Original article

A criterion for measuring the degree of connectedness in linear models of genetic evaluation

JL Foulley, E Hanocq, D Boichard

Institut National de la Recherche Agronomique, Station de Génétique Quantitative et Appliquée, Centre de Recherches de Jouy-en-Josas, 78352 Jouy-en-Josas Cedex, France

Summary – A criterion for measuring the degree of connectedness between factors arising in linear models of genetic evaluation is derived on theoretical grounds. Under normality and in the case of 2 fixed factors (θ, ϕ), this criterion is defined as the Kullback-Leibler distance between the joint distribution of the maximum likelihood (ML) estimators of contrasts among θ and ϕ levels respectively and the product of their marginal distributions. This measure is extended to random effects and mixed linear models. The procedure is illustrated with an example of genetic evaluation based on an animal model with phantom groups.

genetic evaluation / connectedness / Kullback–Leibler's distance / mixed linear model

Résumé – Un critère de mesure du degré de connexion en modèles linéaires d'évaluation génétique. *Cet article établit sur des bases théoriques un critère de mesure du degré de connexion entre facteurs d'un modèle linéaire d'évaluation génétique. Sous l'hypothèse de normalité et dans le cas de 2 facteurs fixés (θ, ϕ), ce critère est défini par la distance de Kullback-Leibler entre, d'une part la densité conjointe des estimateurs du maximum de vraisemblance (ML) de contrastes entre niveaux de θ et ϕ respectivement et, d'autre part, le produit de leurs densités marginales. La mesure est généralisée au cas de facteurs aléatoires et de modèles mixtes. Cette procédure est illustrée par un exemple d'évaluation génétique par modèle animal comportant des effets de groupe fantôme.*

évaluation génétique / connexion / distance de Kullback–Leibler / modèle linéaire mixte

INTRODUCTION

The development of artificial insemination in livestock and the potential for using sophisticated statistical BLUP methodology (Henderson, 1984, 1988) gave new impetus for across-herd or station genetic evaluation and selection procedures, *eg* reference sire systems in beef cattle (Foulley *et al*, 1983; Baker and Parratt, 1988) or sheep (Miraei Ashtiani and James, 1990) and animal model evaluation procedures in swine (Bichard, 1987; Kennedy, 1987; Webb, 1987).

In this context, concern about genetic ties among herds or stations is becoming increasingly important although, from a theoretical point of view, complete disconnectedness among random effects can never occur, as explained in detail by Foulley *et al* (1990).

Petersen (1978) introduced a test for connectedness among sires based on the property of the "sire × sire" information matrix after absorption of herd-year-season equations. Fernando *et al* (1983) proposed an algorithm to search for connected groups in a herd-year-season by sire layout which was based on the physical approach of connection developed by Weeks and Williams (1964). This view was also taken up by Tosh and Wilton (1990) to define an index of degree of connectedness for a factor in an N-way cross classification.

Foulley *et al* (1984, 1990) reviewed the definition and problems relevant to this concept. They offered a method for determining the level of connectedness among 2 levels of a factor by relating the sampling variance of the corresponding contrast under the full model to its value under a model reduced by the factors responsible for unbalancedness.

The purpose of this paper is 2-fold: i) to extend this procedure defined for a specific contrast to a global measure of connectedness among levels of a factor; ii) to set up a theoretical framework to justify such a measure on mathematically rigorous grounds.

METHODOLOGY

Our starting point is the following basic property: if observations in each level of some factor (*ie* θ) are equally distributed across levels of another factor (*ie* ϕ), BLUE estimators of the contrasts $\theta_i - \theta_{i'}$, $\phi_j - \phi_{j'}$ are orthogonal under an additive fixed linear model with independent and homoscedastic errors.

This property is lost under an unbalanced distribution up to an ultimate stage consisting of what is called disconnectedness or confounding between the 2 factors. This suggests the idea of measuring the degree of connectedness by some distance between the current status of the layout and the first "orthonormal" one following the terminology of Calinski (1977) and Gupta (1987). The Kullback–Leibler distance $I_{12}(x) = \int p_1(x) \ln [p_1(x)/p_2(x)]dx$ between 2 probability densities $p_1(x), p_2(x)$ turns out to be a natural candidate for measuring such a distance (Kullback, 1968, 1983).

The model assumed is a linear model with additive fixed effects and NIID (normally, identically and independently distributed) residuals $e \sim N(0, \sigma^2 I_N)$

$$y = X_\theta \theta + X_\phi \phi + X_\lambda \lambda + e \tag{1}$$

where $\mathbf{y}$ is an $N \times 1$ data vector, $\boldsymbol{\theta}$, $\boldsymbol{\phi}$ and $\boldsymbol{\lambda}$ are vectors of fixed effects and $\mathbf{X}_\theta, \mathbf{X}_\phi$ and $\mathbf{X}_\lambda$ are the corresponding incidence matrices.

Without loss of generality, we will assume a full rank parameterization in vectors $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ pertaining to factors θ and ϕ and resulting in contrasts such as $\theta_i - \theta_1$ and $\phi_j - \phi_1$ so that:

$$\boldsymbol{\theta} = \{\theta_i - \theta_1\}; \ i = 2, \ldots, m_\theta \tag{2a}$$

$$\boldsymbol{\phi} = \{\phi_j - \phi_1\}; \ j = 2, \ldots, m_\phi \tag{2b}$$

where m_θ and m_ϕ are the numbers of levels for the factors θ and ϕ respectively.

The vector $\boldsymbol{\lambda}$ in [1] designates remaining effects of the model. In a 2-way cross-classified design (eg mean μ, "treatment" and "block"), one has $\boldsymbol{\lambda} = c\mathbf{1}_N$ with $c = \mu + \theta_1 + \phi_1$ but this parameterization turns out to be more general and may include one or several extra factors.

Degree of connectedness is assessed through the Kullback–Leibler distance between the joint density $f(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\phi}})$ of the ML (maximum likelihood) estimators $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\phi}}$ of $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ defined in [2a] and [2b] respectively, and the product $f(\widehat{\boldsymbol{\theta}})f(\widehat{\boldsymbol{\phi}})$ of their marginal densities which would prevail if the design were orthonormal in θ and ϕ. Then,

$$D = \int \int \ln \left[f(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\phi}})/f(\widehat{\boldsymbol{\theta}})f(\widehat{\boldsymbol{\phi}}) \right] f(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\phi}}) \, d\widehat{\boldsymbol{\theta}} \, d\widehat{\boldsymbol{\phi}} \tag{3}$$

where $d\mathbf{x}$ stands for the symbolic notation $\prod_i dx_i$ (Johnsson and Kotz, 1972).

