

# Simulation analysis to test the influence of model adequacy and data structure on the estimation of genetic parameters for traits with direct and maternal effects

Virginie CLÉMENT<sup>a,\*</sup>, Bernard BIBÉ<sup>a</sup>,  
Étienne VERRIER<sup>b,c</sup>,  
Jean-Michel ELSEN<sup>a</sup>,  
Eduardo MANFREDI<sup>a</sup>,  
Jacques BOUX<sup>a</sup>, Éric HANOCQ<sup>a</sup>

<sup>a</sup> Station d'amélioration génétique des animaux, Institut national de la recherche agronomique, BP 27, 31326 Castanet-Tolosan Cedex, France

<sup>b</sup> Station de génétique quantitative et appliquée, Institut national de la recherche agronomique, 78352 Jouy-en-Josas Cedex, France

<sup>c</sup> Département des sciences animales, Institut national agronomique Paris-Grignon, 16 rue Claude Bernard, 75231 Paris Cedex 05, France

(Received 3 May 2000; accepted 5 May 2001)

**Abstract** – Simulations were used to study the influence of model adequacy and data structure on the estimation of genetic parameters for traits governed by direct and maternal effects. To test model adequacy, several data sets were simulated according to different underlying genetic assumptions and analysed by comparing the correct and incorrect models. Results showed that omission of one of the random effects leads to an incorrect decomposition of the other components. If maternal genetic effects exist but are neglected, direct heritability is overestimated, and sometimes more than double. The bias depends on the value of the genetic correlation between direct and maternal effects. To study the influence of data structure on the estimation of genetic parameters, several populations were simulated, with different degrees of known paternity and different levels of genetic connectedness between flocks. Results showed that the lack of connectedness affects estimates when flocks have different genetic means because no distinction can be made between genetic and environmental differences between flocks. In this case, direct and maternal heritabilities are under-estimated, whereas maternal environmental effects are overestimated. The insufficiency of pedigree leads to biased estimates of genetic parameters.

**genetic parameters / animal model / maternal effects / simulations / connectedness**

\* Correspondence and reprints

E-mail: clement@germinal.toulouse.inra.fr

## 1. INTRODUCTION

The animal model is extensively used for predicting genetic values and estimating genetic parameters, because the optimum combined use of all relationships and performances improves accuracy. However, despite the theoretical advantages of this model, some data and model conditions can affect the validity and precision of the estimation of variance components.

The first source of bias lies in the choice of the genetic model used to analyse data. Concerning maternally influenced traits, there is still discrepancy between the theoretical studies about genetic parameter estimation and practical applications. The reasons for this can be problems of convergence with variance components estimation software, or data structure (for example incomplete pedigree), or unavailability of efficient techniques (software or hardware) as is the case in some developing countries. When traits are governed by both direct and maternal effects, fitting only direct effects leads to an overestimation of direct heritability. For growth traits, most of the estimations of direct heritability with both direct and maternal effects vary between 0.20 and 0.30 [30,38,47]. When maternal effects are ignored, direct heritabilities published can reach 0.73 for daily gain before weaning, [23], 0.48 or 0.50 for birth weight [29], 0.35 for four-month weight [27], 0.56 for weights before weaning [6] or 0.45 for weaning weight [7]. However, the relative part of direct and maternal effects (genetic or environmental) and the nature and magnitude of the relation between these effects are determining conditions for the effectiveness of a selection scheme. Literature on the influence of model adequacy in order to estimate variance components is limited. There are some publications in which various models were tested in order to find the most adapted to analyse data. For example, simulations were used to study biometrical aspects of direct and maternal effects [41,43]. Meyer [33] studied the precision of genetic parameter estimation with different family structures. Robinson [41] and Lee and Pollak [28] tested the sire  $\times$  year variation on the genetic correlation between direct and maternal effects. Quintanilla Aguado [39] studied the importance of the models on maternal effects analysis by fitting an environmental correlation between the dam and the offspring. These previous publications reported biases when using incorrect models. In this article, we quantify this bias for different values of true genetic parameters.

Data structure is the second source of bias likely to affect the estimation of variance components. In traditional farming systems, it is sometimes difficult to identify animals and to record performances and/or genealogy. The amount and the quality of the data are then affected by practical constraints. Although this is often the case in developing countries, this can also concern industrialised countries, in particular as regards hardy breeds managed in large flocks with several males used simultaneously for natural service. One of the consequences

can be the use of a very incomplete pedigree resulting in a less thorough relationship matrix used in the animal model. Moreover, the lack of artificial insemination and a poor exchange of sires across breeding units limit gene flow and cause a partial or complete lack of genetic connectedness. Even in selection schemes under intensive breeding conditions, disconnectedness can be a problem when prediction of genetic values is done on a national scale and artificial insemination is organised into regions, as is the case for instance for the Montbéliarde and Holstein cattle breeds in France [19,20] or in North-American breeds [3,24,44]. The effect of data structure has been extensively studied in the context of genetic evaluation of animals. Absence of connectedness and poor genealogical information are responsible for biases and loss of accuracy in the prediction of genetic values by an animal or sire model [1,21,44]. However, not much is known about the effect of data structure on the estimation of genetic parameters by an animal model, especially in the presence of maternal effects. Diaz *et al.* [10] and Eccleston [11] studied the influence of disconnectedness on models with direct effects and found that it would act only on the precision of the estimation. Now, to propose strategies for improvement, it is necessary to assess the relative importance of deviations from the ideal situation. The second purpose of this article is to test, by simulation, the influence of data structure on the estimation of genetic parameters for traits subject to direct and maternal effects.

## **2. MODEL ADEQUACY**

### **2.1. Data simulation**

#### **2.1.1. Simulated population**

The simulation program was written in Fortran and NAG Libraries were used for all random processes.

As model adequacy can be a real problem in populations under extensive conditions where data structure and unavailability of efficient techniques can be a constraint for the use of the correct model, we used a known African sheep population [12,13,35] to set some parameters of the simulated population (prolificacy, replacement rate, male/female ratio). Compared to the real population, the number of animals per flock was increased in order to avoid confusion between animal and flock effects. The base population consisted of 1 260 unrelated animals (60 males and 1 200 females) assigned randomly to 20 flocks of 63 animals each (3 males and 60 females). Once the base population was created, the simulation was carried out over 6 years. Each year, random mating (no matter what flock animals came from) was practised with a ratio of one male for twenty females. The offspring were generated according to a prolificacy of 115%. Each year, 1/3 of the males and 1/5 of the females were replaced by offspring at random. The remaining offspring was discarded

so that the number of animals per flock and the number of flocks were constant over time. The average number of offspring per female was equal to 2.7. The data set corresponds to a fully connected population with complete pedigree.

### 2.1.2. Models used for simulating data

The simulated models were similar to those used in Robinson's study [41], with  $A$  representing the genetic direct effects,  $M$  the genetic maternal effects,  $R$  the genetic correlation between direct and maternal effects, and  $C$  the maternal environmental effects. Some authors (Hohenboken and Brinks [22], Koch [25], Foulley and Ménéssier [17] and Cantet [8]) have shown that a more complex biological model could exist, this model including a non genetic correlation between maternal effects of dams and daughters. Several biometrical models have been proposed to consider this correlation [8,40,41]. We could have used this model in our simulations, but we wanted to limit this work to the models most frequently used for the study of maternal effects. The models and the corresponding (co)variances are presented in Table I.

For the base population (which represents founder parents), random effects were sampled from normal distributions with zero mean and variances corresponding to each random effect. Direct genetic value  $A_i$  for individuals  $i$  was simulated in a distribution  $N(0, \sigma_{Ao})$  and maternal genetic value  $M_i$  for individuals  $i$  was simulated using:

$$M_i = r_{AoAm} \times (\sigma_{Am}/\sigma_{Ao}) \times A_i + \sqrt{(1 - r_{AoAm}^2)} \times Q_i \times \sigma_{Am}$$

where  $r_{AoAm}$  is the genetic correlation between direct and maternal effects and  $Q_i$  is a random variable sampled from a standard normal distribution  $N(0, 1)$ . Since dams were unknown for these animals, when the simulated model included maternal effects, no record was generated for this base population.

