson R and Brotherstone S 1999. Genetic and environmental smoothing of lactation curves with cubic splines. *Journal of Dairy Science* 82, 632–638.
- Wilmink JBM 1987a. Adjustment of lactation yield for age at calving in relation to level of production. *Livestock Production Science* 16, 321–334.
- Wilmink JBM 1987b. Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. *Livestock Production Science* 16, 335–348.
- Wood PDP 1969. Factors affecting the shape of the lactation curve in cattle. *Animal Production* 11, 307–316.

## RESULTATS COMPLEMENTAIRES

### Modélisation de la numération cellulaire

Les analyses graphiques présentées dans l'article précédent ont toutes été réalisées sur la production laitière. Pour les matières grasse et protéique ainsi que les taux correspondants, les courbes obtenues pour le modèle *Meff* étaient cohérentes avec nos attentes. (NB : Les taux ont fait l'objet de cette analyse préalable même si ces derniers ne seront pas, *a priori*, évalués directement comme c'était le cas dans le modèle lactation mais « ré-estimés » à partir des quantités de matières comme c'est le cas dans les autres pays.) Cependant, pour l'analyse du Score de Cellules Somatiques, la situation était différente. Les courbes prenant en compte l'effet de la gestation présentaient des variations non interprétables comme l'illustre la figure 3.4 (courbes en forme de S couché).

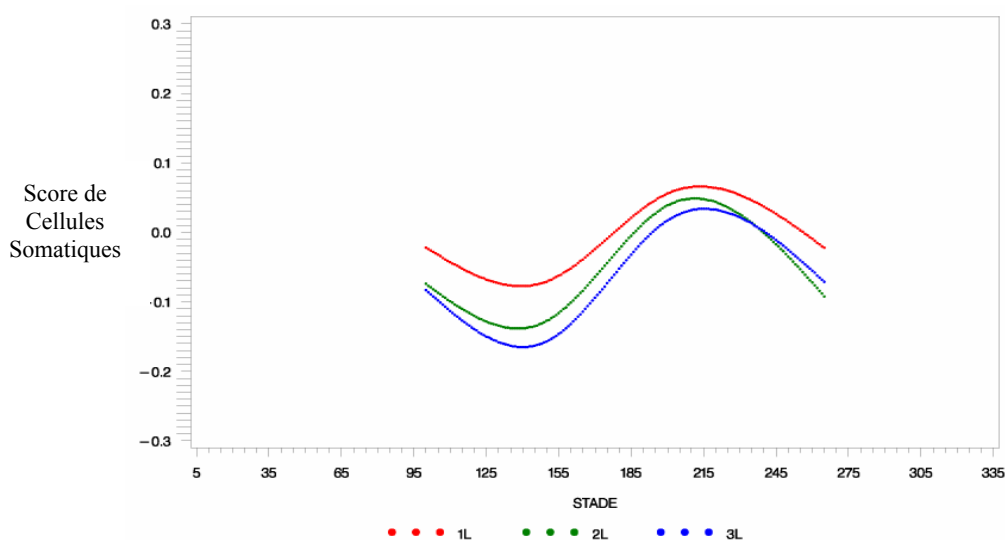


Figure 3.4. Effet du rang de lactation sur le Score de Cellules Somatiques en fonction du stade de gestation (nombre de jours depuis la date supposée de fécondation) en race Montbéliarde (données du Jura) modélisé avec une spline de 4 nœuds à partir du modèle *Meff*.

Des analyses spécifiques ont donc été menées sur le fichier constitué des données du Jura pour la race Montbéliarde, afin de déterminer la modélisation adéquate pour ce caractère. A partir du modèle *Meff*, 6 modèles simplifiés ont été élaborés et comparés. En plus des modèles *Meff-age*, *Meff-dry*, *Meff-month*, *Meff-gest* déjà utilisé pour les caractères de production, nous avons introduit dans les comparaisons un modèle *Meff-agedry* incluant seulement deux splines : l'une sur le stade de lactation définie par l'interaction entre le mois de vêlage et le rang de lactation et l'autre sur le stade de gestation définie uniquement par rang de lactation et enfin un modèle extrême *Meffno* n'incluant pas de courbe pour modéliser la trajectoire phénotypique de la numération cellulaire.

Comme pour l'analyse sur la production laitière, la statistique F de l'analyse GLM des résiduelles pour l'interaction  $yparity \times cl(DIM)$  est significative quel que soit le modèle (Tableau 3.1). En se basant sur les conclusions de l'article précédent qui ne montraient pas d'évolution de la forme des courbes des effets âge, mois de vêlage, durée tarissement et gestation au cours du temps, l'effet de l'année n'a donc pas été pris en compte comme facteur d'interaction des

courbes influençant la trajectoire de SCS au cours de la lactation. La comparaison des différents modèles (Tableau 3.1) permet de voir que la suppression des courbes modélisant les effets de l'âge au vêlage et de la durée de tarissement respectivement *Meff-age*, *Meff-dry* n'a pas d'impact sur la qualité d'ajustement du modèle. Par contre, la suppression des courbes modélisant les effets du mois de vêlage et de la gestation *Meff-month*, *Meff-gest* entraînent des valeurs de statistiques F significatives de 2.28 et 3.35 pour l'interaction entre month x cl(DIM) et parity x cl(DCC) respectivement.

Tableau 3.1 : Statistiques F et R<sup>2</sup> de l'analyse GLM des résidus de 7 modèles pour la numération cellulaire (*Meff* : courbes de lactation sur l'âge au vêlage, le mois de vêlage, la durée de tarissement et la gestation ; *Meff-age* : semblable à *Meff* mais sans la courbe de lactation sur l'âge au vêlage ; *Meff-dry* : semblable à *Meff* mais sans la courbe de lactation sur la durée de tarissement ; *Meff-month* : semblable à *Meff* mais sans la courbe de lactation sur le mois de vêlage ; *Meff-gest* : semblable à *Meff* mais sans la courbe sur l'effet de la gestation ; *Meff-agedry* : semblable à *Meff* mais sans la courbe sur l'effet de l'âge ni la durée de tarissement ; *Meffno* : sans aucune des courbes)

Effect <sup>1</sup>	d.l.	<i>Meff</i>	<i>Meff-age</i>	<i>Meff-dry</i>	<i>Meff-month</i>	<i>Meff-gest</i>	<i>Meff-agedry</i>	<i>Meffno</i>
<b>R<sup>2</sup></b>		0.0010	0.0010	0.0010	0.0017	0.0016	0.0011	0.0161
<b>cl(DIM)</b>	48	0.81	0.97	0.97	1.16	1.79***	1.07	38.30***
<b>cl(DCC)</b>	17	1.09	0.74	1.08	1.13	2.40**	0.75	2.73***
<b>ymonth</b>	198	0.03	0.03	0.02	0.05	0.03	0.03	0.38
<b>ydry</b>	48	0.02	0.02	0.03	0.02	0.02	0.03	0.04
<b>yage</b>	121	0.02	0.02	0.03	0.02	0.02	0.02	0.03
<b>yparity x cl(DIM)</b>	288	1.37***	1.37***	1.37***	1.38***	1.37***	1.37***	1.40***
<b>month x cl(DIM)</b>	528	0.59	0.59	0.59	2.28***	0.61	0.59	2.35***
<b>dry x cl(DIM)</b>	128	0.69	0.70	1.11	0.75	0.69	1.11	1.23*
<b>age x cl(DIM)</b>	336	0.57	0.98	0.57	0.76	0.58	0.98	1.03
<b>parity x cl(DCC)</b>	17	0.63	0.51	0.63	0.70	3.35***	0.51	2.98***

<sup>1</sup> Les classes incluses dans l'analyse sont définies de la façon suivante: classes de 20 jours pour le stade de lactation (cl(DIM)), classes de 10 jours pour le stade de gestation (cl(DCC)), année-rang de lactation-mois de vêlage (ymonth), année-rang de lactation-durée de tarissement (ydry), année-rang de lactation-âge au vêlage (yage), interaction année-rang de lactation (yparity) x cl(DIM), interaction rang de lactation-mois de vêlage (month) x cl(DIM), interaction rang de lactation-durée de tarissement (dry) x cl(DIM), interaction rang de lactation-âge au vêlage (age) x cl(DIM), interaction rang de lactation x cl(DCC).

Le modèle *Meff-agedry* inclut seulement les deux effets ayant un impact sur la qualité d'ajustement et donne des résultats globalement similaires à *Meff* avec un nombre de paramètres réduits (R<sup>2</sup> respectif de 0.0010 et 0.0011). Il est même légèrement meilleur sur cl(DCC) où la statistique F passe de 1.09 à 0.75. Cette amélioration, même si elle paraît faible d'après les statistiques de test, est pourtant importante comme en témoigne la trajectoire des courbes de l'effet de la gestation (Figure 3.5) qui nous paraît plausible d'après les connaissances biologiques. Nous observons une augmentation du score cellulaire du fait d'une baisse de la production et donc d'une moindre dilution lors du dernier tiers de la gestation. Le modèle extrême *Meffno* confirme, par des statistiques F hautement significatives sur cl(DIM) et cl(DCC) l'importance de modéliser l'effet des stades de lactation et de gestation quel que soit le caractère. Les statistiques F significatives pour l'effet yparity x cl(DIM) sont très homogènes entre les modèles. Comme pour le lait étudié précédemment, le grand nombre de classes et de données utilisées pour l'analyse explique que certains effets soient significatifs.

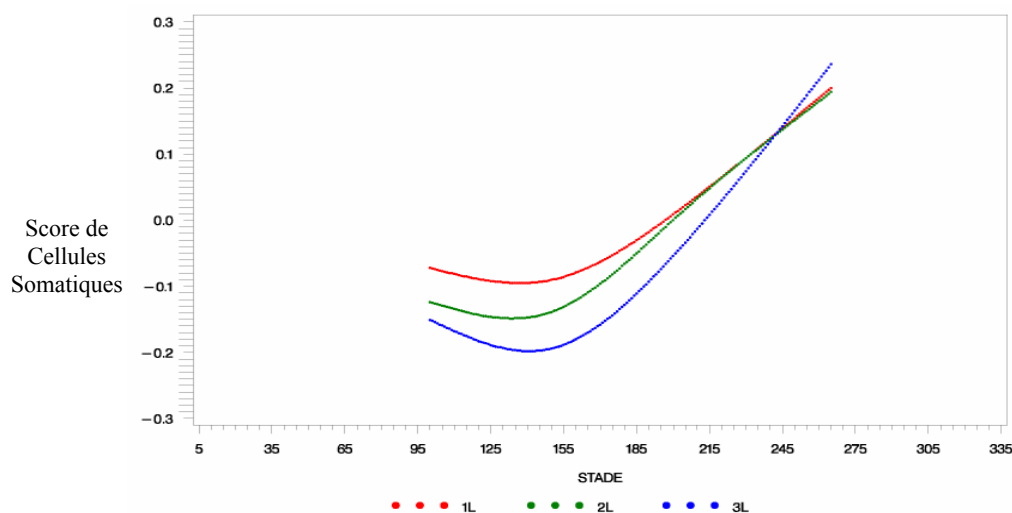


Figure 3.5. Effet du rang de lactation sur le Score de Cellules Somatiques en fonction du stade de gestation en race Montbéliarde (données du Jura) modélisé avec une spline de 4 nœud à partir du modèle *Meffmthdcc*.

### Définition des classes de durée de tarissement

En réponse à un relecteur de l'article précédent, nous nous sommes interrogés sur la qualité de la définition des classes sur laquelle se base l'effet de la durée de tarissement. Initialement, notre étude sur le modèle contrôles élémentaires avait été réalisée à partir de classes définies de façon similaire au modèle lactation. A l'exception des vaches en 1<sup>ère</sup> lactation où il n'y a pas d'effet de tarissement précédent et qui constituent donc une classe à part, 5 classes étaient formées pour chacune des lactations suivantes : une 1<sup>ère</sup> classe entre 0 et 25 jours, une 2<sup>nde</sup> entre 26 et 45 jours, une 3<sup>ème</sup> entre 46 et 60 jours, une 4<sup>ème</sup> entre 61 et 75 jours et enfin une 5<sup>ème</sup> classe entre 76 et 99 jours. Les lactations précédées d'un tarissement de 100 jours ou plus étaient regroupées avec la classe ayant le plus fort effectif, c'est-à-dire la 4<sup>ème</sup>. Les courbes ainsi obtenues montraient de fortes variations de la trajectoire de la lactation entre les classes, surtout pour les durées inférieures à 25 jours dont la courbe était quasi plate avec de légères vagues, alors que les courbes des classes suivantes se superposaient voire étaient confondues au-delà de 61 jours de tarissement (Figure 3.6).

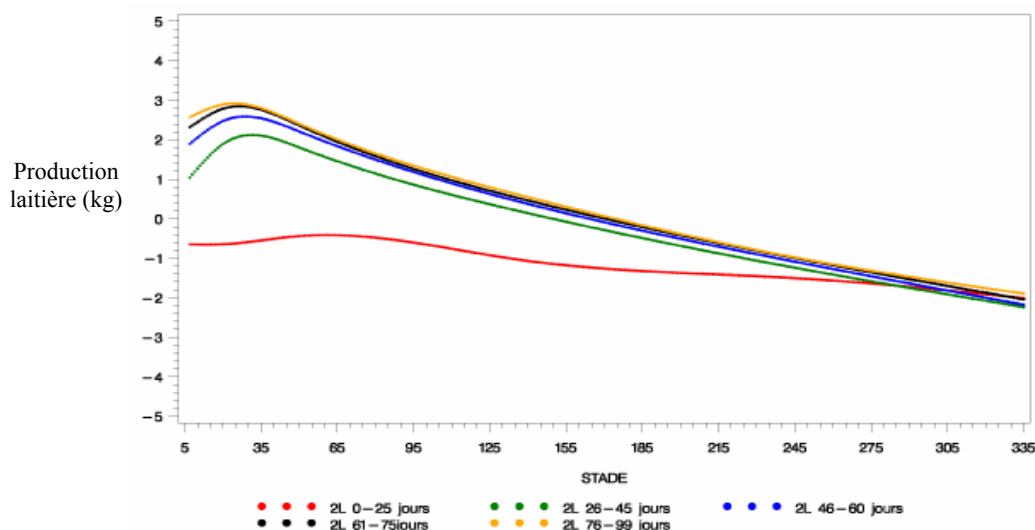


Figure 3.6. Effet de 5 classes de durée de tarissement en 2<sup>ème</sup> lactation sur la production laitière (kg) en fonction du stade de lactation en race Montbéliarde (données du Jura).

La démarche a consisté à redéfinir les classes d'une amplitude plus faible. Dans une première étape, 9 classes ont été formées (0-15 jours ; 16-25 jours ; 26-35 jours ; 36-45 jours ; 46-55 jours ; 56-65 jours ; 66-75 jours ; 76-90 jours et 91 jours et plus). En comparaison à la figure précédente, la figure 3.7 permet de mieux distinguer les effets des tarissements courts grâce à la scission de la classe 0-25 en 2 classes. La classe 0-15 jours montre un effet défavorable de 1.7 kg sur les  $\frac{2}{3}$  de la lactation certainement lié à la forte proportion de vaches ayant avortées dans cette classe. La classe 16-25 jours présente une forme de courbe intermédiaire entre la 1<sup>ère</sup> et la 3<sup>ème</sup> classe avec de légères vagues certainement dues aux faibles effectifs de cette classe. Au-delà de 35 jours de tarissement, le différentiel de production entre classes s'amointrit. La dernière classe : 91 jours et plus se retrouve au milieu du « paquet » certainement à cause de vaches ayant une durée de tarissement calculée très longue (plus d'une année) cachant probablement une lactation non enregistrée entre les deux recensées et donc un tarissement d'une durée inférieure à celle calculée.

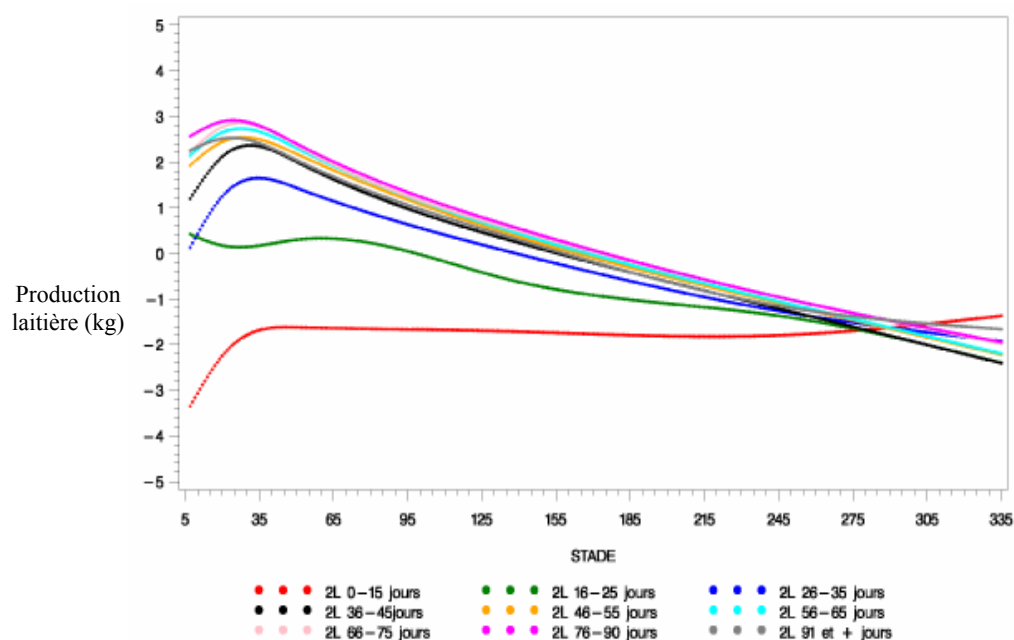


Figure 3.7. Effet de 9 classes de durée de tarissement en 2<sup>ème</sup> lactation sur la production laitière (kg) en fonction du stade de lactation en race Montbéliarde (données du Jura).

Pour éviter une sur-paramétrisation et ainsi assurer une estimation plus robuste des courbes, une deuxième étape a consisté à définir des classes d'amplitude intermédiaire par rapport aux deux précédents cas tout en cherchant à amoindrir les inconvénients précédemment cités. Six classes ont ainsi été constituées: 0-15 jours ; 16-30 jours ; 31-45 jours ; 46-60 jours ; 61-80 jours et 81 jours et plus. Les 4 premières classes ont une amplitude de 15 jours, ce qui permet de bien identifier les effets des tarissements inférieurs à 30 jours. Au-delà de 60 jours, les différences entre courbes étant faibles, seules deux classes ont été formées, ce qui permet de limiter l'impact des tarissements soi-disant très longs. Les courbes ainsi obtenues (Figure 3.8) présentent une trajectoire vraisemblable.



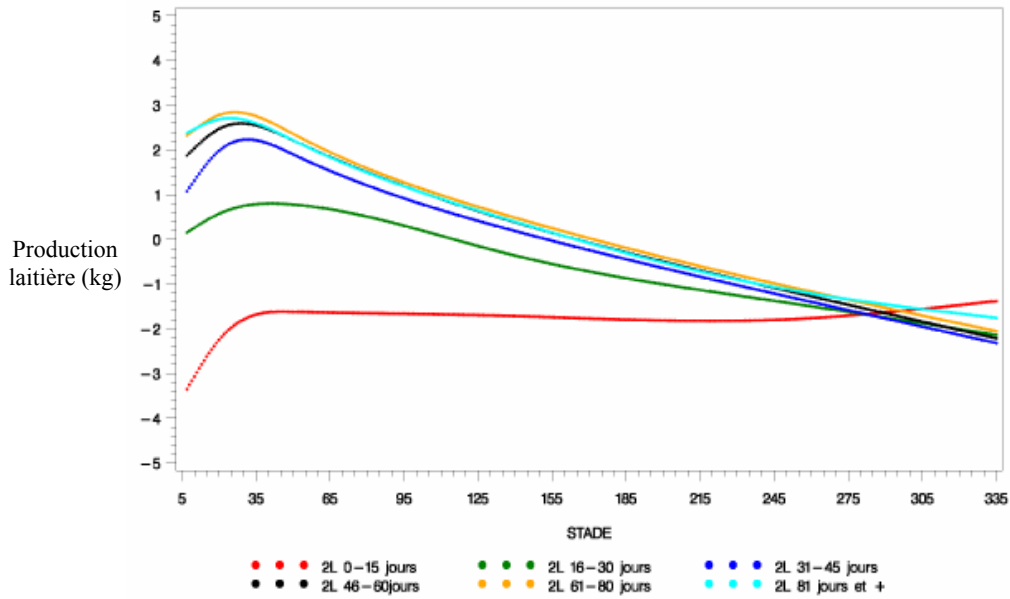


Figure 3.8. Effet de 6 classes de durée de tarissement en 2<sup>ème</sup> lactation sur la production laitière (kg) en fonction du stade de lactation en race Montbéliarde (données du Jura).

## CONCLUSIONS

Cette étude a permis d'affiner le modèle d'évaluation génétique qui peut pour les 3 caractères de quantités ainsi que le taux butyreux et protéique être écrit sous la forme :

$$\begin{aligned}
 Perf_{ijklmnopqt'} = & HTD_i + \\
 & \left. \begin{aligned}
 & région_j \times année_k \times nlac_l \times durée\_tarissement_m + \\
 & région_j \times année_k \times nlac_l \times mois_n + \\
 & région_j \times année_k \times nlac_l \times age_o + \\
 & spl_{DIM}(région_j \times nlac_l \times durée\_tarissement_m, t) + \\
 & spl_{DIM}(région_j \times nlac_l \times mois_n, t) + \\
 & spl_{DIM}(région_j \times nlac_l \times age_o, t) + \\
 & spl_{DCC}(région_j \times nlac_l, t') +
 \end{aligned} \right\} \begin{array}{l} \text{Constantes} \\ \text{Courbes} \end{array} \left. \vphantom{Perf_{ijklmnopqt'}} \right\} \text{Effets fixes} \\
 & \sum_{a=1}^4 (\text{génétique}_{pa} \cdot \nu_{at}) + \sum_{b=1}^4 (\text{environnement\_permanent}_{pb} \cdot \zeta_{bt}) + \\
 & \sum_{c=1}^6 \left( (\text{troupeau}_q \times année_k)_c \cdot \tau_{ct} \right) + e_{ijklmnopqt'} \left. \vphantom{\sum_{c=1}^6} \right\} \text{Effets aléatoires}
 \end{aligned}$$

Pour la numération cellulaire, le modèle est légèrement simplifié puisqu'il ne prend en compte que les 2 courbes modélisant l'effet du mois de vêlage et l'effet de la gestation sur la forme de la courbe de lactation. De plus, il n'inclut pas d'effets aléatoires sur la variable troupeau-année. La performance  $y$  est donc décrite comme :



$$\begin{aligned}
 Perf_{ijklmnopqt'} = & HTD_i + \\
 & \left. \begin{aligned}
 & région_j \times année_k \times nlac_l \times durée\_tarissement_m + \\
 & région_j \times année_k \times nlac_l \times mois_n + \\
 & région_j \times année_k \times nlac_l \times age_o + \\
 & spl_{DIM}(région_j \times nlac_l \times mois_n, t) + \\
 & spl_{DCC}(région_j \times nlac_l, t') +
 \end{aligned} \right\} \begin{array}{l} \text{Constantes} \\ \text{Courbes} \end{array} \left. \vphantom{\begin{aligned} région_j \times année_k \times nlac_l \times durée\_tarissement_m + \\ région_j \times année_k \times nlac_l \times mois_n + \\ région_j \times année_k \times nlac_l \times age_o + \\ spl_{DIM}(région_j \times nlac_l \times mois_n, t) + \\ spl_{DCC}(région_j \times nlac_l, t') + \end{aligned}} \right\} \begin{array}{l} \text{Effets} \\ \text{fixes} \end{array} \\
 & \sum_{a=1}^4 (\text{génétique}_{pa} \cdot \nu_{at}) + \sum_{b=1}^4 (\text{environnement\_permanent}_{pb} \cdot \zeta_{bt}) + \\
 & + e_{ijklmnopqt'} \left. \vphantom{\sum_{a=1}^4 (\text{génétique}_{pa} \cdot \nu_{at}) + \sum_{b=1}^4 (\text{environnement\_permanent}_{pb} \cdot \zeta_{bt}) +} \right\} \begin{array}{l} \text{Effets} \\ \text{aléatoires} \end{array}
 \end{aligned}$$

En race Montbéliarde, la forme des courbes de lactation ne montre pas d'évolution liée aux effets fixes au cours des 15 dernières années. Il est probable que cette constatation soit également valable dans les autres races. Toutefois, elle fera l'objet de vérifications ultérieures. Les changements de forme de courbes de production au cours du temps sont donc probablement imputables aux effets aléatoires tels que l'effet génétique additif, l'effet d'environnement permanent voire celui du troupeau. Cette conclusion nous permet d'envisager la réalisation des évaluations génétiques à partir d'une approche en deux étapes, où les données sont précorrignées dans une première étape pour les effets ne dépendant pas du temps, ce qui contribue à réduire considérablement les besoins informatiques. Les résultats obtenus avec cette approche ne présentent pas de biais par rapport à une évaluation classique où tous les effets sont estimés simultanément. Ce modèle est considéré être celui qui sera utilisé en France pour l'évaluation génétique sur les contrôles élémentaires.

# **Chapitre 4**

## **Analyse des effets fixes du modèle**

## Chapitre 4

L'objet initial de cette thèse, tel que décrit dans la demande d'action innovante qui a permis son financement, était de permettre « l'utilisation en appui technique des résultats des évaluations génétiques des vaches laitières sur la production à chaque contrôle ». Nous nous sommes donc intéressés plus particulièrement à l'analyse des effets fixes qui, dans le cadre de la valorisation de l'évaluation génétique sur les contrôles élémentaires à des fins d'appui technique, sont les effets les plus pertinents puisqu'ils concernent un grand nombre d'animaux.

L'analyse des courbes de lactation telles que définies dans la partie 3.3 a permis à partir des données des trois principales races françaises que sont la Holstein, la Montbéliarde et la Normande de quantifier l'impact de nombreux facteurs de variations sur la quantité de lait, les taux butyreux et protéique et les comptages leucocytaires. Les effets tels que la race, le rang de lactation, la région, le stade de lactation, l'âge au vêlage, le mois de vêlage, la durée de tarissement mais aussi l'effet de la gestation ont ainsi été pris en compte à travers quatre courbes modélisées à partir de splines. La 1<sup>ère</sup> partie de ce chapitre est consacrée à la présentation des principaux résultats obtenus lors de ces analyses. Dans la mesure du possible, nous avons cherché à mettre en évidence à la fois les caractéristiques communes entre les races, les régions, les caractères étudiés... mais également les principales différences en essayant d'en avancer les possibles raisons physiologiques ou environnementales.

La seconde partie de ce chapitre a été consacrée à l'analyse de l'effet troupeau-jour de contrôle. Cet effet permet de caractériser finement les effets d'environnement à court terme tels que l'alimentation, la météorologie... du jour du contrôle. De ce fait, il est unanimement reconnu comme un très bon indicateur de la conduite du troupeau. Afin d'accroître son intérêt à des fins d'appui technique, diverses méthodes permettant de le prédire ont été étudiées et surtout nous avons cherché à mettre en avant les applications potentielles qui pouvaient en être tirées pour la conduite du troupeau.

## **Partie 4.1**

# **Impact des effets d'environnement sur la production**

# Impact des effets environnementaux sur le lait, les taux et la numération cellulaire en race Holstein, Montbéliarde et Normande

H. Leclerc<sup>1,2†</sup>, D. Duclos<sup>1</sup>, A. Barbat<sup>1</sup> et V. Ducrocq<sup>1</sup>

<sup>1</sup>INRA, UR337, Station de Génétique Quantitative et Appliquée, 78352 Jouy-en-Josas

<sup>2</sup>Institut de l'Élevage, Département de Génétique, INRA-SGQA, 78352 Jouy-en-Josas

Courriel : [helene.leclerc@jouy.inra.fr](mailto:helene.leclerc@jouy.inra.fr)

## Résumé

Les facteurs de variations des caractères de production laitière sont nombreux. A travers l'étude des millions de lactations recueillies par le contrôle laitier et stockées depuis les années 1990 à des fins d'évaluations génétiques, il a été possible de modéliser l'impact des principaux facteurs que sont la race, le numéro de lactation, la région, le stade de lactation, l'âge au vêlage, le mois de vêlage, la durée de tarissement et la gestation sur la quantité de lait, les taux butyreux et protéique ainsi que la numération cellulaire. Compte tenu de la quantité de données disponible, il a été possible de définir finement l'ensemble des facteurs de variation et de les modéliser sous forme de courbes. L'évolution des caractères en fonction du stade de lactation est ainsi prise en compte. L'effet du mois de vêlage se révèle être prééminent sur la quantité de lait et le taux butyreux, alors que l'effet de la gestation est conséquent sur la quantité de lait produite surtout en fin de gestation. Par contre, pour l'âge au vêlage et la durée de tarissement, seules les classes fortement pénalisantes comme un vêlage précoce et une durée de tarissement très courte se démarquent fortement et modifient la trajectoire de la courbe de lactation. Le taux protéique et la numération cellulaire présentent quant à eux des formes de courbes de lactation plus homogènes entre classes. Les perspectives de valorisation des courbes de lactation, définies en fonction des caractéristiques propres à l'animal sont nombreuses, notamment pour la prédiction de la production à venir.

---

† Article en préparation pour INRA Productions Animales

Les effets environnementaux comme la région, le rang de lactation, l'âge de l'animal lors du vêlage, le mois de vêlage, la durée de tarissement précédant le vêlage, l'éventuelle gestation en cours sont des facteurs reconnus comme ayant un impact sur la production journalière et constituent donc des éléments nécessaires à la prédiction individuelle de la production sur la lactation.

De nombreuses études ont été consacrées à l'analyse des effets environnementaux ayant un impact sur la courbe de lactation principalement au cours des années 1970 et 1980, parmi lesquelles nous pouvons citer Danell (1982), Wilmink (1987) et Pérochon *et al.* (1996). La plupart du temps, ces études ont été consacrées à la quantité de lait, n'abordant le taux butyreux que ponctuellement (Danell, 1982). Elles ont mis en évidence l'importance des effets d'environnement tels que le troupeau, l'année, la région, le mois de vêlage sur les performances laitières enregistrées, mais aussi des effets physiologiques tels que le rang de lactation, l'effet de la gestation ou des effets de conduite du troupeau tels que l'âge au vêlage ou la durée de la période de tarissement précédent la lactation. Malgré leurs origines différentes, ces effets sont généralement regroupés sous le terme d'« effets environnementaux ».

Selon les études, deux finalités peuvent être distinguées : le développement de facteurs correctifs en vue d'une évaluation génétique dont les effets sont estimés de façon globale sur la lactation (Danell, 1982 ; Wilmink, 1987) ou bien en vue du développement d'outils d'appui technique afin de prédire la production en fonction des effets d'environnement prévisibles. Dans ce cas, la production est modélisée en fonction du stade de lactation de l'animal généralement à l'échelle de la semaine (Pérochon *et al.*, 1996). Pour les évaluations génétiques, les effets sont estimés à partir une population très vaste, généralement à l'échelle nationale, ce qui permet de distinguer de multiples facteurs de variation et d'en étudier les nombreuses modalités définies par des classes d'amplitude limitée. Pour l'appui technique, les données proviennent le plus souvent de troupeaux expérimentaux dont les conditions de milieu sont contrôlées et les performances enregistrées plus fréquemment, mais dont les effectifs sont limités et restreignent l'étude à l'analyse de quelques modalités de facteurs de variation sélectionnés. Toutefois, dans le cadre de la mise en place d'une évaluation génétique basée sur les contrôles élémentaires, il devient possible de concilier ces deux objectifs à travers une modélisation fine de la production.

L'évaluation génétique basée sur les contrôles élémentaires consiste à estimer la valeur génétique additive d'un animal à partir de l'ensemble des contrôles de performances enregistrées par le contrôle laitier (environ 10 contrôles par lactation) plutôt que par les seules variables composites que sont les productions totales sur une lactation. Comme dans l'évaluation génétique basée sur les données de lactation, le modèle contrôles élémentaires inclut de nombreux effets fixes (race, année, région, âge au vêlage, mois de vêlage...) modélisés sous forme de constantes, mais sa principale particularité est de modéliser l'interaction de la plupart de ces effets avec le stade de lactation. Ces interactions se matérialisent par des courbes dont la forme et le niveau varient suivant les différents effets. Les effets dits aléatoires qui dépendent de l'animal tels que l'effet génétique ou l'effet de l'environnement permanent (effet non transmissible qui permet de tenir compte de la répétabilité de la performance) sont aussi modélisés sous forme de courbes dont l'effet varie avec le stade de lactation. Le modèle contrôles élémentaires permet donc une modélisation beaucoup plus fine de la production laitière sur la lactation, notamment grâce aux effets environnementaux décrits finement.

