

# Intégration des caractères fonctionnels dans un index de synthèse pour les races bovines laitières : de la théorie à la pratique

V. DUCROCQ, D. BOICHARD, A. BARBAT, H. LARROQUE

INRA, Station de Génétique Quantitative et Appliquée, Département de Génétique Animale, 78352 Jouy-en-Josas cedex

**RESUME** – Une évaluation génétique de type BLUP - Modèle animal multicaractères a été mise en place afin de se rapprocher de la situation idéale d'analyse simultanée des caractères jouant un rôle dans le nouvel index de synthèse. Ces caractères sont 3 caractères fonctionnels (comptages leucocytaires, longévité fonctionnelle et fertilité femelle) ainsi que leurs prédicteurs potentiels: production laitière, taux protéique, vitesse de traite, un caractère d'aplombs et quatre caractères morphologiques de mamelle. Pour contourner la trop grande complexité statistique et informatique d'une évaluation multicaractères au sens strict, l'évaluation BLUP porte sur des données pré corrigées des effets de milieu et d'environnement permanent estimés au cours des évaluations génétiques actuelles de ces caractères. Ce calcul des données pré corrigées a été adapté pour des caractères tels que la longévité ou la fertilité. Les paramètres génétiques pour les 11 caractères considérés ont été estimés et une évaluation conjointe de ces 11 caractères a été mise en place. Le calcul des coefficients de détermination montre qu'une évaluation des caractères fonctionnels combinant l'information directe sur ces caractères et l'information indirecte des prédicteurs précoces améliore nettement la précision des index des jeunes taureaux.

## Integration of functional traits into a total merit index for dairy cattle breeds: from theory to practice

V. DUCROCQ, D. BOICHARD, A. BARBAT, H. LARROQUE

INRA, Station de Génétique Quantitative et Appliquée, Département de Génétique Animale, 78352 Jouy-en-Josas cedex

**SUMMARY** – A multiple trait BLUP animal model genetic evaluation was developed in order to approximate the ideal situation consisting in a joint analysis of all traits playing a role in the new total merit index. These traits are three functional traits (somatic cells count, functional longevity and female fertility) as well as their potential predictors : milk yield, protein content, milking ease, a feet and legs trait and four udder traits. To get around the statistical and computational complexity that a strict multiple trait analysis represents, the BLUP evaluation was performed on pre-adjusted records. Pre-adjustment factors were the estimates of fixed and permanent environment effects of the current genetic evaluations. The computation of pre-adjusted records for traits like longevity or female fertility required a special treatment. Genetic parameters for the 11 traits considered were estimated and a joint evaluation of these traits was implemented. The computation of reliabilities shows that a functional traits evaluation that combines direct information on these traits with indirect information from early predictors greatly improves the accuracy of the evaluation of young bulls.

## INTRODUCTION

Un nouvel objectif de sélection alliant production et caractères fonctionnels a été mis au point (Colleau et Régaldo, 2001) et entériné par les UPRa Montbéliarde, Normande et Prim'Holstein au printemps 2001. Cet objectif combine les valeurs génétiques des caractères de production à travers leur synthèse (INEL) et les caractères fonctionnels suivants : les comptages leucocytaires (CELL), la fertilité femelle (FER), la longévité fonctionnelle (LGF) et la synthèse morphologique (MORP). Les pondérations économiques associées varient légèrement selon les races : en races Montbéliarde et Prim'Holstein, pour un poids de référence égal à 1 pour l'INEL (ou pour une synthèse laitière légèrement différente en Montbéliarde), les quatre caractères fonctionnels reçoivent chacun un poids de 0,25. En race Normande, un poids supérieur est donné à la morphologie (0,35) et un poids inférieur est attribué à la fertilité (0,20) et à la longévité fonctionnelle (0,15).