In real data, the distribution of flock effects was close to a normal distribution. We then used a random variable distributed according to  $N(0, \sigma_f^2)$  to generate this  $t_k$  effect for flock  $k$  which was considered as fixed in the variance component estimation model.

Over the successive years, genetic effects of the offspring were calculated as the mid-parent values, plus a Mendelian deviation, calculated following the formula [15]:

$$W_i(o) = R_i \sqrt{\frac{1}{2} \left( 1 - \frac{F_p + F_m}{2} \right)} \sigma_{Ao} \quad (1)$$

$$\begin{aligned} W_i(m) &= r_{AoAm} \times (\sigma_{Am}/\sigma_{Ao}) W_i(o) + \sqrt{1 - r_{AoAm}^2} \\ &\times R'_i \times \sqrt{\frac{1}{2} \left( 1 - \frac{F_p + F_m}{2} \right)} \sigma_{Am} \end{aligned} \quad (2)$$

**Table I.** Models used to simulate and analyse data.

Simulation models	(Co)variances fitted
A	$\sigma_e^2 \sigma_{Ao}^2$
AMR0	$\sigma_e^2 \sigma_{Ao}^2 \sigma_{Am}^2$
AMRi*	$\sigma_e^2 \sigma_{Ao}^2 \sigma_{Am}^2 \sigma_{AoAm}$
AMR0C	$\sigma_e^2 \sigma_{Ao}^2 \sigma_{Am}^2 \sigma_C^2$
AMRi*C	$\sigma_e^2 \sigma_{Ao}^2 \sigma_{Am}^2 \sigma_{AoAm} \sigma_C^2$
Analysis models	(Co)variances estimated
A	$\sigma_e^2 \sigma_{Ao}^2$
AMR	$\sigma_e^2 \sigma_{Ao}^2 \sigma_{Am}^2 \sigma_{AoAm}$
AC	$\sigma_e^2 \sigma_{Ao}^2 \sigma_C^2$
AMRC	$\sigma_e^2 \sigma_{Ao}^2 \sigma_{Am}^2 \sigma_{AoAm} \sigma_C^2$

Simulation models:

A: model with direct genetic effects; AMR0: model with uncorrelated direct and maternal genetic effects; AMRi: model with correlated direct and maternal genetic effects; AMR0C: model with uncorrelated direct and maternal genetic effects, and maternal environmental effects; AMRiC: model with correlated direct and maternal genetic effects, and maternal environmental effects;

\* $Ri = \sigma_{AoAm} / \sigma_{Ao} \sigma_{Am}$ ,  $Ri = -0.25$  or  $-0.50$ .

Analysis models:

A: model with direct genetic effects; AC: model with direct genetic effects and maternal environmental effects; AMR: model with direct and maternal genetic effects; AMRC: model with direct genetic effects, maternal genetic effects and maternal environmental effects.

where  $W_i(o)$  and  $W_i(m)$  are Mendelian deviations of the offspring ( $i$ ) for the direct effect ( $o$ ) and the maternal effect ( $m$ ), respectively;  $R_i$  and  $R'_i$  are independent random variables sampled from a standard normal distribution  $N(0, 1)$ ;  $F_p$  and  $F_m$  are coefficients of inbreeding of the sire ( $F_p$ ) and the dam ( $F_m$ ), respectively. The calculation of inbreeding coefficients was made using the algorithm proposed by Meuwissen and Luo [32]. Residual effects were simulated for offspring according to  $N(0, \sigma_e^2)$  for direct effects and  $N(0, \sigma_C^2)$  for maternal environmental effect. Residuals corresponding to records of dams were independent from residuals corresponding to records of their progeny. Finally, a file of about 9 500 animals with a single record per animal (except for the base population) was obtained corresponding to six years of simulation.

**2.1.3. Values of parameters used in the simulation**

Two sets of genetic parameters values were used for the simulations. The first set (called population 1) was supposed to reflect genetic parameter values found

in the literature for growth traits in cattle and sheep of temperate climate [30, 38,47]: 0.20 for direct heritability ( $h_{Ao}^2$ ), 0.30 for maternal heritability ( $h_{Am}^2$ ) and 0.05 for the part of variance due to maternal environmental effects ( $c^2$ ). The second set (called population 2) was chosen to reflect what can be found in countries with high constraints. They were close to genetic parameters estimated on a Tunisian breed of sheep [5]: 0.05 for  $h_{Ao}^2$ , 0.10 for  $h_{Am}^2$  and 0.25 for  $c^2$ . The genetic correlation between direct and maternal effects ( $r_{AoAm}$ ) has often been found to be negative or equal to zero in cattle and sheep [2, 17,31,34]. Consequently, three values, 0,  $-0.25$  and  $-0.50$  were used for both populations. Seven simulation models were used for each population, model A including direct effects only, models AMR0, AMR25 and AMR50 including direct and maternal effects under the three alternative values of the genetic correlation, and models AMR0C, AMR25C and AMR50C which, in addition to the direct and maternal genetic effects, considered the maternal environmental effect. Values of variance components are presented in Table II.

Fifty replicates were made for each population and each of the seven models simulated. A distinct seed for the random number generator was set for each replicate. The same seed was used to simulate the genetic mean of flocks in order to limit the variability of samples.

## 2.2. Data analysis

The VCE program [37] was used to estimate genetic parameters by means of REML methodology. Four models were used for analysing the seven simulated data sets for each population. The first three models included direct effects only (model A), maternal and direct genetic effects (model AMR), maternal and direct genetic effects plus maternal environmental effects (model AMRC). In addition, a model including direct genetic and strictly environmental maternal effects (model AC) was used, and this fourth model assumes that maternal effect has no genetic component in the dam. These four analysis models, presented in Table I, were fitted to each of the seven data sets simulated under the genetic assumptions described above for populations 1 and 2. The average and the empirical standard deviation were calculated over the fifty replicates obtained for each model and each population.

## 2.3. Results and discussion

Results are shown in Tables III and IV for populations 1 and 2, respectively. Empirical standard deviations between replicates varied between 0.02 and 0.04 for heritabilities of direct and maternal effects. They were higher for the genetic correlation, particularly when true values tended to zero (AMR0, AMR0C) and when direct and maternal heritabilities were small (population 2).

**Table II.** Values of variance components and parameters used for simulation.

Variations, covariances and parameters	A	AMR0	AMR25	AMR50	AMR0C	AMR25C	AMR50C
Population 1							
$\sigma_p^2$	200.0	356.3	354.4	355.4	355.9	354.7	354.7
$h_{Ao}^2$	0.2	0.2	0.2	0.2	0.2	0.2	0.2
$h_{Am}^2$	0.3	0.3	0.3	0.3	0.3	0.3	0.3
$r_{AoAm}$	–	0	–0.25	–0.50	0	–0.25	–0.50
$\sigma_{AoAm}/\sigma_p^2$	–	–	–0.06	–0.12	–	–0.06	–0.12
$c^2$	–	–	–	–	0.05	0.05	0.05
Population 2							
$\sigma_p^2$	243.2	377.9	388.2	387.1	385.0	385.0	385.0
$h_{Ao}^2$	0.05	0.05	0.05	0.05	0.05	0.05	0.05
$h_{Am}^2$	0.10	0.10	0.10	0.10	0.10	0.10	0.10
$r_{AoAm}$	–	0	–0.25	–0.50	0	0	0
$\sigma_{AoAm}/\sigma_p^2$	–	–	–0.02	–0.04	–	–0.02	–0.04
$c^2$	–	–	–	–	0.25	0.25	0.25

A: model with direct genetic effects. AMR0: model with uncorrelated direct and maternal genetic effects. AMR25: model with correlated direct and maternal genetic effects ( $r_{AoAm} = -0.25$ ). AMR50: model with correlated direct and maternal genetic effects ( $r_{AoAm} = -0.50$ ). AMR0C: model with uncorrelated direct and maternal genetic effects, and maternal environmental effects. AMR25C: model with correlated direct and maternal genetic effects ( $r_{AoAm} = -0.25$ ), and maternal environmental effects. AMR50C: model with correlated direct and maternal genetic effects ( $r_{AoAm} = -0.50$ ), and maternal environmental effects.

$\sigma_p^2$ : phenotypic variance;  $h_{Ao}^2$ : direct heritability;  $h_{Am}^2$ : maternal heritability;  $r_{AoAm}$ : genetic correlation between direct and maternal effects;  $c^2$ : part of variance due to maternal environmental effects.

