Predicting cumulated response to directional selection in finite panmictic populations

E. Verrier *, J. J. Colleau and J. L. Foulley

INRA, Station de Génétique Quantitative et Appliquée, F-78350 Jouy-en-Josas, France

Summary. Accurate prediction of the cumulated genetic gain requires predicting genetic variance over time under the joint effects of selection and limited population size. An algorithm is proposed to quantify at each generation the effects of these factors on average coefficient of inbreeding, genetic variance, and genetic mean, under a purely additive polygenic model, with no mutation, and under the assumption of absence of inbreeding depression on viability affecting selection differentials. This algorithm is relevant to populations where mating is at random and generations do not overlap. It was tested via Monte Carlo simulation on a population of 3 males and 25 females mass selected out of 50 candidates of each sex, over 30 generations. For two values of the initial heritability of the selected trait, 0.5 and 0.9 (to represent high accuracy in index selection), predicted values of the genetic variance are in agreement with observed results up to the 12th and 19th generations, respectively. Beyond these generations, the variance is overestimated, due to an underestimation of the effect of selection on the rate of inbreeding. Finally, the algorithm provides predictions of the cumulated responses close to the observed values in both selected populations. It is concluded that, as regards the hypotheses of the study, the proposed algorithm is satisfactory, and could be used to optimize selection methods with respect to the cumulated genetic gain in the mid- or long-term. Possible extensions of the algorithm to more realistic situations are discussed.

Key words: Genetic variance – Genetic gain – Inbreeding – Selection – Additive infinitesimal model

* Current address: Institut National Agronomique P-G, Département des Sciences Animales, 16 rue Claude Bernard, F-75231 Paris Cedex 05, France

Introduction

Mid- and long-term responses to selection depend on changes occurring in additive-genetic variance. Accurate prediction over time is, therefore, highly dependent on the quality of prediction for genetic variance. This is all the more true when very high genetic gains are expected, e.g., in some Multiple Ovulation and Embryo Transfer selection systems (Nicholas and Smith 1983; Ruane and Thompson 1989; Colleau and Mocquot 1989). The problem addressed here will be that for a fully polygenic and additive trait, i.e., a very large number of independent additive loci control the genetic variance of this trait.

In this genetic context, directional selection modifies the genetic variance, first by inducing linkage disequilibrium [as first mentioned by Lush, (1945, pp. 141–143) and studied by Bulmer (1971)], and second by enhancing loss of variation through inbreeding if the population is of limited size (Lush 1946; Robertson 1961).

This last effect depends on the genetic variance being submitted to selection and, therefore, depends on time. Some experimental work shows that the effect of selection on the rate of inbreeding decreases with time (Kownacki et al. 1981, 1987). Such a phenomenon was not taken into account in the first algorithms proposed to predict the change in genetic variance with respect to time (Dempfle 1975; Keightley and Hill 1987; Chevalet 1988), since they used the asymptotic concept of effective size and since this value is computed as under pure drift. Robertson (1961) and Burrows (1984a, b) derived formulae to compute the effective size when one generation of selection occurs, and Wray and Thompson (1990) derived a method to compute the asymptotic rate of inbreeding under selection. These algorithms, however, assume the effect of selection on inbreeding to be constant over time, or consider its value only after a large number of generations.

The purpose of this paper is to propose a prediction algorithm overcoming this drawback, limited to a situation with separate generations. A Monte Carlo simulation was used on a mass-selection case to measure the accuracy of the proposed algorithm.

Theory

Hypotheses

The quantitative trait is assumed to be genetically determined by an additive infinitesimal model generating a normal distribution of genetic values at each generation. Environmental effects are also assumed to be normally distributed and independent of genetic values. Generations do not overlap. The population size is constant over generations, with N_m sires and N_f dams. It is assumed that inbreeding does not affect viability or fertility of the animals, nor does it affect the values for the selected trait (it is strictly additive). Total panmixy is assumed, i.e., polygyny and polyandry are allowed, and matings are at random. The population is closed and mutations are not included in the model.

Prediction of the genetic variance

At each generation, the genetic variance (VA) is partitioned into a between-full-sibs family component and a within-family component (VA_w). The first term depends on variances in parents, VA_s and VA_d, respectively, for sires and dams. Due to sampling of parents with replacement, the total genetic variance is:

$$VA^{[t+1]} = \frac{1}{4}\left(1 - \frac{1}{N_m}\right) VA_s^{[t]} + \frac{1}{4}\left(1 - \frac{1}{N_f}\right) VA_d^{[t]} + VA_w^{[t+1]} \tag{1}$$

where the superscripts refer to the generation number at birth.

Assuming that selection criteria of candidates are normally distributed and are independent variates, the genetic variance among selected parents is:

$$VA_s^{[t]} = VA^{[t]} (1 - K_s R_s^{2[t]}) \tag{2a}$$

$$VA_d^{[t]} = VA^{[t]} (1 - K_d R_d^{2[t]}) \tag{2b}$$

with $K = i(i-x)$, where i is the standardized factor for intensity of selection and x is the truncation point, and where R^2 is the square of the correlation between the selection index and the true genetic value. As shown by Bulmer (1971), with such a polygenic determinism of the trait, the reduction of variance in the selected parents corresponds to an increase of linkage disequilibrium. This phenomenon is statistically described by an equation from Pearson (1903), which gives the change in the variance-covariance matrix of a vector, X_2, when a change in the variance-covariance matrix of a vector, X_1, occurs:

$$\Sigma_{22}^* = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} (\Sigma_{11} - \Sigma_{11}^*) \Sigma_{11}^{-1} \Sigma_{12} \tag{3}$$

where Σ_{ij} represents the variance-covariance matrix between vectors X_i and X_j before selection, and the asterisks refer to the corresponding matrix after selection. Under the assumption of normality of the distributions, applying Eq. (3) to the change in genetic variance after one generation of selection leads to Eq. (2).

