HAL open science

Prediction of the response to a selection for canalisation of a continuous trait in animal breeding

Magali San Cristobal, Jean-Michel Elsen, Loys Bodin, Claude Chevalet

To cite this version:

Magali San Cristobal, Jean-Michel Elsen, Loys Bodin, Claude Chevalet. Prediction of the response to a selection for canalisation of a continuous trait in animal breeding. Genetics Selection Evolution, 1998, 30 (5), pp.423-451. hal-00894220

HAL Id: hal-00894220

https://hal.science/hal-00894220v1

Submitted on 11 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Genet. Sel. Evol. 30 (1998) 423–451
© Inra/Elsevier, Paris

Original article

Prediction of the response to a selection for canalisation of a continuous trait in animal breeding

Magali SanCristobal-Gaudy [a*], Jean-Michel Elsen [b], Loys Bodin [b], Claude Chevalet [a]

[a] Laboratoire de génétique cellulaire, Institut national de la recherche agronomique, BP27, 31326 Castanet-Tolosan cedex, France
[b] Station d'amélioration génétique des animaux, Institut national de la recherche agronomique, BP27, 31326 Castanet-Tolosan cedex, France

(Received 7 November 1997; accepted 31 August 1998)

Abstract – Canalising selection is handled by a heteroscedastic model involving a genotypic value for the mean and a genotypic value for the log variance, associated with a single phenotypic value. A selection objective is proposed as the expected squared deviation of the phenotype from the optimum, of a progeny of any candidate for selection. Indices and approximate expressions of parent–offspring regression are derived. Simulations are performed to check the accuracy of the analytical approximation. Examples of fat to protein ratio in goat milk yield and muscle pH data in pig breeding are provided in order to investigate the ability of these populations to be canalised towards an economic optimum. © Inra/Elsevier, Paris

canalising selection / heteroscedasticity / selection index

Résumé – Prédiction de la réponse à une sélection canalisante d'un caractère continu en génétique animale. Le problème de la sélection canalisante est traité grâce à un modèle hétéroscédastique mettant en jeu une valeur génétique pour la moyenne et une valeur génétique pour le logarithme de la variance, toutes deux associées à une seule valeur phénotypique. Pour un objectif de sélection visant à minimiser l'espérance des carrés des différences entre le phénotype et l'optimum, pour un descendant d'un candidat à la sélection, des index sont estimés et des expressions approchées de la régression parent-descendant sont calculées. La précision de ces expressions analytiques est mesurée à l'aide de simulations. Afin d'appréhender la capacité de

* Correspondence and reprints
E-mail: msc@toulouse.inra.fr

ces populations à être canalisées vers un optimum économique, des exemples sont donnés : le rapport entre matière grasse et matière protéique du lait de chèvre, et le pH d'un muscle chez le porc. © Inra/Elsevier, Paris

sélection canalisante / hétéroscédasticité / index de sélection

1. INTRODUCTION

Production homogeneity is an important factor of economic efficiency in animal breeding. For instance, optimal weights and ages at slaughtering exist for broilers, lambs and pigs, and the breeder's profit depends on his ability to send large homogeneous groups to the abattoir; optimal characteristics of meat such as its pH 24 h after slaughtering exist but depend on the type of transformation; ewes lambing twins have the maximum profitability while single litters are not sufficiently productive and triplets or larger litters are too difficult to raise; with extensive conditions where food is determined by climatic situations, genotypes able to maintain the level of production would be of interest.

Hohenboken [22] listed different types of matings (inbreeding, outbreeding, top crossing and assortative matings) and selection (normalising, directional and canalising) which can lead to a reduction in trait variability.

Stabilisation of phenotypes towards a dominant expression has been known for a long time as a major determinant of species evolution, similarly to mutations and genetic drift (e.g. [4] for a review). Different hypotheses explaining these natural stabilising selection forces have been proposed [2, 3, 8, 15, 16, 19, 27, 38, 45–47, 49, 52]. A number of models assume that trait stabilisation is controlled by fitness genes (e.g. [9] for a review), which keeps the mean phenotype at a fixed 'optimal' level, without a necessary reduction of the trait variability. Alternative hypotheses were proposed for canalisation; for instance Rendel et al. [32, 33] assumed that the development of a given organ is under the control of a set of genes, while a major gene controls the effects of these genes within bounds to keep the phenotype roughly constant.

Whatever its origin stabilisation is to be related to the environment(s) in which it is observed, which makes it essential [48] to distinguish stabilisation of a trait in a precise environment (normalising selection) from the aptitude to maintain a constant phenotype in fluctuating environments (canalising selection).

Various artificial stabilising selection experiments have been carried out with laboratory animals: drosophila [17, 23, 29, 30, 34, 40, 41, 44, 48], tribolium [5, 6, 24, 43] and mice [32]. Most often, selection was of a normalising type with a culling of extreme individuals, this selection being applied globally [5, 29, 30, 41, 43, 44], within family [24] or between family [6, 34]. Canalising selection was experimentally applied by Waddington [48] and by Sheiner and Lyman [40], their rule being the selection of individuals less sensitive to breeding temperature and by Gibson and Bradley [17] who applied a culling of extremes in a population bred in unstable environment (fluctuating temperature).

Some general conclusions from these experiments may be proposed: 1) very generally, stabilising selection is efficient, leading to a strong diminution of

phenotypic variance; 2) heritability estimations during and at the end of the selection experiments often showed that the selected trait genetic variance decreased, this conclusion not being general; 3) in many cases it was possible to prove that the environmental variance, or the sensitivity of individuals to environmental fluctuation, was reduced by selection.

In this paper we investigate mathematical tools for the evaluation of the possibility and efficiency of organising canalising selection in animal populations. Existence of a genetic component in variance heterogeneity between groups is a prerequisite for such a selection goal to be feasible. Statistical modelling and estimation procedures have been developed to take account of variance heterogeneity (e.g. [10, 11, 35, 36]), in particular using a logarithmic link between variances and predictive parameters [12, 13, 39].

In the following, we extend such models by introducing a genetic value among these parameters, consider the possibility of estimating this new genetic value, then discuss the efficiency of selection based on this model. Although our objective is to apply such methodologies to continuous and discrete traits, we first concentrate here on continuous traits. Applications to artificial canalising selection towards an economic optimum in goat and pig breeding are given.

2. GENETIC MODEL

2.1. Building of a model

Our approach was motivated by the extensive literature mentioned in the Introduction, and in particular the paper of Rendel and Sheldon [34] shows that artificial canalising selection does work, in the sense that the population mean reaches the optimum and, more importantly, the environmental variance is reduced. Some individuals are less susceptible to environment than others, this particularity being genetically controlled, since it responds to selection. Some genes are now known to control variability, e.g. the Apolipoprotein E locus [31] in humans, the Ubx locus in *Drosophila* [18], the dwarfism locus in chickens (Tixier-Boichard, pers. comm.), and some QTLs with effects on variance are already suspected [1].

Like Wagner et al. [50] in their equation 7, the effect of polymorphism at a given locus on the environmental variance may be expressed by a genotype-dependent multiplicative factor for this variance. The same hypotheses (in particular no interactions between genes) and reasoning as in the Fisher model allow the previous one-locus model to be extended to a polygenic or infinitesimal model, in which each individual has a genetic value governing a multiplicative factor for the environmental variance.

Since the analysis needs the evaluation of phenotypic variances associated with genetic values, it must be based on experimental designs allowing for the repeated expression of the same or of closely related genetic values. Although not necessarily efficient, any population scheme might be considered, but we focus here on two simple situations, repeated measurements on a single individual, and evaluation of one individual from the performances of its offspring.

2.2. Animal model: basic model

A model linking a phenotype y_j of a given animal (from repeated phenotypes $\boldsymbol{y} = (y_1, ..., y_j, ..., y_n)$) with two genetic values u and v is considered. According to the infinitesimal model of quantitative genetics, these genetic values u and v, possibly correlated, are assumed to be continuous normally distributed variables, and contribute to the mean and to the logarithm of the environmental variance. The simplest version of the model can be written as:

$$y_j = \mu + u + \exp\left(\frac{\eta + v}{2}\right)\varepsilon_j \qquad j = 1, ...n \tag{1}$$

where μ is the population mean and η the population log variance mean, while:

$$\begin{pmatrix} u \\ v \end{pmatrix} \sim \mathcal{N}_2\left[\mathbf{0}, \begin{pmatrix} \sigma_u^2 & r\sigma_u\sigma_v \\ r\sigma_u\sigma_v & \sigma_v^2 \end{pmatrix}\right] \tag{2}$$

and the ε_js are independent identically distributed $\mathcal{N}(0,1)$ Gaussian variables, independent of u and v. Additive genetic variances are denoted by σ_u^2 and σ_v^2, and r is the correlation coefficient between u and v. The distribution of the conditional random variable $Y|u, v$ is Gaussian $\mathcal{N}(\mu + u, \exp(\eta + v))$, but the unconditional distribution of Y is not. The unconditional mean and variance (the phenotypic variance σ_Y^2) of the random variable Y are equal to

$$E(Y) = \mu \tag{3}$$

$$Var(Y) = \sigma_u^2 + \exp\left(\eta + \frac{\sigma_v^2}{2}\right) = \sigma_Y^2 \tag{4}$$

Note that the v genetic value and its variance σ_v^2 are dimensionless; $\exp(\eta)$ has the same units as the phenotypic variance, and $\exp(\sigma_v^2/2)$ is the average (genetic) scale factor of the environmental variance.

2.3. Animal model: extensions

More general formulations of the model are needed to cope with real situations. First, introducing permanent environmental effects (denoted by p and t) common to several performances of the same individual is necessary to take account of non-genetically controlled correlations, both on the mean value – as it is usual to deal with repeatability – and on the log variance of the within performance environmental effect. Thus, the jth performance of an individual is modelled as:

$$y_j = \mu + u + p + \exp\left(\frac{\eta + v + t}{2}\right)\varepsilon_j \qquad j = 1, \ldots n \tag{5}$$

where (u, v), (p, t) and ε follow independent Gaussian distributions: the bivariate normal (2), a similar bivariate distribution with components σ_p^2, σ_t^2 and correlation ρ, and $\mathcal{N}(0,1)$, respectively.

