

Aus dem Forschungsbereich Genetik und Biometrie,
des Leibniz Institutes für Nutztierbiologie, Dummerstorf

Contributions to Methods Useful for Optimising Animal Breeding Plans

Dissertation

zur Erlangung des Doktorgrades
der Agrar- und Ernährungswissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Vinzent Börner, MSc.

aus Ueckermünde

Dekanin: Prof. Dr. Karin Schwarz
Erster Berichterstatter: Prof. Dr. Norbert Reinsch
Zweiter Berichterstatter: Prof. Dr. Georg Thaller

Tag der mündlichen Prüfung: 07.02.2011

Gedruckt mit Genehmigung des Dekans der Agrar- und Ernährungswissenschaftlichen
Fakultät der Christian-Albrecht-Universität Kiel

Contents

General Introduction	1
1 Gametic Gene Flow Method Accounts for Genomic Imprinting	9
1.1 Introduction	11
1.2 Methods	13
1.2.1 Components of the Gene Flow Method	13
1.2.2 Gametic Gene Flow Method	16
1.2.3 Inbreeding	20
1.2.4 Application to a Breeding Programme	21
1.3 Results	25
1.3.1 Genetic Net Present Values	27
1.3.2 Inbreeding	32
1.4 Discussion	33
1.4.1 Trait Realisation and Genetic Net Present Values	33
1.4.2 Inbreeding	35
1.5 Summary	36
1.6 Acknowledgement	36
1.7 Bibliography	37
1.8 Appendix 1	40

1.9	Appendix 2	41
2	Decorrelated Selection Indices versus Optimum Selection Indices in Optimising Multistage Dairy Cattle Breeding Schemes regarding Ge- nomic Selection	45
2.1	Introduction	47
2.2	Methods	50
2.2.1	Construction of Selection Indices and the Implementation of GEBV	50
2.2.2	Genetic Gain	51
2.2.3	Breeding Program	53
2.2.3.1	Breeding Costs	57
2.2.4	Parameter Variation	60
2.2.5	Maximisation	60
2.2.5.1	Combination of Stages	60
2.2.5.2	Dynamic Pedigree Information Content	62
2.2.6	Software	63
2.3	Results	63
2.3.1	Comparison of Methods for Calculating the Genetic Gain	63
2.3.2	Genetic Gain of Optimal Indices	72
2.4	Discussion	72
2.4.1	Comparison of the results for the decorrelated and optimal index	72
2.4.2	Results using the optimal index	77
2.5	Conclusions	77
2.6	Acknowledgement	78
2.7	Bibliography	78

3	Optimisation of Genomic Selection Dairy Cattle Breeding Schemes regarding different SNP-Chips and Imputing	81
3.1	Introduction	83
3.2	Methods	86
3.2.1	Construction of Selection Indices and the Implementation of GEBV	86
3.2.2	Genetic Gain	87
3.2.3	Breeding Program	88
3.2.4	Breeding Costs	92
3.2.5	Maximisation	95
3.2.6	Dynamic Pedigree Information Content	98
3.2.7	Breeding Scheme Similarity Indicator	98
3.2.8	Software	99
3.3	Results	99
3.3.1	Optimisation Results	99
3.3.2	Sensitivity of Maximisation Results to Changes in Accuracy and Cost . .	102
3.3.3	Sensitivity of Maximisation Results to Changes of Breeding Scheme Structures	107
3.4	Discussion	111
3.5	Conclusions	116
3.6	Acknowledgement	116
3.7	Bibliography	117
3.8	Appendix	119
	General Discussion	121
	Gametic Gene Flow Method	121
	Genomic Selection in Animal Breeding Plans	122
	Decorrelated Selection Indices in Multistage Selection Schemes	123

Applicability of Maximisation Techniques to Multistage Selection Schemes	124
Function of Genomic Selection in Multistage Dairy Cattle Breeding Programmes .	126
Biological and Technical Parameters of Dairy Cattle Breeding Program Optimisa- tions	127
Summary	129
Zusammenfassung	133
Bibliography	146
Appendix	147