For both populations, average parameters estimated with the true model (same simulation and analysis models) were very close to true values.

### 2.3.1. Simulation model A (only direct effects)

When data simulated according to a direct effect model were analysed with a more complex model (models AMR or AMRC), the direct heritability was unbiased and maternal effects (genetic or environmental) were estimated as equal to zero. Genetic correlation could not be estimated, because the maternal genetic variance was equal to zero in most of the cases.

**Table III.** Model adequacy: estimation of genetic parameters for different simulation models and different analysis models for population 1. *(continued on the next page)*

Genetic Parameters	Simulation models						
	A	AMR0	AMR25	AMR50	AMR0C	AMR25C	AMR50C
True values:							
$h_{Ao}^2$	0.20	0.20	0.20	0.20	0.20	0.20	0.20
$h_{Am}^2$	–	0.30	0.30	0.30	0.30	0.30	0.30
$r_{AoAm}$	–	0	–0.25	–0.50	0	–0.25	–0.50
$c^2$	–	–	–	–	0.05	0.05	0.05
$\sigma_p^2$	200	356.3	354.4	355.4	355.9	354.7	354.7
Analysis Models:							
A	<b>0.20 ± 0.02</b>	0.42 ± 0.02	0.34 ± 0.02	0.25 ± 0.03	0.43 ± 0.02	0.35 ± 0.02	0.27 ± 0.02
AC	$h_{Ao}^2$	0.20 ± 0.02	0.29 ± 0.03	0.23 ± 0.02	0.16 ± 0.02	0.22 ± 0.02	0.16 ± 0.02
	$h_{Am}^2$	0.01 ± 0.01	0.24 ± 0.02	0.21 ± 0.02	0.18 ± 0.02	0.27 ± 0.02	0.23 ± 0.02
	$c^2$						
AMR	$h_{Ao}^2$	0.20 ± 0.03	<b>0.20 ± 0.03</b>	<b>0.20 ± 0.03</b>	<b>0.20 ± 0.03</b>	0.20 ± 0.03	0.21 ± 0.03
	$h_{Am}^2$	0.01 ± 0.01	<b>0.30 ± 0.02</b>	<b>0.30 ± 0.02</b>	<b>0.30 ± 0.03</b>	0.34 ± 0.04	0.36 ± 0.02
	$r_{AoAm}$	ne	<b>0.01 ± 0.11</b>	<b>–0.23 ± 0.08</b>	<b>–0.50 ± 0.06</b>	–0.27 ± 0.09	–0.50 ± 0.06
AMRC	$h_{Ao}^2$	0.21 ± 0.03	0.19 ± 0.02	0.20 ± 0.02	0.20 ± 0.03	<b>0.21 ± 0.02</b>	<b>0.20 ± 0.03</b>
	$h_{Am}^2$	0.01 ± 0.01	0.29 ± 0.03	0.30 ± 0.03	0.29 ± 0.02	<b>0.31 ± 0.03</b>	<b>0.30 ± 0.04</b>
	$r_{AoAm}$	ne	0.02 ± 0.09	–0.27 ± 0.08	–0.51 ± 0.06	<b>–0.24 ± 0.09</b>	<b>–0.51 ± 0.07</b>
	$c^2$	0.01 ± 0.01	0 ± 0.01	0.03 ± 0.02	0 ± 0.01	<b>0.05 ± 0.02</b>	<b>0.05 ± 0.03</b>



**Table III.** Continued.

$h_{Ao}^2$ : direct heritability;  $h_{Am}^2$ : maternal heritability;  $r_{AoAm}$ : genetic correlation between direct and maternal effects;  $c^2$ : part of variance due to maternal environmental effects;  $\sigma_p^2$ : phenotypic variance; ne: cannot be estimated; in bold: true model.

Simulation models:

A: model with direct genetic effects. AMR0: model with uncorrelated direct and maternal genetic effects. AMR25: model with correlated direct and maternal genetic effects ( $r_{AoAm} = -0.25$ ). AMR50: model with correlated direct and maternal genetic effects ( $r_{AoAm} = -0.50$ ). AMR0C: model with uncorrelated direct and maternal genetic effects, and maternal environmental effects. AMR25C: model with correlated direct and maternal genetic effects ( $r_{AoAm} = -0.25$ ), and maternal environmental effects. AMR50C: model with correlated direct and maternal genetic effects ( $r_{AoAm} = -0.50$ ), and maternal environmental effects.

Analysis models:

A: model with direct genetic effects. AC: model with direct genetic effects and maternal environmental effects. AMR: model with direct and maternal genetic effects. AMRC: model with direct genetic effects, maternal genetic effects and maternal environmental effects.

**Table IV.** Estimation of genetic parameters for different simulation models and different analysis models for population 2.  
(continued on the next page)

Genetic Parameters	A	AMR0	Simulation models				AMR50C	
			AMR25	AMR50	AMR0C	AMR25C		
True values								
$h_{Ao}^2$	0.05	0.05	0.05	0.05	0.05	0.05	0.05	
$h_{Am}^2$	–	0.10	0.10	0.10	0.10	0.10	0.10	
$r_{AoAm}$	–	0	–0.25	–0.50	0	–0.25	–0.50	
$c^2$	–	–	–	–	0.25	0.25	0.25	
$\sigma_p^2$	200	356.3	354.4	355.4	355.9	354.7	354.7	
Analysis Models:								
A	$h_{Ao}^2$	<b>0.05 ± 0.02</b>	0.13 ± 0.02	0.11 ± 0.02	0.07 ± 0.02	0.23 ± 0.02	0.18 ± 0.02	0.16 ± 0.03
AC	$h_{Ao}^2$	0.05 ± 0.02	0.09 ± 0.02	0.07 ± 0.02	0.05 ± 0.02	0.07 ± 0.02	0.05 ± 0.02	0.04 ± 0.04
	$c^2$	0.01 ± 0.01	0.09 ± 0.02	0.08 ± 0.02	0.07 ± 0.01	0.34 ± 0.01	0.31 ± 0.01	0.33 ± 0.01
AMR	$h_{Ao}^2$	0.05 ± 0.02	<b>0.05 ± 0.02</b>	<b>0.05 ± 0.02</b>	<b>0.05 ± 0.02</b>	0.06 ± 0.02	0.06 ± 0.02	0.06 ± 0.02
	$h_{Am}^2$	0.01 ± 0.01	<b>0.11 ± 0.02</b>	<b>0.11 ± 0.02</b>	<b>0.11 ± 0.03</b>	0.40 ± 0.02	0.38 ± 0.02	0.40 ± 0.02
	$r_{AoAm}$	n.e.	<b>0.03 ± 0.28</b>	<b>–0.24 ± 0.26</b>	<b>–0.51 ± 0.21</b>	–0.40 ± 0.08	–0.55 ± 0.14	–0.69 ± 0.12
AMRC	$h_{Ao}^2$	0.06 ± 0.02	0.05 ± 0.02	0.05 ± 0.02	0.05 ± 0.02	<b>0.05 ± 0.02</b>	<b>0.05 ± 0.02</b>	<b>0.05 ± 0.02</b>
	$h_{Am}^2$	0.01 ± 0.01	0.10 ± 0.02	0.11 ± 0.02	0.10 ± 0.02	<b>0.10 ± 0.03</b>	<b>0.10 ± 0.03</b>	<b>0.11 ± 0.03</b>
	$r_{AoAm}$	ne	0.01 ± 0.19	–0.24 ± 0.29	–0.51 ± 0.20	<b>0.04 ± 0.31</b>	<b>–0.26 ± 0.24</b>	<b>–0.50 ± 0.16</b>
	$c^2$	0 ± 0.01	0.01 ± 0.01	0.01 ± 0.01	0 ± 0.01	<b>0.25 ± 0.02</b>	<b>0.25 ± 0.03</b>	<b>0.25 ± 0.03</b>

**Table IV.** Continued.

$h_{Ao}^2$ : direct heritability;  $h_{Am}^2$ : maternal heritability;  $r_{AoAm}$ : genetic correlation between direct and maternal effects;  $c^2$ : part of variance due to maternal environmental effects;  $\sigma_p^2$ : phenotypic variance; ne: cannot be estimated; in bold: true model.

