

# Considerations on measures of precision and connectedness in mixed linear models of genetic evaluation

D Laloë, F Phocas, F Ménéssier

*Station de génétique quantitative et appliquée, Institut national de la recherche agronomique, 78352 Jouy-en-Josas cedex, France*

(Received 5 April 1995; accepted 24 May 1996)

**Summary** – Three criteria for the quality of a genetic evaluation are compared: the prediction error variance (PEV); the loss of precision due to the estimation of the fixed effects (degree of connectedness) (IC); and a criterion related to the information brought by the evaluation in terms of generalized coefficient and determination (CD) (precision). These criteria are introduced through simple examples based on an animal model. The main differences between them are the choice of the matrix studied (CD vs PEV, IC), the method used to account for the relationships (CD vs PEV), the use of a reference matrix or model (PEV vs CD, IC), and the data design (IC vs PEV, CD). IC is shown to favor designs with limited information provided by the data and another index is suggested, which minimizes this drawback. The behavior of IC and CD is studied in a hypothetical ‘herd + sire’ model. The precision criteria set a balance between connectedness level and information provided by the data, whereas the connectedness criteria favor the model with minimum information and maximum connectedness level. Genetic relationships between animals decrease both PEV and genetic variability. PEV considers only the favorable effects on PEV; CD accounts for both effects. CD sets a balance between the design and the information brought by the data, the PEV and the genetic variability and is thus a method of choice for studying the quality of a genetic evaluation.

**genetic evaluation / precision / mixed linear model / disconnectedness / genetic progress**

**Résumé** – Quelques considérations à propos des mesures de précision et de connexion dans les modèles linéaires mixtes d'évaluation génétique. Trois critères d'appréciation de la connexion et de la précision des évaluations génétiques sont étudiés et comparés. Le premier critère est la variance d'erreur de prédiction (PEV), le second mesure la diminution de la PEV quand les effets fixés sont connus (indice de connexion ou IC), et le troisième est un critère de précision de l'évaluation, exprimé par le coefficient de détermination généralisé (CD). Ces critères sont présentés à l'aide d'exemples simples basés sur un modèle animal. Ils se distinguent par le choix de la matrice étudiée (CD versus PEV, IC), la prise en compte de la seule structure des données (IC versus PEV, CD), la présence d'une matrice ou d'un modèle de référence (PEV versus IC, CD), et la

*manière de prendre en compte les relations de parenté entre animaux (CD versus PEV). On montre comment IC favorise les situations où l'information apportée par les données est faible. Un nouvel indice de connexion, s'attachant également à la seule structure des données, est proposé, palliant cet inconvénient. L'intérêt d'IC et de CD est étudié sur un exemple de modèle «troupeau-père», où les troupeaux sont de taille fixée, les pères servent dans un seul troupeau, à l'exception d'un père de référence assurant les liaisons génétiques entre troupeaux. CD permet d'optimiser le plan d'expérience par un compromis entre connexion et information contenue dans les données, alors que l'utilisation d'IC aboutit au choix d'un plan où les pères utilisés dans un seul troupeau ont un seul veau par troupeau. Si CD et PEV sont équivalents pour des animaux non apparentés, PEV privilégie les forts apparentements, qui diminuent la variance d'erreur de prédiction. Mais les parentés diminuent également la variabilité génétique, ce que prend en compte CD. Ainsi, on montre, sur un modèle animal strictement aléatoire avec même apparentement entre animaux, comment PEV peut conduire au choix d'un plan minimisant le progrès génétique. On retrouve dans ce cas simple la formule classique du progrès génétique, où le CD généralisé joue le même rôle que le CD individuel d'un indice de sélection. CD, compromis entre structure et quantité de données, d'une part, et variance d'erreur de prédiction et variabilité génétique, d'autre part, est une méthode de choix pour l'analyse de la qualité d'une évaluation génétique.*

**évaluation génétique / précision / modèle linéaire mixte / disconnexion / progrès génétique**

## INTRODUCTION

The problem of precision and especially of disconnectedness in BLUP genetic evaluation, is becoming increasingly important in animal breeding. Since the work of Petersen (1978) and Foulley et al (1984, 1990), three papers have addressed this subject: Foulley et al (1992), Kennedy and Trus (1993), and Laloë (1993).

In the context of genetic evaluation, disconnectedness is not clearly defined. Sometimes, it is the lack of genetic ties between levels of fixed effects, and other times it is defined as the inestimability of contrasts between levels of genetic effects. Both definitions are somewhat incoherent, since, as Foulley et al (1992) wrote "From a theoretical point of view, complete disconnectedness among random effects can never occur". These authors introduced the concept of "level (or degree) of disconnectedness" by relating the prediction error variance (PEV) of the genetic effects to the PEV under a reduced model excluding the fixed effects. They suggested a global measure of connectedness among levels of a factor. Kennedy and Trus (1993) suggested the PEV of differences in predicted genetic values between candidates for selection as the most appropriate measure of connectedness. Laloë (1993) introduced the concept of generalized coefficient of determination (CD), the CD of a linear combination of genetic values, and suggested a new definition of disconnectedness among random effects: a design is disconnected for a random factor if the generalized CD of a contrast between its levels is null. Some global measures of the precision of an evaluation or of a set of evaluated animals were suggested.

The aim of this paper is to compare the three methods, theoretically and with some numerical examples based on animal models and sire models.

## MODELS, NOTATION AND CRITERIA

Consider a mixed model with one random factor (and the residual effect)

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad [1]$$

where  $\mathbf{y}$  is the performance vector of dimension  $n$ ,  $\mathbf{b}$  the fixed effect vector,  $\mathbf{X}$  the pertinent incidence matrix,  $\mathbf{u}$  the random effect vector,  $\mathbf{Z}$  the corresponding incidence matrix and  $\mathbf{e}$  the residual vector.

$$\begin{pmatrix} \mathbf{u} \\ \mathbf{e} \end{pmatrix} \sim N \left[ \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{A}\sigma_a^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\sigma_e^2 \end{pmatrix} \right]$$

where  $\mathbf{A}$  is the numerator relationship matrix, and the scalars  $\sigma_a^2$  and  $\sigma_e^2$  are the additive and residual variance components, respectively. BLUP (best linear unbiased predictor) of  $\mathbf{u}$ , denoted  $\hat{\mathbf{u}}$ , is the solution of  $(\mathbf{Z}'\mathbf{M}\mathbf{Z} + \lambda\mathbf{A}^{-1})\hat{\mathbf{u}} = \mathbf{Z}'\mathbf{M}\mathbf{y}$ , where  $\lambda = \sigma_e^2/\sigma_a^2$ , and  $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$  is a projection matrix orthogonal to the vector subspace spanned by the columns of  $\mathbf{X}$ :  $\mathbf{M}\mathbf{X} = \mathbf{0}$ .

The joint distribution of  $\mathbf{u}$  and  $\hat{\mathbf{u}}$  is multivariate normal, with a null expectation and variance matrix equal to

$$\Sigma_1 = \begin{pmatrix} \mathbf{A}\sigma_a^2 & \Psi \\ \Psi & \Psi \end{pmatrix}$$

with  $\Psi = \mathbf{A}\sigma_a^2 - (\mathbf{Z}'\mathbf{M}\mathbf{Z} + \lambda\mathbf{A}^{-1})^{-1}\sigma_e^2 = \mathbf{A}\sigma_a^2 - \mathbf{C}^{\mathbf{u}\mathbf{u}}\sigma_e^2$

The distributions of  $\mathbf{u}|\hat{\mathbf{u}}$  and  $\mathbf{u} - \hat{\mathbf{u}}$  are multivariate normal:  $N(\hat{\mathbf{u}}, \mathbf{C}^{\mathbf{u}\mathbf{u}}\sigma_e^2)$  and  $N(\mathbf{0}, \mathbf{C}^{\mathbf{u}\mathbf{u}}\sigma_e^2)$ , respectively.

