Models to estimate genetic parameters in crossbred dairy cattle populations under selection

Promotoren: dr. ir. R. D. Politiek
emeritus hoogleraar in de veeteeltwetenschappen

dr. ir. E. W. Brascamp
hoogleraar in de veefokkerij

Models to estimate genetic parameters in crossbred dairy cattle populations under selection

Julius H. J. van der Werf

Proefschrift
ter verkrijging van de graad van
doctor in de landbouw- en milieuwetenschappen,
op gezag van de rector magnificus,
dr. H. C. van der Plas,
in het openbaar te verdedigen
op vrijdag 1 juni 1990
des namiddags te vier uur in de aula
van de Landbouwuniversiteit te Wageningen.

ISN= 519342

Stellingen

1. De werkelijke genetische waarden van dieren die in een populatie worden geimporteerd, zullen bij verwaarlozing van niet-additieve effecten over het algemeen worden overschat.
   dit proefschrift

2. Bij de momenteel toegepaste reproduktietechnieken is de benutting van dominantie-effecten ten bate van geplande paringen bij rundvee van weinig foktechnische betekenis.
   dit proefschrift

3. Analyse van phenotypische gegevens uitsluitend van geselecteerde generaties levert in geen geval een schatting op van de genetische variantie in de ongeselecteerde basispopulatie.
   D.A. Sorensen and B.W. Kennedy, 1984 J. Animal Sci. 59:1213

4. De minimale eisen die de maximum likelihood methoden stellen aan de familiestructuur in een dataset, hebben de behoefte aan betrouwbare criteria betreffende de kwaliteit van schattingen van genetische parameters vergroot.
   dit proefschrift

5. Behalve selektieintensiteit wordt ook het behoud van variantie in een kleine populatie overschat met formule's voor oneindige populaties.
   dit proefschrift

6. Bij kritiek op het gebruik van simulatiestudies wordt vaak ten onrechte geen onderscheid gemaakt tussen simulatiestudies ter verificatie van methoden, en simulatiestudies voor het voorspellen van de werkelijkheid.

7. Het dilemma van de fokkerij, dat bij een toenemende nauwkeurigheid de toekomstige mogelijkheden sterker worden verkleind, wordt met name duidelijk bij gebruik van moleculair-genetische selektie criteria.

8. Veefokkerij-onderzoek zal, naast het gebruik van genetische variatie, relatief meer plaats moeten inruimen voor de problematiek van het behoud van genetische variabiliteit.

9. Toepassing van biotechnologische en medische vindingen zullen moeten leiden tot een explicietere formulering van het mens-zijn, aangezien God of lot in een aantal gevallen wordt vervangen door eigen verantwoordelijkheid.

10. Het terugbrengen van de formele universitaire studieduur tot 4 jaar is in het kader van een toenemende internationale concurrentie een blunder te noemen.

11. Vervangen van het begrip 'instromer' door de term 'doorstromer' getuigt van een afnemend vertrouwen in eigen academisch kunnen.

12. Cultuur heeft vele gelijkenissen met genetica; het repliceert, het evolueert en éénvormigheid is een van haar grootste bedreigingen.

J.H.J. van der Werf
Models to estimate genetic parameters in crossbred dairy cattle populations under selection.
1 juni 1990.

"Although this may seem a paradox, all exact science is dominated by the idea of approximation"

(Bertrand Russel (1872-1970)

Cover design: Wim Valen

Van der Werf, J.H.J. 1990. Models to estimate genetic parameters in crossbred dairy cattle populations under selection (Modellen voor het schatten van genetische parameters in gekruiste en geselekteerde melkvee populaties). Doctoral Thesis, Department of Animal Breeding, Wageningen Agricultural University. P.O. Box 338, 6700 AH Wageningen, The Netherlands.

VOORWOORD

Dit proefschrift kwam tot stand naar aanleiding van mijn onderzoek bij de vakgroep Veefokkerij. Mijn huidige en ex-collega's wil ik bedanken voor de door hen gecreëerde werksfeer, welke ik als plezierig en stimulerend heb ervaren. Een aantal mensen die mij bij de vervaardiging van dit boekje in het bijzonder hebben bijgestaan wil ik graag met name noemen.

Prof.dr.ir. R.D. Politiek bedank ik voor zijn voortvarende en stimulerende begeleiding, en voor de vrijheid die hij me bood bij de invulling van het onderzoek. Prof. dr. ir. Pim Brascamp bedank ik voor de geboden ruimte om het geheel af te ronden, en voor zijn kritische bijdrage. Siem Korver en Johan van Arendonk hebben een grote invloed gehad op het uitzetten van de hoofdlijnen van het onderzoek. Hen wil ik bedanken voor de aansporingen, de vele suggesties en opmerkingen die mij op het juiste spoor hebben gehouden. De besprekingen in de begeleidingscommissie waren altijd stimulerend en leerzaam. Behalve bovengenoemde personen gaat hiervoor mijn dank ook uit naar Rob Verdooren en Hans Wilmink.

De vele melkproduktiegegevens, die welwillend door het NRS ter beschikking waren gesteld, zijn door Willem de Boer met veel oplettendheid en onnavolgbare ijver computermatig verwerkt. Arnoud van der Lugt heeft deze taak bekwaam overgenomen bij de analyses van het vijfde hoofdstuk. Imke de Boer heeft een grote bijdrage geleverd aan de totstandkoming van het vierde hoofdstuk.

I thank Brian Kennedy for his interest in my work and for giving the opportunity to visit Guelph. Karin Meyer provided her very useful REML-programs. Also, Robin Thompson, Mike Grossman, John Pollak, Dick Quaas, and Naomi Wray deserve credit for their suggestions and comments.

Cathy, ik wil je bedanken voor al je steun en geduld, en voor je hulp bij de Engelse formuleringen. Tenslotte wil ik mijn ouders bedanken voor hun steun en belangstelling tijdens mijn studie. Aan hen draag ik dit werk op.

Het LEB-fonds bedank ik voor de financiële ondersteuning bij de afronding van het proefschrift.

CONTENTS

CHAPTER 1

INTRODUCTION

Reliable estimates of genetic variance and heritabilities serve several purposes in animal breeding. First, heritabilities and estimates of non-additive variation provide information about the mechanism of inheritance of phenotypically observed characteristics in animals. Secondly, genetic parameters can be directly applied to estimate breeding values and to design and optimize breeding programs.

Phenotypic expression is affected by changes of environment and, when there is artificial selection, by a change of the underlying genetic variation. Genetic parameters might therefore be subject to change. Accurate and updated genetic parameters are needed to optimize breeding programs. Moreover, accurate and regular parameter estimation allows for the detection of shifts from the assumed pattern of inheritance, as selection programmes become more effective and complex.

Methods to determine genetic variance have been greatly improved over the last two decades. Recently, maximum likelihood based methods have been introduced (Patterson and Thompson, 1971) making use of mixed models (Henderson, 1984). This has proved to be a general framework for the analysis of data and the estimation of genetic parameters. It potentially accounts for all systematic effects, unbalanced data and various forms of selection and non-random sampling, the last being particularly important for animal breeding applications. Analysis of dairy cattle data is often based on systematically recorded field data because, selection experiments are expensive and can only be of limited size. Besides a good method and reliable data, the model actually used to estimate parameters is crucial. It is related to the estimation method and to the properties and the structure of the data. Restrictions on the model have diminished, through improvement of both methods and computing capacity.

High estimates of heritability for milk production traits in Dutch Black and White cattle in recent years (Wilmink and De Graaf, 1986) were direct motivation of this research. The main characteristics of this cattle

population are a large import of American Holstein bull semen and an intensified selection of parents to have progeny. Models currently used for genetic evaluation only assume additivity of gene effects for milk production traits and random sampling of the genetic effects.

Non-additive effects for milk production traits in cattle have been reported to be low and were therefore not utilized in the breeding scheme. However, little is known about the extent of bias in estimation of additive genetic effects, when non-additive effects are ignored. Bias is expected to depend on the magnitude of non-additive effects and the design of mating parents from different breed groups. Therefore, the influence of non-additive effects on estimation of additive parameters will be determined in a simulation study for different mating designs (Chapter 2). Models that account for non-additive and additive genetic effects, will be compared in Chapter 3 and these effects will be estimated, using field data.

Estimates of genetic parameters used in selection programs should preferably not be biased by this selection (Robertson, 1977). Accounting for selection bias is a general problem, that is not only relevant for cattle populations, but also for other species. Several studies have indicated that mixed models account for selection, when all relationships between animals and records on which selection was based are used (Sorensen and Kennedy, 1984; Gianola and Fernando, 1986). However, data based upon field records are usually sampled during a limited time period.

In this thesis, an attempt will be made to determine to what extent estimates of genetic variance are biased, if pedigree information and ancestral records are not available. The problem will be studied using simulated data from a "species neutral" population (Chapter 4). A method that was proposed to account for selection on pedigree (Graser et al., 1987) will be investigated in Chapter 5. The actual effect of ignoring selection will be quantified and discussed for the dairy cattle data.

References

Gianola, D. and R.L. Fernando. 1986. Bayesian methods in animal breeding theory. J. Anim. Sci. 63:217.

Graser, H.U., S.P. Smith and B. Tier. 1987. A derivative free approach for

estimating variance components in animal models by restricted maximum likelihood. J. Anim. Sci. 64:1362.

Henderson, C.R. 1984. Linear models in animal breeding. University of Guelph Press. Guelph, Canada.

Patterson, H.D. and R. Thompson. 1971. Recovery of inter-block information when block sizes are unequal. Biometrika 58:545.

Robertson, A. 1977. The effect of selection on the estimation of genetic parameters. Z. Tierzuchtg. Zuchtungsbiol. 94:131.

Sorensen D.A., and B.W. Kennedy. 1984. Estimation of genetic variances from unselected and selected populations. J. Anim. Sci. 59:1213.

Wilmink, J.B.M. and F.M. de Graaf. 1986. Genetic parameters for 305 days yield in first lactation and adjustment of Net Guilder Index. Rep. 86-676, Dutch Royal Cattle Syndicate, Arnhem. (Dutch with English summary).

Chapter 2

INFLUENCE OF NON-ADDITIVE EFFECTS ON ESTIMATION OF GENETIC PARAMETERS IN DAIRY CATTLE

J.H.J. VAN DER WERF AND W. DE BOER
Department of Animal Breeding
Wageningen Agricultural University
PO Box 338, 6700 AH Wageningen
The Netherlands

Published in: Journal of Dairy Science 72, 2606-2614 (1989)

ABSTRACT

A population was simulated having progeny that descended from sires and dams with various fractions of genes from two breeds. Additive breed effects and non-additive effects from breed crosses were simulated.

Data on performance were analysed using mixed models that accounted for fixed additive genetic effects and random sire effects. Three additive models, with genetic groups defined according to 1) breed composition of the progeny, 2) breed composition of the sire and the dam, or 3) linear regression on breed fraction of the progeny, were compared with a non-additive model with a linear regression on breed fraction, heterozygosity and recombination in the genome of the progeny. Variance components were estimated using restricted maximum likelihood.

Additive genetic variance and heritability were overestimated for an additive model with progeny groups. Additive models gave estimates for breed difference, group effects and breeding values that were not equal to true additive genetic values. Breed differences were overestimated when when sire groups were used. Estimators for each parameter were unbiased with the non-additive model.

INTRODUCTION

Due to increased exchange of semen and embryos, dairy cattle data often result from a mixture of genes from different populations. Therefore, the observed genetic effects in the resulting population might not be solely additive and may contain non-additive genetic effects. In some crossbreeding programs, the objective is explicitly to estimate or to use the non-additive genetic variation. However, non-additive variance is not often used explicitly in dairy cattle breeding. Moreover, models for analysis of data from crossbred dairy populations often consider only additive genetic effects. Ignoring non-additive variance might bias predictions of breeding values, which also may not be of minimum prediction error variance. Estimators of variance components and additive genetic parameters might be biased as well.

Genetic variation in crossbred populations usually is estimated from

within breed variance, i.e., predictions of random effects are adjusted for fixed breed-group effects. Recently, estimates for heritability of milk production traits in crossbred populations, using mixed models with breed-group effects (12,20,22), were higher than published values from pure breeds (7,10). Among other factors, non-additive effects might have caused an increase in heritability estimates.

Experimental evidence for non-additive effects in dairy production traits has been reviewed by several authors, (e.g. 14, 19). Estimates of heterosis, defined as the relative deviation of the F1 mean from the midparent mean, for first lactation 305-days milk yield varied from -1.7% to 8.2%. Recent estimates for percentages of heterosis were 10% for Holstein Friesian x Red Danish (4) and 2% for Swedish Friesian x Swedish Red and White (2).

There are some examples of a linear relationship between degree of heterozygosity and heterosis (15). In most data from farm animals, however, heterosis in F1 crosses is more than twice that in F2 crosses (19). Such a deviation from linearity between degree of heterozygosity and observed heterosis can be due to epistatic effects, which are often estimated as recombination loss (8). There is a considerable variation in estimates of recombination loss for milk production, ranging from -12% (4) to -0.4% (2). Crossbreeding parameters might be hard to estimate, particularly from field data, and well designed experiments are required, at minimum (3,18).

In this paper, the impact of heterosis and recombination loss on the estimation of additive genetic parameters was studied for data from crossbred populations. Computer simulation was used for a comparison of estimates from different models with known values. A half-sib structure, typical for estimation of variances in dairy populations, was considered.

MATERIAL AND METHODS

Simulation of Data

A dairy population from a combination of two breeds, Dutch Friesian (FH) and Holstein Friesian (HF), was simulated. Cows originated from matings of different genetic groups on the basis of the fraction of HF-genes.

Records were simulated according to the following model;

$$y_{ijkmn} = \mu + h_i + ys_j + g_{s_k} + g_{d_m} + NA_{km} + a_{kmn} + e_{ijkmn}$$

where μ is the population mean of reference breed FH; h_i and ys_j are fixed environmental effects of herd and year season with i=1,..,150 and j=1,..,4; g_{s_k} and g_{d_m} are fixed additive effects of breed group of sire and of breed group of dam with k=1,..,9 and m=1,..,9; NA_{km} is the fixed non-additive effect of interaction of sire group k by dam group m; a_{kmn} is the additive genetic effect of the cow making the record; and e_{ijkmn} is a residual effect. Sires and dams were genetically unrelated; only relations between sires and daughters were considered.

Heritability (h^2) was .30 and phenotypic variance (σ_p^2) was 518,400. μ was fixed at 6000 kg. A pseudorandom value for each of herd was sampled from $N[0, 0.1\sigma_p^2]$ and for each year-season effect from $N[0, 0.025\sigma_p^2]$, corresponding to assumed fractions of the observed variance in milk production that is due to herds and year-seasons. The additive effect of the cow (a_{kmn}) was $0.5 * a_{sire} + \sqrt{(0.75)} * a_n$, where a_{sire} and a_n were sampled from $N[0, \sigma_a^2]$ and e_{ijkmn} was sampled from $N[0, \sigma_e^2]$. $\sigma_a^2 = h^2\sigma_p^2$ is additive genetic variance and $\sigma_e^2 = (1-h^2)\sigma_p^2$ is residual variance.

Breed difference (HF - FH) was 800 kg. Additive genetic group effects were linearly related to fraction of HF genes of animals from that group. Let two parents be mated with fraction of HF-genes being p_s for sire and p_d for dam where $p_s = (k-1)/8$ for sires in group k and $p_d = (m-1)/8$ for dams in group m. Group effects were $0.5*p_s*800$ for sire group and $0.5*p_d*800$ for dam group. The fraction HF-genes (p_p) in the progeny was equal to $[(p_s+p_d)/2]$.

Non-additive effects (NA_{km}) originate either through dominance effects, from interactions between HF and FH genes within loci, or epistatic effects from interactions between loci. It is assumed that a very large number of loci contributes to the genetic variance with no linkage in segregation. Coefficients for the heterosis effect (HET) were derived from the degree of heterozygosity of the animals (3). Thus, heterosis represents dominance effects as well as 50% of additive by additive effect that is confounded with dominance. The coefficient for heterosis (b_{HET}) was $[p_s(1-p_d) + p_d(1-p_s)]$.

Dickerson (1) defined recombination (REC) loss as a deviation from linear relation of performance with heterosis, such that "the coefficient

of REC describes the average fraction of independently segregating pairs of loci in gametes from both parents which are expected to be non-parental combinations." The coefficient for a recombination effect (b_{REC}) was derived from the heterozygosity of the parental gametes, representing a within-gamete epistatic effect as $[p_s(1-p_s)+ p_d(1-p_d)]$. Hill (6) argued that Dickerson's definition can be ambiguous and he prefers an explicit parametrization with dominance effects and additive by additive effects (epistatic effects). Assuming that recombination refers to 50 % of the epistatic effects (with no linkage), however, the parametrization that was used by Dickerson (1) defines appropriately the sum of dominance and epistatic effects. Effects of heterosis and recombination were simulated with the level varying at 0, 5 and 10% of the phenotypic mean of the two breeds.

Three data structures were simulated with particular sire group by dam group combinations. For structure I, additive and non-additive effects for each combination of sire and dam group are in Table 1. Matings were such that non-additive effects were unequal within groups of progeny that had equal additive effect. Hence, additive effects and non-additive effects were not confounded. In structure I, data were distributed equally over sire group by dam group combinations. A second structure was created to represent an actual mating situation in a gene importing country. Structure II (Table 2) was based on Dutch data from 399,383 crossbred cows (Dutch Friesian x Holstein Friesian) calving between 1983 and 1986 (21). Structure III (Table 3) represented a future generation of cows, which sires were distributed over groups according to the inseminations in 1987 in The

TABLE 1.Additive (Add) and Non-additive[1] (Nadd) effects (kg) for different combinations of sire- and dam groups for structure I .

group of dam (%HF genes)	effect	group of sire (%HF genes)		
		1 (0%)	5 (50%)	9 (100%)
1 (0%)	Add	0	200	400
	NAdd	0	80	320
5 (50%)	Add	200	400	600
	Nadd	80	0	80

[1] heterosis= 5%, recombination loss= -5%

TABLE 2. Distribution of data over sire- and dam groups[1] for structure II.

	group of sire					
group of dam	1	5	7	8	9	total
1	.251	.059	.034	.010	.495	.849
5	.004	.003	0	0	.014	.021
7	.015	.008	.006	0	.090	.119
8	0	0	0	0	.007	.007
9	0	0	0	0	.004	.004
total	.270	.070	.040	.010	.610	1.000

[1] group number i corresponds to (i-1)*12.5 %HF genes

TABLE 3. Distribution of data over sire- and dam groups[1] for structure III.

	group of sire					
group of dam	1	5	7	8	9	total
1	.034	.030	.016	.009	.158	.247
2	0	0	0	0	.005	.005
3	.004	.008	.004	0	.055	.071
4	0	.005	.003	0	.033	.041
5	.012	.057	.031	.018	.385	.503
6	0	0	0	0	.016	.016
7	0	.010	.006	.003	.074	.093
8	0	0	0	0	.016	.016
9	0	0	0	0	.008	.008
total	.050	.110	.060	.030	.750	1.000

[1] group number i corresponds to (i-1)*12.5 %HF genes

Netherlands. Dams were distributed over groups according to the distribution of progeny that resulted from matings in structure II.

For each mating structure, a data set was simulated for 100 sires having progeny in a number of herds. Total number of records per herd averaged 33, with a standard deviation of 17. Progeny group size was 50 and alternative sizes were 25 and 100. For each herd, daughters were randomly assigned to dam groups and to sire groups according to the distribution of the structure. Such a design was sampled 10 times. To reduce sampling variance,

10 replications within each design resulted in 10 x 10 repetitions for each alternative. Using one design for several samples was attractive computationally .

Models for Evaluation

Additive models A1, A2 and A3 varied in strategy to account for fixed genetic effects. In A1, fixed group effects (g_p) were defined according to breed composition of the progeny making the record. In A2, two breed groups were defined according to the breed composition of the sire (g_s) and of the dam (g_d). A1 will be referred to as progeny group model and A2 as parent group model. Model A3 accounted for breed differences by a linear regression of performance on breed composition of progeny (p_p). In addition, a complete genetic model (NA) was used that included a regression on p_p and on coefficients for heterosis and recombination in the progeny. Models can be represented as;

$$\text{model A1: } y_{ijkn} = HYS_i + g_{p_j} + s_k + r_{ijkn}$$

$$\text{model A2: } y_{ijkmn} = HYS_i + g_{s_j} + g_{d_m} + s_{jk} + r_{ijkmn}$$

$$\text{model A3: } y_{ijkn} = HYS_i + b_1 * p_{p_j} + s_k + r_{ijkn}$$

$$\text{model NA: } y_{ijkn} = HYS_i + b_1 * p_{p_j} + b_2 * b_{HET} + b_3 * b_{REC} + s_k + r_{ijkn}$$

Fixed herd-year-season (HYS) effects were used to account for environmental effects. Effects of HYS were not simulated. Rather, HYS subclasses were used to create subclass sizes as is common in sire evaluation programs. Sire and residual effects were taken as random with variances $var(s) = 0.25 * I_s \sigma_a^2$ and $var(r) = I_n (\sigma_e^2 + 0.75 \sigma_a^2)$.

Variance components for s and r were estimated using REstricted Maximum Likelihood (REML) (11). Solutions were computed for fixed additive group effects from the same model, using REML estimates for the variance components. Breeding values were computed from model A2 as $2(g_{s_j} + s_{jk})$ or from model A1 as $(g_p + 2s_k)$, where g_p is the solution for progeny group, with the same fraction of HF genes as the sire.

RESULTS

Genetic Parameters

Estimates of variance components and heritability were obtained for each model. Results for structure I are in Table 4, by levels of heterosis (HET) and recombination (REC), averaged over 100 repetitions.

When non-additive effects were absent, models showed similar results with respect to estimates of variances and heritability. Hence, differences in expectations of y_{ijklmn} due to fixed additive genetic effects were accounted for by each model; the parent group model was equivalent to the simulated model and the progeny group model reflected the genetic mean of animals making the record. A linear regression on p_p (models A3 and NA) also accounted appropriately for differences due to breed effects.