Simulation models:

A: model with direct genetic effects. AMR0: model with uncorrelated direct and maternal genetic effects. AMR25: model with correlated direct and maternal genetic effects ( $r_{AoAm} = -0.25$ ). AMR50: model with correlated direct and maternal genetic effects ( $r_{AoAm} = -0.50$ ). AMR0C: model with uncorrelated direct and maternal genetic effects, and maternal environmental effects. AMR25C: model with correlated direct and maternal genetic effects ( $r_{AoAm} = -0.25$ ), and maternal environmental effects. AMR50C: model with correlated direct and maternal genetic effects ( $r_{AoAm} = -0.50$ ), and maternal environmental effects.

Analysis models:

A: model with direct genetic effects. AC: model with direct genetic effects and maternal environmental effects. AMR: model with direct and maternal genetic effects. AMRC: model with direct genetic effects, maternal genetic effects and maternal environmental effects.

### **2.3.2. Simulation models AMR0, AMR25, AMR50 (direct and maternal genetic effects)**

When the dam effect was neglected (analysis model A) on data simulated according to a model with direct and maternal genetic effects, the direct heritability was overestimated. Estimates of direct heritability could reach more than twice the true value when genetic correlation was equal to zero ( $\hat{h}_{Ao}^2 = 0.42$  for population 1 and 0.13 for population 2). The importance of the bias increased as the genetic correlation was reached zero. These results agree with those obtained by Waldron *et al.* [47] or Nasholm and Danell [36] on real data, and by Southwood *et al.* [43], Robinson [41] or Quintanilla Aguado [39] on simulated data. Results obtained for a selected population are similar [42]. When maternal effects are partially neglected, it is difficult, with an animal model, to distinguish between maternal effects and the contribution of the dam to the genotype of her offspring, the direct genetic variance being inflated by part of the genetic maternal variance. It seems that another part of maternal heritability is included in the residual variance.

When adding an environmental maternal effect (analysis model AC), results were closer to true values: estimated direct genetic and residual variances and the estimated direct heritability decreased, part of the overall variance being accounted for by the added maternal effect. The direct heritability was slightly overestimated ( $\hat{h}_{Ao}^2 = 0.23$  for population 1 and 0.07 for population 2) for the simulation model AMR25. For the simulation model AMR50, the direct heritability was equal to the true value for population 2 and slightly under-estimated ( $\hat{h}_{Ao}^2 = 0.16$ ) for population 1. The introduction of this non-genetic maternal effect allowed us to take into account a fraction of the genetic maternal effects, which in the previous model was included in the direct genetic and residual variances. However, and particularly for the first population, the estimated environmental maternal variance contained only a part of the genetic maternal variance. Accounting for non-genetic maternal effects does not compensate for the overall overestimation due to the maternal genetic effects being ignored.

With the introduction of both genetic and environmental maternal effects (analysis model AMRC which is an overparameterised model compared to the simulation model) estimates were similar to those estimated with the correct model.

### **2.3.3. Simulation models AMR0C, AMR25C, AMR50C (genetic direct effects, genetic and environmental maternal effects)**

For those more complex simulation models, the direct heritability was overestimated when using analysis model A. As compared to the simulation excluding maternal environmental effects (AMR0, AMR25, AMR50), this

overestimation was similar in population 1, much higher in population 2, because a part of the environmental maternal effect seems to be included in direct genetic variation: when an environmental effect was added into the analysis model (AC) for population 2, direct heritability was not overestimated anymore. As before, the bias found on direct heritability depended on the value of the genetic correlation: the overestimation was less important for a genetic correlation of  $-0.50$ , as if the existence of a negative correlation between direct and maternal effects partially compensates the bias. Hence, as for the cases AMR0, AMR25 and AMR50, we can expect that direct heritability will be even more biased for higher values of the genetic correlation. This is in agreement with the study of Waldron *et al.* [47] using real data: genetic correlation estimated with a model including correlated direct and maternal effects varied between 0.09 and 0.30; with a model excluding maternal genetic effects, direct heritability was 1.3 to 3 times higher than the heritability estimated with the full model. Meyer [33] showed that there is a strong negative correlation (from  $-0.9$  to  $-0.6$ ) between genetic maternal variance and direct-maternal covariance estimators. This result shows that each modification of one of the components leads to a variation of the second one in the opposite direction which could explain why the gap between the true heritability and the estimated direct heritability increases when genetic correlation tends to positive values.

With the analysis model AMR, for the first population, the direct heritability and the genetic correlation were correctly estimated, but the maternal heritability was overestimated ( $\hat{h}_{Am}^2 = 0.36$ ). For the second population, only the direct heritability was correctly estimated. The estimated maternal heritability was four times higher than the true value and the genetic correlation was very negative ( $r_{AoAm} = -0.40$ ). The suppression of the common environment effect due to the dam acted on estimated genetic maternal effects by increasing them above their true value, irrespective of the true value of the genetic correlation. In fact, genetic maternal and environmental maternal effects are confounded depending upon the relationship among mothers. The value of maternal heritability estimated by this reduced model corresponded approximately to the maternal heritability increased by the  $c^2$  value. Moreover, when environmental maternal effects were important, as in population 2, the high increase of maternal heritability was compensated by a decrease of genetic correlation. These results agree with those of Waldron *et al.* [47] and Meyer [34] in cattle and those of Koerhuis and Thompson [26] in broiler chickens. These authors observed a decrease of the maternal heritability on growth traits from an equivalent value of environmental maternal effects, when the latter are included in the model. A strong negative correlation between genetic and environmental maternal variance estimators helps to understand this result.

Generally speaking, a reduced model (with one or several random effects omitted) led to a variable bias, up to more than 50% of the true value, arising

from a confusion between different variance components. On the contrary, fitting unnecessary random effects neither yielded biased estimates (genetic parameters relative to these effects being either equal to zero, or which cannot be estimated) nor substantial losses in the precision of estimates. This result is opposite to Meyer's [33] results who found an influence of unnecessary effects on the precision of variance component estimation.

It seems, as pointed out by Cantet *et al.* [9] that there has been an evolution of estimates for growth traits over the last fifteen years. Estimates of direct heritability tend to increase, whereas those of maternal heritability decrease. The negative values of genetic correlation estimates are less negative than before. The overestimation of maternal heritability when environmental maternal effects are omitted should thus be related to the fact that estimates of maternal heritability obtained with older methods prior to the REML-animal model tend to be higher than those of more recent works. Indeed, the dam-offspring relationship, which was widely used to estimate genetic parameters before the use of the animal model contains a common environment effect provided by the dam to her offspring and leads, when this effect is neglected, to an overestimation of the corresponding parameter [16, 17, 25]. The variance estimated by the dam-offspring relationship also contains a dominance covariance between direct and maternal effects, which can be another source of bias.

### 3. DATA STRUCTURE

#### 3.1. Data simulation

Effects of data structure on the quality of parameter estimation were studied considering characteristics of the population as simulated in the previous part, the only differences being relative to level of connectedness and percentage of sires known. The model used for simulating data included direct effects, genetic and environmental maternal effects. To simplify the model, we assumed that direct and maternal effects were uncorrelated (model AMR0C of the precedent part).

Four levels of connectedness among flocks were simulated (connected, disconnected, 1 link sire, 2 link sires), which corresponded to four types of mating. In each situation, there was a ratio of one male for twenty females, and mating occurred randomly. Whatever the design, the offspring stayed in their dam's flock, so connections, when occurring, were only ensured by link sires. By and large, a part of the males and females were randomly mated within-flock (within-flock mating), the other part being randomly mated with reproducers from other flocks (between-flock mating). The number of animals of each type of mating (within- or between-flock) is presented in Table V. In the connected design, the between-flock mating represented 100% of the mating.

