Estimation of genetics parameters of egg production traits of laying hens by restricted maximum likelihood applied to a multiple-trait reduced animal model

B. Besbes, Vincent Ducrocq, Jean Louis J. L. Foulley, M. Protais, Alice Tavernier, Michèle Tixier-Boichard, Catherine C. Beaumont

HAL Id: hal-02714916
https://hal.inrae.fr/hal-02714916v1

Submitted on 1 Jun 2020

Original article

Estimation of genetic parameters of egg production traits of laying hens by restricted maximum likelihood applied to a multiple-trait reduced animal model

B Besbes[4], V Ducrocq[1*], JL Foulley[1], M Protais[4]

A Tavernier[1], M Tixier-Boichard[2], C Beaumont[3]

[1] *INRA, Station de Génétique Quantitative et Appliquée, Centre de Recherche de Jouy-en-Josas, 78352 Jouy-en-Josas Cedex;*
[2] *INRA, Laboratoire de Génétique Factorielle, 78352 Jouy-en-Josas Cedex;*
[3] *INRA, Station de Recherches Avicoles, 37380 Monnaie;*
[4] *Institut de Sélection Animale, Établissements de Le Foeil, 22800 Quintin, France*

Summary – Variance components for egg production traits (No of eggs produced between 19 and 26, 26 and 38 and 26 and 54 wk of age), egg characteristics (average egg weight at 2 different ages and egg density) and body weight of hens at 40 wk of age were estimated in two strains of a breeding company by univariate and multivariate Restricted Maximum Likelihood (REML) applied to a Reduced Animal Model (RAM). To allow tridiagonalization of the coefficient matrix when RAM is considered, the approach of Thompson and Meyer (1990) using an imaginary random effect with negative variance was implemented. Canonical transformation was also employed. REML estimation, carried out on 15 random samples of ≈ 7 000 recorded hens drawn from each of 2 data files corresponding to 2 different strains, showed a rather small genetic antagonism between the group of egg production traits and egg density on one hand and that of egg weights on the other hand. Weight of hens behaved differently in the 2 strains. It showed also that traits within these groups were positively correlated. Heritabilities obtained by univariate and multivariate analyses were very similar and were lower for egg production traits (from 0.09 to 0.27) than for egg characteristics or weight of hens (from 0.34 to 0.48).

egg production / genetic parameter / restricted maximum likelihood / animal model

* Correspondence and reprints

Résumé − **Estimation des paramètres génétiques des caractères de ponte chez la poule par maximum de vraisemblance restreinte appliqué à un modèle animal réduit multicaractères.** *La production d'œufs (nombre d'œufs produits entre 19 et 26, 26 et 38 et 26 et 54 semaines d'âge), les caractéristiques de l'œuf (le poids moyen des œufs à 2 âges différents et densité de l'œuf) ainsi que le poids des poules à 40 semaines d'âge ont été étudié dans deux souches d'une firme de sélection. Les composantes de la variance de ces caractères ont été estimées à l'aide du maximum de vraisemblance restreinte (REML) uni et multicaractères appliqué à un modèle animal réduit (RAM). Pour tridiagonaliser la matrice des coefficients et réduire, par conséquent, les calculs liés à son inversion à chaque itération de l'algorithme EM (Espérance−Maximisation), nous avons utilisé l'algorithme proposé par Thompson et Meyer (1990). Ce dernier introduit dans le modèle un effet aléatoire imaginaire supplémentaire, ayant une variance négative. Ces calculs ont été également réduits grâce à une décomposition canonique. Les estimées du REML, obtenues à partir de 15 échantillons d'environ 7 000 observations chacun, ont montré un léger antagonisme entre, d'une part, le groupe des caractères de pente et la densité de l'œuf et d'autre part celui du poids des œufs. Les corrélations entre les caractères intragroupe étaient positives. Il est apparu également que les caractères de production d'œufs étaient moins héritables (de 0,09 à 0,27) que ceux liés aux caractéristiques de l'œuf ou de la poule (de 0,34 à 0,48), dans les souches considérées. Enfin, nous signalons que la comparaison entre les héritabilités obtenues avec les analyses uni- et multicaractères n'a pas montré de différences nettes.*

caractère de ponte / paramètre génétique / maximum de vraisemblance restreinte / modèle animal

INTRODUCTION

Best Linear Unbiased Prediction (BLUP) applied to an Animal Model has been recognized as the method of choice for estimating the genetic merit of candidates for selection. Under normality and in the absence of prior knowledge about means and variances, breeding values should be predicted using BLUP methodology, with the unknown variances replaced by their corresponding Restricted Maximum Likelihood (REML; Patterson and Thompson, 1971) estimates (Gianola *et al*, 1986).

REML is preferred to other variance components estimation methods because of its ability to account for selection bias. These methods (BLUP and REML) are nowadays utilized in many countries all over the world and for various domestic species. Surprisingly, they are almost completely ignored in laying hens evaluation systems even though strong selection has been carried out on this species for many generations.

The purposes of this study were: 1) to estimate genetic parameters of 7 correlated egg production traits by REML applied to a multiple-trait reduced animal model; and 2) to show the application of some state-of-the-art techniques which make estimation possible in strains of laying hens with large numbers of birds.

