HAL
open science

La méthode française d'évaluation génétique des reproducteurs laitiers

B. Bonaiti, D. Boichard, E. Verrier, Vincent Ducrocq, A. Barbat, M. Briend

To cite this version:

B. Bonaiti, D. Boichard, E. Verrier, Vincent Ducrocq, A. Barbat, et al.. La méthode française d'évaluation génétique des reproducteurs laitiers. Productions Animales, 1990, 3 (2), pp.83-92. hal-00895891

HAL Id: hal-00895891

https://hal.science/hal-00895891v1

Submitted on 11 May 2020

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.

INRA Prod. Anim.,
1990, 3 (2), 83 - 92

*B. BONAITI, D. BOICHARD, E. VERRIER *, V. DUCROCQ, A. BARBAT, M.BRIEND*

INRA Station de Génétique Quantitative et Appliquée
78352 Jouy-en-Josas Cedex

** Institut National Agronomique Paris-Grignon, 16, rue Claude-Bernard, 75231 Paris Cedex 05*

La méthode française d'évaluation génétique des reproducteurs laitiers

En mars 1990 une nouvelle méthode d'évaluation génétique des animaux laitiers est entrée en application en France. Cet article donne les grandes lignes de cette méthode basée sur le principe du « BLUP » appliqué à un modèle « animal ».

La valeur génétique des animaux est estimée à partir de deux types d'information : les performances mesurées et les généalogies. Ainsi, l'identification des animaux, leur filiation et le contrôle de leurs performances sont à la base de tout protocole d'évaluation génétique (aussi appelé « indexation »). L'index est la meilleure estimation de la valeur génétique d'un reproducteur. C'est aussi sa valeur génétique la plus probable. En moyenne, un reproducteur transmet à ses descendants la moitié de son index. L'évaluation génétique permet donc aux éleveurs et aux opérateurs de la sélection de choisir les meilleurs reproducteurs en vue de leurs objectifs. Par une analyse régulière du présent et du passé, elle permet aussi de mesurer l'efficacité des programmes de sélection.

Dans un article précédent, Ducrocq (1990) fait le point sur la théorie de l'évaluation des reproducteurs en mettant l'accent sur les principes et les propriétés du BLUP et du modèle animal. Nous présentons ici concrètement et plus simplement la méthode basée sur le modèle animal, développée en 1989 par le Département de Génétique Animale de l'INRA, dans le cadre de l'évaluation génétique en routine des reproducteurs laitiers en France. Elle est entrée en application en mars 1990 pour les bovins, et sera utilisée prochainement pour les ovins et caprins.

Résumé

Les principales caractéristiques de la méthode d'évaluation génétique des reproducteurs laitiers, appliquée en France à partir de 1990, sont décrites. Les différences avec la méthode précédente sont soulignées. La méthode BLUP, appliquée à un modèle animal, permet d'utiliser toute l'information disponible de façon optimale. Pour chaque individu évalué, cette information est résumée en trois sources synthétiques : la moyenne des index de ses parents ; deux fois la moyenne pondérée des index de ses descendants, corrigés pour la valeur de ses conjoints ; la moyenne pondérée de ses propres performances, corrigées pour les effets de milieu. Les coefficients de pondération à appliquer à chacune de ces sources sont explicités et justifiés. Quelques exemples illustrent la contribution relative de chacune des trois sources d'information dans des situations courantes.

1 / Généralités sur la méthode

Les caractères de production laitière pour lesquels un index est calculé pour les mâles comme pour les femelles sont au nombre de cinq : la quantité de lait (QL) en kg, la quantité de matière grasse (MG) en kg, la quantité de matière protéique (MP) en kg, le taux butyreux (TB) en g/kg et le taux protéique (TP) en g/kg.

Ceci constitue une première différence avec la méthode précédente, dite IF2 (Poutous *et al* 1981), où les matières grasse et protéique étaient combinées en un caractère unique, la matière utile.

Toutefois, pour faciliter l'utilisation des index et éviter la dispersion de l'effort de sélection, ils sont combinés en un index synthétique, l'INEL (INRA-ITEB 1989), dérivé de la formule prévisible de paiement du lait, et qui vise à maximiser le revenu laitier :

$$\text{INEL} = 0{,}37\ (\text{MG} + 2\ \text{MP} + 6\ \text{TP}) \qquad (1)$$

La méthode utilisée pour chacun des caractères laitiers est le BLUP (en anglais Best Linear Unbiased Predictor, en français Meilleur Prédicteur Linéaire non Biaisé) appliqué à un modèle « animal ».

Le BLUP est une méthode statistique complexe développée par Henderson (ses travaux sont nombreux mais on peut se reporter à Henderson 1973) qui permet d'analyser simultanément des effets de milieu identifiés (voir plus loin) et des effets génétiques. Le BLUP permet

en particulier de tenir compte d'une répartition non aléatoire des valeurs génétiques parmi les différentes modalités des effets de milieu. Par exemple, du fait du progrès génétique, les animaux les meilleurs se trouvent en général parmi les plus jeunes et les effets génétiques sont donc partiellement confondus avec les effets âge ou rang de lactation. Dans les méthodes d'évaluation les plus simples, comme la comparaison aux contemporaines, les données sont simplement corrigées pour la moyenne des animaux soumis aux mêmes conditions de milieu. Les valeurs génétiques obtenues sont donc estimées intra catégorie de milieu (intra troupeau, intra classe d'âge...) et ne sont comparables entre catégories de milieu que sous l'hypothèse d'un même niveau génétique entre catégories (entre troupeaux, entre classes d'âge...), ce qui n'est généralement pas le cas. Le BLUP, au contraire, en estimant simultanément les effets génétiques (intra et entre catégories) et les effets de milieu, se libère de cette hypothèse et les valeurs génétiques estimées, incluant les différences intra et entre catégories, sont tout à fait comparables dans le temps et l'espace.

L'index est calculé pour 5 caractères de production laitière : les quantités de lait, de matière grasse, de matière protéique et les taux butyreux et protéique.

L'expression modèle animal signifie que dans le modèle d'analyse, c'est l'individu réalisant la performance qui est considéré (on parle aussi de modèle individuel) et non son père ou son grand père maternel. On s'affranchit ainsi de plusieurs hypothèses qui sont de moins en moins acceptables, comme l'absence d'accouplements raisonnés ou l'absence de sélection des femelles.