A l'exception de certains caractères morphologiques, les caractères fonctionnels sont moins hérissables que ceux de production, ce qui conduit à une précision des index (les CD – coefficients de détermination) notoirement faible pour les jeunes taureaux. Or il existe parmi la liste des caractères morphologiques des prédictors potentiels de CELL ou LGF (voir par exemple, Weigel et al., 1998, Druet et al., 1999, Rupp et Boichard, 1999 ; Larroque et Ducrocq, 2000, Buenger et al., 2001). Mais comment combiner ces différentes sources d'information ?

La stratégie optimale est l'évaluation génétique multicaractères (MT-BLUP) des données de base à l'aide de la méthodologie BLUP (meilleure prédiction linéaire non biaisée). Les propriétés de cette approche sont connues (voir par exemple, Ducrocq, 1994). On notera en particulier que les poids optimaux des index élémentaires dans l'index de synthèse sont les pondérations économiques des caractères de l'objectif. Corrélations entre caractères et hétérogénéité des CD pour chacun des caractères sont automatiquement pris en compte par l'approche MT-BLUP.

Malheureusement, une évaluation conjointe de l'ensemble des caractères à partir des informations de base n'est pas réaliste. Les modèles d'analyse varient entre caractères ; certains (LGF) ne sont pas linéaires. D'autres comportent des mesures répétées (production, FER, CELL) ou doivent prendre en compte des hétérogénéités de variance (production). Surtout, le volume de données à manipuler lors d'une telle évaluation dépasse largement les capacités informatiques, au moins en race Prim'Holstein.

De nombreux pays calculent des index de synthèse par une combinaison linéaire simple des index élémentaires. Ceci est parfaitement justifiable lorsque les caractères sont peu corrélés et quand les CD sont assez homogènes entre caractères pour un même animal. C'était le cas de l'index synthétique (ISU) précédent qui combinait production, morphologie et vitesse de traite. Ce ne l'est plus ici.

La théorie des index de sélection permet en principe de prendre en compte les corrélations non nulles et les CD variables entre caractères. Mais elle devient rapidement complexe, voire ingérable, si l'on souhaite couvrir l'ensemble des situations possibles (structures de famille variables avec un nombre variable de performances par caractère).

Des approximations ont été proposées : Par exemple, Druet et al. (1999) ou Larroque et Ducrocq (1999) ont utilisé la méthodologie du MACE (Multiple trait Across Country Evaluation ; Schaeffer, 1994) utilisée par Interbull pour les évaluations internationales. Mais cette méthodologie repose sur des hypothèses qui ne sont pas adaptées au cas traité ici.

Nous avons retenu une approche différente qui se rapproche de l'analyse multicaractères des données de base, tout en restant applicable informatiquement.

## MATERIEL ET METHODES

### 1.1 STRATEGIE GENERALE

Une approche MT-BLUP n'est pas directement envisageable compte tenu du volume d'information de base et de la variabi-

lité et la complexité des modèles d'analyse. Une condition nécessaire pour lever cet obstacle est donc de résumer les performances de chaque animal pour chaque caractère par une valeur unique, que l'on puisse décrire par un modèle unique, aussi simple que possible. Pour cela, il faut précorriger les données pour tous les effets de milieu et l'effet d'environnement permanent (employé en cas de performances répétées pour un même animal). Si  $y_{i,m}$  représente la donnée précorrignée de l'animal  $m$  pour le caractère  $i$ , une évaluation MT-BLUP peut être envisagée en supposant le modèle animal suivant :

$$y_{i,m} = \mu_i + a_{i,m} + e_{i,m}$$

où, pour le  $i^{\text{ème}}$  caractère,  $\mu_i$  est la moyenne générale,  $a_{i,m}$  est la valeur génétique additive de  $m$  et  $e_{i,m}$  est la résiduelle. Pour prendre en compte la quantité variable d'information contenue dans  $y_{i,m}$  (par exemple, en cas de données répétées), la variance de la résiduelle est supposée hétérogène :

$$\text{Var}[e_{i,m}] = \frac{1}{w_{i,m}} \sigma_{e,i}^2$$

où  $w_{i,m}$  est le poids de la (des) performance(s) sur le caractère  $i$  de l'animal  $m$  et  $\sigma_{e,i}^2$  est la variance résiduelle de référence pour ce caractère.