MATERIALS AND METHODS

Data and traits description

Data, including records of 165, 748 and 47, 115 survivor laying hens for strains A and B respectively, were supplied by the "Institut de Sélection Animale–ISA". For both strains, these records represented 6 generations of hens.

Traits considered in this analysis were related to egg production (number of eggs produced between 19–26 wk of age (P_1), 26–38 wk (P_2) and 26–54 wk (P_3)) and egg characteristics (average egg weight at 2 different ages (EW_1, EW_2) and egg density (ED)). Weight of hens at 40 wk of age (W_{40}) was also included in the analysis.

P_1 can be considered as being a combination of sexual maturity and early egg production. ED was a measure of the shell strength determined by specific gravity.

P_1, P_2 and P_3 variables exhibited markedly skewed distributions. They were transformed into new variables, satisfying the classical hypotheses for describing traits with polygenic inheritance *via* a linear model with normal error, using a power transformation (Box and Cox, 1964). This transformation relies on a single parameter t and has the following form (Ibe and Hill, 1988):

$$
g_t(y) = \begin{cases} \dfrac{y^t - 1}{t\dot{y}^{(t-1)}} & \text{if} \quad t \neq 0 \\[2mm] \log(y) & \text{if} \quad t = 0 \end{cases} \tag{1}
$$

were y is the geometric mean of the original observations.

The parameter t was empirically chosen in such a way that several normality criteria, such as the low residual sum of squares of the genetic model used to describe the data, the linearity of half-sib on individual regression, the coefficient of symmetry and the Kolmogorov–Smirnov test for normality of the residuals were satisfied as closely as possible and simultaneously, as proposed by Ibe and Hill (1988) and detailed in Besbes *et al* (1992).

Model of analysis

The model describing the records is the following multiple-trait animal model:

$$
\mathbf{Y} = (\mathbf{I}_t * \mathbf{X})\mathbf{b} + (\mathbf{I}_t * \mathbf{Z})\mathbf{a} + \mathbf{e} \tag{2}
$$

where **Y** is the vector ($n \times t$) of observations for the t egg production traits considered in the multiple-trait analysis ($t = 7$),

b is the vector ($f \times t$) of fixed contemporary groups (f was equal to 107 and 70, for strains A and B respectively),

a is the vector ($n \times t$) of random additive genetic effects associated with the animal's traits,

e is the vector ($n \times t$) of residuals,

X and **Z** are known incidence matrices associated with **b** and **a** and (*) indicates direct (Kronecker) product.

It was assumed that

$$E(\mathbf{Y}) = (\mathbf{I}_t * \mathbf{X})\mathbf{b},$$

and
$$\begin{bmatrix} \mathbf{a} \\ \mathbf{e} \end{bmatrix} \approx \text{MVN} \left[\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{G} * \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} * \mathbf{I}_n \end{bmatrix} \right]$$

$\mathbf{A}$ is the relationship matrix between animals. $\mathbf{G}$ and $\mathbf{R}$ are unknown genetic and residual (co)variance matrices between the t traits considered.

Computing strategy for genetic parameter estimation

The multiple-trait animal model [2] had one random effect and equal design matrices for all traits which were recorded for all animals (no missing records). Canonical decomposition was then applied to yield new uncorrelated variables without loss of any information contained in the original variables. This transformation was first suggested for animal breeding problems by Thompson (1976, cited by Jensen and Mao, 1988). The transformation matrix $\mathbf{Q}$, is chosen such that (Quaas *et al*, 1984):

$$\mathbf{Q}\mathbf{R}\mathbf{Q}' = \mathbf{I}_{nt}$$

$$\mathbf{Q}\mathbf{G}\mathbf{Q}' = \mathbf{G}_c$$

where $\mathbf{G}_c$ is a diagonal matrix and $\mathbf{I}_{nt}$ is the identity matrix. After transformation to the canonical scale, model [2] becomes:

$$(\mathbf{Q} * \mathbf{I}_n)\mathbf{Y} = (\mathbf{Q} * \mathbf{I}_n)(\mathbf{I}_t * \mathbf{X})\mathbf{b} + (\mathbf{Q} * \mathbf{I}_n)(\mathbf{I}_t * \mathbf{Z})\mathbf{a} + (\mathbf{Q} * \mathbf{I}_n)\mathbf{e}$$
$$\Leftrightarrow \quad \mathbf{Y}_c = (\mathbf{I}_t * \mathbf{X})\mathbf{b}_c + (\mathbf{I}_t * \mathbf{Z})\mathbf{a}_c + \mathbf{e}_c \qquad [3]$$

with $\quad E(\mathbf{a}_c) = E(\mathbf{e}_c) = \mathbf{0}$,
$\qquad \text{var}(\mathbf{a}_c) = \mathbf{G}_c * \mathbf{A}$,
$\qquad \text{var}(\mathbf{e}_c) = \mathbf{I}_{nt}$

The subscript c refers to the canonical scale. This transformation reduced the multivariate analysis to a series of univariate analyses and consequently, drastically decreased computational costs.