The within-family variance is related to the variance of gene effects in the previous generation, which is not affected by selection (Crow and Kimura 1970, pp. 236–239) but only by drift, and to the probability of gene identity at the loci involved (Langlois 1990). In a pure drift situation, and in the case of an additive model, Foulley and Chevalet (1981) proved that for any pair of parents, VA_w is a function of the genetic variance (V_0) in the base population (assumed to have reached its linkage equi-

librium) and of the inbreeding coefficients of the sire (F_s) and of the dam (F_d):

$$VA_w = \frac{1}{2} V_0 \left(1 - \frac{F_s + F_d}{2}\right). \tag{4}$$

If the number of loci is not large, Eq. (4) does not hold under selection, due to a change in gene frequencies and because the probability of gene identity at any selected locus is greater than the usual inbreeding coefficient, which is defined for neutral genes (Malécot 1948). However, Langlois (1990) showed that, even under selection, VA_w tends to Eq. (4), provided that the number of loci is large enough. Verrier et al. (1989) checked this result via simulation for a situation of very intense mass selection. The small discrepancies they observed were due to a change in gene frequencies, and they concluded that Eq. (4) holds well under selection, within the limits of the additive infinitesimal model. Furthermore, we assume that in a given generation, selection does not affect significantly the average coefficient of inbreeding of the selected parents ($F_s^{[t]} = F_d^{[t]} = F^{[t]}$, where $F^{[t]}$ is the average coefficient of inbreeding of the candidates born at generation t). Therefore,

$$VA_w^{[t+1]} = \frac{1}{2} V_0 (1 - F^{[t]}). \tag{5}$$

Prediction of the average inbreeding coefficient

The proposed method is related to the method described by Burrows (1984a, b), and generalizes it to more than one generation. By definition, $F^{[t]}$ is equal to the average coefficient of relationship (Φ) (Malécot 1948) among mates born and selected at generation $t-1$ ($\Phi[s_{t-1}, d_{t-1}]$). Under the conditions assumed (sampling with replacement), the probability of finding the pair (i, j) among the parents of the next generation corresponds to the probability that both individuals will be selected at the same time. The probability that the second mate is selected given that the first one is selected depends on the correlation between the breeding values of the possible mates. In the population of candidates at generation $t-1$, four categories of pairs (indexed by k) can be found, with an a priori probability π_k, depending on the identity of their parents born and selected at $t-2$:

same sire, same dam $\qquad \pi_1 = \frac{1}{N_m} \frac{1}{N_f}$ (6a)

same sire, different dams $\qquad \pi_2 = \frac{1}{N_m}\left(1 - \frac{1}{N_f}\right)$ (6b)

different sires, same dam $\qquad \pi_3 = \left(1 - \frac{1}{N_m}\right)\frac{1}{N_f}$ (6c)

different sires, different dams $\qquad \pi_4 = \left(1 - \frac{1}{N_m}\right)\left(1 - \frac{1}{N_f}\right)$. (6d)

For simplicity's sake, we will assume that within each class of pairs the coefficients of relationship do not vary and are equal to Φ_k. Due to their probabilistic definition given by Malécot (1948), the Φ_k's at generation $t-1$ are computed from the average coefficients of relationship at generation $t-2$, between sires ($\Phi[s_{t-2}, s_{t-2}]$), between dams ($\Phi[d_{t-2}, d_{t-2}]$), and between sires and dams ($\Phi[s_{t-2}, d_{t-2}] = F^{[t-1]}$):

$$\Phi_1^{[t-1]} = \frac{1}{2}\Phi[s_{t-2}, d_{t-2}] + \frac{1}{4}\left(\frac{1 + F_s^{[t-2]}}{2}\right) + \frac{1}{4}\left(\frac{1 + F_d^{[t-2]}}{2}\right)$$

$$= \frac{1}{2}F^{[t-1]} + \frac{1}{4}(1 + F^{[t-2]}) \tag{7a}$$

$$\Phi_2^{[t-1]} = \frac{1}{2}F^{[t-1]} + \frac{1}{8}(1 + F^{[t-2]}) + \frac{1}{4}\Phi[d_{t-2}, d_{t-2}] \tag{7b}$$

$$\Phi_3^{[t-1]} = \frac{1}{2}F^{[t-1]} + \frac{1}{8}(1 + F^{[t-2]}) + \frac{1}{4}\Phi[s_{t-2}, s_{t-2}] \tag{7c}$$

$$\Phi_4^{[t-1]} = \frac{1}{2}F^{[t-1]} + \frac{1}{4}\Phi[s_{t-2}, s_{t-2}] + \frac{1}{4}\Phi[d_{t-2}, d_{t-2}]. \tag{7d}$$