La stratégie proposée consiste en :

- la définition et le calcul de données précorrignées et de leur poids associé. Les facteurs de correction proviennent des évaluations génétiques propres à chaque caractère. Un traitement spécial des données de longévité est nécessaire, parce que le modèle n'est pas linéaire ;
  - la résolution d'un système d'équations MT-BLUP très simple mais encore très grand ;
  - la combinaison des index MT-BLUP dans l'ISU à partir d'un jeu unique de pondérations économiques.
- Un sous-produit essentiel de cette approche est la production de nouveaux index sur les caractères fonctionnels, combinant de façon naturelle, pour les mâles comme pour les femelles, l'information directe et l'information indirecte provenant d'autres caractères corrélés. C'est le cas pour la longévité fonctionnelle.

### 1.2 CHOIX DES CARACTERES

Il n'était pas envisageable d'analyser ensemble les 5 caractères de production, les caractères fonctionnels et les 23 à 28 caractères de morphologie. Aussi, nous nous sommes limités à ceux *a priori* considérés comme de bons prédictors potentiels de 3 caractères fonctionnels (CELL, LGF, FER). Ce choix a été basé sur les connaissances bibliographiques, les corrélations brutes entre index et le bons sens. Ceci nous a conduit à retenir, outre CELL, LGF et FER :

- deux caractères de production : la quantité de lait (ayant un impact par exemple sur CELL et FER) et le taux protéique ;
- la vitesse de traite (influençant CELL et LGF) ;
- cinq caractères morphologiques, pouvant varier suivant les races : l'un caractérisant les aplombs (Note d'aplombs en Montbéliarde (MO), Aplombs arrière en Normand (NO), et Epaisseur du talon en Prim'Holstein (HF)) et les quatre autres décrivant la mamelle : Attache Avant, Distance Plancher de la Mamelle – Jarret (Développement en MO), Equilibre et un autre caractère (Support en MO, Ecartement des trayons en NO, Hauteur de l'attache arrière en HF). Ces caractères sont souvent décrits comme liés à LGF et CELL.

La fertilité femelle (sur vache) est analysée en France par le succès ou l'échec à chaque IA (Boichard et al., 1998). L'évaluation génétique intègre l'information obtenue sur génisse en tant que caractère corrélé. Cette information a été ignorée ici, afin de ne pas surcharger l'évaluation avec de nombreux animaux n'ayant qu'une performance (une IA génisse). L'objectif de sélection global inclut l'INEL et la synthèse morphologique (MORP) qui ne sont que très partiellement cou-

verts dans la liste de 11 caractères mentionnés pour l'approche MT-BLUP. Mais INEL et MORP correspondent à des caractères assez héréditaires, mieux connus que les autres caractères fonctionnels et dont les index seraient de toute façon très peu influencés par ceux-ci. Par conséquent, dans l'ISU, seuls CELL, LGF et FER sont des index combinés issus de l'approche multicaractères, les deux autres, INEL et MORP proviennent des évaluations "classiques".

### 1.3 POPULATION EVALUEE

Afin de maintenir la taille des fichiers dans des limites acceptables, seules les données des animaux nés à partir de 1988 ont été conservées. Leur généalogie a été remontée sur deux générations pour les femelles et aussi loin que possible pour les mâles. Au delà, les animaux ont été assignés à des groupes de parents inconnus. Le tableau 1 décrit les effectifs évalués et la répartition des données disponibles par caractère.

### 1.4 CALCUL DES DONNEES PRECORRIGEEES

A partir des évaluations génétiques de base des caractères décrits par un modèle linéaire, la résiduelle du modèle (standardisée pour la production, afin de prendre en compte l'existence de variances hétérogènes) a été calculée. En cas de données répétées, c'est une résiduelle moyenne (pondérée, dans le cas de la production, en fonction de la longueur et du numéro de la lactation) qui a été retenue. La donnée précorrigée correspondante est simplement la somme de cette résiduelle et de l'index de la vache. Le poids associé est l'élément diagonal de la matrice des coefficients du BLUP, après absorption de l'effet troupeau-année et de l'effet d'environnement permanent.