L'objectif de cette étude est de montrer, pour les trois principales races laitières françaises, l'impact des différents facteurs environnementaux analysés sur la quantité de lait, les taux butyreux et protéique ainsi que la numération cellulaire, à partir des résultats obtenus sur l'ensemble de la population française dans le cadre de l'évaluation génétique basée sur les contrôles élémentaires.

## **1 / Données**

Dans le cadre des évaluations génétiques nationales, nous avons à notre disposition les données collectées dans l'ensemble des élevages adhérents au contrôle laitier stockés dans la base de données du Système d'Information Génétique (SIG). Ainsi, fin 2007, nous disposions de près de 34.0, 5.7 et 5.2 millions de lactations en races Holstein, Montbéliarde et Normande respectivement. Dans le cadre du contrôle laitier, les troupeaux sont généralement contrôlés 11 fois dans l'année sur une période de 24 heures. En moyenne, nous disposons donc de 10 contrôles par lactation complète, les lactations en cours disposant de moins de 3 contrôles n'étant pas prises en compte dans le modèle contrôles élémentaires. A partir de ces fichiers, seules les données collectées pendant la période figurant dans le tableau 1 au cours de l'une des 3 premières lactations et ayant un stade de lactation compris entre 7 et 335 jours ont été analysées. Pour être inclus dans l'analyse, l'enregistrement devait avoir une production

minimum de 2.0 kg, conformément au règlement technique du contrôle laitier et des taux compris entre 15 et 90 g/kg pour le taux butyreux (TB) et entre 10 et 70 g/kg pour le taux protéique (TP) afin de répondre aux recommandations internationales du Comité International pour le Contrôle des Performances en Élevage (ICAR). Les données de comptages cellulaires (CCS) sont transformées en un score de cellules somatiques (SCS), afin de normaliser leur distribution suivant la formule décrite par Rupp et Boichard (1997) :

$$SCS = \log_2 \left( \frac{CCS}{100\,000} \right) + 3.$$
 Ainsi, un CCS de 12 500 équivaut à un score de 0, celui de 25 000 à un score de 1... toute augmentation du SCS de 1 correspondant à un doublement de la valeur de CCS précédente.

Outre les données relatives aux caractères de production, le SIG dispose également des informations relatives aux inséminations artificielles, aux naissances, aux vêlages... Il est ainsi possible de déterminer la date de fécondation pour chaque gestation. Pour les vaches dont la gestation est théoriquement encore en cours au moment de l'extraction du fichier pour l'évaluation génétique, la dernière insémination est considérée comme fécondante sauf si la date du terme prévu à partir de la durée de gestation de la race est dépassée. Lorsque les lactations sont terminées et qu'elles ont donné lieu à un nouveau vêlage, la date de fécondation est calculée à partir de la date de vêlage suivante et de la durée de gestation moyenne de la race (282 jours en race Holstein, 287 jours en race Normande et 288 jours en race Montbéliarde).

Concernant l'autre facteur essentiel affectant la production laitière que constitue la conduite du troupeau, nous ne disposons pas d'un relevé précis des conditions de conduite des animaux, notamment de l'alimentation et du statut sanitaire. Ces informations ne sont généralement disponibles que dans les troupeaux expérimentaux ou dans certaines bases de données gérées par les organismes de contrôle laitier mais qui n'ont jamais pu, jusqu'à présent être harmonisées ni centralisées au niveau national. Pour palier à cette absence, le modèle définit un effet spécifique au troupeau pour un jour de contrôle donné (appelé HTD pour Herd – Test day) à partir des observations recueillies le même jour dans un même troupeau (sur un minimum de 5 vaches). Cet effet permet de tenir compte de l'environnement de production du jour donné et notamment de la météorologie, facteur rarement intégré aux études visant à modéliser la courbe de production.



## 2 / Le modèle d'analyse

Le modèle d'analyse utilisé pour cette étude correspond à la 1<sup>ère</sup> étape de l'évaluation génétique sur les contrôles élémentaires, dont la mise en place est en cours de préparation. Les effets aléatoires, comme par exemple l'effet génétique lié à l'animal, ne sont pas pris en compte au cours de cette étape. L'analyse est réalisée indépendamment pour chaque race et pour chaque caractère. Les six caractères (quantité de lait, les matières grasse et protéique, le TB et TP ainsi que le SCS) ont été analysés. Les résultats obtenus pour les matières grasse et protéique ne sont toutefois pas restitués dans cette analyse puisque ces caractères dépendent étroitement de la quantité de lait et des taux mesurés. Nous noterons cependant que le TB et le TP ne feront pas, *a priori*, l'objet d'une évaluation génétique directe, mais seront déterminés à partir des évaluations génétiques sur les quantités de matières, comme c'est le cas dans l'ensemble des pays étrangers.

Dans cette 1<sup>ère</sup> étape, la performance (lait 24 heures, TB et TP) y est décrite comme :

$$Perf_{ijklmnopqt'} = HTD_i + \text{région}_j \times \text{année}_k \times \text{nla}_l \times \text{mois}_m + \text{région}_j \times \text{année}_k \times \text{nla}_l \times \text{age}_n + \\ \text{région}_j \times \text{année}_k \times \text{nla}_l \times \text{durée\_tarissement}_o + \\ \text{spl}_{\text{DIM}}(\text{région}_j \times \text{nla}_l \times \text{mois}_m, t) + \text{spl}_{\text{DIM}}(\text{région}_j \times \text{nla}_l \times \text{age}_n, t) + \\ \text{spl}_{\text{DIM}}(\text{région}_j \times \text{nla}_l \times \text{durée\_tarissement}_o, t) + \text{spl}_{\text{DCC}}(\text{région}_j \times \text{nla}_l, t') + \\ e_{ijklmnott'}$$

et le SCS du jour du contrôle est lui analysé suivant un modèle simplifié :

$$Perf_{ijklmnopqt'} = HTD_i + \text{région}_j \times \text{année}_k \times \text{nla}_l \times \text{mois}_m + \text{région}_j \times \text{année}_k \times \text{nla}_l \times \text{age}_n + \\ \text{région}_j \times \text{année}_k \times \text{nla}_l \times \text{durée\_tarissement}_o + \\ \text{spl}_{\text{DIM}}(\text{région}_j \times \text{nla}_l \times \text{mois}_m, t) + \text{spl}_{\text{DCC}}(\text{région}_j \times \text{nla}_l, t') + e_{ijklmnott'}$$

où  $Perf_{ijklmnott'}$  est la performance 24 heures obtenue à un stade de lactation  $t$  et de gestation  $t'$ , dans des conditions troupeau – jour de contrôle  $i$ , dans la région  $j$ , l'année  $k$ , un rang de lactation  $l$ , un mois de vêlage  $m$ , un âge au vêlage  $n$  et une durée de tarissement  $o$ ,

$HTD_i$  est l'effet fixe troupeau – jour de contrôle  $i$ ,

$\text{spl}_{\text{DIM}}$  est une spline avec 6 nœuds utilisée pour modéliser l'effet du stade de lactation,

$\text{spl}_{\text{DCC}}$  est une spline avec 4 nœuds utilisée pour modéliser l'effet du stade de gestation,

$e_{ijklmnott'}$  est la résiduelle de la performance, c'est-à-dire la différence entre la performance

et la somme des effets décrits ci-dessus.

Une spline est une fonction qui dépend de nœuds, c'est-à-dire des points de jonction situés à des positions stratégiques sur la courbe de lactation (7, 20, 50, 135, 245 et 335 jours pour le stade de lactation ; 100, 150, 200 et 265 jours pour l'effet de la gestation (Druet *et al.*, 2003)) permettant de modéliser sa trajectoire avec un maximum de flexibilité. Entre chaque nœud, le segment de la courbe peut être décrit à partir d'une fonction polynomiale cubique. Les propriétés mathématiques des fonctions splines permettent de s'assurer de la continuité de l'ensemble des segments ainsi définis et de limiter le nombre de coefficients nécessaires à celui du nombre de nœuds de la fonction (pour les courbes dépendant du stade de lactation, nous obtenons ainsi 6 coefficients au lieu de 5 segments de 3 coefficients chacun). La modélisation d'une courbe à l'aide d'une spline permet une bien meilleure flexibilité que lorsqu'une seule fonction polynomiale est retenue sur l'ensemble de la lactation.

Outre l'effet HTD qui permet de tenir compte de l'effet de la conduite du troupeau spécifique au jour du contrôle, nous pouvons observer que 3 constantes sont incluses dans le modèle, afin d'ajuster le niveau de la courbe de lactation en fonction des conditions de milieu (race, région, année, mois de vêlage) et des caractéristiques physiologiques de l'animal en production (rang de lactation, âge au vêlage, durée de tarissement précédente). L'avantage est de pouvoir modéliser la forme de la courbe de lactation en fonction des caractéristiques propres à l'animal à partir de la somme de 4 courbes élémentaires pour les quantités et les taux, à savoir : trois courbes définies par région et rang de lactation modélisant l'effet du mois de vêlage, de l'âge au vêlage et de la durée de tarissement précédente en fonction du stade de lactation noté DIM pour « Days In Milk » à l'aide d'une spline et une courbe définie par région et rang de lactation modélisant l'effet de la gestation. Dans le cas de SCS, seule une courbe modélisant l'effet du mois de vêlage est prise en compte, ainsi qu'une sur l'effet de la gestation. D'après les éléments bibliographiques et les résultats obtenus dans de précédentes études (Druet *et al.*, 2003), la gestation pénalise la production laitière au-delà de 100 – 150 jours après fécondation. Cette spline dépend donc du stade de gestation appelé DCC pour « Days Carried Calf ».

Les données analysées s'appuyant sur la base nationale, les fichiers sont donc représentatifs de l'ensemble des conditions de conduite rencontrées usuellement dans la population. Compte tenu du nombre de données disponibles, les effets analysés peuvent être définis beaucoup plus finement que dans les études se basant sur des données expérimentales. Un effet régional a ainsi été inclus dans l'analyse à travers des interactions avec les effets de l'âge, du mois de vêlage, de la durée de tarissement et de la gestation, comme c'est généralement le cas dans les

évaluations génétiques. La constitution de ces régions s'appuie sur la notion de petites régions agricoles, à l'exception des caractères de production de la race Holstein où 39 régions sont formées sur une base départementale. Toutefois, compte tenu du volume de données à traiter et de la complexité du modèle, cette étude a été réalisée sur la base de la précédente définition des régions en race Holstein utilisée jusqu'en 1997 qui n'en comptait que 8 (Institut de l'Élevage-INRA, 1997). En race Normande, 7 régions étaient initialement formées. Toutefois, les deux dernières régions ont été fusionnées afin d'obtenir des effectifs suffisants pour permettre d'estimer de façon fiable l'ensemble des courbes. En race Montbéliarde, 3 régions sont définies. Les cartes relatives à la constitution des régions pour chacune des races figurent en annexe. Concernant la définition des autres effets, l'effet de la période de vêlage est défini pour l'ensemble des races suivant le mois calendaire. L'âge au vêlage est défini en 1<sup>ère</sup> lactation par 10 classes de 2 mois à partir de l'âge de 22 mois, la dernière classe étant constituée de vaches vêlant à 40 mois ou au-delà. Seule la race Holstein dispose d'une classe supplémentaire pour les vêlages très précoces (21 mois). Pour les multipares, il y a 7 classes de 3 mois pour chaque rang de lactation :  $\leq 36$ ,  $37 - 39$ , ...,  $\geq 52$  pour les 2<sup>èmes</sup> lactations et  $\leq 48$ ,  $49 - 51$ , ...,  $\geq 64$  pour les 3<sup>èmes</sup> lactations. Ces classes sont valables pour l'ensemble des races. Concernant la durée de tarissement précédant la lactation, 6 classes ont été définies pour les 2<sup>èmes</sup> et 3<sup>èmes</sup> lactations :  $\leq 15$  jours,  $16 - 30$  jours,  $31 - 45$  jours,  $46 - 60$  jours,  $61 - 80$  jours,  $81$  jours et plus. Les 1<sup>ères</sup> lactations constituent une classe à part. Un minimum de 100 performances a été considéré nécessaire pour estimer l'effet des constantes faisant intervenir l'année, et de 250 pour les courbes. Les caractéristiques des fichiers de données analysés figurent dans le tableau 1.

### 3 / Analyse des effets inclus dans le modèle

Dans une précédente étude sur la construction du modèle d'évaluation génétique (Leclerc *et al.*, 2008), nous avons mis en évidence que la forme de la courbe de lactation ne dépendait pas de l'année, contrairement à son niveau global (la forme de la courbe dépend néanmoins des caractéristiques propres à l'animal tels que l'effet génétique additif et d'environnement permanent ainsi que l'effet du troupeau). Les effets de l'âge au vêlage, du mois de vêlage et de la durée de tarissement précédent sont donc modélisés d'une part à travers une courbe et d'autre part à travers une constante fonction de l'année. La courbe permet ainsi de décrire la trajectoire de la lactation liée à l'effet considéré et la constante module le niveau de la courbe en fonction de l'année. Pour faciliter l'interprétation des résultats de cette étude, les courbes

représentées ici incluent l'effet moyen de la constante sur l'ensemble des années exprimé par rapport à une classe de référence. Cette classe de référence est constituée des vaches en 1<sup>ère</sup> lactation, vêlant en septembre à l'âge de 30 mois, pour la race et la région considérée.

Parmi les nombreux facteurs de variations pris en compte dans cette étude, l'effet de la race et du rang de lactation font partie de ceux dont l'impact est perceptible à l'aide de simples moyennes de la production journalière (Tableau 1). En comparaison à la race Holstein, numériquement la plus importante, la race Montbéliarde produit 15% de lait en moins avec un TB légèrement inférieur, mais un TP un peu supérieur et surtout un niveau de numération cellulaire inférieur. Toujours par rapport à la race Holstein, la Normande produit en moyenne 25% de moins, mais avec des taux très largement supérieurs et une numération cellulaire quasi-équivalente. Le rang de lactation a un effet majeur sur le niveau de production (Tableau 2). En comparaison à une 1<sup>ère</sup> lactation, une 2<sup>ème</sup> lactation produira entre 12 et 15% de plus suivant la race et la production d'une 3<sup>ème</sup> lactation sera de 20% supérieure à celle d'une 1<sup>ère</sup> lactation quelle que soit la race. Les taux progressent légèrement entre la 1<sup>ère</sup> et 2<sup>ème</sup> lactation, mais se stabilisent voire régressent en 3<sup>ème</sup> lactation. Ces résultats sont similaires aux observations faites par Schutz *et al.* (1990). Le score cellulaire augmente lui aussi d'environ 0.5 point entre la 1<sup>ère</sup> et la 3<sup>ème</sup> lactation, sauf pour la race Montbéliarde dont le niveau s'accroît de seulement 0.3 point.

### 3-1 / Effet des années

L'analyse de la constante tenant compte de l'interaction avec l'année montre que globalement les différences de niveau entre courbes évoluent peu, comme en témoignent les faibles écarts-types obtenus sur les effets de l'âge au vêlage et du tarissement (entre 0.1 et 0.2 kg de lait quelle que soit la race à l'exception des vêlages très précoces et de tarissements inférieurs à 1 mois ; résultats non montrés). Seul le mois de vêlage entraîne une variabilité plus importante selon l'année pour tous les caractères et est plus marquée en race Holstein. L'effet des vêlages sur la période septembre – février présente un écart-type compris entre 0.1 et 0.2 kg de lait, 0.1 à 0.2 g/kg de TB et TP et 0.03 à 0.05 point de SCS, qui double généralement pour les vêlages sur la période mars – août certainement à cause d'une plus grande variabilité alimentaire sur la 1<sup>ère</sup> moitié de la lactation suivant les années. En race Holstein, l'écart type peut atteindre 0.9 kg de lait par jour et 0.5 g/kg de TB et TP pour les vêlages en période estivale.

### 3-2/ Effet de l'âge au vêlage

L'effet de l'âge au vêlage a été fréquemment étudié, notamment par Danell (1982) et Wilmink (1987), afin de quantifier son influence sur la production au cours de la lactation. Quelle que soit la race, les profils sont similaires et montrent un accroissement du niveau de production avec l'âge de l'animal, comme l'illustre la figure 1 en race Montbéliarde. Entre un premier vêlage à 2 ans vs 3 ans, la production d'une lactation de 305 jours augmente de 733 kg pour cette race dans la région 1 (734 kg en Holstein et 600 kg en Normande dans la région 2). La différence de production journalière s'amointrit avec l'avancement de la lactation (3.6 kg par jour au pic de lactation contre 1.8 kg après 305 jours de lactation). La différence de production intra rang de lactation entre des vaches ayant une différence d'âge de 1 an s'amenuise avec les lactations puisque cette différence n'est plus que de 360 kg en 2<sup>ème</sup> et de 210 kg en 3<sup>ème</sup> lactation en race Montbéliarde. Pour le TB, l'effet de l'âge au vêlage sur la 1<sup>ère</sup> lactation est différent suivant les races (Figure 2). En race Montbéliarde, l'âge entraîne des différences sur les 2 premiers mois de lactation, mais pas par la suite. En race Normande, cette période est un peu plus longue et se prolonge jusqu'au milieu de la lactation. En race Holstein, des différences notables perdurent durant toute la 1<sup>ère</sup> lactation, et se traduisent par un taux supérieur de 2 g/kg au moment du pic de lactation et de près de 1 g/kg en fin de lactation en faveur des animaux vêlant tardivement. Des hypothèses explicatives peuvent être émises à défaut de disposer de connaissances bibliographiques pour expliquer ce gradient observé sur les profils de taux TB en fonction de l'âge au vêlage et de la race de l'animal. Nous pouvons supposer que les animaux de race Holstein présentent une bonne aptitude à mobiliser leurs réserves d'autant plus qu'ils vêlent tardivement, ce qui se traduit par un TB plus élevé. Au contraire, les vaches vêlant précocement disposent de réserves plus réduites et synthétisent donc moins de matières grasses qui, même si le volume de lait est inférieur, se traduit par un TB plus faible. Les vaches de races Montbéliarde et Normande ont un métabolisme différent. Leur maturité est plus tardive et leur aptitude à mobiliser les réserves corporelles est plus limitée. Il est donc probable que ces animaux, même lorsqu'ils disposent de réserves corporelles, parviennent moins bien à les mobiliser au cours de la lactation, ce qui se traduit par des différences faibles de TB entre les différentes classes d'âge au vêlage. Comme pour la production laitière, les différences entre classes s'amenuisent sans toutefois se résorber complètement en race Holstein. Concernant le TP, nous n'avons pas remarqué d'écarts notables entre les différentes classes d'âge au vêlage, quelle que soit la race, comme l'illustre la figure 3 en race Montbéliarde. Il en est de même sur le rang de lactation. En début de lactation, les animaux les plus jeunes auraient tendance à avoir un TP légèrement

supérieur, certainement à cause d'une moindre dilution dans le lait, qui s'estompe toutefois au cours de la lactation. Ceci suggère donc un rôle différent des réserves corporelles vis-à-vis du métabolisme de synthèse des matières grasses et des matières protéiques. Enfin, la numération cellulaire ne justifie pas la modélisation de l'effet de l'âge au vêlage sous forme d'une courbe évoluant suivant le stade de lactation. Cependant, l'analyse de l'effet de l'âge au vêlage montre qu'en race Normande, l'âge au premier vêlage n'entraîne pas de variation du niveau cellulaire, contrairement à la race Holstein où un vêlage à 3 ans entraîne en moyenne une augmentation du SCS de 0.11 point. Pour les lactations suivantes, un écart d'âge au vêlage d'un an se traduit par une augmentation du SCS de 0.07 point en 2<sup>ème</sup> et 0.05 point en 3<sup>ème</sup> lactation pour la race Normande et de 0.10 et 0.05 points respectivement en race Holstein. La race Montbéliarde a une situation intermédiaire entre les deux précédentes races.

### **3-3/ Effet du mois de vêlage**

Le mois de vêlage a un effet important sur la performance enregistrée, variant en fonction du rang de lactation et de la race de l'animal ainsi que de la région de production reflétant des différences de conditions de milieu. La courbe de l'effet du mois de vêlage ne tient pas compte des conditions de production du troupeau spécifiques au jour de contrôle intégrées dans le modèle à travers l'effet HTD. Ce dernier prend en compte l'effet de l'alimentation et notamment les variations saisonnières en terme de quantité et de qualité tels que celles subies au moment de la mise à l'herbe, mais aussi les effets de la photopériode commune à l'ensemble des animaux du troupeau. Les variations observées dans cette étude correspondent bien à la description des effets du mois de contrôle d'une part et du mois de vêlage d'autre part faite par Wood (1969), Danell (1982) et Wilmink (1987). Les résultats obtenus montrent un pic de production plus élevé pour les vêlages au cours de l'automne (octobre – novembre), quels que soient la race et le rang de lactation. Au contraire, les vaches vêlant l'été (juillet – août) ont une production inférieure, comme l'illustre la figure 4 pour des Holstein en 2<sup>ème</sup> lactation. Dans cette race, la différence de production sur une lactation de 305 jours entre ces mois de vêlage extrêmes est de 263 et 423 kg pour les 2<sup>èmes</sup> et 3<sup>èmes</sup> lactations en région Bretagne. Pour les primipares, la situation est légèrement différente lorsque l'environnement est favorable à la pousse de l'herbe. En effet, dans les régions avec un climat océanique, ce sont les vaches vêlant en mars – avril qui présentent une production inférieure de 200 kg (figure 5), les vêlages estivaux étant moins pénalisés grâce à une pousse de l'herbe plus régulière. En 1<sup>ère</sup> lactation, le pâturage d'une herbe abondante (entre avril et juin) se traduit

par un aplanissement de la courbe pour les lactations en cours (vêlage d'automne) et une phase ascendante plus longue pour les lactations démarrant au cours de cette période. Dans les conditions climatiques plus marquées comme celle de la race Montbéliarde avec un climat continental (région 1), les différences entre mois de vêlage extrêmes (Juillet – Août vs Octobre - Novembre) atteignent 575, 700 et près de 800 kg pour les 1<sup>ères</sup>, 2<sup>èmes</sup> et 3<sup>èmes</sup> lactations respectivement. La forme de la courbe de la lactation dépend donc étroitement du mois de vêlage, du fait d'une interaction très marquée entre le mois de vêlage et le stade de lactation. L'effet de la mise à l'herbe varie ainsi fortement en fonction du stade de lactation de l'animal. Comme le mentionne Wood (1969), il est difficile d'interpréter ces variations et d'identifier avec certitude les facteurs causaux que peuvent être la photopériode, la température, les différences de conduite entre l'hiver et l'été, puisque l'ensemble de ces facteurs sont étroitement imbriqués avec le mois de production, le mois de vêlage, le stade de lactation. Concernant le TB, la trajectoire des courbes de lactation dépend largement du mois de vêlage, comme en témoigne la figure 6 obtenue pour des 1<sup>ères</sup> lactations de race Normande. L'effet de la mise à l'herbe y est encore plus marqué que pour la production laitière avec une différence d'environ 3 g/kg de TB entre les mois extrêmes. Profitant de conditions favorables, les vêlages de printemps montrent des variations de TB entre le pic de lactation et la fin de lactation de moins de 3 g/kg, contre 5 g/kg pour un vêlage en janvier. Globalement, le TP suit une trajectoire beaucoup plus lisse que le TB, ce qui est cohérent avec nos attentes, la variabilité du TP étant nettement plus faible que celle du TB. De forme opposée à celle de la production de lait, la courbe du TP peut être interprétée comme un effet de dilution de la matière protéique dans le lait (résultats non montrés). Nous remarquons tout de même une vague semblable à celle observée sur le TB, dont l'amplitude est cependant réduite, ce qui se traduit par une augmentation du niveau des courbes d'environ 1 g/kg sur la période printanière. Pour la numération cellulaire, la courbe de l'effet du mois de vêlage sur le SCS montre une trajectoire opposée à la production laitière (figure 7), sans forte variation liée à la période de vêlage. Les lactations successives se traduisent par un accroissement de l'amplitude des courbes pour l'ensemble des races. En effet, entre la valeur minimale du SCS obtenue autour du pic de lactation et celle de la fin de lactation, l'augmentation est de 1.2 en 1<sup>ère</sup> lactation, de 2.0 en 2<sup>ème</sup> lactation et de 2.5 en 3<sup>ème</sup> lactation, soit un quadruplement de la numération cellulaire au cours de la lactation pour les multipares.

### 3-4/ Effet de la durée de tarissement

L'effet de la durée de tarissement précédant la lactation est limité à la seule quantité de lait. Sur le TB et le TP, les courbes obtenues sont opposées à celles de la production laitière et suggèrent un simple effet de dilution des matières grasses et protéiques dans le lait. La figure 8 représente les courbes obtenues en 2<sup>ème</sup> lactation en race Holstein pour la production de lait, ainsi que la courbe des vaches en 1<sup>ère</sup> lactation, qui correspond à une fraction de leur courbe de lactation globale. Les courbes des 3<sup>èmes</sup> lactations suivent une trajectoire similaire à celles des 2<sup>èmes</sup> lactations avec toutefois un niveau légèrement supérieur. La classe des tarissements inférieurs ou égal à 15 jours montre une pénalisation importante de ces vaches avec une perte d'environ 7.0 kg de lait au pic de lactation par rapport à une durée de tarissement classique (45 jours ou plus) et plus de 2.0 kg en fin de lactation, soit 1460 kg de lait sur une lactation de 305 jours. Cette interprétation est toutefois à nuancer. En effet, cette classe inclut des vaches ayant avorté et dont le vêlage n'a pas été préparé. Une amélioration du modèle consisterait à identifier les animaux ayant avorté, notamment à travers les données de conditions de naissance disponibles dans le SIG et à les regrouper dans une classe spécifique différente de celle des vaches n'ayant pas été tarées volontairement. Les classes comprenant les durées comprises entre 16 et 30 jours et entre 31 et 45 jours entraînent respectivement une perte de production au cours de la lactation de 550 kg et de 150 kg en comparaison à un tarissement de plus de 45 jours. Au-delà de cette durée, l'accroissement de la période de tarissement ne se traduit pas par une augmentation significative de la production (67 kg entre la classe 46 – 60 jours et les classes suivantes). La durée de tarissement est étroitement liée à l'effet de la gestation et ses besoins. Lors des deux derniers mois de gestation, les besoins énergétiques de la vache sont accrus, pour couvrir les besoins nécessaires à la croissance du veau mais aussi pour reconstituer ses réserves corporelles qui seront mobilisées au début de la lactation suivante. Lorsque le tarissement est inférieur à 45 jours, il y a compétition entre les besoins liés à la production de lait et ceux précédemment mentionnés. La reconstitution des réserves n'est alors que partielle et se traduit donc par une production journalière amoindrie lors de la lactation suivante.

La durée de tarissement a un impact fort sur le niveau cellulaire, dont pâtissent les lactations où la période sèche précédente a été courte. Un tarissement inférieur à 15 jours entre la 1<sup>ère</sup> et la 2<sup>ème</sup> lactation se traduit par une augmentation de 0.15 à 0.25 point de SCS suivant la race et entre 0.37 et 0.46 point de SCS entre les 2<sup>ème</sup> et 3<sup>ème</sup> lactations. L'objectif du tarissement est non seulement de reconstituer les réserves corporelles mais aussi de renouveler le tissu



mammaire et d'induire la production de colostrum. Durant cette phase, les infections résiduelles de la mamelle généralement responsables d'un niveau cellulaire élevé peuvent être soignées grâce à un traitement antibiotique adéquat. Lorsque le tarissement est très court, voire inexistant, les infections s'amplifient avec la nouvelle lactation, ce qui explique les niveaux cellulaires observés pour les classes de tarissement inférieurs à 15 jours.

### 3-5/ Effet de la gestation

L'effet de la gestation sur la lactation au-delà de 100 jours de gestation est représenté pour les trois lactations des trois races sur la figure 9. La valeur de 100 jours est arbitraire, comme en témoigne cette figure. Suivant les races et le rang de lactation, la gestation semble commencer à affecter la production laitière de l'animal entre 100 et 130 jours après fécondation, ce qui est cohérent par rapport aux résultats mentionnés par Danell (1982) ou Coulon *et al.* (1995). Après 230 jours de gestation, soit deux mois avant le vêlage suivant si l'intervalle vêlage-vêlage est de 1 an, la perte de production liée à la gestation varie de façon importante suivant le rang de lactation et la race. En races Holstein et Montbéliarde, la gestation entraîne à cette date une baisse d'environ 3.0 kg en 1<sup>ère</sup> lactation, contre 4.0 kg pour les lactations ultérieures. La race Normande est moins pénalisée, avec une baisse atteignant 1.7 kg en 1<sup>ère</sup> lactation après 230 jours de gestation et d'environ 2.0 kg pour les lactations suivantes. Cette baisse de production durant la gestation est attribuée à des modifications hormonales liées à la croissance fœtale, notamment à l'augmentation des oestrogènes au cours du second tiers de la gestation (Bauman and Currie, 1980). De plus, au cours de la gestation, l'énergie ingérée n'est plus seulement utilisée pour couvrir les besoins d'entretien de l'animal et ceux de sa production laitière. Elle doit aussi répondre aux besoins liés à la croissance du fœtus et la reconstitution de ses réserves corporelles en vue de la lactation suivante. La part de l'énergie utilisée pour la production laitière s'en trouve donc amoindrie, ce qui se traduit par une baisse du niveau de production. Sur les taux, une augmentation progressive est observée, probablement liée à une plus faible dilution des matières dans le lait. Pour le TB, l'augmentation à 230 jours de gestation est quasi identique, quel que soit le rang de lactation, et tourne autour de +2.5 g/kg en Normande, +2.8 g/kg en Holstein et +3.5 g/kg en Montbéliarde, avec des fluctuations régionales. Pour le TP, l'augmentation est d'amplitude limitée et varie suivant le rang de lactation. En 1<sup>ère</sup> lactation, elle est maximale en race Montbéliarde avec 1.0 g/kg. Pour les lactations suivantes, le TP augmente de 1.0 g/kg en Normande à 1.5 g/kg en Holstein et Montbéliarde. Les plus faibles augmentations de taux

observées en Normandie peuvent s'expliquer par la chute inférieure de la quantité de lait au cours de la gestation. Pour le SCS, les courbes montrent de très légères augmentations du niveau cellulaire (entre 0.1 et 0.2 SCS) au-delà de 180 jours de gestation.

### **3-6/ Effet de la région**

L'effet de la région est plus ou moins marqué suivant le caractère et l'effet analysé. La quantité de lait et le TB sont les caractères qui montrent les plus fortes variations contrairement au TP et SCS. Les courbes obtenues pour les classes d'âge au vêlage et la durée de tarissement sont homogènes entre régions intra race, comme l'illustrent les figures 10 et 11. L'effet du mois de vêlage montre des variations régionales plus importantes (Figure 12), surtout pour la période printanière et estivale où les différences climatologiques entre régions peuvent être assez marquées. Toutefois, à quelques exceptions près, l'effet des régions sur les trajectoires des courbes de lactation est faible. Pour cette raison, il est fort probable que l'évaluation génétique sur les contrôles élémentaires en race Holstein se base sur ce niveau de définition de l'effet région et non sur la définition beaucoup plus fine (39 régions) actuellement utilisée, au moins pour l'effet région inclus dans la modélisation des courbes de lactation.