These computational costs were also lowered by reducing the number of equations. Since parents represented only 9% and 8% of the total number of animals, for strains A and B respectively, the reduced animal model (RAM) of Quaas and Pollak (1980) was used. With RAM, all the equations corresponding to animals which were nonparents were absorbed into the remaining equations. As a consequence, the size of the system was brought down to the number of parents.

Considering model [3] for the μth trait, with RAM, the vector $\mathbf{Y}_{c\mu}$ was divided into parents (denoted by subscript p) and nonparents (denoted by subscript n) as follows (subscript μ is dropped for clarity):

$$\begin{bmatrix} \mathbf{Y}_{pc} \\ \mathbf{Y}_{nc} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_p \\ \mathbf{X}_n \end{bmatrix} \mathbf{b}_c + \begin{bmatrix} \mathbf{Z}_p \\ 1/2\mathbf{P}_n \end{bmatrix} \mathbf{a}_{pc} + \begin{bmatrix} \mathbf{e}_{pc} \\ \mathbf{e}_{nc}^* \end{bmatrix} \qquad [4]$$

with $\mathbf{e}_{nc}^* = \mathbf{e}_{nc} + \boldsymbol{\phi}_{nc}$ ($\boldsymbol{\phi}_{nc}$ being the Mendelian samplign on the canonical scale) and $\mathbf{P}_n$ is a matrix of 0 and 1 relating nonparents to their parents. Random variables in [4] have the following (co)-variance structure:

$$
\mathrm{var}
\begin{bmatrix}
\mathbf{a}_{pc} \\
\mathbf{e}_{pc} \\
\mathbf{e}_{nc}^*
\end{bmatrix}
=
\begin{bmatrix}
\mathbf{A}_p \sigma_a^2 & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{I}_p & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{I}_n + \Delta_n \sigma_a^2
\end{bmatrix}
\qquad [5]
$$

σ_a^2 is the μth diagonal element of $\mathbf{G}_c$. Δ_n is a diagonal matrix whose elements are equal to 1/2 if both parents are known and 3/4 if one parent if known. It was assumed that there is equal parental information for all nonparents and that parents are not inbred. If this assumption is not satisfied, nonparents with unknown parents can have dummy parents (Thompson and Meyer, 1990).

Since REML is an iterative procedure, the major cost was the need, at each iteration, of direct inversion of a matrix of size equal to the number of parents. This burden was reduced by tridiagonalizing the coefficient matrix through a series of orthogonal transformations, as proposed by Smith and Graser (1986). This transformation, however, was not directly applicable because, as shown in [5], RAM generates heterogeneous residual variances between parents and nonparents. To overcome this problem, Thompson and Meyer (1990) reparameterized [4] adding an imaginary effect, $\mathbf{e}_I$ with negative variance;

$$
\begin{bmatrix}
\mathbf{Y}_{pc} \\
\mathbf{Y}_{nc}
\end{bmatrix}
=
\begin{bmatrix}
\mathbf{X}_p \\
\mathbf{X}_n
\end{bmatrix}
\mathbf{b}_c +
\begin{bmatrix}
\mathbf{Z}_p \\
1/2\mathbf{P}_n
\end{bmatrix}
\mathbf{a}_{pc} +
\begin{bmatrix}
i\omega\mathbf{I}_p \\
\mathbf{0}
\end{bmatrix}
\mathbf{a}_d +
\begin{bmatrix}
\mathbf{e}_{pc} - \mathbf{e}_I \\
\mathbf{e}_{nc}^*
\end{bmatrix}
\qquad [6]
$$

$\mathbf{e}_I = i\,\omega\mathbf{a}_d$, (with $i^2 = -1$), was chosen such that $\mathbf{e}_{pc} - \mathbf{e}_I$ and $\mathbf{e}_{nc}^*$ had the same variance, hence $\mathrm{var}(\mathbf{e}_I) = -\omega^2\sigma_a^2\mathbf{I}_p$ (with ω^2 being an element of Δ_n). Hence, the mixed model equations corresponding to model [6] can be written as:

$$
\begin{bmatrix}
\mathbf{X}_p'\mathbf{X}_p + \mathbf{X}_n'\mathbf{X}_n & \mathbf{X}_p'\mathbf{Z}_p + \dfrac{1}{2}\mathbf{X}_n'\mathbf{P}_n & i\omega\mathbf{X}_p' \\[2mm]
\mathbf{Z}_p'\mathbf{X}_p + \dfrac{1}{2}\mathbf{P}_n'\mathbf{X}_n & \mathbf{Z}_p'\mathbf{Z}_p + \dfrac{1}{4}\mathbf{P}_n'\mathbf{P}_n + \alpha\mathbf{A}_p^{-1} & i\omega\mathbf{Z}_p' \\[2mm]
i\omega\mathbf{X}_p & i\omega\mathbf{Z}_p & -\omega^2\mathbf{I}_p + \alpha\mathbf{I}_p
\end{bmatrix}
$$

$$
\begin{bmatrix}
\mathbf{b}_c \\
\mathbf{a}_{pc} \\
\mathbf{a}_d
\end{bmatrix}
=
\begin{bmatrix}
\mathbf{X}_p'\mathbf{Y}_{pc} + \mathbf{X}_n'\mathbf{Y}_{nc} \\[1mm]
\mathbf{Z}_p'\mathbf{Y}_{pc} + \dfrac{1}{2}\mathbf{P}_n'\mathbf{Y}_{nc} \\[1mm]
i\omega\mathbf{Y}_{pc}
\end{bmatrix}
\qquad [7]
$$

where α is the μth trait's variance ratio $\sigma_\Psi^2 \big/ \sigma_a^2$, with $\sigma_\Psi^2 = 1 + \omega^2\sigma_a^2$ an element of the new canonical residual matrix $\mathbf{R}_c$.