When non-additive effects were different from zero, the progeny group model (model A1) gave biased estimates for additive genetic variance and heritability (Table 4). Overestimation of σ_a^2 and h^2 was relatively small for

TABLE 4. Estimates for variance components (σ_a^2, σ_e^2) and heritability (h^2) with additive and non-additive models for analysis of different levels of heterosis (HET) and recombination effects (REC) simulated for structure I.

model	HET(%)	REC(%)	σ_a^2	σ_e^2	h^2
A1	0	0	148543	479780	0.287
	10	0	196263	480569	0.370
	5	-5	208168	480573	0.390
	10	-5	270781	481009	0.493
	10	-10	395279	481117	0.681
A2	0	0	148904	479768	0.287
	5	-5	148640	484100	0.285
	10	-10	148380	496976	0.277
A3	0	0	148576	479766	0.287
	5	-5	195545	484932	0.365
	10	-10	328522	498578	0.565
NA	0	0	148791	479791	0.287
	5	-5	148791	479791	0.287
	10	-10	148791	479791	0.287
sampled values			151955	479220	0.293

levels of heterosis and recombination of 5% and smaller. Bias increased with increasing level of heterosis and recombination. Estimates of σ_a^2 did not increase. Hence, the progeny group model accounted for non-additive genetic effects by other effects in the model.

Estimates for additive variance were not biased using the parent group model (A2). In A2, however, the estimate of residual variance was higher resulting in a slight underestimation of heritability. A model with linear regression on p_p (A3) yielded an overestimation of both genetic and residual variance. Model NA accounted for all additive and non-additive effects and showed unbiased estimates for each parameter at each level of non-additive effect.

Structure I is an example of a balanced situation. To make inferences toward 'real life' situations, structures II and III were investigated, reflecting a first and second generation of cows in a country that imports semen. Table 5 gives estimates for variance components for data structures II and III, for moderate levels of heterosis and recombination effects. With structure II, each model, except A3, gave estimates for variance components that were close to simulated values. Differences in heritability estimates were small as well. Differences between models were larger for structure III; models A1 and A3 gave considerable bias in estimates of additive genetic variance and heritability.

TABLE 5. Estimates for variance components (σ_a^2 and σ_e^2) and heritability (h^2) with additive and non-additive models for analysis for structures II and III.[1]

structure	model	σ_a^2	σ_e^2	h^2
II	A1	160430	480281	.308
	A2	156139	481161	.300
	A3	245806	484272	.449
	NA	156574	479669	.301
III	A1	184720	480559	.350
	A2	154825	481225	.297
	A3	210494	482100	.393
	NA	155344	480143	.299

[1] heterosis= 5%, recombination loss= -5%

Estimation of Genetic Effects

Sire effects are usually interpreted as additive genetic effects within populations, and group effects as additive genetic differences between populations. Therefore, solutions for sire effects and for group effects were considered to be biased when they also contained non-additive effects. Solutions were determined for a moderate level of non-additive effects (5% for heterosis and -5% for recombination). Group solutions for animals with 0%HF genes were restricted to zero. Hence, additive genetic effects were expressed relative to the FH population.

Table 6 gives group solutions and average sire effects for structure I. Sire effects were averaged for groups of sires with equal breed composition. In the progeny group model (A1), sires were cross-classified with groups. Sire effects and group effects were each biased by non-additive effects. Non-additive effects were partly accounted for by solutions for groups, but residual bias caused overprediction of effects from 100%HF sires. Sire effects from other groups were underestimated. Hence, variance between sires was biased upward. The deviation of group solution from additive genetic value was largest for the 50%HF group (Table 6). The group effect for 100%HF cows was biased downward due to overestimation of the effects of their sires.

TABLE 6. Solutions[1] from different models for genetic groups of progeny(g_p), sire(g_s) or dam(g_d) and average predicted sire effects ($s_{j..}$) within sire groups for structure I.[2]

		Group Solutions				Average Sire Effects		
model	effect	i= 3	5	7	9	$s_{1..}$	$s_{5..}$	$s_{9..}$
A1	g_{p_i}	301	497	511	-	-45	-80	121
A2	$2g_{s_i}$	-	409	-	1113	0	0	0
A2	$2g_{d_i}$	-	235	-	-			
A3	$b_1 * p_{p_i}$	173	347	520	694	-65	-40	103
NA	$b_1 * p_{p_i}$	197	394	591	787	-2	2	0
additive values		200	400	600	800	2	6	0

[1] $g_{p_1} = g_{s_1} = g_{d_1} = 0$
[2] heterosis= 5%, recombination loss= -5%

In model A2, sires are nested within sire groups. Predictions of sire effects within groups were unbiased because sires were equally distributed over dam groups. Group solutions contained additive effects as well as average non-additive effects (Table 6). Non-additive effects not accounted for by groups contributed to an increase only in the residual variance. Solutions for the group of 100%HF sires considerably overestimated the additive genetic effect. The group effect of 0% HF dams was overestimated, i.e., group solution for 50%HF dams was lower than expected additive value (Table 6).

Model A3 gave biased estimates of group effects as well as biased prediction of sire effects. The estimate for linear regression of performance on p_p (b_1) was 697 for structure I, which is lower than true additive genetic value (Table 6).

The non-additive model (NA) yielded estimates for group effects and sire effects that were close to true additive genetic effects. Solutions for group effects (Table 6) were derived from the estimate for b_1.

Results for structures II (Table 7) and III (Table 8) were comparable with those from structure I. Estimators from the non-additive model were empirically unbiased. Regression on portion of HF genes was 793 for structure II and 773 for structure III. The estimate for heterosis was,

TABLE 7. Solutions[1] from different models for genetic groups of progeny(g_p), sire(g_s) or dam(g_d) and average predicted sire effects ($s_{j..}$) within sire groups for structure II.[2]

		Group Solutions				Average Sire Effects		
model	effect	i= 3	5	7	9	$s_{1..}$	$s_{5..}$	$s_{9..}$
A1	g_{p_i}	320	688	647	760	-10	-77	15
A2	$2g_{s_i}$		528	940	1380	0	0	0
A2	$2g_{d_i}$	-4	34	10	158			
A3	$b_1 * p_{p_i}$	172	343	515	686	-183	-115	94
NA	$b_1 * p_{p_i}$	198	397	595	793	0	-6	0
additive values		200	400	600	800	2	-20	-8

[1] $g_{p_1} = g_{s_1} = g_{d_1} = 0$
[2] heterosis= 5%, recombination loss= -5%

TABLE 8. Solutions[1] from different models for genetic groups of progeny(g_p), sire(g_s) or dam(g_d) and average predicted sire effects ($s_{j..}$) within sire groups for structure III.[2]

	Group Solutions					Average Sire Effects		
model	effect	i= 3	5	7	9	$s_{1..}$	$s_{5..}$	$s_{9..}$
A1	g_{p_i}	355	583	583	668	-64	-165	36
A2	$2g_{s_i}$		561	877	1220	0	0	0
A2	$2g_{d_i}$	-12	16	115	256			
A3	$b_1 * p_{p_i}$	75	151	226	301	-302	157	48
NA	$b_1 * p_{p_i}$	193	387	580	773	-7	-11	3
additive values		200	400	600	800	-10	-22	1

[1] $g_{p_1} = g_{s_1} = g_{d_1} = 0$
[2] heterosis= 5%, recombination loss= -5%

averaged over repetitions, 321 for structure II and 318 for structure III with empirical standard errors 5.8 and 5.4. Estimates for the recombination effect were -332 for structure II and -302 for structure III with empirical standard errors 15.0 and 8.7. Simulated values were 320 for heterosis and -320 for recombination effect.

The effect of size of the progeny groups on bias is given for model A1 (Table 9). A greater fraction of non-additive effects was assigned to the random sire effects with increasing progeny-group size. Bias in estimation of additive genetic variances increased with more progeny per sire. Bias

TABLE 9. Effect of progeny group size (ND) on bias in estimation of additive genetic parameters with model A1 for structure I.[1]

	parameters		group effects			sire effects		
ND	σ_a^2	h^2	g_{p_3}	g_{p_5}	g_{p_7}	$s_{1..}$	$s_{5..}$	$s_{9..}$
25	195708	.369	310	525	565	-25	-70	92
50	208168	.390	301	502	514	-45	-85	127
100	223044	.415	297	464	453	-70	-95	160

[1] heterosis= 5%, recombination loss= -5%

in estimation of group effects for progeny with 25%HF and 50%HF, which is due to non-additive effects, was smaller and bias for the 75%HF progeny group, which is due to overestimation of 100%HF sires, was larger with increasing number of progeny.

DISCUSSION

Non-additive models have been proposed in the literature, including those with random non-additive genetic effects due to dominance variation within breeds (e.g., 5). Those effects were not considered in this study because they are likely to be less important than fixed non-additive effects due to interactions between breeds. Another simplification of the simulated model was the assumption of equal genetic variances across populations and across crossbred groups before genetic equilibrium was reached. Hence, additive genetic variation was expected to be homogeneous, i.e. differences in gene frequency were small. This is justifiable for two closely related breeds with no inbreeding and a trait that is determined by an infinite number of loci. However, it may be worthwhile to investigate models that account for heterogeneity of additive genetic variance across crossbred groups.

More family information is often used in sire evaluation by incorporating numerator additive genetic relationships between sires. Pollak and Quaas (17) have pointed out the association of relationships with genetic groups, i.e. groups can account for selection differentials between generations not accounted for by relationships. Those groups, however, are defined on the basis of time to account for short-term selection. Differences between populations or breeds are based on long-term selection and group effects reflect effects of gene substitution and gene interaction. Hence, use of genetic relationships between generations is not expected to change the interpretation of group effects. However, bias in random effects, due to not accounting for non-additive effects, may decrease when information from relatives is used across genetic groups of sires.

An alternative strategy for analysis of data from crossbred populations could be the use of an animal model so that additive genetic variance would be estimated from the random additive values of animals. Correction for

genetic means could be by groups that are defined on a basis of fraction of HF- genes of animals making the records. However, an additive animal model would not account for different non-additive effects within groups, so that additive genetic variance would likely be overestimated.

In this simulation study, distribution of sire's progeny over dam groups was balanced and distribution of sire and dam groups over environmental effects was random i.e. all herds have dams with on average equal breed composition. Meyer (13) showed that environmental effects might account for differences between genetic groups when groups are partially confounded with environmental effects.

Heterosis levels in crossbred populations, in particular in crosses between different strains of Friesian cattle, might be small in temperate climates (2,14). However, in addition to levels of heterosis and recombination, the genetic structure of the data influences bias in estimating genetic variance and breeding values. Bias depends on the distribution of data over sire group by dam group combinations. In countries where sires are imported from other populations, a progeny group model gives more bias in estimation of additive genetic variance as dams have higher fraction of HF-genes.

European countries recently have been interested in making comparisons of genetic merit between and within dairy cattle populations. Philipsson (16) mentioned some problems of comparing breeding stock internationally, such as possible bias due to selective mating and to special treatment of progeny. Bias in prediction due to non-additive effects was not considered. However, accounting for non-additive effects affects international comparisons of breeding stock. Estimating additive genetic values with a sire group model gives maximum bias through comparison of bull groups of the pure breeds, and mating them with dams from one of the breeds. Breed difference would be estimated as twice the group difference, hence overestimation would be equal to twice the heterosis. The impact of non-additive effects on sire ranking, based on total genetic effect, is dependent upon the heterogeneity of the cow population. However, a correct sire evaluation procedure should specify the components of the total genetic merit.

CONCLUSION

Low levels for heterosis and recombination affected estimators of additive genetic variance in crossbred populations. Predictions of breeding values and estimation of breed difference were considerably biased with additive models. It is necessary to estimate heterosis and recombination loss in actual populations to assess the problem of bias in genetic evaluations.

REFERENCES

1. Dickerson, G.E. 1973. Inbreeding and heterosis in animals. Pages 54-77 in Proc. Anim. Breeding and Genet. Symp.in Honor of Dr. J.L. Lush. Am. Soc. of Anim. Sci. and Am. Dairy Sci. Assoc., Blacksburg V.A.
2. Ericson, K. and B. Danell. 1986. Crossbreeding effects between Swedish dairy breeds for production and reproduction traits. 37th Ann. Meeting EAAP, Budapest, Hungary, vol 1:17 (summary).
3. Fimland, E.A. 1983. Methods of estimating the effect of heterosis. Z. Tierz. Zuchtungsbiol. 100:3.
4. Gjøl-Christensen, L. and J. Pedersen. 1988. Krydsning af malkekvaegracer. 650 Beretning fra Statens Husdyrbrukforsøg. København.
5. Henderson, C.R. 1985. Best linear unbiased prediction of non-additive genetic merits in noninbred populations. J. Anim. Sc. 60:111.
6. Hill, W.G. 1982. Dominance and epistasis as components of heterosis. Z. Tierz. Zuchtungsbiol.99:161.
7. Hill, W.G., M.R. Edwards, M.K. Ahmed and R. Thompson. 1983. Heritability of milk yield and composition at different levels and variability of production. Anim. Prod. 36:59.
8. Kinghorn, B. 1980. The expression of recombination loss in quantitative traits. Z. Tierz. Zuchtungsbiol. 97:138.
9. Kinghorn, B. 1982. Genetic effects in crossbreeding. I:Models of merit. Z. Tierz. Zuchtungsbiol.99:59.
10. Maijala, K., and M. Hanna, 1974. Reliable phenotypic and genetic parameters in dairy cattle. Proc. 1st World Congr. Genet. Appl. to Livest. Prod. Madrid. vol I:541.

11.Meyer, K. 1986. Restricted Maximum Likelihood estimation of genetic parameters-in practice. Proc. 3rd World Congr. Genet. Appl. to Livest. Prod., Lincoln. vol XII:454.

12.Meyer, K. 1987. Estimates of variances due to sire x herd interactions and environmental covariances between paternal half sibs for first lactation dairy production. Livest. Prod. Sci. 17:95.

13.Meyer, K. 1987. Mixed model analysis when genetic and environmental groups are closely linked. Z. Tierz. Zuchtungsbiol. 104:321.

14.McAllister, A.J. 1986. The role of crossbreeding in breeding programs for intensive milk production in temperate climates. Proc. 3rd World Congr. Genet. Appl. to Livest. Prod., Lincoln, vol IX: 47.

15.McGloughin, P. 1980. The relationship between heterozygosity and heterosis in reproductive traits of mice. Anim. Prod. 30:69.

16.Philipsson, J., 1987. Standards and procedures for international genetic evaluations of dairy cattle. J. Dairy Sci. 70:418.

17.Pollak, E.J., and R.L. Quaas. 1983. Definition of group effects in sire evaluation models. J. Dairy Sci. 58:1503.

18.Schulte-Coerne, H. and D. Boie. 1986. Heterosis in crosses between European and American dairy breeds. 37th Ann. Meeting EAAP, Budapest, Hungary, vol 1:21 (summary).

19.Sheridan, A.K. 1981. Crossbreeding and heterosis. Anim. Br. Abstr. 49:131.

20.Van der Werf, J.H.J, A.E. van Veldhuizen and S. Korver. 1987. Relationship between young bull performance and dairy performance of progeny. In: S. Korver, G. Averdunk and B.B. Andersen (Editors), Performance testing of AI bulls for efficiency and beef production in dairy and dual purpose breeds. EAAP-publication no. 34, Pudoc, Wageningen, The Netherlands. ISBN 90-220-0924-6.

21.Van der Werf, J.H.J. and W. de Boer. 1989. Estimation of genetic parameters in a crossbred population of Dutch Black and White dairy cattle. J. Dairy Sci. 72:2615.

22.Wilmink, H. and F. de Graaf. 1986. Genetic parameters for 305-days yield in first lactation and adjustment of Net Guilder Index. Report 86-676, Dutch Royal Cattle Syndicate, Arnhem. (Dutch with English summary).

Chapter 3

ESTIMATION OF GENETIC PARAMETERS IN A CROSSBRED POPULATION OF BLACK AND WHITE DAIRY CATTLE

J.H.J. VAN DER WERF AND W. DE BOER
Department of Animal Breeding
Wageningen Agricultural University
PO Box 338, Wageningen
The Netherlands

Published in Journal of Dairy Science 72: 2615-2623 (1989)

ABSTRACT

Genetic parameters were estimated using data of cows with variable proportions of genes from two breeds: Dutch Friesian and Holstein Friesian. The data set contained 92,333 first lactation records (305 days milk production) from 675 young sires and 307,050 records from 202 proven sires.

Data were analyzed using four additive mixed models with genetic groups defined according to 1) breed composition of the cow, 2) breed composition of sire and dam, 3) linear regression on the fraction of Holstein Friesian genes of the cow and 4) breed composition of sire. A non-additive model included a linear regression on breed fraction, heterozygosity and recombination in the cow's genome.

Estimates for heterosis varied from 2.5% (fat yield) to 0% (protein percentage). Recombination effects varied from -1.9% (protein yield) to 1.5% (fat percentage). Additive models with progeny groups overestimated genetic variance by 6%. Models with sire groups overestimated additive genetic values of imported Holstein Friesian sires by 33%. Using a non-additive model, heritability estimates were .38 for milk yield, .80 for fat percentage and .70 for protein percentage. It was concluded that a non-additive model was preferable for estimation of genetic variance and prediction of breeding values in crossbred dairy populations.

INTRODUCTION

Numerous estimates of genetic parameters for production traits have appeared in the dairy breeding literature (1,2,9,12,14,16,22,23). Such focus can be justified because production traits have high economic importance, and heritabilities and variances may change due to selection, migration of genes from one population to another, or changing environmental conditions. Recently, genes from North American dairy breeds have been imported to European populations. Genetic parameters should be estimated regularly in such a process of upgrading from data collected in well organized milk recording schemes.

Genetic parameters for milk production traits were summarized by Maijala and Hanna (12). The average heritability estimates were .27 for first

lactation milk yield and .24 for fat yield. Estimates using more sophisticated methods (i.e. REML) showed no major differences for first lactation traits (14, 18). More recently, higher heritabilities were found for data from Canada (2) and from the United States (22). Hill et al. (9) have shown that heritabilities tended to increase with production. Those authors used data mainly from single breed populations and thus parameters may not necessarily be valid for upgraded European dairy populations.

Heritability estimates in crossbred populations were adjusted to various degree for fixed additive effects of breed contributions. Meyer (14) found higher heritability for first lactation milk and fat yield when data from imported Canadian sires were included and breed effects were not accounted for. Other authors accounted for breed of sire but found also high heritabilities in crossbred data (1,16,23). Van der Werf et al. (20) found heritabilities of .41 for milk yield and .79 for fat yield after accounting for breed of progeny.

Models for estimating additive genetic variance and heritabilities from data of crossbred populations usually do not include non-additive effects. Heterosis and recombination effects are considered to be small for milk production traits (13). However, not accounting for these non-additive effects may cause upward bias in the estimation of additive genetic parameters (19). Elzo and Famula (4) proposed a general strategy to estimate sire effects, accounting for fixed additive and non-additive breed effects. Other models proposed for analysis of crossbred populations were intended to estimate effects of breed differences, heterosis and recombination (6,10). However, random genetic effects within breed groups were ignored.

The objectives of this paper were to estimate additive and non-additive genetic parameters for milk production traits in a crossbred population. Models accounting differently for fixed additive and non-additive breed effects were compared. Differences in estimates of genetic parameters from sub-populations were investigated.

MATERIAL AND METHODS

First lactation records on 305 days milk, fat and protein yield were obtained from the Dutch Dairy Cattle Syndicate (NRS). Records were sampled from Black and White heifers freshening between August 1983 and September 1986. Records had been previously adjusted for month of calving, and incomplete lactations of 90 days or more had been extrapolated to 305 days according to the methods described by Wilmink (24). Data on the following seven traits were obtained for each cow; milk yield (kg), fat yield (kg), protein yield (kg), fat percentage, protein percentage, fat protein corrected milk (FPCM kg) (11), and carrier (milk-fat-protein) (kg).

Crossbred AI sire progeny of the Dutch Friesian (FH) and the Holstein Friesian (HF) breed were used. Breed composition (% HF genes) was known for all sires and their progeny. Editing of data included checks on age of calving (21 to 32 months) and on breed codes of sire and progeny. Total number of valid records was 451,261.

Heifers were considered to belong to the sire's first group of daughters when the sire was younger than 70 months at the start of their lactation. Young sires were included that had a minimum of 75 progeny distributed over at least 50 herds. Omission of young sires was mostly due to not having a complete first batch of daughters within the period considered. Therefore, elimination was assumed to be unrelated to sire's genetic merit. In total, 92,333 records from 675 young sires were selected.

Records of proven sires improve connectedness in the data and therefore contribute to more accurate estimation of the herd-year-season effects and the within sire variance (21). Therefore, records of herdmates of young bull's progeny were also included in the analysis. Herdmates were required to descend from proven sires that were older than 80 months at time of initiation of lactation and had a minimum of 100 daughters in 75 herds. Herdmate records of 307,050 heifers from 202 proven sires were added to the dataset.

Nine genetic groups were defined according to percentage of HF genes at intervals of 12.5%. The distribution of progeny over groups of sire and groups of dam is expressed in Table 1. The distribution of sires over genetic groups is expressed in Table 2. Male ancestry was known for young sires, which consisted of 93 base animals.

TABLE 1 Distribution of data (%) over sire and dam groups[1,2].

group of dam	group of sire 1	2	3	4	5	6	7	8	9	total
1	24.7	0.1	0.2	0.0	5.7	0.1	3.2	0.9	49.1	84.1
2	-	0.0	-	0.0	-	0.0	-	0.2	-	0.2
3	0.4	-	0.0	-	0.3	-	0.2	-	1.4	2.2
4	-	0.0	-	0.0	-	0.0	-	0.1	-	0.1
5	1.5	-	0.0	-	0.8	-	0.6	-	8.9	11.9
6	-	-	-	-	-	0.0	-	0.2	-	0.2
7	0.0	-	0.0	-	0.0	-	0.0	-	0.7	0.8
8	-	-	-	-	-	0.0	-	0.0	-	0.0
9	0.0	-	0.0	-	0.0	0.0	0.0	0.0	0.4	0.4
total	26.7	0.1	0.3	0.0	6.8	0.1	4.0	1.4	60.5	100 [3]

[1] group number i corresponds to (i-1)*12.5 % Holstein Friesian genes
[2] classes with "-" have no records
[3] 100% = 399,383 records.

TABLE 2. Distribution of number of sires over genetic groups[1].

sires	group of sire 1	2	3	4	5	6	7	8	9	total
base	32	-	-	-	-	-	-	-	61	93
young	174	2	10	1	127	4	111	35	211	675
proven	98	-	-	-	9	-	2	1	92	202

[1] group number i corresponds to (i-1)*12.5 % HF genes

MODELS FOR ANALYSIS

Several grouping strategies can be followed to attempt to account for fixed genetic effects of subpopulations (19). Four additive models that varied in their definition of genetic groups were considered. A non-additive model (NA) also was considered, which accounted for interactions between breed groups due to effects of heterosis and recombination loss.

