

Utilisation des données de contrôles élémentaires pour la modélisation et l'estimation des interactions génotype x milieu : Etude en bovins laitiers

Bérénice Huquet

▶ **To cite this version:**

Bérénice Huquet. Utilisation des données de contrôles élémentaires pour la modélisation et l'estimation des interactions génotype x milieu : Etude en bovins laitiers. Génétique animale. AgroParisTech, 2012. Français. NNT : 2012AGPT0060 . pastel-00947191

HAL Id: pastel-00947191

https://pastel.hal.science/pastel-00947191

Submitted on 14 Feb 2014

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.

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

AgroParisTech
INSTITUT DES SCIENCES ET INDUSTRIES DU VIVANT ET DE L'ENVIRONNEMENT
PARIS INSTITUTE OF TECHNOLOGY FOR LIFE, FOOD AND ENVIRONMENTAL SCIENCES

Ecole Doctorale ABIES
Agriculture
Agriculture
Alimentation
Food
Biologie
Biology
Environnement
Environment
Santé
Health

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

L'Institut des Sciences et Industries du Vivant et de l'Environnement

(AgroParisTech)

Spécialité : Génétique animale

présentée et soutenue publiquement par

Bérénice HUQUET

le 8 octobre 2012

Utilisation des données de contrôles élémentaires pour la modélisation et l'estimation des interactions génotype x milieu

Étude en bovins laitiers

Directeur de thèse : **Vincent DUCROCQ**

Avec l'appui d' **Hélène LECLERC** (Institut de l'Elevage)

Représentant CIFRE Institut de l'Elevage : **Sophie MATTALIA**

Jury

M. Etienne VERRIER, Professeur, AgroParisTech	Président
M. Roger CUE, Professeur, McGill University, Canada	Rapporteur
M. Loys BODIN, Ingénieur de recherche, INRA	Rapporteur
Mme Christèle ROBERT-GRANIÉ, Directrice de recherche, INRA	Examinateur
Mme Sophie MATTALIA, Responsable du service génétique laitière, Institut de l'Elevage	Examinateur
M. Vincent DUCROCQ, Directeur de recherche, INRA	Examinateur

AgroParisTech
INRA GABI
Domaine de Vilvert, 78350 Jouy-en-Josas, France

Cette thèse s'inscrit dans le cadre du **programme CASDAR GENESYS** associant : l'Institut de l'Elevage (idele), l'Institut National de la Recherche Agronomique (INRA) et France Conseil ELevage (FCEL).



Ce travail a été effectué dans le cadre d'une thèse sous **contrat CIFRE** (966/2009). L'Agence Nationale de la Recherche et de la Technologie (ANRT) a subventionné l'Institut de l'Elevage (idele) pour favoriser le développement de la recherche partenariale publique privée entre idele et l'INRA. La thèse a été réalisée au sein du laboratoire INRA UMR 1313 Génétique et Biologie Intégrative (GABI) et de l'équipe « G2B » Génétique et Génomique Bovine sur le centre de Jouy en Josas.



Les devises Shadok
En essayant continuellement on finit par réussir.
Donc : plus ça rate, plus on a de chances que ça marche.
Jacques Rouxel (studio aaa production)

Remerciements

Voilà 3 ans déjà, je me suis lancée dans cette drôle d'aventure : une thèse ! A l'époque, je ne savais pas vraiment ce que cela représentait. J'y voyais surtout une première expérience professionnelle, dans le secteur agronomique, au sein d'un projet ambitieux à finalité pratique, avec des partenaires variés avec qui j'avais apprécié travailler pendant mon stage de fin d'études. Je ne m'étais pas trompée mais cela allait bien au delà de ça ! Trois ans de découvertes : découverte de toute une filière qui m'était inconnue, découverte de la recherche, première expérience du monde professionnel... Ce que je n'avais pas envisagé, c'était la découverte de moi-même, moi en tant qu'actrice du monde professionnel ! Alors cette thèse, c'est certes 3 ans de travail de recherche mais c'est surtout, et avant tout, 3 ans d'épanouissement personnel et professionnel.

Merci à Jean-Pierre Bidanel (directeur de l'unité GABI du département de génétique animale à l'INRA) d'une part, à Pierre-Louis Gastinel (responsable du département de génétique à l'Institut de l'Élevage) d'autre part, d'avoir uni leurs forces pour monter un contrat CIFRE au sein de l'UMT 3G : cette synergie entre les organismes est un élément clé de la réussite de cette thèse.

*Les défis scientifiques et méthodologiques n'ont pas manqué lors de cette aventure, c'est ce qui l'aura rendue palpitante et difficile à la fois. Sur ce chemin semé d'embûches, j'ai toujours pu compter sur des hommes et des femmes bienveillants, prêts à me montrer la lumière au bout du tunnel ! Vincent, Hélène, Sophie, Ingrid, Luc, Jocelyn, Yann, Étienne, Philippe, Marion en tant que membres de mon comité de thèse ainsi que tous les membres du comité de pilotage de Genesys, des groupes de travail de G2B, je tiens à vous remercier pour les conseils, idées, remarques que vous avez pu m'apporter lors des comités de thèse ou réunions de travail, et surtout pour le regain de motivation qu'ont engendré chez moi ces réunions. Une thèse est longue, ces réunions permettent de faire le point, de se rendre compte du travail effectué et d'envisager plus sereinement la suite. La méthode c'est bien, mais il faut qu'elle soit utile ! Alors merci aux membres de l'équipe idele-FCEL de Genesys grâce à qui mon travail a toujours gardé un ancrage terrain. Cette thèse a répondu à certains des défis que nous nous étions lancés. Nous avons exploré une piste innovante qui a abouti, mais grâce aux rapports et aux questions des membres du jury de thèse, je me suis bien rendue compte que tout était améliorable et que les nouvelles pistes d'exploration des interactions Génotype*Milieu ne manquaient pas ! Nous sommes loin d'avoir percé le mystère des interactions Génotype*Milieu ! Alors même si la critique est parfois difficile à entendre, je pense que vos commentaires et questions m'ont beaucoup aidé à prendre à nouveau du recul sur mon travail, à aborder ces problématiques d'un autre point de vue et je vous en remercie.*

Parce qu'une thèse, ce n'est pas seulement de la science, j'ai aussi beaucoup appris sur la conduite de projet, la gestion d'équipe, la communication, la formation... Pour cela, j'ai eu la chance d'avoir des encadrants suffisamment ouverts d'esprit pour me laisser développer ce type de compétences à travers des formations (parfois longues !) ou des prises de responsabilité au sein de l'entreprise et du monde associatif, qui m'éloignaient du coeur scientifique du projet de thèse. Merci à vous, Sophie et Vincent, de m'avoir accompagnée dans ces démarches. En étant présidente de Doc'J, j'ai bien pu me rendre compte que tous les doctorants ne bénéficiaient pas de cette liberté, de cette confiance et de ce soutien que vous m'avez accordés.

Et enfin, ces trois ans ont été une belle aventure humaine. Il y a eu des jours où les seuls mots « interactions » et « TJC » pouvaient me faire hérisser les poils, où j'ai eu envie de passer mon écran par la fenêtre, où je me demandais ce que je faisais bien là... mais l'idée de vous retrouver à Jouy au bâtiment 211 m'a toujours donné envie de me lever le matin (malgré le long trajet qui m'attendait chaque matin et chaque soir !). Merci à vous tous ! J'ai une pensée toute particulière pour la communauté des thésards du bâtiment 211 et toute l'équipe de Doc'J ; c'est une véritable chance de pouvoir compter sur vous. On évalue mal la force que peut donner un groupe, et pourtant qu'elle est grande ! Je n'oublie pas les équipes de la MNE au 7ème comme au 3ème qui m'ont accueillie les bras grands ouverts, tout particulièrement pendant la rédaction du manuscrit (pendant laquelle je n'étais pas toujours très calme...) et, bien évidemment, l'équipe de foot du jeudi qui m'a supportée comme joueuse et présidente pendant 3 ans !

Il reste deux personnes à qui je souhaite adresser des remerciements tout particulier. Hélène, tout d'abord, pour m'avoir tout appris sur ce qu'était une vache (!), pour m'avoir écoutée, soutenue et motivée pendant ces trois ans. Et ... Vincent bien-sûr, à qui je pourrais remettre un diplôme de « super directeur de thèse » ou encore le diplôme de « Jayden, chef des power rangers » ! Merci à toi, mon mentor scientifique ! Tu pourrais faire mentir toutes les caricatures (ou presque :.)) du méchant directeur de thèse de PhD comics. Tu prends le temps d'expliquer, d'écouter, tu corriges tout en un temps record, tu m'as laissée être autonome tout en étant présent aux moments-clés (ce que j'ai tout particulièrement apprécié !). Alors merci beaucoup à vous deux.

Cette aventure s'est terminée le jour de ma soutenance de thèse. Ce fut une très belle journée, tant du point de vue professionnel que personnel. Merci à vous tous, famille, amis, collègues d'avoir fait de cette journée un moment si particulier. Aujourd'hui, je pars pour de nouvelles aventures mais cette journée restera, pour moi, le symbole des 3 années précédentes : une journée scientifiquement dense et pleine de convivialité !

Table des matières

Remerciements **7**

Liste des figures **11**

Liste des tables **14**

1 Contexte de la thèse : adaptation de la génétique à la diversité des systèmes d'élevage via l'étude des interactions Génotype*Milieu **17**
1.1 Qu'appelle-t'on un système d'élevage ? . . . 17
1.2 La diversité des systèmes d'élevage : une réalité internationale et nationale 18
1.3 Une demande d'animaux mieux adaptés à la diversité des systèmes d'élevage 19
1.4 Comment répondre à cette demande ? Le point de vue de la génétique . . . 20
1.4.1 Choisir la race en fonction du contexte local . . . 20
1.4.2 Modifier les objectifs de sélection au sein d'une race . . . 21
1.4.3 Utiliser le croisement . . . 22
1.5 Prise en compte des interactions Génotype*Milieu : une perspective d'adaptation . . . 23
1.5.1 Définition des interactions Génotype*Milieu . . . 24
1.5.2 Objectifs de la thèse . . . 25
1.5.3 Une thèse au sein du projet GENESYS . . . 26
1.5.4 Structuration de la thèse . . . 26

2 Comment définir le milieu ? **29**
2.1 Étude bibliographique . . . 29
2.1.1 Les différentes facettes du milieu . . . 29
2.1.2 Les approches de description du milieu dans l'étude des interactions Génotype*Milieu . . . 33
2.1.3 Les limites des définitions existantes . . . 35
2.2 Vers une nouvelle définition du milieu basée sur les profils Troupeau Jour de Contrôle . . . 35
2.2.1 Prérequis sur les modèles d'évaluation génétique . . . 36
2.2.2 Du modèle d'évaluation génétique « lactation » au modèle « contrôles élémentaires » . . . 36
2.2.3 Les profils Troupeau Jour de Contrôle, reflets de la conduite de troupeau . . . 38
2.2.4 Les effets TJC et autres co-produits du modèle « contrôles élémentaires » déjà valorisés dans des outils de conseil . . . 40
2.2.5 Utilisation des profils Troupeau-Jour de Contrôle comme base de la définition du milieu : les avantages . . . 45

2.2.6 Profils Troupeau-Jour de Contrôle comme base de la définition du milieu : les défis . . . 46
2.3 Comment définir le milieu à partir des profils Troupeau Jour de Contrôle ? 47
2.3.1 Objectifs et Méthodologie générale . . . 47
2.3.2 Les descripteurs des profils TJC . . . 49
2.3.3 Définition du milieu à partir des profils Troupeau Jour de Contrôle : **Article I** . . . 53
2.3.4 Conclusions de l'article I . . . 68

3 Discussion : les effets Troupeau Jour de Contrôle, nouvelles perspectives de valorisation 71
3.1 Une connaissance plus précise des conduites de troupeau pour mieux interpréter les profils TJC . . . 71
3.1.1 Article I : des premiers éléments sur le lien entre profils TJC et conduite . . . 71
3.1.2 Un partenariat entre GENESYS et PhénoFinlait pour aller plus loin 72
3.2 Utiliser les méthodes d'analyse des séries temporelles : une piste pour la valorisation des profils TJC ? . . . 73
3.2.1 Principe des méthodes d'analyse des séries temporelles . . . 73
3.2.2 Utilisation de la composante saisonnière des profils TJC pour décrire un système de conduite . . . 75
3.2.3 Utilisation de la composante conjoncturelle dans l'analyse a posteriori des profils TJC . . . 76
3.2.4 Prédiction de l'effet TJC dans le cadre d'outils de prédiction de la production laitière . . . 76
3.3 Les profils TJC : outil de comparaisons de l'efficacité des conduites . . . 77

4 Étude des interactions Génotype * Milieu 83
4.1 Introduction . . . 83
4.2 Les différentes formes d'interactions Génotype* Milieu . . . 84
4.3 Modélisations des interactions Génotype*Milieu : méthodes et indicateurs . 86
4.3.1 Analyse de (Co)Variance avec un terme d'interaction . . . 86
4.3.2 Régressions aléatoires : adaptation des modèles de norme de réaction en génétique quantitative . . . 87
4.3.3 Modèle multicaractères : adaptation du modèle « character state » en génétique quantitative . . . 89
4.4 Les études déjà menées en France . . . 91
4.5 Les études menées dans la thèse . . . 92
4.6 Première étude sur les interactions Génotype*Milieu en France sur les caractères laitiers . . . 92
4.6.1 Méthodologie générale . . . 92
4.6.2 Mise en évidence d'un effet d'échelle pour les caractères laitiers en races Normande et Holstein : **Article II** . . . 94
4.6.3 Mise en évidence d'un effet d'échelle pour les caractères laitiers en race Montbéliarde . . . 122
4.6.4 Atouts et limites de la première étude . . . 126
4.7 Deuxième étude des interactions Génotype*Milieu en France sur les caractères laitiers et fonctionnels . . . 126
4.7.1 Objectifs . . . 126

4.7.2 Données . . . 127
4.7.3 Méthodes . . . 128
4.7.4 Résultats . . . 131
4.7.5 Discussion/Conclusion . . . 141

5 Discussion générale **143**
5.1 Les enjeux des interactions Génotype*Milieu à l'échelle nationale . . . 143
5.2 Mise en évidence d'un effet d'échelle en France . . . 144
5.3 Mise en évidence de reclassements au niveau international dans la bibliographie . . . 146

6 Conclusion générale **149**

Bibliographie **152**

Liste des publications **167**

Formations suivies **169**

Résumé **173**

Abstract **175**

Table des figures

1.1 Un système d'élevage sous différents angles, sources de nombreuses définitions du milieu dans le cadre d'études des interactions Génotype*Milieu . . 18

2.1 Une lactation décryptée par le modèle génétique « lactation » 37
2.2 Une lactation décryptée par le modèle génétique basé sur les « contrôles élémentaires » 37
2.3 Profils TJC Lait TB et TP d'un élevage de 2005 à 2010 39
2.4 Profil Troupeau Jour de Contrôle Lait d'un élevage 41
2.5 Évolution du prix du lait au cours du temps *(source : Institut de l'Élevage-Département Économie)* 41
2.6 Poster ICAR 2011 présentant les différents outils de conseil en élevage mis en place à partir du modèle génétique « contrôles élémentaires » 43
2.7 Différents profils TJC 47
2.8 Descripteurs profil TJC Lait d'un élevage de 2005 à 2010 par saison 51
2.9 Les 3 fonctions sinusoïdes utilisées pour lisser les profils TJC 52
2.10 Reconstitution de profils TJC à partir de 3 sinusoïdes en appliquant différents jeux de coefficients (la tendance et le niveau moyen ne sont pas pris en compte dans cette reconstitution) 52

3.1 Schéma explicatif de la typologie sur un grand nombre d'élevages uniquement à partir des profils TJC 81
3.2 Schéma explicatif de la typologie à partir des profils TJC sur un faible nombre d'élevages dont la conduite est connue 81
3.3 Schéma explicatif de la typologie ancrée à la réalité de terrain mais utilisant les profils TJC 82

4.1 Formes des relations entre 3 génotypes (représentés par 3 couleurs) et le milieu décrit par une variable continue. a) pas de plasticité phénotypique mais variation dans le niveau génétique b) plasticité mais pas de variation de plasticité selon les génotypes c) plasticité variable mais pas de classement différent des génotypes d) plasticité variable et classement différent des génotypes. 85
4.2 Héritabilité (h2), variance résiduelle et variance génétique en fonction de l'environnement pour le caractère Lait 136
4.3 Héritabilité (h2), variance résiduelle et variance génétique en fonction de l'environnement pour le caractère Matière Grasse (MG) 137
4.4 Héritabilité (h2), variance résiduelle et variance génétique en fonction de l'environnement pour le caractère Matière Protéique (MP) 138

4.5 Héritabilité (h2), variance résiduelle et variance génétique en fonction de l'environnement pour le caractère Score de Cellules Somatiques (SCS) . . . 139
4.6 Héritabilité (h2), variance résiduelle et variance génétique en fonction de l'environnement pour le caractère réussite à la première insémination artificielle (ResIA) 140

Liste des tableaux

4.1 Caractéristiques du jeu de données parangon pour la race Montbéliarde utilisées pour l'analyse multicaractères (écart-types entre parenthèses). C1 : un groupe d'élevages dont la conduite a une intensité moyenne plutôt spécialisée en lait. C2 : un groupe d'élevages dont la conduite est intensive mais plutôt dirigée vers des taux protéique et butyreux élevés. C3 : un groupe d'élevages dont la conduite est peu intensive. 123
4.2 Caractéristiques des jeux de données parangon et diversité utilisés pour l'analyse avec les normes de réaction (écart-types entre parenthèses) 123
4.3 Résultats du modèle multicaractères (écart-types entre parenthèses) 124
4.4 Variances résiduelles utilisées pour calculer les héritabilités pour les modèles de norme de réaction 124
4.5 Corrélations génétiques entre les groupes d'élevages avec le modèle de norme de réaction 125
4.6 Héritabilité pour les différents groupes d'élevages avec le modèle de norme de réaction 125
4.7 Présentation des données de performance 128
4.8 Valeurs propres (exprimées en pourcentage) des matrices de variance/covariance génétique pour les différents caractères et modèles 134
4.9 Les vecteurs propres ($u1$ et $u2$) de la matrice de variance/covariance génétique obtenus avec le modèle 2 qui suppose que la matrice génétique est de rang 2 134
4.10 Bayesian Information Criterion obtenu pour chaque modèle et chaque caractère 135
4.11 Les corrélations génétiques obtenues entre environnements (minimum, moyenne et maximum) par caractère 135

Chapitre 1

Contexte de la thèse : adaptation de la génétique à la diversité des systèmes d'élevage via l'étude des interactions Génotype*Milieu

1.1 Qu'appelle-t'on un système d'élevage ?

Un système d'élevage est un triptyque composé de l'Homme, de l'animal et de ressources (information, énergie, moyens financiers et matériels) (Landais, 1992). Le système est sous la dépendance d'un pilote : l'Homme, qui met en place un projet afin d'atteindre des objectifs économiques, sociaux et environnementaux. Ce concept est illustré par l'étage supérieur de la figure 1.1. Les décisions prises par l'éleveur seront mises en œuvre sur le terrain par un ensemble d'activités finalisées qui est appelé « pratiques d'élevage » ou plus simplement « conduite de troupeau ». Pour atteindre un même objectif, des éleveurs peuvent choisir des conduites de troupeau différentes qui prennent en compte les contraintes spécifiques à l'élevage telles que les conditions pédoclimatiques, la structure des bâtiments, les surfaces agricoles disponibles. Les pratiques d'élevages correspondent aux activités d'entretien des animaux (alimentation, gestion sanitaire), de constitution et de renouvellement du troupeau (politique de reproduction et de réforme), de valorisation des produits (en bovins laitiers : abattage, traite, fabrication de fromages, vente directe...) ainsi qu'à toutes les actions agronomiques effectuées sur les surfaces destinées à l'alimen-

tation des animaux. Elles sont représentées dans la figure 1.1 par l'étage intermédiaire.

Depuis l'article fondateur Landais (1992), la notion de système d'élevage a évolué au cours du temps (Dufumier, 1996; Dedieu *et al.*, 2008) mais les fondamentaux restent les mêmes : les objectifs de l'éleveur se traduisent par une conduite de troupeau choisie sous contraintes qui mène à des performances ; c'est cet ensemble qui définit un système d'élevage. La figure 1.1 p.18 schématise ces relations entre système, conduite, contraintes et performances.

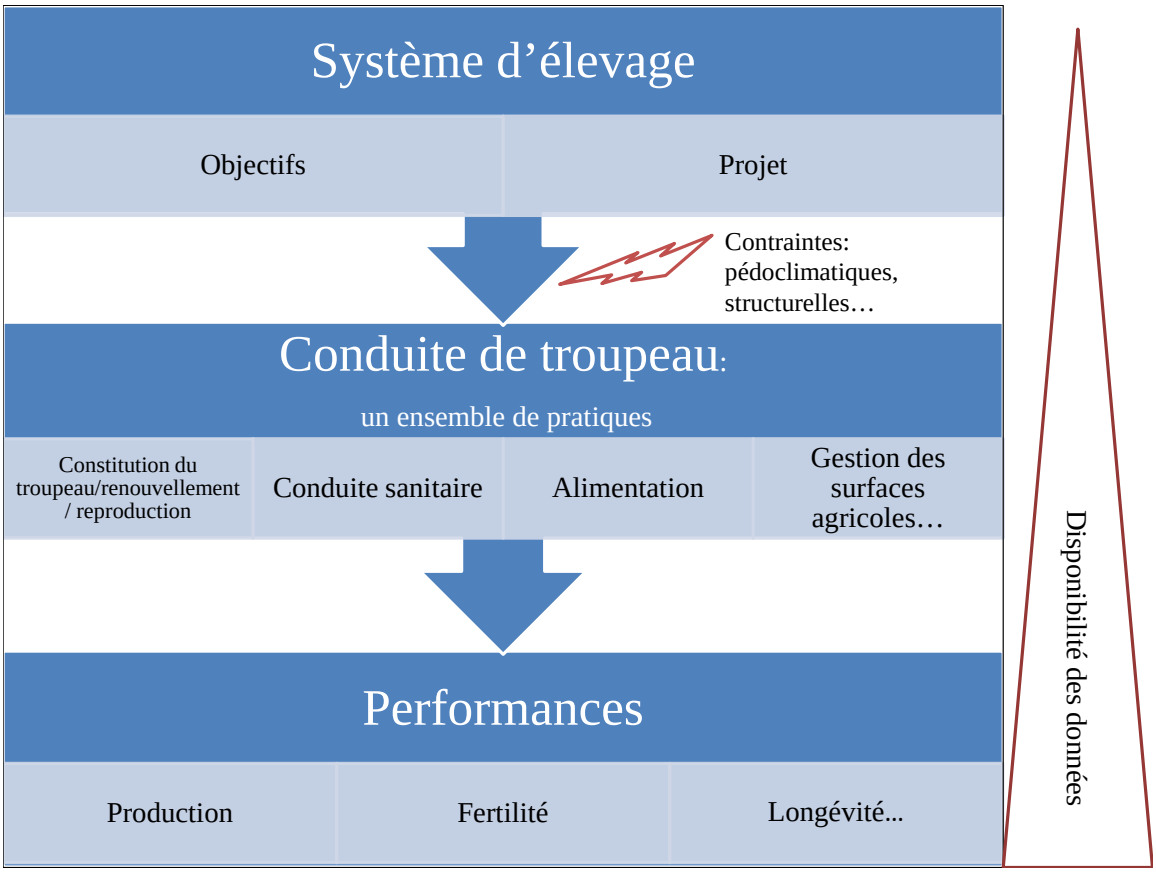


FIGURE 1.1 – Un système d'élevage sous différents angles, sources de nombreuses définitions du milieu dans le cadre d'études des interactions Génotype*Milieu

1.2 La diversité des systèmes d'élevage : une réalité internationale et nationale

A l'échelle mondiale, il existe une grande diversité des milieux dans lesquels évoluent les animaux d'élevage. Les conditions pédoclimatiques sont très diverses en matière de température, hygrométrie, pluviométrie. Cela a pour conséquence une alimentation variée

du bétail aussi bien pour le type que la qualité de l'alimentation. D'autre part, la grande palette des conditions socio-économiques à travers le monde se traduit à l'échelle de l'élevage laitier par des structures et des techniques d'élevage variées : nombre d'animaux par élevage, conditions de logement, équipement de traite... Afin d'étudier cette réalité complexe, nous avons recours à des typologies qui consistent à définir, par une étude, un certain nombre de types de systèmes d'élevage. C'est une simplification de la réalité mais cela facilite l'analyse. Par exemple, Pflimlin *et al.* (2005) proposent pour l'Europe une typologie en 8 classes en fonction des conditions pédoclimatiques et du système fourrager utilisé. Chatellier & Pflimlin (2007) prennent en compte d'autres critères comme la productivité du travail pour décrire la diversité des systèmes en Europe océanique.

En France, toutes ces différences existent également, bien qu'elles soient moins marquées. Quatre types de climat tempéré existent en France : les climats océanique, continental, méditerranéen et montagnard. Il existe plusieurs typologies à dire d'experts des exploitations françaises qui illustrent la diversité des systèmes d'exploitation existants au sein du territoire français. Chatellier *et al.* (1997) propose un typologie basée sur la localisation géographique de l'exploitation (plaine vs montagne/piémont), le système fourrager (degré d'utilisation de maïs ou d'herbe) et du type d'animaux élevés (spécialisés lait, vaches laitières + taurillons...). Plus récemment, l'observatoire de l'alimentation des vaches laitières créé par l'Institut de l'Elevage en partenariat avec le Centre National Interprofessionnel de l'Économie Laitière (CNIEL) a proposé une typologie des systèmes bovins laitiers français en 15 classes en fonction du relief, de la localisation géographique et du type d'alimentation (pourcentage de maïs dans la Surface Fourragère Principale) (Brunschwig *et al.*, 2011). Enfin, Barbin *et al.* (2009) mettent en évidence la diversité des exploitations en matière de résultats économiques et d'organisations du travail.

1.3 Une demande d'animaux mieux adaptés à la diversité des systèmes d'élevage

Pour que l'animal soit en adéquation avec son milieu, deux approches sont possibles : modifier le milieu pour qu'il convienne au mieux à l'animal ou choisir des animaux adaptés au milieu dans lequel ils seront élevés.

Jusqu'à présent, c'est principalement le système de production qui a été modifié pour

que les animaux expriment au mieux leur potentiel : la part de maïs dans la ration a augmenté, l'utilisation des concentrés est devenue massive, les équipements (logements des animaux, matériel de traite) se sont perfectionnés. Cependant, aujourd'hui, compte tenu des avancées déjà réalisées dans ce domaine et des contraintes environnementales (telles que celles imposées par la directive Nitrate ou le plan maîtrise des pollutions d'origine agricole dans les exploitations d'élevage bovin), ce levier semble moins utilisable : « *Une [des] nouveaux défis est [...] de transférer la maîtrise des effets de l'environnement par le système de production à leur maîtrise par l'animal lui-même* » (Bodin *et al.*, 2010).

C'est pourquoi des questions émergent sur la manière de choisir des animaux adaptés à tous les environnement ou adaptés au mieux à l'environnement dans lequel ils évoluent ; sachant que les systèmes d'élevage peuvent être très variés. Peu d'écrits existent sur le sujet, cette préoccupation des éleveurs semble se transmettre plus particulièrement à l'oral. Elles sont tout de même soulevées dans des revues techniques comme l'article « Les Grisons veulent adapter leurs vaches bio aux conditions locales » (Schmid & Pedotti, 2010) ou dans les témoignages d'éleveurs, comme lors de la conférence organisée par l'Institut Technique de l'Agriculture Biologique (ITAB) intitulée « Sélection Animale, Diversité Génétique et Agriculture Biologique » (De Saint Vaury, 2010).

1.4 Comment répondre à cette demande ? Le point de vue de la génétique

Les éleveurs ont plusieurs manières de choisir les animaux les mieux adaptés à leur système d'élevage à travers des choix génétiques.

1.4.1 Choisir la race en fonction du contexte local

Il existe une dizaine de races bovines laitières en France. Chacune présente des potentialités différentes ; le choix de la race à utiliser dans un élevage est donc une première façon d'adapter la génétique au système d'élevage.

Les éleveurs situés dans des zones de montagne au climat difficile avec une topographie accidentée privilégient des races comme l'Abondance, la Tarentaise, la Vosgienne. Il s'agit de races rustiques capables de supporter de grands changements de température, de longues marches (grâce à la qualité de leurs aplombs) et de valoriser des fourrages grossiers.

Ces races compensent un déficit de précocité par une bonne longévité (cf les sites internet des UPRA Abondance; UPRA Vosgienne; UPRA Tarentaise et d'AgroParisTech). La race Montbéliarde possède dans une moindre mesure ces aptitudes (sites de l'OS Montbéliarde et d' AgroParisTech).

Les races Normande et Simmental sont deux races mixtes (comme la Montbéliarde) dont la qualité fromagère du lait est reconnue (sites de l'OS Normande et de l'OS Simmental France). Enfin, les races Brune et Holstein sont deux races spécialisées laitières qui par leur grande taille sont moins bien adaptées aux milieux difficiles, elles sont néanmoins utilisées dans de nombreux systèmes d'élevages (sites de l'OS Brune et de Prim'Holstein France) et bien adaptées aux systèmes intensifs.

AGRIDEA, une association suisse pour le développement de l'agriculture et de l'espace rural propose une grille d'évaluation de l'exploitation en matière de système de pâture, de main d'œuvre disponible, de type bâtiments... pour choisir au mieux la race à utiliser dans cette exploitation (site de l'Agridea).

Plusieurs projets de recherche passés et présents cherchent à déterminer la race ou la lignée d'une race la mieux adaptée à un type de système d'élevage. Les projets présentés ci-après en sont quelques exemples. Le projet « Quelle vache pour la pâture ? » en Suisse (Piccand *et al.*, 2011) compare les performances globales des trois principales races laitières suisses (Tachetée rouge, Brown Swiss et Holstein) aux performances des Holstein-Friesian néo-zélandaises, sur des exploitations pratiquant le pâturage intégral avec vêlages saisonniers de fin d'hiver. Le projet « Quelles vaches pour quels systèmes ? », mené à l'Institut National de Recherche Agronomique du Pin au Haras, compare les performances de Normande et de Holstein en système intensif ou herbager (Conté, 2011). Enfin, un projet intitulé « Quels sont les types génétiques de vache laitière les plus adaptés aux différents types de systèmes d'élevage pour assurer la pérennité de ces systèmes ? » a été déposé par l'Unité de recherche sur les systèmes d'élevage de l'École Supérieure d'Agriculture d'Angers (ESA) (cf site de l'Ecole Supérieure d' Agriculture d'Angers).

1.4.2 Modifier les objectifs de sélection au sein d'une race

L'objectif de sélection d'une race est l'ensemble des caractères à améliorer, avec leurs pondérations relatives. Les objectifs incluent le plus souvent des caractères de production (quantité et qualité du lait), de robustesse et de fonctionnalité dont les pondérations

sont choisies par les Organismes de Sélection (OS) de chaque race. Les OS des races de montagne dont les animaux évoluent dans un milieu difficile accordent une grande importance aux caractères de rusticité comme la longévité. Les animaux des autres races évoluent dans des systèmes d'élevage très variés, pourtant les OS ne définissent qu'un seul objectif de sélection par race sans prendre en compte le système d'élevage dans lequel vont évoluer les animaux. Quelques travaux de recherche remettent en cause cet état de fait notamment dans le cadre de l'agriculture biologique. Ce système à bas intrants entraînerait des problèmes de santé comme des cétoses, c'est pourquoi Hardarson (2002) propose d'adapter les objectifs de sélection dans le cadre de l'agriculture biologique en mettant plus de poids sur la longévité, la persistance, la résistance aux maladies tout en acceptant de moins sélectionner sur la production laitière. Nauta *et al.* (2005) a réalisé une enquête et monté des forums de discussion auprès d'une cinquantaine d'éleveurs en agriculture biologique pour partager leur vision de la sélection en agriculture biologique. Certains d'entre eux montrent un intérêt pour la construction d'un objectif de sélection adapté à leur système, les critères de sélection restant à préciser. Quelques exemples d'objectifs de sélection adaptés à l'agriculture biologique existent. L'institut de recherche de l'agriculture biologique suisse (FIBL Bio Suisse) propose une valeur génétique (ou valeur d'élevage) globale écologique pour différentes races (tachetée rouge, brune, blonde d'allemagne) (Bapst & Bär, 2001) basés sur les travaux de recherche de Postler (1998). Des objectifs de sélection adaptés au système d'élevage irlandais, principalement basé sur le pâturage, ont été proposés pour faire face à 3 scénarios mimant 3 différents systèmes de quota. Ils accordent un poids plus ou moins important aux caractères de fertilité (Veerkamp *et al.*, 2002).

1.4.3 Utiliser le croisement

La plupart des élevages laitiers dans le monde sont conduits en race pure. Pourtant le croisement se développe un peu en France (Bougoin & Le Mezec, 2010) et dans le monde comme en Nouvelle-Zélande (LIC & DairyNZ, 2010). Le croisement a pour objectif de réunir les atouts de plusieurs races et de profiter d'un effet d'hétérosis. Il pourrait être utilisé comme moyen d'adaptation de la génétique aux systèmes d'élevage. Plusieurs études ont été menées en Nouvelle Zélande où le système d'élevage est principalement

basé sur le pâturage pour comparer les performances des animaux et les performances économiques d'élevages utilisant des animaux de race pure ou des animaux croisés selon différents schémas (Lopez-Villalobos *et al.*, 2000). Cependant, la majorité des études sur le croisement se focalisent sur la mesure de l'effet d'heterosis et l'intérêt économique qui peut en découler sans chercher à mieux adapter l'animal à un milieu particulier. En Allemagne, une nouvelle race issue d'un croisement 3 voies a été créée dans les années 60 : la Schwartzbuntes Milchrind der Deutsche Demokratische Republik (SMR) (Freyer *et al.*, 2008). A l'époque, l'objectif n'était pas de créer des animaux bien adaptés à leur environnement mais cette nouvelle race a été la base de nombreuses recherches sur le croisement (comparaison des performances). Des études menées au Danemark (Sørensen *et al.*, 2008), au Kenya (Thorpe *et al.*, 1994) et aux Etats-Unis (VanRaden & Sanders, 2003) ont montré l'intérêt d'utiliser le croisement pour améliorer les performances des animaux sur des caractères de production ou fonctionnels. En revanche, en Allemagne (Blöttner *et al.*, 2011b,a), peu de différences ont été montrées entre les animaux de race pure et les animaux croisés (Hostein*Brown Swiss). Dans tous les cas, le croisement est peu souvent étudié en tant que moyen d'adaptation, d'ailleurs Bodin *et al.* (2010) considèrent que ce moyen est peu envisageable pour le choix d'animaux adaptés à toute une gamme d'environnements.

L'adaptation de la génétique à la diversité des systèmes est un sujet d'étude très large tant les points de vues possibles pour répondre à cette question sont divers. La plupart des exemples cités ci-avant concernent les systèmes en agriculture biologique car ces éleveurs forment un groupe facilement reconnaissable et disposent d'organisations pour communiquer. Cependant, les questions soulevées par ces éleveurs peuvent être généralisées à l'adaptation de la génétique à différents types de système d'élevage (pas seulement celle adoptée en agriculture biologique).

1.5 Prise en compte des interactions Génotype*Milieu : une perspective d'adaptation

Une alternative non évoquée précédemment serait de choisir des animaux en fonction de leur génotype pour qu'ils soient adaptés spécifiquement à un milieu ou qu'ils soient

adaptés à tout type de milieu. Cela consiste à tirer partie de l'éventuelle existence d'interactions Génotype*Milieu. Cette thèse s'inscrit dans cette voie. Cette piste est souvent négligée car elle est moins intuitive et se situe plus en amont de la réflexion sur l'adaptation de la génétique à la diversité des systèmes que le choix de la race, la rénovation des objectifs de sélection ou l'utilisation du croisement.

1.5.1 Définition des interactions Génotype*Milieu

Les animaux sont capables de s'adapter à des modifications de leur environnement par la modification de l'expression de leurs gènes, c'est ce qu'on appelle la plasticité phénotypique ou la sensibilité à l'environnement (Scheiner & Lyman, 1989). Pourtant, à ce jour, les modèles d'évaluation génétique ne prennent pas en compte cette plasticité. Ils considèrent que la part de la performance d'un animal qui est d'origine génétique - c'est-à-dire la valeur génétique d'un animal- est la même quel que soit le système d'élevage dans lequel il évolue. Cela suppose que l'expression des gènes responsables des performances des animaux s'expriment de la même façon quel que soit leur environnement. Si cette hypothèse est infirmée, il devient possible de prendre en compte cette capacité d'adaptation. Le potentiel génétique d'un animal s'exprimerait alors différemment selon le milieu dans lequel il évolue ; en pratique un animal pourrait avoir plusieurs valeurs génétiques.

Les recherches entreprises pour étudier la valeur génétique d'un animal en fonction de l'environnement sont regroupées sous le terme d' « étude des interactions Génotype*Milieu ». Le terme « interactions Génotype*Système d'élevage » semblerait plus approprié puisque l'objectif final est toujours de savoir si la valeur génétique des animaux dépend du système d'élevage. Cependant, comme indiqué dans le paragraphe 1.1, le système d'élevage est une notion difficile à cerner. Il est donc souvent étudié via la description des conduites de troupeau voire même par la mesure des performances des animaux dans l'élevage et/ou les conditions pédoclimatiques de l'élevage (cf figure 1.1 p.18). Toutes ces notions sont regroupées sous le terme de « milieu » dans le cadre de l'étude des interactions Génotype*Milieu en sélection animale et sont présentées dans le chapitre 2. Les différentes formes d'interactions Génotype*Milieu (effet d'échelle, reclassement) et les méthodes pour les mesurer (modèle multicaractère, normes de réaction...) seront présentées dans le chapitre 4.

1.5.2 Objectifs de la thèse

L'étude des interactions Génotype*Milieu comme moyen d'adaptation des animaux à la diversité des systèmes est une voie assez technique et peu intuitive ; elle est donc plus rarement évoquée par les éleveurs que la modification des objectifs de sélection. Néanmoins, un groupe d'éleveurs a montré un intérêt pour l'étude des interactions Génotype*Milieu entre les élevages conventionnels et ceux en agriculture biologique (Nauta *et al.*, 2005), les questions soulevées par un éleveur sur la pertinence des méthodes d'évaluation génétique sous-entendent la nécessité d'étudier les interactions Génotype*Milieu (Le Dû, 2010).

Cette thèse participe à la réflexion sur la manière d'améliorer les caractères laitiers et fonctionnels des bovins laitiers français. Elle se focalise sur la voie qui consisterait à choisir un génotype (un animal) adapté au milieu dans lequel il va évoluer c'est-à-dire sur l'étude des interactions Génotype*Milieu. Ces dernières ont fait l'objet de nombreuses études, principalement à l'étranger ; cette thèse a pour ambition d'apporter un regard nouveau sur la définition du milieu et d'étudier ce phénomène à l'échelle française en bovins laitiers.

L'objectif de cette thèse est d'une part de proposer une nouvelle définition du milieu à partir d'un co-produit du modèle d'évaluation génétique sur contrôles élémentaires : les profils Troupeau Jour de Contrôle ; d'autre part de mesurer à l'échelle française les interactions Génotype*Milieu sur les caractères laitiers et fonctionnels pour les 3 races bovines laitières majoritaires en France (Holstein, Montbéliarde, Normande) en utilisant les profils Troupeau Jour de Contrôle comme descripteurs du milieu.

En pratique, cela revient à se demander si le choix d'un taureau reproducteur doit dépendre du milieu dans lequel évolueront ses descendants c'est-à-dire si le classement des meilleurs animaux du point de vue génétique est le même quel que soit le milieu dans lequel seront élevés leurs descendants. Si ce n'est pas le cas, faut-il songer à calculer et diffuser des valeurs génétique d'animaux par milieu en prenant en compte ces interactions dans les modèles d'évaluation génétique ? Quand bien même le classement des animaux serait le même quel que soit le milieu, la variabilité des valeurs génétiques est-elle la même d'un milieu à l'autre ? Si ce n'est pas le cas la précision des valeurs génétiques et le progrès génétique seraient différents selon le milieu.

Pour répondre à cette problématique, nous disposons des données enregistrées dans le cadre du contrôle de performance, des profils Troupeau Jour de Contrôle pour tous les élevages adhérant à une Entreprise de Conseil ÉLevage (ECEL) et des données de conduite

de troupeau pour un nombre limité d'élevages suivis dans le cadre du projet GENESYS (via des enquêtes de terrain menées par les Entreprises de Conseil en Élevage ou les bases de données des Réseaux d'Élevage).

1.5.3 Une thèse au sein du projet GENESYS

Cette thèse est un des maillons d'un projet plus vaste : le projet GENESYS (GENEtique, conduite et SYStème d'élevage chez les bovins laitiers), un projet pluridisciplinaire qui cherche à combiner la génétique et l'étude des systèmes d'élevage. L'Institut de l'Elevage, l'INRA (l'Unité Mixte de Recherche (UMR) Génétique Animale et Biologie Intégrative (GABI) et l'UMR « Production du Lait ») et France Conseil ELevage (FCEL) participent à ce projet. Une fiche synthétique sur le projet est disponible sur le site internet de l'Institut de l'Elevage (http://idele.fr/filieres/bovins-lait/publication/idelesolr/recommends/casdar-genesys.html). Un premier volet du projet consiste à valoriser des données issues du modèle génétique basé sur les « contrôles élémentaires » pour améliorer les outils de conseil en élevage. Le second étudie l'adaptation de la génétique à la diversité des systèmes d'élevage par 2 voies : étude de l'importance et de l'intérêt du croisement en France et l'étude des interactions Génotype*Milieu en ferme expérimental et au niveau national. Cette dernière approche fait l'objet de cette thèse.

1.5.4 Structuration de la thèse

Les différentes définitions du milieu proposées dans la littérature et à la définition originale retenue dans cette thèse à partir des profils Troupeau Jour de Contrôle sont présentées dans le chapitre 2. Il est suivi d'une discussion sur les perspectives d'utilisation des profils Troupeau Jour de Contrôle dans le cadre du conseil en élevage (chapitre 3) donnant suite aux résultats du projet GENESYS.

Ensuite, le chapitre 4 présente, d'une part les différentes méthodes existantes pour mesurer les interactions Génotype*Milieu, d'autre part les résultats obtenus dans cette thèse en modèle multicaractères et avec des normes de réaction pour les caractères laitiers et fonctionnels en utilisant les profils Troupeau Jour de Contrôle pour décrire le milieu. Une discussion sur les enjeux des interactions Génotype*Milieu en France et à l'international s'en suit (chapitre 5).

Enfin, le chapitre 6 clôt la discussion sur une réflexion sur les apports de la thèse et les perspectives pour de futures études sur les interactions Génotype*Milieu.

Chapitre 2

Comment définir le milieu ?

La question sous-jacente aux études sur les interactions Génotype*Milieu est la suivante : « la valeur génétique (additive) d'un animal dépend-elle du milieu dans lequel il l'exprime ? ». Pour répondre à cette question, il faut d'abord définir le milieu ou système d'élevage.

2.1 Étude bibliographique

2.1.1 Les différentes facettes du milieu

Le milieu défini comme une entité géographique

Dans le cadre du commerce international des semences de taureaux, des études ont été menées pour savoir si le classement des taureaux des pays exportateurs de semences est le même dans le pays importateur sachant que les conditions climatiques et d'élevage peuvent y être très différentes. Dans ce cas, le milieu est défini comme une région, un pays ou un groupe de pays ; aucun critère agronomique n'est pris en compte mais on suppose que derrière les pays se cachent des systèmes d'élevage différents. Des études ont été menées impliquant les pays suivants : les États Unis et l'Espagne (Carabaño *et al.*, 1989), la Colombie, le Mexique, le Puerto Rico et les États Unis (Stanton *et al.*, 1991), le Canada et la Nouvelle Zélande (Charagu & Peterson, 1998), le Mexique et les États Unis (Cienfuegos-Rivas *et al.*, 1999), le Brésil et les États Unis (Costa *et al.*, 2000), l'Angleterre et le Kenya (Ojango & Pollott, 2002), la Colombie et le Brésil (Cerón-Muñoz *et al.*, 2004b), le Luxembourg et la Tunisie (Hammami *et al.*, 2008) ainsi que trois régions des États-Unis

(Carabaño *et al.*, 1990; Van Vleck *et al.*, 1988), deux régions d'Allemagne (König *et al.*, 2005) ou plusieurs régions d'Australie (Haile-Mariam *et al.*, 2008).

Toutes ces études sont inspirées de la méthode d'évaluation génétique internationale appelée Multiple Across Country Evaluation (MACE,(Schaeffer, 1994)) mise en place par Interbull (l'organisation responsable des évaluations génétiques internationales). Le MACE permet de comparer les valeurs génétiques des animaux estimées dans différents pays. Chaque animal possède une valeur génétique particulière par environnement (pays). Le MACE peut être considéré comme une étude des interactions Génotype*Milieu où le milieu est défini comme un pays. Une corrélation génétique entre deux pays inférieure à 1 témoigne d'une interaction Génotype*Pays. Cependant, cette interaction n'est pas seulement une interaction Génotype*Système d'élevage, elle peut aussi être due à différents types de mesures d'un même caractère dans différents pays, différents systèmes de collecte, différentes méthodes d'évaluation génétique nationale.

Définir le milieu comme un pays (ou une autre entité géographique) est très pratique mais assez simpliste puisqu'elle gomme les différences de milieux au sein d'un même pays. En effet, deux exploitations d'un même pays peuvent avoir des systèmes d'élevage qui diffèrent plus que deux exploitations semblables de pays différents. La définition du milieu par une entité géographique a notamment été remise en cause par Weigel & Rekaya (2000). De plus, Weigel *et al.* (2001) ont montré que les corrélations génétiques entre pays ayant le même système d'élevage (exemple : système pâturage vs système intensif) sont plus élevées que la moyenne. La définition du milieu devrait alors être précisée pour mieux prendre en compte le système d'élevage.

Compte tenu des limites évoquées sur la définition du milieu comme une entité géographique, cette approche n'a pas été envisagée dans cette thèse.

Le milieu défini comme un système d'élevage

Pour avoir la meilleure description possible d'un système d'élevage, il faudrait pouvoir connaître, comme présenté dans le chapitre 1 (figure 1.1 p.18), l'ensemble des facettes qui le composent : contraintes pédoclimatiques et structurelles de l'exploitation, conduite de troupeau (alimentation, reproduction, constitution du troupeau...), performances des animaux . Cependant la faible disponibilité de certaines données (comme la connaissance précise de l'alimentation) rend impossible une description aussi complète du milieu. Selon

le caractère étudié pour la mesure des interactions Génotype*Milieu et l'échelle de l'étude (étude internationale ou en ferme expérimentale), les définitions du milieu utilisées dans la littérature sont très diverses. Elles font l'objet de revues bibliographiques dans les thèses de Calus (2006) et Hammami (2009).

Ci-après sont présentées successivement les définitions du milieu construites à partir du système, de la conduite d'élevage, des contraintes pesant sur l'élevage et des performances des animaux.

• Le milieu décrit par un système de production

Nauta *et al.* (2006) définissent le milieu en opposant les élevages en agriculture biologique à ceux en agriculture conventionnelle. Cette étude répondait spécifiquement à une demande des agriculteurs en agriculture biologique qui craignaient que les schémas de sélection ne soient pas adaptés à leur système propre.

• Le milieu décrit par des éléments de la conduite de troupeau

Les études définissant le milieu à partir de la conduite de troupeau sont nombreuses. Elles se basent sur :

– l'alimentation

  La définition de cette dernière est plus ou moins précise selon les études. Plusieurs expérimentations (Boettcher *et al.*, 2003; Kearney *et al.*, 2004a,b) ont été menées aux États-Unis et au Canada en opposant les élevages utilisant majoritairement le pâturage et ceux en conduite classique dans ces pays (basée sur l'ensilage), en opposant le 100% pâturage à la ration totale mélangée (Kolver *et al.*, 2002), ou encore en définissant différents niveaux d'utilisation du pâturage (Horan *et al.*, 2005; McCarthy *et al.*, 2007a). D'autres auteurs se basent spécifiquement sur différentes quantités de concentrés dans la ration (Cromie *et al.*, 1998; Pryce *et al.*, 1999; Buckley *et al.*, 2000; Berry *et al.*, 2003; Horan *et al.*, 2005; McCarthy *et al.*, 2007a; Fulkerson *et al.*, 2008), la charge à l'hectare (Buckley *et al.*, 2000; Berry *et al.*, 2003) ou la qualité de l'ensilage d'herbe (Berry *et al.*, 2003) pour définir le milieu.

– le type de logement des animaux

  Il est exploité comme descripteur du milieu dans les études sur le taux cellulaire (SCC) (Wicks & Leaver, 2006) ou les caractères de morphologie des membres (Fatehi *et al.*, 2003).

– la structure du troupeau

Le nombre d'animaux dans l'exploitation est souvent utilisé (Calus *et al.*, 2002; Kolmodin *et al.*, 2004; Haile-Mariam *et al.*, 2008; König *et al.*, 2005).