### **3-7/ Les perspectives de valorisation pour la prédiction de la production**

De nombreux auteurs ont envisagé le développement d'outils de prédiction des caractères laitiers et plus particulièrement de la quantité de lait à partir des modèles utilisés pour l'évaluation génétique (Wood, 1969). D'autres ont développé des outils à partir de données expérimentales (Pérochon *et al.*, 1996 ; Coulon et Pérochon, 2000) ou de données recueillies au cours de contrôles laitiers mensuels (Baraton et Masselin, 1990). Deux applications principales peuvent être distinguées, valorisant d'une part l'objectif prospectif de la prédiction et d'autre part son objectif analytique. Concernant le premier objectif, les courbes de lactations présentées précédemment constituent un des éléments de référence pour la prédiction de la production d'un animal en fonction de ses propres caractéristiques. Des prédictions similaires peuvent être obtenues à l'échelle du troupeau en cumulant les courbes correspondant à l'ensemble des animaux. Elles peuvent alors constituer une base pour l'appui technique réalisé par le contrôle laitier auprès de l'éleveur par exemple pour quantifier l'impact de modifications souhaitées sur la structure (par exemple en augmentant la proportion de primipares) et la conduite du troupeau (évolution de l'âge au vêlage, de la

saison de vêlage, de la durée de tarissement). Comme le mentionne Wood (1969), l'objectif est alors d'optimiser la conduite du troupeau afin de dégager une marge maximale sur l'atelier laitier, qui dépend d'une part de la grille de paiement du lait et d'autre part du coût alimentaire. Le second objectif lié aux prédictions basées sur les courbes de lactation est la valorisation possible à des fins de diagnostic. La comparaison de la production réalisée à celle prédite peut faciliter une détection dynamique d'éventuels problèmes et l'identification de leurs causes, comme par exemple des situations non optimales d'alimentation, de conduite au sens large ou de maladies (Coulon et Pérochon, 2000).

Dans notre modèle, la courbe de lactation est partitionnée en 4 effets additifs que sont l'âge au vêlage, le mois de vêlage, la durée de tarissement précédente et l'effet de la gestation en cours s'il a lieu d'être. La forme de la courbe de lactation spécifique d'un animal donné peut être obtenue en sommant l'ensemble des courbes correspondantes. Les formes de courbes de lactation ainsi obtenues pour des vaches de race Holstein, Montbéliarde et Normande dans des conditions de milieu similaire (région Bretagne) en 1<sup>ère</sup>, 2<sup>ème</sup> et 3<sup>ème</sup> lactation vêlant pour la 1<sup>ère</sup> fois à 30 mois en septembre puis tous les ans à la même saison sont représentées à la figure 13. Pour respecter cette périodicité, l'intervalle vêlage – insémination fécondante est supposé être de 80 jours et la durée de tarissement entre lactation est de 45 jours. Seule la race Holstein se distingue avec des pics de production plus marqués en 2<sup>ème</sup> et 3<sup>ème</sup> lactation, ainsi qu'une chute plus importante en fin de lactation due à l'effet de la gestation. La race Normande s'illustre quant à elle par sa faible perte de production en fin de gestation. Pour obtenir la courbe de production de l'animal, il convient d'y ajouter le niveau moyen de production mensuelle du troupeau (HTD) estimé ou prévu, ainsi que les effets liés à l'animal (effet génétique additif et effet d'environnement permanent). L'analyse et la prédiction de l'effet HTD n'ont toutefois pas été abordées au cours de cette étude mais sont essentielles pour le développement des outils d'appui technique proposés.

## **Conclusion et perspectives**

A travers la modélisation des performances laitières journalières dans le cadre de l'évaluation génétique, il a été possible de mettre en évidence et de quantifier l'impact des principaux facteurs de variation de la courbe de lactation que sont, outre le stade de lactation, le rang de lactation, la région de production, l'âge au vêlage, le mois de vêlage, la durée de tarissement précédente et la gestation. L'analyse complète de ces facteurs de variation de la production laitière a été réalisée sur la quantité de lait classiquement étudiée mais aussi sur les taux

protéique et butyreux ainsi que sur la numération cellulaire, en essayant de mettre en avant les possibles raisons physiologiques et environnementales expliquant ces variations. L'effet du mois de vêlage est particulièrement marqué sur la forme de la courbe de lactation pour le taux butyreux et la quantité de lait, cette dernière étant aussi fortement affectée par l'effet de la gestation. Les trajectoires des courbes du taux protéique et de la numération cellulaire semblent plus liées à un effet de dilution dans le lait. Le large champ de l'étude ouvre de nombreuses perspectives de valorisation, notamment en termes de développement d'outils d'appui technique. En intégrant à la forme de la courbe de lactation spécifique à l'animal la production moyenne mensuelle du troupeau correspondant dans notre modèle à l'effet HTD, il devient possible de prédire plus finement la production à court (mois suivant), moyen (campagne laitière en cours) et long terme (campagne laitière suivante). L'inclusion des effets génétique additif et d'environnement permanent lié à l'animal ouvre des perspectives encore plus large en permettant une sélection des animaux en fonction de la production attendue et de la marge dégagée par ceux-ci.

### **Remerciements**

Cette étude a été réalisée dans le cadre de l'Unité Mixte Technologique (UMT) INRA - Institut de l'Élevage sur l'évaluation génétique des bovins. Le ministère de l'agriculture est remercié pour avoir contribué à son financement à travers l'action innovante UTILEG.

### **Références**

Barraton Y. et Masselin S., 1990. Etude des facteurs de variation des courbes de lactation et d'évolution des taux. Compte rendu d'étude, Institut Technique de l'Élevage Bovin. 122p.

Bauman D.E. and Currie W.B., 1980. Partitioning of nutrients during pregnancy and lactation: A review of mechanisms involving homeostasis and homeorhesis. *J. Dairy Sci.* 63, 1514-1529.

Bonaïti B., Boichard D., Verrier E., Ducrocq V., Barbat A. et Briend M., 1990. La méthode française d'évaluation génétique des reproducteurs laitiers. *INRA Prod. Anim.*, 3, 83-92.

Coulon J.B. et Pérochon L., 2000. Evolution de la production laitière au cours de la lactation : modèle de prédiction chez la vache laitière. *INRA Prod. Anim.*, 13, 349-360.

Coulon J.B., Pérochon L. and Lescourret F., 1995. Modelling the effect of the stage of pregnancy on dairy cow's milk yield. *Anim. Sci.*, 60, 401-408.

Danell B., 1982. Studies on lactation yield and individual test-day yields of swedish dairy cows. I. Environmental influence and development of adjustment factors. *Acta Agric. Scan.*, 32, 65-81.

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

Institut de l'Élevage-INRA. 1997. Carte des nouvelles régions en race Prim'Holstein. Notes Indexation Bovine Laitière. 1997-8.

Leclerc H., Duclos D., Barbat A., Druet T. and Ducrocq V., 2008. Environmental effects on lactation curves included in a test-day model genetic evaluation. *Animal*, 2, 344-353.

Pérochon L., Coulon J.B. and Lescourret F., 1996. Modelling lactation curves of dairy cows with emphasis on individual variability. *Anim. Sci.*, 63, 189-200.

Rupp R. et Boichard D., 1997. Evaluation génétique des bovins laitiers sur les comptages de cellules somatiques pour l'amélioration de la résistance aux mammites. *Renc. Rech. Rum.*, 4, 211-214.

Schutz M.M., Hansen L.B., Steuernagel G.R. and Kuck A.L., 1990. Variation of milk, fat, protein, and somatic cells for dairy cattle. *J. Dairy Sci.*, 73, 484-493.

Wilmink J.B.M., 1987. Adjustment of Test-Day milk, fat and protein yield for age, season and stage of lactation. *Livest. Prod. Sci.*, 16, 335-348.

Wood P.D.P. 1969. Factors affecting the shape of the lactation curve in cattle. *Anim. Prod.* 11, 307-316.

## Chapitre 4 - Partie 4.1

Tableau 1. Caractéristiques des données utilisées dans les races Holstein, Montbéliarde et Normande.

	<b>Holstein</b>	<b>Montbéliarde</b>	<b>Normande</b>
Nombre de données	148 088 140	24 432 153	22 237 790
Période d'étude (mois/année)	09/1990 – 08/2007	09/1988 – 12/2007	09/1988 – 12/2007
Ratio 1 <sup>ère</sup> / 2 <sup>ème</sup> / 3 <sup>ème</sup> lactation en %	46.6 / 32.9 / 20.5	45.3 / 32.6 / 22.1	47.4 / 32.5 / 20.1
Nombre de régions	8	3	6
Répartition des données en fonction des régions en %	14.7 / 24.5 / 14.7 / 14.0 / 6.0 / 8.0 / 5.7 / 12.4	74.6 / 17.3 / 8.1	7.8 / 12.3 / 50.6 / 13.5 / 9.1 / 6.7
Nombre de courbes sur le mois de vêlage	288	108	216
Nombre de courbes sur l'âge au vêlage	200	72	144
Nb de courbes sur durée du tarissement préc.	104	39	78
Nb courbes sur l'effet gestation	24	9	18
Nb moyen de vaches par contrôle (écart-type)	19.7 (11.3)	16.8 (9.6)	17.5 (10.4)
Lait moyen en kg/jour (écart-type)	24.6 (7.3)	21.0 (6.3)	18.3 (5.5)
TB moyen en ‰ (écart-type)	40.9 (6.8)	39.4 (5.8)	44.2 (6.5)
TP moyen en ‰ (écart-type)	32.0 (3.4)	33.1 (3.4)	34.6 (3.5)
Taux cellulaire moyen en score (écart-type)	2.94 (1.76)	2.58 (1.73)	3.07 (1.65)

Tableau 2. Moyenne des performances journalières dans les races Holstein, Montbéliarde et Normande selon le rang de lactation pour la production de lait, le taux butyreux et protéique et le Score de Cellules Somatiques (SCS).

	<b>Rang</b>	<b>Lait (kg)</b>	<b>TB (g/kg)</b>	<b>TP (g/kg)</b>	<b>SCS</b>
Holstein	1 <sup>ère</sup> lact.	22.6	40.6	31.7	2.73
	2 <sup>ème</sup> lact.	25.9	41.1	32.3	2.98
	3 <sup>ème</sup> lact.	27.2	41.4	32.1	3.31
Montbéliarde	1 <sup>ère</sup> lact.	19.3	39.4	33.0	2.51
	2 <sup>ème</sup> lact.	22.0	39.5	33.3	2.51
	3 <sup>ème</sup> lact.	23.1	39.4	33.1	2.81
Normande	1 <sup>ère</sup> lact.	17.0	44.0	34.2	2.93
	2 <sup>ème</sup> lact.	19.1	44.5	35.1	3.03
	3 <sup>ème</sup> lact.	20.3	44.2	34.9	3.44

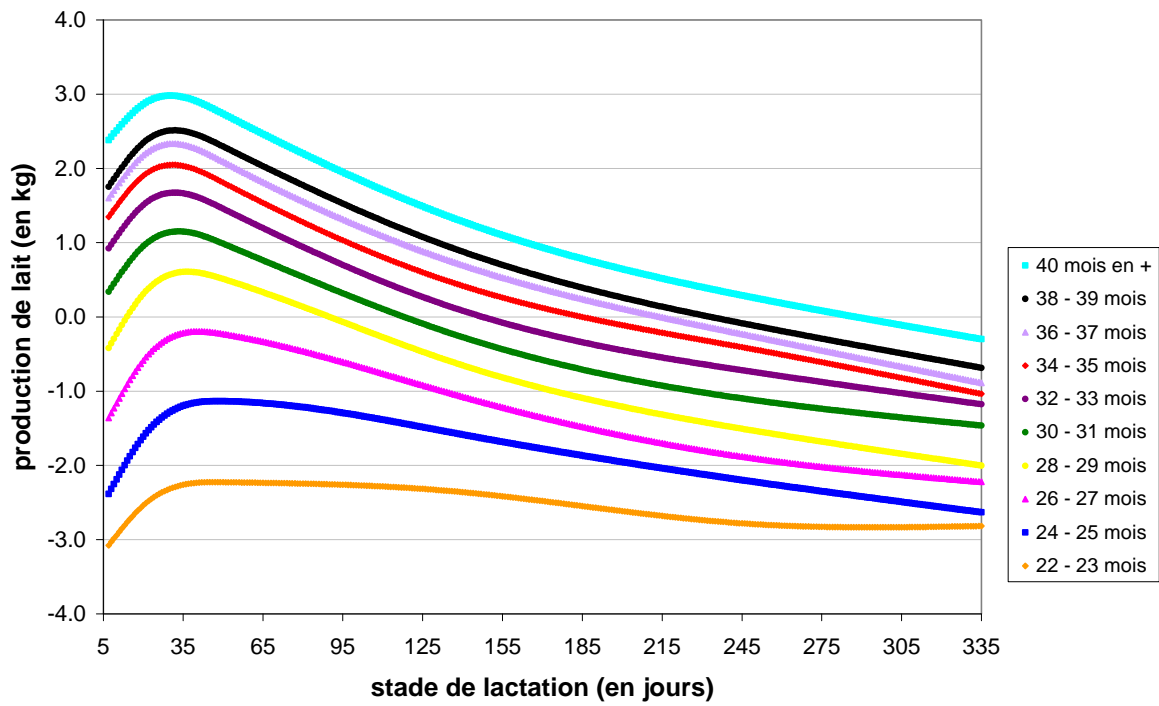


Figure 1. Effet des classes d'âge au vêlage (à partir de 22 mois) sur l'évolution de la production laitière (en kg) en fonction du stade de lactation (en jours) pour les 1<sup>ères</sup> lactations de race Montbéliarde de la région 1.

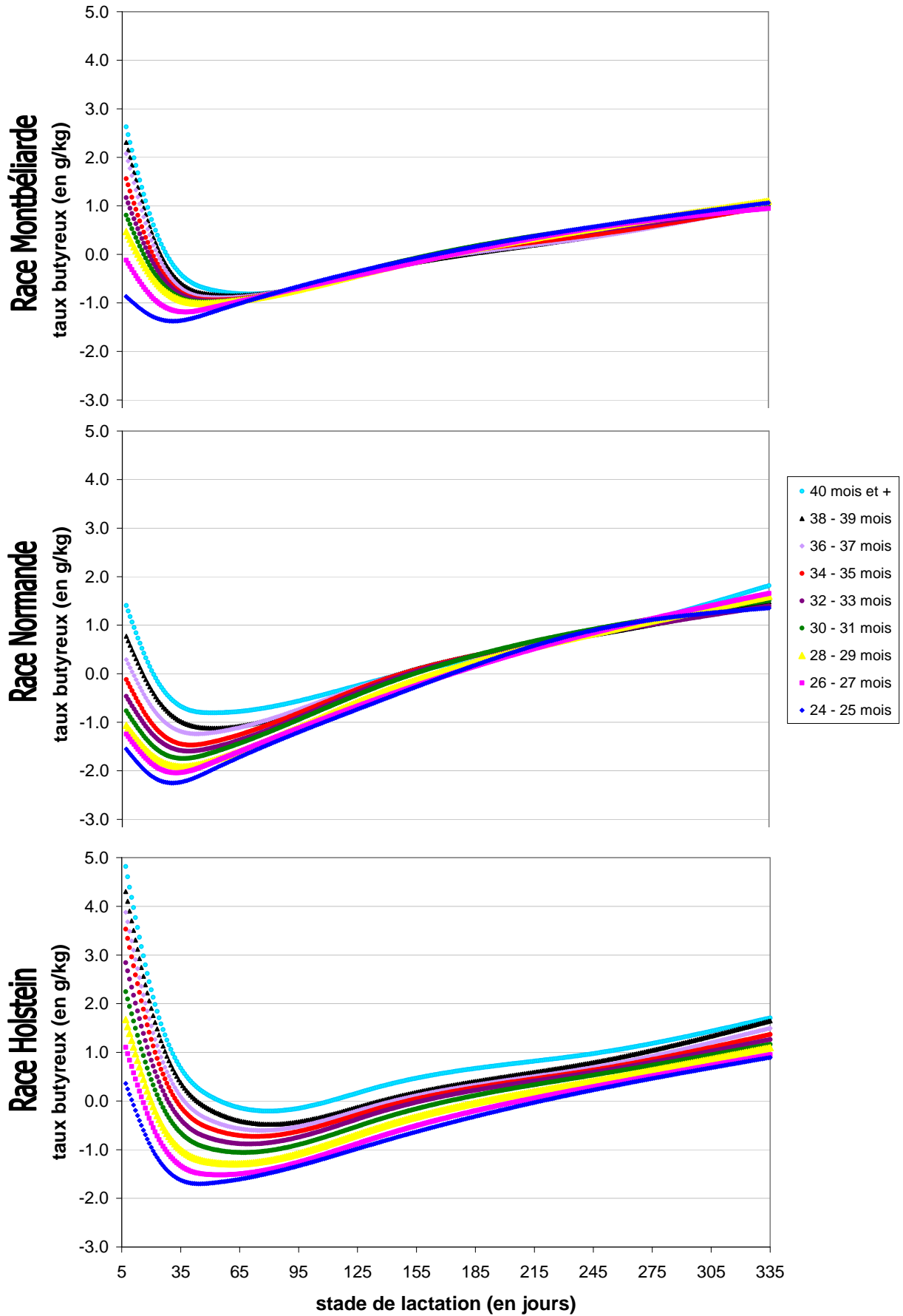


Figure 2. Effet des classes d'âge au vêlage (à partir de 24 mois) sur l'évolution du taux butyreux (en g/kg) en fonction du stade de lactation (en jours) pour des 1<sup>ères</sup> lactations dans les races Montbéliarde (région 1), Normande (région 2) et Holstein (région 2).



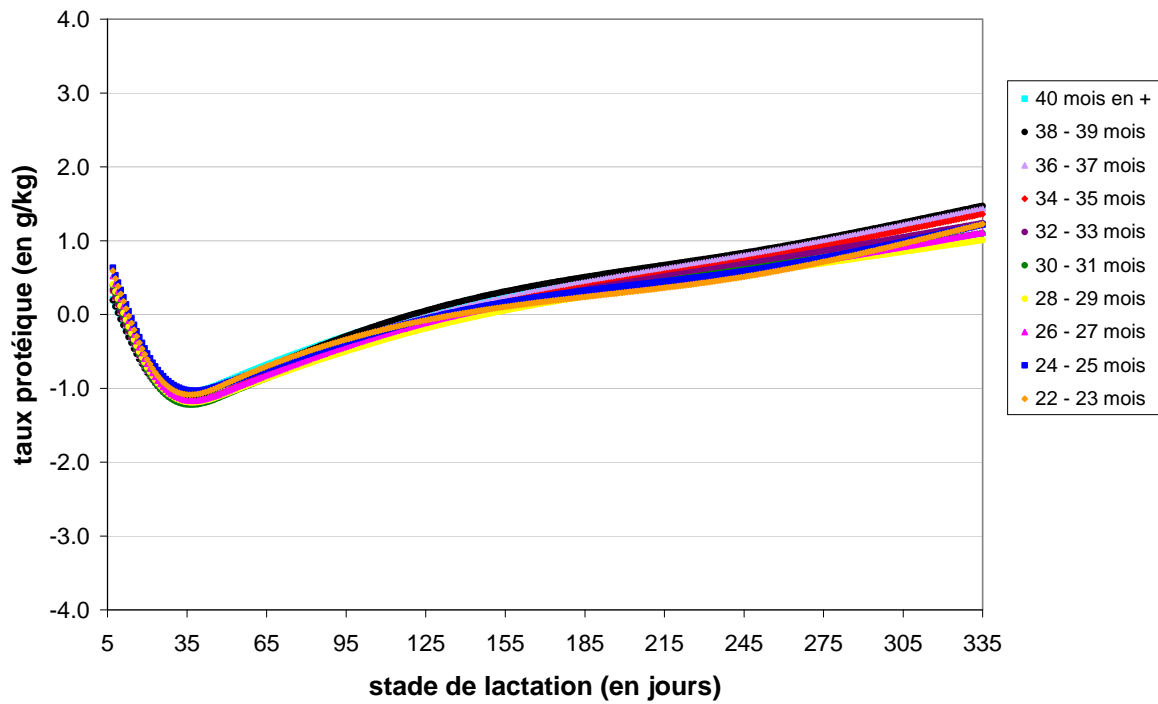


Figure 3. Effet des classes d'âge au vêlage (à partir de 22 mois) sur l'évolution du taux protéique (en g/kg) en fonction du stade de lactation (en jours) pour les 1<sup>ères</sup> lactations de race Montbéliarde de la région 3.

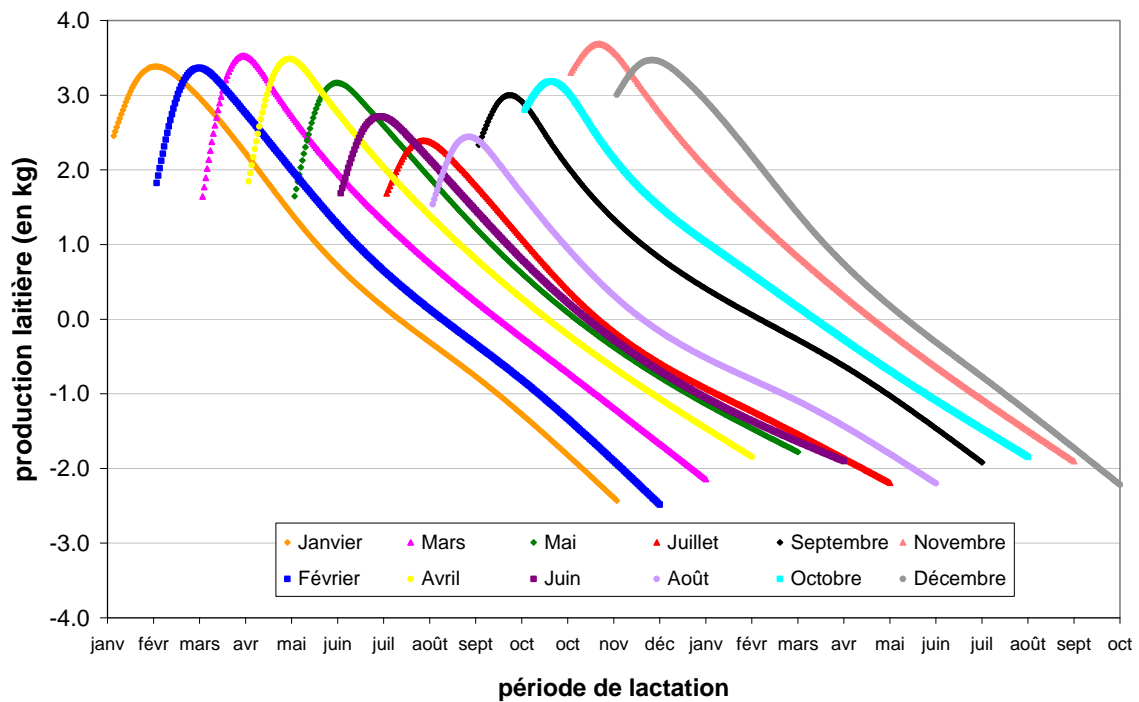


Figure 4. Effet du mois de vêlage sur la production laitière (en kg) en fonction de la période de lactation pour les 2<sup>èmes</sup> lactations de race Holstein de la région 2.

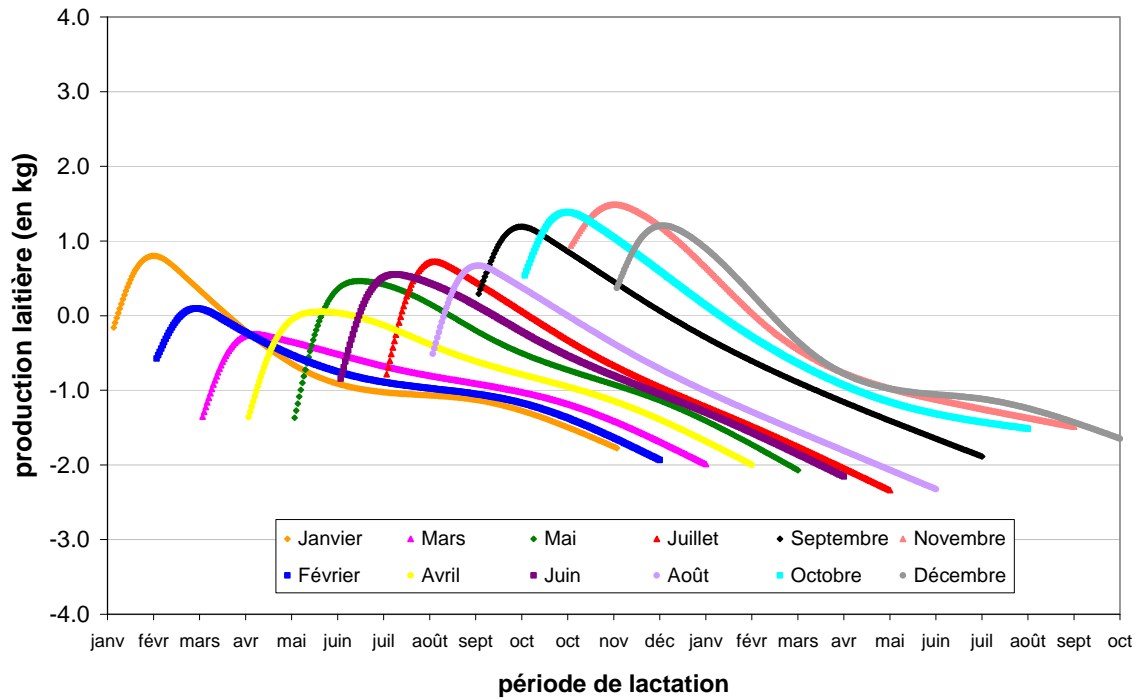


Figure 5. Effet du mois de vêlage sur la production laitière (en kg) en fonction de la période de lactation pour les 1<sup>ères</sup> lactations de race Holstein de la région 2.

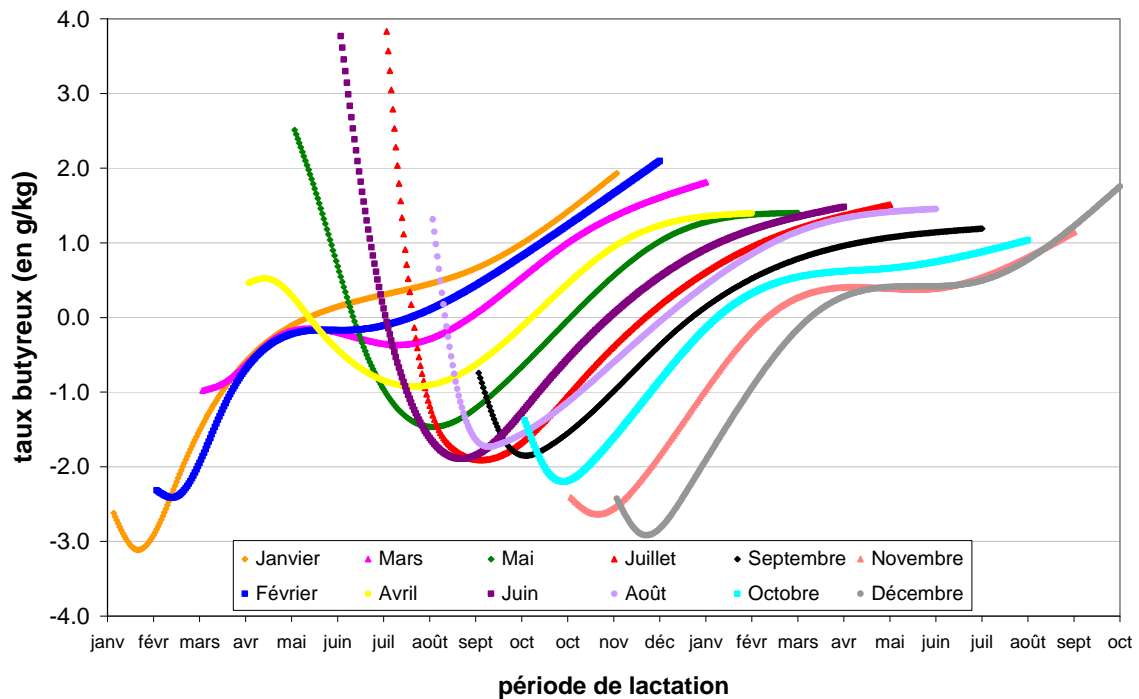


Figure 6. Effet du mois de vêlage sur le taux butyreux (en g/kg) en fonction de la période de lactation pour les 1<sup>ères</sup> lactations de race Normande de la région 2.

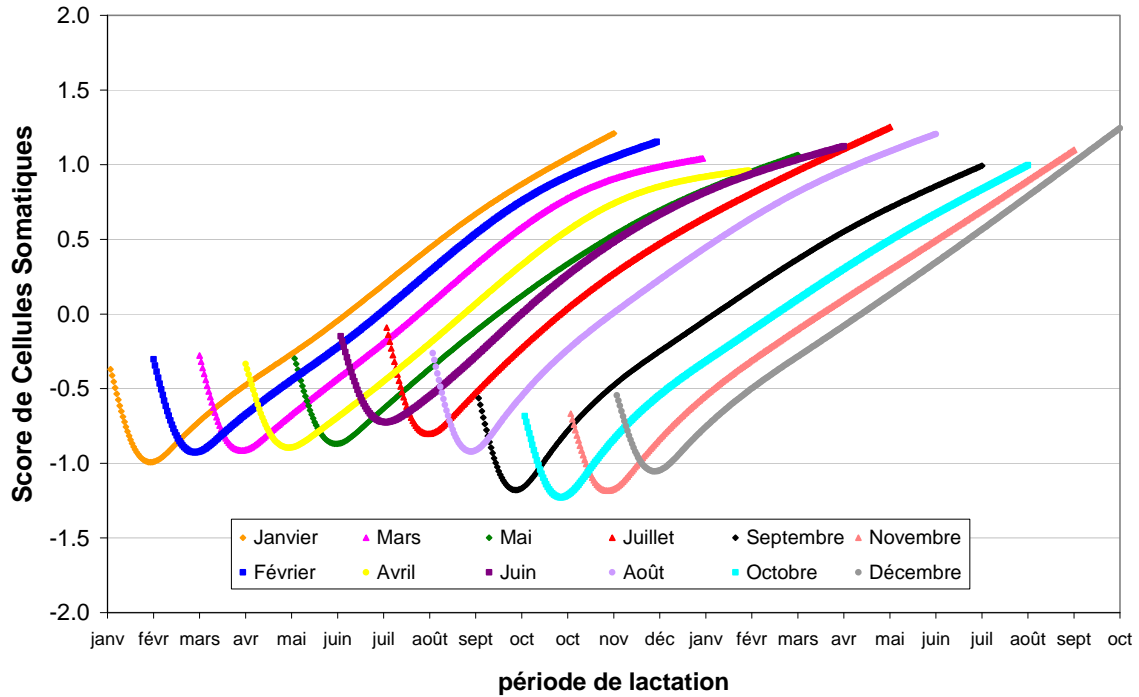


Figure 7. Effet du mois de vêlage sur le Score de Cellules Somatique en fonction de la période de lactation pour les 2<sup>èmes</sup> lactations de race Montbéliarde de la région 2.

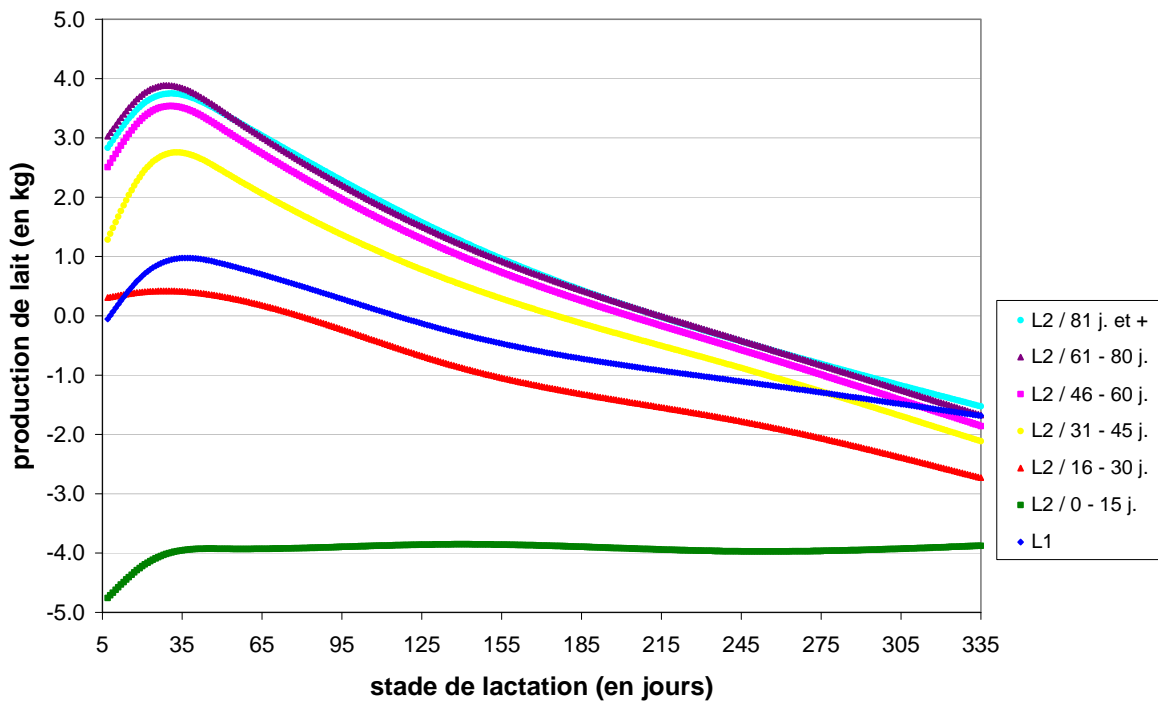


Figure 8. Effet de la durée de tarissement sur la production laitière (en kg) en fonction du stade de lactation (en jours) pour les 1<sup>ères</sup> lactations et 2<sup>èmes</sup> lactations de race Holstein de la région 1.

