



HAL
open science

Le point sur la notion de connexion en génétique animale

Eric Hanocq, Laurent Tiphine, Bernard B. Bibé

► **To cite this version:**

Eric Hanocq, Laurent Tiphine, Bernard B. Bibé. Le point sur la notion de connexion en génétique animale. *Productions Animales*, 1999, 12 (2), pp.101-111. hal-02697483

HAL Id: hal-02697483

<https://hal.inrae.fr/hal-02697483>

Submitted on 1 Jun 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.,
1999, 12 (2), 101-111*

E. HANOCQ¹, L. TIPHINE^{}, B. BIBÉ
INRA Station d'Amélioration Génétique
des Animaux, BP 27,
31326 Castanet-Tolosan Cedex*

^{} Institut de l'Élevage, Département Génétique,
Identification et Contrôle de Performances,
149 rue de Bercy, 75595 Paris*

*¹Correspondance : LGAP, Domaine
Brunehaut 80200 Estrées-Mons*

hanocq@mons.inra.fr

Le point sur la notion de connexion en génétique animale

Les schémas de sélection des animaux domestiques se sont progressivement développés pour souvent atteindre une dimension nationale, voire internationale. L'évaluation génétique des animaux nécessite de dissocier, dans les performances des animaux, les effets génétiques des effets environnementaux. Pour cela, une répartition adéquate des performances dans les différents milieux est primordiale. Elle doit permettre la comparaison, ou connexion, de tous les animaux du schéma, et ainsi dégager un progrès génétique optimal.

La performance d'un animal est sous l'influence combinée de facteurs génétiques (le génotype de l'animal) et de facteurs environnementaux (les conditions de milieu). Seule une partie des caractéristiques du génotype des parents est transmissible aux descendants.

Le sélectionneur a besoin d'avoir une estimation précise de cette partie génétique transmissible non observable afin de choisir les animaux qu'il destine à la reproduction. En génétique animale, cette partie génétique transmissible – dite valeur génétique additive – est souvent simplement appelée "valeur génétique".

Pour gagner en efficacité, les schémas de sélection, en particulier ceux dont au moins une partie est développée en ferme, doivent reposer sur une base de sélection la plus large possible. Avec l'extension des schémas de sélection, les performances des animaux prises en compte dans l'évaluation génétique sont donc réalisées dans des conditions de milieu très différentes, qu'elles soient spatiales (zones géographiques) ou temporelles (période et âge des reproducteurs). Pour pouvoir comparer le niveau génétique des animaux candidats à la sélection, il est absolument indispensable de dissocier et d'estimer sans confusion possible l'importance des différents facteurs de variation génétiques et environnementaux connus et identifiés.

Les méthodes d'évaluation génétique, comme les structures, ont évolué. La méthodologie "BLUP" (Henderson 1973) (meilleure prédiction linéaire non biaisée, en français) s'est progressivement imposée depuis la fin des années 70. Dans sa version actuelle, elle s'applique à un modèle "animal" qui permet de prendre en compte l'ensemble des généalogies ; elle offre par ailleurs dans le cadre d'un

Résumé

La notion de connexion est présentée dans son contexte statistique d'origine, puis définie dans le cadre plus spécifique de la génétique animale. Une connexion satisfaisante, à savoir une répartition suffisamment équilibrée des performances dans les différents niveaux des facteurs de variation génétiques ou environnementaux, est indispensable pour pouvoir estimer et comparer les facteurs (effets "troupeau", valeurs génétiques ...) inclus dans un modèle d'évaluation génétique. Elle est mesurée à l'aide d'un critère dont les valeurs varient de façon continue entre 0 (absence de connexion) et 1 (équilibre optimal). Par ailleurs, si les estimations sont réalisables, elles doivent être suffisamment précises pour pouvoir être exploitées et donc s'appuyer sur un nombre de performances assez important. Un critère est aussi présenté pour mesurer cette précision. Ces deux critères sont utilisés à titre d'exemple dans le cadre d'un dispositif théorique. Une telle approche permet aussi d'illustrer la perte de progrès génétique qu'entraîne l'absence de connexion. Dans le contexte d'un schéma de sélection, les facteurs géographiques et temporels représentent des facteurs de risques quant à l'absence de connexion. Les difficultés de perception par le sélectionneur d'un défaut de connexion sont soulignées. Elles amènent à préconiser des études spécifiques de la connexion et de manière plus générale de la nature et de la représentativité des performances incluses dans une évaluation génétique. Enfin, les principaux cas de figure auxquels peuvent être confrontés les schémas de sélection (connexion absente, insuffisante ou satisfaisante) et la stratégie correspondante du sélectionneur sont discutés.

La notion de connexion intègre les valeurs génétiques des animaux, leurs performances et les milieux dans lesquels celles-ci sont obtenues.

modèle polygénique strictement additif des propriétés d'optimalité d'estimation des valeurs génétiques (Ducrocq 1990).

En théorie, l'outil statistique est donc adapté à l'évaluation génétique nationale, mais la réalisation des propriétés théoriques n'est effective que si la structure de l'information disponible (performances et relations de parenté) pour effectuer les estimations le permet. Dans cet article, le terme structure désigne la répartition des performances, ou des individus apparentés, dans les différents milieux. Une bonne répartition permet une bonne comparaison de la valeur génétique des animaux ou, de façon équivalente, des effets de milieu (élevage par exemple) ; on parle alors d'une bonne connexion. La notion de connexion ne doit pas être dissociée de la notion de précision des estimations. Même si les comparaisons sont réalisables, pour qu'elles puissent être exploitées, elles doivent s'appuyer sur un nombre suffisant de données.

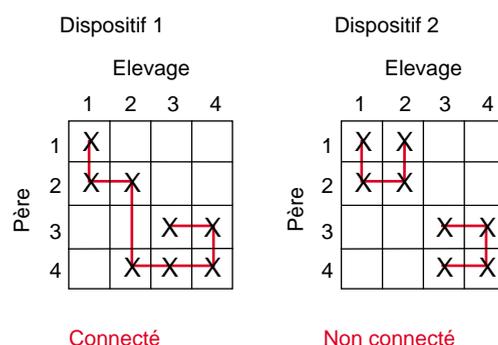
L'objectif de cet article est de faire le point sur la notion de connexion pour communiquer aux acteurs de la sélection animale les principaux enseignements tirés des travaux de la recherche menés sur ce thème. Des outils d'analyse ont été développés, une illustration de leur utilisation sera présentée à travers quelques exemples. L'importance de la notion de précision est toujours rappelée. L'accent est aussi mis sur les conséquences que peut avoir un déficit de connexion sur les estimations réalisées et sur l'efficacité de la sélection. Enfin, quelques solutions sont proposées pour contrôler le niveau de connexion dans un schéma de sélection.