Model A1 was an additive model with grouping according to fraction of

HF genes in the progeny from which the records originated (g_p). The model was described as:

$$y_{ijklmn} = hys_{in} + b_{1n}(A_{ijklm}-\bar{A}_{.....}) + b_{2n}(A_{ijklm}-\bar{A}_{.....})^2 + m_{jn} + g_{p_{kn}} + s_{1n} + e_{ijklmn} \quad [1]$$

where:

y_{ijklmn}= ijklmth observation for the nth trait (n=1,..,7),

hys_{in}= the fixed effect of the ith herd-year-season class for the nth trait (i=1,...,62605). Two seasons were distinguished per herd-year: February to August and September to January,

A_{ijklm} = the calving age (months) of the ijklmth cow

$\bar{A}_{.....}$ = the mean calving age,

b_{1n}= the linear regression coefficient of age for the nth trait,

b_{2n}= the quadratic regression coefficient of age for the nth trait,

m_{jn}= the fixed effect of the jth month of calving class for the nth trait (j=1,..,12),

$g_{p_{kn}}$= the fixed effect of the kth progeny group class for the nth trait; group referring to percentage HF genes (k=1,...,9),

s_{1n}= the effect of the 1th sire for the nth trait. Effects of young sires were considered random while proven sires were treated as fixed effects,

e_{ijklmn}= the random residual effect for the nth trait associated with the ijklmth cow.

Model A1 was referred to as additive progeny group model. The model can be written in matrix notation as,

$$y = Xb + Qg_p + Zs + e \quad [2]$$

where b is a vector of fixed environmental effects, g_p is a vector of genetic effects of the progeny group, s is a vector of sire effects and e is a vector of residual effects. Design matrices for fixed effects, group effects and sire effects are X, Q and Z, respectively. The matrix Z and s' was partitioned into $[Z_1\ Z_2]$ and $[s_1'\ s_2']$, where Z_1 and Z_2 are matrices relating records to proven (s_1) and young sires (s_2), respectively.

First moments for model A1, treating s_1 sires as fixed, are;

$$E(y) = Xb + Qg_p + Z_1s_1 \quad [3].$$

Model A2 was a parent group model. It can be written as,

$$y = Xb + Q_1g_s + Q_2g_d + Zs + e \quad [4]$$

with g_s and g_d being additive genetic effect of sire group and dam group and Q_1 and Q_2 design matrices for sire and dam group, respectively. Model A3 is a model with a linear regression of performance on the fraction HF genes of the progeny,

$$y = Xb + b_1p_p + Zs + e \quad [5]$$

where p_p is a vector with the fraction HF genes for each animal, and b_1 is the regression of y on breed composition of the animal making the record. It should be noted that models A1, A2 and A3 account for additive genetic differences between crossbred groups.

Model A4 was an additive sire group model,

$$y = Xb + Q_1g_s + Zs + e. \quad [6]$$

This model is used in many countries for sire evaluation.

Expectations for y were similar to [3] for models A2, A3 and A4, except for the second term which was replaced by $Q_1g_s + Q_2g_d$, b_1p_p, and Q_1g_s respectively.

The NA model was defined as,

$$y = Xb + b_1p_p + b_2h + b_3r + Zs + e \quad [7]$$

where b_2 and b_3 are regressions on vectors with coefficients for heterozygosity effect (h) and recombination effect (r) for each animal (3,19). Additive genetic differences between crossbred groups were accounted for by a linear regression on the breed composition of the progeny making the record.

The expectation of y under the NA model, treating s_1 as fixed, is

$$E(y) = Xb + b_1p_p + b_2h + b_3r + Z_1s_1 \quad [8].$$

The dispersion matrix of y, treating s_1 as fixed, was for each model,

$$var(y) = Z_2AZ_2'\sigma_s^2 + I_n\sigma_e^2 \quad [9]$$

with $var(s_2) = A\sigma_s^2$, $var(e) = I_n\sigma_e^2$ and $cov(e,s_2) = 0$. The matrix A contains additive genetic relationships between the young sires. The vector s_2 was extended to include base animals (sires and maternal grandsires of the young bulls). Sire variances and covariances were assumed to be homogeneous across populations.

Variance components were estimated by REML. A univariate procedure was used for analysis of each trait. For estimation of genetic correlations between traits, a multivariate REML procedure was used. Computations were made feasible by transforming the data to canonical variates, possible because using design matrices were the same for each trait (15). Sampling errors for the variances were approximated using the estimates as true values. Sampling errors for the genetic parameters were derived from a linear approximation using Taylor series expansion (14). Breeding values of sires were computed from mixed model solutions for fixed additive genetic effects (groups or regression) and from predicted sire effects.

RESULTS AND DISCUSSION

Estimation of Non-additive Genetic Parameters

Overall mean and solutions for heterosis and recombination effects for each trait from model NA are in Table 3. Estimates for heterosis were small but significant for yield traits (about 2.5%). When considering the genetic distance between the Dutch Friesian and the Holstein Friesian breed, estimates for heterosis were in agreement with other estimates described in literature (5, 13). Estimates for milk composition traits were smaller than 1%.

Estimates for effect of recombination were negative and smaller than heterosis (Table 3). Heterosis was assumed to represent dominance effects

and half of additive by additive effects, whereas the recombination effect represented half of the additive by additive effects (19). Literature values for recombination loss often are not significant (13) or are significant but small (5). The recombination effect for fat percentage was positive, which results from a smaller recombination effect for fat yield in comparison with milk yield. The heterosis for fat percentage was smaller than the recombination effect, which implies that the dominance effect of fat percentage was negative.

TABLE 3. Estimates of heterosis, recombination effects, with standard errors and overall means for milk production traits.

trait	heterosis	SE	recombination	SE	mean
milk yield	122.9	5.4	-101.2	13.5	5299
FPCM[1]	136.93	5.1	-75.93	12.8	5502
carrier	112.44	5.0	-95.72	12.6	4891
fat yield	5.959	.22	-1.325	.55	229.6
protein yield	4.367	.17	-3.457	.42	178.4
fat%	.013	.003	.0640	.006	4.34
protein%	.001	.001	-.0056	.003	3.37

[1] fat protein-corrected milk

Estimates for non-additive parameters and their low standard errors were consistent and in agreement with analysis of data simulated (19), using a distribution over sire- and dam groups similar to that described in this study. Low sampling errors might have been due to the use of regression to estimate heterosis, recombination, and breed effects. Regression was used rather than interactions between subclass effects, which have larger sampling errors. Results showed that it was feasible to obtain reliable estimates of crossbreeding effects from field data. Although the distribution of data over crossbred groups was unbalanced, many combinations were represented with considerable amounts of data.

Comparison of Models for Milk Yield

Estimates of variance components and heritability for milk yield are in Table 4 for each model. Van der Werf and De Boer (19) have shown by simulation that all models give similar results in absence of non-additive

effects. With models A1 and A3, however, non additive effects caused inflated estimates for sire variance and heritability. Results from the NA model were assumed to be unaffected by non-additive genetic effects (19).

The overestimation of σ_s^2 and h^2 was 6% using model A1. Simulation results, using a sire group by dam group structure as expressed in Table 1, showed an overestimation of σ_s^2 by 2.5% when heterosis was 5% and recombination loss was -5% (19). In a simulated structure with increased heterozygosity of dams, bias increased dramatically to 19%. In the data of this study, young sires (random) were mated to relatively more F1 dams, which explains the significant bias in spite of low levels of heterosis and

TABLE 4. Estimates of sire variance ($\hat{\sigma}_s^2$), residual variance ($\hat{\sigma}_e^2$) and heritability ($\hat{h}^2$) for first lactation milk yield (kg) with different models.

model	$\hat{\sigma}_s^2$ [1]	$\hat{\sigma}_e^2$ [2]	$\hat{h}^2$ [3]
A1	49,525	443,285	.402
A2	45,586	443,438	.373
A3	53,351	443,473	.430
A4	45,647	443,986	.373
NA	46,553	443,320	.380

[1] approximated SE varied from 2968 kg^2 (A4) to 3422 kg^2 (A3).
[2] approximated SE were 1082 kg^2 for each model.
[3] approximated SE were .02 for each model.

recombination loss. Estimates for the residual component differed only slightly for all models. In accordance with sire variance, heritability was biased upward for models A1 and A3.

Estimates for additive genetic differences between crossbred groups and average breeding values of sires within groups are shown in Table 5. To make solutions comparable over models, groups with 0% HF were restricted to 0 for each model and solutions of sire and dam groups were multiplied by 2. For models A3 and NA, additive genetic differences between groups were derived from the estimated regression on corresponding HF percentage.

Solutions for 50% HF and 100% HF groups were relatively higher with models A2 and A4 than with other models. Compared to the non-additive model (NA), breed differences were overestimated by 70% using sire group

solutions. Differences between dam groups were small using model A2. Neither sire group nor dam group solutions from an additive model could be interpreted as representing half of additive genetic differences between subpopulations. Group solutions from the additive progeny group model A1 were more in agreement with the NA model. The 50% HF progeny group was overestimated and the 100% HF progeny group was underestimated. Results from Table 5 agreed with previously reported results from simulation (19).

Estimates of breeding values were similar for additive models A2, A3 and A4 (Table 5). Average estimated breeding values of sires from those models were higher than estimates from the NA model; by about 20% for 50% HF sires and about 30% for 100% HF sires. The progeny group model underestimated 75% HF sires, whereas 100% HF sires were overestimated by 10%.

Breed difference between HF and FH was estimated at 530 kg with the NA model, whereas while differences between breeding values of sires averaged 680, i.e., random effects of 100% HF sires were positive after correcting progeny records for fixed additive and non-additive genetic effects. The mean of effects of 100% HF sires was about 5 times its standard deviation. This may have been caused by assortative mating of young 100%HF bulls or by favorable treatment of their progeny. Such effects would be confounded with sire group effects in models A2 and A4. Models with sire groups gave therefore lower estimates for the sire variance than the NA model (Table 4).

TABLE 5. Estimates of fixed additive genetic effects and average breeding values for sires of crossbred groups from different models (milk yield).

model	group effect						average sire breeding value		
	50%HF[1]		75%HF		100%HF		50%HF	75%HF	100%HF
		SE		SE		SE			
A1	298	4	346	6	432	20	326	356	753
A2-sires[2]	434	11	582	13	904	7	367	531	900
A2-dams[2]	135	8	280	27	304	37			
A3	154	3	231	5	308	6	368	524	874
A4[1]	437	11	585	13	906	7	371	534	902
NA	265	5	397	7	530	10	302	408	680

[1] Holstein Friesian
[2] solutions for sire and dam groups were multiplied by 2

Estimation of Additive Genetic Parameters for all Traits

Variance components and heritability of each trait from model NA are in Table 6. Estimates were corrected for fixed effects of breed and for interaction between breeds. Within trait parameters from single trait models differed only slightly (<.2%) from estimates from a multivariate procedure. Heritability for milk yield was higher than that given by Maijala and Hanna (12) and somewhat higher than more recent estimates (1, 16, 22). Heritabilities for milk composition traits were substantially higher than most literature values given for single breeds. However, heritability estimates for milk yield and fat percentage were very similar to those from a crossbred population of Holstein Friesian x European Friesian cows (1).

TABLE 6. Estimates of sire variance ($\hat{\sigma}_s^2$), residual variance ($\hat{\sigma}_e^2$), and heritability ($\hat{h}^2$) for milk production traits with a non-additive model (NA).

trait	$\hat{\sigma}_s^2$	$\hat{\sigma}_e^2$	$\hat{h}^2$	SE
milk yield	46,553	443,320	.380	.02
FPCM[1]	34,941	400,893	.321	.02
carrier	42,264	389,960	.391	.02
fat yield	74.84	753.8	.361	.02
protein yield	37.87	422.9	.329	.02
fat %	.024	.0965	.799	.03
protein %	.0046	.0216	.701	.03

[1] Fat protein-corrected milk

Estimates for genetic and phenotypic correlations between traits are in Table 7. Genetic correlations agreed with literature values, except for the correlation between milk yield and fat yield, which was lower in this study, and between milk yield and milk composition traits, which were more negative than most values in literature. Correlations were quite similar to those found by Boichard and Bonaïti (1)

The lower correlation between milk yield and fat yield agreed with the more negative correlation between milk yield and the ratio of fat to milk. The correlation between milk yield and the ratio of protein to milk was more negative due to a relative higher genetic variability for milk yield.

Phenotypic correlations were lower between milk and fat yield and higher between milk yield and milk composition traits.

TABLE 7. Estimates of phenotypic (above diagonal) and genetic (below diagonal) correlations for milk production traits with a non-additive model (NA).

	kg milk	FPCM	carrier	kg fat	kg prot.	fat%	prot.%
milk, kg	---	.944	.999	.796	.929	-.388	-.409
FPCM[1]	.883	---	.932	.947	.960	-.075	-.190
carrier, kg	.999	.861	---	.776	.920	-.417	-.432
fat, kg	.583	.889	.548	---	.853	.237	-.057
protein, kg	.868	.927	.851	.722	---	-.192	-.052
fat %	-.515	-.062	-.550	.393	-.222	---	.580
protein %	-.524	-.189	-.551	.059	-.035	.657	---

[1] fat protein-corrected milk

Estimation of Genetic Parameters from Subpopulations

Because variance among sires may not be equal for different breed groups, the data set was divided into subsets according to breed composition of the sire. Three subsets with progeny from 174 young FH sires in set I, from 127 young 50% HF sires for set II, and from 211 young HF sires for set III were analyzed. Herdmate records from proven sires were used in each subset, irrespective of breed composition. Progeny within subset were not necessarily from identical subpopulations because dams were from different breed groups. The NA model was used to correct for possible bias due to non-additive effects.

Estimates of sire variance and heritabilities in subpopulations are given in Table 8. Sire variances for milk yield, carrier yield, fat yield, and composition traits were significantly larger in subset III than in other subsets. Standard errors on heritabilities in subset III varied from .03 to .07. Results suggested that variances and heritabilities were not equal for the different populations.

TABLE 8. Estimates of sire variance and heritability for milk production traits from subsets of data with progeny of sires with 0% Holstein Friesian, 50% HF, and 100% HF genes (model NA).

	heritability			sire variance		
trait	0%HF	50%HF	100%HF	0%HF	50%HF	100%HF
milk, kg	.339	.307	.444	34,743	35,960	59,499
FPCM	.331	.280	.332	31,136	29,535	38,842
carrier, kg	.341	.313	.463	30,601	32,245	54,865
fat, kg	.332	.307	.384	57.24	61.42	85.55
protein, kg	.358	.288	.336	36.23	32.23	41.35
fat %	.490	.570	1.00	.0106	.0164	.0338
protein %	.507	.582	.852	.0029	.0038	.0058

[1] fat protein corrected milk

However, heritabilities from subset III were also higher than recent estimates from North American and Canadian HF populations (2, 22). Variances and heritabilities from subsets I and II better matched literature values. Differences between subsets were relatively small for protein yield and FPCM.

Breed differences between subsets were partly confounded with other effects. Progeny from 100% HF sires freshened more in later years. They also might have performed in herds with better management. Residual variance was 375,194 for subset I and 476,706 for subset III. Allowing for differences in scale, there was still a clear distinction between sire variance from subsets I and III. Increase in genetic variation has been correlated with level of production (9) or with herd type (16).

Overestimation of sire effects of 100% HF bulls increases the estimate for sire variance. Differences between subsets might have been caused by assortative mating and preferential treatment of progeny of the young bulls from subset III. However, models with sires nested within groups (A2 and A4) also showed high heritabilities. Another bias might have arisen from selection of young bulls based on pedigree indexes. This type of selection is not accounted for when information of sires' ancestors is not included in the analysis (17). Although single trait selection would have reduced genetic variation (17), the estimated variance among the 100% HF sires was larger than in the American HF population. In contrast to sires from subsets I and II, sires from subset III descended from imported cows or

were imported themselves. Breeding values for imported sire may have been extreme either for milk yield or fat percentage, which are negatively correlated. As a result, genetic variance among imported sires would have been increased for each trait. Methods to accommodate for selection occurring prior to the formation of the base population (7,8) could be particularly important for populations that import sires.

CONCLUSION

Small effects of heterosis and recombination were shown for Holstein Friesian x Dutch Friesian crosses. Nevertheless, differences between the non-additive model and additive models were substantial for estimates of breed differences, genetic parameters and breeding values across breeds. The use of non-additive models was therefore warranted for analysis of crossbred populations.

Estimates of genetic parameters differed from known values, in particular for milk and fat yield, fat percentage and protein percentage. Analyses of subpopulations revealed higher genetic variances for data from progeny of imported sires. More research is needed to determine to what extent variances among imported sires are biased by selection on pedigree.

ACKNOWLEDGMENTS

The authors acknowledge Karin Meyer for the use of her REML programs and Brian Kennedy for helpful suggestions.

REFERENCES

1. Boichard, D., and B. Bonaïti. 1987. Genetic parameters for first lactation dairy traits in Friesian, Montbeliarde and Normande Breeds. Génét. Sél. Evol. 19:337.
2. Cue, R.I., H.G. Monardes, and J.F. Hayes. 1987. Correlations between production traits in first lactation Holstein cows. J.Dairy Sc. 10:2132.

3. Dickerson, G.E. 1972. Inbreeding and heterosis in animals. Pages 54-77 in Proc. Anim. Breeding and Genet. Symp. in honor of Dr. J.L. Lush. Am. Soc. Anim. Sci. and Am. Dairy Sci. Assoc., Blacksburg, VA.
4. Elzo, M.A., and T.R. Famula. 1985. Multibreed sire evaluation procedures within a country. J. Anim. Sci. 60:942.
5. Ericson, K. 1987. Crossbreeding effects between two Swedish dairy breeds for production and reproductive traits. Report 76. Dep. of Anim. Breeding Genet. Uppsala, Sweden. ISBN 91-576-3217-0.
6. Fimland, E.A. 1983. Methods of estimating the effect of heterosis. Z. Tierz. Zuchtungsbiol. 100:3.
7. Graser, H.U., S.P. Smith, and B.Tier. 1987. A derivative free approach for estimating variance components in animal models by restricted maximum likelihood. J. Anim. Sci. 64:1362.
8. Henderson, C.R. 1985. Best linear unbiased prediction using relationship matrices derived from selected base populations. J.Dairy Sci. 68:443.
9. Hill, W.G., M.R.Edwards, M.K. Ahmed and R.Thompson. 1983. Heritability of milk yield and composition at different levels and variability of production. Anim. Prod. 36:59.
10. Kinghorn, B. 1982. Genetic effects in crossbreeding. I: Models of merit. Z. Tierz. Zuchtungsbiol. 99:59.
11. Korver, S. 1982. Feed intake and production in dairy breeds dependent on the ration. Ph.D. Thesis, Agric. University, Wageningen, The Netherlands.
12. Maijala, K., and M. Hanna. 1974. Reliable phenotypic and genetic parameters in dairy cattle. Proc. 1st World Congr. Genet. Appl. to Livest. Prod., Madrid. I:542.
13. McAllister, A.J. 1986. The role of crossbreeding in breeding programs for intensive milk production in temperate climates. Proc. 3rd World Congr. on Genet. Appl. Livest. Prod., Lincoln. IX:47.
14. Meyer, K. 1984. Estimates of genetic parameters for milk and fat yield for the first three lactations in British Friesian cows. Anim. Prod. 38:313.
15. Meyer, K. 1985. Maximum Likelihood estimation of variance components for a multiple trait mixed model with equal design matrices. Biometrics 41:153.

16. Meyer, K. 1987. Estimates of variances due to sire x herd interactions and environmental covariances between paternal half sibs for first lactation dairy production. Livest. Prod. Sci. 17:95.
17. Sorensen, D.A. and B.W. Kennedy. 1984. Estimation of genetic variances from unselected and selected populations. J. Anim. Sci. 59:1213.
18. Tong, A.W., B.W. Kennedy, and J.E. Moxley. 1979. Heritabilities and genetic correlations for the first three lactations from records subject to culling. J. Dairy Sci.62: 1784.
19. Van der Werf, J.H.J and W. de Boer. 1989. Influence of non-additive effects on the estimation of additive genetic parameters in dairy cattle. J. Dairy Sci. 72:2602.
20. Van der Werf, J.H.J., A.E. van Veldhuizen and S.Korver. 1987. Relationship between young bull performance and dairy performance of progeny. S. Korver, G. Averdunk and B.B. Andersen (Editors), Performance testing of AI bulls for efficiency and beef production in dairy and dual purpose breeds. Eur. Assoc. Anim. Prod. publ. no. 34, Pudoc, Wageningen, Neth. ISBN 90-220-0924-6.
21. Van Vleck, L.D. 1985. Including records of daughters of selected bulls in estimation of sire components of variance. J. Dairy Sci. 68:2396.
22. Van Vleck, L.D., M.C. Dong and G.R. Wiggans. 1988. Genetic (co)variances for milk and fat yield in California, New York, and Wisconsin for an animal model by restricted maximum likelihood. J. Dairy Sci. 71:3053.
23. Wilmink, J.B.M. and F.M. de Graaf, 1986. Genetic parameters for 305 days yield in first lactation and adjustment of Net Guilder Index, Report 86-676, Dutch Royal Cattle Syndicate, Arnhem, Neth. (Dutch with English summary).
24. Wilmink, J.B.M. 1987. Studies on test day and lactation milk, fat and protein yield of dairy cows. Ph.D. Thesis. Wageningen Agric. Univ., Neth.

Chapter 4

ESTIMATION OF ADDITIVE GENETIC VARIANCE WHEN BASE POPULATIONS ARE SELECTED

J.H.J.VAN DER WERF AND I.J.M. DE BOER
Department of Animal Breeding
Wageningen Agricultural University
PO Box 338, 6700 AH Wageningen
The Netherlands

Accepted for publication in Journal of Animal Science

ABSTRACT

A population of size 40 was simulated 1000 times for ten generations. Five out of twenty males were selected each generation and each male was mated to four females to have two progeny. The additive genetic variance (σ_a^2) before selection was 10 and the initial heritability was .5. Due to covariances among animals, inbreeding and gametic disequilibrium, the genetic variance was reduced to 6.72 after ten generations of selection. Reduction of variance was lower in another population simulated with size 400 and ten percent of the males selected. Restricted Maximum Likelihood was used to estimate σ_a^2 using an animal model. The estimate of σ_a^2 was empirically unbiased when all data and all relationships were used. Omitting data from selected ancestors caused biased estimates of σ_a^2 due to not accounting for all gametic disequilibrium. Including additional relationships between assumed base animals adjusted for inbreeding and for covariances. Bias from gametic disequilibrium decreased slightly with the use of more relationship information, and it was smaller in the small population, and when selection had been practiced for just a few generations.

