

Interaction génotype-milieu et modèle mixte

I. Modélisation

J. MALLARD, J.P. MASSON*, Madeleine DOUAIRE

Laboratoire de Recherches de la Chaire de Génétique
**Laboratoire de Recherches de la Chaire de Statistique et Informatique*
Ecole nationale supérieure agronomique
65, rue de S^t-Brieuc, F 35042 Rennes

Résumé

Un dispositif expérimental souvent utilisé par les généticiens pour prendre en compte les effets du milieu est étudié par deux modèles.

Dans le premier modèle, la variable est décomposée en :

$$Y_{ij, 1}^k = \mu^k + A_i + B_{ij} + \theta_{ij}^k + E_{ij, 1}^k \quad (I)$$

où μ^k représente l'effet moyen de la modalité k du facteur milieu;
 A_i représente l'effet aléatoire du facteur père;
 B_{ij} représente l'effet aléatoire du facteur mère intra père;
 θ_{ij}^k représente l'interaction aléatoire (père, mère) \times milieu.

Ce modèle résulte d'une procédure d'échantillonnage qui rend aléatoires les effets génétiques; il nous donne des espérances et covariances dépendantes du milieu. L'introduction de conditions supplémentaires entraîne des liaisons entre les covariances.

Le deuxième modèle donne la décomposition intrinsèque suivante :

$$Y_{ij, 1}^k = \mu^k + A_i^k + B_{ij}^k + E_{ij, 1}^k \quad (II)$$

avec $E(Y_{ij, 1}^k) = \mu^k$
 $\forall i, \text{cov}(A_i) = (\sigma_A^{(k, k)})$
 $\forall (i, j), \text{cov}(B_{ij}) = (\sigma_B^{(k, k)})$
 $\forall (i, j, 1), \text{cov}(E^k) = \sigma_{E^k}^2 \cdot I_k$.

Les autres covariances sont supposées nulles. Les espérances et les covariances sont fonctions de l'environnement.

Le modèle (I) contient le modèle (II) et permet d'explicitier par identification le contenu des paramètres du modèle (II).

Ces deux modèles apparaissent comme beaucoup plus généraux que des modèles classiquement utilisés sur de tels dispositifs expérimentaux.

Mots-clés : Interaction génotype-milieu; échantillonnage; modèle mixte.

Summary

Genotype X environment interaction. I. Establishing models

An experimental design often used by geneticists to account for environmental effects has been studied setting two models. In the first model, the variable is defined as follows :

$$Y_{ij, 1}^k = \mu^k + A_i + B_{ij} + \theta_{ij}^k + E_{ij, 1}^k \quad (I)$$

where μ^k represents the fixed effect of the k^{th} environment
 A_i represents the random effect of the i^{th} sire
 B_{ij} represents the random effect of the j^{th} dam within i^{th} sire
 θ_{ij}^k represents the interaction of random effects (sire, dam) with environment.

This model results from a sampling procedure which implies that the genetic effects are random. The model gives expectations and covariances which are dependent of the environmental effects.

The second model gives the following intrinsic breakdown :

$$Y_{ij, l}^k = \mu^k + A_i^k + B_{ij}^k + E_{ij, l}^k \tag{II}$$

with $E(Y_{ij, l}^k) = \mu^k$
 $V_i, \text{COV}(A_i) = (\sigma A^{(k, k)})$
 $V(i, l), \text{COV}(B_{ij}) = (\sigma B^{(k, k)})$
 $V(i, j, l), \text{COV}(E^k) = \sigma^2 E^k \cdot I_k$.

The other covariances are assumed to be equal to zero. Again, the expected values and covariances are functions of the environment.

The first model includes the second and by identification enables the parameters of the second to be defined more explicitly. These two models appear to be much more general than those classically used in such experimental designs.

Key-words : Genotype X environment interaction, sampling, mixed model.

I. Introduction

Lorsqu'on exploite différents génotypes (deux, par exemple) dans des milieux différents (deux), on peut obtenir des résultats de types divers qu'on représentera en utilisant la classification de Haldane (1946) (cf. fig. 1).

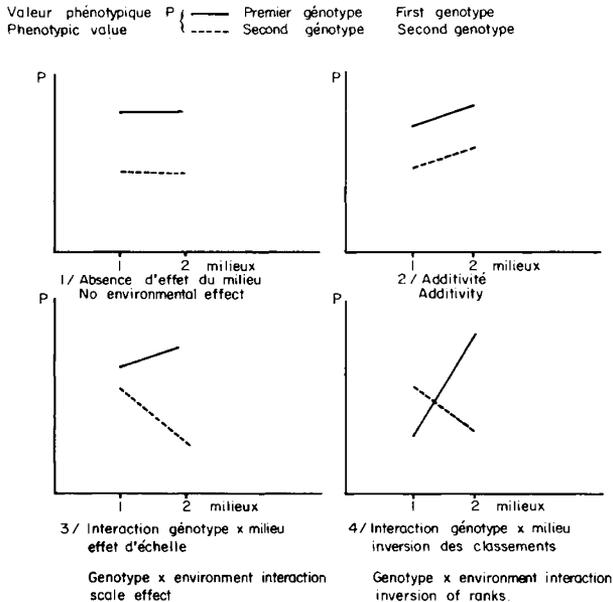


FIG. 1

*Classification de Haldane.
 Haldane's classification.*

- Sur cette figure 1, les cas 1 et 2 correspondent à l'additivité des effets des génotypes et des milieux (en 1, il n'y a pas d'effets du milieu).

- L'interaction décrite en 3 correspond à un phénomène d'échelle, le milieu 2 induisant une plus grande variabilité dans la performance mesurée : le classement des deux génotypes reste inchangé. Ce type d'interaction ne modifie en rien les résultats obtenus par le sélectionneur qui classe des génotypes dans un des milieux.

- Le dernier cas (4), par contre, illustre le « cauchemar » du sélectionneur puisque le changement de milieu modifie l'intérêt relatif des différents génotypes. Dans le pire des cas, il peut l'inverser complètement.

Les divers génotypes issus de schémas de sélection sont aujourd'hui confrontés à une gamme de milieux toujours plus large et le risque de rencontrer certaines inadaptations d'un génotype particulier à un milieu donné — une interaction génotype x milieu (cas 4) — va croissant. Aussi, l'attention portée à l'étude de l'interaction génotype-milieu pour la connaissance et pour l'action se traduit-elle par une bibliographie très importante (BRUN, 1982).