For estimating the (co)variance components, equations for fixed effects are absorbed. The resulting system has equations of the form (Thompson and Meyer, 1990):

$$\begin{bmatrix} \mathbf{B} + \alpha \mathbf{A}_p^{-1} & i\mathbf{C} \\ i\mathbf{C}' & \mathbf{D} + \alpha \mathbf{I}_p \end{bmatrix} \begin{bmatrix} \mathbf{a}_{pc} \\ \mathbf{a}_d \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ i\mathbf{g} \end{bmatrix} \qquad [8]$$

where:

$$\mathbf{B} = \mathbf{Z}_p'\mathbf{Z}_p + \frac{1}{4}\mathbf{P}_n'\mathbf{P}_n - (\mathbf{Z}_p'\mathbf{X}_p + \frac{1}{2}\mathbf{P}_n'\mathbf{X}_n)(\mathbf{X}_p'\mathbf{X}_p + \mathbf{X}_n'\mathbf{X}_n)^{-1}(\mathbf{X}_p'\mathbf{Z}_p' + \frac{1}{2}\mathbf{X}_n'\mathbf{P}_n)$$

$$\mathbf{C} = \omega\mathbf{Z}_p' - \omega(\mathbf{Z}_p'\mathbf{X}_p + \frac{1}{2}\mathbf{P}_n'\mathbf{X}_n)(\mathbf{X}_p'\mathbf{X}_p + \mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{X}_p'$$

$$\mathbf{D} = -\omega^2\mathbf{I}_p + \omega^2\mathbf{X}_p(\mathbf{X}_p'\mathbf{X}_p + \mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{X}_p'$$

$$\mathbf{f} = \mathbf{Z}_p'\mathbf{Y}_{pc} + \frac{1}{2}\mathbf{P}_n'\mathbf{Y}_{nc} - (\mathbf{Z}_p'\mathbf{X}_p + \frac{1}{2}\mathbf{P}_n'\mathbf{X}_n)(\mathbf{X}_p'\mathbf{X}_p + \mathbf{X}_n'\mathbf{X}_n)^{-1}(\mathbf{X}_p'\mathbf{Y}_{pc} + \mathbf{X}_n'\mathbf{Y}_{nc})$$

$$\mathbf{g} = \omega\mathbf{Y}_{pc} - \omega\mathbf{X}_p(\mathbf{X}_p'\mathbf{X}_p + \mathbf{X}_n'\mathbf{X}_n)^{-1}(\mathbf{X}_p'\mathbf{Y}_{pc} + \mathbf{X}_n'\mathbf{Y}_{nc})$$

Further simplification of [8] was achieved by eliminating $\mathbf{A}_p^{-1}$. This involved the Cholesky decomposition of the numerator relationship matrix $(\mathbf{A}_p = \mathbf{LL}')$ and pre- and post-multiplication of the left hand side by $\mathbf{L}'$ and $\mathbf{L}$ and multiplication of the right hand side by $\mathbf{L}'$ (Smith and Graser, 1986). The solutions for $\mathbf{a}_{pc}$ and $\mathbf{a}_d$ are of the form:

$$\begin{bmatrix} \widehat{\mathbf{a}}_l \\ \widehat{\mathbf{a}}_d \end{bmatrix} = (\mathbf{H} + \alpha \mathbf{I}_{p'})^{-1} \begin{bmatrix} \mathbf{L}'\mathbf{f} \\ i\mathbf{g} \end{bmatrix} \qquad [9]$$

with $p' = 2p, \widehat{\mathbf{a}}_l = \mathbf{L}^{-1}\widehat{\mathbf{a}}_{pc}$ and $\mathbf{H} = \begin{bmatrix} \mathbf{L}'\mathbf{BL} & i\mathbf{L}'\mathbf{C} \\ i\mathbf{C}'\mathbf{L} & \mathbf{D} \end{bmatrix}$

As shown in [9], this method led to the tridiagonalization of a complex matrix of size twice the number of parents. After the tridiagonal matrix $\mathbf{T}$ (for $\mathbf{T}$ such that $\mathbf{T} = \mathbf{PHP}'$, where $\mathbf{P}$ is an orthogonal matrix) was found and using algebra similar to that of Smith and Graser (1986) system [9] becomes:

$$\widehat{\mathbf{a}}_m = (\mathbf{T} + \alpha \mathbf{I}_{p'})^{-1}\mathbf{q}^* \qquad [10]$$

where $\widehat{\mathbf{a}}_m = \mathbf{P}' \begin{bmatrix} \widehat{\mathbf{a}}_l \\ \widehat{\mathbf{a}}_d \end{bmatrix}$ and $\mathbf{q}^* = \mathbf{P}' \begin{bmatrix} \mathbf{L}'\mathbf{f} \\ i\mathbf{g} \end{bmatrix}$

Then, we iteratively calculate the canonical REML estimates of the (co)variance matrices, $\widehat{\mathbf{G}}_c$ and $\widehat{\mathbf{R}}_c$, between all traits using an Expectation–Maximization (EM) type algorithm (Dempster *et al*, 1977; Harville, 1977).