When q individuals are measured in several environments, a more general heteroscedastic model can be stated as:

$$y_{ij} = \mathbf{x}_i'\boldsymbol{\beta}+\mathbf{z}_i'\boldsymbol{u}+\mathbf{z}_i'\boldsymbol{p}+\exp\left[\frac{1}{2}\left(\mathbf{q}_i'\boldsymbol{\delta}+\mathbf{z}_i'\mathbf{v}+\mathbf{z}_i'\boldsymbol{t}\right)\right]\varepsilon_{ij} \quad j=1,\ldots n_i \quad i=1,\ldots k \tag{6}$$

where y_{ij} is the jth performance of a particular animal in a particular (animal $\times$ environment) combination i. This full model (6) is a generalisation of model (1) introducing environmental and genetic parameters to be estimated: location parameters $(\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{p})$ and dispersion parameters $(\boldsymbol{\delta}, \mathbf{v}, \boldsymbol{t})$ with incidence matrices $(\mathbf{x}_i, \mathbf{z}_i, \mathbf{z}_i)$ and $(\mathbf{q}_i, \mathbf{z}_i, \mathbf{z}_i)$, respectively. Vectors $\boldsymbol{u}$, $\boldsymbol{p}$, $\mathbf{v}$ and $\boldsymbol{t}$ have the same length q. $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ denote fixed effects, while $\boldsymbol{u}$, $\mathbf{v}$ and $\boldsymbol{p}$, $\boldsymbol{t}$ are random genetic and random permanent environmental effects attached to individuals, respectively. The vectors of genetic values $\boldsymbol{u}$ and $\mathbf{v}$ have then a joint normal distribution:

$$\begin{pmatrix}\boldsymbol{u}\\ \mathbf{v}\end{pmatrix} \sim \mathcal{N}_{2q}\left[\mathbf{0}, \begin{pmatrix}\sigma_u^2 & r\sigma_u\sigma_v\\ r\sigma_u\sigma_v & \sigma_v^2\end{pmatrix}\otimes \mathbf{A}\right] \tag{7}$$

where $\otimes$ denotes the Kronecker product and $\mathbf{A}$ is the relationship matrix between the animals present in the analysis. Permanent environmental effects $\boldsymbol{p}$ and $\boldsymbol{t}$ are similarly distributed as:

$$\begin{pmatrix}\boldsymbol{p}\\ \boldsymbol{t}\end{pmatrix} \sim \mathcal{N}_{2q}\left[\mathbf{0}, \begin{pmatrix}\sigma_p^2 & \rho\sigma_p\sigma_t\\ \rho\sigma_p\sigma_t & \sigma_t^2\end{pmatrix}\otimes \boldsymbol{I}\right] \tag{8}$$

where $\boldsymbol{I}$ is the identity matrix, independently of $(\boldsymbol{u}, \mathbf{v})$.

This general way of setting up the model needs, however, some caution when applied to actual data, to assess which parameters are estimable, taking account of the structure of the experimental design. Specifically, analysing a possible genetic determinism of heteroscedasticity needs a sufficient number of repeated measures to be available for the same (or related) genotypes.

2.4. Sire model

In a progeny test scheme, the phenotypic values attached to an individual are the performances of its offspring. From the previous animal model, the performance y_{ij} of the jth offspring of sire i can be written as follows, conditional on the genetic values u_i and v_i of the sire and assuming unrelated dams:

$$y_{ij} = \mu + \frac{1}{2}u_i + \alpha_{ij} + p_i + \exp\left(\frac{\eta + \frac{1}{2}v_i + \beta_{ij} + t_i}{2}\right)\varepsilon_{ij} \tag{9}$$

It is assumed here that the terms α_{ij} and β_{ij} include the genetic effects in offspring not accounted for by the part transmitted by the sire. Permanent environmental effects in the offspring (the p and t variables of model 5 are possible.

This can be rewritten as

$$y_{ij} = \mu + \frac{1}{2}u_i + p_i + \left[\frac{3}{4}\sigma_u^2 + \exp(\eta + \frac{3}{8}\sigma_v^2 + \frac{1}{2}v_i + t_i)\right]^{1/2} \varepsilon_{ij} \qquad (10)$$

with $E(\varepsilon_{ij}) = 0$, $Var(\varepsilon_j) = 1$. The distribution of ε_j is only approximately normal $\mathcal{N}(0,1)$. Models (9) and (10) are not strictly equivalent, but, since the first two moments of y_j are equal under both models, they are equivalent in the sense of Henderson [21] (see e.g. [37] for an application of this concept). For example, for large numbers of offspring per sire, the mean sire's performances and sample within sire variances have asymptotically the same structure of variances and covariances between relatives under both models.

The corresponding generalised approximate sire model is written as

$$y_{ij} = \boldsymbol{x}_i'\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{z}_i'\boldsymbol{u} + \boldsymbol{z}_i'\boldsymbol{p} + \left[\frac{3}{4}\sigma_u^2 + \exp(\boldsymbol{q}_i'\boldsymbol{\delta} + \frac{3}{8}\sigma_v^2 + \frac{1}{2}\boldsymbol{z}_i'\boldsymbol{v} + \boldsymbol{z}_i'\boldsymbol{t})\right]^{1/2} \varepsilon_{ij}$$
$$j = 1, \ldots n_i \quad i = 1, \ldots k \qquad (11)$$

with the joint densities (7) for $\boldsymbol{u}$ and $\boldsymbol{v}$, and (8) for $\boldsymbol{p}$ and $\boldsymbol{t}$.

Methods needed to estimate parameters are outlined in Appendix A. In particular, they allow the genetic values of individuals to be estimated, as the conditional expectations of genetic values, given observed phenotypes $\boldsymbol{y}$: $\widehat{\boldsymbol{u}} = E(\boldsymbol{u}|\boldsymbol{y})$ and $\widehat{\boldsymbol{v}} = E(\boldsymbol{v}|\boldsymbol{y})$, if variance components are known. Estimation of variance components was similarly developed to make the method possible to apply.

In the following we first focus on developments of the basic model, which is simple enough to derive approximate analytical predictions of the response to selection and to compare several selection criteria. In a second step we check the validity of the theoretical approach by means of simulations and test the ability of the extended models and corresponding numerical procedures to tackle actual data and evaluate the potential for canalising selection.

3. SELECTION OBJECTIVE AND CRITERION

3.1. Objective and criterion

One objective that summarises the breeding goal (progeny performances close to the optimum and with low variability around it) is the minimisation of the expected squared deviation of offspring performances from the optimum y_0. This is the one we have chosen. For an individual characterised by a set $\boldsymbol{y}$ of performances (on itself and on its relatives), a selection criterion is defined as the expectation of the squared deviation $E\left[(Y_d - y_0)^2|\boldsymbol{y}\right]$ of offspring performance Y_d, conditional on $\boldsymbol{y}$, and selection will proceed by keeping individuals with minimal values of this index, such that:

$$E\left[(Y_d - y_0)^2|\boldsymbol{y}\right] = Var(Y_d|\boldsymbol{y}) + \left[E(Y_d|\boldsymbol{y}) - y_0\right]^2 \qquad (12)$$

is lower than a threshold $t(i)$ depending on the chosen selection intensity i.

In classical linear theory, it is equivalent to giving an individual a merit with respect to the selection objective, defined as the expectation of its offspring performance, or to consider its genetic value u, since the former is just equal to half the latter. Breeding animals are ranked according to their estimated genetic value.

In the present context, due to the non-linearity of the model, we define, for a candidate to selection with given genetic values u and v, its merit for canalising selection as the expected squared deviation of an *offspring performance*:

$$M^*(u,v) = E\left[(Y_d - y_0)^2 | u, v\right] \tag{13}$$

Its conditional expectation $E(M^*|\boldsymbol{y})$ is equal to the index

$$I^*(\boldsymbol{y}) = E\left[(Y_d - y_0)^2 | \boldsymbol{y}\right] \tag{14}$$

With complications due to the non-linear setting of our model, we derive in the following the mean and variance of an individual's phenotype distribution, conditional on the performances of a relative.

3.2. Conditional mean and variance

We need the distribution of a phenotype Y_d of a progeny d, given performances $\boldsymbol{y}$ of a relative F. Let u_d, v_d be the genetic values of d, $\boldsymbol{y} = \{y_j\}$, $j = 1, ...n$, u and v the phenotypic and genetic values of animal F. Performances of animals F and d follow model (1), with:

$$\begin{pmatrix} u \\ u_d \\ v \\ v_d \end{pmatrix} \sim \mathcal{N}_4 \left[\mathbf{0}, \begin{pmatrix} \sigma_u^2 & r\sigma_u\sigma_v \\ r\sigma_u\sigma_v & \sigma_v^2 \end{pmatrix} \otimes \begin{pmatrix} 1 & a \\ a & 1 \end{pmatrix} \right] \tag{15}$$

where a is the relationship coefficient between animals F and d ($a = 0.5$ if d is the progeny of F).

The density $f(y_d|\boldsymbol{y})$ describing the distribution of Y_d, conditional on $\boldsymbol{y}$ cannot be explicitly derived, but its moments are calculable or can be approximated. We have:

$$E(Y_d|\boldsymbol{y}) = \int y_d f(y_d|u_d, v_d) f(u_d, v_d|u, v) f(u, v|\boldsymbol{y})\, du_d\, dv_d\, du\, dv\, dy_d \tag{16}$$

This is first integrated over y_d, owing to

$$Y_d|u_d, v_d \sim \mathcal{N}\left[\mu + u_d, \exp(\eta + v_d)\right]$$

then with respect to u_d and v_d with

$$u_d, v_d|u, v \sim \mathcal{N}_2 \left[a \begin{pmatrix} u \\ v \end{pmatrix}, \ (1 - a^2) \begin{pmatrix} \sigma_u^2 & r\sigma_u\sigma_v \\ r\sigma_u\sigma_v & \sigma_v^2 \end{pmatrix} \right]$$

and finally the distribution of u and v conditional on $\boldsymbol{y}$ is approximated as:

$$u, v|\boldsymbol{y} \dot{\sim} \mathcal{N}_2 \left[\begin{pmatrix} \widehat{u} \\ \widehat{v} \end{pmatrix}, \begin{pmatrix} \widehat{C}_{uu} & \widehat{C}_{uv} \\ \widehat{C}_{vu} & \widehat{C}_{vv} \end{pmatrix} \right] \tag{17}$$

where $\widehat{u} = E(u|\boldsymbol{y})$, $\widehat{v} = E(v|\boldsymbol{y})$, $\widehat{C}_{uu} = Var(u|\boldsymbol{y})$, $\widehat{C}_{vv} = Var(v|\boldsymbol{y})$, $\widehat{C}_{uv} = Cov(u, v|\boldsymbol{y})$, are the estimated first and second moments of the genetic values (see Appendix A for the estimation method).

It follows that

$$E(Y_d|\boldsymbol{y}) = \mu + a\widehat{u} \tag{18}$$

and that

$$Var(Y_d|\boldsymbol{y}) \doteq (1 - a^2)\sigma_u^2 + a^2\widehat{C}_{uu} + \exp\left(\eta + \frac{1 - a^2}{2}\sigma_v^2 + a\widehat{v} + \frac{a^2}{2}\widehat{C}_{vv}\right) \tag{19}$$

These expressions are given numerical values after estimates of genetic values and of variance components are available.

General formulae can be derived that take into account all performances of the whole pedigree, not only performances of a single relative. The explicit forms of the extensions of equations (18) and (19) are given in Appendix B.

The combination of equations (18) and (19) gives the index $I^*(\boldsymbol{y})$ in equation (14), equal to the conditional expectation $E(M^*|\boldsymbol{y})$ of the genetic merit M^*, as in Goffinet and Elsen [20].

3.3. Approximate criteria

When the conditional variance terms $(\widehat{C})$ can be neglected, for instance when n is large, I^* is approximately equal to the maximum likelihood estimate of the merit M^*:

$$\begin{aligned} \widehat{I}(\boldsymbol{y}) &= M^*(\widehat{u}, \widehat{v}) \\ &= (\mu - y_0 + \frac{1}{2}\widehat{u})^2 + \frac{3}{4}\sigma_u^2 + \exp(\eta + \frac{3}{8}\sigma_v^2 + \frac{1}{2}\widehat{v}) \end{aligned} \tag{20}$$

where hats denote, in this case, modes of the density of $u, v|\boldsymbol{y}$. This is to be related to the work of Wilton et al. [51], who developed a quadratic index for a quadratic merit, by "minimising the expectation of the squared difference between total merit and index, both expressed as deviations from their expectations". In their setting, normality was assumed for the distributions of genetic values and of performances, so that this criterion was equal to the maximum likelihood estimate of the merit.

