



POUR MIEUX COMPRENDRE LES INDEX BLUP – MODELE ANIMAL

Florence Phocas (INRA-GABI)

L'index d'un animal pour un caractère représente la prédiction de sa valeur génétique. La valeur génétique vraie n'est jamais connue, mais sa valeur la plus probable peut être estimée avec une certaine précision. Le coefficient de détermination (CD) constitue la mesure de cette précision et indique le degré de fiabilité qu'on peut accorder à l'index. Un index sur ascendance seule a un CD inférieur à 0,5, un index basé sur une seule performance a un CD inférieur ou égal à l'héritabilité du caractère, tandis que le CD d'un index sur descendance peut atteindre théoriquement 1 si les produits sont très nombreux. L'index est calculé en fonction de l'ensemble des informations disponibles à un moment donné.

1/ TROIS SOURCES D'INFORMATION POUR L'INDEX \hat{a}_i D'UN ANIMAL I

La performance individuelle (y_i)

- traduit en index individuel ($y_i^* = y_i - X_i b$) par la correction des effets non génétiques (b).

Les performances des descendants (\hat{a}_d)

- traduites par les index des n descendants (\hat{a}_i) corrigés des index des conjoints (\hat{a}_c) de l'animal i .

Les performances des ascendants (\hat{a}_a)

- traduites en index des parents \hat{a}_p et \hat{a}_m .

Rappelons qu'un index sur ascendance, même avec des parents connus avec une précision extrême, ne fournit que l'espérance de la valeur génétique vraie d'un animal.

Cette estimation est d'une précision égale à $CD_a = (CD_{\text{père}} + CD_{\text{mère}}) / 4$. En pratique, un CD sur ascendance de 0,3 s'obtient pour des animaux issus de taureaux largement diffusés par IA.

2/ CONTRIBUTION DES TROIS SOURCES D'INFORMATION A L'INDEX \hat{a}_i

[équation 1]

$$\hat{a}_i = [w_1 y_i^* + w_2 \hat{a}_d + w_3 \hat{a}_a] / [w_1 + w_2 + w_3]$$

où $w_s/[w_1+w_2+w_3]$ est appelée la contribution relative de la source d'information s à l'index \hat{a}_i .

Ci-dessous sont données les valeurs des \hat{a}_s et w_s de l'index prédit pour l'animal i **en supposant que les 2 parents et tous les conjoints de l'animal sont connus et qu'une seule performance propre est mesurable sur l'animal.**

$$w_1 = h^2 / (1 - h^2)$$

où h^2 est l'héritabilité du caractère ($w_1=0$ si l'animal n'a pas de performance individuelle).

$$\hat{a}_d = 2 \sum_f (\hat{a}_f - \frac{1}{2} \hat{a}_c) / n$$

est l'index sur descendance de l'animal i.

$$w_2 = n/2$$

est égal à la moitié du nombre de descendants contrôlés si tous les conjoints sont connus.

N.B. : On peut écrire une formulation plus générale (Bonaiti et al., 1990) de \hat{a}_d et w_2 en fonction de la connaissance ou non de chacun des conjoints de l'animal i :

$$\hat{a}_d = 2 \sum_f d_f (\hat{a}_f - \frac{1}{2} \hat{a}_c) / \sum_f d_f$$

$$\text{et } w_2 = \sum_f d_f / 4$$

avec $d_f=2$ si le conjoint est connu et $d_f=4/3$ si le conjoint est inconnu.

Donc, si tous les conjoints sont inconnus, \hat{a}_d s'écrit comme si tous les conjoints étaient connus mais $\hat{a}_c=0$ et le poids w_2 devient égal à $n/3$.

$$\hat{a}_a = \hat{a}_p/2 + \hat{a}_m/2$$

est l'index sur ascendance de l'animal i.

$$w_3 = 2$$

est toujours le poids accordé à l'index sur ascendance quand les 2 parents sont connus. Si un seul parent est connu, $w_3=4/3$. Si aucun parent n'est connu, $w_3=1$. Les valeurs inconnues sont mises à 0 au niveau des index.

Une formulation équivalente de l'équation 1 montre le rôle respectif de l'ascendance et du reste des informations (individu et descendants) qui permet d'estimer l'aléa de méiose :

[équation 2]

$$\hat{a}_i = \hat{a}_a + [w_1 (y_i^* - \hat{a}_a) + w_2 (\hat{a}_d - \hat{a}_a)] / [w_1 + w_2 + w_3]$$

► L'index sur ascendance donne de bons résultats en terme de niveau génétique moyen d'une population animale, mais ne peut en aucun cas suffire à prédire de manière fiable la valeur génétique d'un individu donné, en raison même de la redistribution des gènes parentaux au moment de la méiose (phénomène de l'aléa de méiose).

► La prédiction d'une valeur génétique proche de la réalité pour un animal donné nécessite la mise à l'épreuve de l'animal, c'est-à-dire le contrôle d'un nombre important de descendants pour quantifier cet aléa de méiose.