1 / La notion de connexion en statistique

1.1 / La connexion : un concept "tout ou rien"

La notion de connexion a été décrite et formalisée depuis une cinquantaine d'années dans le domaine des statistiques. Il s'agit d'un concept relativement abstrait ; nous l'illustrerons par un exemple. Le tableau 1 représente la descendance de 4 taureaux répartie dans

Tableau 1. Représentation schématique d'un jeu de données soumis à deux facteurs de variation (père et élevage). Les croix matérialisent la présence de descendants du père n dans l'élevage n .



plusieurs élevages. L'existence de données y est représentée par une croix.

Un dispositif est dit connecté s'il permet l'estimation de l'incidence de tous les effets. Dans le tableau 1, cela suppose de pouvoir estimer pour chaque facteur l'incidence des niveaux 1 à 4.

Searle (1986), reprenant un certain nombre de travaux statistiques antérieurs, propose une règle graphique pour déterminer si un dispositif est connecté ou non : un dispositif est déclaré connecté si toutes les croix (cellules avec performances, ou cellules pleines) peuvent être reliées par une ligne continue formée de segments horizontaux et/ou verticaux dont les changements de direction s'effectuent sur une cellule pleine. Les traits verticaux matérialisent la connexion entre deux pères. En pratique, ces deux pères, ayant des descendants dans un même élevage, la différence entre la valeur moyenne des performances des descendants de chacun de ces pères reflète alors, en espérance, une différence génétique. Dans ce cas, la différence génétique entre les deux pères est estimable, leur comparaison a un sens, ils sont dits connectés entre eux. De manière équivalente, les traits horizontaux matérialisent la connexion entre élevages. Dans le cas du tableau 1, il est possible de conclure que le dispositif 1 est connecté, tous les pères et tous les élevages sont comparables, alors que le dispositif 2 ne l'est pas. Dans ce dernier cas, il est par exemple impossible de comparer la valeur génétique des pères 2 et 3 : la comparaison confond les effets "père" et "élevage". La structure des données, c'est-à-dire la répartition des performances sur lesquelles sont basées les estimations, n'est pas adéquate. Dans le cas du dispositif 2, les pères (et les élevages) 1 et 2, d'une part, et 3 et 4, d'autre part, sont tout de même connectés deux à deux.

1.2 / La connexion : un concept graduel

Dans le cas du dispositif 1, la structure des données permet la connexion de tous les pères entre eux et de tous les élevages entre eux. Toutefois, certaines comparaisons ne peuvent pas être faites directement à cause de l'absence de données dans certaines cellules. Par exemple, il est possible de comparer directement les pères 1 et 2, car ils ont des performances dans le même élevage 1, les pères 2 et 4, car ils ont des performances dans le même élevage, par contre les pères 1 et 4 peuvent être comparés l'un à l'autre uniquement de façon indirecte, car ils peuvent l'être tous les deux au père 2. Le fait que la comparaison soit indirecte a pour conséquence de diminuer la précision des estimations envisagées. Caractériser un dispositif de façon "tout ou rien" n'est donc pas pleinement satisfaisant. Même connecté, il peut être plus ou moins bien adapté aux estimations envisagées.

Le concept doit être étendu. La connexion caractérise encore la structure des données, mais, dans le cas d'un dispositif connecté, elle doit permettre en plus de distinguer la qualité

de la répartition des données dans les différents niveaux des facteurs de variation, encore appelée "équilibre ou orthogonalité du dispositif". Cette extension permet une mesure continue de la connexion.

2 / La connexion en sélection animale

2.1 / L'approche continue

Foulley *et al* (1984) introduisent le concept de connexion dans le domaine de la génétique animale à travers cette notion de mesure continue de l'état d'équilibre d'un dispositif. Le dispositif non connecté représente alors un cas extrême de déséquilibre. Le dispositif parfaitement connecté, ou parfaitement équilibré, est un dispositif où toutes les cellules sont pleines et où les données sont également réparties dans les niveaux des facteurs. Dans l'exemple du tableau 1, tous les pères devraient avoir le même nombre de descendants par élevage. Le dispositif est alors qualifié de complet et équilibré.

2.2 / Critères de mesure

Pour juger de la précision des estimations réalisées, les généticiens utilisent le concept de "variance d'erreur" (VE). Ils estiment ainsi la variance de la différence entre les valeurs estimées et les valeurs vraies. Plus cette variance est importante, plus l'erreur faite sur les estimations risque d'être grande et donc la précision faible. La valeur de cette variance est difficilement interprétable en tant que telle à cause de sa dépendance à l'unité de mesure ; il est commode de l'exprimer de façon relative, en la comparant à une variance de référence.

De façon générale, le facteur dont on mesure la connexion est appelé "facteur d'intérêt", le facteur qui peut créer un manque de connexion étant appelé "facteur de nuisance". En pratique, la VE observée est calculée en tenant compte de tous les facteurs de variation (d'intérêt et de nuisance), en d'autres termes avec un modèle statistique complet. Elle est notée VE_C . A l'opposé, la VE de référence est calculée en ne tenant pas compte du facteur de nuisance, donc à l'aide d'un modèle statistique réduit. Elle est notée VE_R . Le critère (γ) proposé par Foulley *et al* (1984, 1990, 1992) pour mesurer la connexion s'exprime sous la forme d'un rapport de variances : $\gamma = VE_R / VE_C$

La notation "VE" utilisée ici correspond à la notation anglaise "PEV" introduite par les auteurs pour désigner la "Variance d'Erreur de Prédiction". Ce critère, appelé coefficient de connexion, mesure l'augmentation relative de la variance d'erreur du fait de la présence d'un facteur de nuisance et d'une structure de données plus ou moins bonne. Il varie de façon continue entre 0 et 1. Une valeur de 1, indiquant que la VE observée (VE_C) est égale à la VE de référence (VE_R), traduit le fait que la structure du dispositif observé (ou des données) est optimale quant à la connexion. En

d'autres termes, dans un dispositif orthogonal, la présence du facteur de nuisance n'affecte pas la précision d'estimation du facteur d'intérêt. Une valeur proche de 0 traduit le fait que la structure du dispositif n'est pas adaptée aux comparaisons envisagées. Le facteur de nuisance dégrade de façon importante la précision des estimations, voire empêche totalement leur réalisation.

Dans le cas de l'étude de la connexion, on s'intéresse non pas à la VE relative à l'estimation d'un niveau donné de facteur, mais à celle relative à la comparaison de deux (ou plusieurs) niveaux. Dans les dispositifs du tableau 1, la comparaison peut concerner soit les troupeaux, soit les pères. Considérer le facteur "troupeau" permet de se placer dans le contexte général de l'analyse de données expliquées par un certain nombre de facteurs de variation. L'effet du facteur est estimé à partir des seules données, le facteur est dit "fixe". En génétique animale le facteur "troupeau" est souvent considéré comme tel. L'étude du facteur "père" permet d'illustrer les spécificités que procure l'analyse d'un facteur génétique. Classiquement, les animaux sont supposés appartenir à une même population. Les caractéristiques de cette population peuvent être synthétisées à l'aide d'outils statistiques (distribution des valeurs génétiques, moyenne, variance ...). En génétique quantitative, un caractère est supposé influencé par de nombreux gènes ayant chacun un effet faible. Ce déterminisme polygénique permet d'expliquer le fait que c'est le plus souvent la loi normale qui est choisie pour représenter la distribution des valeurs génétiques des animaux. On caractérise donc la connaissance *a priori* que l'on a de la valeur génétique d'un animal à évaluer par cette distribution. Ce procédé équivaut en statistique classique à supposer que les effets génétiques sont des effets aléatoires. Ainsi, outre l'information apportée par les données, des informations génétiques (relations de parenté ...) et statistiques, dites "*a priori*", sont utilisées. La valeur de l'héritabilité du caractère étudié détermine le poids donné à l'information *a priori*.