L'aviculture a été pionnière en la matière du fait de sa précoce vocation « mondialiste », mais tous les domaines de l'amélioration génétique sont maintenant concernés.

A. Interprétation des dispositifs de mesure

Pour l'interprétation des résultats de dispositifs expérimentaux visant à quantifier l'importance d'interactions génotype-milieu, deux conceptions — deux modèles explicatifs différents — ont été utilisés quand les milieux sont fixés et les génotypes, « aléatoires ».

Considérons, pour aborder le problème, le plus simple des dispositifs utilisables (schéma croisé génotype x milieu).

- Le premier modèle — le plus « naturel » — consiste à décrire les données par le modèle linéaire mixte de la forme :

$$P_{i, k}^k = \mu + e^k + G_i + (eG)_i^k + Z_{i, k}^k \tag{a}$$

où μ représente la moyenne générale;

e^k représente l'effet — considéré comme fixé — du milieu k ;

G_i représente l'effet — considéré comme aléatoire — du génotype i ;

$(eG)_i^k$ représente l'interaction génotype x milieu (aléatoire);

$Z_{i, k}^k$ représente l'aléatoire résiduelle.

avec les conditions supplémentaires et les hypothèses classiques sur les différents termes; on suppose notamment que toutes les variables aléatoires introduites ($G_i, (eG)_i^k, Z_{i, k}^k$) sont centrées, de variances indépendantes des indices

$$(\sigma_{G_i}^2 = \sigma_G^2, \sigma_{(eG)_i^k}^2 = \sigma_{eG}^2, \sigma_{Z_{i, k}^k}^2 = \sigma_Z^2)$$

et de covariances nulles (cf. tabl. 1).

L'importance de l'interaction est chiffrée par le coefficient de corrélation intraclasse $\sigma_{eG}^2 / \sigma_P^2$ correspondant au terme (eG) .

TABLEAU 1

Matrice de variances-covariances des aléatoires du modèle (a).

Covariance matrix of the random variables of model (a).

	G_i	$(eG)_i^k$	$Z_{i,1}^k$
G_i	σ_G^2	0	0
$(eG)_i^k$	0	σ_{eG}^2	0
$Z_{i,1}^k$	0	0	σ_Z^2

• Le deuxième modèle explicatif (FALCONER, 1952) propose une attitude *a priori* tout à fait différente : considérer l'expression d'un caractère donné dans deux milieux différents comme l'expression de deux caractères génétiquement liés. La corrélation génétique qui mesure leur liaison est utilisée pour chiffrer l'importance de l'interaction. Le modèle multivariable s'écrit :

$$P_{i,1}^k = \mu^k + G_i^k + Z_{i,1}^k \quad (b)$$

Pour un même milieu (indiqué par k), les variables centrées G_i^k et $Z_{i,1}^k$ sont supposées de variances égales notées σ_G^2 et σ_Z^2 et de covariances nulles.

Pour deux milieux différents, k et k', il existe une covariance entre G_i^k et $G_i^{k'}$ notée $\sigma_G^{(k,k')}$, toutes les autres variables étant de covariances nulles (cf. tabl. 2).

B. Comparaison « qualitative » des deux modèles

Les modèles (a) et (b) sont utilisés pour « mesurer » des interactions génotype x milieu; en sélection animale, ils permettent d'estimer des valeurs génétiques à partir des corrélations entre performances d'un même génotype dans différents milieux. Quelles sont alors les conséquences de l'utilisation de l'un ou l'autre de ces modèles?

La comparaison des tableaux 1 et 2 montre que le modèle (b) contient le modèle (a) : moins de covariances sont supposées nulles. La différence essentielle réside dans la possibilité offerte par le modèle (b) de variances inégales pour des variables observées dans des milieux différents. Cela a de nombreuses conséquences :

• Lorsque l'interaction est du type (3) (effet d'échelle), le modèle (b) conduit à une corrélation génétique de 1, alors qu'elle est inférieure pour le modèle (a). DICKERSON (1962) le montre bien sur un exemple. ROBERTSON, (1959) propose une façon de calculer cette corrélation à partir d'un modèle (a) que YAMADA, (1962) présente comme une correction pour prendre en compte l'inégalité des variances pour des dispositifs équilibrés.

TABLEAU 2

Hypothèses relatives au modèle (b).

Hypothesis about the model (b).

Variances et covariances							
	G_i^k		$G_i^{k'}$		$Z_{i,1}^k$		$Z_{i,1}^{k'}$
G_i^k	0	$\sigma_{G^k}^2$	0	$\sigma_{G^{k,k'}}^{(k, k')}$	0	0	0
$G_i^{k'}$	0	$\sigma_{G^{k,k'}}^{(k, k')}$	0	$\sigma_{G^{k'}}^2$	0	0	0
$Z_{i,1}^k$	0	0	0	0	$\sigma_{Z^k}^2$	0	0
$Z_{i,1}^{k'}$	0	0	0	0	0	$\sigma_{Z^{k'}}^2$	0

$E(P_{i,1}^k) = \mu^k$
 $\forall k, E(G_i^k) = E(Z_{i,1}^k) = 0.$

• Le modèle (a) ne permet pas d'obtenir une corrélation génétique négative (puisqu'elle est donnée par un rapport de variances), même dans le cas (4) [inversion des classements] (YAMADA, 1962).

• Le modèle (b) permet bien sûr d'estimer autant de valeurs génétiques qu'il y a de conditions d'exploitation; cela aide en particulier à résoudre le difficile problème du choix du milieu de sélection. Mais surtout, les estimateurs des valeurs génétiques sont construits différemment: le modèle (a) conduit à une égale pondération des performances enregistrées dans des milieux différents, qu'il y ait ou non interaction; tandis que le modèle (b) conduit à des pondérations différentes, dès que la corrélation est différente de 1.

On voit bien que ces modèles traduisent des conceptions différentes; certains auteurs optent pour le modèle (a), comme VEZELY *et al.*, (1970); d'autres ne choisissent pas et utilisent les deux modèles sans étudier véritablement le problème de l'estimation de la covariance génétique (modèle (b)) (PANI *et al.*, 1977).

C. Objectifs de cette étude

Pour aborder l'interaction génotype x milieu dans un schéma de sélection, il nous a paru utile de préciser les différences entre ces deux modèles et d'explicitier d'éventuelles hypothèses par trop restrictives de l'un ou de l'autre.