3/ CALCUL DE LA PRECISION CD_i DE L'INDEX \hat{a}_i

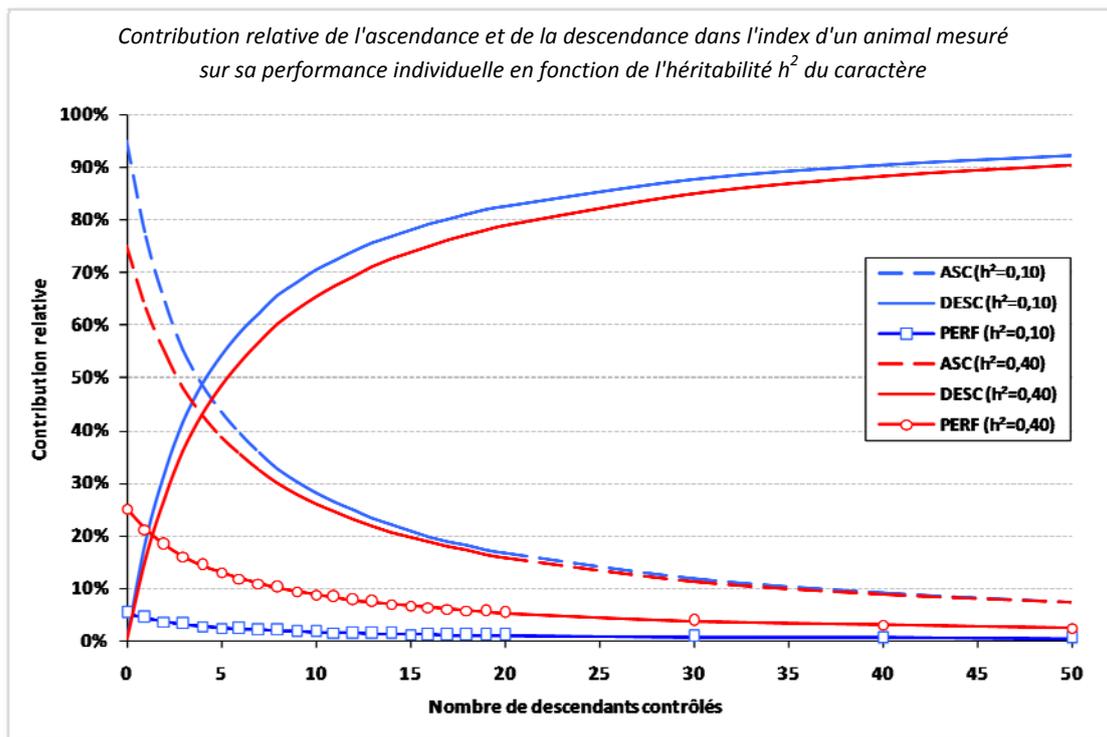
La précision peut être calculée à partir de celles des trois sources d'information relatives à l'index \hat{a}_i (Harris et Johnson, 1994).

Voici, ci-contre, la formulation dans le cas de base, où l'on suppose une seule performance par animal, toutes les performances étant contrôlées dans le même milieu (b se résume à une moyenne des performances).

$CD_1 = h^2$ est le CD sur performance individuelle
$CD_2 = n h^2 / [4 + (n-1) h^2]$ est le CD sur descendance
$CD_3 = (CD_p + CD_m) / 4$ est le CD sur ascendance (calculé en ignorant l'individu i)
$CD_{1+2} = [CD_1 + CD_2 - 2CD_1CD_2] / [1 - CD_1CD_2]$
D'où $CD_i = CD_{1+2+3} = [CD_{1+2} + CD_3 - 2 CD_{1+2}CD_3] / [1 - CD_{1+2}CD_3]$

4/ QUELQUES CAS D'ECOLE

A/ Illustration des contributions relatives des trois sources d'information dans un index où l'animal, ses ascendants et ses descendants sont contrôlés sur le caractère indexé



La descendance explique 50% de l'index d'un animal à partir de 5 veaux contrôlés.

La performance individuelle explique alors seulement 2% de l'index d'un caractère à $h^2=0,1$ et 12% de l'index d'un caractère à $h^2= 0,4$.

Quelle que soit l'héritabilité du caractère (entre 0,01 et 0,50), il faut entre 40 et 50 descendants pour qu'ils contribuent à 90% de l'index d'un animal. Il faut entre 80 et 100 descendants pour expliquer 95% de l'index d'un animal.

B/ Illustrations du nombre de descendants nécessaires pour compenser des différences de niveau génétique d'ascendance paternelle dans l'index d'un animal

B.1/ Les hypothèses

On mesure

dans un même milieu le poids au sevrage de tous les individus. Le poids moyen est de 300 kg. L'écart type phénotypique σ_p est de 30 kg. Les index sont centrés sur 0 kg.