The following is a second model:

$$\mathbf{y} = \mathbf{1}\mathbf{m} + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad [2]$$

With this random model,  $\mathbf{u}|\hat{\mathbf{u}} \sim N(\hat{\mathbf{u}}, \mathbf{C}_r^{\mathbf{u}\mathbf{u}}\sigma_e^2)$  and  $\mathbf{u} - \hat{\mathbf{u}} \sim N(\mathbf{0}, \mathbf{C}_r^{\mathbf{u}\mathbf{u}}\sigma_e^2)$ , with  $\mathbf{C}_r^{\mathbf{u}\mathbf{u}} = (\mathbf{Z}'\mathbf{M}_r\mathbf{Z} + \lambda\mathbf{A}^{-1})^{-1}$  and  $\mathbf{M}_r = \mathbf{I} - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'$ , the projection matrix orthogonal to the vector  $\mathbf{1}$ . This model can be considered to exhibit the information provided by the data in order to predict genetic values, without any loss due to the estimation of fixed effects, except the mean.

### Criteria

Three criteria are proposed to judge the quality of the prediction of a contrast, ie, a linear combination of the breeding values  $\mathbf{x}'\mathbf{u}$ , where  $\mathbf{x}$  is a vector whose elements sum to 0:

– PEV( $\mathbf{x}$ ) (Kennedy and Trus, 1993). Comparisons between animals that are poorly connected would have higher prediction error than those that are well connected.

This method is denoted PEV.

$$\text{PEV}(\mathbf{x}) = \mathbf{x}'\mathbf{C}^{\text{uu}}\mathbf{x}$$

– IC( $\mathbf{x}$ ), the connectedness index (Foulley et al 1992), ie, the relative decrease in PEV when fixed effects are exactly known or do not exist (reduced model). It varies between 0 and 1, and is close to 1 when the animals are well connected. This method is denoted IC.

$$\text{IC}(\mathbf{x}) = \mathbf{x}'\mathbf{C}_r^{\text{uu}}\mathbf{x}/\mathbf{x}'\mathbf{C}^{\text{uu}}\mathbf{x} \quad [4]$$

– CD( $\mathbf{x}$ ), the generalized CD (Laloë, 1993), which corresponds to the square of the correlation between the predicted and the true difference of genetic values. This method is denoted CD.

$$\text{CD}(\mathbf{x}) = 1 - (\text{var}(\mathbf{x}'\mathbf{u}|\hat{\mathbf{u}})/\text{var}(\mathbf{x}'\mathbf{u})) = \mathbf{x}'(\mathbf{A} - \lambda\mathbf{C}^{\text{uu}})\mathbf{x}/\mathbf{x}'\mathbf{A}\mathbf{x} \quad [5]$$

## AN ANIMAL MODEL EXAMPLE

The examples from Kennedy and Trus (1993) are used to illustrate the three measures. Consider an animal model for which there are two management unit effects that are estimated from the data jointly with the genetic values of four animals. All animals have single records. The first two animals ( $u_1$  and  $u_2$ ) are in unit 1, and the last two ( $u_3$  and  $u_4$ ) are in unit 2. Heritability equals 0.5 and  $\sigma_e^2 = \sigma_a^2 = 1$  ( $\lambda = 1$ ). Two cases are considered: (i) the animals are unrelated, and (ii) animals are unrelated within management unit, but each animal has a full sib in the other management unit; ( $u_1, u_3$ ) and ( $u_2, u_4$ ) are full-sib pairs. Obviously, there are no genetic ties between management units in case (i), and the corresponding design is genetically disconnected. Four contrasts between animals are considered: animals within a management unit ( $u_1 - u_2$ ), animals from different management units ( $u_1 - u_3$  and  $u_2 - u_3$ ) and genetic levels of the units ( $u_1 + u_2 - u_3 - u_4$ ). For each contrast, the above three criteria were calculated, and their values are presented in table I. Some comments about these values allow the identification of following problems.

First, IC could not detect any lack of genetic links between units. Its value was 0.5 in case (i) (unrelated animals) for  $u_1 + u_2 - u_3 - u_4$ . Kennedy and Trus (1993) showed that PEV could detect lack of genetic links between units by a covariance of 0 between the BLUE (best linear unbiased estimator) of these units.

Second, disconnectedness was detected by CD, which delivered null CD for the unit comparison, whatever the case, ie, even if the units were genetically linked. Here, the design was such that a difference of genetic levels between units could not be predicted:  $u_1 + u_2 - u_3 - u_4$  was always null, whatever the data, as proven in *Appendix 1*. This concept of connectedness is not equivalent to the lack of genetic links between management units, but to the lack of information provided by the data ( $\text{var}(\mathbf{x}'\mathbf{u}|\hat{\mathbf{u}}) = \text{var}(\mathbf{x}'\mathbf{u})$ ). However, PEV showed that the genetic levels of the units were more likely to be the same in case (ii) than in case (i), due to the genetic links between units in case (ii): PEV = 4 in case (i) and PEV = 2 in case (ii).

**Table I.** Values of the criteria for the animal model example.<sup>a</sup>

<i>Contrast</i>	$u_1 - u_2$	$u_1 - u_3$	$u_2 - u_3$	$u_1 + u_2 - u_3 - u_4$
IC				
(i)	1	0.67	0.67	0.5
(ii)	1	0.80	0.85	0.67
CD				
(i)	0.5	0.25	0.25	0
(ii)	0.53	0.17	0.45	0
PEV				
(i)	1	1.5	1.5	4
(ii)	0.93	0.83	1.1	2

<sup>a</sup> IC, CD and PEV defined in text. (i) Unrelated animals; (ii) ( $u_1, u_3$ ) and ( $u_2, u_4$ ) are full-sib pairs.

Finally, the two methods (PEV and CD) accounted for relationships between animals in different ways. Genetic links between units increased the CD of  $u_2 - u_3$  (unrelated animals of different units), 0.45 (case (ii)) vs 0.25 (case (i)), but the CD of  $u_1 - u_3$  (related animals of different units) decreased, 0.17 (case (ii)) vs 0.25 (case (i)). PEV decreased in both cases. This decrease was higher for related animals, 0.83 (case (ii)) vs 1.5 (case (i)) than for unrelated ones, 1.1 (case (ii)) vs 1.5 (case (i)). The two methods give, therefore, contradictory results. Indeed, the more the animals were related, the lower the genetic variability of their comparison; PEV( $\mathbf{x}$ ) decreased, but so did  $\mathbf{x}'\mathbf{A}\mathbf{x}$ . The variance of  $\mathbf{x}'\hat{\mathbf{u}}$  was proportional to  $\mathbf{x}'\mathbf{A}\mathbf{x} - \lambda\text{PEV}(\mathbf{x})$ . If the relative decrease of PEV( $\mathbf{x}$ ) were smaller than the relative decrease of  $\mathbf{x}'\mathbf{A}\mathbf{x}$ , the variance of  $\mathbf{x}'\hat{\mathbf{u}}$  would decrease, and hence the probability that high differences between animals could be exhibited by the evaluation. For instance, in case (i) (unrelated animals), PEV( $\mathbf{x}$ ) = 1.5 and  $\mathbf{x}'\mathbf{A}\mathbf{x} = 2$ , while in case (ii) (related animals), PEV( $\mathbf{x}$ ) = 0.83 and  $\mathbf{x}'\mathbf{A}\mathbf{x} = 1$ . The decrease of PEV( $\mathbf{x}$ ) did not compensate for the loss of genetic variability, and CD( $\mathbf{x}$ ) went from 0.25 (case (i)) to 0.17 (case (ii)).

## OVERALL INDICES

The best model was different according to the contrasts; when CD was used, we chose case (ii) for considering the contrasts  $u_1 - u_2$  and  $u_2 - u_3$ , but case (i) was the best for the contrast  $u_1 - u_3$ . It could be interesting to extend these procedures, defined here for a specific contrast, to a global measure of precision of an evaluation. An overall criterion could be useful when optimizing a design or comparing the precisions of different evaluations. Such overall criteria are derived on the basis of the means of quadratic ratios. As shown in *Appendix 2*, the ratio of the quadratic forms  $\mathbf{x}'\mathbf{B}\mathbf{x}/\mathbf{x}'\mathbf{C}\mathbf{x}$  is related to the generalized eigenvalue problem  $[\mathbf{B} - \mu_i\mathbf{C}]\mathbf{c}_i = 0$ , and two global means of these ratios of quadratic forms are the geometric and the arithmetic means of the corresponding eigenvalues  $\mu_i$ .