INTRODUCTION

Parameters for production traits often are estimated from data on selected animals. It has been shown that, in principle, selection can be accommodated by an appropriate model that includes all data upon which selection decisions were based (Henderson, 1975; Sorensen and Kennedy, 1984b, Gianola and Fernando, 1986). However, the time span covered by data generally is limited and in practice all data since the start of selection are not available.

Sorensen and Kennedy (1984b) simulated several generations of selection and omitted data from earlier generations. They concluded that the estimate of the additive genetic variance before selection was nearly unbiased when their model acknowledged all relationships that developed in previous generations. This seems inconsistent with the condition that all data are needed for unbiased estimation.

Since information on pedigree usually does not date back to a true base population, an assumption concerning a base population is made, e.g. the first generation of animals with data are considered as unrelated, unselected and noninbred base animals. Sorensen and Kennedy (1984b) indicated that such a model would estimate the genetic variance in the implied base generation. However, they analyzed data from only two generations with an animal model. The estimate would be the equivalent of offspring on parent regression which is unbiased by selection of parents. This may not be true if more generations are included.

The aim of this research was to study estimates of additive genetic variance when base populations were selected. We examine how estimates are affected by omitting data from selected ancestors and by ignoring known relationships of selected animals and then systematically including more relationship information among animals.

MATERIAL AND METHODS

Simulated Data

This study follows the Monte Carlo simulation strategy described by Sorensen and Kennedy (1984a). They assumed a genetic model where a large number of unlinked loci contribute to the genetic variance. A number of s males and m females, all assumed unrelated and unselected, were randomly sampled from a conceptually infinite base population. The base animals were mated at random (m/s= d females per male) to produce m males and m females, which is generation 1. The s phenotypically best males were selected in each of the subsequent generations. Each of the females was randomly mated to one of the selected males. Selection was only on males and generations did not overlap.

The model, used for simulation of a record for the i^{th} animal was

$$y_i = \mu + a_i + e_i$$

where y_i is the phenotypic value of the i-th animal, a_i is its additive genetic value, and e_i is a residual random value for possible nonadditive and environmental effects. The parameter μ is the phenotypic mean of generation 1. Random values for e_i were drawn from a normal distribution with mean zero and variance 10. The additive genetic variance before selection (σ_a^2) was 10 and initial heritability was .5. For the base animals

(generation 0) genetic values are drawn from N[0, 10]. Genetic values for animals of later generations were simulated as,

$$a_i = \tfrac{1}{2}a_{s_i} + \tfrac{1}{2}a_{d_i} + \phi_i \qquad [1]$$

where a_{s_i} and a_{d_i} are genetic values of sire and dam of individual i. The value ϕ_i results from Mendelian sampling, which is independent of a_s and a_d (Bulmer, 1971). The coefficient of inbreeding as defined by Falconer (1989) for the i-th animal is F_i. The variance of a_i can be presented as

$$\begin{aligned} var(a_i) &= (1+F_i)\sigma_a^2 \\ &= \tfrac{1}{4}var(a_s) + \tfrac{1}{4}var(a_d) + \tfrac{1}{2}cov(a_s, a_d) + var(\phi) \end{aligned} \qquad [2]$$

The variance of ϕ can be given as

$$\begin{aligned} var(\phi) &= (1+F_i)\sigma_a^2 - \tfrac{1}{4}var(a_s) - \tfrac{1}{4}var(a_d) - \tfrac{1}{2}cov(a_s, a_d) \\ &= [(1+F_i) - \tfrac{1}{4}(1+F_s) - \tfrac{1}{4}(1+F_d) - F_i]\sigma_a^2 \\ &= \tfrac{1}{2}[1 - \tfrac{1}{2}(F_s+F_d)]\sigma_a^2 \end{aligned} \qquad [3]$$

where F_s and F_d are inbreeding coefficients for sire and dam, respectively. Inbreeding coefficients were computed using Quaas' (1976) algorithm. The residual genetic value for each animal (ϕ) was drawn from a normal distribution with mean 0 and variance according to [3]. The number of replicates per population depended on the size and number of generations (=g) simulated. Per sample, there was one record available for each of 2*m*g animals and the (s+m) unselected and unrelated base animals were identified as parents without records (generation 0).

Analyses of data sets

To study the effect of omitting data from ancestral generations, additive genetic variance was estimated from data sets that differed in number of generations with records known. We used the complete relationship matrix, i.e. all relations were known since the start of selection, to account for inbreeding and covariances between animals.

In another set of analyses we assumed data known for a limited number

of generations after animals had been selected for several generations. Using this data set we compared models that differed in amount of covariances known between animals by varying the generation that was assumed to consist of unrelated, unselected and noninbred animals.

Inbreeding and covariances among animals in a given population depend upon the size of the population. We have therefore simulated two population sizes, with 40 and 400 animals per generation, respectively. Parameters for s and m were 5 and 20 for the small population and were 20 and 200 for the large population, respectively

The true genetic variance in generation t was defined as

$$\sigma^2_{a_t} = {}_{n=1}\tfrac{1}{} \; (\mathbf{a}_t'\mathbf{a}_t - n\bar{a}_t^2) \qquad [4]$$

where $\mathbf{a}_t$ is a vector with true breeding values for n animals in generation t. The variance $\sigma^2_{a_t}$ is expected to be smaller than σ^2_a since [4] does not adjust for inbreeding and gametic disequilibrium (Bulmer, 1971). Furthermore, [4] does not adjust for covariances between animals, which particularly occur in small populations.

In a balanced hierarchical design for n animals having s sires and d dams per sire (sires and dams unrelated), each dam having p progeny, the expectation of [4] is equal to

$$\tfrac{1}{4}\left(\frac{n-d*p}{n-1}\right) \text{var}(a_s) + \tfrac{1}{4}\left(\frac{n-p}{n-1}\right) \text{var}(a_d) + \text{var}(\phi) \qquad [5]$$

Family structures in later generations are more complicated due to covariances between sires and dams. We can write [4] more generally as ${}_{n=1}\tfrac{1}{}\; \mathbf{a}'(\mathbf{I}-\tfrac{1}{n}\mathbf{J})\mathbf{a} = {}_{n=1}\tfrac{1}{}\; \mathbf{a}'\mathbf{Q}\mathbf{a}$, with J being an n by n matrix with all elements equal to 1, and I is the identity matrix of order n. With no selection, the expectation of [4] is

$$E(\sigma^2_{a_t}) = {}_{n=1}\tfrac{1}{}\; \text{tr}(\mathbf{Q}\mathbf{A}_t)\sigma^2_a \qquad [6]$$

with $\mathbf{A}_t$ being the matrix with additive genetic relationships between animals in generation t. Sorensen and Kennedy (1984b) used $\sigma^2_{a_t} / \sigma^2_a(1-F_t)$ to indicate the reduction of genetic variance due to gametic disequilibrium. F_t

represented the average inbreeding coefficient in generation t. This ratio, however, does not adjust for an additional reduction due to covariances among animals. Therefore, we used $\sigma^2_{a_t}/_{n-1}\text{tr}(QA_t)\sigma^2_a$ as a measure of disequilibrium due to selection.

Estimation of Genetic Variance with Restricted Maximum Likelihood

The following model was used for analysis of the data:

$$y = Xb + Za + e \qquad [7]$$

with

$$E(y) = Xb$$
$$\text{var}(a) = A\sigma^2_a$$
$$\text{var}(y) = V = V_1\sigma^2_a + V_0\sigma^2_e$$
$$= ZAZ'\sigma^2_a + I_n\sigma^2_e$$

Mixed model equations after absorbing fixed effects are

$$[Z'MZ + \alpha A^{-1}]\,\hat{a} = Z'My \qquad [8]$$

where $M = I - X(X'X)^{-}X'$ and $\alpha = \sigma^2_e/\sigma^2_a$

We want to maximize the likelihood of the parameter vector r $(=[\sigma^2_a\ \sigma^2_e])$ in the space of error contrasts, hence maximize $(r|K'y)$, where $K'K = M$.

The log likelihood function of $K'y$ can be written as (Searle, 1979; Smith and Graser, 1986):

$$f(K'y) = \text{constant} + [N - \text{rank}(C)]\log\sigma^2_e + q\log\sigma^2_a + \log|C| + y'Py/\sigma^2_e, \qquad [9]$$

where C is the full rank coefficient matrix of the mixed model equations before absorption of the fixed effects, q is equal to the number of random animal effects, and $P = V^{-1} - V^{-1}X(X'V^{-1}X)^{-}X'V^{-1}$.

Procedures to determine the maximum of [9] have been presented by Graser et al. (1987) and Meyer (1989).

RESULTS

Reduction of genetic variance due to selection

To demonstrate the changes in genetic variance after selection in a small population, we show the mean variance at each generation averaged over 1000 replicates. Table 1 shows the results from random mating and

TABLE 1. Means ($\bar{a}_t$) and variances ($\sigma^2_{a_t}$) of true additive genetic values, expected variance, average inbreeding coefficient (F_t) and gametic disequilibrium (DE) averaged over 1000 replicates for 5 generations of random mating [a].

generation	$\bar{a}_t$	$\sigma^2_{a_t}$	$\frac{1}{n}\sum_{i=1}^{n} tr(QA_t)\sigma^2_a$ [b]	F_t	DE[c]
0	-.01	10.03 (2.92)[d]	10.00	.000	1.00
1	-.02	9.66 (2.65)	9.49	.000	1.02
2	-.04	9.20 (2.69)	9.17	.016	1.01
3	-.01	8.91 (2.53)	8.93	.037	1.00
4	-.03	8.71 (2.57)	8.69	.059	1.00
5	-.01	8.46 (2.44)	8.47	.083	1.00

[a] initial genetic variance was 10 and each generation contained 40 animals
[b] expected variance after adjusting for covariances and inbreeding
[c] disequilibrium computed as $\sigma^2_{a_t} / \frac{1}{n}\sum_{i=1}^{n} tr(QA_t)\sigma^2_a$
[d] empirical standard deviation

TABLE 2. Means ($\bar{a}_t$) and variances ($\sigma^2_{a_t}$) of true additive genetic values, expected variance, average inbreeding coefficient (F_t) and selection disequilibrium (DE) averaged over 1000 replicates for 10 generations of selection [a] [b].

generation	$\bar{a}_t$	$\sigma^2_{a_t}$	$\frac{1}{n}\sum_{i=1}^{n} tr(QA_t)\sigma^2_a$ [c]	F_t	DE[d]
0	.05	9.96 (2.83)[e]	10.00	.000	1.00
1	.06	9.37 (2.71)	9.49	.000	.99
2	1.31	8.47 (2.34)	9.14	.017	.93
3	2.45	8.05 (2.27)	8.85	.040	.91
4	3.55	7.83 (2.07)	8.64	.065	.91
5	4.65	7.59 (2.05)	8.39	.089	.90
6	5.76	7.32 (1.79)	8.17	.113	.90
7	6.81	7.10 (1.79)	7.98	.137	.89
8	7.86	6.97 (1.84)	7.76	.159	.90
9	8.86	6.80 (1.83)	7.55	.181	.90
10	9.84	6.72 (1.86)	7.37	.203	.91

[a] initial genetic variance was 10 and each generation contained 40 animals
[b] five out of twenty males were selected each generation
[c] expected variance after adjusting for covariances and inbreeding
[d] disequilibrium computed as $\sigma^2_{a_t} / \frac{1}{n}\sum_{i=1}^{n} tr(QA_t)\sigma^2_a$
[e] empirical standard deviation from 1000 replicates

Table 2 shows these from selection in a population of size 40 (=2m).

Mean additive genetic values were close to zero for all generations with random mating. The additive genetic variance declined due to the establishment of covariances between animals and an increase of the average inbreeding coefficient (F_t). The expected genetic variance according to [6] has been adjusted for both covariances and inbreeding. Deviations of DE from unity represented gametic disequilibrium. In case of no selection, this deviation is small and generated by chance only. Gametic equilibrium (or selection disequilibrium) has also been referred to as linkage disequilibrium, which we find a confusing term since loci are assumed to be unlinked.

Table 2 shows the change of means and variance of additive genetic values for 5 generations with selection in the small population. There was a clear response to selection and the reduction of σ_a^2 was significant. As expected, the average inbreeding was somewhat higher with selection. The genetic variance decreased more compared to the situation with random mating. An additional decline of genetic variance was due to the establishment of disequilibrium. After some generations of selection, the coefficient for disequilibrium (DE) stabilized at a constant value because new disequilibrium and recombination offset each other.

For infinite population size, the reduction of variance after truncation selection is equal to k= i(i-x) where x is the truncation point and i is the intensity of selection. Becker (1975) gives i= 1.27 and k= .759 for a selected portion of 25%. Selection intensities for small sample sizes can be derived from order statistics, e.g. i=1.2145 when selecting five out of twenty (Beyer, 1968; Becker, 1975). The effective selection intensity is further reduced when selection is on correlated variables. The first generation of our population consisted of five half-sib families and twenty full sib families with intraclass correlation equal to $\frac{1}{4}h^2$ and $\frac{1}{2}h^2$, respectively. From the selection response we can derive the obtained selection intensity to be equal to 1.12. This is in accordance with Table 1 from Hill (1977).

We used order statistics to compute the variance among a selected group as well. The variance among the five highest ranking out of twenty was 20.7% of the variance before selection, giving k= .795. Apparently, the reduction of variance after selection in small samples is higher compared

to infinite population size. However, selection in small populations involves usually also selection on correlated variables because animals are more related to each other, and this gives a smaller reduction in genetic variance. Since values for DE had been corrected for covariances and inbreeding, we used $\sigma^2_{a_t} = DE_t\sigma^2_{a_0}$ in Bulmers' (1971) formula for an infinite population,

$$\sigma^2_{a_t} = \tfrac{1}{4}(1-kh^2)\sigma^2_{a_{t-1}} + \tfrac{1}{4}\sigma^2_{a_{t-1}} + \tfrac{1}{2}\sigma^2_{a_0},$$

to derive empirically a k-value of about .6 for t= 2 and of about .5 in later generations.

Estimation of genetic variance when data from selected animals are missing

Genetic variance was estimated using different models and data sets. Average true genetic variances are given for one thousand replications for the small population (2m= 40) in Table 2 and for twenty replications for the large population (2m= 400) in Table 3. Selection response and effect of gametic disequilibrium on genetic variance was higher in the large population, due to the higher selection intensity. However, reduction of genetic variance was smaller because of less inbreeding in this population (Table 3).

TABLE 3. Means ($\bar{a}_t$) and variances ($\sigma^2_{a_t}$) of true additive genetic values, expected variance, average inbreeding coefficient (F_t) and selection disequilibrium (DE) averaged over 20 replicates for 5 generations of selection [a] [b].

generation	$\bar{a}_t$	$\sigma^2_{a_t}$	$n^{-1}_{-1}tr(QA_t)\sigma^2_a$ [c]	F_t	DE[d]
1	.05	9.82 (1.19)[e]	9.88	.000	.99
2	1.90	8.92 (.78)	9.79	.005	.91
3	3.72	8.55 (.76)	9.72	.013	.88
4	5.47	8.19 (.88)	9.61	.020	.85
5	7.14	8.12 (.79)	9.54	.027	.85

[a] initial genetic variance was 10 and each generation contained 400 animals
[b] Twenty out of 200 males were selected each generation.
[c] expected variance after adjusting for covariances and inbreeding
[d] disequilibrium computed as $\sigma^2_{a_t}/n^{-1}_{-1}tr(QA_t)\sigma^2_a$
[e] empirical standard deviation from 20 replicates

TABLE 4. Estimated genetic variance ($\hat{\sigma}^2_a$) after omitting data from an increasing number of selected generations but including the complete relationship matrix.

data used from	$\hat{\sigma}^2_a$ (small pop[a].)	n	$\hat{\sigma}^2_a$ (large pop[b].)
gen 1-5 (all data)	9.81 (2.78)[c]	1000	10.09 (.81)[d]
gen 2-5	9.49 (3.22)	997	9.90 (.83)
gen 3-5	9.54 (3.83)	980	9.58 (1.17)
gen 4-5	9.60 (4.76)	901	8.93 (1.94)

[a] Five out of twenty males were selected each generation
[b] Twenty out of 200 males were selected each generation
[c] empirical standard deviations from n replicates
[d] empirical standard deviations from 20 replicates

Estimated additive genetic variances by subsequently omitting data from parental generations are given for the two populations in Table 4. Some of the replicates for the small population did not converge to a solution and the estimate for σ^2_a approached 0. Those replicates were omitted for calculating the mean result.

The estimate of additive genetic variance using REML was close to the initial value when all data and the complete relationships matrix were used (Table 4). This agrees with the result of REML accounting for selection using the complete mixed model (Sorensen and Kennedy, 1984b;, Gianola and Fernando, 1986).

Subsequently omitting data in the small population from generations 1, 1 to 2 and 1 to 3, gave estimates for σ^2_a equal to 9.49, 9.54, and 9.60, respectively (Table 4). This suggests that the relationships matrix accounted for most of the selection, even though the analysis did not include records on which selection decisions had been based. The empirical standard deviation of the estimates increased considerably, however, when data were omitted from the analysis. Omitting data from selected generations in the large population had more effect on the mean estimated genetic variance; estimates of σ^2_a decreased more when data from more previous generations were omitted. Decrease of mean estimate was also larger when considered in proportion to the higher coefficient of disequilibrium for the large population.

Table 5 shows the effect of assuming generation 5 as the base population

(unselected, unrelated and non-inbred). Note that in this case the assumed base population had a smaller variance than twice the variance generated by Mendelian Sampling. The genetic variance was estimated omitting data from subsequent generations. Omitting data had little effect on the mean estimate. Similar results were found when generation 6 was assumed to be the base generation. Adding data from more subsequent generations affected estimate of genetic variance in terms of standard error, but not in terms of the mean. Notice that with data from generation 6 and 7, and considering generation 6 as the base, an estimate was obtained, which was higher than $\sigma^2_{a_6}$. This is not in agreement with the result of Sorensen and Kennedy (1984b).

TABLE 5. Estimated genetic variance ($\hat{\sigma}^2_a$) after omitting data from an increasing number of selected generations for a given set of covariances between animals [a].

data used from	base generation[b]	$\hat{\sigma}^2_a$	n
gen 9-10	5	8.56 (4.29)[c]	855
gen 8-10	5	8.15 (3.42)	973
gen 7-10	5	8.20 (2.84)	990
gen 6-10	5	8.30 (2.42)	999
gen 6-7	6	7.81 (3.85)	918
gen 6-8	6	7.78 (3.21)	982
gen 6-9	6	7.80 (2.61)	994
gen 6-10	6	7.92 (2.31)	1000

[a] Five out of twenty males were selected each generation
[b] Generation assumed to consist of unrelated, non-inbred and unselected animals
[c] empirical standard deviation from n replicates

Including additional relationships between base animals

Usually in data analysis, an arbitrary generation is treated as consisting of unselected, unrelated and noninbred base animals. Results from Table 5 suggest that bias from prior selection could be (partly) removed when relationships between these assumed base animals are included in the model. The effect of including relationships established in earlier

generations is shown in Tables 6 and 7 for the small and the large population, respectively. Data were used from animals of two generations and relationships known from an increasing number of previous generations were included.

Estimates of genetic variance were higher than the true variance of the

TABLE 6. Estimated genetic variance ($\hat{\sigma}^2_a$) using data from two selected generations and including the relationships generated from a various number of generations [a].

data from	base gen.[b]	$\hat{\sigma}^2_a$	bias(%)[c]	n
gen 4-5	3	8.71 (4.33)[d]	-7.3	924
gen 4-5	2	9.05 (4.56)	-5.9	892
gen 4-5	1	9.34 (4.68)	-5.6	896
gen 4-5	0	9.60 (4.76)	-4.6	901
gen 9-10	8	7.76 (3.97)	-7.4	903
gen 9-10	7	8.07 (4.08)	-7.1	864
gen 9-10	6	8.33 (4.19)	-7.5	859
gen 9-10	5	8.56 (4.29)	-6.3	855

[a] Five out of twenty males were selected each generation
[b] Generation assumed to consist of unrelated, non-inbred and unselected animals
[c] reference values computed as $\hat{\sigma}^2_a$ for the case of no selection
[d] empirical standard deviation from n replicates

TABLE 7. Estimated genetic variance ($\hat{\sigma}^2_a$) using data from generations 4 and 5 and including the relationships generated from an increasing number of generations [a].

base generation[b]	$\hat{\sigma}^2_a$	bias(%)[c]
3	8.58 (1.87)[d]	-12.3
2	8.75 (1.91)	-10.6
1	8.87 (1.93)	-9.8
0	8.93 (1.94)	-9.8

[a] Twenty out of 200 males were selected each generation
[b] Generation assumed to consist of unrelated, non-inbred and unselected animals
[c] reference values computed as $\hat{\sigma}^2_a$ for the case of no selection.
[d] empirical standard deviations from 20 replicates

assumed base population. Estimates were compared with values, that werecomputed using the same model, but with data obtained for the case of no selection of males. Results indicate that adding additional relationships between base animals accounted for covariances and inbreeding. Table 6 shows that estimates generally are only slightly less biased by disequilibrium from selection, when additional relationships between the assumed base animals are included.

In the large population there was less accumulation of inbreeding and the effect of including more relationships was smaller (Table 7). However, the bias from gametic disequilibrium was larger than in the small population. This bias was only slightly reduced by including additional relationships between base animals.

DISCUSSION

We considered selection for a metric trait, which is equally affected by many unlinked loci. The change of genetic variance due to changes in gene frequencies is small and can be ignored in such an infinitesimal model.

Changes due to inbreeding are irreversible. However, the coefficient of inbreeding is only a relative measure, i.e. a base population is a population that has by definition an average inbreeding coefficient of zero (Falconer, 1989). In practice, choosing a generation is arbitrary and accounting for inbreeding in previous generations does not provide estimates that are better able to predict future genetic gain because newly generated Mendelian Sampling variance will consistently be reduced through inbreeding.

Bulmer (1971) has pointed out that gametic disequilibrium vanishes after selection is ceased. Moreover, the variance that is generated at each generation by recombination is not affected by disequilibrium. Estimates of genetic variance should therefore be corrected for the effect of disequilibrium.

Robertson (1977) indicated that estimates of genetic variance based on half and full-sib analysis will be biased due to a reduction of genetic variance among selected parents. Parent-offspring regression is not

affected by this type of selection (Hill and Thompson, 1977). A maximum likelihood estimator, like REML, combines information from contrasts within a generation, with parent offspring covariances and with information from contrasts between families (i.e. between base parents) (Thompson, 1977). Little is known about the weighing of information from these different sources. It was empirically shown in this study that the mean estimate of additive genetic variance was primarily determined by the generation that was assumed to be the base generation (Tables 5 to 7). This is supported by results from Table 5, which show no change in mean estimate when information on more subsequent generations is available to estimate variance from Mendelian sampling. However, this result needs theoretical verification.