For the $(\eta + 1)$ round of iteration, estimators of the elements in $\mathbf{G}_c$ are:

$$\widehat{g}_{cjk}^{(\eta+1)} = \begin{cases} \widehat{\mathbf{a}}_{mj}'\widehat{\mathbf{a}}_{mk} \Big/ \left[N - \text{tr}\,(\mathbf{T} + \alpha_j \mathbf{I}_{p'})^{-1} \right] & \text{if} \quad j = k \qquad [11] \\[1em] \widehat{\mathbf{a}}_{mj}\widehat{\mathbf{a}}_{mk} \Big/ N & \text{if} \quad j \neq k \qquad [12] \end{cases}$$

and those in $\mathbf{R}_c$ are:

$$\hat{\mathbf{r}}_{cjk}^{(\eta+1)} = \begin{cases} \left[\mathbf{Y}_{cj}^{*'}\mathbf{Y}_{ck}^* - \hat{\mathbf{a}}_{mj}'\mathbf{q}_k^* - \alpha_k\hat{\mathbf{a}}_{mj}'\hat{\mathbf{a}}_{mk}\right]\Big/\left[N - f - p' + \text{tr}\left(\mathbf{T} + \alpha_j\mathbf{I}_{p'}\right)^{-1}\right] \\ \\ \left[\mathbf{Y}_{cj}^{*'}\mathbf{Y}_{ck}^* - \hat{\mathbf{a}}_{mj}'\mathbf{q}_k^* - \alpha_k\hat{\mathbf{a}}_{mj}'\hat{\mathbf{a}}_{mk}\right]\Big/\left[N - f - p'\right] \end{cases}$$

$$\text{if} \quad j = k \quad [13]$$

$$\text{if} \quad j \neq k \quad [14]$$

where N is the total number of observations, f is the number of levels of the fixed effect, p' is twice the number of parents and $\mathbf{Y}_c^{*'}\mathbf{Y}_c^*$ is the quadratic form of the canonical observations after absorption of the fixed effects. It has the following expression:

$$\mathbf{Y}_c^{*'}\mathbf{Y}_c^* = \mathbf{Y}_{pc}'\mathbf{Y}_{pc} + \mathbf{Y}_{nc}'\mathbf{Y}_{nc} - (\mathbf{Y}_{pc}'\mathbf{X}_p + \mathbf{Y}_{nc}'\mathbf{X}_n)$$
$$(\mathbf{X}_p'\mathbf{X}_p + \mathbf{X}_n'\mathbf{X}_n)^{-1}(\mathbf{X}_p'\mathbf{Y}_{pc} + \mathbf{X}_n'\mathbf{Y}_{pc})$$

Once the variance components in [11] through [14] are obtained, we go back to the original scale by performing back-transformation as follows:

$$\hat{\mathbf{G}}_{(\eta+1)} = \mathbf{Q}^{-1}\hat{\mathbf{G}}_c\mathbf{Q}^{-T}$$

$$\hat{\mathbf{R}}_{(\eta+1)} = \mathbf{Q}^{-1}\hat{\mathbf{R}}_c\mathbf{Q}^{-T} \qquad [15]$$

then, the new estimates of $\mathbf{G}$ and $\mathbf{R}$ are used to obtain a new $\mathbf{Q}$ transformation matrix and new $\mathbf{G}_c$ and $\mathbf{R}_c$, so the process from [10] to [15] is iterated until convergence is reached.

The single trait analysis corresponds to the special case where $\mathbf{Q} = \mathbf{I}$ (Hence $\mathbf{G}$ and $\mathbf{R}$ are diagonal), and the EM–REML equations are those in [11] and [13].

Sampling procedure

Despite these cost-reducing techniques (canonical decomposition, use of RAM, tridiagonalization) the time-consuming tridiagonalization and above all the amount of computer memory required, prohibited the application of REML estimation to the whole population. Therefore, 15 samples, reflecting as well as possible the population's structure and selection, were drawn in each strain from the data file in the following manner:

1. Choose S sire ($S = 5$) at random among the youngest parents.

2. Include the sire and dam of each selected sire.

3. For each of the selected sires S_i in (1) and (2), choose 3 sires S_j at random among those whose offspring are contemporary to those of S_i.

4. For each of the sire selected in (1) to (3), choose 3 females at random among its mates.

5. Repeat step (2) to (4) until the generation which is assumed to be the base population has been reached.

6. Include all offspring of all the selected matings in the sample.

As mentioned above, the system to solve is of size twice the number of parents. This, depending on the computer capacity, limits the sample's size. The values 5 and 3 in steps 1,3 and 4 were chosen by trial and error in order to obtain a maximum of 800 parents, corresponding to sampling rates of $\approx$ 5 and 17% respectively for strains A and B, which is as many as can be handled.