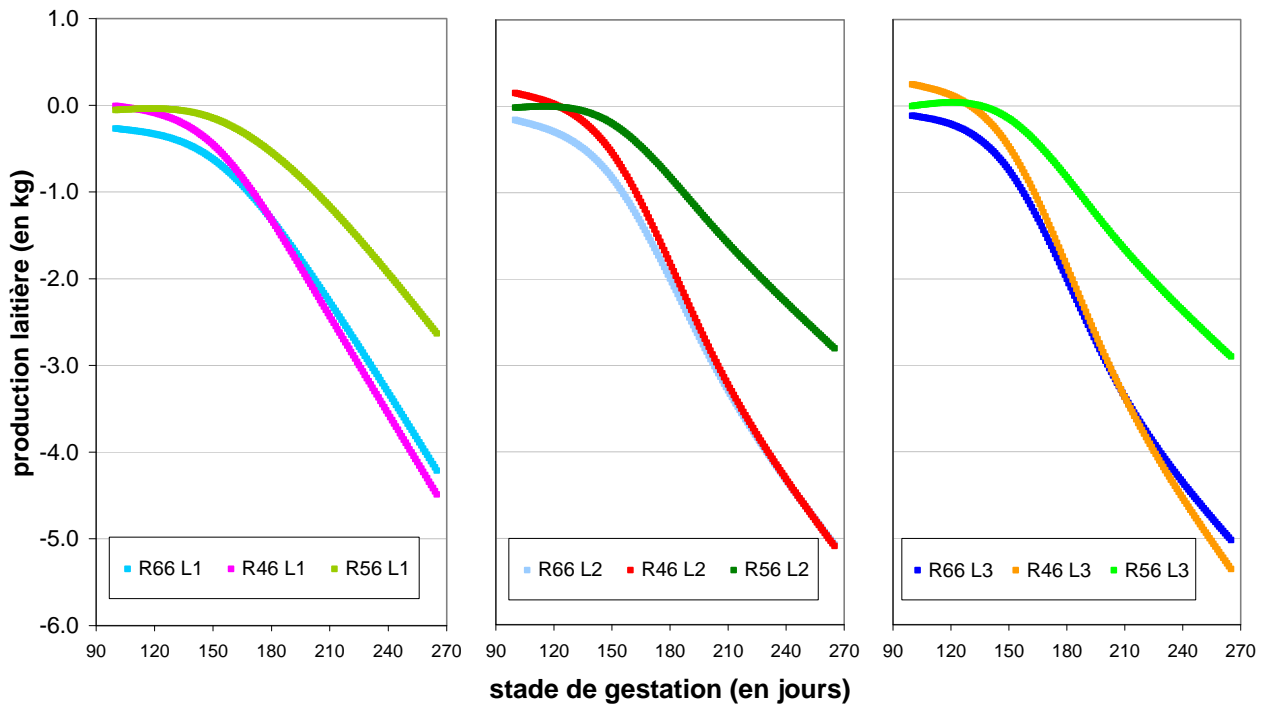


Figure 9. Effet de la gestation sur la production laitière (en kg) en fonction du stade de gestation (en jours) pour les 1<sup>ères</sup>, 2<sup>èmes</sup> et 3<sup>èmes</sup> lactations de race Holstein (R66 région 2), Montbéliarde (R46 région 1) et Normande (R56 région 3).

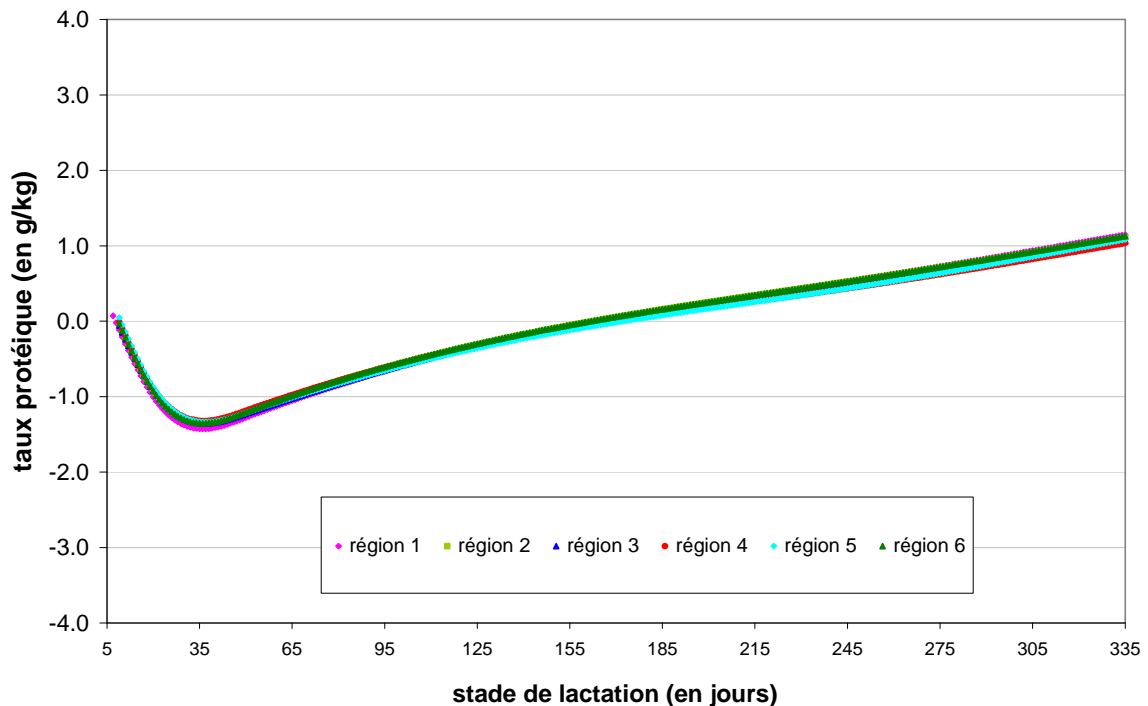


Figure 10. Effet de la région de la classe d'âge au vêlage 30-31 mois en 1<sup>ère</sup> lactation sur le taux protéique (en g/kg) en fonction du stade de lactation (en jours) en race Normande.

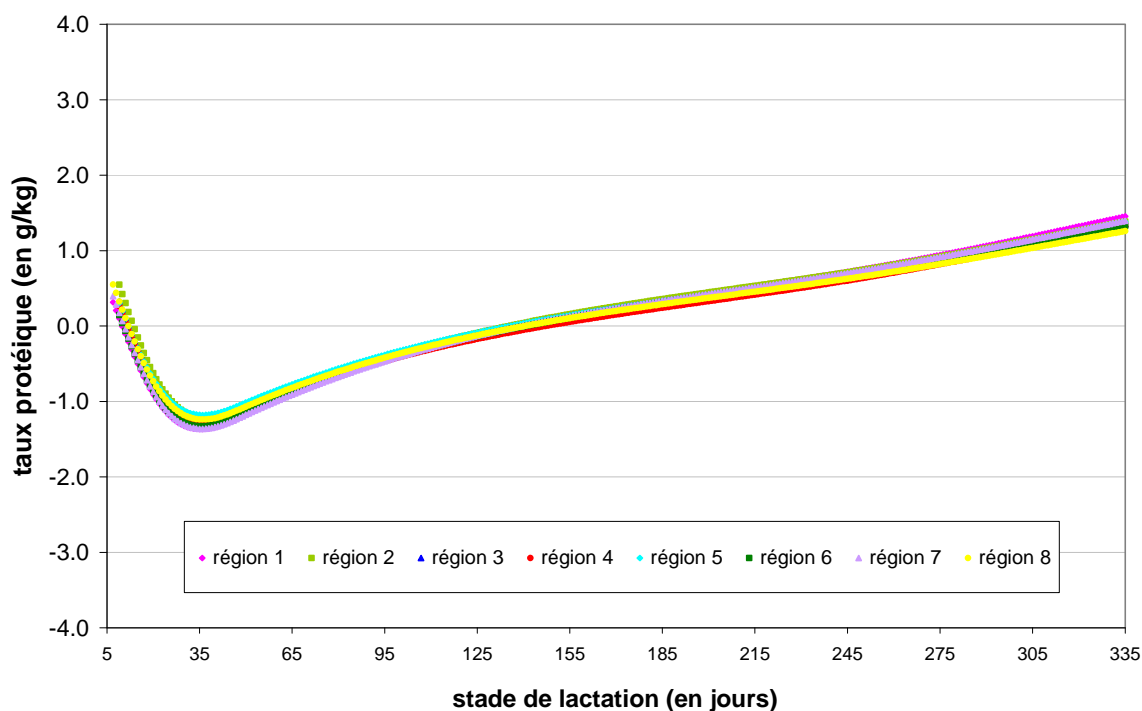


Figure 11. Effet de la région de la classe de durée de tarissement 46-60 jours en 2<sup>ème</sup> lactation sur le taux protéique (en g/kg) en fonction du stade de lactation (en jours) en race Holstein.

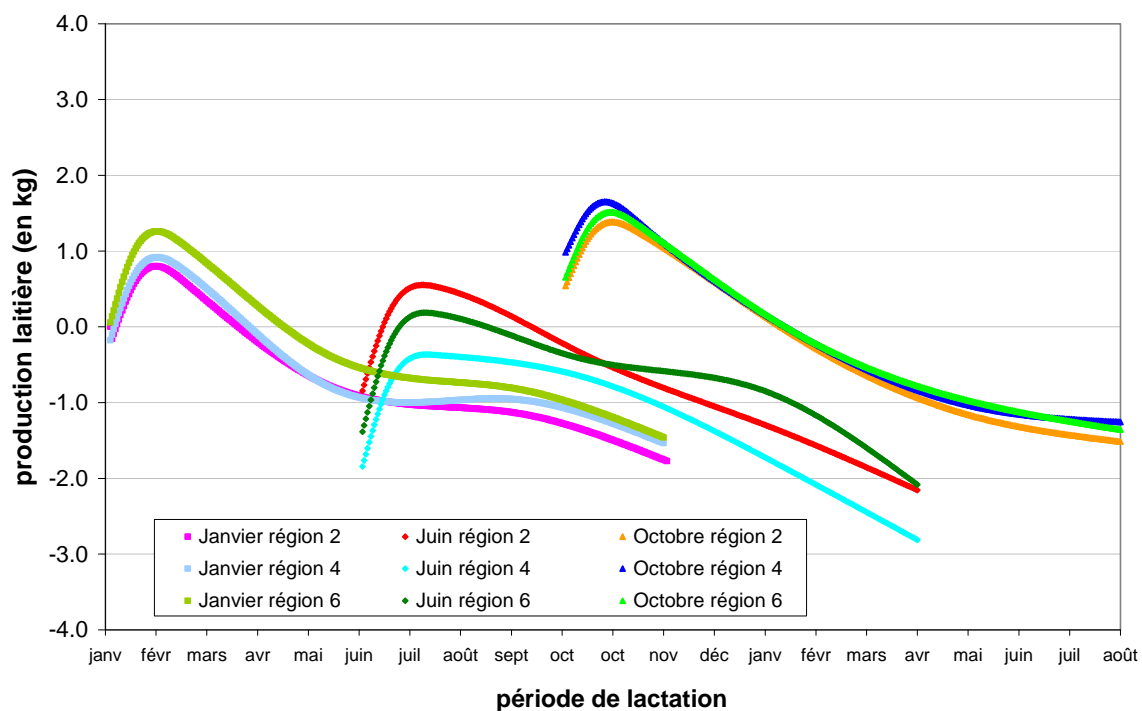


Figure 12. Effet du mois de vêlage (janvier, juin et octobre) sur la production laitière (en kg) en fonction de la période de lactation suivant la région (2, 4 ou 6) sur les 1<sup>ères</sup> lactations de race Holstein.

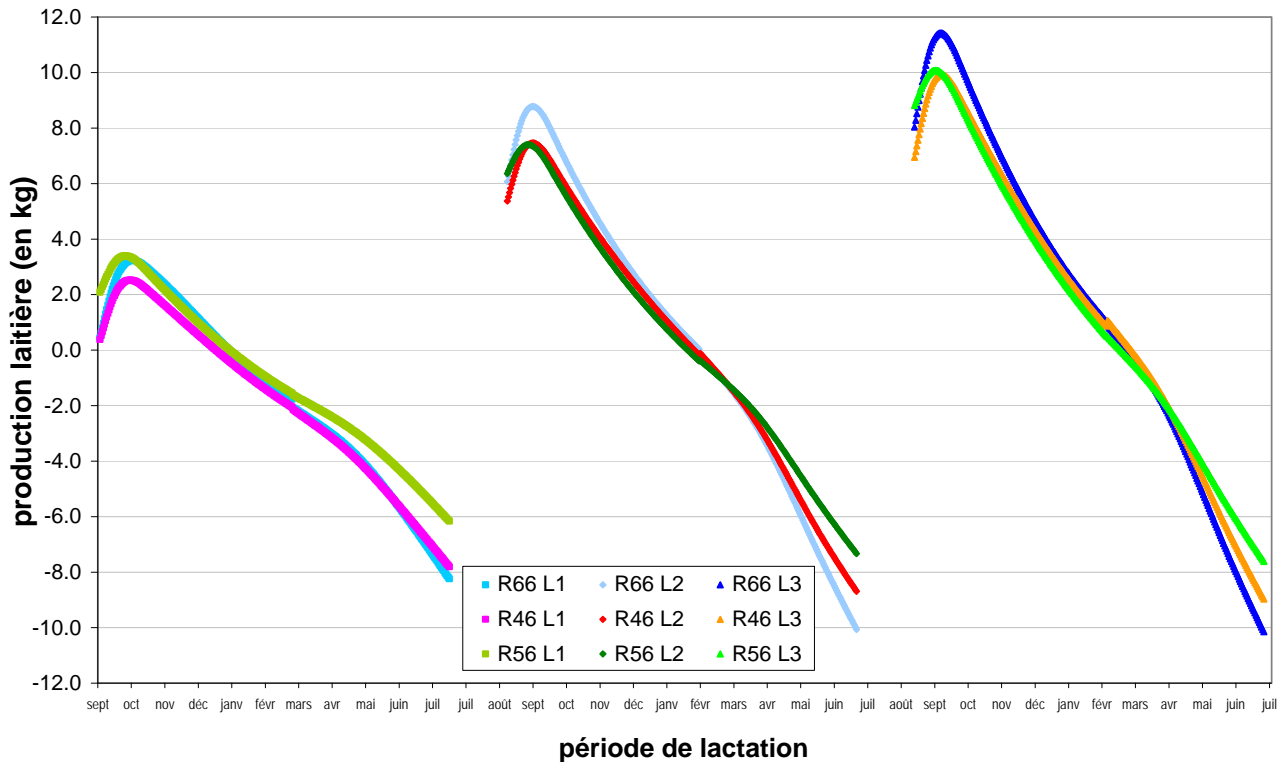


Figure 13. Courbes de lactations obtenues en cumulant les effets du mois, de l'âge au vêlage de la durée de tarissement et de la gestation pour des vaches de race Holstein (R66 région 2), Montbéliarde (R46 région 3) et Normande (R56 région 2) au cours de leurs 3 premières lactations (L1, L2 et L3). (vêlage septembre ; âge au vêlage : L1 = 30 mois, L2 = 42 mois, L3 = 54 mois ; durée tarissement entre lactation = 45 jours ; intervalle vêlage – insémination fécondante = 80 jours).

## Annexe

Définition des régions françaises prises en compte dans la constitution des effets liés au mois de vêlage, à l'âge au vêlage, à la durée de tarissement et à l'effet de la gestation en fonction de la race. Elles sont dérivées des régions décrites par Bonaiti *et al.* (1990) avec un regroupement de régions en race Normande afin d'obtenir un nombre de données suffisant pour l'estimation des effets. Elles sont définies à partir des régions du Groupe d'Economie Bovine de l'Institut de l'Élevage, constituées de regroupement de petites régions agricoles de l'Institut National de la Statistique et des Etudes Economiques (INSEE).

Race Holstein : 8 régions



Race Normande : 6 régions



Race Montbéliarde : 3 régions



## **Partie 4.2**

# **L'effet troupeau-jour de contrôle**



# Prédiction de l'effet troupeau-jour de contrôle et applications potentielles pour la conduite du troupeau

H. Leclerc<sup>1,2†</sup> et V. Ducrocq<sup>1</sup>

<sup>1</sup>INRA, UR337, Station de Génétique Quantitative et Appliquée, 78352 Jouy-en-Josas

<sup>2</sup>Institut de l'Élevage, Département de Génétique, INRA-SGQA, 78352 Jouy-en-Josas  
Courriel : [helene.leclerc@jouy.inra.fr](mailto:helene.leclerc@jouy.inra.fr)

## Résumé

Le développement de modèles d'évaluation génétique basés sur les données de contrôles élémentaires offre de nombreuses perspectives de valorisation à travers le développement d'outils d'appui technique pour les élevages, par rapport aux modèles basés sur les données lactation. Cet aspect, même s'il a initialement constitué une des motivations pour le développement des modèles contrôles élémentaires, a fait l'objet de peu d'études jusqu'à récemment. Au cours des 5 dernières années, deux méthodes ont toutefois été proposées afin de prédire l'effet troupeau – jour de contrôle (HTD pour Herd – Test Day), qui caractérise l'effet d'environnement et principalement l'effet de la conduite du troupeau le jour du contrôle. La première est basée sur une décomposition de l'effet HTD en une somme de termes prédictibles. La seconde repose sur la méthodologie du modèle mixte avec des effets fixes définis intra troupeau et des effets aléatoires intra troupeau - année. La méthode dite de Holt-Winters basée sur l'observation de séries temporelles paraît également adaptée à l'objectif recherché. Cette étude s'est attachée à comparer ces trois méthodes de prédiction de l'effet HTD sur la base des effets observés antérieurement pour le lait, les matières grasses et protéiques ainsi que les taux butyreux et protéiques. Deux applications peuvent découler de ce travail. D'une part, la prédiction de l'effet HTD peut aider à cerner la production des mois futurs et ainsi permettre d'adapter la conduite du troupeau en fonction des besoins de l'exploitation agricole. D'autre part, l'effet HTD peut être analysé comme un élément de suivi technique du troupeau permettant de détecter d'éventuels problèmes grâce à la comparaison de la valeur prédite et de la valeur réalisée. Exprimé en écart par rapport à un groupe de référence, par exemple une moyenne départementale, l'effet HTD permet également de situer le niveau technique du troupeau, de repérer ses points forts ou faibles et ainsi de mieux chiffrer les marges de progrès techniques possibles.

---

† Article en vue d'une soumission à Journal of Dairy Science

## Introduction

Jusqu'à maintenant, les outils d'appui technique ont principalement été développés par les organismes de contrôle de performance pour valoriser les données individuelles collectées et apporter des conseils aux élevages adhérents. Dans la plupart des pays, ces données sont parallèlement exploitées dans le cadre des évaluations génétiques des reproducteurs portant sur les caractères de production laitière : quantité de lait, taux butyreux et protéique, matières grasse et protéique, mais aussi santé de la mamelle à travers la numération cellulaire. Toutefois, cette valorisation « génétique » étant réalisée en France à partir de données agrégées sur l'ensemble de la lactation (une valeur par caractère), elle ne peut pas contribuer à enrichir les données de performances utilisées et être exploitée par le contrôle laitier, à l'exception des estimées d'effets « troupeau » (HY pour Herd – Year). En effet, l'effet HY est un co-produit de l'évaluation génétique sur les données « lactation » qui traduit l'influence de tous les facteurs de milieu communs à l'ensemble des vaches d'une même race produisant dans le même élevage au cours de la même année, indépendamment du niveau génétique, des caractéristiques individuelles de la vache et des autres facteurs de milieu intégrés dans le modèle d'évaluation. Il mesure globalement les conditions d'élevage dans le troupeau durant l'année laitière, qu'il s'agisse de contraintes pédo-climatiques ou de la conduite du troupeau par l'éleveur. Il est habituellement exprimé en écart à un groupe de référence qui est dans la mesure du possible une base régionale. Seegers et Journot (1992) ont montré l'intérêt de cet effet pour l'appui technique. Ils ont notamment mis en évidence l'effet prédominant de la conduite et du « savoir-faire » de l'éleveur sur l'effet HY par rapport au milieu de production, appréhendé à travers la typologie des élevages basée d'une part sur les contraintes pédo-climatiques et d'autre part sur le système alimentaire. L'effet HY est diffusé aux centres régionaux informatiques depuis Août 1991 (Balvay, 1992) dès lors que les troupeaux disposent d'au moins 10 lactations prises en compte dans l'analyse.

Avec le développement des outils informatiques, l'intérêt pour les évaluations génétiques basées sur les performances élémentaires s'est accru (Swalve, 2000 ; Jensen, 2001). Les principaux pays laitiers, comme le Canada (Schaeffer *et al.*, 2000), l'Allemagne (Reinhardt *et al.*, 2002), les Pays-Bas (De Roos *et al.*, 2001), l'Italie (Samorè *et al.*, 2002), les pays nordiques (Lidauer *et al.*, 2006) se sont orientés vers cette méthodologie qui offre de nombreux avantages. Parmi les principaux, nous pouvons citer la prise en compte d'effets d'environnement spécifiques du jour du contrôle et la modélisation de nombreux effets sous forme de courbes dont la trajectoire varie suivant le stade de lactation. La meilleure prise en

compte des effets d'environnement permet d'améliorer la précision des valeurs génétiques estimées. En effet, à la place d'un groupe de contemporaines définissant l'effet moyen du troupeau sur une année particulière (HY) dans un modèle « lactation », le groupe de contemporaines est défini par un effet troupeau – jour de contrôle (HTD pour Herd – Test day) dans un modèle contrôles élémentaires. La plupart des recherches se sont concentrées sur les aspects en lien direct avec les évaluations génétiques tels que la modélisation des courbes de lactation, la définition du modèle, l'estimation des paramètres génétiques... Quelques études ont cependant permis de montrer que les valorisations du modèle contrôles élémentaires à des fins d'appui technique étaient très prometteuses (Everett *et al.*, 1994 ; Caccamo *et al.*, 2008) notamment grâce à l'exploitation des estimées de l'effet HTD (Mayeres *et al.*, 2004 ; Koivula *et al.*, 2007). Celui-ci constitue un bon indicateur de la conduite du troupeau, puisque les données sont corrigées pour l'ensemble des autres effets du modèle, notamment le niveau génétique ainsi que les effets dits « environnementaux » comme le stade de lactation, l'âge et le mois de vêlage. L'effet HTD est donc un élément caractérisant l'environnement à court terme, notamment l'effet de l'alimentation le jour de contrôle, de la météorologie, de l'état sanitaire général du troupeau. Pour ces raisons, il est beaucoup plus précis que l'effet HY qui amalgamait l'ensemble des effets au cours de l'année.

Dans le cadre du développement de l'évaluation génétique sur les contrôles élémentaires, nous cherchons à enrichir les données du contrôle laitier afin de fournir des éléments permettant le déploiement de nouveaux outils d'appui technique. Dans ce contexte, nous nous sommes intéressés à deux méthodes proposées pour la prédiction de l'effet HTD (Mayeres *et al.*, 2004 ; Koivula *et al.*, 2007). Une autre approche de prédiction, basée sur la méthode d'extrapolation de Winters (1960) (aussi appelé Holt-Winters) disponible dans SAS à travers la procédure Forecast (SAS Institute, 2004) a également été testée. L'objectif de cette étude a été de comparer les différentes approches proposées pour la prédiction de l'effet HTD. Nous avons également cherché à mettre en évidence l'intérêt du développement d'outils pour le suivi technique des troupeaux d'une part à partir de la prédiction des performances à venir et d'autre part à partir de l'analyse *a posteriori* des écarts observés entre HTD prédits et HTD réels.

## Matériels et Méthodes

### Données

Pour réaliser cette étude, nous avons utilisé les effets HTD estimés pour les groupes de contemporaines, définis comme l'ensemble des animaux de lactation 1 à 3 contrôlés le même jour dans un même élevage, obtenus au cours de l'évaluation génétique sur les données de production journalière. Les évaluations génétiques considérées ici ont été réalisées à partir de performances journalières sur le lait, la matière grasse (MG) et protéique (MP), le taux butyreux (TB) et protéique (TP) de vaches Montbéliardes ayant vêlé entre septembre 1988 et août 2004 dans le département du Jura. Seules les données ayant un stade de lactation compris entre 7 et 335 jours ont été prises en compte. Au total, 2,4 millions de données élémentaires provenant de 135 743 vaches ont été utilisées. Ces données ont été obtenues au cours de 137 158 contrôles dans 1170 troupeaux. Dans le modèle, l'effet HTD est défini à partir d'un minimum de 5 performances.

En France, l'année laitière est définie de septembre de l'année  $N$  à août de l'année  $N+1$ . Afin de simplifier les notations, les données obtenues au cours de cette période sont considérées de l'année  $N$ .

### Modélisation de l'effet HTD

Comme décrit par Leclerc *et al.*, (2008), l'évaluation génétique française sur les contrôles élémentaires est réalisée en 2 étapes. Au cours de la 1<sup>ère</sup> étape, les performances sont précorrignées pour l'effet du mois de vêlage, de l'âge au vêlage, de la durée de tarissement et de la gestation sur les courbes de lactation définies par région et rang de lactation. Dans la 2<sup>ème</sup> étape, les performances précorrignées sont décrites par un effet HTD, par 3 constantes pour tenir compte de l'effet de l'année sur le niveau moyen des courbes pour le mois, l'âge au vêlage, la durée de tarissement, ainsi que par les effets génétiques additifs, les effets d'environnements permanents et les effets aléatoires troupeau – année. Cette étude se base sur l'analyse et la prédiction de l'effet HTD estimé au cours de la 2<sup>ème</sup> étape, soit un effet corrigé pour l'ensemble des effets fixes et aléatoires du modèle comme l'effet du mois de vêlage, de l'âge au vêlage, de la durée de tarissement et de la gestation sur le stade de lactation mais aussi les effets dépendant directement de l'animal tel que l'effet génétique.

Afin d'exploiter les informations relatives à la conduite et au contexte de production sous-jacents à l'effet HTD pour l'appui technique, Mayeres *et al.* (2004) ont proposé de le

décomposer dans le modèle d'évaluation génétique en 3 éléments de sorte qu'il soit directement utilisable à des fins de prédiction. L'effet fixe HTD du modèle d'évaluation génétique est ainsi décrit comme la somme d'un effet troupeau – année (HTY), d'un effet troupeau – mois de contrôle (HTM) et enfin d'un effet correspondant à la spécificité de l'effet troupeau avec le jour de contrôle (HTDr) qui peut aussi être vu comme l'effet de l'interaction entre le mois de contrôle et l'année pour ce troupeau. Cette décomposition résulte de la volonté de prendre en compte les tendances saisonnières observées sur l'effet HTD. La somme des effets HTY et HTM est la production attendue pour un mois de contrôle donné, l'effet HTDr étant en espérance nulle. Pour permettre d'intégrer une évolution des tendances saisonnières, Mayeres *et al.* (2004) ont défini l'effet HTM au sein de périodes de 5 ans. Contrairement à ces auteurs où l'effet HTD inclus dans le modèle d'évaluation génétique était directement remplacé par les 3 termes décrits précédemment, nous avons dans cette étude décomposé *a posteriori* l'effet HTD estimé lors de l'évaluation génétique. Afin d'améliorer la prise en compte de l'évolution possible des tendances saisonnières au cours du temps, la période sur laquelle se basent les prédictions n'a pas été définie à travers des groupes fixes d'années mais comme une période mobile évoluant chaque année. Dans la suite de cette étude, l'appellation « Mayeres » est utilisée pour désigner cette méthode.

La seconde méthode d'analyse des HTD utilisée dans cette étude est la méthode de Koivula *et al.* (2007). L'effet HTD n'y est pas analysé en tant que tel, mais à travers son écart par rapport à la moyenne des effets HTD du pays pour le mois de contrôle considéré. Grâce à ce mode d'expression, qui gomme l'effet des fluctuations saisonnières de production, un troupeau qui suit les tendances moyennes aura un effet « conduite » décrit par une ligne horizontale mois après mois. L'écart de l'effet HTD pour l'année à venir est prédit à partir de la moyenne des écarts observés lors des 5 précédentes années. Dans le modèle retenu, cette écart est modélisé à partir de la somme d'une tendance linéaire mensuelle et de 3 fonctions sinus sur le mois de contrôle traitées d'une part comme effets fixes intra troupeau et d'autre part comme effets aléatoires intra troupeau – année. Deux des fonctions sinus ont une amplitude annuelle avec un décalage de phase de 2 mois, la dernière fonction ayant une amplitude bisannuelle. Les composantes de variance des effets aléatoires ont été estimées préalablement par un maximum de vraisemblance restreinte (REML) en utilisant un algorithme AI (Druet, 2006). La dénomination « Koivula » fait référence aux résultats obtenus avec cette méthode.

Enfin, la dernière méthode proposée pour réaliser des prévisions repose sur l'utilisation de méthodes de lissages exponentiels à partir de l'observation d'une série temporelle telle que

développée dans une procédure SAS issue du « package » intitulé Forecast (SAS, 2004). Cette procédure permet de prédire des données à partir d'extrapolations basées uniquement sur la date et les valeurs prises par la série dans le passé. Parmi les méthodes d'extrapolation disponibles, nous avons retenu la méthode de Holt-Winters, généralisation de la méthode de lissage exponentiel double, qui est utilisée lorsque la série présente à la fois une tendance linéaire et une saisonnalité (Winters, 1960). Cette méthode possède deux variantes : l'une multiplicative et l'autre additive. Une étude préliminaire a permis de montrer que le modèle additif présentait globalement des résultats légèrement meilleurs au modèle multiplicatif. Dans le modèle additif, la méthode Holt-Winters estime la  $p^{\text{ème}}$  prédiction de la performance  $F_{t+p}$  à partir de la date  $t$  du dernier contrôle connu de la série temporelle  $y_1, \dots, y_t$  comme :

$$F_{t+p} = a_t + p \cdot b_t + S_{t-T+p} + \varepsilon_t$$

- où  $a_t$  est la constante obtenue au temps  $t$ ,  
 $b_t$  est la tendance linéaire obtenue au temps  $t$ ,  
 $S_t$  est la composante saisonnière de période  $T$  obtenue au temps  $t$ ,  
 $t$  représente le temps (en mois),  
 $T$  est le nombre de périodes constituant l'effet saisonnier,  
 $p$  est le numéro de la période pour laquelle une prédiction est souhaitée  
 $\varepsilon_t$  est un terme d'erreur.

Les formules récursives de mise à jour des paramètres  $a_t$ ,  $b_t$  et  $S_t$  sont de la forme :

$$\hat{a}_t = \alpha (y_t - \hat{S}_{t-T}) + (1-\alpha) [\hat{a}_{t-1} + \hat{b}_{t-1}],$$

$$\hat{b}_t = \beta (\hat{a}_t - \hat{a}_{t-1}) + (1-\beta) \hat{b}_{t-1},$$

$$\hat{S}_t = \gamma (y_t - \hat{a}_t) + (1-\gamma) \hat{S}_{t-T}.$$

- où  $\alpha$  est le paramètre de lissage de la constante,  
 $\beta$  est le paramètre de lissage de la tendance linéaire,  
 $\gamma$  est le paramètre de lissage de l'effet saisonnier.

Pour ces 3 paramètres de lissage, des valeurs élevées accorderont un poids important aux données les plus récentes, contrairement à des poids faibles qui privilégieront les données anciennes. Les valeurs courantes sont généralement situées entre 0.1 et 0.3 d'après la documentation sur cette procédure (SAS, 2004). Un minimum de 15 contrôles dans le troupeau est requis dans cette étude pour prédire les effets HTD à venir. Dans la suite de cette étude, les résultats provenant de cette méthode sont intitulés « Addwinters ».