- le choix des animaux via la composition génétique du troupeau (Weigel & Rekaya, 2000; Zwald *et al.*, 2003; Cerón-Muñoz *et al.*, 2004a; Maltecca *et al.*, 2004)
- le matériel utilisé

  Par exemple, au Pays-Bas, suite à l'introduction des robots de traite en ferme, le milieu a été défini en opposant les fermes utilisant des robots de traite aux autres (Mulder *et al.*, 2004).
- la gestion des surfaces agricoles comme le niveau de fertilisation azotée (Buckley *et al.*, 2000)
- la gestion des maladies au sein du troupeau (Raffrenato *et al.*, 2003)

Ces données sur la conduite de troupeau doivent être collectées dans le cadre d'enquêtes ponctuelles ou pérennes comme le suivi d'exploitations types dans le cadre des Réseaux d'élevages, qui est un dispositif de recherche déployé par les Chambres d'agriculture et coordonné par l'Institut de l'Élevage depuis le début des années 80 (Charroin *et al.*, 2005). La disponibilité de ces données est donc soumise à des contraintes financières et matérielles (mise à disposition de personnel pour réaliser les enquêtes ponctuelles en ferme, entretien et exploitation de la base de données des Réseaux d'Élevage). Elles sont valorisées dans le cadre d'études à moyenne échelle quand la conduite alimentaire, par exemple, est décrite très finement. Par conséquent, ces études sont réalisées, d'une part sur un faible nombre d'animaux, et d'autre part sur un éventail de systèmes d'élevage limité. C'est encore plus marqué pour les études réalisées en fermes expérimentales où le milieu est très contrôlé mais où le nombre de milieux qui peuvent être testés est faible.

• Le milieu décrit par les contraintes pesant sur l'exploitation

Les contraintes qui pèsent sur l'exploitation en terme de position géographique et de conditions météorologiques sont utilisées pour décrire le milieu même si ces informations ne reflètent qu'une partie des éléments descriptifs du milieu (Kolmodin *et al.*, 2004; Haile-Mariam *et al.*, 2008).

• Le milieu décrit par les performances des animaux

Certaines études (dont les plus anciennes) décrivent le milieu non pas par les caractéristiques de celui-ci mais par les performances des animaux comme la production de lait

moyenne de l'élevage (De Veer & Van Vleck, 1987; McDaniel & Corley, 1967) ou les quantités de matières grasse et protéique, les taux butyreux et protéique ou la quantité de cellules dans le lait (Castillo-Juarez *et al.*, 2000, 2002), l'effet moyen de l'élevage sur la production ou la fertilité (Bonaiti, 1982; Strandberg *et al.*, 2000; Calus *et al.*, 2002; Berry *et al.*, 2003; Kolmodin *et al.*, 2004; Haile-Mariam *et al.*, 2008; Hammami *et al.*, 2009), sa variabilité (Raffrenato *et al.*, 2003), la forme des courbes de lactation des animaux (Raffrenato *et al.*, 2003), la durée de lactation, la persistance laitière (Fikse *et al.*, 2003; Calus *et al.*, 2005) ou encore le score cellulaire, l'état corporel et le rapport matière grasse sur matière protéique (Calus *et al.*, 2005).

Les différentes définitions du milieu utilisées pour étudier les interactions Génotype*Milieu sont donc variées et offrent des points de vues différents sur la notion de système d'élevage. Pourtant, quelle que soit la définition du milieu utilisée pour étudier les interactions Génotype*Milieu, un compromis doit être atteint entre la précision des données qui permettent de décrire le système d'élevage et leur disponibilité. En effet, l'étude des interactions Génotype*Milieu requiert de travailler sur un grand nombre d'élevages mais il est difficile d'obtenir un descriptif fin du milieu sur un grand nombre d'élevages.

2.1.2 Les approches de description du milieu dans l'étude des interactions Génotype*Milieu

Une ou plusieurs variables décrivant une ou plusieurs facettes de l'environnement (alimentation, performances...) peuvent être utilisées pour décrire un système d'élevage. Le plus souvent, ces variables doivent être transformées pour l'étude des interactions Génotype*Milieu. En effet, pour être utilisé, le milieu doit être synthétisé sous la forme d'une variable quantitative continue ou d'une variable qualitative (ordonnée comme une quantité faible, moyenne ou élevée de concentrés dans la ration, ou nominale comme des groupes d'élevages).

Premier cas : une seule variable disponible pour décrire le milieu

Dans le cas où une seule variable qualitative (par exemple : utilisation versus non utilisation du pâturage) est utilisée pour décrire l'environnement, aucune transformation n'est nécessaire pour l'utiliser dans le cadre des interactions Génotype*Milieu. Si une

seule variable quantitative (quantité de concentrés dans la ration) est utilisée, elle peut être utilisée sans transformation ou après mise en classe.

Néanmoins, si une seule variable est disponible pour décrire le milieu, la vision globale des caractéristiques du système d'élevage est absente. C'est pourquoi certains auteurs ont cherché à combiner plusieurs descripteurs du milieu pour créer soit une variable soit des classes (d'élevages ou d'animaux) qui résume(nt) les informations des différentes variables décrivant le milieu.

Deuxième cas : plusieurs variables disponibles pour décrire le milieu

Buckley *et al.* (2000) se basent sur trois descripteurs de la conduite de troupeau au pâturage pour définir le milieu : la quantité de concentrés distribuée dans l'alimentation, le chargement à l'hectare et le niveau de fertilisation azotée. Horan *et al.* (2005) et McCarthy *et al.* (2007a) s'intéressent à trois types de conduite de troupeau décrites par différents niveaux d'utilisation du pâturage et des concentrés et des chargements plus ou moins importants. Raffrenato *et al.* (2003) utilisent une variable synthétisant les pratiques de l'élevage en terme d'alimentation, de traite et de gestion des maladies. D'autres études combinent plusieurs paramètres comme des descripteurs du climat, de la structure du troupeau, des performances du troupeau et de la composition génétique du troupeau (Weigel & Rekaya, 2000; Zwald *et al.*, 2003; Cerón-Muñoz *et al.*, 2004a; Maltecca *et al.*, 2004).

Quand le nombre de descripteurs devient important, certains auteurs utilisent des variables synthétiques comme les premières composantes principales décrivant l'ensemble des descripteurs (Windig *et al.*, 2006). Haskell *et al.* (2007) proposent une méthode originale qui consiste à créer une variable synthétique décrivant le milieu à partir de variables disponibles à l'échelle nationale (descripteurs du troupeau, des performances et du climat). Cette dernière est construite comme la variable la plus corrélée possible aux principaux facteurs de variation de descripteurs fins de la conduite de troupeau (comme le type d'alimentation, la quantité de main d'œuvre ou la gestion des génisses de renouvellement).

2.1.3 Les limites des définitions existantes

L'ensemble des définitions proposées ci-dessus présentent des limites. La définition du milieu comme une entité géographique gomme les différences de systèmes d'élevage qui peuvent exister au sein de celle-ci et sépare des exploitations appartenant à des entités géographiques différentes mais qui auraient des systèmes semblables. N'utiliser qu'une seule variable pour décrire le milieu ne peut pas refléter toute sa complexité. Le fait d'utiliser les performances brutes du troupeau comme descripteurs du milieu revient à considérer que la variabilité de ces performances est uniquement due à des différences de milieu sans influence du niveau génétique des animaux, ce qui va à l'encontre de l'idée des études sur les interactions Génotype*Milieu.

De plus, les données utilisées pour décrire le milieu dans ces publications sont des données annuelles, à part pour certaines qui utilisent la saison principale de vêlage. Elles donnent une vision globale de la conduite de troupeau adoptée sur l'année mais ne décrivent pas les conduites saisonnières (comme les périodes de pâturage).

C'est pourquoi dans le cadre de la définition du milieu pour les exploitations bovines laitières françaises, nous avons cherché un descripteur de l'environnement qui reflète un maximum de caractéristiques du milieu à un instant t et qui soit disponible en routine : les effets Troupeaux Jour de Contrôle.

2.2 Vers une nouvelle définition du milieu basée sur les profils Troupeau Jour de Contrôle

Le développement en France d'une évaluation bovine laitière basée sur les performances élémentaires offre l'opportunité de définir autrement le milieu grâce aux profils Troupeau Jour de Contrôle (TJC). Nous présenterons d'abord les prérequis sur les modèles d'évaluation génétique puis les spécificités du modèle « contrôles élémentaires » et enfin nous introduirons la notion de profil TJC valorisée dans des outils de conseil et comme descripteur potentiel du milieu.

2.2.1 Prérequis sur les modèles d'évaluation génétique

Les modèles d'évaluation génétique visent à expliquer les performances d'un animal par l'ensemble des facteurs connus pour avoir un impact sur la performance, dès lors que ceux-ci ont pu être enregistrés lors du contrôle de performances. Comme leur nom l'indique, l'objectif premier de ces modèles est d'estimer la part de la performance d'un animal qui est d'origine génétique c'est-à-dire transmise à sa descendance. Elle est appelée valeur génétique (additive) de l'animal. Les éleveurs et leurs conseillers se basent sur cette valeur pour choisir parmi l'ensemble des animaux existants les reproducteurs à utiliser pour procréer la nouvelle génération de vaches laitières. Pour estimer au mieux les valeurs génétiques, les autres facteurs pouvant avoir un impact sur la performances doivent être pris en compte et estimés. Il s'agit notamment de la race,de l'élevage, de l'année, de la zone géographique, du rang de lactation, de l'âge et du mois de vêlage, du stade de gestation, de la durée de tarissement du vêlage précédent de la vache.

2.2.2 Du modèle d'évaluation génétique « lactation » au modèle « contrôles élémentaires »

Le modèle génétique dit « modèle lactation » consiste à expliquer la production totale de lait d'un animal sur une lactation par un ensemble de facteurs génétiques et non génétiques (durée de tarissement, rang de lactation..). La lactation est souvent définie sur une durée de 305 jours et estimée suivant la méthode dite Fleischmann qui permet d'interpoler la production entre deux dates de contrôle. Le modèle « contrôles élémentaires », quant à lui, explique chaque performance journalière enregistrée lors du contrôle de performances (le plus couramment une fois par mois). La spécificité du modèle basé sur les « contrôles élémentaires » est donc la prise en compte des facteurs jouant sur la performance journalière en fonction du stade de lactation contrairement au modèle « lactation » qui estime seulement leur effet global sur la lactation (Leclerc *et al.*, 2008). Les figures 2.2 et 2.1 présentent des vues simplifiées de la façon dont chacun des modèles décrypte une même lactation.

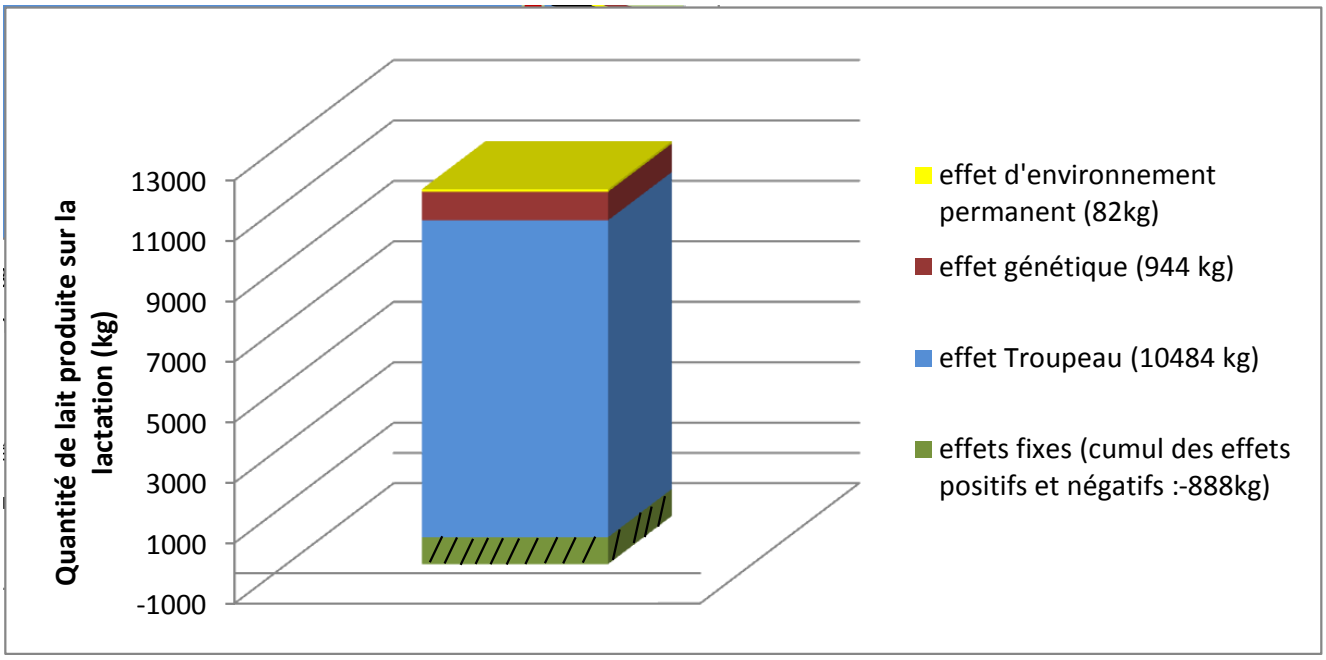


FIGURE 2.1 – Une lactation décryptée par le modèle génétique « lactation »

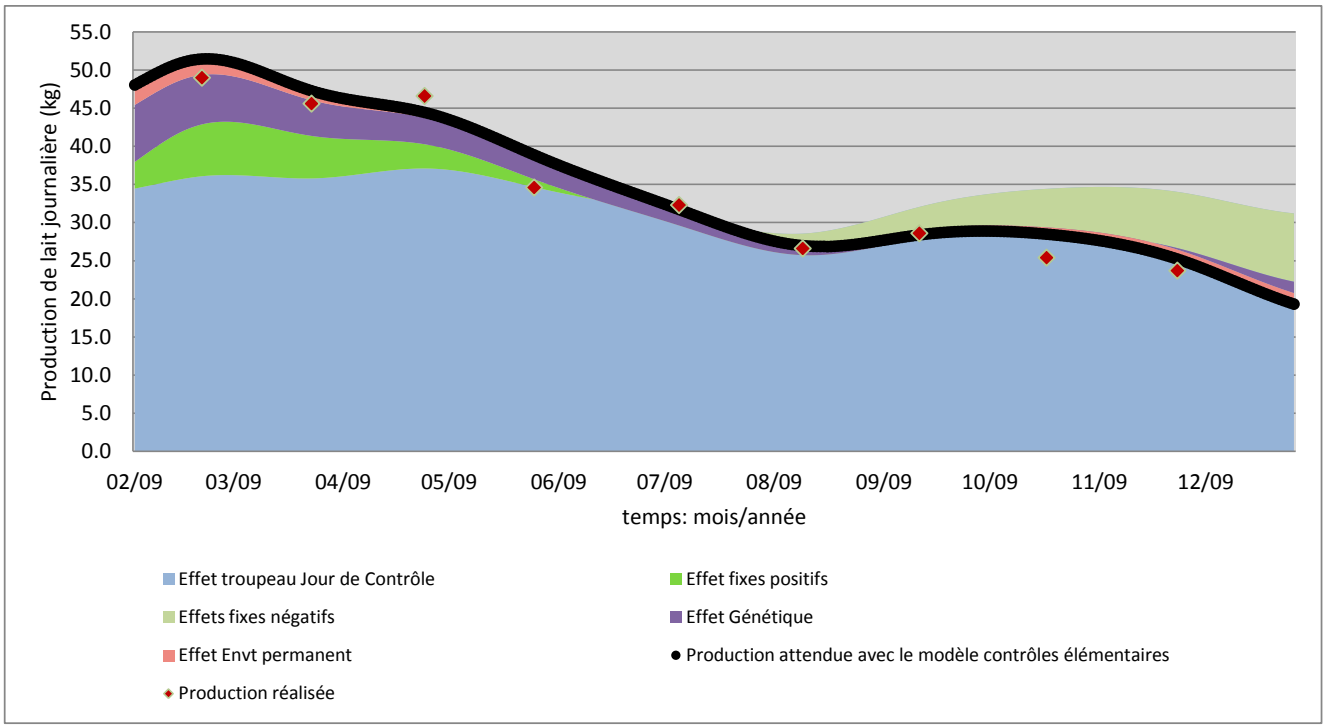


FIGURE 2.2 – Une lactation décryptée par le modèle génétique basé sur les « contrôles élémentaires »

Le modèle s'écrit comme suit ((Leclerc *et al.*, 2008)) :

$Y_{ijklnott'vw} = TJC_i + annee_mois_{jklm} + annee_age_{jkln} + annee_tar_{jklo} + spl_{SL}(mois_{jlm}, t) + spl_{SL}(age_{jln}, t) + spl_{SL}(tar_{jlo}, t) + spl_{GEST}(ranglactation_{jl}, t') + \sum_{r=1}^{4}(\nu_{tqr} a_{vr}) + \sum_{r=1}^{4}(\zeta_{tqr} p_{vr}) + \sum_{r=1}^{6}(\tau_{tqr} h_{wr}) + e_{ijklmnott'vw}$

où $Y_{ijklnott'vw}$ représente la performance de l'animal v au stade de lactation (SL) t du rang de lactation q et au stade de gestation (GEST) t'. TJC_i est l'effet TJC spécifique à cet élevage pour cette date de contrôle. Les termes d'interaction $annee - mois$, $annee - tar$, $annee - age$ correspondent respectivement aux combinaisons $annee_k - region_l - rangdelactation_j - moisdevelage_m$, $annee_k - region_l - rangdelactation_j - dureedelaperiodedetarissementprecedente_o$, $annee_k - region_l - rangdelactation_j - ageauvelage_n$. Ces termes sont constants sur la lactation. $spl_{var}(eff)$ est une régression spline de la variable var spécifique à l'effet eff. Trois fonctions splines permettent de modéliser l'effet du stade de lactation dépendant des facteurs $rangdelactation_j - region_l - moisdevelage_m$, $rangdelactation_j - region_l - classedeladureedelaperiodedetarissement_o$ et $rangdelactation_j - region_l - classed'ageauvelage_n$. Une quatrième fonction spline spl_{GEST} modélise l'effet de la gestation au $rangdelactation_j - region_l$ comme une fonction du stade de gestation t'. e est le terme résiduel.

Des régressions aléatoires modélisent les effets génétique (a_{vr}), d'environnement permanent (p_{vr}) et troupeau-année de vêlage (h_{wr}) de l'animal v, troupeau-année de vêlage w avec les coefficients ν_{tqr}, ζ_{tqr}, τ_{tqr}, au stade de lactation t du rang de lactation q.

2.2.3 Les profils Troupeau Jour de Contrôle, reflets de la conduite de troupeau

Pour un élevage donné, le modèle génétique basé sur les « contrôles élémentaires » estime un effet TJC pour chaque date de contrôle qui est commun à l'ensemble des vaches d'une même race présentes dans le troupeau au jour du contrôle. Il représente la part de la performance de toutes les vaches du troupeau permise par l'ensemble des éléments communs au troupeau en ce jour de contrôle de performance comme les conditions climatiques du jour (élément subi par l'éleveur) et la conduite de troupeau du jour (l'alimentation, les conditions de logement... éléments maîtrisés par l'éleveur). Cet effet est corrigé pour tous les autres effets du modèle « contrôles élémentaires » (valeur génétique, âge au vêlage, race, stade de lactation...).

Pour un élevage, la courbe représentant l'effet TJC au cours du temps est appelée profil TJC. Il est obtenu pour chacun des caractères évalués avec le modèle « contrôles élémentaires » tel qu'il est prévu à ce jour c'est à dire pour le lait, le taux butyreux (TB), le taux protéique (TP) et les matières grasse (MG) et protéique (MP). Cependant, en France, les profils de matières ne sont pas valorisés car elles sont peut utilisées pour le conseil en élevage. Un profil TJC pourrait dans l'avenir être obtenu pour le score cellulaire. La figure 2.3 p.39 montre les profils TJC lait, TB, TP pour un élevage sur cinq années.

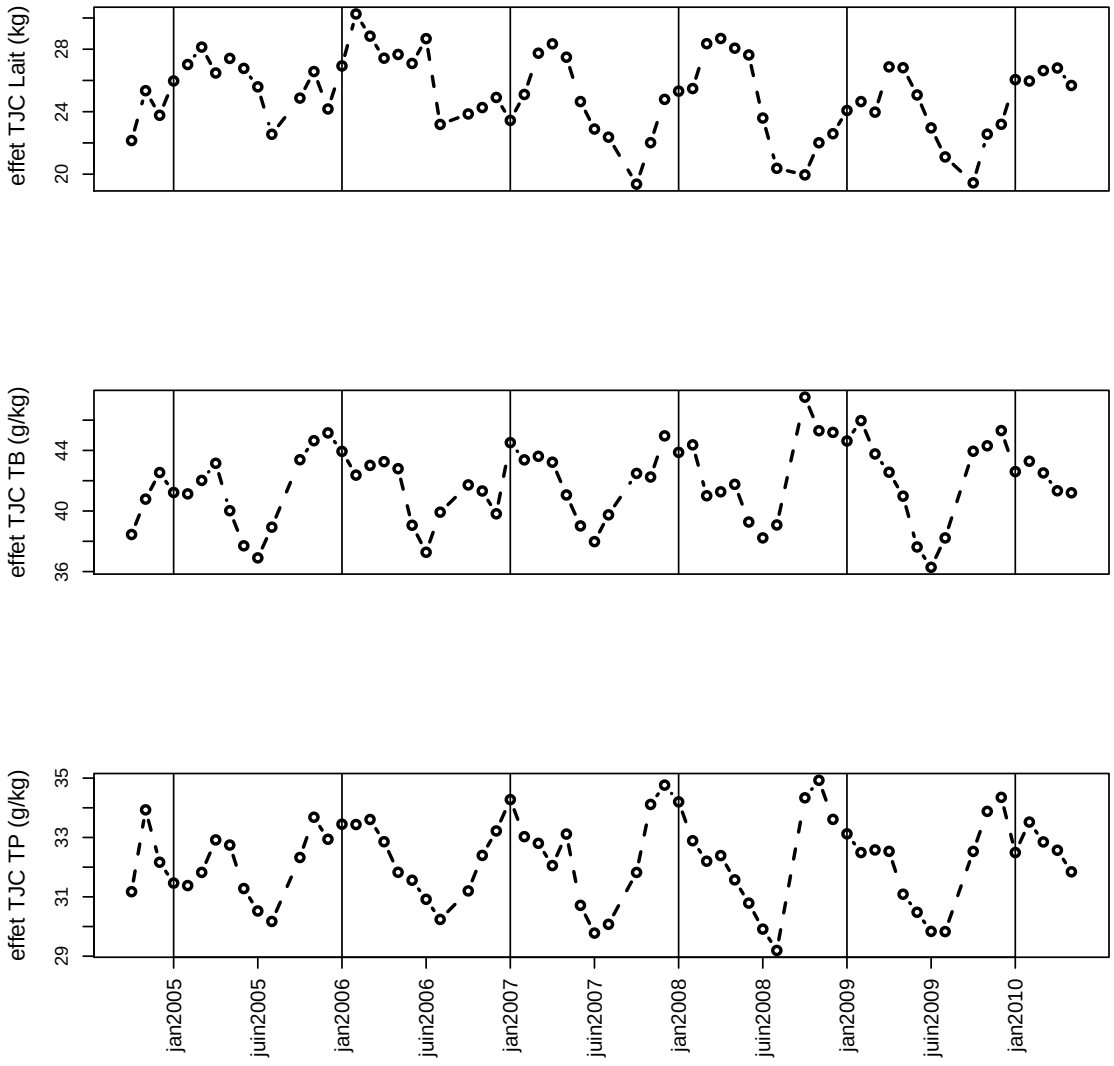


FIGURE 2.3 – Profils TJC Lait TB et TP d'un élevage de 2005 à 2010

En supposant que les conditions climatiques et la conduite saisonnière du troupeau sont répétables entre années dans une exploitation, la forme globale de ses profils TJC

devrait être répétable d'une année sur l'autre. D'une façon générale, l'aspect global du profil est dû, d'une part aux conditions climatiques générales spécifiques à une région (élément non maîtrisable par l'éleveur), d'autre part à la conduite de troupeau adoptée par l'éleveur (élément maîtrisé). Les conditions climatiques difficiles (froid, neige), notamment en zones montagneuses, entraînent chaque année une baisse de l'effet TJC de ces élevages pendant l'hiver ce qui est dû à une alimentation limitée durant cette période. Ces conditions climatiques sont subies et ne peuvent pas être modifiées. En revanche, des élevages qui subissent les mêmes conditions climatiques peuvent avoir des effets TJC différents du fait de l'adaptation de la conduite de troupeau à ces conditions. C'est pourquoi nous faisons l'hypothèse, comme dans (Mayeres *et al.*, 2004; Koivula *et al.*, 2007), que les profils TJC (lait, TP, TB) d'un élevage permettent de mesurer l'efficacité de la conduite du troupeau. Ils sont le reflet de l'impact de la conduite de troupeau (principalement l'alimentation,) au cours du temps, sur la production de chaque animal du troupeau. Les éléments du profil TJC non répétables d'une année sur l'autre sont dus par exemple à des modifications ponctuelles des conditions climatiques et/ou de la situation économique. Par exemple, la canicule de 2003 a entraîné une chute des effets TJC à cette période, et les rallonges de quotas en 2009 ont incité certains éleveurs à modifier l'alimentation de leur troupeau pour leur permettre de produire plus, ce qui a entrainé une hausse ponctuelle des effets TJC de ces élevages. Le profil TJC présenté en figure 2.4 p.41 présente peu d'éléments répétables d'une année sur l'autre. En revanche, l'éleveur semble adapter sa conduite aux fluctuations du prix du lait notamment lors de l'hiver 2007/2008 (cf figure 2.5 p.41).

2.2.4 Les effets TJC et autres co-produits du modèle « contrôles élémentaires » déjà valorisés dans des outils de conseil

L'ensemble des outils présentés ici utilisent les co-produits du modèle génétique basé sur les « contrôles élémentaires ». Ils ont été pressentis suite à la thèse d'Hélène Leclerc (Leclerc, 2008; Leclerc *et al.*, 2009) et développé au sein du projet GENESYS. Ils ont fait l'objet d'une valorisation sous forme de poster au congrès ICAR 2011 (Huquet *et al.*, 2011a) présenté en figure 2.6 p.43 ainsi que d'un article et une présentation orale au congrès Rencontres Recherches Ruminants (Huquet *et al.*, 2011b). Les objectifs des outils puis les moyens mis en place pour les construire vont maintenant être successivement

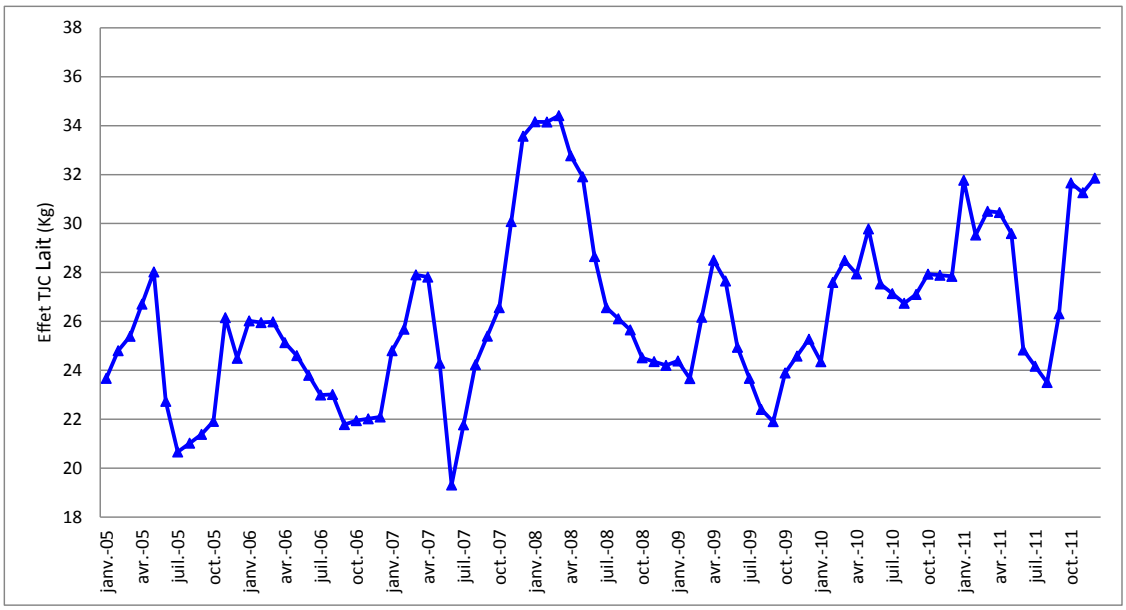


FIGURE 2.4 – Profil Troupeau Jour de Contrôle Lait d'un élevage

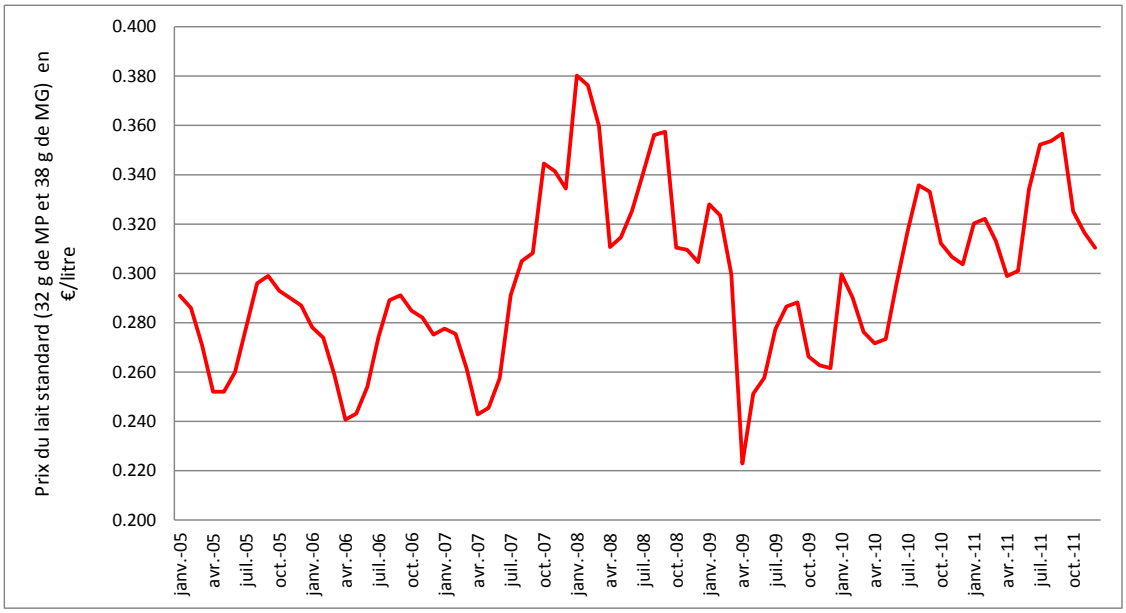


FIGURE 2.5 – Évolution du prix du lait au cours du temps *(source : Institut de l'Élevage-Département Économie)*

présentés.

• Les objectifs des outils de conseil

Les outils sont construits à des fins prospective (analyse a priori) ou rétroactive (analyse a posteriori).

Dans l'approche prospective, un nouvel outil de prédiction de la production laitière est développé. Ce type d'outil est déjà utilisé au sein des Entreprises de Conseil en ELevage mais ils s'appuient sur un référentiel de courbes de lactation (ARCOLA), qui a été élaboré au début des années 1990 et qui intègre uniquement le niveau de production annuel attendu du troupeau, le rang de lactation et le mois de vêlage. Le modèle génétique basé sur les « contrôles élémentaires » permet d'améliorer ces outils. En effet, il permet d'estimer, pour chaque date de contrôle, chacun des facteurs qui jouent sur la performance de l'animal. Il « décortique » la performance pour comprendre quelle est la part due à chaque effet (génétique, mois de vêlage, conduite de troupeau (TJC)). La connaissance de ces effets à l'échelle du stade de lactation et non plus à l'échelle de la lactation complète (comme c'était le cas avec le modèle génétique « lactation ») permet de prédire plus finement la production de chaque vache et par cumul, la production du troupeau entier (application 1 de la figure 2.6). L'ensemble de ces facteurs peut aussi être utilisé pour simuler un changement de conduite touchant précisément un des effets estimés par le modèle « contrôles élémentaires » comme l'âge au vêlage ou le mois de vêlage (application 2 de la figure 2.6).

Dans une approche analytique, l'éleveur et le conseiller peuvent comparer les effets TJC prévisionnels aux effets TJC réalisés. S'il existe une déviation entre ces effets, il faut en comprendre la cause : était-ce dû à un changement volontaire de la conduite de troupeau de la part de l'éleveur (exemple de l'application 3 de la figure 2.6) ou à une autre cause qu'il faut alors diagnostiquer (mauvaises conditions climatiques, dégradation de la qualité de l'aliment...) ?

• Les moyens mis en place pour la construction des outils de conseil

Ce travail a été réalisé le cadre du projet GENESYS.

Monthly milk records: the basis of a new valued extension service for an efficient management

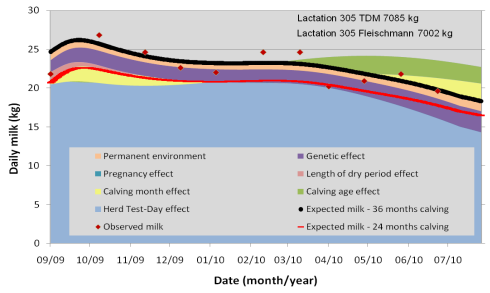
Bérénice Huquet [1,2], Hélène Leclerc [1,2], Vincent Ducrocq [2]
(1) Institut de l'Elevage • 149, rue de Bercy • 75012 Paris • France – (2) UMR 1313 (GABI) • INRA • 78352 Jouy-en-Josas • France

❑ Context:

- Needs for new management tools
- Test Day Models : monthly milk records explained by factors such as age, month of calving, length of dry period, gestation, genetic level depending on lactation stage and a measure of production environmental effect : the **Herd Test-Day** effect (HTD)

❑ Four main applications of Test-Day Model results:

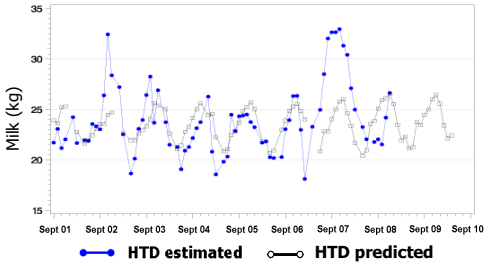
① To forecast future animal production from environmental effects (predicted HTD, age and month of calving…) and animal effects



② To simulate the impact of management changes (calving period, age at first calving…) through a precise decomposition of lactation

Ex: simulation of a cow calving at 2 instead of 3 years of age (red line in the upper figure)

③ To compare predicted HTD with real ones



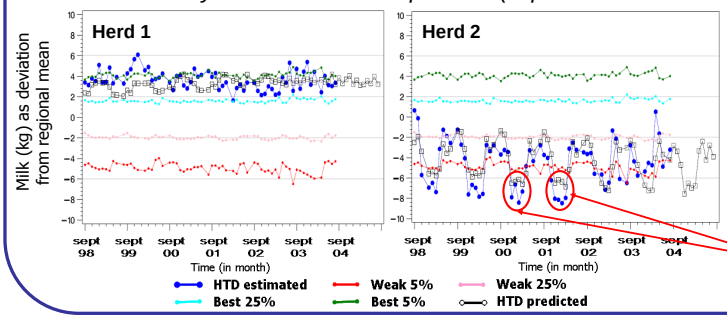
Real vs expected production due to short term environment effect (real vs expected HTD).

Observation : increase of HTD (till +8kg of Milk) at the end of 2007
Interpretation : adaptation of herd management to produce more milk due to a large increase of milk price

.

④ To highlight herd management strengths and weaknesses through Herd Test-Day pattern comparison

2 herds with very different milk HTD patterns (expressed as a deviation from the regional mean)



HTD pattern comparison:
Herd 1 : stable milk production, high level : 4kg/d more than regional means
Herd 2 : Poor management (between -2kg/d and -8kg/d), especially during the winter

⇨ **displays areas for improvements**

❑ Conclusions & perspectives:

- Test-Day Models give keys for the development of new herd management tools.
- The diversity of HTD patterns reflects the diversity of herd management.
- HTD patterns are the basis of a new typology of herd management

We (Institut de l'Elevage, INRA, FCEL) gratefully acknowledge the financial support from CASDAR - GENESYS

ICAR 2011 Annual Workshop: new technologies and new challenges for breeding and herd management

INSTITUT DE L'ELEVAGE

INRA

FIGURE 2.6 – Poster ICAR 2011 présentant les différents outils de conseil en élevage mis en place à partir du modèle génétique « contrôles élémentaires »

Le modèle génétique basée sur les « contrôles élémentaires » est appliqué sur les données de performances au sein de l'Unité Mixte Technologique Génétique, Génomique et Gestion des populations bovines (UMT 3G). Les valeurs des effets aléatoires et fixes du modèle d'évaluation génétique dont l'effet Troupeau Jour de Contrôle (dit « estimé »), sont stockées dans des bases du Système d'Information Génétique. L'ensemble de ces informations peut ensuite être utilisé par les Entreprises de Conseil en ELevage pour construire des outils de prédiction de la production laitière à l'échelle de l'animal et par cumul à l'échelle du troupeau ou pour simuler un changement de conduite.

Les évaluations ont lieu 3 fois par an. Pourtant, le conseil en élevage est une activité en continu auprès des éleveurs qui s'opère plus particulièrement au moment du contrôle de performance (une fois par mois). A cette occasion, éleveur et conseiller peuvent diagnostiquer d'éventuels problèmes de conduite et décider d'adapter la conduite générale du troupeau ou de mettre en œuvre une nouvelle stratégie. Cette réaction doit être rapide, c'est pourquoi les éléments qui leur permettent de faire leurs choix, notamment les effets TJC estimés et les effets TJC prévisionnels qu'il faut comparer pour cette date de contrôle, doivent être disponibles rapidement. En effet, le passage post contrôle de performance a lieu généralement 2 à 5 jours après ce dernier, dès lors que les analyses de taux (TB, TP) ont été effectuées par le laboratoire d'analyse.

Aujourd'hui, la prévision des effets TJC pour les 3 années à venir est réalisée au sein de l'UMT 3G à chaque évaluation génétique. Elle ne fait pas encore l'objet d'une fonction informatique développée par le CTIG (Centre de Traitement de l'Information Génétique de l'INRA de Jouy en Josas) car la méthode mérite d'être améliorée. Actuellement les effets TJC prévisionnels sont calculés comme la moyenne des effets TJC du même mois sur les 5 années précédentes.

En ce qui concerne les TJC estimés, ils ne sont disponibles que pour les jours précédents la dernière évaluation génétique. Par conséquent, entre 2 évaluations, les conseillers ne disposent pas d'effet TJC estimé pour le jour du contrôle de performance. La fréquence des évaluations (3 fois par an) n'est donc pas suffisante pour les outils de conseil et pour des considérations pratiques (temps de calcul), il n'est pas envisageable que les évaluations génétiques soient plus fréquentes. C'est pourquoi, une solution intermédiaire a été mise en place. Une fonction informatique a été créée par le CTIG pour calculer à la demande des effets TJC dits « provisoires » pour toutes les dates de contrôles depuis

la dernière évaluation. Des performances corrigées sont calculées pour ces dates en soustrayant aux performances mesurées lors du contrôle de performance la somme de tous les effets estimés lors de l'évaluation précédente pour les caractéristiques propres de l'animal (rang de lactation, stade de lactation, mois de vêlage, âge au vêlage, effets génétiques et d'environnement permanent) et en supposant une résiduelle moyenne nulle.

L'ensemble de ces outils est actuellement disponible. Cependant, à l'heure actuelle, ils ne gèrent que les cas de figure pris en compte dans l'évaluation génétique : seuls les animaux en première, deuxième ou troisième lactation, avec des stades de lactation compris entre 7 et 305 jours et de races Normande, Montbéliarde et Holstein sont pris en compte. Ces conditions sont jugées trop restrictives pour une utilisation de conseil en élevage. C'est pourquoi une extension de ces outils pour prendre en compte ces nouveaux critères est prévue.

Au sein de cette thèse, nous proposons de valoriser les profils Troupeau-Jour de Contrôle comme base de la définition du milieu dans le cadre de l'étude des interactions Génotype*Milieu.

2.2.5 Utilisation des profils Troupeau-Jour de Contrôle comme base de la définition du milieu : les avantages

Nous proposons d'utiliser les profils TJC Lait, TB et TP comme descripteurs du milieu à la place des définitions proposées précédemment. Cette définition présente plusieurs avantages majeurs : les profils TJC sont disponibles pour tous les élevages adhérents au contrôle de performances et dont les données peuvent être utilisées dans le modèle d'évaluation génétique « contrôles élémentaires », contrairement à d'autres descripteurs de la conduite qui ne sont disponibles que sur un faible nombre d'élevages. Les performances des animaux (ex : production moyenne de lait par animal dans un élevage) sont dues à l'effet du milieu mais aussi à ceux de la génétique et des caractéristiques propres de l'animal comme l'âge et le mois de vêlage... Utiliser les effets TJC au lieu des performances moyennes des animaux d'un élevage permet de se focaliser sur la part de la production uniquement due au milieu (conditions climatiques et conditions d'élevage). Enfin, les profils TJC sont des données temporelles qui apportent une nouvelle dimension à la définition du milieu par rapport aux définitions proposées jusqu'à présent. En effet, les variations de la conduite de troupeau au cours de l'année peuvent être prises en compte. Néanmoins, il

est nécessaire de résumer les effets TJC sous forme de variables synthétiques pour pouvoir les utiliser dans l'étude des interactions Génotype*Milieu.

2.2.6 Profils Troupeau-Jour de Contrôle comme base de la définition du milieu : les défis

L'originalité de cette thèse consiste à définir le milieu à partir de profils TJC des élevages en faisant l'hypothèse que ces profils TJC sont le reflet de leur conduite de troupeau. Cela représente un défi en terme de méthodologie et d'interprétation. D'une part, il faut trouver des méthodes adaptées pour classifier des élevages -c'est à dire pour définir des milieux contrastés- non pas sur la base d'une ou plusieurs variables décrivant la conduite de troupeau mais à partir de séries temporelles (dans notre cas : les profils TJC lait, TB, TP). Les élevages d'un même groupe doivent avoir des profils TJC les plus semblables possibles et les élevages de groupes différents doivent avoir des profils TJC les moins homogènes possibles. D'autre part, les groupes d'élevages obtenus sur la base de leurs profils TJC doivent avoir la même conduite de troupeau qu'il faut réussir à interpréter.

Une étape préliminaire à la classification des élevages est la réflexion sur la définition des critères représentant la similitude entre 2 profils TJC. Plusieurs critères pourraient être pris en compte. Ils sont explicités ci-dessous et la figure 2.7 page 47 les illustrent :

- le niveau moyen du profil : la conduite permet-elle de produire beaucoup ou peu en moyenne sur l'année ? (cf le profil TJC vert vs les autres) mais aussi sa forme :
- l'amplitude du profil : l'effet de la conduite sur les performances est-il très différent d'une saison à l'autre ce qui entraîne un profil TJC avec une forte amplitude (profil orange) ou au contraire est-il très plat (profil rouge) ?
- la saison des pics (maximum) et des creux (minimum) du profil TJC : en quelle saison, la conduite de troupeau permet-elle de produire le plus ou le moins (cf profil bleu vs orange) ?
- la durée des périodes de pics et de creux dans le profil TJC

Une fois ces critères établis, doit-on considérer que deux élevages ayant exactement le même niveau moyen et la même forme de profil TJC mais décalée d'un mois ou deux dans le temps (ex : avec un pic du profil TJC lait en juin plutôt qu'en avril) doivent faire partie du même milieu, c'est à dire considérer que leurs conduites de troupeau sont semblables ?

La méthode de classification doit aussi pouvoir prendre en compte le fait que chaque élevage n'est pas défini par un seul mais par trois profils TJC (pour les trois caractères laitiers). Ce qui revient à faire de la classification multiple de séries temporelles.

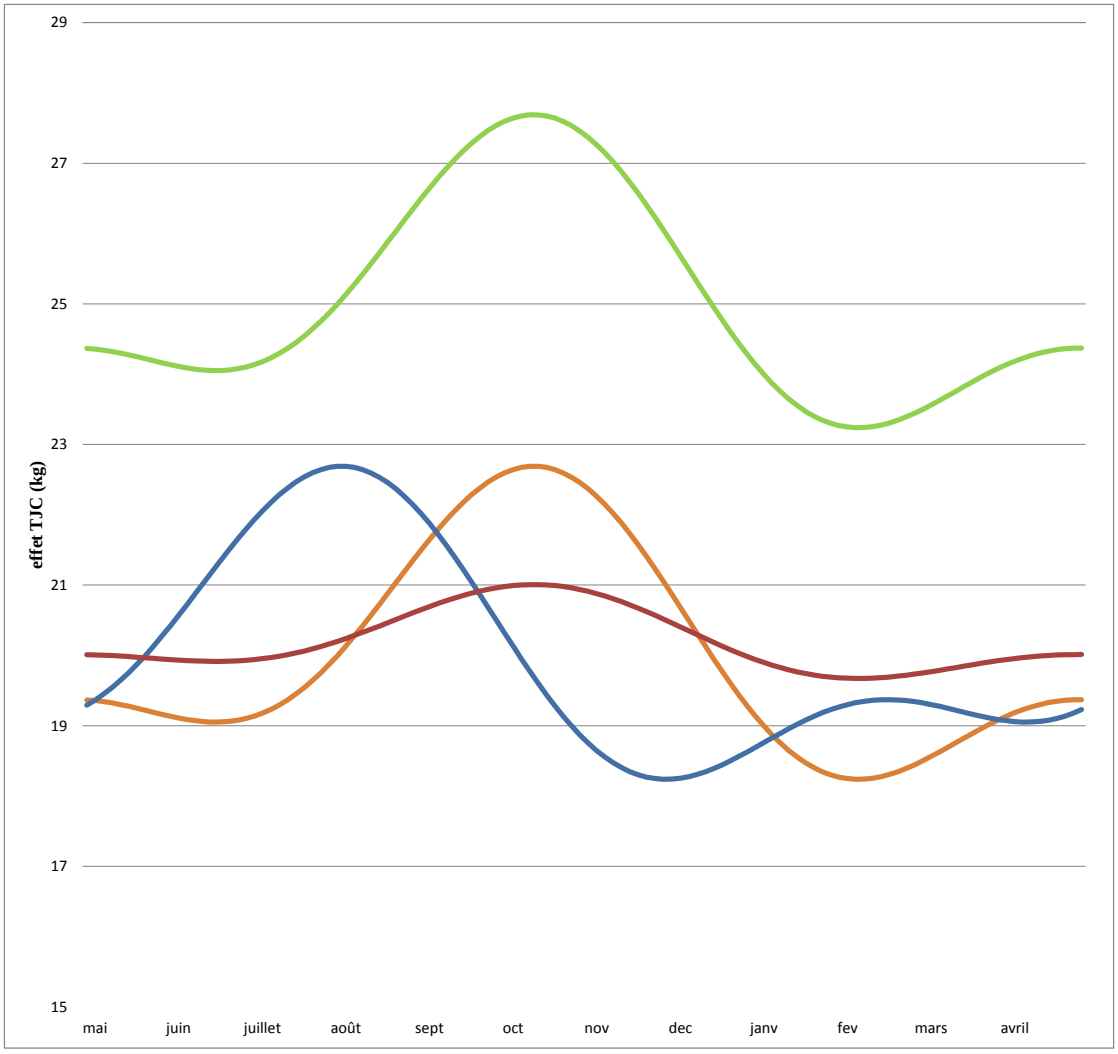


FIGURE 2.7 – Différents profils TJC

2.3 Comment définir le milieu à partir des profils Troupeau Jour de Contrôle ?

2.3.1 Objectifs et Méthodologie générale

Dans le cadre de la thèse, l'objectif de l'article I est de mettre en place la méthodologie de description du milieu qui sera utilisée dans l'étude sur les interactions Génotype*Milieu

réalisée à l'échelle nationale (article II). Cette contrainte impose d'utiliser, dès l'article I, des variables disponibles dans les bases de données nationales pour décrire le milieu : les profils TJC. Le milieu sera d'abord décrit par les principaux facteurs discriminant les élevages selon leurs profils TJC puis par une typologie des élevages qui permettra d'analyser la réalité complexe de la diversité des systèmes d'élevage qui se cache derrière les profils TJC.

Un deuxième objectif est d'étudier le lien entre profils TJC et conduite de troupeau. Pour atteindre le deuxième objectif, il est nécessaire de disposer de données de conduite de troupeau. En effet, pour pouvoir interpréter le lien entre profils TJC et conduite de troupeau, il est indispensable de baser l'étude sur des élevages pour lesquels à la fois les profils TJC et des données de conduite de troupeau sont disponibles. C'est pourquoi cette partie de la thèse se base uniquement sur les élevages dits « genesys » c'est à dire des élevages dont la conduite de troupeau est connue grâce aux bases de données des Réseaux d'Élevage ou aux enquêtes réalisées par les Entreprises de Conseil en Élevage (ECEL) de l'Orne, de Bretagne et du Jura, partenaires du projet GENESYS.

Il n'a pas été envisagé de réaliser cette étude par race. En effet, le nombre d'élevages « genesys » étant limité pour les races normande et montbéliarde, l'analyse prend en compte simultanément les races normande, monbéliarde et holstein ; un même type de conduite de troupeau pouvant être partagé par des élevages utilisant des races différentes.

L'approche choisie pour atteindre ces objectifs est constituée de cinq étapes :

1. Définir les variables (appelées « descripteurs ») qui permettront de décrire les profils TJC. Les profils TJC étant des données complexes, il était nécessaire de définir des variables plus simples. La définition de ces variables a fait l'objet d'études courtes qui sont développées dans le prochain paragraphe.

2. Décrire la variabilité des profils TJC au sein des élevages « genesys » par une analyse factorielle. Elle permet de décrire les profils TJC, en vue de réaliser par la suite une classification des élevages, en se posant les questions suivantes : « Quels sont les élevages qui se ressemblent le plus (d'après leurs profils TJC) ? », « Quels sont ceux qui diffèrent le plus ? », « Quelles sont les principales dimensions de variabilité des élevages décrits par leurs profils TJC ? » (Escofier & Pagès, 2008). Autrement dit « Quelles sont les principales caractéristiques des profils TJC qui permettent de discriminer au mieux les élevages ? » .