### Overall connectedness index

The ratio of quadratic forms here is  $\mathbf{x}'\mathbf{C}_r^{\mathbf{uu}}\mathbf{x}/\mathbf{x}'\mathbf{C}^{\mathbf{uu}}\mathbf{x}$ . The overall index suggested by Foulley et al (1992) is the geometric mean of the eigenvalues of  $[\mathbf{C}_r^{\mathbf{uu}} - \mu_i\mathbf{C}^{\mathbf{uu}}]\mathbf{c}_i = 0$ , or

$$\text{IC} = \left[ \frac{\det(\mathbf{C}_r^{\mathbf{uu}})}{\det(\mathbf{C}^{\mathbf{uu}})} \right]^{\frac{1}{n}} \quad [6]$$

This index is suggested, using the Kullback information (Kullback, 1983) between the joint density of the maximum likelihood estimator of  $\mathbf{b}$  and  $\mathbf{u} - \hat{\mathbf{u}}$  and the product of their marginal densities that would prevail if the design were orthonormal in  $\mathbf{b}$  and  $\mathbf{u}$ . All the indices of connectedness (IC and  $\text{IC}(\mathbf{x})$ ) are strictly positive and  $\leq 1$ . The null value never occurs when dealing with random factors, because the random effects are always estimable and the rank of both matrices equals  $n$  (eg, Foulley et al, 1990). An  $\text{IC}(\mathbf{x})$  equal to 1 demonstrates that  $\mathbf{x}'(\mathbf{u} - \hat{\mathbf{u}})$  is orthogonal to the fixed effects and, for the global IC, that  $\mathbf{u} - \hat{\mathbf{u}}$  is orthogonal to the fixed effects.

Application of the overall connectedness index among sires in a reference sire system based on planned artificial inseminations with link bulls has already been undertaken in France (Foulley et al, 1990; Hanocq et al, 1992; Laloë et al, 1992).

### Criteria of precision

Here, we devote our attention to the CDs of the contrasts between genetic values, which could be summarized in the  $(n - 1)$  greatest eigenvalues  $\mu_i$  of the generalized eigenvalue problem (Laloë, 1993):

$$[(\mathbf{A} - \lambda\mathbf{C}^{\mathbf{uu}}) - \mu_i\mathbf{A}]\mathbf{c}_i = \mathbf{0} \quad [7]$$

Some properties of the solutions, written in ascending order, are briefly given here. The  $\mu_i$ s are located between 0 and 1:  $\mu_2 \leq \text{CD}(\mathbf{x}) \leq \mu_n$ ;  $\mu_1$  is always null, and the associated eigenvector  $\mathbf{c}_1$  is proportional to  $\mathbf{A}^{-1}\mathbf{1}$ ; the other eigenvectors correspond to contrasts, since (cf, *Appendix 2* [A2.12]):  $\mathbf{c}'_1\mathbf{A}\mathbf{c}_i = 0$  for  $i > 1 \Leftrightarrow \mathbf{1}'\mathbf{A}^{-1}\mathbf{A}\mathbf{c}_i = 0 = \mathbf{1}'\mathbf{c}_i$ , ie, the definition of a contrast;  $\text{CD}(\mathbf{c}_i) = \mu_i$ .

Eigenvalues and eigenvectors for case (ii) are reported in table II. It could be verified that eigenvectors corresponding to a null eigenvalue are respectively  $\mathbf{c}_1$ , proportional to  $\mathbf{A}^{-1}\mathbf{1}$ , and  $\mathbf{c}_2$ , which corresponds to the genetic level comparison of the units. The other eigenvectors correspond to contrasts. Moreover, any contrast  $\mathbf{x}'\mathbf{u}$  can be written as a linear combination of the  $\mathbf{c}_i$ s ( $i$  ranging from 2 to  $n$ ) (cf, *Appendix 2* [A2.15]).

$$\mathbf{x}'\mathbf{u} = \left[ \sum_{i=2}^n \mathbf{a}_i\mathbf{c}_i \right]' \mathbf{u} \quad [8]$$

**Table II.** Eigenvectors and eigenvalues for case (ii).

<i>i</i>	1	2	3	4
$\mu_i$	0	0	0.333	0.6
$\mathbf{c}_i$	0.408	0.707	0.707	0.408
	0.408	0.707	-0.707	-0.408
	0.408	-0.707	-0.707	0.408
	0.408	-0.707	0.707	-0.408

From *Appendix 2* [A2.6], the CD of any contrast is a weighted mean of the eigenvalues of [7]:

$$CD(\mathbf{x}) = \frac{\sum_{i=1}^n a_i^2 \mu_i}{\sum_{i=1}^n a_i^2} \quad [9]$$

Two overall indices of precision can be computed:

$$\rho_1 = \sum_{i=2}^n \frac{\mu_i}{n-1} \quad [10]$$

$$\rho_2 = \left[ \prod_{i=2}^n \mu_i \right]^{\frac{1}{n-1}} \quad [11]$$

These criteria have been used to validate the rule of publication of French beef bull genetic values from field data evaluation (Laloë and Ménessier, 1995).

### PEV

Kennedy and Trus (1993) did not suggest any overall criterion of precision. By analogy, use of  $\det(\mathbf{C}^{\text{uu}})^{1/n}$  is suggested.

The values of the different criteria are reported in table III. Null values of  $\rho_2$  showed that both designs were disconnected.  $\rho_1$  was the same for both cases, as IC and  $\det(\mathbf{C}^{\text{uu}})^{1/n}$  favored the design where animals are related.

**Table III.** Values of the various criteria.

Case	$\rho_1$	$\rho_2$	IC	$\det(\mathbf{C}^{\text{uu}})^{1/n}$
(i)	0.333	0	0.841	0.707
(ii)	0.333	0	0.904	0.622

## CONCEPT OF (DIS)CONNECTEDNESS AND RANDOMNESS OF GENETIC EFFECTS

Disconnectedness, as defined in the linear fixed model context ( $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$ ) (use of a generalized inverse of  $\mathbf{X}'\mathbf{X}\sigma_e^2$  as the variance matrix of BLUE ( $\mathbf{b}$ ) –  $\mathbf{b}$ , occurrence of non-estimable contrasts, ‘all or none’ characteristic), never occurs when dealing with a random factor.  $\text{Var}(\hat{\mathbf{u}} - \mathbf{u}) = \mathbf{C}^{\text{uu}}\sigma_e^2$  is always positive definite. However,  $\lambda\mathbf{C}^{\text{uu}}$  is upwardly bound by  $\mathbf{A}$ , in the sense that, whatever  $\mathbf{x}$ ,  $\lambda\mathbf{x}'\mathbf{C}^{\text{uu}}\mathbf{x} \leq \mathbf{x}'\mathbf{A}\mathbf{x}$ .

If the PEV of a contrast  $\mathbf{x}'\mathbf{u}$  reaches the upper bound  $\mathbf{x}'\mathbf{A}\mathbf{x}$ ,  $\text{CD}(\mathbf{x}) = 0$  and:

$$\mathbf{x}'(\mathbf{A} - \lambda\mathbf{C}^{\text{uu}})\mathbf{x} = 0 \quad [12]$$

$$\mathbf{x}'\hat{\mathbf{u}} = 0 \text{ with probability } 1 \quad [13]$$

Equation [13] implies that  $\mathbf{x}'\hat{\mathbf{u}}$  does not follow a normal distribution, but a point-mass distribution at 0:  $P(\mathbf{x}'\hat{\mathbf{u}} = 0) = 1$ . In that sense, disconnectedness for a random factor is an ‘all or none’ characteristic concerning the distribution of the predictors in the same way as for a fixed factor. If a fixed factor is disconnected, ie, if a contrast between its levels is not estimable, then the CD of a contrast between its levels is null when it is treated as random. Thus the following definition of disconnectedness for random factors is proposed: a random factor is disconnected when at least one contrast between its levels has a null CD. With this definition, the status of a factor with respect to connectedness does not depend on the fixed or random nature of this factor. Connectedness leads to the same consequences in terms of the decrease of a matrix rank or probability laws in both random and fixed cases. Because IC and PEV deal with  $\mathbf{C}^{\text{uu}}$  instead of  $\mathbf{A} - \lambda\mathbf{C}^{\text{uu}}$ , they cannot exhibit this kind of disconnectedness for a random factor. As shown below, IC is devoted to the orthogonality between random and fixed factors and can detect perfectly connected contrasts or designs, but not disconnected ones.