Une méthode appliquant le principe du BLUP à un modèle animal est considérée actuellement comme la procédure optimale d'évaluation des reproducteurs car elle bénéficie à la fois des propriétés théoriques du BLUP et du modèle animal. Non biaisées, les estimations de valeur génétique permettent de comparer entre eux tous les animaux de la population analysée, même s'ils ne sont pas contemporains. En conséquence, elles permettent aussi de mesurer des différences génétiques entre régions ou entre troupeaux, ainsi que le progrès génétique. D'autre part, si toute la population est considérée depuis son origine avant toute sélection, les estimations de valeur génétique sont en théorie insensibles à la sélection réalisée, aux types d'accouplements réalisés et même à l'évolution des paramètres génétiques au cours du temps.

Ces propriétés remarquables, associées à une apparente simplicité théorique, expliquent que le BLUP appliqué à un modèle animal soit devenu la référence internationale en matière d'évaluation des reproducteurs. Son emploi est donc tout à fait important dans la perspective de compétition accrue avec nos partenaires étrangers dans le commerce des reproducteurs et des semences ou embryons.

Bien qu'IF2 présente également certaines de ces propriétés, la nouvelle méthode constitue un progrès important. Précisons les analogies et les différences entre les deux méthodes. Une méthode d'évaluation peut être définie par deux caractéristiques, le modèle de description des données et le plus ou moins grand nombre d'approximations dans la résolution des équations.

Du point de vue du modèle, IF2 et la nouvelle méthode sont proches, puisque les deux analysent les performances sur plusieurs lactations en estimant des effets troupeau x année, des effets du mois, de l'âge à la mise bas et de l'intervalle entre mise bas, définis intra région, année et numéro de lactation, et prennent en compte la valeur génétique des conjointes (mères supports) lors du calcul de la valeur des mâles.

C'est surtout dans la stratégie de résolution que les deux méthodes diffèrent. La nouvelle méthode ne fait aucune approximation, respectant ainsi parfaitement la théorie. IF2, pour des raisons de moyens de calcul, ne prenait pas en compte l'ensemble des parentés. L'information apportée par les filles était négligée pour estimer la valeur des mères. De même, l'index sur ascendance des taureaux ne tenait compte que du père et du grand père maternel, négligeant l'information apportée par les mères. Les pondérations appliquées aux différentes sources d'information étaient estimées et, de ce fait, pouvaient différer légèrement des pondérations optimales théoriques. Les index des mâles et des femelles étaient calculés selon deux modèles différents partiellement reliés, rendant difficile leur comparaison. La prise en compte des biais résultant de l'élimination progressive des femelles au cours de leur carrière était simplifiée. A chaque traitement trimestriel n'étaient calculés que les effets les plus récents (portant sur les lactations des trois dernières années), ce qui avait pour conséquence de considérer le progrès génétique estimé comme acquis.

Si le modèle animal constitue un progrès, il convient cependant de souligner que la meilleure méthode est nécessaire mais pas suffisante pour obtenir des index fiables et précis. Les propriétés remarquables du modèle animal ne sont pleinement réalisées que si les données analysées le permettent. En particulier, la comparaison des animaux dans l'espace et dans le temps n'est possible que si toutes les généalogies et toutes les performances sont disponibles et inclues dans l'analyse. Dans le cadre des schémas laitiers impliquant l'ensemble du cheptel français, le modèle animal ne trouve donc tout son intérêt que s'il est appliqué à un fichier national, rassemblant toute l'information disponible, présente et passée.

En aucun cas, la méthode ne sait corriger ou « amortir » des erreurs dans les données ou les généalogies. Les problèmes potentiels sont de deux ordres :

- les effets de milieu parasites : si un facteur engendrant des écarts de production est non identifié, il n'est pas pris en compte dans le modèle et provoque des biais sur les index.
- la disconnexion, c'est-à-dire l'absence d'information permettant la comparaison des animaux : ainsi, les différences génétiques entre régions, entre troupeaux, entre classes d'âge, ne sont correctement estimables que par l'intermédiaire de connexions génétiques (par exem-

ple échanges de semences entre régions, testage de taureaux sur plusieurs régions, utilisation prolongée dans le temps de taureaux servant de témoins) abondantes, fiables et non soumises à des traitements préférentiels.

Signalons enfin que l'application de cette méthode nécessite un important travail de développement ainsi que des moyens de calcul très puissants. Pour une description détaillée de l'algorithme original de résolution, on se reportera à Ducrocq *et al* (1990). Encore maintenant, malgré les progrès de l'informatique, sa réalisation pratique, demeurée longtemps impossible, demande de nombreuses heures de calcul.

2 / Données et modèle d'analyse

2.1 / Précorrection et extrapolations des lactations

L'information élémentaire est constituée de contrôles laitiers mensuels réalisés sur l'ensemble d'une lactation. Cette information est d'abord combinée pour calculer des performances par lactation (production totale, taux moyens). Chaque lactation fait ensuite l'objet d'une précorrection et éventuellement d'une extrapolation, qui visent à standardiser les performances de quantités (lait et matières), les taux (TB et TP) restant inchangés. Les méthodes de précorrection et d'extrapolation ne subissent aucune modification par rapport à IF2. Mises au point il y a 15 ans, elles demandent à être validées et éventuellement adaptées, ce qui sera fait prochainement.

a / La quantité totale de lait est précorrigée pour la durée (Poutous et Mocquot 1975) et pour le rang de lactation, selon la formule suivante :

$$QL305 = QL \times C \times 385 / (DL + 80) \qquad (2)$$

avec QL305 la quantité de lait précorrigée (kg), QL la quantité de lait brute (kg), DL la durée de lactation (jours) et C un coefficient multiplicatif visant à homogénéiser les moyennes (toutes les lactations sont exprimées en équivalent adulte) mais surtout les variances.

rang de lactation :	valeur de C :
1	1,30
2	1,17
3	1,10

La durée de lactation est un caractère peu héritable qui détermine une partie importante des variations de production. Un ajustement est donc nécessaire. En effet, l'absence de correction favoriserait les animaux peu fertiles, à lactation longue, et pénaliserait trop les animaux présentant une lactation courte. A l'inverse, une correction complète favoriserait les animaux à mauvais potentiel, présentant souvent des lactations courtes. La méthode de correction pour la durée de lactation présentée en (2) se situe volontairement entre ces deux extrêmes. Pour les lactations de plus de 270 jours, la production corrigée est indépendante de la durée. Par contre, en dessous de cette limite, elle reste positivement corrélée à la durée de lactation.