3. Décrire les élevages par leurs coordonnées sur les premiers axes de l'analyse factorielle qui sont les « principales dimensions de variabilité » des profils TJC. Cette étape décrit les élevages uniquement par les éléments qui permettent de les discriminer au mieux.
4. Classifier les élevages sur la base de leurs coordonnées sur les premiers axes de l'analyse factorielle. L'objectif est de créer des groupes d'élevages au sein desquels les profils TJC sont les plus semblables possibles et entre lesquels les profils TJC sont les plus différents possibles.
5. Décrire les groupes d'élevages obtenus par leurs profils TJC (à partir desquels ont été créés les groupes) et par leur conduite de troupeau afin d'interpréter le lien entre profils TJC et conduite de troupeau.

2.3.2 Les descripteurs des profils TJC

L'objectif est d'obtenir des descripteurs des profils TJC qui à la fois soient simples et concentrent le maximum d'informations contenues dans les profils TJC et qui soient le reflet de la conduite globale du troupeau (pas d'éléments ponctuels). Ces descripteurs seront ensuite à la base de la définition du milieu dans l'étude des interactions Génotype*Milieu (soit sous forme de groupes d'élevages, soit sous forme d'un gradient).

Cette étude a aussi permis de se familiariser avec les profils TJC. Les différentes étapes de la réflexion sur la construction de ces descripteurs sont présentées ci-après.

• Remarque préliminaire

Les effets TJC utilisés pour construire les profils TJC étant estimés intra race par le modèle génétique « contrôles élémentaires », il est nécessaire de s'affranchir de l'effet race sur les profils TJC. Comme indiqué précédemment, les profils TJC sont utilisés pour décrire des types de systèmes d'élevages ; un même système pouvant être utilisé pour des races différentes. Tous les profils TJC ont donc été centrées par race. Les descripteurs des profils TJC ont ensuite été définis sur ces profils centrés.

• Les effets TJC comme descripteurs des profils TJC

Chaque profil TJC d'un élevage a été décrit par 72 variables correspondant aux effets TJC estimés à partir des contrôles de performance réalisés chaque mois pendant 6 ans de 2005 à 2010 (c'est à dire lors de 72 contrôles). Chaque élevage avait ainsi un profil TJC Lait, TB, TP ; il était donc décrit par 72*3=216 variables très corrélées entre elles. Cette description

des profils TJC ne permettait toutefois pas de discriminer les élevages autrement que sur le niveau moyen de leurs profils (leur conduite permet-elle de produire peu ou beaucoup ?) et pas sur leur forme ; c'est pourquoi il fallait trouver d'autres descripteurs des profils.

• Les moyennes saisonnières des effets TJC comme descripteurs des profils TJC

Afin de prendre en compte la forme des profils TJC, plusieurs nouveaux descripteurs des profils TJC ont donc été construits. Il s'agissait :

- du niveau moyen annuel des effets TJC par caractère (en moyenne sur l'année, la conduite permet de produire beaucoup ou peu ?)
- du niveau moyen par saison des effets TJC par caractère (pour chaque saison, la conduite permet de produire beaucoup ou peu ?)
- des écarts entre saisons (la conduite permet-elle de produire des quantités très différentes selon les saisons ?)
- de la variabilité des effets TJC au cours des mois, estimée à partir de leur variance

Ces variables ont été calculées par année dans un premier temps puis en moyenne sur 5 ans. La figure 2.8 p.51 illustre la manière dont les saisons et les niveaux moyens annuel ou par saison ont été définis. Comme la saison de pâturage ou de retour à l'étable est variable d'une région à l'autre en France, la définition des saisons a nécessité d'exclure certains mois. Nous avons fait le choix de définir le printemps par les mois de mai et juin (en vert), l'été par les mois de juillet et août (en rouge) et l'hiver par les mois de décembre, janvier et février (en bleu) afin de ne pas prendre en compte des mois de transition en terme de conduite de troupeau.

Ces descripteurs étaient sensibles à des données extrêmes ou non représentatives de l'ensemble du profil. C'est pourquoi cette définition n'a pas été conservée par la suite.

• Conclusion de ces premières études sur les descripteurs potentiels des profils TJC

L'ensemble des définitions proposées ci-avant présentaient l'inconvénient majeur de se baser sur les profils TJC qui sont certes le reflet de la conduite de troupeau globale dans l'exploitation mais qui comportent aussi une part d'éléments de conduite ponctuels pouvant être dus à des évènements climatiques non maîtrisables, des aléas non maîtrisés dans la conduite de troupeau, ou encore des modifications ponctuelles du contexte économique laitier. Pour décrire la conduite de troupeau d'un élevage par les profils TJC, il faut essayer de s'affranchir de ces évènements ponctuels afin de se focaliser sur les caractéristiques du

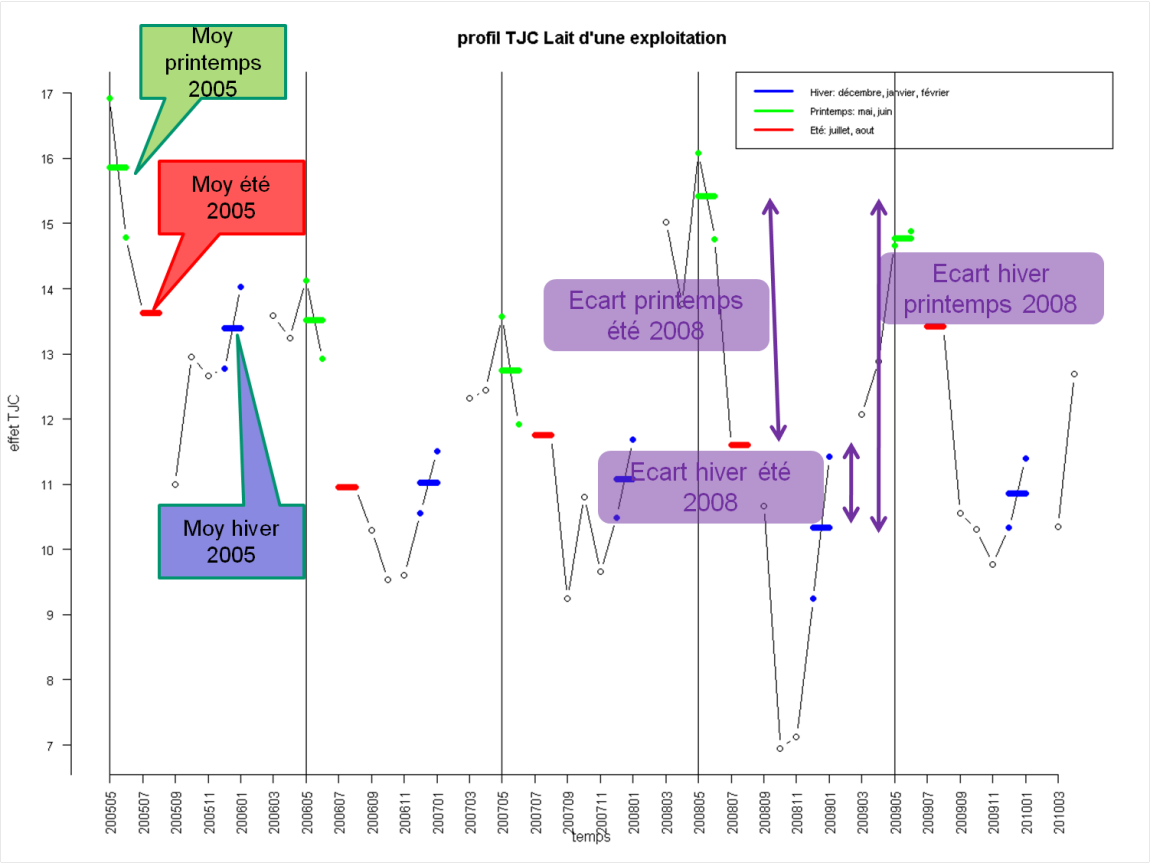


FIGURE 2.8 – Descripteurs profil TJC Lait d'un élevage de 2005 à 2010 par saison

profil qui se répètent chaque année et sont vraiment le reflet de la stratégie de conduite de troupeau de l'éleveur.

- Les descripteurs des profils TJC retenus

Afin de prendre en compte l'aspect temporel des profils TJC et uniquement leurs caractéristiques répétables d'une année sur l'autre, les profils TJC ont été lissés. La méthode utilisée est inspirée de (Koivula *et al.*, 2007) et consiste à modéliser un profil TJC comme une combinaison linéaire d'un niveau moyen, d'une tendance et de 3 fonctions sinusoïdales (présentées sur la figure 2.9 p.52). La description exacte du modèle utilisé figure dans l'article I.

La tendance permet de prendre en compte des élevages dont les profils ont une allure semblable d'une année sur l'autre mais dont le niveau moyen diminue ou augmente année après année. Deux fonctions sinusoïdales ont pour période 1 an avec un décalage de phase de 2 mois et une sinusoïde a pour période 6 mois ce qui permet de modéliser toute sorte de profils (avec un ou 2 pics dans l'année par exemple) en appliquant à ces courbes une

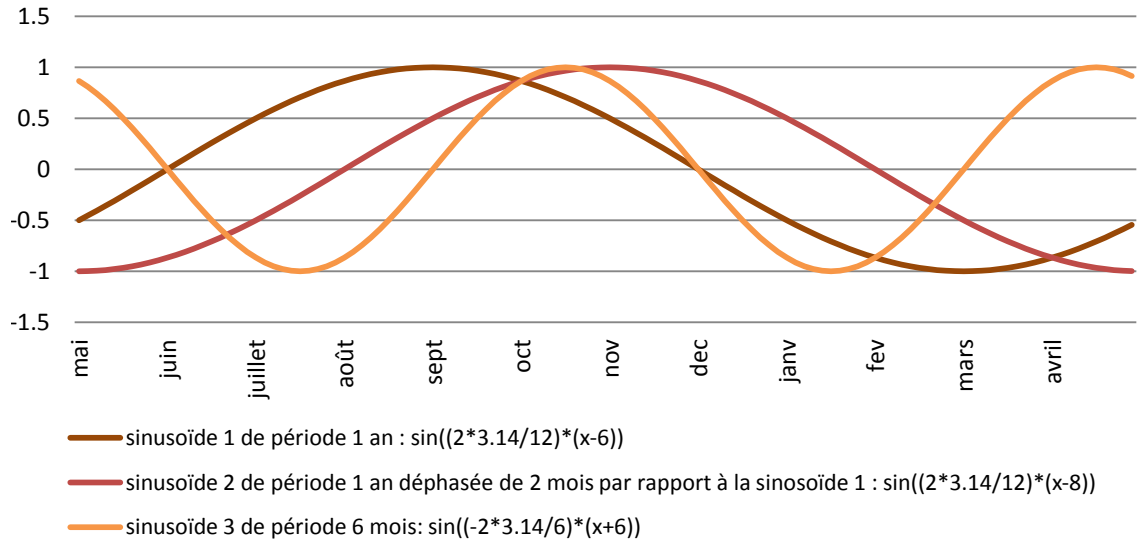


FIGURE 2.9 – Les 3 fonctions sinusoïdes utilisées pour lisser les profils TJC

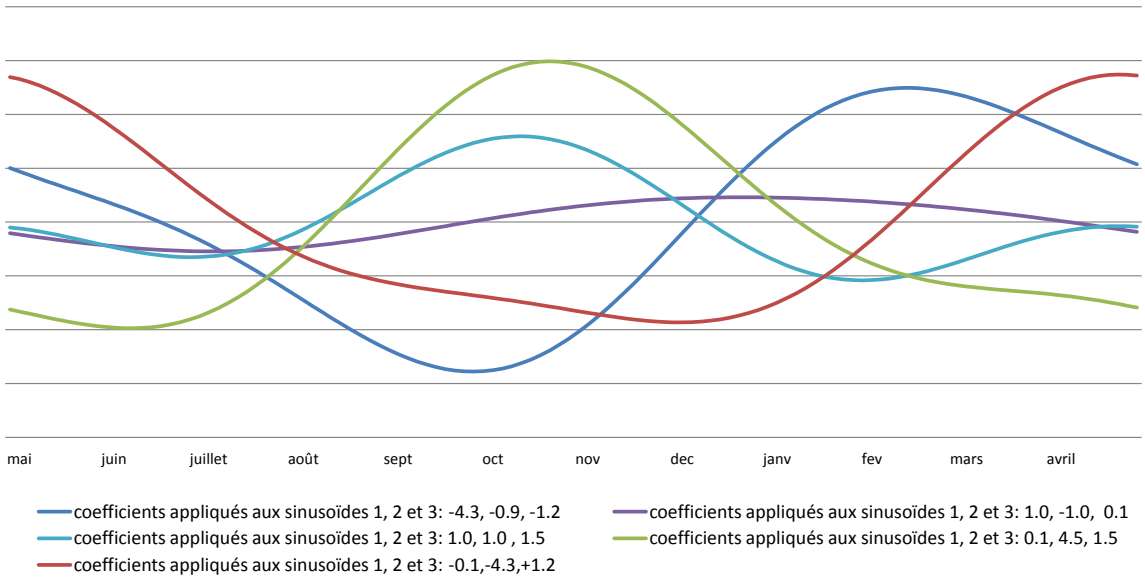


FIGURE 2.10 – Reconstitution de profils TJC à partir de 3 sinusoïdes en appliquant différents jeux de coefficients (la tendance et le niveau moyen ne sont pas pris en compte dans cette reconstitution)

gamme variable de coefficients. La figure 2.10 p.52 donne des exemples de reconstitution de profils TJC à partir des 3 sinusoïdes (sans prendre en compte le niveau moyen et la tendance). Les profils TJC reconstitués peuvent avoir des amplitudes importantes ou faibles (courbe verte vs courbe violette) et différentes périodes favorisant la production (mai, novembre et février pour les courbes rouge, verte et bleue respectivement).

En modélisant les profils TJC de cette manière, ils peuvent être décrits individuellement par 5 paramètres correspondant au niveau moyen, à la tendance et à la forme du profil. Ces paramètres ont d'abord été pressentis comme descripteurs des profils. Cependant, les coefficients relatifs aux 3 sinusoïdes étaient difficiles à interpréter en terme de forme de courbe. C'est pourquoi les descripteurs des profils TJC finalement retenus ont été des variables décrivant les profils TJC lissés : la moyenne (niveau moyen intra-élevage), le minimum et le maximum du profil TJC lissé ainsi que les mois qui y sont associés (mois min et mois max) et l'amplitude du profil. Cette définition permet de se baser sur la part des profils répétable d'une année sur l'autre et de travailler sur des variables facilement interprétables. Un graphe inclus dans l'article I explicite ces variables. Ces descripteurs des profils TJC sont à la base de la définition du milieu proposée dans les articles I et II.

2.3.3 Définition du milieu à partir des profils Troupeau Jour de Contrôle : Article I

Huquet B., Leclerc H. and Ducrocq V.,2012 Characterization of French dairy farm environments from herd-test-day profiles. *J Dairy Sci*, 95 :4085-4098.

J. Dairy Sci. 95:4085–4098
http://dx.doi.org/10.3168/jds.2011-5001
© American Dairy Science Association®, 2012.

Characterization of French dairy farm environments from herd-test-day profiles

B. Huquet,*†[1] H. Leclerc,† and V. Ducrocq*
*INRA, UMR1313 Génétique Animale et Biologie Intégrative (GABI), 78352 Jouy-en-Josas, France
†Institut de l'Elevage, 149 rue de Bercy, 75012 Paris, France

ABSTRACT

Genotype by environment interactions are ignored in national genetic evaluations of dairy cattle. However, some breeders consider that the genetic merit of top sires is not correctly estimated for use in their own particular environment or with their own herd management. With the objective of later investigating genotype by environment interactions at the national level, we studied the relationship between herd management (e.g., feeding system, herd size, production, workforce) and herd-test-day (HTD) profiles for milk yield and contents of fat and protein using a database of 934 herds. Herd-test-day profiles, estimated using a test-day model, are thought to reflect herd management because they represent the part of production due only to month-to-month variations in environmental conditions of production; that is, those related mainly to feeding and climate conditions. Herd clustering based on results from a factor analysis of descriptors of HTD profiles was performed. Each cluster of herds was then characterized in terms of herd management. Three herd clusters were identified. Cluster 1 was characterized by low HTD profiles for milk yield and protein and fat contents as well as a low level of intensification. Cluster 2 was defined by a high milk yield HTD profile and a high level of intensification. Cluster 3 consisted of herds with HTD profiles showing high protein and fat contents but it was difficult to link this cluster to any specific herd management. The use of clusters based on HTD profiles as descriptors of distinct herd management systems will be used in a genotype by environment interaction study in France.

Key words: genotype by environment interaction, herd environment, herd-test-day

INTRODUCTION

What is the "best cow" for a breeder? For a long time it was thought that the same cow was good for everyone. It now appears that the best cow may be the one best adapted to its own local environment. Two main levers exist to adapt animals to their environment, which is defined in terms of, for example, feeding, climate, and soil fertility. The first one is not addressed in this paper and concerns adapting breeding goals for an entire breed to its environment; for example, by putting more weight on functional or morphological traits in breeding programs to breed more robust animals when they live in more demanding environments. The second is to take advantage of genotype by environment (**G×E**) interactions.

The French genetic evaluation assumes, as do all national evaluations for dairy breeds, that G×E interactions do not exist (Interbull, 2011). However, because of the diversity of pedoclimatic conditions in France, dairy farms have very diversified herd management systems, with different local uses of pasture and maize, for example. For this reason, some breeders question the efficiency of the existing breeding schemes for their own management system. To overcome these concerns, a G×E interaction study at the French national level has been considered necessary.

A crucial point in any G×E interaction study is the precise categorization of the environment. Since the 1990s, a better focus on G×E interactions has led to several definitions of the environment. In 1990, Carabaño et al. (1990) defined the environment by referring to the different regions of the United States. In 1994, the International Bull Service Evaluation (Interbull) implemented multiple-trait across country evaluation (**MACE**), in which each country is considered as a distinct environment (Schaeffer, 1994). Later, to characterize the environment, Weigel and Rekaya (2000) and Zwald et al. (2003) defined multiple-trait herd clusters based on herd management, genetic, and climatic descriptors.

A test-day model is already used in different national evaluations; its current development for French genetic

Received September 30, 2011.
Accepted February 22, 2012.
[1]Corresponding author: berenice.huquet@jouy.inra.fr

evaluation (Leclerc et al., 2008, 2009) makes it possible to draw up a new definition of the environment. Indeed, the French test-day model describes monthly milk records as functions of systematic factors such as age, month of calving, length of dry period, stage of gestation, time-dependent additive genetic and permanent environment effects at the cow level, along with measurement of the local production environment effect: the herd-test-day effect (**HTD**). This effect is shared by all cows of a herd and represents the part of production (milk, protein, and fat yields or contents) on the test-day that can be attributed specifically to feeding and climatic conditions, corrected for all other effects. Fluctuations in the HTD effect over time are the consequences of month-to-month variations in environmental conditions of production and are hereafter called HTD profiles (Figure 1). Herd-test-day profiles (milk yield and protein and fat contents) of a herd are assumed representative of its herd management (essentially feeding), as described in Mayeres et al. (2004) and Koivula et al. (2007).

This study represents the first phase of a larger project in which G×E interaction will be investigated at the French national level. The objectives of the present study were 2-fold: first, to define contrasted environments based on HTD profiles (data that are available at the national level) to use them for G×E interaction studies; and second, to relate these environments to farm-level characteristics of management and feeding systems.

The approach consisted of clustering a sample of herds according to the shape of their HTD profiles for 3 traits (milk yield and fat and protein contents). We assumed that new herds with similar HTD profiles but for which there are no known herd management characteristics available in national databases could be allocated to a particular cluster and used in future G×E interaction studies.

MATERIALS AND METHODS

Herd Management Descriptors and HTD Profiles

In 2009, a survey was carried out on 934 French herds for which herd management was well known by extension specialists. Information on feeding systems and dairy structures (herd size, production, and workforce) was collected. These herds were representative of the 3 major French dairy breeds: Montbéliarde, Normande, and Holstein. We verified that overall these herds were representative of French production systems (Table 1). Additional descriptors of the herds such as age at first calving, calving interval and lactation length for primiparous and multiparous cows, and distribution of calving dates and lactations were included using the national database. Table A1 in the Appendix presents the descriptive statistics of the parameters used to characterize the herd management of the 934 farms.

During spring 2010, a within-breed genetic evaluation using a test-day model described in Leclerc et al. (2008, 2009) was carried out based on records from 12,916,854 Montbéliarde, Normande, and Holstein cows born after 1987 in 84,975 French herds, including the 934 herds of this study. Each trait (milk yield and protein and fat contents) was evaluated using a single-trait random regression test-day model in which the first 3 parities were considered as correlated traits. In addition to the estimation of HTD effects, the shape of the lactation curve related to calving month, calving age, length of dry period and gestation (defined by parity and region of indexation) were taken into account cumulating cubic splines for each level of these effects. Genetic and permanent environment effects were included using a reduced rank model. Each herd was then described by its 3 HTD profiles for milk yield and protein and fat contents from September 2004 to May 2010. These profiles were assumed to be the footprint of the herd management (Figure 1).

Smoothing HTD Profiles

Herd-test-day profiles were smoothed to identify repeated annual features and to buffer outlier test-day effects. Within-breed analyses inspired by the model of Koivula et al. (2007) were carried out using the WOMBAT (Meyer, 2007) software by trait (milk yield and protein and fat contents) and breed. For these analyses, the data set included HTD records from September 2004 to May 2010 of herds with at least 10 records per year (1,714 Montbéliarde herds, 3,780 Normande herds, 19,075 Holstein herds). Herd-test-day records were described using the following random regression model:

$$\begin{aligned}\mathrm{HTD}(t,h,cm,y) = {} & herd_h + d_h t + a_h\,\mathrm{s1(cm)} + b_h\,\mathrm{s2(cm)} \\ & + c_\mathrm{h}\,\mathrm{s3(cm)} + \alpha_{hy}\,\mathrm{s1(cm)} + \beta_{hy}\,\mathrm{s2(cm)} \\ & + \gamma_{hy}\,\mathrm{s3(cm)} + e,\end{aligned}$$

where HTD(t,h,cm,y) is the HTD record of herd h at time t (t represents time in months across years, equal to 1 in September 2004, 2 in October 2004, to 13 in September 2005). A fixed effect for the herd ($herd_h$) was included, and d_h, a_h, b_h, and c_h are the coefficients of the fixed regressions of HTD on the different covariates estimated within herd. The covariates of the model were a linear time trend t and 3 sine functions (named

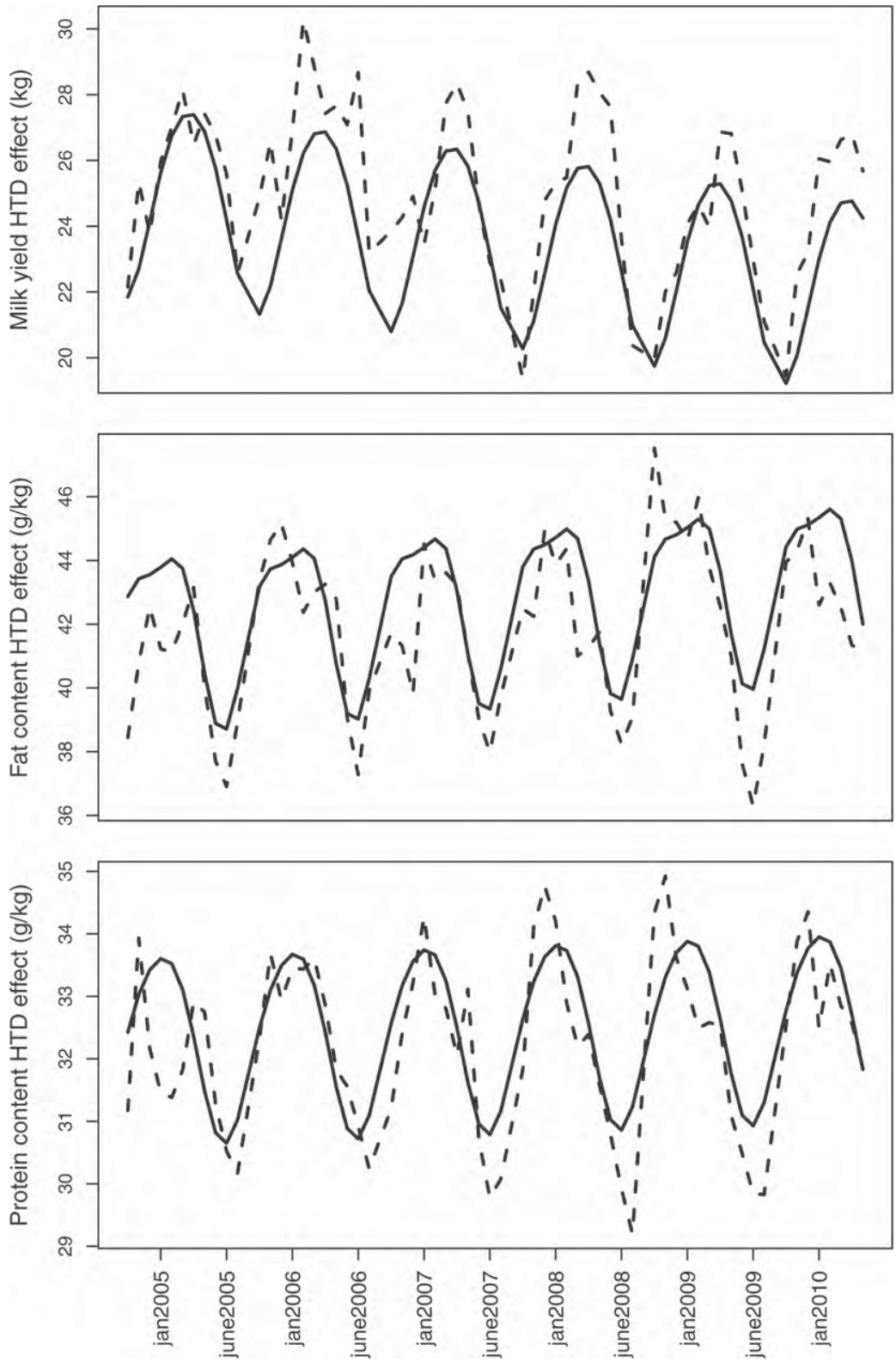


Figure 1. Example of herd-test-day (HTD) profiles (dashed line) and smoothed HTD profiles (solid line) of a herd for milk yield (R^2 milk yield = 0.71), fat content (R^2 fat content = 0.66), and protein content (R^2 protein content = 0.60).

Table 1. Number of herds in the national data set, the data set of the survey, and the data set used for the factorial and cluster analysis for Montbéliarde, Normande, and Holstein breeds and their distribution (%) in their different regions of indexation in France (3, 6, and 8 regions, respectively)

		Region							
Item	Herds (no.)	1	2	3	4	5	6	7	8
Montbéliarde									
National	9,708	57.6	27.6	14.8					
Survey	210	74.5	16.8	8.7					
Analysis	42	35.7	23.8	40.5					
Normande									
National	6,623	7.4	15.2	49.7	12.8	8.5	6.4		
Survey	90	2.2	22.2	42.2	22.2	5.6	5.6		
Analysis	83	3.6	25.5	37.3	19.3	6	7.2		
Holstein									
National	39,458	12.2	24.7	17.5	12.4	7.7	9.2	4.5	11.8
Survey	634	7.3	38.2	13.9	11.6	6.8	8.7	6.2	7.3
Analysis	327	8.9	36.3	20.3	11.1	5.5	0	8.3	9.5

s1, s2, s3) estimated within herd. The values of these sine functions depended on the calendar month (cm) of the HTD record, with cm = 1 in January and cm = 2 in February:

$$\text{s1}(\text{cm}) = \sin\left[\frac{2}{12}\pi \times \left(\text{cm} - 2\right)\right],$$

$$\text{s2}(\text{cm}) = \sin\left[\frac{2}{12}\pi \times \left(\text{cm} - 4\right)\right];$$

that is, with a 2-mo phase difference from s1, and

$$\text{s3}(\text{cm}) = \sin\left[\frac{-2}{6}\pi \times \left(\text{cm} + 10\right)\right].$$

The random regression part included 3 coefficients (α, β, γ) of the same 3 sine functions within herd × year (hy):

$$with\ var\begin{bmatrix}\alpha_{hy}\\ \beta_{hy}\\ \gamma_{hy}\end{bmatrix} = \begin{bmatrix}\sigma^2_{\alpha} & 0 & 0\\ 0 & \sigma^2_{\beta} & 0\\ 0 & 0 & \sigma^2_{\gamma}\end{bmatrix}.$$

The smoothed HTD profiles of the 934 herds for the 3 traits (Figure 1) were built from the estimated fixed effects part of the model using $\hat{a}_h$, $\hat{b}_h$, $\hat{c}_h$, $\hat{d}_h$, $\widehat{herd}_h$. This part represents the repeated annual pattern of the HTD profile. It is the combination of 5 components: the 3 sine curves (2 with a period of 12 mo, 1 with a period of 6 mo), a linear trend, and the herd effect.

Selection of 452 Herds with Consistent HTD Profiles

The coefficient of determination (R^2) of the model for each herd and trait was calculated as the squared correlation between the HTD effects and the smoothed HTD effects. A low R^2 means that the model based on sine curves did not fit the data well. Consequently, a low R^2 reflects a HTD profile that was not consistent from one year to another. In this study, herds for which at least one out of the 3 HTD (milk yield, protein and fat contents) profiles had a low R^2 were discarded. An arbitrary threshold of 0.3 for the R^2 led to 452 herds for which the HTD patterns were considered sufficiently consistent from one year to another from 2004 to 2010. These herds remain representative of the 3 breeds and the diversity of French regions (Table 1).

Describing Smoothed Herd-Test-Day Profiles

The 5 fixed coefficients $\left(\hat{a}_h,\ \hat{b}_h,\ \hat{c}_h,\ \hat{d}_h,\ \widehat{herd}_h\right)$ of the smoothed HTD profiles were hard to interpret and far from intuitive, in particular due to the large overlap between the annual sine functions. Consequently, 7 new descriptors were built to visually describe the smoothed HTD profiles of year 2009: minimum, maximum, and the months where these extrema were observed, as well as the mean, the range, and the width at half the maximum (Figure 2). Values for year 2009 were chosen to facilitate the interpretation of links between HTD profile clusters and data on herd management also collected for year 2009 in the cluster analysis. Table 2 shows the summary statistics for quantitative descriptors for Montbéliarde, Normande, and Holstein herds, and Table 3 shows the distribution of each categorical descriptor. Initially, each categorical descriptor had 12

levels (one per month). Some levels were merged in the level "other" to obtain a minimum number of herds per level, which explains why each categorical descriptor has only 3 or 4 levels.

Factor Analysis

In the data set, the 452 herds were described by 21 variables: the 2009 smoothed HTD profile descriptors (Figure 2) for milk yield and protein and fat contents ($3 \times 7 = 21$ variables). These descriptors were centered within breed to work with the 3 breeds simultaneously. A multiple factorial analysis (**MFA**) determined the main factors explaining the variation in the smoothed HTD profiles (Escoffier and Pagès, 1994). Multiple factorial analysis is a factor analysis similar to principal components analysis, which facilitates the joint use of quantitative and categorical variables, such as months of maximum HTD effect. The number of axes in the MFA is equal to the number of quantitative variables plus the number of levels of each categorical variable. This analysis included 15 quantitative variables and 6 categorical variables (months of extrema) with 3 or 4 levels each; that is, 23 levels in total (Table 3). This leads to 38 factors in the analysis. Eight factors out of the 38 factors of the MFA explained 67% of the total variation in the 21 original smoothed HTD profile descriptors (Figure 3). The coordinates of the herds on these 8 factors were used in the cluster analysis instead of the initial descriptors to eliminate noise. Indeed, MFA can be viewed as a way to separate signal and noise in a data set: the first dimensions extract the most of the information, whereas the last ones are restricted to noise.

Cluster Analysis

This step consisted in creating several herd clusters based on the first 8 factors of the MFA, which summarize the descriptors of the smoothed HTD profiles. Several methods can be used to carry out a cluster analysis on a multivariate data set. For example, Ward (1963) proposed hierarchical clustering algorithms, which break up the data into a hierarchy of clusters. The most commonly used is agglomerative hierarchical clustering (**AHC**). Hartigan and Wong (1979) proposed the K-means algorithm, a partitional algorithm that aims to partition the data set into K groups.

The drawback of hierarchical clustering is that herds that have been incorrectly grouped at an early stage of the algorithm cannot be reallocated to another cluster. It is particularly problematic when a large number of herds have to be clustered. Such a drawback does not exist with partitional algorithms. However, in contrast to hierarchical algorithms, the number of clusters has to be chosen a priori and the method is not deterministic (the choice of initial partition affects the final clusters). We chose to use a hybrid clustering method that combined the advantages of both clustering methods.

First, a K-means clustering of 452 herds into 50 clusters was obtained to reduce the dimension of the data set. Then, an AHC was performed on the centers of these 50 clusters and, according to the maximum inertia gain, 3 clusters of herds were retained. Finally, to reallocate herds that could have been incorrectly grouped by the AHC, a K-means was carried out, choosing the result of the AHC as initial partition of the algorithm.

The HTD profiles of the 452 herds were assumed representative of the continuous nature of the environment. To obtain herd clusters with clearly typical features, the 40 most representative herds of each cluster, namely the 40 nearest to the cluster centers, were selected (Figure 4). Clusters based on these typical herds were described by their 2009 HTD profile descriptors for milk yield and protein and fat contents (active variables of the factor analysis) and by herd management descriptors (variables not included in the building of the 8 factors of the MFA).

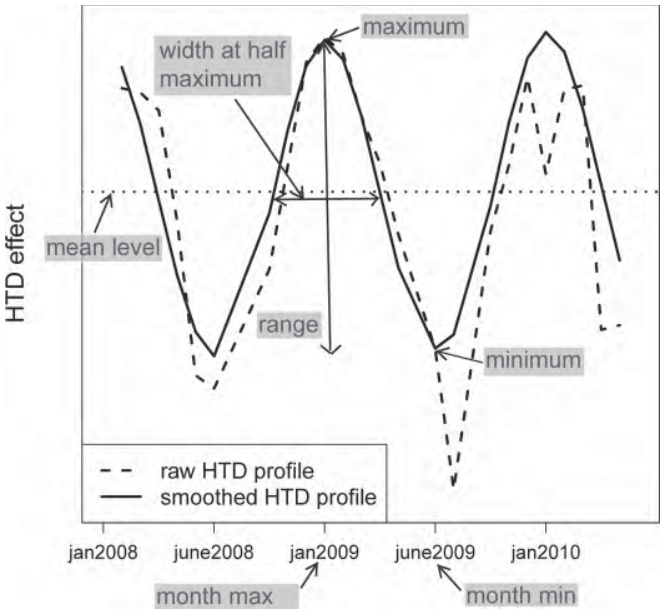


Figure 2. Herd-test-day (HTD) profile, smoothed HTD profile, and descriptors of the smoothed HTD profile used for the factor analysis.

Table 2. Summary statistics of the smoothed herd-test-day (HTD) profile quantitative descriptors for Montbéliarde, Normande, and Holstein herds taken into account in the factorial analysis

HTD profile	Minimum	Maximum	Mean	SD	CV
Montbéliarde (n = 42 herds)					
Milk yield (kg/d)					
Maximum	13.86	27.82	20.91	3.44	0.16
Minimum	9.58	25.06	17.04	3.6	0.21
Mean	12.25	26.07	18.89	3.51	0.19
Range	1.09	6.34	3.88	1.23	0.32
Width (mo)	4	9	7.1	1.14	0.16
Fat content (g/kg per day)					
Maximum	35.45	49.85	42.01	2.78	0.07
Minimum	31.22	42.4	36.5	2.49	0.07
Mean	33.4	46.27	39.38	2.52	0.06
Range	3.11	9.55	5.51	1.44	0.26
Width (mo)	5	8	6.86	0.95	0.14
Protein content (g/kg per day)					
Maximum	31.92	37.43	34.45	1.27	0.04
Minimum	29.4	34.58	31.52	1.33	0.04
Mean	30.49	36.02	33	1.23	0.04
Range	1.45	5.2	2.92	0.86	0.29
Width (mo)	4	10	7.38	1.23	0.17
Normande (n = 83 herds)					
Milk yield (kg/d)					
Maximum	11.69	23.55	19.31	2.19	0.11
Minimum	8.23	20.94	16.20	2.33	0.14
Mean	9.7	22.02	17.67	2.25	0.13
Range	1.64	4.92	3.11	0.8	0.26
Width (mo)	4	9	6.84	1.21	0.18
Fat content (g/kg per day)					
Maximum	38.38	53.28	46.44	2.47	0.05
Minimum	32.64	46.59	40.68	2.14	0.05
Mean	35.61	48.87	43.83	2.17	0.05
Range	1.96	10.08	5.76	1.61	0.28
Width (mo)	5	10	6.63	1.06	0.16
Protein content (g/kg per day)					
Maximum	31.28	38.71	36.06	1.14	0.03
Minimum	29.33	35.75	33.44	1.06	0.03
Mean	30.06	36.95	34.78	1.07	0.03
Range	1.17	5.34	2.62	0.76	0.29
Width (mo)	5	10	7.37	1.13	0.15
Holstein (n = 327 herds)					
Milk yield (kg/d)					
Maximum	16.32	36.8	26.32	3.53	0.13
Minimum	11.63	34.27	22.29	3.83	0.17
Mean	13.65	35.96	24.38	3.64	0.15
Range	0.86	9.65	4.03	1.28	0.32
Width (mo)	5	10	6.93	1.1	0.16
Fat content (g/kg per day)					
Maximum	30.53	49.05	42.69	2.16	0.05
Minimum	26.28	43.69	37.51	2.25	0.06
Mean	28.66	47.08	40.35	2.09	0.05
Range	1.35	9	5.18	1.34	0.26
Width (mo)	5	11	6.92	1.05	0.15
Protein content (g/kg per day)					
Maximum	29.62	37	33.72	1.06	0.03
Minimum	27.02	34.25	31.2	1.11	0.04
Mean	28.67	34.92	32.52	1.01	0.03
Range	0.81	5.32	2.52	0.77	0.31
Width (mo)	4	11	7.19	0.94	0.13

RESULTS

Consistency of HTD Profiles

The goodness of fit of the model describing HTD profiles as a sum of sine functions was better for protein and fat contents than for milk yield in both the complete data set with 24,569 herds (mean R^2 of 0.50, 0.51, and 0.44 for fat content, protein content, and milk yield, respectively) and that with 452 herds (mean R^2 of 0.54, 0.54, and 0.50 for fat content, protein content,

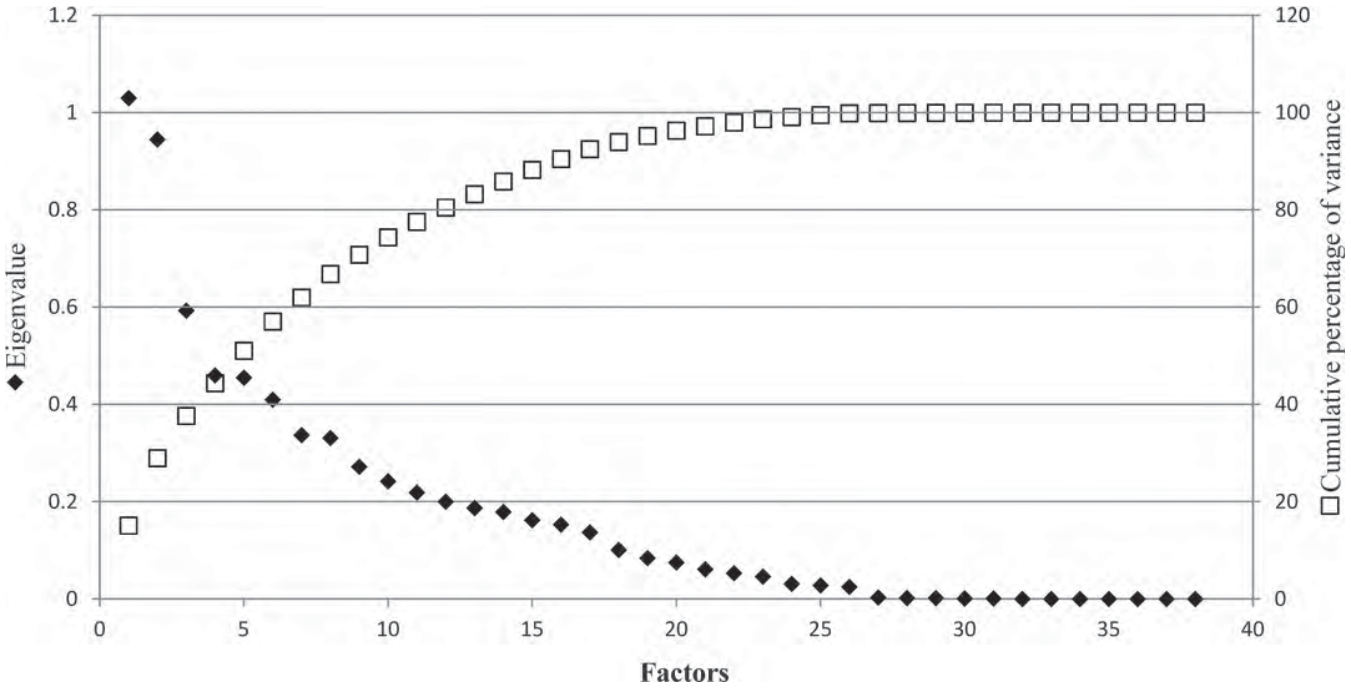


Figure 3. Eigenvalues corresponding to each factor (scale on the left) and the cumulative proportion of the variation that they comprise (scale on the right).

Table 3. Distribution of each categorical descriptor of the smoothed herd-test-day (HTD) profiles for the 452 herds taken into account in the factorial analysis

	Percentage of herds for each level			
Categorical descriptor/level	All herds	Montbéliarde	Normande	Holstein
Month of minimum milk yield HTD profile				
January	34	10	24	40
February	26	31	22	26
March	19	36	22	17
Other	21	23	32	17
Month of maximum milk yield HTD profile				
July	30	19	8	37
August	26	21	18	29
September	26	48	42	20
Other	18	12	32	14
Month of minimum fat content HTD profile				
September	12	14	2	15
October	58	40	55	61
November	22	36	28	19
Other	8	10	15	5
Month of maximum fat content HTD profile				
April	32	40	22	34
May	47	40	53	46
Other	21	20	25	20
Month of minimum protein content HTD profile				
August	22	17	11	25
September	48	29	43	52
October	19	36	30	14
Other	11	18	16	9
Month of minimum protein content HTD profile				
March	22	24	16	23
April	56	57	54	56
May	15	17	25	13
Other	7	2	5	8

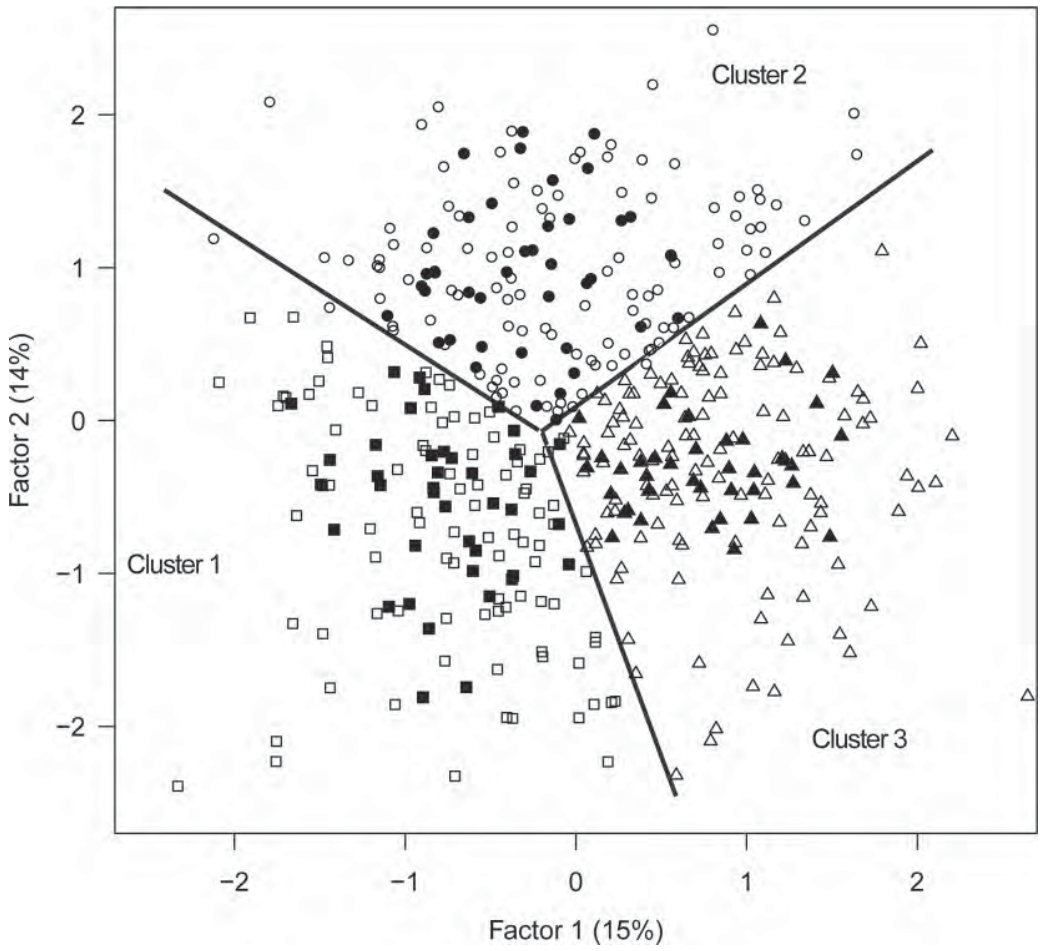


Figure 4. First factorial map of the multiple factor analysis representing the 452 herds in 3 clusters and the 40 most representative herds of each cluster (solid symbols).

and milk yield, respectively). The shape of the HTD profiles for fat and protein contents was more consistent from one year to another than the shape of milk yield HTD profiles, undoubtedly due to a lower effect of herd management changes on protein and fat contents than on milk yield.

Description of Smoothed HTD Profiles

Table 2 shows that the mean level of milk yield HTD profile was greater for the Holstein breed (24 kg/d) than for both Montbéliarde and Normande breeds (19 and 18 kg/d, respectively). For fat content, the mean level was greater for Normande (44 g/kg per day) than for the 2 other breeds (40 g/kg per day for Montbéliarde and Holstein breeds). The same hierarchy was observed for protein content (35 g/kg per day for Normande and 33 g/kg per day for Montbéliarde and Holstein breeds). Range of HTD profiles and width at half the maximum for the 3 traits were of the same order of magnitude for the 3 breeds. These mean differences between breeds show the need to correct for a breed effect in the factorial and cluster analysis to focus on elements that are shared by the 3 breeds. Table 3 shows that whatever the traits, for more than 75% of the herds, the extrema of HTD profiles were located in only one season. For example, for 79% of the herds, the environment (feeding and climate conditions) was mainly unfavorable during winter for milk production: the minimum of milk yield HTD profiles were in January, February, or March. For fat and protein contents, the favorable season was spring and the unfavorable season was the end of summer and the beginning of autumn. However, some herds had other shapes of HTD profiles, with the favorable season for milk production being in spring, for example.

Table 4. Correlations between factors of the multiple factorial analysis and quantitative variables describing herd-test-day (HTD) profiles for milk yield, fat content, and protein content: maximum, minimum, mean level, range, and width at half maximum

	Factor							
HTD profile	1	2	3	4	5	6	7	8
Milk yield								
Maximum	−0.09	0.88	−0.06	0.21	0.27	0.19	0.16	−0.04
Minimum	−0.04	0.93	−0.01	0.26	0.18	0.14	0.03	0.07
Mean	−0.07	0.92	−0.05	0.23	0.23	0.17	0.1	0
Range	−0.12	−0.31	−0.13	−0.19	0.23	0.06	0.38	−0.34
Width	0.08	0.15	0.45	0.15	0.09	−0.45	−0.34	0.37
Fat content								
Maximum	0.84	−0.23	−0.02	0.3	−0.04	0.18	0.24	0.18
Minimum	0.76	−0.16	0.23	0.48	−0.16	0	0.13	−0.18
Mean	0.85	−0.2	0.06	0.39	−0.12	0.11	0.2	0.03
Range	0.13	−0.13	−0.41	−0.27	0.18	0.32	0.14	0.6
Width	−0.02	−0.09	0.63	−0.22	0.41	0.08	0.18	0.11
Protein content								
Maximum	0.78	0.23	−0.13	−0.46	0.21	−0.08	−0.05	−0.01
Minimum	0.65	0.49	0.06	−0.44	−0.22	−0.12	−0.09	−0.16
Mean	0.75	0.4	−0.04	−0.48	0	−0.14	−0.06	−0.07
Range	0.17	−0.4	−0.25	−0.01	0.62	0.08	0.05	0.25
Width	0.02	−0.08	0.16	0.02	0.13	0.59	−0.45	−0.08

MFA Results

The MFA determined the major factors of variation of the shapes of HTD profiles defined by the smoothed HTD profile descriptors. Figure 3 shows the eigenvalues corresponding to each factor. Each factor of the MFA can be interpreted through the variables with high absolute value coordinates and high correlation with the factor (Table 4). Factor 1 was related to the minimum, maximum, and mean of the 2009 smoothed HTD profiles for protein and fat contents, whereas factor 2 was related to the same variables for milk yield. In a factorial analysis, because each consecutive factor is defined to maximize the variability not captured by the preceding factors, consecutive factors (as factor 1 and factor 2) are independent from each other. This indicates that herds with a high mean level of milk HTD profile included herds with either low or high mean levels for HTD protein and fat contents. Factors 1 and 2 explained 29% of the total variation in the HTD profiles (Figure 3), indicating that mean levels in milk yield and protein and fat contents were a major source of variation in HTD profiles. Factors 3 to 8 were defined by variables such as range, full width at half maximum of the 2009 smoothed HTD profiles, and months of maximum and minimum smoothed HTD effects. For example, these factors discriminated herds with HTD protein and fat content profiles with a high range but narrow peak for milk HTD profile from herds with HTD protein and fat content profiles with a low range but wide peak for milk HTD profile. Factors 3 to 8 explained 38% of the total variation in the data, which supports the conclusion that the pattern of the HTD profile (especially the HTD content profiles) is a source of diversity. Altogether, factors 1 to 8 explained 67% of the total variation in the HTD profiles (Figure 3) and were used in the cluster analysis.