The joint and the marginal distributions arising in [3] are as follows:

$$(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\phi}}) \sim N[(\boldsymbol{\theta}, \boldsymbol{\phi}), \mathbf{C}] \tag{4}$$

$$\widehat{\boldsymbol{\theta}} \sim N[\boldsymbol{\theta}, \mathbf{C}_{\theta\theta}] \tag{5a}$$

$$\widehat{\boldsymbol{\phi}} \sim N[\boldsymbol{\phi}, \mathbf{C}_{\phi\phi}] \tag{5b}$$

where $\mathbf{C}$ is the variance-covariance matrix of the ML estimators of $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ under model [1] and such that:

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{\theta\theta} & \mathbf{C}_{\theta\phi} \\ \mathbf{C}_{\phi\theta} & \mathbf{C}_{\phi\phi} \end{bmatrix} \tag{6}$$

This matrix and its block components can be obtained from the information matrix $\mathcal{I}$ in $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ after absorption of the $\boldsymbol{\lambda}$ equations.

$$\mathcal{I} = \begin{bmatrix} \mathcal{I}_{\theta\theta \cdot \lambda} & \mathcal{I}_{\theta\phi \cdot \lambda} \\ \mathcal{I}_{\phi\theta \cdot \lambda} & \mathcal{I}_{\phi\phi \cdot \lambda} \end{bmatrix} \tag{7}$$

A typical expression for $\mathcal{I}_{\theta\phi \cdot \lambda}$ in [7] is $\mathbf{I}_{\theta\phi \cdot \lambda} = \mathbf{X}'_\theta \mathbf{M}_\lambda \mathbf{X}_\phi$ where $\mathbf{M}_\lambda = \mathbf{I}_N - \mathbf{X}_\lambda (\mathbf{X}'_\lambda \mathbf{X}_\lambda)^- \mathbf{X}'_\lambda$ is the usual orthogonal projector.

Relationships between elements in [6] and [7] are as follows:

$$\mathbf{C} = \mathcal{I}^{-1} \tag{8}$$

$$\mathbf{C}_{\theta\theta} = (\mathcal{I}_{\theta\theta \cdot \lambda} - \mathcal{I}_{\theta\phi \cdot \lambda} \mathcal{I}_{\phi\phi \cdot \lambda}^{-1} \mathcal{I}_{\phi\theta \cdot \lambda})^{-1} \tag{9a}$$

$$= \mathcal{I}_{\theta\theta \cdot \phi\lambda}^{-1} \tag{9b}$$

By putting formulae [8], [9a] and [9b] into the expressions in [5a] and [5b], using those in [3] and letting $\alpha = (\theta', \phi')'$, one gets

$$D = \frac{1}{2} \left\{ \left[\ln \frac{|\mathcal{I}|}{|\mathcal{I}_{\theta\theta\cdot\phi\lambda}| \, |\mathcal{I}_{\phi\phi\cdot\theta\lambda}|} \right] + E[(\hat{\alpha} - \alpha)'\mathbf{Q}(\hat{\alpha} - \alpha)] \right\}$$

where $(\hat{\alpha} - \alpha)'\mathbf{Q}(\hat{\alpha} - \alpha)$ is a quadratic form in $(\hat{\alpha} - \alpha)$, the matrix $\mathbf{Q}$ of which being:

$$\mathbf{Q} = - \begin{bmatrix} \mathcal{I}_{\theta\theta\cdot\lambda} - \mathcal{I}_{\theta\theta\cdot\phi\lambda} & \mathcal{I}_{\theta\phi\cdot\lambda} \\ \mathcal{I}_{\phi\theta\cdot\lambda} & \mathcal{I}_{\phi\phi\cdot\lambda} - \mathcal{I}_{\phi\phi\cdot\theta\lambda} \end{bmatrix}$$

Now $E(\hat{\alpha}) = \alpha$ since the ML estimator of α is unbiased. Moreover, $\mathrm{tr}\,(\mathbf{QC}) = 0$ since, from [8] and [9a] and [9b]:

$$(\mathcal{I}_{\theta\theta\cdot\lambda} - \mathcal{I}_{\theta\theta\cdot\phi\lambda})\mathbf{C}_{\theta\theta} + \mathcal{I}_{\theta\phi\cdot\lambda}\mathbf{C}_{\phi\theta} = 0 \qquad [9c]$$

and ditto for the other term in ϕ.

Then, D reduces to:

$$D = \frac{1}{2} \ln \frac{|\mathcal{I}|}{|\mathcal{I}_{\theta\theta\cdot\phi\lambda}| \, |\mathcal{I}_{\phi\phi\cdot\theta\lambda}|} = -\frac{1}{2} \ln \frac{|\mathbf{C}|}{|\mathbf{C}_{\theta\theta}| \, |\mathbf{C}_{\phi\phi}|} \qquad [10]$$

Alternative expressions to [10] can be derived using the conditional distribution of the ML estimator of one vector ($\hat{\theta}$ or $\hat{\phi}$) given the value of the other due to the following equality:

$$\frac{f(\hat{\theta}, \hat{\phi})}{f(\hat{\theta})f(\hat{\phi})} = \frac{f(\hat{\theta}|\hat{\phi})}{f(\hat{\theta})} = \frac{f(\hat{\phi}|\hat{\theta})}{f(\hat{\phi})} = \frac{f(\hat{\theta}|\hat{\phi})f(\hat{\phi}|\hat{\theta})}{f(\hat{\theta}, \hat{\phi})} \qquad [11]$$

Now:

$$\hat{\theta}|\hat{\phi} \sim N\left[\theta + \mathbf{C}_{\theta\phi}\mathbf{C}_{\phi\phi}^{-1}(\hat{\phi} - \phi), \mathbf{C}_{\theta\theta\cdot\phi}\right]$$

where

$$\mathbf{C}_{\theta\theta\cdot\phi} = \mathbf{C}_{\theta\theta} - \mathbf{C}_{\theta\phi}\mathbf{C}_{\phi\phi}^{-1}\mathbf{C}_{\phi\theta}$$

Because

$$\mathbf{C}_{\theta\theta\cdot\phi}^{-1} = \mathcal{I}_{\theta\theta\cdot\lambda} \qquad [12]$$

one has:

$$D = \frac{1}{2} \ln \frac{|\mathcal{I}_{\theta\theta\cdot\lambda}|}{|\mathcal{I}_{\theta\theta\cdot\phi\lambda}|} = -\frac{1}{2} \ln \frac{|\mathbf{C}_{\theta\theta\cdot\phi}|}{|\mathbf{C}_{\theta\theta}|} \qquad [13]$$