General Introduction

The goal of breeding activities in commercial livestock populations is the increase of the mean of the genetically based performance capacity concerning one or numerous traits being summarised in the aggregate genotype via weighting factors. Budgeting breeding activities is carried out in order to predict this increase per generation and per year if a certain selection strategy is applied. Furthermore, transforming the genetic gain from natural units into monetary units due to using economic weights as weighting factors for the traits in the aggregate genotype, and relating it to the summarised costs of all the activities, the outcome in terms of the breeding profit can be used to evaluate the economics of a breeding scheme. Given a certain selection intensity, the genetic gain is a linear function of the accuracy of breeding value estimates, where this accuracy relies on two different parameters. First, the amount of information gathered about the selection candidate, where these information can be phenotypic measurements or genetic information (direct or marker genotypes) originating from the candidate itself or its relatives, and the correlations among information sources and with the aggregate genotype. Second, the capability of the applied statistical model to explain the genetic variation. Furthermore, since the amount of information is a function of economical resources as well as of time, budgeting breeding activities has to regard this relation when optimising selection schemes in order to maximise breeding profit.

The discovery of “genomic imprinting” challenged a major paradigm of animal breeding

assuming the effect of an allele on the descendants phenotype as independent of parent-of-origin. This biochemical mechanism modifies the expression of genes in an individual due to DNA methylation during gametogenesis in an individual's parents. This leads to a partial or complete deactivation of the effect of an affected gene on the phenotype (Reik *et al.*, 1987; Sapienza *et al.*, 1987). The pattern of this DNA methylation depends on the sex of the ancestor. Analyses in livestock found imprinted quantitative trait loci (QTL) and genes (*IGF2* in pigs, *Callipyge* in sheep) as well as significant parts of the genetic variance to be induced by genomic imprinting, where imprinting affected traits were related to growth and carcass quality (de Koning *et al.*, 2001a,b, 2000; de Vries *et al.*, 1994; Stella *et al.*, 2003). Thus, accounting for genomic imprinting in genetic evaluations for such traits leads to increased accuracy of breeding values and, therefore, a more realistic prediction of the genetic gain. A possible approach to account for genomic imprinting is the estimation of two different breeding values for each individual, the first is the breeding value of the individual if it acts as a dam and the second if it acts as a sire (Neugebauer *et al.*, 2010a,b).

The genetic gain of each trait in the aggregate genotype is affected by its economic weight, and in addition to that, by the time lag between the selection for a trait and the realisation of that trait. Since these time lags can be different for each trait in the breeding goal, and are strongly affected by the breeding scheme structure, they should be regarded in the breeding goal definition. The realised genetic superiority of trait realising individuals measured in trait units as a function of time can be derived via the gene flow method (Hill, 1974; McClintock & Cunningham, 1974), which is a Markov chain based algorithm modelling the distribution of desired genes across tiers of a population (e.g. nucleus, multiplier, production level) as a function of time and breeding scheme structure. The results of this method in terms of standard discounted expressions can be used as weighting factors for traits in the aggregate genotype having

the same unit (e.g. direct genetic and maternal genetic effect (Balcerzak *et al.*, 1989)), or as correction factors for economic weights (McClintock & Cunningham, 1974; Reinsch, 1995). As the selection candidate's breeding values as a sire and a dam are related to the same trait, they are characterised by the same unit (e.g. kg milk), and can be summarised by a weighted sum, where the respective weighting coefficients are the standard discounted expressions due to maternally and paternally inherited genes. In order to derive these weighting coefficients, the first part of this thesis deals with an extension of the gene flow method accounting for the probability that the genes of a selected individual are inherited to a trait realising individual via its dam or sire, and, therefore, derives standard discounted expressions for a trait in the aggregate genotype being partitioned into breeding values as sire and as dam.

As mentioned before, the accuracy of breeding values and, therefore, the genetic gain is also a function of the information gathered about the selection candidate, where this amount of information is mostly a function of time. Thus, a higher accuracy of estimated breeding values is often accompanied by an increased generation interval and might decrease the genetic gain per time unit. Budgeting breeding schemes elaborate on optimising this interaction, and possible solutions depend on the correlations among the available information and their correlation to the aggregate genotype. Although progeny performance as an information source is lengthening the generation interval always in all species, the correlation of all earlier information to the breeding goal may be so small that the marginal benefit of progeny testing on the genetic gain per time unit is still positive. This is especially the case for male selection in dairy cattle where selection candidates have no own performance.