Dans le cas d'un facteur fixe, pour quantifier l'influence de la structure des données sur la précision des comparaisons envisagées, la VE observée est exprimée par rapport à la VE qui serait obtenue en l'absence de problèmes de connexion. Avec un même nombre total de données, une telle situation peut être représentée, soit par un dispositif complet et équilibré comme décrit précédemment, soit par un dispositif dans lequel il n'y aurait qu'un seul père dans le cas d'une étude de la connexion des troupeaux. Dans ce dernier cas l'effet "père" n'a plus lieu d'être pris en compte (puisque'il n'y en a plus qu'un) car il ne va pas engendrer de différences entre performances.

Dans le cas d'un effet aléatoire, pour mesurer la précision de comparaison permise par les données, sans tenir compte de l'information statistique *a priori*, la VE observée est relativisée à la variance génétique additive suivant le même principe que pour le coefficient de détermination (CD) individuel. Le CD concerne une comparaison (Laloë 1993) : par

Pour comparer les valeurs génétiques des animaux candidats à la sélection, il faut s'affranchir des effets du milieu sur leurs performances.

exemple dans le cas du tableau 1, le CD (noté CD_x) ne mesure pas la précision de l'estimation de la valeur génétique d'un des deux pères mais celle de la différence de valeur génétique entre ces pères. Comme pour le critère γ , le CD varie entre 0 et 1.

3 / Illustration numérique

3.1 / Utilisation du critère de mesure de la connexion

Le tableau 2 décline plusieurs cas d'un dispositif très simple : un dispositif croisé pères (2 niveaux) x troupeaux (2 niveaux). Les descendants des deux pères sont plus ou moins bien répartis dans les deux troupeaux. Leurs performances peuvent servir à la comparaison des effets "élevage" ou des valeurs génétiques des pères. Les pères 1 et 2 ont chacun 40 descendants (80 dans le dispositif F). La répartition de ces descendants diffère d'un dispositif à l'autre. Le dispositif A montre une situation équilibrée, chaque père a la moitié de ses descendants dans chacun des deux troupeaux. A l'opposé, dans le dispositif E chaque père a ses descendants dans un seul des deux troupeaux.

En s'intéressant à la connexion des troupeaux, ou des pères si on les considère en tant qu'effet fixe (on ne tient compte que de l'information apportée par les données), la valeur prise par le critère γ varie entre 1, quand le dispositif est complet et équilibré (dispositif A), et 0, quand le déséquilibre est extrême (dispositif E non connecté). En intégrant l'information a priori sur les pères, le CD de contraste entre les valeurs génétiques

des deux pères est compris entre 0,78 (dispositif A) et 0 (dispositif E). Dans ce dernier cas, les effets de milieu (ici troupeau) et génétique qui conditionnent la valeur de la performance sont confondus. Le dispositif ne permet pas d'estimer la part de l'une et l'autre de ces composantes. Par un phénomène de vase communicant, des différences dues à un effet de milieu peuvent ainsi être considérées comme des différences génétiques et vice versa. La comparaison des effets "élevage" ou des valeurs génétiques des pères n'est pas valable avec un tel dispositif. La diminution progressive de la valeur du CD du contraste entre les deux pères traduit la diminution du nombre de performances sur lesquelles est basée la comparaison et, dans le cas d'une absence de connexion, la diminution du nombre de pères pouvant réellement être comparés.

La notion de continuité introduite par Foulley et al (1984) permet de rendre compte de la détérioration progressive de la structure des dispositifs considérés (en passant du dispositif A au dispositif E) pour comparer la valeur génétique des pères ou les effets "troupeau". De façon intuitive, il est clair que la comparaison des pères ou des troupeaux est de plus en plus difficile de la situation A à la situation D. Elle n'a aucun sens en E ($\gamma = 0$ et $CD = 0$). La comparaison, par exemple de la valeur génétique des pères, repose sur la présence simultanée dans un même troupeau de certaines de leurs performances. Dans le cas B par exemple, si l'on suppose l'appartenance du père 1 au troupeau 1 et du père 2 au troupeau 2, les 10 performances du père 1 dans le troupeau 2 et les 10 performances du père 2 dans le troupeau 1 permettent la comparaison directe de leurs valeurs génétiques. Cette

Les comparaisons sont les plus précises possible si tous les animaux ont leurs performances également réparties dans tous les milieux considérés : le dispositif est alors parfaitement connecté.

Tableau 2. Les différents dispositifs étudiés et leurs caractéristiques.

		Père	Père	γ	CD_x^a	Dispositif équilibré équivalent ^b		Dispositif équivalent au A	
		1	2						
Dispositif A	Troupeau 1	20 ^c	20	1	0,78	20	20	20	20
	Troupeau 2	20	20			20	20	20	20
Dispositif B	Troupeau 1	30	10	0,75	0,73	15	15	40	13
	Troupeau 2	10	30			15	15	13	40
Dispositif C	Troupeau 1	35	5	0,44	0,61	9	9	80	11
	Troupeau 2	5	35			9	9	11	80
Dispositif D	Troupeau 1	39	1	0,10	0,26	2	2	400	10
	Troupeau 2	1	39			2	2	10	400
Dispositif E	Troupeau 1	40	0	0	0	0	0	/	/
	Troupeau 2	0	40			0	0	/	/
Dispositif F	Troupeau 1	40	40	1	0,88	40	40	40	40
	Troupeau 2	40	40			40	40	40	40

^a CD de la comparaison des valeurs génétiques des pères 1 et 2 (héritabilité : 0,33)

^b Équivalence en terme de précision de comparaison entre les deux troupeaux ou pères

^c Nombre de descendants du père 1 dans le troupeau 1

comparaison, à effectif total constant (40 descendants par père) ne repose donc que sur 2x20, 2x10, 2x5, 2x1 et 0 données, pour les situations A, B, C, D et E, respectivement.

Le critère γ prend la valeur de 1 pour les deux dispositifs complets et équilibrés, A et F, différenciés uniquement par le nombre de performances. Le CD quant à lui vaut 0,78 pour le dispositif A et 0,88 pour le F. La juxtaposition de ces résultats illustre le fait que, par construction (choix de la variance de référence), le critère γ ne s'intéresse qu'à la structure des données, alors que le CD mesure une quantité d'information.