Description of Herd Clusters by HTD Profiles and Herd Management Descriptors

Application of the cluster analysis to these 8 factor scores led to 3 clusters of, respectively, 138, 149, and 165 herds. The relatively homogeneous partition of the 3 breeds among the 3 clusters proved that clustering was not due to breed effects (Table 5). The 40 most representative herds of each cluster are indicated by solid symbols in Figure 4. Note that the factorial map does not allow perfect visualization of these herds because only 2 factors are in evidence, whereas herds were represented by 8 factors, which explains why they do not appear exactly on the center of their cluster.

In the following description, a high mean level of the HTD profile must be interpreted within breed, and a

Table 5. Number of herds per breed and cluster

	Cluster			
Breed	1	2	3	Total
Montbéliarde	17	11	14	42
Normande	24	22	37	83
Holstein	97	116	114	327
Total	138	149	165	452

high HTD profile means one with a high mean level. We found one cluster with low HTD profiles for the 3 traits (cluster 1), one cluster with a high milk yield HTD profile (cluster 2), and one cluster with high protein and fat content HTD profiles (cluster 3).

Table 6 shows the variables that characterized the 40 most representative herds of each cluster. Cluster 1 was made up of herds for which the mean levels of milk yield, fat content, and protein content were low. In this group of herds, the level of intensification was low: milk production per cow, density per hectare, quantity of concentrate per cow, and percentage of first-lactation cows were lower than the overall mean. Cluster 2 consisted of herds with a high mean level of milk yield and a low mean level of fat content. The range of the 3 HTD profiles and the width of the peak of protein and fat contents HTD profiles were lower than in the other clusters, showing relatively homogeneous management across seasons. This may be explained by a high level of intensification (contrary to cluster 1); milk production per cow, density per hectare, and quantity of concentrate per cow were higher than the overall mean, and first calvings occurred at an earlier age. Cluster 3 was made up of herds with a low mean level for their milk HTD profile and a high mean level for fat and protein content HTD profiles. The ranges of fat and protein contents HTD profiles were higher than in the other clusters. For this cluster, it was difficult to link HTD profiles to herd management: only the percentage of first-lactation cows was higher than the overall mean. The variables potentially needed to describe this type of herd management might not have been included in the survey or in the variables built from the national database.

Some key data on herd management, such as the duration of fodder distribution or the length of grazing period, were thought to have been collected but the amount of missing data unfortunately prevented their use in the analysis. This could explain the difficulty in assessing the relationship between HTD profile patterns and herd management in each cluster.

Describing the diversity of the environment with only 3 clusters may seem too simplistic. Hence, the same methodology was used to create 6 clusters (results not shown). Cluster 1 did not change; cluster 2 (high milk yield HTD profile) was divided into 2 groups (named 2.1 and 2.2) according to the level and range of protein and fat contents; and cluster 3 (high contents HTD profiles) was divided into 2 groups (named 3.1 and 3.2) according to the range of profiles. Cluster 2.1 included herds with a low mean level and range for fat and protein contents HTD profiles. In contrast, cluster 2.2 contained herds with a high mean level for protein content and a low range for milk yield and protein content. Cluster 3.1 was composed of herds with a small range of fat content HTD profiles, whereas cluster 3.2 consisted of herds with a high range for fat and protein contents. None of the variables describing HTD profiles or herd management characterized the sixth cluster. The interpretation in terms of herd management of

Table 6. Means (SD) of characteristic herd descriptors for each herd cluster

Descriptor	Cluster 1	Cluster 2	Cluster 3	Overall[1]
Herd-test-day profiles				
Milk yield mean level[2] (kg/d)	−1.44*** (1.84)	3.47 *** (1.90)	−0.74 (1.25)	0.43 (2.75)
Fat content mean level[2] (g/kg per day)	−1.23*** (1.00)	−0.82** (1.00)	1.53*** (0.99)	−0.17 (1.58)
Protein content mean level[2] (g/kg per day)	−0.75*** (0.50)	0.18* (0.56)	0.48*** (0.47)	−0.05 (0.73)
Milk yield range (kg/d)	3.89 (1.09)	3.4* (0.8)	3.81 (0.95)	3.7 (1.0)
Fat content range (g/kg per day)	2.62 (0.47)	4.94 (0.84)	5.7* (1.1)	5.3 (1.1)
Protein content range (g/kg per day)	5.26 (1.23)	2.1*** (0.5)	2.8*** (0.5)	2.5 (0.60)
Herd management				
Milk production per cow (kg)	7,428*** (982)	8,260** (1,091)	7,840 (759)	7,865 (1,019)
Fat yield quota (g/L)	42 (1.9)	42 (1.8)	40* (7.0)	42 (4.4)
Total forage area (ha)	78 (38)	61* (22)	78 (53)	72 (41)
Grass area (ha)	52 (40)	36* (17)	49 (42)	45 (35)
Total conserved forages (10^3 kg of DM/LU[3])	3.43** (0.92)	3.97 (0.85)	3.86 (1.07)	3.8 (0.97)
Hay (10^3 kg of DM/LU)	0.8** (1.1)	0.35 (0.40)	0.41 (0.28)	0.5 (0.71)
Maize silage (10^3 kg of DM/LU)	2.31*** (1.23)	3.15 (0.89)	3.12 (1.00)	2.9 (1.11)
Concentrate (kg/cow)	1,232** (387)	1,546** (490)	1,344 (444)	1,378 (461)
Stocking rate (LU/ha of forages)	1.43* (0.51)	1.75* (0.40)	1.53 (0.38)	3.7 (0.44)
Percentage of first lactation	27** (15)	34 (10.5)	37* (16)	17 (15)
Age at first calving (mo)	31 (3.4)	29* (2.8)	31 (4.9)	30 (0.26)

[1]Overall mean of the 120 (3 × 40) most representative herds used for the description of clusters.

[2]Average deviation from the breed mean.

[3]Livestock unit.

*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$: significance of the Student test between the overall mean and the mean of the cluster.

these 6 clusters was difficult and questionable, meaning that the variables used to describe the HTD profiles or herd management were not accurate enough to properly characterize such a number of clusters.

DISCUSSION

This study aimed to find a new description of the environment based on milk yield and fat and protein content HTD profiles. The strategy consisted of using HTD profiles assuming that they were appropriate variables representative of the herd management (essentially feeding) even though they also included the effect of climatic conditions. This description, based on information available at the national level, will be used to more precisely assess the extent of G×E interactions in French dairy herds during the next phase of the project.

Smoothing the HTD profiles using sine functions gave results similar to those of Koivula et al. (2007). All types of combinations of HTD profiles for milk yield and fat and protein contents were found. The diversity of combinations of HTD profiles is undoubtedly related to the diversity of herd management and climate in France. A clear continuum exists in both herd management and HTD profile variability. The major factors of variation in the HTD profiles were the mean levels, whereas variations in profile shapes were less substantial. Three clusters of herds were obtained from the description of their 3 HTD profiles: the first with a low mean level for the 3 traits, the second with a high mean level for milk yield, and the third with a high mean level for fat and protein contents. The first 2 were clearly related to different levels of intensification. Summarizing the variability of herd management in 3 categories may be too crude, and interpretation of the 3 clusters in terms of herd management is quite simplistic. However, surprisingly, a description of the environment through 6 clusters instead of 3 did not give a clearer picture of the existing variability. This may be due to a description of herd management or HTD profile descriptors that is not accurate enough. More precise descriptors of herd management, such as diet fed in each season, would improve the precision of interpretation of these HTD profiles. Indeed, HTD profiles are based on monthly records, whereas the herd management descriptors available in the survey were annual. Moreover, cluster analysis results could be improved by further enrichment of the parameters used to describe the shape of HTD. For example, the method did not detect, although they exist, herds with seasonal variations; for example, herds in which herd management makes it possible to produce more in winter than in summer (or vice versa).

The correct definition of the environment is a crucial point in G×E interaction studies. Four major types of definition exist in the literature. In small-scale studies (with a limited number of herds or animals), the environment can be controlled accurately. For example, it can correspond to different levels of concentrate used (Fulkerson et al., 2008) or to the use of pasture versus confinement (Boettcher et al., 2003; Kearney et al., 2004). However, such criteria do not make it possible to study G×E interactions at a national level. The second definition is used at an international level: in the MACE procedure (Schaeffer, 1994) carried out by Interbull, each country is considered as a distinct environment. This is also the case in the studies of Cienfuegos-Rivas et al. (1999), Weigel et al. (2001), Ojango and Pollott (2002), and Hammami et al. (2008). This approach is a rather extreme approximation because it is known that several environments co-exist in most countries, whereas some herds in different countries could be considered to be in the same environment. For this reason, other researchers have worked on a trans-national definition of environments based on herd management, climate, and genetic merit descriptors (Weigel and Rekaya, 2000; Zwald et al., 2003). Their descriptors of herd management are quite different from those used in this paper. They used, among others, peak yield, days to peak yield, persistency, standard deviation of milk yield, and fat-to-protein ratio. These factors are more a consequence of herd management and genetic level than descriptive elements of the herd management itself. However, as in our study, they defined the environment as a group of herds with similar characteristics. The advantage of such a strategy is that environments are supposedly easier to interpret. In contrast, clusters do not model the existing continuum of the environment. The last way of defining the environment considers this continuum, characterizing it through one or several "environmental parameters." They can be quantitative variables (Fikse et al., 2003; Calus et al., 2005), such as peak milk yield, or variables that combine information on different descriptors (Haskell et al., 2007). Such a definition could be the most realistic one but involves the use of reaction norm models, which are more complex than character state models (in which environments are defined as clusters). One major choice to make is the type of function (e.g., linear or quadratic) to use in modeling the link between the environmental parameter and the genetic merit.

The originality of this study was the use of HTD profiles to define the environment. No similar study has previously been carried out. Our approach is attractive in that it is potentially applicable to a large population without necessitating the collection of a lot of extra information on herd management. Indeed, HTD profiles

are available for all herds participating in the French milk recording system. It thus gives the possibility of accounting for variation in herd management across seasons, in contrast to the definition of Zwald et al. (2003) or Weigel and Rekaya (2000).

Different choices made in this study led to a great simplification of the existing diversity of HTD profiles shapes. These include the arbitrary threshold over which HTD profiles were considered to be consistent from one year to another, the way of summarizing the shape of smoothed HTD profiles by 7 descriptors, the selection of only the first 8 factors of the factorial analysis in the cluster analysis, and the interpretation of 3 clusters on the basis of only 40 herds each. All these elements could be modified to obtain a more precise description of HTD profiles. However, even though the 3 herd clusters obtained can be considered as a simplistic view of the reality, they provide a useful way to study G×E interactions among contrasting environments.

In the next phase of the project, 2 strategies will be implemented to more precisely assess the extent of existing G×E interactions in French dairy herds, not only on production traits but also on important functional traits (e.g., fertility, udder health, longevity). The first one will allocate a large number of herds to a cluster through a discriminant analysis and then apply trait analysis or a character state model using only a selection of records from representative herds of the 3 distinct clusters. The second will use a reaction norm model, taking the first axis of the factor analysis on the descriptors of HTD profiles as environmental parameters.

CONCLUSIONS

In this study, we proposed a new definition of the environment based on HTD profiles (of milk yield and fat and protein contents), estimated from a test-day model, and used it for the creation of herd clusters. These clusters could be interpreted in terms of the level of intensification of herd management. Further analysis using more specific herd management descriptors such as seasonal feed rations could improve the interpretation of HTD profiles. It is hoped that a more meaningful assessment of G×E will result from studies based on a careful use of by-products of national genetic evaluations such as HTD profiles.

ACKNOWLEDGMENTS

This study was financially supported by the French Ministry of Agriculture (Programme CASDAR GENESYS). France Conseil Elevage, the 3 partners Bretagne Conseil Elevage, Jura Conseil Elevage, Orne Conseil Elevage, and les Réseaux d'Elevage (Database DIAPASON, Institut de l'Elevage) are acknowledged for providing the data and for their suggestions. The 3 anonymous reviewers are acknowledged for their constructive evaluation of the manuscript.

REFERENCES

Boettcher, P. J., J. Fatehi, and M. M. Schutz. 2003. Genotype × environment interactions in conventional versus pasture-based dairies in Canada. J. Dairy Sci. 86:383–389.

Calus, M. P. L., J. J. Windig, and R. F. Veerkamp. 2005. Associations among descriptors of herd management and phenotypic and genetic levels of health and fertility. J. Dairy Sci. 88:2178–2189.

Carabaño, M. J., K. M. Wade, and L. D. Van Vleck. 1990. Genotype by environment interactions for milk and fat production across regions of the United States. J. Dairy Sci. 73:173–180.

Cienfuegos-Rivas, E. G., P. A. Oltenacu, R. W. Blake, S. J. Schwager, H. Castillo-Juarez, and F. J. Ruiz. 1999. Interaction between milk yield of Holstein cows in Mexico and the United States. J. Dairy Sci. 82:2218–2223.

Escoffier, B., and J. Pagès. 1994. Multiple factor analysis (AFMULT package). Comput. Stat. Data Anal. 18:121–140.

Fikse, W. F., R. Rekaya, and K. A. Weigel. 2003. Assessment of environmental descriptors for studying genotype by environment interaction. Livest. Prod. Sci. 82:223–231.

Fulkerson, W. J., T. M. Davison, S. C. Garcia, G. Hough, M. E. Goddard, R. Dobos, and M. Blockey. 2008. Holstein-Friesian dairy cows under a predominantly grazing system: Interaction between genotype and environment. J. Dairy Sci. 91:826–839.

Hammami, H., B. Rekik, H. Soyeurt, C. Bastin, J. Stoll, and N. Gengler. 2008. Genotype × environment interaction for milk yield in Holsteins using Luxembourg and Tunisian populations. J. Dairy Sci. 91:3661–3671.

Hartigan, J. A., and M. A. Wong. 1979. Algorithm AS 136: A K-means clustering algorithm. J. R. Stat. Soc. Ser. C Appl. Stat. 28:100–108.

Haskell, M. J., S. Brotherstone, A. B. Lawrence, and I. M. S. White. 2007. Characterization of the dairy farm environment in Great Britain and the effect of the farm environment on cow life span. J. Dairy Sci. 90:5316–5323.

Interbull. 2011. Description of National Genetic Evaluation Systems for dairy cattle traits as applied in different Interbull member countries. Accessed Dec. 13, 2011. http://www-interbull.slu.se/national_ges_info2/framesida-ges.htm.

Kearney, J. F., M. M. Schutz, P. J. Boettcher, and K. A. Weigel. 2004. Genotype × environment interaction for grazing versus confinement. I. Production traits. J. Dairy Sci. 87:501–509.

Koivula, M., J. I. Nousiainen, J. Nousiainen, 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.

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

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.

Meyer, K. 2007. WOMBAT—A tool for mixed model analyses in quantitative genetics by REML. J. Zhejiang Univ. Sci. B 8:815–821.

Ojango, J. M. K., and G. E. Pollott. 2002. The relationship between Holstein bull breeding values for milk yield derived in both the UK and Kenya. Livest. Prod. Sci. 74:1–12.

Schaeffer, L. R. 1994. Multiple-country comparison of dairy sires. J. Dairy Sci. 77:2671–2678.
Ward, J. H. 1963. Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc. 58:236–244.
Weigel, K. A., and R. Rekaya. 2000. A multiple-trait herd cluster model for international dairy sire evaluation. J. Dairy Sci. 83:815–821.
Weigel, K. A., R. Rekaya, N. R. Zwald, and W. F. Fikse. 2001. International genetic evaluation of dairy sires using a multiple-trait model with individual animal performance records. J. Dairy Sci. 84:2789–2795.
Zwald, N. R., K. A. Weigel, W. F. Fikse, and R. Rekaya. 2003. Application of a multiple-trait herd cluster model for genetic evaluation of dairy sires from seventeen countries. J. Dairy Sci. 86:376–382.

APPENDIX

Table A1. Summary statistics of descriptors of herd management for the 934 herds in the study

Item	Minimum	Maximum	Mean	SD	CV	Missing data (%)
Areas						
Total area (ha)	10.5	470.4	106.7	61.85	0.58	4.82
Total area per cow (ha)	0.66	113.5	10.17	14.37	1.41	5.78
Total forage area (ha)	6.1	348.8	72.52	40.09	0.55	4.71
Total forage area per cow (ha)	0.2	112	7.2	10.87	1.51	5.67
Total forage area dedicated to dairy cattle (ha)	0	250.5	67	34.43	0.51	28.37
Grass area (ha)	2.27	300	53.4	41.33	0.77	4.82
Maize area (ha)	0	103.5	18.7	15.89	0.85	4.07
Maize area per cow (ha)	0	33.6	1.78	2.99	1.68	5.03
Cereal, oilseed and protein seed area (ha)	0	253	35.28	40.56	1.15	9.96
Maize (ha)/total forage (ha) area ratio	0	0.89	0.31	0.23	0.74	4.71
Production						
Milk production (10^3 kg/herd per year)	88	4,150	410,4	240,2	0.59	29.23
Milk production per cow (kg)	3,751	97,890	7,637	3,421	0.45	13.28
Fat yield quota (g/L)	0	50.1	40.92	3.57	0.09	15.95
Manpower	0.33	20	2.24	1.13	0.5	3.32
Livestock						
Livestock unit (LU)	19.2	444.8	97.74	48.64	0.5	2.89
Dairy cattle LU	15.6	313.3	83.79	36.24	0.43	2.36
Dairy cows LU	13.4	243.7	56.35	24.16	0.43	2.36
Heifer LU	0	110.93	26.54	14.73	0.56	0.32
Other LU	0	282	13.37	25.42	1.9	2.46
Number of dairy cows	13.1	211.9	55.04	23.02	0.42	2.89
Number of 2-yr-old heifers	0	56	12.63	8.83	0.7	40.47
Number of heifers between 1 and 2 yr old	0	66	19.92	10.2	0.51	40.47
Number of 1-yr-old heifers	0	84	20.75	11.45	0.55	40.47
Stocking rate (LU/ha of forages)	0.18	7.11	1.47	0.55	0.37	4.82
Percentage of the main breed in the herd	50	100	98.6	6.93	0.07	20.88
Feeding						
Total conserved forages (10^3 kg of DM/LU)	0.41	22.93	3.68	1.33	0.36	16.7
Maize silage (10^3 kg of DM/LU)	0	5.95	2.46	1.36	0.55	25.05
Grass silage (10^3 kg of DM/LU)	0	2.2	0.33	0.46	1.39	34.15
Hay (10^3 kg of DM/LU)	0	6.57	0.92	0.99	1.08	18.52
Concentrate (kg/cow)	307	6,563	1,468	600	0.41	6.96
Concentrate (g/L)	29	798	202.62	76	0.38	7.28
Duration of grazing (d)	0	348	212	40	0.19	72.91
Duration of forage retaining (d)	0	245	47	55	1.17	71.95
Lactation distribution (no.)						
First	0	115	10.32	11.04	1.07	0.96
Second	0	82	8.14	8.47	1.04	0.96
Third	0	37	5.79	5.8	1	0.96
Fourth and greater	0	50	8.07	7.79	0.97	0
Lactation distribution (%)						
First	0	100	28.99	16.08	0.55	0.96
Second	0	100	23.97	14.27	0.6	0.96
Third	0	100	18.62	13.18	0.71	0.96
Fourth and greater	0	100	28.42	17.63	0.62	0.96
Lactation length (d)						
Primiparous	41	796	344	48	0.14	0.32
Multiparous	198	539	332	40	0.12	0.11
Calving interval (d)						
Primiparous	336	581	410	34	0.08	0.32
Multiparous	343	595	412	30	0.07	0.11

Continued

Table A1 (Continued). Summary statistics of descriptors of herd management for the 934 herds in the study

Item	Minimum	Maximum	Mean	SD	CV	Missing data (%)
Percentage of calvings by month						
January	0	87.5	8.27	7.2	0.87	0
February	0	66.67	5.84	5.77	0.99	0
March	0	66.67	6.25	6.12	0.98	0
April	0	42.86	4.63	4.85	1.05	0
May	0	33.33	5.03	4.8	0.95	0
June	0	39.34	6.05	5.49	0.91	0
July	0	51.28	7.35	6.14	0.84	0
August	0	56.1	11.13	7.69	0.69	0
September	0	55.17	13.96	9.11	0.65	0
October	0	50	11.75	7.01	0.6	0
November	0	35.29	10.64	5.87	0.55	0
December	0	28.85	8.76	5.43	0.62	0
Number of calvings by month						
January	0	26	4.11	3.19	0.78	0
February	0	26	2.88	2.73	0.95	0
March	0	25	3.07	2.92	0.95	0
April	0	20	2.44	2.6	1.07	0
May	0	19	2.64	2.82	1.07	0
June	0	29	3.18	3.2	1.01	0
July	0	34	3.92	3.83	0.98	0
August	0	38	6.04	5.02	0.83	0
September	0	35	7.34	5.47	0.75	0
October	0	35	6.19	4.39	0.71	0
November	0	23	5.64	3.8	0.67	0
December	0	25	4.67	3.46	0.74	0
Other						
Age at first calving (mo)	1.96	3.86	2.58	0.3	0.12	0.43
Mean lactation number	1.7	7.5	2.6	0.48	0.18	29.01
Number of milking machines	1	36	9.68	3.68	0.38	20.66
Number of cows by milking machines	1.64	20	5.92	2.0	0.34	21.73
Replacement rate (%)	4	79	32.62	10.21	0.31	18.52

2.3.4 Conclusions de l'article I

Dans le cadre de l'étude des interactions Génotype*Milieu, le milieu a souvent été défini à partir du niveau de performances brutes des animaux. Or, ces performances sont dues aux effets de milieu mais aussi aux effets propres à l'animal : valeur génétique, mois de vêlage, rang de lactation... Utiliser les effets TJC pour définir le milieu permet de se focaliser sur la part de la production uniquement due aux conditions d'élevage. Il n'y a alors pas de confusion entre effet du milieu et effet génétique au sein même de la définition du milieu pour l'étude de leur interaction. De plus, les profils TJC sont disponibles pour tous les élevages adhérents au contrôle laitier, ce qui permet de réaliser une étude à grande échelle contrairement aux études qui utilisent des variables qui décrivent bien la réalité de terrain de la conduite (comme l'alimentation) mais qui ne sont disponibles que sur un faible nombre d'animaux. Ces éléments démontrent l'intérêt d'utiliser les profils TJC pour décrire le milieu dans le cadre de l'étude des interactions Génotype*Milieu. Néanmoins, les effets TJC sont estimés à partir du modèle basé sur les contrôles élémentaires qui néglige les interactions Génotype*Milieu. Par conséquent, la mise en évidence d'interactions Génotype*Milieu à partir de cette définition du milieu peut être plus difficile.

L'article I a donné des premières connaissances sur les profils TJC. D'une part, il existe une grande diversité des profils Troupeau Jour de Contrôle particulièrement en terme de niveau moyen et d'amplitude : certaines conduites de troupeau permettent de produire plus que d'autres et certaines sont plus saisonnalisées que d'autres. D'autre part, l'allure des profils Troupeau Jour de Contrôle est plus répétable d'une année sur l'autre pour les TB et TP que pour le lait.

La méthodologie pour décrire le milieu à partir des profils TJC pour l'étude des interactions Génotype*Milieu a été mise en place. Le milieu peut être décrit soit par les composantes principales d'une analyse factorielle synthétisant les descripteurs des profils TJC, soit par des groupes d'élevages obtenus par une classification mixte (combinaison de la Classification Ascendante Hiérarchique et de la classification par k-means) à partir des coordonnées des élevages sur ces composantes principales. Cette méthode présente néanmoins un certain nombre de limites : seuls les élevages ayant des profils TJC considérés comme stables sont pris en compte dans la classification. Il serait intéressant de s'intéresser à ces élevages pour connaître la cause de l'instabilité de leurs profils TJC. On peut imaginer des causes telles qu'un changement important de conduite (passage en

agriculture biologique par exemple), des effets TJC qui représentent plus les conditions météorologiques que la conduite alimentaire (cad que les effets TJC représentent plus la part non maîtrisable de la conduite que la part maîtrisable). De plus, la méthode n'utilise que les profils TJC lait, TB et TP pour décrire le milieu (les seuls disponibles au moment de l'analyse). Pourtant, on peut imaginer qu'il faudrait étendre le nombre de profils TJC (ex : les profils TJC concernant le score cellulaire) tout particulièrement pour étudier les interactions Génotype*Milieu sur des caractères qui ne soient pas de production laitière.

Cet article a montré que les profils Troupeau Jour de Contrôle peuvent être à la base d'une définition du milieu pour étudier les interactions génotype*milieu mais les 3 classes d'élevages mises en évidence ne sont pas directement utilisables. En effet, l'effectif de la base de données « genesys » a été considérée comme insuffisant pour construire une équation de prédiction permettant d'affecter un élevage qui ne fait pas partie des élevages « genesys » à une classe à partir de ses profils TJC. En revanche, la méthodologie développée dans l'article I pour décrire les profils TJC et classifier des élevages sur cette base a été réutilisée dans l'étude des interactions Génotype*Milieu (article II) qui est réalisée sur une base de données plus importante.

Chapitre 3

Discussion : les effets Troupeau Jour de Contrôle, nouvelles perspectives de valorisation

Au sein du projet GENESYS, les effets Troupeau-Jour de Contrôle ont déjà fait l'objet de plusieurs valorisations présentées dans le chapitre précédent : des outils de conseil en élevage et une nouvelle définition du milieu dans le cadre dans le cadre de l'étude des interactions Génotype*Milieu (article I). Fort des expériences apportées par le projet GENESYS, cette discussion tend à mettre en avant les nouvelles pistes de valorisation des profils TJC ou des améliorations possibles des outils existants.

3.1 Une connaissance plus précise des conduites de troupeau pour mieux interpréter les profils TJC

3.1.1 Article I : des premiers éléments sur le lien entre profils TJC et conduite

Le niveau moyen des profils TJC est révélateur de l'intensité de la conduite qui se traduit par différentes utilisations de l'ensilage maïs, des fourrages, des concentrés... Ces résultats confirment que les profils TJC sont bien le reflet de la conduite mais seuls les aspects de niveau moyen et d'amplitude de profils ont été pris en compte. L'atout principal des profils TJC, c'est à dire le reflet de la conduite *au cours du temps*, reste à exploiter.

La connaissance du lien entre profils TJC et conduite pourrait être améliorée en affinant la connaissance de la conduite associée à ces profils.

3.1.2 Un partenariat entre GENESYS et PhénoFinlait pour aller plus loin

L'étude de l'article I a été menée à partir des élevages, suivis dans le cadre du projet GENESYS, pour lesquels nous ne disposions que d'une description *annuelle* de la conduite de troupeau (surfaces agricoles, stock fourrager annuel, quantité de concentrés moyenne distribuées sur l'année...). Pour mieux interpréter l'allure des profils Troupeau Jour de Contrôle d'un élevage, il faudrait aussi exploiter leur composante temporelle. Pour cela, il est nécessaire de connaître la variation de la conduite de troupeau au cours du temps qui leur est associée. Il s'agit notamment des dates d'entrée et de sortie de pâturage, le calendrier des rations au cours des saisons...

Une étude est actuellement en cours pour analyser le lien entre profils Troupeau Jour de Contrôle et conduite à partir des données d'alimentation recueillies en 2009/2010 dans le cadre du programme PhénoFinlait (Phénotypage Fin du Lait : http://www.phenofinlait.fr/spip.php). PhénoFinlait est un projet de recherche entrepris par le CNIEL (Centre National Interprofessionnel de l'Economie Laitière), l'INRA, l'Institut de l'Elevage, l'UNCEIA (Union Nationale des Coopératives d'Elevage et d'Insémination Animale), Capgenes (organisme et entreprise de Sélection pour l'espèce caprine), Actilait (Institut technique du lait et des produits laitiers), France Conseil Elevage, le Comité National Brebis Laitières et Labogena (laboratoire d'analyses génétiques pour les espèces animales). Il a pour objectif de mettre en relation la composition fine du lait avec le génotype des animaux et les conditions de milieu (l'alimentation des animaux en particulier). Dans ce cadre, 6 relevés d'alimentation sur un an (2009/2010) ont été effectués dans 1000 élevages bovins laitiers : 3 en période hivernale et 3 en période estivale. L'exploitation de ces données dans le cadre d'un partenariat entre les projets GENESYS et PhénoFinlait permettra de mieux faire correspondre l'allure des profils TJC et l'évolution de la conduite sur l'année grâce aux relevés d'alimentation : la composante temporelle des profils TJC pourra alors être exploitée plus avant.

L'interprétation des profils TJC à partir des relevés d'alimentation PhénoFinlait présente certaines difficultés notamment dues au fait que sur les 11 ou 12 effets TJC durant

l'année 2009/2010 ne correspondent pas tous à un relevé d'alimentation. L'intrapolation des 6 relevés d'alimentation aux mois pour lesquels il n'y a pas eu de relevé est délicate car il s'agit des périodes de transition alimentaire. Si les prédictions sont fausses, cela peut mener par la suite à une mauvaise interprétation du lien entre effets Troupeau Jour de Contrôle et conduite. Par conséquent, l'analyse du lien entre effets TJC et conduite doit s'effectuer uniquement à partir des effets TJC pour lesquels il existe des données de conduite.

L'idée serait d'expliquer l'effet TJC de l'élevage h à un temps t par les différentes variables de conduites relevées dans l'élevage h au jour t. Dans un premier temps, les données pourront être traitées par un modèle linéaire : $TJC_{th} = b_0 + b_1 * X_{1th} + ... + b_m * X_{mth}$ avec $X_1 ... X_m$ les variables de conduite. Cependant, la relation entre l'effet TJC et certaines variables de conduite pourraient prendre d'autres formes de type logarithmique ou exponentielle, dans ce cas il faudrait utiliser un modèle linéaire généralisé pour pouvoir utiliser une fonction de lien f entre la variables à prédire et les variables explicatives : $TJC_{th} = f(b_0 + b_1 * X_{1th} + ... + b_m * X_{mth})$. Les modèles additifs généralisés ou generalized additive model (GAM) (Hastie & Tibshirani, 1990) permettent de décrire la relation entre l'effet TJC et chacune des variables explicatives par une fonction non paramétrique de type spline ou régression locale (loess : lissage par interpolation). Ce type de modèle peut néanmoins s'avérer trop compliqué pour décrire le lien entre effet TJC et conduite. Les résultats de ces modèles pourraient permettre d'améliorer les outils existants pour connaitre l'effet de l'alimentation sur la production.

3.2 Utiliser les méthodes d'analyse des séries temporelles : une piste pour la valorisation des profils TJC ?

3.2.1 Principe des méthodes d'analyse des séries temporelles

Par définition, une série temporelle est un ensemble d'observations (X_t) d'une variable statistique faites à intervalles réguliers (année, trimestre, mois, jour...). Considérer les profils TJC comme des séries temporelles est donc une approximation. En effet, un effet TJC est estimé par contrôle de performance. Or, d'une part ces derniers n'ont pas lieu

à date fixe, d'autre part il existe souvent des données manquantes puisque, par exemple, sur la période estivale (juillet/août), il n'y a souvent qu'un seul contrôle réalisé.

Les méthodes d'analyse des séries temporelles ont évolué avec l'avancée des connaissances mathématiques et des capacités de calcul ; il en existe 2 types : les méthodes empiriques (aussi appelées non paramétriques) et les méthodes paramétriques.

Les premières méthodes à avoir été développées sont des méthodes empiriques. Elles décomposent la série temporelle X_t en plusieurs composantes dites tendancielle T_t, saisonnière S_t et conjoncturelle C_t (ou irrégulière). Ces composantes peuvent être combinées de façon additive ($X_t = T_t + S_t + C_t$) ou multiplicative ($X_t = T_t * S_t * C_t$). Une des méthodes empiriques les plus utilisées est la méthode « X11 » (Ladiray & Quenneville, 1999). Elle permet de désaisonnaliser une série temporelle afin de n'étudier que la partie conjoncturelle. Elle se base sur des méthodes de lissage du type des moyennes mobiles. En économie, elle a l'avantage de pouvoir prendre en compte des effets particuliers de calendrier comme l'effet de Pâques qui ne s'applique pas tous les ans à la même date. Chaque série temporelle peut alors être décomposée en 3 courbes décrivant la tendance, la saisonnalité et la conjoncture.

Les méthodes paramétriques ont été développées plus tard. La plus célèbre est le modèle ARIMA (Auto Regressive Integrated Moving Average), un modèle stochastique développé par Box *et al.* (1976). Nous commencerons par introduire un cas particulier du modèle ARIMA : le modèle ARMA (Auto Regressive Moving Average). Ce modèle est une combinaison de 2 processus : un processus autorégressif (AR) et un modèle de moyenne mobile (MA).

Le premier processus consiste à expliquer une observation à l'instant t par une combinaison linéaire des p observations précédentes et d'un aléa, il s'écrit :

$$X_t = c + \sum_{i=1}^{p} (\Phi_i X_{t-i}) + \epsilon_t$$

où c est une constante, ϵ représente la composante aléatoire et les Φ_i représentent les paramètres d'autorégression à estimer.

Le second processus (MA, pour Moving Average) consiste à expliquer une observation à l'instant t par une moyenne pondérée des précédentes ; ce qui revient en fait à l'expliquer par une moyenne à laquelle est ajoutée une somme pondérée des erreurs ayant entaché les

valeurs précédentes (Delignières, 2007). Il s'écrit :

$$X_t = c + \sum_{i=1}^{q} (\Theta_i \epsilon_{t-i}) + \epsilon_t$$

où les Θ_i représentent les paramètres de moyennes mobiles à estimer et ϵ_t la composante aléatoire (erreur) à l'instant t.

Cependant les modèles ARMA ne s'appliquent qu'aux séries stationnaires, c'est à dire aux séries dont la moyenne et la variance sont constantes au cours du temps. Ce n'est pas le cas des profils TJC qui présentent à la fois une tendance et une saisonnalité (c'est à dire une périodicité de période 1 an). Dans ce cas, il faut utiliser le modèle SARIMA (Seasonal Auto Regressive Integrated Moving Average) qui est un cas particulier des modèles ARIMA (Box *et al.*, 1976) plus souvent cités. Dans ce cas 3 paramètres décrivent le modèle : p et q décrits ci-avant et le paramètre d'intégration d qui représente le degré d'intégration nécessaire pour obtenir une série stationnaire.

Les méthodes présentées ici n'ont pas été testées mais la méthode développée dans l'article I est une façon de traiter les profils comme des séries temporelles via une méthode paramétrique. Elle consiste à décrire les profils TJC par une combinaison linéaire d'une tendance et de 3 fonctions sinusoïdales.

Les méthodes d'analyse des séries temporelles paramétriques ou non paramétriques peuvent apporter un regard différent sur les différents outils construits au sein de GENESYS à partir des profils TJC.

3.2.2 Utilisation de la composante saisonnière des profils TJC pour décrire un système de conduite

Dans l'article I, nous nous sommes attachés à extraire des profils TJC les éléments qui se répétaient chaque année via une combinaison linéaire de 3 fonctions sinusoïdales. Cette composante des profils TJC était considérée comme le reflet de la conduite de troupeau globale adoptée par l'éleveur.

Dans la même idée, nous aurions pu utiliser la méthode non paramétrique de désaisonnalisation « X11 ». Cependant, comme dans la méthode de l'article I, chaque élevage aurait été décrit par 3 courbes correspondant à la composante saisonnière des 3 profils

TJC. Cette méthode ne met pas à disposition des paramètres permettant de décrire la courbe de saisonnalité pour réaliser une classification des élevages mais uniquement des courbes. Par conséquent, la méthode « X11 » ne nous permet pas de nous affranchir de l'étape critique de la définition de descripteurs de ces courbes (de type moyenne, amplitude...).

Les méthodes non paramétriques qui extraient la composante saisonnière des séries temporelles semblaient à première vue une bonne piste pour l'étude des conduites de troupeau via les profils TJC. Cependant, leur objectif est d'éliminer les composantes tendancielle et saisonnière pour ensuite se concentrer sur la composante conjoncturelle. Au contraire, notre objectif, dans le cadre de l'étude des types de conduite via les profils TJC, est d'étudier la composante saisonnière.

3.2.3 Utilisation de la composante conjoncturelle dans l'analyse a posteriori des profils TJC

Un des outils développé au sein du projet GENESYS consiste à comparer les effets TJC prédits et les effets TJC estimés pour un jour de contrôle. Il permet à l'éleveur et à son conseiller de mesurer l'effet d'un changement de conduite (comme un ajout de concentré dans la ration) réalisé volontairement depuis le dernier contrôle ou de se rendre compte d'un problème si l'effet TJC estimé est plus faible que l'effet TJC prédit et que ce n'est pas volontaire. L'éleveur peut tenter de diagnostiquer la cause de cet évènement qui peut être, par exemple, météorologique ou technique. La composante conjoncturelle des profils TJC aurait dans cette optique pu être un autre élément du diagnostic puisqu'elle ne révèlerait que la part de la production due aux éléments de la conduite de troupeau mis spécifiquement en place ce jour là mais qui ne correspond pas aux activités classiques réalisées tous les ans à la même époque.

3.2.4 Prédiction de l'effet TJC dans le cadre d'outils de prédiction de la production laitière

Actuellement dans les outils de prédiction de la production laitière mis en place au sein de GENESYS, les effets TJC prédits sont calculés comme la moyenne des effets TJC du même mois sur les 5 années précédentes. D'autres méthodes sont envisagées

pour mieux prendre en compte les tendances au cours des années précédentes. Il s'agit d'une méthode inspirée de la méthode de Koivula *et al.* (2007) comme dans l'article I ou encore du modèle SARIMA présenté ci-avant. L'utilisation des modèles ARIMA pour décrire des séries temporelles a déjà montré son efficacité notamment dans le domaine de l'économétrie. Cependant, il ne faut pas négliger la complexité de ces modèles qui peut être un frein à leur utilisation dans le cadre d'outils de prédiction de la production laitière.

Pour appliquer un modèle SARIMA aux profils TJC, il faut déterminer les ordres des processus autorégressif p, d'integration d (qui permet de retirer la tendance de la série) et de moyennes mobiles q. Pour les profils TJC, la tendance est, dans la majorité des cas, linéaire donc d peut être considéré comme égal à 1. Il reste à déterminer le nombre d'effets TJC précédents à utiliser pour expliquer l'effet TJC à un instant t dans le processus autorégressif (p) et dans le processus de moyenne mobile (q). Plus ces paramètres sont élevés, mieux le modèle s'ajustera aux données. En contrepartie, les paramètres seront plus difficiles à estimer et la modélisation sera sensible à quelques données particulières non révélatrices de l'ensemble du profil. La surparamétrisation est donc un piège dans le cadre de l'utilisation d'un tel modèle dans un but prédictif : si le modèle s'ajuste de façon excessive aux données, sa qualité réelle de prédiction diminue.

3.3 Les profils TJC : outil de comparaisons de l'efficacité des conduites

Actuellement, les éleveurs comparent leurs performances techniques pour trouver des marges de progrès, d'une part en se basant sur les performances brutes des animaux, d'autre part en se comparant à la moyenne nationale ou à la moyenne des élevages de leur zone géographique. Cette situation n'est pas idéale puisque les performances brutes ne sont pas uniquement le reflet de la conduite de troupeau. De plus, au sein d'une zone géographique plusieurs systèmes d'élevage peuvent cohabiter. Or, un éleveur trouvera plus facilement des marges de progrès pour son élevage en se comparant à des élevages adoptant le même système que lui (à moins qu'il veuille totalement changer de système). A l'heure actuelle, il existe des typologies des systèmes d'élevage comme celle proposée par l'Observatoire de l'alimentation des vaches laitières (Brunschwig *et al.*, 2011) qui décrit les systèmes bovins laitiers français en 15 classes en fonction du relief, de la localisation

géographique et du type d'alimentation (pourcentage de maïs dans la Surface Fourragère Principale). Elles permettent à chaque éleveur de se positionner à l'intérieur d'une gamme de systèmes d'élevage. En revanche, les données de performances utilisées pour le conseil en élevage ne peuvent pas être analysées à l'échelle nationale en fonction de ces systèmes puisque les bases nationales ne contiennent pas l'information « système d'élevage ».

Les profils TJC représentent une opportunité pour créer un nouveau référentiel de conduite de troupeau. D'une part, ils se focalisent sur la part de la production uniquement due à l'effet de la conduite contrairement aux performances brutes. D'autre part, ils sont disponibles à l'échelle nationale dans le Système d'Information Génétique ce qui permettrait une exploitation par les Entreprises de Conseil en ELevage (ECEL).

Une typologie des élevages basée sur les profils TJC pourrait donc être à l'origine d'un nouveau référentiel des systèmes d'élevage ou enrichir l'existant. Il reste à définir comment elle sera construite ; l'idéal étant de réaliser la typologie la plus proche possible de la réalité, c'est-à-dire à partir d'un grand nombre d'élevages bien représentatifs de la diversité des systèmes, et suffisamment « intuitive » pour qu'elle soit facilement utilisable par les éleveurs et leurs conseillers. Cette typologie soulève donc deux problèmes majeurs :

- sur quels élevages et descripteurs du système d'élevage peut-on se baser ? Un grand nombre d'élevages dont les profils TJC sont connus mais pas leur conduite ou bien un faible nombre d'élevages dont les profils TJC et leur conduite sont connus ?
- avec quelle méthode définir les types de conduite ? La classification des élevages doit-elle être réalisée à dire d'experts ou automatiquement (comme à travers une Classification Ascendante Hierarchique) ?

Différentes stratégies peuvent être proposées :

• Une typologie sur un grand nombre d'élevages uniquement à partir des profils TJC

La première solution est de réaliser une classification des élevages à dire d'experts à partir de leurs seuls profils TJC. Un arbre de décision permettrait alors de définir l'appartenance d'un élevage à un système sur la base de critères définis par les experts à partir des profils TJC.

La deuxième solution (cf figure 3.1) est de réaliser une classification automatique (non

supervisée) sur l'ensemble des élevages à partir des descripteurs des profils TJC pour les affecter à un groupe qui est censé représenter un système (comme dans l'article II). Il faut ensuite réaliser une analyse discriminante où la variable à expliquer est l'appartenance à un groupe et les variables explicatives sont les descripteurs des profils TJC. Pour définir la qualité du modèle qui permettra d'affecter un nouvel élevage à un groupe à partir des descripteurs de ses profils TJC, il sera nécessaire de couper le jeu de données en données d'apprentissage et de validation ou d'utiliser les méthodes de bootstrap ou de validation croisée. Comme le jeu de données est grand, la méthode de validation sera de bonne qualité (l'équation de prédiction pourra ne pas être testée sur les données qui ont permis d'ajuster le modèle).

Cependant, dans les 2 cas, la classification des élevages obtenue ne peut pas être reliée à une réalité de terrain, et n'est pas donc pas envisageable dans le cadre de la création d'un outil de conseil en élevage. Cette stratégie n'a pas été envisagée dans l'article I puisqu'un des objectifs était d'étudier le lien entre profils TJC et conduite de troupeau.

- Une typologie à partir des profils TJC sur un faible nombre d'élevages dont la conduite est connue (comme dans l'article I)

Dans ce cas, le jeu de données est constitué d'élevages dont les profils TJC et la conduite sont connus (cf figure 3.2). Il s'agit donc, par définition, d'un jeu de donnée d'effectif limité. La méthodologie est la même que dans le paragraphe précédent : une classification des élevages est réalisée sur la base des descripteurs des profils TJC. En revanche, les groupes d'élevages obtenus peuvent être décrits par les variables de conduite. C'est ce qui a été réalisé dans l'article I. Cette typologie est plus ancrée à la réalité de terrain que celle présentée précédemment mais les classes obtenues ne seront pas forcement facilement interprétables en termes de conduite ; cela a d'ailleurs été le cas pour un groupe d'élevages dans l'article I. L'utilisation de données de conduite plus fines que dans l'article I dans le cadre du partenariat GENESYS PhenoFinlait pourrait néanmoins faciliter l'interprétation des groupes d'élevages en termes de conduite.

De plus, le fait de travailler sur un faible nombre d'élevage pose problème pour établir et valider l'équation de prédiction construite lors de l'analyse discriminante (qui permet d'affecter un nouvel élevage à un groupe à partir de ses données TJC).

- Une typologie ancrée à la réalité de terrain mais utilisant les profils TJC

Comme dans la typologie précédente, le jeu de données est constitué d'élevages dont les profils TJC et la conduite sont connus (cf figure 3.2). La stratégie consisterait à réaliser une classification des élevages à partir des données de conduite (classification automatique ou à dire d'experts). Les groupes obtenus seraient alors facilement interprétables en termes de conduite. Ils peuvent aussi être décrits par leurs profils TJC. La variable à expliquer dans l'équation de prédiction définie par l'analyse discriminante serait alors le groupe de l'élevage (préalablement défini à partir des données de conduite) et les variables explicatives seraient les descripteurs des profils TJC.

Cette stratégie est le point de vue opposé de celui pris dans l'article I pour étudier le lien entre profils TJC et conduite. Dans cette stratégie, on détermine des types de conduite qu'on essaie ensuite de lier à des profils TJC particuliers. Alors que dans l'article I, on cherchait à déterminer des profils TJC types et à les interpréter en terme de conduite. Cette stratégie n'a pas été envisagée dans l'article I car l'objectif était à la fois d'étudier le lien entre profils TJC et conduite mais aussi de mettre en place une méthode qui permette de définir des milieux pour l'étude des interactions Génotype*Milieu. Cette contrainte nous imposait donc de faire une classification des élevages sur la base de leurs profils TJC, données disponibles dans les bases de données nationales.