With the discovery of hundreds of thousands of SNP markers in strong linkage disequilibrium with nearby QTL or genes, ideas rose and methods were developed in order to estimate highly accurate breeding values from individual marker genotypes (genomically

estimated breeding values) shortly after the birth of selection candidates (Meuwissen *et al.*, 2001). This was promising especially for dairy cattle breeders since overcoming the trade off between a shortened generation interval and a sufficiently accurate breeding value seemed feasible (Schaeffer, 2006). The selection of individuals on the basis of their genomically estimated breeding values can be seen as a preselection stage in order to reduce the number of test bulls in a multi stage progeny testing selection scheme, or, alternatively, as an information source on the selection candidate replacing progeny testing (Lillehammer *et al.*, 2010; Spelman *et al.*, 2010; Winkelman & Spelman, 2010). Furthermore, in conventional dairy cattle breeding schemes with progeny testing selection paths are only interrelated due to the allocation of limited financial resources if bull dams are performance tested in certain test stations, or if they are used for embryo transfer. In all other cases the overwhelming part of the breeding expenditures are focused on selecting sires. Genomic selection intensifies this interrelation because both males and females can be genotyped (König *et al.*, 2009; Schaeffer, 2006). Thus, optimisation of breeding schemes in terms of maximising the genetic gain per year given a certain investment has become much more complex since different kinds of one-, two or multistage selection can be applied in each selection path, but all are dependent on the same limited economical resource.

Calculations in order to find such breeding schemes can be carried out using deterministic or stochastic approaches. Both need to take account of the underlying sources of genetic and environmental variation (e.g. direct genetic, maternal genetic, imprinting, common environment, error). Stochastic simulations create each individual of the population and derive the genetic gain from the difference between the mean of known true breeding values of the selected individuals and the population mean. The major advantage of the stochastic approach is the possibility of calculating the variation of the genetic gain due to repeated simulations. On the contrary, deterministic calculations

do not regard individuals. Thus, the mean of the true breeding values of the selected individuals has to be derived from integrating the distribution of estimated breeding values. Since the computational burden is lower compared to stochastic simulations, it allows for more different breeding schemes to be evaluated. Additionally, if certain assumptions are made, multidimensional Newton method can be used to find breeding schemes producing the maximum genetic gain given the input parameters (Ducrocq & Colleau, 1989).

In multistage breeding schemes the selection indices of successive selection stages are correlated, requiring multidimensional integration techniques for the derivation of the selection intensity. The solutions from such algorithms were reported to be unstable and dependent on the correlation structure between stages, and, moreover, to be very time consuming for a higher number of stages (Ducrocq & Colleau, 1989; Saxton, 1989). In order to circumvent this problem, stage selection indices can be derived such that they are uncorrelated allowing for the calculation of the selection intensity at each stage via one-dimensional integration and the total genetic gain of a selection path as the sum over the genetic gains obtained at each selection stage (Xu & Muir, 1991, 1992). Although the maximisation of multistage breeding schemes using this approach is very fast and possible for an almost unlimited number of stages within each path, the imposed constraint of uncorrelatedness among indices at successive stages causes a loss in the variance of the estimated breeding values and, therefore, a loss in predicted genetic gain (Xu & Muir, 1992). Since new developments for the integration of multivariate normal distributions yield stable and valid results independent of the correlation structure (Genz, 1992), the second part of this thesis investigates the extend of the loss in predicted genetic gain due to using decorrelated selection indices compared to optimal selection indices if both methods are applied to a multistage multi-path dairy cattle breeding scheme with genomic selection and under the complex interactions between selection paths resulting

from cost restriction and path specific selection.