### Eléments de comparaison

Les trois méthodes de prédiction de l'effet HTD (Mayeres, Koivula et Addwinters) ont été appliquées sur les 5 caractères laitiers (quantité de lait, MG, MP, TB et TP). Deux variables ont été comparées : d'une part, l'effet HTD *sensu stricto* et d'autre part l'écart entre l'effet HTD pour un troupeau et un mois de contrôle donné et la moyenne des HTD observés sur le département pour ce mois de contrôle, variable d'analyse suggérée par Koivula *et al.* (2007). Outre le nom de la méthode ayant servi à obtenir ces résultats, les deux variables seront distingués dans la suite par un « *d* » précédant l'appellation lorsque le caractère est exprimé en écart par rapport à la moyenne départementale soit *dMayeres*, *dKoivula*, *dAddwinters*.

Les comparaisons entre méthodes ont été réalisées en confrontant les prédictions faites pour chacune des méthodes avec les effets HTD « réels », estimés *a posteriori*, sur la base de 8 années (entre 1996 et 2003). Pour être pris en compte dans l'analyse, le troupeau devait disposer d'un minimum de 25 dates de contrôle au cours des 3 années précédant la prédiction, dont un minimum de 8 sur la dernière année. Cette dernière condition permet d'exclure les troupeaux n'adhérant plus au moment de la prédiction aux services du Contrôle Laitier. Enfin, pour faciliter l'analyse des résultats, seuls les effets ayant pu être prédits avec les 3 méthodes ont été pris en compte. Les méthodes de Mayeres et Koivula ont initialement été décrites en utilisant un pas de temps de 5 ans pour prédire la campagne suivante. Afin de réduire le temps nécessaire à la prise en compte des données de nouveaux troupeaux adhérents et d'accroître la flexibilité des prédictions, nous avons également étudiés ces méthodes en prédisant les variables sur la base des 3 années précédentes. Les résultats présentés sont donc issus de 60 182 effets HTD pour les prédictions basées sur les 5 années antérieures et 59 673 effets HTD lorsque cette période est limitée à 3 ans.

Outre les analyses graphiques, 4 critères ont été retenus pour juger des performances de chacune des méthodes : les corrélations entre l'effet prédit et l'effet « réel », le pourcentage de cas où la prédiction s'écartait significativement de l'effet « réel » et enfin, les moyennes et écarts-types des erreurs de prédictions. Concernant la détection des mois de contrôle s'écartant significativement de la prédiction, le critère retenu dans cette étude est celui proposé par Koivula *et al.* (2007). L'ensemble des effets HTD présentant un écart à la prédiction supérieur à 1 écart-type de l'effet HTD pour le troupeau analysé est considéré comme déviant. Statistiquement, l'intervalle de confiance formé par 1 écart-type inclut 68% des données. Ces auteurs ont volontairement choisi un seuil d'alerte très strict. Afin

d'encourager l'identification des facteurs de causalité de ces variations inattendues, les contrôles concernés peuvent être visuellement identifiés à travers des graphiques retraçant l'évolution de l'effet HTD pour un troupeau donné dans l'application Web Finlandaise dédiée à l'appui technique des troupeaux laitiers.

## Résultats

Quelles que soient les méthodes de prédiction utilisées, elles se basent sur la valorisation des données antérieures et notamment des tendances qui peuvent s'en dégager pour prédire l'avenir. Elles supposent donc une certaine régularité dans les cycles observés ou du moins une évolution progressive. Les variables laitières analysées dans le cadre de cette étude semblent bien se prêter à ces modélisations. En effet, les tendances saisonnières observées par Mayeres *et al.*, (2004) sur l'effet HTD sont aussi constatées sur les données françaises, comme en témoigne la figure 1 représentant la moyenne mensuelle des effets HTD observés sur les troupeaux du Jura entre septembre 1989 et août 2004 pour la quantité de lait. Le niveau de production moyen sur l'année (HTY), corrigé pour tous les autres effets, y compris génétiques, a peu évolué sur cette période, puisqu'il était de 20.9 kg en 1989 et de 21.7 kg en 2003, avec un maximum de 22.0 kg en 1997. La variation mensuelle observée suit assez bien le photopériodisme avec un minimum de près de 3.0 kg de lait en novembre, et un maximum de 3.2 kg de lait en mai. La variation spécifique du mois de contrôle avec l'année (HTDr) est assez faible, avec un écart-type de 0.37 kg. L'impact le plus négatif est observé pour août 2003, été de canicule en France avec - 1.2kg, et le plus favorable en juin 1991, printemps exceptionnel dans cette région pour la qualité et la quantité d'herbe produite avec + 1.0 kg par jour de lactation.

Au cours d'une analyse préliminaire de la méthode proposée par Winters (1960), différents effets saison ont été testés. Un effet mois ou trimestre sont apparus d'emblée être les plus adéquats. Un effet semestre ne permettait pas de s'ajuster suffisamment aux variations observées. D'autre part, un effet quinzaine ne pouvait pas être déterminé, les contrôles ayant en moyenne lieu entre 10 et 11 fois par an. En affinant la comparaison entre les différents effets saison envisageables, il est apparu qu'un effet mois offrait une plus grande flexibilité. Ce dernier a donc été retenu pour la comparaison globale entre les 3 méthodes. La procédure SAS sur laquelle se base le modèle Addwinters propose l'utilisation de paramètres de lissage standards. Bien que les résultats obtenus dans ce cas soient corrects, l'utilisation de



pondérations choisies librement permet toutefois d'améliorer les résultats. Le modèle retenu et présenté dans cette étude a été obtenu avec les poids de 0.10 pour la constante, 0.02 pour la tendance linéaire et 0.22 pour l'effet de la saison pour la quantité de lait. Une pondération élevée accordera un poids important aux données les plus récentes, contrairement à un poids faible qui privilégiera les données anciennes. Les valeurs courantes sont généralement situées entre 0.1 et 0.3 d'après la documentation sur cette procédure (SAS, 2004). Ainsi, dans le modèle proposé, l'évolution générale de l'effet HTD est lente alors que les variations saisonnières sont sujettes à des évolutions plus rapides.

La représentation graphique de la quantité de lait prédite selon les 3 méthodes testées en comparaison à l'effet HTD « réel », estimé *a posteriori*, montre une bonne cohérence entre les différentes méthodes (corrélation autour de 0.95), comme l'illustre la figure 2. Les corrélations obtenues entre l'effet HTD « réel » et l'effet HTD prédit par chacune des méthodes varient entre 0.681 et 0.845 suivant le caractère laitier analysé, la méthode utilisée, et enfin la période sur laquelle se basent les prédictions (Tableau 1a). La quantité de lait est le caractère le plus facilement prédictible avec des corrélations autour de 0.84, quelle que soit la méthode de prédiction. Les quantités de matières présentent des corrélations variant entre 0.80 et 0.84 suivant la méthode employée alors que les taux présentent eux des corrélations nettement inférieures, surtout pour le TB. Ceci est probablement lié à la variabilité plus importante de ce caractère. Les corrélations obtenues avec la méthode Koivula sont meilleures que les 2 autres méthodes surtout pour la MG et la MP. Concernant la période sur laquelle se base les prédictions, l'impact est différent suivant les méthodes. Il est faible sur la méthode Addwinters qui ne dépend pas directement de cette période. La réduction de cette période pénalise légèrement la qualité des prédictions réalisées avec la méthode Mayeres. Au contraire, elle a tendance à avantager les prédictions faites avec la méthode Koivula, à l'exception des taux pour lesquels les tendances à long terme semblent à privilégier. Pour la prédiction de l'écart entre l'effet HTD d'un troupeau pour un mois de contrôle donné et la moyenne départementale des effets HTD de ce mois de contrôle ( $dMéthode$ ), nous aboutissons aux mêmes conclusions que précédemment sur les méthodes employées et l'influence de la période sur laquelle se basent les prédictions (Tableau 1b). Cependant, les corrélations obtenues sont variables suivant les caractères. Pour la quantité de lait, nous observons une baisse de la corrélation intra mois de l'ordre de 0.07 à 0.09 lorsque l'effet HTD est exprimé en écart par rapport à la moyenne départementale par rapport à l'effet HTD *sensu*

*stricto* (comparaison des tableaux 1a et 1b). Sur les MG et MP, les corrélations obtenues sont équivalentes. Par contre, sur les taux, les corrélations chutent de façon importante par rapport à celles obtenues sur l'effet HTD *sensu stricto* pour atteindre des valeurs comprises entre 0.510 et 0.591.

La qualité de la détection des mois de contrôles s'écartant significativement de la prédiction peut être mesurée à travers le nombre de cas où l'effet HTD « réel » est en dehors de l'intervalle de confiance à 68%. Les pourcentages figurant dans le tableau 2 montrent des variations importantes suivant le caractère. La quantité de lait présentent de faibles pourcentages par rapport au niveau attendu avec l'utilisation d'un seuil à 1 écart-type quelle que soit la méthode utilisée et la durée de la période sur laquelle se base les prédictions. Pour le TP et le TB, nous obtenons également de faibles pourcentages (18.1 – 21.6%) lorsque la période de référence est de 5 ans. Par contre, lorsque cette période est réduite à 3 ans, le pourcentage de données s'écartant significativement de la prédiction atteint des valeurs comprises entre 24.1 et 29.6%. A l'extrême, le pourcentage de données en dehors de l'intervalle de confiance de la prédiction peut atteindre 44.8% avec la MG. Les variations entre les méthodes suivent des tendances similaires à l'analyse des corrélations. Les différences liées à la durée de la période sur lesquelles se basent les prédictions sont toutefois plus marquées. Lorsque l'effet HTD est exprimé en écart à la moyenne départementale (*dMéthode*), les bons résultats observés sur le lait et les taux tendent à s'amenuiser. Pour les 5 caractères laitiers, les pourcentages observés varient entre 32.1 et 49.6% (résultats non montrés). Les erreurs moyennes de prédiction (Tableau 3) sont globalement proches de 0. Le signe négatif de la méthode Addwinters montre une tendance à sous-estimer les performances contrairement aux 2 autres méthodes. Les écarts-types restent cependant élevés, avec près de 1.9 kg de lait, 1.9 g/kg de TB et 1.0 g/kg de TP, les différences entre méthodes étant toujours en faveur de la méthode Koivula.

La figure 3 illustre la mise en évidence des variations inattendues de l'effet HTD, c'est-à-dire présentant une différence entre l'effet HTD prédit et l'effet HTD « réel » supérieure à un écart-type pour le troupeau considéré. Ces alertes poussent ainsi à s'interroger sur les raisons de ces variations, principalement lorsqu'il s'agit d'une chute.

## Discussion

L'objectif principal de cette étude était de comparer différentes méthodes de prédiction de l'effet HTD. Malgré des performances prédictives un peu inférieures aux autres méthodes de

prédictions testées, la méthode Mayeres présente l'avantage d'être très simple, facilement compréhensible et de ne nécessiter aucune information complémentaire. Compte tenu de son faible niveau de complexité, les résultats obtenus sont jugés satisfaisants. Elle demande toutefois une période de référence longue pour accroître la fiabilité des prédictions, ce qui limite son utilisation dans le cas de l'adhésion de nouveaux troupeaux. La méthode basée sur la procédure Forecast de SAS (2004) permet d'atteindre une qualité de prédiction satisfaisante, nécessitant seulement l'ajustement des paramètres de modélisation afin d'optimiser les prédictions. Cet ajustement, réalisé manuellement dans le cadre de cette étude, doit pouvoir être automatisé puisque la plage de variation des pondérations est limitée. Un autre intérêt de cette procédure est qu'elle est utilisable dès lors qu'un troupeau dispose d'au moins 15 contrôles. Parmi les 3 méthodes testées, la méthode Koivula est sans conteste la meilleure en terme de résultats obtenus. Elle nécessite cependant l'estimation préalable des composantes de la variance avec un logiciel spécifique pour disposer des paramètres nécessaires à la procédure Mixed de SAS (2004), ce qui complique un peu son utilisation. L'utilisation de la procédure Mixed de SAS est également informatiquement plus contraignante que les méthodes précédentes. Le choix de la méthode doit donc être raisonné suivant la taille du fichier et la fréquence de la prédiction.

La prédiction de l'effet HTD peut faire l'objet de 2 principales applications : d'une part en exploitant son caractère prospectif, et d'autre part en l'utilisant comme outil d'analyse afin de suivre et étudier l'évolution de la conduite du troupeau. Sur le premier aspect, l'effet HTD est un élément majeur de la performance puisque dans le modèle d'évaluation, il joue le rôle de moyenne. En additionnant l'effet HTD prédit et les autres effets (ceux décrivant l'environnement tels que la région, l'année, l'âge au vêlage, le mois de vêlage... ainsi que ceux dépendant de l'animal comme l'effet génétique) il est possible de prédire une performance. En utilisant des effets HTD prédits suivant la méthode éponyme, Mayeres *et al.* (2004) obtenaient des corrélations entre production « réelle » et production prédite variant entre 0.786 et 0.821 suivant le rang de lactation pour la quantité de lait et un écart type d'erreur de prédiction oscillant entre 3.12 et 4.43 kg. Ces auteurs montraient qu'il était ainsi possible de prédire les performances attendues pour un animal lors des contrôles futurs. Le développement d'outils similaires à l'échelle du troupeau permettrait l'utilisation de divers leviers techniques comme l'adaptation de la ration alimentaire, la modulation de la durée de tarissement, ou la décision d'effectuer ou non d'éventuelles réformes de manière à ajuster la production au contingentement défini par les quotas.

Contrairement à l'effet troupeau disponible à partir des évaluations génétiques sur les données lactation, l'effet HTD inclus dans le modèle contrôles élémentaires est le reflet des conditions spécifiques au jour du contrôle. L'évolution de l'effet HTD mois après mois permet donc de réaliser un suivi technique de l'élevage. L'écart entre l'effet HTD et la valeur moyenne d'un groupe de référence (le département dans cette étude) permet de juger du niveau de performance du troupeau, car l'effet du groupe de référence tient compte des variations saisonnières, par exemple liées au climat ou aux cycles alimentaires. De plus, ce mode d'expression facilite la détection et l'interprétation des déviations observées, notamment des éventuelles périodes de carence. Les graphiques de la figure 4 représentent quelques exemples des différentes situations rencontrées dans le département du Jura, d'une part sur la quantité de lait (colonne de gauche) et d'autre part sur le taux protéique (colonne de droite). S'appuyant sur les niveaux obtenus par les 5% et 25% meilleurs et moins bons troupeaux, il est facile de situer le niveau de conduite du troupeau par rapport aux autres troupeaux du groupe de référence. Ainsi, le graphique A1 de la figure 4 montre un troupeau conduit de façon stable, dont le profil suit les variations saisonnières moyennes. Il bénéficie d'un bon niveau technique puisqu'il est situé entre les 25% et 5% meilleurs troupeaux du département, ce qui se traduit par une production supérieure de 4 kg de lait par jour et par vache en moyenne. Le graphique A2 montre au contraire un troupeau d'un niveau faible avec en moyenne -4 kg de lait par rapport à la moyenne départementale sur la période septembre 1998 – août 2004. La conduite du troupeau est stable d'une année sur l'autre, comme en témoignent les faibles différences entre les prédictions réalisées avec la méthode Koivula et les effets HTD réels, observés *a posteriori*. La cyclicité observée des écarts à la moyenne départementale montre cependant que la conduite hivernale est très pénalisante pour le niveau de production du troupeau (-6 à -8 kg), alors que la conduite estivale fluctue autour de -2 kg. A l'opposé, le graphique A3 montre un troupeau se situant dans la moyenne pendant la période d'alimentation estivale et dans les 25% meilleurs au cours de la phase hivernale. Les marges de progrès techniques sont ainsi clairement visualisées et les efforts d'appui technique peuvent alors se concentrer sur la période identifiée. Sur le TP, le graphique B1 montre un niveau stable à l'exception de la période septembre 2003 – février 2004, où le troupeau chute d'un point de TP comparativement à la moyenne départementale. Cette variation par rapport au groupe de référence témoigne d'un impact plus marqué de l'été caniculaire dans ce troupeau. L'analyse de la qualité des ressources fourragères aurait certainement pu constituer un élément de réponse. Comme pour la quantité de lait, les graphiques B2 et B3 montrent des troupeaux opposés sur la cyclicité des écarts par rapport au groupe de référence avec d'une

part un troupeau supérieur à la moyenne d'environ 1 g/kg de TP sur la période estivale et négatif sur la période hivernale (graphique B2) et d'autre part, un troupeau se situant dans les 5% meilleurs sur la période hivernale et au contraire dans les 25% plus faibles sur la période estivale (graphique B3).

Pour affiner l'analyse, il est envisageable de définir des groupes de référence à partir d'une classification typologique des exploitations basée sur la région pédo-climatique et le système alimentaire, comme cela avait été proposé par Seegers et Journot (1992). Ainsi, les différences observées sont systématiquement imputables à la conduite du troupeau. Ces indicateurs permettraient de juger des forces et faiblesses d'un troupeau en tenant compte de l'environnement de production, sans remettre en cause son système d'alimentation, et de chiffrer clairement les marges de progrès technique. L'objectif est de donner, à l'éleveur et à son technicien, l'ensemble des éléments permettant d'améliorer la conduite du troupeau grâce à la maîtrise des facteurs techniques, reflet du « savoir-faire ».

### **Remerciements**

Cette étude a été réalisée dans le cadre de l'Unité Mixte Technologique (UMT) INRA - Institut de l'Élevage sur l'évaluation génétique des bovins. Le ministère de l'agriculture est remercié pour avoir contribué à son financement à travers l'action innovante UTILEG. Minna Koivula est également sincèrement remerciée pour les explications fournies sur la méthode éponyme.

### **Références**

Balvay, B. 1992. L'effet troupeau : définition, caractéristiques et propositions d'utilisation. Compte rendu de l'Institut de l'Élevage N°1951, 9p.

Caccamo, M., R.F. Veerkamp, G. De Jong, M.H. Pool, R. Petriglieri, and G. Licitra. 2008. Variance components for test-day milk, fat, and protein yield and somatic cell score for analyzing management information. *J. Dairy Sci.* 91, 3268-3276.

De Roos, A.P.W., A.G.F. Harbers, and G. De Jong. 2001. Random regression test-day model in the Netherlands. *Interbull Bull.* 27, 155-158.

Druet T., and V. Ducrocq. 2006. Innovations in software packages in quantitative genetics. Proc. World Congress on Genetic Applied to Livestock Production, Belo Horizonte, Brazil. Communication 27-10.

Everett, R.W., F. Schmidt, and L.H. Wadell. 1994. A test-day model for monitoring management and genetics in dairy cattle. *J. Dairy Sci.* 77 (Suppl.1), 267. (Abstr.)

Jensen J. 2001. Genetic evaluation of dairy cattle using test-day models. *J. Dairy Sci.* 84, 2803-2812.

Koivula, M., J.I. Nousianinen, J. Nousianinen, and E.A. Mäntysaari. 2007. Use of herd solutions from a random regression test-day model for diagnostic dairy herd management. *J. Dairy Sci.* 90, 2563-2568.

Leclerc, H., D. Duclos, A. Barbat, T. Druet, and V. Ducrocq. 2008. Environmental effects on lactation curves included in a test-day model genetic evaluation. *Animal*, 2, 344-353.

Lidauer, M., J. Pedersen, J. Pösö, E.A. Mäntysaari, I. Strandén, P. Madsen, U.S. Nielsen, J.-Å. Eriksson, K. Johansson, and G.P. Aamand. 2006. Joint Nordic test-day model: evaluation model. *Interbull Bull.* 35, 103-107.

Mayeres, P., J. Stoll, J. Bormann, R. Reents, and N. Gengler. 2004. Prediction of daily milk, fat, and protein production by a random regression test-day model. *J. Dairy Sci.* 87, 1925-1933.

Reinhardt, F., Z. Liu, A. Bünger, L. Dopp, J. Bormann, and R. Reents. 2002. Impact of application of a random regression test-day model to production trait genetic evaluations in dairy cattle. *Interbull Bull.* 29, 103-107.

Samorè, A.B., P. Boettcher, J. Jamrozik, A. Bagnato, and A.F. Groen. 2002. Genetic parameters for production traits and somatic cell score estimated with a multiple trait random regression model in Italian Holstein. Proc. World Congress on Genetic Applied to Livestock Production, Montpellier, France. XXIX: 63-66.

SAS Institute. 2004. Base SAS and SAS/ETS<sup>®</sup> software. Version 9.1. Sas Inst. Inc., Cary NC.

Schaeffer, L.R., J. Jamrozik, G.J. Kistemaker, and B.J. Van Doormaal. 2000. Experience with a test-day model. *J. Dairy Sci.* 83, 1135-1144.

Seegers, J., et V. Journot. 1992. Effets « troupeau » : cohérence avec les indicateurs technico-économiques de l'appui technique OPTILAIT ; Relations avec les indicateurs techniques de la maîtrise du taux protéique en élevage. Essai Institut de L'Élevage – ONILAIT N°92042, 24p.

Swalve, H.H. 2000. Theoretical basis and computational methods for different test-day genetic evaluation methods. *J. Dairy Sci.* 83, 1115-1124.

Winters, P.R. 1960. Forecasting sales by exponentially weighted moving averages. *Management Sci.* 6, 324-342.

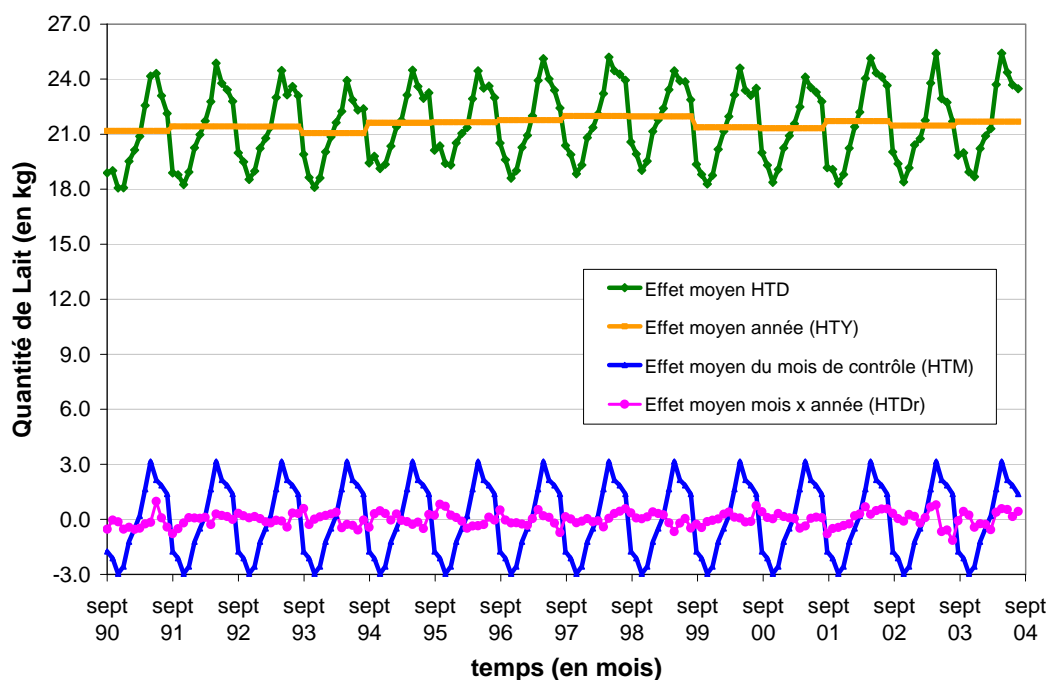


Figure 1. Moyenne des effets HTD observés sur la quantité de lait produite dans le département du Jura en race Montbéliarde entre septembre 1989 et Août 2004, et sa décomposition en 3 effets : l'effet de l'année (HTY), l'effet du mois de contrôle (HTM) et l'effet de l'interaction entre le mois et l'année (HTDr).

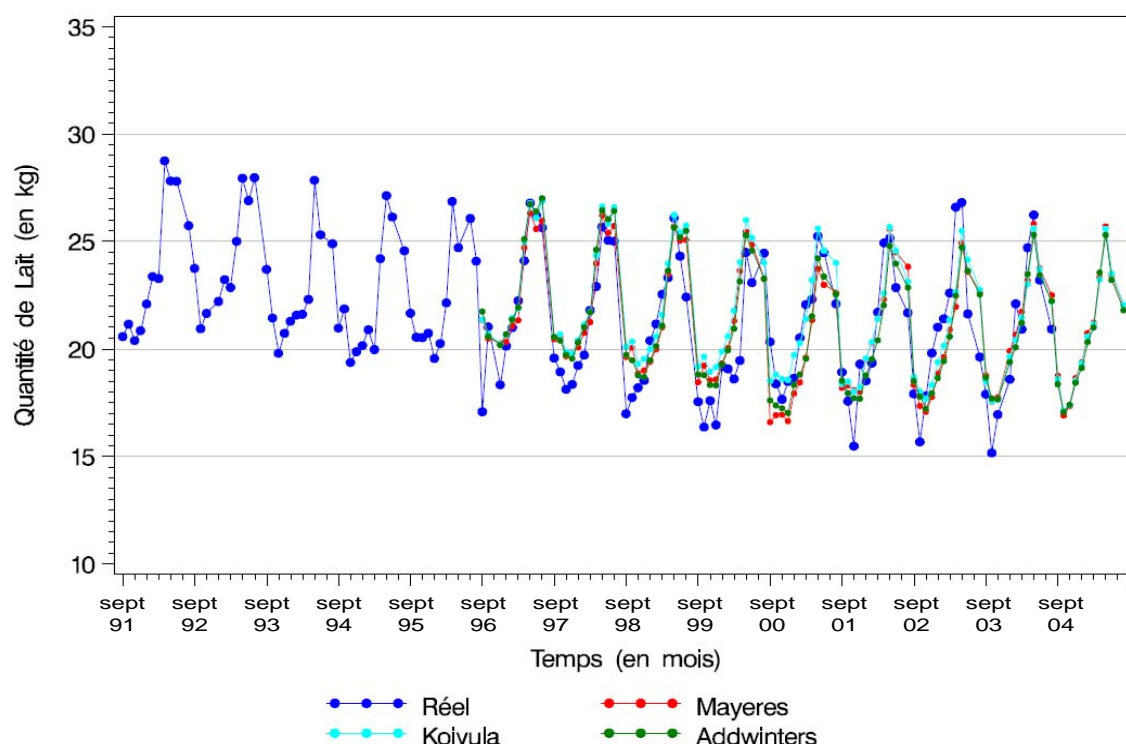


Figure 2. Comparaison des effets HTD prédits entre septembre 1996 et Août 2005 en race Montbéliarde sur la quantité de lait selon les méthodes Mayeres, Koivula et Addwinters sur la base des 5 années précédentes et de l'effet HTD « réel » estimé *a posteriori* pour un troupeau particulier entre septembre 1991 et Août 2004.



Chapitre 4 - Partie 4.2

Tableau 1a. Corrélation entre l'effet troupeau – jour de contrôle (HTD) prédit par la méthode Mayeres, Koivula et Addwinters avec l'effet HTD « réel » (*a posteriori*) sur la quantité de lait, matière grasse (MG), matière protéique (MP), taux butyreux (TB) et protéique (TP) sur la base des 5 ou 3 années antérieures (respectivement 60182 et 59673 effets HTD).

	Mayeres		Koivula		Addwinters	
	5 ans	3 ans	5 ans	3 ans	5 ans	3 ans
Lait (en kg)	0.840	0.833	0.844	0.845	0.841	0.841
MG (en g)	0.811	0.801	0.839	0.842	0.808	0.807
MP (en g)	0.805	0.796	0.838	0.841	0.805	0.803
TB (en g/kg)	0.706	0.681	0.723	0.708	0.713	0.700
TP (en g/kg)	0.758	0.747	0.771	0.762	0.763	0.755

Tableau 1b. Corrélation entre l'écart prédit entre l'effet troupeau – jour de contrôle (HTD) et la moyenne départementale par la méthode Mayeres, Koivula et Addwinters et l'écart réel entre l'effet HTD et la moyenne départementale effective (*a posteriori*) sur la quantité de lait, matière grasse (MG), matière protéique (MP), taux butyreux (TB) et protéique (TP) sur la base des 5 ou 3 années antérieures (respectivement 60182 et 59673 effets HTD).

	<i>d</i> Mayeres		<i>d</i> Koivula		<i>d</i> Addwinters	
	5 ans	3 ans	5 ans	3 ans	5 ans	3 ans
Lait (en kg)	0.766	0.756	0.754	0.755	0.766	0.767
MG (en g)	0.813	0.803	0.837	0.840	0.810	0.810
MP (en g)	0.796	0.787	0.826	0.830	0.796	0.795
TB (en g/kg)	0.530	0.510	0.530	0.522	0.539	0.531
TP (en g/kg)	0.591	0.576	0.591	0.586	0.589	0.584

Chapitre 4 - Partie 4.2

Tableau 2. Pourcentage d'effet troupeau – jour de contrôle (HTD) prédit par la méthode Mayeres, Koivula et Addwinters en dehors de l'intervalle défini par l'effet HTD « réel » (*a posteriori*)  $\pm 1$  écart-type sur la quantité de lait, matière grasse (MG), matière protéique (MP), taux butyreux (TB) et protéique (TP) sur la base des 5 ou 3 années antérieures (respectivement 60182 et 59673 effets HTD).

	Mayeres		Koivula		Addwinters	
	5 ans	3 ans	5 ans	3 ans	5 ans	3 ans
Lait (en kg)	18.1	19.8	16.8	18.0	18.2	19.0
MG (en g)	33.6	44.8	28.7	36.9	34.4	44.3
MP (en g)	30.3	39.9	25.0	31.6	30.6	39.1
TB (en g/kg)	21.8	29.6	20.1	26.6	21.6	28.2
TP (en g/kg)	19.3	25.8	18.1	24.1	18.8	25.5

Tableau 3. Moyenne et écart-type (entre parenthèse) des erreurs de prédictions entre les effets HTD prédits selon les méthodes Mayeres, Koivula et Addwinters sur la base d'une prédiction basées sur les 5 années antérieures et les effets HTD « réels » estimé *a posteriori* pour la quantité de lait, matière grasse (MG), matière protéique (MP), taux butyreux (TB) et protéique (TP).

	Mayeres		Koivula		Addwinters	
	Moyenne	(écart-type)	Moyenne	(écart-type)	Moyenne	(écart-type)
Lait (en kg)	0.043	(1.897)	0.082	(1.857)	-0.150	(1.909)
MG (en g)	0.32	(72.11)	3.52	(65.86)	-7.66	(73.49)
MP (en g)	0.16	(67.88)	1.88	(61.43)	-4.28	(68.57)
TB (en g/kg)	0.008	(1.916)	0.027	(1.836)	-0.029	(1.898)
TP (en g/kg)	0.040	(1.051)	0.151	(1.013)	0.102	(1.036)

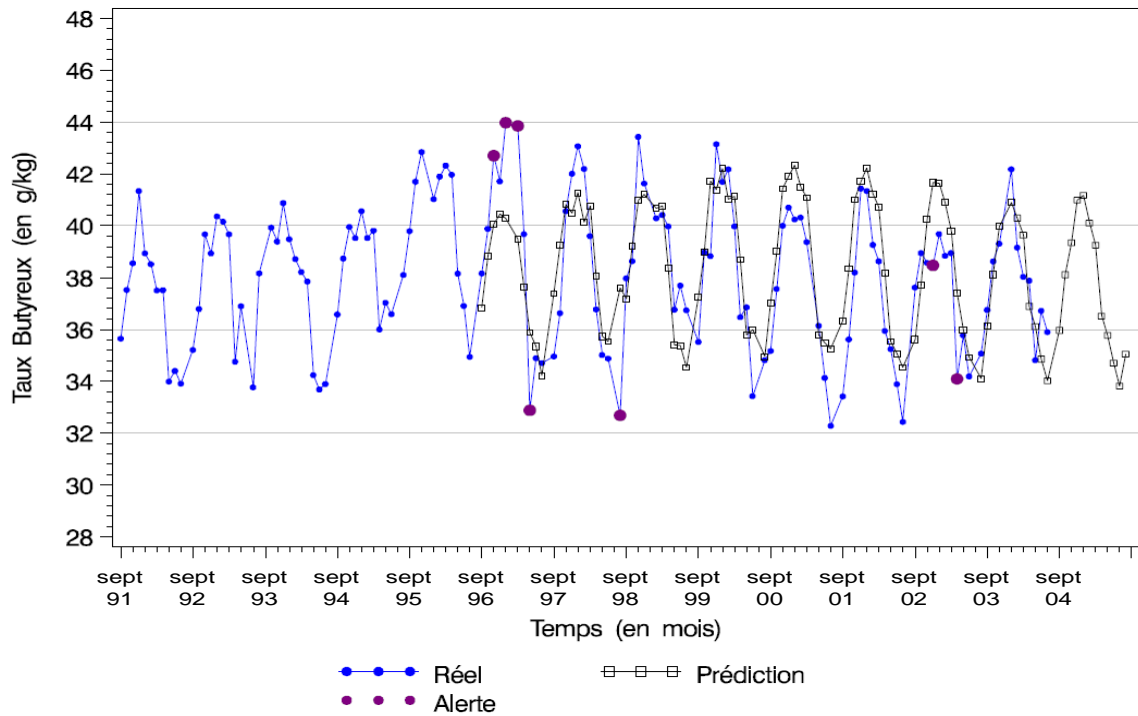


Figure 3. Comparaison des effets HTD prédits entre septembre 1996 et Août 2005 en race Montbéliarde sur le taux butyreux selon la méthode Koivula (Prédiction) sur la base des 5 années précédentes et des effets HTD « réels » estimés *a posteriori* pour un troupeau particulier entre septembre 1991 et Août 2004. Les contrôles déviant significativement de la prédiction sont mis en évidence par des alertes.