After selection, the average coefficient of relationship is:

$$\Phi[s_{t-1}, d_{t-1}] = F^{[t]} = \frac{\sum\limits_{k=1}^{4} \pi_k Q_k^{[t-1]} \Phi_k^{[t-1]}}{\sum\limits_{k=1}^{4} \pi_k Q_k^{[t-1]}} \tag{8}$$

where Q_k is the conditional probability that a random pair of the category k will be selected simultaneously. If ϱ_k represents the correlation between the values for the selection criterion in the male (I_m) and the female (I_f) candidates with the category k, Q_k is:

$$Q_k = pr\left(\frac{I_m}{\sigma_{I_m}} \geq x_s \quad \text{and} \quad \frac{I_f}{\sigma_{I_f}} \geq x_d \mid \varrho_k\right),$$

where σ_I is the standard deviation of the selection criteria. Following classical notation (Johnson and Kotz 1972) Q_k can be written:

$$Q_k^{[t]} = L(x_s, x_d; \varrho_k^{[t]}).$$

The computation of the correlations (ϱ_k) is rather difficult because of the limited size of the population, which implies negative terms in the covariance in each category. For example, the covariance between mates of the fourth category is a priori negative and not null, as it would be if the population size was infinite. The computation of the ϱ_k's is described in detail in the Appendix for the case of mass selection, and its extension for other kinds of selection criterion is outlined. Conditional probabilities (Q_k) are computed by numerical integration of a bivariate normal distribution using Dutt's method (see Ducrocq and Colleau 1986, for details). These calculations are well suited to relatively large populations but, in small populations, they ignore classical order and joint order statistics, which implies that a random event affecting a male i, e.g., affects all the pairs involving this male.

The same method is used to compute the average coefficients of relationship between selected males $(\Phi[s, s])$ and between selected females $(\Phi[d, d])$, which are used in Eq. (7), probabilities Q_k being $L(x_s, x_s; \varrho_k)$ and $L(x_d, x_s; \varrho_k)$, respectively. The F values obtained by the proposed algorithm can be compared with the observed F, during Monte Carlo simulations, and the F obtained in a pure drift situation, which can simply be computed from the equation 7.1.4 of Crow and Kimura (1970, p. 231):

$$F^{[t]} = \frac{1}{2N_e} + \left(1 - \frac{1}{N_e}\right) F^{[t-1]} + \frac{1}{2N_e} F^{[t-2]} \tag{9}$$

with

$$N_e = \frac{4 N_m N_f}{N_{m+} N_f}.$$

Algorithms

Three algorithms using the same expression for VA [Eqs. (1), (3), and (5)] but differing by the F value used, can be compared: (i) the proposed algorithm, denoted (C), where F is computed from Eq. (8); (ii) a "pure drift" algorithm, denoted (N), where F is computed from Eq. (9); (iii) a pseudo-prediction algorithm, denoted (T), where F is the average value observed in the simulations.

Simulation processes

The initial genetic variance (V_0) was chosen as unity. Genetic values in this initial generation were, therefore,

randomly chosen out of an $N(0,1)$ distribution. For the other generations, the genetic value of an offspring (A_i) was generated from the values of its sire (A_s) and dam (A_d) according to the following formula:

$$A_i = \frac{1}{2} A_s + \frac{1}{2} A_d + \alpha_i \left[\frac{1}{2} V_0 \left(1 - \frac{F_s + F_d}{2}\right)\right]^{1/2} \tag{10}$$

The corresponding phenotypic values (Y_i) were:

$$Y_i = A_i + \beta_i \sigma_E \tag{11}$$

In Eqs. (10) and (11), α_i and β_i are two independent random numbers taken from an $N(0,1)$ distribution, and σ_E is the environmental standard deviation, assumed constant over generations. Equation (4) was used for predicting the within-family variance in the simulation algorithm according to the theoretical results of Langlois (1990) and the simulation results of Verrier et al. (1989), both showing the adequacy of Foulley and Chevalet's (1981) equation within the additive infinitesimal model, i.e., under the hypotheses of this study. The simulation process does not generate data according to the approximation made in Eq. (5) for the parental inbreeding coefficients. Therefore, any comparison between observed and predicted results would be affected by possible errors originating from this approximation.

At each generation, 3 males and 25 females were selected, on the basic of their own performance, out of 50 animals of each sex. The parents of each offspring were randomly sampled with replacement in the list of the selected possible parents. The corresponding effective size (computed as with pure drift) was 10.7. At each generation, and for each replicate, the genetic mean $(\bar{A})$ and the genetic variance (VA) were computed from the individual values. The average inbreeding coefficient was computed from individual coefficients, for all the offspring, and for the males and the females selected as parents for the next generation. Three values of initial heritability (h_0^2) were considered: 10^{-6}, 0.5, and 0.9; the corresponding populations were denoted S_0, S_5, and S_9, respectively. The population S_0 was intended to simulate pure drift results. The population S_9 was used to test situations where the accuracy of selection could be very high. Each simulation set was run with 500 replicates over 30 generations.