b / Quand la lactation n'est pas terminée au moment de l'indexation, la quantité de lait est extrapolée selon la méthode proposée par Mocquot (1978). Bien entendu, les lactations courtes mais terminées (tarissement précoce) ne sont pas modifiées.

c / Les quantités de matière précorrigées sont obtenues simplement à partir de la quantité de lait précorrigée et des taux.

$$MG305 = QL305 \times TB \qquad (3a)$$
$$MP305 = QL305 \times TP \qquad (3b)$$

2.2 / Modèle d'analyse

Les performances (P) précorrigées et éventuellement extrapolées sont analysées selon le modèle suivant :

$$P = \mu + F + A + V + E \qquad (4)$$

avec μ : moyenne de la population
F : somme des effets de milieu identifiables
A : valeur génétique additive (transmissible) de l'animal
V : valeur de l'animal non transmissible, appelée effet permanent. Cet effet traduit le fait que les performances d'un même animal ont tendance à se ressembler au cours de sa carrière. En effet, toutes les lactations sont influencées par certaines phases marquantes comme l'élevage de la génisse. Généralement, cet effet est dû à des effets de milieu non identifiables, variant d'un animal à l'autre mais se répétant d'une lactation à l'autre d'un même animal. Il peut s'y ajouter des effets génétiques non transmissibles. Il est illustré par le fait que la répétabilité est plus élevée que l'héritabilité. Si la valeur génétique d'une vache, c'est-à-dire ce qu'elle est capable de transmettre à sa descendance, est prédite au mieux par A, sa valeur économique, c'est-à-dire ce qu'elle est capable de produire dans le futur, est prédite au mieux par A + V.
E : valeur résiduelle, due au milieu, non expliquée par le modèle, variant aléatoirement d'un animal à l'autre et d'une lactation à l'autre chez un même animal.
Sept facteurs de milieu sont identifiés et pris en compte pour l'estimation de la valeur de F :

- le troupeau dans lequel l'animal réalise sa lactation,
- la région, définie par des groupes de petites régions agricoles INSEE (voir cartes en annexe),
- l'année au cours de laquelle a commencé la lactation (la campagne est définie du 1er septembre au 31 août),
- le rang de lactation,
- l'âge au vêlage, pour les premières lactations (par classe de 3 mois),
- le mois de vêlage,
- l'intervalle entre les deux derniers vêlages, pour les lactations autres que la première (par classe de 1 mois).

Ces différents facteurs de milieu sont analysés conjointement et certaines interactions sont considérées. En effet, à titre d'exemple, l'effet du mois de vêlage ou du troupeau n'est pas

identique d'une année sur l'autre. Finalement, le modèle peut être réécrit de façon explicite :

$$P_{ijklmnpr} = \mu + T_{jk} + L_{rkl} + N_{rkm} + I_{rkln} + M_{rklp} + A_i + V_i + E_{ijklmnpr} \quad (5)$$

avec $P_{ijklmnpr}$ la performance,

A_i la valeur génétique de la vache,

V_i l'effet permanent, propre à la vache i, mais non transmissible,

T_{jk} l'effet du troupeau j lors de la campagne k,

L_{rkl} l'effet du numéro de lactation l lors de la campagne k dans la région r,

N_{rkm} l'effet de l'âge m à la première mise bas lors de la campagne k dans la région r (pour les lactations 1 uniquement),

I_{rkln} l'effet de l'intervalle n entre les deux dernières mise bas en lactation l (2 et 3 uniquement) lors de la campagne k dans la région r,

M_{rklp} l'effet du mois p de mise bas lors de la campagne k en lactation l et dans la région r,

$E_{ijklmnpr}$ la résiduelle.

Dans les faits, l'effet ($\mu + T_{jk}$) est estimé globalement, ce qui est bien entendu rigoureusement équivalent à une estimation séparée de μ et de T_{jk}. Il en est de même de l'effet ($L_{rkl} + N_{rkm}$) en première lactation et de ($L_{rkl} + I_{rkln}$) en lactations 2 et 3. Enfin, les effets mois sont supposés identiques en lactations 2 et 3.

Les effets A, V et E sont dits aléatoires, du fait des hypothèses concernant leur distribution. Les autres sont des effets fixés, considérés comme parasites et sources de biais s'ils étaient omis.

Ce modèle assez complexe de description des données est assez différent des modèles choisis à l'étranger, en particulier aux USA. Ce choix est lié à l'impossibilité, compte tenu de la petite taille moyenne des troupeaux, de définir la plus petite cellule élémentaire (troupeau x année x saison x numéro de lactation...) comme effet fixé unique. En effet, dans de nombreux cas, les effectifs intra combinaison élémentaire seraient trop faibles pour permettre une comparaison fiable.

Le calcul de l'index d'un individu prend en compte : ses performances propres, les index de ses parents et ceux de ses descendants corrigés pour la valeur des conjoints.

2.3 / Les données prises en considération

Toutes les généalogies connues sont prises en compte. Cette règle ne souffre qu'une exception, en race bovine Frisonne : lorsque qu'un animal a un index calculé à l'étranger (USA ou Canada) et que cette information est intégrée dans le calcul, la généalogie de cet animal n'est pas considérée car elle est déjà prise en compte dans son index étranger.

Le cas des performances est plus complexe. Seules sont considérées les lactations de rang 1, 2 ou 3, correspondant à une mise bas postérieure au 1er septembre 1974. Pour que la sélection en première lactation (25 % d'éliminations sur des critères laitiers) soit correctement prise en compte sans engendrer de biais, une lactation de rang 2 ou 3 n'est considérée que si la première lactation est connue. Enfin, pour des raisons techniques liées aux capacités de calcul, une lactation de rang 2 ou 3 est exclue de l'analyse quand elle n'est pas réalisée dans le même troupeau que la première lactation.

Ces contraintes entraînent donc une perte d'informations. Si cette perte est négligeable pour les mâles, elle peut être importante pour certaines femelles. Dans une deuxième phase, les lactations non prises en compte dans le modèle animal au sens strict sont donc réintégrées. Toutefois, elles ne sont utilisées que pour l'évaluation des femelles réalisant ces performances ; autrement dit, cette information complémentaire n'est pas utilisée pour le calcul des index de leurs apparentés.