The previous calculations make it numerically possible to set up a selection scheme, but do not allow analytical predictions of the efficiency of selection according to the values of variance components σ_u^2, σ_v^2 and r. Some insight can be obtained using a simpler selection criterion, as follows.

In the individual model (1), assuming that repeated measures are available for the candidates for selection, we consider the following selection index I

$$I(\mathbf{y}) = (\overline{y} - y_0)^2 + S_y^2 \tag{21}$$

which is equal to the sample mean square deviation, $\overline{y}$ denoting the sample mean and S_y^2 the sample variance of the performance set of an individual, $S_y^2 = \frac{1}{n}\sum_{j=1}^{n}(y_j - \overline{y})^2$. Note, however, that this index measures the value of a candidate, not directly the expected value of its future offspring. Truncation selection would be accordingly characterised by a step fitness function w_t defined as:

$$w_t(\mathbf{y}) = I\!I\{I(\mathbf{y}) < t\} \tag{22}$$

Instead, we consider a continuous fitness function

$$w(\mathbf{y}) = 1 - s\, I(\mathbf{y}) \tag{23}$$

where s is a selection coefficient which can be adjusted to obtain the same selection differential as equation (22). The positivity of $w(y)$ in equation (23) necessitates a small s value. Hence we assume that selection is weak, allowing first-order approximation of the response to selection.

For progeny test selection the model for $\mathbf{y}$ is equation (10), but without p and t, and yields a similar selection index, $\mathbf{y}$ values being made up of the performances of the offspring of the candidate for selection. The selection criterion (21) is then a true measure of the candidate's value, and can be considered as an approximation of the criterion (12) for this simple population structure.

4. RESPONSE TO CANALISING SELECTION

We seek the responses to selection for the genotypic values u and v, the genetic merit, and the performance $(Y - y_0)^2$. We quantify the effects of selection by the regression of offspring on the selected parent (e.g. [9]), in a general way as:

$$b(w, X) = \frac{R(w, X)}{S(w, X)} = \frac{E_d^w(X) - E(X)}{E^w(X) - E(X)} \tag{24}$$

where X is any trait of interest, $E^w(X)$ its expectation in the selected part in the candidate population, and $E_d^w(X)$ the expectation of phenotypes among the offspring of the w-selected parents. The numerator is the response $R(w, X)$ to selection based on the fitness function w in the trait X of interest, measured in the next generation. The denominator is the selection differential $S(w, X)$, measured among parents. As a rule, we restrict the following derivations to selection in one sex only in the parent population.

4.1. Analytical approximations

4.1.1. Animal model

We first derive the distribution of u and v in the parent population after selection according to the fitness function w, then calculate the corresponding distribution in the offspring population.

Let $f(\boldsymbol{y})$ be the unconditional distribution of $\boldsymbol{Y}$, and $f(u,v)$ the joint density of u and v. The density of $\boldsymbol{Y}$ in the selected parental population is

$$f^w(\boldsymbol{y}) = \frac{w(\boldsymbol{y})f(\boldsymbol{y})}{\int w(\boldsymbol{y})f(\boldsymbol{y})d\boldsymbol{y}} \tag{25}$$

Following Gavrilets and Hastings [14], we introduce the mean fitness of the genotype (u,v):

$$\begin{aligned} wg(u,v) &= E[w(\boldsymbol{Y})|u,v] \\ &= 1 - s\,(\mu + u - y_0)^2 - s\,e^{\eta+v} \\ &= 1 - s\,M(u,v) \end{aligned} \tag{26}$$

As with $M^*(u,v)$ in equation (13), this function $M(u,v) = E(I(\boldsymbol{Y})|u,v)$ can be considered as a genetic merit referring to a candidate's own value and not as in equation (13) to that of a future offspring. The mean fitness of the population is the proportion of selected individuals:

$$\begin{aligned} \overline{w} &= E[w(\boldsymbol{Y})] \\ &= 1 - s\,E(M) \end{aligned}$$

where

$$E(M) = E(Y - y_0)^2 = (\mu - y_0)^2 + \sigma_u^2 + \exp\left(\eta + \frac{\sigma_v^2}{2}\right) \tag{27}$$

We obtain the distribution of genetic values among selected parents:

$$f^w(u,v) = \frac{wg(u,v)f(u,v)}{\overline{w}} \tag{28}$$

4.1.1.1. Genetic response

Since genetic values are transmitted linearly to the offspring, the genetic responses to selection, $R(w,u)$ and $R(w,v)$, are the differences of expected genotypic values u and v, respectively, between candidates and selected individuals (assuming that selection occurs in a single sex, only half of this progress is transmitted to the next generation):

$$\begin{aligned} R(wg,u) &= \int u f^w(u,v)du\,dv \\ &= -\frac{s}{\overline{w}}2(\mu - y_0)\sigma_u^2 - \frac{s}{\overline{w}}r\sigma_u\sigma_v\exp\left(\eta + \frac{\sigma_v^2}{2}\right) \end{aligned} \tag{29}$$

$$R(wg,v) = -\frac{s}{\overline{w}}\sigma_v^2\exp(\eta + \frac{\sigma_v^2}{2}) - \frac{s}{\overline{w}}2r\sigma_u\sigma_v(\mu - y_0) \tag{30}$$

where wg refers to equation (26). The effects of non-linearity are seen in the above equations.

Note that if genotypes are correlated (if r is not zero), the efficiency of selection is reduced if r and $(\mu - y_0)$ are of opposite signs.

4.1.1.2. Parent–offspring regression

The efficiency of individual canalising selection towards y_0 is evaluated by the regression coefficient (24) calculated for the trait $X = \Pi(Y) = (Y - y_0)^2$. The fact that the expectation of the trait Π of interest is equal to the expectation of the index I involved in the fitness function w defined in equation (23) makes the following derivations feasible. Summarising the detailed calculations given in Appendix C, we state that the numerator of equation (24) is equal to the w-selection response in the genetic merit M:

$$\begin{aligned} R(w, \Pi(Y)) &= E^w\left[E(\Pi(Y_d|\mathbf{y}))\right] - E[\Pi(Y)] \\ &= -\frac{s}{\overline{w}} Cov\left[I(\mathbf{Y}), I(Y_d)\right] \\ &= -\frac{s}{\overline{w}} Cov\left[M, M_d\right] \qquad (31) \\ &= R(w, M) \qquad (32) \end{aligned}$$

since $M = E(\Pi|u, v)$. The denominator of equation (24) is the selection differential:

$$\begin{aligned} S(w, \Pi(Y)) &= E^w[\Pi(Y)] - E[\Pi(Y)] \\ &= -\frac{s}{\overline{w}} Var[I(\mathbf{Y})] \\ &= -\frac{s}{\overline{w}} Var[E(I(\mathbf{Y})|u, v)] - \frac{s}{\overline{w}} E[Var(I(\mathbf{Y})|u, v)] \qquad (33) \end{aligned}$$

This leads to, if $r = 0$,

$$b\left[w, \Pi(Y)\right] = \frac{\frac{1}{2}\sigma_u^4 + 2(y_0 - \mu)^2\sigma_u^2 + V^2\left[\exp(\frac{1}{2}\sigma_v^2) - 1\right]}{2(1-\frac{1}{n})\sigma_u^4 + 4(1-\frac{1}{n})(y_0-\mu)^2\sigma_u^2 + (1+\frac{2}{n})V^2\left[\exp(\sigma_v^2) - 1\right] + \frac{2}{n}\sigma_Y^4 + \frac{4}{n}\sigma_Y^2(y_0-\mu)^2} \qquad (34)$$

where V stands for $\exp(\eta + \sigma_v^2/2)$. If genotypes are correlated, an extra term $2r\sigma_u\sigma_v V\left(\mu - y_0 + \frac{1}{4}r\sigma_u\sigma_v\right)$ is added to the numerator, and $4(1 + \frac{2}{n})r\sigma_u\sigma_v V \left(\mu - y_0 + \frac{1}{2}r\sigma_u\sigma_v\right)$ is added to the denominator.

The response to selection can be written as:

$$R(w, \Pi) = -\imath\, b(w, \Pi)\, \sigma_\Pi \qquad (35)$$

i.e. as the product of selection intensity $\left(\imath = \frac{s}{\overline{w}}\right)$, of a realized heritability, the ratio $b(w, \Pi)$ defined in equation (34), and of the standard deviation σ_Π of the selection index.

4.1.2. Sire model

As for the individual model, the genetic merit for the sire model is defined as:

$$M(u,v) = E\left[I(\mathbf{Y})|u,v\right] = \frac{3}{4}\sigma_u^2 + (\mu - y_0 + \frac{1}{2}u)^2 + \exp(\eta + \frac{3}{8}\sigma_v^2 + \frac{1}{2}v)$$

and the fitness

$$wg(u,v) = 1 - s\,M(u,v)$$

The expectation $E(M) = E[I(\mathbf{Y})]$ is the same as given in equation (27). The response to selection in the trait $\Pi(Y)$ among male parents is

$$E_d^w[\Pi(Y)] - E[\Pi(Y)] = -\frac{s}{\overline{w}}\left\{\frac{1}{32}\sigma_u^4 + \frac{1}{2}\sigma_u^2(\mu - y_0)^2 + V^2\left[\exp(\sigma_v^2/8) - 1\right]\right.$$
$$\left. + \frac{1}{2}r\sigma_u\,\sigma_v V\left[(\mu - y_0) + \frac{1}{16}r\sigma_u\,\sigma_v\right]\right\} \tag{36}$$

and the selection differential is

$$E^w[\Pi(Y)] - E[\Pi(Y)] = -\frac{s}{\overline{w}}\left\{\frac{1}{8}(1 - \frac{1}{n})\sigma_u^4 + \left(1 - \frac{1}{n}\right)\sigma_u^2(\mu - y_0)^2\right.$$
$$+ (1 + \frac{2}{n})V^2\left[\exp(\sigma_v^2/4) - 1\right] + \frac{2}{n}\sigma_Y^4 + \frac{4}{n}\sigma_Y^2(\mu - y_0)^2$$
$$\left. + (1 + \frac{2}{n})r\sigma_u\sigma_v V\left[(\mu - y_0) + \frac{1}{8}r\sigma_u\sigma_v\right]\right\} \tag{37}$$

The regression coefficient b giving the response to canalising selection in a progeny test scheme is equal to the ratio of (36) to (37). *Figure 1* plots the response given in equation (36) in units of selection intensity and phenotypic variance, from an equation similar to equation (35).