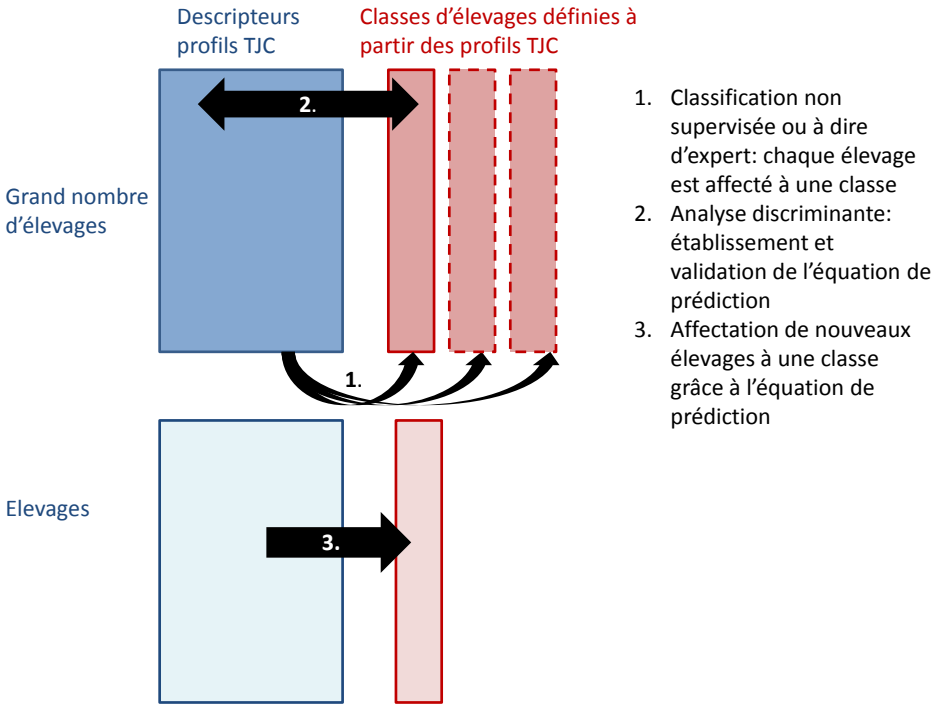


FIGURE 3.1 – Schéma explicatif de la typologie sur un grand nombre d'élevages uniquement à partir des profils TJC

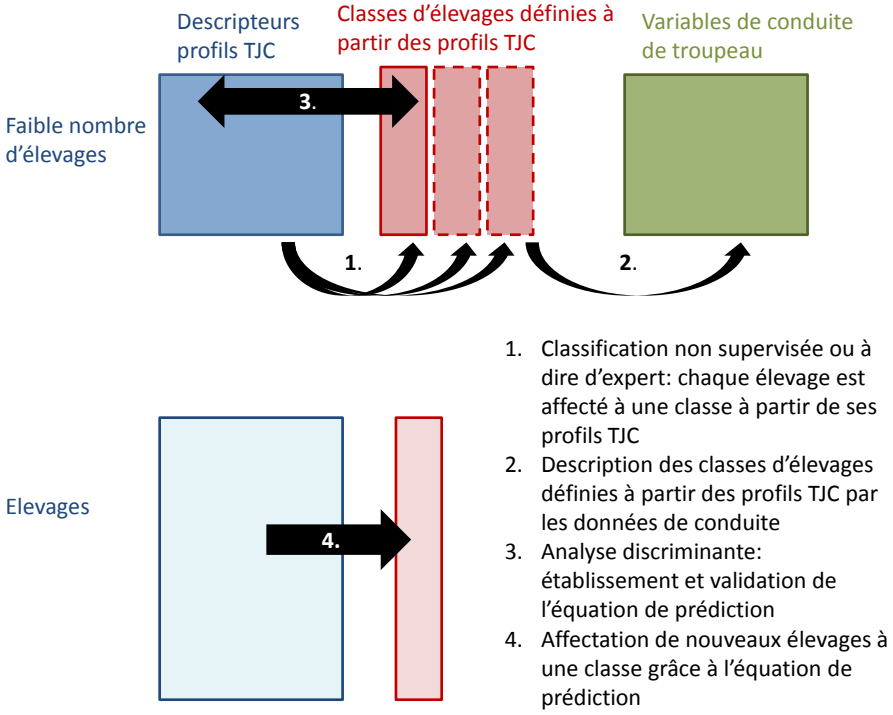


FIGURE 3.2 – Schéma explicatif de la typologie à partir des profils TJC sur un faible nombre d'élevages dont la conduite est connue

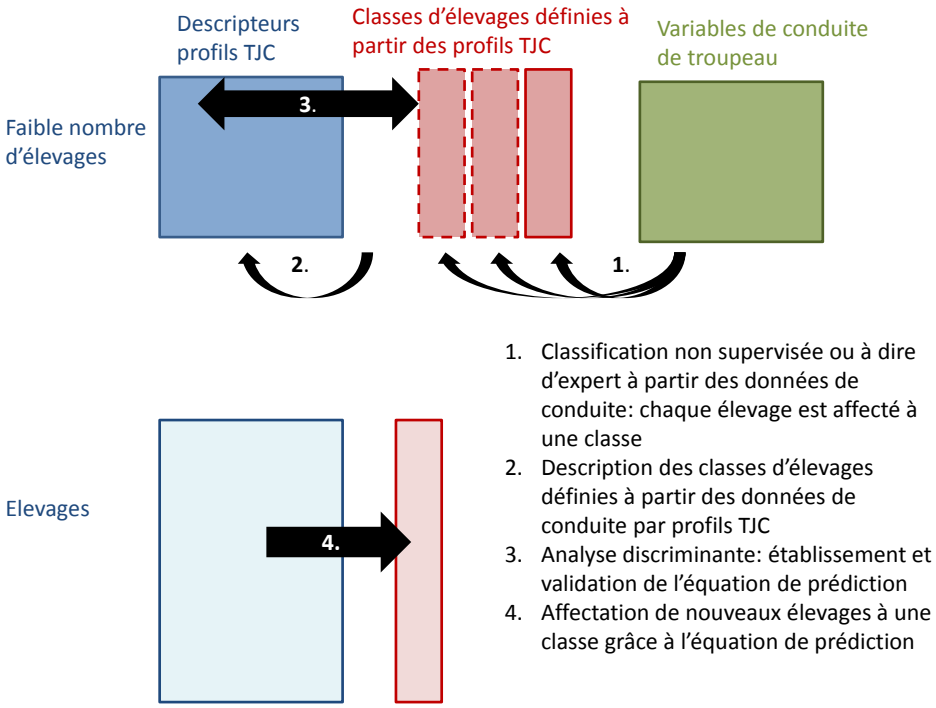


FIGURE 3.3 – Schéma explicatif de la typologie ancrée à la réalité de terrain mais utilisant les profils TJC

Chapitre 4

Étude des interactions Génotype * Milieu

Les études des interactions Génotype*Milieu sont très variées de part la définition du milieu (présentée dans le chapitre 2), la définition du génotype et le type de modélisation utilisée. Ce chapitre débute par une revue bibliographique sur les définitions possibles du génotype et les méthodes d'estimation des interactions Génotype*Milieu existantes. Il présente ensuite les résultats obtenus lors de cette thèse sur la mesure des interactions Génotype*Milieu en France sur les caractères laitiers et fonctionnels.

4.1 Introduction

Les interactions Génotype*Milieu sont le reflet de la plasticité phénotypique des êtres vivants (animaux ou végétaux) c'est-à-dire de la capacité des génotypes à répondre différemment à une modification de l'environnement (Scheiner & Lyman, 1989). Elles existent lorsque la différence entre les performances (ou valeurs phénotypiques) de deux génotypes n'est pas la même dans deux environnements différents. La notion d'environnement ou de milieu a été largement discutée dans le chapitre 2. La manière de définir le génotype est un autre paramètre de l'étude des interactions Génotype*Milieu mais il est moins problématique. Le génotype peut être défini comme une race (Steinheim *et al.*, 2008; Cutullic *et al.*, 2010), une lignée génétique (Pryce *et al.*, 1999; McCarthy *et al.*, 2007b), un groupe d'animaux aux caractéristiques communes comme leur valeur génétique (Buckley *et al.*, 2000) ; c'est le cas dans la plupart des études réalisées en élevages expérimentaux en géné-

tique animale. Dans le cadre d'études sur un plus grand nombre d'animaux, le génotype est défini au niveau de l'animal (c'est le cas dans le MACE : Multiple Across Country Evaluation (Schaeffer, 1994) réalisé par Interbull). Enfin, la définition du génotype au niveau d'un gène est surtout utilisé en génétique humaine (Padyukov *et al.*, 2004).

Pour mesurer les interactions Génotype*Milieu, les performances d'un même génotype doivent être mesurées dans différents environnements. L'utilisation de clones en végétal permet de tester les performances d'un même génotype dans différents environnements ; ce qui n'est pas possible en animal puisqu'un individu (c'est-à-dire un génotype) n'a des performances que dans un seul milieu. Cependant, la large utilisation de l'insémination animale chez les bovins laitiers permet d'associer à un taureau les performances de ses filles qui, elles, peuvent être mesurées dans différents environnements. De plus, la filière « bovin lait » profite de l'existence de grandes bases de données sur les performances des animaux actualisées régulièrement (le Système d'Information Génétique) contrairement à la filière végétale. Ce volume important de phénotypes enregistrés permet de réaliser des études sur les interactions Génotype*Milieu sur un grand nombre d'animaux, ce qui rend les résultats représentatifs de la situation réelle.

4.2 Les différentes formes d'interactions Génotype* Milieu

La présence d'interaction Génotype*Milieu signifie que la valeur génétique d'un animal dépend de son environnement. Les différentes formes d'interactions entre deux génotypes et deux environnements ont été présentées par Haldane (1946). Depuis, les méthodes permettent de prendre en compte un bien plus grand nombre de génotypes et de milieux. Strandberg (2011) reprend ces définitions et les illustrent par des graphes dans le cadre d'études où le milieu est décrit par une variable continue. La figure 4.1 p.85 en est inspirée.

Dans le cas *a* de la figure 4.1, il existe une variabilité génétique mais pas de plasticité phénotypique : les 3 génotypes n'ont pas les mêmes valeurs phénotypiques et génétiques mais ne réagissent pas à une modification de l'environnement. Dans le cas *b*, les 3 génotypes réagissent de la même façon à une modification de l'environnement, ils possèdent donc une sensibilité à l'environnement (ils sont plastiques) mais elle n'est pas variable. Les cas a et b ne permettent pas de choisir des animaux (ou génotypes) particulièrement mieux

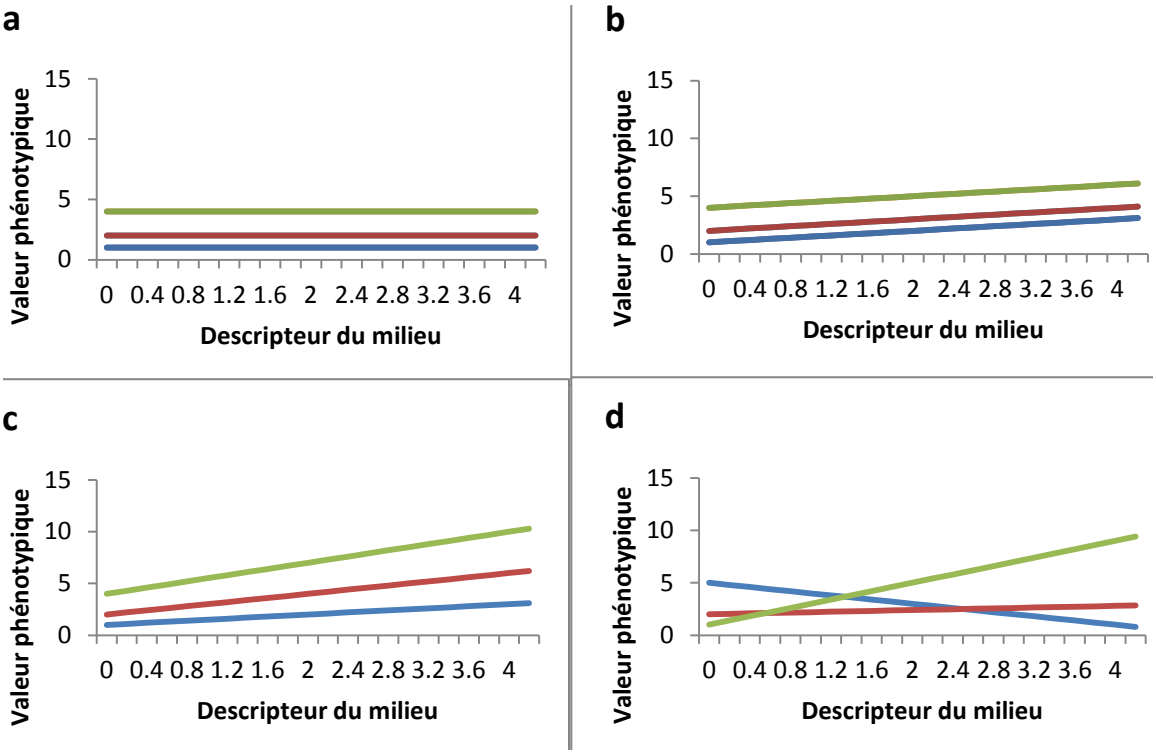


FIGURE 4.1 – Formes des relations entre 3 génotypes (représentés par 3 couleurs) et le milieu décrit par une variable continue. a) pas de plasticité phénotypique mais variation dans le niveau génétique b) plasticité mais pas de variation de plasticité selon les génotypes c) plasticité variable mais pas de classement différent des génotypes d) plasticité variable et classement différent des génotypes.

adaptés à un milieu donné. Le cas *c* présente des génotypes plastiques et leur plasticité est variable : certains répondent plus à une modification de l'environnement que d'autres. Le classement des génotypes reste le même quelle que soit la valeur du milieu. Ce phénomène est un type d'interactions génotype*milieu appelé « effet d'échelle ». Le dernier cas (*d*) correspond aux interactions génotype*milieu conduisant à un « reclassement » (reranking en anglais) : les génotypes sont plastiques avec des plasticités différentes ce qui conduit à des classements des génotypes différents selon le milieu.

Les effets d'échelle et les reclassements sont les phénomènes que l'on tend à mettre en évidence lors d'études sur les interactions génotype*milieu.

4.3 Modélisations des interactions Génotype*Milieu : méthodes et indicateurs

Les méthodes utilisées pour décrire et illustrer les interactions Génotype*Milieu ont évolué au cours du temps, principalement pour pouvoir prendre en compte des descriptions du milieu de plus en plus complexes (variable qualitative puis quantitative et enfin variable quantitative de synthèse). Les premières revues bibliographiques sur le thème (Freeman, 1973) prennent surtout des exemples en sélection végétale.

Par définition, les modèles utilisés pour l'étude des interactions Génotype*Milieu s'opposent au modèle qui les néglige (modèle additif) : $P_i = G_i + E_j$ où la performance de l'animal i (P_i) est expliquée par l'effet de son génotype (G_i) et l'effet de l'environnement j (E_j). Il existe plusieurs manières de modéliser l'interaction selon les données disponibles et la nature de la variable permettant de décrire l'environnement. Elles sont décrites ci-après.

4.3.1 Analyse de (Co)Variance avec un terme d'interaction

L'analyse de variance (ANOVA) (Pagès, 2010) est la méthode la plus ancienne pour détecter des interactions Génotype*Milieu. Elle permet d'expliquer une variable quantitative (la performance) par des variables qualitatives (le génotype, l'environnement) et leurs interactions. Le modèle s'écrit alors $P_i = G_i + E_j + G_i * E_j + \Delta_{ij}$. Le terme d'interaction G*E signifie que l'effet du génotype sur la performance dépend de la modalité prise par la variable d'environnement. Si le modèle contient aussi des variables explicatives quantitatives, le modèle est alors un mélange d'analyse de variance et de régression que l'on appelle analyse de la covariance (ANCOVA). Dans ce genre d'études, réalisées sur un faible nombre de génotypes et d'environnements, le génotype et l'environnement sont le plus souvent considérés comme deux effets fixes. Ces modèles permettent d'étudier le comportement d'un nombre fini de génotypes dans un nombre fini d'environnements. Ils s'appliquent très bien à des plans d'expérience équilibrés, c'est-à-dire où tous les génotypes sont étudiés dans tous les environnements. Ces modèles sont assez courants dans le cadre de la sélection variétale, (Denčić *et al.*, 2011) à titre d'exemple. Les exemples sont moins courants en génétique animale car il est plus difficile de tester le même génotype dans différents environnements. Cependant, il existe certains exemples comme Wicks & Leaver (2006) ou Pryce *et al.* (1999) qui définissent un génotype par un groupe d'animaux ayant

des valeurs génétiques semblables (élevées vs faibles).

En génétique animale, les matrices de parenté sont souvent utilisées pour passer outre l'impossibilité de mesure le même génotype dans différents environnements. Le terme d'interaction est alors aléatoire, il est le plus souvent sous la forme élevage*taureau (herd*sire) ; c'est le cas dans les études de Dimov *et al.* (1995) et Haskell *et al.* (2007) en bovins laitiers et de Wilson *et al.* (1972) en bovins allaitants.

Une fois l'interaction modélisée via un terme d'interaction, un test de significativité de ce terme peut être réalisé dans les ANOVA ou ANCOVA ; sinon les estimations des paramètres de la variance sont utilisés pour mesurer l'interaction entre génotype et milieu.

4.3.2 Régressions aléatoires : adaptation des modèles de norme de réaction en génétique quantitative

En biologie évolutive, la plasticité phénotypique (ou sensibilité à l'environnement) est la modification du phénotype issu d'un génotype soumis à un changement d'environnement. Quand la variable d'environnement est continue, elle est étudiée par des normes de réaction (Via *et al.*, 1995). Le cas où la variable d'environnement n'est pas continue est traité dans le prochain paragraphe.

Les normes de réaction modélisent le phénotype comme une fonction de l'environnement qui peut prendre diverses formes comme des fonctions polynomiales (simple, de Legendre...). Ce concept est repris en génétique quantitative pour estimer des valeurs génétiques en fonction d'une variable d'environnement. La régression aléatoire est utilisée.

Si une fonction polynomiale de degré n est utilisée pour la norme de réaction, le modèle en génétique quantitative peut alors s'écrire de la manière suivante :

$$y_i = \sum(effets_fixes_i) + \sum_{p=0}^{n}(a_{ip}(\lambda_p x^p)) + \epsilon \tag{4.1}$$

avec y_i= performance de l'animal i

n= le degré de la fonction utilisée pour illustrer la norme de réaction

a_{ip}= la valeur génétique de l'animal i associée au p^{eme} terme de la fonction polynomiale. L'effet génétique suit une distribution proportionnelle au numérateur de la matrice de parenté.

λ_p=le p^{eme} coefficient de la fonction polynomiale. Par exemple, pour les fonctions les plus simples : $\forall p \in [1, n]$ λ_p =1 ou pour un polynôme de Legendre de degré 2 : $\lambda = [0, -3/2, 0, 5/2]$, x=la valeur de la variable d'environnement, ϵ= la résiduelle du modèle

Les modèles de normes de réaction peuvent prendre en compte :

- des fonctions qui illustrent le lien entre environnement et phénotype qui soient plus ou moins complexes : comme des pôlynomes simples (Calus & Veerkamp, 2003), des polynômes de Legendre (Windig *et al.*, 2006), des fonctions sigmoïdes (Moreteau *et al.*, 1997) ou des fonctions créées par l'utilisateur (comme ce sera le cas dans l'article II).
- plusieurs caractères (comme le lait et les cellules) qui varient en fonction de l'environnement (Windig *et al.*, 2006)

Les termes a_0, a_1,..., a_p sont parfois considérés comme des caractères à part entière : des caractères de plasticité phénotypique qui peuvent potentiellement être sélectionnés. Les indicateurs de présence d'interaction Génotype*Milieu sont calculés à partir des paramètres de variance de ces termes. Si des animaux adaptés à tous les types d'environnement sont recherchés, les animaux dont la norme de réaction a une faible pente seront privilégiés.

D'autres indicateurs de présence d'interactions Génotype*Milieu peuvent être calculés comme la corrélation génétique entre 2 milieux ou l'héritabilité dans chaque milieu ; le milieu étant défini par des valeurs spécifiques de la variable d'environnement « x ». Pour l'exemple, les 2 milieux sont notés ici « k » et « l » respectivement pour x=k et x=l.

Soit G de dimension n*n, la matrice de (co)variance pour l'effet aléatoire « animal ».

La variance génétique dans le milieu « k » s'écrit :

$$m'_k G m_k \tag{4.2}$$

avec m_k un vecteur colonne correspondant au milieu « k ». Il contient les n éléments ($\lambda_p * k^p$, toujours dans le cas d'une norme de réaction de type polynomial) avec $p \in [1, n]$, n le degré de la fonction utilisée pour la norme de réaction et k la valeur prise par la variable d'environnement x dans le milieu « k ».

La corrélation génétique entre les milieux « k » et « l » s'écrit :

$$\frac{m'_k G m_l}{\sqrt{m'_k G m_k} * \sqrt{m'_l G m_l}} \tag{4.3}$$

L'héritabilité de l'environnement k s'écrit :

$$h^2 = \frac{m'_k G m_k}{m'_k G m_k + \sigma^2_{\epsilon,k}} \tag{4.4}$$

avec $\sigma^2_{\epsilon,k}$ la variance résiduelle dans le milieu k.

Si la corrélation génétique entre 2 milieux est faible, cela signifie qu'il existe un reclassement des animaux selon leur valeur génétique (reranking). Si la corrélation est proche de 1, le classement des animaux est le même dans les 2 milieux. Cependant, il peut exister un effet d'échelle (scale effect) c'est-à-dire que la variabilité des valeurs génétiques dans un milieu soit plus importante que dans un autre (proportionnellement à la variance residuelle), l'héritabilité du caractère dépend du milieu. Dans le cadre des études des interactions Génotype*Milieu qui utilisent les normes de réaction : l'héritabilité et les corrélations génétiques sont le plus souvent calculées pour les milieux extrêmes. Un exemple est donné dans (Calus & Veerkamp, 2003) : le milieu est défini par une variable d'environnement continue, l'effet génétique est modélisé par une régression aléatoire sur cette variable d'environnement ; enfin, la corrélation génétique entre les milieux extrêmes c'est-à-dire les milieux correspondants au quartile inférieur et au quartile supérieur de la variable d'environnement est calculée.

Ces indicateurs de présence des interactions Génotype*Milieu (héritabilité et corrélations génétiques) sont aussi utilisés dans le modèle multicaractères.

4.3.3 Modèle multicaractères : adaptation du modèle « character state » en génétique quantitative

Le modèle « character state » est, en biologie évolutive, la norme de réaction dans laquelle la réponse phénotypique d'un génotype n'est étudiée que pour certaines valeurs ou gammes de valeurs de la variable d'environnement. Le milieu est donc une variable quantitative discrète qui est une approximation de la variable continue qui décrit l'environnement. Les différents « character states » c'est-à-dire les valeurs du phénotype pour

différentes valeurs de l'environnement sont considérées comme différents caractères.

En génétique quantitative, ce modèle se traduit par un modèle dit multicaractères. A l'origine, ce modèle consistait à évaluer simultanément des animaux pour différents caractères (lait et cellules par exemple) pour prendre en compte la corrélation qui pouvait exister entre ces caractères. L'analogie avec le modèle « character state » vient de cette particularité (prise en compte de plusieurs carcatères).

Cependant, le modèle multicaractères a ensuite été un outil utilisé dans d'autres circonstance. Le MACE (Multiple Across Country Evaluation) (Schaeffer, 1994) dans les évaluations internationales utilise un modèle multicaractères pour pouvoir comparer les palmarès des animaux dans différents pays. L'expression phénotypique d'un caractère dans deux pays (milieux) différents par un même génotype est traitée dans le modèle comme deux caractères différents mais corrélés, bien que ce soit le même phénotype.

On remarque que dans cette définition, le milieu n'est plus du tout décrit par une variable quantitative discrète (comme dans le modèle « character state ») mais par une variable qualitative nominale (les modalités de la variable « pays » ne sont pas ordonnées). Cela ouvre alors la possibilité d'utiliser le modèle multicaractères pour l'étude d'interactions Génotype*Milieu où le milieu est décrit par une variable qualitative ordonnée (comme trois niveaux de production (De Veer & Van Vleck, 1987)) ou une variable qualitative nominale (petites exploitations vs grandes exploitations (König *et al.*, 2005), utilisation d'un robot de traite vs d'une salle de traite (Mulder *et al.*, 2004), pâturage vs confinement (Kearney *et al.*, 2004a,b), différents pays (Ojango & Pollott, 2002; Costa *et al.*, 2000), différents groupes d'élevages (Zwald *et al.*, 2003)).

Le modèle multicaractère s'écrit :

$$y_{im} = \sum(effets_fixes_i) + a_{im} + \epsilon \tag{4.5}$$

avec y_{im} la performance de l'animal i dans le milieu m , a_{im} la valeur génétique de l'animal i dans le milieu m (effet aléatoire) et ϵ la résiduelle du modèle. L'effet génétique suit une distribution proportionnelle au numérateur de la matrice de parenté. La matrice de variance covariance de l'effet aléatoire animal a pour dimension n*n avec n, le nombre de milieux étudiés. Elle contient sur la diagonale les variances génétiques dans chaque milieu σ^2_m, les termes non diagonaux sont les covariances génétiques entre deux milieux (m et m') $\sigma_{mm'}$. L'interaction Génotype*Milieu est alors évaluée en comparant les valeurs génétiques

des animaux estimés dans les différents milieux. Comme dans le paragraphe précédent, la corrélation génétique entre chaque couple de milieux m et m' ($\sigma_{mm'}/\sqrt{\sigma^2_m * \sigma^2_{m'}}$) et les héritabilités de chaque milieu ($\sigma^2_{m'}/(\sigma^2_{m'} + \sigma^2_\epsilon)$) sont estimés pour mettre en avant un reclassement des animaux selon le milieu ou un effet d'échelle (héritabilité hétérogène selon les milieux).

4.4 Les études déjà menées en France

Bonaiti (1982) a réalisé, au début des années 80, une première étude sur les interactions Génotype*Milieu en France. Il utilisait une approche multicaractères avec un modèle père grand-père. Le milieu était défini par 3 classes correspondant à trois niveaux d'étable ; ces derniers étant définis comme « *la valeur de l'effet sur la production laitière du milieu propre à une étable* ». Cette étude avait mis en évidence une héritabilité de la quantité moyenne de matière utile croissante avec le niveau de production (de 0.31 à 0.42) et des corrélations génétiques entre milieux allant de 0.86 à 0.98. La variation d'héritabilité était due à une augmentation plus importante de la variance génétique que de la variance résiduelle avec le niveau de production. Aujourd'hui, les avancées méthodologiques permettent de mettre à jour cette étude.

Des études s'apparentant aux études Génotype*Milieu, bien qu'elles n'en portent pas le nom, ont également été menées. Il s'agit de toutes les études centrées sur la prise en compte des variances hétérogènes. Elles ne portent pas toutes ce nom car l'hétérogénéité peut être étudiée suivant différents facteurs décrivant ou non le milieu au sens de « système d'élevage ». Il peut s'agir du troupeau, de la région, de l'année pour les caractères laitiers des bovins (Robert-Granié *et al.*, 1999), du pointeur, du type de logement, du type d'alimentation ou du stade de lactation pour les caractères de morphologie des bovins (Robert-Granié *et al.*, 1997). La méthodologie permettant d'étudier cette hétéroscédasticité a notamment été développée dans la thèse de Robert (1996). Elle consiste à simplifier un modèle dit « saturé » qui correspond au modèle multicaractères pour tester l'homogénéité des corrélations génétiques entre milieux ou à la fois l'homogénéité des corrélations génétiques et des héritabilités entre milieux. Ces études ont conduit à la prise en compte, en France, des hétérogénéités de variance dans les modèles d'évaluation génétique en bovins laitiers pour les caractères laitiers (IBL, 1999) et de morphologie (IBL, 2005).

Ces études ne sont pas pour autant spécifiques aux bovins laitiers ou au cas français. Par exemple, David *et al.* (2011) ont récemment étudié en ovins allaitants les variances hétérogènes en fonction du nombre d'agneaux par portée. Les hétérogénéités de variance ont aussi été étudiées en bovins laitiers en Allemagne (Dodenhoff & Swalve, 1998) en fonction de différentes régions, en Espagne (Ibáñez *et al.*, 1999) en fonction du niveau de production, de la région, de la taille du troupeau et de l'année, en Uruguay (Urioste *et al.*, 2001) en fonction du mois de vêlage, la durée de lactation, du niveau de production...

4.5 Les études menées dans la thèse

L'objectif dans cette partie de la thèse est de mesurer d'éventuelles interactions Génotype*Milieu en France sur les caractères de production laitière et sur des caractères fonctionnels pour les trois grandes races bovines laitières de France : Normande, Montbéliarde, Holstein. Elle entre dans la lignée des études sur les variances hétérogènes présentées précédemment avec la particularité de se focaliser sur l'hétéroscédasticité qui a pour origine le milieu ; d'où l'appellation spécifique d'étude des interactions Génotype*Milieu. Une première étude a été réalisée sur les 3 races sur les caractères laitiers. Elle utilise les 2 types de modèles permettant l'étude des interactions Génotype*Milieu : le modèle multicaractères et les normes de réaction et a fait l'objet d'un article scientifique en cours de révision (article II). Compte tenu de l'expérience acquise lors des analyses de la première étude, une deuxième étude a été menée à bien sur les caractères laitiers et fonctionnels. La méthode retenue repose sur une variante du modèle de norme de réaction proposé dans l'article II et n'a été appliquée que sur la race Montbéliarde.

4.6 Première étude sur les interactions Génotype*Milieu en France sur les caractères laitiers

4.6.1 Méthodologie générale

Le milieu est défini à partir des descripteurs des profils Troupeau Jour de Contrôle présentés dans l'article I (chapitre 2). Cette définition avait l'avantage de pouvoir être utilisée sur l'ensemble des élevages adhérents à France Conseil ELevage. Dans l'étude

sur les caractères laitiers, les caractères suivants ont été étudiés : la production laitière, les matières grasse et protéique, les taux butyreux et protéique. Ils sont tous mesurés sur 305 jours. Deux types de modélisation des interactions ont été testés : un modèle multicaractères et un modèle de régression aléatoire (normes de réaction). Les différentes étapes de l'approche méthodologique sont les suivantes :

1. Une description de la variabilité du milieu Les premières composantes principales d'une analyse factorielles obtenues à partir des descripteurs des profils Troupeau Jour de Contrôle (TJC) d'environ 15.000 élevages normands, montbéliards et holsteins ont été utilisées (comme dans l'article I, illustrée par l'étape 1 de la figure 3.1 page 81).

   L'estimation des interactions Génotype*Milieu a ensuite été réalisée par race et sur deux jeux de données distincts appelés respectivement « parangon » et « diversité ». Dans le premier, seuls les élevages considérés comme les plus caractéristiques des milieux retenus ont été pris en compte, les élevages aux profils TJC atypiques n'ont donc pas été intégrés. Au contraire, le jeu de données « diversité » contenait des élevages représentant l'ensemble de la diversité des profils TJC (c'est-à-dire des milieux).

2. L'utilisation du modèle multicaractères pour estimer les interactions Génotype*Milieu
   - La description du milieu par des groupes d'élevages définis par une classification non supervisée à partir de leurs coordonnées sur les premières composantes principales (de l'étape 1)
   - La mise en place d'un modèle multicaractères pour estimer les paramètres des composantes de la variance
   - L'estimation des corrélations génétiques entre milieux et des héritabilités dans chaque milieu comme présenté dans le paragraphe 4.3.3
3. L'utilisation du modèle de norme de réaction pour estimer les interactions Génotype*Milieu
   - La description du milieu à travers les coordonnées des élevages sur les premières composantes principales (de l'étape 1)
   - La mise en place d'un modèle de norme de réaction pour estimer une valeur

génétique indépendante du milieu et des valeurs génétiques dépendant du milieu (une par composante principale prise en compte dans le modèle pour décrire le milieu) par animal. Des modèles prenant en compte 1, 2 ou 3 composantes principales pour décrire le milieu ont été testés.

– L'estimation des héritabilités pour différentes valeurs de la variable d'environnement et des corrélations génétiques entre ces environnements comme présenté dans le paragraphe 4.3.2

L'article II ne présente les résultats que pour les races Holstein et Normande, les résultats pour la Montbéliarde sont ajoutés à la suite.

4.6.2 Mise en évidence d'un effet d'échelle pour les caractères laitiers en races Normande et Holstein : Article II

Huquet B., Leclerc H. and Ducrocq V., 2012 Modelling and estimation of genotype by environment interactions for production traits in French dairy cattle. *Genet Sel Evol*, (accepté le 7 novembre 2012).

Modelling and estimation of genotype by environment interactions for production traits in French dairy cattle

Bérénice Huquet [*1] [2] , Hélène Leclerc[2] and Vincent Ducrocq[1]

[1]INRA, UMR1313 Génétique Animale et Biologie Intégrative, F-78302 Jouy-en-Josas, France
[2]Institut de l'Elevage, 149 rue de Bercy, 75012 Paris, France

Email: Bérénice Huquet *- berenice.huquet@jouy.inra.fr; Hélène Leclerc- helene.leclerc@idele.fr; Vincent Ducrocq - vincent.ducrocq@jouy.inra.fr;

*Corresponding author

Abstract

Background

Genotype by environment interactions are currently ignored in national genetic evaluations of dairy cattle. However, this is often questioned, especially when environment or herd management is wide-ranging. The aim of this study was to assess genotype by environment interactions for production traits (milk, protein, fat yields and fat and protein contents) in French dairy cattle using an original approach to characterize the environments.

Methods

Genetic parameters of production traits were estimated for three breeds (Holstein, Normande and Montbéliarde) using multiple-trait and reaction norm models. Variables derived from Herd Test Day profiles obtained after a test day model evaluation were used to define herd environment.

Results

Multiple-trait and reaction norm models gave similar results. Genetic correlations were very close to unity for all traits, except between some extreme environments. However, a relatively wide range of heritabilities by trait and breed was found across environments. This was more the case for milk, protein and fat yields than for protein and fat contents.

Conclusions

No real reranking of animals was observed across environments. However, a significant scale effect exists: the more intensive the herd management for milk yield, the larger the heritability.

Background

Two main opportunities are available to improve production traits in dairy cattle: through the modification of herd management and/or the genetic level. Except when it is necessary to choose a local breed for a specific environment (such as the Abondance breed in the French Alps), these two opportunities are generally considered separately, as in genetic evaluation. Indeed, they imply the absence of genotype by environment (G*E) interactions, i.e., the breeding value of an animal is assumed to be the same regardless of the environment in which it will be raised. Dealing with this situation, some breeders question the efficiency of current breeding schemes for their own particular management system. Thus, the objective of this study was then to estimate genotype by environment interactions for production traits (milk, protein, fat yields and fat and protein contents) in French dairy cattle. The overall objective was to assess whether these interactions could be an opportunity to better adapt animals to their environment. Genotype by environment interaction studies raise three main questions: How to define the genotype? How to describe the environment? Which model to choose in order to estimate genotype by environment interactions? This study used an innovative description of herd environment: Herd Test Day profiles, which are by-products of a test day model evaluation. Two models, a multiple-trait and a reaction norm model were tested.

Methods

The approach consisted of two steps. The first step dealt with the definition of herd environment through Herd Test Day profiles. This was done across breeds (Holstein, Normande and Montbéliarde) rather than within breed because two herds with different breeds could share the same type of environment. The second step was a genotype by environment interaction analysis. As genetic evaluations are within breed, genotype by environment (G*E) parameters were estimated within breed.

Description of the environment: Herd Test Day profiles

The methodology used to describe herd environment from Herd Test Day profiles was described in [1]. The main difference with this previous study is that we worked here with a larger dataset. A short description

of the main steps involved and results obtained follows.

Herd environments were described through Herd Test Day (HTD) profiles for milk yield, fat and protein contents between 2005 and 2010. Herd Test Day profiles represent the evolution of HTD effects over time, as HTD effects are obtained from a test day model evaluation which aims at predicting the breeding value of animals at any day of the lactation period. The test day model uses each test day record available in national databases, in contrast to the 305-day lactation model which relies on the performance of an animal cumulated over 305 days. In order to improve the accuracy of daily breeding value estimation, other factors affecting the performance such as age and month of calving, length of dry period and gestation are estimated over the whole lactation through splines. Similarly, genetic and permanent environment effects throughout the lactation are predicted using continuous functions and the detailed description of the French test day model is given in [2]. The HTD effect is independent from all other effects and it estimates the effect of all features common to all cows of the herd on a particular test-day, i.e., essentially the effect of herd management (feeding, housing) of the test day. Therefore, the HTD effect can be interpreted as the herd management level of a herd on a given test-day. The HTD profile is a continuous function showing changes in HTD effects over time and can be interpreted as the changes in the herd management level over time. In previous studies, genetic evaluation using a test day model carried out for milk yield and for fat and protein contents on French national data bases, separately for Holstein, Normande and Montbéliarde, the three major dairy breeds. This made it possible to describe herds by their three HTD profiles (milk yield, protein and fat contents) from 2005 to 2010 (see dashed curves in Figure 1).

HTD profiles, reflecting changes in HTD effects over time, can be decomposed into a systematic within year change that will be assumed to reveal practices related to the global herd management during the year as in [3], and a deviation from this global component due to specific characteristics (unusual weather conditions, feedstuffs availability, etc.) that cannot be related to regular management activities. Therefore, HTD profiles had to be corrected for these occasional features in order to be used as the definition of the environment in a genotype by environment interaction study. For this purpose, HTD profiles were smoothed to focus on their repeated annual features using a model inspired by the model of Koivula et al [3] and described in [1]. Basically, the method consisted of describing HTD profiles by a continuous function involving a linear trend and three sine curves. Examples of HTD profiles before and after smoothing are shown in Figure 1. Note that in the rest of the study, only herds for which smoothing was obtained with a minimum coefficient of determination were retained (see [1] for details).

Each HTD profile was then summarized by seven descriptors, as shown in Figure 2, leading to 21 de-

scriptors (7 descriptors times 3 traits) for each herd. These descriptors were reduced with Multiple Factor Analysis (MFA) to 10 Principal Components (PC) by retaining all PC that explained more than 4% of the total variance. MFA is similar to principal components analysis, which enables the joint use of categorical and quantitative data [4]. The MFA was performed on data from 12 061 Holstein, 2 591 Normande and 1 104 Montbéliarde herds. The descriptors were centered within breed in order to correct for breed effects.

The first principal component (PC1, explaining 15% of the total variance) was interpreted as a measure of the specialisation of the herd management; it discriminated herds with herd management favouring high milk production (low PC1 score) from the herds favouring high fat content (high PC1 score). The second PC (13%) was interpreted as a measure of the intensity of production related to herd management; it discriminated herds with high HTD effects for milk yield, and for fat and protein contents (high PC2 score) from herds with low HTD effects for milk yield, fat and protein contents (low PC2 score). Principal component 3 (8%) was interpreted as related to the seasonality of herd management. It differentiated herds for which the range of HTD profiles for the three traits was small (high PC3 score) from those with large ranges (low PC3 score), that is, PC3 discriminated herds in which herd management led to more or less similar milk yield and composition between seasons from herds in which herd management led to variable milk yield and composition across seasons. The next 7 PC explained 46% of the total variance but their interpretation was less intuitive. In total, these 10 first PC explained 76% in HTD effects of the total variance and were mainly related to the mean level of HTD effects and the range of HTD profiles, rather than to periods during which minimum and maximum HTD effect were obtained.

Based on the MFA, herd environment was characterized by the 10 PC scores for 15 756 Holstein, Normande and Montbéliarde herds. These PC scores and their interpretations were the basis of the different herd environment definitions that were used to describe genotype by environment interactions.

Estimation of genotype by environment interactions

G*E interactions were estimated for the three breeds based on data from two different sets of herds within breed ("paragon" or "diversity", see below) with two different models (multiple-trait and reaction norm models). For clarity, only analyses and results for the Normande and Holstein breeds, which are respectively a national dual purpose and an international dairy breed, are presented here. The following paragraphs describe the methodology used.

Herd selection for G*E estimation

Studies on the estimation of G*E interactions require the estimation of genetic parameters, which was carried out within breed (Holstein, Normande). A substantial but not excessively large data set is required for this purpose in order to obtain accurate estimates but limit computation time. Among the 12 061 Holstein and 2 591 Normande available herds, herds used in the G*E interaction study were selected by two strategies. Both were based on the PC scores of the herds. The first strategy consisted in selecting only typical herds for which characteristics were representative of a majority of the French herds, leading to what will be referred hereafter as the "paragon herd data set" (a paragon is regarded as a perfect example of a particular feature). The second strategy aimed at representing the complete diversity of French herds, leading to the "diversity herd data set".

For the first strategy, three herd groups were created using a clustering method based on the first 10 factors of the MFA. Note that this was not a classification approach. See [1] for more methodological details. At this stage, each cluster included herds with each of the three breeds. However, since the G*E analysis was carried out within breed, herd clusters were then defined per breed. Practically, Normande and Holstein herds of cluster 1 in the joint breed dataset were respectively assigned to cluster 1 in the Normande and Holstein datasets, and the same for clusters 2 and 3. These herd clusters were also used as classes for the definition of fixed effects in the reaction norm model. The 400 Normande and 750 Holstein herds that were most representative herds of each cluster, i.e. those nearest to the center of each cluster, were selected as paragons and included in the "paragon dataset". They are represented in pink, orange and red in Figures 3 and 4 for the Holstein and Normande breeds, respectively. This strategy was primarily used to obtain clearly distinct and well-defined clusters for the G*E interaction estimation by a multiple-trait model. Indeed, defining clusters of herds along a continuum and then choosing the paragons in each cluster forced level of homogeneity within and heterogeneity between clusters. This increased the power to detect possible genotype by environment interactions with the multiple-trait model. The "paragon dataset" was also used with the reaction norm models in order to compare both models on the same dataset.

For the second strategy, the first factorial map (PC 1 and 2 of the MFA) was divided into 25 equal-sized cells. Up to 40 or 60 herds (for the Holstein and Normande breeds, respectively) were selected at random for each cell. This led to two "diversity data sets" containing respectively 539 Holstein and 472 Normande herds (in blue in Figures 3 and 4, respectively).

Data selection for estimation of G*E interactions

The traits analysed were 305-day milk, fat and protein yield, as well as 305-day fat and protein contents. Data consisted of first lactations from cows which had a first calving between 2000 and 2011. Records were edited on several criteria: performances deviating by more than three phenotypic standard deviations from the breed average, herds with less than 100 first lactations from 2000 to 2011, records from daughters of bulls with less than three daughters that were located in just one herd were deleted, as well as records from cows which had an age at first calving lower than 23 months or greater than 40 months, or a lactation length lower than 180 days. These steps reduced the size of the datasets. Tables 1, 2 and 3 show the final number of herds and records used to estimate G*E interactions for the different models, breeds and herd data sets.

Models

Two models to estimate G*E interactions were tested: a multiple-trait model, in which the environment was considered to be specific to a group of herds, and a reaction norm model, in which the environment varied continuously as a function of PC scores of the herds. Both were animal models from the perspective of fitting breeding values. The pedigree files contained three generations (see Tables 1, 2 for the multiple-trait model and Table 3 for the reaction norm model). The only difference between the two models was the modelling of breeding values and additive genetic variance; fixed effects were the same for the two models. All analyses were carried out using the WOMBAT software [5], separately for each trait and breed. A more detailed description of the two models follows:

- Multiple-trait model

The multiple-trait across country evaluation (MACE) [6] can be seen as a G*E analysis for which the environment is country-specific. A given trait (e.g. milk production) recorded in different countries is modelled as distinct traits. Consequently, it is named a multiple-trait model, although a single phenotype is analysed. Using this model, genetic correlations can then be estimated between countries/environments. The model used in this paper mimics the model proposed in [7], in which the environment is defined by herd clusters. It was applied to the paragon datasets of each breed. Environments were defined based on the three herd clusters identified in the first strategy of herd selection (basis of the "paragons herd data set"). These are shown in pink, orange and red in Figures 3 and 4 for the Holstein and Normande breeds, respectively. Tables 1 and 2 summarize the average performances for these three herd clusters in the paragon dataset for the Hosltein and the Normande breed respectively.

The multiple-trait model was:

$$Y_{ih} = HerdYear_i + AgeYearCluster_i$$

$$+MonthYearCluster_i + a_{ic} + e_c.$$

where: Y_{ih} = 305-days first lactation performance of animal i in herd h, HerdYear is the fixed effect of Herd-Year class, AgeYearCluster is the fixed effect of Age at calving-Year-herdCluster class, MonthYearCluster is a fixed effect of Month at calving-Year-herdCluster subclass, a_{ic} is the random genetic effect of cow i in the herd cluster c, and e_c is the random residual effect. Note that genetic and residual variances were different for each herd cluster. This model showed problems of convergence whatever the Restricted Maximum Likelihood (REML) estimation algorithm used (combinations of options Average Information (AI), Parameter Expanded (PX) and Expectation Maximisation (EM) of the WOMBAT software [5]). Some eigenvalues of the genetic variance/covariance matrix were close to zero, because all genetic correlations estimated between herd clusters were very close to one. Consequently, the following single-trait model was implemented instead:

$$Y_{ih} = HerdYear_i + AgeYearCluster_i$$

$$+MonthYearCluster_i + a_i + e_{cy}.$$

The trait is assumed to be the same genetic trait in each cluster but allowing for different genetic variances by cluster. The same fixed effects were estimated as in the previous model were used. Heterogeneous residual variance (e_{cy}) by the combination of herd cluster and birth year period was allowed for rather than only by herd cluster. Birth year periods were based on years 2000 to 2002, 2003 to 2005, 2006 to 2008 and 2009 to 2011. Hence, four residual variances were obtained for each herd cluster; one for each group of three birth years. The mean of these four variances was used as the overall residual variance of the herd cluster. Finally, heritabilities for each herd cluster were calculated as the ratio of the genetic variance to the sum of the genetic and residual variances for that herd cluster.

• Reaction norm model

In reaction norm models, the additive genetic effect is described as a continuous function of environmental parameters. In this study, PC scores of herds were used as environmental parameters. The reaction norm model was applied both to the "paragon" and "diversity" herd datasets. Table 3 summarizes average performances and environment descriptors (PC scores) for both sets of herds and breeds.

The model was :

$$Y_{ih} = HerdYear_i + AgeYearCluster_i$$

$$+MonthYearCluster_i + a_i + e.$$

with $a_i = a_{0i} + \sum_{j=1}^{p}(a_{ji} * PC_{jh})$

The three herd clusters defined in the first herd selection strategy were included as fixed effects subclasses, whereas the breeding value of a cow was estimated as a function of the PC scores of the cow's own herd, rather than the PC scores of the herd cluster. This study focused on random genetic effects across environments for the estimation of G*E interactions and not on the fixed effects. Using exactly the same fixed effects allowed comparison of estimates of G*E interactions between the single-trait and reaction norm models. Heterogeneous residual variances by group of birth year period (2000 to 2002, 2003 to 2005, 2006 to 2008 and 2009 to 2011) were estimated. Animal breeding values a_i were modelled in two parts: one (a_{0i}) expressing the animal's average effect across environments and one ($\sum_{j=1}^{p}(a_{ji} * PC_{jh})$) that depended on the PC scores of the animal's herd, i.e., on the environment. PC_{jh} is the PC score of herd h on the j^{th} axis of the factor analysis. a_{ji} is the coefficient applied for animal i on the j^{th} PC score of its herd. Note that linear reaction norms were assumed. Analyses were carried out taking into account the first one, two or three PC scores (p=1, 2 or 3). Within breed and trait, these three models were compared using the Bayesian Information Criterion (BIC). Only results from the best model (the one with the smallest BIC value) will be presented.

The estimated covariance matrix (V) of the reaction norm model combined variances and covariances of the random genetics effects a_0, a_1,..., a_p of, for example, milk yield:

$$V = \begin{bmatrix} \sigma^2_{a_0(milk)} & & & \\ \sigma_{a_0(milk),a_1(milk)} & \sigma^2_{a_1(milk)} & & \\ \vdots & & \ddots & \\ \sigma_{a_0(milk),a_p(milk)} & \sigma_{a_1(milk),a_p(milk)} & \cdots & \sigma^2_{a_p(milk)} \end{bmatrix}.$$

Genetic variances and covariances for each herd cluster were obtained using the expression MVM' with :

$$M = \begin{bmatrix} 1 & PC1score_1 & PC2score_1 & \ldots & PCpscore_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & PC1score_n & PC2score_n & \ldots & PCpscore_n \end{bmatrix}.$$

A row of matrix M represented one "state" of the environment gradient, i.e. specific values of the environment, described by p PC scores, with p depending on the model. The number of environment states

analysed were arbitrarily chosen to be respectively 25 and 625 for models with one or two PCs, respectively, to describe the environment. Four residual variances were estimated per breed and trait; one for each of four birth year period. The mean of these four residual variances was used as the reference residual variance. Finally, heritabilities for each state of the environment were calculated as the ratio of the genetic variance over the genetic plus residual variance for that state of the environment. Genetic correlations between two environment states were calculated as the ratio of the genetic covariance between these two environments to the product of their genetic standard deviations.

Results

Multiple and single-trait analysis with the paragon dataset

Description of the three herd clusters/environments

With the multiple-trait model, environment was described by three herd clusters which represented three types of herd management. They were built based on the first 10 PC scores, summarizing the features of their three HTD profiles. Therefore, herd clusters were built based on production level only due to herd management rather than the global production level (that includes herd management but also genetic effect for example). Thus in the following, the production level must be interpreted as the level of milk yield and protein and fat contents due to herd management only. The interpretation of herd clusters was very similar for both breeds. Tables 1 and 2 show the number of herds in each cluster and means of PC scores in each herd cluster for each breed. Cluster 1 was made up of herds with management that resulted in an average production intensity (mean PC2 score close to zero) but that was more geared towards milk production than towards protein and fat contents (negative PC1 score). The management of the herds in this cluster was rather insensitive to season of production for the Holstein breed (mean PC3 score is positive) but not for the Normande breed (mean PC3 score is negative). Cluster 2 consisted of herds with a higher herd management intensity (high PC2 score) and rather specialised in protein and fat contents (positive PC1 score). Management of herds in this cluster was affected by the season of production for the Normande breed but not for the Holstein breed. Cluster 3 was composed of herds with a low herd management intensity (negative PC2 score).

Genetic parameters in the three environments

In the multiple-trait model, which had problems of convergence, the smallest genetic correlation was 0.98 for protein content in the Normande breed between cluster 1 and cluster 3. This indicated that there was no reranking of animals across herd clusters, i.e., environments. The single-trait model assumed that genetic

correlations were one between environments but allowed for heterogeneous genetic and residual variances, leading to different heritabilities across herd clusters (Table 4).

Except for fat yield in the Normande breed and fat content in the Holstein breed, heritabilities decreased from cluster 1 to cluster 3, for all traits and breeds: heritabilities were greater for herds with a high herd management intensity that favoured milk production. The largest ranges of heritabilities were found in the Holstein breed (Table 4): with a decrease in heritability from herd cluster 1 to herd cluster 3 by 15% for milk yield, 15% for protein yield, 22% for fat yield. In most cases, these decreases in heritability were due to a greater decrease of genetic variance than of residual variance.

In conclusion, the multiple-trait model did not reveal evidence of significant reranking of animals between environments. However, although herd clusters in the paragon dataset did not reflect extreme herd managements, heritabilities were found to differ between environments.

Reaction norm model

Reaction norm models were tested using one, two or three PCs as environmental parameter(s) within breed, trait and dataset (paragon or diversity herd sets). According to the Bayesian Information Criterion, the best model used only PC1 as environmental parameter (i.e., herd management specialisation: milk yield versus protein and fat contents) for protein and fat contents, whereas the best models for milk, protein and fat yields used both PC1 and PC2 (herd management specialisation and herd management intensity) as environmental parameters. This was the case for all breeds and both datasets.

The first eigenvalue of the covariance matrix (V) that combined variances and covariances of the random genetics effects a_0, a_1,..., a_p (where p=1,2,3 is the number of PCs included as environmental parameters) was very high whatever the breed, trait and dataset. This eigenvalue corresponded mainly to the random genetic effect a_0, which represents the part of the breeding value that does not depend on the environment. This eigenvalue represented a minimum of 99% of the sum of eigenvalues of the covariance matrix (V). This was a further argument supporting the quasi absence of G*E interactions for production traits, in terms of reranking.