3 / Le calcul de l'index

Tous les calculs décrits dans cette partie s'appuient sur la théorie de la Génétique Quantitative, qui précise les règles de transmission héréditaire des caractères quantitatifs, ainsi que les liaisons qui peuvent exister entre performances d'individus apparentés. En particulier, de nombreux calculs découlent de la simple équation exprimant la valeur génétique d'un individu i (A_i) en fonction de celles de son père (A_p) et de sa mère (A_m)

$$A_i = 1/2\, A_p + 1/2\, A_m + W_i \quad (6)$$

W_i étant un écart intra famille traduisant l'aléa de méiose, d'espérance nulle et variant aléatoirement entre les descendants d'un même couple de parents. Notons que dans la population d'origine et en l'absence de consanguinité, la variance de cet écart W_i représente la moitié de la variance génétique. L'existence de W_i explique la précision faible ou moyenne du choix sur ascendance.

3.1 / Ecriture générale de l'index

La méthode utilisée permet de prendre en compte les performances individuelles ainsi que toutes les informations en provenance de tous les animaux apparentés, de près ou de loin, à l'animal à évaluer. Au travers de la méthode d'indexation, toute l'information disponible qui, pour un même individu, peut être très volumineuse, est résumée, sans aucune approximation, en trois sources d'information synthétiques :
- l'individu lui-même,
- son ascendance directe, c'est-à-dire son père et sa mère,
- sa descendance directe, c'est-à-dire ses fils et ses filles, corrigée pour le niveau de ses conjoint(e)s.

Ce résultat fondamental découle du fait que si la matrice de parenté est une matrice relativement « pleine » (la covariance entre deux individus est non nulle dès qu'ils sont apparentés, aussi éloignés soient-ils), son inverse est au contraire très creuse, un individu n'étant relié qu'à ses parents, ses descendants et ses conjoints.

L'individu contribue à son propre index au travers de ses performances. Ses ascendants, ses descendants directs et ses conjoints ne contribuent qu'au travers d'index.

Le modèle animal conduit à un système linéaire explicite comptant autant d'équations

que d'inconnues (donc plus d'équations que d'animaux à évaluer). L'équation particulière de l'individu i fournit la formule de calcul de l'index d'un individu i (A_i) estimateur de sa valeur génétique vraie A_i :

$$A_i = \frac{p_1 A_a + p_2 A_d + p_3 P_c}{p_1 + p_2 + p_3} \qquad (7)$$

avec

A_a : moyenne des index des parents (ascendance).

A_d : deux fois la moyenne pondérée des index des descendants corrigés pour l'index des conjoints de l'individu i.

P_c : moyenne pondérée des performances de l'individu, corrigées pour les effets de milieu identifiés et pour l'effet permanent non transmissible de l'individu.

p_1, p_2, p_3 : coefficients de pondération, ou poids, accordés à chaque source d'information, en fonction de sa qualité et de son abondance. La division par la somme des poids permet de pondérer en fait chaque source par sa contribution relative ($p_j / \Sigma\, p_j$), la somme des contributions relatives étant égale à 1.

L'index peut être exprimé sous deux autres formes équivalentes, présentant une certaine similitude avec les indices de sélection classiques :

$$A_i = A_a + \frac{p_2 (A_d - A_a) + p_3 (P_c - A_a)}{p_1 + p_2 + p_3} \qquad (8)$$

$$A_i = \frac{p_1 A_a + p_2 A_d}{p_1 + p_2} + \frac{p_3}{p_1 + p_2 + p_3} \left(P_c - \frac{p_1 A_a + p_2 A_d}{p_1 + p_2} \right) \qquad (9)$$

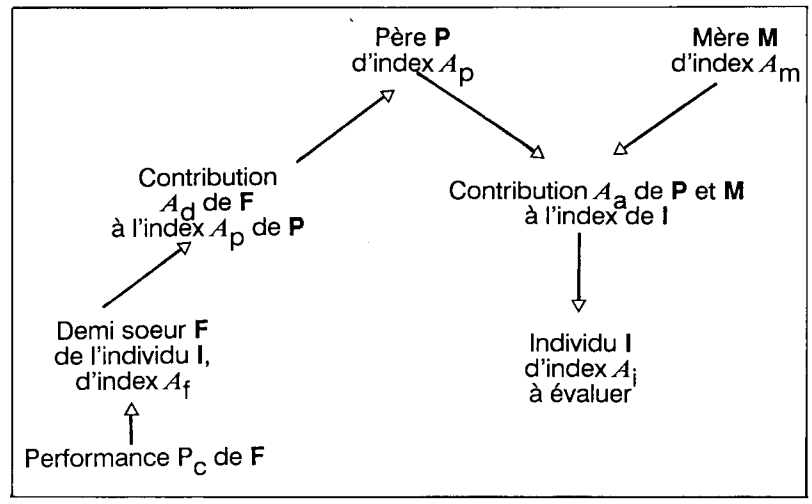
La formule (8) illustre le fait que l'index est constitué d'un index sur ascendance (son espérance) et du reste de l'information (performances propres et descendants) régressée sur l'index sur ascendance et estimant l'aléa de méiose W_i (on dit régressé car les poids $p_2/(p_1 + p_2 + p_3)$ et $p_3/(p_1 + p_2 + p_3)$ sont plus petits que 1).

La formule (9) illustre que l'index est constitué d'une information sur apparentés (parents et descendants) et de performances propres régressées sur cette information généalogique.

Cette formalisation illustrée en (7), tout en restant simple et parlante, permet bien d'utiliser toute l'information disponible. Par exemple, les performances des demi-soeurs de i (figure 1), bien que n'apparaissant pas explicitement dans les formules ci-dessus, contribuent à l'index de i, les règles de calcul permettant de leur affecter le coefficient de pondération approprié. En effet, les performances des demi-soeurs f de l'individu i à évaluer contribuent à leur propre index A_f. Elles contribuent donc au titre de la descendance à l'index A_p du père p de i. Finalement, elles contribuent donc au titre de l'ascendance, à l'index A_i de i. Il en est de même pour tous les apparentés de i.

La formule de calcul de l'index est donc simple mais nécessite de connaître A_a, A_d et P_c. Toute nouvelle performance individuelle modifie l'index (A_i). Par contre coup, elle modifie l'index des apparentés de i, puisque i contribue à l'index de ses ascendants, descendants et

Figure 1. *Exemple de liaison entre l'index d'un individu à évaluer et les performances d'apparentés.*



conjoints. Une nouvelle performance modifie donc indirectement la fraction ascendance (A_a) et descendance (A_d) de l'index de l'animal i. Tous les index sont dépendants les uns des autres. Cet équilibre, traduit dans le système d'équations, montre que l'on ne peut pas considérer une équation seule, indépendamment des autres. Ceci explique qu'à chaque traitement, tous les individus sont réanalysés et que leur index est susceptible de changer, même pour les individus vieux.