On va supposer

que les mères des animaux i et j sont de même valeur génétique moyenne, c'est-à-dire $\hat{a}_m=0$, toutes deux ont un seul descendant, sont connues sur performance individuelle et non pas d'ascendance connue.

On va considérer

que les animaux i et j ont des descendants avec des conjoints tous inconnus ($\hat{a}_c = 0$ et $w=2=n/3$) et que tous les descendants d'un animal sont connus sur performance individuelle et ont tous la même valeur \hat{a}_f .

$$\hat{a}_{fi} = [2/3 \hat{a}_i + h^2(y_f-300) / (1-h^2)] / [4/3 + h^2 / (1-h^2)] \text{ et } \hat{a}_{di} = 2 \hat{a}_{fi}$$

On va aussi considérer

deux cas extrêmes sur l'ascendance paternelle :

1/ Le père de i est un taureau dont l'index est estimé $\hat{a}_{pi}=0$ kg.

$$\hat{a}_i = [h^2(y_i-300) / (1-h^2) + n/3 \hat{a}_{di} + 2*0/2] / [h^2 / (1-h^2) + n/3 + 2]$$

$$\hat{a}_i = [h^2(y_i-300) / (1-h^2) + 2nh^2(y_f-300) / (4-h^2)] / [h^2 / (1-h^2) + nh^2 / (4-h^2) + 2]$$

2/ Le père de j est un taureau dont l'index est estimé $\hat{a}_{pj}= +1$ écart type génétique σ_a .

$$\hat{a}_j = [h^2 (y_j-300) / (1-h^2) + n/3 \hat{a}_{dj} + 2*\sigma_a/2] / [h^2 / (1-h^2) + n/3 + 2]$$

$$\hat{a}_j = [h^2(y_j-300) / (1-h^2) + 2nh^2(y_f-300) / (4-h^2) + \sigma_a] / [h^2 / (1-h^2) + nh^2 / (4-h^2) + 2]$$

Par la suite on notera I_p l'index paternel \hat{a}_p . Une valeur d'un écart type génétique σ_a pour cet index correspond à environ 9,5 kg si le caractère est supposé d'héritabilité 10% et à 19,0 kg si le caractère est supposé avoir une héritabilité de 40%.

On va faire varier

les performances de l'animal et de ses descendants (toutes les performances sont mises à une même valeur) d'un écart type phénotypique autour de la moyenne du caractère, c'est-à-dire entre 285 et 315 kg.

B.2/ Selon l'index paternel et l'héritabilité du caractère, quel est le nombre de descendants nécessaires à l'obtention pour un animal d'un index proche de la valeur vraie ?

NB1 : Si tous les descendants sont à la moyenne phénotypique de la population ($y=300\text{kg}$), alors la valeur génétique vraie de l'animal est de 0kg .

NB2 : Si tous les descendants sont à un demi-écart type phénotypique de la moyenne ($\pm 15\text{kg}$), alors la valeur génétique vraie de l'animal est de 1 écart type phénotypique en valeur absolue, soit 30kg au-dessus ou en-deçà de la moyenne génétique de la population.

Ci-dessous les valeurs d'index prédit (en kg) pour un caractère d'héritabilité $h^2=0,4$:

$h^2=0,4$	$Ip=0\text{kg}$		$Ip=19,0\text{ kg}$	
	$y=285\text{kg}$ Index	$y=315\text{kg}$ Index	$y=285\text{kg}$ Index	$y=315\text{kg}$ Index
0	-3,75	3,75	3,37	10,87
1	-4,80	4,80	2,03	11,63
2	-5,77	5,77	0,80	12,34
3	-6,67	6,67	-0,34	12,99
4	-7,50	7,50	-1,40	13,60
5	-8,28	8,28	-2,39	14,16
6	-9,00	9,00	-3,31	14,69
7	-9,68	9,68	-4,17	15,19
8	-10,31	10,31	-4,98	15,65
9	-10,91	10,91	-5,73	16,08
10	-11,47	11,47	-6,45	16,49
20	-15,68	15,68	-11,80	19,56
50	-21,49	21,49	-19,18	23,79
100	-24,92	24,92	-23,54	26,30
1 000	-29,38	29,38	-29,22	29,55
10 000	-29,94	29,94	-29,92	29,95