In the 2 link sires design, 2 males per flock contributed to connectedness by between-flock mating, the other mating being within flocks. In the 1 link sire design, between-flock mating concerned only 1 male and 20 females per flock. In the disconnected design, mating occurred within flock only. The coefficient of connectedness  $\gamma$  proposed by Foulley *et al.* [14] was applied in this paper. Nevertheless, in the same way as Hanocq and Boichard [20], this applied to measure connectedness between flocks instead of connectedness between sires. Thus with such an approach,  $\gamma$  can be defined as the ratio of the value of the prediction error variance of a contrast between two flocks using the full model to its value calculated under the reduced model. The full model involved both a fixed flock effect and a random sire effect. The reduced model, obtained by excluding the sire effect from the model, represented the optimal statistical situation from a connectedness viewpoint. The value of  $\gamma$  and the percentage of offspring coming from a link sire are presented in Table V. These levels of connectedness concern the direct effects: only the sires can have offspring in different flocks. Connectedness from a maternal viewpoint is only the consequence of connectedness at direct level ensured by sires.

Four pedigree structures were used to estimate genetic parameters: 100, 50, 20 or 10% of sires known. To simulate unknown paternity, a given percentage of sire identifications was randomly set to zero. Preliminary simulations on similar simulated populations (results not presented) have shown that below 10% of known sires, the number of different relationships was insufficient to estimate all variance components and estimated parameters depend on the starting value of the variance component estimation programme. With at least 10% of known sires, simulated and estimated parameters were equal.

Two situations were modelled. In the first one, the population was genetically homogeneous, whereas in the second one, flocks initially had different genetic means.

Connectedness level and knowledge of paternity corresponded to two different problems. Connectedness level was performed first by a mating plan assuming that all sires were known. When all relationships and performance were simulated, some sires were randomly sampled and supposed unknown for variance component estimation. Therefore, knowledge of paternity is an additional problem, independent of the population structure, but liable to hide the lack of connectedness.

### **3.1.1. Base population**

#### *Flocks with different genetic levels (cases GD10, GD20, GD30)*

When flocks had different genetic levels, the genetic variance could be divided into two parts, the within-flock genetic variance ( $\sigma_{Ao\text{ within}}^2$  for direct effects and  $\sigma_{Am\text{ within}}^2$  for maternal effects) and the between-flock genetic variance

**Table V.** Mating designs used to model different degrees of connectedness. Population simulated: 20 flocks, 60 males and 1 200 females.

Design	Connected	2 link sires	1 link sire	Disconnected
Within-flock mating	–	1 male per flock 20 females per flock	2 males per flock 40 females per flock	3 males per flock 60 females per flock
Between-flock mating	60 males 1 200 females	40 males 800 females	20 males 400 females	– 0
Percentage of offspring arising from a link sire	100	66	33	0
$\gamma$ coefficient Foulley <i>et al.</i> [12]	1	0.64	0.32	0



( $\sigma_{A_o \text{ between}}^2$  for direct effects and  $\sigma_{A_m \text{ between}}^2$  for maternal effects). Thus the (direct or maternal) genetic value of an animal of the base population was the sum of its genetic value in the population, plus the genetic mean of the flock in which it was born. No record was performed for animals of the base population because dams are unknown.

The same genetic means were used for all replicates in order to limit the variability of the sample, and after checking that, in spite of the small number of flocks, the distribution was not too different from a normal distribution. For the relative part of within and between-flock variances, three cases were simulated: direct and maternal within-flock variances were equal to 90% (GD10), 80% (GD20) and 70% (GD30) of the total variance, and direct and maternal between-flock variances were equal to 10% (GD10), 20% (GD20) and 30% (GD30) of the total variance.

*Same genetic level for all flocks (case GD0)*

In this extreme situation, direct and maternal within-flock variances were equal to 100% of the total variance.

### **3.1.2. Subsequent years**

The genetic simulation model was

$$Y_{ijk} = \mu + A_i + M_j + C_j + t_k + E_{ijk}$$

where  $Y_{ijk}$  is the record of animal  $i$  with dam  $j$  in flock  $k$ ;

$\mu$  is the phenotypic mean of the population;

$A_i$  is direct genetic value of animal  $i$ ;

$M_j$  is the maternal genetic value of dam  $j$ ;

$C_j$  is the maternal environmental effect;

$t_k$  is the environmental effect of flock  $k$ ;

$E_{ijk}$  is the residual.

For offspring, residual effects were randomly generated and direct and maternal genetic values were equal to the average value of parents, plus the Mendelian deviation calculated using formulae (1) and (2), but the value used for variances  $\sigma_{A_o}^2$  and  $\sigma_{A_m}^2$  varied according to the mating system. The within-family genetic variance (due to meiosis) depends on the gene pool to which it is possible to refer in the base population [15,46]. For a fully connected system, one may consider all flocks as a single population, and consequently, the within-family variance was computed from the total genetic variance in the base population. In contrast, for a completely disconnected system, one may consider each flock as a separate sub-population, and the within-family variance was computed only from the within-flocks genetic variance in the base population. For a partially disconnected system (1 or 2 link sires), the situation

is intermediate between the previous two. Therefore, in this case, the within-family variance was computed by combining both within- and between-flock genetic variances according to the probabilities of gene origin of the considered parents.

A hundred replicates was run for each type of data set tested, with a distinct seed for the random number generator for each replicate. The same seed was used to simulate the genetic mean of flocks in order to limit the variability of samples.

Values of the variance components and parameters correspond to situation AMROC of population 1 (Tab. II).

### 3.2. Data analysis

Data were analysed using the model with direct and maternal genetic effects and environmental maternal effects (model AMRC). A fixed flock effect was fitted in all situations. The data sets studied corresponded to the four connectedness designs (connected, 1 or 2 link sires, unconnected), the four levels of pedigree information (100, 50, 20 and 10% of paternity known) and degree of genetic differences among flocks (GD0, GD10, GD20 and GD30).

The average and the empirical standard deviation over replicates are calculated for variance components:  $\sigma_E^2$ ,  $\sigma_{Ao}^2$ ,  $\sigma_{Am}^2$ ,  $\sigma_{AoAm}$  and  $\sigma_C^2$ .

### 3.3. Results and discussion

Variances and covariance are presented in Tables VI and VII.

#### 3.3.1. Precision of the estimates

As shown in Table VI, and in contrast to what has been observed by several authors for the estimation of genetic parameters [10, 11] or for the prediction of genetic value [1, 21, 45], no clear pattern of reduction of precision related to a lack of connectedness was observed in this study, but it is possible that the number of replicates was insufficient to see an effect of connectedness on the precision of the estimates. However, standard deviation between replicates of covariance increased when genetic difference between flocks became more important.

Regardless of the level of connectedness, the alteration of genealogical information, through a progressive elimination of paternity, acted on the precision of variances and covariance, as shown in Table VII. Including the complete pedigree *via* the relationship matrix allows for a better dissociation, on, one hand, of genetic and environmental effects, and on the other hand, genetic effects among the latter, and provides greater precision.

**Table VI.** Estimated parameters with different levels of connectedness for a situation with full pedigree information and same genetic levels between flocks (GD0) or for situations with a genetic difference between flocks equal to 10% (GD10), 20% (GD20) and 30% (GD30) of the overall genetic variance.