3.2 / Explicitation des sources d'information

a / Moyenne des ascendants

$$A_a = 1/2\, A_p + 1/2\, A_m \qquad (10)$$

A_p et A_m représentent les index respectifs du père et de la mère de l'individu à évaluer. Les coefficients 1/2 traduisent le fait qu'un parent transmet en espérance la moitié de sa valeur génétique à ses descendants.

Tous les animaux ont une contribution de l'ascendance, même quand leurs parents sont inconnus. Un raffinement de la méthode réside dans le fait que si un parent quelconque est non identifié (ce qui est le cas d'un quart des animaux environ), on remplace son index par son espérance, c'est-à-dire sa valeur la plus probable. Cette valeur est estimée à partir de l'ensemble des animaux ayant des parents inconnus et présentant des caractéristiques voisines (même sexe, même année de naissance, même origine). Par exemple, on suppose que les parents des vaches sans généalogie, nées en Bretagne en 1985, ont le même niveau génétique espéré : ils appartiennent à un même « groupe de parents inconnus ». Ce niveau génétique espéré (l'effet « groupe ») est estimé à partir de l'information apportée par les index de tous leurs descendants. C'est cette valeur de l'effet groupe qu'on utilisera comme contribution de l'ascendance pour les animaux sans généalogie. Cette méthode complexe permet de s'affranchir de l'hypothèse d'un niveau généti-

que identique pour l'ensemble des animaux sans généalogie et permet donc de considérer l'existence de plusieurs populations de base de niveaux génétiques différents.

b / Deux fois la moyenne de la descendance

$$A_d = 2 \Sigma d_f (A_f - A_c / 2) / \Sigma d_f \quad (11)$$

la somme s'effectuant sur tous les descendants, mâles et femelles, de i, avec
A_f : index d'un fils ou d'une fille
A_c : index de l'autre parent de f, conjoint de i
d_f : coefficient de pondération dépendant de la connaissance du conjoint (voir explication au paragraphe 3.3).
Dans le cas particulier où tous les conjoints sont connus, A_d vaut deux fois la simple moyenne des index des n descendants, mâles et femelles, de l'individu i, corrigés pour les n conjoint(e)s de i.

$$A_d = 2 \Sigma (A_f - A_c/2) / n \quad (12)$$

On peut constater que :
- 1/2 A_c représente l'espérance de ce qu'a transmis au descendant f le second parent, soit la moitié de sa valeur génétique,
- A_f - 1/2 A_c représente l'espérance de ce qu'a transmis le seul individu i, à évaluer,
- 2 (A_f - 1/2 A_c) représente l'espérance de la valeur génétique de i, au vu des seuls index de f et c
- A_d représente l'espérance de la valeur génétique de i, au vu de tous ses descendants.

c / Moyenne pondérée des performances propres

$$P_c = \Sigma w (P - \mu - F - V) / \Sigma w \quad (13)$$

la somme s'effectuant sur les lactations de i, avec
P : performance de l'individu à évaluer
μ : moyenne de la population
F : somme de tous les effets de milieu estimés
V : effet permanent estimé de la femelle
w : coefficient de pondération de la lactation (voir paragraphe 3.3)

Cette formule montre qu'il est indispensable de connaître μ et les effets F et V pour calculer P_c. Les effets de milieu sont simplement estimés par la moyenne des performances réalisées dans une catégorie donnée, corrigées pour la valeur de l'animal (A+ V) et des autres effets de milieu. Ainsi, par exemple, l'effet « milieu » troupeau x année est estimé par la moyenne des performances réalisées l'année considérée et dans le troupeau considéré, corrigées pour les variations de rang de lactation, d'âge, de mois de mise bas, d'intervalle entre mise bas, pour les différences de niveau génétique (correction pour l'index) et pour les différences dans la politique de sélection massale des femelles (effet permanent). L'effet du milieu troupeau, à valeur génétique et conditions d'âge, de mois de mise bas et d'intervalle entre vêlages constants, mesure l'effet conjoint de la technicité de l'éleveur et des ressources disponibles pour le troupeau (alimentation, climat...). Cet effet ne dépend pas de la politique d'élimination après la première ou deuxième lactation, politique qui peut aussi faire varier la moyenne de production, ni du niveau génétique du troupeau, estimé par la moyenne des index des animaux. Il est bien sûr tout à fait différent de la simple moyenne du troupeau.

Comme précédemment, on montre ainsi que tous les paramètres sont interdépendants : pour calculer un index, il faut connaître l'index des apparentés et les effets de milieu ; pour calculer un effet de milieu, il faut connaître les index et les autres effets de milieu. De nouveau, cette interdépendance est traduite par le système d'équations. Chaque inconnue (index, effet du troupeau, effet de l'âge...) a une équation, impliquant les autres paramètres inconnus. Le nombre d'équations atteint souvent plusieurs millions. C'est tout l'intérêt du BLUP que d'estimer tous les effets, génétiques et de milieu, en même temps ; c'est aussi ce qui le rend complexe à résoudre. Cette interdépendance entre les effets est prise en compte par le calcul simultané, au travers d'itérations, de tous les facteurs de variation et de toutes les valeurs génétiques. Mais, à la différence d'IF2, le nombre nécessaire d'itérations pour stabiliser les solutions (à des valeurs constantes d'équilibre) est réalisé à chaque traitement.

d / Intégration des lactations exclues du modèle animal

On a vu que de nombreuses lactations étaient exclues de l'analyse par le modèle animal et qu'il était important de réintégrer cette information dans les index des femelles. Les performances sont tout d'abord corrigées pour les effets de milieu estimées par le modèle animal. Soit P'_c une performance de ce type, de poids p_3'. Connaissant l'index A_i de la vache i, issu du modèle animal, qui s'écrit

$$A_i = \frac{p_1 A_a + p_2 A_d + p_3 P_c}{p_1 + p_2 + p_3} \quad (14)$$

on obtient aisément le nouvel index A_i'

$$A_i' = \frac{p_1 A_a + p_2 A_d + p_3 P_c + p_3' P_c'}{p_1 + p_2 + p_3 + p_3'} \quad (15)$$

e / Intégration des index étrangers

En race Frisonne, compte tenu de l'origine des animaux et de nos importations en provenance d'Amérique du Nord, il est intéressant de prendre en compte l'information correspondant aux index américains et canadiens dans l'évaluation française. Toutefois, une information ne peut être intégrée que sous forme de performances et de généalogies. La méthode est la suivante. L'index étranger d'un individu correspond à une espérance de supériorité S de ses descendants, estimée avec une certaine précision. Cette information est donc traduite en générant des filles fictives, réalisant une lactation chacune, de niveau S. Le nombre de filles fictives est fonction de la précision de l'index étranger mais est limité à 100. Ces filles sont placées dans des troupeaux fictifs dont l'effet est fixé à zéro. Les mères de ses femelles sont supposées inconnues et sont placées dans des groupes particuliers. L'effet estimé de ces groupes, équivalent à la valeur moyenne des mères, a la propriété intéressante d'ajuster automatiquement (entre autres) pour les différences de niveau génétique entre pays, qu'il

n'est donc pas nécessaire de connaître précisément a priori.