As already mentioned, the tremendous impact of genomic selection on dairy cattle breeding schemes is based on shortening the generation interval together with a sufficiently accurate estimated breeding value. For the selection paths “sire-dam” and “sire-sire” this accuracy is reported to be still below that from progeny testing for milk traits, but the genetic gain per year is predicted to be higher (Hayes *et al.*, 2009b; Schaeffer, 2006). On the contrary, the accuracy of genomically estimated breeding values for females in the path “dam-sire” might be even higher than the conventional one since the selection of bull dams relied mainly on information about ancestors and own performance. Thus, in this path genomic selection might increase the genetic gain per year in two ways, due to a more accurate breeding value and due to shortening the generation interval (Schaeffer, 2006). The major obstacle for realising such a scenario is based on the interrelation between breeding costs, the demand for bull dams and the selection intensity in that path. Since the number of selection candidates in the path “sire-dam” dictates, via reproduction parameters, the demand for bull dams, this number might be as high that in order to achieve a suitable selection intensity, and, therefore, a genetic gain, a sufficient proportion of a high number of initial selection candidates has to be genotyped leading to high investments, and, possibly, an overdraw of limited economical resources. In this case - and even under the assumption of unlimited economical resources - the additional genetic gain from genotyping potential bull dams will not justify the financial efforts. A possible way out of this dilemma is genotyping individuals with low-cost SNP chips carrying only a small subset of SNPs used on high density SNP panels (Habier *et al.*, 2009). SNP genotypes from such chips can be used differently according to its design. If the SNPs at this chips are a subset of those being used by a more expensive type, and consists of evenly spaced SNPs, so called imputation algorithms can be applied in order to calculate the most probable high density genotype for

each individual, which already has been genotyped with a low density SNP panel (Howie *et al.*, 2009; VanRaden, 2010; Weigel *et al.*, 2010). Despite of a certain loss in accuracy, such techniques are expected to allow for applying genomic selection in the situation outlined above. As the accuracy of genomically estimated breeding values from high density and low density SNP genotyping as well as the future costs of such SNP chips is under steady development, the third part of this thesis, in an optimisation approach, investigates how a variation of these accuracies and cost will affect the genetic gain per year in a multistage dairy cattle breeding scheme where both low and high density SNP panels are used, and all selection paths are linked together by restricted overall breeding costs.

1 Gametic Gene Flow Method Accounts for Genomic Imprinting

Gametic Gene Flow Method Accounts for Genomic Imprinting

Vinzent Börner & Norbert Reinsch
Leibniz Intsitut for Farm Animal Biology, FBN
Research Unit Genetic and Biometry
18196 Dummerstorf, Germany

Abstract

Findings within the last fifteen years emphasise the possible role of genomic imprinting for trait expression in livestock species. In genetic evaluation, genomically imprinted traits can be treated by models with two different breeding values per animal; one accounts for the paternal and the other for the maternal expression pattern. Relative weighting factors for these breeding values were derived by a generalised version of the discounted gene flow method, which was extended to a gametic level to account for parent-of-origin effects. The gametic approach proved also useful for calculating the expected increase in inbreeding induced by one round of selection and its dynamics over time. The gametic gene flow method was applied to a hypothetical pig breeding programme. Relative weighting factors were higher for the paternally inherited genetic effect even in female selection paths, but depend on the breeding scheme heavily. The maximum medium-term increase in inbreeding due to selection exceeded the long term increase in a range of 20 % to 100 %.

Keywords: economic weights, parent-of-origin effects, breeding plan, inbreeding

1.1 Introduction

The economic success of commercial breeding programmes depends heavily on the transfer of genes of selected breeding animals to the production stage where traits of interest are realised.

Until the 1980's it was a paradigm that the expression of genes in offspring is independent of parent-of-origin. First indications that this is not the case arose from transgenic mouse experiments (Reik *et al.*, 1987; Sapienza *et al.*, 1987). A process called genomic imprinting (Reik *et al.*, 1987) was found to be responsible for a total or partial silencing of gene expression through DNA methylation during gametogenesis. The possible evolutionary advantage of genomic imprinting was summarised in the parental conflict theory (Haig, 2004; Wilkins & Haig, 2003). According to this theory, imprinting affected genes are expected to be prenatal growth related, whereby maternally inherited alleles tend to be inactivated if they promote growth and activated if they retard growth. Together with an opposite male deactivation/activation pattern, genomic imprinting is assumed to protect maternal resources against excessive exploitation through paternally inherited growth promoting alleles of progeny. The DNA methylation during gametogenesis depends on the parental sex and is reversible only during gametogenesis in offspring (Whitehead, 2004, pp.32). Thus, if the underlying genes are polymorphic and related to an economically important trait, genomic imprinting is of special interest for breeders. First indications about the possible role of genomic imprinting in livestock came from variance component estimations in pigs describing significant parts of the phenotypic variance as induced by imprinted loci (de Vries *et al.* (1994); in a magnitude of roughly 10% for growth and backfat thickness). Also in pigs, variance component estimation of litter size suggests that paternal imprinting has an effect on this trait (Stella *et al.*, 2003). An analysis of cattle data (Engellandt & Tier, 2002) found significant variance