4.1.3. Extensions

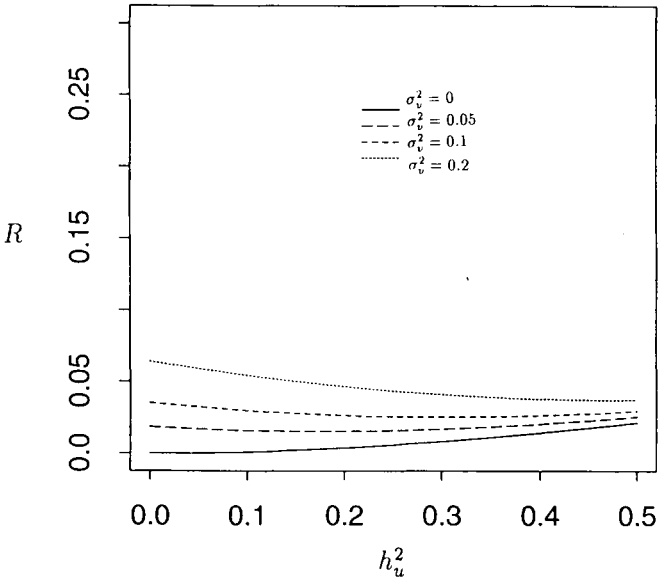
The previous *exact* results, obtained using the fitness function (23) and analogous for the sire model, hold for weak selection, and their expressions as ratios of a covariance to a variance indicate that they can also be obtained from a linear approximation. This comment makes it possible to extend easily the approximate prediction of response in cases when different weights are given to the variance of performances and to their deviation from the optimum.

Considering the animal model with repeated measurements (5), let us denote $\Pi_1(y) = (\overline{y} - y_0)^2$, $\Pi_2(y) = S_y^2$, the two components of $\Pi = (\Pi_1(y), \Pi_2(y))'$, $s = (s_1, s_2)'$ a vector of selective values, $\alpha = (\alpha_1, \alpha_2)'$ a vector of weights. We are interested in the response for the trait $\alpha'\Pi$, when using the index $s'\Pi$ as selection criterion. The parent–offspring regression is equal to

$$b = \frac{s'G\alpha}{s'P\alpha} \tag{38}$$

where G and P are 2×2 symmetric matrices of elements

(a)



(b)

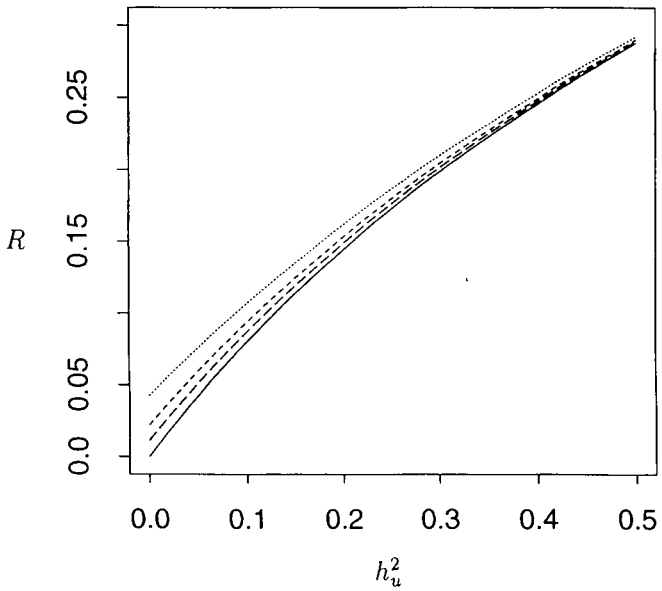


Figure 1. Expected one-generation response R to canalising selection in a progeny test scheme, with $n = 20$ observations per sire, $r = 0$, in units of selection intensity and phenotypic variance (equations (36)/(37)), with varying heritability h_u^2. (a) The population mean is at the optimum. (b) The population mean is 1 phenotypic standard deviation away from the optimum.

$$
\begin{aligned}
G_{11} = Cov(\Pi_{1,d}, \Pi_1) &\propto \frac{1}{2}h_u^4 + 2\lambda^2 h_u^2 \\
G_{12} = Cov(\Pi_{1,d}, \Pi_2) &= G_{21} = Cov(\Pi_{2,d}, \Pi_1) \\
&\propto rh_u\sigma_v(1-h_u^2)(\lambda + \frac{1}{4}rh_u\sigma_v) \\
G_{22} = Cov(\Pi_{2,d}, \Pi_2) &\propto (1-h_u^2)^2\left[\exp(\sigma_v^2/2) - 1\right] \\
P_{11} = Var(\Pi_1) &\propto 2c^4 + 4\lambda^2c^2 + \frac{4}{n}(1-h_u^2)^2(\lambda^2+c^2) \\
&\quad + \frac{1}{n^2}(1-h_u^2)^2\left[3\exp(\sigma_v^2) - 1\right] \\
&\quad + \frac{12}{n}rh_u\sigma_v(1-h_u^2)(\lambda + \frac{1}{2}rh_u\sigma_v) \\
P_{12} = P_{21} = Cov(\Pi_1, \Pi_2) &\propto 2(1-\frac{1}{n})rh_u\sigma_v(1-h_u^2)(\lambda + \frac{1}{2}rh_u\sigma_v) \\
&\quad + \frac{1}{n}(1-\frac{1}{n})(1-h_u^2)^2\left[\exp(\sigma_v^2/2) - 1\right] \\
P_{22} = Var(\Pi_2) &\propto (1-h_u^2)^2\left[(1-\frac{1}{n^2})\exp(\sigma_v^2+\sigma_t^2) - (1-\frac{1}{n})^2\right]
\end{aligned}
\tag{39}
$$

introducing the following notations $h_u^2 = \sigma_u^2/\sigma_Y^2$, $c^2 = (\sigma_u^2 + \sigma_p^2)/\sigma_Y^2$, $\lambda = (\mu - y_0)/\sigma_Y$. From equation (38), parent–offspring regressions for the mean and for the variance can be written separately. With $s_1 = 0$ and $\alpha_1 = 0$ for instance, b tends to

$$
\frac{1}{1+\exp(\sigma_v^2/2)} \tag{40}
$$

as n tends to infinity and if $\sigma_t^2 = 0$. This parent–offspring regression is lower than a half, and tends to $1/2$ as σ_v^2 tends to zero.

Note that the parent–offspring regression for $\overline{y}$ is

$$
\frac{1}{2}\frac{\sigma_u^2}{\sigma_u^2 + \sigma_p^2 + \dfrac{V}{n}} \tag{41}
$$

which tends to $1/2$ as n tends to infinity and if $\sigma_p^2 = 0$.

If the unbiased estimate of variance $\Pi_2' = \dfrac{1}{n-1}\sum_{j=1}^{n}(y_j - \overline{y})^2$ is used in the index, then the variance term $P_{22}' = Var(\Pi_{22}')$ is proportional to

$$
(1-h_u^2)^2\left[\left(1+\frac{2}{n-1}\right)\exp(\sigma_v^2+\sigma_t^2) - 1\right]
$$

When $\sigma_v^2 = 0$, the response in Π_2' is null and the selection differential is equal to $2/(n-1)$, taking into account $n-1$ degrees of freedom. For $n = 2$, it

corresponds to the variance of the trait $(Y - \overline{y})^2$ (squared deviation from the mean), up to a multiplicative term. When $\sigma_u^2 = 0$, the response in $\overline{Y}$ is null and the selection differential is equal to V/n,

More generally, this extension shows that a selection index (weights $s = (s_1, s_2)'$) can be adjusted to optimise the response in a given objective specified by weights $\alpha = (\alpha_1, \alpha_2)'$.

4.2. Simulations

Simulations were used to check the accuracy of previous analytical expressions of response as proposed in equations (34) and (36)/(37), in more general situations:

– intermediate selection intensity, since the analysis assumes only weak selection;

– behaviour of the population parameters (mean, variance) during several generations of selection;

– comparison of the relative efficiencies of different selection criteria, replacing in the simulation the theoretical continuous selection scheme (23) by truncation selection according to the simplified index (22) and by the likelihood based index $\widehat{I}$ (20).

Simulations were restricted to the case of the sire model with no genetic correlation ($r = 0$).

4.2.1. Selection scheme

The selection scheme was as follows.

1) Genetic values of sires and dams of the base population were randomly drawn from the joint distribution (7) with no relationships ($\boldsymbol{A}$ is the identity matrix), giving the sets $\{(u_i, v_i), i = 1, \ldots, S\}$ for the sires, and $\{(u_j, v_j), j = 1, \ldots, D\}$ for the dams.

2) Sires and dams were mated at random.

3) For each couple (i, j), the performance y_{ij} of a daughter was generated according to:

$$y_{ij} = \mu + \frac{u_i + u_j}{2} + \alpha_{ij} + \exp\left[(\eta + \frac{v_i + v_j}{2} + \beta_{ij})/2\right] \varepsilon_{ij}$$

where ε_{ij}, α_{ij} and β_{ij} were drawn from the Gaussian distributions $\mathcal{N}(0, 1)$, $\mathcal{N}(0, \sigma_u^2/2)$ and $\mathcal{N}(0, \sigma_v^2/2)$, respectively. The terms α_{ij} and β_{ij} represent Mendelian sampling.

4) An index for each sire was computed and elite sires were selected.

5) The elite sires produced S sons with the same female cohort used in steps 1–2.

Step 2 (with sons of step 5 and daughters of step 3) to step 5 were repeated until the 10th generation.

The sire selection of step 4 was a truncation selection based either on the simplified index $I(\boldsymbol{y}_i) = (\overline{y}_i - y_0)^2 + S_i^2$ or on the maximum likelihood estimate

of the merit $\widehat{I}(\mathbf{y}_i) = M(\widehat{u}_i, \widehat{v}_i) = \frac{3}{4}\sigma_u^2 + \exp(\eta + \frac{\widehat{v}_i}{2} + \frac{3}{8}\sigma_v^2) + (\mu - y_0 + \frac{\widehat{u}_i}{2})^2$, with $\widehat{u}_i$ and $\widehat{v}_i$ maximum likelihood estimates of u_i and v_i respectively, according to model (10), but allowing for no permanent environmental effect, and assuming that variance components were known.

4.2.2. Simulation experiments

For a constant phenotypic standard deviation for the base population, several values of variance components were tested: $\sigma_u^2 = 0.033$ and 0.114, corresponding to a 'low' ($h_u^2 = 0.10$) and a 'high' heritability ($h_u^2 = 0.3$); a 'low' variability variance $\sigma_v^2 = 0.03$ and a 'high' $\sigma_v^2 = 0.15$ (corresponding to ratios of maximum to minimum variance equal to 3 and 10, respectively). Three base phenotypic means were considered: $\mu_{t=0} = 1$, 1.8 and 2, for an optimum equal to $y_0 = 2$ (giving discrepancies $\lambda_{t=0} = (\mu_{t=0} - y_0)/\sigma_{Y,t=0}$ between population mean and optimum, expressed in phenotypic standard deviations, equal to 1.75, 0.35 and 0).

For given values of the set σ_u^2, σ_v^2, μ and y, of the numbers of sires and dams and of selection intensity, 100 selection experiments were performed, and statistics averaged over the runs. The evolution of phenotypic mean and variance, estimated merit over the ten generations and parent–offspring regression are highlighted.

4.2.3. Results

Figure 2 displays the curves given by the analytical approximation of the response, with point estimates and confidence intervals obtained with 100 simulated selection experiments, showing good agreement of the approximation with truncation selection on the simplified index I (not shown), but also with the likelihood based index $\widehat{I}$, except for intermediate values of λ for which the theory provides underestimates.

Figure 3 plots the evolution of phenotypic means and standard deviations over generations of canalising selection. Several aspects appear:

– with a high heritability h_u^2, the population mean tends in a linear manner towards the optimum in a very efficient way;

– the convergence of the mean is slightly better if σ_v^2 is low;

– the decrease in phenotypic variance has a linear tendency, although more fluctuating than the evolution of the mean;

– this decrease is even more evident as σ_v^2 is higher and h_u^2 is lower.