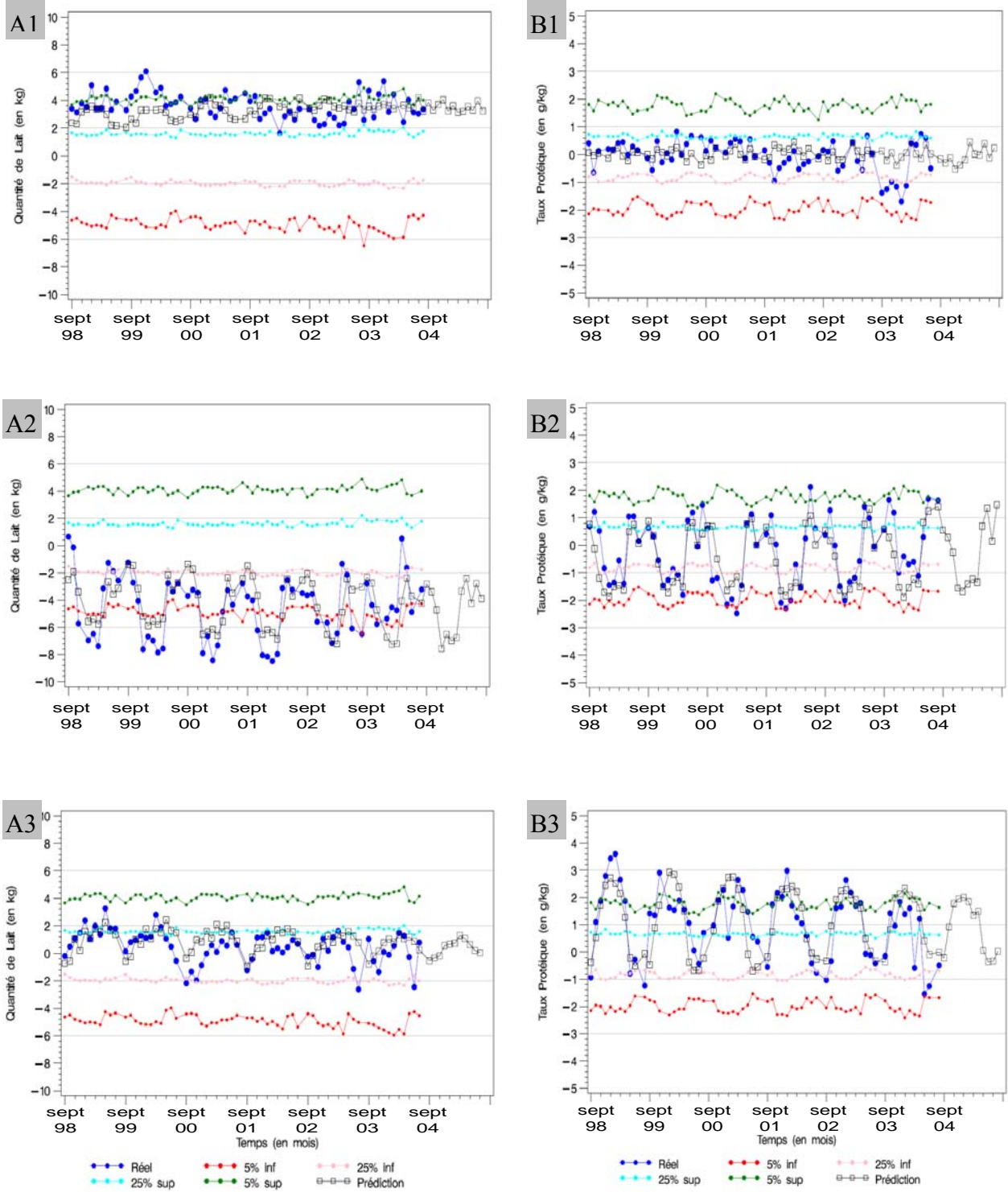


Figure 4. Effets HTD exprimés en écart à la moyenne départementale entre septembre 1998 et Août 2005 en race Montbéliarde, avec d'une part l'écart prédits selon la méthode Koivula (Prédiction) et d'autre part l'écart « réels » estimés *a posteriori*, sur la quantité de lait (A1 à A3) et le taux protéique (B1 à B3) pour six troupeaux particuliers.



# **Chapitre 5**

## **Discussion générale**

Les recherches visant à améliorer la quantité et la qualité des caractères de production laitière des animaux d'élevages trouvent leur principale justification dans l'importance alimentaire et économique que la filière laitière a dans de nombreux pays développés et est en train d'acquiescer dans de nombreux pays en développement tels qu'en Asie.

La sélection génétique constitue un moyen efficace de faire progresser le niveau de production et la composition du lait, grâce à une héritabilité élevée de ces caractères (autour de 0.3 pour les quantités de lait et de matières et de 0.5 pour les taux). Dans le cadre des caractères laitiers, les progrès accomplis ont largement profité du fait que les variables d'intérêt étaient « homogènes » sur l'ensemble du globe. Cela a pour conséquences des choix proches des animaux reproducteurs, mais également de la méthodologie d'analyse. En effet, il existe une émulation certaine entre les différents groupes de la communauté scientifique. Ces quelques éléments contribuent à expliquer les progrès obtenus grâce aux recherches effectuées dans ce domaine depuis un demi-siècle.

La maîtrise technique des facteurs de production tels que l'alimentation et le suivi sanitaire des animaux constitue un autre levier d'action pour garantir un bon niveau de production et de qualité du lait. Pour aider à l'acquisition et au maintien de cette maîtrise technique, l'éleveur s'appuie généralement sur des organismes de conseils en élevage tels que les organismes de contrôle laitier. Le développement d'outils d'appui technique performants tirant profit des dernières avancées méthodologiques et technologiques constitue alors un enjeu majeur de diffusion et de valorisation des recherches entreprises.

Ce dernier chapitre est l'occasion de mettre en perspective les travaux effectués au cours de cette thèse portant sur chacun des deux aspects évoqués précédemment. La première partie est consacrée à la mise en place de l'évaluation génétique sur les contrôles élémentaires. Un état des lieux de l'avancement de cette évaluation est présenté afin de faire le point sur l'ensemble du travail accompli et restant à réaliser avant son officialisation. La deuxième partie est consacrée aux perspectives de valorisation des résultats de l'évaluation génétique pour l'appui technique, et notamment à la mise en évidence des points susceptibles de poser problème ou nécessitant des études complémentaires.

## **1. La mise en place de l'évaluation génétique sur les contrôles élémentaires**

La mise en place d'une nouvelle évaluation génétique est le plus souvent le travail de toute une équipe qui conçoit et met en œuvre l'ensemble des maillons d'une chaîne de calcul depuis la collecte des données jusqu'à l'exploitation des informations nouvelles. Dans le cadre de l'évaluation génétique sur les données de contrôles élémentaires des caractères de production laitière, nous ne pouvons pas, à strictement parler, employer le terme de « nouvelle » évaluation puisque les caractères laitiers sont évalués génétiquement en France depuis les années 1960 (Colleau, 1996). Dans notre cas, l'évaluation génétique des caractères laitiers que nous proposons est seulement « nouvelle » en terme de modèle statistique employé. En effet, nous nous appuyons sur les mêmes données que dans le modèle lactation. Seul le mode d'expression a évolué : production cumulée sur une durée de référence *vs* production par jour de contrôle. De plus, le contrôle de performance réalisé en élevage laitier présente l'avantage

de ne pas avoir subi de profonde évolution dans la collecte des données depuis 20 ans, date à laquelle les informations leucocytaires ont été introduites. Nous disposons donc de 20 années de données au cours desquelles les données de contrôles élémentaires de plus de 2,5 millions de vaches ont été enregistrées chaque année.

Nous proposons de faire évoluer le modèle français vers un modèle dit de « contrôles élémentaires », modèle qui depuis 15 ans s'est progressivement imposé comme « LE » modèle d'évaluation génétique des caractères laitiers dans l'ensemble des grands pays producteurs. Grâce à la modélisation fine des performances où la plupart des effets sont décrits sous forme de courbes pour tenir compte d'un effet variable suivant le stade de lactation, le modèle d'évaluation génétique sur les contrôles élémentaires permet de valoriser au mieux la richesse des informations recueillies au cours du contrôle de performance. Outre l'intérêt méthodologique d'une telle évaluation, cette évolution présente également un aspect commercial (par exemple en termes « d'image ») non négligeable pour les entreprises de sélection qui attendent impatiemment sa mise en place. Ce travail de thèse répond donc à la fois à un besoin scientifique mais aussi aux attentes de ce monde professionnel.

La qualité d'une évaluation génétique dépend étroitement de la qualité des données sur lesquelles elle est basée. Pour la quantité de lait, les mesures sont réalisées à l'aide d'appareils homologués pour cette utilisation (Tru-test ou compteur à lait) et vérifiés annuellement. L'opérateur réalisant la mesure est une des sources d'imprécision possible, notamment car le relevé des quantités de lait produit reste une mesure visuelle dans le cas des Tru-tests. Différentes sources d'erreur sont également possibles lors de la saisie dans la base de données car celle-ci reste manuelle quel que soit le type d'appareil d'utilisé. Pour le moment, les mesures et échantillons restent majoritairement effectués par des agents du contrôle laitier, formés spécifiquement à cette pratique. Les protocoles faisant intervenir l'éleveur dans la réalisation du contrôle de performance représentent à l'heure actuelle moins de 3% des lactations contrôlées (ICAR, 2008). Mais compte tenu de la conjoncture actuelle cette proportion est amenée à augmenter. L'analyse des échantillons est effectuée dans un laboratoire de référence selon des méthodes homologuées. Cependant de simples approximations dans l'étalonnage des appareils peuvent entraîner de légères variations des taux butyreux, protéique ou leucocytaire. Malgré ces quelques sources d'imprécision, les données de production laitière sont considérées très fiables en France. Les pertes de précisions liés à l'utilisation de protocoles simplifiés, notamment les protocoles de types T, Z ou Robot où l'échantillon est réalisé au cours d'une seule des traites quotidiennes (et ayant un intervalle variable avec la traite précédente dans le cas des robots) sont prises en compte dans le modèle d'évaluation. L'évaluation génétique sur contrôles élémentaires apporte également plus de souplesse vis à vis de l'intervalle entre contrôles. Il serait en effet envisageable de réduire la fréquence des contrôles, 6 à 7 contrôles au cours de la lactation étant suffisant pour une évaluation génétique aussi précise que le modèle « lactation » actuel. Mais les stades de lactation de ces contrôles doivent être bien choisis de sorte que la production au pic de lactation puisse être correctement estimée. Trois contrôles au cours des 100 premiers jours de lactation, les autres étant réparties sur le reste de la lactation seraient un idéal. Malheureusement, contrairement à la situation ovine où cet allègement du contrôle de



performance fonctionne parfaitement, ce type de simplification est difficile à mettre en œuvre en raison de la dispersion des vêlages chez les bovins laitiers.

La quasi-totalité des informations utilisées dans le cadre de l'évaluation génétique, telles que les dates de naissance et de vêlage successifs, ont un bon niveau de fiabilité compte tenu du contexte réglementaire. Seule l'information sur la date de tarissement effective, si elle était transmise au SIG comme cela est prévu, permettrait d'améliorer la définition des classes de durée de tarissement inclus dans le modèle. En effet, les classes dont l'amplitude est généralement de 15 jours sont définies à partir d'une durée de tarissement « théorique » estimée à partir de l'intervalle entre la date du dernier contrôle enregistré de la lactation précédente à laquelle est ajoutée une durée forfaitaire (14 ou 28 jours suivant l'intervalle avec le contrôle suivant dans l'élevage) et la date de vêlage de la lactation considérée. Par exemple, si une vache a une durée théorique de tarissement de 50 jours sachant que l'intervalle entre les deux contrôles est d'un mois (30 jours), cette vache a une durée effective de tarissement de 63 jours si elle a été tarie le lendemain du dernier contrôle, mais seulement de 34 jours si elle a été tarie la veille du contrôle suivant. Cette imprécision est encore plus grande lorsque l'intervalle entre deux contrôles est de deux mois.

La mise en place du nouveau modèle d'évaluation génétique peut être décomposée en six phases. La 1<sup>ère</sup> étape consistait, à travers l'étude préliminaire réalisée par Druet *et al.* (2003), à définir le type de modélisation employée pour décrire les effets fixes variant en fonction du stade de lactation, qui a résulté en un choix de fonctions splines.

La 2<sup>ème</sup> étape était axée sur la définition du modèle d'évaluation génétique. Comme nous avons pu le voir dans le chapitre 3, ce dernier a considérablement évolué entre le début et la fin de cette thèse. La modélisation de la partie fixe du modèle d'évaluation abordée dans la partie 3.3 a été simplifiée grâce à la suppression de l'interaction avec l'année initialement prise en compte dans la modélisation des courbes d'effets fixes (âge, mois de vêlage, durée de tarissement, gestation). Le modèle simplifié implique désormais que le changement des formes de courbes de lactation au cours du temps est strictement dû à une évolution des effets aléatoires. Ces résultats ont permis de scinder le modèle d'évaluation en deux. Dans une première étape où seuls les effets fixes sont évalués (par exemple, une fois par an), les coefficients définissant les courbes d'effets fixes sont mis à jour et permettent de précorriger les données pour ces effets avant la deuxième étape. Au cours de cette dernière, l'ensemble des effets fixes dépendant du temps et les effets aléatoires sont pris en compte. Grâce au découpage de l'évaluation génétique en deux étapes, les besoins informatiques et les temps de calculs ont été considérablement réduits. Ces résultats, validés à partir de données de race Montbéliarde nécessiteront toutefois d'être vérifiés sur d'autres races, notamment la Holstein. La partie du modèle consacrée aux effets aléatoires a également été simplifiée (partie 3.2) grâce à une réduction de la dimension des matrices de (co)variances d'un tiers de leur taille sans réel impact sur l'estimation des effets qui nous intéressent, c'est-à-dire les effets génétique et d'environnement permanent. Une étude complémentaire est actuellement en cours afin de réduire au maximum le nombre de paramètres nécessaires pour décrire l'effet aléatoire troupeau – année. Dans le modèle finalement retenu à l'issue de ce travail de thèse,

l'effet troupeau – année reste modélisé par six variables élémentaires, nombre supérieur aux effets génétique ou d'environnement permanent alors que la part de variance qu'il explique est très faible par rapport à la variance totale du modèle et que cet effet ne semble pas intéressant pour l'appui technique. Pour aller plus loin dans la diminution du nombre de variables pour cet effet, il sera toutefois nécessaire de valider une condition *sine qua non*. Le modèle avec un effet aléatoire troupeau – année décrit par un très petit nombre de variables (en plus de la correction incluse à travers les hétérogénéités de variances dans l'évaluation génétique à l'aide d'un modèle autorégressif intra année – troupeau) devrait obtenir des résultats pratiquement équivalents sur l'index « global » ou les index « élémentaires » des effets génétiques et d'environnement permanent à notre modèle actuel.

La 3<sup>ème</sup> étape de cette mise en place était l'estimation des paramètres génétiques pour huit des races laitières évaluées en France (non décrite dans cette thèse). Cette phase était achevée au début de cette thèse dans les trois principales races que sont la Holstein, la Montbéliarde et la Normande. Pour les cinq races régionales que sont la Brune, la Pie Rouge des plaines, la Simmental, l'Abondance et la Tarentaise, les estimations sont bien avancées pour l'ensemble des caractères évalués mais nécessitent d'être finalisées car des problèmes importants de convergence ont été rencontrés pour certaines combinaisons race – caractère.

La 4<sup>ème</sup> étape consistait au développement du logiciel d'évaluation génétique et sa validation. Le travail de développement du logiciel *Genekit* entamé par Vincent Ducrocq au début de cette thèse a permis d'aboutir à un logiciel répondant à l'ensemble des spécificités du modèle français tels que le recours aux splines ou aux régressions aléatoires. Pour permettre l'évaluation simultanée de plus de 200 millions de données de contrôles élémentaires en race Holstein, différentes astuces techniques ont dû être mises en œuvre notamment pour éviter de devoir stocker l'ensemble des coefficients correspondant au stade de lactation dans le fichier de données.

Le logiciel inclut également le calcul des coefficients de détermination (Ducrocq and Schneider, 2007), une extension multicaractères de la méthode dite « des sources d'information » proposée par Harris et Johnson (1998).

Une autre spécificité de ce logiciel est l'inclusion d'une approche permettant de résumer les données anciennes en les réduisant à un nombre limité de données dites « résumées ». En effet, nous pouvons considérer que les valeurs génétiques des vaches ayant par exemple effectuées leur dernière lactation avant 2000 ne bougent plus puisqu'il n'y a pas d'apport d'informations nouvelles récemment. Il est possible de réduire de façon importante le volume des données historiques en synthétisant l'information correspondante à ces animaux sous forme de 2 à 4 contrôles résumés (2 dans le cas d'une seule lactation, 4 pour les trois premières lactations) au lieu des 30 données de contrôles élémentaires potentielles. Il est ensuite possible de combiner ces informations aux données de contrôles élémentaires plus récentes en réduisant de façon importante les temps de calcul tout en conservant des corrélations supérieure à 0.999 pour les effets génétiques (Ducrocq and Schneider, 2008).

La dernière spécificité de ce logiciel d'évaluation génétique est l'intégration de la correction pour les hétérogénéités de variances réalisée au cours de l'été 2008. Ce changement majeur

introduit dans le modèle d'évaluation sur les données lactation en 1999 (Robert-Granié *et al.*, 1999) est ainsi conservé. De ce point de vue toutefois, quelques améliorations, décrites dans la partie 2.4 ont été introduites dans le modèle d'évaluation sur les contrôles élémentaires. Ainsi, la résiduelle combine à la fois une hétérogénéité dépendant du stade de lactation et des sources d'hétérogénéité plus classiques telles que l'interaction troupeau – année intégrant un terme autorégressif entre années et une interaction région – année – rang de lactation – type de protocole de contrôle laitier. Ce dernier élément permet de tenir compte automatiquement des différences de précision entre types de protocole de contrôle laitier contrairement à ce qui est fait dans la plupart des pays, où lorsque ce facteur est intégré au modèle, il se traduit par une diminution du poids accordé aux données dans l'évaluation génétique augmentant avec le niveau de simplification du protocole (Schaeffer *et al.*, 2000).

Il faut noter que l'étape de validation du logiciel (partie 3.1) avait seulement pour objectif de s'assurer de l'exactitude des calculs mis en œuvre dans *Genekit*, pour les aspects relevant directement de l'utilisation d'un modèle de régression aléatoire sur des données de contrôles élémentaires. Les extensions présentées précédemment n'ont pas fait l'objet de validation formelle.

L'étape suivante, la 5<sup>ème</sup>, consistait à réaliser une évaluation test. D'abord effectuée sur un « échantillon » de la race Montbéliarde (le département du Jura), elle a ensuite été étendue à l'ensemble des données de cette race. Cette 5<sup>ème</sup> étape est étroitement liée à la 2<sup>ème</sup>. En effet, les résultats des évaluations tests nous ont amené à modifier de façon progressive le modèle d'évaluation pour le faire évoluer jusqu'au modèle retenu actuellement, sans toutefois qu'il soit définitivement stabilisé. Ces évolutions successives du modèle n'ont pas remis en cause les estimations des paramètres génétiques réalisés auparavant dans la mesure où ils touchent principalement les effets fixes, et que la réduction du rang des matrices de (co)variances des effets aléatoires est faite en maintenant la variance totale constante par l'ajout de la variance « perdue » à la variance résiduelle. Pour le moment, seule la race Montbéliarde a fait l'objet d'une évaluation génétique. Dans les races Holstein et Normande, seule la 1<sup>ère</sup> étape concernant l'évaluation des effets fixes a été réalisée à l'heure actuelle, mais un test complet devrait être effectué très prochainement

Enfin, la dernière étape nécessaire avant la réalisation d'une évaluation officielle touche principalement des aspects plutôt en aval des évaluations génétiques. Le premier concerne des points tels que la définition des caractères diffusés, la définition des bases de référence, le mode d'expression... ainsi que le travail de communication nécessaire pour expliquer les changements par rapport à l'évaluation actuelle et leurs conséquences. Le deuxième aspect intéresse plus spécifiquement les races pour lesquelles le commerce de matériel génétique, principalement de la semence, constitue un enjeu économique. Dans ces races, les comparaisons génétiques sont effectuées à partir des index déterminés dans le cadre des évaluations internationales. Or, pour participer à celles-ci, il est nécessaire de valider à partir d'au moins une des trois méthodes agréées (Boichard *et al.*, 1995), l'absence de biais dans le modèle d'évaluation. Cette étape de validation du progrès génétique reste à effectuer. Nous espérons néanmoins ne pas être confronté aux difficultés rencontrées notamment par les pays nordiques pour lesquels cette étape de validation a nécessité plus d'une année. De plus, il sera nécessaire de participer, comme lors de chaque modification de modèle, à une évaluation test

(« test run » d'Interbull) afin d'estimer les nouvelles corrélations entre la France et les autres pays participants avant toute évaluation internationale officielle. Ces test runs ayant lieu deux fois par an (mai et septembre), il serait intéressant que la France puisse envoyer les données correspondantes au modèle sur les contrôles élémentaires lors de la prochaine évaluation test.

Afin de limiter le temps de calcul nécessaire à la réalisation de l'évaluation génétique, il a été décidé de n'évaluer que les trois caractères de quantités (lait, matière grasse et matière protéique) ainsi que le taux leucocytaire contrairement aux analyses réalisées au cours de cette thèse. Cette décision s'appuie sur la pratique de l'ensemble des autres pays qui n'évaluent pas directement les taux butyreux et protéique mais les « ré-estiment » à partir des valeurs génétiques estimées sur les matières. Pour cela, il est nécessaire d'établir des formules de conversion. Pour faciliter leur détermination, nous nous appuierons sur les valeurs génétiques estimées directement.

Les contraintes en terme de nombre de données et de mémoire nécessaire ont constitué un frein important à la mise en place de ce nouveau modèle d'évaluation. Nous espérons que l'ensemble des barrières sont désormais levées et qu'il va être possible de réaliser rapidement une évaluation test incluant les effets aléatoires en race Holstein, notamment grâce à l'acquisition très prochaine d'un nouvel ordinateur plus puissant.

## **2. Les perspectives de valorisations des résultats de l'évaluation à des fins non génétiques**

Le paysage agricole français est actuellement en profonde mutation. Les éleveurs sont faces à des marchés de plus en plus fluctuants (phénomène que la levée possible des quotas risque d'amplifier). Avec l'évolution de la demande sociétale, ils ont besoin d'intégrer de nouvelles exigences environnementales et de bien-être animal dans leur système de production. Avec la baisse des revenus agricoles, il est nécessaire de mettre en place des indicateurs techniques systématiques et efficaces pour repérer rapidement des problèmes alimentaires ou sanitaires dans leur troupeau dont la taille ne cesse de s'accroître. De nouveaux outils de conseil en élevage doivent être développés pour répondre à ces attentes et s'adapter à la diversité grandissante des systèmes de production.

### **2.1. Prédire la production : une fonction clé**

Le développement de l'évaluation génétique sur les contrôles élémentaires nous semble constituer une excellente opportunité pour développer ces nouveaux outils. En effet, le modèle d'évaluation génétique français est tout à fait original par sa simplicité d'interprétation et par sa flexibilité d'utilisation. La courbe de production de l'animal y est décomposée en une somme d'effets d'environnement (troupeau, race, région, rang de lactation, âge et mois de vêlage, durée de tarissement, gestation...) et d'effets propres à la vache (caractéristiques génétiques et non génétiques). Ces effets sont calculés pour les stades de lactation allant de 7 à 335 jours et sont corrigés pour tous les autres facteurs de variation ce qui les rend directement interprétables.

Comme nous avons pu le voir dans la partie 1.3 présentant les principaux objectifs techniques des outils de conseil en élevage présents sur le territoire français, de nombreuses utilisations sont basées sur la prédiction de la production d'un animal et par extension de la production du troupeau aussi bien pour l'analyse des résultats passés que pour la prévision de la production future (FCL, 2005a et 2005b). La prédiction des caractères relatifs à la quantité de lait et à sa qualité (TB et TP) est basée sur l'utilisation de courbes-types du référentiel ARCOLA (Bouloc, 1993) dont les facteurs de variation pris en compte sont limités et les valeurs estimées dans les années 1990 sont vieillissantes. Afin d'illustrer par un exemple les prédictions réalisées à partir du référentiel ARCOLA, nous avons retenu la 2<sup>ème</sup> lactation d'une vache Montbéliarde. Nous avons prédit sa production suivant les courbes-types du référentiel ARCOLA, en tenant compte de ses caractéristiques (un premier vêlage à l'âge de 32 mois, un second vêlage quinze mois plus tard au mois de décembre, après 80 jours de tarissement, dans un troupeau dont la moyenne de production se situe autour de 8000 kg de lait par vache et par lactation) (Figure 5.1). A partir des productions moyennes observées sur les cinq premiers contrôles, la prédiction de production réalisée au mois d'avril sur cette vache se base sur une courbe-type avec un niveau de production de 10 000 kg. Cependant, les contrôles des mois de mai, juin et juillet sont nettement supérieurs aux prédictions réalisées. En basant les prédictions sur des courbes-types, la courbe de production prédite conserve la forme d'une courbe de lactation théorique et ne s'adapte absolument pas à l'environnement de production, dépendant largement de l'alimentation offerte à la vache. En effet, les productions journalières obtenues par cette vache au cours des mois de juin et juillet, avec respectivement 32.1 kg et 30.9 kg de lait, correspondent à une courbe-type dont le niveau est approximativement de 11 500 kg compte tenu des stades de lactation de ces contrôles.

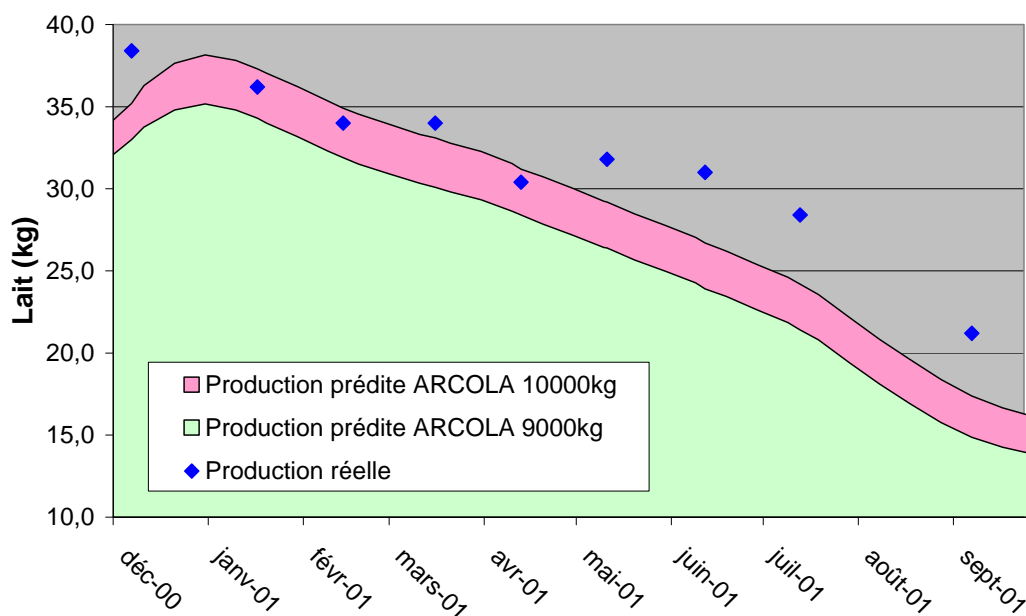


Figure 5.1. Quantités « réelles » de lait mesurées mensuellement sur une vache Montbéliarde en 2<sup>ème</sup> lactation, vêlant le 2 décembre 2000 à l'âge de 47 après plus de 80 jours de tarissement, confrontées aux courbes-types de lactation obtenues selon le référentiel ARCOLA pour des niveaux de production de 9000 et 10000 kg de lait par lactation 305 jours.

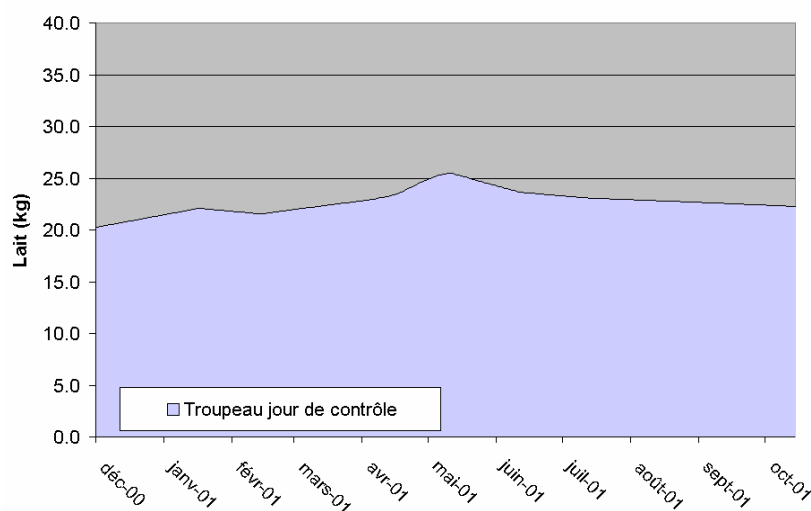
Au contraire, avec un outil d'appui technique basé sur les effets estimés ou prédits à partir de l'évaluation sur les données de contrôles élémentaires, la courbe de production obtenue en sommant les différents effets (Figure 5.2) tient compte de l'ensemble des facteurs influençant sa forme et son niveau. Ainsi, si nous reprenons l'ensemble des effets du modèle d'évaluation défini à la fin du chapitre 3, nous avons :

- ✓ L'effet HTD prédit suivant la méthode de Koivula (étape 1 de la figure 5.2) à partir des données des années antérieures (Koivula *et al.*, 2007) comme décrit dans la partie 4.2.
- ✓ Les courbes liés aux effets de l'âge, du mois de vêlage, de la durée de tarissement et de la gestation présentées dans la partie 4.1, qui lorsqu'elles sont cumulées définissent la forme générale de la courbe de lactation en fonction du stade de lactation (étape 2 de la figure 5.2). Ces effets ne dépendant pas de l'année, ils sont donc disponibles en permanence.
- ✓ Les constantes définissant le niveau des courbes précédentes n'évoluant pas de façon significative entre les années d'après nos observations, il est envisageable de baser les prédictions sur les valeurs des constantes obtenues l'année précédente dans des conditions équivalentes (incluses dans les courbes des effets de la figure 5.2).
- ✓ L'effet génétique est prédit à partir des valeurs génétiques estimées lors de la précédente évaluation génétique si l'animal avait un minimum de trois performances enregistrées. Dans le cas contraire, la valeur génétique obtenue à partir de son ascendance est utilisée.
- ✓ Comme pour l'effet génétique, l'effet d'environnement permanent est prédit à partir des valeurs estimées lors de la précédente évaluation génétique. Si l'animal n'a jamais été contrôlé ou l'a été moins de trois fois avant la précédente évaluation, l'effet est, par défaut, supposé nul.
- ✓ L'effet troupeau – année n'est pas pris en compte dans les prédictions, d'une part parce que son effet est très faible proportionnellement aux autres ; et d'autre part car il semble, selon nos premières observations, très variable d'une année sur l'autre et donc difficilement prédictible. Ceci revient donc à supposer cet effet nul.