La qualité globale d'un dispositif dépend à la fois des aspects structure et quantité. En terme de quantité, il est intéressant de mesurer l'effort qu'il faudrait consentir (sur la taille de la descendance dans cet exemple) pour compenser la perte de précision observée en raison de la dégradation de la structure des données. Pour cela, l'expression littérale de la VE a été développée ; elle intègre à la fois les paramètres de taille de la descendance et de sa répartition. A l'inverse, la précision observée en situation de déséquilibre pourrait être obtenue avec une taille moindre si le dispositif était équilibré. Nous appellerons cette taille, la taille de descendance efficace. Par exemple, en conservant la structure du dispositif B (l'effectif des différentes cases varie dans une même proportion), il faudrait 53 descendants par père pour compenser le déséquilibre dans la répartition des données et retrouver la précision obtenue avec le dispositif A, parfaitement équilibré avec 40 descendants par père. L'approche inverse permet de voir que le dispositif B, déséquilibré, procure la même précision qu'un dispositif complet équilibré (du type A) dans lequel chacun des pères aurait 30 descendants (15 dans chaque troupeau). Le seul déséquilibre aboutit au fait que, en terme de précision, les 40 descendants "n'en valent" finalement que 30 efficaces. Le dispositif D, et a fortiori le E, illustrent combien peut être grande la perte d'efficacité d'un dispositif du seul fait de sa structure, et donc d'une connexion plus ou moins déficiente.

Dans l'absolu, si A et F ont la même structure, le dispositif F est bien sûr préférable car les comparaisons sont basées sur 2x40 données au lieu de 2x20. La traduction des VE en terme de nombre de descendants efficaces est explicite (40 pour le dispositif A et 80 pour

le F). Les valeurs des CD de contraste illustrent bien. Cette supériorité du dispositif F est aussi traduite par la valeur des CD individuels des animaux (0,39 et 0,44 dans les dispositifs A et F, respectivement). Il est important de noter que, malgré un nombre correct de données par père, les CD individuels sont assez faibles en raison du faible nombre de pères inclus dans le dispositif.

3.2 / Spécificité de la génétique animale

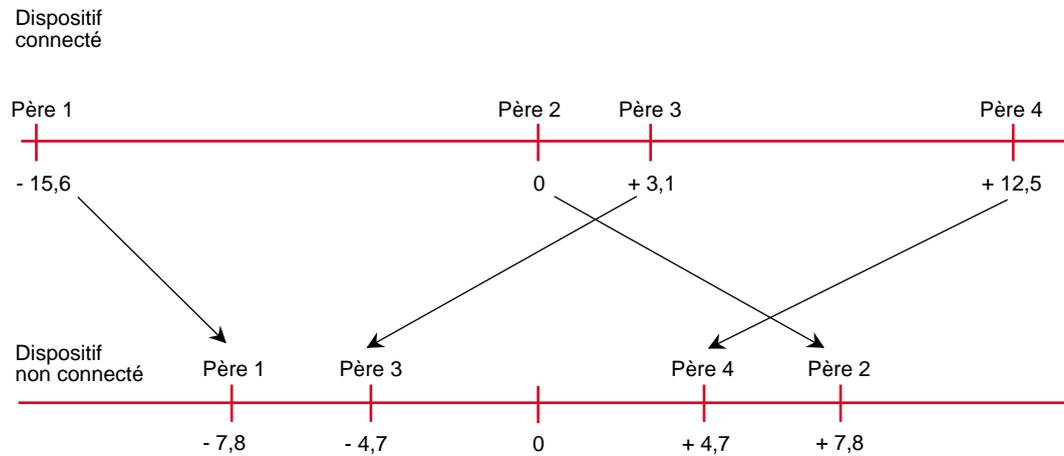
Lorsque qu'un facteur est considéré comme aléatoire, comme précisé précédemment, les estimations reposent à la fois sur l'information *a priori* (génétique et statistique) et l'information apportée par les données. Dans le cas du dispositif E, il n'y a pas d'information apportée par les données pour la comparaison des effets "père" ou "troupeau" ($\gamma = 0$ et $CD_x = 0$). En revanche, si l'effet "père" est considéré comme aléatoire, l'information *a priori* est toujours présente. Elle permet donc la réalisation pratique des estimations, mais la présence de cette seule information minimale peut avoir des conséquences néfastes sur la qualité de ces estimations si la connexion par les données n'est pas assurée. L'exemple du tableau 3 permet d'illustrer ces propos.

Deux dispositifs sont considérés : l'un est complet et équilibré ; le second, comme le montrerait l'application de la règle de Searle (1986) n'est pas connecté par les données. Dans ce dernier dispositif, seuls les pères 1 et 2 d'une part, et 3 et 4 d'autre part sont connectés deux à deux. Les performances des descendants des 4 pères sont sous l'influence du facteur "troupeau" à deux niveaux. Pour le dispositif connecté, les valeurs génétiques centrées sur 0, estimées à l'aide d'un modèle prenant en compte les deux facteurs de variation, augmentent du père 1 au père 4 (figure 1). Dans le dispositif non connecté, afin d'illustrer l'incidence sur les valeurs estimées de la seule structure du dispositif, le nombre de descendants de chaque père (40) est identique à celui du dispositif connecté, mais leur répartition est différente. Cette seule répartition entraîne des biais d'estimation très importants, le classement des pères est en effet largement modifié. Par exemple, le père 2 qui était troisième devient premier. La structure des données ne permet pas d'es-

Tableau 3. Répartition dans deux troupeaux et moyenne des performances des descendants de pères pour deux dispositifs utilisés pour l'évaluation génétique sur le gain moyen quotidien entre 3 et 70 jours (GMQ 30-70, en grammes) chez les ovins.

Dispositif	Troupeau	Répartition des descendants				GMQ moyen par père et troupeau			
		Père 1	Père 2	Père 3	Père 4	Père 1	Père 2	Père 3	Père 4
Dispositif connecté	Troupeau 1	20	20	20	20	360	380	384	396
	Troupeau 2	20	20	20	20	330	350	354	366
Dispositif non connecté	Troupeau 1	40	40	0	0	360	380	/	/
	Troupeau 2	0	0	40	40	/	/	354	366

Figure 1. Influence de la connexion sur l'estimation de la valeur génétique des pères (l'héritabilité du GMQ 30-70 a été fixée à 0,33).