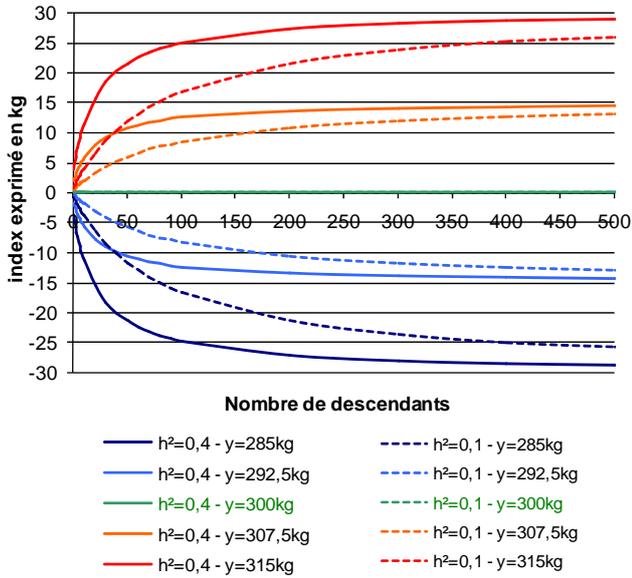
Ci-dessous les valeurs d'index prédit pour un caractère d'héritabilité $h^2=0,1$:

$h^2=0,1$	$Ip=0\text{kg}$		$Ip=9,5\text{ kg}$	
	$y=285\text{kg}$ Index	$y=315\text{kg}$ Index	$y=285\text{kg}$ Index	$y=315\text{kg}$ Index
0	-0,79	0,79	3,70	5,28
1	-1,14	1,14	3,30	5,58
2	-1,48	1,48	2,90	5,87
3	-1,82	1,82	2,52	6,15
4	-2,14	2,14	2,14	6,43
5	-2,46	2,46	1,77	6,70
6	-2,77	2,77	1,41	6,96
7	-3,08	3,08	1,06	7,22
8	-3,38	3,38	0,72	7,47
9	-3,67	3,67	0,38	7,72
10	-3,95	3,95	0,05	7,96
20	-6,50	6,50	-2,88	10,11
50	-11,83	11,83	-9,03	14,62
100	-16,81	16,81	-14,78	18,84
1 000	-27,78	27,78	-27,44	28,12
10 000	-29,76	29,76	-29,72	29,80

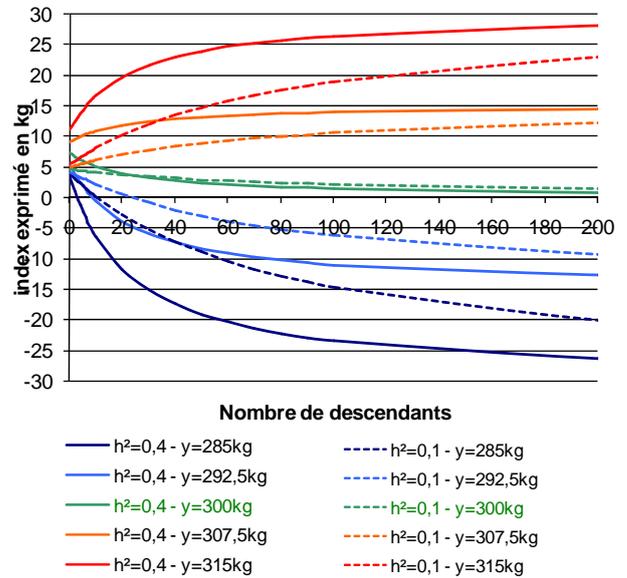
Les 4 graphes suivants montrent l'évolution de l'index d'un animal en fonction du nombre de ses descendants. Ils illustrent qu'il est totalement illusoire d'espérer connaître

d'après les performances de ses quelques produits (<10) la valeur génétique vraie d'une femelle, en se dégageant de l'influence du niveau d'index paternel.

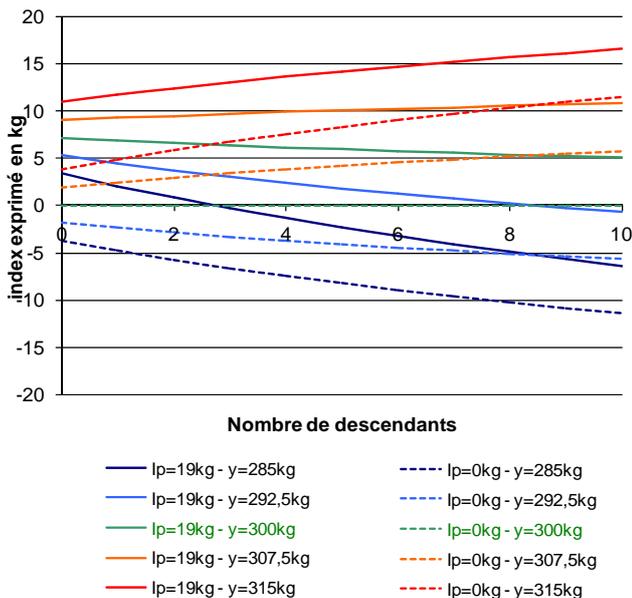
Graph 1 : Evolution de l'index d'un animal, de niveau sur ascendance moyen, en fonction du nombre et des performances de ses descendants.