Nous construirons un modèle mixte par échantillonnage selon la technique classique utilisée encore récemment par G. LEFORT (1977) pour une modélisation dialléle à effets aléatoires. Nous l'appliquerons d'abord à un dispositif simple (deux facteurs croisés) pour expliciter la méthode; nous aborderons ensuite un dispositif très fréquemment

utilisé en sélection animale, comprenant trois facteurs dont deux constituent une hiérarchie aléatoire. Sur ce dispositif, nous comparerons deux modèles : l'un obtenu par échantillonnage, appelé modèle (I), l'autre relevant de la conception (b) précédente, appelé modèle (II). Nous constaterons que ce modèle (I) « contient » le modèle (II) et identifierons les paramètres des deux modèles. L'ensemble des résultats sera alors repris dans le cas particulier de deux milieux (les sexes mâle et femelle) : modèle « covariance frères-sœurs ».

II. Construction d'un modèle par échantillonnage

A. Modèle à deux facteurs croisés : conséquences d'un tirage équiprobable sur les modalités d'un des facteurs

1) Modèle avant tirage

Considérons une population initiale d'objets, partitionnée en K populations correspondant aux K modalités d'un premier facteur fixé (le milieu, par exemple) de mêmes tailles N. Cet effectif est le nombre de modalités d'un deuxième facteur croisé avec le premier. (Nommons génotype ce second facteur.)

Nous posons le modèle :

$$Z_{v, i}^k = m_v^k + \varepsilon_{v, i}^k$$

m_v^k est l'espérance d'une aléatoire $Z_{v, i}^k$ relative au « milieu » k et au « génotype » v. Nous supposons que les erreurs $\varepsilon_{v, i}^k$ sont centrées et de variance $\sigma_v^2(k, v)$.

Nous effectuons la décomposition classique de m_v^k pour faire apparaître les effets des deux facteurs :

$$m_v^k = \mu^k + \beta_v + \theta_v^k$$

avec les contraintes de définition :

$$\sum_{v=1}^N \beta_v = 0; \quad \forall k, \sum_{v=1}^N \theta_v^k = 0; \quad \forall v, \sum_{k=1}^K \theta_v^k = 0.$$

2) Tirage sur les génotypes

Effectuons un tirage exhaustif équiprobable de taille J sur les N modalités du génotype; nous construisons donc un modèle où le milieu est fixé et le génotype aléatoire, souvent rencontré en production animale. (On aurait pu constituer l'inverse : des génotypes (lignées végétales) fixés et des milieux aléatoires.)

Soient X_{jv} les aléatoires associées au tirage :

$$X_{jv} = \begin{cases} 1 & \text{si la modalité } v \text{ est retenue au } j^{\text{ième}} \text{ tirage} \\ 0 & \text{sinon} \end{cases} \quad (j = 1, \dots, J)$$

avec notamment $P\{X_{jv} = 1\} = E\{X_{jv}\} = \frac{1}{N}$

$$\sum_{v=1}^N X_{jv} = 1.$$

Si $Y_{j, i}^k$ est la 1^{ière} observation du k^{ième} niveau du premier facteur et du niveau du deuxième facteur obtenu lors du j^{ième} tirage, le modèle devient :

$$Y_{j, i}^k = \sum_{v=1}^N X_{jv} Z_{v, i}^k$$

$$= \mu^k + \sum_{v=1}^N X_{jv} \beta_v + \sum_{v=1}^N X_{jv} \theta_v^k + \sum_{v=1}^N X_{jv} \varepsilon_{v, i}^k.$$

Soit :

$$Y_{j, i}^k = \mu^k + B_j + \Theta_j^k + E_{j, i}^k,$$

$$\text{où } B_j = \sum_{v=1}^N X_{jv} \beta_v$$

$$\Theta_j^k = \sum_{v=1}^N X_{jv} \theta_v^k$$

$$E_{j, i}^k = \sum_{v=1}^N X_{jv} \varepsilon_{v, i}^k.$$

Les propriétés suivantes de ce modèle résultent de l'échantillonnage :

- L'espérance de $Y_{j, i}^k$ dépend seulement de la modalité du facteur fixé (le milieu, par exemple); en effet, $E(B_j) = E(\Theta_j^k) = E(E_{j, i}^k) = 0$, donc

$$E(Y_{j, i}^k) = \mu^k.$$

- Les différents effets interviennent dans la structure de variance-covariance :

→ $\text{var}(B_j) = \frac{1}{N} \sum_v \beta_v^2 = \sigma_\beta^2$, variance empirique sur la population (inconnue) des β_v ; elle est indépendante de j.

$\text{cov}(B_j, B_{j'}) = -\frac{1}{N-1} \sigma_\beta^2$. Elle est indépendante des indices j et j' et tend vers zéro quand N tend vers l'infini.

→ $\text{var}(\Theta_j^k) = \frac{1}{N} \sum_{v=1}^N (\theta_v^k)^2 = \sigma_\theta^2(k)$, variance empirique sur la population des θ_v^k , à k fixé.

$\text{cov}(\Theta_j^k, \Theta_{j'}^{k'}) = \frac{1}{N} \sum_{v=1}^N \theta_v^k \theta_v^{k'} = \sigma_\theta(k, k')$, covariance empirique sur la population des couples $(\theta_v^k, \theta_v^{k'})$, $v = 1, \dots, N$.

$$\left. \begin{aligned} \text{cov}(\Theta_j^k, \Theta_j^{k'}) &= -\frac{1}{N-1} \sigma_\theta^2(k) \\ \text{cov}(\Theta_j^k, \Theta_{j'}^{k'}) &= -\frac{1}{N-1} \sigma_\theta(k, k') \end{aligned} \right\} \begin{array}{l} \text{qui tendent vers } 0 \\ \text{quand } N \text{ tend vers l'infini} \end{array}$$

→ $\text{var}(E_{j, i}^k) = \frac{1}{N} \sum_v \sigma_\varepsilon^2(k, v) = \sigma_\varepsilon^2(k)$, variance dépendant du milieu k.

→ $\text{cov}(B_j, \Theta_j^k) = \frac{1}{N} \sum_v \beta_v \theta_v^k = \sigma_{\beta\theta}(k)$, covariance empirique sur la population des couples (β_v, θ_v^k) , $v = 1, \dots, N$, à k fixé.

$$\text{cov}(B_j, \Theta_j^k) = 0.$$

La décomposition classique de m_k^k utilisée fait apparaître sur les aléatoires du modèle les contraintes suivantes :

$$\sum_k \Theta_j^k = 0 \quad (\text{var } \Theta_j^k = 0)$$

et par suite : $\sum_k \text{cov}(B_j, \Theta_j^k) = 0$.