## BOUNDARIES AND RELATIVE EVOLUTION OF CRITERIA

### *Lower boundary of the index of connectedness*

Since  $\mathbf{C}^{\text{uu}}$  is positive definite,  $\text{IC}(\mathbf{x})$  is never null and the index of connectedness never reaches the null value. It is interesting to characterize the lower boundary of this index, and how it varies.

Consider a contrast  $\mathbf{x}'\mathbf{u}$ , and denote the generalized coefficient of determination of  $\mathbf{x}'\mathbf{u}$  obtained with model [2] as  $\text{CD}_r(\mathbf{x})$ .  $\text{CD}_r(\mathbf{x})$  can be considered as the amount of information provided by data, independent of the design. A formula relating  $\text{IC}(\mathbf{x})$ ,  $\text{CD}(\mathbf{x})$  and  $\text{CD}_r(\mathbf{x})$  could be derived from [4] and [5]:

$$\text{IC}(\mathbf{x}) = \frac{1 - \text{CD}_r(\mathbf{x})}{1 - \text{CD}(\mathbf{x})} \quad [14]$$

$\text{IC}(\mathbf{x})$  has a minimal value when  $\mathbf{x}$  is disconnected in the complete model [1] ( $\text{CD}(\mathbf{x}) = 0$ ) and is equal to  $1 - \text{CD}_r(\mathbf{x})$ , by applying [14]. Thus, the index of

connectedness of a disconnected contrast increases as the amount of data decreases, contrary to the assumption of IC accounting only for the design.

The connectedness index of a contrast  $\mathbf{x}'\mathbf{u}$  is then located in the interval  $[1 - CD_r(\mathbf{x}), 1]$ . Particularly, when  $CD_r(\mathbf{x}) = 0$ ,  $IC(\mathbf{x}) = 1$ . This case occurs, for instance, when considering a contrast between a sire and a dam known only by their common progeny. Their predicted genetic values will always be equal whatever the performances. Thus, the question of whether there is any assortative mating cannot be answered.  $IC(\mathbf{x})$ , however, is always equal to 1 and these animals would be declared as perfectly connected and then comparable.

The same kind of result can be found again when working with a design as a whole; consider a nested, balanced 'herd/sire' model, with  $t$  progeny per sire,  $h$  herds and  $n$  different sires per herd. This design is clearly disconnected.

Some values of  $\rho_1$  and IC in relation to  $t$  are indicated in table IV, where  $h$  and  $n$  are equal to 5 and 2, respectively. Heritability equals 0.2. Though all these designs are disconnected, IC varies from 0.980 ( $t = 1$ ) to 0 ( $t = \infty$ ). The greater the amount of data, the lower IC. The design where  $t = 1$  seemed to be very well connected, the index of connectedness can not exhibit any disconnectedness and favors designs with low precision. The variation of this index for similar disconnected situations makes it unreliable for use.

**Table IV.** Values of IC and  $\rho_1$  in relation to  $t$ .<sup>a</sup>

$t$	$\rho_1$	IC
1	0.028	0.980
5	0.116	0.911
10	0.192	0.844
30	0.340	0.685
50	0.403	0.597
100	0.467	0.480
$\infty$	0.556	0

<sup>a</sup> $n = 2$ ,  $h = 5$ ; heritability equals 0.2.

Another index of connectedness is proposed, in order to study the causes of low precision of an evaluation. This low precision could be caused by a lack of information provided by the data or the design structure. It would be interesting to determine the main cause of this low precision. This would allow the precisions obtained in both reduced and complete models to be compared, on the basis of the matrices  $\mathbf{A} - \mathbf{C}^{uu}$  and  $\mathbf{A} - \mathbf{C}_r^{uu}$  in order to avoid the above-described drawback of IC. This new index is denoted  $\phi(\mathbf{x})$  for a contrast  $\mathbf{x}'\mathbf{u}$  and is equal to  $CD(\mathbf{x})/CD_r(\mathbf{x})$  or to the ratio of quadratic forms  $\mathbf{x}'(\mathbf{A} - \mathbf{C}^{uu})\mathbf{x}/\mathbf{x}'(\mathbf{A} - \mathbf{C}_r^{uu})\mathbf{x}$ .  $\phi(\mathbf{x})$  is located between 0 (disconnectedness) and 1 (no impact of the fixed effects), whatever  $CD_r(\mathbf{x})$ . The overall indices of connectedness are:

$$\phi_1 = \rho_1/\rho_{1r} \quad [15]$$

$$\phi_2 = \rho_2/\rho_{2r} \quad [16]$$

where  $\rho_{1r}$  and  $\rho_{2r}$  are the overall criteria of precision  $\rho_1$  and  $\rho_2$  obtained with the reduced model, respectively.

In the above sire model example,  $\phi_2 = 0$ , revealing again that the design is disconnected. It can be shown in this example that  $\phi_1 = (n-1)h/(nh-1)$ , ie, the proportion of connected contrasts among all the contrasts. It does not depend on the heritability or the amount of information provided by the data, ie, the number of progeny per sire. For the situations reported in table IV, the values of  $\phi_1$  and  $\phi_2$  are constant, and equal to 0.556 and 0, respectively, as the value of IC varies from 0 to 0.980.

These new indices obviously have the same limitations as the original one (they only take into account the impact of the fixed effects, orthogonality is favored) and can not be the only criterion used to judge a design. They could be used, however, to see if a low value of a CD is caused by a small amount of data or by a poor design, and also to evaluate the global loss of information due to the design.

### ***Upper boundary of the index of connectedness: complete connectedness***

Consider a completely connected design, ie, one whose overall index of connectedness is 1. Then, for any  $\mathbf{x}$ ,  $\mathbf{x}'\lambda\mathbf{C}_r^{\mathbf{uu}}\mathbf{x} = \mathbf{x}'\lambda\mathbf{C}^{\mathbf{uu}}\mathbf{x}$ . Since both matrices are positive definite,  $\mathbf{C}_r^{\mathbf{uu}} = \mathbf{C}^{\mathbf{uu}}$  and, consequently,  $\mathbf{Z}'\mathbf{M}\mathbf{Z} = \mathbf{Z}'\mathbf{M}_r\mathbf{Z}$ . It can be seen that the condition of complete connectedness is independent of the relationship matrix. This equality characterizes a design where, in a fixed effects model context,  $\mathbf{u}$  is orthogonal to all other effects (except the mean). This kind of orthogonal design must be complete with proportional frequencies (Coursol, 1980; Mukhopadhyay, 1983). All the levels of the random factor must then be identically distributed among all levels of all the fixed factors. For instance, for a sire model, the following equality must be satisfied for any sire and any level of factors included in the model:

$$n_{ij(k)} = \frac{n_{i0}n_{0j(k)}}{n_{00}} \quad [17]$$

where  $n_{00}$  is the total number of progeny,  $n_{i0}$  the sire  $i$  number,  $n_{0j(k)}$  the number of the level  $j$  of the  $k$ th fixed factor, and  $n_{ij(k)}$  the sire  $i$  number in the level  $j$  of the  $k$ th fixed factor.

### ***Boundaries of the criteria of precision***

The CD of a contrast is the square of correlation between  $\mathbf{x}'\mathbf{u}$  and  $\mathbf{x}'\hat{\mathbf{u}}$ , which varies between 0 and 1. A value of zero indicates that the data does not provide any information about the comparison:  $\text{var}(\mathbf{x}'\mathbf{u}|\hat{\mathbf{u}}) = \text{var}(\mathbf{x}'\mathbf{u})$ . The contrast between genetic values cannot be predicted, and there is a disconnectedness, according to Laloë definitions (1993). A value of 1 (which is never reached) would indicate that the correlation between predicted and exact values was equal to 1, or that no more information could be obtained from the data.

### ***PEV***

IC and CD measure the discrepancy between the real situation and a reference situation. The values of the index of connectedness and of the criteria of precision

are located between 0 and 1. The theoretical interpretation of these values is that the nearer a value is to 1, the better the situation would be. An IC of a contrast equal to 1 demonstrates that there is no influence of the fixed effects on the prediction of this contrast; a CD is the squared correlation between the predictor and the real value; these values are interpretable. However, a value of a PEV alone cannot be interpreted in itself. It must be compared with values of the same contrast in other situations, or with other contrasts. For instance, in case (ii) of the theoretical animal model example where the PEV between individual units was 2, this must be compared to the value of the same PEV in case (i) ( $PEV = 4$ ), or the covariance between units must be considered.