Similarly, by substituting θ to ϕ:

$$D = \frac{1}{2} \ln \frac{|\mathcal{I}_{\phi\phi\cdot\lambda}|}{|\mathcal{I}_{\phi\phi\cdot\theta\lambda}|} = -\frac{1}{2} \ln \frac{|\mathbf{C}_{\phi\phi\cdot\theta}|}{|\mathbf{C}_{\phi\phi}|} \qquad [14]$$

and, finally from the last term in [11], one has:

$$D = \frac{1}{2} \ln \frac{|\mathcal{I}_{\theta\theta \cdot \lambda}| \, |\mathcal{I}_{\phi\phi \cdot \lambda}|}{|\mathcal{I}|} = -\frac{1}{2} \ln \frac{|\mathbf{C}_{\theta\theta \cdot \phi}| \, |\mathbf{C}_{\phi\phi \cdot \theta}|}{|\mathbf{C}|} \qquad [15]$$

Four remarks are worth mentioning at this stage:

1) As shown by formulae [10] [13], [14] and [15], one may talk equivalently about connectedness between θ and ϕ as well as connectedness of θ (or among θ levels) due to the incidence of ϕ (or connectedness of ϕ due to the incidence of θ) in a model including θ, ϕ and λ using the terminology of Foulley *et al* (1984, 1990). This terminology is also in agreement with that taken up by statisticians (Shah and Yadolah, 1977).

2) It is interesting to notice that the variance $\mathbf{C}_{\theta\theta \cdot \phi}$ of the conditional distribution of $\widehat{\theta}$ given $\widehat{\phi}$ is also the variance of the marginal distribution of $\widehat{\theta}$ under the reduced model (θ, λ). This leads to view the ratio of determinants in [13] in the same way as Foulley *et al* (1990) *ie* using their notation:

$$D = -\frac{1}{2} \ln \frac{|\mathbf{C}_R|}{|\mathbf{C}_F|} \qquad [16]$$

where $\mathbf{C}_R$ and $\mathbf{C}_F$ are $\mathbf{C}$ matrices pertaining to $\widehat{\theta}$ under the full (F) model in [1] and the reduced model (R) without ϕ respectively. Moreover, the γ coefficient defined as:

$$\gamma = \exp(-2D) = |\mathbf{C}_R| \, / \, |\mathbf{C}_F| \qquad [17]$$

generalizes the $\gamma_{ii'}$ coefficient of connectedness introduced by Foulley *et al* (1990) for the contrast $\widehat{\theta}_i - \widehat{\theta}_{i'}$; it varies similarly from $\gamma = 0$ (or $D = +\infty$) in the case of complete disconnection to $\gamma = 1$ (or $D = 0$) in the case of perfect connection (*ie* orthogonality).

3) Let us consider the characteristic equation:

$$|\mathbf{C}_{\theta\theta \cdot \phi} - k\mathbf{C}_{\theta\theta}| = 0 \qquad [18]$$

The roots k_i of [18] are the eigenvalues of $\mathbf{C}_{\theta\theta}^{-1}\mathbf{C}_{\theta\theta \cdot \phi}$ or $\mathbf{C}_F^{-1}\mathbf{C}_R$ so that:

$$\left|\mathbf{C}_{\theta\theta}^{-1}\mathbf{C}_{\theta\theta \cdot \phi}\right| = \prod_i k_i = \overline{k}_g^{r_\theta} \qquad [19]$$

where $\overline{k}_g$ is the geometric mean of the $k_i's$ and $r_\theta = \dim(\mathbf{C}_{\theta\theta})$.

Hence $\ln\gamma = r_\theta\overline{k}_g$ which is the justification to standardize D and γ to:

$$D^* = D/r_\theta \qquad [20a]$$

and

$$\gamma^* = \gamma^{1/r_\theta} \qquad [20b]$$

so as to take into account the numbers of elements in θ to be estimated when comparing degree of connectedness of factors differing in number of levels. This standardization procedure is analogous to that proposed by Sölkner and James (1990) for comparing statistical efficiency of crossbreeding experiments involving different numbers of parameters. In that respect γ_θ^* can be interpreted as a kind of average measure of connectedness for $(\widehat{\theta}_i, \widehat{\phi})$ among all pairs of levels of the factor θ due to the incidence of the nuisance factor ϕ for a fixed effect model (see the Appendix). Since γ is equal to both $\left| C_{\theta\theta}^{-1} C_{\theta\theta\cdot\phi} \right|$ and $\left| C_{\phi\phi}^{-1} C_{\phi\phi\cdot\theta} \right|$, one can standardize with respect to r_θ or as well as to r_ϕ depending on the factor which we are interested in.

4) An alternative form to [18] is:

$$\left| C_{\theta\phi} C_{\phi\phi}^{-1} C_{\phi\theta} - (1-k) C_{\theta\theta} \right| = 0 \tag{21}$$

the roots of which $\rho_i^2 = 1 - k_i$ turn out to be the squared canonical sampling correlations between $\widehat{\theta}$ and $\widehat{\phi}$. Since the (non zero) roots of [21] are also the (non zero) roots of $\left| C_{\phi\theta} C_{\theta\theta}^{-1} C_{\theta\phi} - \rho^2 C_{\phi\phi} \right| = 0$, they satisfy the equation $\left| C_{\phi\phi\cdot\theta} - (1-\rho^2) C_{\phi\phi} \right| = 0$. Thus γ can be expressed as:

$$\gamma = \prod_{i=1}^{r_\theta} (1-\rho_i^2) = \prod_{i=1}^{r_\phi} (1-\rho_i^2) \tag{22}$$

with $\rho_i = 0$ (*ie* $k_i = 1 - \rho_i^2 = 1$) for $i = r_\theta + 1, r_\theta + 2, \ldots, r_\phi$ if $r_\theta < r_\phi$ or for $i = r_\phi + 1, r_\phi + 2, \ldots, r_\theta$ if $r_\phi < r_\theta$.

5) The presentation was restricted to 2 factors θ and ϕ. It can be extended to more than 2 classifications. For instance, with 3 factors θ, ϕ, Ψ, one can consider the Kullback–Leibler distance between $f(\widehat{\theta}, \widehat{\phi}, \widehat{\Psi})$ and $f(\widehat{\theta}) f(\widehat{\phi}, \widehat{\Psi})$. The resulting D coefficient can be expressed as $D = \dfrac{1}{2} \ln \left(|\mathcal{I}_{\theta\theta\cdot\lambda}| / |\mathcal{I}_{\theta\theta\cdot\phi\Psi\lambda}| \right)$ and interpreted as the degree of connectedness of θ due to fitting ϕ and Ψ in the complete model $(\theta, \phi, \Psi, \lambda)$.