REML using all relationships and all data, has been shown to account for all selection. The argument, that accounting for selection is possible only when all data is used (Gianola and Fernando, 1986), has been demonstrated for sequential selection within a generation, or with selection on a correlated trait (Meyer and Thompson, 1984). However, some of the genetic variance lost through selection is regained at each generation through gene segregation during meiosis (Mendelian sampling). Contributions from selected parents and those due to the Mendelian recombination can be treated as random and independently acting terms. Thompson (1977) has shown the contributions of parental generations acting on the total additive genetic variance.

Let us consider to have data on generation 2 and at least one later generation. If we assume the covariance between sires and dams can be ignored, the variance in generation 2 can be written as

$$\mathrm{var}(a_2) = \tfrac{1}{4}\mathrm{var}(a_{s_1}) + \tfrac{1}{4}\mathrm{var}(a_{d_1}) + \mathrm{var}(\phi_2) \qquad [10]$$

The variance in this generation is reduced due to selection of their sires only. Other terms are not affected by selection. Selection is on records from sires in generation 1, which is $y_{s_1} = \mu + a_{s_1} + e_1 = \mu + \tfrac{1}{2}a_{s_0} + \tfrac{1}{2}a_{d_0} + \phi_{s_1} + e_1$ (subscripts refer to the generation number). Assuming normality, the variance of the genetic variables after selection on y_{s_1} is $H_s = H - BP^{-1}(I - P^{-1}P_s)B'$ (Pearson, 1903), with

$$H = \text{var}\begin{bmatrix} \tfrac{1}{2}a_{s_0} \\ \tfrac{1}{2}a_{d_0} \\ \phi_{s_1} \end{bmatrix} = \begin{bmatrix} \tfrac{1}{4} & 0 & 0 \\ 0 & \tfrac{1}{4} & 0 \\ 0 & 0 & \tfrac{1}{2} \end{bmatrix}\sigma_a^2, \quad B = \text{cov}\begin{bmatrix} & \tfrac{1}{2}a_{s_0} \\ y_{s_1} & , \tfrac{1}{2}a_{d_0} \\ & \phi_{s_1} \end{bmatrix} = \begin{bmatrix} \tfrac{1}{4} \\ \tfrac{1}{4} \\ \tfrac{1}{2} \end{bmatrix}\sigma_a^2$$

$P = \text{var}(y_1)$ and $P_s = (1-k)P$. Then

$$H_s = \begin{bmatrix} \tfrac{1}{4}-\tfrac{1}{16}kh^2 & -\tfrac{1}{16}kh^2 & -\tfrac{1}{8}kh^2 \\ -\tfrac{1}{16}kh^2 & \tfrac{1}{4}-\tfrac{1}{16}kh^2 & -\tfrac{1}{8}kh^2 \\ -\tfrac{1}{8}kh^2 & -\tfrac{1}{8}kh^2 & \tfrac{1}{2}-\tfrac{1}{4}kh^2 \end{bmatrix}\sigma_a^2 \qquad [11]$$

Note that the variance of a_{s_1} after selection is $1'H_s1 = (1-kh^2)\sigma_a^2$, which is equal to what was predicted by Bulmer (1971). From [11] it can be seen that the Bulmer effect, i.e. the reduction of genetic variance after selection, is due to the reduction of variance of each of the components of the total genetic variance, and, moreover, due to negative covariances between these components. In genetic terms this has been referred to as ' negative covariances between loci' (Bulmer, 1971).

Eventhough the variance among the sires of generation 2 is reduced by $kh^2\sigma_a^2$, the observed bias in estimation of genetic variance, using data from generation 2, was smaller than $-\frac{1}{4}kh^2\sigma_a^2$. Estimated values were 9.64 and 9.93 for the small and for the large population, respectively. A mixed model, explaining the breeding values in terms of genetic values of parental generations, reduces bias from the Bulmer effect, since it does not include covariances between these parental contributions. Note that the sum of the diagonal elements of [11] is equal to $(1-.375kh^2)\sigma_a^2$, rather than $1-kh^2\sigma_a^2$.

In the large population the bias from selection was considerably greater when data from more generations were omitted, even when the complete relationship matrix was used. After repeated cycles of selection, expression [11] consists of more ancestral contributions, which are redundantly affected by selection, e.g. the sum of all diagonal elements after

selection on y_2 would be $(1-.45kh^2)$. Bias from disequilibrium was small in the small population, and the difference with bias in the large population was not in proportion to the difference in coefficient for disequilibrium. Apparently, the effect of selection on different genetic components is dependent on the population size, e.g. in the small population, there is a greater chance of selecting related animals (Robertson, 1961), and the variance among Mendelian sampling terms is relatively not as much reduced. This tendency towards 'family selection' is greater when selection was on estimated breeding values using an animal model.

IMPLICATIONS

Including additional relationships between the assumed base animals accounted for covariances and inbreeding, and partly for bias from gametic disequilibrium as well. Relationships accounted for more bias from gametic disequilibrium when selection had been practiced for just a few generations, or when the population size was small. The mean estimate of additive genetic variance was primarily determined by which generation was assumed as unrelated, noninbred and unselected and having more subsequent generations with data did not affect the mean estimates. Variance components used for animal evaluation should be estimated with a model that coincides with the same base population, possibly even with different values for the components of variance between selected base animals, and variance within families. Estimates of genetic variance to determine expected genetic progress are less biased by covariances and gametic disequilibrium, when more relationships are included in the model.

ACKNOWLEDGEMENTS

The authors wish to thank Brian Kennedy and Dick Quaas for valuable suggestions and Karin Meyer for using her DFREML program.

LITERATURE CITED

Becker, W.A. 1975. Manual of Quantitative Genetics. Washington State University Press, Washington.

Beyer, W.H. (ed). 1968. Handbook of tables for probability and statistics. 2nd edition. Chem. Rubber Co. Cleveland Ohio.

Bulmer, M.G. 1971. The effect of selection on genetic variability. Ameri. Nat. 105:201.

Falconer, D.A. 1989. Introduction to quantitative genetics (3rd Ed.). Longman, London.

Gianola, D. and R.L. Fernando. 1986. Bayesian methods in animal breeding theory. J. Anim. Sci. 63:217.

Graser, H.U., S.P. Smith and B. Tier. 1987. A derivative free approach for estimating variance components in animal models by restricted maximum likelihood. J. Anim. Sci. 64:1362.

Henderson, C.R. 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics 31:423.

Hill, W.G. 1977. Order statistics of correlated variables and implications in genetic selection programmes. II. Response to selection. Biometrics 33:703.

Hill, W.G. and R. Thompson. 1977. Design of experiments to estimate offspring-parent regression using selected parents. Anim. Prod.24:163.

Meyer, K. and R. Thompson. 1984. Bias in variance and covariance component estimators due to selection on a correlated trait. Z. Tierz. Züchtgsbiol. 101:33-50.

Meyer, K. 1989. Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative free algorithm. Génét. Sél. Evol. 21:317-340.

Quaas, R.L. 1976. Computing the diagonal elements and inverse of a large numerator relationship matrix. Biometrics 32:949.

Robertson, A. 1961. Inbreeding in artificial selection programmes. Genet. Res. 2:189.

Robertson, A. 1977. The effect of selection on the estimation of genetic parameters. Z. Tierz. Züchtgsbiol. 94:131.

Pearson, K. 1903. Mathematical contributions to the theory of evolution. XI. On the influence of natural selection on the variability and

correlation of organs. Phil. Trans. R. Soc. London (A) 200:1.

Searle, S. R. 1979. Notes on variance component estimation: A detailed account of Maximum Likelihood and kindred methodology. Paper BU-673M, Biometrics Unit, Cornell Univ., Ithaca, N.Y.

Smith, S.P. and H.U. Graser. 1986. Estimating variance components in a class of mixed models by restricted maximum likelihood. J. Dairy Sci. 69:1156.

Sorensen D.A., and B.W. Kennedy. 1984a. Estimation of response to selection using least squares and mixed model methodology. J. Anim. Sci.59:1097.

Sorensen D.A., and B.W. Kennedy. 1984b. Estimation of genetic variances from unselected and selected populations. J. Anim. Sci.59:1213.

Thompson, R. 1977. The estimation of heritability with unbalanced data. II. Data available on more than two generations. Biometrics 33:497.

Chapter 5

RESTRICTED MAXIMUM LIKELIHOOD ESTIMATION OF ADDITIVE GENETIC VARIANCE IN SELECTED POPULATIONS USING A CONDITIONAL MODEL

J.H.J VAN DER WERF
Department of Animal Breeding
Wageningen Agricultural University
PO Box 338, 6700 AH Wageningen
The Netherlands

To be submitted to: Journal of Animal Science

ABSTRACT

A method to estimate genetic parameters conditional to selection occuring before formation of the base population was investigated. The method assumes base parents as fixed and a conditional variance is based upon the Mendelian sampling of gametes from the base parents. In a simulation study, s sires were selected and each was mated to m/s females to create 2m animals for the next generation. Selection was for five generations but only animals of generations 4 and 5 were assumed to have performance records and parents known. Simulated values for additive genetic and residual variance were 10. When s=20 and m=200, estimated genetic variance was 8.58 when base animals were assumed random, and it was 6.03 when they were fixed. Residual variance was overestimated in the latter case. When males of generation 4 were not selected to have progeny, estimated genetic variance was 9.91. It was concluded that estimates for genetic parameters with the conditional model were not biased by selection of base animals. However, the model introduced a new bias when descendants of base animals were selected to have progeny.

INTRODUCTION

Directional selection decreases the additive genetic variance due to the establishment of covariances between animals, inbreeding and gametic disequilibrium (Sorensen and Kennedy, 1984; Van der Werf and De Boer, 1990). A mixed model accounts for the change of variance when additive genetic relationships are known, tieing animals back to a certain base population that consists of unrelated, unselected and noninbred animals (Sorensen and Kennedy, 1984). However, data used for estimation of genetic parameters arise generally from recording during a limited time period and a group of animals with unknown parents is treated as the base population. Therefore, base population animals may be selected and estimates of genetic variance, ignoring the history of data and pedigree, may be influenced by selection of ancestors (Van der Werf and De Boer, 1990).

Henderson (1985, 1988) proposed to apply a 'conditional relationships matrix' to account for selected base populations. Graser et al (1987) have

suggested to treat base animals as fixed and to estimate genetic variance independent from the variance among selected base animals. The latter procedure is in effect equivalent to that of Henderson, but the approach is more tractable from a genetic point of view. Graser et al. (1987) divided the genetic variance into a part coming from variance between selected base animals and a part due to Mendelian Sampling. The method uses the fact that variance from Mendelian Sampling is assumed to be not affected by prior selection (Bulmer, 1971).

The statistical and genetic properties of the conditional model proposed by Graser et al. (1987) have not been extensively investigated. The authors used actual data on beef cattle, but they did not find significantly different results compared to a method treating base animals as random. A comparison based on simulated data was performed in this study because true variances as well as selection differentials are known. The method treating base animals as fixed was also applied to first lactation production records of Dutch Black and White dairy cows. Heritabilities from previous analysis of this data were suspected to be biased by selection of base sires, when a model with random base animals were used (Van der Werf and De Boer, 1989). A comparison with a conditional model was used to test this hypothesis.

MATERIAL

Simulated data

A Monte Carlo simulation study was carried out as described by Van der Werf and De Boer (1990), considering a trait that was affected by a large number of loci. A base generation (generation 0) of s males and m females, all assumed unrelated and unselected, was mated at random to produce m males and m females. In each of the subsequent generations, the s phenotypically best males were selected and each of them was mated to m/s females. Animals were only mated within generation and the only fixed effect simulated was the mean. Two random effects were simulated; an additive genetic and a residual effect, both distributed as N[0,10]. Additive genetic values for animals not belonging to the base generation were;

$$a_i = \tfrac{1}{2}a_{s_i} + \tfrac{1}{2}a_{d_i} + \phi_i \qquad [1]$$

where s_i is the sire, d_i is the dam, and ϕ_i can be seen as a random variable representing Mendelian sampling,

with $\phi_i \sim N[0, \text{var}(\phi_i)]$

and $\text{var}(\phi_i) = \tfrac{1}{2}[1-\tfrac{1}{2}(F_s + F_d)]\sigma_a^2$

with F_s and F_d being inbreeding coefficients of sire and dam, respectively.

Two population sizes were simulated to vary inbreeding and selection intensities. Parameters were s= 5 and m= 20 for population A, and s= 20 and m= 200 for population B. The total number of generations simulated was 10 for A and 5 for B.

METHODS

Accounting for selected base animals

Consider a data vector

$$y = X\beta + Za + e, \qquad [2]$$

with ß,a and e being vectors of fixed effects, breeding values and residual effects, respectively. The matrices X and Z are design matrices. Breeding values can be expressed according to [1] in terms of contributions from parents, and a random term due to Mendelian sampling. Graser et al. (1987) partitioned a vector with breeding values in a_b for base animals (which they assumed unrelated) and a_r as a vector of random additive values for descendents of the base animals. In matrix notation, [1] is expressed as,

$$\begin{bmatrix} a_b \\ a_r \end{bmatrix} = \begin{bmatrix} I & 0 \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} a_b \\ a_r \end{bmatrix} + \begin{bmatrix} 0 \\ \phi \end{bmatrix} \qquad [3]$$

where P_{21} and P_{22} are submatrices with each row having one half in the colomn for each of its two parents. The mixed model can be written as

$$y = X\beta + Z_1 a_b + Z_2 a_r + e \qquad [4]$$

with

$$a_r = (I-P_{22})^{-1}P_{21}a_b + (I-P_{22})^{-1}\phi, \qquad [5]$$

Note that $(I-P_{22})^{-1}$ exists because the diagonals of P_{22} are all zero and each row has only two non-zero elements. Graser et al. (1987) have written [4] as

$$y = X\beta + [Z_1 + Z_2Q]a_b + Z_2 s^* + e. \qquad [6]$$

with $Q = (I-P_{22})^{-1}P_{21}$

$s^* = (I-P_{22})^{-1}\phi$.

Graser et al. (1987) assumed that if additive genetic values of the base animals (a_b) are treated as fixed the variances can be written as:

$$\text{var}(y) = Z_2\text{var}(s^*)Z_2 + \text{var}(e)$$

and $$\text{var}(s^*) = (I-P_{22})^{-1}D((I-P_{22})')^{-1}\sigma_a^2 - G\sigma_a^2 \qquad [7]$$

with $D\sigma_a^2 =$ var(ϕ), a diagonal matrix.

The matrix Q is what Quaas (1988) called a 'base ancestor- descendant' matrix and s^* can be seen as genetic values for animals with records after correction for the breeding values of their parents.

Graser et al. (1987) have given suggestions for a 'derivative free' maximization of the likelihood function for [6]. Quaas (1984) has proposed an equivalent model to [2]. This model has been often used in algorithms for variance component estimation that utilize a tridiagonalization of the coefficient matrix (Meyer, 1986). It is presented here to show how the equations for a conditional model can be set up for those algoritms.

Let A be the matrix containing the additive genetic relationships between animals. The matrix A has been written as LL' (Quaas 1976, 1984). we can also partition L in unrelated base animals, and their offspring;

$$L = \begin{bmatrix} I & 0 \\ (I-P_{22})^{-1}P_{21} & (I-P_{22})^{-1}D^{\frac{1}{2}} \end{bmatrix} \qquad [8]$$

where the diagonals of $D^{\frac{1}{2}}$ contain $d_i^{\frac{1}{2}}$ and in our notation this is equal to;

$$L = \begin{bmatrix} I & 0 \\ Q & G^{\frac{1}{2}} \end{bmatrix} \qquad [9]$$

Analogously to Quaas (1984) we can write [6] as

$$y = X\beta + [Z_1 + Z_2Q]a_b + Z_2G^{\frac{1}{2}}\theta + e.$$

which we write as

$$y = X\beta + Wa_b + Z^*\theta + e. \qquad [10]$$

where $\theta = G^{-\frac{1}{2}}s^* = D^{-\frac{1}{2}}\phi$. Notice that $var(\theta) = I_r\sigma_a^2$, where I_r is an identity matrix of order r, and θ represents a vector with Mendelian sampling variables, standardized to having a variance of σ_a^2. The variance of y is

$$V = Wvar(a_b)W' + Z^*var(\theta)Z^{*\prime} + var(e)$$

When base animals are unselected, unrelated and noninbred, $var(a_b) = I_b\sigma_a^2$ where σ_a^2 is the additive genetic variance in absence of selection. It can easily be verified that in case of no selection $V = (WW' + Z^*Z^{*\prime})\sigma_a^2 + I_n\sigma_e^2 = ZAZ'\sigma_a^2 + I_n\sigma_e^2$, what is generally assumed. However, when base animals are selected, $var(a_b) = \delta I_b\sigma_a^2$ with $\delta<1$ whereas $var(\theta) = I_r\sigma_a^2$ is not affected by selection. The genetic variance unaffected by selection can be computed by REML as the variance of random breeding values, conditional upon effects of selected parents the base generation. Now, since $cov(a_b,y) = \delta W\sigma_a^2$ and $var(a_b) = \delta I_b\sigma_a^2$, the variance of y conditional upon a_b is

$$\begin{aligned} var(y|a_b) &= V - WW'\delta\sigma_a^2 \\ &= Z^*Z^{*\prime}\sigma_a^2 + I\sigma_e^2 \end{aligned}$$

Analogously to Graser et al. (1987), when a_b is assumed fixed, the variance of y can be written as;

$$var(y) = var(s^*) + var(e)$$

or $\quad var(y) = Z^*Z^{*\prime}\sigma_a^2 + I\sigma_e^2$

The mixed model equations for [10] assuming a_b fixed, are

$$\begin{bmatrix} X'X & X'W & X'Z^* \\ W'X & W'W & W'Z^* \\ Z^{*\prime}X & Z^{*\prime}W & Z^{*\prime}Z^*+\alpha I \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{a}_b \\ \hat{\theta} \end{bmatrix} = \begin{bmatrix} X'y \\ W'y \\ Z^{*\prime}y \end{bmatrix} \qquad [11]$$

Mixed model equations according to [11] are exactly the equivalent mixed model equations as proposed by Quaas(1984) except for not adding α $(= \sigma_e^2/\sigma_a^2)$ to the diagonal of the base animals in [11]. Equations [11] were applied to a sire model in this study to analyze milk production data. In the sire model, the vector a was replaced by a vector of sire effects, and σ_a^2 was replaced by the sire variance.

Equations [11] are not advantageous for variance component estimation with an animal model using a so called 'derivative free algorithm' (Meyer, 1989). In that case tridiagonalization is not used, and for very large pedigrees it is not easy to set up W'W. Modified equations can be derived from [11], as indicated by Graser et al. (1987). Using

$$B = \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & -G^{-\frac{1}{2}}Q & G^{-\frac{1}{2}} \end{bmatrix}$$

[11] can be modified by premultiplying both sides of the equations by B and inserting $B'(B')^{-1}$ between the coefficient matrix and the solution vector. The equations obtained are,

$$\begin{bmatrix} X'X & X'Z_1 & X'Z_2 \\ Z_1'X & Z_1'Z_1 + \alpha Q'G^{-1}Q & -\alpha Q'G^{-1} \\ Z_2'X & -\alpha G^{-1}Q & Z_2'Z_2+\alpha G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{a}_b \\ \hat{a}_r \end{bmatrix} = \begin{bmatrix} X'y \\ Z_1'y \\ Z_2'y \end{bmatrix} \qquad [12]$$

which are mixed model equations for [4], treating a_b as fixed. Equations [12] can be rapidly constructed using the well known rules to set up the inverse of the relationships matrix (Quaas, 1976). This is easily seen by showing that

$$L^{-1} = \begin{bmatrix} I & 0 \\ -G^{-\frac{1}{2}}Q & G^{-\frac{1}{2}} \end{bmatrix}$$

and

$$A^{-1} = L^{-t}L^{-1} = \begin{bmatrix} I+Q'G^{-1}Q & -Q'G^{-1} \\ -G^{-1}Q & G^{-1} \end{bmatrix}$$

Analysis of simulated data

Mixed linear models were used for data analysis and Restricted Maximum Likelihood was used to estimate components of variance. To estimate additive genetic variance, records from selected generations were used and pedigree information was assumed known back to one previous generation only. An animal model was applied to the simulated data set and the maximum likelihood was determined with a derivative free approach (Graser et al., 1987; Meyer, 1989). The likelihood equation evaluated was

$$L = -\tfrac{1}{2}[(N-\text{rank}(C))\log\hat{\sigma}_e^2 + \log|C| + q\log\hat{\sigma}_a^2 + y'Py/\hat{\sigma}_e^2] \qquad [13]$$

with N being the number of observations, $|C|$ is the determinant of the full rank coefficientmatrix for the mixed model equations according to the applied model, q is the number of random levels in a and $P = V^{-1} - V^{-1}X(X'V^{-1}X)^{-}X'V^{-1}$. The term yP'y is a quadratic to estimate the residual variance and is equal to $y'(y - X\hat{\beta} - Z\hat{a})$.

Data sets were created from animals with selected base parents. We considered data from generations 4 and 5 (or 9 and 10), with pedigree information known back to the previous generation only. An analysis with

all breeding values assumed random was compared with an analysis treating breeding values of the base sires as fixed effects. In the likelihood function [13], C and q are replaced by C^* and q_r, representing the coefficientmatrix of [12] and the number of random levels in a_r, respectively. Because family sizes were small in the population A, dams were treated as unrelated and unselected random base animals to avoid too many fixed effects in the model. In fact, the derivative free algoritm did not converge to a constant value for σ_e^2/σ_a^2 in most of the samples of population A, when all base animals were treated as fixed. In populations B sires, as well as dams were treated as fixed.

RESULTS

In Table 1 the average results from 1000 replicates of simulation of population A are shown. The genetic variance within generations decreased from 10 to 6.67 due to inbreeding and gametic disequilibrium after 10 generations of selection in the small population. Genetic variance was 8.03 at generation 3 and 7.02 at generation 8. Genetic variance at generation t was defined as $(a_t'a_t - n\bar{a}_t)/(n-1)$, where a_t is a n*1 vector with breeeding values of animals in the t^{th} generation. More results of this simulation were described by Van der Werf and De Boer (1990).

The result of estimating genetic variance using only records from generations 4 and 5, and assuming generation 3 as base animals, are shown in Table 2. The estimate of genetic variance ($\hat{\sigma}_a^2$) was 8.71 in the small population (A), when considering base animals as random. Using the same dataset, but treating base sires as fixed, gave an estimate for σ_a^2 of 9.38, which underestimated the true value of 10. Using data from generations 8 and 9 an estimate of 7.76 when generation 8 was considered as a random base, and the estimate was 8.77 when the same animals were treated as fixed.