Dans le cas de la fertilité, il a fallu prendre en compte le fait que le modèle utilisé est un modèle père-grand-père maternel et que l'effet vache comprend une partie génétique et une partie d'environnement permanent. Le cas de la longévité est plus compliqué, car le modèle est non linéaire, la résiduelle n'est pas distribuée selon une loi normale et coexistent des données complètes et des données "censurées", correspondant aux animaux encore en vie (Ducrocq, 1997). On a tout d'abord développé des index LGF approchés avec un modèle animal d'analyse de survie, puis les équations conduisant à ces index ont été transformées pour se rapprocher d'un système d'équations BLUP unicaractère. Les éléments constitutifs du membre de droite du système ont alors été assimilés à des performances précorrigées. Les détails de la méthodologie et de sa validation ont été décrits par ailleurs (Ducrocq, 2001).

**Tableau 1**  
Effectifs et répartition des données lors de l'évaluation de Juin 2001

	MO	NO	HF
Animaux évalués	1133192	1141081	6447180
Vaches avec donnée laitière connue (A)	802896	809600	5454847
(A) et avec morphologie connue <sup>(a)</sup>	56%	30% 26%	37% 27% 18%
(A) et avec comptage leucocytaire connu	88%	94%	91%
(A) et avec longévité connue	87%	83%	91%
(A) et avec donnée de fertilité connue	73%	81%	71%

(a) les tables de pointages ayant été modifiées en Normande et Prim'Holstein, les nouveaux caractères ne sont pas disponibles pour les animaux les plus âgés.

### 1.5 PARAMETRES GENETIQUES

Les (co)variances génétiques et résiduelles entre les 11 caractères évalués ont été estimées par la technique du maximum de vraisemblance restreinte (REML) appliquée à un modèle père. Des fichiers plus réduits ont été créés pour que cette partie reste abordable. On n'a retenu que les animaux ayant au moins une performance (quel que soit son poids) sur chacun des caractères. Ces fichiers comprenaient respectivement en races Montbéliarde, Normande et Prim'Holstein: 287866 vaches et 3114 pères, 161615 vaches et 1968 pères et 269523 et 4567 pères.

### 1.6 EVALUATION GENETIQUE

Même avec un modèle aussi simple, la résolution des équations du BLUP multicaractères avec les effectifs indiqués au tableau 1 reste une tâche très lourde. Une approche particulière et un logiciel de résolution spécifique ont été développés pour ce calcul. L'approche fait appel à (une adaptation de) la décomposition canonique des données de base en 11 systèmes indépendants, suivie d'une résolution de ces systèmes par un algorithme itératif (gradients conjugués avec préconditionneur - voir Ducrocq et al., 2001, pour plus de détails). Un coefficient de détermination pour chaque caractère, intégrant toutes les informations sur les parents, les performances propres et les descendants pour les taureaux comme pour les vaches a aussi été entrepris. Son calcul est basé sur une extension multicaractères de l'approche de Harris et Johnson (1998).

**Tableau 2**  
Corrélations génétiques(a) entre les comptages leucocytaires ou la longévité fonctionnelle et les autres caractères en Montbéliarde (MO), Normande (NO) et Prim'Holstein (HF)

	Comptages Leucocytaires			Longévité Fonctionnelle		
	MO	NO	HF	MO	NO	HF
Comptages Leucocytaires	-	-	-	<b>0,33</b>	<b>0,42</b>	<b>0,42</b>
Longévité Fonctionnelle	<b>0,33</b>	<b>0,42</b>	<b>0,42</b>	-	-	-
Fertilité Femelle	0	0	-0,12	-0,17	<b>-0,30</b>	<b>-0,42</b>
Vitesse de Traite	0	0,14	<b>0,49</b>	<b>-0,37</b>	-0,15	0
Caractère d'aplombs <sup>(b)</sup>	0	-0,09	0	-0,21	-0,19	-0,11
Attache avant Mamelle	-0,17	-0,19	-0,05	-0,27	<b>-0,32</b>	-0,16
Distance <sup>(b)</sup> plancher-Jarret	<b>0,44</b>	-0,28	-0,27	<b>0,43</b>	<b>-0,42</b>	<b>-0,45</b>
Equilibre Mamelle	-0,21	-0,24	-0,10	-0,23	<b>-0,36</b>	-0,23
Autre caractère de Mamelle <sup>(b)</sup>	-0,13	-0,14	0,12	-0,15	<b>-0,35</b>	0,29
Lait	0,15	0	0,10	0,10	<b>-0,35</b>	-0,05
Taux protéique	0	0	0	0,23	0,10	0