Comparison between observed and predicted results

Figure 1 shows the change in genetic variance (VA) in the three simulated populations. An important decrease in VA from generations 0–1 was observed in the selected populations. As pointed out by Bulmer (1971), the higher h_0^2, the larger is this initial decrease: -22% in S_5, -36% in S_9. On the other hand, in the population S_0, managed as with pure drift, the initial decrease was only -6.5%.
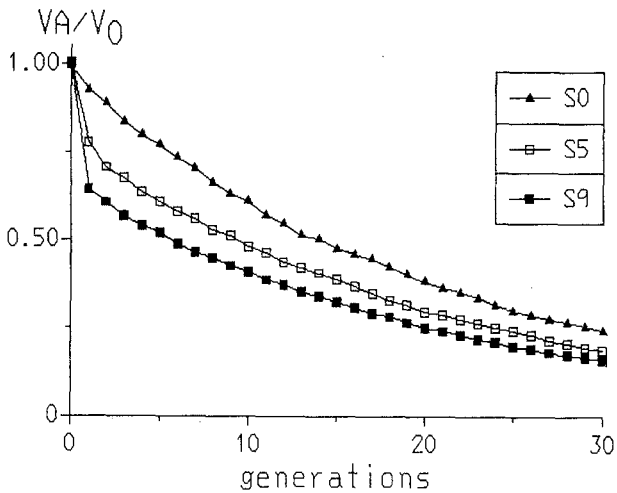
Fig. 1. Proportion of genetic variance retained (VA/V$_0$) in the three simulated populations. Mean of 500 replicates
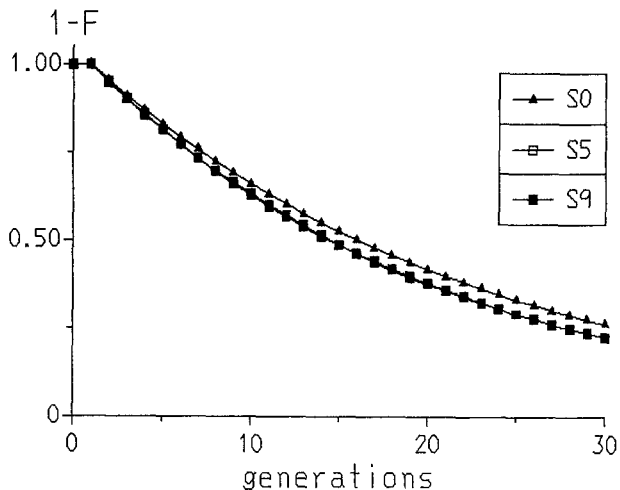

Fig. 2. Evolution of 1-F, where F is the average coefficient of inbreeding, in the three simulated populations (1-F curves for S$_5$ and S$_9$ overlap). Mean of 500 replicates

Fig. 3. Comparison of different algorithms for predicting the change in average coefficient of inbreeding (F) in the two selected populations. *Solid lines* represent differences between expected and observed values of F, as a percentage of observed values. *Dotted lines* represent critical values for a statistical significance at the 5% level, computed using the observed standard deviations between replicates

Next, VA decreased more regularly in the selected populations, and was always smaller than in the unselected one. However, the absolute difference in VA in the three populations decreased over time. By generation 30, the remaining variance was only 24%, 19%, and 16% of V$_0$ in S$_0$, S$_5$, and S$_9$, respectively.

In the three simulated populations, no significant differences were observed between the average inbreeding coefficients of candidates and selected animals. In fact, it could be shown, from the values observed for the standard deviation between replicates, that a number of about 6,000 or 14,000 replicates would have been required to assess the significance of the small differences observed in the males and females, respectively. Figure 2 shows the results obtained for the average inbreeding coefficient of the candidates. Inbreeding was higher in
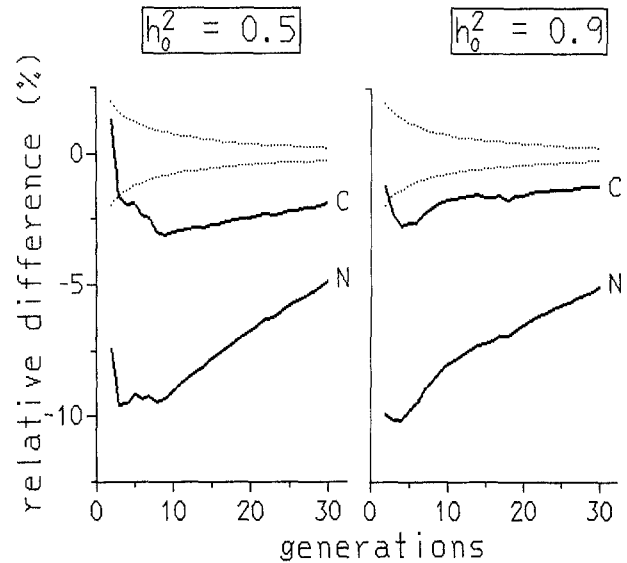
selected populations than under drift, with maximum difference in the second generation, i.e., the first generation when inbreeding was observed: +7% in S$_5$ and +10% in S$_9$. Furthermore, after this second generation, differences were small and not significant between S$_5$ and S$_9$ (the two curves overlap). The change in F at a given generation differed between selected and unselected populations, and this difference decreased over time. A comparison between Figs. 1 and 2 shows that VA was always smaller than V$_0$(1 − $F^{[t]}$), especially in the selected populations, and even in a pure drift situation (selfing was excluded, since the population was dioecious).