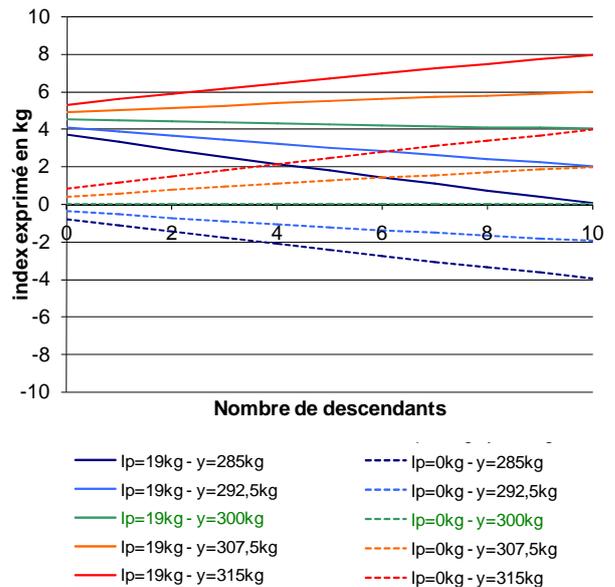
Graph 2 : Evolution de l'index d'un animal, dont l'index paternel est un écart type génétique au-dessus de la moyenne de la population, en fonction du nombre et des performances de ses descendants.



Grappe 3 : Evolution de l'index d'une reproductrice pour un caractère à h²=0,4, moyenne=300kg et écart type phénotypique=30kg, en fonction de l'index paternel (Ip) et des performances (y) de ses descendants.



Graph 4 : Evolution de l'index d'une reproductrice pour un caractère à h²=0,1, moyenne=300kg et écart type phénotypique=30kg, en fonction de l'index paternel (Ip) et des performances (y) de ses descendants.



C/ Calcul de la précision CD_i de l'index \hat{a}_i

Etant donné que nous n'observons pas plus de 10 descendants d'une femelle et rarement même plus de 5, il s'agit de calculer l'intervalle de confiance des index prédits pour ces femelles. Cet intervalle de confiance se calcule à partir des CD des index.

Sous l'hypothèse d'un caractère pour lequel chaque animal est mesuré sur performance individuelle et dont la mère à ce seul descendant et pas d'ascendant connu ($CD_m=h^2$), on obtient le CD de l'animal en fonction de l'héritabilité du caractère, du CD paternel (CD_p) et du nombre de descendants contrôlés :

Caractère à $h^2 =$	0,1		0,4	
	CD _p =0,9 CD _i	CD _p =0,5 CD _i	CD _p =0,9 CD _i	CD _p =0,5 CD _i
0	0,35	0,27	0,53	0,49
1	0,36	0,28	0,56	0,52
2	0,37	0,29	0,58	0,54
3	0,38	0,30	0,60	0,56
4	0,39	0,32	0,61	0,58
5	0,40	0,33	0,63	0,60
6	0,41	0,34	0,64	0,62
7	0,42	0,35	0,66	0,63
8	0,43	0,36	0,67	0,65
9	0,44	0,37	0,68	0,66
10	0,44	0,38	0,69	0,67
20	0,51	0,47	0,77	0,76
50	0,65	0,63	0,87	0,87
100	0,76	0,75	0,92	0,92

On calcule l'intervalle de confiance à 95% de la valeur d'un index I comme l'intervalle centré sur I avec pour borne minimale $I-\Delta$ et borne maximale $I+\Delta$, où

$$\Delta = 1,96 * \sigma_a * (1-CD_i)^{0,5}$$

La mention « à 95% » signifie statistiquement que le risque demeure dans 5% des cas que la valeur génétique vraie ne soit pas comprise dans l'intervalle donné.



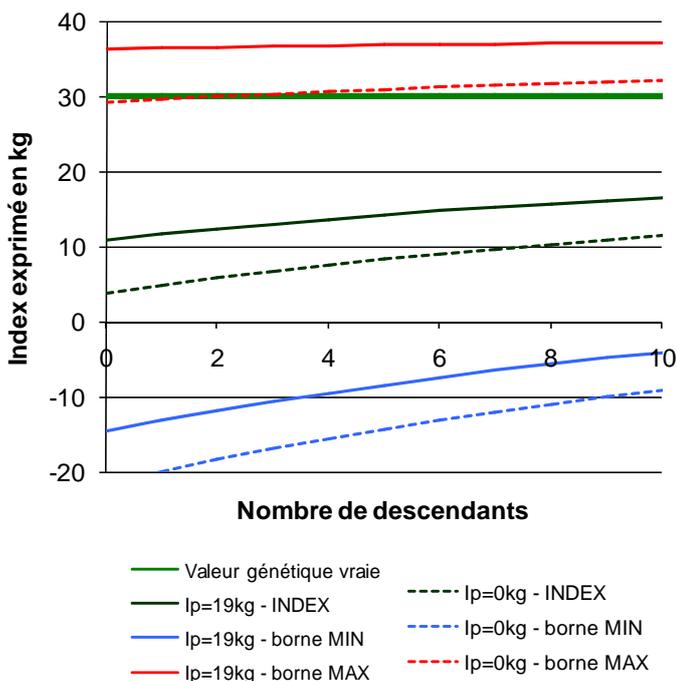
On note sur le graphe 5 ci-dessous que :

1/ les intervalles de confiance (IC) sont très larges et contiennent le plus souvent la valeur génétique vraie de l'animal même sans descendant contrôlé sur un caractère $h^2=0.4$.