The derivative free REML algorithm did not converge in about 9% of the replicates for population A. Due to the small sample sizes there was a greater chance of the within family variance being smaller than the between family variance. In all cases, the genetic variance adapted small values (<1) with further iteration. Those replicates were discarded from

TABLE 1. Means ($\bar{a}_t$) and variances ($\sigma^2_{a_t}$) of true additive genetic values, average inbreeding coefficient (F_t) for several generations in a small and a large population [a].

generation	$\bar{a}_t$	$\sigma^2_{a_t}$	F_t
small population[b] (1000 replicates)			
0	.05	9.96 (2.83)[d]	.000
1	.06	9.37 (2.71)	.000
2	1.31	8.47 (2.34)	.017
3	2.45	8.05 (2.27)	.040
4	3.55	7.83 (2.07)	.065
5	4.65	7.59 (2.05)	.089
6	5.76	7.32 (1.79)	.113
7	6.81	7.10 (1.79)	.137
8	7.86	6.97 (1.84)	.159
9	8.86	6.80 (1.83)	.181
10	9.84	6.72 (1.86)	.203
large population[c] (20 replicates)			
1	.05	9.82 (1.19)	.000
2	1.90	8.92 (.78)	.005
3	3.72	8.55 (.76)	.013
4	5.47	8.19 (.88)	.020
5	7.14	8.12 (.79)	.027

[a] initial genetic variance was 10
[b] five out of twenty males were selected each generation
[c] twenty out of 200 males were selected each generation.
[d] empirical standard deviation

TABLE 2. Estimated genetic variance ($\hat{\sigma}^2_a$) using simulated data from selected generations of different populations and using a model treating base animals as random or as fixed [a].

Pop	data (gen)	base generation	n	$\hat{\sigma}^2_a$	$\hat{\sigma}^2_e$
A[b]	4-5	3 random	924	8.71 (4.33)[e]	10.17
A	4-5	3 sires fixed	872	9.38 (4.95)	10.04
A	9-10	8 random	903	7.76 (3.97)	10.00
A	9-10	8 sires fixed	865	8.77 (4.75)	9.69
B[c]	4-5	3 random	20	8.58 (1.86)	10.44
B	4-5	3 sires fixed	20	9.14 (2.26)	10.19
B	4-5	3 sires +dams fix.	18	6.03 (2.14)	11.77

[a] initial genetic and residual variance was 10
[b] five out of twenty males were selected each generation
[c] twenty out of 200 males were selected each generation
[d] empirical standard deviations from n replicates

determination of the mean estimate. Empirical standard deviations were calculated over valid replicates and appeared to be 15 to 20% higher when sires were treated as fixed.

Estimates were expected to be affected by the average inbreeding coefficient of the base animals, since the model assumes non-inbred base animals. We expected additionally a downward bias from the assumption of dams being random and unselected. Such expected values would therefore be a function of the terms $(1-F_{base})\sigma^2_a$ and $\sigma^2_{a_{dams}}$. Those terms were equal to 9.6 and 8.05, respectively, when animals from generation 3 were the base, and the estimate was 9.38. The same values were 8.41 and 6.97, respectively, when the base is at generation 8, wereas the estimate was 8.77. Hence, the estimate of genetic variance was higher than expected, and the estimate of the residual variance was lower than 10 in the small population using a model with fixed sires.

Table 2 also shows the estimates for the population B with 200 females per generation. Average inbreeding coefficient at generation 3 was 0.012. Genetic variance among animals of generation 3 was 8.53. Estimating genetic variance and only treating sires as fixed resulted in 9.14. The estimate from a model treating both base-sires and -dams as fixed was much lower; 6.03 for population B. The estimate with all base animals fixed was therefore significantly lower than 2 times the Mendelian Sampling variance at generation 4, which was 9.88 for population B. Estimates of residual variance were higher than the simulated value of 10.

To study further the large bias observed in the large population, a data set with records only from the first two generations was analysed. Note that base animals from generation 0 were noninbred and unselected. Results are shown in Table 3, indicating that estimates for genetic variance were also biased when a conditional model and data from generations 1 and 2 was used. Apparently, having a selected base population was not the cause of bias in the conditional model.

We also analysed data for the case of no selection of parents to have progeny. Results from no selection, however, were not biased, indicating that selection of animals with records did cause biased estimates. This was even more clearly demonstrated in a simulation, in which parents with records known were randomly chosen, but those from previous generations had been selected. Analyzing records from generations 4 and 5, and selecting

TABLE 3. Estimated genetic variance ($\hat{\sigma}^2_a$) using simulated data from selected generations with varying heritabilty and a model treating base animals as fixed [a] [b].

base gen.		data (gen)	$\hat{\sigma}^2_a$	$\hat{\sigma}^2_e$
selection				
0	random	1-2	9.97 (1.80)[c]	10.17
0	fixed	1-2	6.51 (3.00)	11.79
3	random	4-5	8.58 (1.87)	10.44
3	fixed	4-5	6.03 (2.14)	11.77
no selection				
0	random	1-2	9.92 (1.76)	10.35
0	fixed	1-2	9.95 (4.16)	10.26
3	random	4-5	9.79 (2.18)	10.01
3	fixed	4-5	10.08 (3.81)	9.83
selection, except in generation 4				
3	fixed	4-5	9.91 (3.77)	9.93

[a] initial genetic variance was 10
[b] twenty out of 200 males were selected each generation
[c] empirical standard deviations from n replicates

parents, except in generation 4, gave an unbiased estimate of genetic and residual variance (Table 3). Hence, the conditional model accounted for previous selection before forming the base generation, but estimates were biased when animals with records were selected.

Bias from a model with fixed base animals reduced, when more generations had data (Table 4). Bias from analysis of a data set with base animals having records was similar to the case when these records were omitted, but emperical standard errors were slightly lower (Table 4). Note that the term $\frac{1}{2}\hat{\sigma}^2_a + \hat{\sigma}^2_e$ was always close to 15. This terms represents the 'within family variance' or, with more generations, the total variance given the variance among base animals. This estimate was rather robust to the models and data sets used.

TABLE 4. Estimated genetic variance ($\hat{\sigma}^2_a$) using simulated data from different selected generations and a model treating base animals as fixed [a] [b].

data (gen)	base gen.	$\hat{\sigma}^2_a$	$\hat{\sigma}^2_e$
4-5	3	6.03 (2.14)[c]	11.77
3-5	2	8.28 (1.79)	10.76
2-5	1	8.96 (1.23)	10.49
1-5	0	9.24 (0.99)	10.34
4-5	3	6.03 (2.14)	11.77
3-4	2	7.94 (2.72)	10.99
2-3	1	7.05 (3.45)	11.47
1-2	0	6.51 (3.00)	11.79
including records from base animals [d]			
(3) 4-5	3	6.02 (1.94)	11.87
(2) 3-5	2	8.37 (1.60)	10.82
(1) 2-5	1	8.93 (1.14)	10.59

[a] Initial genetic and residual variance was 10
[b] Twenty out of 200 males were selected each generation
[c] Empirical standard deviations from n replicates
[d] Data on parents only for the generation between brackets.

DISCUSSION

Properties of the conditional model

When only base sires were treated as fixed, the procedure of Graser et al. (1987) seemed to reduce the bias in $\hat{\sigma}^2_a$ considerably, by applying a different model to the same information on data and pedigree (Table 2). However, treating all base animals as fixed gave an estimate of genetic variance which was much lower than two times the variance of Mendelian Sampling in generation 4; the component that was assumed to be estimated.

Dams were nested within sires. Therefore, to create a full rank coëfficient matrix in [1], equations for sires and for μ were set to zero in a model treating base animals fixed. It should be noted that the procedure of Graser et al. (1987) can not be applied to an animal model with single records available on animals from one generation only, since there is no covariance between animals within full sib classes after

correction for family effects.

Estimates of effects of base dams, when fixed and nested within sires, can be seen as estimates of family means. The variance of the random genetic component is then derived from deviations from family means. Curnow (1961) has shown that the log-likelihood can be written as the sum of the log-likelihood of parental values (y_b) plus the log-likelihood of offspring values given parental values (z_p). However, offspring values were derived from offspring records deviated from a regression of parental values on offspring values ($z_p = y_p - [\text{cov}(y_b, y_p)/\text{var}(y_b)]y_b$). When base animals are taken as fixed, the deviations considered are $z_p = y_p - \frac{1}{2}y_b$ and therefore parent offspring regressions may not be appropriate.

The following example, suggested by Thompson (1989, personal communication), shows how fixed base animals can lead to biased estimates. Suppose records are available only on females from two generations (1 and 2) and no male pedigree known. Furthermore, dams from generation 0 are base dams and they each have one female offspring. Dams from generation 1 are selected and have n offspring each. If base dams are assumed fixed, there are only two contrasts that provide information about the variance components: the within family variance is derived from the contrast ($y_{2_{ij}} - y_{2_{i.}}$), estimating $\sigma^2_w = \sigma^2_e + .75\ \sigma^2_a$. The other quadratic is based upon contrasts between family means, corrected for the fixed effect of the base dam; it uses ($y_{2_{i.}} - \frac{1}{2}y_{1_i}$). Let the variance of selected 1st generation dams with progeny be $k\sigma^2_{p1}$, and σ^2_{a1} is the genetic variance in that generation, which is reduced due to selection of base parents (gen 0). The term ($\bar{y}_{2_{i.}} - \frac{1}{2}y_{1_i}$) can also be written as ($-\frac{1}{2}(1-h_1^2)y_{1_i} + \bar{e}_{2i.}$), where $\frac{1}{2}h_1^2$ is the regression of y_{1_i} on $\bar{y}_{2_{i.}}$. The variance of ($-\frac{1}{2}(1-h_1^2)y_{1_i} + \bar{e}_{2i.}$) is $\frac{1}{4}(1-h_1^2)^2\text{var}(y_{1_i}) + \sigma^2_w/n + \frac{1}{4}\sigma^2_{a1}(1-h_1^2)$, and it can be shown that this is equal to $\sigma^2_w/n + \frac{1}{4}c\sigma^2_a$, with $c = (k\sigma^2_a + \sigma^2_{a1})/\sigma^2_{p1}$. Hence, estimates are not biased by variance among base dams, since σ^2_{a1} does not appear in the expectations. However, when there is selection in generation 1, c becomes smaller than 1 and the residual variance will be underestimated and consequently the estimate of σ^2_a will be biased upward. This result was confirmed by a simulation with 200 dams in generations 0 (no records), and selecting 100 dams (out of 200) in generation 1 to have 4 progeny each. The estimate of σ^2_a (100 replicates), considering base dams fixed, was 0.692 times the true value. If base dams were not fixed, information of unselected dams about

the selected dam mean could be used to account for this selection (Thompson, 1973; 1976).

The example with one sex recorded was an easy structure, and used to show that in a conditional model the expectations can be biased by selection of progeny from fixed base animals. Magnitude and sign of bias are not general but depend on population structure, parameters and selection intensity. In the population simulated in this study, records were known for two sexes. In a hierarchical structure with two sexes known, there are three quadratics available to estimate variance components; the within family variance, the variance among full sib groups within half sib groups, and the variance among half sib groups. It is difficult to derive a general formula for the bias for more complicated structures. Results from simulation with two base parents fixed showed an under- rather than overestimation of the genetic variance (Table 4). The bias decreased with a higher heritability, and with the number of base sires in generation 0 smaller than 20 (Table 5).

With no selection and no inbreeding, the variance among the components of genetic variance is assumed to be constant, i.e. V(sires): V(dams): V(Mendelian Sampling)= δ_s: δ_d: δ_m= .5: .5: 1. The variance among sires and among dams will be reduced by selection of parents and grandparents.Therefore, we assumed in an alternative analysis a fixed ratio between the three components of additive genetic variance. For example assuming a reduction

TABLE 5.Estimated genetic variance ($\hat{\sigma}^2_a$) using simulated data from selected generations with varying number of base sires and varying heritability and a model treating base animals as fixed [a] [b].

data (gen)	base gen.	$\hat{\sigma}^2_a$	$\hat{\sigma}^2_e$
no. of base sires(ns), h_2=.50			
20 1-2	0	6.51 (3.00)	11.79
10 1-2	0	7.34 (3.90)	11.64
1 1-2	0	7.73 (3.82)	11.36
heritability, ns=20			
.50 1-2	0	6.51 (3.00)[c]	11.79
.80 1-2	0	8.95 (1.99)	3.00

[a] Initial genetic variance was 10
[b] ns out of 200 males were selected each generation
[c] empirical standard deviations from n replicates

TABLE 6. Estimated genetic variance ($\hat{\sigma}^2_a$) using simulated data from selected generations [b] and a model assuming a fixed ratio between different components of genetic variance.

ration V_s :V_d :V_m [a]	n	$\hat{\sigma}^2_a$	$\hat{\sigma}^2_e$
.3 :.4 : 1	20	9.71 (1.97)[c]	9.99
.25 :.5 : 1	20	8.74 (1.86)	10.34

[a] V_s, V_d and V_m= variance among sires, dams and among Mendelian sampling terms, respectively.

[b] Data used from generations 4 and 5 of population B assuming generation 3 as base.

[c] Emperical standard error between brackets.

of variance among sires of 50% would give a ratio of δ_s: δ_d: δ_m=.25: .5 : 1. A better approximation of the true ratio for the animals of generation 4 was assumed to be .3: .4: 1. Operationally we added in [12] α/δ_s to the diagonals for sires and α/δ_d to the diagonals for dams in a_b. Results in Table 6 show that assuming a reduced variance among base animals gave a better estimate of the additive genetic variance than treating base animals as fixed. Bias of estimated variances was smaller when the assumed parameter ratio approached the true values more accurately.

Comparison with Westell grouping

Henderson (1988) has shown a more general way to account for selected base populations. He described selection on a vector of breeding values as $\underset{\sim}{M}'a$, and proposed to account for this kind of selection by substracting $\alpha M(M'AM)^{-1}M'$ from the random part of the mixed model equations $(Z'Z+ \alpha A^{-1})$. Considering selection on base animals as selection on

$$M'a=[M_b' \quad 0] \begin{bmatrix} a_b \\ a_r \end{bmatrix}$$

a term of $\alpha M_b(M_b'A_{bb}M_b)^{-1}M_b'$ would be subtracted from the partition of base animals from the coefficient matrix. Henderson remarked that assuming base animals as fixed is equal to assuming M_b=I and $\alpha(A_{bb})^{-1}$= αI_b would be subtracted from the diagonal block pertaining to base animals, which gives equation [12].

Henderson (1988) indicated that to predict random effects his M'a selection could also be accomodated with a 'phantom grouping' strategy as proposed by Westell et al. (1988) and further described by Quaas (1988). Using our notation we can show this analogy as follows. The equations with groups for phantom parents, assuming var(a)= $A\sigma_a^2$ and var(e)=$I_n\sigma_e^2$, are (Quaas, 1988)

$$\begin{bmatrix} X'X & X'Z & 0 \\ Z'X & Z'Z+\alpha A^{-1} & -\alpha A^{-1}Q_g \\ 0 & -\alpha Q_g'A^{-1} & \alpha Q_g'A^{-1}Q_g \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{a} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ 0 \end{bmatrix} \qquad [14]$$

where Q_g relates animals to groups and Q_g is defined as $(I-P)^{-1}P_bQ_b$. P describes the gene flow among animals and is equal to

$$\begin{bmatrix} I & 0 \\ P_{21} & P_{22} \end{bmatrix}$$

as in [3], P_b relates animals to phantom parents and Q_b relates phantom parents to groups. When two phantom parents are assigned to each base animal, and those two phantom parents are assigned to the same group, P_bQ_b is a matrix (with order nr. of animals x nr. of groups) equal to $[I_b : 0]'$. Since A^{-1} can be written as $(I-P')D^{-1}(I-P)$, the term $A^{-1}Q_g$ in [14] can be replaced by $(I-P')D^{-1}P_bQ_b$ which is equal to $[I_b : 0]'$, and $Q_g'A^{-1}Q_g = I_b$. Therefore, using this grouping strategy and absorbing equations for groups in [14] is in effect equal to subtracting αI_b from the base animals partition of the coefficient matrix, giving again equations [12]. Hence, assuming one group for each base animal is a grouping strategy equivalent to treating base animals as fixed.

Base animals could be combined in one group, e.g. according to their age- or breed, and equation [14] can be modified according to the definition of Q_b. Contrasts between base animals from different groups are not used to estimate genetic variances, but contrasts within groups are. It seems therefore that Henderson's approach, or equivalently the phantom grouping strategy (eq'n [14]) can be a general strategy to account for

diferent expectations among base animals. A requirement for unbiased estimation of σ_a^2 is that the variance among base animals within groups is not altered by selection, and, as shown in this study, that progeny of base animals are not selected to have records.

Analysis of field records

First lactation records (305 days milk production) from Dutch Black and White heifers calving between September 1983 and September 1986 were used. Records were on crossbred animals carrying various portions of Holstein Friesian and Dutch Friesian genes. Details on data edits and data structure are in Van der Werf and De Boer (1989). The data set is summarized in Table 7. Data were split up in three subsets according to the breed of the sire, because selection strategies may have been different for the base populations in the different breeds.

TABLE 7. Characteristics of sub sets of field records; number of proven sires (NPS), number of young sires (NYS) and average effective number of daughters per young sire (ED).

	total records	NPS	NYS	ED
subset I (FH)[a]	89,576	202	174	82
subset II (50% HF)	87,217	202	127	91
subset III (HF)	163,132	202	211	120

[a] refers to breed of young sires where FH= Dutch Friesian, HF= Holstein Friesian.

Records of field data were analyzed using a mixed linear model that accounted for fixed effects of herd-year-season, month of calving, a second dergree polynomial for age of calving, fixed genetic effects and random effects of sires. Fixed genetic effects consisted of linear regression on percentage of HF genes, on heterozygosity, and on recombination in the genome of the progeny. Sires of unproven sires were considered to be base animals. We compared analyses treating base sires as random (model RB) with an analysis were they were treated as fixed (model FB). Because young bulls were assumed to be not selected to have progeny, the estimate of genetic variance was not biased by selection.

Variance components were estimated using the REML-EM algorithm on a tridiagonal set of equations (eq'n [11], and Meyer, 1986). In order to utilize the orthogonality of the tridiagonalizing matrix in computing the sums of squares necessary for the EM-algoritm, fixed effects and base sires, when treated as fixed, had to be absorbed into equations of their random sons before the coefficientmatrix was tridiagonalized.

Table 8 shows estimates for different subsets and for overall data using a model with base animals assumed fixed (model FB). Estimates are given proportional to the estimate from a model assuming base animals random. Deviations from 100 represent bias from the random model due to selection of base animals. Directional selection on one trait is expected to lead to an underestimation of the genetic variance using a model RB. For example, with a proportion selected of 60% with an accuracy of .90, the expected reduction in genetic variance among sires of young bulls is about 50% and the variance among young bulls is expected to be reduced by 12.5%. The value for the relative variance in Table 8 would be (1/.875* 100%)= 114.4. Since estimates of residual variance were almost identical for both models, relative values for heritability were close to relative values for the genetic variance.

In field data it is difficult to predict the change of variance after selection because the realized selection intensity is not precisely known. Selection is practiced on several traits and independent culling type of selection on two negatively correlated traits might even have given rise to an increase of genetic variation. Graser et al. (1987) applied the conditional model to a field dataset and also found only small differences compared to a model assuming unselected base animals. From Table 8 it

TABLE 8 Estimates of genetic variance with base animals fixed relative to estimates with base animals random

	subdatasets (%HF in sires)		
	0%	50%	100%
milk yield (kg)	101.4	99.9	107.8
fat yield (kg)	101.7	101.8	97.9
protein yield (kg)	100.9	100.6	106.1
fat content	104.7	96.3	106.4
protein content	103.4	87.9	97.1

appears that the reduction in variance among sires was small in our dataset. This might be due to the fact that selection was for more traits and the selection intensity per trait was low. Genetic trends estimated for tested bulls in the Netherlands (1980-1988) were lower for the FH sires compared with sires with HF genes (Boukamp, 1989). Results from Table 8 show that the change of variance after selection was also lower for FH sires.

CONCLUSION

Conditional models treating base animals as fixed can be useful to account for selection prior to the base population. However, estimates will be biased by selection of progeny of base animals, even when data of culled animals are available. When base animals as well as progeny are selected, different components of genetic variance have to be estimated, e.g. the variance among base animals, and the variance conditional to the base parents. However, this is not feasible with many data sets. Estimation of genetic variance based on field recorded milk production data could be accounted for selection of sires of test bulls, and estimated variance for milk yield was about 8% higher compared to a model that assumes no selection.

ACKNOWLEDGEMENTS

The author is very grateful to Robin Thompson for his valuable suggestions, to Karin Meyer for providing her DFREML programs, and to Arnoud van der Lugt and Imke de Boer for their help in the analysis.

REFERENCES

Boukamp, M. 1989. De genetische trend en voorspelbaarheid van fokwaarden bij in Nederland ingezette proefstieren. Student thesis, Wageningen Agricultural University (Dutch with Eng. summary).

Bulmer, M.G. 1971. The effect of selection on genetic variability. Ameri. Nat. 105:201.

Curnow, R.N. 1961. The estimation of repeatability and heritability from records subject to culling. Biometrics 17:553.

Graser, H.U., S.P. Smith and B. Tier. 1987. A derivative free approach for estimating variance components in animal models by restricted maximum likelihood. J. Anim. Sci. 64:1362.

Henderson, C.R. 1985. Best Linear unbiased prediction using numerator relationship matrices derived for selected base populations. J. Dairy Sci. 68:443.

Henderson, C.R. 1988. A simple method to account for selected base populations. J. Dairy Sci. 71:3399.

Maijala, K. and M. Hanna. 1974. Reliable phenotypic and genetic paramters in dairy cattle. Proc. 1st World Congr. Genet. Appl. To. Livest. Prod., Madrid. vol. 1:542.

Meyer, K. 1986. Restricted maximum likelihood to estimate genetic parameters - In practice. Proc. 3rd World Congr. Genet. Appl. to Livest. Prod. Linclon. vol 12:454.

Meyer, K. 1989. Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative free algorithm. Génét. Sél. Evol.21:317-340.

Quaas, R.L. 1976. Computing the diagonal elements and inverse of a large numerator relationship matrix. Biometrics 32:949.

Quaas, R.L. 1984. Equivalent models. In: BLUP School Handbook: use of mixed models for prediction and for estimation of (co)variance components. Anim. Genet. and Breed. Unit, Univ. of New England, N.S.W. 2351, Australia. p. 21-23.