## AN OPTIMIZATION PROBLEM

Consider a model including the fixed effect 'herd' and a random effect 'sire'. The number of observations  $N$  is the same per herd (here,  $N = 60$ ). There are two natural service intraherd sires ( $t$  observations per sire) and a reference sire ( $m$  observations per herd and sire) used in each herd, as shown in table V;  $N = 2t + m$ . The sires are not related and heritability equals 0.2. The problem is how to choose  $m$  and  $t$  in order to obtain the most precise genetic values of the ten natural service sires. In that context, where animals are unrelated, PEV and CD are equivalent. If normed contrasts  $\mathbf{x}'\mathbf{u}$  (such as  $\mathbf{x}'\mathbf{x} = 1$ ) are considered, without loss of generality, the following results:

$$CD(\mathbf{x}) = 1 - \frac{PEV(\mathbf{x})}{\sigma_a^2} \quad [18]$$

An increase of CD then corresponds to a decrease of PEV, and the use of both methods leads to the same results. For this reason, we used IC and CD. IC,  $\phi_1$ ,  $\phi_2$ ,  $\rho_1$  and  $\rho_2$  were computed for the set of the ten natural service sires, and IC( $\mathbf{x}$ ),  $\phi(\mathbf{x})$  and CD( $\mathbf{x}$ ) were computed for a contrast between genetic levels of two herds, and with respect to different values of  $t$  and  $m$ . These results are given in table VI.

Table V. Data structure.

Herd	Sire											
	1	2	3	4	5	6	7	8	9	10	11	
A	$t$	$t$										$m$
B			$t$	$t$								$m$
C					$t$	$t$						$m$
D							$t$	$t$				$m$
E									$t$	$t$		$m$

### Criteria of connectedness

IC,  $\phi_1$  and IC( $\mathbf{x}$ ) increase with  $m$ , starting from strictly positive values and reaching their maximum value near 1 when  $m = 58$  (table VI).  $\phi_2$  and  $\phi(\mathbf{x})$  also increase

**Table VI.** Indices of connection and precision with respect to different values of  $t$  and  $m$  for the set ten natural services and for a contrast between genetic levels of two herds.<sup>a</sup>

$t$	$m$	$IC$	$\phi_1$	$\phi_2$	$\rho_1$	$\rho_2$	$IC(\mathbf{x})$	$\phi(\mathbf{x})$	$CD(\mathbf{x})$
30	0	0.685	0.556	0	0.340	0	0.388	0	0
29	2	0.704	0.591	0.326	0.357	0.197	0.416	0.080	0.058
28	4	0.723	0.622	0.431	0.371	0.256	0.444	0.150	0.089
25	10	0.774	0.696	0.600	0.396	0.341	0.527	0.317	0.180
22	16	0.819	0.751	0.694	0.403	0.372	0.607	0.440	0.236
20	20	0.846	0.781	0.739	0.400	0.379	0.658	0.506	0.260
15	30	0.905	0.841	0.821	0.371	0.362	0.779	0.641	0.283
10	40	0.952	0.890	0.882	0.307	0.304	0.885	0.753	0.260
5	50	0.986	0.939	0.937	0.196	0.195	0.986	0.863	0.180
2	56	0.997	0.973	0.973	0.093	0.093	0.994	0.939	0.089
1	58	0.999	0.986	0.986	0.049	0.049	0.998	0.968	0.049

<sup>a</sup>  $IC$ ,  $CD$ ,  $\phi_1$ ,  $\phi_2$ ,  $\rho_1$ ,  $\rho_2$ ,  $IC(\mathbf{x})$  and  $CD(\mathbf{x})$  are defined in text.

with  $m$ , but start from 0 when  $m = 0$ , exhibiting a disconnectedness, to a maximum value near 1 when  $m = 58$ . All these criteria favor the less incomplete design, which is also the design where the natural service sires have only one progeny. Whatever the criteria used, studying only the structure of the design was insufficient to judge the precision of an evaluation or to optimize a design.

### Criteria of precision

Criteria of precision range from 0 when  $m = 0$ , exhibiting a disconnectedness, to several maxima ( $m = 20$  for  $\rho_2$ ,  $m = 16$  for  $\rho_1$  and  $m = 30$  for  $CD(\mathbf{x})$ ). It was not surprising that the maxima were different depending to the criteria because  $\rho_2$  is more sensitive to a poor connectedness than  $\rho_1$ , and reached its maximum value for a more connected design than  $\rho_1$ . The contrast of genetic levels between herds was the less connected one, and it was most precise for a greater value of  $m$ . The values of the criteria then decreased; the enhancement of connectedness no longer compensated for the loss of information provided by the data. Unlike the indices of connectedness, the use of criteria of precision led to optima that were compromises between information from the data and the structure of the design.

Consider the contrast between genetic levels of two different herds,  $CD(\mathbf{x}) = 0.180$  in two cases:

- for  $t = 25$ ,  $\phi(\mathbf{x}) = 0.317$  ( $IC(\mathbf{x}) = 0.527$ ), with a poor level of connectedness, about two-thirds of the information is lost, due to the design structure:
- for  $t = 5$ ,  $\phi(\mathbf{x}) = 0.863$  ( $IC(\mathbf{x}) = 0.986$ ), the restrictive factor here is the amount of information that can be obtained from the data.

This conclusion is obvious without using these criteria on simple designs, but the interpretation of the indices needs to be as clear as possible when dealing with more complicated ones.

## LINKS BETWEEN IC, PEV, CD AND EXPECTED GENETIC PROGRESS

### *Maximization of IC and genetic progress*

Hanocq et al (1966) showed in a simulation study that a high level of connectedness only slightly increases the genetic trend. In the extreme, if the factor 'year' is included in the model and the corresponding design is completely connected ( $IC = 1$ ), all the sires must be used the same way in all the years (equation [17]). Such a design surely cannot lead to any genetic progress, since animals born in different years would be bred from exactly the same sires in the same proportions.

### *Behavior of PEV and CD on a hypothetical animal model where animals are equally related*

As noted before, CD and IC are equivalent when dealing with contrasts involving unrelated animals, but they account for relationships differently. It would be interesting to see what the differences are when one method is compared to the other, particularly with respect to the genetic progress. Indeed, Kennedy and Trus (1993) wrote "... minimization of PEV does not necessarily maximize rate of genetic improvement because it may come at a cost of reduced selection intensity associated with selection among related as opposed to unrelated individuals". We will use a hypothetical and unrealistic model to study the behavior of both indices according to the relationships between animals.

For a 'mean + animal' model, where the animals are equicorrelated with a relationship coefficient  $r$ , and the number  $n$  of animals is large, we have (cf, *Appendix 3*)

$$PEV(\mathbf{x}) \propto (1 - r) / \{1 - r + \lambda\}, \text{ for any contrast } \mathbf{x}'\mathbf{u} \quad [19]$$

$$\rho_1 = \rho_2 = (1 - r) / \{1 - r + \lambda\} \quad [20]$$

Here,  $PEV(\mathbf{x})$  and  $\rho_1$  vary in exactly the same way according to  $r$ . To optimize the design with PEV (minimization of  $PEV(\mathbf{x})$ ) or with CD (maximization of  $\rho_1$ ) leads to a maximal  $r$  or a null  $r$ , respectively.

The expression of the expected genetic progress is (cf, *Appendix 3*):

$$R_{p,r} = i_{p,r} \rho_1^{0.5} \sigma_a \quad [21]$$

where  $i_{p,r} = i_p(1 - r)^{0.5}$  can be viewed as the reduced selection intensity associated with selection among related animals (Kennedy and Trus, 1993), and  $\rho_1$  is the global criterion of precision. This expression is similar to the expression of the expected genetic progress in the case of a classical selection index and made on a large population of unrelated animals:

$$R_p = i_p CD^{0.5} \sigma_a \quad [22]$$

where  $i_p$  is the selection intensity and CD the coefficient of determination of the animal selection index.  $\rho_1$  plays the same role in [21] as CD in [22]. The increase of

$r$  induces a decrease of  $R$ , initially because of the decrease in the selection intensity, as noted by Kennedy and Trus (1993), and secondly because of a decrease in the precision  $\rho_1$ . At the same time, the PEV decreases. In this situation, PEV and genetic progress are in conflict.