This general balance was encountered throughout the simulation experiments: a particular aspect was maximally improved when the other aspects were not under selection pressure. Variances are best reduced when the population mean is at the optimum. The optimum is more rapidly reached when no genetic variability of the variances is present.

Figure 4 compares the performances of the two indices I and $\widehat{I}$. The likelihood based index gives more efficient results for the trait mean μ_t, probably because heterogeneous variances were taken into account in the evaluation of the animal genetic values u, giving less biased estimates. On the contrary, the

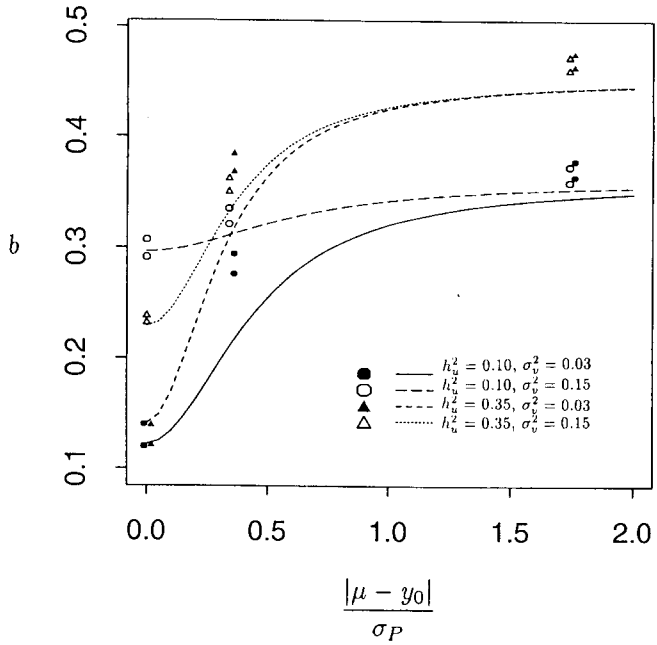


Figure 2. Parent–offspring regression b based on the approximate index I and continuous fitness (lines) and on truncation selection with the index $\widehat{I}$ (confidence intervals over 100 replicates at size 95 %). The selection scheme involves 500 sires (candidates for selection) and 100 elite sires.

phenotypic variance $\sigma^2_{Y,t}$ is best reduced with the simplified index, presumably due to the lack of robustness of v estimation by maximum likelihood. A full Bayesian estimation procedure with marginal posterior expectation of parameters might be more appropriate. It was nevertheless not performed because of the heaviness of the algorithm, since numerical integrations are then needed. The two indices give, however, equal values of the global criterion $(\mu_t - y_0)^2 + \sigma^2_{Y,t}$ at any time t.

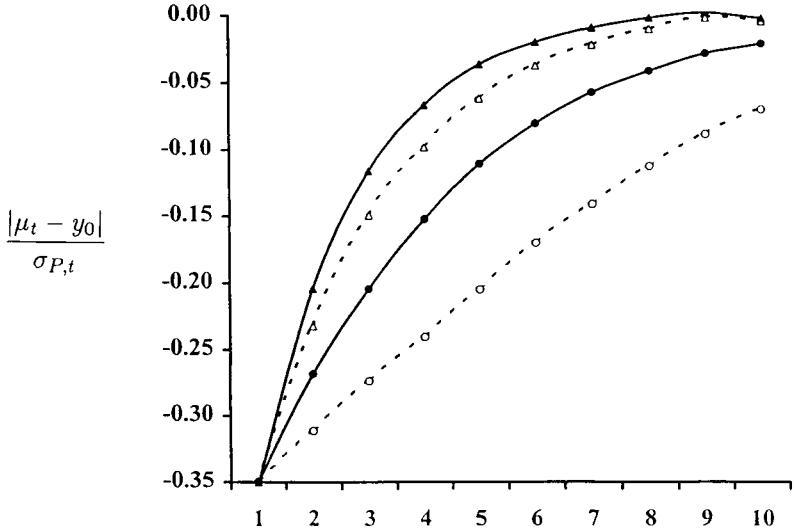
The phenotypic variance and squared difference between mean and optimum are lowered more and more as selection intensity is increased, while the parent–offspring regression remains constant in the simulations as in the approximate theory (not shown).

5. DISCUSSION

5.1. Model for the variance

The introduction of a log linear model is an easy way to handle a multiplicative model on the variance. It is known that the distribution of $\ln S^2$, the logarithm of the sample variance estimator, is approximately normal (e.g. [25]). Similarly, Bayesian considerations on prior/posterior densities show that the Gaussian distribution is a good approximation to a log inverted chi-square

(a)



(b)

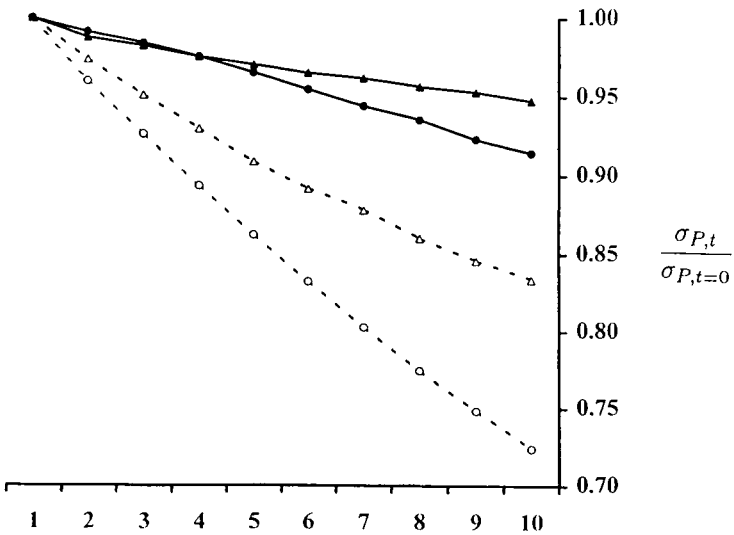


Figure 3. Response to selection. Population mean (a) and standard deviation (b) over ten generations of truncation canalising selection with index $\widehat{I}$. The selection scheme involves 500 sires (candidates for selection) and 100 elite sires. The mean at generation 0 is 0.35 phenotypic standard deviation away from the optimum. Plain line: $\sigma_v^2 = 0.03$; dotted line: $\sigma_v^2 = 0.15$. Circles: $h_u^2 = 0.1$; triangles: $h_u^2 = 0.35$.

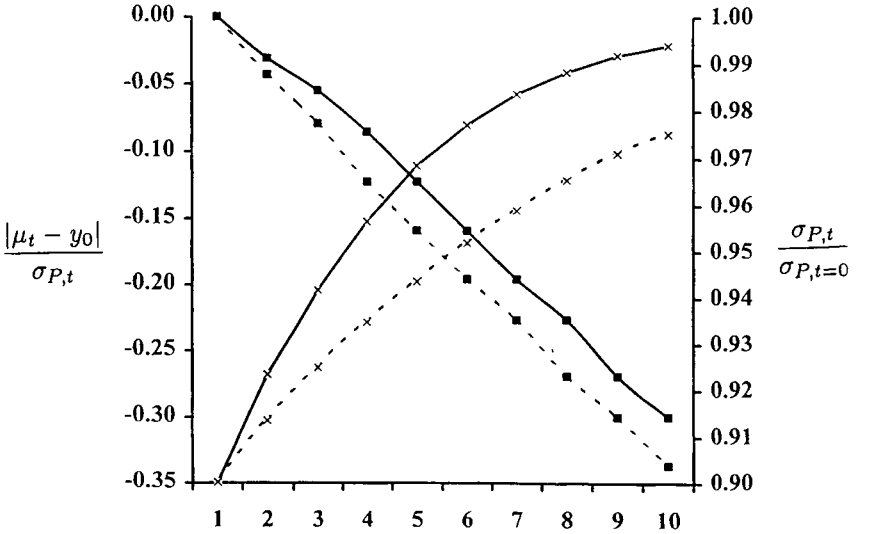


Figure 4. Approximate (I, dotted line) versus optimal ($\widehat{I}$, plain line) index selection. The selection scheme involves 500 sires (candidates for selection) and 100 elite sires. The mean at generation $t = 0$ is 0.35 phenotypic standard deviation away from the optimum, $h_u^2 = 0.1$ and $\sigma_v^2 = 0.03$. Squares: $(\mu_t - y_0)/\sigma_{Y,t}$, left scale; stars: $\sigma_{Y,t}/\sigma_{Y,t=0}$, right scale.

(see [13]). This led us to focus all analytical derivations on the first two moments of distributions, assimilating when needed any distribution to the Gaussian distribution sharing these same moments. Although this may be a crude approximation if it is used for prediction of genetic response over several generations, it allows first order solutions to be derived, and makes it possible to build statistical evaluation procedures.

The model allows estimation of the importance of genetic determinism in the heterogeneity of variances, and hence prediction of how the population may respond to selection against variability. For example, the proportion of the selection response due to the genetic variability in the v-component is given by the ratio

$$\frac{R(w, \Pi_2)}{R(w, \Pi)} = \frac{G_{22}}{G_{11} + G_{22} + 2G_{12}}$$

where the Gs are given in equation (40). It is all the more important as the population mean is closer to the optimum, the u-genetic variance is lower, and the v-genetic variance is larger.

Estimation of genetic parameters (σ_u^2, r, σ_v^2) may be somewhat imprecise, especially for σ_v^2 and r. Hence it may be worth considering the robustness of predictions with respect to badly known parameters. As far as a simple global criterion is used, the question can be dealt with easily, considering the expected responses as functions of parameter values. The situation would be more difficult to handle for selection schemes that would rely on the knowledge

of parameter values, for example if a balance between selection for the mean or for the variance were adjusted each generation.

5.2. Data

The generalised version of the sire model (11), including fixed and random permanent environmental effects, was applied to actual data in goats (dairy production) and in pigs (pH of muscles after slaughtering).

5.2.1. Milk data

Protein and fat contents were measured on milk from 2 383 first lactation goats between 1992 and 1995. The goats were daughters of 54 artificial insemination sires, with 20 observations at least in the data set. The trait of interest is the ratio of fat to protein contents, with a desired optimum equal to 1.3. This objective would be complementary to yield traits such as milk yield or protein yield. The phenotypic mean and variance are equal to 1.1 and 0.0135, respectively, i.e. the population mean is 1.7 phenotypic standard deviations away from the optimum. Data are normally distributed. For computational ease, data were pre-corrected with the additive model including herd, season, lactation length and age, on a much larger data set including all lactations of all herds where the 2 383 kept daughters had been producing. The variance components were estimated, leading to a null correlation coefficient ($\widehat{r} \simeq 0$) and zero variability variance ($\widehat{\sigma}_v^2 \simeq 0$), and a heritability $\widehat{h}_u^2 = 0.44$ of the same order as those for the protein and fat.

A canalising selection experiment is expected to drive the population mean rapidly towards the optimum, but without change in environmental variance. For example, assuming selection of the best 10 % of sires, a reduction of 1.5 phenotypic standard deviations of the population quadratic deviation $(\mu - y_0)^2$ would be expected in one generation.