Estimates of residual variances are shown in Table 5 and genetic correlations and heritabilities in Tables 6 and 7. Note that for these results, only environment states corresponding to herds in the dataset were taken into account. Indeed, Figures 3 and 4 show that some areas defined by PC1 and PC2 included no herds (e.g. in the top left corner of the figures). On average, genetic correlations between environments were very high on average for all breeds, traits, and datasets (see Table 6), supporting again the absence of

reranking of animals across environments. Genetic correlations were higher between environments defined for the paragon dataset than between environments defined for the diversity dataset. This was as expected because herds included in the paragon dataset were chosen to reflect herd managements that were common in France. In contrast, the diversity dataset also included herds that represented extreme environments. In the diversity dataset, the average genetic correlations were lower for the Holstein than for the Normande breed. This is due to the fact that the Normande herds available in the study reflected herd managements less extreme than the Holstein herds: the range of their PC scores was narrower than the one for Holstein herds (see Figures 3 and 4). The lowest genetic correlations were obtained with the diversity dataset for milk, fat and protein yields: between 0.60 and 0.68 for the Holstein breed and between 0.86 and 0.92 for the Normande breed. These correlations were obtained between extreme environments.

Although no reranking was shown, heterogeneity of heritabilities was again found for milk, fat and protein yields for both breeds (Table 7), demonstrating a scale effect. Since the residual variance was the same across environments within breed, trait and dataset, this was due to a heterogeneity of genetic variances across environments. In contrast, heritabilities for protein and fat contents were more homogeneous across environments. Similar to what was observed for genetic correlations, the range of heritabilities was higher in the diversity dataset than in the paragon dataset and was even higher for the Holstein breed than for the Normande breed.

For the yield traits, environment was described simultaneously by two environmental parameters (PC1 and PC2 scores). The shape of heritabilities across environments was the same for all breeds and datasets for milk, protein and fat yields. Figures 5 and 6 show estimates of heritability for milk and fat yield, respectively, for the Holstein breed as functions of the PC1 and PC2 herd scores using the "diversity" dataset. Heritabilities increased with increasing PC1 and PC2 herd scores. This gradient was more important for herd management intensity (PC2) than for herd management specialisation (milk yield versus fat and protein contents, PC1).

For clarity, only analyses and results for the Normande and Holstein breeds were presented. However, estimates of G*E interactions in the Montbéliarde breed led to exactly the same conclusions with multiple-trait, single-trait and reaction norm models (results not shown).

Discussion

The aim of this study was to estimate genotype by environment interactions for production traits in the Holstein and Normande breeds in France. Genetic correlations between environments were very close to unity, except between very extreme environments for all breeds, models, and datasets, demonstrating that reranking of animals for production traits across environments does not exist in France. Such a result was previously reported in studies in France that used herd production level as definition of herd environment [8] as well as in other studies [9–11] that used different definitions of the environment. Yet, other studies did report genetic correlations less than one between environments, i.e., with reranking of animals. These studies dealt with data from different countries [12–15], that is, for ranges of environments that were greater than in this national study.

Variable genetic variances across environments for production traits were found in this study and have been reported before [9, 16]. Genetic variances increased with the capacity of the herd management to promote milk production. These results are in agreement with [8, 9, 17], in which genetic variances increased with increasing production level.

In a G*E interaction study, the definition of the herd environment is crucial. Definitions used in the literature are extremely diverse; they depend on the scale of the study (experimental farm versus national or international studies) and on the traits analysed. In the case of production traits, definition of the herd environment can be based on specific features of the feeding system, such as the level of concentrate in the diet [18, 19], grazing severity and silage quality [20], features of the reproduction system, such as the calving system (seasonal or uniform) [11], features of the herd structure, such as herd size [11], features of the climate such, as temperature humidity index [16], rainfall [7], or features of genetic background (percent of Holstein genes) [7]. Many studies have described the environment based on observed performances of the animals, such as herd milk production level [8, 21], fat and protein yields [22], peak milk yield, or persistency [23]. In these cases, environmental and genetic factors are combined. In this study, herd environment was described based on HTD profiles. This definition focuses on the part of the production due to herd management only. This improves the study of G*E interactions because environmental and genetic factors are no longer confounded in the definition of the environment. Moreover, HTD profiles are available from national databases and do not require extra recording, in contrast to many herd management descriptors. Using HTD profiles allows analysis of large datasets. Finally, HTD profiles summarize all impacts of environment on production and offer a general overview of the environment, whereas some other herd management descriptors reduce environment to a limited number of features (temperature, average performances, herd size).

Summarizing HTD profiles descriptors by first PC scores allowed correlations between the 21 descriptors to be taken into account and a focus on the main causes of variability among HTD profiles. However, by limiting the analysis to the first PC scores, part of the diversity of HTD profiles that reflect differences in herd management (i.e., environment diversity) was not accounted for, regardless of the model used (multiple-trait, single-trait or reaction norm model). In fact, the three herd clusters used to describe the environment in the multiple-trait model were built based on the first 10 PC scores only, which explained about 76% of the total variance in HTD profiles. Moreover, we selected paragon herds for each cluster, which reduced the within-environment diversity. For the reaction norm models, the environment was described only through one or two PC scores. This may seem reductive but reaction norm models based on three PC scores gave poorer goodness of fit in terms of the Bayesian Information Criterion.

The models that were used to estimate G*E interactions were animal models with pedigree information over three generations, in contrast with other studies that used simpler models such as sire models [13, 16] or sire-maternal grand sire models [17]. Two types of models were tested: multiple-trait and reaction norm models. A drawback of multiple-trait models is that they require classification of environments, which cannot represent the full diversity of environments. Moreover, in this study, the multiple-trait model was applied to the "paragon" dataset, in which extreme environments were not represented, which led to a reduction in environmental variance. Despite this reduced diversity of environments, a clear heterogeneity of heritabilities among the three herd clusters was identified. In contrast, reaction norms model an "infinite" number of environments, which more precisely depicts the existing continuum of the environment. Generally, the parameter that describes the environment in reaction norm models is a single measure such as age at calving, herd size [24], or herd-year averages of protein yield [25]. This environment parameter can also be a synthetic variable that summarizes information of several environmental variables. In [17], 65 environmental variables were reduced into four PCs by a factor analysis and were used separately. Hence, one major improvement in the current study was that several PCs were used simultaneously to describe the environment. The number of parameters to estimate in the model was limited by using linear reaction norms rather than more sophisticated functions such as polynomials. The next step will be to study the possibility to simultaneously account for a larger number of PCs to describe the environment. In particular, a reduced rank genetic matrix could be used to summarize the effect of several PCs on the genetic effect. Reaction norm models applied to the diversity dataset allowed the investigation of extreme environments (for one or two PCs). Here, an average residual variance which did not depend on the environment was used to estimate heritabilities with the reaction norm model. Consequently, differences of heritabilities across environments

were only due to differences in genetic variances, which may have exacerbated differences in heritabilities between environments. These differences of heritabilities across environments might be exacerbated by the use of linear reaction norms and of an average residual variance. Thus, the reaction norm model could be improved by allowing different residual variances across environments.

No reranking of animals was shown for production traits. These traits have been selected for a long time, and thus, animals may be well adapted to all herd managements that currently exist in France. Nevertheless, within the context of the development of a sustainable agriculture, new ecological constraints appear such as controlling the use of phytosanitary products or protecting some agricultural areas. Also, new economical constraints due to reorganization of agricultural areas, with a decrease in the number of farmers or the end of quotas and liberalisation of milk production could raise new types of herd managements. Depending on how breeders react to these constraints, the range of environments could get larger. Thus, G*E interaction studies will have to be updated in order to assess whether animals remain well adapted to all herd environments.

Greater G*E interactions might exist for more recently selected traits. For these traits, the processes may not have already removed the animals' capacity to be specifically adapted to a particular environment. Thus, on a follow-up study, we will investigate G*E interactions on functional traits.

Conclusions

Presence of genotype by environment interactions was evaluated for production traits (milk, protein and fat yields, protein and fat contents) using multiple-trait (which was eventually converted to a single-trait model) and reaction norm animal models for the Holstein, Normande and Montbéliarde breeds, and using herd environment descriptors derived from Herd Test Day profiles. No reranking of animals between environments was found for any breed or model. Therefore, it can be concluded that existing breeding schemes are efficient regardless of the environment in which animals are raised and produce. However, a heterogeneity of heritabilities across environments was apparent. In most cases: the more intensive the herd management for milk yield, the larger the heritability. Ignoring this heterogeneity makes reliabilities of estimated breeding values inaccurate. Moreover, the heritability and the genetic variance gradients across environments could entail a higher genetic response in the most intensive herd managements.

Competing interests

The authors declare that they have no competing interests.

Author's contributions

BH and VD jointly conceived the design of the study. BH, HL and VD discussed the results. BH wrote and checked programs. BH wrote the draft of the manuscript, VD and HL made suggestions and corrections. All authors read and approved the final manuscript.

Acknowledgements

This study was financially supported by the Ministry of Agriculture (Programme CASDAR GENESYS). Authors thanks Anne Barbat (INRA, France) for her help with the data preparation.

References

1. Huquet B, Leclerc H, Ducrocq V: **Characterization of French dairy farm environments from herd-test-day profiles.** *J Dairy Sci* 2012, **95**:4085–4098.
2. Leclerc H, Duclos D, Barbat A, Druet T, Ducrocq V: **Environmental effects on lactation curves included in a test-day model genetic evaluation**. *Animal* 2008, **2**:344–353.
3. Koivula M, Nousiainen JI, Nousiainen J, Mäntysaari EA: **Use of herd solutions from a random regression test-day model for diagnostic dairy herd management.** *J Dairy Sci* 2007, **90**:2563–2568.
4. Escofier B, Pagès J: **Multiple factor analysis (AFMULT package)**. *Comput Stat Data Anal* 1994, **18**:121–140.
5. Meyer K: **WOMBAT: a tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML).** *J Zhejiang Univ Sci B* 2007, **8**:815–821.
6. Schaeffer LR: **Multiple-country comparison of dairy sires.** *J Dairy Sci* 1994, **77**:2671–2678.
7. Weigel KA, Rekaya R: **A multiple-trait herd cluster model for international dairy sire evaluation.** *J Dairy Sci* 2000, **83**(4):815–821.
8. Bonaiti B: **Interaction génotype milieu sur la production laitière chez les bovins II.-Corrélation génétique entre niveaux de production et variation de l'héritabilité**. *Ann Genet Sel Animal* 1982, **14**:441–452.
9. Raffrenato E, Blake RW, Oltenacu PA, Carvalheira J, Licitra G: **Genotype by environment interaction for yield and somatic cell score with alternative environmental definitions.** *J Dairy Sci* 2003, **86**:2470–2479.
10. Castillo-Juarez H, Oltenacu PA, Blake RW, Mcculloch CE, Cienfuegos-Rivas EG: **Effect of herd environment on the genetic and phenotypic relationships among milk yield, conception rate, and somatic cell score in Holstein cattle.** *J Dairy Sci* 2000, **83**:807–814.
11. Haile-Mariam M, Carrick MJ, Goddard ME: **Genotype by environment interaction for fertility, survival, and milk production traits in Australian dairy cattle.** *J Dairy Sci* 2008, **91**:4840–4853.
12. Cerón-Muñoz MF, Tonhati H, Costa CN, Rojas-Sarmiento D, Echeverri Echeverri DM: **Factors that cause genotype by environment interaction and use of a multiple-trait herd-cluster model for milk yield of Holstein cattle from Brazil and Colombia.** *J Dairy Sci* 2004, **87**:2687–2692.
13. Weigel KA, Rekaya R, Zwald NR, Fikse WF: **International genetic evaluation of dairy sires using a multiple-trait model with individual animal performance records.** *J Dairy Sci* 2001, **84**:2789–2795.
14. Ojango JMK, Pollott GE: **The relationship between Holstein bull breeding values for milk yield derived in both the UK and Kenya**. *Livest Prod Sci* 2002, **74**:1–12.
15. Hammami H, Rekik B, Soyeurt H, Bastin C, Stoll J, Gengler N: **Genotype x environment interaction for milk yield in Holsteins using Luxembourg and Tunisian populations.** *J Dairy Sci* 2008, **91**:3661–3671.
16. Hayes BJ, Carrick M, Bowman P, Goddard ME: **Genotype x environment interaction for milk production of daughters of Australian dairy sires from test-day records.** *J Dairy Sci* 2003, **86**:3736–3744.
17. Windig JJ, Calus MPL, Beerda B, Veerkamp RF: **Genetic correlations between milk production and health and fertility depending on herd environment.** *J Dairy Sci* 2006, **89**:1765–1775.
18. Cromie A, Kelleher D, Gordan F, Rath M: **Genotype by environment interaction for milk production traits in Holstein-Friesian dairy cattle in Ireland**. *Interbull Bull* 1998, **17**:100–104.
19. Fulkerson WJ, Davison TM, Garcia SC, Hough G, Goddard ME, Dobos R, Blockey M: **Holstein-Friesian dairy cows under a predominantly grazing system: interaction between genotype and environment.** *J Dairy Sci* 2008, **91**:826–839.
20. Berry DP, Buckley F, Dillon P, Evans RD, Rath M, Veerkamp RF: **Genetic parameters for body condition score, body weight, milk yield, and fertility estimated using random regression models.** *J Dairy Sci* 2003, **86**:3704–3717.
21. De Veer JC, Van Vleck LD: **Genetic parameters for first lactation milk yields at three levels of herd production.** *J Dairy Sci* 1987, **70**:1434–1441.

22. Castillo-Juarez H, Oltenacu PA, Cienfuegos-Rivas EG: **Genetic and phenotypic relationships among milk production and composition traits in primiparous Holstein cows in two different herd environments**. *Livest Production Sci* 2002, **78**:223–231.

23. Fikse WF, Rekaya R, Weigel KA: **Assessment of environmental descriptors for studying genotype by environment interaction**. *Livest Production Sci* 2003, **82**:223–231.

24. Calus MPL, Veerkamp RF: **Estimation of environmental sensitivity of genetic merit for milk production traits using a random regression model.** *J Dairy Sci* 2003, **86**:3756–3764.

25. Kolmodin R: **Reaction norms for the study of genotype by environment interaction in animal breeding**. *PhD thesis*, Swedish University of Agricultural Sciences; 2003.

Figures

Figure 1 - Herd Test Profiles (figure extracted from [1])

This figure shows an example of a herd described by its three Herd Test Day profiles (for milk yield, and protein and fat contents) before (dashed line) or after (solid line) smoothing.

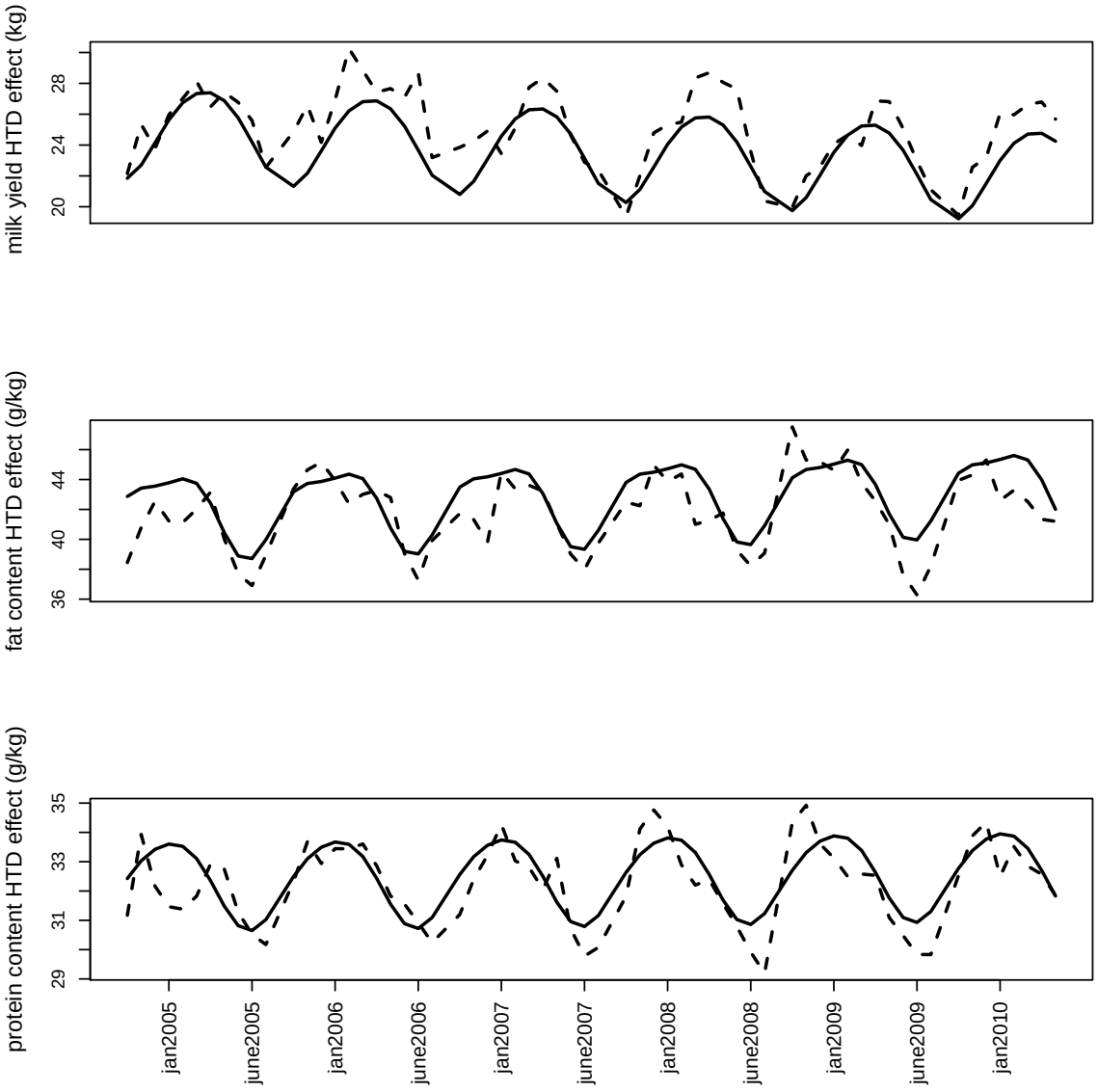


Figure 2 - Descriptors of Herd Test Day Profiles (figure extracted from [1])

This figure shows the descriptors of smoothed Herd Test Day profiles used in the Multiple Factorial Analysis.

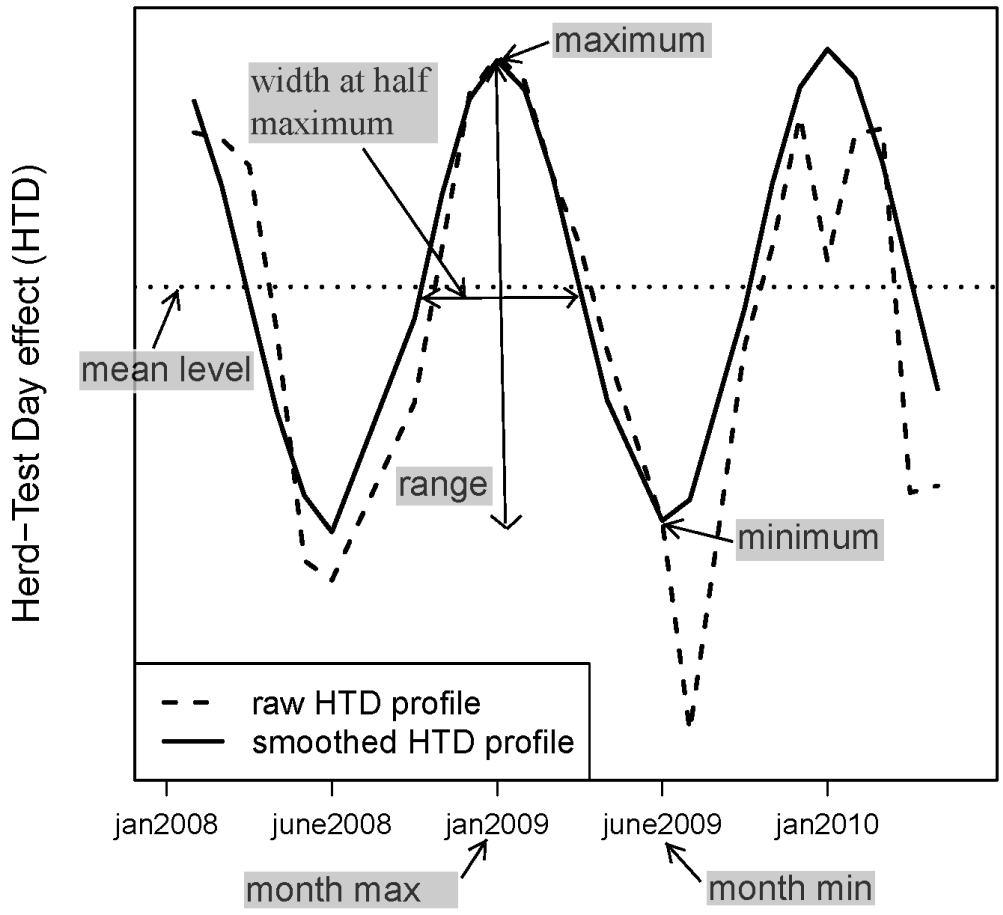


Figure 3 - Diversity and paragon herds for the Holstein breed on the first factorial map

This figure shows how Holstein herds from the diversity and paragon lists were distributed over the first factorial map. Herds in pink, orange and red are the "paragon" herds and depict the three herd clusters used in the multiple-trait model for the paragon dataset.

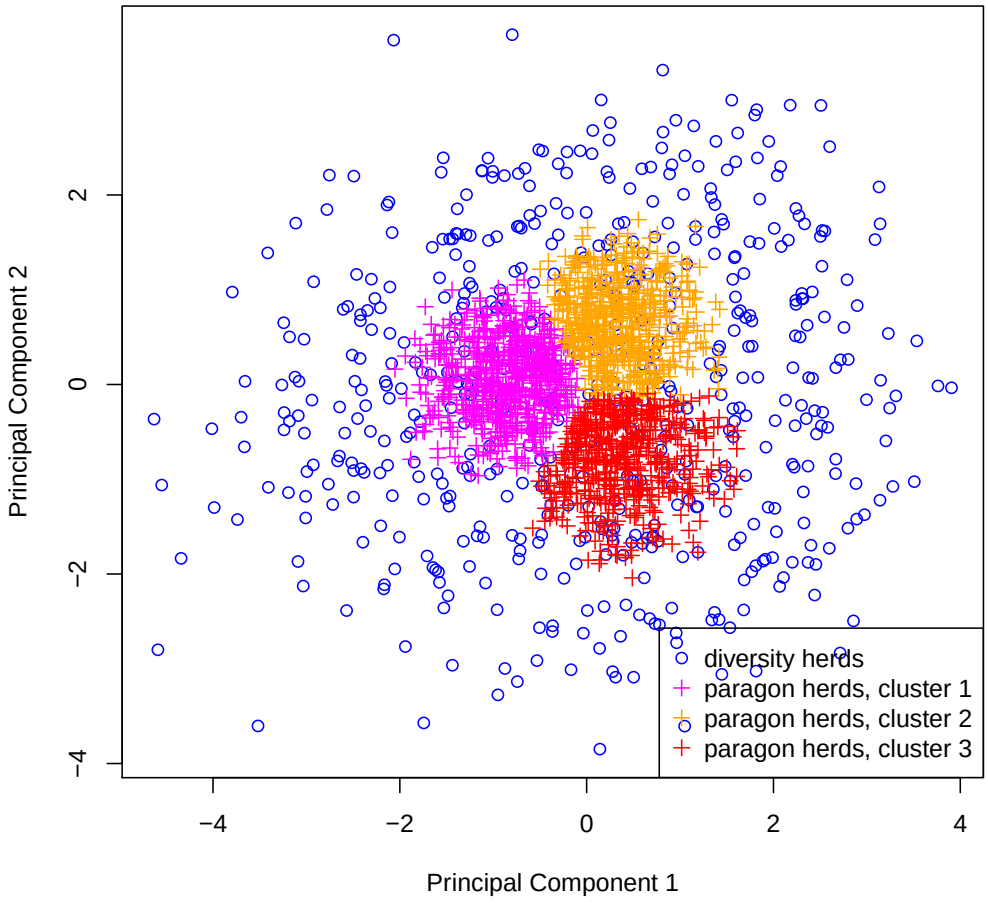


Figure 4 - Diversity and paragon herds for the Normande breed on the first factorial map

This figure shows how Normande herds from the diversity and paragon lists were distributed over the first factorial map. Herds in pink, orange and red are the "paragon" herds and depict the three herd clusters used in the multiple-trait model for the paragon dataset.

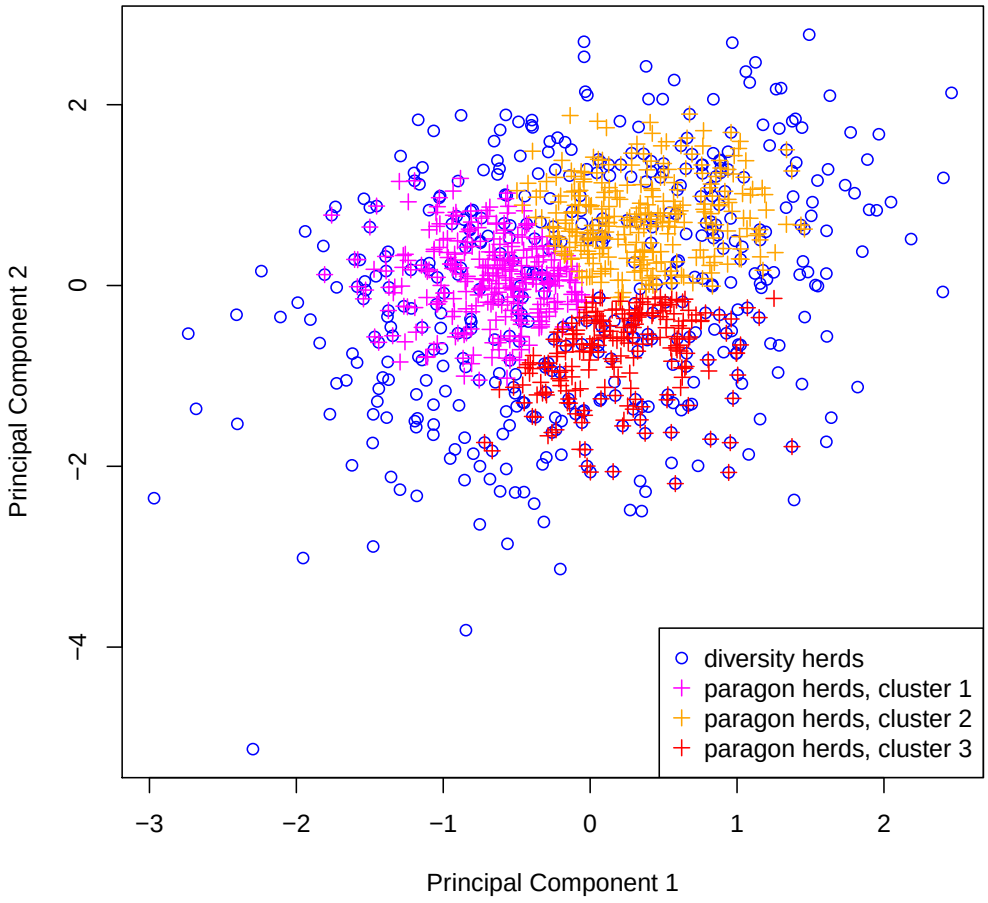


Figure 5 - Heritabilities for milk yield in the Holstein breed based on the reaction norm model

This figure shows heritabilities of milk yield calculated from the reaction norm model applied on the Holstein diversity dataset.

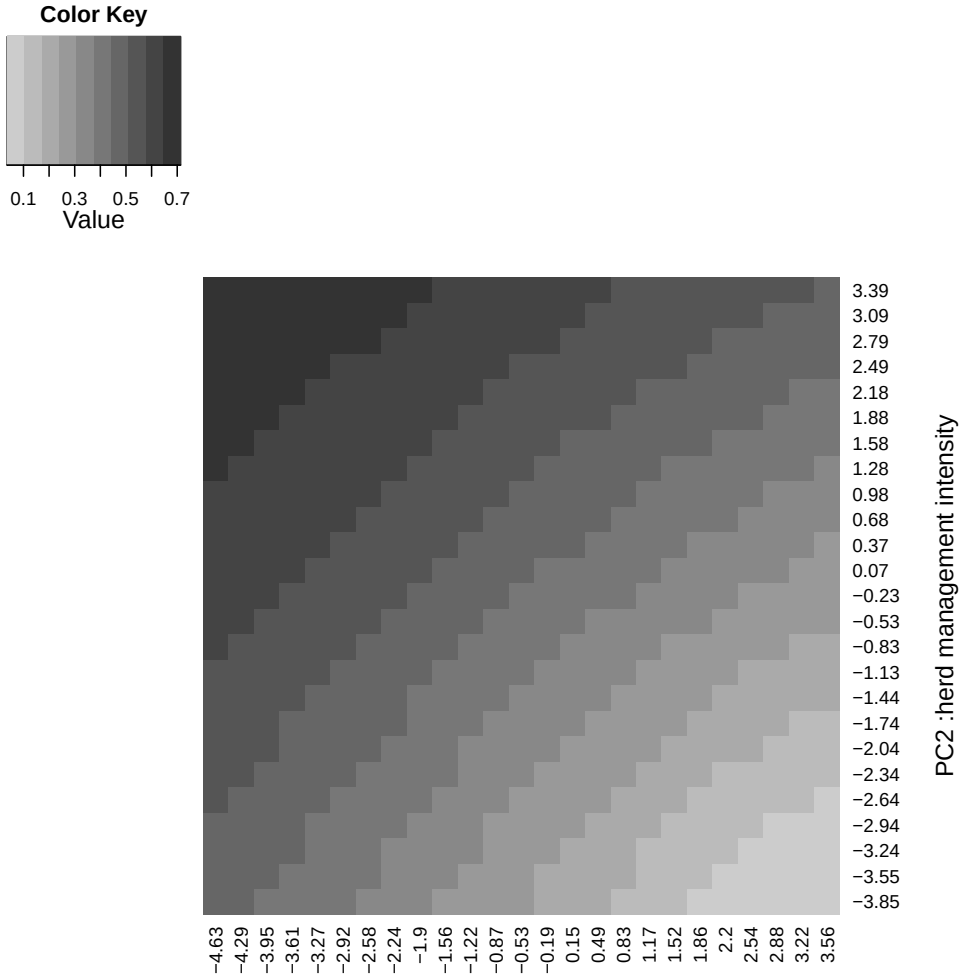
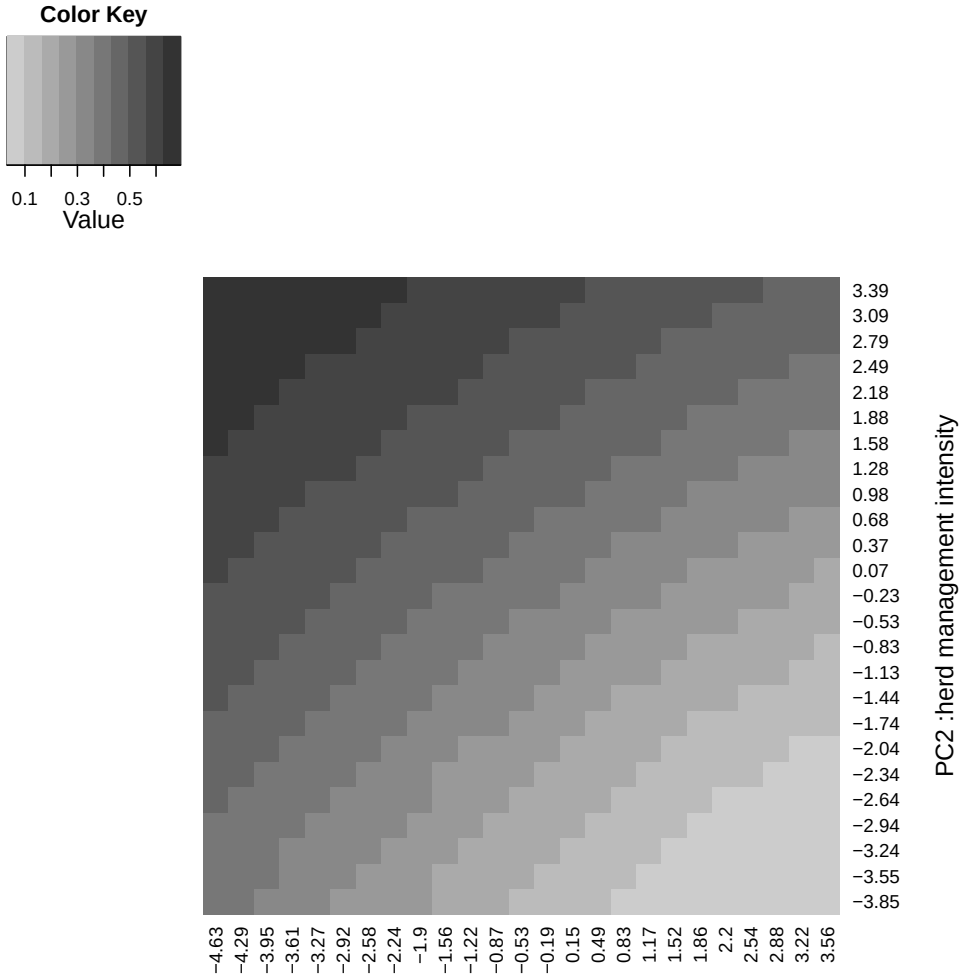


Figure 6 - Heritabilities for protein yield in the Holstein breed based on reaction norm model

This figure shows heritabilities of protein yield estimated from the reaction norm model applied to the Holstein diversity dataset.



Tables

Table 1 - Characteristics of the Holstein paragon dataset for the multiple-trait model

This table presents the number of herds and first lactation records in the paragon dataset per breed and the number of animals included in the pedigree file. Mean and standard deviation (between brackets) of analysed traits and PC scores (only for the 3 first PC; herd clusters were created based on the first 10 PC) for each herd cluster and breed are given.

	herd cluster	C1	C2	C3
Data	Number of herds	629	555	506
	Number of records	99,862	79,104	67,774
	Pedigree		412,732	
Environment descriptors	PC 1 score: herd management specialisation milk yield vs protein and fat contents	-0.88 (0.4)	0.36 (0.4)	0.38 (0.5)
	PC 2 score: herd management intensity	0.06 (0.4)	0.69 (0.4)	-0.79 (0.4)
	PC 3 score: herd management seasonality	0.15 (0.4)	0.05 (0.5)	-0.02 (0.5)
Phenotypes	Milk yield (kg)	10,163 (1,550)	9,818 (1,553)	9,152 (1,449)
	Fat yield (kg)	399 (60)	392 (62)	364 (59)
	Protein yield (kg)	325 (49)	316 (50)	291 (47))
	Fat content (%)	3.95 (0.43)	4.01 (0.42)	4.0 (0.42)
	Protein content (%)	3.20 (0.20)	3.23 (0.20)	3.19 (0.20)

Table 2 - Characteristics of the Normande paragon dataset for the multiple-trait model

This table presents the number of herds and first lactation records in the paragon dataset per breed and the number of animals included in the pedigree file. Mean and standard deviation (between brackets) of analysed traits and PC scores (only for the 3 first PC; herd clusters were created based on the first 10 PC) for each herd cluster and breed are given.

	herd cluster	C1	C2	C3
Data	Number of herds	291	287	218
	Number of records	49,131	45,352	30,982
	Pedigree		197,039	
Environment descriptors	PC 1 score: herd management specialisation milk yield vs protein and fat contents	-0.70 (0.4)	0.36 (0.4)	0.21 (0.4)
	PC 2 score: herd management intensity	0.08 (0.5)	0.75 (0.4)	-0.75 (0.5)
	PC 3 score: herd management seasonality	-0.14 (0.5)	-0.23 (0.5)	-0.07 (0.6)
Phenotypes	Milk yield (kg)	7,649 (1,219)	7,438 (1,230)	6,935 (1,162)
	Fat yield (kg)	327 (54)	322 (55)	297 (52)
	Protein yield (kg)	266 (43)	259 (44)	239 (41)
	Fat content (%)	4.29 (0.37)	4.34 (0.36)	4.30 (0.36)
	Protein content (%)	3.48 (0.20)	3.49 (0.20)	3.15 (0.20)

Table 3 - Characteristics of the paragon and diversity datasets for the reaction norm model

This table displays the number of herds and first lactation records in the paragon and diversity datasets per breed and the number of animals used in the pedigree for the reaction norm models. Mean and standard deviation (between brackets) of traits and environmental descriptors are given.

	breed	Holstein		Normande	
	dataset	paragons	diversity	paragons	diversity
Data	Number of herds	1,690	539	796	472
	Number of records	246,740	75,173	125,465	70,105
	Pedigree	412,732	133,310	197,039	112,276
Environment descriptors	PC1 score: herd management specialisation milk yield vs protein and fat contents	-0.13 (0.7)	-0.12 (1.7)	-0.10 (0.6)	-0.07 (1.0)
	PC2 score: herd management intensity	0.03 (0.8)	0.08 (1.5)	0.12 (0.7)	0.01 (1.2)
	PC3 score: herd management seasonality	0.07 (0.5)	0.10 (0.8)	-0.16 (0.5)	-0.12 (0.8)
Phenotypes	Milk yield (kg)	9,775 (1,577)	9,777 (1,863)	7,397 (1241)	7,294 (1,321)
	Fat yield (kg)	387 (62)	386 (73)	318 (55)	314 (59)
	Protein yield (kg)	313 (51)	313 (60)	257 (44)	253 (48)
	Fat content (%)	3.99 (0.43)	3.98 (0.45)	4.31 (0.36)	4.32 (0.37)
	Protein content (%)	3.21 (0.20)	3.20 (0.20)	3.48 (0.20)	3.46 (0.21)

Table 4 - Residual variance, genetic variance and heritabilities for the three herd clusters with the single-trait model

This table shows residual (var res) and genetic (var gen) variances and heritabilities (h^2) estimated with the single-trait model applied to the paragon dataset, as well as standard errors of estimates of genetic variances (between brackets). Results are presented for each trait by breed and herd cluster (C1, C2, C3).

		Holstein			Normande		
Trait		C1	C2	C3	C1	C2	C3
	gen var	835,690	802,710	590,920	519,830	473,220	420,700
		(17,204)	(18,570)	(15,387)	(15,473)	(15,302)	(15,638)
Milk yield	res var	952,294	970,338	887,510	636,713	643,032	586,439
	h^2	0.47	0.45	0.40	0.45	0.42	0.42
	gen var	1232.5	1193.7	939.62	822	821	744
		(26)	(29)	(25)	(27)	(29)	(28)
Fat yield	res var	1507	1605	1443	1260	1284	1141
	h^2	0.45	0.43	0.39	0.39	0.39	0.39
	gen var	608	580	416	513	472	428
		(15)	(16)	(13)	(17)	(17)	(17)
Protein yield	res var	967	1013	896	757	771	685
	h^2	0.39	0.36	0.32	0.40	0.38	0.38
	gen var	15.5	15.4	15.0	8.4	8.4	7.8
		(0.18)	(0.20)	(0.21)	(0.16)	(0.17)	(0.19)
Fat content	res var	3.9	4.2	4.2	3.6	3.9	3.9
	h^2	0.80	0.78	0.78	0.70	0.69	0.67
	gen var	2.6	2.5	2.5	2.5	2.5	2.4
		(0.038)	(0.041)	(0.042)	(0.051)	(0.052)	(0.061)
Protein content	res var	1.3	1.4	1.4	1.2	1.3	1.3
	h^2	0.67	0.64	0.64	0.67	0.66	0.64

Table 5 - Residual variances with the reaction norm model

This table presents residual variances used to calculate heritabilities based on the reaction norm model for the paragon and diversity datasets.

	Holstein		Normande	
Trait	paragon	diversity	paragon	diversity
Milk yield	931,367	945,995	623,825	620,922
Fat yield	1,506	1,531	1,233	1,207
Protein yield	948	940	740	723
Fat content	4.12	4.63	3.74	3.75
Protein content	1.33	1.35	1.27	1.29

Table 6 - Genetic correlations between environments with the reaction norm model

This table provides statistics on genetic correlations estimated with the reaction norm model for the paragon dataset and diversity datasets. Minimum, mean and maximum estimates if genetic correlations between environments are given for each trait by breed and dataset.

		Holstein						Normande					
		paragon			diversity			paragon			diversity		
Trait	model	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
Milk yield	RN_2	0.96	0.99	1	0.68	0.97	1	0.99	0.99	1	0.92	0.99	1
Fat yield	RN_2	0.97	0.99	1	0.60	0.96	1	0.99	0.99	1	0.89	0.99	1
Protein yield	RN_2	0.93	0.99	1	0.64	0.96	1	0.99	0.99	1	0.86	0.99	1
Fat content	RN_1	0.99	0.99	1	0.94	0.99	1	0.98	0.99	1	0.99	0.99	1
Protein content	RN_1	0.99	0.99	1	0.99	0.99	1	0.99	0.99	1	0.99	0.99	1

Table 7 - Heritabilities across environments with the reaction norm model

This table presents heritabilities estimated with the reaction norm model applied to the paragon and diversity datasets. Minimum, mean and maximum of heritabilities across environments are given for each production trait per breed and dataset.

		Holstein						Normande					
	dataset	paragon			diversity			paragon			diversity		
Trait	model	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
Milk yield	RN_2	0.26	0.43	0.55	0.12	0.42	0.66	0.28	0.43	0.52	0.17	0.40	0.53
Fat yield	RN_2	0.28	0.42	0.52	0.15	0.40	0.63	0.27	0.39	0.47	0.14	0.37	0.50
Protein yield	RN_2	0.80	0.35	0.49	0.07	0.36	0.64	0.24	0.39	0.48	0.12	0.36	0.52
Fat content	RN_1	0.79	0.79	0.79	0.76	0.76	0.77	0.67	0.69	0.70	0.68	0.69	0.71
Protein content	RN_1	0.65	0.66	0.67	0.63	0.64	0.66	0.65	0.66	0.67	0.64	0.65	0.66

4.6.3 Mise en évidence d'un effet d'échelle pour les caractères laitiers en race Montbéliarde

L'étude des interactions Génotype*Milieu a également été réalisée en race Montbéliarde à partir des 1104 élevages montbéliards inclus dans l'analyse factorielle multiraciale qui a permis de définir les principales dimensions de variabilité des profils Troupeau Jour de Contrôle. Les tableaux 4.1 et 4.2 présentent les données utilisées pour les modèles multicaractères et de norme de réaction.

Comme pour les races Normande et Holstein, le modèle multicaractère a mis en évidence un gradient d'héritabilité entre les milieux et des corrélations génétiques entre milieux très proches de 1. Ceci démontre l'absence de reclassement entre milieux et l'existence d'un effet d'échelle : l'héritabilité est d'autant plus forte que la conduite est spécialisée en lait (tableau 4.3).

Le modèle de norme de réaction (dont les particularités sont présentées dans le tableau 4.4), a lui aussi mis en évidence des corrélations génétiques entre milieux très élevées (tableau 4.5) ce qui montre que le classement des animaux est globalement le même quel que soit le type de conduite. Cependant, on observe certaines corrélations génétiques faibles qui correspondent aux corrélations génétiques entre milieux vraiment extrêmes ce qui montrerait une interaction Génotype*Milieu sous forme de reclassement entre ces milieux. De plus, il existe un effet d'échelle puisque les héritabilités diffèrent selon le milieu (tableau 4.6).

Pour conclure, la Montbéliarde présente les mêmes tendances que les races Normande et Holstein décrites dans l'article II : quel que soit le modèle et le jeu de données utilisé, on observe des interactions Génotype*Milieu sous la forme d'un effet d'échelle mais pas sous la forme d'un reclassement des animaux selon le milieu, sauf entre milieux tout à fait extrêmes.

TABLE 4.1 – Caractéristiques du jeu de données parangon pour la race Montbéliarde utilisées pour l'analyse multicaractères (écart-types entre parenthèses). C1 : un groupe d'élevages dont la conduite a une intensité moyenne plutôt spécialisée en lait. C2 : un groupe d'élevages dont la conduite est intensive mais plutôt dirigée vers des taux protéique et butyreux élevés. C3 : un groupe d'élevages dont la conduite est peu intensive.

	Groupes d'élevages	C1	C2	C3
Données	Nombre d'élevages	136	103	70
	Nombre de données	20,410	14,862	8,339
	Nombre d'élevages dans le pedigree	71,789		
Descripteurs de l'environnement	Coordonnées sur le premier facteur : spécialité de la conduite de troupeau lait vs taux protéique et butyreux	**-0.76** (0.4)	0.25 (0.4)	0.21 (0.4)
	Coordonnées sur le deuxième facteur : intensité de la conduite de troupeau	**0.09** (0.5)	**0.67** (0.5)	**-0.74** (0.5)
	Coordonnées sur le troisième facteur : saisonnalité de la conduite de troupeau	0.09 (0.5)	0.01 (0.5)	0.02 (0.5)
Phénotypes	Lait (kg)	8,649 (1,361)	8,363(1,306)	7,631 (1,296)
	Matière grasse (kg)	340 (56)	335 (55)	301 (55)
	Matière protéique (kg)	290 (46)	281 (44)	252 (44))
	Taux butyreux (%)	3.94 (0.31)	4.01 (0.31)	3.95 (0.31)
	Taux protéique (%)	3.36 (0.17)	3.37 (0.18)	3.31 (0.18)

TABLE 4.2 – Caractéristiques des jeux de données parangon et diversité utilisés pour l'analyse avec les normes de réaction (écart-types entre parenthèses)

	breed	Montbéliarde	
	jeu de données	parangon	diversité
Données	Nombre d'élevages	309	497
	Nombre de données	43,611	69,659
	Pedigree	71,789	112,800
Descripteurs de l'environnement	Coordonnées sur le premier facteur : spécialité de la conduite de troupeau lait vs taux protéique et butyreux	-0.22 (0.6)	-0.22 (0.7)
	Coordonnées sur le deuxième facteur : intensité de la conduite	0.13 (0.7)	0.20 (0.8))
	Coordonnées sur le troisième facteur : saisonnalité de la conduite	0.07 (0.5)	0.12 (0.7)
Phénotypes	Lait (kg)	8,356 (1,382)	8,375 (1,433)
	Matière grasse (kg)	331 (57)	332 (59)
	Matière protéique (kg)	280 (47)	281 (49)
	Taux butyreux (%)	3.97 (0.31)	3.97 (0.32))
	Taux protéique (%)	3.35 (0.18)	3.35 (0.18))

TABLE 4.3 – Résultats du modèle multicaractères (écart-types entre parenthèses)

		Montbéliarde		
Caractère		C1	C2	C3
Lait	var gen	527,476	407,107	337,441
		(27,943)	(28,158)	(26,535)
	var res	706,251	747,296	667,955
	h^2	0.43	0.36	0.34
Matière grasse	var gen	810.1	692.0	526.9
		(43)	(46)	(43)
	var res	1143	1242	1100
	h^2	0.42	0.36	0.32
Matière protéique	var gen	466	390	302
		(27)	(28)	(26)
	var res	760	791	703
	h^2	0.38	0.33	0.30
Taux butyreux	var gen	6.4	6.2	5.7
		(0.19)	(0.22)	(0.27)
	var res	2.6	3.0	3.0
	h^2	0.71	0.67	0.66
Taux protéique	var gen	2.0	1.8	1.8
		(0.07)	(0.07)	(0.10))
	var res	0.99	1.06	1.20
	h^2	0.67	0.63	0.60

TABLE 4.4 – Variances résiduelles utilisées pour calculer les héritabilités pour les modèles de norme de réaction

	Montbéliarde	
Caractère	parangon	diversité
Lait	703,175	693,485
Matière grasse	1,145	1,128
Matière protéique	743	733
Taux butyreux	2.80	2.80
Taux protéique	1.05	1.00

TABLE 4.5 – Corrélations génétiques entre les groupes d'élevages avec le modèle de norme de réaction

		Montbéliarde					
	jeu de données	parangon			diversité		
Caractère	modèle	Min	Moy	Max	Min	Moy	Max
Lait	RN_2	0.89	0.99	1	0.82	0.98	1
Matière grasse	RN_2	0.87	0.99	1	0.72	0.99	1
Matière protéique	RN_2	0.82	0.99	1	0.62	0.99	1
Taux butyreux	RN_1	0.97	0.99	1	0.97	0.99	1
Taux protéique	RN_1	0.99	0.99	1	0.99	0.99	1

TABLE 4.6 – Héritabilité pour les différents groupes d'élevages avec le modèle de norme de réaction

		Montbéliarde					
	jeu de données	parangon			diversité		
Caractère	modèle	Min	Moy	Max	Min	Moy	Max
Lait	RN_2	0.22	0.38	0.52	0.12	0.41	0.59
Matière grasse	RN_2	0.19	0.38	0.52	0.10	0.40	0.62
Matière protéique	RN_2	0.17	0.35	0.51	0.09	0.37	0.60
Taux butyreux	RN_1	0.69	0.69	0.70	0.68	0.70	0.73
Taux protéique	RN_1	0.64	0.64	0.65	0.65	0.66	0.67

4.6.4 Atouts et limites de la première étude

Cette première étude présente des avancées par rapport aux études déjà existantes. Elle présente une définition novatrice du milieu à partir des profils TJC. De plus, elle a été réalisée sur des volumes de données importants avec un modèle animal plus complet que les modèles habituellement utilisés pour ce genre d'étude.

Cependant, comme indiqué dans le chapitre 2, le milieu est défini d'une part à partir d'effets TJC issus d'un modèle qui néglige les interactions Génotype*Milieu, d'autre part seul les profils TJC lait, TB et TP sont utilisés alors que l'on peut imaginer que d'autres profils TJC auraient pu être utilisés.

En ce qui concerne les modèles utilisés pour estimer les interactions Génotype*Milieu, il s'agit de modèles lactation alors que les profils TJC sont issus d'un modèle contrôles élémentaires. Cette situation n'est pas idéale mais étudier les interactions Génotype*Milieu par un modèle basé sur les contrôles élémentaires auraient été trop complexe et difficile à mettre en place sur nos volumes de données.