This sampling procedure has the following characteristics: only the pedigrees of males are complete. Step 3 leads to the inclusion of hens contemporary to those hens whose records were used to select sires. Hence, selection on the male side is (approximately) accounted for. This is not the case on the female sire. However, the change from one generation to the next in the expected value $E(\mathbf{a}) = \mathbf{g}$ of the genetic merit of hens, due to selection, is accounted for since the levels of the fixed effect (contemporary groups) are defined within generation. In other words, this expected value, considered as fixed, using the approach of Westell (1984) or Quaas (1988) is completely confounded with the contemporary group effect. Only the effect of selection on the female side on the genetic variance is not accounted for as usually indicated when using animal models.

As already mentioned, only the pedigrees of males are complete. Hence, neither Henderson's rules (1976) for computing the numerator relationship matrix $\mathbf{A}$ or its inverse $\mathbf{A}^{-1}$, nor Quaas's algorithm (1989) to eliminate $\mathbf{A}^{-1}$ from mixed model equations used in REML could be utilized. $\mathbf{A}$ and its Cholesky factor $\mathbf{L}(\mathbf{A} = \mathbf{LL}')$ were therefore calculated directly considering the pedigree available for the entire population (Tier, 1990).

RESULTS AND DISCUSSION

The average numbers of parents per sample were 710 and 600 for strains A and B respectively, which represented samples of $\approx$ 7 000 and 6 500 recorded hens partitioned into 107 and 70 contemporary groups. The tridiagonalization of the coefficient matrix of size twice the number of parents required respectively 113 and 60 min of CPU time on an IBM 3090-17T. This represented 98% of the total computational time needed for genetic parameter estimation for such samples.

As a consequence of a smaller number of sires per generation for strain B, the corresponding samples overlapped in a higher proportion. On average, 38% of parents were common to any 2 samples *vs* 17% for strain A. This overlap was neglected when computing means and standard deviations of the estimates obtained from the 15 samples.

As a consequence of working with imaginary terms, the algorithm proposed by Thompson and Meyer produced a tridiagonal matrix $\mathbf{T}$ with some negative eigenvalues. When by chance, during the EM iteration, the variance ration α was very close to one of those negative eigenvalues, the matrix of system [10] was singular. This led to an infinite trace, preventing the EM–REML from converging. When this occurred, it was necessary to "jump" over the value α to avoid numerical problems.

In any case, after the tridiagonal matrix was found, the equations of the EM could be solved in linear time and any number of iterations was easily performed (Smith and Graser, 1986). However, the number of rounds needed to reach convergence varied from sample to sample (from 120 to 400 here).

Single trait genetic parameter estimates

Heritability values (table I) exhibited very small differences between strains A and B. The highest difference was observed for the number of eggs produced between 26 and 38 wk of age: 0.09 for strain A *vs* 0.13 for strain B. However, it should be noted that for similar heritabilities, strain B presented, in general, larger additive genetic and residual variances.

These estimates also showed that egg production traits (P_1, P_2 and P_3) are much less heritable than egg characteristics or body weight.

Heritabilities of egg production traits were lower than usually reported in the literature, especially those for P_2 and P_3. Regarding heritability of P_1, a combined trait of both sexual maturity and early egg production, it was roughly within the range of Liljedahl and Weyde (1980) estimates but closer to that of egg production. We should, however, be careful with such comparisons since most reported heritabilities and variance components were obtained on the original scale without performing any transformation but using methods assuming normality of the data distribution. Strictly speaking, such results should be interpreted with caution.

The purpose of the Box–Cox transformation, applied for P_1, P_2 and P_3 is then to change the scale of measurements in order to make the analysis more valid. Besbes *et al* (1992) showed that this transformation resulted in an increase of all heritabilities without drastically modifying the genetic and residual correlations between these traits.

Heritability values of egg and hen weight, though rather small, remained within the range of estimates reported by King and Henderson (1954, cited by Kolstad, 1980), Kolstad (1980) and Sorensen *et al* (1980). The same trend was observed for the estimates of egg density (specific gravity).

Multiple-trait genetic parameter estimates

As shown in table II, heritabilities obtained by multivariate EM–REML were very close to those obtained by single trait analysis. This result was in agreement with that of Colleau *et al* (1989).

The comparison of genetic correlations of both strains revealed quite similar global trends concerning the sign of these correlations. But those of strain A were, in general, smaller in absolute values.

These correlations showed a rather small antagonism between the group of egg production traits (P_1, P_2 and P_3) and egg density (ED) on one hand and that of average egg weights on the other hand. The largest antagonism was observed between egg weight (EW_1) and egg density (-0.17 and -0.23 respectively for strain A and B) but remained rather small.

In the literature, there is a large variation in the reported scale of the genetic correlation between number of eggs produced and average egg weight. For Sorensen *et al* (1980), this correlation was -0.32 and -0.17 depending on the population. Liljedahl and Weyde (1980) reported an evolution of this correlation from slightly positive and non significant values in the base population to markedly negative ones in selected lines. These results were in contrast with those of Kolstad (1980)

Table I. Univariate REML estimates of heritabilities (h^2) and genetic (σ_a^2) and residual (σ_e^2) variances of egg production traits[1,3] (P_1, P_2 and P_3), egg characteristics (EW_1, EW_2 and ED) and weight of hens (W_{40}), for 2 pure lines A and B.