Variances		Genetic difference between flocks			
		GD0	GD10	GD20	GD30
Connected	$\sigma_{Ao}^2$	69.7 ± 9.0	68.9 ± 10.7	66.7 ± 9.8	67.6 ± 10.5
	$\sigma_{Am}^2$	105.1 ± 14.9	95.3 ± 12.5	91.5 ± 9.7	82.1 ± 11.0
	$\sigma_{AoAm}$	0.9 ± 1.7	0.2 ± 2.7	2.9 ± 5.2	-5.4 ± 9.2
	$\sigma_E^2$	161.3 ± 5.3	161.1 ± 7.1	163.0 ± 5.6	161.5 ± 6.3
	$\sigma_c^2$	18.1 ± 8.8	21.1 ± 7.9	22.8 ± 6.5	27.9 ± 8.3
	$\sigma_p^2$	355.0 ± 9.8	346.6 ± 4.2	341.1 ± 7.3	333.7 ± 2.5
2 link sires	$\sigma_{Ao}^2$	69.5 ± 9.8	67.7 ± 9.7	66.4 ± 11.1	59.5 ± 10.9
	$\sigma_{Am}^2$	101.4 ± 14.9	97.1 ± 11.9	85.1 ± 12.0	74.4 ± 11.9
	$\sigma_{AoAm}$	3.4 ± 1.3	-2.1 ± 8.6	-3.3 ± 8.9	-4.6 ± 11.8
	$\sigma_E^2$	161.1 ± 5.9	160.0 ± 6.6	157.0 ± 5.7	156.3 ± 6.0
	$\sigma_c^2$	18.6 ± 8.2	20.3 ± 6.3	22.7 ± 7.9	23.3 ± 7.1
	$\sigma_p^2$	353.9 ± 10.0	342.9 ± 7.8	328.0 ± 9.2	308.9 ± 7.3
1 link sire	$\sigma_{Ao}^2$	69.6 ± 10.1	65.8 ± 9.8	64.4 ± 8.9	59.1 ± 7.5
	$\sigma_{Am}^2$	99.5 ± 16.7	93.6 ± 12.1	82.4 ± 11.1	76.2 ± 10.5
	$\sigma_{AoAm}$	4.5 ± 1.3	-0.2 ± 2.1	-2.5 ± 7.2	-3.8 ± 9.0
	$\sigma_E^2$	161.2 ± 6.3	158.3 ± 5.9	159.0 ± 4.8	158.5 ± 5.4
	$\sigma_c^2$	20.7 ± 9.9	19.9 ± 8.0	24.4 ± 8.3	25.2 ± 7.6
	$\sigma_p^2$	355.5 ± 1.9	337.3 ± 3.7	327.7 ± 6.1	315.2 ± 3.7
Disconnected	$\sigma_{Ao}^2$	69.3 ± 11.7	65.5 ± 9.7	54.6 ± 10.9	48.3 ± 9.8
	$\sigma_{Am}^2$	107.9 ± 13.5	94.2 ± 13.5	83.3 ± 12.8	75.9 ± 11.3
	$\sigma_{AoAm}$	-1.3 ± 1.3	0.1 ± 4.8	0.8 ± 8.0	-0.4 ± 14.1
	$\sigma_E^2$	159.7 ± 6.0	159.6 ± 5.5	161.2 ± 7.0	161.1 ± 6.1
	$\sigma_c^2$	18.1 ± 7.3	18.4 ± 8.2	18.2 ± 8.1	16.8 ± 5.4
	$\sigma_p^2$	353.8 ± 2.2	337.0 ± 3.7	318.1 ± 5.6	301.8 ± 9.0

$\sigma_{Ao}^2$ : direct genetic variance;  $\sigma_{Am}^2$ : genetic maternal variance;  $\sigma_{AoAm}$ : genetic covariance between direct and maternal effects;  $\sigma_c^2$ : part of variance due to maternal environmental effects;  $\sigma_p^2$ : phenotypic variance.

GD0: no genetic difference between flocks. GD10: genetic difference between flocks equal to 10% of the overall genetic variance. GD20: genetic difference between flocks equal to 20% of the overall genetic variance. GD30: genetic difference between flocks equal to 30% of the overall genetic variance.

**Table VII.** Estimated parameters with different percentages of known paternity for different situations of connectedness and a genetic difference between flocks equal to 30% of overall genetic variance (GD30).

Genetic Parameters		Levels of connectedness					
		Connectedness	2 link sites	1 link site	Disconnectedness		
Known paternity	100% paternity	$\sigma_{Ao}^2$	67.6 ± 10.5	59.5 ± 10.9	59.1 ± 7.5	48.3 ± 9.8	
		$\sigma_{Am}^2$	82.1 ± 11.0	74.4 ± 11.9	76.2 ± 10.5	75.9 ± 11.3	
		$\sigma_{AoAm}$	-5.4 ± 2.7	-4.6 ± 3.4	-3.8 ± 2.4	-0.4 ± 3.9	
		$\sigma_e^2$	161.5 ± 6.3	156.3 ± 6.0	158.5 ± 5.4	161.1 ± 6.1	
		$\sigma_c^2$	27.9 ± 8.3	23.3 ± 7.1	25.2 ± 7.6	16.8 ± 5.4	
		$\sigma_p^2$	333.7 ± 2.5	308.9 ± 7.3	315.2 ± 3.7	301.8 ± 9.0	
		50% paternity	$\sigma_{Ao}^2$	64.0 ± 11.3	56.6 ± 11.9	49.4 ± 11.3	35.7 ± 8.9
			$\sigma_{Am}^2$	80.3 ± 13.2	77.6 ± 14.0	75.0 ± 13.9	73.6 ± 14.0
			$\sigma_{AoAm}$	-0.9 ± 4.9	-2.4 ± 6.0	0.7 ± 2.9	5.4 ± 4.3
			$\sigma_e^2$	164.3 ± 7.8	160.4 ± 6.9	162.2 ± 6.9	166.6 ± 6.5
			$\sigma_c^2$	26.6 ± 7.2	25.8 ± 8.2	21.2 ± 8.7	17.5 ± 8.1
			$\sigma_p^2$	334.3 ± 4.9	318.1 ± 13.1	308.4 ± 8.5	298.9 ± 2.4
		20% paternity	$\sigma_{Ao}^2$	63.5 ± 18.6	54.2 ± 13.6	43.5 ± 15.4	35.8 ± 12.4
			$\sigma_{Am}^2$	76.1 ± 17.2	75.8 ± 14.6	72.5 ± 14.9	72.8 ± 13.2
			$\sigma_{AoAm}$	0.3 ± 8.2	-1.2 ± 12.6	2.1 ± 7.2	6.3 ± 7.1
			$\sigma_e^2$	163.8 ± 13.5	161.9 ± 9.8	164.9 ± 10.0	166.0 ± 9.5
			$\sigma_c^2$	29.9 ± 10.0	26.6 ± 9.1	24.2 ± 9.9	17.0 ± 8.1
			$\sigma_p^2$	336.7 ± 7.9	317.3 ± 8.9	307.3 ± 8.5	298.0 ± 3.8
		10% paternity	$\sigma_{Ao}^2$	61.8 ± 21.6	54.1 ± 22.8	43.6 ± 20.8	28.4 ± 15.3
			$\sigma_{Am}^2$	74.9 ± 13.1	77.2 ± 15.0	71.4 ± 14.7	69.8 ± 15.0
			$\sigma_{AoAm}$	2.7 ± 11.9	-1.7 ± 11.8	2.8 ± 10.0	8.5 ± 13.1
			$\sigma_e^2$	164.9 ± 15.9	161.9 ± 17.5	165.0 ± 15.0	172.0 ± 11.5
			$\sigma_c^2$	31.3 ± 9.7	26.2 ± 9.2	23.8 ± 9.2	18.6 ± 9.1
			$\sigma_p^2$	335.4 ± 3.1	317.7 ± 4.9	306.6 ± 12.4	297.2 ± 7.3

$\sigma_{Ao}^2$ : direct genetic variance;  $\sigma_{Am}^2$ : genetic maternal variance;  $\sigma_{AoAm}$ : genetic covariance between direct and maternal effect;  $\sigma_c^2$ : part of variance due to maternal environmental effects;  $\sigma_p^2$ : phenotypic variance.

GD0: no genetic difference between flocks. GD10: genetic difference between flocks equal to 10% of the overall genetic variance. GD20: genetic difference between flocks equal to 20% of the overall genetic variance. GD30: genetic difference between flocks equal to 30% of the overall genetic variance.

### **3.3.2. Bias of the estimates**

#### *Influence of disconnectedness and genetic difference between flocks*

With all sires known, and no initial genetic difference between flocks and the connected design (Tab. VI), the means over replicates of estimated variance components were close to true values. A decrease of gene flow across flocks (1 or 2 link sire designs) had no effect on direct genetic variance estimates, but led to a decrease of estimated maternal genetic variance. In the GD0 case, estimated environmental maternal and residual effects remained stable whatever the connectedness level was.

In the connected design, when genetic difference between flocks increased (GD10, GD20, GD30) estimated direct and maternal genetic variances decreased, but the trend was much more marked for maternal variance which went down to 77% of the true value for a genetic difference of 30%, whereas direct variance was equal to 95% of the true value. In parallel, estimated maternal environmental variance increased when genetic difference between flocks became more important.