6) This approach developed for models with fixed effects can be extended to mixed models as well. A first obvious extension consists of taking λ in [1] (or part of it) as a vector of random effects. The only change to implement in computing the matrix in [7] is to carry out an absorption of λ equations which takes into account the appropriate structure of this vector. Actually this can be easily done using the mixed model equations of Henderson (1984).

In more general mixed models, one has to keep in mind that from a statistical point of view, connectedness is an issue only for factors considered as fixed (Foulley *et al*, 1990). In other words, in a model without group effects, BLUP of sire transmitting abilities or individual genetic merits always have solutions whatever the distribution of records across herd-year-seasons and other fixed effects. Nevertheless, the phenomenon of non orthogonality between the estimation of a contrast of fixed effects and the error of prediction in some level of a random

effect still exists and may be addressed in the same way as outlined previously. For instance to measure degree of connectedness between one random factor $\mathbf{u} = \{u_i\}$; $i = 1, 2, \ldots, m_u$ (*eg* sire) and one fixed factor $\boldsymbol{\phi}$ (*eg* herd), it suffices to consider in [3] its error of prediction from BLUP *ie* replace $\hat{\boldsymbol{\theta}}$ in [2a] by $\Delta = \{\Delta_i = \hat{u}_i - u_i\}$. All the above formulae apply since the derivation of [10] or [16] requires tr $(\mathbf{QC}) = 0$ (see [9c]) which results from general properties of the $\mathcal{I}$ and $\mathbf{C}$ matrices ([8], [9a] and [9b]) that do not refer to any particular structure (fixed or random) of the vectors of parameters. Again, the only computational adjustment to make is to view the corresponding $\mathcal{I}$ matrices as coefficient matrices of Henderson's mixed model equations (Henderson, 1984) after absorption of the equations in $\boldsymbol{\lambda}$. In fact, this extension fully agrees with the role played by $|\mathbf{C}|$ in the the theory of Bayes D-optimality (see *eg* DasGupta and Studden, 1991).

NUMERICAL EXAMPLE

A small hypothetical data set is employed to illustrate the procedure.

Table I. Layout used in the numerical example.

Individual*	Sire*	Dam*	Sex**	Herd	Year
A***	a	b	–	–	–
B	c	d	M	1	1
C	A	e	F	1	1
D	A	f	M	2	2
E	g	h	F	2	2
F	B	C	F	1	2
G	B	i	F	1	2
H	D	E	F	2	3

*Upper case = identified individuals; lower case = unknown base parents grouped into 3 classes: 1 = acg, 2 = bfh and 3 = dei; ** M = male, F = female; *** No performance recorded.

The layout (table I) consists of a pedigree of 8 individuals (A to H) with performance records on 7 of them (B to H) varying according to sex (s_i; $i = 1, 2$), year (a_j; $j = 1, 2, 3$) and herd (h_k; $k = 1, 2$). Unknown base parents (a to h) were assigned to 3 levels of a group factor (g_l; $l = 1, 2, 3$). Data of this layout are analyzed according to an individual (or "animal") genetic model (Quaas and Pollak, 1980) accomodated to the so-called accumulated grouping procedure of Thompson (1979), Quaas and Pollak (1982), Westell (1984) and Robinson (1986) (see Quaas, 1988 for a synthetic approach to this procedure). Using classical notations, this model can be written as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \qquad [23a]$$

or, using distributions

$$\mathbf{y}|\boldsymbol{\beta}, \mathbf{u} \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \mathbf{I}\sigma_e^2) \qquad [23b]$$

$$\mathbf{u}|\mathbf{g}, \sigma_a^2 \sim N(\mathbf{Q}\mathbf{g}, \mathbf{A}\sigma_a^2) \qquad [23c]$$

where $\mathbf{y}$ is the data vector, $\boldsymbol{\beta}$ is the vector of fixed effects (sex, year, herd), $\mathbf{u}$ is the random vector of breeding values, and $\mathbf{X}$ and $\mathbf{Z}$ are the corresponding incidence matrices. The vector $\mathbf{u}$ of breeding values has expectation $\mathbf{Qg}$ and variance $\mathbf{A}\sigma_a^2$ where $\mathbf{Q}$ defined as in Quaas (1988) assigns proportions of genes from the 3 levels of group (vector $\mathbf{g}$) to the 8 identified individuals, $\mathbf{A}$ is the so-called numerator relationship matrix among those individuals and σ_a^2 is the additive genetic variance. Using Quaas' notations, $\mathbf{u}$ can be alternatively written as:

$$\mathbf{u} = \mathbf{Qg} + \mathbf{u}^* \qquad\qquad [23d]$$

with $\mathbf{u}^* \sim N(\mathbf{0}, \mathbf{A}\sigma_a^2)$ being the random vector of the within-group breeding values.

The (full rank) parameterization chosen here is:

$$\boldsymbol{\beta} = (\mu + s_1 + a_1 + h_1 + g_1;\ s_2 - s_1;\ a_2 - a_1, a_3 - a_1;\ h_2 - h_1)' \qquad [24a]$$

$$\mathbf{g} = (g_2 - g_1;\ g_3 - g_1)' \qquad [24b]$$

$$\mathbf{u} = (u_A, u_B, \ldots, u_H)', \mathbf{u}^* = (u_A^*, u_B^*, \ldots, u_H^*)' \qquad [24c]$$

The grouping strategy of base animals is an issue of great concern for animal breeders due to the possible confounding or poor connectedness with other fixed effects in the model (Quaas, 1988). Therefore, it is of interest to look at the degree of connectedness between this group factor and other fixed effects, or equivalently to degree of connectedness among group levels due to the incidence of other fixed effects. In this example, 3 fixed factors (in addition to group) were considered which are sex (S), year (A) and herd (H) and their incidence on connectedness of groups can be assessed separately (S, A, H) or jointly $(S + A, A + H, H + S, S + A + H)$. From notations in [1], degree of connectedness of G due to A is based on:

$$\boldsymbol{\theta} = (g_2 - g_1;\ g_3 - g_1)' \qquad [25a]$$

$$\boldsymbol{\phi} = (a_2 - a_1;\ a_3 - a_1)' \qquad [25b]$$

$$\boldsymbol{\lambda} = (\mu + s_1 + a_1 + h_1 + g_1;\ s_2 - s_1;\ h_2 - h_1;\ \mathbf{u}^{*\prime})' \qquad [25c]$$

The corresponding information matrix is obtained from the coefficient matrix derived by Quaas (1988) for a mixed model having the structure described in [23a], [23b] and [23c]. Letting the vector of unknowns be $(\boldsymbol{\beta}', \mathbf{g}', \mathbf{u}')'$, this coefficient matrix is given by:

$$\mathbf{W}'\mathbf{W} + \mathbf{A}^* t \qquad\qquad [26]$$

where $\mathbf{W} = (\mathbf{X}|\mathbf{0}|\mathbf{Z}), \mathbf{A}^* = \mathbf{U}'\mathbf{A}^{-1}\mathbf{U}, \mathbf{U}$ being $(\mathbf{0}| - \mathbf{Q}|\mathbf{I})$ and $t = \sigma_e^2/\sigma_a^2$ (here $t = 1.5$).