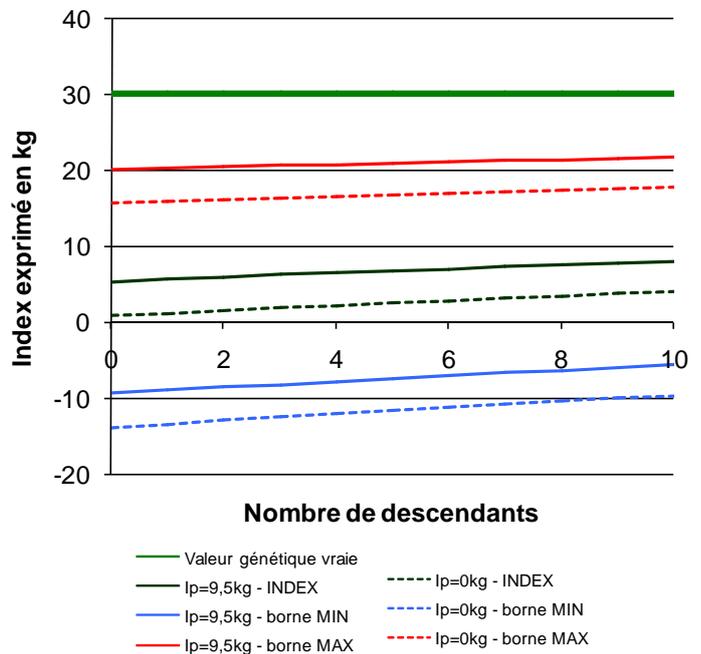
2/ toutefois, dans le cas d'un index paternel moyen ($I_p=0\text{kg}$) pour un caractère $h^2=0.4$, il faut attendre le troisième veau connu pour que l'IC de l'index d'une excellente reproductrice contienne sa valeur génétique vraie ($a_i=30\text{kg}$).

Ceci rend compte de la faible probabilité ($p=0,01$) qu'une femelle dont l'index sur ascendance est à la moyenne de la population ait un aléa de méiose si favorable que sa valeur génétique vraie soit supérieure de 30kg à la valeur génétique moyenne de ses parents et de la population pour un caractère d' $h^2=0.4$ et d'écart type phénotypique 30kg.

Graph 5 : Intervalle de confiance à 95% de l'index d'un animal sur un caractère de $h^2=0.4$, moyenne 300 kg et écart type phénotypique 30kg, en fonction de l'index paternel I_p ($CD_p=0.9$) et du nombre de descendants ayant tous une performance $\gamma=315\text{kg}$.



Graph 6 : Intervalle de confiance à 95% de l'index d'un animal sur un caractère de $h^2=0.1$, moyenne 300 kg et écart type phénotypique 30kg, en fonction de l'index paternel I_p ($CD_p=0.9$) et du nombre de descendants ayant tous une performance $\gamma=315\text{kg}$.



Le graphe ci-dessus montre que la valeur génétique vraie d'une femelle exceptionnelle ne peut absolument pas être prédite à partir du nombre limité de descendants qu'elle peut obtenir pour une faible héritabilité du caractère. Dans ce cas, la probabilité que la valeur génétique vraie de l'animal soit de +30kg (soit plus de 3 écarts type génétiques) est quasi nulle (inférieure à 1/10 000 même pour $I_p=9,5\text{kg}$). La probabilité d'un aléa de méiose de +15kg n'est que de 6% pour $I_p=9,5\text{kg}$ et de 1% pour $I_p=0\text{kg}$!

QUID DE L'IMPORTANCE DU CD DE L'INDEX PATERNEL SUR L'EVALUATION D'UNE DE SES FILLES ?

Il faut avant tout bien comprendre que l'index d'un animal dépend certes de la valeur des index de ses parents, mais non de leur précision. Seule la précision de l'index de l'animal est affectée par la précision des index de ses ascendants. C'est donc l'intervalle de confiance de l'index qui est réduit par une meilleure connaissance des index des parents, et ce, dans une amplitude relativement faible comme le montre les deux graphes ci-contre.