Quaas, R.L. 1988. Additive genetic model with groups and relationships. J. Dairy Sci. 71:1338.

Sorensen, D.A., and B.W. Kennedy. 1984. Estimation of genetic variance from selected and unselected populations. J. Anim. Sci. 59:1213.

Thompson, R. 1973. The estimation of variance and covariance components with an application when records are subject to culling. Biometrics 29: 527.

Thompson, R. 1976. Design of experiments to estimate heritability when observations are availble on parents and offspring. Biometrics 32:283.

Van der Werf, J.H.J. and W. de Boer. 1989. Estimation of genetic parameters in a crossbred population of Black and White Dairy Cattle. J. Dairy Sci.72:2615.

Van der Werf, J.H.J. and I.J.M. de Boer, 1990. Estimation of additive genetic variance when base populations are selected. Accepted for publication in J. Animal Sci.

Westell, R.A. R.L. Quaas and L.D. Van Vleck. 1988. Genetic groups in an animal model. J. Dairy Sci. 71:1310.

Chapter 6

GENERAL DISCUSSION

Results from studying genetic as well as statistical assumptions made in analysis of data to estimate genetic parameters were presented in this thesis. Non-additive genetic effects between breeds were estimated and its influence on the estimation of additive genetic parameters has been discussed. In this chapter some general comments will be made concerning the estimation of non-additive parameters from field data and the need to account for non-additive effects in practice. Furthermore, the effects of selection and inbreeding on estimates of additive genetic variance and methods to account for these effects were studied. Genetic and non-genetic aspects of estimation of genetic variance of milk production traits in crossbred populations under selection as well as the choice of variance and heritability estimates, used in genetic evaluation and optimization of a breeding program, will be discussed.

1 Non-additive effects in crossbreeding data

1.1 Estimation of crossbreeding parameters from field data

Estimates of non-additive parameters presented in this thesis were based on field data. The advantage of such data is that many observations were available. Estimates from literature arise mostly from crossbreeding experiments in which the design can be made optimal. The question is how estimates from field data compare with those from carefully designed experiments.

Let Y_i be the phenotypic mean of the i^{th} combination of sire- and dam group (i=1,..,n), β_j is the j^{th} crossbreeding parameter (j=1,..p) and K is a (n x p) matrix with k_{ij} being the coefficient for the j^{th} parameter in the i^{th} mating combination. Then $Y = K\beta + e$ and $p = \text{rank}(K)$. The matrix K can be constructed for the breed combinations found in this study (Table 1). Parameters are estimated as $\hat{\beta} = (K'V^{-1}K)^{-1}K'V^{-1}y$ and $\text{var}(\hat{\beta}) = (K'V^{-1}K)^{-1}$. Approximate estimates and measures of accuracy can be easily made by

TABLE 1. Mating combinations used in this study (Chapter 3, Table 1) with actual and optimal number of observations.

mating type[a]		actual	optimal[b]
p_s	p_d	% records	% records
0	0	24.7	5.8
0	.5	1.5	8.0
.5	0	5.7	8.0
.5	.5	.8	9.1
.5	.75	.6	8.8
.75	0	3.2	10.2
.75	1	.7	7.2
.875	0	.9	7.1
1	0	49.1	13.1
1	.5	8.9	10.9
1	.25	1.4	11.6

[a] p_s and p_d refer to fraction HF genes of sire and dam, respectively
[b] according to criterium of Fimland (1983)

assuming the matrix V to be a diagonal matrix, with v_i is σ_e^2/n_i, where n_i is the number of observations for each combination. Such a variance-covariance matrix V gives the Weighted Least Squares Solution (WLS) of ß. The sampling variances and correlations between estimates were obtained from var($\hat{\beta}$), after obtaining $\hat{\beta}$ as WLS (Table 2). Note that the WLS standard errors are lower than the standard errors from the GLS analysis (Chapter 3), because covariances due to animal effects were ignored. Table 2 shows estimates of breed effects and non-additive parameters to be negatively correlated, whereas heterosis and recombination estimates had positive sampling correlation.

TABLE 2. Variances and correlation matrix of parameter estimates from Weighted Least Squares for actual data set[1].

parameter	variance	correlation matrix		
		breed	heterosis	recomb
mean	4.38	-.08	-.35	-.27
breed effect	76.67	1	-.86	-.57
heterosis	24.43		1	.55
recombination	143.40			1

[1] data set was described in Chapter 3 (Table 1)

Fimland (1983) presented a method to determine a design with approximate optimal number of observations for each mating combination, given a set of mating combinations and a number of total observations. Such a design should give maximum information about the parameters of interest. When k_i is the i^{th} row of K, the optimal number of observations (n_i) as a fraction of N observations for the i^{th} mating type can be approximated as $n_i/N \sim [1/k_i' k_i]/\Sigma_{j=1}^{n}[1/k_j' k_j]$ (Table 1).

Sampling variances and correlations for an optimal structure with the same total number of observations and the same mating types are given in Table 3. The size of the data set is inversely related to sampling variances whereas the design determines sampling variances as well as sampling correlations. For example, an optimal designed experiment of size 4000 would give sampling correlations as in Table 3 but sampling variances would

TABLE 3.Variances and correlation matrix of parameter estimates from Weighted Least Squares for optimal data structure.[2]

parameter	variance	correlation matrix		
		breed	heterosis	recomb
mean	11.74	-.47	-.42	-.59
breed effect	37.37	1	-.43	-.12
heterosis	19.05		1	.08
recombination	52.82			1

[1] size of data set as in Table 2

be 100 times larger than those of Table 2. It can therefore be concluded that estimates for crossbreeding parameters from field data give low sampling errors, relative to estimates from small scale experiments, but sampling correlations can be substantially lower using a designed experiment.

It should be noticed that Fimland's approximate method for an optimal design does not consider sampling covariances. Other optimality criteria to determine optimal designs have been used, e.g. the D-optimality, maximizing the determinant of $(X'V^{-1}X)$ (Cameron and Thompson, 1986; Sölkner and James, 1989). Using D-optimality, Sölkner (1989) showed the efficiency

of estimations to be quite robust for suboptimal numbers of observations allocated to each group but concluded that choice of the right genetic groups (i.e. mating combinations) are more important.

1.2 Parametrization of crossbreeding effects

Although nine genetic effects can be derived from a two-locus model (Anderson and Kempthorne, 1954) the model is usually restricted to only a few of these. Fimland (1983) gives the expectation of parameters when some other factors are ignored. Let K contain the coefficients for the parameters in Table 2, and K_m be a column of coefficients for the maternal effect for each mating combination. The expectation of the parameters in the model ignoring the maternal effect (m) is then

$$E(\hat{\beta}) = \beta + Qm \qquad (1)$$

where $Q = (K'V^{-1}K)^{-1}K'V^{-1}K_m$. For the data set in this study, Q' = [-.004, 0.64, -.324, .743]. Hence, ignoring a significant positive maternal additive effect would give an underestimation of the heterosis component whereas the recombination effect would be overestimated.

Using (1), a Q' for the paternal effect of [.007, 1.18, 0.401, -0.236] can be obtained. Although the paternal genetic effect is expected to be low, there could be a 'paternal effect' due to favorable treatment of daughters of (imported) HF sires or imported semen could have been used to breed the best dams. Preferential treatment of cows related to the percentage of Holstein genes of the sire would therefore give an upward bias for the breed effect and the effect of heterosis whereas the recombination effect would be underestimated.

Heterosis and recombination loss have been related to genetic effects at locus level by Hill (1982). Let δ represent the dominance effect, $\alpha\alpha$ the additive by additive epistatic effect and $\delta\delta$ the dominance by dominance interaction. Using Hill (1982), it can be shown for a two-locus model that heterosis represents $(2\delta-\alpha\alpha)$ and recombination can be replaced by $(-\alpha\alpha-\delta\delta)$. When dominance by dominance interactions are ignored, the vector of parameters (β) that includes heterosis and recombination can be transformed to produce a vector c with parameters δ and $\alpha\alpha$; $c = T\beta$. Now $Var(\hat{c}) = T\,Var(\hat{\beta})T'$ and since T is invariant to the design, genetic effects

from different model of the same design are simply linear functions of each other. The solution for δ was then 112.1 (s.e.= 7.6) and for $\alpha\alpha$ it was 101.2 (s.e.= 12.0). Sampling correlation of $\hat{\delta}$ with estimated breed effect was .17 and with $\alpha\alpha$ it was .96. Correlation of $\alpha\alpha$ with breed effect was .57.

1.3 Non-additive effects within breeds

The existence of positive heterosis in a cross of two strains of Friesian cattle indicates that dominance effects exist. However, Beckett et al. (1979) analyzed crosses between inbred lines of Holstein cows and concluded that the possibility of using specific combining abilities of inbred lines was not very likely. Under an 'infinitesimal' model with many loci affecting the trait under selection, the change in gene frequency after several generations of selection is small compared to differences between breeds and the variation within breeds. It would be interesting if the hypothesis concerning the underlying genetic model could be supported by molecular genetic techniques to measure genetic distance, as recently proposed (e.g. Sharp, 1987), but within breed applications of these methods are unknown.

Dominance might not be used in dairy cattle breeding, but it could be nuisance variance. Estimates for dominance variance from sire by maternal grandsire components have often been found insignificant or with high standard errors (Allaire and Henderson, 1965; Lee and Henderson, 1969). Studies on twins showed higher covariances than expected from additive variance, but this can be due to environmental effects as well (Johannson and Rendel, 1968). Recently in dairy cattle the number of full sibs has increased substantially due to embryo transfers (ET). Tempelman (1989) found high estimates of dominance variance (as a ratio to total variance); 0.08-0.31 for milk yield and 0.08-0.50 for fat yield. However, about 20% of the data arose from ET coming from highly selected parents and families were partly confounded with herd groups. Furthermore, standard errors and sampling correlation with additive variance were high.

Henderson (1985a,b) has proposed animal models including additive as well as non-additive genetic variances. Components of additive and dominance variance could be obtained from such models using a Restricted Maximum Likelihood procedure. Lin and Lee (1989) advocated a full-sib

model, arguing that an animal model is often not practical because it requires inversion of a large coefficient matrix. Although the computational argument diminishes due to the introduction of derivative free methods (Meyer, 1989) and/or sparse matrix techniques (Misztal, 1989), setting up a full-sib model could be useful because it provides more explicit information on whether different components of variance are separable from a given data structure.

Inbreeding depression depends on dominance as well. The change of the mean is proportional to the degree of inbreeding (Bulmer, 1980). Estimates of effects per percent of inbreeding range from -15kg to -40kg for milk yield and from -0.5 to -1.0 kg for fat yield (Hodges et al., 1979; Hudson and Van Vleck, 1984). Dairy cattle populations generally have low coefficients for average inbreeding, but coefficients increase with more intensive use of MOET schemes, e.g. in nucleus herds. A mixed model using additive and dominance relationships matrix accounts for the effect of inbreeding on the variance-covariance structure, but not for inbreeding depression (Mäki-Tanila and Kennedy, 1986). Inbreeding depression could be accommodated by inclusion of the inbreeding coefficient (F) as a covariate in the model (Kennedy and Sorensen, 1987).

1.4 Non-additive effects and genetic evaluation

The influence of non-additive genetic effects on estimation of breeding values has been clearly demonstrated in this thesis. Non-additive effects from crossbreeding should be seen as nuisance effects which have to be accounted for in breeding value estimation. As demonstrated in Chapter 3, this can be done simply by including heterosis and recombination as covariables in the model. It might be useful to determine sampling variances and correlations for a particular design because effects may be confounded in some data sets.

Ignoring significant non-additive variance within breeds would have an impact on genetic evaluation (Chapter 3). For example, covariances between full sibs would be underestimated, which is undesirable because breeding values of sires and elite cows are often based on information from full sib ET progeny. Predicted breeding values, however, would not be biased by dominance as long as the appropriate heritability is used.

Using dominance effects to predict performance of planned matings does

not seem to be opportune in the current dairy cattle breeding. The coefficient for dominance relationship diminishes rapidly with distant relationships. With sufficient genetic progress, it will mostly not be interesting to reproduce dominance effects at the time it has been accurately estimated. Predicting dominance effects would be more useful when more links exist between animals (e.g. with heavier inbreeding) or if identical genotypes (clones) would be generated.

2 Estimation of additive genetic variance

2.1 Genetic aspects of the model

It has been assumed throughout this thesis that genetic variance was equal for the different populations and that different crossbred groups showed homogeneous variances. This was justified assuming very small differences in gene frequency between parental populations, which was suggested to be possible under an infinitesimal model. Little is known about validation of this assumption. Meyer and Hill (1990) have tested the 'infinitesimal model' hypothesis using data from a long term selection experiment in mice. They suggested some of the change of genetic variance was due to changes in gene frequencies. Simulations at one-locus level show heterogeneity of variance for moderate (>.1) gene frequency differences. Analysis by subpopulation showed some heterogeneity of genetic variance (Chapter 3), but this could also be caused by environmental effects or by genotype by environment interaction.

Although the model corrected for systematic population effects and selection, heritability estimates for milk production traits continued to stay higher as previously assumed. Other forms of non-random sampling are assortative mating of the best sires to the best dams or sampling of the best genotypes at the high yielding herds. The fact that some sires have daughters only in the best herds could be a problem when there is a sire by herd interaction. Meyer (1987) found a small reduction in sire variance after accounting for this interaction. Assortative mating would lead to an overestimation of the sire component of additive variance using a sire model which assumes all dams unrelated and their merit to be uncorrelated with the sires genotype. An animal model should account for this bias, at

least when there are records available on dams and when assortative mating is based on these records.

Preliminary analysis using an animal model was done for 4 different data sets of about 10,000 records each. Records of Black and White cows (Dutch x Holstein Friesian crosses) were obtained from the Dutch Cattle Syndicate (NRS) for the years 1983-1989. Heritability estimates for milk yield averaged 0.47. Estimates from a sire model in similar data were lower (Chapter 3). Variances obtained from animal models (Swalve and Van Vleck, 1987; Van Vleck and Dong, 1988) have been found to be slightly higher than previous estimates from sire models, particularly because sire models often ignore relationships among sires and include effects of selected sires. However, sire models used in this study accounted for these effects. Recently, high heritabilities for milk production have also been found

TABLE 4 Recent estimates of heritability ($\hat{h}^2$), and genetic ($\hat{\sigma}^2_a$) and environmental variance ($\hat{\sigma}^2_e$) for 1st lactation milk yield.

authors	herd type	method[1]	breed	$\hat{h}^2$	$\hat{\sigma}^2_a$	$\hat{\sigma}^2_e$
Hill	low mean	SM	BF	.24	85360	264480
et al. 1983	high mean		BF	.30	153630	355070
	low var.		BF	.24	72132	233510
	high var.		BF	.30	164500	382010
Boichard and Bonaïti ,1987	all	SM	FF/HF	.37	242060	410286
Van Vleck	low mean	AM	HF	.23	75780	247220
et al., 1988	moderate mean		HF	.29	118600	288480
	high mean		HF	.36	150380	280920
Van der Werf &	all	SM	FH	.34	138970	269769
De Boer, 1989			HF	.44	238000	302904

[1] SM= sire model AM= animal model
[2] BF, FF, FH, HF = British, French, Dutch and Holstein Friesian, respectively

using a sire model (Cue et al., 1987) or dam-daughter regression (Wade and Van Vleck, 1989), the last method being unaffected by selection of dams.

Heritabilities as high as 0.47 have not been often published for milk

yield in American Holstein populations. Variance in this study could be higher because genes from two strains of Friesian cows are mixed, or because the imported Holstein population expresses other genes in a new environment. On the other hand, high genetic variances and heritabilities have been reported in high yielding herds (Table 4; Wade and Van Vleck, 1989). Estimates of heritability, and genetic and environmental variance are summarized in Table 4 for different herd production levels and different Friesian populations. More detailed analysis is needed using data stratified by subpopulation and herd production level.

2.2 Aspects of the data structure

Parent-offspring regression and half-sib analysis put requirements on the data structure. Since variance components from animal breeding data are estimated using methods that are based on mixed model equations (MIVQUE, REML), requirements to be set on the data structure have become more loose. A REML analysis can essentially be applied to almost any data structure, and it will mostly provide estimates for genetic variance as long as there is data on related animals. Information to estimate the additive genetic component of variance, coming from different sources such as contrasts among parents, parent-offspring covariances and contrasts within families, is combined to a final REML estimate. It would be useful to know how those different sources of information contribute to the final estimate. Possible bias from ignoring maternal-, dominance- or permanent environmental effects, or bias from selection could be assessed. An example has been described in Chapter 4 where we compared data sets relating to the same base but with data on a different number of subsequent generations. We showed empirically that having more generations with data did not reduce the bias from selected base animals. However, there is a lack of theory about how the 'between-' and the 'within-'base animal components of variance are weighted.

There is also little known about the information available from data from a complex family structure. Thompson (1989) suggested to find eigenvalues of the coefficient matrix of an animal model by diagonalization of the mixed model equations. The number of nonzero eigenvalues shows how many contrast between animal are available. Each eigenvalue can be seen as the coefficient for the animal component of variance in the expectation of

the quadratic pertaining to the information available on a certain animal. He named the list of latent roots a 'familyogram'. It is not yet clear what exactly are properties of a 'good' familyogram.

The information provided by a given data set is determined by the number of contrasts that can be used to estimate each component of variance. Using ANOVA methods, the estimability of components is relatively easy to see from the degrees of freedom that is associated with fitting certain effects. The information coming from the data determines the shape of the curve from plotting the likelihood against parameter values. The second derivative of the likelihood function is summarized in the information matrix, containing information about approximate variances and covariances of the parameter estimates. REML algorithms often use only first derivatives or use a direct derivative free maximization and the information matrix is not explicitly set up. The matrix can be approximated (Meyer, 1989) but this is more difficult when more then two components have to be estimated. Sometimes accuracy of estimation and speed of convergence can be improved by a reparametrization of the variance components, e.g between- and within- family variance is estimated rather than additive genetic and residual components (Thompson and Meyer, 1986). REML estimation of several genetic components using an animal model therefore requires careful inspection of the data structure.

2.3 Which variances should be used?

Best Linear Unbiased Predictions of breeding values only exists when true variances are known and they are better approximated as used variances approach the true values. It has been shown in Chapters 4 and 5 that different estimates for genetic variance can be obtained, depending upon the data and the model used. Hence the question arises which heritability is actually needed for animal breeding applications?

Mixed models for genetic evaluation assume the base population to be noninbred and unrelated. It has been argued in Chapter 5 that inbreeding does not have to be accounted for to estimate genetic variance, but the genetic variance used in the evaluation model should relate to the assumed base population. Accounting for relationships among "base" animals uses more information on relatives and can account for some bias from gametic disequilibrium, particularly in small populations. Animal models are

expected to produce higher estimates of genetic variance when they account for more relationships. This has been shown in chapter 4 and by Dong et al. (1988). It would be consistent in this respect to use higher heritabilities in animal model prediction of breeding values.

The genetic variance was divided in a part coming from variation between base parents and a part resulting from Mendelian sampling at formation of parental gametes; between- and within-family variance, respectively. Between- and within-family components of variance are only equal in case of no selection. Should the heritability, in case of selection, be accounted for the reduction of the between family component due to gametic disequilibrium? For the prediction of long term response of selection the Mendelian sampling variance is the relevant parameter. Both within- and between-family variance are needed to predict immediate response. Prediction of response to selection may than be complicated by selection on correlated variables (Hill, 1977). For the estimation of breeding values in populations under selection, including different parameters for the within- and between family component could be considered because it is relatively easy to do, as demonstrated in Chapter 5. Estimating those different components may be difficult when data sets don't contain more than two generations with data, which is often the case. Alternatively, one could assume a certain ratio between the two components. This approach is rather pragmatic, although basically not more so than assuming ratios that refer to a situation without selection.

The effect of accounting for different values for the between- and within-family variance has not been quantified. Using a compound value would give an underestimation of information on sibs whereas parental information would obtain too much weight. However, selection based on family information is quite robust to errors in heritability.

A reason to have a conservative policy with respect to the heritability used could aim at a lower weight for an animals's own performance. This could be desirable if it is believed that preferential treatment of records occurs. On the other hand, a higher heritability as actually estimated has been advocated to optimize long term selection (James, 1972). The reason is that long term index (or BLUP) selection and conserving genetic variability is antagonistic and individual (or mass) selection would make more use of the available variability. BLUP procedures maximize genetic

progress on short term, given the data, but it does not maximize long term response because it capitalizes on family selection, in particular with low heritabilities. Another aspect in long term selection can be creation of new variation by mutation (Hill and Keightley, 1988) or new genes becoming more important with a rise of production level. In both cases a higher heritabilities would be appropriate since the family information would need less weighing (Dempfle and Gründl, 1988).

Methods to account for heterogeneous variance in animal models may be exhaustive, although simplifications exist when heritabilities are assumed equal for different environments, and when genetic correlations are assumed to be unity (Henderson, 1984). Meyer (1987) reported a small sire by herd interaction for milk yield but Merks (1988) found surprisingly low genetic correlations for growth and backfat in pigs comparing similar traits at nucleus level and farm level. Winkelman and Schaeffer (1988) accounted for heterogeneity of genetic and residual variance but they did not find an effect for sire evaluation. The effect on cow evaluation is expected to be larger. A problem with heterogeneous variance models might be the amount of information available per stratum to estimate variances. A Bayesian approach might be useful and practical to handle such problems (Gianola et al., 1989)

3 Recommendations

Not accounting for non-additive effects in genetic evaluation of crossbred populations causes biased estimates of additive genetic differences and breeding values between crossbred groups. Analysis of crossbreeding data with linear regression on breed effect, heterosis and recombination effect is therefore recommended to adjust for systematic additive and non-additive breed effects.

Variance among parents for a certain trait will be affected by intensive selection on that trait or on highly correlated traits. Estimates of genetic parameters will be less affected by selection of parents if more relationship information among parents is used. A conditional model could be used to account for previous selection, only if progeny of base parents are not selected.

Variance components used for animal evaluation should be estimated with

a model that coincides with the same base population. Estimates of genetic variance are often higher with an animal model, compared to a sire model, and those animal model estimates should be used if genetic evaluation is done with such a model. Estimates of genetic parameters to determine expected genetic progress are less biased by covariances and gametic disequilibrium, when more relationships are included in the model.

High estimates of heritability for milk production traits in crossbred dairy cattle justify a raise of parameters currently used for genetic evaluation of the Dutch Black and White dairy population. However, before using values higher than .40 for yield traits, more research is suggested to investigate whether the increased genetic variance is due to genetic or to environmental changes, and whether these high values will persist.