a) les corrélations font référence au modèle d'analyse des données de base: des valeurs positives du caractères sont donc défavorables pour les comptages leucocytaires (plus de cellules), la longévité (plus de réforme), le développement en Montbéliarde et la hauteur de l'attache arrière en Prim'Holstein.

b) voir texte

**Tableau 3**  
**Coefficient de détermination moyen des index de longévité fonctionnelle des taureaux d'IA suivant le modèle d'évaluation utilisé (race Normande)**

Année de Naissance	Information directe		Information combinée
	Modèle père grand-père	Modèle animal	BLUP Modèle animal multicaractères
1991	0,77	0,78	0,80
1992	0,72	0,74	0,77
1993	0,53	0,58	0,66
1994	0,40	0,47	0,58

## RESULTATS ET DISCUSSION

### 2.1 PARAMETRES GÉNÉTIQUES

La plupart des corrélations résiduelles sont faibles (<0.20), à l'exception de celles entre caractères de mamelle. Le tableau 2 présente les estimées des corrélations génétiques entre CELL d'une part, et LGF d'autre part et l'ensemble des 11 caractères considérés. En gras sont indiquées les corrélations supérieures à 0,30 en valeur absolue. Il ressort que CELL et LGF sont bien corrélés entre eux, quelle que soit la race. Les caractères de mamelle, et surtout la distance plancher de la mamelle – jarret (« développement » en Montbéliarde) sont assez corrélés à LGF. Cette distance est également liée à CELL, surtout en race Montbéliarde. Le caractère choisi pour décrire les membres est un prédicteur très moyen de LGF en Normandie et Montbéliarde et mauvais en Prim'Holstein. Ces résultats sont en accord avec ceux de Larroque et Ducrocq (2001).

La fertilité s'ajoute à la liste des bons prédicteurs de LGF en Holstein mais pas en Montbéliarde, la Normandie étant intermédiaire. Ces corrélations suivent d'ailleurs la liaison production laitière-FER (-0,25, -0,20, -0,08 pour HF, NO et MO respectivement).

Des différences raciales nettes apparaissent avec la vitesse de traite: la liaison entre vitesse de traite et LGF est favorable (-0,37) en Montbéliarde, pour laquelle traite et CELL ne sont pas liées. A l'inverse, vitesse de traite et longévité ne sont pas liées en Holstein, alors que la liaison avec CELL est nettement défavorable, confirmant ainsi les résultats de Rupp et Boichard (1999). Là encore, la Normandie est dans une situation intermédiaire. Une interprétation possible serait de considérer qu'un optimum au delà duquel une trop grande facilité de traite devient gênant a été atteint en Holstein, mais pas encore en Montbéliarde.

### 2.2. EVALUATION GENETIQUE

L'utilisation de données précorrrigées a été partiellement validée en comparant les résultats d'analyses BLUP modèle animal unicaractère à partir de ces données aux résultats des évaluations nationales (voir Ducrocq, 2001 et Ducrocq et al., 2001, pour plus de détails). Dans tous les cas, les résultats sont très encourageants.

Le calcul des CD par l'approche de Harris et Johnson (1998) permet de chiffrer le gain de précision apporté par les prédicteurs des caractères fonctionnels. Ainsi, le tableau 3 présente les CD moyens pour LGF des taureaux d'IA récents en race Normande. On constate que le passage au modèle animal pour l'analyse de la longévité permet à lui seul une augmentation pouvant aller jusqu'à 18% du CD des taureaux les plus jeunes.