## CONCLUSION

Methods PEV and CD answer different questions. If the predicted value of a contrast is null, PEV allows the appreciation of the likelihood of this result. The probability that  $\mathbf{x}'\mathbf{u}|_{\mathbf{x}'\hat{\mathbf{u}}=0}$  will be near 0 increases as  $\text{PEV}(\mathbf{x})$  decreases, because  $\mathbf{x}'\mathbf{u}|_{\mathbf{x}'\hat{\mathbf{u}}=0} \sim N(0, \text{PEV}(\mathbf{x}))$ . The CD permits the determination of whether the predicted value will be different from 0. In general terms, the probability that  $\mathbf{x}'\hat{\mathbf{u}}$  will be different from 0 increases with  $\text{CD}(\mathbf{x})$ , because  $\mathbf{x}'\hat{\mathbf{u}} \sim N(0, \text{CD}(\mathbf{x})\mathbf{x}'\mathbf{A}\mathbf{x}\sigma_a^2)$ . PEV is more related to the likelihood of the hypothesis 'all the animals are equal', and CD could be linked to the power of the test 'are the animals different?'. This distinction is very important, since the main aim of genetic evaluation is to discriminate between animals on the basis of their predicted genetic values, in order to select the best ones.

While both methods are equivalent when animals are unrelated, they can, however, be in conflict in other situations. Genetic relationships decrease the PEV, and also decrease the selection intensity and the genetic variability. PEV is minimized when  $\text{var}(\mathbf{x}'\mathbf{u}|_{\mathbf{x}'\hat{\mathbf{u}}})$  is a minimum, and CD is minimized when  $\text{var}(\mathbf{x}\mathbf{u}|_{\mathbf{x}'\hat{\mathbf{u}}})/\text{var}(\mathbf{x}'\mathbf{u})$  is a minimum. PEV then favors contrasts between related animals, where  $\text{var}(\mathbf{x}'\mathbf{u}|_{\mathbf{x}'\hat{\mathbf{u}}})$  is small, as CD accounts for the decrease of  $\text{var}(\mathbf{x}'\mathbf{u})$ . CD combines both aspects, genetic variability and PEV, and is therefore more related to genetic progress, as shown in the theoretical example in the previous section.

The problem of (dis)connectedness is formulated differently according to a priori knowledge about differences between the evaluated populations or genetic levels of management units. First, if the differences are known or supposed to be, can they be exhibited in the evaluation? This question can be answered by CD. Second, a priori, there are no differences. Disconnectedness is then only a source of a decrease of precision, and its study has no inherent interest. Its study may permit the choice of a strategy for precision increase, either by connectedness increase or by an increase of the amount of information provided by data. IC is not very appropriate to this kind of study, mainly because it does not always exhibit disconnectedness and because it decreases with the amount of information obtained from data. Large values of this index could be due either to a good connectedness or to poor information. Another index, devoted to the design structure and independent of the information obtained from the data, was suggested to minimize this drawback. To look only at the data structure is not sufficient. An orthogonal design could not lead to any genetic progress. A genetic evaluation must be precise and discriminatory. CD, which combines data structure and amount of information and also accounts for both PEV and genetic variability, is a good method to select for judging the precision of a genetic evaluation or to optimize corresponding designs.

## REFERENCES

- Coursol J (1980) *Techniques statistiques des modèles linéaires. I. Aspects théoriques*. CIMPA/ICPAM, Nice
- 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) Colloque 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, Heidelberg, 302-337
- Foulley JL, 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
- Gomez-Raya L (1992) Prediction of genetic progress with different accuracies among selection candidates. *J Anim Breed Genet* 109, 347-357
- Hanocq E, Foulley JL, Boichard D (1992) Measuring connectedness in genetic evaluation with an application to Limousin and Maine-Anjou sires. In: *43rd Ann Meet EAAP. Madrid, Spain, Sept 13-17 1992*, Ministerio de Agricultura, Pesca y Alimentacion, Spain, 242 (Abstr)
- Hanocq E, Foulley JL, Boichard D (1996) A simulation study of the effect of connectedness on genetic trend. *Genet Sel Evol* 28, 67-82
- Kennedy BW, Trus D (1993) Considerations on genetic connectedness between management units under an animal model. *J Anim Sci* 71, 2341-2352
- 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
- Laloë D (1993) Precision and information in linear models of genetic evaluation. *Genet Sel Evol* 25, 557-576
- Laloë D, Sapa J, Méniissier 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
- Laloë D, Méniissier F (1995) Validation of the rule of publication of French beef bulls genetic values from field data evaluation. In: *46rd Ann Meet EAAP, Prague, Czech Republic, Sept 4-8 1995*, Wageningen Pers, 28 (Abstr)
- Mukhopadhyay AC (1983) Orthogonal arrays and applications. In: *Encyclopedia of Statistical Sciences* (Kotz S, Johnson NL, eds), John Wiley and Sons, New York, vol 6, 523-527
- Phocas F, Colleau JJ (1995) Approximating selection differentials and variances for correlated selection indices. *Genet Sel Evol* 27, 551-565
- Owen DB, Steck GP (1962) Moments of order statistics from the equicorrelated multivariate normal distribution. *Ann Math Stat* 33, 1286-1291
- Petersen PH (1978) A test for connectedness fitted for the two-way BLUP sire evaluation. *Acta Agric Scand* 28, 360-362

## Appendix 1.

**Proof that**  $CD(\mathbf{x}) = 0 \Rightarrow \mathbf{x}'\hat{\mathbf{u}} = 0$

The CD of contrast  $\mathbf{x}$  is null if and only if the expectation under the distribution of  $\hat{\mathbf{u}}$  of the Kullback information between the respective distributions of  $\mathbf{x}'\mathbf{u}|_{\hat{\mathbf{u}}}$  and  $\mathbf{x}'\mathbf{u}$  is null (Laloë, 1993, formula 24):  $CD(\mathbf{x} = 0 \Leftrightarrow E_{\hat{\mathbf{u}}}(I[g(\mathbf{x}'\mathbf{u}|_{\hat{\mathbf{u}}}); f(\mathbf{x}'\mathbf{u})]) = 0$ .

$I [g(\mathbf{x}'\mathbf{u}|\hat{\mathbf{u}}):f(\mathbf{x}'\mathbf{u})]$  is a null or positive random variable (Kullback, 1983), and its expectation is null if and only if it takes only null values. Then, whatever  $\hat{\mathbf{u}}$ ,  $I [g(\mathbf{x}'\mathbf{u}|\hat{\mathbf{u}}):f(\mathbf{x}'\mathbf{u})] = 0$ . Then, the distributions of  $\mathbf{x}'\mathbf{u}|\hat{\mathbf{u}}$  and  $\mathbf{x}'\mathbf{u}$  are the same (Kullback, 1983). Notably, their expectations are equal:  $E(\mathbf{x}'\mathbf{u}|\hat{\mathbf{u}}) = E(\mathbf{x}'\mathbf{u}) = 0 = \mathbf{x}'\hat{\mathbf{u}}$ , whatever  $\hat{\mathbf{u}}$ . Then  $CD(\mathbf{x}) = 0 \Rightarrow \mathbf{x}'\hat{\mathbf{u}} = 0$ .

**Appendix 2.**

**Ratio of quadratic forms and generalized eigenvalue problem**

Let us consider a positive semi-definite matrix  $\mathbf{B}$  and a positive definite matrix  $\mathbf{C}$ . We are interested in the ratio of quadratic forms  $\mathbf{x}'\mathbf{B}\mathbf{x}/\mathbf{x}'\mathbf{C}\mathbf{x}$ , and we want to characterize this ratio in some manner.