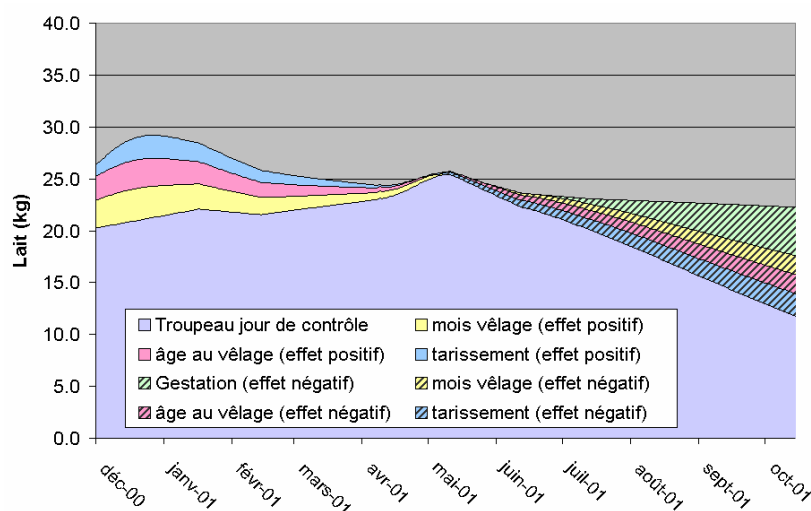
Ainsi, il est possible de construire la courbe de production (Figure 5.2) de cette vache selon l'approche présentée. La ligne supérieure de la zone de surface bleue correspond à la production prédite de cette vache au cours de sa deuxième lactation. La courbe de production obtenue peut être comparée à la production mesurée lors des contrôles (Figure 5.3). L'adéquation entre la production prédite et la production réelle est meilleure que celle obtenue avec des prédictions basées sur les courbes-types du référentiel ARCOLA actuellement utilisées dans les outils d'appui technique de nombreux OCL.

## **2.2. De nombreuses valorisations possibles mais aussi quelques limites...**

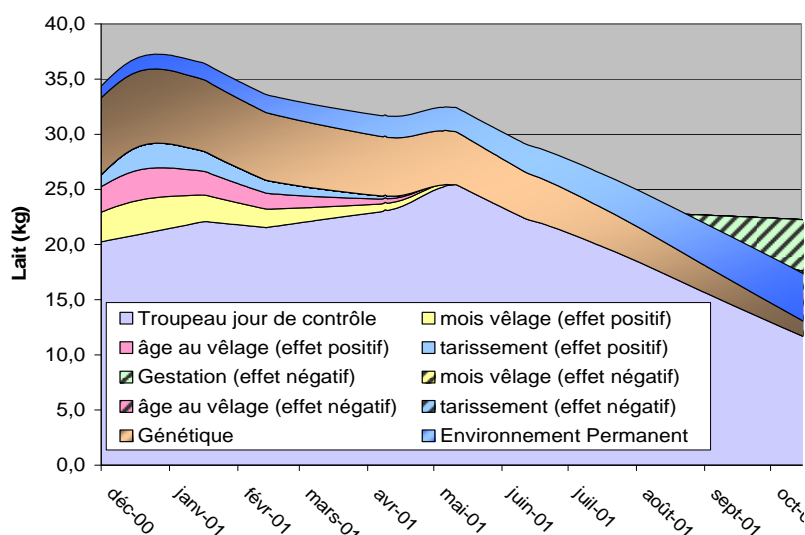
La diversité des effets estimés à partir du modèle d'évaluation génétique sur les contrôles élémentaires, s'adaptant à toutes sortes de situations de race, région, pratique d'élevage... permet d'envisager l'établissement d'une courbe de production « standard » pour n'importe quelle vache grâce à une fonction faisant appel aux effets spécifiques de l'animal. Cette nouvelle génération d'outils d'appui technique intégrerait des fonctions avec des objectifs similaires à ceux actuellement utilisés dans les logiciels d'appui technique puisque ceux-ci répondent aux principaux besoins. De façon similaire au référentiel ARCOLA dans l'outil



**Etape 1**



**Etape 2**



**Etape 3**

Figure 5.2. Construction de la courbe de production prédite pour la vache Montbéliarde en 2<sup>ème</sup> lactation, vêlant le 2 décembre 2000 à l'âge de 47 après plus de 80 jours de tarissement (identique à la figure 5.1) en cumulant les effets HTD prédits pour cet élevage à partir des années antérieures (1<sup>ère</sup> étape), les effets des courbes pour l'âge, le mois de vêlage, la durée de tarissement et la gestation du stade de lactation (intégrant les constantes estimées l'année précédente) déterminés à partir des caractéristiques de l'animal (étape 2) et enfin des effets aléatoires que sont l'effet génétique et l'effet d'environnement permanent estimés à partir de la 1<sup>ère</sup> lactation de l'animal (étape 3).

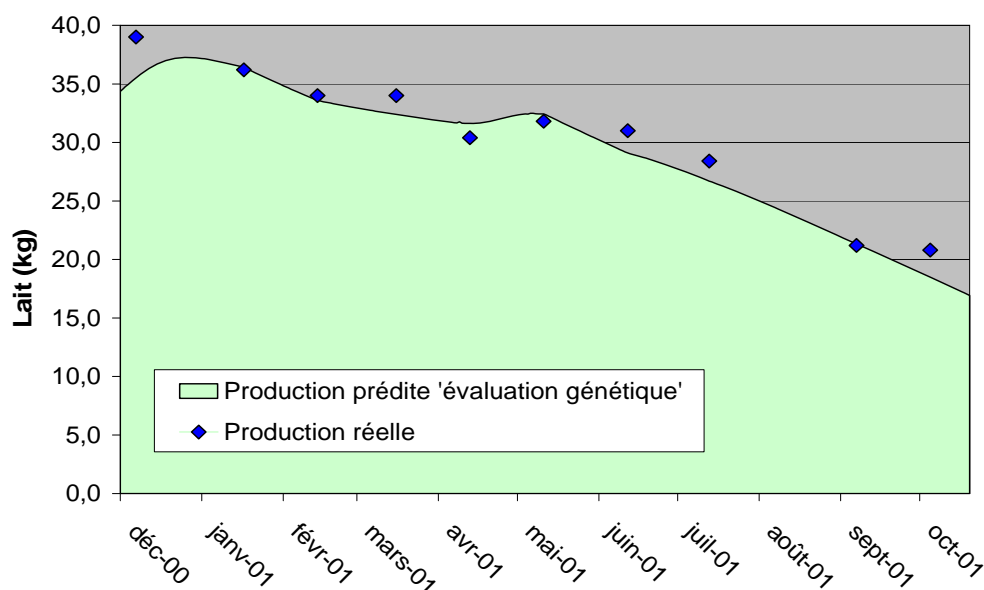


Figure 5.3. Quantités « réelles » de lait mesurées mensuellement sur une vache Montbéliarde en 2<sup>ème</sup> lactation, vêlant le 2 décembre 2000 à l'âge de 47 après plus de 80 jours de tarissement, confrontées à la production prédite à partir des effets estimés dans le cadre de l'évaluation génétique sur les contrôles élémentaires.

SIEL, la fonction principale établissant une courbe de production de l'animal présentée précédemment constituerait le socle du nouvel outil. Elle permettra de prédire les productions futures à partir de celle-ci, mais aussi de confronter les productions réelles aux productions attendues, étape essentielle à la détection de problèmes alimentaires ou sanitaires et à l'établissement d'un diagnostic. Les données à partir desquelles ces courbes seraient établies feraient l'objet de mises à jour de façon régulières (lors de chaque évaluation génétique, soit trois fois par an).

Une fonction d'analyse dynamique de l'effet de la conduite du troupeau à partir de l'effet HTD basée sur l'approche présentée dans la partie 4.2 pourrait également être incluse dans cette nouvelle génération d'outils. Ces effets HTD constitueraient des indicateurs des éventuels décrochages du niveau de production et de taux y compris cellulaires. Afin de maintenir l'aspect dynamique de cette analyse et permettre une véritable utilisation au quotidien de ces indicateurs, il nous paraît nécessaire de s'affranchir de la fréquence des évaluations génétiques nationales en développant un module de calcul des effets HTD « provisoires », mis à jour mensuellement afin de permettre aux OCL de disposer d'éléments réactifs.

Les effets HTD estimés pourraient également être utilisés pour déterminer le niveau de production potentielle de l'animal, en se basant sur les meilleurs niveaux HTD observés dans le troupeau et sur les caractéristiques de l'animal. La notion de production potentielle est essentielle pour déterminer la capacité d'ingestion de l'animal (Faverdin *et al.*, 2007) sur laquelle se base le logiciel d'alimentation INRation utilisé par les OCL pour le conseil et le calcul des rations.



L'utilisation des effets estimés lors de l'évaluation génétique sur les contrôles élémentaires à des fins d'appui technique présente toutefois quelques limites. En effet, seules les trois premières lactations sont incluses dans l'évaluation génétique, et la durée de lactation est pour le moment limitée à 335 jours. Or, dans un élevage, toutes les vaches productrices doivent être prises en compte dans les conseils apportés par l'agent du contrôle laitier, quel que soit le stade ou le rang de lactation de l'animal. Dans une optique de prédiction de la production du troupeau ou dans une optique de gestion de quota, il s'agit une condition essentielle. Afin de répondre à ces besoins, il sera nécessaire d'envisager les modalités d'extension aux lactations de rang supérieur à 3 et de prolongation de la durée de lactation. En effet, cette durée ne cesse de s'accroître au fil des ans (entre 2 et 5 jours par an en moyenne depuis 10 ans), tendance qui ne semble malheureusement pas prête à s'inverser compte tenu de la dégradation du niveau de fertilité. En race Holstein par exemple, la durée moyenne pour l'année 2007 (Institut de l'Élevage, 2008) était de 356 jours en première lactation avec un écart-type d'environ 120 jours.

Comme nous avons pu le voir dans la première partie de cette discussion consacrée à l'évaluation génétique, le modèle d'évaluation sur les données de contrôles élémentaires apporte un surplus de flexibilité vis-à-vis de la variabilité des protocoles de contrôle laitier par rapport à un modèle d'évaluation basé sur des données lactation. Toutefois, pour la valorisation des résultats pour l'appui technique, la situation est bien différente. Par exemple, dans le cadre d'un protocole de type T où une seule des deux traites est contrôlée en alternance, la quantité de lait est actuellement multipliée par 2 et les taux mesurés sur la traite considérée sont censés représenter le taux moyen journalier. Cependant, les deux traites sont rarement effectuées avec un intervalle strict de 12 heures et les différences d'activité entre la période diurne et nocturne expliquent les variations observées sur la quantité de lait et le TB. Par exemple, dans une étude réalisée par Leclerc et Delacroix (2004) portant sur plus de 300 lactations avec des mesures hebdomadaires, le lait 24 heures était de 19.7 kg et le TB moyen de 42.2 g/kg. L'analyse séparée des productions du matin et du soir avait cependant mis en évidence de grosses différences entre les deux traites puisque le matin il y avait en moyenne 11.8 kg de lait avec un TB de 38.1 g/kg et le soir 7.9 kg avec un TB de 48.4 g/kg. Dans l'évaluation génétique, ces différences ne posent pas de problèmes insurmontables pour estimer les effets aléatoires qui intéressent les généticiens : l'ensemble des animaux étant soumis à des conditions similaires pour une date de contrôle, l'effet HTD absorbe ces variations à travers son rôle de « moyenne » des contemporaines. Par contre, il n'est pas envisageable de valoriser directement l'effet HTD obtenu sur la quantité de lait comme indicateur de la conduite du troupeau dans le cas de protocole T et obtenu sur le TB dans le cas de protocole T, Z et Robot. Pour cela, il faudrait développer des formules de correction permettant de se ramener à des variables équivalentes à celles qui auraient été obtenues si les mesures avaient été faites sur 24 heures comme cela a été proposé par DeLorenzo et Wiggans (1986) ou Liu *et al.* (2000). Les éleveurs adhérant à des protocoles simplifiés doivent donc être conscients des conséquences de leur choix sur les possibilités d'appui technique. À l'avenir, les protocoles de contrôle laitier seront donc de moins en moins dictés par les exigences liées à l'évaluation génétique mais par les attentes des éleveurs en terme de conseils.

### 2.3. Les autres perspectives de valorisation

Suite à ce travail sur la mise en place des évaluations génétiques sur les contrôles élémentaires, de nombreuses perspectives de valorisation ont pu être mises en avant. Afin de les mettre en œuvre, un groupe de travail constitué de personnes des départements Génétique, Techniques d'Elevage et Qualité et Actions Régionales de l'Institut de l'Elevage, des départements Génétique et Physiologie Animale et Système d'Elevage de l'INRA, de FCL et d'OCL pilotes ont répondu à un appel d'offre CASDAR (Compte d'Affectation Spécial pour le Développement Agricole et Rural) pour trouver les financements nécessaires à la poursuite de ce travail. Malheureusement, le projet n'a pas été retenu par manque d'adéquation avec le sujet de l'appel d'offre qui était en lien avec le « Grenelle de l'environnement ». Toutefois, afin de ne pas rompre la dynamique créée autour de ce sujet transdisciplinaire, les actions initiales de ce projet (notamment le développement des indicateurs de conduite des troupeaux laitiers à travers l'analyse dynamique des effets HTD et la prédiction de la production d'une vache) ont été déposées dans le cadre de l'appel d'offre « Actions innovantes » du ministère de l'Agriculture. L'examen des projets est actuellement en cours, mais des réunions de travail du groupe de travail sont d'ores et déjà prévues pour le premier semestre 2009, quel que soit le résultat de cet appel d'offre.

Outre les perspectives déjà évoquées, d'autres objectifs nous semblent pertinents à aborder. Dans le cadre d'une thèse actuellement en cours à l'Unité Mixte de Recherches sur la Production Laitière de Rennes, il est prévu de développer un outil modélisant le fonctionnement d'une exploitation laitière afin d'aider l'éleveur à s'adapter à une conjoncture en perpétuelle évolution. Ces deux dernières années en sont le parfait exemple. Après un second semestre 2007 où le prix du lait s'est envolé et où les suppléments de quotas ont atteint 15%, l'année 2008 contraste fortement avec un prix du lait revenu à un niveau semblable à 2006 et un niveau de production trop important par rapport à la référence nationale. L'objectif est de simuler les conséquences des pratiques d'élevage et leur modification dynamique (alimentation, fréquence de traite, gestion de la reproduction) sur le fonctionnement et les productions d'un troupeau laitier. Le développement d'outils d'aide à la décision basés sur les effets estimés dans les évaluations génétiques faciliterait les choix de conduite des systèmes laitiers afin d'accroître les capacités d'anticipation pour répondre aux attentes des éleveurs et des transformateurs.

La caractérisation des effets HTD, reflets de la conduite du troupeau, en fonction des systèmes de production pourrait constituer un axe de recherche intéressant. Pour cela, il serait nécessaire de croiser les effets HTD avec des variables décrivant le système de production incluses dans les bases de données régionales fournies par les OCL pilotes et de la base de données « Diapason » qui regroupe les données de plus de 300 exploitations laitières françaises suivies dans le cadre des réseaux d'élevage (Charroin *et al.*, 2005). Nous pourrions ainsi identifier quelques variables permettant de caractériser les modes de conduite afin de les inclure à terme dans le SIG et permettre ainsi d'affiner les outils d'appui technique proposés.

Enfin, une dernière direction de recherche, en se basant sur les différents systèmes de production mis en évidence, serait d'aborder la question de l'adaptation de la génétique aux

systèmes d'élevage. En effet, faute d'information relative au milieu de production, l'approche génétique suppose des réponses homogènes à la sélection. A partir de ces éléments, il serait possible d'entamer une réflexion sur l'adaptation de la génétique à la diversité des systèmes d'élevages présents en France.

## Références

- Boichard D., Bonaïti B., Barbat A. and Mattalia S. 1995. Three methods to validate the estimation of genetic trend for dairy cattle. *Journal of Dairy Science*. 78: 475-486.
- Bouloc N. 1993. Guide d'utilisation du référentiel courbe de lactation. *Compte rendu Institut de L'Elevage* N°2086, 68p.
- Charroin T., Palazon R., Madeline Y, Guillaumin A. et Tchakerian E. 2005. Le système d'information des réseaux d'élevage français sur l'approche globale de l'exploitation : Intérêt et enjeux dans une perspective de prise en compte de la durabilité. *Rencontres Recherches Ruminants*. 12: 335-338.
- Colleau J.J. 1996. Evaluation génétique des animaux d'élevages. *INRA Productions animales*. N°HS « 50 ans de recherches en production animales »: 27-40.
- DeLorenzo M.A. and Wiggans G.R. 1986. Factors for Estimating Daily Yield of Milk, Fat, and Protein from a Single Milking for Herds Milked Twice a Day. *Journal of Dairy Science*. 69: 2386-2394.
- Druet T., Jaffrézic F., Boichard D. and 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*. 86: 2480-2490.
- Ducrocq V. and Schneider M. del P. 2007. Generalisation of the information source method to compute reliabilities in test day models. *Interbull Bulletin*. 37:82-87.
- Ducrocq V. and Schneider M. del P. 2008. How to summarize historic test-day record information?. *Interbull Bulletin*. 38:27-31.
- Faverdin P., Delaby L. et Delagarde R. 2007. L'ingestion d'aliments par les vaches laitières et sa prévision au cours de la lactation. *INRA Productions Animales*, 20 : 151-162.
- France Contrôle Laitier. 2005a. SIEL : Fiche domaine valorisation du contrôle de performance. V1.7.0. 28p.
- France Contrôle Laitier. 2005b. SIEL : Fiche domaine production laitière. V1.7.0. 35p.
- Harris B. and Johnson D. 1998. Approximate reliabilities of genetic evaluation under an animal model. *Journal of Dairy Science*. 81: 2723-2728.
- ICAR (International Committee for Animal Recording). 2008. Yearly Milk Enquiry – On-line Database. <http://www.waap.it/enquiry/> Accessed September. 1, 2008.
- Koivula M., Nousianinen J.I, Nousianinen J. and Mäntysaari E.A. 2007. Use of herd solutions from a random regression test-day model for diagnostic dairy herd management. *Journal of Dairy Science*. 90: 2563-2568.
- Leclerc H. et Delacroix J. 2004. Méthodes d'enregistrement des performances du contrôle bovin – comparaison des méthodes Z4 et T4 avec la méthode de référence et les performances réelles. *Compte rendu Institut de L'Elevage* N°0477010, 38p.
- Liu Z., Reents R., Reinhardt F. and Kuwan K. 2000. Approaches to estimating daily yield from single milk testing schemes and use AM-PM records in test-day model genetic evaluation in dairy cattle. *Journal of Dairy Science*. 83: 2672-2682.
- Robert-Granié C., Bonaïti B., Boichard D. and Barbat A. 1999. Accounting for variance heterogeneity in French dairy cattle genetic evaluation. *Livestock Production Science*. 60: 343-357.
- Schaeffer L.R., Jamrozik J., Van Dorp R., Kelton D.F. and Lazenby D.W. 2000. Estimating daily yields of cows from different milking schemes. *Livestock Production Science*. 65: 219-227.

## Conclusion générale

Les travaux effectués au cours de cette thèse ont permis tout d'abord de mettre en évidence la complexité et les limites rencontrées pour la mise en place d'une évaluation française sur les données de production du jour de contrôle. En effet, la mise en œuvre de l'évaluation telle qu'elle était initialement prévue se serait avérée extrêmement délicate dans le cadre d'évaluations officielles en routine compte tenu du volume de données à traiter et du nombre d'effets inclus dans le modèle. Les travaux réalisés se sont donc attachés à lever ces barrières techniques. Le nombre total d'effets aléatoires du modèle a été réduit d'un tiers grâce à l'utilisation d'une approche de rang réduit entraînant des variations négligeables des effets estimés. La définition des effets fixes modélisés sous forme de courbes dont l'impact varie suivant le stade de lactation a été revue. Il est apparu que l'effet de l'année influence uniquement le niveau de la courbe et non sa forme. Cette modification a permis d'accroître la stabilité des effets estimés, de réduire de façon importante les temps de calcul mais surtout de permettre la réalisation de l'évaluation génétique en deux étapes. Les données, précorrignées pour les effets liés aux courbes de lactation dépendant de l'âge, du mois de vêlage, de la durée de tarissement et de la gestation, sont analysées dans une seconde étape suivant un modèle incluant l'ensemble des effets aléatoires et les effets fixes dépendant du temps.

D'autre part, les travaux réalisés ont permis de mettre en avant de multiples perspectives de valorisation des résultats de l'évaluation génétique à des fins d'appui technique. En effet, grâce à une modélisation originale des effets fixes comme somme de courbes spécifiques, le modèle d'évaluation génétique sur les contrôles élémentaires permet d'envisager le développement de nouveaux outils d'appui technique pour le conseil en élevage où les prédictions de production seraient réalisées à partir des effets estimés lors des évaluations génétiques sur les contrôles élémentaires. Ces nouveaux outils pourraient également inclure un module d'analyse dynamique de l'effet de la conduite du troupeau à travers l'effet HTD.

Pour répondre de façon optimale aux attentes des éleveurs, le développement d'outils d'appui technique doit être pensé avec attention à la fois en terme d'analyses et conseils apportés mais aussi du point de vue du mode de communication proposé. En effet, l'explosion des technologies au cours de 20 dernières années offrent de nouvelles perspectives. Désormais, un grand nombre d'éleveurs sont équipés d'outils informatiques, envoient leur déclaration de naissance via des interfaces web, et les plus « portés » sur la technologie sont équipés d'assistants personnels pour enregistrer les événements tels que les chaleurs observées sur les vaches du troupeau. Les outils d'appui technique proposés aux éleveurs français ne doivent pas rester en dehors de cette vague technologique. Ainsi, au Québec, les concepteurs des outils de conseils en élevage ont développé des modules facilitant la visualisation et l'interprétation des données (Lefebvre *et al.*, 2004a et 2004b). L'objectif est de simplifier la détection des données hors normes et d'identifier les aspects à améliorer. En détectant les problèmes le plus tôt possible, leur incidence économique est limitée. Le seul titre des documents présentant ces outils - « Une image vaut mille mots... sans compter les chiffres » - reflète pleinement les objectifs poursuivis. Parmi ces modules, l'un est dédié à l'analyse des courbes de lactation et un autre à l'analyse des résultats du jour du contrôle. En Finlande, les

## Conclusion générale

éleveurs ont la possibilité de se connecter à une application Web regroupant l'ensemble des résultats obtenus lors des contrôles de performance mais proposant également une analyse dynamique de la conduite du troupeau à travers les effets HTD de l'évaluation génétique sur les contrôles élémentaires (Koivula *et al.*, 2007). Enfin, un dernier exemple concerne les Pays-Bas où début 2006, des tests étaient en cours sur des outils d'appui technique basés sur les solutions des effets de l'évaluation sur les données de contrôles élémentaires (De Roos, *communication personnelle*). Afin de valoriser au mieux les résultats, deux modules concernant la production et l'alimentation, étaient développés pour cet outil.

Outre les possibilités de valorisation des résultats de l'évaluation génétique sur les contrôles élémentaires en lien direct avec les caractères de production, il semble concevable de développer des applications à la détection de troubles de la santé tels que les maladies métaboliques, les mammites, la fertilité. Par exemple, l'analyse du ratio TP/TB, utilisé comme indicateur du bilan énergétique de l'animal, permet de détecter des risques d'acétonémie lorsque ce ratio est faible ou d'acidose lorsqu'il est élevé. En collaboration avec des nutritionnistes, il serait possible d'élaborer des conseils pour remédier rapidement à ces situations avant qu'elles ne se dégradent. L'analyse conjointe des données d'occurrence de mammites cliniques, de comptages leucocytaires et de production laitière disponibles dans le cadre des évaluations génétiques peut également ouvrir de nouvelles perspectives pour réduire l'incidence de cette maladie, économiquement très onéreuse au niveau d'un élevage. Enfin, un dernier exemple d'utilisation des co-produits des évaluations génétiques nationales serait celui de la fertilité dont la dégradation a des conséquences multiples allant de l'augmentation des taux de réformes à l'accroissement de la durée de lactation et donc la réduction de la production par jour de vie de l'animal sans compter les coûts de multiples inséminations infructueuses. En croisant les informations disponibles dans le cadre des évaluations génétiques sur les contrôles élémentaires et celles de réussite ou d'échec à l'IA, nous pouvons imaginer des outils indiquant aux éleveurs les périodes favorables pour la réussite à l'insémination en se basant sur les effets néfastes d'un déficit énergétique sur la reprise de l'activité ovarienne et la fertilité.

Les perspectives de ces travaux sont donc extrêmement nombreuses et dépassent largement le cadre de la génétique. En effet, ces axes de recherche impliquent l'ensemble des professionnels gravitant autour du monde de l'élevage. Ces travaux ouvrent la porte à de futures collaborations entre généticiens, nutritionnistes, physiologistes, vétérinaires... permettant d'incorporer l'aspect multidimensionnel de la production laitière. La dimension de la base nationale dont la vocation était initialement uniquement génétique, associée à la richesse de bases de données plus spécifiques offre de multiples possibilités d'études appliquées très utiles.

## Références

- Koivula M., Nousianinen J.I, Nousianinen J. and Mäntysaari E.A. 2007. Use of herd solutions from a random regression test-day model for diagnostic dairy herd management. *Journal of Dairy Science*. 90: 2563-2568.
- Lefebvre D., Lacroix R., Pietersma D. and Wade K. 2004a. Analyses + : pour visualiser vos données. *Le producteur de lait québécois*. Mars 2004: 23-25.
- Lefebvre D., Lacroix R., Pietersma D. and Wade K. 2004b. Analyses + (la suite) : Une image vaut mille mots... sans compter les chiffres\$. *Le producteur de lait québécois*. Juin 2004: 20-22.

# Annexes

## Principe de l'évaluation génétique

La génétique quantitative est utilisée pour l'amélioration des individus d'une espèce, en sélectionnant comme reproducteurs les meilleurs sujets selon les critères recherchés (production laitière, morphologie...). Cette sélection ne se base pas sur des performances brutes aussi appelées phénotypes (production de lait journalière, taille de l'animal) mais sur leur valeur génétique additive (ou transmissible). L'évaluation génétique fournit une prédiction de la valeur génétique vraie (qui reste inconnue) à partir de deux sources d'informations : les performances mesurées sur les animaux ainsi que leurs relations de parentés.

### A/ Le modèle linéaire

Pour déterminer la part de la performance liée à la composante génétique avec un maximum de précision et un minimum de biais, il est nécessaire de modéliser l'ensemble des facteurs de variation majeurs des performances. Le modèle retenu doit répondre à deux contraintes :

- ✓ Etre opérationnel : c'est-à-dire suffisamment proche de ce que nous connaissons de la réalité biologique, tout en restant suffisamment simple pour obtenir des résultats fiables et sensés.
- ✓ Etre cohérent avec le modèle génétique : c'est-à-dire prendre en compte les connaissances sur le déterminisme génétique du caractère et en particulier les ressemblances des performances entre apparentés.

Toute observation phénotypique peut être décomposée en la somme d'effets du milieu aussi appelé environnement, d'effets génétiques et d'un effet résiduel, et peut être écrit sous la forme :

$$y_{ij} = \mu_i + g_i + e_{ij} \quad (\text{A.1})$$

où  $y_{ij}$  est l'observation  $j$  de l'animal  $i$ ,

$\mu_i$  se rapporte à l'ensemble des effets de milieu identifiés et enregistrés (fixes) de l'animal  $i$ , tels que son sexe, son âge, son troupeau

$g_i$  est la somme des effets génétiques qui regroupe dans un modèle polygénique les effets additifs (effets moyens des gènes contrôlant l'expression du caractère), de dominance (effets d'interaction entre les deux allèles de chaque locus impliqué dans l'expression du caractère) et d'épistasie (effets d'interaction entre allèles de loci différents) du génotype de l'animal  $i$ ,

$e_{ij}$  est l'écart entre  $y_{ij}$  et son espérance selon le modèle  $\mu_i + g_i$ . Il représente la somme des effets d'environnement non identifiés ou non enregistrables (aléatoires) de l'animal  $i$ , il constitue l'erreur du modèle.

La valeur génétique additive dans le terme  $g$  ci-dessus est fonction des gènes transmis par les parents au descendant. C'est l'élément d'intérêt car elle constitue la composante essentielle sur laquelle la sélection animale peut s'appuyer. Les effets de la dominance et de l'épistasie n'étant pas transmissibles et généralement supposés peu importants, ils sont inclus dans la résiduelle du modèle, tel que :

$$y_{ij} = \mu_i + g_{ai} + e_{ij}^* \quad (\text{A.2})$$

où  $e_{ij}^*$  est la somme des effets d'environnement aléatoires, des effets génétiques de dominance et d'épistasie. L'équation (A.2) est celle d'un modèle linéaire, généralement employé dans la plupart des modèles de génétique animale. Classiquement, il est supposé que  $y$  suit une distribution normale multivariée. Les caractères quantitatifs dépendent d'une infinité de gènes sur des locus indépendants et ayant chacun un effet faible. La somme des effets moyens de ces gènes suit une distribution normale de moyenne nulle et de variance  $\sigma_a^2$ . Ce modèle est communément appelé modèle infinitésimal (Fisher, 1918 ; Bulmer, 1985). La résiduelle étant la somme d'effets individuels faibles (les plus importants devant être inclus dans le modèle), elle peut être considérée comme la réalisation d'une variable aléatoire obtenue par tirage dans une population d'effets ayant une distribution donnée de moyenne nulle et de variance  $\sigma_e^2$ . Nous supposons donc que les erreurs sont normalement et indépendamment distribuées.

La prédiction des effets génétiques peut être réalisée suivant différents modèles qui varient en fonction du type de population évalué. Les modèles « simplifiés », ne visant qu'à prédire la valeur génétique d'une partie seulement des animaux de la population étudiée, ont été développés afin d'obtenir des temps de calculs ou des besoins de stockage raisonnables à une époque où la puissance des ordinateurs n'était pas suffisante pour traiter l'ensemble des données. Certains de ces modèles simplifiés sont encore utilisés car ils sont bien adaptés à la structure des données disponibles.

➤ Le modèle père se contente de prédire les effets génétiques des animaux mâles utilisés en tant que père, en utilisant l'information des ses descendants. La taille du système à résoudre est pour obtenir les prédictions de valeur génétique alors considérablement réduite. Les principales limites de ce modèle sont les hypothèses d'absence d'accouplements raisonnés et d'absence de sélection sur les femelles, ce qui représente, en fait, une source importante de biais dans l'évaluation lorsque le progrès génétique est important. La description de la partie génétique étant moins précise, une fraction importante des effets génétiques se retrouve dans la partie non expliquée (résiduelle).

$$\begin{aligned} \text{Dans ce cas : Variance génétique} &= \frac{1}{4} \sigma_a^2 \\ \text{Variance résiduelle} &= \sigma_e^2 + \frac{3}{4} \sigma_a^2 \end{aligned}$$

➤ Le modèle père grand-père maternel permet de prendre partiellement en compte l'origine maternelle à travers le grand-père.

$$\begin{aligned} \text{Dans ce cas : Variance génétique} &= \left(\frac{1}{4} + \frac{1}{16}\right) \sigma_a^2 = \frac{5}{16} \sigma_a^2 \\ \text{Variance résiduelle} &= \sigma_e^2 + \frac{11}{16} \sigma_a^2 \end{aligned}$$

Ce modèle a constitué une première amélioration du modèle père. Il est utilisé pour l'évaluation génétique de nombreux caractères fonctionnels comme la longévité, les facilités de naissance et de vêlage, la mortinatalité.

➤ Le modèle père mère permet d'évaluer le niveau génétique des parents des animaux avec performances et donc d'alléger les calculs par rapport au modèle animal.

$$\text{Dans ce cas : Variance génétique} = \frac{1}{4} + \frac{1}{4} \sigma_a^2 = \frac{1}{2} \sigma_a^2$$

$$\text{Variance résiduelle} = \sigma_e^2 + \frac{1}{2} \sigma_a^2$$

Cette situation est très intéressante lorsqu'un même couple père mère a beaucoup de descendants, ce qui est fréquemment le cas chez les porcs et les volailles.