5.2.2. pH data

pH values of semi-membranous muscle were measured on 947 piglets from 25 Large White sires. Data were normally distributed. Each sire had at least 20 piglets. Data were pre-corrected by the usual linear model accounting for sex, line, year and slaughtering date effects on the trait mean, on a much larger data set, in order to simplify further computations. Thereafter, a sire model for the residuals of the previous model was fitted.

Estimated values of variance components under model (11) with 'permanent environmental effects' (non-genetic-sire effects) were equal to $\widehat{\sigma}_v^2 = 0.15$, $\widehat{h}_u^2 = 0.26$ (with $\widehat{\sigma}_Y^2 = 0.037$), $\widehat{r} = 0.79$, $\sigma_p^2 = 0.00045$, $\sigma_t^2 = 0.046$ and $\rho = 0.79$. With an optimum value $y_0 \simeq 5.7$ not different from the overall mean $\mu = 5.75$, the estimated variance components should allow a high response to canalising selection to be obtained through a strong reduction of the genetically controlled part of environmental variance: assuming that selection sorts out the best 10 % of male parents, a reduction of about 12 % of the initial phenotypic variance in one generation. A null correlation would give a reduction of 11 % (*figure 1*).

It must be stressed that predictions derived from the above analysis of fat to protein ratio in goats and of pig pH muscle data are only indicative. For example, the effect of a wrongly estimated correlation value r remains to be assessed, even if – in the goat example – no significant genetic component of variance was found for variances. Also, although precision of the previous early estimates was not evaluated, larger data sets are probably needed. A proper prediction of expected response to selection cannot be proposed until these analyses are carried out.

So far we do not have results from an actual selection experiment, based on our index selection rules, which would be necessary to completely validate the approach through the comparison of observed realised heritabilities with our predictions. It is one of the perspectives of the current work to organise such selection experiments.

5.3. Selection criteria

We have considered a single global criterion that combines selection for the mean and selection against the variance of the trait.

Shnol and Kondrashov [42] considered the action of selection with fitness $w(y)$ on a quantitative trait y. They concluded that truncation selection minimises the genetic load and the variance of the trait after selection. Linear selection (corresponding to our continuous fitness with low selection) gives minimal variance of the relative fitness and is less efficient than truncation selection. However, linear selection gave us the opportunity for robust analytical approximations of realised heritability. Calculations were impossible for truncation selection, even with the simpler index. Within the limits of the present comparisons with simulations, the fitness approximation proved useful, even in cases with strong departure from linearity, and with a rather strong selection intensity (proportion of selected individuals equal to 20 %).

More sophisticated selection criteria may be defined, allowing selection to be differentially directed towards changing the mean value of the trait or reducing the environmental variance. In fact using a global genetic merit to be maximised in the next generation is a way to distribute selection intensity between both parameters. It is possible that a higher multi-generation response could be obtained if selection were controlled each generation in view of the objective. For example, the index $(\overline{y} - y_0)^2 + S_y^2$ can be generalised into $s_1(\overline{y} - y_0)^2 + s_2 S_y^2$, allowing a greater selection pressure either on the location near the optimum, or on the dispersion, as illustrated in the above theoretical section. The same remark is available for the index $(E(Y_d|\boldsymbol{y}) - y_0)^2 + Var(Y_d|\boldsymbol{y})$. More generally, the selection criterion might be based on the economic worth of offspring. The criterion would then be defined as the expected economic value of offspring, a function depending on the distribution of expected phenotypes and on the economic value of phenotypic values. But of course other types of indices and mating systems are potentially interesting to consider, for instance a linear index when σ_v^2 is small, mate selection or group selection.

Managing the balance between location and scale could be interesting in a long-term selection process, provided some analytical approximation is available in order to include one-generation expected response in a dynamic programming approach. Evaluation of the approximation for mid-term objectives

remains, however, to be considered. While the present paper focused on short-term selection (one generation), such developments would require some analytical approximation of the response during several generations. At variance with the present work, changes in genetic variances and covariances should be taken into account. Further research is needed in this area, keeping in mind that the approach used, according to which most distributions are approximated by the Gaussian ones that share the same first and second moments, is known to be a rather poor approximation in genetic models as soon as multi-generation problems are considered. It may be, however, a useful approach for predictions over five to ten generations ([7]).

Another extension of this work concerns discrete characters. For example, a concrete demand of sheep breeders is obtaining exactly two lambs per lambing, with reduced variability around this economic optimum (SanCristobal-Gaudy et al., in prep.). More generally, the innovation of this work – the introduction of two groups of polygenes, possibly not independent, acting respectively on the trait mean and log variance – could be useful in other areas of applied quantitative genetics in which heterogeneities of variance arise. Also, while the two sources of genetic variability were studied within the framework of the infinitesimal model, extensions might include major genes which control either the mean or the variability of a trait. For example, using the present setting, a segregation analysis could be conducted to decide whether polygenes and/or major genes act on the log variance, as was carried out for the mean [26].

ACKNOWLEDGEMENTS

We thank Eduardo Manfredi, who kindly read the manuscript, and Pascale Le Roy with Thierry Tribout for providing milk and pH data, respectively. We also would like to thank Robin Thompson, Daniel Sorensen, Philippe Baret, Agustin Blasco and Alan Templeton for interesting discussions on the topic.

REFERENCES

[1] Allain D., Lantier I., Elsen J.M., Francois D., Brunel J.C., Weisbecker J.L., Schibler L., Vaiman D., Cribiu E., Gautier A., Berthon P., Lantier F., A design aiming at detecting QTL controlling wool traits and other traits in the INRA401 sheep line, in: Proceedings of the 6th World Congress on Genetics Applied to Livestock Prodution, Armidale, NSW, Australia, 11–16 January, vol. 24, 1998, pp. 51–54.

[2] Bulmer M.G., The effect of selection on genetic variability, Am. Nat. 105 (1971) 201–211.

[3] Bulmer M.G., The effect of selection on genetic variability: a simulation study, Genet. Res. 28 (1976) 101–117.

[4] Bulmer M.G., Maintenance of genetic variability by mutation-selection balance: a child's guide through the jungle, Genome 31 (1989) 761–767.

[5] Campo J.L., Garcia Gill M., Assortative mating and directional or stabilizing selection for a non-linear function of traits in *Tribolium*, J. Anim. Breed. Genet. 110 (1993) 74–80.

[6] Cardin S., Minvielle F., Selection on phenotype variation of pupa weight in *Tribolium Castaneum*, Can. J. Genet. Cytol. 28 (1986) 856–861.

[7] Chevalet C., An approximate theory of selection assuming a finite number of quantitative trait loci, Genet. Sel. Evol. 26 (1994) 379–400.

[8] Falconer D.S., The problem of environment and selection, Am. Nat. 86 (1952) 293–298.

[9] Falconer D.S., Introduction to Quantitative Genetics, 3rd ed., Longman Scientific and Technical, Essex, UK, 1989.

[10] Foulley J.L., Quaas R.L., Heterogeneous variances in Gaussian linear mixed models, Genet. Sel. Evol. 27 (1995) 211–228.

[11] Foulley J.L., Hébert D., Quaas R.L., Inferences on homogeneity of between family components of variance and covariance among environments in balanced cross-classified designs, Genet. Sel. Evol. 26 (1994) 117–136.

[12] Foulley J.L., Gianola D., San Cristobal M., Im S., A method for assessing extent and sources of heterogeneity of residual variances in mixed linear models, J. Dairy Sci. 73 (1990) 1612–1624.

[13] Foulley J.L., San Cristobal M., Gianola D., Im S., Marginal likelihood and Bayesian approaches to the analysis of heterogeneous residual variances in mixed linear Gaussian models, Comput. Stat. Data Anal. 13 (1992) 291–305.

[14] Gavrilets S., Hastings A., A quantitative genetic model for selection on developmental noise, Evolution 48 (1994) 1478–1486.

[15] Gavrilets S., Scheiner S.M., The genetics of phenotypic plasticity. V. Evolution of reaction norm shape, J. Evol. Biol. 6 (1993) 31–48.

[16] Gavrilets S., Scheiner S.M., The genetics of phenotypic plasticity. VI. Theoretical predictions for directional selection, J. Evol. Biol. 6 (1993) 49–68.

[17] Gibson J.B., Bradley B.P., Stabilising selection in constant and fluctuating environments, Heredity 33 (1974) 293–302.

[18] Gibson G., Hogness D.S., Effect of polymorphism in the *Drosophila* regulatory gene Ultrabithorax on homeotic stability, Science 271 (1986) 200–203.

[19] Gillepsie J.H., Turelli M., Genotype-environment interaction and the maintenance of polygenic variation, Genetics 121 (1989) 129–138.

[20] Goffinet B., Elsen J.M., Critère optimal de sélection : quelques résultats généraux, Génét. Sél. Evol. 16 (1984) 307–318.

[21] Henderson C.R., Applications of Linear Models in Animal Breeding, University of Guelph, Guelph, Canada, 1984.

[22] Hohenboken W.D., The manipulation of variation in quantitative traits: a review of possible genetic strategies, J. Anim. Sci. 60 (1985) 101–110.

[23] Imasheva A.G., Zhivotovskii L.A., Lazebnyi O.E., Coupled effects of directional and stabilizing selection, Biol. Sci. 309 (1990) 729–732.

[24] Kaufman P.K., Enfield F.D., Comstock R.E., Stabilizing selection for pupa weight in *Tribolium castaneum*, Genetics 87 (1977) 327–341.

[25] Layard M.W.J., Large sample tests for homogeneity of variances, J. Am. Stat. Assoc. 68 (1973) 195–198.

[26] Le Roy P., Méthodes de détection de gènes majeurs. Application aux animaux domestiques, thèse, Université de Paris-Sud, Orsay, France, 1989.

[27] Lerner I.M., Genetic Homeostasis, Oliver and Boyd, Edinburgh, 1954.

[28] Numerical Algorithms Group, The NAG Fortran Library Manual, NAG Ltd, Oxford, 1990.

[29] Piniero R., Selection for canalization at two dorsocentral bristles in *Drosophila melanogaster*, J. Heredity 83 (1992) 49–55.

[30] Prout T., The effects of stabilizing selection on the time of development in *Drosophila melanogaster*, Genet. Res. Camb. 3 (1962) 364–382.

[31] Reilly S.L., Ferrell R.E., Kottke B.A., Kamboh M.I., Sing C.F., The gender-specific apolipoprotein E genotype influence of the distribution of lipids and apolipoproteins in the population of Rochester, MN. I. Pleiotropic effects on means and variances, Am. J. Hum. Genet. 49 (1991) 1155–1166.

[32] Rendel J.M., Canalisation and selection, in: Thompson Jr J.N., Thoday J.M. (Eds.),Quantitative Genetic Variation, Academic Press, Inc., 1979, pp. 139–156.

[33] Rendel J.M., Sheldon B.L., Finlay D.E., Selection for canalization of the scute phenotype. II, Am. Nat. 100 (1966) 13–31.

[34] Rendel J.M., Sheldon B.L., Selection for canalization of the scute phenotype in *Drosophila melanogaster*, Austr. J. Biol. Sci. 13 (1960) 36–47.

[35] Robert C., Foulley J.L., Ducrocq V., Genetic variation of trait measured in several environments. I. Estimation and testing of homogeneous genetic and intraclass correlations between environments, Genet. Sel. Evol. 27 (1995) 111–123.