Le classement des animaux sur un index de sélection peut être modifié par l'absence de connexion du dispositif d'évaluation.

timer sans confusion les effets "père" et "troupeau". Des performances moindres en raison d'un effet "troupeau" défavorable sont attribuées à la valeur génétique considérée alors à tort comme moindre. A l'inverse, un effet "troupeau" favorable aboutit à une surestimation des valeurs génétiques. Le manque de connexion par les données peut donc avoir pour conséquence de sélectionner des animaux qui ne devraient pas l'être, les meilleurs animaux pouvant être éliminés. Le progrès génétique alors réalisé peut être fortement diminué, uniquement à cause d'une répartition non optimale des données. La figure 1 permet d'illustrer que, même dans le dispositif globalement non connecté, les pères connectés entre eux, puisque utilisés dans un même milieu, sont comparables. Leur classement respectif est bon et la différence de valeur génétique est bien estimée. L'écart entre les pères 1 et 2 est de 15,6 dans les deux dispositifs. Le problème se situe dans la comparaison du groupe des pères 1 et 2, avec le groupe des pères 3 et 4. En l'absence de connexion par les données, par hypothèse les pères sont supposés appartenir à une même population, donc ils deviennent théoriquement comparables. Mais comme il n'y a pas de données permettant de préciser le niveau génétique relatif des deux groupes de pères connectés (1 et 2 / 3 et 4), la différence génétique est supposée nulle (figure 1).

Les pères 1 et 2 d'une part, et 3 et 4 d'autre part, étant regroupés, le CD du contraste entre ces deux groupes est de 0,78 pour le dispositif connecté alors qu'il est de 0 dans le cas contraire. Même si la réalisation pratique des estimations est possible, la valeur nulle du CD de contraste montre qu'elles ne sont pas exploitables car insuffisamment précises.

En conclusion, sans information a priori sur les effets, quand le dispositif n'est pas connecté par les données, l'estimation conjointe des facteurs de variation n'est pas mathématiquement réalisable. A l'opposé, si l'effet est considéré comme aléatoire, la comparaison des valeurs génétiques peut théoriquement être faite. Toutefois, si les animaux ne sont pas apparentés, l'absence de données ne permet en rien de distinguer deux animaux considérés comme issus d'une même popula-

tion, ils sont donc supposés avoir la même valeur génétique. Bien que réalisable, la comparaison peut être fortement biaisée. Seul un contrôle du niveau de connexion du dispositif par les données permet de déceler une telle éventualité.

Les deux dispositifs envisagés ici pour l'estimation des valeurs génétiques sont extrêmes quant à la connexion. La connexion, même si elle n'est pas totalement absente, peut aussi s'avérer très faible. Les estimations reposent dans ce cas sur un nombre de données très restreint, ce qui augmente les risques que les données soient soumises à des facteurs non contrôlés et inconnus. Des risques élevés de biais sont donc à craindre aussi dans ce type de situation.

4 / Implications dans les schémas de sélection

4.1 / Des risques de mauvaise connexion

Les dispositifs présentés ci-dessus sont volontairement très simples, mais dans leur principe ils sont tout à fait généralisables. Replacés dans le contexte d'un schéma de sélection, les notions de connexion, de facteurs de nuisance et d'intérêt conservent toute leur signification.

La notion de connexion recouvre les caractéristiques communes (environnementales ou génétiques) des individus, milieux ou performances sur lesquelles sont basées les comparaisons. Les facteurs de nuisance et d'intérêt varient selon le schéma de sélection et l'objectif de l'étude. D'une façon générale, la présence de facteurs de nuisance est liée au fait que les performances sont effectuées dans des lieux différents et/ou à des moments différents. A titre d'exemple, les facteurs "troupeau", "station de contrôle", "région", "pays" peuvent apparaître comme facteurs de nuisance spatiaux, alors que les facteurs "saison", "année", "numéro de la bande de contrôle" sont des facteurs de nuisance temporels. Le plus souvent d'ailleurs, les deux types de

facteurs coexistent. Dans le cadre des évaluations génétiques, les facteurs "père", "animal" ou encore "groupe d'animaux" sont des facteurs classiques. Cependant, pour étudier l'incidence de facteurs non génétiques, il est aussi possible d'inverser le rôle donné à ces facteurs et de s'intéresser à la connexion entre troupeaux, pour l'étude des effets "élevage" par exemple.

En conclusion, quel que soit le schéma de sélection, dès qu'il y a une certaine compartimentation de la population selon des facteurs temporels ou spatiaux, la connexion peut être déficiente. Avec une mauvaise connexion, la perte d'efficacité est inévitable.

4.2 / Effet d'un manque de connexion sur l'efficacité d'un schéma de sélection

L'évaluation génétique des animaux représente une étape clé du processus de sélection. Elle permet le classement des animaux candidats à la sélection sur la base des valeurs génétiques estimées. Il doit être le plus proche possible de celui qui pourrait être établi sur les valeurs génétiques réelles si celles-ci étaient connues. La qualité de l'évaluation génétique dépend de nombreux facteurs, en particulier de la structure des données.

Ainsi qu'illustré précédemment, un manque de connexion peut entraîner des erreurs de classement des animaux, la sélection n'est alors pas optimale et le potentiel génétique du schéma de sélection est sous-utilisé. Hanocq *et al* (1996) ont illustré ce phénomène par simulation dans un schéma de testage sur descendance compartimenté en sous-populations (centres de production de semence). L'héritabilité était de 0,25 et, à la première génération, la différence de niveaux génétiques minimale entre centres était de 0,4 écart type génétique. La perte de progrès génétique s'élevait, à chaque génération, à 20 % du progrès qui aurait pu être réalisé si il n'y avait pas eu de problème de connexion. Ils ont aussi montré que ce mauvais choix se répercutait sur les générations suivantes : le manque de connexion à un instant donné a aussi un effet néfaste durable. Ces résultats montrent bien qu'il est impératif de savoir diagnostiquer l'existence d'une connexion déficiente dans un schéma de sélection.

4.3 / Bilan du niveau de connexion dans un schéma de sélection

a / Difficulté de diagnostic

- au niveau du schéma lui-même

Le fait de procéder par simulation facilite l'étude de l'incidence d'un mauvais niveau de connexion sur le fonctionnement d'un schéma de sélection. En effet, même si les valeurs génétiques sont estimées comme en réalité, les valeurs génétiques "vraies" sont elles aussi connues. La validité des estimations peut donc être vérifiée par comparaison aux vraies valeurs. Dans la réalité, le sélectionneur ne dispose que d'une seule source d'information,

les estimations de valeurs génétiques. Si cette estimation est erronée comme dans l'exemple de la figure 1, l'image qu'il a de la réalité n'est pas fidèle, mais il ne peut pas s'en apercevoir au seul examen des résultats obtenus. Pour cela il doit faire une étude spécifique de la connexion dans le schéma. Il est aussi important de rappeler que, pour permettre une estimation correcte du progrès génétique, les connexions doivent être assurées dans l'espace, à un moment donné (pour comparer entre eux les animaux de l'ensemble du schéma) et dans le temps (pour comparer d'une année sur l'autre les valeurs obtenues). Contrairement à une idée reçue, le "BLUP-modèle animal" n'est pas garant d'une estimation sans biais et précise du progrès génétique (Ollivier 1994) ; celle-ci requiert en particulier certaines contraintes sur la structure du dispositif analysé (générations chevauchantes, accouplements répétés, lot témoin) qui concourent à la réalisation d'une bonne connexion.