Traits[3]	h^2		σ_a^2		σ_e^2	
	A	B	A	B	A	B
P_1	0.25 ± 0.01[2]	0.27 ± 0.01	16.8 ± 0.6	26.3 ± 1.1	50.3 ± 0.7	70.0 ± 0.9
P_2	0.09 ± 0.00	0.13 ± 0.01	2.4 ± 0.1	4.1 ± 0.3	22.7 ± 0.2	27.9 ± 0.5
P_3	0.18 ± 0.01	0.16 ± 0.01	44.6 ± 2.3	38.9 ± 2.0	210.7 ± 2.0	207.4 ± 3.9
EW_1	0.47 ± 0.01	0.48 ± 0.01	464.4 ± 13.1	631.5 ± 28.4	528.7 ± 7.2	667.3 ± 16.0
EW_2	0.47 ± 0.01	0.44 ± 0.01	600.1 ± 19.8	781.5 ± 38.0	674.2 ± 13.1	995.1 ± 31.9
$W40$	0.50 ± 0.01	0.51 ± 0.01	133.5 ± 3.4	164.7 ± 4.3	130.8 ± 2.3	170.2 ± 13.0
ED	0.34 ± 0.01	0.34 ± 0.02	119.3 ± 1.9	107.0 ± 7.6	231.7 ± 2.7	202.2 ± 2.8

[1] Normalized using Box–Cox transformation (the values of t were 3.0, 6.5 and 4.8 respectively for P_1, P_2 and P_3). [2] Approximate error SD $= \sigma/\sqrt{n}$, where n is the number of replicates. [3] $P_{1,2,3}$ = number of eggs produced between 19 and 26, 26 and 38 and 26 and 54 wk of age; $EW_{1,2}$ = average egg weight (g) at 2 different ages; ED = egg density; W_{40} = weight (g) of hens at 40 wks of age.

Table II. Multivariate EM–REML estimates[1] of heritabilities[2] (diagonal elements) and genetic correlations (of diagonals) between egg production traits[3] (P_1, P_2 and P_3), egg characteristics (EW_1, EW_2 and ED) and weight of hens (W_{40}), for 2 pure lines A (first row) and B (second row).

Trait	P_1	P_2	P_3	EW_1	EW_2	ED	W_{40}
P_1[4]	0.26	0.27	0.06	−0.09	−0.04	0.04	−0.09
	0.27	0.36	0.21	−0.18	−0.16	0.09	−0.12
P_2[4]		0.10	0.66	−0.09	−0.15	0.08	0.02
		0.13	0.78	−0.08	−0.14	0.12	−0.26
P_3[4]			0.18	−0.09	−0.14	0.03	0.25
			0.16	−0.13	−0.22	0.04	−0.12
EW_1				0.47	0.92	−0.17	0.29
				0.51	0.94	−0.23	0.25
EW_2					0.49	−0.13	0.32
					0.47	−0.22	0.33
ED						0.33	−0.03
						0.35	−0.08
W_{40}							0.51
							0.51

[1] These estimates are averages of 15 samples. [2] The error SD $= \sigma/\sqrt{n}$ (where n is the replications number) associated with the heritabilities was ≈ 0.01, except for ED whose value was 0.02. [3] $P_{1,2,3}$ = number of eggs produced between 19 and 26, 26 and 38 and 26 and 54 wk of age. $EW_{1,2}$ = average egg weight at 2 different ages. ED = egg density. W_{40} = weight of hens at 40 wk of age. [4] Transformed using the Box-Cox method.

which showed some evidence for a decline in the genetic correlation from −0.18 in the base population to −0.11 in the selected lines. Nearly all these correlations and heritabilities were estimated by analyzing the resemblance between paternal half-sibs and separately for each generation and trait. Hence, they were biased by selection. This was theoretically not the case with our REML estimation. However, it must be said that our "base" population was already a selected one and, also, the sampling procedure could not take all the selection process into account.

From table II, it can be seen that correlations between egg number and body weight of hens varied considerably among strains (the correlation between P_3 and W40 was 0.25 for strain A and −0.12 for strain B) which was also found by Kolstad (1980). This author reported that an intermediate body weight is optimum for reproduction, of which egg production is a component, and hence one should expect these correlations to be positive or negative depending on whether the weight is below or above the optimum of the population. Therefore, it might be thought that strain A was closer to an optimum for body size than strain B. Great variations of these correlations were also observed among the 3 recording periods (they varied, in these case of strain A, from −0.09 for P_1 to 0.25 for P_3 while correlations between egg and body weight were positive and consistently moderate in all cases.

Table II also shows that traits within each group of traits were positively correlated. The highest correlations were between P_2 and P_3 for the first group of traits, respectively 0.66 and 0.78 and EW_1 and EW_2 for the second group, 0.92 and 0.94. To some degree, EW_1 and EW_2 were measurements of the same characteristic.