3.3 / Explicitation des poids.

Toute l'information relative à un individu à évaluer est donc résumée en trois sources qu'il faut pondérer entre elles :

- une moyenne d'index des ascendants directs (père et mère)
- deux fois la moyenne pondérée des index des descendants directs (fils et filles) corrigés pour la valeur des conjoints
- une moyenne pondérée des performances propres.

Il n'est malheureusement pas possible de justifier de façon parfaitement rigoureuse les poids impliqués dans une équation particulière (celle de l'individu i) sans considérer l'ensemble du système d'équations. Une analyse de ce type dépasserait le cadre de cet article. Pour une démonstration complète, on pourra se reporter à Ducrocq (1990). Toutefois, quand on analyse une équation particulière du système (celle de l'individu i), on peut montrer que les poids sont ceux qu'on obtiendrait s'il n'existait aucune incertitude sur les autres inconnues impliquées dans l'équation et si les différents prédicteurs étaient indépendants les uns des autres. Sous cette hypothèse qu'on admettra, le poids d'une information est alors proportionnel à la précision qu'apporte cette information sur le paramètre en question. La précision est mesurée par l'inverse de la variance d'erreur. Si la variance d'erreur est faible et donc la précision forte, l'information correspondante est un bon prédicteur et sa pondération est élevée. Dans la suite, on s'attachera à déterminer la variance d'erreur de chaque type d'information pour expliciter et justifier sa pondération dans l'index de l'individu i.

Enfin, il convient de ne pas confondre le poids d'une information et le coefficient de détermination (CD) ou précision (R^2) de l'index qui sont deux choses différentes. Ainsi, par exemple, si un individu a ses deux parents connus et s'il est indexé uniquement sur ascendance, le poids de l'information sur ascendance est constant (100 % en relatif) et indépendant de la précision de l'index des parents. Par contre le CD de l'individu est variable et dépend du CD des parents : $CD_i = 0,25\ CD_p + 0,25\ CD_m$. La notion de CD ne sera pas abordée ici.

a / Moyenne des ascendants

$$p_1 = d_i \qquad (16)$$

d_i est un facteur traduisant la qualité de la généalogie de i. En effet, nous savons d'après (6) que :

$$A_i = 1/2\ A_d + 1/2\ A_m + W_i$$

La variance de A_i (la variance génétique de la population, Var A) peut s'exprimer en fonction des variances parentales (Var A_p et Var A_m) et de la variance intra famille de pleins frères soeurs (Var W). Sous certaines hypothèses, on considère généralement que :

$$\text{Var } A_p = \text{Var } A_m = \text{Var } A$$

Donc

$$\text{Var } W = \text{Var } A - 1/4 \text{ Var } A_p - 1/4 \text{ Var } A_m = 1/2 \text{ Var } A$$

Quand aucun parent n'est connu, toute la variabilité de A_i, Var A, reste à expliquer. A l'autre extrême, quand les deux parents sont connus, la variabilité qui reste à expliquer et qui ne dépend pas des parents, est la variance de A_i connaissant p et m, soit la variance de W : 1/2 Var A. C'est la variabilité entre pleins frères. Enfin, quand un seul parent est connu, p par exemple, la variabilité restant à expliquer connaissant p, est de 3/4 Var A.

Ainsi nous pouvons construire le tableau suivant, donnant les valeurs de d_i selon la connaissance de la généalogie à évaluer :

Connaissance de la généalogie	Variance de ce qui n'est pas expliqué par la généalogie	d_i
0 parent connu	Var A	1
1 parent connu	3/4 Var A	4/3
2 parents connus	1/2 Var A	2

b / Poids de la descendance

$$p_2 = \Sigma\ d_f / 4 \qquad (17)$$

la somme s'effectuant sur tous les descendants, mâles et femelles, connus. d_f dépend de la connaissance du conjoint de l'animal i à évaluer, conjoint qui est le second parent du descendant f. d_f se définit comme d_i et deux cas peuvent être rencontrés :

conjoint inconnu $d_f = 4/3$ (1 parent connu)
conjoint connu $d_f = 2$ (2 parents connus)

En effet, en appliquant la formule (6) au descendant f, on a :

$$A_f = 0,5\ A_i + 0,5\ A_c + W_f$$

que l'on peut réécrire :

$$A_i = 2\ A_f - A_c - 2\ W_f \qquad (18)$$

Par définition, l'individu i est connu puisqu'on cherche à l'évaluer. L'autre parent, c peut être connu ou non.

En supposant le conjoint c connu, la variance de A_i non expliquée par f et c est Var (2 W) = 4 (Var A)/2 = $(4/d_f)$ Var A puisque $d_f = 2$

Si, par contre, le conjoint c est inconnu, la variance de A_i non expliquée par f vaut

$$\text{Var } A_c + \text{Var } (2W) = \text{Var } A + 4\ (\text{Var } A)/2 = 3 \text{ Var } A = 4\ (3/4) \text{ Var } A = (4/d_f) \text{ Var } A$$

puisque d_f vaut 4/3.

Dans les deux cas, la pondération de l'information d'un descendant, inversement proportionnelle à la variance d'erreur, est donc égale à $d_f/4$.

c / Poids de la moyenne des performances propres

$$P_3 = \Sigma\ w\ h^2 / (1 - r) \qquad (19)$$

la somme s'effectuant sur les performances prises en compte, avec h^2 l'héritabilité du caractère et r sa répétabilité.

Les coefficients de pondération appliqués à chaque information contenue dans le calcul de l'index correspondent à la précision de cette information.