L'apport de l'information indirecte provenant de CELL, FER et des caractères de morphologie permet ensuite à la plupart de ces jeunes taureaux de franchir la limite imposée pour la publication des index fonctionnels en France (0,50). Pour les taureaux plus vieux, comme prévu, la contribution des parentés entre femelles comme celle de l'information indirecte est beaucoup plus limitée. En race Prim'Holstein, 97% des taureaux nés en 1994 et 59% de ceux nés en 1995 dépassaient cette limite de 0,50 de CD lors du calcul de Juin 2001 contre respectivement 86% et 16% avec la seule information directe. Pour les taureaux Holsteins nés en 1990, les index MT-BLUP pour la longévité fonctionnelle avaient une corrélation de 0,97 avec les index du modèle père-grand-père d'analyse de survie (sans information indirecte). Cette corrélation diminue pour les taureaux plus jeunes (0,95 ; 0,94 ; 0,93 ; 0,84 et 0,60 les années suivantes), traduisant l'importance croissante de l'apport des informations indirectes. Ces diverses constatations ont conduit au remplacement de la publication (mais pas du calcul !) des index LGF du modèle non linéaire par ceux de l'approche multicaractères. Ces nouveaux index LGF ainsi que les nouveaux index de synthèse ont fait l'objet d'une première publication officielle en Juin 2001.

## 3. CONCLUSION

L'approche proposée permet une prise en compte simultanée dans l'index de synthèse de caractères corrélés et pour lesquels la quantité d'information disponible est très variable. Elle permet l'utilisation dans l'index de synthèse d'un jeu de pondérations unique et clair : celles de l'objectif de sélection. Cette démarche reste conceptuellement simple et s'applique de la même façon à tous les individus (mâles et femelles). Elle fournit des sous-produits intéressants, comme un index longévité fonctionnel plus précis et disponible plus précocement. Elle peut malgré tout renforcer l'image de « boîte noire » qu'ont parfois les évaluations génétiques. Une utilisation sérieuse du nouvel index synthétique dans la sélection des taureaux comme des vaches devrait contribuer à des progrès génétiques plus équilibrés et « durables », sans l'épée de Damoclès que pourrait constituer une dégradation génétique continue des qualités d'élevage.

**Boichard D., Barbat A., Briend M. 1998.** In: Renc. Rech. Ruminants, 5: 103-106.

**Boichard D., Rupp R. 1997.** Interbull Bulletin 18: 54-60.

**Buenger A., Ducrocq V., Swalve H.H., 2001.** J. Dairy Sci., 84: 1531-1541.

**Colleau J.J., Régaldo D., 2001.** In: Renc. Rech. Ruminants, 8. (ce volume).

**Druet T., J. Soelkner, N. Gengler, 1999.** J. Dairy Sci. 82. Online.

**Ducrocq V., 1994.** 5th World Cong. Genet. Appl. Livest. Prod., August 6-11, Guelph, Canada, 18: 455-462

**Ducrocq V., 1997.** In: Renc. Rech. Ruminants, 4: 215-218.

**Ducrocq V., 2001.** Interbull meeting, 30-31 Août. Budapest, Hongrie.

**Ducrocq V., Boichard D., Barbat A., Larroque H., 2001.** 52<sup>nd</sup> EAAP meeting, 26-29 Août, Budapest, Hongrie.

**Harris B., Johnson D., 1998.** J. Dairy Sci., 81:2723-2728.

**Larroque H, Ducrocq V., 2001.** Genet. Sel. Evol., 33, 39-59.

**Rupp R., Boichard D., 1999.** J. Dairy Sci., 82:2198-2204.

**Schaeffer L.R., 1994.** J. Dairy Sci., 77:2671-2678.

**Weigel K.A., Lawlor T.J., Van Raden P.M., Wiggans G. R., 1998.** J. Dairy Sci., 81: 2040-2044.