In this example, the matrices involved in [26] are:

$$\mathbf{Q} = \begin{bmatrix} (0.50) & 0.50 & 0 \\ (0.50) & 0 & 0.50 \\ (0.25) & 0.25 & 0.50 \\ (0.25) & 0.75 & 0 \\ (0.50) & 0.50 & 0 \\ (0.375) & 0.125 & 0.50 \\ (0.25) & 0 & 0.75 \\ (0.375) & 0.625 & 0 \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \end{bmatrix} \quad \mathbf{Z} = \begin{bmatrix} 0 & \mathbf{0}' \\ \mathbf{0} & \mathbf{I}_7 \end{bmatrix}$$

Elements in the first column of $\mathbf{Q}$ within brackets are deleted in the computations due to the parameterization chosen in [24a] and [24b]. $\mathbf{A}^{-1}$ is half stored with non zero elements being:

$(11) = 5/3$; $(13) = (14) = (27) = -2/3$; $(22) = (33) = (44) = 11/6$;
$(23) = (45) = 1/2$; $(26 = (36) = (48) = (58) = -1$; $(55) = 3/2$;
$(66) = (88) = 2$ and $(77) = 4/3$.

$\mathbf{A}^*$ may also be calculated directly from Quaas' rule (Quaas, 1988).

Connectedness between groups due to the incidence of the other fixed effects was assessed under the full model using Quaas' system in [26], and also for an $\mathbf{u}^*$ deleted model ($\mathbf{y} = \mathbf{X\beta} + \mathbf{ZQg} + \mathbf{e}$), then using the ordinary least squares equations. Numerical results are given in table II. In this example, the main sources of disconnectedness are by decreasing order: herd, year and sex, the first factor being by far the most important one since the γ^* values associated with herd are 0.312, 0.247, 0.272 and 0.239 when this factor is considered alone, and with year, sex and year plus sex respectively. Actually, this result is not surprising on account of the grouping procedure based on parents in groups 2 and 3 coming out of different herds. One may also notice that D values for combinations of factors exceed the sum of D values for single factors. For instance, D is equal to 1.433 for $S + A + H$ vs $\Sigma D = 1.316$ for each factor taken separately. Results for the purely fixed model ($\mathbf{u}^*$ deleted) are in close agreement with those of the full model. This procedure of ignoring $\mathbf{u}^*$ effects for investigating linkage among groups was first advocated by Smith et al (1988) due to its relative ease of computation in large field data sets.

The extension of the theory to the measure of degree of connectedness of random factors is illustrated in this example by calculations of D and γ^* for breeding values (table II). Sources of unbalancedness rank as previously, but the average level of connectedness ($\gamma^* = 0.574$) for breeding values in higher than for groups ($\gamma^* = 0.239$) due to prior information (Foulley et al, 1990).

The theory also applies to specific contrasts among effects as originally proposed by Foulley et al (1984, 1990). The degree of connectedness for pair comparisons among breeding values then reduces, simply to the ratio of prediction error variance of the pair comparison under a reduced model (R) with some effects deleted (in table III, all fixed effects except mean and group) and under the full model (F), ie:

Table II. Degree of connectedness among groups and among breeding values due to the incidence of sex (S), year (A) and herd (H).

| Source | Groups | | | | Breeding values | |
| | u^*-included | | u^*-deleted | | | |
	D	γ^*	D	γ^*	D	γ^*
S	0.002	0.998	0.000	1.000	0.230	0.944
A	0.148	0.862	0.172	0.842	0.409	0.903
H	1.166	0.312	1.234	0.291	1.458	0.695
$S + A$	0.200	0.819	0.203	0.817	0.633	0.854
$A + H$	1.398	0.247	1.518	0.219	2.019	0.604
$H + S$	1.303	0.272	1.337	0.263	1.793	0.639
$S + A + H$	1.433	0.239	1.532	0.216	2.220	0.574

Table III. Degree of connectedness for pair comparisons among breeding values computed exactly (I) or approximated *via* their group-components (II).

II-I[a]	A	B	C	D	E	F	G	H
A		0.095	0.146	0.788	0.847	0.102	0.111	0.797
B	0.064		0.828	0.092	0.058	0.840	0.441	0.064
C	0.136	0.808		0.058	0.088	0.792	0.296	0.059
D	0.808	0.148	0.064		0.809	0.060	0.047	0.822
E	1.000	0.064	0.136	0.808		0.057	0.061	0.802
F	0.072	0.808	0.808	0.095	0.072		0.443	0.047
G	0.087	0.399	0.064	0.064	0.087	0.136		0.045
H	0.808	0.095	0.072	0.808	0.808	0.064	0.064	

[a]I: above the diagonal II: below the diagonal (see definitions in [28a] and [28b] respectively).

$$\gamma_{ii'} = \text{Var}_R(\widehat{\delta}_{ii'} - \delta_{ii'})/\text{Var}_F(\widehat{\delta}_{ii'} - \delta_{ii'}) \qquad [27]$$

where $\delta_{ii'} = u_i - u_{i'}$.

Table III gives such results for specific pair comparisons among breeding values either defined exactly (I):

$$u_i - u_{i'} = \sum_\ell (q_{i\ell} - q_{i'\ell})g_\ell + u_i^* - u_{i'}^* \qquad [28a]$$

or approximated (II) *via* their group component:

$$u_i - u_{i'} \cong \sum_\ell (q_{i\ell} - q_{i'\ell})g_\ell. \qquad [28b]$$

Figures shown reflect a great heterogeneity in the pattern of degree of connectedness. This diversity can usually be well explained by looking at the levels of factors

which differ or are shared by individuals compared. For instance, B and F are closely connected ($\gamma^* = 0.840$ and 0.808 in I and II respectively) because they are in the same herd and share close proportions of genes from the 3 groups of base parents (0.5, 0 and 0.5 from groups 1, 2 and 3 respectively in B vs 0.375, 0.125 and 0.5 in F). On the contrary, D and G who are coming from different herds and for whom, $3/4$ of their genes are originating from different groups (groups 2 and 3 respectively) are poorly connected ($\gamma^* = 0.047$ and 0.064 in I and II respectively). Moreover, γ^* values computed according to both procedures (exact or approximate definition) are in good agreement in this example although it is difficult to draw general conclusions from such a limited example.