En effet, d'après (4), une performance est décrite par le modèle suivant :

$P = \mu + F + A + V + E$

donc $A = P - \mu - F - V - E$ (20)

Connaissant P, μ, F et V, ce qui n'est pas expliqué dans A est donc E, de variance

$(1-r)\, V_p / w = (1-r)\, VarA / w\, h^2$ (21)

en notant V_p la variance phénotypique du caractère.

Le rapport $h^2/(1-r)$ indique l'intérêt qu'il y a de considérer les performances plutôt que la généalogie de l'animal à évaluer. Il augmente avec l'héritabilité du caractère : en effet la part d'origine génétique transmissible dans les différences observées de performances augmente avec l'héritabilité. Ce rapport est aussi un facteur de traduction qui permet de combiner des performances et des index. Pour les caractères laitiers, les valeurs suivantes sont retenues :

Caractères	h^2	r	$h^2/(1-r)$
Quantités	0,30	0,50	0,60
Taux	0,50	0,70	1,67

Le coefficient w dépend du rang de lactation :

rang de lactation	w
1	1
2	0,8
3	0,8

Les poids des lactations 2 et 3 sont inférieurs à 1 car les performances sont moins héritables qu'en première lactation et leurs corrélations génétiques avec la première lactation est légèrement inférieure à 1.

En outre, si une lactation n'est pas terminée, et donc si elle est extrapolée, w est d'autant plus réduit que la lactation est courte. Ceci traduit le fait qu'une lactation extrapolée est connue avec moins de précision qu'une lactation complète.

Σ w est donc proche du nombre de lactations effectuées, une lactation non terminée comptant pour une fraction de lactation.

Précisons enfin, qu'évidemment, $P_3 = 0$ pour les mâles. Il en est de même d'un certain nombre de femelles sans lactation mais incluses dans l'analyse car elles constituent un maillon d'une chaîne de parenté.

4 / Examen de cas concrets

Un exemple numérique complet est présenté par Ducrocq (1990). Sans détailler les calculs et afin de comprendre l'importance accordée à chaque source d'information, on peut examiner leur contribution relative dans le cadre de quelques cas concrets.

L'index des mâles dépend surtout de la descendance ; l'index des femelles dépend surtout de l'ascendance et des performances propres.

Tableau 1. *Contributions relatives des différentes sources d'information dans le calcul des index des mâles.*

Situation considérée	p_1	p_2	Contribution relative (%)	
			Ascendance	Descendance
20 filles de mère connue	2	10	17	83
50 filles de mère connue	2	25	7	93
1 000 filles de mère connue	2	500	0,4	99,6

P_1 : poids de l'ascendance. P_2 : poids de la descendance.
Hypothèse : les deux parents du mâle sont connus.

Tableau 2. *Contributions relatives des différentes sources d'information dans le calcul des index des femelles.*

Situation considérée	p_1	p_2	Quantité				Taux			
			p_3	Contributions relatives (%)			p_3	Contributions relatives (%)		
				Ascendance	Descendance	Performance		Ascendance	Descendance	Performance
1 parent, 1 lactation	1,33	0	0,60	69	0	31	1,67	44	0	56
1 parent, 2 lactations	1,33	0	1,08	55	0	45	3,00	31	0	69
1 parent, 3 lactations	1,33	0	1,56	46	0	54	4,33	23	0	77
2 parents, 1 lactation	2,00	0	0,60	77	0	23	1,67	54	0	46
2 parents, 2 lactations	2,00	0	1,08	65	0	35	3,00	40	0	60
2 parents, 3 lactations	2,00	0	1,56	56	0	44	4,33	32	0	68
1 parent, 1 lactation 1 fille de père connu	1,33	0,5	0,60	55	20	25	1,67	38	14	48
2 parents, 1 lactation 1 fille de père connu	2,00	0,5	0,60	65	16	19	1,67	48	12	40
2 parents, 3 lactations 1 fille de père connu	2,00	0,5	1,56	50	12	38	4,33	30	7	63
2 parents, 3 lactations 3 filles de père connu	2,00	1,5	1,56	40	30	30	4,33	26	19	55

P_1 : poids de l'ascendance. P_2 : poids de la descendance. P_3 : poids des performances.

4.1 / Ascendance seule connue

Dans ce cas, l'ascendance a évidemment une contribution relative de 100 % et l'index obtenu est un index sur ascendance classique :

$$A_i = 1/2\ A_p + 1/2\ A_m \qquad (21)$$

4.2 / Cas des mâles

Le tableau 1 présente quelques exemples de pondération de l'ascendance avec la descendance dans le cas de mâles d'insémination (descendance nombreuse) ou de monte naturelle.

On constate que l'ascendance n'a une contribution importante à l'index que si le nombre de filles reste faible.

4.3 / Cas des femelles

Les situations rencontrées chez les femelles laitières sont plus variées que celles des mâles. Elles diffèrent selon le caractère, quantité ou taux. Du tableau 2 se dégagent cependant quelques tendances générales :

- l'ascendance a une contribution relative toujours très importante pour les quantités ;
- les performances propres ont un poids prépondérant pour les taux dans la plupart des situations, du fait de l'héritabilité élevée de ces caractères ;
- en l'absence de transfert d'embryons, la descendance a un poids faible mais parfois non négligeable (jusqu'à 1/3).

Le poids relatif de l'index sur ascendance est donc toujours élevé, comparé à celui des performances propres, en particulier pour les quantités. Ceci s'explique essentiellement par le fait que l'index de chacun des parents prend en compte l'information de tous les apparentés, y compris les performances de l'individu à évaluer. L'importance réelle des performances individuelles est donc supérieure à leur pondération dans l'index. D'autre part, malgré un poids relatif modéré, l'influence réelle des performances est souvent prépondérante du fait de leur variabilité nettement plus élevée que celle des index.

Conclusion

La méthode d'indexation, définie par l'application du BLUP à un modèle individuel, permet de tirer parti de l'ensemble de l'information disponible afin de :

- corriger au mieux les performances pour les effets de milieu identifiables
- combiner de façon optimale performances individuelles et information généalogique.

Le changement de méthode intervenu début 1990 pour les bovins doit ainsi permettre d'accroître la fiabilité des index laitiers, tant mâles que femelles, qui, rappelons-le encore, était déjà tout à fait satisfaisante.

Alors que la méthode IF2 était difficile à expliquer et justifier et que la complexité des calculs tend à éloigner les concepteurs des utilisateurs, un atout important du modèle animal est sa relative simplicité théorique qui peut permettre de présenter les principes et les résultats de façon aussi transparente que possible.