Since  $\mathbf{C}$  is positive definite, a lower triangular and non-singular matrix  $\mathbf{L}$  exists such that  $\mathbf{C} = \mathbf{L}\mathbf{L}'$ . Hence  $\mathbf{x}'\mathbf{B}\mathbf{x}/\mathbf{x}'\mathbf{C}\mathbf{x} = \mathbf{x}'\mathbf{B}\mathbf{x}/\mathbf{x}'\mathbf{L}\mathbf{L}'\mathbf{x} = \mathbf{y}'\mathbf{L}^{-1}\mathbf{B}\mathbf{L}'^{-1}\mathbf{y}/\mathbf{y}'\mathbf{y}$  where  $\mathbf{y} = \mathbf{L}'\mathbf{x}$ . This ratio of quadratic forms is related to the standard eigenvalue problem:

$$\mathbf{L}^{-1}\mathbf{B}\mathbf{L}'^{-1}\mathbf{d}_i = \mu_i\mathbf{d}_i \tag{A2.1}$$

where  $\mu_i$  and  $\mathbf{d}_i$  are the eigenvalues and the eigenvectors of  $\mathbf{L}^{-1}\mathbf{B}\mathbf{L}'^{-1}$ , respectively. The following equations recall the properties of the eigenvectors:

$$\mathbf{d}'_i\mathbf{d}_j = \mathbf{d}'_i\mathbf{L}^{-1}\mathbf{B}\mathbf{L}'^{-1}\mathbf{d}_j = 0 \text{ for } i \text{ different from } j \tag{A2.2}$$

$$\mathbf{d}'_i\mathbf{L}^{-1}\mathbf{B}\mathbf{L}'^{-1}\mathbf{d}_i = \mu_i \tag{A2.3}$$

$$\mathbf{d}'_i\mathbf{d}_i = 1 \tag{A2.4}$$

**Eigenvalues and ratio of quadratic forms**

The vector  $\mathbf{y}$  can be written as a linear combination of the  $\mathbf{d}_i$ s:

$$\mathbf{y} = \sum_{i=1}^n a_i\mathbf{d}_i \tag{A2.5}$$

From the above properties of the eigenvectors, we get:

$$\mathbf{x}'\mathbf{B}\mathbf{x}/\mathbf{x}'\mathbf{C}\mathbf{x} = \mathbf{y}'\mathbf{L}^{-1}\mathbf{B}\mathbf{L}'^{-1}\mathbf{y}/\mathbf{y}'\mathbf{y} = \frac{\sum_{i=1}^n a_i^2\mu_i}{\sum_{i=1}^n a_i^2} \tag{A2.6}$$

Thus, these ratios of quadratic forms are the weighted means of the eigenvalues of  $\mathbf{L}^{-1}\mathbf{B}\mathbf{L}'^{-1}$ . They are located in the interval  $[\mu_1, \dots, \mu_n]$  of the eigenvalues sorted

in ascending order of  $\mathbf{L}^{-1}\mathbf{B}\mathbf{L}'^{-1}$ . It seems natural to choose some means of the eigenvalues as the global means of the ratio of quadratic forms, eg, the arithmetic and the geometric means:

The geometric mean of the eigenvalues is:

$$\left[ \prod_{i=1}^n \mu_i \right]^{1/n} = \left[ \frac{\det(\mathbf{B})}{\det(\mathbf{C})} \right]^{1/n} \quad [\text{A2.7}]$$

The arithmetic mean of the eigenvalues is:

$$\frac{1}{n} \sum_{i=1}^n \mu_i = \frac{\text{tr}(\mathbf{C}^{-1}\mathbf{B})}{n} \quad [\text{A2.8}]$$

[A2.1] can be written as:

$$\mathbf{L}^{-1}\mathbf{B}\mathbf{c}_i = \mu_i \mathbf{L}'\mathbf{c}_i \quad [\text{A2.9}]$$

where  $\mathbf{c}_i = \mathbf{L}'^{-1}\mathbf{d}_i$  or  $\mathbf{d}_i = \mathbf{L}'\mathbf{c}_i$ , where  $\mathbf{c}_i$  is the eigenvector associated with  $\mu_i$ .

Then, by premultiplication of both sides of equation [A2.9] by  $\mathbf{L}$ , we get the so-called generalized eigenvalue problem:

$$\mathbf{B}\mathbf{c}_i = \mu_i \mathbf{C}\mathbf{c}_i \quad [\text{A2.10}]$$

or

$$[\mathbf{B} - \mu_i \mathbf{C}]\mathbf{c}_i = 0 \quad [\text{A2.11}]$$

### *Properties of the eigenvectors of [A2.11]*

From the properties [A2.2] to [A2.4] of the eigenvectors of the standard eigenvalue problem we deduce:

$$\mathbf{c}'_i \mathbf{C}\mathbf{c}_j = \mathbf{c}'_i \mathbf{B}\mathbf{c}_j = 0 \text{ for } i \text{ different from } j \quad [\text{A2.12}]$$

$$\mathbf{c}'_i \mathbf{B}\mathbf{c}_i = \mu_i \quad [\text{A2.13}]$$

$$\mathbf{c}'_i \mathbf{C}\mathbf{c}_i = 1 \quad [\text{A2.14}]$$

Finally, if  $\mathbf{y} = \sum_{i=1}^n a_i \mathbf{d}_i$ , then  $\mathbf{x} = \mathbf{L}'^{-1}\mathbf{y} = \sum_{i=1}^n a_i \mathbf{L}'^{-1}\mathbf{d}_i = \sum_{i=1}^n a_i \mathbf{c}_i \quad [\text{A2.15}]$

### **Appendix 3.**

#### **Expected genetic response from a BLUP evaluation on a 'mean + animal' model with equicorrelated relationship matrix**

The general expression of the expected response to selection is (eg, Gomez-Raya, 1992):

$$R_p = E_{\mathbf{u}} \left\{ E(u_i | \hat{u}_{i,s_i,p} = 1) \right\}$$

where  $p$  is the proportion of animals selected, and  $s_{i,p} = 1$  if  $i$  is among the selected animals.  $E(u_i|\hat{u}_i)$  being equal to  $\hat{u}_i$ , we have:

$$R_p = E_{\hat{\mathbf{u}}}(u_i | s_{i,p} = 1) \tag{A3.1}$$

If the animals are unrelated and if the selection is based on a classical selection index with the same information per animal (mass selection, progeny selection with the same number of progeny per animal), the  $u_i$  are normally, independently and identically distributed:  $\hat{\mathbf{u}} \sim N(\mathbf{0}, \mathbf{I}\sigma_u^2)$  or  $\hat{\mathbf{u}} \sim N(\mathbf{0}, \mathbf{ICD}\sigma_a^2)$ ; and we get the well-known formula  $R_p = i_p\sigma_{\hat{\mathbf{u}}} = i_p\text{CD}^{0.5}\sigma_a$ , where  $i_p$  is the selection intensity and CD the coefficient of determination of the animal selection index.

We determine the expected response to selection in a simple random animal model, where the animals are linked by the same relationship, in order to see the impact of this relationship on the expected genetic progress, the criteria of precision and the PEV.

First, we will report here some properties of special patterned matrices that will be useful in the following. Let us denote  $\mathbf{K}_{a,b,n}$  the matrices of order  $n$  such that  $\mathbf{K}_{a,b,n}(i, i) = a$  whatever  $i$  and  $\mathbf{K}_{a,b,n}(i, j) = b$  for  $i$  different from  $j$ :  $\mathbf{K}_{a,b,n} = (a - b)\mathbf{I}_n + b\mathbf{J}_n$ .  $K_n$  denotes the set of the positive semi-definite matrices  $\mathbf{K}_{a,b,n}$ . Then, if  $\mathbf{K}_{a,b,n}$  and  $\mathbf{K}_{c,d,n}$  belong to  $K_n$ :

- $\mathbf{K}_{a,b,n} + \mathbf{K}_{c,d,n}$  belongs to  $K_n$
- $\mathbf{K}_{a,b,n}\mathbf{K}_{c,d,n}$  belongs to  $K_n$
- $r\mathbf{K}_{a,b,n}$  belongs to  $K_n$ , if  $r > 0$
- if  $\mathbf{K}_{a,b,n}^{-1}$  exists, it belongs to  $K_n$
- $\mathbf{K}_{a,b,n}$  has two eigenvalues (eg, Laloë, 1993):

$$\left. \begin{aligned} \mu_1 &= a + (n - 1)b \\ \mu_2 &= a - b \end{aligned} \right\} \tag{A3.2}$$

The multiplicity of  $\mu_1$  is 1, and the corresponding eigenvector is proportional to  $\mathbf{1}$ ; the multiplicity of  $\mu_2$  is  $n - 1$ , and the corresponding eigenvectors  $\mathbf{c}$ 's are contrasts ( $\mathbf{c}'\mathbf{1} = 0$ ).