Dans les modèles de norme de réaction, la variance résiduelle ne dépendait pas du milieu ; c'est un point à améliorer dans les prochaines études. Enfin, nous avons utilisé plusieurs variables continues pour décrire le milieu mais la relation entre le milieu et la valeur génétique a été considérée comme linéaire. Cette forme a été choisie pour sa simplicité, des relations plus complexes auraient entraîné un nombre de paramètres trop importants (surtout en utilisant plusieurs variables de milieu). Cette relation linéaire pourrait exacerber l'interaction Génotype*Milieu sous forme d'effet d'échelle qui a été mise en évidence.

4.7 Deuxième étude des interactions Génotype*Milieu en France sur les caractères laitiers et fonctionnels

4.7.1 Objectifs

Cette étude est un autre volet de l'étude des interactions Génotype*Milieu en France. Elle complète la précédente par le développement d'une nouvelle méthode de description

du milieu au sein du modèle d'étude des interactions et un élargissement des types de caractères étudiés.

L'objectif est de mesurer les interactions Génotype*Milieu, via un modèle de norme de réaction, sur les caractères laitiers (Lait, MG, MP) et les caractères fonctionnels (Score de Cellules Somatiques (SCS), Intervalle Vêlage-première Insémination Artificielle (IVIA1) et la réussite à la première insémination artificielle (ResIA). L'étude a été menée sur la race Montbéliarde. Les caractères fonctionnels sont étudiés car ils ont un poids de plus en plus important dans l'Index de Synthère Upra (ISU) notamment ceux qui concernent la santé de la mamelle et la reproduction (IBL, 2012).

4.7.2 Données

Les caractères étudiés sont la quantité de lait, de matière grasse (MG), de matière protéique (MP) et le score cellulaire (SCS) mesurés sur 305 jours de lactation. L'Intervalle Vêlage- première Insémination Artificielle (IVIA1) correspond à la durée entre le premier vêlage d'une vache et sa première insémination artificielle pour un deuxième vêlage. La réussite à la première insémination artificielle est une variable binaire (0/1). Tous ces caractères sont étudiés sur des vaches en première lactation.

La sélection des données de performances de production laitière a été réalisée sur les mêmes critères que ceux retenus dans l'article II. Des différences avec l'article II peuvent néanmoins exister puisque les données utilisées sont issues d'extractions du Système d'Information Génétique à des dates différentes. Sur ce jeu de données, les performances pour le score cellulaire, la réussite à la première insémination artificielle et l'intervalle vêlage-première insémination artificielle ne présentaient pas de données aberrantes. Il n'y a donc pas eu de critères de sélection supplémentaires. Le tableau 4.7 donne un résumé des performances utilisées.

Comme dans l'article II, l'environnement a été décrit par les coordonnées de chaque élevage sur les composantes principales des profils TJC obtenues après l'application de l'approche décrite dans l'article II. Pour rappel, l'analyse factorielle était réalisée sur les descripteurs des profils TJC (comme définis dans l'article I) de 15756 élevages holsteins, normands et montbéliards.

Le nombre d'élevages retenus dans l'analyse (306) a été réduit pour limiter le volume

TABLE 4.7 – Présentation des données de performance

	1% inf	moy	99% sup	sd
Lait (kg)	5403	8425	11560	1334
MG (kg)	209	333	462	55
MP (kg)	179	282	388	45
SCS	0.41	2.61	5.68	1.2
IVIA1 (j)	36	75	172	27
ResIA	réussite :17406 échec :15161			

du jeu de données en utilisant la même méthode que la sélection des élevages pour les jeux de données « diversité » de l'article II : les élevages sélectionnées représentent toute la diversité du milieu. Cela correspond à 32567 vaches en première lactation.

4.7.3 Méthodes

Les analyses réalisées pour l'article II ont montré que les temps de calcul étaient moins importants pour l'estimation des interactions Génotype*Milieu via un modèle de norme de réaction que via un modèle multicaractères. De plus, la description du milieu via une ou plusieurs variables continues permet de mieux modéliser le continuum des systèmes d'élevage existants en France que la définition arbitraire de classes. C'est pourquoi nous avons privilégié les normes de réaction dans cette nouvelle étude.

Différentes alternatives existent sur la manière de prendre en compte le milieu au sein du modèle. Dans la bibliographie, la plupart des auteurs décrivent l'environnement à travers une seule variable, cette variable pouvant être synthétique ((Windig *et al.*, 2006) par exemple). Dans l'article II, nous avons comparé les modèles prenant en compte 1, 2 ou 3 variables d'environnement qui étaient les composantes principales synthétisant les descripteurs des profils TJC. Pour les caractères laitiers, il s'est avéré que les meilleurs modèles de norme de réaction (au sens du BIC : Bayesian Information Criterion) prenaient en compte 1 ou 2 composantes principales. Pourtant, instinctivement, il semblerait plus raisonnable de prendre en compte un maximum de composantes principales pour décrire l'environnement afin de perdre le moins possible d'information contenue dans les profils TJC. De plus, certaines solutions des modèles de norme de réaction utilisées dans l'article II étaient à la bordure de l'espace des paramètres. Cela avait pour conséquence

des matrices de variance/covariance génétiques qui n'étaient pas de plein rang. Le modèle pouvait donc être amélioré.

Trois modèles de norme de réaction ont été comparés afin d'explorer de nouvelles façons de prendre en compte l'environnement dans l'estimation de la part génétique du modèle. Ils ont été utilisés séparément pour chaque caractère étudié. L'ensemble des analyses a été réalisé avec le logiciel WOMBAT Meyer (2007). Pour les 3 modèles, il s'agit d'un modèle animal prenant en compte une généalogie remontant sur 3 générations ce qui correspond à 58539 animaux. Les 3 modèles comportent les mêmes effets fixes quel que soit le caractère : Troupeau_Année, Age au vêlage_Année et Mois de vêlage_Année. En revanche, la manière de modéliser les effets génétique et résiduel diffère.

Le premier modèle est un modèle de norme de réaction prenant en compte simultanément et de manière disjointe les 3 premières composantes principales des profils TJC pour décrire l'environnement :

$$Y_{it} = Troupeau_Annee_i + AgeVel_Annee_i + MoisVel_Annee_i + a_{it} + e$$

avec $a_{it} = a_{0i} + \sum_{j=1}^{3}(a_{ij} * CP_{jt})$.
Y_{it} est la performance de l'animal i dans le troupeau t. a_i est l'effet génétique de l'animal i dans le troupeau (environnement) t. Il est considéré comme aléatoire. Il est composé d'un effet génétique indépendant de l'environnement a_{0i} et d'effets génétiques dépendant de l'environnement dans lequel l'animal évolue. L'environnement est représenté par les coordonnées de l'élevage auquel il appartient sur les 3 premières composantes principales CP_{1t}, CP_{2t}, CP_{3t}. La matrice de variance/covariance génétique associée à ce modèle est donc une matrice de dimension 4*4 dont le rang devrait être égal à 4.

Le modèle suppose pour la résiduelle e, une variance hétérogène qui dépend de l'environnement. Dans notre cas, il faudrait que la variance résiduelle puisse dépendre des valeurs des composantes principales mais le logiciel WOMBAT ne permet pas encore d'établir une structure de variance résiduelle qui soit une fonction définie par l'utilisateur. Par conséquent, on a fait dépendre la variance résiduelle de groupes d'élevages (qui représentent différents environnements) et ce, bien que le modèle de norme de réaction permette théoriquement de traiter l'environnement comme une ou des variables continues. Cette stratégie est inspirée de celle utilisée par Windig *et al.* (2006). Elle n'est pas idéale mais représente une amélioration par rapport à la manière de modéliser l'hétérogénéité

des variances résiduelles utilisée dans l'article II. Ainsi, la variance résiduelle est estimée pour 8 groupes d'élevages définis en fonction de leurs coordonnées sur les 3 premières composantes principales. L'espace en 3 dimensions défini par les coordonnées des élevages sur les 3 premières composantes principales a été divisé en 8 espaces de volume égal ; chaque espace détermine un groupe d'élevages.

Les résultats du premier modèle ont montré que la matrice de variance/covariance estimée n'était pas de plein rang. Ainsi, un deuxième modèle a été mis en place. Il est le même que le précédent mais il suppose a priori une matrice de variance/covariance génétique de dimension 4 et de rang 2. Les 2 vecteurs propres de cette matrice sont de dimension 4 et sont nommés $u1$ et $u2$. L'option « PC » des modèles de régressions aléaoires (RR pour random regression) du logiciel WOMBAT Meyer (2007) a été utilisée pour réaliser ce modèle.

Deux nouvelles variables nommées $V1$ et $V2$ ont alors été créées pour définir de nouvelles variables d'environnement. Elles sont des combinaisons linéaires d'un effet moyen et des 3 premières composantes principales. Les poids des différentes composantes sont attribués par les valeurs des vecteurs propres $u1$ et $u2$. Ces variables sont définies pour l'élevage t de la façon suivante :

$$V1_t = u1[1] + \sum_{j=1}^{3}(u1[j+1] * CP_{jt})$$

et

$$V2 = u2[1] + \sum_{j=1}^{3}(u2[j+1] * CP_{jt})$$

Le troisième modèle utilise ces variables directement comme nouvelles variables d'environnement et s'écrit :

$$Y_{it} = Troupeau_Annee_i + AgeVel_Annee_i + MoisVel_Annee_i + a_i + e$$

avec $a_{it} = a_{i1} * V1_t + a_{i2} * V2_t$.
Ce troisième modèle définit des variances résiduelles hétérogènes qui dépendent de 9 groupes d'élevages ; ce nombre est arbitraire. Les 9 groupes d'élevages sont définis selon les 9 quantiles de la variable $V2$ (la variable $V1$ n'a pas été prise en compte dans la définition des groupes pour la variance résiduelle car les 3 premières composantes princi-

pales de l'analyse factorielle décrivant les profils Troupeau Jour de Contrôle contribuaient peu à sa construction). Il faut bien noter que ces 9 groupes sont définis à partir de la variable $V2$ alors que les groupes d'élevages utilisés dans les modèles 1 et 2 étaient construits directement à partir des 3 composantes principales. Pour ce dernier modèle, les héritabilités dans chaque groupe et les corrélations génétiques entre ces groupes ont été calculées comme cela a été présenté dans le paragraphe 4.3.2.

Les 3 modèles ont été comparés pour chaque caractère à l'aide du Bayesian Index Criterion (BIC).

4.7.4 Résultats

Pour le premier modèle, les corrélations estimées entre les termes a_0, a_1, a_2, a_3 sont très élevées pour tous les caractères (jusqu'à 0.99). Cela a entraîné une vitesse de convergence assez lente pour ce modèle. Le tableau 4.8 montre la décomposition des 4 valeurs propres de la matrice génétique : la matrice de matrice génétique de variance/covariance est de rang 3 voire de rang 2. Le tableau 4.9 montre que la contribution de la quatrième valeur propre est nulle et celle de la troisième est très faible (inférieure à 5%). C'est cette observation qui nous a amené à tester le modèle 2 qui consiste à forcer la matrice génétique à être de rang 2.

Dans le modèle 2, les 2 vecteurs propres résument l'information de la matrice génétique de dimension 4 (cf le tableau 4.9). Cette réduction de rang facilite grandement la convergence du modèle. On remarque que la contribution de la première valeur propre est très peu modifiée entre les modèles 1 et 2.

La première valeur associée au vecteur propre $u1$ est très élevée (supérieure à 0.97) pour tous les caractères mis à part l'intervalle vêlage-première insémination artificielle. Cela signifie que la valeur génétique indépendante de l'environnement a_0 contribue très largement à la construction de ce premier vecteur.

Le deuxième vecteur propre ($u2$) est, quant à lui, essentiellement construit à partir des valeurs génétiques dépendant des 3 composantes principales décrivant l'environnement a_1, a_2, a_3 ; la contribution de a_0 étant quasi nulle (tableau 4.9). Pour tous les caractères, sauf -encore une fois- l'intervalle vêlage-première insémination artificielle, le deuxième vecteur propre est principalement construit à partir des composantes principales. Pour les trois caractères laitiers, les poids se répartissent approximativement de la même façon avec un

coefficient positif très important pour la troisième composante principale qui discrimine les élevages selon « la saisonnalité de leur conduite », c'est à dire selon l'homogénéité ou l'hétérogénéité de l'efficacité de leur conduite pour produire du lait et des taux tout au long de l'année. Il faut tout de même noter le coefficient quasi nul de la deuxième composante principale (qui discrimine les élevages selon leur intensité de production) pour la matière protéique. Pour le Score de Cellules Somatiques, les poids des 2 premières composantes sont inversés par rapport aux caractères laitiers. Enfin, pour la réussite à la première insémination artificielle, le poids le plus important est celui de la première composante principale qui discrimine les élevages selon la spécialité de la conduite (lait vs taux).

La réduction du rang de la matrice génétique a permis de mettre en exergue une nouvelle décomposition de l'effet génétique : un effet génétique que l'on peut toujours interpréter comme indépendant de l'environnement et un effet génétique qui dépend de l'environnement (et non pas trois effets comme dans le modèle 1). L'environnement y est défini par une combinaison linéaire des 3 composantes principales utilisées dans le modèle 1, les poids de chaque composante étant définis par le vecteur propre $u2$ (et non pas imposés par l'utilisateur).

Cette nouvelle approche de l'environnement est utilisée dans le modèle 3. Le tableau 4.9 montre que comme entre les modèles 1 et 2, la contribution de la première valeur propre est quasi la même entre les modèles 2 et 3 (excepté pour l'intervalle vêlage-première insémination artificielle). Le tableau 4.10 montre le BIC obtenu pour chaque modèle et chaque caractère. Le modèle ayant le BIC le plus faible est considéré comme le meilleur. Excepté pour l'intervalle vêlage-première insémination artificielle, le modèle retenu est toujours le troisième.

Les résultats obtenus pour l'intervalle vêlage-premier insémination artificielle étant très différents des autres caractères, nous ne nous attarderons pas sur leur interprétation. Une étude plus approfondie de ce caractère est nécessaire pour vérifier l'origine de ce comportement atypique.

Le tableau 4.11 montre les corrélations génétiques obtenues entre les différents groupes d'élevages définis à partir de la nouvelle variable d'environnement (V2), combinaison linéaire des 3 premières composantes principales. Pour les caractères de production laitière, les corrélations génétiques sont très élevées à part entre environnements extrêmes où la corrélation peut descendre jusqu'à 0.83 en lait. Ces résultats confirment les résultats de

l'article II plaidant pour l'absence d'interaction Génotype*Milieu sous forme de reclassement. Le même phénomène est observé pour le score cellulaire. En revanche, la corrélation génétique moyenne entre les environnements pour la réussite à la première insémination artificielle est plus faible (0,88) ce qui mettrait en évidence une interaction Génotype*Milieu qui n'est pas seulement un effet d'echelle mais conduit à des reclassements.

Pour tous les caractères, les héritabilités varient en fonction de l'environnement : de 0,37 à 0,43 pour le lait, de 0,29 à 0,46 pour la matière grasse, de 0,33 à 0,39 pour la matière protéique, de 0,17 à 0,23 pour le score cellulaire et de 0,015 à 0,027 pour la réussite à la première insémination artificielle (cf les figures 4.2, 4.3, 4.4, 4.5, 4.6). Cette variabilité de l'héritabilité ne peut pas être attribuée à une plus forte augmentation de la variabilité résiduelle par rapport à la variabilité génétique ou inversement car les variances génétique et résiduelle ne varient pas de manière linéaire en fonction de la nouvelle variable d'environnement, combinaison linéaire des 3 composantes principales (cf les figures 4.2, 4.3, 4.4, 4.5, 4.6). L'héritabilité évolue donc de manière essentiellement non linéaire en fonction de la nouvelle variable d'environnement.

TABLE 4.8 – Valeurs propres (exprimées en pourcentage) des matrices de variance/covariance génétique pour les différents caractères et modèles

		valeur propre 1	valeur propre 2	valeur propre 3	valeur propre 4
Lait	modèle 1	93.21	5.61	1.18	0.00
	modèle 2	94.90	5.10		
	modèle 3	94.08	5.92		
MG	modèle 1	96.65	2.93	0.42	0.00
	modèle 2	97.04	2.96		
	modèle 3	97.36	2.64		
MP	modèle 1	92.77	5.36	1.88	0.00
	modèle 2	95.36	4.64		
	modèle 3	94.61	5.39		
SCS	modèle 1	95.44	3.92	0.64	0.00
	modèle 2	96.30	3.70		
	modèle 3	95.46	4.54		
IVIA	modèle 1	68.43	30.69	0.87	0.00
	modèle 2	69.18	30.82		
	modèle 3	79.20	20.80		
ResIA	modèle 1	82.13	13.80	4.06	0.01
	modèle 2	81.19	18.81		
	modèle 3	77.84	22.16		

TABLE 4.9 – Les vecteurs propres ($u1$ et $u2$) de la matrice de variance/covariance génétique obtenus avec le modèle 2 qui suppose que la matrice génétique est de rang 2

	Lait		MG		MP	
	u1	u2	u1	u2	u1	u2
niveau génétique moyen	**0.9904**	0.0392	**0.9769**	-0.0297	**0.9850**	0.0421
spécialité de la conduite	-0.0701	0.1949	-0.0692	0.3463	-0.0734	0.3287
intensité de la conduite	0.1178	-0.3380	0.1944	0.0057	0.1502	-0.3595
stabilité de la conduite	0.0159	**0.9199**	0.0553	**0.9376**	0.0420	**0.8723**
	SCS		IVIA1		ResIA	
	u1	u2	u1	u2	u1	u2
niveau génétique moyen	**0.9991**	-0.0295	0.3703	**0.6947**	**0.9913**	-0.0988
spécialité de la conduite	-0.0152	-0.3202	**0.6703**	0.1520	0.0348	**0.8333**
intensité de la conduite	-0.0177	0.3831	0.2350	0.1935	-0.0477	-0.1833
stabilité de la conduite	0.0363	**0.8659**	**-0.5985**	**0.6760**	0.1176	0.5121

TABLE 4.10 – Bayesian Information Criterion obtenu pour chaque modèle et chaque caractère

Lait	modèle 1	455030.16
	modèle 2	455001.41
	modèle 3	**454991.38**
MG	modèle 1	260334.36
	modèle 2	260303.89
	modèle 3	**260295.02**
MP	modèle 1	246036.36
	modèle 2	246009.13
	modèle 3	**246004.99**
SCS	modèle 1	43388.84
	modèle 2	43358.03
	modèle 3	**43330.51**
IVIA	modèle 1	232575.43
	modèle 2	**232545.31**
	modèle 3	232656.54
ResIA	modèle 1	-5182.47
	modèle 2	-5671.22
	modèle 3	**-5703.92**

TABLE 4.11 – Les corrélations génétiques obtenues entre environnements (minimum, moyenne et maximum) par caractère

	min	moy	max
Lait	0.83	0.97	1
MG	0.93	0.99	1
MP	0.85	0.97	1
SCS	0.87	0.97	1
IVIA1	0.01	0.77	1
ResIA	0.43	0.88	1

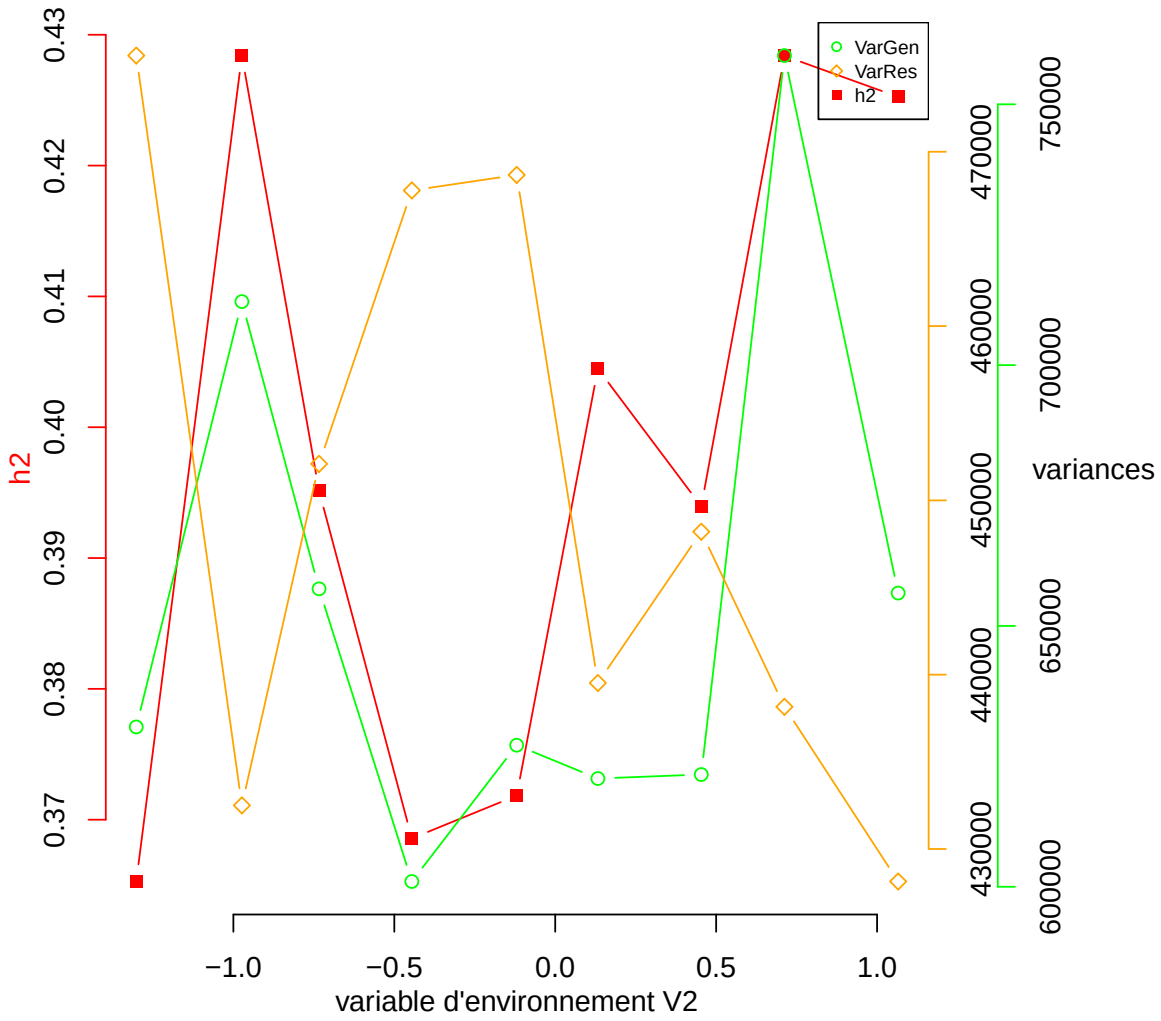


FIGURE 4.2 – Héritabilité (h2), variance résiduelle et variance génétique en fonction de l'environnement pour le caractère Lait

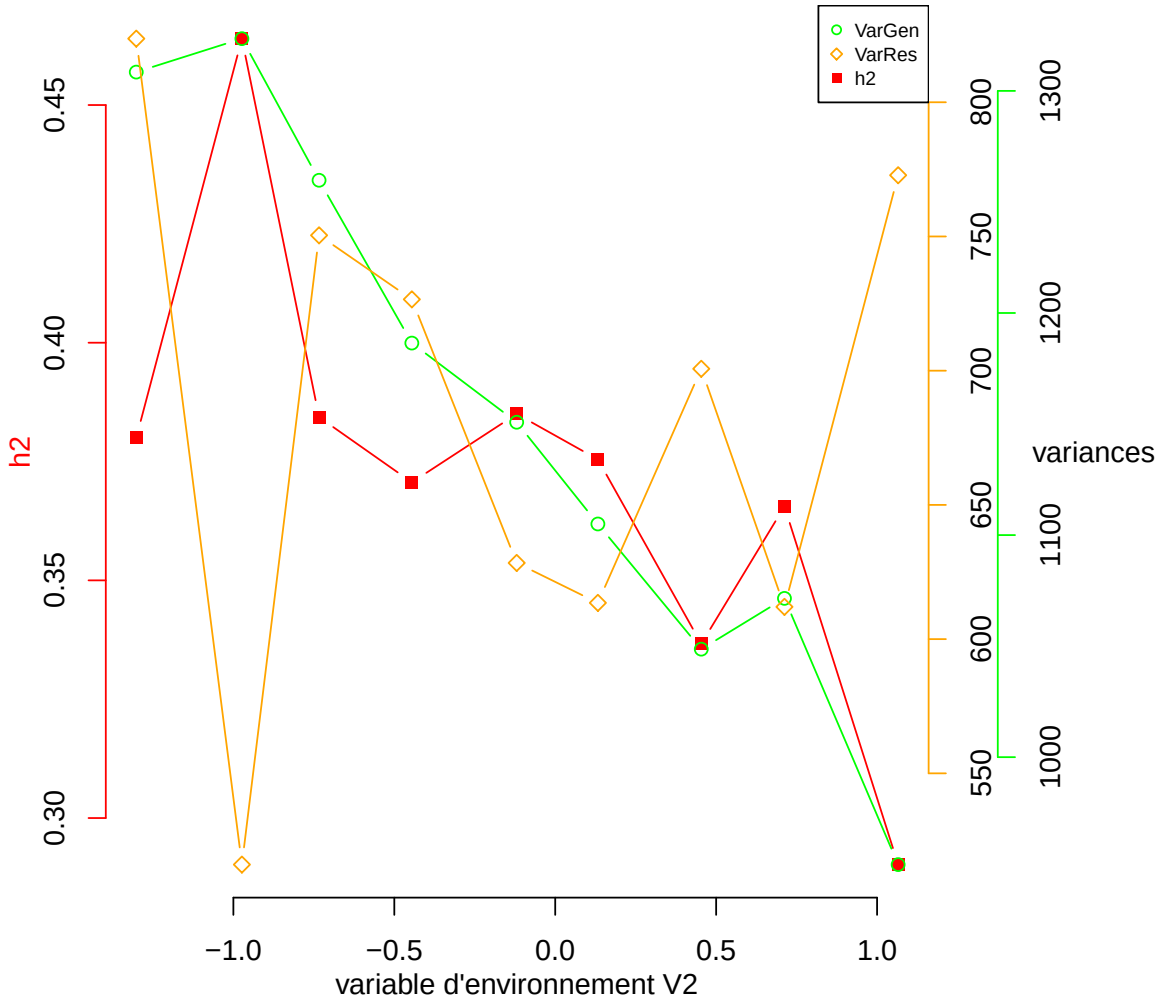


FIGURE 4.3 – Héritabilité (h2), variance résiduelle et variance génétique en fonction de l'environnement pour le caractère Matière Grasse (MG)

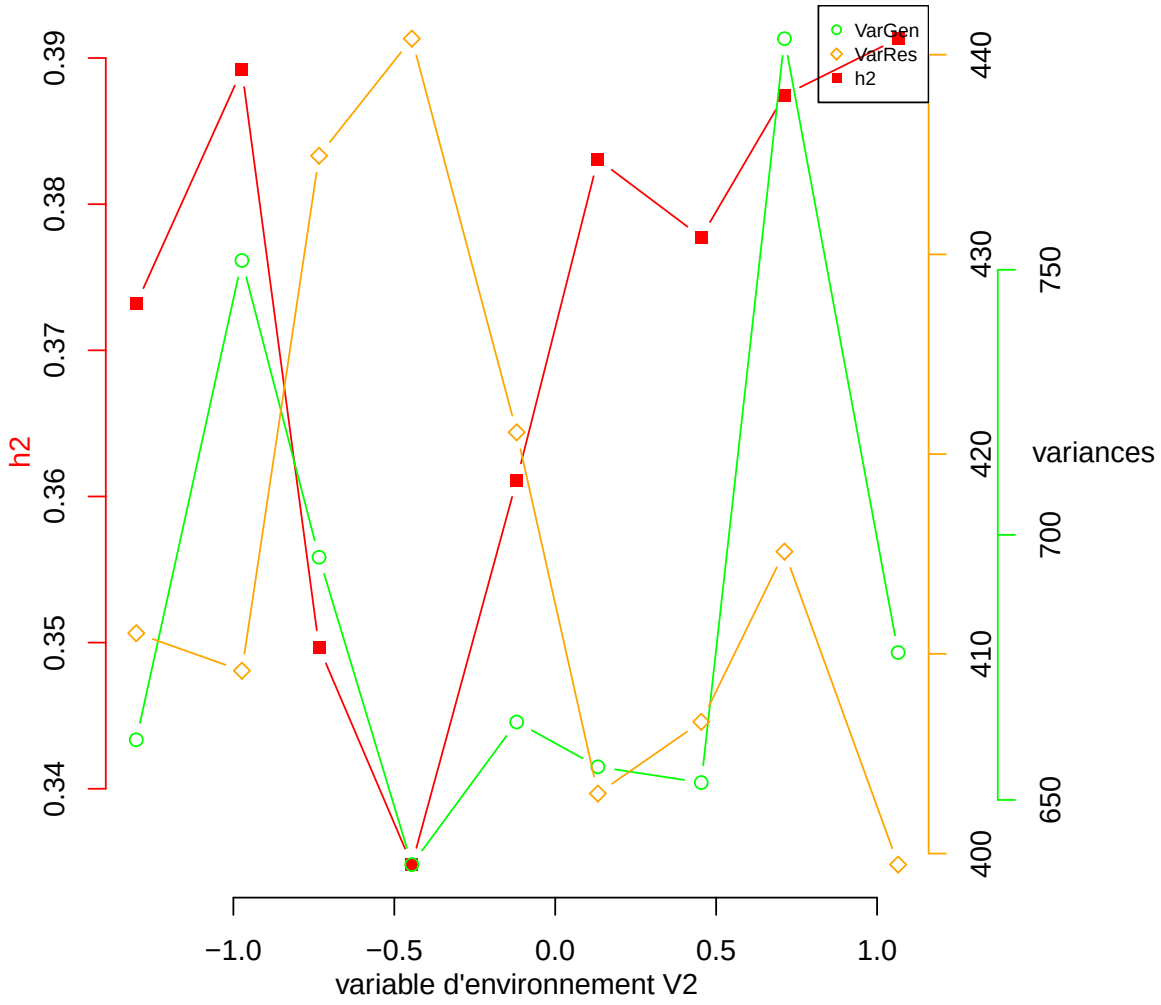


FIGURE 4.4 – Héritabilité (h2), variance résiduelle et variance génétique en fonction de l'environnement pour le caractère Matière Protéique (MP)

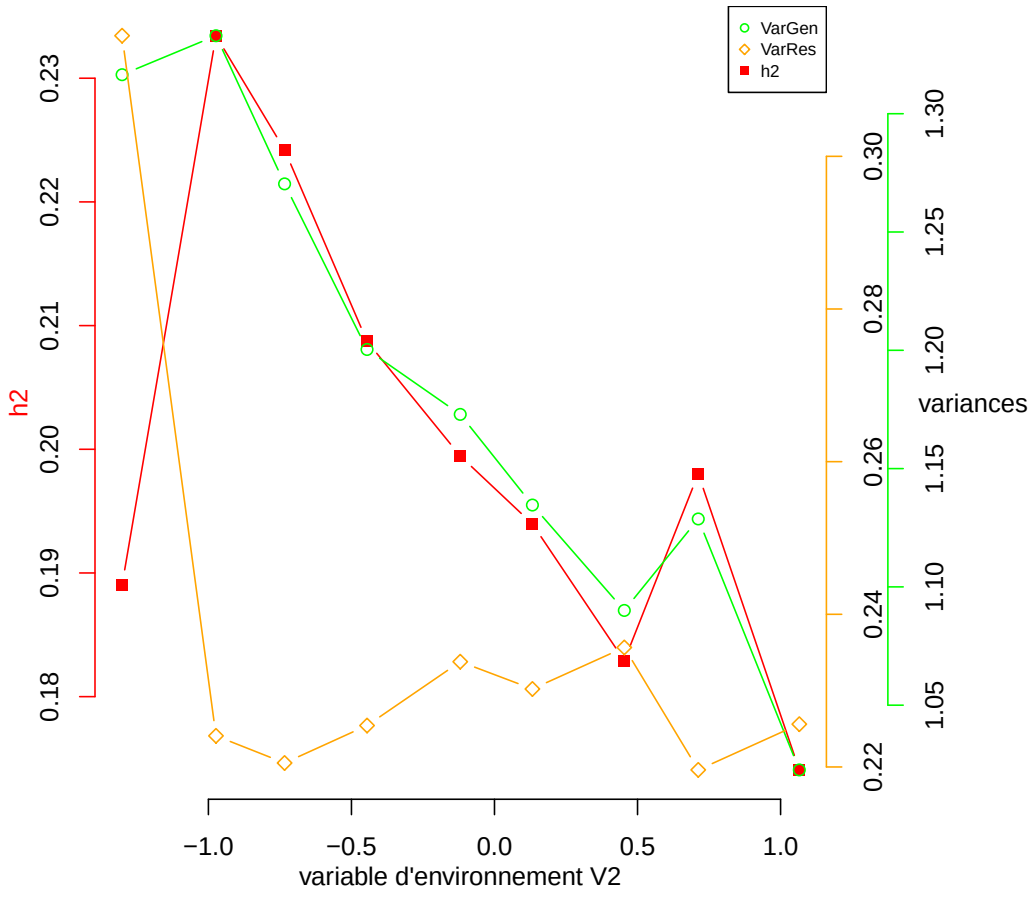


FIGURE 4.5 – Héritabilité (h2), variance résiduelle et variance génétique en fonction de l'environnement pour le caractère Score de Cellules Somatiques (SCS)

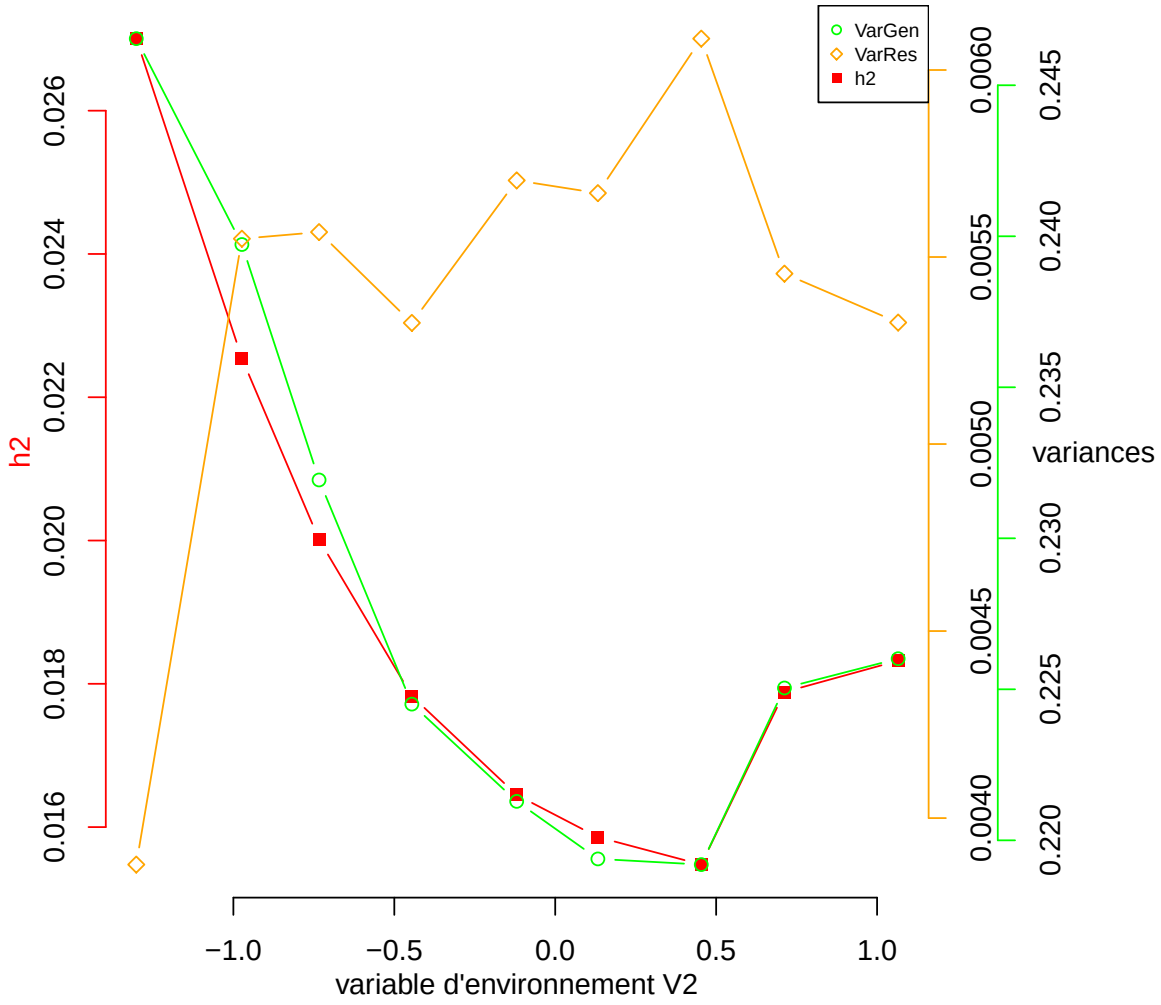


FIGURE 4.6 – Héritabilité (h2), variance résiduelle et variance génétique en fonction de l'environnement pour le caractère réussite à la première insémination artificielle (ResIA)

4.7.5 Discussion/Conclusion

Cette étude présente les même limites que la précédente : l'utilisation d'une définition du milieu issue d'un modèle négligeant les interactions Génotype*Milieu et qui ne prend en compte que les profils TJC de caractères laitiers, l'utilisation d'un modèle lactation, une relation linéaire entre le milieu et les valeurs génétiques qui pourrait exacerber les effets d'échelle. Cependant, une amélioration par rapport à l'étude précédente : la prise en compte d'une variance résiduelle hétérogène.

Cette étude a mis en évidence un reclassement pour la réussite à la première insémination artificielle. En revanche, aucun reclassement n'a été montré pour le lait, la matière grasse, la matière protéique et le score cellulaire. Pour ces caractères, des interactions Génotype*Milieu ont été trouvées sous forme d'une héritabilité qui dépend d'une combinaison linéaire des 3 composantes principales du milieu sachant que le poids de chaque composante dans cette combinaison linéaire dépend du caractère. Cependant, cette étude a été réalisée sur un faible nombre d'élevages et uniquement en race Montbéliarde. Les résultats demandent à être confirmés sur un plus grand nombre d'élevages, la prochaine étude devra aussi être réalisée sur les races Normande et Holstein.

L'intérêt principal de cette étude est son aspect méthodologique. Plusieurs modèles de normes de réaction ont été comparés, ils prennent en compte différemment l'environnement dans la mesure de l'effet génétique. Il a été montré que la modélisation de l'effet génétique, en fonction d'une combinaison linéaire non fixée a priori des descripteurs de l'environnement, est possible en imposant à la matrice de variance/covariance des composantes de l'effet génétique (un effet moyen et des normes de réaction) à être de rang réduit (de rang 2 dans cette étude). La combinaison linéaire qui détermine la nouvelle variable d'environnement dépend du caractère étudié et n'est pas déterminée par l'utilisateur. Ce type d'approche n'avait, jusqu'à présent, jamais été testé. Il a été montré que c'est principalement l'effet de l'hétérogénéité de la production permise par la conduite de troupeau au cours d'une année qui entre en compte pour décrire les interactions Génotype*Milieu pour les caractères laitiers et le Score de Cellules Somatiques alors qu'il s'agit de l'intensité de la conduite pour la réussite à la première insémination artificielle. L'utilisation de ces variables synthétiques est intéressante mais complique l'interprétation des résultats.

De plus, le modèle 3 est meilleur que le modèle 1 selon le critère BIC. Cela montre que l'effet de l'environnement sur l'effet de la génétique est mieux révélé selon un seul axe

synthétisant différentes informations sur l'environnement (les 3 premières composantes principales) que selon différents axes (les composantes principales elles-même), bien que l'information contenue dans ces axes semble la même.

Dans cette étude, seules 3 variables d'environnement ont été résumées en 1 variable synthétique ($V2$). Cependant, elle ouvre la voie à de nouvelles études prenant en compte éventuellement un nombre plus élevé de variables d'environnement dont l'information serait résumée à travers la contrainte de rang réduit de la matrice génétique. Dans notre étude par exemple, on pourrait envisager d'intégrer conjointement les 10 premières composantes principales dans la définition de l'effet génétique.

Chapitre 5

Discussion générale

5.1 Les enjeux des interactions Génotype*Milieu à l'échelle nationale

Les études des interactions Génotype*Milieu tendent à mettre en évidence un reclassement des animaux selon le milieu (les meilleurs animaux ne sont pas les mêmes selon le milieu dans lequel ils évoluent) ou un effet d'échelle, c'est-à-dire que le classement des animaux est le même quel que soit le milieu mais les valeurs génétiques sont plus variables dans un milieu que dans un autre. Les impacts des interactions Génotype*Milieu ont déjà été développés dans la thèse de Mulder (2007). Les principaux sont résumés ici.

Si les interactions Génotype*Milieu sont négligées à tort, les schémas de sélection actuels ne sont pas optimaux, le progrès génétique réel est surestimé et la définition des index de synthèse actuels pourrait ne pas être la plus appropriée. En effet, les corrélations génétiques entre caractères pourraient ne pas être les mêmes selon les milieux. Cependant, les prendre en compte au niveau national entraînerait de nombreuses modifications. Toutes les chaînes d'évaluation génétique concernées par ces interactions devraient être adaptées, parfois lourdement, pour les inclure. De nouvelles questions surgiraient : les schémas de sélection devraient-ils être modifiés pour avoir un schéma adapté à chaque type d'environnement ? Comment la diffusion des valeurs génétiques devrait-elle être adaptée ? Faudrait-il multiplier le nombre de catalogues de taureaux en fonction des systèmes d'élevage existants ? L'ensemble de ces modifications auraient-elles des conséquences sur le fonctionnement des Entreprises de Sélection dans un contexte très concurrentiel ? Se

focaliseraient-elles sur un type de système ou se diversifieront-elles ? Quel serait l'impact sur le prix des semences de taureaux, la consanguinité de la population bovine ? Il est donc essentiel de trouver un compromis réaliste entre avantages et inconvénients résultant de la prise en compte des interactions Génotype*Milieu.

5.2 Mise en évidence d'un effet d'échelle en France

Cette thèse apporte un nouveau point de vue sur la manière de définir l'environnement dans l'étude des interactions Génotype*Milieu via l'utilisation des profils Troupeau-Jour de Contrôle. Dans l'article II, la méthodologie mise en place dans l'article I pour définir le milieu à partir des profils TJC a été appliquée sur 15000 élevages environ (vs 450 élevages dans l'article I). Les composantes principales qui synthétisent les informations portées par les profils TJC ne sont pas exactement les mêmes dans les 2 études puisqu'elles n'ont pas été construites sur le même jeu de données mais elles montrent les mêmes tendances.

Les premières composantes principales des profils TJC, c'est-à-dire celles qui discriminent au mieux les élevages, définies dans l'article II à partir d'un grand nombre d'élevage sont les suivantes :

- une composante représentative de la « spécialisation de la conduite » : la première composante principale oppose les élevages dont la conduite favorise plutôt la production laitière à ceux dont la conduite favorise des taux protéique et butyreux élevés.
- une composante représentative de l' « intensité de la conduite » : la deuxième composante principale oppose les élevages dont la conduite permet de produire (du lait et des taux) plus ou moins que la moyenne.
- une composante représentative de la « saisonnalité de la conduite » : la troisième composante principale oppose les élevages dont la conduite permet de produire de façon homogène sur l'année à ceux dont la conduite favorise la production pendant une période particulière de l'année.

Ces composantes principales ont été utilisées telles quelles comme variables d'environnement dans les modèles de normes de réaction de l'article II.

Elles ont aussi été à la base de la création des 3 groupes d'élevages qui illustrent 3 types de milieu contrastés dans le modèle multicaractères (article II) :

- un groupe d'élevages dont la conduite a une intensité moyenne plutôt spécialisée en lait
- un groupe d'élevages dont la conduite est intensive mais plutôt dirigée vers des taux protéique et butyreux élevés
- un groupe d'élevages dont la conduite est peu intensive.

Outre l'introduction d'une innovation dans la définition du milieu, cette thèse répond aux interrogations des éleveurs français sur la pertinence des schémas de sélection. En ce qui concerne la mesure des interactions Génotype*Milieu, aucun reclassement n'a été mise en évidence quelle que soit la race (Normande, Montbéliarde, Holstein) à part pour la réussite à la première insémination artificielle en Montbéliarde. En revanche, un effet d'échelle plus ou moins important selon les races et les caractères existe.

La présence d'un effet d'échelle dans les études des interactions Génotype*Milieu réalisées dans différents pays a souvent été montrée (Raffrenato *et al.*, 2003; Haile-Mariam *et al.*, 2008). Cependant, la définition du milieu utilisée dans cette thèse ne facilite pas l'exacerbation de la présence d' interactions Génotype*Milieu. En effet, la définition du milieu est basée sur les profils TJC qui sont le coproduit d'un modèle génétique qui néglige ces interactions. Cette définition du milieu est donc une approximation tout comme celle qui consiste à définir le milieu à partir de performances brutes moyennes. De plus, les profils TJC sont issus d'un modèle génétique basé sur les contrôles élémentaires mais les interactions Génotype*Milieu ont été étudiées via un modèle lactation.

L'effet d'échelle ne sera pas pris en compte dans les modèles d'évaluation génétique, pourtant il serait utile que les éleveurs et leurs conseillers soient conscients qu'un animal avec une valeur génétique donnée ne donnera pas les mêmes performances selon le milieu dans lequel il évolue. Un indicateur pourrait être imaginé pour qu'un éleveur évalue la variation (positive ou négative) de performance à laquelle il peut s'attendre sachant la valeur génétique de l'animal choisi et le milieu dans lequel il produit. La mise en place d'un tel indicateur doit prendre en compte plusieurs contraintes comme la nécessité de simplicité de l'outil et la robustesse de la définition du milieu. En effet, les typologies présentées dans le chapitre 3 pourraient permettre de définir à quel milieu appartient

chaque élevage mais cette typologie doit être robuste pour qu'un élevage ne change pas de milieu au cours du temps et doit être mise à jour régulièrement pour prendre en compte les nouveaux élevages. De plus, cet indicateur devra permettre d'associer une valeur génétique à un écart de performances attendu pour les différents caractères évalués.

5.3 Mise en évidence de reclassements au niveau international dans la bibliographie

Au niveau international, des reclassements d'animaux entre pays ont été mis en évidence. Cela ne pose pas de problème majeur pour la commercialisation des semences entre pays membres d'Interbull (International Bull Evaluation Service) puisque l'évaluation internationale prend en compte des corrélations génétiques différentes de 1 entre les pays. Par conséquent, le classement des animaux est différent selon les pays. En revanche, cela peut avoir des conséquences lors de l'exportation de semences d'un pays membre d'Interbull vers un pays non membre. Des taureaux considérés comme les meilleurs aux Etats Unis ou en Europe peuvent ne pas être les meilleurs ou ne pas donner les performances espérées s'ils sont utilisés dans des pays importateurs en Afrique ou en Amérique du Sud qui ne sont pas membres d'Interbull. En effet, il n'existe pas de classement national pour ces pays (ex : le Kenya (Ojango & Pollott, 2002) ou la Tunisie (Hammami *et al.*, 2008)). Les pays non membres pourraient se baser sur le classement d'autres pays qui leur semble le plus proche de leur contexte local. Cependant, ce n'est pas toujours évident. A titre d'exemple, seule l'Afrique du Sud est membre d'Interbull sur le continent africain.

Des possibilités d'amélioration de l'évaluation internationale ont été proposées par Weigel & Rekaya (2000); Weigel *et al.* (2001). Elles consistent à réaliser une évaluation internationale selon des milieux plutôt que selon les pays. Une combinaison de variables facilement disponibles au niveau international était proposée pour décrire le milieu. Cependant, la mise en place d'un tel modèle est complexe. Il relève d'une grande complexité informatique car Interbull devrait avoir accès à l'élevage d'appartenance de chaque animal pour pouvoir l'affecter à un milieu. De plus, la définition des effets Troupeau*Année*Saison commune à l'ensemble des élevages peut devenir non pertinente. En effet, pour estimer cet effet, le nombre de performances par modalité doit être suffisant, cette condition n'est pas facilement réalisable dans les troupeaux de faible effectif. Pour

obtenir l'effectif nécessaire, la définition de la saison pourrait être modifiée en fonction de la taille de l'élevage mais cela rendrait la définition de l'effet Troupeau*Année*Saison incohérente. L'évaluation génétique internationale considère des corrélations génétiques non égales à 1 entre pays ce qui permet de prendre en compte le fait que les performances puissent être collectées de façons différentes et traiter par des modèles génétiques différents entre les pays. Cette contrainte ne pourrait pas être prise en compte dans le modèle proposé par Weigel & Rekaya (2000). Enfin, même si techniquement il serait intéressant de faire une évaluation internationale basée sur des systèmes d'élevage plutôt que sur des pays, ce concept pourrait être mal accepté sur le terrain. Le concept d'évaluation génétique internationale basée sur des systèmes d'élevage plutôt que sur des pays est donc un défi demandant beaucoup de travail pour être relevé. Or, à l'heure actuelle, Interbull met plutôt l'accent sur toutes les valorisations possibles des nouvelles données moléculaires (via les puces) dans les évaluations génétiques nationales et internationales.

Chapitre 6

Conclusion générale

Les profils Troupeau Jour de Contrôle, un des co-produits du modèle d'évaluation génétique basé sur les contrôles élémentaires (Leclerc, 2008) sont valorisés dans le projet GENESYS à des fins de conseil en élevage. Dans cette thèse, ils ont été valorisés pour définir le milieu dans les études sur les interactions Génotype*Milieu (article I). Ils ont l'avantage d'être disponibles dans les bases de données nationales. De plus, contrairement à d'autres définitions du milieu utilisées dans la bibliographie, les profils Troupeau Jour de Contrôle se focalisent uniquement sur la part de la production permise par les conditions d'élevage (c'est à dire corrigée des effets génétiques, de l'âge au vêlage...).