Il n'y a pas d'interaction si et seulement si $\sum_k \sigma_\theta^2(k) = 0$ et alors quel que soit (k, k') , $\sigma_\theta(k, k') = \sigma_{\beta\theta}(k) = \sigma_{\beta\theta}(k') = 0$; σ_β^2 est la variance de l'aléatoire associée au facteur échantillonné quand le modèle est additif.

Dans la suite, nous ne présenterons plus les covariances qui tendent vers zéro quand N tend vers l'infini.

En fin de compte, si nous supposons N assez grand, nous obtenons :

$$\begin{cases} E(Y_{j, 1}^k) = \mu^k \\ \text{var}(Y_{j, 1}^k) = \sigma_\beta^2 + \sigma_\theta^2(k) + 2\sigma_{\beta\theta}(k) + \sigma_E^2(k) \\ \text{cov}(Y_{j, 1}^k, Y_{j, 1}^{k'}) = \sigma_\beta^2 + \sigma_\theta(k, k') + \sigma_{\beta\theta}(k) + \sigma_{\beta\theta}(k'). \end{cases}$$

Le tableau 3 explicite ces résultats dans un cas particulier simple. Il convient de souligner que l'espérance et la matrice de variances-covariances du vecteur des aléatoires observables dépendent de la modalité du facteur fixé; en revanche, il n'y a pas de corrélation entre des variables associées à des modalités différentes du facteur aléatoire. On retrouve donc les hypothèses du modèle mixte de SCHEFFE (1956, 1959).

TABLEAU 3

Modèle mixte à deux facteurs croisés obtenu par échantillonnage.

Mixed model with two crossed factors obtained by sampling.

		Variances et covariances					
	Espérances	y_1^1	y_1^2	y_2^1	y_2^2	y_3^1	y_3^2
y_1^1	μ^1	σ_1^2	σ_{12}	0		0	
y_1^2	μ^2	σ_{21}	σ_2^2				
y_2^1	μ^1	0		σ_1^2	σ_{12}	0	
y_2^2	μ^2			σ_{21}	σ_2^2		
y_3^1	μ^1	0		0		σ_1^2	σ_{12}
y_3^2	μ^2					σ_{21}	σ_2^2

K = 2
J = 3
N grand.

Remarquons enfin que cette présentation met bien en évidence que le terme d'interaction est de nature aléatoire et que sa loi dépend de la modalité du facteur fixé.

On peut donner à l'interaction θ_j^k différentes formes et en déduire les conséquences au niveau de la structure de variances-covariances [modèles de TUKEY, 1949; GOFFINET & VINCOURT, 1980].

3) *Modèle à 3 facteurs croisés; conséquences d'un tirage « hiérarchique »*

Nous avons appliqué ces principes à un modèle plus utile dans la pratique de la sélection animale où les génotypes sont structurés par deux facteurs hiérarchisés (hiérarchie père, mère/père, par exemple). Une étude statistique de ce modèle sera proposée dans la deuxième partie de cette étude.

a) *Modèle initial*

Considérons le modèle

$$Z_{uv, l}^k = \frac{\mu^k + \alpha_u + \beta_v + (\alpha\beta)_{uv} + (\alpha\beta\mu)_{uv, k} + \varepsilon_{uv, l}^k}{m_{uv}^k}$$

à trois facteurs croisés (facteurs milieu, père, mère, par exemple) $k = 1, \dots, K$; $u = 1, \dots, P$; $v = 1, \dots, M$. L'équivalent du paramètre N de l'étude précédente est, dans ce cas, le produit $P \times M$. Les paramètres inconnus $\mu^k, \alpha_u, \beta_v, (\alpha\beta)_{uv}, (\alpha\beta\mu)_{uv, k}$ (tous fixés pour le moment) sont définis par les conditions supplémentaires :

$$\sum_{u=1}^P \alpha_u = 0; \quad \sum_{v=1}^M \beta_v = 0$$

$$\forall u, \sum_v (\alpha\beta)_{uv} = 0; \quad \forall v, \sum_u (\alpha\beta)_{uv} = 0$$

$$\forall k, \sum_{u, v} (\alpha\beta\mu)_{uv, k} = 0; \quad \forall (u, v), \sum_k (\alpha\beta\mu)_{uv, k} = 0.$$

Les erreurs $\varepsilon_{uv, l}^k$ sont centrées, indépendantes.

L'interaction entre le premier facteur indicé par k et les deux autres facteurs est notée $(\alpha\beta\mu)_{uv, k}$; nous la décomposerons parfois en :

$$(\alpha\beta\mu)_{uv, k} = (\alpha\mu)_u^k + (\beta\mu)_v^k + (\alpha\beta\mu)_{uv}^k$$

avec les contraintes habituelles.

b) *Tirage « hiérarchique »*

Indépendamment du premier facteur indicé par k (facteur milieu, par exemple), nous effectuons :

a) Un premier tirage exhaustif relatif au deuxième facteur indicé par u. L'aléatoire associée au $i^{\text{ème}}$ de ces tirages ($i = 1, \dots, I$) sera notée S_{iu} avec :

$$S_{iu} = \begin{cases} 1 & \text{si la modalité } u \text{ du deuxième facteur est retenue} \\ 0 & \text{sinon.} \end{cases}$$

b) Un deuxième tirage exhaustif conditionnel relatif au troisième facteur indicé par v retenant I échantillons distincts de J modalités du troisième facteur. A chaque modalité retenue du deuxième facteur est associée un ensemble de J modalités du troisième facteur. Les aléatoires associées aux tirages sont notées T_{jv}^i avec :

$$T_{jv}^i = \begin{cases} 1 & \text{si la modalit  v du troisi me facteur est retenue} \\ & \text{au j me tirage dans la i me modalit  du second} \\ 0 & \text{sinon.} \end{cases}$$