Equivalently, we have:

$$\left. \begin{aligned} a &= \frac{\mu_1 + (n - 1)\mu_2}{n} \\ b &= \frac{\mu_1 - \mu_2}{n} \end{aligned} \right\} \tag{A3.3}$$

The eigenvalues of the product (or the sum) of two matrices belonging to  $K_n$  are the product (or the sum) of the homologous eigenvalues of both matrices. Moreover, the eigenvalues of the inverse of a positive definite matrix are the inverse of the eigenvalues of this matrix.

Because of all these properties, working with the eigenvalues of this kind of matrices greatly simplifies the algebra.

The model is  $\mathbf{y} = \mathbf{1}m + \mathbf{Z}\mathbf{u} + \mathbf{e}$ , where the  $u$  are equicorrelated, with a correlation  $r$ .  $n$  animals are included in the evaluation and are recorded.  $\mathbf{Z} = \mathbf{I}_n$ . The variance

matrix of  $\mathbf{u}$  is equal to  $\mathbf{A}\sigma_a^2 \mathbf{K}_{1,r,n}\sigma_a^2$  and

$$\mathbf{Z}'\mathbf{M}\mathbf{Z} = \mathbf{K}_{\frac{n-1}{n}, \frac{-1}{n}, n}$$

The matrix variance of  $\hat{\mathbf{u}}, (\mathbf{A} - \lambda\mathbf{C}^{\mathbf{uu}})\sigma_a^2$ , must be expressed in order to get the parameters used in [A3.1]. On the other hand, we need the eigenvalues of  $\mathbf{L}^{-1}(\mathbf{A} - \lambda\mathbf{C}^{\mathbf{uu}})\mathbf{L}^{-1}$ , where  $\mathbf{A} = \mathbf{L}\mathbf{L}'$  in order to get the precision criteria. These eigenvalues are also the eigenvalues of  $\mathbf{L}'^{-1}\mathbf{L}^{-1}(\mathbf{A} - \lambda\mathbf{C}^{\mathbf{uu}})$ , ie,  $\mathbf{A}^{-1}(\mathbf{A} - \lambda\mathbf{C}^{\mathbf{uu}})$ . Finally, the coefficients of the PEV matrix  $\mathbf{C}^{\mathbf{uu}}$  are required.

Matrices  $\mathbf{C}^{\mathbf{uu}}, (\mathbf{A} - \lambda\mathbf{C}^{\mathbf{uu}})\sigma_a^2$  and  $\mathbf{A}^{-1}(\mathbf{A} - \lambda\mathbf{C}^{\mathbf{uu}})$  belong to  $K_n$ , since they are simple functions of  $\mathbf{A}$  and  $\mathbf{Z}'\mathbf{M}\mathbf{Z}$ , which are matrices belonging to  $K_n$ . The calculation will be as follows:

- (i) to get the eigenvalues of  $\mathbf{Z}'\mathbf{M}\mathbf{Z}$  and  $\mathbf{A}$ , using [A3.2];
- (ii) to get the eigenvalues of all intermediate matrices, using the above properties;
- (iii) to get the eigenvalues of  $\mathbf{A}^{-1}(\mathbf{A} - \lambda\mathbf{C}^{\mathbf{uu}})$ , and the coefficients of  $(\mathbf{A} - \lambda\mathbf{C}^{\mathbf{uu}})\sigma_a^2$ , using [A3.3].

Eigenvalues of all these matrices are reported in table A1.

**Table A1.** Eigenvalues of matrices.

<i>Matrix</i>	$\mu_1$	$\mu_2$
$\mathbf{A}$	$1 + (n - 1)r$	$1 - r$
$\mathbf{A}^{-1}$	$\frac{1}{1 + (n - 1)r}$	$\frac{1}{1 - r}$
$\mathbf{Z}'\mathbf{M}\mathbf{Z}$	0	1
$\mathbf{Z}'\mathbf{M}\mathbf{Z} + \lambda\mathbf{A}^{-1}$	$\frac{\lambda}{1 + (n - 1)r}$	$\frac{1 - r + \lambda}{1 - r}$
$\mathbf{C}^{\mathbf{uu}}$	$\frac{1 + (n - 1)r}{\lambda}$	$\frac{1 - r}{1 - r + \lambda}$
$\lambda\mathbf{C}^{\mathbf{uu}}$	$1 + (n - 1)r$	$\lambda \frac{1 - r}{1 - r + \lambda}$
$(\mathbf{A} - \lambda\mathbf{C}^{\mathbf{uu}})\sigma_a^2$	0	$\frac{(1 - r)^2}{1 - r + \lambda} \sigma_a^2$
$\mathbf{A}^{-1}(\mathbf{A} - \lambda\mathbf{C}^{\mathbf{uu}})$	0	$\frac{1 - r}{1 - r + \lambda}$

The space of contrasts between the  $n$  animals is  $(n - 1)$ -dimensional vectorial, which is spanned by the  $(n - 1)$  eigenvectors  $\mathbf{d}_2, \dots, \mathbf{d}_n$  corresponding to the second eigenvalue of  $\mathbf{C}^{\mathbf{uu}}$ . Any contrast  $\mathbf{y}'\mathbf{u}$  corresponds to a linear combination of these eigenvectors:

$$\mathbf{y} = \sum_{i=2}^n a_i \mathbf{d}_i$$

From [A2.6] we get:

$$\frac{\mathbf{y}'\mathbf{C}^{\mathbf{uu}}\mathbf{y}}{\mathbf{y}'\mathbf{y}} = \frac{\sum_{i=2}^n a_i^2 \mu_2}{\sum_{i=2}^n a_i^2} = \frac{\sum_{i=2}^n a_i^2}{\sum_{i=2}^n a_i^2} \mu_2 = \mu_2$$

The PEV of any contrast between animals is proportional to the second eigenvalue of  $\mathbf{C}^{\mathbf{uu}}$ :

$$\text{PEV}(\mathbf{y}) \propto (1-r)/\{1-r+\lambda\} \quad [\text{A3.4}]$$

The  $(n-1)$  greatest eigenvalues of  $\mathbf{A}^{-1}(\mathbf{A}-\lambda\mathbf{C}^{\mathbf{uu}})$  are all equal. Thus their geometric and arithmetic means are equal to this eigenvalue:

$$\rho_1 = \rho_2 = (1-r)/\{1-r+\lambda\} \quad [\text{A3.5}]$$

The following can be deduced from the eigenvalues of  $(\mathbf{A}-\lambda\mathbf{C}^{\mathbf{uu}})\sigma_a^2$  and of [A3.3]:

$$(\mathbf{A}-\lambda\mathbf{C}^{\mathbf{uu}})\sigma_a^2 = \mathbf{K}_{a,b,n}$$

where  $a = \{(n-1)/n\}(1-r)\rho_1\sigma_a^2$  and  $b = (-1/n)(1-r)\rho_1\sigma_a^2$ .

When  $n$  tends to infinity,  $a$  tends to  $(1-r)\rho_1\sigma_a^2$  and  $b$  tends to 0. The  $\hat{u}_i$ s become independently and identically distributed, with a variance equal to  $(1-r)\rho_1\sigma_a^2$ . Then:

$$R = i_p \sigma_{\hat{u}} = i_p \sqrt{1-r} \sqrt{\rho_1} \sigma_a$$

After the results of Owen and Steck (1962), as recently discussed by Phocas and Colleau (1995),  $i_p \sqrt{1-r}$  is the expectation of the upper  $p$ -fraction of a large sample of equicorrelated multinormal variates, where each variate is with mean 0, variance 1 and with a correlation  $r$  between variates.

The expected genetic gain is then equal to:

$$R = i_{p,r} \rho_1^{0.5} \sigma_a \quad [\text{A3.6}]$$

where  $i_{p,r} = i_p(1-r)^{0.5}$  could be viewed as a selection intensity accounting for the genetic relationship  $r$  between animals (Kennedy and Trus, 1993) and  $\rho_1$  is the overall precision criterion.