Les interactions Génotype*Milieu ont été étudiées en utilisant un modèle multicaractères et des modèles de normes de réaction. Des interactions Génotype*Milieu ont été mises en évidence en France sous forme d'effet d'échelle pour les caractères de production laitière et le score cellulaire (article II et étude complémentaire). Elles ne remettent pas en cause la pertinence des schémas de sélection actuels. Une approche innovante consistant à modéliser l'effet génétique en fonction de l'environnement en contraignant la matrice génétique à être de rang réduit a été proposée. Elle permet de prendre en compte, au sein du modèle génétique, un plus grand nombre de variables pour décrire l'environnement tout en les résumant. L'effet génétique est alors modélisé en fonction d'un seul axe qui résume toutes les variabilités de l'environnement. Cela permet aussi de diminuer le nombre de paramètres à estimer dans le modèle.

La plupart des études montrent la présence d'interaction Génotype*Milieu sous forme d'effet d'échelle dans des environnements peu variables et des reclassements dans les études

internationales. Cependant, rien n'est pour l'instant mis en place pour prendre en compte ces résultats au niveau national. Cela semble demander un coût trop important ou avoir des conséquences trop lourdes (comme la modification des schémas de sélection) par rapport aux bénéfices qu'apporterait leur intégration dans les modèles génétiques (c'est-à-dire des valeurs génétiques d'animaux prenant en compte la plasticité phénotypique de animaux). Cela peut aussi s'expliquer par le fait qu'aucun reclassement n'avait été mis en évidence et que les hétérogénéités de variance résiduelle sont déjà prises en compte (bien que l'héritabilité soit considérée constante). Par conséquent, les impacts des interactions Génotype*Milieu sur les schémas de sélection sont faibles.

Malgré le grand nombre d'études réalisées sur ce thème et la relative concordance des résultats obtenus, la question de la présence d'interaction Génotype*Milieu semble être toujours en suspens. Les études réalisées en ferme expérimentale bénéficient d'un environnement très contrôlé mais les résultats peuvent souvent être considérés comme non extrapolables à une situation plus générale. Au contraire, les études réalisées à grande échelle ne répondent pas aux contextes locaux ou au cas de systèmes particuliers. Jusqu'à présent, le développement d'une nouvelle méthode comme l'utilisation des normes de réaction dans la thèse de Kolmodin (2003), d'une nouvelle définition du milieu comme l'utilisation des profils Troupeau Jour de Contrôle dans cette thèse ou un questionnement local comme dans la thèse d'Hammami (2009) (qui s'intéresse aux interactions Génotype*Milieu entre le Luxembourg et la Tunisie) ou dans celle de Nauta (2009) (qui s'intéresse plus particulièrement aux élevages en agriculture biologique) ont été la source de nouvelles études sur les interactions Génotype*Milieu. Elles avaient pour objectif de donner une réponse plus claire ou plus adaptée à une demande locale que celles données par les études précédentes. Aujourd'hui, le paysage agricole étant en constante mutation, de nouveaux systèmes d'élevage peuvent voir le jour à l'avenir, ce qui posera à nouveau la question des interactions Génotype*Milieu. La question est alors la suivante : « Quelles pourraient-être les nouvelles innovations dans le cadre de l'étude des interactions Génotype*Milieu ? ». Elles pourraient naître des nouveaux indicateurs de milieu apportés par l'élevage de précision ou de la connaissance de nouvelles données moléculaires.

A l'avenir, de plus en plus d'informations seront disponibles sur les phénotypes mais aussi sur l'environnement dans lequel ils sont mesurés. Cette évolution est née d'avancées technologiques qui ont permis le développement de l'agriculture de précision. La notion

d'agriculture de précision a d'abord été développée dans le secteur végétal où de nouveaux outils tel que le GPS (Global Positioning System) ont permis de prendre en compte l'existence de variabilités intra-parcellaires, comme dans la gestion des intrants. La notion d'élevage de précision est apparue plus récemment ; le premier congrès européen sur l'élevage de précision a eu lieu en 2003 alors que le premier congrès européen sur l'agriculture de précision avait eu lieu en 1997 (http://www.ecplf2011.cz/). Actuellement, le nombre de capteurs d'information utilisés en ferme augmente. Ils sont de natures diverses : analyseur de lait, mesure de progestérone dans le lait, podomètres... Ils sont utilisés dans des outils d'aide à la décision pour une meilleure gestion de l'alimentation, de la reproduction et des problèmes sanitaires. En effet, l'élevage de précision peut être un moyen *« d'épanouir les différences entre animaux ou de prendre en compte les besoins personnalisés des animaux »* (Berducat, 2007). Les outils développés dans le cadre de l'élevage de précision apportent une mine d'informations qui pourrait, à terme, être utilisée en génétique pour une meilleure description de l'environnement. A l'heure actuelle, ces données ne sont disponibles que pour les fermes équipées. Le développement de l'agriculture de précision est freiné par des contraintes économiques et techniques : le coût du capteur doit être couvert par les bénéfices apportés par ses produits, l'utilisation d'un capteur nécessite une bonne identification des animaux (l'utilisation du système RFID Radio Frequency IDentification pour l'identification des animaux est actuellement optionnelle en bovins laitiers alors qu'elle est obligatoire en petits ruminants), le système de gestion de l'exploitation doit pouvoir gérer les informations générées par les différents capteurs présents sur l'exploitation... En acceptant que les freins à l'utilisation de ces capteurs soient un jour levés, ces outils d'aide à la décision pourraient se généraliser et les informations qu'ils génèrent pourraient être remontées dans le Système d'Information Génétique et utilisées dans de nouvelles études à grande échelle des interactions Génotype*Milieu.

Les avancées, sources de nouvelles études sur les interactions Génotype*Milieu, pourraient venir de l'amélioration de la définition du milieu mais aussi du génotype. Des données moléculaires sont aujourd'hui disponibles pour un grand nombre d'animaux grâce aux génotypages réalisés en vue de la réalisation et de l'utilisation des évaluations génomiques. Elles pourraient donner naissance à une nouvelle génération d'études des interactions Génotype*Milieu où la définition du génotype serait basée sur les données moléculaires : les SNP (Single Nucleotide Polymorphism). L'étude des interactions QTL (Quantitative

Trait Loci)*Milieu a déjà été commencé en bovins laitiers (Lillehammer *et al.*, 2009) et chez les végétaux (Li *et al.*, 2003), elle prendrait une toute autre ampleur avec l'utilisation des SNP. Au delà d'une nouvelle définition du génotype, les données moléculaires offrent une nouvelle opportunité pour mieux comprendre le phénomène de plasticité phénotypique (ou sensibilité à l'environnement) au niveau moléculaire. En effet, derrière l'objectif partagé par l'ensemble des chercheurs en génétique quantitative qui consiste à mesurer les interactions Génotype*Milieu, se cache la question du phénomène biologique qui est à l'origine de ces interactions : la sensibilité à l'environnement. La compréhension de la modification de l'expression génétique en fonction de l'environnement n'est pas totale. Scheiner & Lyman (1991) repris par Bodin *et al.* (2010) résument les modèles proposés : le modèle de surdominance, le modèle pléiotropique et le modèle épistatique. Le modèle de surdominance suppose que la sensibilité à l'environnement dépend du nombre de locus hétérozygotes mais ce modèle est peu souvent mis en avant (Scheiner, 1993). Les deux autres modèles ont été plus souvent cités. Le modèle pléiotropique suppose que la sensibilité à l'environnement est due à une expression différente, dans différents environnements, des mêmes gènes alors que le modèle épistatique suppose qu'il existe des gènes qui contrôlent l'expression moyenne du caractère et d'autres gènes qui contrôlent l'expression des premiers en fonction de l'environnement. Ces mécanismes ont par exemple été mis en évidence par Wu (1997). Les nouvelles données moléculaires représentent donc une piste pour mieux comprendre ses différents modèles.

Ces éléments nouveaux montrent que, si cette thèse fait suite à une longue liste de travaux internationaux sur les interactions Génotype*Milieu, elle n'est certainement pas la dernière ! Elle apporte simplement une contribution à l'étude et la compréhension des mécanismes conduisant à une expression différenciée des caractères d'intérêt chez les bovins laitiers en fonction du milieu dans lequel ils sont mesurés.

Bibliographie

1999 (Mars). *Note Indexation Bovine Laitière : La prise en compte des variances hétérogènes dans l'indexation laitière*. Tech. rept. 99-2. Institut de l'Elevage, Institut National de Recherche Agronomique. 91

2005 (Juin). *Note Indexation Bovine Laitière : La prise en compte des variances hétérogènes dans l'évaluation génétique de la morphologie*. Tech. rept. 2005-5. Institut de l'Elevage, Institut National de Recherche Agronomique. 91

2012 (Février). *Note Indexation Bovine Laitière : L'ISU 2012 des races laitières nationales*. Tech. rept. 2012-02. Institut de l'Elevage, Institut National de Recherche Agronomique. 127

Agridea. *Sélection vaches laitières*. http://www.agridea-lausanne.ch/files/11.3.11-18.pdf. [en ligne] 28 juin 2012. 21

AgroParisTech. *Les races bovines françaises*. http://www.agroparistech.fr/svs/genere/especes/bovins.htm. [en ligne] 21 Juillet 2010. 21

Bapst, B., & Bär, M. 2001. *La valeur d'élevage globale écologique (VEGE). Un grand pas vers la sélection d'une vache bio*. Tech. rept. FiBL, BIO SUISSE, SVKB, SBZV. https://www.fibl-shop.org/shop/pdf/mb-vege.pdf [en ligne] 8 Novembre 2010. 22

Barbin, G., M., Chaumet J., P., Chotteau, C., Guesdon J., B., Lelyon, C., Monniot, A., Mottet, & G., You. 2009. *France Laitière 2015, Vers une accentuation des contrastes régionaux*. Tech. rept. Institut de l'Elevage, http://idele.fr/recherche/publication/idelesolr/recommends/france-laitiere-2015.html. 19

Berducat, M. 2007. *Capteurs et echnologies au service de l'elevage de précision*. Tech. rept.

Cemagref Cetim. [en ligne, consulté le 23 juillet 2012] ftp://ftp.clermont.cemagref.fr/pub/Applications/Cemagref-NoteCetim_ECPA2007_elevage.pdf. 151

Berry, D. P., Buckley, F., Dillon, P., Evans, R. D., Rath, M., & Veerkamp, R. F. 2003. Genetic parameters for body condition score, body weight, milk yield, and fertility estimated using random regression models. *J Dairy Sci*, **86**(11), 3704–3717. 31, 33

Blöttner, S., Heins, B. J., Wensch-Dorendorf, M., Hansen, L. B., & Swalve, H. H. 2011a. Brown Swiss * Holstein crossbreds compared with pure Holsteins for calving traits, body weight, backfat thickness, fertility, and body measurements. *J Dairy Sci*, **94**(2), 1058–1068. 23

Blöttner, S., Heins, B. J., Wensch-Dorendorf, M., Hansen, L. B., & Swalve, H. H. 2011b. Short communication : A comparison between purebred Holstein and Brown Swiss * Holstein cows for milk production, somatic cell score, milking speed, and udder measurements in the first 3 lactations. *J Dairy Sci*, **94**(10), 5212–5216. 23

Bodin, L., Bolet, G., Garcia, M., Garreau, H., Larzul, C., & David, I. 2010. Robustesse et canalisation : vision de généticiens. *INRA Productions Animales*, **23**, 11–22. 20, 23, 152

Boettcher, P. J., Fatehi, J., & Schutz, M. M. 2003. Genotype x environment interactions in conventional versus pasture-based dairies in Canada. *J Dairy Sci*, **86**(1), 383–389. 31

Bonaiti, B. 1982. Interaction génotype milieu sur la production laitière chez les bovins II.- Corrélation génétique entre niveaux de production et variation de l'héritabilité. *Pages 441–452 of : Annales de Génétique et de Sélection Animale*, vol. 14. 33, 91

Bougoin, M-H., & Le Mezec, P. 2010. Le croisement entre races laitières en France vu par les bases de données. *Page 464 of : Rencontre Recherche Ruminants*, vol. 17. Institut de l'Elevage-INRA. 22

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. 1976. *Time series analysis*. Holden-day San Francisco. 74, 75

Brunschwig, P., Picard, S., & Ballot, N. 2011 (juillet). *Observatoire de l'alimentation des vaches laitières*. http://idele.fr/recherche/publication/idelesolr/

recommends/observatoire-de-lalimentation-des-vaches-laitieres.html. 19, 77

Buckley, F., Dillon, P., Rath, M., & Veerkamp, R. F. 2000. The relationship between genetic merit for yield and live weight, condition score, and energy balance of spring calving Holstein Friesian dairy cows on grass based systems of milk production. *J Dairy Sci*, **83**(8), 1878–1886. 31, 32, 34, 83

Calus, M. P. L. 2006. *Estimation of genotype * environment interaction for yield, health and fertility in dairy cattle*. Ph.D. thesis, Wageningen University. 31

Calus, M. P. L., & Veerkamp, R. F. 2003. Estimation of environmental sensitivity of genetic merit for milk production traits using a random regression model. *J Dairy Sci*, **86**(11), 3756–3764. 88, 89

Calus, M. P. L., Groen, A. F., & De Jong, G. 2002. Genotype x environment interaction for protein yield in Dutch dairy cattle as quantified by different models. *J Dairy Sci*, **85**(11), 3115–3123. 32, 33

Calus, M. P. L., Windig, J. J., & Veerkamp, R. F. 2005. Associations among descriptors of herd management and phenotypic and genetic levels of health and fertility. *J Dairy Sci*, **88**(6), 2178–2189. 33

Carabaño, M. J., Van Vleck, L. D., Wiggans, G. R., & Alenda, R. 1989. Estimation of genetic parameters for milk and fat yields of dairy cattle in Spain and the United States. *J Dairy Sci*, **72**(11), 3013–3022. 29

Carabaño, M. J., Wade, K. M., & Van Vleck, L. D. 1990. Genotype by environment interactions for milk and fat production across regions of the United States. *J Dairy Sci*, **73**(1), 173–180. 30

Castillo-Juarez, H., Oltenacu, P. A., Blake, R. W., Mcculloch, C. E., & Cienfuegos-Rivas, E. G. 2000. Effect of herd environment on the genetic and phenotypic relationships among milk yield, conception rate, and somatic cell score in Holstein cattle. *J Dairy Sci*, **83**(4), 807–814. 33

Castillo-Juarez, H., Oltenacu, P. A., & Cienfuegos-Rivas, E. G. 2002. Genetic and phenotypic relationships among milk production and composition traits in primiparous

Holstein cows in two different herd environments. *Livestock production science*, **78**(3), 223–231. 33

Cerón-Muñoz, M. F., Tonhati, H., Costa, C. N., Rojas-Sarmiento, D., & Echeverri Echeverri, D. M. 2004a. Factors that cause genotype by environment interaction and use of a multiple-trait herd-cluster model for milk yield of Holstein cattle from Brazil and Colombia. *J Dairy Sci*, **87**(8), 2687–2692. 32, 34

Cerón-Muñoz, M. F., Tonhati, H., Costa, C. N., Maldonado-Estrada, J., & Rojas-Sarmiento, D. 2004b. Genotype x environment interaction for age at first calving in Brazilian and Colombian Holsteins. *J Dairy Sci*, **87**(8), 2455–2458. 29

Charagu, P., & Peterson, R. 1998. Estimates of GxE effects for economic efficiency among daughters of Canadian and New Zealand sires in Canadian and New Zealand dairy herds. *Interbull Bulletin*, **17**(17), 105–109. 29

Charroin, T., Palazon, R., Madeline, Y., Guillaumin, A., & Tchakerian, E. 2005. Le système d'information des Réseaux d'Elevage françaises sur l'approche globale de l'exploitation. Intérêt et enjeux dans une perspective de prise en compte de la durabilité. *Rencontres autour des recherches sur les ruminants*, 335–338. 32

Chatellier, V., & Pflimlin, A. 2007. Les systèmes laitiers des régions européennes de l'Espace Atlantique. Eléments de réflexion économique complémentaire au projet « Green Dairy ». *In : Colloque « systèmes laitiers et environnement dans l'espace Atlantique »*. 19

Chatellier, V., Colson, F., Arnaud, F., Guesdon, J. C., Kempf, M., Legendre, J., & Perrot, C. 1997. La diversité des systèmes d'élevage bovin en France et leur contribution à la production de viande bovine. *INRA Prod. Anim*, **10**(3), 227–240. 19

Cienfuegos-Rivas, E. G., Oltenacu, P. A., Blake, R. W., Schwager, S. J., Castillo-Juarez, H., & Ruiz, F. J. 1999. Interaction between milk yield of Holstein cows in Mexico and the United States. *J Dairy Sci*, **82**(10), 2218–2223. 29

Conté, A. 2011. Quelle vache pour quel système ? *RéussirLait*, **250**(Sep), 114–115. 21

Costa, C. N., Blake, R. W., Pollak, E. J., Oltenacu, P. A., Quaas, R. L., & Searle, S. R. 2000. Genetic analysis of Holstein cattle populations in Brazil and the United States. *J Dairy Sci*, **83**(12), 2963–2974. 29, 90

Cromie, AR, Kelleher, DL, Gordan, FJ, & Rath, M. 1998. Genotype by environment interaction for milk production traits in Holstein-Friesian dairy cattle in Ireland. *In : Proceedings INTERBULL open management, 18-19 January, Rotorau, New Zealand.* 31

Cutullic, E., Delaby, L., Gallard, Y., & Disenhaus, C. 2010. L'effet de la stratégie d'alimentation sur la reproduction des vaches laitières varie selon la race et les différentes phases du cycle de reproduction. *Pages 149–152 of : Rencontre Recherche Ruminants*, vol. 17. Institut de l'Elevage-INRA. 83

David, I, Bouvier, F., François, D., Poivey, J.P., & Tiphine, L. 2011. Heterogeneity of variance components for preweaning growth in Romane sheep due to the number of lambs reared. *Genetics Selection Evolution*, **43**, 32. 92

De Saint Vaury, G. 2010. Réflexion et témoignage d'éleveur. *Pages 49–50 of : Journées techniques Elevage Biologique : Sélection Animale, Diversité Génétique & Agriculture Biologique.* Institut Technique de l'Agriculture Biologique. 20

De Veer, J. C., & Van Vleck, L. D. 1987. Genetic parameters for first lactation milk yields at three levels of herd production. *J Dairy Sci*, **70**(7), 1434–1441. 33, 90

Dedieu, B., Faverdin, P., Dourmad, J.Y., & Gibon, A. 2008. Système d'élevage, un concept pour raisonner les transformations de l'élevage. *INRA Productions animales*, **21**(1), 45–57. 18

Delignières, D. 2007 (09). *Analyse des séries temporelles.* Support de cours de la Faculté des Sciences du Sport et de l'Education Physique (Université Montpellier I) http://didier.delignieres.perso.sfr.fr/Supports-doc/Analysetemporelle.pdf. 75

Denčić, S., Mladenov, N., & Kobiljski, B. 2011. Effects of genotype and environment on breadmaking quality in wheat. *International Journal of Plant Production*, **5**, 71–82. 86

Dimov, G., Albuquerque, L. G., Keown, J. F., Vleck, L. D. Van, & Norman, H. D. 1995. Variance of interaction effects of sire and herd for yield traits of Holsteins in California, New York, and Pennsylvania with an animal model. *J Dairy Sci*, **78**(4), 939–946. 87

Dodenhoff, J., & Swalve, H.H. 1998. Heterogeneity of variances across regions of northern Germany and adjustment in genetic evaluation. *Livestock Production Science*, **53**(3), 225–236. 92

Dufumier, M. 1996. *Les projets de développement agricole : Manuel d'expertise*. KARTHALA Editions. 18

Ecole Supérieure d' Agriculture d'Angers. http://www.groupe-esa.com/projet-ur-urse-quels-sont-les-types-genetiques-de-vache-laitiere-les-plus-adaptes-a kjsp. [en ligne] 14 Novembre 2011. 21

Escofier, B., & Pagès, J. 2008. *Analyses factorielles simples et multiples*. 4 edn. Dunod. 48

Fatehi, J., Stella, A., Shannon, J. J., & Boettcher, P. J. 2003. Genetic parameters for feet and leg traits evaluated in different environments. *J Dairy Sci*, **86**(2), 661–666. 31

Fikse, W. F., Rekaya, R., & Weigel, K. A. 2003. Assessment of environmental descriptors for studying genotype by environment interaction. *Livestock production science*, **82**(2-3), 223–231. 33

Freeman, G. H. 1973. Statistical methods for the analysis of genotype-environment interactions. *Heredity*, **31**(3), 339–354. 86

Freyer, G., König, S., Fischer, B., Bergfeld, U., & Cassell, B. G. 2008. Invited review : crossbreeding in dairy cattle from a German perspective of the past and today. *J Dairy Sci*, **91**(10), 3725–3743. 23

Fulkerson, W. J., Davison, T. M., Garcia, S. C., Hough, G., Goddard, M. E., Dobos, R., & Blockey, M. 2008. Holstein-Friesian dairy cows under a predominantly grazing system : interaction between genotype and environment. *J Dairy Sci*, **91**(2), 826–839. 31

Haile-Mariam, M., Carrick, M. J., & Goddard, M. E. 2008. Genotype by environment interaction for fertility, survival, and milk production traits in Australian dairy cattle. *J Dairy Sci*, **91**(12), 4840–4853. 30, 32, 33, 145

Haldane, J. B. S. 1946. The interaction of nature and nurture. *Annals of Eugenics*, **13**(3), 197–205. 84

Hammami, H. 2009. *Genotype by Environment Interaction for Production Traits of Holsteins Using Two Countries as Model : Luxembourg and Tunisia*. Ph.D. thesis, Faculté universitaire des sciences agronomiques de Gembloux. 31, 150

Hammami, H., Rekik, B., Soyeurt, H., Bastin, C., Stoll, J., & Gengler, N. 2008. Genotype x environment interaction for milk yield in Holsteins using Luxembourg and Tunisian populations. *J Dairy Sci*, **91**(9), 3661–3671. 29, 146

Hammami, H., Rekik, B., Bastin, C., Soyeurt, H., Bormann, J., Stoll, J., & Gengler, N. 2009. Environmental sensitivity for milk yield in Luxembourg and Tunisian Holsteins by herd management level. *J Dairy Sci*, **92**(9), 4604–4612. 33

Hardarson, G. H. 2002. Is the modern high potential dairy cow suitable for organic farming conditions ? *Acta Veterinaria Scandinavica*, **43**(Suppl 1), S63. 22

Haskell, M. J., Brotherstone, S., Lawrence, A. B., & White, I. M. S. 2007. Characterization of the dairy farm environment in Great Britain and the effect of the farm environment on cow life span. *J Dairy Sci*, **90**(11), 5316–5323. 34, 87

Hastie, T.J., & Tibshirani, R.J. 1990. *Generalized additive models*. Chapman & Hall/CRC. 73

Horan, B., Dillon, P., Faverdin, P., Delaby, L., Buckley, F., & Rath, M. 2005. The interaction of strain of Holstein-Friesian cows and pasture-based feed systems on milk yield, body weight, and body condition score. *J Dairy Sci*, **88**(3), 1231–1243. 31, 34

Huquet, B., Leclerc, H., & Ducrocq, V. 2011a (06). Montly milk records : the basis of a new valued extension service for an efficient management. *In : ICAR meeting*. Poster. 40

Huquet, B., Leclerc, H., & Ducrocq, V. 2011b. Valorisation des coproduits du modèle génétique basé sur les contrôles élémentaires pour améliorer les outils de conseil. *Pages 23–26 of : Rencontre Recherche Ruminants*, vol. 18. Institut de l'Elevage-INRA. 40

Ibáñez, M. A., Carabaño, M. J., & Alenda, R. 1999. Identification of sources of heterogeneous residual and genetic variances in milk yield data from the Spanish Holstein-Friesian population and impact on genetic evaluation. *Livestock Production Science*, **59**(1), 33–49. 92

Kearney, J. F., Schutz, M. M., Boettcher, P. J., & Weigel, K. A. 2004a. Genotype x environment interaction for grazing versus confinement. I. Production traits. *J Dairy Sci*, **87**(2), 501–509. 31, 90

Kearney, J. F., Schutz, M. M., & Boettcher, P. J. 2004b. Genotype x environment interaction for grazing vs. confinement. II. Health and reproduction traits. *J Dairy Sci*, **87**(2), 510–516. 31, 90

König, S., Dietl, G., Raeder, I., & Swalve, H. H. 2005. Genetic relationships for dairy performance between large-scale and small-scale farm conditions. *J Dairy Sci*, **88**(11), 4087–4096. 30, 32, 90

Koivula, M., Nousiainen, J. I., Nousiainen, J., & Mïitysaari, E. A. 2007. Use of herd solutions from a random regression test-day model for diagnostic dairy herd management. *J Dairy Sci*, **90**(5), 2563–2568. 40, 51, 77

Kolmodin, R. 2003. *Reaction norms for the study of genotype by environment interaction in animal breeding*. Ph.D. thesis, Swedish University of Agricultural Sciences. 150

Kolmodin, R., Strandberg, E., Danell, B., & Jorjani, H. 2004. Reaction norms for protein yield and days open in Swedish Red and White dairy cattle in relation to various environmental variables. *Acta Agriculturae Scandinavica, Section A-Animal Science*, **54**(3), 139–151. 32, 33

Kolver, E. S., J. R. Roche, De Veth, M. J., Thorne, P. L., & Napper, A. R. 2002. Total mixed rations versus pasture diets : Evidence for a genotype x diet interaction in dairy cow performance. *Pages 246–251 of : Proc. N.Z. Soc. Anim. Prod.*, vol. 62. 31

Ladiray, D., & Quenneville, B. 1999 (Juillet). *Comprendre la méthode X11*. Tech. rept. United States Census Bureau. http://www.census.gov/ts/papers/x11doc.pdf. 74

Landais, E. 1992. Les trois pôles des systèmes d'élevage. *Les Cahiers de la Recherche Développement*. 17, 18

Le Dû, A. 2010. Rélexion et témoignage d'éleveur : gestion de la sélection sur l'élevage. *Pages 41–42 of : Journées techniques Elevage Biologique : Sélection Animale, Diversité Génétique & Agriculture Biologique*. Institut Technique de l'Agriculture Biologique. 25

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*. Ph.D. thesis, AgroParisTech. 40, 149

Leclerc, H., Duclos, D., Barbat, A., Druet, T., & Ducrocq, V. 2008. Environmental effects on lactation curves included in a test-day model genetic evaluation. *animal*, **2**(03), 344–353. 36, 38

Leclerc, H., Barbat-Leterrier, A., & Ducrocq, V. 2009. Mise en place d'une évaluation génétique sur les caractères laitiers à partir des performances élémentaires en remplacement des données lactation en race bovine. *Pages 291–294 of : Rencontre Recherche Ruminants*. Institut de l'élevage. 40

Li, Z. K., Yu, S. B., Lafitte, H. R., Huang, N., Courtois, B., Hittalmani, S., Vijayakumar, C. H M, Liu, G. F., Wang, G. C., Shashidhar, H. E., Zhuang, J. Y., Zheng, K. L., Singh, V. P., Sidhu, J. S., Srivantaneeyakul, S., & Khush, G. S. 2003. QTL x environment interactions in rice. I. heading date and plant height. *Theor Appl Genet*, **108**(1), 141–153. 152

LIC, & DairyNZ. 2010. *New Zealand Dairy Statistics 2009-10*. Tech. rept. LIC and DairyNZ. [en ligne] consulté le 6 mars 2012. 22

Lillehammer, M., Hayes, B. J., Meuwissen, T. H E, & Goddard, M. E. 2009. Gene by environment interactions for production traits in Australian dairy cattle. *J Dairy Sci*, **92**(8), 4008–4017. 152

Lopez-Villalobos, N., Garrick, D. J., Holmes, C. W., Blair, H. T., & Spelman, R. J. 2000. Effects of selection and crossbreeding strategies on industry profit in the New Zealand dairy industry. *J Dairy Sci*, **83**(1), 164–172. 23

Maltecca, C., Bagnato, A., & Weigel, K. A. 2004. Comparison of international dairy sire evaluations from meta-analysis of national estimated breeding values and direct analysis of individual animal performance records. *J Dairy Sci*, **87**(8), 2599–2605. 32, 34

Mayeres, P., Stoll, J., Bormann, J., Reents, R., & Gengler, N. 2004. Prediction of daily milk, fat, and protein production by a random regression test-day model. *J Dairy Sci*, **87**(6), 1925–1933. 40

McCarthy, S., Horan, B., Dillon, P., O'Connor, P., Rath, M., & Shalloo, L. 2007a. Economic comparison of divergent strains of Holstein-Friesian cows in various pasture-based production systems. *J Dairy Sci*, **90**(3), 1493–1505. 31, 34

McCarthy, S., Berry, D. P., Dillon, P., Rath, M., & Horan, B. 2007b. Influence of Holstein-Friesian strain and feed system on body weight and body condition score lactation profiles. *J Dairy Sci*, **90**(4), 1859–1869. 83

McDaniel, B. T., & Corley, E. L. 1967. Relationships between sire evaluations at different herdmate levels. *J Dairy Sci*, **50**(5), 735–741. 33

Meyer, Karin. 2007. WOMBAT : a tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). *J Zhejiang Univ Sci B*, **8**(11), 815–821. 129, 130

Moreteau, B., Morin, J. P., Gibert, P., Pétavy, G., Pla, E., & David, J. R. 1997. Evolutionary changes of nonlinear reaction norms according to thermal adaptation : a comparison of two Drosophila species. *Comptes Rendus de l'Académie des Sciences-Series III-Sciences de la Vie*, **320**(10), 833–841. 88

Mulder, H. 2007. *Methods to optimize livestock breeding programs with genotype by environment interaction and genetic heterogeneity of environmental variance.* Ph.D. thesis, Wageningen University. 143

Mulder, H. A., Groen, A. F., De Jong, G., & Bijma, P. 2004. Genotype x environment interaction for yield and somatic cell score with automatic and conventional milking systems. *J Dairy Sci*, **87**(5), 1487–1495. 32, 90

Nauta, W. J. 2009. *Selective Breeding in Organic Dairy Production.* Ph.D. thesis, Wageningen University. 150

Nauta, W. J., Groen, A. F., Veerkamp, R. F., Roep, D., & Baars, T. 2005. Animal breeding in organic dairy farming : an inventory of farmers' views and difficulties to overcome. *NJAS-Wageningen Journal of Life Sciences*, **53**(1), 19–34. 22, 25

Nauta, W. J., Veerkamp, R. F., Brascamp, E. W., & Bovenhuis, H. 2006. Genotype by environment interaction for milk production traits between organic and conventional dairy cattle production in The Netherlands. *J Dairy Sci*, **89**(7), 2729–2737. 31

Ojango, J. M. K ., & Pollott, G. E. 2002. The relationship between Holstein bull breeding values for milk yield derived in both the UK and Kenya. *Livestock Production Science*, **74**(1), 1–12. 29, 90, 146

OS Brune. *La Brune, la laitière de demain, performances et atouts.* http://www.brune-genetique.com/fr/performances.php. [en ligne] 18 Octobre 2010. 21

OS Montbéliarde. *La race Montbéliarde.* http://www.montbeliarde.org/race.php. [en ligne] 11 avril 2007. 21

OS Normande. *La race Normande, Description.* http://www.lanormande.com/web/presentation_generale_de_la_race_normande.html. [en ligne] 5 Mars 2010. 21

OS Simmental France. *La race Simmental, aptitudes et performances.* http://www.simmentalfrance.fr/aptitudes-et-performances.html. [en ligne] 24 Novembre 2008. 21

Padyukov, L., Silva, C., Stolt, P., Alfredsson, L., & Klareskog, L. 2004. A gene–environment interaction between smoking and shared epitope genes in HLA–DR provides a high risk of seropositive rheumatoid arthritis. *Arthritis & Rheumatism*, **50**(10), 3085–3092. 84

Pagès, J. 2010. *Statistiques générales pour utilisateurs 1-Méthodologie.* Presses Universitaires de Rennes. 86

Pflimlin, A., Buczinski, B., & Perrot, C. 2005. Proposition de zonage pour préserver la diversité des systèmes d'élevage et des territoires européens. *Fourrages*, 311–329. 19

Piccand, V., Schori, F., Troxler, J., Wanner, M., & Thomet, P. 2011. Projet : "Quelle vache pour la pâture ?" : problématique et description de l'essai. *Recherche Agronomique Suisse*, **2**(5), 200–205. 21

Postler, G. 1998. Der Ökologische Gesamtzuchtwert. *Kultur und Politiek*, **2**, 16–22. 22

Prim'Holstein France. *La race Prim'Holstein.* http://www.primholstein.com/primolstein_fr/race/. [en ligne] 12 Janvier 2012. 21

Pryce, J. E., Nielsen, B. L., Veerkamp, R. F., & Simm, G. 1999. Genotype and feeding system effects and interactions for = health and fertility traits in dairy cattle. *Livestock Production Science*, **57**(3D3), 193–201. 31, 83, 86

Raffrenato, E., Blake, R. W., Oltenacu, P. A., Carvalheira, J., & Licitra, G. 2003. Genotype by environment interaction for yield and somatic cell score with alternative environmental definitions. *J Dairy Sci*, **86**(7), 2470–2479. 32, 33, 34, 145

Robert, C. 1996. *Etude de quelques problèmes lies à la mise en oeuvre du REML en génétique quantitative.* Ph.D. thesis, Université Paul Sabatier. 91

Robert-Granié, C., Ducrocq, V., & Foulley, J.L. 1997. Heterogeneity of variance for type traits in the Montbeliarde cattle breed. 91

Robert-Granié, C., Bonaīti, B., Boichard, D., & Barbat, A. 1999. Accounting for variance heterogeneity in French dairy cattle genetic evaluation. *Livestock Production Science*, **60**(2-3), 343–357. 91

Schaeffer, L. R. 1994. Multiple-country comparison of dairy sires. *J Dairy Sci*, **77**(9), 2671–2678. 30, 84, 90

Scheiner, S. M., & Lyman, R. F. 1991. The genetics of phenotypic plasticity. II. Response to selection. *Journal of Evolutionary Biology*, **4**(1), 23–50. 152

Scheiner, S.M. 1993. Genetics and evolution of phenotypic plasticity. *Annual review of ecology and systematics*, 35–68. 152

Scheiner, S.M., & Lyman, R.F. 1989. The genetics of phenotypic plasticity I. Heritability. *Journal of Evolutionary Biology*, **2**(2), 95–107. 24, 83

Schmid, A., A. Spengler-Neff, & Pedotti, R. 2010. Les Grisons veulent adapter leurs vaches bio aux conditions locales. *bio actualités*, 12, 10–11. 20

Sørensen, M. K., Norberg, E., Pedersen, J., & Christensen, L. G. 2008. Invited review : crossbreeding in dairy cattle : a Danish perspective. *J Dairy Sci*, **91**(11), 4116–4128. 23

Stanton, T. L., Blake, R. W., Quaas, R. L., Vleck, L. D. Van, & Carabaño, M. J. 1991. Genotype by environment interaction for Holstein milk yield in Colombia, Mexico, and Puerto Rico. *J Dairy Sci*, **74**(5), 1700–1714. 29

Steinheim, G., Ødegård, J., Ådnøy, T., & Klemetsdal, G. 2008. Genotype by environment interaction for lamb weaning weight in two Norwegian sheep breeds. *Journal of animal science*, **86**(1), 33–39. 83

Strandberg, E. 2011. Animal Genetic * Environment Interaction. *Chap. 343 of :* Meyers, Robert. A. (ed), *Encyclopedia of Sustainability Science and Technology.* Springer Science+Business Media, LLC. 84

Strandberg, E., Kolmodin, R., Madsen, P., Jensen, J., & Jorjani, H. 2000. Genotype by environment interaction in Nordic dairy cattle studied by use of reaction norms. *Interbull Bull*, **25**, 41–45. 33

Thorpe, W., Morris, C. A., & Kang'ethe, P. 1994. Crossbreeding of Ayrshire, brown Swiss, and Sahiwal cattle for annual and lifetime milk yield in the lowland tropics of Kenya. *J Dairy Sci*, **77**(8), 2415–2427. 23

UPRA Abondance. *La race Abondance, caractéristiques.* http://www.abondance.asso.fr/siab/caracteristiques.htm. [en ligne] 22 Février 2007. 21

UPRA Tarentaise. *La race Tarentaise, caractéristiques raciales.* http://www.race-tarentaise.com/web/item/francais/caracteristiques_raciales/portrait/index.php. [en ligne] 28 Février 2011. 21

UPRA Vosgienne. *La race Vosgienne, caractéristiques.* http://www.racevosgienne.com/standard.htm. [en ligne] 15 Février 2012. 21

Urioste, J.I., Gianola, D., Rekaya, R., Fikse, W.F., & Weigel, K.A. 2001. Evaluation of extent and amount of heterogeneous variance for milk yield in Uruguayan Holsteins. *Animal Science*, **72**(2), 259–268. 92

Van Vleck, L. D., Dong, M. C., & Wiggans, G. R. 1988. Genetic (co) variances for milk and fat yield in California, = New York, and Wisconsin for an animal model by restricted maximum = likelihood. *J Dairy Sci*, **71**(3D11), 3053–3060. 30

VanRaden, P. M., & Sanders, A. H. 2003. Economic merit of crossbred and purebred US dairy cattle. *J Dairy Sci*, **86**(3), 1036–1044. 23

Veerkamp, R. F., Dillon, P., Kelly, E., Cromi, A .R., & Groen, A. F. 2002. Dairy cattle breeding objectives combining yield, survival and calving interval for pasture-based systems in Ireland under different milk quota scenarios. *Livestock Production Science*, **76**(1-2), 137–151. 22

Via, S., Gomulkiewicz, R., De Jong, G., Scheiner, S. M., Schlichting, C. D., & Tienderen, P. H. Van. 1995. Adaptive phenotypic plasticity : consensus and controversy. *Trends in Ecology & Evolution*, **10**(5), 212–217. 87

Weigel, K. A., & Rekaya, R. 2000. A multiple-trait herd cluster model for international dairy sire evaluation. *J Dairy Sci*, **83**(4), 815–821. 30, 32, 34, 146, 147

Weigel, K. A., Rekaya, R., Zwald, N. R., & Fikse, W. F. 2001. International genetic evaluation of dairy sires using a multiple-trait model with individual animal performance records. *J Dairy Sci*, **84**(12), 2789–2795. 30, 146

Wicks, H. C. F., & Leaver, J. D. 2006. Influence of genetic merit and environment on somatic cell counts of Holstein-Friesian cows. *Vet J*, **172**(1), 52–57. 31, 86

Wilson, L. L., Rishel, W. H., & Harvey, W. R. 1972. Influence of Herd, Sire and Herd× Sire Interactions on Live and Carcass Characters of Beef Cattle. *Journal of Animal Science*, **35**(3), 502–506. 87

Windig, J. J., Calus, M. P. L., Beerda, B., & Veerkamp, R. F. 2006. Genetic correlations between milk production and health and fertility depending on herd environment. *J Dairy Sci*, **89**(5), 1765–1775. 34, 88, 128, 129

Wu, R. L. 1997. Genetic control of macro-and micro-environmental sensitivities in Populus. *Theor Appl Genet*, **94**(1), 104–114. 152

Zwald, N. R., Weigel, K. A., Fikse, W. F., & Rekaya, R. 2003. Application of a multiple-trait herd cluster model for genetic evaluation of dairy sires from seventeen countries. *J Dairy Sci*, **86**(1), 376–382. 32, 34, 90

Liste des publications

Articles scientifiques

Huquet B., Leclerc H. and Ducrocq V. Modelling and estimation of genotype by environment interactions for production traits in French dairy cattle. *Genet Sel Evol* (accepté le 7 novembre 2012).

Huquet B., Leclerc H. and Ducrocq V. Characterization of French dairy farm environments from herd-test-day profiles. 2012, *J Dairy Sci*, 95, 4085-4098.

Communications orales

Huquet B., Leclerc H. and Ducrocq V. Genotype by environment interactions when environments are defined from Herd-Test-Day profiles. 2012, 63rd annual meeting of the European Association for Animal Production, Bratislava, Slovakia.

Huquet B., Leclerc H. and Ducrocq V. Performances laitières mensuelles : les valoriser au mieux dans un nouvel outil de conseil. 2011, 18èmes journées Rencontres Recherches Ruminants, 7-8 Décembre, Paris, France.

Poster

Huquet B., Leclerc H. and Ducrocq V. Monthly milk records : the basis of a new valued extension service for an efficient management. 2011, ICAR (International Committee for Animal Recording) Annual Workshop, Bourg-en-Bresse, France.

Formations suivies

Catégorie : Formations scientifiques

- classification, données manquantes: méthodes et application 2010 - 2011 paris, société française de statistiques (SFDS) 14 heures
- formation interne : utilisation des marqueurs 2001 - 2002 21 heures
- Introduction to genome based analysis of quantitative traits with machine learning and non parametric methods Daniel Gianola 2009 - 2010 PARIS agroparistech 14 heures

Total du nombre d'heure pour la catégorie Formations scientifiques : *49 h*

Catégorie : Langues et Techniques de communication

- Building Your Base, novembre 2010 (23 novembre 2010) 2010 - 2011 AgroParisTech - av du Maine 4 heures

- Write Right Niveau 1- Rédaction et structuration de l'article scientifique - juin 2011 (07 juin 2011) 2010 - 2011 AgroParisTech - avenue du Maine 21 heures

Total du nombre d'heure pour la catégorie Langues et Techniques de communication : *25 h*

Catégorie : Méthodes, outils, langages

- apprentissage du langage awk (théorie et pratique) 2010 - 2011 jouy en josas, INRA 7 heures
- formation à unix 2001 - 2002 jouy en josas, INRA 7 heures
- Comment utiliser les infocentres ? 2009 - 2010 jouy en josas, INRA 7 heures

Total du nombre d'heure pour la catégorie Méthodes, outils, langages : *21 h*

Catégorie : Ouverture scientifique et culturelle

- Journées ABIES 1 er 2 avril 2010 (01 avril 2010) 2009 - 2010 AgroParisTech - 16 rue Claude Bernard, Paris 14 heures
- Journées ABIES 29 et 30 Mars 2011 (29 mars 2011) 2010 - 2011 AgroParisTech - 16 rue Claude Bernard, Paris 14 heures

Total du nombre d'heure pour la catégorie Ouverture scientifique et culturelle : *28 h*

Catégorie : Réunion d'information

- Sélection génomique et changements dans l'indexation des bovins laitiers : évolutions ou révolution ? Journée technique du 13 octobre 2009 2009 - 2010 PARIS agroparistech 7 heures

- Développements récents et perspectives d'application dans différentes filières Journée Sélection Génomique 12 mai 2011 2010 - 2011 PARIS agroparistech 8 heures

Total du nombre d'heure pour la catégorie Réunion d'information : *15 h*

Catégorie : Séminaires

- colloque Doc'J 2010 2009 - 2010 jouy en josas, INRA 8 heures
- colloque Doc'J 2011 2010 - 2011 jouy en josas, INRA 8 heures
- colloque Doc'J 2012 2011 - 2012 jouy en josas, INRA 8 heures
- séminaire COSADD: « Elevage et Développement Durables : quels objectifs et stratégies de sélection ? » 2009 – 2010 8heures
- 3R 2009 Rencontre autour de la Recherche sur les Ruminants 2009 - 2010 14 heures
- 3R 2010 Rencontre autour de la Recherche sur les Ruminants 2010 - 2011 14 heures
- 3R 2011 Rencontre autour de la Recherche sur les Ruminants 2010 - 2011 14 heures
- colloque épigénétique 2010 - 2011 Les Colloques de l'Académie d'Agriculture de France
- congrès ICAR 2010 International Committee for Animal Recording 2009 - 2010 RIGA, LETTONIE 21 heures
- congrès ICAR 2011 (workshop) 2010 - 2011 bourg en bresse 14 heures
- EAAP 2011 European Association for Animal Production 2010 - 2011 35 heures
- workshop INTERBULL2010 (mars) 2001 - 2002 salon international de l'agriculture, paris 14 heures
- INTERBULL 2011 2010 - 2011 STAVANGER, NORWAY 14 heures
- INTERBULL 2010 2009 - 2010 RIGA, LETTONIE 14 heures
- séminaire des thésards du département génétique animale de l'inra 2010 2009 - 2010 pornichet, inra 14 heures
- séminaire des thésards du département génétique animale de l'inra 2011 2010 - 2011 limoges, inra 14 heures
- séminaire des thésards du département génétique animale de l'inra 2012 2011 - 2012 Oléron, inra 14 heures
- séminaire du département inra génétique animale 2010 - 2011 nice, inra 32 heures

Total du nombre d'heures pour la catégorie Séminaires *: 245 h*

Catégorie : Management et relation interpersonnelle

- Doctoral Management Program (avec ParisTech) 2009 - 2010 ParisTech, paris 150 heures

Total du nombre d'heure pour la catégorie Management et relation interpersonnelle : *150 h*

Catégorie : Valorisation des compétences

- Nouveau Chapitre de la Thèse – 2011, Association Bernard Gregory, abies - *40 heures*

Total du nombre d'heure pour la catégorie Valorisation des compétences: *40 h*

Résumé

La France présente une grande diversité de conditions pédoclimatiques et par conséquent des systèmes d'élevage très variés. Pourtant, les schémas de sélection actuels en bovins laitiers considèrent que les meilleurs reproducteurs sont les mêmes quel que soit le type de système d'élevage, c'est-à-dire qu'il n'existe pas d'interactions Génotype*Milieu. L'objectif de cette thèse est de mesurer les interactions Génotype*Milieu en France sur les caractères laitiers et fonctionnels dans les 3 races bovines laitières principales (Normande, Montbéliarde, Holstein).

Un point crucial dans ce type d'étude est la façon de définir le milieu. L'innovation de cette thèse est l'utilisation des profils Troupeau-Jour de Contrôle. Ce sont des coproduits du modèle génétique basé sur les contrôles élémentaires. Ils reflètent la production permise par la conduite de troupeau au cours du temps. Ils présentent l'avantage d'être disponibles à partir des bases de données nationales et d'être uniquement le reflet de la conduite, contrairement à d'autres définitions qui mêlent effet génétique et effet d'environnement au sein de la définition du milieu ou qui se focalisent sur certains points précis de la conduite sans prendre en compte son effet global. La description des profils Troupeau-Jour de Contrôle de plus de 15000 élevages normands, montbéliards et holsteins par des méthodes de lissage de séries temporelles, d'analyses factorielles et de classification a permis de créer 2 définitions du milieu en vue de l'étude des interactions Génotype*Milieu : des milieux définis comme des groupes d'élevages aux conduites distinctes ou un milieu défini comme un continuum à travers une ou des variables synthétiques.

L'importance des interactions Génotype*Milieu a été estimées à partir de 2 types de modèles : un modèle multicaractères qui valorise la définition du milieu sous forme de groupes d'élevages et un modèle de norme de réaction qui valorise, quant à lui, le milieu défini comme un continuum. Les avancées méthodologiques proposées dans cette thèse concernent les modèles de normes de réaction. Des approches permettant de prendre en compte plusieurs variables d'environnement au sein d'un même modèle et de les résumer au sein d'une matrice génétique de rang réduit sont mises en avant.

Aucun reclassement n'a été mis en évidence : les meilleurs reproducteurs sont les mêmes quel que soit le système d'élevage. Les schémas des sélections actuels sont donc performants. Il existe tout de même une interaction Génotype*Milieu significative sous forme d'effet d'échelle : la variabilité des valeurs génétiques des animaux est plus importante dans les systèmes d'élevage plus intensifs. Cet effet d'échelle ne sera pas pris en compte dans les modèles d'évaluation génétique, en revanche, il est possible d'imaginer un indicateur utilisable sur le terrain pour mesurer les écarts de performances, dus à cet effet, auxquels il faut s'attendre.

Mots clés : bovins laitiers, interactions génotype*milieu, profil Troupeau-Jour de Contrôle (TJC), modèle multicaractères, modèle de norme de réaction

Abstract

Because of the diversity of pedoclimatic conditions in France, dairy farms have very diversified herd management systems. For this reason, some breeders question the efficiency of the existing breeding schemes for their own management system. To overcome these concerns, a genotype by environment interaction study at the French national level has been considered necessary. The aim of this thesis is to assess the presence of genotype by environment interactions in Normande, Montbéliarde and Holstein breeds for production and functional traits.

A tricky point in genotype by environment interaction studies is the environment definition. The innovation of this thesis deals with the use of Herd-Test Day profiles. They are co-products of the French test day model. They reflect the production due to herd management over time. They are available in national databases and only reflect herd management effect contrary to other definitions in which there is a confusion between genetic and environmental effects in the environment definition or which focus on specific features of the herd management without taking into account its global effect. Herd Test Day profiles of more than 15,000 herds have been studied through time series smoothing, factor analysis and clustering methods. It led to 2 definitions of the environment for the genotype by environment interaction study : environments defined as herd groups or one to several environmental gradients.

Genotype by environment interactions were assessed with 2 models : the multitrait model and the reaction norm model. The first one uses herd groups as definition of the environment whereas reaction norm model considers the environment as a gradient. Several methodological improvements have been suggested for reaction norm models : taking into account several environmental gradients in a reaction norm model and summarizing them through a reduced rank genetic matrix.

No reranking has been shown : the best parents are the same whatever the herd management system. Consequently, current breeding schemes are relevant. However, a scale effect exists : the variability of animal breeding values is higher in intensive herds. Genetic models will not account for this scale effect. However, a tool useful for breeders could indicate the deviation between expected performances and actual performances due to this scale effect.

Keywords : dairy cattle, genotype by environment interaction, Herd-Test Day (HTD) profile, multitrait model, reaction norm model