Des calculs du m me type que ceux effectu s au § II.A permettent d' crire le mod le (I) ci-dessous et d'en d duire ces propri t s

$$Y_{ij,l}^k = \mu^k + A_i + \underbrace{U_{ij} + V_{ij} + \theta_{ij}^k + E_{ij,l}^k}_{B_{ij}} \tag{I}$$

avec
$$A_i = \sum_{u=1}^P S_{iu} \alpha_u,$$

al atoire correspondant au 2^e facteur maintenant devenu facteur hi rarchisant (facteur p re, par exemple)

$$U_{ij} = \sum_{v=1}^M T_{jv}^i \beta_v,$$

al atoire correspondant   l'effet principal du troisi me facteur (facteur m re, par exemple)

$$V_{ij} = \sum_{v=1}^M T_{jv}^i \left\{ \sum_{u=1}^P S_{iu} (\alpha\beta)_{uv} \right\},$$

al atoire correspondant   l'interaction des 2^e et 3^e facteurs

$$\theta_{ij}^k = \sum_{v=1}^M T_{jv}^i \left\{ \sum_{u=1}^P S_{iu} (\alpha\beta\mu)_{uv}^k \right\},$$

al atoire correspondant   l'interaction du premier et des 2 derniers facteurs

$$E_{ij,l}^k = \sum_{v=1}^M T_{jv}^i \left\{ \sum_{u=1}^P S_{iu} e_{u,l}^k \right\}.$$

c) * tude du mod le mixte obtenu et cons quences*

L'esp rance de la variable al atoire observable $Y_{ij,l}^k$ ne d pend que de la modalit  k du facteur fix  (le milieu, par exemple) : $E(Y_{ij,l}^k) = \mu^k$; les autres al atoires A_i , U_{ij} , V_{ij} , θ_{ij}^k et $E_{ij,l}^k$ sont centr es.

En ce qui concerne la structure de variances-covariances, on obtient ais ment les termes :

$$\begin{cases} \longrightarrow \text{var}(A_i) = \frac{1}{P} \sum_u \alpha_u^2 = \sigma_{\alpha}^2, & \text{variance empirique de l'effet principal associ  au} \\ & \text{2^e facteur (p re, par exemple).} \\ \longrightarrow \text{var}(U_{ij}) = \frac{1}{M} \sum_v \beta_v^2 = \sigma_{\beta}^2, & \text{variance empirique de l'effet principal associ  au} \\ & \text{3^e facteur (m re, par exemple).} \\ \text{var}(V_{ij}) = \frac{1}{PM} \sum_u \sum_v (\alpha\beta)_{uv}^2 = \sigma_{\alpha\beta}^2, & \text{variance empirique de l'interaction } 2 \times 3 \text{ (p re} \times \\ & \text{m re).} \end{cases}$$

$$\rightarrow \begin{cases} \text{var}(\theta_{ij}^k) = \frac{1}{PM} \sum_u \sum_v (\alpha\beta\mu)_{uv, k}^2 = \sigma_\theta^2(k) \\ \text{cov}(\theta_{ij}^k; \theta_{ij'}^{k'}) = \frac{1}{PM} \sum_u \sum_v (\alpha\beta\mu)_{uv, k} (\alpha\beta\mu)_{uv, k'} = \sigma_\theta(k, k') \\ \text{cov}(\theta_{ij}^k; \theta_{ij'}^{k'}) = \frac{1}{P} \sum_u (\alpha\mu)_u^k (\alpha\mu)_u^{k'} = \sigma_{\alpha\mu}(k, k') \end{cases}$$

$$\rightarrow \begin{cases} \text{cov}(A_i, \theta_{ij}^k) = \frac{1}{P} \sum_u \alpha_u (\alpha\mu)_u^k = \sigma_{\alpha, \alpha\mu}(k) \\ \text{cov}(U_{ij} + V_{ij}; \theta_{ij}^k) = \frac{1}{PM} \sum_u \sum_v (\beta_v + (\alpha\beta)_{uv}) (\alpha\beta\mu)_{uv, k} \\ = \sigma_{\beta, \beta\mu}(k) + \sigma_{\alpha\beta, \alpha\beta\mu}(k) \end{cases}$$

avec

$$\sigma_{\beta, \beta\mu}(k) = \frac{1}{M} \sum_v \beta_v (\beta\mu)_v^k$$

$$\sigma_{\alpha\beta, \alpha\beta\mu}(k) = \frac{1}{PM} \sum_u \sum_v (\alpha\beta)_{uv} (\alpha\beta\mu)_{uv, k}$$

La décomposition classique de m_{uv}^k utilisée fait apparaître sur les aléatoires du modèle les contraintes suivantes :

$$\sum_k \theta_{ij}^k = 0 \quad (\text{var } \theta_j = 0)$$

et par suite les sommes sur l'indice k (milieu) de chacune des covariances sont nulles.

Il n'y a pas d'interaction si et seulement si $\sum_k \sigma_\theta^2(k) = 0$ et donc quel que soit (k, k'), $\sigma_\theta(k, k')$, $\sigma_{\alpha\mu}(k, k')$, $\sigma_{\alpha, \alpha\mu}(k)$, $\sigma_{\beta, \beta\mu}(k) + \sigma_{\alpha\beta, \alpha\beta\mu}(k)$ sont nuls.

Les autres termes de covariance soit sont nuls, soit présentent $\frac{1}{M}$, $\frac{1}{P}$ ou $\frac{1}{PM}$ en facteur et sont de ce fait négligeables pour M et P assez grands.

On notera sans surprise que l'ensemble de ces résultats est identique (au niveau de complexité près) aux résultats présentés sur le modèle à 2 facteurs croisés.

Nous résumons dans le tableau 4 l'ensemble de ces résultats quand M et P sont assez grands.

TABEAU 4

*Modèle à 3 facteurs croisés; conséquences d'un tirage hiérarchique. Tableau des résultats.
Three way crossed lay out model ; consequences of a nested sampling. Table of results.*

$j \neq j'; k \neq k'$	$\sigma_\alpha^2 + \sigma_{\alpha\mu}(k, k') + \sigma_{\alpha, \alpha\mu}(k) + \sigma_{\alpha, \alpha\mu}(k')$
$j \neq j'; k = k'$	(même milieu) (same environment) $\sigma_\alpha^2 + \sigma_{\alpha\mu}^2(k) + 2\sigma_{\alpha, \alpha\mu}(k)$

$j = j'; k \neq k'$ (même mère)
(same dam)

$$\sigma_{\alpha}^2 + \sigma_{\beta/\alpha}^2 + \sigma_{\theta}(k, k') + \sigma_{\alpha, \alpha\mu}(k) + \sigma_{\beta, \beta\mu}(k) + \sigma_{\alpha\beta, \alpha\beta\mu}(k) + \sigma_{\alpha, \alpha\mu}(k') + \sigma_{\beta, \beta\mu}(k') + \sigma_{\alpha\beta, \alpha\beta\mu}(k')$$