- dans un contexte concurrentiel

Le manque de connexion pose problème pour le fonctionnement interne d'un schéma, mais il devient encore plus pénalisant dans un contexte concurrentiel. Le progrès génétique n'est pas un indicateur suffisant de la performance d'un schéma, il faut aussi pouvoir situer le niveau génétique de la population par rapport aux concurrents. Si chaque schéma fonctionne indépendamment des autres, il est impossible de comparer le niveau génétique des différents schémas et de faire un classement unique des animaux de l'ensemble de la population. Disposer de ce classement est pourtant primordial pour l'élaboration d'une stratégie de sélection et pour l'instauration et l'orientation des échanges commerciaux nationaux et internationaux.

b / Mesure objective de la connexion dans les schémas de sélection

Le critère γ permet de mesurer objectivement la connexion dans les schémas de sélection même si ceux-ci sont parfois très complexes. L'utilisation de ce critère paraît disproportionnée pour des dispositifs aussi simples que ceux présentés dans les tableaux 2 et 3, mais devient très vite impérative dans les situations réelles. Des algorithmes informatiques (Weeks et Williams 1964, Fernando *et al* 1983) permettent d'identifier les éléments (génétiques ou environnementaux) non connectés, ou encore les sous-ensembles connectés de la population. La mesure continue de la connexion suppose le calcul de très nombreux coefficients de connexion γ entre tous les animaux ou troupeaux pris deux à deux. La simple moyenne de tous les coefficients calculés étant très peu informative, il est commode de les analyser à l'aide de techniques statistiques de classification. Ces techniques permettent une visualisation graphique synthétique de la structure de la population et par conséquent l'identification aisée des éléments pas ou peu connectés au reste de la population. Hanocq *et al* (1992) ont utilisé ce type d'approche en races bovines allaitantes Limousine et Maine Anjou et ont ainsi

pu illustrer le rôle majeur que tenaient les taureaux d'insémination artificielle dans l'établissement de la connexion des taureaux de monte naturelle, utilisés pour la plupart dans un seul troupeau.

La démarche générale utilisée pour étudier la connexion consiste à identifier les facteurs de nuisance pouvant exister dans le schéma analysé. Bien que pouvant être multiples, souvent un seul d'entre eux s'avère être limitant pour la connexion des facteurs d'intérêt. L'utilisation du critère γ permet d'ailleurs de tester l'importance respective des différents facteurs de nuisance. Cette étude préliminaire étant faite, il est possible de ne focaliser l'étude que sur le ou les facteurs prépondérants (Hanocq *et al* 1992). Ainsi, l'utilisation d'un modèle père prenant en compte la majeure partie des sources de connexion est souvent commode à utiliser. Par ailleurs, si dans les dispositifs présentés dans ce texte, la connexion est assurée par la diffusion des mâles, il y a souvent plusieurs sources de connexion (vente d'animaux, ...) (Hanocq et Boichard 1999). Il s'avère qu'il est là aussi possible de ne prendre en compte que la source principale de connexion. Les autres sources de connexion ignorées pour l'analyse sont additionnelles et peuvent éventuellement améliorer la situation minimaliste alors décrite. Globalement, la connaissance approfondie de la population étudiée permet souvent la simplification des études.

Des études ont déjà été menées avec le critère γ ou des critères au principe similaire dans plusieurs populations bovines, ovines et porcines (Foulley *et al* 1990, Hanocq *et al* 1992, Laloë *et al* 1992, Banos 1993, Hofer 1994, Hanocq 1995). Kennedy et Trus (1993), Hanocq *et al* (1996) et Hanocq (1996) l'ont fait sur des exemples de taille réduite ou lors de simulations. Elles sont parfois complétées d'une analyse spécifique des aspects quantitatifs à l'aide du CD de contraste (Hanocq et Boichard 1999). L'utilisation du CD de contraste s'avère très intéressante pour mesurer la précision de comparaison des niveaux génétiques de groupes d'animaux. Ces groupes peuvent être constitués volontairement (appartenance à une même unité de sélection ...) ou reconnus comme tels grâce à l'étude de la connexion dans la population avec le critère γ .

Globalement, le critère γ permet de décrire la structure des données dans un dispositif (ou une population). Il peut ainsi être utilisé, soit pour une description *a posteriori* de données déjà réalisées, soit dans le cadre de la planification expérimentale. Le CD de contraste, quant à lui, permet plutôt d'indiquer dans le cadre des évaluations génétiques la précision permise par la connexion du dispositif.

4.4 / Utiliser les résultats des études de la connexion

En terme de planification expérimentale, les aspects quantitatifs et structuraux relatifs aux données sont à envisager simultanément et doivent être totalement intégrés dans la conception d'un protocole. Dans le cadre de

l'évaluation génétique des animaux basée sur un contrôle de performances en ferme, la situation est totalement différente. Dans la grande majorité des cas, la réalisation des performances est beaucoup moins planifiée, et les caractéristiques des données ne sont découvertes qu'au moment d'effectuer l'évaluation génétique. L'étude de la connexion permet une description objective de la structure de la population (niveau global de connexion et homogénéité). Les règles de décision quant à l'utilisation pratique des résultats de l'étude peuvent toutefois avoir une part d'arbitraire. Pour se fixer un seuil minimum d'acceptabilité pour le critère γ , comme pour l'exemple du tableau 2 il est possible de relier les valeurs obtenues à ce qu'elles traduisent pour la population étudiée, c'est-à-dire le nombre de données sur lequel reposent les comparaisons correspondantes. L'étude de la connexion peut aboutir à différents types de conclusions.

a / Le niveau de connexion est jugé suffisant

La structure des données ne s'avère pas limitante - la structure de la population est homogène et les quantités de données mises en jeu sont suffisantes - cela garantit une bonne précision des comparaisons mais ne renseigne pas sur l'éventuelle existence de biais qui dépendent de la représentativité des données de connexion sur lesquelles sont basées les comparaisons. Cette représentativité n'est permise qu'en l'absence de facteurs parasites affectant spécifiquement et de façon non contrôlée les performances, il peut s'agir par exemple de traitements préférentiels. Hanocq (1995) a spécifiquement étudié cette notion de représentativité des données en race bovine laitière Holstein. Même en l'absence de problème de connexion, il insiste sur le côté primordial de la planification rigoureuse et contrôlée d'une certaine part des connexions (ici grâce à l'utilisation conjointe de jeunes taureaux mis en testage) afin de disposer d'un certain nombre de données dont il est possible de contrôler *a priori* la représentativité.

b / Le niveau de connexion doit être amélioré

Pour que les estimations puissent être faites avec une précision acceptable, le niveau de connexion doit parfois être amélioré. La représentation graphique de la population, à l'aide des techniques de classification par exemple, permet d'identifier les points faibles d'un schéma de sélection quant à la connexion (Hanocq *et al* 1992). Certaines dispositions de planification sont à prendre pour renforcer ou même établir les connexions là où elles s'avèrent déficientes. Les actions envisageables et la marge de manœuvre du sélectionneur dépendent de l'espèce concernée, de ses caractéristiques biologiques (reproduction en particulier...), de la structure du schéma de sélection et parfois de son contexte historique.