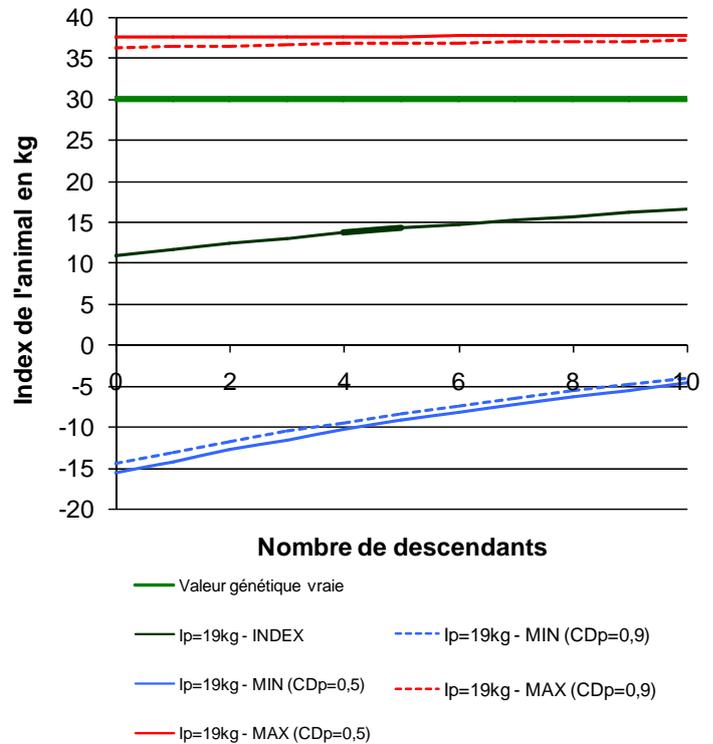
Nous allons considérer 2 CD différents pour l'index paternel I_p d'un taureau améliorateur ($I_p=+19$ kg) :

1/ un $CD_p = 0.5$ qui correspond à l'index d'un jeune mâle non testé, mais connu sur performance individuelle ($h^2=0.4$) et sur ascendance paternelle (avec un CD du père à taureau pouvant varier de 0.5 à 0.9).

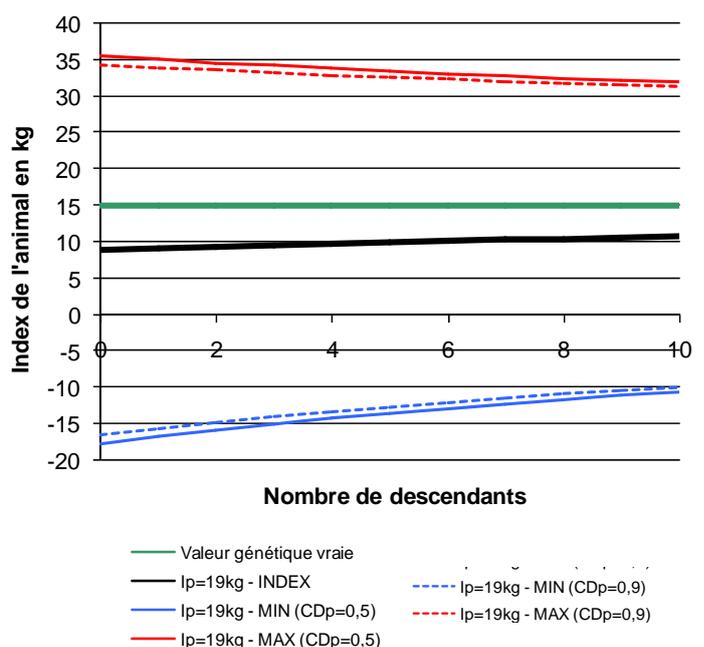
2/ un $CD_p = 0.9$ qui correspond forcément à un taureau testé sur 50 à 100 descendants pour un caractère à $h^2=0.4$.

NB. Le CD_p théorique d'un taureau connu sur ascendance, performance individuelle et seulement une vingtaine de descendants (cas d'un taureau de MN après une campagne d'utilisation ou d'un taureau d'IA en sortie de station de contrôle sur descendance) est de 0.76 pour un caractère à $h^2=0.4$.

Graph 7 : Intervalle de confiance à 95% de l'index d'un animal sur un caractère de $h^2=0.4$, moyenne 300kg et écart type phénotypique 30kg, en fonction du CD de l'index paternel I_p et du nombre de descendants ayant tous une performance $y=315$ kg.



Graph 8 : Intervalle de confiance à 95% de l'index d'un animal sur un caractère de $h^2=0.4$, moyenne 300kg et écart type phénotypique 30kg, en fonction du CD de l'index paternel I_p et du nombre de descendants ayant tous une performance $y=307,5$ kg.