Under a pure drift siutation (population S$_0$), the (C) and (N) algorithms provided exactly the same values for the average inbreeding coefficient in every generation. Consequently, the values of VA provided by (C) and (N) were exactly the same. Moreover, with drift, the predictions of F and VA made from (C) and (N) were in good agreement with the observed values in every generation.

Figure 3 shows a comparison between values of F observed in selected populations and predictions based on (C) and (N). The average inbreeding coefficient was always underestimated, except from algorithm (C) in generation 2 (the first generation when inbreeding appeared). However, the relative difference between predicted and observed values was about three and four times larger when F was computed as under pure drift (N) than when F was computed using the proposed method. The relative differences increased in the first generations and de-
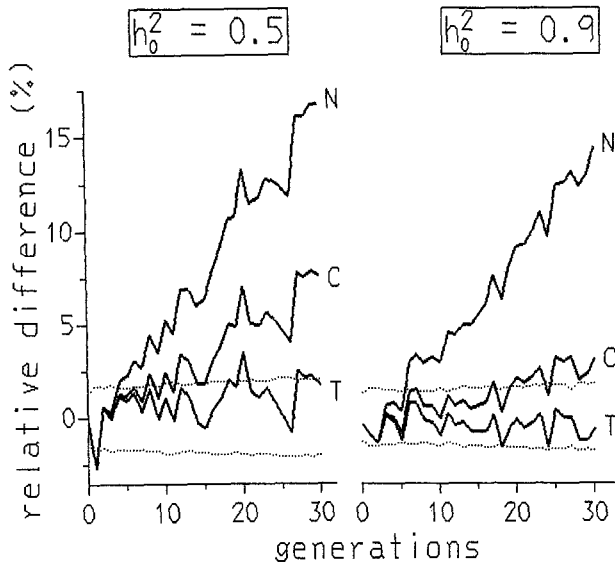
Fig. 4. Comparison of different algorithms for predicting the change in genetic variance in the two selected populations, represented as in Fig. 2
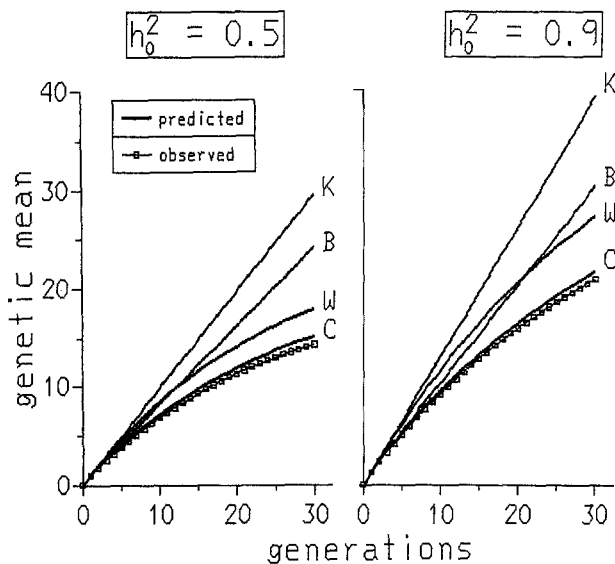

Fig. 5. Change in genetic mean, expressed in initial genetic standard deviation unit, in the two selected populations (mean of 500 replicates). Comparison of predictions using different algorithms for predicting the genetic variance (VA): K – VA = constant = V_0; B – Bulmer (1971); W – Wright (1931); C – proposed algorithm

creased from the tenth generation on: the effect of selection on inbreeding decreased over time, and the intrinsic effect of drift prevailed in the mid term.

A comparison between values of VA observed in selected populations (S_5 and S_9) and predicted values from algorithms (C), (N), and (T) is shown in Fig. 4. For algorithm (C), significant differences were observed only from the 13th and the 20th generation in S_5 and S_9,

respectively. The relative difference between observed and predicted values was larger when the initial heritability was smaller, and it increased over time: using (C), VA was overestimated in the 30th generation, with an amount of +7% and +3% in S_5 and S_9, respectively. Computing F under pure drift [algorithm (N)] resulted in relative differences that were about 2 and 3.5 times larger than the ones observed when using (C), in populations S_5 and S_9, respectively, and that were significant from the fifth generation on, in both selected populations. As expected, predicting genetic variance by using in Eq. (5) the observed value of F (T) led to results very close to the observed ones.

Finally, Fig. 5 shows that the proposed algorithm provided predicted responses to selection close to the observed ones: in the 30th generation, the genetic mean was overestimated by an amount of +5.4% and +3.4% only in S_5 and S_9, respectively. On the other hand, assuming a constant genetic variance or using equations taking into account the effect of selection or drift only induced very high biases.

Discussion

The joint effects of selection and drift

The genetic variance of a selected population is clearly smaller than that of an unselected one, with the same numbers of parents, where only drift occurs. Our results are in agreement with other simulation studies (e.g., Sirkkomaa and Lindstrom 1981; Mueller and James 1983; Murrmann-Kahl and Dempfle 1984). Particularly, these results show that under selection, the higher the accuracy of selection, the quicker the decrease of VA is. It should be noted that the effect of selection on VA is maximum in the first generations. Beyond this, the effect of drift per se tends to prevail. This phenomenon results directly from the decrease in accuracy of selection when VA decreases, thus explaining the lower impact of selection.