DISCUSSION AND CONCLUSION

This paper provides a theoretical framework to the definition of an objective criterion for measuring the degree of connectedness between factors involved in Gaussian linear models of genetic evaluation. The procedure proposed herein is based upon the assessment of non-orthogonality between estimators of contrasts (or errors of prediction for random effects) via the Kullback–Leibler distance. This measure offers great flexibility since it can be employed for a particular comparison among levels of some factor or for a global evaluation of their degree of connectedness. Applications of these criteria to degree of connectedness among sires in a reference sire system based on planned artificial inseminations with link bulls have already been made in France (Foulley $et\ al$, 1990; Hanocq $et\ al$, 1992; Laloe $et\ al$, 1992).

The criterion derived is invariant to one-to-one linear transformations on the vector of parameters θ or ϕ. Letting $\theta^* = S\theta$ with S being a full rank transformation matrix, the characteristic equation in [18] becomes $|SC_{\theta\theta.\phi}S' - kSC_{\theta\theta}S'| = 0$ which reduces to the original equation by factorizing $|S| \neq 0$. This property ensures that D does not depend on the contrasts chosen among the θ_i's provided the parameterization in θ (for fixed effects) consists of the maximum number of linearly independent estimable functions.

Other criteria may be envisioned. Foulley $et\ al$ (1990) suggested using as a measure of disconnectedness the criterion:

$$E = 1/2[\text{tr}(\mathbf{C}_R^{-1}\mathbf{C}_F) - \ln|\mathbf{C}_R^{-1}\mathbf{C}_F| - \dim(\mathbf{C}_R)] \qquad [29]$$

where $\mathbf{C}_R$ and $\mathbf{C}_F$ are the same as in [16]. This criterion appears also in statistical inference on variance-covariance matrices as the so-called Stein loss function (Anderson, 1984; Loh, 1991). Here, it can be interpreted as the Kullback–Leibler distance between the marginal density $f(\widehat{\theta})$ of $\widehat{\theta}$, and its conditional density, $f(\widehat{\theta}|\phi)$, given the value of the parameter ϕ.

The feasibility of our procedure is determined by the ability to compute the logarithm of the determinant of a coefficient matrix after possible absorption of some factors as required by other statistical procedures based on the likelihood function. In the current context of genetic evaluation with the animal model, an application of this procedure to phantom groups might be feasible using, at least, the model ignoring u^* as a first approximation.

In that respect, it has also been suggested (Kennedy and Trus, 1991) to look at the elements of the coefficient matrix $\mathbf{X'ZQ}$ whose relative values in row k provides the expected proportions of genes out of the different levels of groups contributing to the corresponding level of the kth fixed effect. In our example, these values are as follows:

sex	1:	0.375	0.375	0.250
sex	2:	0.350	0.300	0.350
year	1:	0.375	0.125	0.500
year	2:	0.344	0.344	0.312
year	3:	0.375	0.625	0.000
herd	1:	0.344	0.094	0.562
herd	2:	0.375	0.625	0.000

These figures show a more unbalanced distribution across herd and/or year than across sex levels. Notice that this matrix gives the distribution of data according to groups for each factor separately. No account is taken of the joint distribution of data between those factors. In this model, this means that the factors sex and group are not perfectly connected due to slighty unbalanced proportions observed. As a matter of fact, $\widehat{s}_2 - \widehat{s}_1$ is correlated to $\widehat{g}_2 - \widehat{g}_1$ and $\widehat{g}_3 - \widehat{g}_1$ in the "sex + group" model whereas they are uncorrelated in the full model (see table II).

The γ^* criterion applied to breeding values measures how the $\mathbf{C}$ matrix of variances of prediction errors is reshaped due to the incidence of an unbalanced distribution of data across the nuisance factors. This change in $\mathbf{C}$ implies a related change in the variance covariance matrix of estimated breeding values which influences the selection differential. Accuracy of selection is also expected to be altered. In this respect, insufficient connectedness can be compared to some extent to some non-optimum selection procedure which ignores, or does not weight properly, some sources of information, *eg*, within family selection *vs* index selection. More research is needed in this field to quantify the amount of genetic progress which may be lost due to reduction in the degree of connectedness.

For fixed effects, connectedness is directly related to the unbiasedness requirement. This is especially true for group effects in the animal model for which much concern has been raised (Smith *et al*, 1988; Quaas, 1988; Canon *et al*, 1992). The criterion developed here may help to check whether differences between groups in a particular model can be reasonably captured by the data structure. If not, one will have to reconsider the grouping procedure, or one may be tempted to put prior information on group effects *ie* to treat them as random as suggested by Foulley *et al* (1990). In any case, one will have to compare different models and there are now specific statistical procedures available to do that in animal breeding (Wada and Kashiwagi, 1990).

APPENDIX

Another look at the standardization procedure

The starting point consists of decomposing the joint density $f(\widehat{\theta}, \widehat{\phi})$ according to the elements in $\widehat{\theta}$. Let us consider for the sake of simplicity the case of 2 elements

$\widehat{\theta}_1$ and $\widehat{\theta}_2$. One has

$$f(\widehat{\theta}_1, \widehat{\theta}_2, \widehat{\phi}) == f(\widehat{\theta}_1, \widehat{\phi}) f(\widehat{\theta}_2 | \widehat{\theta}_1, \widehat{\phi}) \qquad [A.1a]$$

Now $f(\widehat{\theta}_2 | \widehat{\theta}_1, \widehat{\phi})$ can be rewritten as:

$$f(\widehat{\theta}_2 | \widehat{\theta}_1, \widehat{\phi}) = f(\widehat{\theta}_2, \widehat{\phi} | \widehat{\theta}_1) / f(\widehat{\phi} | \widehat{\theta}_1) \qquad [A.1b]$$

Putting [A.1b] into [A.1b] and dividing both sides by $f(\widehat{\theta}_1, \widehat{\theta}_2) f(\widehat{\phi})$ gives

$$\frac{f(\widehat{\theta}_1, \widehat{\theta}_2, \widehat{\phi})}{f(\widehat{\theta}_1, \widehat{\theta}_2) f(\widehat{\phi})} = \frac{f(\widehat{\theta}_1, \widehat{\phi})}{f(\widehat{\theta}_1) f(\widehat{\phi})} \frac{f(\widehat{\theta}_2, \widehat{\phi} | \widehat{\theta}_1)}{f(\widehat{\theta}_2 | \widehat{\theta}_1) f(\widehat{\phi} | \widehat{\theta}_1)}$$

or, in shorter notations,

$$R(\widehat{\theta}_1, \widehat{\theta}_2, \widehat{\phi}) = R(\widehat{\theta}_1, \widehat{\phi}) R(\widehat{\theta}_2, \widehat{\phi} | \widehat{\theta}_1) \qquad [A.2]$$

where $R(x, y | z) = f(x, y | z) / f(x | z) f(y | z)$.