$j = j'; k = k'; l \neq m$ (même mère; même milieu)
(same dam; same environment)

$$\sigma_{\alpha}^2 + \sigma_{\beta/\alpha}^2 + \sigma_{\theta}^2(k) + 2\sigma_{\alpha, \alpha\mu}(k) + 2\sigma_{\beta, \beta\mu}(k) + 2\sigma_{\alpha\beta, \alpha\beta\mu}(k)$$

$j = j'; k = k'; l = m$ (même mère; même milieu; même répétition)
variance
(same dam; same environment; same repetition)

$$\sigma_{\alpha}^2 + \sigma_{\beta/\alpha}^2 + \sigma_{\theta}^2(k) + 2\sigma_{\alpha, \alpha\mu}(k) + 2\sigma_{\beta, \beta\mu}(k) + 2\sigma_{\alpha\beta, \alpha\beta\mu}(k) + \sigma_E^2(k)$$

$$E(Y_{ij, l}^k) = \mu^k.$$

Les aléatoires observables ne dépendent pas du même indice i (pères différents) sont non corrélées. Pour les autres, la covariance est :

The random variables which don't depend of the same subscript i (different sires) are uncorrelated. For the others ones the covariance matrix is :

$$\text{cov}(Y_{ij, l}^k; Y_{ij', m}^k) =$$

III. Modèle multivariable

Nous avons présenté en introduction une façon d'aborder l'étude de l'interaction génotype-milieu qui est de considérer l'expression d'un caractère dans deux milieux différents comme l'expression de deux caractères génétiquement liés. Sur le dispositif précédemment étudié (trois facteurs dont deux constituent une hiérarchie aléatoire), le modèle b de l'introduction — qui relève de cette conception — devient le modèle II suivant :

$$Y_{ij, l}^k = \mu^k + A_i^k + B_{ij}^k + E_{ij, l}^k \tag{II}$$

- $k = 1, \dots, K$ indice les modalités du facteur « milieu »
- $i = 1, \dots, I$ indice les modalités du facteur « père »
- $j = 1, \dots, J_i$ indice les modalités du facteur « mère/père »
- l : indice de répétition à l'intérieur d'une modalité $\frac{k}{ij}$.

avec $E(Y_{ij, l}^k) = \mu^k.$

$$\begin{aligned} \forall i, \text{cov}(A_i) &= (\sigma_A^{(k, k')}) \\ \forall(i, j), \text{cov}(B_{ij}) &= (\sigma_B^{(k, k')}) \\ \forall(i, j, l), \text{cov}(E^k) &= \sigma_E^2 k \cdot I_k. \end{aligned}$$

I_k est la matrice identité d'ordre égal au nombre d'aléatoires observables pour la modalité k .

A_i est le vecteur des A_i^k ; B_{ij} est le vecteur des B_{ij}^k et E_k est le vecteur des $E_{ij, l}^k$.

IV. Identification

Nous disposons des deux modèles (I) et (II) proposés pour décrire une même réalité. La méthode de construction du modèle (I) que nous avons adoptée nous permet de procéder à un essai d'identification des termes de ces deux modèles.

A. Identification des modèles (I) et (II)

L'écriture des espérances est rigoureusement identique dans les deux cas; elles dépendent de la seule modalité du facteur fixé (milieu).

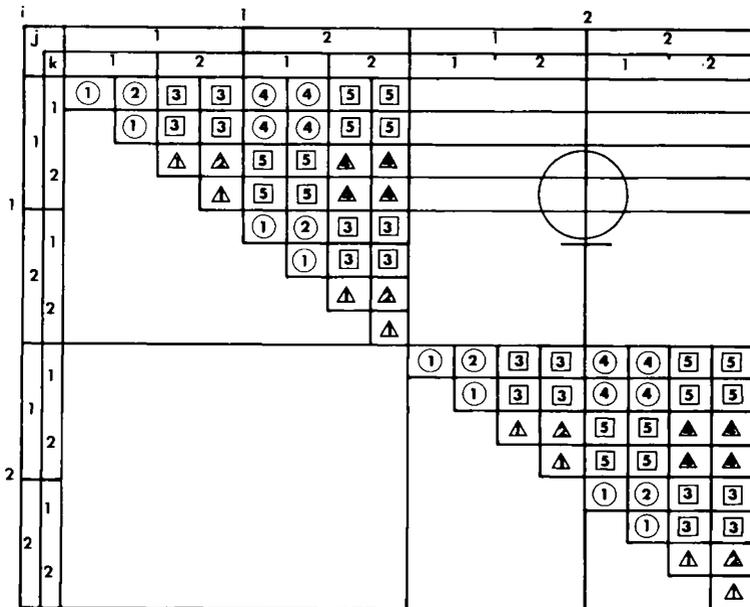
Nous allons maintenant identifier les structures de covariances des aléatoires introduites. La figure 2 illustre cette identification. Les membres de gauche contiennent les termes du modèle (II).

FIGURE 2

Structure de covariance des aléatoires observables d'un modèle mixte obtenu par échantillonnage « hiérarchisé » des génotypes.

Covariance matrix of the random variables of a mixed model obtained by « nested » sampling of the genotypes.

a) *Matrice des variances-covariances — Covariance matrix.*



Nous prenons $K = 2$ $I = 2$ $J = 2$ $L = 2$.
 Nous supposons M et P grands.