La méthode n'est bien entendu pas figée : le modèle, la définition des groupes (nécessaires pour l'estimation de l'index sur ascendance quand les parents sont inconnus), les performances prises en compte, sont susceptibles d'évoluer en fonction des circonstances, des besoins et des résultats d'études ultérieures.

Toutefois, une limite mérite d'être rappelée : une bonne méthode n'est pas suffisante en soi. La fiabilité et la précision des résultats dépendent avant tout de la qualité de l'information brute : généalogies, performances, facteurs de variation.

Références bibliographiques

DUCROCQ V., 1990. Les techniques d'évaluation génétique des bovins laitiers. INRA Productions Animales, 3 (1), 3-16

DUCROCQ V., BOICHARD D., BONAITI B., BARBAT A., BRIEND M., 1990. A pseudo absorption strategy for solving animal model equations for large data files. J. Dairy Sci., 73 (sous presse).

HENDERSON C.R., 1973. Sire evaluation and genetic trend. In Proceeding of the Animal Breeding and Genetics Symposium in honor of Dr J.L.Lush, Blacksburgh, Virginia. August 1972, pp10-41, American Society of Animal Science, Champaign, Illinois.

INRA-ITEB, 1989. Index Economique Laitier (INEL). Compte rendu 1676. 13p.

MOCQUOT J.C., 1978. Etude de la production laitière des bovins. Intérêt des lactations partielles pour la sélection. Prévision de la production totale corrigée pour sa durée. Ann. Génét. Sél. Anim., 10, 359-376.

POUTOUS M., MOCQUOT J.C., 1975. Etude de la production laitière des bovins. III. Relations entre critères de production, durée de lactation et intervalle entre premier et deuxième vêlages. Ann. Génét. Sél. Anim., 7, 181-189.

POUTOUS M., BRIEND M., CALOMITI S., DOAN D., FELGINES C., STEIER G., 1981. Méthode de calcul des index laitiers. Bull. Tech. Inf., 361, 433-446

Summary

The french method of genetic evaluation of dairy animals.

The main characteristics of the dairy traits evaluation system applied in France from 1990 are presented, with emphasis on the differences with respect to the previous system. The BLUP method, applied to an animal model, takes into account all the available information, which is summarized for each animal into three synthetic components : its pedigree value (i.e. the average of its parents breeding values (BV)) ; twice the weighted average of the BV of its progeny, adjusted for its mate ; the weighted average of its own records, adjusted for environmental effects. The weight of each component is explained, justified and illustrated with some examples.

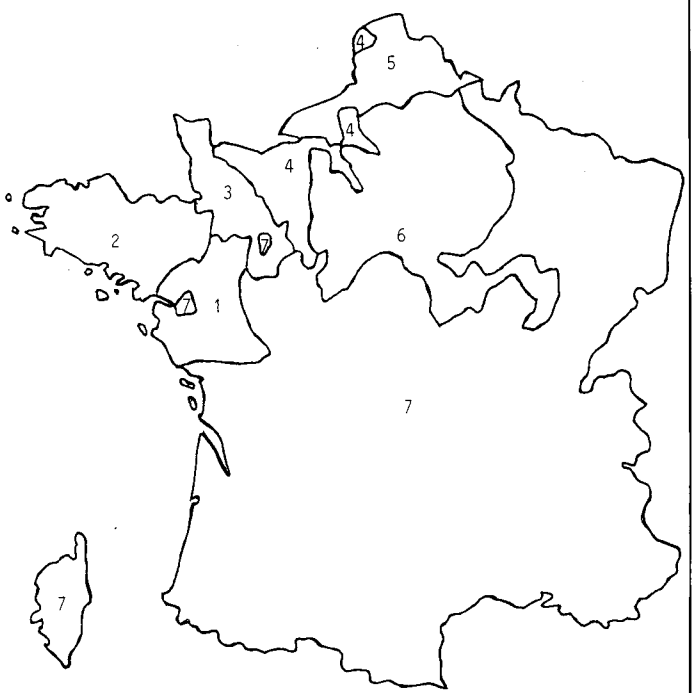
BONAITI B., BOICHARD D., VERRIER E., DUCROCQ V., BARBAT A., BRIEND M.,1990. La méthode française d'évaluation génétique des reproducteurs laitiers. INRA Prod. Anim., 3(2), 83-92.

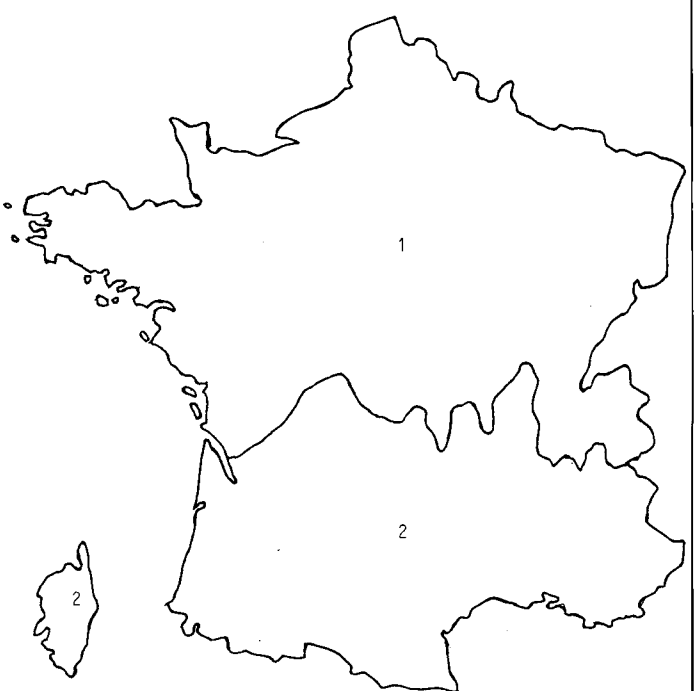
Annexe. *Carte des régions*

Les régions sont définies intra race pour le calcul des effets de l'âge, du mois de mise bas et de l'intervalle entre mise bas par année et numéro de lactation. Elles constituent des zones à la fois assez importantes pour une bonne estimation des différents effets, tout en restant aussi homogènes que possible. Elles sont basées sur les régions du Groupe d'Economie Bovine de l'ITEB, constituées de regroupement de petites régions agricoles INSEE. Elles sont bien sûr à interpréter en fonction de la densité des animaux de chaque race. Pour les races non mentionnées, une seule région est définie.

Race Frisonne : 8 régions

Race Normande : 7 régions



Race Montbéliarde : 3 régions

Race Brune : 2 régions