Using [A.2], the Kullback-Leibler distance $D(\widehat{\theta}_1, \widehat{\theta}_2, \widehat{\phi})$ defined in [3] can be expressed as the sum of 2 terms:

$$D(\widehat{\theta}_1, \widehat{\theta}_2, \widehat{\phi}) = \int \int \int \ln\left[R(\widehat{\theta}_1, \widehat{\phi})\right] f(\widehat{\theta}_1, \widehat{\phi}) f(\theta_2 | \widehat{\theta}_1, \widehat{\phi}) d\widehat{\theta}_1 d\widehat{\theta}_2 d\widehat{\phi} \qquad [A.3a]$$

$$+ \int \int \int \ln\left[R(\widehat{\theta}_2, \widehat{\phi} | \widehat{\theta}_1)\right] f(\widehat{\theta}_2, \widehat{\phi} | \widehat{\theta}_1) f(\widehat{\theta}_1) d\widehat{\theta}_1 d\widehat{\theta}_2 d\widehat{\phi} \qquad [A.3b]$$

After integrating out $\widehat{\theta}_1$ and $\widehat{\phi}$, the first term [A.3a] can be written as

$$\int D(\widehat{\theta}_1, \widehat{\phi}) f(\widehat{\theta}_2 | \widehat{\theta}_1, \widehat{\phi}) d\widehat{\theta}_2$$

which reduces to $D(\widehat{\theta}_1, \widehat{\phi})$ since, according to [10], this term is a constant.

The second term [A.3b] can be viewed as:

$$E_{\widehat{\theta}_1} \left\{ E\left[\ln R(\widehat{\theta}_2, \widehat{\phi} | \widehat{\theta}_1) \right] \right\}$$

ie the expectation with respect to the distribution of $\widehat{\theta}_1$ of the conditional expectation of $\ln R(\widehat{\theta}_2, \widehat{\phi} | \widehat{\theta}_1)$ taken with respect to the distribution of $\widehat{\theta}_2, \widehat{\phi}$ given $\widehat{\theta}_1$. This conditional expectation is by definition a D-measure noted $D(\widehat{\theta}_2, \widehat{\phi} | \widehat{\theta}_1)$; because this is again a constant (see [10]):

$$D(\widehat{\theta}_2, \widehat{\phi} | \widehat{\theta}_1) = -\frac{1}{2} \ln \frac{|\mathbf{C}_{\theta_2\phi \; \theta_2\phi \cdot \theta_1}|}{|\mathbf{C}_{\theta_2\theta_2 \cdot \theta_1}| |\mathbf{C}_{\phi\phi \cdot \theta_1}|}$$

which does not depend upon $\widehat{\theta}_1$, [A.3b] reduces to that term.

Hence, after regrouping the expressions for [A.3a] and [A.3b], one has:

$$D(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\phi}}) = D(\widehat{\theta}_1, \widehat{\boldsymbol{\phi}}) + D(\widehat{\theta}_2, \widehat{\boldsymbol{\phi}}|\widehat{\theta}_1).$$ [A.4]

Similarly, $\gamma(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\phi}}) = \exp\left[-2D(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\phi}})\right]$ can be expressed as the following product:

$$\gamma(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\phi}}) = \gamma(\widehat{\theta}_1, \widehat{\boldsymbol{\phi}})\gamma(\widehat{\theta}_2, \widehat{\boldsymbol{\phi}}|\widehat{\theta}_1)$$ [A.5a]

and equivalently after permutation of $\widehat{\theta}_1$ and $\widehat{\theta}_2$, as:

$$\gamma(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\phi}}) = \gamma(\widehat{\theta}_2, \widehat{\boldsymbol{\phi}})\gamma(\widehat{\theta}_1, \widehat{\boldsymbol{\phi}}|\widehat{\theta}_2).$$ [A.5b]

Thus, letting $\Gamma(\widehat{\theta}_i, \widehat{\boldsymbol{\phi}})$ such that:

$$\left[\Gamma(\widehat{\theta}_i, \widehat{\boldsymbol{\phi}})\right]^2 = \gamma(\widehat{\theta}_i, \widehat{\boldsymbol{\phi}})\gamma(\widehat{\theta}_i, \widehat{\boldsymbol{\phi}}|\widehat{\theta}_j) \quad \text{for} \quad i = 1, 2 \quad \text{and} \quad j \neq i$$ [A.6]

one has:

$$\gamma(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\phi}}) = \prod_i \Gamma(\widehat{\theta}_i, \widehat{\boldsymbol{\phi}})$$ [A.7]

and $\gamma^*(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\phi}}) = \left[\gamma(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\phi}})\right]^{1/2}$ can be interpreted as, either the geometric mean of the $\Gamma(\widehat{\theta}_i, \widehat{\boldsymbol{\phi}})$ coefficients in [A.6], or as the geometric mean of all possible $\gamma(\widehat{\theta}_i, \widehat{\boldsymbol{\phi}}|\widehat{\theta}_j)$ coefficients (including the unconditional ones). For three elements $\theta_i, \theta_j, \theta_k$, in $\boldsymbol{\theta}$, one would have:

$$\left[\Gamma(\widehat{\theta}_i, \widehat{\boldsymbol{\phi}})\right]^3 = \gamma(\widehat{\theta}_i, \widehat{\boldsymbol{\phi}})\left[\prod_{j \neq i}\gamma(\widehat{\theta}_i, \widehat{\boldsymbol{\phi}}|\widehat{\theta}_j)\right]^{1/2}\gamma(\widehat{\theta}_i, \widehat{\boldsymbol{\phi}}|\widehat{\theta}_j, \widehat{\theta}_k)$$ [A.8a]

and similarly for 4 elements $\theta_i, \theta_j, \theta_k, \theta_l$

$$\left[\Gamma(\widehat{\theta}_i, \widehat{\boldsymbol{\phi}})\right]^4 = \gamma(\widehat{\theta}_i, \widehat{\boldsymbol{\phi}})\left[\prod_{j \neq i}\gamma(\widehat{\theta}_i, \widehat{\boldsymbol{\phi}}|\widehat{\theta}_j)\right]^{1/3}$$

$$\left[\prod_{j \neq i, k > j}\gamma(\widehat{\theta}_i, \widehat{\boldsymbol{\phi}}|\widehat{\theta}_j, \widehat{\theta}_k)\right]^{1/3}\gamma(\widehat{\theta}_i, \widehat{\boldsymbol{\phi}}|\widehat{\theta}_j, \widehat{\theta}_k, \widehat{\theta}_\ell)$$ [A.8b]

These formulae can be easily extended to any number of elements r_θ in $\boldsymbol{\theta}$. Formula [A.7] applies and the coefficient of the power pertaining to the $\gamma(\widehat{\theta}_i, \widehat{\boldsymbol{\phi}}|\widehat{\theta}_j, \ldots)$ term given k variables $\widehat{\theta}_j$ in $\left[\Gamma(\widehat{\theta}_i, \widehat{\boldsymbol{\phi}})\right]^{r_\theta}$ is then $1/C_{r_\theta - 1}^k$ *ie* the inverse of the coefficient for the kth power in the binomial expansion of order $r_\theta - 1$.