[36] Robert C., Foulley J.L., Ducrocq V., Genetic variation of trait measured in several environments. II. Inference on between-environments homogeneity of intraclass correlations, Genet. Sel. Evol. 27 (1995) 125–134.

[37] Robert C., Étude de quelques problèmes liés à la mise en œuvre du REML en génétique quantitative, thèse, Université Paul Sabatier, Toulouse, France, 1996.

[38] Robertson A., The effect of selection against extreme deviants based on deviation or on homozygosis, J. Genet. 54 (1956) 236–249.

[39] San Cristobal M., Foulley J.L., Manfredi E., Inference about multiplicative heteroskedastic components of variance in a mixed linear Gaussian model with an application to beef cattle breeding, Genet. Sel. Evol. 25 (1993) 3–30.

[40] Scheiner S.M., Lyman R.F., The genetics of phenotypic plasticity. II Response to selection, J. Evol. Biol. 4 (1991) 23–50.

[41] Scharloo W., Hoogmoed M.S., Ter Kuile A., Stabilizing and disruptive selection on a mutant character in drosophila. I. The phenotypic variance and its components, Genetics 56 (1967) 709–726

[42] Shnol E.E., Kondrashov A.S., Some relations between different characteristics of selection, J. Math. Biol. 32 (1994) 835–840.

[43] Soliman M.H., Directional and stabilizing selection for developmental time and correlated response in reproductive fitness in *Tribolium castaneum*, Theor. Appl. Genet. 63 (1982) 111–116.

[44] Tantawi A.O., Tayel A.A., Studies on natural populations of *Drosophila*. X. Effects of disruptive and stabilising selection on wing length and the correlated response in *Drosophila melanogaster*, Genetics 65 (1969) 121–132.

[45] Via S., Lande R., Genotypic-environment interaction and the evolution of phenotypic plasticity, Evolution 39 (1985) 505–522.

[46] Via S., Lande R., Evolution of genetic variability in a spatially heterogeneous environment: effects of genotype-environment interaction, Genet. Res. 49 (1987) 147–156.

[47] Waddington C.H., The Strategy of the Genes, Cambridge University Press, Cambridge, 1957.

[48] Waddington C.H., Experiments on canalizing selection, Genet. Res. 1 (1960) 140–150.

[49] Waddington C.H., Robertson E., Selection for developmental canalization, Genet. Res. 7 (1966) 303–312.

[50] Wagner G.P., Booth G., Bagheri-Chaichian H., A population genetic theory of canalization, Evolution 51 (1997) 329–347.

[51] Wilton J.W., Evans D.A., Van Vleck L.D., Selection indices for quadratic models of total merit, Biometrics 24 (1968) 937–949.

[52] Wright S., The analysis of variance and the correlations between relatives with respect to deviations from an optimum, J. Genet. 30 (1935) 243–256.

APPENDIX A: Parameter estimation

The genetic evaluation of the animals needs the expectation of $\boldsymbol{u}$ and $\boldsymbol{v}$ given performances $\boldsymbol{y}$ of relatives, namely $\widehat{\boldsymbol{u}} = E(\boldsymbol{u}|\boldsymbol{y})$ and $\widehat{\boldsymbol{v}} = E(\boldsymbol{v}|\boldsymbol{y})$, which depend on the variance components. Approximations are obtained by replacing expectations by modes, since $\boldsymbol{u}$ and $\boldsymbol{v}$ are Gaussian, and $\boldsymbol{y}$ is nearly so, and using a Newton-Raphson iterative scheme, which involves first (ω) and second ($\boldsymbol{C}^{-1}$) derivatives of the log likelihood. At iteration t, the current estimate of $\boldsymbol{\tau} = (\boldsymbol{\beta}', \boldsymbol{u}', \boldsymbol{p}', \boldsymbol{\delta}', \boldsymbol{v}', \boldsymbol{t}')'$ is equal to

$$\widehat{\boldsymbol{\tau}}^{[t]} = \widehat{\boldsymbol{C}}^{[t-1]}\widehat{\boldsymbol{\omega}}^{[t-1]} + \widehat{\boldsymbol{\tau}}^{[t-1]} \tag{42}$$

Most derivatives involved in $\boldsymbol{C}$ and ω can be found in e.g. [13]. The extension to permanent environmental effects presents no difficulties. Note that in the present context, dispersion parameters $\boldsymbol{v}$ are of prime interest, as well as location parameters $\boldsymbol{u}$, and so are estimated together.

When the variance components are unknown, they can be estimated by their conditional expectation $\widehat{\boldsymbol{\sigma}}^2 = E(\boldsymbol{\sigma}^2|\boldsymbol{y})$, with $\boldsymbol{\sigma}^2 = (\sigma_u^2, \sigma_v^2, r, \sigma_p^2, \sigma_t^2, \rho)$. An EM algorithm can be proposed for the estimation of variance components $\boldsymbol{\sigma}^2$, which are replaced by their current estimates $\widehat{\boldsymbol{\sigma}}^{2[t-1]}$ in $\widehat{\boldsymbol{C}}^{[t-1]}$ and $\widehat{\boldsymbol{\omega}}^{[t-1]}$ in the iterative system (42). Equations relative to the animal model (6)–(7) have known forms:

$$\begin{aligned}
\widehat{\sigma}_u^{2[t]} &= \frac{1}{q}\left(\widehat{\boldsymbol{u}}'\boldsymbol{A}^{-1}\widehat{\boldsymbol{u}} + \mathrm{tr}(\boldsymbol{A}^{-1}\widehat{\boldsymbol{C}}_{uu})\right)^{[t]} \\
\widehat{\sigma}_v^{2[t]} &= \frac{1}{q}\left(\widehat{\boldsymbol{v}}'\boldsymbol{A}^{-1}\widehat{\boldsymbol{v}} + \mathrm{tr}(\boldsymbol{A}^{-1}\widehat{\boldsymbol{C}}_{vv})\right)^{[t]} \\
\widehat{\sigma}_{uv}^{[t]} &= \frac{1}{q}\left(\widehat{\boldsymbol{u}}'\boldsymbol{A}^{-1}\widehat{\boldsymbol{v}} + \mathrm{tr}(\boldsymbol{A}^{-1}\widehat{\boldsymbol{C}}_{uv})\right)^{[t]} \\
\widehat{\sigma}_p^{2[t]} &= \frac{1}{q}\left(\widehat{\boldsymbol{p}}'\widehat{\boldsymbol{p}} + \mathrm{tr}(\widehat{\boldsymbol{C}}_{pp})\right)^{[t]} \\
\widehat{\sigma}_t^{2[t]} &= \frac{1}{q}\left(\widehat{\boldsymbol{t}}'\widehat{\boldsymbol{t}} + \mathrm{tr}(\widehat{\boldsymbol{C}}_{tt})\right)^{[t]} \\
\widehat{\sigma}_{pt}^{[t]} &= \frac{1}{q}\left(\widehat{\boldsymbol{p}}'\widehat{\boldsymbol{t}} + \mathrm{tr}(\widehat{\boldsymbol{C}}_{pt})\right)^{[t]}
\end{aligned} \tag{43}$$

where $\widehat{\boldsymbol{u}}^{[t]}$ (resp. $\widehat{\boldsymbol{v}}^{[t]}$) are the current estimates of $\boldsymbol{u}$ (resp. $\boldsymbol{v}$),

$$\widehat{\boldsymbol{C}} = \begin{pmatrix} \widehat{\boldsymbol{C}}_{uu} & \widehat{\boldsymbol{C}}_{uv} & \widehat{\boldsymbol{C}}_{up} & \widehat{\boldsymbol{C}}_{ut} \\ \widehat{\boldsymbol{C}}_{vu} & \widehat{\boldsymbol{C}}_{vv} & \widehat{\boldsymbol{C}}_{vp} & \widehat{\boldsymbol{C}}_{vt} \\ \widehat{\boldsymbol{C}}_{pu} & \widehat{\boldsymbol{C}}_{pv} & \widehat{\boldsymbol{C}}_{pp} & \widehat{\boldsymbol{C}}_{pt} \\ \widehat{\boldsymbol{C}}_{tu} & \widehat{\boldsymbol{C}}_{tv} & \widehat{\boldsymbol{C}}_{tp} & \widehat{\boldsymbol{C}}_{tt} \end{pmatrix} \tag{44}$$

and $\sigma_{uv} = r\sigma_u\sigma_v$, $\sigma_{pt} = \rho\sigma_p\sigma_t$.

EM equations for variance components in a sire model need numerical integration and are now presented.

The log likelihood of the sire model (11) is written as

$$\begin{aligned}\mathcal{L} = &-\frac{1}{2}\sum_{i=1}^{k}\left(\frac{\mathbf{e}_i'\mathbf{e}_i}{\sigma_i^2} + n_i \ln\sigma_i^2\right)\\ &-\frac{1}{2(1-r^2)}\left(\frac{\mathbf{u}'\mathbf{A}^{-1}\mathbf{u}}{\sigma_u^2} - 2r\,\frac{\mathbf{u}'\mathbf{A}^{-1}\mathbf{v}}{\sigma_u\sigma_v} + \frac{\mathbf{v}'\mathbf{A}^{-1}\mathbf{v}}{\sigma_v^2}\right)\\ &-\frac{q}{2}\ln\left[\sigma_u^2\,\sigma_v^2\,(1-r^2)\right] + const\end{aligned}$$

where $\mathbf{e}_i = \{y_{ij} - \mathbf{x}_i'\boldsymbol{\beta} - \frac{1}{2}\mathbf{z}_i'\mathbf{u}\}_{j=1,\ldots,n_i}$

and $\sigma_i^2 = \frac{3}{4}\sigma_u^2 + \exp\left[\mathbf{p}_i'\boldsymbol{\delta} + \frac{3}{8}\sigma_v^2 + \frac{1}{2}\mathbf{q}_i'\mathbf{v}\right]$

If there is an overall mean effect in $\boldsymbol{\delta}$, the term $\frac{3}{8}\sigma_v^2$ is part of the constant term. We consider this parameterisation in the following.

Variance components are estimated by maximisation of the marginal log likelihood l_m (or the marginal posterior distribution $p(\boldsymbol{\sigma}^2|\mathbf{y})$ in a Bayesian setting):

$$l_m = \ln p(\boldsymbol{\sigma}^2|\mathbf{y}) = \ln\int p(\boldsymbol{\sigma}^2, \boldsymbol{\tau}|\mathbf{y})\,d\boldsymbol{\tau} \tag{45}$$

Because explicit integration with respect to $\boldsymbol{\delta}$ and $\mathbf{v}$ is not analytically feasible (to our knowledge), we chose the implementation of an iterative algorithm involving first $\dot{l}_m$ and second $\ddot{l}_m$ derivatives of l_m (e.g. [12]):

$$\dot{l}_m = E_c\left[\frac{\partial}{\partial\boldsymbol{\sigma}^2}\mathcal{L}\right]$$

$$\ddot{l}_m = E_c\left[\frac{\partial}{\partial\boldsymbol{\sigma}^2}\mathcal{L}\right] + \mathrm{Var}_c\left[\frac{\partial^2}{\partial(\boldsymbol{\sigma}^2)^2}\mathcal{L}\right]$$

where E_c and Var_c denote expectation and variance, respectively, with respect to $\boldsymbol{\tau}|\mathbf{y}, \boldsymbol{\sigma}^2$. Let us denote $\dot{l}_{m,\sigma^2}$ (resp. $\ddot{l}_{m,\sigma^2,\sigma'^2}$) the element of the vector $\dot{l}_m$ (resp. matrix $\ddot{l}_m$) pertaining to σ^2 (resp. σ^2 and σ'^2), with $\sigma^2 = \sigma_u^2$, σ_v^2 or r: $\dot{l}_{m,\sigma^2} = E_c\left[\frac{\partial}{\partial\sigma^2}\mathcal{L}\right]$.