REFERENCES

Allaire, F.R. and C.R. Henderson. 1965. Specific combining ability among dairy sires. J. Dairy Sci. 48:1096-1100.

Anderson, V.L. and O.Kempthorne. 1954. A model for the study of quantitative inheritance. Genetics, 39, 883-898.

Beckett, R.C., T.M. Ludwick, E.R. Rader, H.C. Hines and R. Pearson. 1979. Specific and general combining abilities for production and reproduction among lines of Holstein cattle. J. Dairy Sci. 62:613-620.

Boichard, D., and B. Bonaïti. 1987. Genetic parameters for first lactation dairy traits in Friesian, Montbeliarde and Normande Breeds. Génét. Sél. Evol. 19:337.

Bulmer, M.G. 1980. The mathematical theory of quantitative genetics. Clarendon press, Oxford, Eng.

Cameron, N.D. and R. Thompson. 1986. Design of multivariate selection experiments. Theor. Appl. Genet. 72:466-476.

Cue, R.I., H.G. Monardes, and J.F. Hayes. 1987. Correlations between production traits in first lactation Holstein cows. J. Dairy. Sci. 70-2132.

Dempfle, L., and E. Gründl. 1988. Identification of superior animals and their use in improvement programmes. In: S. Korver et al. (compilers). Advances in animal breeding. Proc. of the world symposium in honor of

prof. R.D. Politiek. Pudoc, Wageningen, Neth.

Dong, M.C., L.D. Van Vleck and G.R. Wiggans. 1988. Effect of relationships on estimation of variance components with an animal model and restricted maximum likelihood. J. Dairy. Sci. 71:3047-3052.

Fimland, E. 1983. Methods of estimating the effects of heterosis. Z. Tierzüchtg. Züchtungsbiol. 100:3-8.

Gianola, D., J.L. Foulley, R.L. Fernando and C.R. Henderson. 1989. Estimation of heterogeneous variances using empirical Bayes methods: Theoretical considerations. Submitted to J. of Dairy Sci.

Henderson, C.R. 1984. Applications of linear models in animal breeding. University of Guelph, Guelph.

Henderson, C.R., 1985a. Best linear unbiased prediction of non-additive genetic merits in noninbred populations. J. Anim. Sci. 60:111-117.

Henderson, C.R., 1985b. MIVQUE and REML estimation of additive and non-additive genetic variances. J. Anim. Sci. 61:113-121.

Hill, W.G. 1977. Order statistics of correlated variables and implications in genetic selection programmes. II. Response to selection. Biometrics 33:703.

Hill, W.G. 1982. Dominance and heterosis as components of heterosis. Z. Tierzüchtg. Zuchtungsbiol. 99:161-168.

Hill, W.G., M.R. Edwards, M.K. Ahmed and R. Thompson. 1983. Heritability of milk yield and composition at different levels and variability of production. Anim. Prod. 36:59.

Hill, W.G. and P.D. Keightley. 1988. Interaction between molecular and quantitative genetics. In: S. Korver et al. (compilers). Advances in animal breeding. Proc. of the world symposium in honor of prof. R.D. Politiek. Pudoc, Wageningen, Neth.

Hodges, J., L. Tannen, B. McGillivray, P.G. Hiley and S. Ellis. 1979. Inbreeding levels and their effect on milk, fat, and calving interval in Holstein Friesian cows. Can. J. Anim. Sci. 59:153.

Hudson, G.F. and L.D. Van Vleck. 1984. Inbreeding and artificially bred dairy cattle in the NorthEastern United States. J. Dairy Sci.67:161.

James, J.W. 1972. Optimum selection intensity in breeding programmes. Anim. Prod. 14:1-9.

Johannson, I. and J.Rendel. 1968. Genetics and Animal Breeding. Oliver and Boyd. Edinburgh.

Kennedy, B.W. and D. Sorensen. 1987. Properties of mixed model methods for prediction of genetic merit. in. Proc. of the Second Int. Conf. on Quant. Genet. Ed. Weir, B.S. et al. Sinauer Publ. Sunderland Mass.

Lee, A.J. and C.R. Henderson, 1969. Components of genetic variance in milk yield. J. Dairy Sci. 52:780

Lin, C.Y. and A.J. Lee. 1989. Estimation of additive and non-additive genetic variances in noninbred populations under sire and full sib model. Can. J. Anim. Sci. 69:61-68.

Mäki-Tanila, A., and B.W. Kennedy. 1986. Mixed model methodology under genetic models with a small number of additive and non-additive loci. In Proc. 3rd World Congr. Genet. Appl. Livestock Prod. 12:443-448.

Merks, J.W.M. 1988. Genotype x environment interactions in pig breeding programmes. Thesis, Wageningen Agric. Univ. Neth.

Meyer, K. 1987. Estimates of variances due to sire by herd interactions and environmental covariances between paternal half-sibs for first lactation dairy production. Livest. Prod. Sci. 17:95-115.

Meyer, K. 1989. Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative free algorithm. Génét. Sél. Evol. 21:317-340.

Meyer, K., and W.G. Hill. 1990. Mixed model analysis of a selection experiment for food intake in mice. Submitted to Genet. Res.

Misztal, I. 1989. Restricted maximum likelihood estimation of variance components in animal models using sparse matrix inversion and a supercomputer. J. Anim. Sci 67-Suppl.1:30.

Sharp, P. M., 1987. Molecular genetic distances among animal populations-estimation and application. Proc. 38th Annual Meeting of the Eur. Assoc. Anim. Prod.

Sölkner, J. 1989. The impact of different genetic models on the optimum design of crossbreeding experiments. Presented at the symposium on statistical methods and genetic models in animal breeding, 16-18 October, Achselschwang, FRG.

Sölkner, J. and J.W. James. 1989. Optimum design of crossbreeding experiments. I. A basic sequential procedure. J. Anim. Breed. Genet. (in press).

Swalve, H. and L.D. Van Vleck. 1987. Estimation of genetic (co)variances for milk yield in first three lactations using an animal model and

restricted maximum likelihood. J. Dairy Sci. 70:842.

Tempelman, R.J., 1989. Additive and dominance genetic variation of production and conformation traits in Canadian Holsteins. MSc. Thesis, University of Guelph.

Thompson, R. and K. Meyer. 1986. Estimation of variance components. What is missing in the EM algoritm? J. Stat. Comput.

Thompson, R. 1989. Information obtained from selection experiments. Presented at the symposium on statistical methods and genetic models in animal breeding, 16-18 October, Achselschwang, FRG.

Van der Werf, J.H.J. and W. de Boer. 1989. Estimation of genetic parameters in a crossbred population of Black and White Dairy Cattle. J. Dairy Sci. 72:2615-2623.

Van Vleck, L.D. and M.C. Dong. 1988. Genetic (co)variances for milk, fat, and protein yield in Holsteins using an animal model. J. Dairy. Sci. 71:3040-3046.

Van Vleck, L.D., M.C. Dong and G.R. Wiggans. 1988. Genetic (co)variances for milk and fat yield in California, New York and Wisconsin for an animal model by restricted maximum likelihood. J. Dairy. Sci. 71:3053-3060.

Winkelman, A. and L.R. Schaeffer. Effect of heterogeneity of variance on dairy sire evaluation. J. Dairy Sci. 71:3033.

Wade, K.M. and L.D. Van Vleck. 1989. Genetic parameters of Holsteins in California, New York, and Wisconsin. J. Dairy Sci. 72:1280-1285.

SUMMARY

Estimates of genetic parameters needed to control breeding programs, have to be regularly updated, due to changing environments and ongoing selection and crossing of populations. Restricted maximum likelihood methods optimally provide these estimates, assuming that the statistical-genetic model used is correct.

Generally, a model for analysis of milk production data assumes only additive genetic effects and random sampling. These assumptions are rarely met. In many animal populations genetic material from other populations is used. Crossing of lines or breeds often gives rise to non-additive effects. Furthermore, most of the data used for genetic analysis come from populations under selection. The subject of this thesis was to determine whether or not models for genetic evaluation of dairy populations should account for non-additive effects and selection, and how this should be done.

The influence of non-additive effects on the estimation of heritabilities and breeding values was studied in Chapter 2. A population having progeny that descended from sires and dams with various fractions of genes from two breeds was simulated. Additive breed effects and non-additive effects from breed crosses, were simulated. Data on performance were analyzed using mixed models, that accounted for fixed additive genetic group and random sire effects. Three additive models, with genetic groups defined according to 1) breed composition of the progeny, 2) breed composition of the sire and dam, or 3) linear regression on breed fraction, were compared with a non-additive model, with a linear regression on breed fraction, heterozygosity and recombination in the genome of the progeny. Variance components were estimated using restricted maximum likelihood.

Additive genetic variance and heritability were overestimated for an additive model with progeny groups. Additive models gave biased estimates for breed differences, group effects and breeding values. Breed differences were overestimated when sire groups were used. Estimates for each parameter were unbiased using the non-additive model.

In Chapter 3, the same models were applied to data of cows with variable proportions of genes from the Dutch Friesian and the Holstein Friesian (HF) populations. The data set contained 92,333 first lactation records (305

days milk production) of cows from 675 young sires and 307,050 records of cows from 202 proven sires. Estimates for heterosis varied from 2.5% (fat yield) to 0% (protein percentage). Recombination effects varied from -1.9% (protein yield) to 1.5% (fat percentage). Additive models with progeny groups overestimated genetic variance by 6%. Models with sire groups overestimated additive genetic values of imported HF sires by 33%. Using a non-additive model, heritability estimates were .38 for milk yield, .80 for fat percentage and .70 for protein percentage. It was concluded that a non-additive model was preferable for estimation of genetic variance and prediction of breeding values in crossbred dairy populations.

In the fourth chapter, the effect of selection on estimation of additive genetic variance was studied. A population of size 40 was simulated 100 times, for ten generations. Five out of twenty males were selected at each generation and each male was mated to four females and had two progeny. The additive genetic variance (σ_a^2) before selection was 10 and the initial heritability was .5. The genetic variance was reduced to 6.72 after ten generations of selection, due to covariances among animals, inbreeding and gametic disequilibrium. Reduction of variance was lower in another population simulated with size 400 and ten percent of the males selected. Restricted Maximum Likelihood was used to estimate σ_a^2 using an animal model. The estimate of σ_a^2 was empirically unbiased, when all data and all relationships were used. Omitting data from selected ancestors caused biased estimates of σ_a^2 due to the fact that not all gametic disequilibrium was accounted for. Inbreeding and covariances were adjusted for, when additional relationships between assumed base animals were considered. Bias from gametic disequilibrium decreased slightly with the use of more relationship information. Estimates from data based on later generations only, were biased by selection. Mean estimates of genetic variance depended on the assumed base population and were insensitive to the number of subsequent generations with data.

A method to estimate genetic parameters conditional to selection occurring before formation of the base population was investigated in Chapter 5. For this, simulated data from the same populations as in Chapter 4 was used. The method assumes base parents as fixed and a conditional variance is based upon the Mendelian sampling of gametes from the base parents. Selection was for five generations but only animals of generations

4 and 5 were assumed to have performance records and parents known. Additive genetic and residual variance were assumed to be 10. When 20 out of 200 sires were selected per generation, estimated genetic variance was 8.58 when base animals were assumed random, and it was 6.03 when they were fixed. Residual variance was overestimated in the latter case. When males of generation 4 were not selected to have progeny, estimated genetic variance was 9.91. It was concluded that estimates for genetic parameters with the conditional model were not biased by selection of base animals. However, the procedure with fixed base parents was biased when descendants of base animals were selected to have progeny.

Genetic variance of milk production traits was estimated with a conditional model to account for selection of sires. In the HF subpopulation, which had been selected more intensively, genetic variance for milk yield was estimated about 8% higher compared to a random models that assumes no selection.

Estimates of heritability for milk production traits were found to be high with a sire model, after correction for non-additive effects (Chapter 3) and selection of parents (Chapter 5). Preliminary results with an animal model, which accounted for non-random mating of sires, did not show lower estimates. More research is suggested to determine whether the cause for high heritabilities is genetic or environmental.

Main conclusions

- By not accounting for non-additive effects in genetic evaluation of crossbred populations, biased estimates of breeding values and additive genetic differences between crossbred groups are found. Records of crossbred dairy cattle should therefore be adjusted for systematic additive and non-additive breed effects.
- Estimation of crossbreeding parameters from field data can provide low standard errors, although sampling correlation may be high for certain mating designs.
- Estimates of genetic variance based on data from selected generations only were biased by selection. Mean estimates of genetic variance depended mostly on the assumed base population and were insensitive to the number of subsequent generations with data. Additional relationships adjust genetic variance estimates for covariances among animals, and for

some of the gametic disequilibrium.

- Estimates for genetic parameters with a conditional model are not biased by selection of base animals, but a bias will be introduced when descendants of base animals have been selected to have progeny.
- Heritability estimates of milk production traits in crossbred dairy cattle data were found to be higher as parameters currently assumed for genetic evaluation.

SAMENVATTING

Genetische parameters, benodigd voor fokwaardeschattingen en optimalisatie van fokprogramma's, moeten regelmatig opnieuw worden geschat omdat milieuomstandigheden steeds veranderen, en populaties voortdurend onderhevig zijn aan selectie en inkruising. De restricted maximum likelihood (REML) methode geeft hiervoor optimale schattingen indien de statistische en genetische aannames in het gebruikte model correct zijn.

In het algemeen wordt bij de analyse van melkproduktiekenmerken een additief genetisch model aangenomen en de gegevens worden verondersteld op toevallige manier verzameld te zijn. Aan deze aanames wordt meestal niet voldaan. Tussen populaties van landbouwhuisdieren wordt vaak genetische materiaal uitgewisseld en bij het kruisen van rassen treden vaak ook niet-additief genetische effecten op. Verder wordt er in de meeste populaties geselekteerd. In dit proefschrift wordt ingegaan op de vraag of modellen, gebruikt voor de genetische evaluatie van melkvee populaties, rekening dienen te houden met niet-additieve effecten en met selectie.

De invloed van niet-additieve effecten op de schatting van erfelijkheidsgraden en fokwaarden komt in hoofdstuk 2 aan de orde. Er werden produktiegegevens gesimuleerd voor een populatie van koeien die afstamden van ouders met verschillende fracties genen van twee uitgangsrassen. Hierbij werden additieve raseffecten en niet-additieve kruisingseffecten gesimuleerd. De gegevens werden geanalyseerd met behulp van gemengde modellen waarin vaste effecten van genetische groepen en random stier effecten werden opgenomen. In drie additieve modellen werden genetische groepen gedefinieerd volgens 1) rassamenstelling van de koe 2) rassamenstelling van de moeder en van de vader, of 3) een lineaire regressie op de rassamenstelling van de koe. Deze modellen werden vergeleken met een niet-additief model waarin een lineaire regressie op coëfficiënten voor rasaandeel, heterozygotie en recombinatie in het genotype van de koe. Variantie componenten werden geschat met de REML-methode.

Additief genetische variantie en erfelijkheidsgraden werden overschat met een additief model met nakomelingengroepen. Alle additieve modellen gaven onzuivere schattingen voor rasverschil, groepseffecten en fokwaarden. Rasverschil werd overschat in een model met stiergroepen. Schattingen voor alle parameters waren zuiver bij gebruik van een niet-additief model.

In hoofdstuk 3 werden bovenstaande modellen toegepast voor de analyse van een praktijkdataset met melkproduktiegegevens van koeien met variërende aandelen Holstein Friesian (HF) en Fries Hollands (FH) bloed. De gegevensset bestond uit 92.333 eerste lactatielijsten (305 dagen produktie) van dochters van 675 proefstieren, en 307.050 lactaties van dochters van 202 fokstieren. Schattingen voor heterosis varieerden van 2.5% voor vetproduktie tot 0% voor percentage eiwit in de melk. Recombinatie effecten varieerden van -1.9% voor eiwit produktie tot 1.5% voor vet percentage. Een additief model met nakomelingen groepen overschatte de genetische variatie met 6%. Modellen met stiergoepen overschatten de addititief genetische waarden van geimporteerde stieren met 33%. Geschatte erfelijkheidsgraden met een niet-additief model waren .38 voor melkproduktie, .80 voor vetproduktie en .70 voor eiwitproduktie. De conclusie was dat bij gekruiste melkvee populaties gebruik moet worden gemaakt van een niet-additief model voor het schatten van genetische variantie en fokwaarden.

In het vierde hoofdstuk is het effect van selectie op de fokwaardeschatting bestudeerd. Een populatie ter grootte van 40 dieren werd 1000 maal gesimuleerd voor 10 generaties. In iedere generatie werden van de twintig mannetjes 5 geselekteerd. Alle vrouwtjes werden met één van de geselekteerde mannetjes gepaard om 2 nakomelingen te produceren. De additief genetische variatie voor selectie was 10 en de oorspronkelijke erfelijkheidsgraad was .5. De genetische variantie nam af to 6.72 na 10 generaties van selectie door het onstaan van covarianties tussen dieren, door inteelt en door een onevenwichtige verdeling van genen over de gameten (disequilibrium). De reductie van de variantie was kleiner in een tweede populatie met 400 dieren per generatie, waarbij steeds 10% van de mannetjes werd geselekteerd. De genetische variantie werd geschat met REML, gebruikmakend van een diermodel. De schatting van σ_a^2 was empirisch zuiver bij gebruik van alle gegevens en alle genetische relaties. Weglaten van gegevens van geselekteerde ouders veroorzaakte een onderschatting van σ_a^2 door disequilibrium. Het gebruik van alle afstammingsgegevens zorgde voor schattingen die waren gecorrigeerd voor covarianties en inteelt, terwijl een gedeeltelijke correctie plaatsvond voor disequilibrium. De onzuiverheid veroorzaakt door disequilibrium verdween enigszins bij het gebruik van meer verwantschapsinformatie. Schattingen van de genetische variantie waren dus onzuiver door selectie als ze uitsluitend gebaseerd zijn op gegevens van

latere geselekteerde generaties. Het niveau van de schatting werd bepaald door de veronderstelde basispopulatie en de schatting werd slechts in geringe mate beïnvloed door het aantal latere generaties dieren met gegevens.

Een methode voor het schatten van genetische parameters welke niet wordt beinvloed door selectie in eerdere generaties is in hoofstuk 5 onderzocht. Hierbij werd gebruik gemaakt van dezelfde gesimuleerde populaties als in hoofstuk 4. De onderzochte methode gaat uit van een voorwaardelijke kansverdeling die is gebaseerd op de variatie als gevolg van Mendeliaanse recombinatie van gameten van de ouderdieren. Na vijf jaren van selektie werden alleen gegevens en ouderinformatie van de dieren van generatie 4 en 5 gebruikt voor de analyse. De werkelijke additief genetische variantie en restvariatie waren gelijk aan 10. Bij selektie van 20 van de 200 mannetjes in iedere generatie was de schatting van de genetische variantie gelijk aan 8.58 in een model met random basis dieren, en de schatting was 6.03 als de basis dieren fixed werden verondersteld. De restvariantie werd overschat in het laatste geval. De geschatte genetische variantie was gelijk aan 9.91 als dieren van de vierde generatie niet werden geselekteerd. De conclusie was dat schattingen van genetische parameters met het voorwaardelijke model niet onzuiver waren door selektie van basisdieren, maar een onzuiverheid werd geintroduceed indien nakomelingen van basisdieren werden geselekteerd.

De genetische variantie van melkproduktiekenmerken werd geschat met een voorwaardelijk model om te corrigeren voor selektie van stiervaders. In de HF deelpopulatie, waarin een intensievere selektie had plaatsgevonden, werd de genetische variantie 8% hoger ingeschat in vergelijking met een model dat random basis dieren aanneemt.

Schattingen voor erfelijkheidsgraden voor melkproduktie bleken hoog te zijn bij gebruik van een stiermodel, ook na correctie voor niet-additieve effecten en voor selectie. Voorlopige resultaten op basis van een diermodel, waarbij nog wordt gecorrigeerd voor selectieve inzet van stieren, bleken geen verlaging van de geschatte erfelijkheidsgraad op te leveren. Vervolgonderzoek moet meer duidelijkheid brengen omtrent het punt of een verhoogde erfelijkheidsgraad een genetische oorzaak heeft, dan wel door omgevingsomstandigheden wordt bepaald.

De belangsrijkste conclusies waren

- Modellen die niet corrigeren voor niet-additieve effecten bij de genetische evaluatie van gekruiste populaties veroorzaken een onzuiverheid in de schatting van fokwaarden en additief genetische verschillen tussen kruisingsgroepen. Gegevens van gekruiste populaties dienen daarom te worden gecorrigeerd voor systematische additieve en niet-additieve raseffecten.
- Schattingen van niet additieve parameters op basis van praktijkmateriaal leveren nauwkeurige schattingen op, hoewel correlaties tussen schattingen hoog kunnen zijn bij bepaalde datastructuren.
- Schattingen van genetische variantie die slechts gebaseerd zijn op data van geselekteerde generaties dieren, zijn onzuiver door deze selektie. Het niveau van de schatting wordt vooral bepaald door de veronderstelde basisgeneratie, en wordt weinig beïnvloed door het aantal volgende generaties waarvan gegevens bekend zijn. Opnemen van meer verwantschapsinformatie bij het schatten van de genetische variantie geeft een correctie voor covarianties tussen dieren, en in enige mate ook voor disequilibrium.
- Schattingen van genetische parameters met behulp van een voorwaardelijk model zijn niet onzuiver door selektie van basisdieren, maar er wordt een onzuiverheid geïntroduceerd indien nakomelingen van basisdieren geselekteerd zijn.
- Erfelijkheidgraden van melkproduktiekenmerken in gekruiste melkveepopulaties werden hoger bevonden dan de parameters die momenteel in de fokwaardschatting worden toegepast.

Curriculum Vitae

Julius Herman Jozef van der Werf werd op 4 maart 1960 geboren te Scharsterbrug (Friesland). In 1978 behaalde hij het Atheneum B diploma aan de Rijksscholengemeenschap te Heerenveen. In dat jaar begon hij met zijn studie Zoötechniek aan de Landbouwhogeschool te Wageningen. Het doctoraal examen werd afgelegd in september 1984, met als verzwaard hoofdvak de Veeteelt en als hoofdvak de Algemene Agrarische Economie. Na zijn afstuderen trad hij in dienst van de Landbouwhogeschool waar hij gedurende een jaar werkte aan een vooronderzoek naar de relatie van de fokrichting met de biologische en economische doelmatigheid van rundveehouderij. Sinds 1 december 1985 is hij werkzaam als Universitair Docent bij de vakgroep Veefokkerij van de Landbouwuniversiteit.