ACKNOWLEDGMENTS

REFERENCES

Anderson TW (1984) *An Introduction to Multivariate Statistical Analysis.* John Wiley and Sons, New York

Baker RL, Paratt AC (1988) Evaluation, selection and diffusion of sires using artificial insemination and/or natural mating. *In: 3rd World Congr Sheep and Beef Cattle Breeding, Paris, 19-23 June 1988*, INRA, Paris, vol 1, 121-140

Bichard M (1987) Problems of across-population genetic evaluation within improvement schemes. *In: Proc 6th Conf the Australian Association of Animal Breeding and Genetics. Perth, WA, 9-11 February, 1987*, 103-111

Calinski T (1977) On the notion of balance in block designs. *In: Recent Development in Statistics* (Barra JR, ed) North Holland, 365-374

Canon J, Gruand J, Gutierrez JP, Ollivier L (1992) Expérience de sélection sur la croissance du tissu maigre chez le porc avec des pères répétés: Évolution génétique des caractères soumis à la sélection. *Genet Sel Evol* 24 (5) (in press)

DasGupta A, Studden WJ (1991) Robust Bayesian experimental designs in normal linear models. *Ann Stat* 19, 1244-1256

Fernando R, Gianola D, Grossman M (1983) Identifying all connected subsets in a two-way classification without interaction. *J Dairy Sci* 66, 1399-1402

Foulley JL, Schaeffer LR, Song H, Wilton JW (1983) Progeny group size in an organized progeny test program of AI beef bulls using reference sires. *Can J Anim Sci* 63, 17-26

Foulley JL, Bouix J, Goffinet B, Elsen JM (1984) Comparaison de pères et connexion. *In: Insémination Artificielle et Amélioration Génétique: Bilan et Perspectives Critiques* (Elsen JM, Foulley JL, eds) Colloq de l'INRA 29, 131-176

Foulley JL, Bouix J, Goffinet B, Elsen JM (1990) Connectedness in genetic evaluation. *In: Advances in Statistical Methods for Genetic Improvement of Livestock* (Gianola D, Hammond K, eds) Springer-Verlag, Heidelberg, 277-308

Gupta S (1987) A note on the notion of balance in designs. *Calcutta Stat Assoc Bull* 36, 85-89

Hanocq E, Foulley JL, 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 1992*, 4 p

Henderson CR (1984) *Applications of Linear Models in Animal Breeding.* University of Guelph, Guelph

Henderson CR (1988) Theoretical basis and computational methods for a number of different animal models. *J Dairy Sci* 71, 1-16 (suppl 2)

Johnson NL, Kotz S (1972) *Distributions in Statistics: Continuous Multivariate Distributions.* John Wiley and Sons, New York

Kennedy BW (1987) Genetic evaluation in swine using the animal model. *In: 38th Annu Meet EAAP. Lisbon, Portugal, Sept 28–Oct 1st, 1987*, vol 1, 38-39 (abstr)

Kennedy BW, Trus D (1991) Measures of connectedness among management units under an animal model. *In: 86th ADSA Annu Meet Logan, Utah, August 12-15, 1991. J Dairy Sci* 44, suppl 1, 159 (abstr)

Kullback S (1968) *Information Theory and Statistics.* Dover, New York

Kullback S (1983) Kullback information *In: Encyclopedia of Statistical Sciences* (Kotz S, Johnson NL, eds) John Wiley and Sons, New York, vol 4, 421-425

Laloe D, Sapa J, Menissier F, Renand G (1992) Use of the relationship matrix and planned matings in the evaluation of natural service sires of French beef breeds. *Genet Sel Evol* 24, 137-145

Loh WL (1991) Estimating covariance matrices. *Ann Stat* 19, 283-296

Miraei Ashtiani SR, James JW (1990) Efficient use of link rams in merino sire reference schemes. *In: Proc Austr Assoc Anim Breed Genet* 9, 388-391

Petersen PH (1978) A test for connectedness fitted for the two-way BLUP sire evaluation. *Acta Agric Scand* 28, 360-362

Quaas RL (1988) Additive genetic model with groups and relationships. *J Dairy Sci* 71, 1338-1345

Quaas RL, Pollak EJ (1980) Mixed model methodology for farm and ranch beef cattle testing programs. *J Anim Sci* 51, 1277-1287

Quaas RL, Pollack EJ (1982) Thompson's accumulated group model for sire evaluation. *J Anim Sci* 55 (suppl 1) 160 (abstr)

Robinson GK (1986) Group effects and computing strategies for models for estimating breeding values. *J Dairy Sci* 69, 3106-3111

Shah KR, Yadolah D (1977) On the connectedness of designs. *Sankhya* 39, 284-287

Smith S, Scarth RD, Tier B (1988) *Genetic Groups and the Feasability of Reference Sire Schemes.* Tec Rep, AGBU, Univ New England, Armidale

Sölkner J, James JW (1990) Optimum design of crossbreeding experiments. I-A basic sequential procedure. *J Anim Breed Genet* 107, 61-67

Thompson R (1979) Sire evaluation. *Biometrics* 35, 339-353

Tosh JJ, Wilton JW (1990) Degree of connectedness in mixed models. *In: 4th World Congr Genetics Applied to Livestock Production, Edinburgh, 23–27 July 1990, vol 13* (Hill WG, Thompson R, Wooliams JA, eds) 480-483

Wada Y, Kashiwagi N (1990) Selecting statistical models with information statistics. *J Dairy Sci* 73, 3575-3582

Webb AJ (1987) Choice of selection objectives in specialized sire and dam lines for commercial crossbreeding. *In: 38th Annu Meet EAAP, Lisbon, Portugal, Sept 28–Oct 1st, 1987*, vol 2, 1162-1163, (abstr)

Weeks DL, Williams DR (1964) A note on the determination of connectedness in an N-way cross-classification. *Technometrics* 6, 319-324

Westell RA (1984) Simultaneous evaluation of sires and cows for a large population. PhD thesis, Cornell University, Ithaca, NY