Observed results depended on the connectedness level but also on the replacement rate of the females (each year, 1/5 of the females was replaced). At the beginning of the simulation, the direct and maternal genetic variances corresponded to within-flock genetic variance. When no connection was generated, the genetic variance estimated remained equal to within-flock genetic variance. This is what is observed in the disconnected design: estimated direct and maternal genetic variances for the cases GD10, GD20 and GD30 were equal to approximately 10, 20 and 30% of the true value, respectively. We can conclude that part of the genetic variability was eliminated with flock effect, so for the disconnected situation, the totality of the genetic variability between flocks disappeared. In the three other cases, the situation was different. Progressively, the establishment of connections across flocks took place by way of link sires which allowed estimability of genetic differences between flocks and to take them into account in the overall genetic variance estimation. This latter increased and got closer to the true value. According to Kennedy and Trus [24], relationships across flocks make possible to reduce the sampling error for genetic difference between flocks, by adding a sampling positive covariance between them. When no link sires exist, neither the between-flock genetic variance nor the environmental variance between flocks are accounted for and, with such a data structure, the animal model is not able to dissociate both components [24]. Thus across flock relationships, in order to restore connection, is required to separate correctly fixed and random effects and to estimate between-flock genetic variability. As connectedness was ensured by sires, the gap between true and estimated direct genetic variance decreased rapidly as simulation progressed. However, our simulation demonstrated that

when genetic difference between flocks was high (GD30), the genetic connectedness achieved during six years was insufficient for a correct estimation of the variance components. Other results (not presented) showed that with a larger number of years of simulation, estimated values were closer to true values (for example, when the simulation was conducted during three more years, direct and maternal genetic variances were equal to 69.9 and 88.9, respectively, in the connected situation for the GD30 case).

Concerning maternal genetic variance, more time is required to ensure that simulated connected design is efficient. For example, at the end of the simulation, the maternal grand-dams had still passed on 78% of the gene of their flock and the dams, 39%. Therefore, six years were not sufficient to take into account the overall genetic variability, which was still close to within-flock genetic variance.

#### *Influence of percentage of knowledge of paternity and disconnectedness*

Because of the similarity of the results between the different cases tested, only the GD30 situation is described (Tab. VII) for different levels of connectedness and different percentages of sires known. In the connected population, with all sires known, both direct and maternal genetic variances were underestimated, whereas maternal environmental variance was overestimated. When connectedness was incomplete, estimated genetic variances still decreased, for the disconnected design, down to 68% of the true value for direct genetic variance, and to 71% of the true value for genetic maternal variance, with part of the genetic variability being eliminated with flock effect. Elimination of a part of paternity accentuated the under-estimation of direct and maternal variances: when only 10% of sires were known, in the disconnected population, estimated direct genetic variance was equal to only 40% of the true value. Thus disconnecting the system and discarding sires correspond to two different mechanisms (one of which is due to population structure, the other to data recording), but lead to the same result in terms of variance component estimation. Recording an incomplete pedigree can mask a connectedness problem. Whatever the percentage of known sires, when the connectedness level decreased, estimated covariance between direct and maternal genetic effects increased. Estimated maternal environmental variance was higher than its true value, except for the disconnected designs where this component was unbiased.

It seems that with the disconnectedness situation and incomplete sire identification, some additional variability is attributed to the dam-offspring genetic covariance and direct and maternal heritabilities are under-estimated. These results are in accordance with Gerstmayr's observations [18]: when one of the heritability estimates (direct or maternal) increases, the dam-offspring genetic correlation decreases, and inversely.

The optimal situation for which the flocks all have the same genetic level is probably rather rare, in particular when the animals can have various genetic origins. The choice of the threshold beyond which the number of known sires will become sufficient to obtain an unbiased and relatively precise estimate depends on the data structure of the studied population. For extensive systems, one reliable solution could be the use of DNA markers. The conditions for successful DNA fingerprinting depend not only on the cost of the method, but also on the breeding system of the animals (number of males, numbers of animals per flock, grazing area, etc.). Barnett [4] showed that, recently, the cost generated when applying these methods to determine maternal pedigree in Australian flocks of Merino sheep for the prediction of genetic values, was higher than the return from extra productivity.

#### 4. CONCLUSION

The aim of this article was to study some of the genetic parameter estimation difficulties. The adequation of the estimation model is of particular interest since a reduced model leads to biased estimates. The importance of the bias depends on the true values of genetic parameters. When maternal heritability is low, exclusion of the dam effect does not affect the estimates much. However, estimated direct heritability can be more than doubled when maternal effects with a high influence on the trait are ignored. The bias was accentuated when the true genetic correlation was close to zero. When maternal effects with an environmental origin have a low influence on the trait, as it has been found in literature concerning temperate climates, the consequences on the estimates are only minor. However, when part of the variance due to maternal environmental effect reached 0.20, estimates of the other parameters were biased.

Data structure can affect the precision of variance components. Insufficient paternal genealogical information increases empirical standard deviations between replicates of the estimates, whereas insufficient genetic connectedness do not seem to act on the precision of the estimates. Data structure can also affect the unbiasedness of variance components: estimations are biased by the absence of genetic connection and unknown paternity. The fact that flocks have different genetic means highly accentuates the bias, so the use of link sires to establish connections is a major concern. Finally, extreme cases where sires are totally unknown or with no genetic connection between flocks, can make parameters non-estimable.

However, while data structure and analysis model affect the quality of the estimation, some situations will not be greatly affected. A bias of 10% for example will not be a problem for estimating genetic parameters, while it will have serious effects on the prediction of genetic values in a selection scheme.

It might be useful, under such conditions to consider the application of DNA fingerprinting for pedigree determination.

Even if the animal model has the capacity to thoroughly describe genealogical relationships, the analysis of a variance model and the structure of the animal population must be carefully controlled to get correct variance component estimations. In conclusion, the animal model is able to correctly dissociate variance components provided that all the necessary information is available.

These results were obtained in simplified simulation conditions. With real data, the problem becomes more complex and there are several additional causes of bias – due in particular to incorrect definition of the model – which are likely to interact. The statistical model can be inadequate for example if (co)variances are not well-described, as in the case of discrete traits or heterogeneous variances. A good definition of the biological model is important because direct and maternal effects can interact depending on environmental conditions. Finally, the genetic model can be different or more complex than the models used in the present experiment. This is the case, for example when trait expression is governed by a limited number of genes, or in a mixed inheritance situation or when dominance and epistatic effects are present.

## REFERENCES

- [1] Analla M., Sanchez-Palma A., Munoz-Serrano A., Serradilla J.M., Simulation analysis with BLUP methodology of different data structures in goat selection schemes in Spain, *Small Rumin. Res.* 17 (1995) 51–55.
- [2] Baker R.L., The role of maternal effects on the efficiency of selection in beef cattle. A review, *Proc. N. Z. Soc. Anim. Prod.* 40 (1980) 285–303.
- [3] Banos G., Schaeffer L.R., Burnside B., Genetic relationships and linear model comparisons between United States and Canada Ayrshire and Jersey bull populations, *J. Dairy. Sci.* 74 (1991) 1060–1068.
- [4] Barnett N., Optimising the use of pedigree information in Merino sheep improvement programs. Ph. D. thesis, University of New England, 1998.
- [5] Bedhiaf S., Bouix J., Clément V., Bibé B., François D., Importance du choix du modèle d'analyse dans l'estimation des paramètres génétiques de la croissance des ovins à viande en Tunisie, *Renc. Rech. Ruminants* 7 (2000) 169–172.
- [6] Ben Gara A., Rouissi H., Jurado J.J., Bodin L., Gabina D., Boujenane I., Mavrogenis A.P., Djemali M., Serradilla J.M., Étude de la simplification et de la standardisation du protocole de pesées chez les ovins à viande, *Options Méditerranéennes, FAO-CIHEAM-INRA* 33 (1997) 11–34.
- [7] Brash L.D., Fogarty N.M., Gilmour A.R., Genetic parameters for Australian maternal and dual-purpose meatsheep breeds. III. Liveweight, fat depth and wool production in Coopworth sheep, *Austr. J. Agric. Res.* 45 (1994) 481–486.
- [8] Cantet R.J.C., Kress D.D., Anderson D.C., Doornbos D.E., Burfening P.J., Blackwell R.L., Direct and maternal variances and covariances and maternal