The correlation between P_2 and P_3 is lower than the literature average values (0.79 and 0.85, respectively for selected and unselected White Leghorn breed) obtained by Fairfull and Gowe (1990). This is probably due to a low correlation between P_2 and the residual egg production (38–54 wk of age). Hence, P_2 and P_3 cannot be considered as measuring the same trait even if they overlap.

The positive correlations between P_1 and the other egg production traits (P_2 and P_3) indicate that this trait has to be considered as an egg production trait rather than a sexual maturity trait (sexual maturity is usually negatively correlated with the number of eggs). The nonsignificant correlation between these traits and shell quality was also reported by Kolstad (1980). However Fairfull and Gowe (1990) obtained, for selected White Leghorn, a negative correlation of ≈ -0.2 between survivor egg production from housing to 40 or 55 wk of age and specific gravity.

CONCLUSION

The present study showed a rather small antagonism between egg production traits and egg weights and a positive correlation between traits within the same category.

Comparison of the heritabilities obtained with univariate and multivariate REML estimation showed no significant differences.

By employing a series of transformations in the variance components estimation process, especially the tridiagonalization of the system proposed by Thompson and Meyer (1990), the problem of slow convergence of the EM algorithm was largely alleviated since the computations necessary at each round of iteration were performed in linear time. This allowed us to study large samples with an acceptable computational time. However, this algorithm encountered numerical problems, depending on the matrix structure, that potential users should be aware of.

The animal model used involved additive genetic effects only. The complexity of calculations precluded the use of a more sophisticated model including *eg* non-additive, maternal and common environmental effects. Comparison of sire and dam components estimated elsewhere (Besbes, 1992) revealed very small differences between these 2, which support the choice of model [2].

ACKNOWLEDGMENTS

REFERENCES

Besbes B (1993) Application de la méthodologie du BLUP à l'évaluation génétique de deux souches de poules pondeuses. Thèse Doc Sci, INA, Paris-Grignon, France

Besbes B, Ducrocq V, Foulley JL, Protais M, Tavernier A, Tixier-Boichard M, Beaumont C, (1992) Box–Cox transformation to egg production traits of laying hens to improve genetic parameter estimation and breeding evaluation. *Livest Prod Sci*, 32 (in press)

Box GPE, Cox DR (1964) An analysis of transformation. *JR Stat Soc B* 26, 211-243

Colleau JJ, Beaumont C, Regaldo D (1989) Restricted maximum likelihood (REML) estimation of genetic parameters for type traits in Normande cattle breed. *Livest Prod Sci* 23, 47-66

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *JR Stat Soc B* 39, 1-38

Fairfull RW, Gowe RS (1990) Genetics of egg population in chickens. *In: Poultry Breeding and Genetics* (Crawford RD ed) Elsevier, ch 29, 705-759

Gianola D, Foulley JL, Fernando CL (1986) Prediction of breeding values when variances are not known. *Genet Sel Evol* 18, 485-498

Harville PA (1977) Maximum likelihood approaches to variance component estimation and to related problems. *J Am Stat Assoc* 72, 320-338

Henderson CR (1976) A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* 32, 69-83

Ibe SN, Hill WG (1988) Transformation of poultry egg production data to improve normality, homoscedaticity and linearity of genotypic regression. *J Anim Breed Genet* 105, 231-240

Jensen J, Mao IL (1988) Transformation algorithms in analysis of single trait and of multitrait models with equal design matrices and one random factor per trait: a review. *J Dairy Sci* 66, 2750-2761

Kolstad N (1980) Scandinavian selection and crossbreeding experiment with laying hens. III. Results from the Norwegian part of the experiment. *Acta Agric Scand* 30, 261-287

Liljedahl LE, Weyde K (1980) Scandinavian selection and crossbreeding experiment with laying hens. II. Results from the Swedish part of the experiment. *Acta Agric Scand* 30, 237-260

Patterson HD, Thompson R (1971) Recovery of interblock information when block sizes are unequal. *Biometrika* 58, 545-554

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

Quaas RL (1989) Transformed mixed model equations: a recursive algorithm to eliminate A^{-1} *J Dairy Sci* 72, 1937-1941

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

Quaas RL, Anderson RD, Gilmour AR (1984) Use of mixed models for prediction and for estimation of (co)variance components. *BLUP School Handbook*. Univ New Engl, Australia

Smith SP, Graser HU (1986) Estimating variance components in a class of mixed models by restricted maximum likelihood. *J Dairy Sci* 69, 1156-1165

Sorensen P, Ambosen T, Peterson A (1980) Scandinavian selection and crossbreeding experiment with laying hens. IV. Results from the Danish part of the experiment. *Acta Agric Scand* 30, 288-308

Thompson R, Meyer K (1990) Estimating genetic parameters using an animal model with imaginary effects. *Genet Sel Evol* 21, 1-15

Tier B (1990) Computing inbreeding coefficients quickly. *Genet Sel Evol* 22, 419-430

Westell RA (1984) Simultaneous genetic evaluation of sires and cows for a large population of dairy cattle. Ph D Diss, Cornell Univ, Ithaca, NY, Univ Microfilms Inc, Ann Arbor, MI