Results for average coefficient of inbreeding (F) showed that selection changes the family structure. These results are in agreement with the observation of Lush (1946) and the experimental studies by Barria and Bradford (1981) and Kownacki et al. (1981, 1987). However, the observed values of F did not differ with different values of the initial heritability (0.5 and 0.9). These results are in agreement with the simulation results of Wray and Thompson (1990): for mass-selected populations of larger size, the rate of inbreeding increased when the initial heritability increased from 10^{-6} to 0.4, but the same rate or a smaller rate was observed between the values of 0.4 and 0.6. The theories of Robertson (1961) and Burrows (1984a, b), and the method proposed in this paper postulate that the effect of selection on the family structure

should increase with the correlation (ϱ) between the values for the selection criterion of related individuals. Selection decreases the heritability of the trait but, in our simulations, large differences between both populations still held: in the 30th generation, the corresponding values were 0.16 and 0.59 for initial values of 0.5 and 0.9, respectively. In such a circumstance, differences in inbreeding would have been expected, and the observed results are puzzling. A clear understanding of this phenomenon would probably be useful for constructing more accurate algorithms.

Several other factors could affect the genetic variance notably. For example, assortative mating could compensate the reduction in variance due to selection (see, e.g., Fernando 1984; Langlois 1990), and avoidance of close mating could decrease the rate of inbreeding in the short term. In the case of overlapping generations, the evolution of genetic variance is more complex, because it depends on within-cohort variance and on between-cohorts variance. Some simulation results (Verrier 1989) showed the erratic decrease of genetic variance in this situation, and also showed a much more dramatic effect of selection on the rate of inbreeding than in the present simulation study.

The validity of the proposed algorithm

The proposed method provided predicted values of the variance in agreement with observed values in the first 12 or 19 generations of selection. Using the observed values of F for the prediction of VA led to predicted values in agreement with observed values. Therefore, prediction errors in VA, which appeared in the long term from the proposed algorithm, are essentially due to prediction errors in the rate of inbreeding, which was underestimated as early as the first generations.

The errors of prediction for F are substantially smaller than those resulting from ignoring the effect of selection. They could be explained by some deliberate simplifications of the algorithm, such as ignoring grandparental subclasses, assuming that the population is large enough to ignore any order statistics, assuming that perfect normality of selection criteria holds during the selection and, finally, ignoring relationships between candidates when calculating the genetic variance among the selected parents. With such a large number of approximations, larger discrepancies could well have been obtained.

Finally, the most relevant criterion for testing the proposed method might be prediction of cumulated genetic gain. The proposed algorithm provided a good approximation for genetic change in the long run for the simulated populations, although genetic gains were slightly overestimated. Closer predictions would have been obtained if the selection differentials had been better

predicted. This problem is very difficult to solve since order statistics are involved, candidates are not independent from one another, and strict normality does not hold (Dempfle 1987).

Possible extensions to other situations

Extenting the algorithm to situations where the numbers of animals are not constant over time and to different selection criteria is straightforward. The second case would correspond to manipulations of the R^2 and ϱ values. Some examples are presented by Verrier (1989). His predicted results for different criteria, given different weights to individual and family information, are in agreement with previous simulation results (e.g., Dempfle 1975; Hill 1985; Toro et al. 1988; Wray and Thompson 1990). However, accuracy of the algorithm in a situation of combined selection has not yet been tested.

The proposed method could be extended to a multiple-trait situation to investigate the change in genetic correlations with selection. This requires considering matrices of genetic variance-covariance between the traits. The parental matrix could be obtained using Eq. (3) from Pearson (1903), and the within-family matrix from Eq. (4), and from an extension of this equation to the within-family covariance between traits, proposed by Foulley and Chevalet (1981). The computation of the average coefficient of inbreeding could be made in adapting the method proposed in this paper to the selection index used.

Hierarchical matings prevent us from considering the different pairs of mates as independent events. For instance, if females are mated to one male only, there are only three categories of mates: categories 1, 2, and 4, with $\pi_4^* = \pi_3 + \pi_4$ [Eqs. (6a)–(6d)]. With only this modification, the algorithm would be usable. Avoidance of close mating could be traited in a similar way, e.g., by suppressing category 1 (fullsibs mating). Very strong assortative (or disassortative) hierarchical matings could be approximated by dividing the selected parts of the binormal distribution, each Q_k and π_k corresponding to a weighted mean of different probabilities. Other cases, weak assortative hierarchical matings or assortative nonhierarchical matings, would be virtually impossible to treat with this algorithm.

An examination of practical schemes requires an adaptation of the algorithm to a situation where generations overlap. This case is rather difficult, since the family structure is very much more complex and the theoretical and numerical problems outlined for the separate generations case are amplified. The extensions of the algorithm we tried to develop turned out to be unsatisfactory; perhaps due to the large number of approximations made, F values were notably underestimated when compared to Monte Carlo simulation results (Verrier 1989).