b) *Légende et contenu des symboles*

Meaning of abbreviations

○ k = 1 1^{er} milieu first environment
 △ k = 2 2^e milieu second environment

	modèle (II)	modèle (I)	
①	$\sigma_{E_1}^2 + \sigma_A^{2(1)} + \sigma_B^{2(1)}$	$\sigma_{E_1}^2(1)$	$\sigma_{\theta}^2(1) + 2[\sigma_{\alpha, \alpha\mu}(1) + \sigma_{\beta, \beta\mu}(1) + \sigma_{\alpha\beta, \alpha\beta\mu}(1)]$
△	$\sigma_{E_2}^2 + \sigma_A^{2(2)} + \sigma_B^{2(2)}$	$\sigma_{E_2}^2(2)$	$\sigma_{\theta}^2(2) + 2[\sigma_{\alpha, \alpha\mu}(2) + \sigma_{\beta, \beta\mu}(2) + \sigma_{\alpha\beta, \alpha\beta\mu}(2)]$
②	$\sigma_A^{2(1)} + \sigma_B^{2(1)}$	$\sigma_{\alpha}^2 + \sigma_{\beta/\alpha}^2 +$	$\sigma_{\theta}^2(1) + 2[\sigma_{\alpha, \alpha\mu}(1) + \sigma_{\beta, \beta\mu}(1) + \sigma_{\alpha\beta, \alpha\beta\mu}(1)]$
△	$\sigma_A^{2(2)} + \sigma_B^{2(2)}$	$\sigma_{\alpha}^2 + \sigma_{\beta/\alpha}^2 +$	$\sigma_{\theta}^2(2) + 2[\sigma_{\alpha, \alpha\mu}(2) + \sigma_{\beta, \beta\mu}(2) + \sigma_{\alpha\beta, \alpha\beta\mu}(2)]$
③	$\sigma_A^{(1, 2)} + \sigma_B^{(1, 2)}$	$\sigma_{\alpha}^2 + \sigma_{\beta/\alpha}^2 + \sigma_{\theta}^{(1, 2)} + \sigma_{\alpha, \alpha\mu}(1) + \sigma_{\beta, \beta\mu}(1) + \sigma_{\alpha\beta, \alpha\beta\mu}(1)$	$+ \sigma_{\alpha, \alpha\mu}(2) + \sigma_{\beta, \beta\mu}(2) + \sigma_{\alpha\beta, \alpha\beta\mu}(2)]$
④	$\sigma_A^{2(1)}$	$\sigma_{\alpha}^2 +$	$\sigma_{\alpha\mu}^2(1) + 2\sigma_{\alpha, \alpha\mu}(1)$
△	$\sigma_A^{2(2)}$	$\sigma_{\alpha}^2 +$	$\sigma_{\alpha\mu}^2(2) + 2\sigma_{\alpha, \alpha\mu}(2)$
⑤	$\sigma_A^{(1, 2)}$	$\sigma_{\alpha}^2 + \sigma_{\alpha\mu}(1, 2) + \sigma_{\alpha, \alpha\mu}(1)$	$+ \sigma_{\alpha, \alpha\mu}(2)$

Remarque :

Dans ce cas particulier où il n'y a que deux milieux, les contraintes de définition du modèle impliquent :

$$\begin{aligned} \sigma_{\theta}^2(1) &= \sigma_{\theta}^2(2) = -\sigma_{\theta}(1, 2) \\ \sigma_{\alpha\mu}^2(1) &= \sigma_{\alpha\mu}^2(2) = -\sigma_{\alpha\mu}(1, 2) \\ \sigma_{\alpha, \alpha\mu}(1) + \sigma_{\alpha, \alpha\mu}(2) &= 0 \\ [\sigma_{\beta, \beta\mu}(1) + \sigma_{\alpha\beta, \alpha\beta\mu}(1)] + [\sigma_{\beta, \beta\mu}(2) + \sigma_{\alpha\beta, \alpha\beta\mu}(2)] &= 0. \end{aligned}$$

1. *Identification au 1^{er} niveau de la hiérarchie*

$$\sigma_A^{(k, k')} \equiv \sigma_{\alpha}^2 + \sigma_{\alpha\mu}(k, k') + \sigma_{\alpha, \alpha\mu}(k) + \sigma_{\alpha, \alpha\mu}(k')$$

et de même,

$$\sigma_A^{2(k)} \equiv \sigma_{\alpha}^2 + \sigma_{\alpha\mu}^2(k) + 2\sigma_{\alpha, \alpha\mu}(k).$$

On remarquera que $\sigma_A^{2(k)}$ est, compte tenu de la définition des différents termes du modèle (I), identifié à $\frac{1}{p} \sum_u (\alpha_u + (\alpha\mu)_u^k)^2 \geq 0$.

2. *Identification au 2^e niveau de la hiérarchie*

Nous rappelons que, dans le modèle (I), nous avons la décomposition :

$$\begin{aligned} \sigma_{\theta}(k, k') &= \frac{1}{PM} \sum_u \sum_v (\alpha\beta\mu)_{uv, k} (\alpha\beta\mu)_{uv, k'} \\ &= \sigma_{\alpha\mu}(k, k') + \sigma_{\beta\mu}(k, k') + \sigma_{\alpha\beta\mu}(k, k') \\ \sigma_B^{(k, k')} &\equiv \sigma_{\beta/\alpha}^2 + \sigma_{\beta\mu}(k, k') + \sigma_{\alpha\beta\mu}(k, k') + 2[\sigma_{\beta, \beta\mu}(k, k') + \sigma_{\alpha\beta, \alpha\beta\mu}(k, k')] \end{aligned}$$

et de même,

$$\sigma_B^{2(k)} \equiv \sigma_{\beta/\alpha}^2 + \sigma_{\beta\mu}^2(k) + \sigma_{\alpha\beta\mu}^2(k) + 2[\sigma_{\beta, \beta\mu}(k) + \sigma_{\alpha\beta, \alpha\beta\mu}(k)]$$

avec $\sigma_{\beta/\alpha}^2 = \sigma_{\beta}^2 + \sigma_{\alpha\beta}^2$.

On remarquera que $\sigma_{\beta}^{2(k)}$ est, compte tenu de la définition des différents termes du modèle (I), identifié à :

$$\frac{1}{M} \sum_{\nu} [\beta_{\nu} + (\beta\mu)_{\nu}^k]^2 + \frac{1}{PM} \sum_{\mu} \sum_{\nu} [(\alpha\beta)_{\mu\nu} + (\alpha\beta\mu)_{\mu\nu}^k]^2 \geq 0.$$

B. Interprétations et conclusions

Le modèle d'échantillonnage proposé (modèle (I)) (avec M et P infiniment grands) permet de justifier par des arguments d'échantillonnage simples le modèle (II) proposé a priori par le généticien. Il donne une forme aux différentes variances et covariances du modèle (II).

Les seuls paramètres « estimables » sont ceux du modèle (II) ou des combinaisons de ces paramètres. Les décompositions plus fines du modèle (I) dépendent des conditions supplémentaires imposées aux paramètres pour définir les effets principaux et les interactions avant tirage et les aléatoires résultant de l'échantillonnage, par la suite. Certains de ces termes présentent cependant un intérêt appréciable pour le généticien. Ainsi, σ_{α}^2 représente la « vraie » variance père « indépendante » des effets du milieu : une valeur génétique générale; de même, $\sigma_{\beta/\alpha}^2$ représente la « vraie » variance mère intra père; ils décrivent la partie du génotype capable de s'exprimer dans toute la gamme des milieux envisagés, c'est-à-dire très précisément ce que le sélectionneur vise à améliorer.