La planification a pour objectif la réalisa-

tion contrôlée de connexions dans la population. Dans les populations sélectionnées d'animaux domestiques, elle concerne surtout l'utilisation raisonnée des mâles. Les mâles ont un pouvoir de diffusion dans le temps et l'espace plus grand que les femelles. La technique de l'Insémination Artificielle (IA), en semence congelée ou fraîche, constitue un outil privilégié et très performant. Son utilisation à grande échelle comme dans les races bovines laitières permet l'établissement d'un niveau de connexion satisfaisant (Hanocq et Boichard 1999). De nombreux mâles y sont utilisés et chacun dans de nombreux troupeaux sur plusieurs années, ce qui permet un nombre important de connexions directes (les plus efficaces) entre animaux et de même entre troupeaux.

Pour des raisons techniques ou historiques, l'utilisation spontanée de l'IA est souvent limitée pour les races ovines et bovines allaitantes. Elle ne suffit pas à garantir la connexion de l'ensemble des troupeaux et doit donc être planifiée. Un taux même important d'IA dans un schéma de sélection peut en effet être moins efficace pour l'établissement de connexions qu'un taux plus réduit accompagné de planification (Hanocq *et al* 1992). Même si les animaux ont une capacité de diffusion relativement réduite (en terme de nombre de troupeaux), une planification rigoureuse permet la réalisation des connexions de proche en proche. L'utilisation de semence fraîche pour la création de connexions peut être très efficace. Historiquement, en France, les premières planifications effectuées en ferme ont été instaurées en races bovines allaitantes (Foulley et Sapa 1982) sur la base d'études antérieures (Foulley et Clerget-Darpoux 1978, Foulley et Ménissier 1978, Foulley et Bibé 1979). Elles ont permis l'augmentation de la taille du noyau de sélection et l'amélioration de l'efficacité du schéma sélection. Dans des situations particulières, il est aussi possible d'utiliser les relations de parenté pour créer des connexions (par exemple grâce à l'utilisation de mâles apparentés dans différents lieux ou à différentes époques), mais la maîtrise de la planification est beaucoup plus délicate. Cela pourrait être éventuellement le cas dans des dispositifs où une grande maîtrise du fonctionnement est possible, comme par exemple dans les stations de contrôle de performances.

En pratique, l'établissement de connexions suppose un minimum d'échanges génétiques. Dans des situations de compétition intense entre organismes de sélection, il peut y avoir une certaine réticence à effectuer ces échanges. Pour les raisons évoquées précédemment, ces échanges sont pourtant primordiaux et les organismes ont de toute façon tout intérêt à utiliser les meilleurs reproducteurs de la population du moment, quelle que soit l'origine de ces reproducteurs. Les différences génétiques entre organismes ne sont généralement pas suffisamment importantes pour que les meilleurs reproducteurs ne puissent pas être dans l'un quelconque des organismes quel que soit son niveau génétique moyen (Banos et Smith

1991). Hanocq *et al* (1996) ont montré que les connexions entre organismes concurrents ne pénalisaient pas les meilleurs du moment si leur stratégie de renouvellement était bien raisonnée. Par contre, elles permettent d'augmenter le progrès génétique global à terme, en particulier en augmentant les origines génétiques et la taille du noyau de sélection. Ce gain d'efficacité est primordial pour des organismes nationaux (même en concurrence) dans un contexte de compétition internationale. Plus généralement, une bonne maîtrise de la connexion doit permettre de se situer entre la sélection en population unique et la sélection en lignées comparées, combinant ainsi les avantages de la première pour l'efficacité et ceux de la seconde pour la préservation de la variabilité.

c / Des éléments de la population sélectionnée ne sont pas connectés

Quand l'analyse de la connexion indique que la population sélectionnée est très hétérogène, à tel point qu'elle est constituée de sous-ensembles non connectés les uns aux autres, pour une plus grande efficacité, des connexions planifiées doivent absolument être instaurées à terme. Toutefois, cette planification des données ne prend pas effet immédiatement, seules les évaluations génétiques futures en bénéficieront. En attendant, il convient de prendre les dispositions nécessaires à une exploitation correcte des résultats de l'évaluation génétique déjà effectuée.

Malgré un manque de connexion, il est important d'inclure une quantité maximale d'information dans l'analyse en prenant en compte l'ensemble des données (à moins que celles-ci ne soient pas représentatives de la population). L'évaluation génétique ainsi effectuée, les résultats fournis doivent cependant être considérés différemment selon qu'ils concernent la partie connectée de la population ou non. Ce qui est vrai pour un manque de connexion l'est aussi pour un niveau de connexion insuffisant. De même que les animaux qui ont un CD insuffisant n'ont pas d'index publié, ceux qui ne sont pas connectés au reste de la population ou insuffisamment ne peuvent pas participer au schéma de sélection. Un CD minimum garantit une quantité d'information minimale nécessaire à la publication d'un index, un γ et un CD de contraste minimum garantissent une connexion suffisante pour que les index publiés puissent être comparés à ceux des animaux du noyau de sélection.

Conclusion

L'évaluation génétique des animaux utilise des informations de nature diverse (performances, généalogies, hypothèses statistiques et génétiques ...) collectées dans des conditions de milieu variées qui influencent la valeur des performances enregistrées. Il y a donc un risque de confusion entre les effets génétiques et environnementaux. Les propriétés théoriques des outils statistiques doivent permettre de s'affranchir de ces difficultés,

Dans les schémas de sélection, il est parfois nécessaire d'augmenter le niveau de connexion, par exemple en planifiant des inséminations artificielles.

mais pour cela, la structure de l'information doit être adéquate et la quantité de données mises en jeu suffisante. Si c'est le cas, la comparaison de la valeur génétique des animaux (et celle des effets de milieu) inclus dans l'évaluation génétique a un sens : ils sont dits "connectés" et les comparaisons sont exploitables. La notion de connexion recouvre les caractéristiques communes (environnementales et génétiques) des individus, milieux et performances sur lesquelles sont basées les comparaisons.

La taille des schémas de sélection est importante et leur structure complexe, des outils sont disponibles pour étudier la connexion dans de telles situations. La bonne connaissance de la population et du schéma permet en outre de simplifier les analyses en ne se focalisant que sur les éléments limitants. Un bon niveau de connexion est nécessaire pour que le progrès génétique créé soit optimal et puisse être utilisé au mieux. La quantité de données doit être suffisante pour pouvoir estimer correctement les valeurs génétiques (mesure grâce au CD), mais, de surcroît, les données doivent être suffisamment bien structurées (mesure grâce au critère γ).

Les efforts effectués pour l'obtention d'une quantité suffisante de données peuvent être ruinés si la structure n'est pas bonne. L'absence de traitements préférentiels agissant sur les performances doit aussi être vérifiée. Si le niveau de connexion dans un schéma de sélection n'est pas suffisant, il est possible de l'améliorer progressivement ; l'utilisation raisonnée de l'insémination artificielle devient alors primordiale. De manière générale, une méthodologie statistique aussi sophistiquée soit-elle ne saurait pallier des insuffisances et lacunes du dispositif. La maîtrise de la connexion participe donc à cet effort indispensable de valorisation des données et des contrôles de performances.