As for the animal model, the estimation procedure proceeds in two steps: at each iteration, a current estimate $\widehat{\boldsymbol{\tau}}^{[t]}$ of $\boldsymbol{\tau}$ is obtained in solving the linear system (42), in which variance components are replaced by their current estimates $\widehat{\boldsymbol{\sigma}}^{2[t]}$. Then, variance component estimates are updated, as detailed in the following. At convergence, $\widehat{\boldsymbol{\sigma}}^2$ maximises the marginal log likelihood (45), and maximum a posteriori estimates of $\boldsymbol{\tau}$ are obtained as a by-product.

At iteration t, the current estimate $\widehat{r}^{[t+1]}$ of the correlation r satisfies the equation $\dot{l}_{m,r} = 0$. It is the solution, lying in $[-1, 1]$, of the third order equation

$$-qr^3 + r^2 E_c^{[t]}\Big(\frac{\boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{v}}{\sigma_u\sigma_v}\Big) + r\left[q - E_c^{[t]}\Big(\frac{\boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{u}}{\sigma_u^2} + \frac{\boldsymbol{v}'\boldsymbol{A}^{-1}\boldsymbol{v}}{\sigma_v^2}\Big)\right] + E_c^{[t]}\Big(\frac{\boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{v}}{\sigma_u\sigma_v}\Big) = 0 \quad (46)$$

The current standard deviation $\widehat{\sigma}_v^{[t+1]}$ is also obtained directly from the equation $\dot{l}_{m,\sigma_v^2} = 0$. It is the positive solution of a second order equation:

$$\widehat{\sigma}_v^{[t+1]} = \frac{1}{2q(1-r^2)}\left[-rE_c^{[t]}\Big(\frac{\boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{v}}{\sigma_u}\Big) + \sqrt{(\widehat{\Delta}^{[t]}}\Big)\right] \quad (47)$$

where $\Delta = r^2 E_c^2\Big(\frac{\boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{v}}{\sigma_u}\Big) + 4q(1-r^2)E_c(\boldsymbol{v}'\boldsymbol{A}^{-1}\boldsymbol{v})$.

No explicit solution is found for the equation $\dot{l}_{m,\sigma_u^2} = 0$. So second derivatives $\ddot{l}_m$ are used to provide an iterative solution. More exactly, we chose to implement an EM-type algorithm involving only the E_c-part of the Hessian matrix. The current variance $\widehat{\sigma}_u^{2[t+1]}$ is equal to

$$\widehat{\sigma}_u^{2[t+1]} = \widehat{\sigma}_u^{2[t]} - \widehat{\dot{l}}_{m,\sigma_u^2}^{[t]} / \widehat{\ddot{l}}_{m,\sigma_u^2,\sigma_u^2}^{[t]} \quad (48)$$

with

$$\begin{aligned}
\dot{l}_{m,\sigma_u^2} &= \frac{3}{8}E_c\left[\sum_i\Big(\frac{\boldsymbol{e}_i'\boldsymbol{e}_i}{\sigma_i^2} - n_i\Big)/\sigma_i^2\right] \\
&\quad + \frac{1}{2\sigma_u^3(1-r^2)}E_c\Big(\frac{\boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{u}}{\sigma_u} - r\,\frac{\boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{v}}{\sigma_v}\Big) - \frac{q}{2\sigma_u^2} \\
\ddot{l}_{m,\sigma_u^2,\sigma_u^2} &= -\frac{9}{32}E_c\left[\sum_i\Big(\frac{2\boldsymbol{e}_i'\boldsymbol{e}_i}{\sigma_i^2} - n_i\Big)/\sigma_i^4\right] \\
&\quad - \frac{1}{2\sigma_u^5(1-r^2)}E_c\Big(2\frac{\boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{u}}{\sigma_u} - \frac{3}{2}r\,\frac{\boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{v}}{\sigma_v}\Big) + \frac{q}{2\sigma_u^4}
\end{aligned}$$

Numerical integration is performed for the expectations in $\widehat{\dot{l}}_{m,\sigma_u^2}$ and $\widehat{\ddot{l}}_{m,\sigma_u^2,\sigma_u^2}$ indexed by i: vectors $\boldsymbol{\tau}_s$ are randomly drawn from a $\mathcal{N}(\widehat{\boldsymbol{\tau}}^{[t]}, \widehat{\boldsymbol{C}}^{[t]})$ (which has an asymptotic justification) and these expectations are approximated by

$$\frac{1}{S}\sum_i\sum_{s=1}^{S}\Big(\frac{\boldsymbol{e}_{i,s}'\boldsymbol{e}_{i,s}}{\sigma_{i,s}^2} - n_i\Big)/\sigma_{i,s}^2$$

$$\frac{1}{S}\sum_i\sum_{s=1}^{S}\Big(\frac{2\boldsymbol{e}_{i,s}'\boldsymbol{e}_{i,s}}{\sigma_{i,s}^2} - n_i\Big)/\sigma_{i,s}^4$$

When permanent environment effects are present in the model, the estimation equations are the same as those in the animal model (equation 43).

A Fortran 77 program was written, using the NAG library [28], and is available on request.

APPENDIX B: Conditional mean and variance of progeny phenotypes

$\boldsymbol{u}$ and $\boldsymbol{v}$ denote k-vectors of genotypic values of all animals considered, related by the relationship matrix $\boldsymbol{A}$. Future offspring of these animals have $\boldsymbol{u_d}$ and $\boldsymbol{v_d}$ genotypic values, and are related by $\boldsymbol{A_{dd}}$. Genotypic values of parents and offsprings are related via $\boldsymbol{A_d}$.

It can be shown that the conditional expectation of a performance $Y_{d,i}$ of a future offspring of some animal i of the parent population is equal to

$$E(Y_{d,i}|\boldsymbol{y}) = \mu + (\boldsymbol{A}_d\boldsymbol{A}^{-1}\widehat{\boldsymbol{u}})_i = \mu + \sum_{j=1}^{k}(\boldsymbol{A}_d\boldsymbol{A}^{-1})_{i,j}\widehat{u}_j \tag{49}$$

and the variance given the performances of all the animals is

$$\begin{aligned} Var(Y_{d,i}|\boldsymbol{y}) &= (\boldsymbol{A}_{dd} - \boldsymbol{A}_d\boldsymbol{A}^{-1}\boldsymbol{A}_d)_{ii}\sigma_u^2 + (\boldsymbol{A}_d\boldsymbol{A}^{-1}\widehat{\boldsymbol{C}}_{uu}\boldsymbol{A}^{-1}\boldsymbol{A}_d')_{ii} \\ &+ \exp\left[\eta + \frac{1}{2}(\boldsymbol{A}_{dd} - \boldsymbol{A}_d\boldsymbol{A}^{-1}\boldsymbol{A}_d)_{ii}\sigma_v^2 + (\boldsymbol{A}_d\boldsymbol{A}^{-1}\widehat{\boldsymbol{v}})_i \right. \\ &\left. + \frac{1}{2}(\boldsymbol{A}_d\boldsymbol{A}^{-1}\widehat{\boldsymbol{C}}_{vv}\boldsymbol{A}^{-1}\boldsymbol{A}_d')_{ii}\right] \end{aligned} \tag{50}$$

where $\widehat{u}$s and $\widehat{v}$s are parts of equation (42), and $\widehat{\boldsymbol{C}}$s are submatrices of equation (44). Note that all the individuals in the analysis are involved in these formulae.

APPENDIX C: Derivations of $R(w,\Pi)$ and $S(w,\Pi)$ (equations (31)–(34) in the text)

First, we write

$$f^w(\boldsymbol{y}) = \frac{1 - sI(\boldsymbol{y})}{\overline{w}} f(\boldsymbol{y})$$

Now,

$$\begin{aligned} R(w,\Pi) &= R(w,I) \\ &= E^w\left[E\left(I(Y_d)|y\right)\right] - E\left(I(Y)\right) \\ &= E\left[\frac{1 - sI(y)}{\overline{w}} E\left(I(Y_d)|y\right)\right] - E\left(I(Y)\right) \\ &= \frac{1}{\overline{w}} E\left[E\left(I(Y_d)|y\right)\right] - \frac{s}{\overline{w}} E\left[I(Y)I(Y_d)\right] - \frac{\overline{w}}{\overline{w}} E\left(I(Y)\right) \\ &= -\frac{s}{\overline{w}} Cov\left[I(Y), I(Y_d)\right] \end{aligned}$$

with $\overline{w} = 1 - s\,E(I) = 1 - s\,E(M)$. Furthermore,

$$\begin{aligned} E[I(Y_d)I(\boldsymbol{Y})] &= E\left[(I(Y)I(Y_d))\,|u, v, u_d, v_d\right] \\ &= E[M(u,v)M(u_d,v_d)] \end{aligned}$$

so

$$R(w,\Pi) = -\frac{s}{\overline{w}} Cov[M, M_d]$$

This covariance is equal to

$$Cov(u^2, u_d^2) + 4(\mu - y_0)^2 Cov(u, u_d) + \exp(2\eta) Cov(\exp(v), \exp(v_d))$$

plus covariances that are null if $r = 0$. The above three covariances are respectively equal to

$$\frac{1}{2}\sigma_u^4, \quad \frac{1}{2}\sigma_u^2, \quad \exp(\sigma_v^2)[\exp(\sigma_v^2/2) - 1]$$

leading to the numerator of equation (34).

The selection differential is equal to

$$\begin{aligned} S(w,\Pi) &= S(w, I) \\ &= \int I(\mathbf{y}) f^w(\mathbf{y})\, d\mathbf{y} - E(I) \\ &= (\frac{1}{\overline{w}} - 1)E(I) - \frac{s}{\overline{w}} E(I^2) \\ &= -\frac{s}{\overline{w}} Var(I) \end{aligned}$$

This variance is calculated from the two terms

$$Var\left\{E[I(\mathbf{y})|u,v]\right\} + E\left\{Var[I(\mathbf{Y})|u,v]\right\}$$

with

$$I(\mathbf{Y}) = \frac{1}{n}\sum_j Y_j^2 - 2y_0 \frac{1}{n}\sum_j Y_j + y_0^2$$

and Y_j following the Gaussian model (1). One has $E(I|u,v) = M(u,v)$, then

$$Var\left\{E[I(\mathbf{Y})|u,v]\right\} = Var(M) = 2\sigma_u^4 + 4(\mu - y_0)^2 \sigma_u^2 + V^2[\exp(\sigma_v^2) - 1]$$

on one hand, and on the other hand

$$Var(I|u,v) = \frac{1}{n}\left[4(\mu + u - y_0)^2 \exp(\eta + v) + 2\exp(2\eta + 2v)\right]$$

then the expectation of the above expression equals

$$\frac{1}{n}\left[4(\mu - y_0)^2 V + 4\sigma_u^2 V + 2V^2 \exp(\sigma_v^2)\right]$$

Putting these terms together leads to the denominator of equation (34).