C. Cas particulier : le modèle « covariances frères-sœurs »

Dans de nombreux cas — recensés dans la seconde partie de cette étude — le facteur fixé n'a que deux modalités (par exemple, le sexe de l'individu mesuré). Notre exemple de la figure 2 correspond à cette situation.

L'écriture s'en trouve simplifiée et permet de mettre plus aisément en évidence les conclusions de l'identification.

Nous pouvons écrire :

— au premier niveau de la hiérarchie

$$\sigma_A^2 \sigma^{\alpha} = \sigma_{\alpha}^2 + \Phi(\sigma^{\alpha}, \sigma^{\alpha})$$

où σ_{α}^2 est la « vraie » variance père et $\Phi(k, k')$ est une partie de l'interaction génotype x milieu;

$$\begin{aligned} \Phi(k, k') &= \sigma_{\alpha\mu}(k, k') + \sigma_{\alpha, \alpha\mu}(k) + \sigma_{\alpha, \alpha\mu}(k') \\ \sigma_{A, \alpha}^2 \sigma_{\alpha}^{\alpha} &= \sigma_{\alpha}^2 + \Phi(\sigma_{\alpha}^{\alpha}, \sigma_{\alpha}^{\alpha}) \\ \sigma_{A, \alpha}^2 \sigma_{\alpha}^{\alpha} &= \sigma_{\alpha}^2 + \Phi(\sigma_{\alpha}^{\alpha}, \sigma_{\alpha}^{\alpha}) \end{aligned}$$

— de même, au deuxième niveau de la hiérarchie

$$\sigma_B^2 \sigma^{\alpha} = \sigma_{\beta/\alpha}^2 + \psi(\sigma^{\alpha}, \sigma^{\alpha})$$

où $\sigma_{\beta/\alpha}^2$ est la « vraie » variance mère intra père et $\psi(k, k')$ est une partie de l'interaction génotype x milieu.

$$\begin{aligned} (\psi(k, k') &= \sigma_{\beta\mu}(k, k') + \sigma_{\alpha\beta\mu}(k, k') + 2[\sigma_{\beta, \beta\mu}(k, k') + \sigma_{\alpha\beta, \alpha\beta\mu}(k, k')]) \\ \sigma_{B, \alpha}^2 \sigma_{\alpha}^{\alpha} &= \sigma_{\beta/\alpha}^2 + \psi(\sigma_{\alpha}^{\alpha}, \sigma_{\alpha}^{\alpha}) \\ \sigma_{B, \alpha}^2 \sigma_{\alpha}^{\alpha} &= \sigma_{\beta/\alpha}^2 + \psi(\sigma_{\alpha}^{\alpha}, \sigma_{\alpha}^{\alpha}). \end{aligned}$$

V. Conclusions

Le modèle $Y_{ij, 1}^k = \mu_k + A_i + B_{ij} + (A\mu)_i^k + (AB\mu)_{ij}^k + E_{ij, 1}^k$ (dédit du modèle (a) vu au début) assorti des hypothèses classiques sur les variances est un sous-modèle du modèle (II). Les différentes variances (covariances) ne dépendent pas de la (des) modalité(s) du facteur fixé, ce qui ne constitue pas toujours un modèle adapté; ce modèle chiffre une inversion de classement des génotypes ou un simple effet d'échelle par une corrélation intra-classe également positive et ne satisfait pas le sélectionneur. Les corrections proposées pour corriger les biais les plus gênants liés à l'usage abusif de ce modèle présentent des imperfections, comme le souligne YAMADA, (1962).

Nous avons, utilisant une procédure simple d'échantillonnage, justifié le modèle (II) posé a priori par le généticien dans le droit fil des idées de Falconer. Cette étude nous a conduit à une décomposition plus riche, mais non intrinsèque; l'identification des paramètres des modèles (I) et (II) nous permet d'affiner le sens et l'interprétation, notamment, en montrant que l'espérance, mais aussi la structure de variances-covariances dépendent du milieu.

Cette démarche est applicable à des situations plus compliquées; d'autres types de tirage sont possibles.

Il reste à développer ces modèles au plan statistique. C'est le but de la deuxième partie de ce travail (Interaction génotype-milieu et modèle mixte-Statistiques pour un modèle à deux niveaux de hiérarchie).

Reçu le 18 octobre 1982.

Accepté le 29 avril 1983.

Références bibliographiques

- BRUN J.M., 1982. Interactions géneur X population des partenaires. I. Définition d'indicateurs. *Ann. Génét. Sél. Anim.*, **4**, 463-480.
- DICKERSON G.E., 1962. Implications of genetic-environmental interaction in animal breeding. *Anim. Prod.*, **4**, 47-67.
- FALCONER D.S., 1952. The problem of environments and selection. *Am. Nat.*, **86**, 293-298.
- GOFFINET B., VINCOURT P., 1980. Prise en compte de l'interaction génotype x milieu en sélection végétale. In : LEGAY J.M. et al., *Biométrie et Génétique*, vol. 4, 90-97, Société Française de Biométrie. I.N.R.A., Département de Biométrie.
- HALDANE J.B.S., 1946. The interaction of nature and nurture. *Ann. Eugen.*, **19**, 197-205.
- LEFORT G., 1977. Remarques sur la modélisation et l'interprétation des dispositifs dialèles. *Ann. Amélior. Plant.*, **27**, 171-202.
- LEFORT G., 1980. Le modèle de base de la sélection, justification et limites. In : LEGAY J.M. et al., *Biométrie et Génétique*, vol. 4, 1-14, Société Française de Biométrie. I.N.R.A., Département de Biométrie.
- PANI S.N., KRAUSE G.F., LASLEY J.F., 1977. The importance of sire x sex interaction for preweaning and weaning traits. *J. Anim. Sci.*, **43**, 72-85.
- ROBERTSON A., 1959. The sampling variance of the genetic correlation coefficient. *Biometrics*, **15**, 469-485.
- SCHEFFE H., 1956. A «mixed model» for the analysis of variance. *Ann. Math. Stat.*, **27**, 23-36.
- SCHEFFE H., 1959. *The analysis of variance*. J. Wiley; New York.
- TUKEY J.W., 1949. One degree of freedom for non-additivity. *Biometrics*, **5**, 232-242.
- VEZELY J.A., ROBINSON O.W., 1970. Genotype-sex interaction in sheep. *J. Anim. Sci.*, **31**, 469-485.
- YAMADA Y., 1962. Genotype by environmental interaction and genetic correlation of the same trait under different environments. *Jap. J. Genet.*, **37**, 498-509.