Remerciements

Les auteurs tiennent à remercier D. Laloë (INRA-SGQA, Jouy en Josas) et L. Journaux (Institut de l'Élevage - DGICP, Paris) pour leur lecture du manuscrit et leurs commentaires, ainsi que J.L. Foulley (INRA - SGQA, Jouy en Josas), qui de plus est à l'origine des travaux sur la connexion en génétique animale. Leurs remerciements s'adressent aussi aux lecteurs anonymes mandatés par la revue pour leurs commentaires très constructifs.

Références

- Banos G., 1993. Measure of genetic relationships and ties among three nordic Black-and-White dairy bull populations. In: 44th Annu. Meet. EAAP, Aarhus (Denmark), August 16-19, 4 p.
- Banos G., Smith C., 1991. Selecting bulls across countries to maximize genetic improvement in dairy cattle. *J. Anim. Breed. Genet.*, 108, 174-181.
- Ducrocq V., 1990. Les techniques d'évaluation génétique des bovins laitiers. *INRA Prod. Anim.*, 3, 3-16.
- Fernando R.L., Gianola D., Grossman M., 1983. Identifying all connected subsets in a two-way classification without interaction. *J. Dairy Sci.*, 66, 1399-1402.
- Foulley J.L., Bibé B., 1979. Utilisation conjointe de l'insémination artificielle et de la saillie naturelle en vue de l'indexation des reproducteurs. In : 5ème journée de la recherche ovine et caprine, Paris (France), 5-6 décembre 1979, INRA-ITOVIC, 205-211.
- Foulley J.L., Clerget Darpoux F., 1978. Progeny group size for evaluating natural service bulls using AI reference sires. *Ann. Genet. Sel. Anim.*, 10, 541-556.
- Foulley J.L., Ménessier F., 1978. Arguments in favour of an evaluation system for natural service bulls belonging to beef breeds, by taking into account the diffusion of the best AI bulls. XVth World Charolais convention, Harrogate (Great Britain), Sept. 16-29, Charolais, 52, 53-57.
- Foulley J.L., Sapa J., 1982. The French evaluation program for natural service beef bulls using AI sire progeny as herd ties. *British Cattle Breeders Club, Winter Conference, Cambridge (Great Britain), January 11-12, Digest*, 37, 64-67.
- Foulley J.L., Bouix J., Goffinet B., Elsen J.M., 1984. Comparaison de pères et connexion. In : J.M. Elsen et J.L. Foulley (eds), *Insémination artificielle et amélioration génétique : Bilan et perspectives critiques, Les colloques de l'INRA*, 29, 131-176.
- Foulley J.L., Bouix J., Goffinet B., Elsen J.M., 1990. Connectedness in genetic evaluation. In: D. Gianola et K. Hammond (eds), *Advances in Statistical Methods for Genetic Improvement of Livestock*, Springer-Verlag, Heidelberg, 277-308.
- Foulley J.L., Hanocq E., Boichard D., 1992. A criterion for measuring the degree of connectedness in linear models of genetic evaluation. *Genet. Sel. Evol.*, 24, 315-330.
- Hanocq E., 1995. Etude de la connexion en sélection animale. Thèse Institut National Agronomique Paris Grignon, Paris (France), 126 p.
- Hanocq E., 1996. Deux critères pour décrire la structure des données d'un dispositif : principe et éléments d'interprétation. In : J.M. Elsen et E. Hanocq (eds), *Séminaire du Département de Génétique Animale de l'INRA, St Martin de Ré*, 2-4 avril, 13-24.
- Hanocq E., Boichard D., 1999. Connectedness in the French Holstein cattle population. *Genet. Sel. Evol.*, 31, 163-176.
- Hanocq E., Foulley J.L., Boichard D., 1992. Measuring connectedness in genetic evaluation with an application to Limousin and Maine Anjou sires. In: 43rd Annu. Meet. EAAP, Madrid (Spain), Sept. 13-17, 4 p.
- Hanocq E., Boichard D., Foulley J.L., 1996. A simulation study of the effect of connectedness on genetic trend. *Genet. Sel. Evol.*, 28, 67-82.

- Henderson C.R., 1973. Sire evaluation and genetic trends. In: Proc. Anim. Breed. Genet. Symp. in Honor of Dr J.L. Lush, Blacksburg (Virginia), August 1972, 10-41.
- Hofer A., 1994. Precision of comparisons of estimated breeding values of centrally tested pigs across herds of origin. In: 5th World Congress of Genetics Applied to Livestock Production, Guelph, Ontario (Canada), August 7-12, 18, 447-450.
- Kennedy B.W., Trus D., 1993. Considerations on genetic connectedness between management units under an animal model. J. Anim. Sci., 71, 2341-2352.
- Laloë D., 1993. Precision and information in linear models of genetic evaluation. Genet. Sel. Evol., 25, 557-576.
- Laloë D., Sapa J., Ménissier F., Renand G., 1992. Use of the relationship matrix in the evaluation of natural service Limousin bulls. Genet. Sel. Evol., 24, 137-145.
- Ollivier L., 1994. Utilisation du modèle animal pour l'analyse des expériences de sélection. In : J.L. Foulley et M. Molénat (eds), Séminaire modèle animal, La Colle sur Loup (France), 26-29 septembre, 37-40.
- Searle S.R., 1986. Linear Models. Wiley, New York.
- Weeks D.L., Williams D.R., 1964. A note on the determination of connectedness in an N-way cross classification. Technometrics, 6, 319-324.

Abstract

An overview of connectedness in animal breeding.

The connectedness was defined through a statistical approach, and under the specific animal breeding situation. A good level of connectedness, *ie* a balanced distribution of the records across the levels of the genetic and environmental factors, is absolutely necessary for estimating and comparing the factors (herd effect, breeding values ...) involved in a linear model of genetic evaluation. A criterion, varying continuously from 0 (complete disconnection) to 1 (perfect connection), was used for measuring the level of connectedness. The estimations provided are only useful when their accuracy is sufficient, thus only if the number of records upon which they rely is large enough. A criterion, used for measuring such accuracy, is also presented. The properties of these criteria were presented in a

single theoretical design. Such an approach was also used for illustrating the decrease in the selection scheme efficiency due to a lack of connectedness. In the context of a selection scheme, geographical and environmental factors are known to greatly influence the degree of connectedness. A lack of connectedness is not easy to be detected by the breeder. A specific study of the connectedness and of the reliability of the data is then necessary before performing a genetic evaluation. The main situations encountered in the selection schemes (disconnectedness, insufficient or good connectedness) and their consequences on the breeders' strategy are discussed.

Hanocq E., Tiphine L., Bibé B., 1999. Le point sur la notion de connexion en génétique animale. INRA Prod. Anim., 12, 101-111.