➤ Au niveau de la prise en compte des parentés, le modèle animal est le modèle le plus complet et constitue la référence. Il utilise au mieux l'ensemble des informations disponibles. Il permet d'évaluer le niveau génétique de l'ensemble des animaux ayant des performances mais aussi de l'ensemble de ses ascendants et de ses descendants avec en tenant compte des coefficients de parenté. Ce modèle, qui est le plus gourmand en ressource informatique est utilisable sur de grandes populations grâce au progrès qu'a connu l'informatique et au développement d'algorithmes de calcul de plus en plus performants.

$$\text{Dans ce cas : Variance génétique} = \sigma_a^2$$

$$\text{Variance résiduelle} = \sigma_e^2$$

Il est utilisé en France pour les évaluations génétiques des caractères de production, de numération cellulaire, de morphologie, de fertilité et de l'ISU.

## **B/ La meilleure prédiction linéaire non biaisée de la valeur génétique**

Les modèles statistiques les plus utilisés en génétique quantitative sont appelés "modèles mixtes" car les effets fixes (effets de milieu), et les effets aléatoires (valeurs génétiques, résiduelle) sont estimés simultanément (Eisenhart, 1947). En 1953, Henderson a publié dans une méthode générale d'estimation des composantes de la variance dans le cadre d'un modèle mixte. En 1963, il en a donné une formulation générale sous forme matricielle d'un prédicteur de l'ensemble des valeurs génétiques des animaux d'une population appelé BLUP (meilleure prédiction linéaire non biaisée) pour 'Best Linear Unbiased Prediction' (Henderson, 1963). 'Best' car il minimise l'erreur quadratique moyenne liée à l'estimation, 'Unbiased' dans le sens où l'espérance de la valeur génétique vraie est égale à cet index ( $E(a|\hat{a}) = \hat{a}$ ), et 'Linear' parce que cet estimateur peut être écrit comme une combinaison linéaire des données.

Jusqu'au milieu des années 1970, ces résultats théoriques n'ont pu faire l'objet d'une pleine application du fait des limites informatiques. Henderson (1976) et Quaas (1976) ont montré que sous certaines conditions, la méthode BLUP met en œuvre des matrices d'une structure particulière (matrices très simples et creuses, dont la plupart des éléments sont nuls et pour laquelle l'inverse peut être obtenu directement sans difficultés), ce qui permet de limiter le nombre d'éléments à stocker en mémoire et d'améliorer la performance des algorithmes itératifs. Ces résultats ont permis la généralisation de l'utilisation d'un modèle père d'abord,



puis d'un modèle animal. Avec les progrès informatiques, les modèles se sont complexifiés et peuvent prendre en compte simultanément plusieurs variables (modèle multicaractères).

### B1/ Le modèle animal unicaractère avec un seul effet aléatoire

Dans un modèle animal unicaractère visant à estimer simultanément les effets fixes et prédire l'effet aléatoire, la performance peut être écrite sous la forme matricielle comme :

$$\mathbf{y} = \mathbf{X} \mathbf{b} + \mathbf{Z} \mathbf{a} + \mathbf{e} \quad (\text{A.3})$$

où :  $\mathbf{y}$  est le vecteur des observations élémentaires,

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

$\mathbf{a}$  est le vecteur des valeurs génétiques additives de tous les animaux que nous cherchons à évaluer, qu'ils possèdent une performance ou non,

$\mathbf{e}$  est le vecteur des résiduelles, distribuées suivant une loi normale de moyenne nulle,

$\mathbf{X}$  et  $\mathbf{Z}$  sont les matrices d'incidence qui relient les observations aux effets fixes et aléatoires qui les ont influencées.

où :  $E(\mathbf{a}) = \mathbf{0}$        $\text{Var}(\mathbf{a}) = \mathbf{A} \sigma_a^2$       où  $\mathbf{A}$  est la matrice de parenté,

$E(\mathbf{e}) = \mathbf{0}$        $\text{Var}(\mathbf{e}) = \mathbf{I} \sigma_e^2$       où  $\mathbf{I}$  est la matrice identité, les résiduelles sont donc indépendantes (non corrélées).

Les solutions du BLUP sont obtenues en résolvant les « équations du modèle mixte » [membre de gauche = membre de droite] qui s'écrivent :

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \alpha \mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix} \quad (\text{A.4})$$

Où  $\alpha = \sigma_e^2 / \sigma_a^2 = (\sigma_y^2 - \sigma_a^2) / \sigma_a^2 = (1 - h^2) / h^2$  où  $h^2$  est l'héritabilité du caractère, c'est à dire la part de la variabilité totale qui est génétiquement transmissible.

### B2/ Le modèle animal unicaractère avec plusieurs effets aléatoires

Dans le modèle animal décrit précédemment (A.3), les résiduelles sont supposées non corrélées et homoscédastiques (de variance homogène). Or, dans certains cas, les facteurs d'environnement constituent une composante importante de la covariance entre performances. C'est notamment le cas entre les différentes lactations d'une vache ou entre les différentes portées d'une truie. Ces effets environnementaux sont pris en compte en décomposant la résiduelle  $\mathbf{e}$  du modèle (A.3) en un effet appelé environnement permanent  $\mathbf{p}$  qui regroupe les effets non génétiques (et génétiques non additifs) propres à un animal et influençant de la même manière les performances d'un même animal et une nouvelle résiduelle de variance inférieure, tel que :

$$\mathbf{y} = \mathbf{X} \mathbf{b} + \mathbf{Z} \mathbf{a} + \mathbf{W} \mathbf{p} + \mathbf{e} \quad (\text{A.5})$$

où :  $\mathbf{p}$  est le vecteur des valeurs d'effet permanent de tous les animaux avec performance,

$\mathbf{W}$  est la matrice d'incidence qui relie les observations aux effets d'environnement permanent qui les ont influencées.

où :  $E(\mathbf{p}) = \mathbf{0}$        $\text{Var}(\mathbf{p}) = \mathbf{I} \sigma_p^2$

Les équations du modèle mixte s'écrivent alors :

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} & \mathbf{X}'\mathbf{W} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \alpha\mathbf{A}^{-1} & \mathbf{Z}'\mathbf{W} \\ \mathbf{W}'\mathbf{X} & \mathbf{W}'\mathbf{Z} & \mathbf{W}'\mathbf{W} + \gamma\mathbf{I} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{p}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \\ \mathbf{W}'\mathbf{y} \end{bmatrix} \quad (\text{A.6})$$

Où  $\gamma = \sigma_e^2 / \sigma_p^2 = (\sigma_y^2 - (\sigma_a^2 + \sigma_p^2)) / \sigma_p^2 = (1-r) / (r-h^2)$  et  $\alpha = (1-r) / h^2$  où  $r$  est la répétabilité du caractère, part de la variabilité totale qui est spécifiquement rattachée à l'animal en question.

### B3/ Le modèle animal unicaractère avec des groupes de parents inconnus

Certains animaux figurant dans la généalogie ont des parents inconnus, généralement considérés comme appartenant à la population de base. L'utilisation de la matrice de parenté  $\mathbf{A}$  dans un modèle animal suppose que ces animaux proviennent d'une seule population ayant une valeur génétique moyenne nulle et une variance génétique commune  $\sigma_a^2$ . Les valeurs génétiques des animaux des générations suivantes sont exprimées relativement à cette population de base. Or, lorsque les animaux de cette population de base proviennent de populations avec des niveaux génétiques différents, par exemple des taureaux de différents pays ou nés à des périodes éloignées, cette différence doit être prise en compte dans le modèle pour éviter de possibles biais.

Quaas (1988) a proposé une procédure de constitution des groupes basée sur l'origine géographique, l'année de naissance de l'animal, le sexe et éventuellement la voie de sélection (pères à taureaux, mères à vaches...) afin de refléter le niveau génétique d'un groupe homogène d'individus. L'écriture matricielle de (A.3) incluant les groupes de parents inconnus est :

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{Q}\mathbf{g} + \mathbf{Z}\mathbf{a}^* + \mathbf{e} \quad (\text{A.7})$$

où :  $\mathbf{g}$  est le vecteur des valeurs génétiques additives de groupes de parents inconnus,  
 $\mathbf{Q}$  est la matrice d'incidence qui relie les ancêtres inconnus aux groupes ainsi formés.

où :  $E(\mathbf{a}^*) = \mathbf{0}$  ou bien  $E(\mathbf{a}) = \mathbf{Q}\mathbf{g}$ , avec  $\mathbf{a} = \mathbf{Q}\mathbf{g} + \mathbf{a}^*$

Les équations du modèle mixte s'écrivent alors :

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z}\mathbf{Q} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Q}'\mathbf{Z}'\mathbf{X} & \mathbf{Q}'\mathbf{Z}'\mathbf{Z}\mathbf{Q} & \mathbf{Q}'\mathbf{Z}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z}\mathbf{Q} & \mathbf{Z}'\mathbf{Z} + \alpha\mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{g}} \\ \hat{\mathbf{a}}^* \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Q}'\mathbf{Z}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix} \quad (\text{A.8})$$

### B4/ Le modèle animal multicaractères

La sélection animale repose généralement sur une combinaison de caractères ayant un intérêt économique qui peuvent être phénotypiquement et/ou génétiquement liés. L'évaluation multicaractères est l'approche de choix dans ce type de situation, puisque elle permet, à partir des observations sur un premier caractère, d'enrichir nos connaissances des effets génétiques et de milieu sur les autres caractères et vice versa. Les principaux avantages du modèle multicaractères sont d'une part la prise en compte de la sélection réalisée sur les caractères corrélés et d'autre part de permettre l'amélioration de la précision des prédictions. Le principal inconvénient est quant à lui ses larges besoins en terme de ressources informatiques. Ce modèle peut s'écrire sous la forme :

$$\mathbf{y}_i = \mathbf{X}_i \mathbf{b}_i + \mathbf{Z}_i \mathbf{a}_i + \mathbf{e}_i \quad (\text{A.9})$$

avec  $i$  correspondant au caractère 1 à  $n$ , les équations du modèles mixtes étant :

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{A}^{-1} \otimes \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \quad (\text{A.10})$$

où :  $E(\mathbf{a}) = 0$        $\text{Var} \begin{bmatrix} \mathbf{a} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{A} \otimes \mathbf{G} & 0 \\ 0 & \mathbf{I} \otimes \mathbf{R} \end{bmatrix}$       où  $\mathbf{G}$  est la matrice de covariance génétique  
 $E(\mathbf{e}) = 0$        $\mathbf{R}$  est la matrice de covariance résiduelle

$\otimes$  est le produit de kronecker

avec  $\mathbf{A} \otimes \mathbf{G} = \text{Var} \begin{bmatrix} a_i \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \mathbf{A} \sigma_{a_i}^2 & \cdots & \mathbf{A} \sigma_{a_i a_n} \\ \vdots & \ddots & \vdots \\ \mathbf{A} \sigma_{a_n a_i} & \cdots & \mathbf{A} \sigma_{a_n}^2 \end{bmatrix} \quad (\text{A.11})$

et  $\mathbf{I} \otimes \mathbf{R} = \text{Var} \begin{bmatrix} e_i \\ \vdots \\ e_n \end{bmatrix} = \begin{bmatrix} \mathbf{I} \sigma_{e_i}^2 & \cdots & \mathbf{I} \sigma_{e_i e_n} \\ \vdots & \ddots & \vdots \\ \mathbf{I} \sigma_{e_n e_i} & \cdots & \mathbf{I} \sigma_{e_n}^2 \end{bmatrix} \quad (\text{A.12})$

## B5/ De nombreuses propriétés

Le BLUP reprend les principales propriétés des méthodes développées antérieurement telles que les indices de sélection (Hazel, 1943) comme la maximisation de la corrélation entre les valeurs génétiques vraies et les valeurs génétiques prédites, la maximisation de la probabilité d'un classement correct des animaux deux à deux, la maximisation de l'espérance du progrès génétique pour un cycle de sélection. Cependant, les méthodes antérieures ne possédaient ces propriétés que lorsque certaines hypothèses très restrictives étaient faites. Le BLUP permet de s'affranchir d'une d'entre elles (connaissance des effets fixes), grâce à l'estimation simultanée de l'ensemble des effets génétiques et des effets de milieu inclus dans le modèle d'analyse. Les biais liés à une mauvaise connaissance *a priori* des effets de l'environnement sont ainsi éliminés. La répartition non aléatoire des valeurs génétiques parmi les différentes modalités des effets de milieu y est aussi prise en compte.

L'utilisation d'un modèle animal permet, lorsque l'analyse tient compte de l'ensemble de la population depuis son origine (avant toute sélection) et jusqu'aux animaux les plus jeunes, ainsi que de l'ensemble des relations de parentés entre animaux, de prendre en compte l'évolution de la moyenne et de la variance génétique sous l'effet de la sélection, de la consanguinité, des accouplements dirigés ou du hasard (Sorensen and Kennedy, 1982). En effet, l'évolution de la variance génétique au cours des générations, suite à la sélection, à la dérive génétique et à la consanguinité est directement prise en compte à travers l'inverse de la matrice de covariance  $\mathbf{G}$  de valeurs génétiques additives.

Toute indexation suppose qu'au préalable, les paramètres génétiques (les variances et covariances) des différents effets aléatoires (génétiques, environnement permanent et résiduels) du modèle aient été estimés.

## **C/ Les méthodes d'estimation des paramètres génétiques**

Généralement, un nombre conséquent de données (fonction de leur structure et des conditions de mesures) est nécessaire pour l'estimation des composantes de variances : de quelques centaines pour des expériences où le milieu est contrôlé, à plusieurs centaines de milliers pour des données où les effets de milieu peuvent être très variables comme la fertilité.

L'estimation des paramètres génétiques n'est pas toujours jugée nécessaire lorsque des analyses sur des populations similaires ont déjà été réalisées. En effet, les estimations réalisées à partir d'un nombre restreint de données sont généralement moins pertinentes que des données bibliographiques obtenues dans des conditions similaires sur de grands effectifs. Toutefois, cette étape est nécessaire lorsque nous nous intéressons à de nouveaux caractères ou que la population à estimer a une origine génétique différente des autres. De même, du fait de la sélection, les variances et covariances évoluent, ce qui implique des ré-estimations régulières de celles-ci.

Les méthodes d'analyse de variances type ANOVA (pour ANalysis Of VAriance) ne sont plus utilisées, car les données de génétique animale sont déséquilibrées, et proviennent généralement de populations sélectionnées. Les résultats obtenus seraient alors biaisés car l'ANOVA est basée sur l'hypothèse d'échantillons aléatoires. Les méthodes actuelles sont généralement basées sur la notion de vraisemblance développée par Fisher (1925). Elle se définit comme une fonction de paramètres, les données étant considérées comme des paramètres connus. L'application de la théorie du maximum de vraisemblance à l'estimation des composantes de la variance en modèle linéaire a été proposée initialement par Hartley et Rao (1967). La méthode la plus couramment utilisée est l'approche REML (Restricted Maximum Likelihood : maximum de vraisemblance restreinte) modèle animal qui permet d'aboutir aux estimées de composantes de variance-covariance ayant la vraisemblance maximale, en prenant en compte la perte d'information liée à l'estimation des effets de milieu. D'autres méthodes, comme les méthodes Bayésiennes, constituent des alternatives possibles aux approches REML. Elles consistent à baser l'estimation des paramètres sur une distribution marginale *a posteriori* mélangeant l'information provenant des données à une information sur la distribution *a priori*. Ces méthodes, très simples à mettre en œuvre et nécessitant que peu de ressources mémoires offrent une alternative à l'utilisation d'approches REML dans le cadre de fichiers ayant une taille importante (Sorensen and Gianola, 2002). N'ayant pas été utilisée au cours de ces travaux, les méthodes Bayésiennes ne seront pas développées dans cette brève présentation.

### **C1/ Maximum de vraisemblance (ML)**

Les méthodes basées sur le maximum de vraisemblance (ML - Maximum Likelihood) permettent de prendre en compte, dans certains cas, l'effet de la sélection, si toutes les données qui ont contribué aux décisions de sélection et corrélées avec le ou les caractères analysés sont incluses dans l'analyse. Cependant, même si ces conditions ne sont que partiellement remplies, l'estimation ML est souvent considérée être moins biaisée par la

sélection que l'estimation par ANOVA. Les estimateurs du maximum de vraisemblance sont asymptotiquement convergents (sans biais), efficaces et normaux (Harville, 1977).

L'estimation ML repose sur un concept très simple, mais nécessite des hypothèses sur la distribution des données (par exemple, distribution normale). Pour un modèle d'analyse particulier, nous pouvons décrire la probabilité sachant les paramètres d'observer les données récoltées. L'estimation du maximum de vraisemblance est, par définition, la valeur des paramètres pour laquelle la densité de probabilité d'observer les données réelles est maximale. Cependant, l'estimation ML a un inconvénient majeur : elle maximise la vraisemblance globale (pour l'ensemble des paramètres du modèle : effets fixes et composante de (co)variances). Elle sous-estime donc les composantes de (co)variances lorsque les effectifs par niveau d'effets fixes sont limités, car elle ignore la perte de degrés de liberté liée à l'estimation de ceux-ci.

## **C2/ Maximum de Vraisemblance Restreint (REML) et les méthodes dérivées**

Le maximum de vraisemblance restreint (REML – Restricted Maximum Likelihood) décrit par Patterson et Thompson (1971), est une méthode qui reprend des propriétés du maximum de vraisemblance, mais qui contrairement au ML prend en compte la perte de degrés de liberté due aux effets fixes, en maximisant uniquement la vraisemblance d'une fonction des données corrigées des effets fixes. Les estimateurs REML supposent que la distribution des observations suit une loi normale. En cas d'écart important à la normalité, un biais non négligeable peut affecter les estimations des composantes de variance et de covariance.

Le REML, comme le ML, est une méthode relativement flexible puisqu'elle peut être appliquée à tous les types de modèle linéaire mixte.

Les besoins informatiques pour le calcul sont encore plus élevés que pour les méthodes ML. Les applications ont donc été limitées pendant de nombreuses années du fait de la nécessité d'inverser de façon répétée des matrices creuses de grande taille intervenant dans les algorithmes de recherche des maxima de vraisemblances restreintes. Depuis les années 1980, de nombreux algorithmes efficaces ont été développés pour faciliter le calcul du REML. Les avancées du REML ont suivi le développement des évaluations BLUP et à l'utilisation du modèle animal. Ces algorithmes sont généralement classés selon leur niveau d'utilisation des dérivées de la fonction log-vraisemblance résiduelle (allant de la non-utilisation de dérivés à l'emploi des dérivées premières et secondes).

Parmi les synthèses réalisées sur l'estimation des composantes de la variance en génétique animale, Bidanel (1994) et Hofer (1998) ont présenté les différents algorithmes d'optimisation existant en explicitant l'intérêt et les critères de choix de chacun d'entre eux. Ils permettent tous l'utilisation des logiciels de traitement des matrices creuses. Ils sont caractérisés par leur simplicité algorithmique, leur performance sur le plan informatique (taille maximale susceptible d'être traitée, efficacité d'utilisation des ressources informatiques), leur robustesse (aptitude à trouver effectivement le maximum de la fonction dans toutes les situations), leur stabilité numérique et leur flexibilité (aptitude à traiter différents types de modèles). De façon

générale, la vitesse de convergence, la stabilité numérique, la robustesse et les temps de calcul par itération augmentent avec une utilisation croissante de l'information sur les dérivées.

Les méthodes DF (Derivative Free : sans dérivée) développées par Graser *et al.* (1987) ne nécessitent pas le calcul explicite des dérivées. L'optimum de la fonction est localisé sans avoir recours à leur calcul. D'après Robert (1996), les composantes estimées sont obtenues par l'évaluation explicite de la fonction de vraisemblance dont nous cherchons le maximum par des techniques générales d'optimisation à travers la recherche de meilleures solutions dans toutes les directions de l'espace de paramètres. Ces méthodes présentent une grande flexibilité vis-à-vis du modèle utilisé. Cependant, la vitesse de convergence est généralement faible et des problèmes de robustesse et de stabilité numérique sont parfois rencontrés sur les applications multicaractères (Misztal, 1994).

L'algorithme EM (Expectation Maximisation) a été développé par Dempster *et al.* (1977) en vue d'obtenir les solutions du maximum de vraisemblance en présence de données manquantes, mais son cadre dépasse très largement le calcul de composantes de variance. Il nécessite le calcul à chaque itération des dérivées premières (inversion de blocs correspondant aux effets aléatoires de la matrice des coefficients du modèle mixte). La vitesse de convergence peut être relativement faible, mais cet algorithme est plus robuste que les méthodes précédemment citées. Les estimées des paramètres génétiques sont toujours situées dans l'espace des paramètres (Harville, 1977). La possibilité d'utiliser des méthodes de traitement des matrices creuses et des procédures d'accélération de la convergence le rendent extrêmement performant dans certaines situations.

Les méthodes basées sur les dérivées secondes (méthodes de Newton-Raphson ou des scores de Fisher) sont en général fortement pénalisées par la lourdeur des calculs qu'implique le calcul de dérivées secondes ou de leurs espérances, qui dépendent de l'inverse de la matrice des coefficients du modèle mixte. Cependant, elles présentent l'avantage d'avoir une vitesse de convergence quadratique (nombre d'itérations très faible). D'après Robert (1996), un autre inconvénient de ces deux algorithmes est qu'ils ne garantissent pas que les paramètres estimés appartiennent à l'espace des paramètres, mais ceci peut être modifié en adaptant l'algorithme de sorte qu'il contraigne les estimations à y rester (Harville, 1977).

Il existe d'autres méthodes, comme l'AI-REML (information moyenne pour Average Information) (Johnson and Thompson, 1995) qui est un mélange des deux algorithmes précédents, mais dont les besoins en calcul sont proches de ceux des algorithmes nécessitant le calcul de dérivées premières, tout en conservant une vitesse de convergence très intéressante (généralement moins de 10 itérations).

## Références

- Bidanel J.P., 1994. Estimation des paramètres génétiques sous un modèle animal. In Séminaire *Modèle animal*, J.L. Foulley et M. Molénat eds, pp. 25-36.
- Bulmer M.G. 1985. The mathematical theory of quantitative genetics. Clarendon Press, Oxford.
- Dempster A.P., Laird N.M. and Rubin D.B. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*. B39: 1-38.
- Eisenhart C. 1947. The assumptions underlying the analysis of variance. *Biometrics*. 3: 1-21.
- Fisher R.A. 1918. The Correlation Between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh*. 52: 399-433.
- Fisher R.A. 1925. Statistical methods for research workers. Oliver and Boyd. Edinburgh and London.
- Graser H.-U., Smith S.P. and Tier B. 1987. A derivative-free approach for estimating variance components in animal models by restricted maximum likelihood. *Journal of Animal Science*. 64: 1362-1370.
- Hartley H.O. and Rao J.N.K. 1967. Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika*. 54: 93-108.
- Harville D.A. 1977. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of American Statistics Association*. 72: 320-338.
- Hazel L.N. 1943. The genetic basis for constructing selection index. *Genetics*. 28: 476-490.
- Henderson C.R. 1963. Selection index and expected genetic advance. In *Statistical Genetics and Plant Breeding*. W.D. Hanson and H.F. Robinson eds., pp. 141-163.
- Henderson C.R. 1976. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics*. 32: 69-83.
- Hofer A. 1998. Variance component estimation in animal breeding: a review. *Journal of Animal Breeding and Genetics*. 115: 247-265.
- Johnson D.L. and Thompson R. 1995. Restricted maximum likelihood estimation of variance components for univariate animal models using sparse matrix techniques and average information. *Journal of Dairy Science*. 78: 449-456.
- Misztal I. 1994. Comparison of computing properties of derivative and derivative-free algorithms in variance-component estimation by REML. *Journal of Animal Breeding and Genetics*. 111: 346-355.
- Patterson H.D. and Thompson R. 1971. Recovery of inter-block estimation when block sizes are unequal. *Biometrika*, 58: 545-554.
- Quaas R.L. 1976. Computing the diagonal elements and inverse of a large numerator relationship matrix. *Biometrics*. 32: 949-953.
- Robert C. 1996. Etude de quelques problèmes liés à la mise en oeuvre du REML en génétique quantitative. Thèse de Doctorat, Université Paul Sabatier Toulouse.
- Sorensen D. and Gianola D. 2002. Likelihood, Bayesian and MCMC methods in quantitative Genetics. Berlin: Springer.
- Sorensen D. and Kennedy B.W. 1982. Estimation of genetic variances in control and selected populations. In: Proceedings of the *World Congress on Genetic Applied to Livestock Production*, Madrid, Spain. VII: 220-225.

## FORMATIONS SUIVIES

### **Approfondissements scientifiques et techniques, connaissance du milieu professionnel**

2004	Module ABIES « Write Right » (D. White)	3 jours
2005	Formation Bibliographie (INRA Jouy-en-Josas)	4 jours
	Séminaire département Génétique Animale sur les méthode MCMC (C. Guihenneuc)	4 jours
	Cours Supérieur d'Amélioration Génétique des Animaux Domestiques (CSAGAD) « Amélioration génétique des ruminants laitiers » (Rennes)	5 jours
	Formation Unix (INRA Jouy-en-Josas)	2 jours
2006	Module ABIES « La thèse, au service de mon projet professionnel et personnel » (E. Birlouez)	3 jours
	Cours de programmation « Computational techniques in animal breeding » (INRA Jouy-en-Josas ; I. Misztal)	5 jours
	CSAGAD « Après la Loi d'Orientation Agricole, quelle organisation Et quelle réglementation pour la sélection animale » (Paris)	1 jour
2007	Formation SAS « Production d'états » (INRA Jouy-en-Josas ; T. Brun)	2 jours

### **Ouverture scientifique et culturelle**

2005	Journée ABIES	2 jours
	Séminaire des thésards du département de Génétique Animale (Toulouse)	2 jours
2006	Congrès Mondial « <i>World Congress on Genetic Applied to Livestock Production</i> » (Brésil)	5 jours
	Séminaire des thésards du département de Génétique Animale (Limoges)	2 jours
2007	Congrès Interbull (Irlande)	2 jours
	Congrès de la Fédération Européenne de Zootechnie (Irlande)	4 jours
	Séminaire des thésards du département de Génétique Animale (Jouy-en-Josas)	2 jours
2008	Journées Scientifiques du département de Génétique Animale (Lacanau)	4 jours





## RESUME

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.*

Les modèles d'évaluation génétique des caractères de production laitière et de comptages cellulaires chez les bovins ont considérablement évolué depuis 10 ans. La plupart des pays ont mis en place (Allemagne, Canada, Danemark, Italie, Nouvelle-Zélande, Pays-Bas...) ou sont en train de mettre en place (France) des évaluations exploitant directement la production à chaque contrôle au lieu de la production cumulée en 305 jours. Le principal intérêt des modèles de contrôles élémentaires réside dans une meilleure prise en compte des effets systématiques de l'environnement de chacun des contrôles. Les valeurs génétiques estimées sont donc plus précises. La France a fait le choix d'un modèle original par sa simplicité d'interprétation et sa flexibilité d'utilisation. Les courbes de lactation sont décomposées en une somme d'effets propres à la vache (caractéristiques génétiques et non génétiques) et d'environnement (troupeau, race, région, rang de lactation, âge et mois de vêlage, durée de tarissement, gestation...). Ces effets sont calculés pour les stades de lactation allant de 7 à 335 jours et sont corrigés pour tous les autres facteurs de variation ce qui les rend directement interprétables. Toutefois, ces modèles sont informatiquement très coûteux compte tenu de leur niveau de sophistication et du nombre considérable de données élémentaires à analyser simultanément.

Le premier objectif de cette thèse a été de valider le logiciel d'évaluation génétique développé pour cette application et d'étudier les possibilités de simplification du modèle afin de limiter les besoins informatiques. L'exactitude des calculs mis en œuvre a été validée à l'aide d'une approche novatrice basée sur la reconstruction de données simulées répondant à des caractéristiques spécifiques telles que les effets estimés par le logiciel sont exactement ceux simulés si celui-ci est correct. Compte tenu des limites informatiques, le nombre total d'effets aléatoires du modèle a été réduit en utilisant une approche de réduction du rang des matrices de (co)variances de ces effets aléatoires sans toutefois entraîner de variations notables des valeurs génétiques estimées. Nous avons d'autre part mis en évidence que seul le niveau (et non la forme) des effets fixes modélisés par des courbes fonctions du stade de lactation dépend de l'année du contrôle. Cette constatation permet une précorrection des données pour les effets liés aux courbes de lactation dépendant de l'âge, du mois de vêlage, de la durée de tarissement et de la gestation. Cette modification a permis d'accroître la stabilité des effets estimés et de réduire de façon importante les temps de calcul en rendant possible la réalisation de l'évaluation génétique en deux étapes : l'étape de précorrection est suivie d'une analyse des données précorrigées à partir d'un modèle incluant l'ensemble des effets aléatoires et les seuls effets fixes qui dépendent de l'année.

Outre des gains de précision des valeurs génétiques estimées – probablement assez modestes –, ces modèles sophistiqués et informatiquement très contraignants offrent des perspectives extrêmement intéressantes en matière de valorisation des données collectées en appui technique dont l'étude constituait le deuxième objectif de cette thèse. Les travaux réalisés ont permis de mettre en avant de multiples perspectives de ce point de vue. En effet, grâce à la modélisation originale des effets fixes comme somme de courbes spécifiques, le modèle d'évaluation génétique sur les contrôles élémentaires permet d'envisager le développement de nouveaux outils pour le conseil en élevage où les prédictions de production seraient réalisées à partir des effets estimés lors des évaluations. Ces nouveaux outils pourraient également inclure un module d'analyse dynamique de l'effet de la conduite du troupeau à travers l'effet « troupeau – jour de contrôle » dont nous avons pu mettre en évidence la bonne prédictibilité (corrélation autour de 0.8 entre valeur prédite et valeur estimée *a posteriori*). L'association des résultats d'analyses réalisées dans le cadre des évaluations génétiques nationales sur divers caractères (production, fertilité, santé de la mamelle) ouvre de nouvelles voies de recherche regroupant l'ensemble des thématiques ayant trait à la production laitière : génétique, nutrition, physiologie, santé...

## ABSTRACT

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

Genetic evaluation models for milk production traits and cell counting in cattle have evolved considerably over the last 10 years. Most countries (Germany, Canada, Denmark, Italy, New Zealand, The Netherlands ...) have developed or are currently settling (France) evaluations that directly use each test-day production instead of using cumulated production over 305 days. The main interest of test-day models is in the fact that they have an increased consideration for systematic environmental effects of each test-day. The estimated genetic values are therefore more accurate. France has chosen an original model that is easy to interpret and flexible. The lactation curves are made up of a sum of cow's (genetic and non-genetic traits) and environment's effects (herd, breed, region, lactation number, age and calving month, length of dry period, pregnancy...). These effects are determined for the days in milk from 7 to 335 days and are corrected for all other factors of variation making them directly interpretable. However, the models are costly in terms of computing needs due to their sophistication and the large number of data to be analysed simultaneously.

The first objective of this thesis was to validate the software for genetic evaluation developed for this application and to study the possibilities of simplifying the model in order to limit computing needs. The exactness of the calculations performed were validated using a new approach based on the rebuilding of simulated data fulfilling specific conditions so that the effects estimated by the software are exactly those simulated if the evaluation software is correct. Considering the computing limits, the total number of random effects of the model was reduced using a reduced rank approach on the (co)variances matrix of these random effects without leading to important variation of the estimated genetic values. We also show that only the level (and not the form) of the fixed effects modelled by days-in-milk curves depend on test-day year. This observation allows a pre-correction of the data for the effects associated with lactation curves depending on age, calving month, length of dry period and pregnancy. This modification has allowed to increase the stability of the effects estimated and to decrease in an important way the computing time, by making two-step genetic evaluation possible: the pre-correction step is followed by an analysis of the pre-corrected data in a model including all the random effects and the fixed effects that depend on the year.

Besides the gain in accuracy of the estimated genetic values – which are probably modest –, the main benefits of this sophisticated and highly computationally demanding models are related to prediction and herd management support which was the second objective of this thesis. The studies performed show the different perspectives of this view point. Indeed, due to the original modelling of the fixed effects as sums of specific curves, the genetic evaluation test-day model favours the development of new tools for herd management where production predictions will be made using estimated effects from genetic evaluation. These new tools could also include a dynamic analysis of the herd management via the “herd – test-day” effect for which we show a good predictability (a correlation around 0.8 between the predicted and estimated *a posteriori* value). The association of results of national genetic evaluations for different traits (production, fertility, mammary gland health) lead to new research that regroups all fields related to dairy production: genetic, nutrition, physiology, health...