Conclusions

The simulations presented confirm the complexity of the joint effects of selection and drift on change in inbreeding and genetic variance in populations of limited size. They refute the validity of the concept of effective size in such a situation. A probabilistic approach to predict the change in the average coefficient of inbreeding, taking into account all the parameters of selection, should be used. Comparison between observed and predicted results clearly showed that the most critical point for research lies here. Within our hypothesis (additive infinitesimal model, random mating, separate generations), the algorithm we tested proved to be rather satisfactory. It could be used under simplified conditions to compare different strategies of selection with respect to accumulated genetic gain in the mid- or long-term. However, adaptation of the proposed method to overlapping generations is difficult, and further research is clearly needed to provide an efficient algorithm for prediction well suited to practical selection schemes.

Appendix

Computation of the correlation between values for the selection criterion

In the case of mass selection, the correlation (ϱ) between values for the selection criterion of two individuals, i and i', is the correlation between their own performance, Y_i and $Y_{i'}$:

$$\varrho = \frac{\text{Cov}(Y_i, Y_{i'})}{\sigma_{Y_i} \sigma_{Y_{i'}}}. \tag{A1}$$

The genetic model is strictly additive, and we consider no common environmental effect. Therefore, in a population of infinite size, the covariance between Y_i and $Y_{i'}$ is a simple function of the covariances beween the breeding values of the sires (A_s and $A_{s'}$ respectively) and of the dams (A_d and $A_{d'}$ respectively). Due to random mating, the covariances between the breeding values of parents of different sexes (A_s, $A_{d'}$) and ($A_{s'}$, A_d) are null; therefore,

$$\text{Cov}(Y_i, Y_{i'}) = \tfrac{1}{4}\text{Cov}(A_s, A_{s'}) + \tfrac{1}{4}\text{Cov}(A_d, A_{d'}). \tag{A2}$$

The limited size of the parental groups induces negative terms in the covariances in any situation. Therefore, the computation of ϱ has to take into account the random fluctuation of the means of the variables involved.

Consider, for any random variable X, observed on N individuals, the mean $\bar{X}$, and for any individual the value X_i^* defined as:

$$X_i^* = X_i - \bar{X}. \tag{A3}$$

We have to compute in fact:

$$\frac{\tfrac{1}{4}\text{Cov}(A_s^*, A_{s'}^*) + \tfrac{1}{4}\text{Cov}(A_d^*, A_{d'}^*)}{\sigma_{Y_i^*}\sigma_{Y_{i'}^*}}. \tag{A4}$$

By construction, X_i^* and $\bar{X}$ are independent. If we note $\sigma_x^2 = \text{Var}\, X_i$, then,

$$\text{Var}\, X_i^* = \text{Cov}(X_i^*, X_i^*) = \sigma_x^2\left(1 - \frac{1}{N}\right), \tag{A5}$$

and

$$\text{Cov}(X_i^*, X_{i'}^*) = -\text{Var}\, \bar{X} = \frac{-\sigma_x^2}{N}. \tag{A6}$$

Applying Eqs. (A5) and (A6) to Eq. (A4) leads to the expression of the correlations between mates of the four categories defined in the text [Eqs. (6a)–(6d)]:

$$\varrho_1^{[t-1]} = \frac{1}{4D}\left[\left(1 - \frac{1}{N_m}\right)\text{VA}_s^{[t-2]} + \left(1 - \frac{1}{N_f}\right)\text{VA}_d^{[t-2]}\right] \tag{A7a}$$

$$\varrho_2^{[t-1]} = \frac{1}{4D}\left[\left(1 - \frac{1}{N_m}\right)\text{VA}_s^{[t-2]} - \frac{1}{N_f}\text{VA}_d^{[t-2]}\right] \tag{A7b}$$

$$\varrho_3^{[t-1]} = \frac{1}{4D}\left[-\frac{1}{N_m}\text{VA}_s^{[t-2]} + \left(1 - \frac{1}{N_f}\right)\text{VA}_d^{[t-2]}\right] \tag{A7c}$$

$$\varrho_4^{[t-1]} = \frac{1}{4D}\left[-\frac{1}{N_m}\text{VA}_s^{[t-2]} - \frac{1}{N_f}\text{VA}_d^{[t-2]}\right] \tag{A7d}$$

with,

$$D = \left[\left(1 - \frac{1}{T_m}\right)\left(1 - \frac{1}{T_f}\right)\right]^{1/2}[\text{VA}^{[t-1]} + VE]. \tag{A7e}$$

In these expressions, N represents the numbers of parents and T the numbers of candidates, with the indices m and f for the males and the females, respectively.

The extension of Eq. (A7) to any criterion selection is easy because, on the one hand, the variance of the criterion is always known and because, on the other hand, it is possible to express the covariance between criteria as a function of the covariance between breeding values (Colleau and Poutous 1973). However, Eqs. (A7a)–(A7e) consider only the covariance between parental breeding values, and do not consider the individual deviations from the family mean. Following the previous arguments, it can be shown that, for two fullsibs, there is a negative covariance between the deviations of their own performance (Y_{ij} and $Y_{ij'}$) from their family mean ($Y_{i.}$):

$$\text{Cov}(Y_{ij} - Y_{i.}, Y_{ij'} - Y_{i.}) = \frac{-\text{Var}\, Y}{n}$$

where n is the family size. Therefore, the complete expression of ϱ is more complex (see Dempfle 1987, for the correlation between fullsibs). However, some numerical results (Verrier 1989) show that the negative covariance between individual deviations has a significant effect on the values of ϱ only when the deviation has a great weight in the criterion. In practice, it is necessary to take this covariance into account only when selection is based on the deviation from the family mean, with no attention to the number of selected parents in each family ("deviation selection" of Hill 1985, or "unrestricted within-family selection" of Dempfle 1987). On the other hand, Eqs. (A7a)–(A7e) or the extension of these equations are a good approximation in the case of mass selection or combined selection.