CONCLUSION PRATIQUE A PARTIR DE L'ANALYSE DES CAS D'ECOLE

La descendance toujours limitée d'une femelle ne permet jamais d'avoir un index très précis ($CD > 0,6$). Par contre, l'examen des performances des produits, au fur et à mesure de sa carrière, permet de déceler si la vache va oui ou non confirmer son niveau sur ascendance. Sur les graphiques précédents, les courbes montrent l'évolution des index de 2 vaches ayant le même niveau génétique vrai (permettant aux veaux une performance moyenne de 315 kg, par exemple), mais l'une d'entre elles étant issue d'un taureau améliorateur et l'autre d'un taureau moyen. On observe bien la convergence des valeurs génétiques prédites au cours de la carrière (l'index d'une vache diminue et l'autre augmente), mais on n'arrivera jamais au même index proche de la valeur génétique vraie, même si 10 veaux étaient connus pour chacune des reproductrices. Cela étant, l'examen de la production et l'évolution des index au cours du temps donnent une réponse.

L'étude faite montre que pour un caractère à $h^2=0,4$, il faut attendre de connaître la performance de 3 veaux (à la moyenne de la population) pour que l'index d'une femelle génétiquement moyenne, mais de bonne ascendance ($l_p=1$ écart type génétique) devienne inférieur à l'index d'une femelle d'ascendance moyenne ($l_p=0$) mais dont les veaux ont une performance supérieure d'un écart type phénotypique à la moyenne de la population. Si la performance de ces derniers veaux n'est supérieure que d'un quart d'écart type phénotypique ou bien si le caractère a une $h^2=0,1$, il faudra attendre la performance de 5 veaux pour observer le changement de classement.

On conclut donc que les index, à partir de 5 veaux connus, permettent de discriminer raisonnablement bien des reproductrices aux valeurs génétiques vraies très différentes, quelles que soient leurs valeurs d'index sur ascendance.

En revanche, les prédictions peuvent être très régressées vers la moyenne de la population et donc très éloignées des valeurs génétiques vraies.

RECOMMANDATIONS POUR UNE BONNE UTILISATION DES INDEX DES VACHES

Rappelons que l'évaluation génétique BLUP va fournir le meilleur classement possible entre les individus en minimisant, sur l'ensemble de la population, un critère statistique qui est l'erreur quadratique moyenne (ce critère est une combinaison du biais et de l'imprécision de la prédiction).

Cependant, au niveau de la comparaison de deux index particuliers, l'évaluation génétique ne peut (et ne pourra jamais) garantir un classement sans erreur des animaux : il faudrait sinon que les CD de tous les animaux soient égaux à 1, c'est-à-dire que tous les animaux aient au moins plusieurs centaines de descendants chacun !

En particulier, le classement des reproductrices entre elles est difficile car les femelles ont peu de descendants pour quantifier correctement l'aléa de méiose.

Il s'agit alors de réfléchir l'utilisation des index sur la voie femelle, en connaissance de cause :

1/ L'index sur ascendance et performance propre est très utile pour le choix des femelles de renouvellement du troupeau.

2/ La sélection des vaches (poursuite de carrière dans l'élevage, choix des mères à vaches et taureaux) est forcément améliorée par la prise en compte de leurs produits dans l'indexation.

a/ A index sur ascendance de niveau proche, les index des vaches permettent de les classer de la meilleure façon possible.

b/ A index sur ascendance très différents et faible nombre de descendants contrôlés

(moins de 5 veaux), les index prédisent souvent que les femelles issues des meilleures ascendances ont en espérance une meilleure valeur génétique que les femelles issues d'ascendance de moindre niveau. Le nombre de descendants contrôlés est insuffisant pour apprécier l'importance de l'aléa de méiose, qui est donc estimé à une valeur proche de zéro. Ce raisonnement statistique permet d'estimer sans biais le niveau génétique moyen de la population de vaches. Toutefois, il importe pour chaque éleveur de raisonner de manière multicritère le choix entre deux vaches. Les phénotypes de leur descendance sont à confronter aux valeurs et précisions des index sur ascendance pour se forger une opinion personnelle sur le choix d'une vache par rapport à une autre. Ce raisonnement là ne peut être fait qu'à titre individuel et donc relève entièrement de l'expertise de l'éleveur.

3/ Et si on doit regarder la production d'une vache pour affiner l'analyse individuelle, l'analyse au niveau d'une race ou d'une population, par exemple le choix des mères à taureaux doit se faire à partir des index et ne nécessite pas d'ajouter des critères de production, car en moyenne, on ne se trompe pas !

 **POUR PLUS D'ELEMENTS, SE REFERER :**

→ au cours d' E. Verrier (AgroParisTech) et J.P. Bidanel (INRA) du CSAGAD :

« L'évaluation génétique des reproducteurs à l'aide de la méthode BLUP » (mars 2006, 69p.),

→ à l'article de B. Bonaïti et al. :

« La méthode française d'évaluation génétique des reproducteurs laitiers », INRA Productions animales, 1990, 3(2) 83-92,

→ à l'article de D. Boichard et al. :

« Le modèle sous-jacent à l'évaluation des valeurs génétiques », INRA Productions animales, 1992, hors série « Eléments de Génétique quantitative et application aux populations animales », 185-195.