- phenotypic effects on preweaning growth on beef cattle, *J. Anim. Sci.* 66 (1988) 648–660.
- [9] Cantet R.J.C., Gianola D., Misztal I., Fernando R.L., Estimates of dispersion parameters and of genetic and environmental trends for weaning weight in Angus cattle using a maternal animal model with genetic grouping, *Livest. Prod. Sci.* 34 (1993) 203–212.
- [10] Diaz C., Carabano M.J., Hernandez D., Connectedness in genetic parameters estimation and BV prediction, in: 46th Annual Meeting of the European Association for Animal Production, 4–8 September 1995, Prague, EAAP.
- [11] Eccleston J.A., Variance components and disconnected data, *Biometrics* 34 (1978) 479–481.
- [12] Faugère O., Dockes A.C., Perrot C., Faugère B., L'élevage traditionnel des petits ruminants au Sénégal. I. Pratiques de conduite et d'exploitation des animaux chez les éleveurs de la région de Kolda, *Rev. Élev. Méd. Vét. Pays Trop.* 43 (1990) 249–259.
- [13] Faugère O., Dockes A.C., Perrot C., Faugère B., L'élevage traditionnel des petits ruminants au Sénégal. II. Pratiques de conduite et d'exploitation des animaux chez les éleveurs de la région de Louga, *Rev. Élev. Méd. Vét. Pays Trop.* 43 (1990) 261–273.
- [14] Foulley J.L., Bouix J., Goffinet B., Elsen J.M., Connectedness in genetic evaluation, in: Gianola D., Hammond K. (Eds.), *Advances in statistical methods for genetic improvement of livestock*, Springer-Verlag, Berlin, 1990, pp. 277–308.
- [15] Foulley J.L., Chevalet C., Méthode de prise en compte de la consanguinité dans un modèle simple de simulation de performances, *Ann. Génét. Sél. Anim.* 13 (1981) 189–196.
- [16] Foulley J.L., Lefort G., Méthodes d'estimation des effets directs et maternels en sélection animale, *Ann. Génét. Sél. Anim.* 10 (1978) 475–496.
- [17] Foulley J.L., Méniissier F., Variabilité génétique des caractères de production des femelles Charolaises contrôlées en station. Résultats préliminaires, in: VI<sup>e</sup> journées d'information du « grenier de Theix ». L'exploitation des troupeaux de vaches allaitantes, *Suppl. Bull. Tech. CRZV Theix*, 1974, pp. 171–191.
- [18] Gerstmayr S., Impact of the data structure on the reliability of the estimated genetic parameters in an animal model with maternal effects, *J. Anim. Breed. Genet.*, 109 (1992) 321–336.
- [19] Hanocq E., Étude de la connexion en sélection animale. Thèse de doctorat, Institut national agronomique Paris-Grignon, 1995.
- [20] Hanocq E., Boichard D., Connectedness in the French Holstein cattle population, *Genet. Sel. Evol.* 31 (1999) 163–176.
- [21] Hanocq E., Boichard D., Foulley J.L., A simulation study of the effect of connectedness on genetic trend, *Genet. Sel. Evol.* 28 (1996) 67–82.
- [22] Hohenboken W.D., Brinks J.S., Relationships between direct and maternal effects on growth in Herefords. II. Partitioning of covariance between relatives, *J. Anim. Sci.* 32 (1971) 26–34.
- [23] Jurado J.J., Alonso A., Alenda R., Selection response for growth in a Spanish Merino flock, *J. Anim. Sci.* 72 (1994) 1433–1440.
- [24] Kennedy B.W., Trus D., Considerations on genetic connectedness between management unit under an animal model, *J. Anim. Sci.* 71 (1993) 2341–2352.

- [25] Koch R.M., The role of maternal effects in animal breeding. VI. Maternal effects in beef cattle, *J. Anim. Sci.* 35 (1972) 1316–1323.
- [26] Koerhuis A.N.M., Thompson R., Models to estimate maternal effects for juvenile body weight in broiler chickens, *Genet. Sel. Evol.* 29 (1997) 225–249.
- [27] Kumar N., Reheja K.L., Genetic and phenotypic parameters of growth and reproduction traits in UAS strain of sheep estimated by multitrait animal model, *Indian J. Anim. Sci.* 63 (1993) 978–983.
- [28] Lee C., Pollak E.J., Relationship between sire  $\times$  year interactions and direct-maternal genetic correlation for weaning weight of Simmental cattle, *J. Anim. Sci.* 75 (1997) 68–75.
- [29] Maria G.A., Boldman K.G., van Vleck L.D., A note on heritability estimates for growth traits in male and female Romanov sheep, *Anim. Prod.* 57 (1993) 326–328.
- [30] Ménissier F., Comments on optimisation of cattle breeding scheme: beef breeds for suckling herds. A review, *Ann. Génét. Sél. Anim.* 8 (1976) 71–87.
- [31] Ménissier F., Sapa J., Poivey J.P., Les qualités maternelles des ruminants allaitants : exemple de facilités de vêlage et de l’allaitement, in: *Éléments de génétique quantitative et application aux populations animales*, INRA Prod. Anim. (Suppl.) (1992) 135–145.
- [32] Meuwissen T.H.E., Luo Z., Computing inbreeding coefficients in large populations, *Genet. Sel. Evol.* 24 (1992) 305–313.
- [33] Meyer K., Bias and sampling covariances of estimates of variance components due to maternal effects, *Genet. Sel. Evol.* 24 (1992) 487–509.
- [34] Meyer K., Variance components due to direct and maternal effects for growth traits of Australian beef cattle, *Livest. Prod. Sci.* 31 (1992) 179–204.
- [35] Moulin C.H., Faugère O., Faugère B., L’élevage traditionnel des petits ruminants au Sénégal. III. Pratiques de conduite et d’exploitation des animaux chez les éleveurs de la communauté rurale de Kaymor (Siné-Saloum, Sénégal), *Rev. Élev. Méd. Vét. Pays Trop.* 47 (1994) 223–234.
- [36] Nasholm A., Danell O., Genetic relationships of lamb weight, maternal ability, and mature ewe weight in Swedish Finewool Sheep, *J. Anim. Sci.* 74 (1996) 329–339.
- [37] Neumaier A., Groenveld E., Restricted maximum likelihood of covariances in sparse linear models, *Genet. Sel. Evol.* 30 (1998) 3–26.
- [38] Poivey J.P., Niare T., Cournut J., Bibé B., Étude des possibilités d’allègement du protocole des pesées des agneaux dans les élevages français, in: *38<sup>e</sup> Réunion Annuelle FEZ, 28 septembre–1<sup>er</sup> octobre 1987*, Lisbonne, EAAP.
- [39] Quintanilla Aguado R., Efectos maternos en el peso al destete del ganado vacuno de carne: implicaciones en la estimación de parámetros y la evaluación genética. Ph. D. thesis, Université de Barcelone, 1998.
- [40] Quintanilla R., Varona L., Pujol M.R., Piedrafitra J., Maternal animal model with correlation between maternal environmental effects of related dams, *J. Anim. Sci.* 77 (1999) 2904–2917.
- [41] Robinson D.L., Models which might explain negative correlations between direct and maternal genetic effects, *Livest. Prod. Sci.* 45 (1996) 111–122.
- [42] Roehe R., Kennedy B.W., The influence of maternal effects on accuracy of evaluation of litter size in swine, *J. Anim. Sci.* 71 (1993) 2353–2364.

- [43] Southwood O.I., Kennedy B.W., Meyer K., Gibson J.P., Estimation of additive maternal and cytoplasmic genetic variances in animal models, *J. Dairy Sci.* 72 (1972) 3006–3012.
- [44] Tong A.K.W., Kennedy B.W., Moxley J.E., Potential errors in sire evaluation from regional genetic differences, *J. Dairy Sci.* 63 (1980) 627–633.
- [45] Tosh J.J., Wilton J.W., Effects on data structure on variance of prediction error and accuracy of genetic evaluation, *J. Anim. Sci.* 72 (1994) 2568–2577.
- [46] Verrier E., Colleau J.J., Foulley J.L., Effect of mass selection on the within-family genetic variance in finite populations, *Theor. Appl. Genet.* 77 (1989) 142–148.
- [47] Waldron D.F., Morris C.A., Baker R.L., Johnson D.L., Maternal effects for growth traits in beef cattle, *Livest. Prod. Sci.* 34 (1993) 57–70.