References

Barria N, Bradford GE (1981) Long-term selection for rapid gain in mice. I: Genetic analysis at the limit response. J Anim Sci 52:729–738

Bulmer MG (1971) The effect of selection on genetic variability. Am Nat 105:201–211

Burrows PM (1984a) Inbreeding under selection from unrelated families. Biometrics 40:357–366

Burrows PM (1984 b) Inbreeding under selection from related families. Biometrics 40:895–906

Chevalet C (1988) Control of genetic drift in selected populations. Proc 2nd Int Conf Quant Genet, May 31–June 5, 1987, Raleigh/NC 379–394

Colleau JJ, Mocquot JC (1989) Using embryo transfer in cattle breeding. 5th Annu Meet Embryo Transfer Assn, Sept 8–9, Lyon, France

Colleau JJ, Poutous M (1973) Méthode de calcul des index de production latière des femelles en France. Ann Genet Sel Anim 5:73–82

Crow JF, Kimura M (1970) An introduction to population genetics theory. Harper and Row, New York

Dempfle L (1975) A note on increasing the limit of selection through selection within families. Genet Res 24:127–135

Dempfle L (1987) Statistical aspects of design of animal breeding programmes: a comparison among various selection strategies. In: Gianola D, Hammond K (eds) Advances in statistical methods for genetic improvement of livestock. Int Symp, Armidale, Australia, February 16–20

Ducrocq V, Colleau JJ (1986) Interest in quantitative genetics of Dutt's and Deak's methods for numerical computation of multivariate normal probability integrals. Genet Sel Evol 18:447–474

Fernando RL (1984) Selection and assortative mating. PhD thesis, University of Illinois, Champaign-Urbana

Foulley JL, Chevalet C (1981) Méthode de prise en compte de la consanguinité dans un modèle simple de simulation des performances. Ann Genet Sel Anim 13:189–196

Hill WG (1985) Fixation probabilities of mutant genes with artificial selection in finite populations. Genet Sel Evol 17:351–358

Johnson NL, Kotz S (1972) Distributions in statistics: continuous multivariate distributions. Wiley, New York, p 94

Keightley D, Hill WG (1987) Directional selection and variation in finite populations. Genetics 117:573–582

Kownacki M, Jaszczak K, Jezierski T (1981) The development of inbreeding in selected and unselected populations of different sizes. Rocz Nauk Roln Ser B 100:45–49 (in Polish)

Kownacki M, Laudanski Z, Wojcik R (1987) Long-term response to mass selection in relation to population size in laboratory mice. Genet Pol 28:131–144

Langlois B (1990) Réflexions sur l'incidence de la sélection et des croisements raisonnés sur les paramètres du modèle génétique additif. Genet Sel Evol (in press)

Lush JL (1945) Animal breeding plans, 3rd edn. Iowa State University Press, Ames/IA

Lush JL (1946) Chance as a cause of gene frequency within pure breeds of livestock. Am Nat 80:318–342

Malecot G (1948) Les mathématiques de l'hérédité. Masson, Paris

Mueller JP, James JW (1983) Effect on linkage disequilibrium of selection for a quantitative character with epistasis. Theor Appl Genet 65:25–30

Murrmann-Kahl S, Dempfle L (1984) Effektive Populationsgröße und Änderung der additiv genetischen Varianz in kleinen Populationen bei Selektion. 35th Annu Meet E.A.A.P., The Hague, August 6–9

Nicholas FW, Smith C (1983) Increased rates of genetic change in dairy cattle by embryo transfer and splitting. Anim Prod 36:341–353

Pearson K (1903) Mathematical contributions to the theory of evolution. XI: On the influence of natural selection on the variability and correlations of organs. Philos Trans R Soc London Ser A 200:1–66

Robertson A (1961) Inbreeding in selection programmes. Genet Res 2:189–194

Ruane J, Thompson R (1989) Simulation of an adult MOET nucleus breeding scheme in dairy cattle. In: Kalm E, Liboriussen T (eds), New selection schemes in cattle: nucleus progammes. Pudoc, Wageningen, pp 72–80

Sirkkomaa S, Lindstrom UB (1981) Simulation response to selection for body weight in rainbow trout. Acta Agric Scand 31:426–432

Toro MA, Silio L, Rodriganez J, Dobao TM (1988) Inbreeding and family index selection for prolificacy in pigs. Anim Prod 46:79–85

Verrier E (1989) Prédiction de l'évolution de la variance génétique dans les populations animales d'effectif limité soumises à sélection. Thèse de l'Institut National Agronomique Paris-Grignon

Verrier E, Colleau JJ, Foulley JL (1989) Effect of mass selection on the within-family genetic variance in finite populations. Theor Appl Genet 77:142–148

Wray NR, Thompson R (1990) Prediction of inbreeding in selected populations. Genet Res (in press)

Wright S (1931) Evolution in mendelian populations. Genetics 16:97–159