components due to genomic imprinting for fattening traits (meat content, kidney fat weight, pelvic fat weight). Beside these quantitative approaches, quantitative trait loci (QTL) and genes coding for traits of economic relevance have been found to be influenced by genomic imprinting. In sheep, the so called "*Callipyge*" phenotype, exhibiting an abnormal growth of the economically interesting hind quarters, is caused by an imprinting affected single base polymorphism (Cockett *et al.*, 1999; Freking *et al.*, 2002). In pig, imprinted QTL were found for muscle depth, meat quality, growth rate and backfat thickness (de Koning *et al.*, 2001a,b, 2000). Furthermore, a genomic imprinting affected nucleotide substitution explaining high proportions of phenotypic variances of muscle mass and fat deposition traits was mapped in an intron of the *IGF2* gene (Van Laere *et al.*, 2003). Imprinting may also play a role for economically important traits in poultry (Tuiskula-Haavisto *et al.*, 2004).

According to these findings, it is may be necessary to account for genomic imprinting in defining the breeding goal and in genetic evaluation. A flexible and general approach might be the simultaneous calculation of two breeding values for a single animal by an extended BLUP-model (Neugebauer *et al.*, 2010; Reinsch & Guiard, 2005), whereby one reflects the breeding value of the animal as a dam and the other of the animal as a sire. When the breeding values as a sire and a dam for a number of animals are written as respective vectors a_s and a_d , then their covariance is

$$Var \begin{pmatrix} a_s \\ a_d \end{pmatrix} = \begin{pmatrix} A\sigma_s^2 & A\sigma_{sd} \\ A\sigma_{sd} & A\sigma_d^2 \end{pmatrix} = A \begin{pmatrix} \sigma_s^2 & \sigma_{sd} \\ \sigma_{sd} & \sigma_d^2 \end{pmatrix} \quad (1.1)$$

and the imprinting variance can be expressed as $\sigma_s^2 + \sigma_d^2 - 2\sigma_{sd}$. In the case of Mendelian inheritance both effects coincide, expressed by having the same variance and being perfectly correlated. Thus, the imprinting variance is zero. In the case of genomic imprinting the correlation is below one and, in some cases, the additive genetic variance can not be

portioned equally among sire and dam. As this leads to an individual-specific difference between the conventional breeding value and the aggregation of these two parent-of-origin dependent breeding values, the ranking of the selection candidates might change. However, for calculating a total merit index out of these two breeding values, economic weights for both are a necessary prerequisite.

Economically derived weighting factors evaluate the marginal benefit of increasing the trait realisation by one unit (Groen, 1989). Weighting the traits of interest in an overall breeding value by these weight factors implies the assumption of realising all traits of interest with the same time lag, which is an oversimplification of real production circumstances. The discounted gene flow method (Hill, 1974; McClintock & Cunningham, 1974) can be used to account for these time lags by using the discounted relative trait realisations as correction factors for the economic weights (Balcerzak *et al.*, 1989).

The aim of our work was to derive relative economic weights for both breeding values of a genomic imprinting affected trait. For this purpose, the gene flow method (Hill, 1974) was extended to a gametic level and applied to a hypothetical pig breeding programme. Additionally, it is demonstrated that this approach is also useful for modeling the loss of heterozygosity by a single round of selection in populations with overlapping generations.

1.2 Methods

1.2.1 Components of the Gene Flow Method

This section gives a brief introduction to the gene flow method and the notation we use in this article. For a more detailed description and an example without and with selection paths the reader is referred to Hill (1974).

A breeding population can be divided into classes (C) of animals. The main char-

acteristics of these classes are sex ($s = 1[\textit{male}], 2[\textit{female}]$) and use ($u = 1, \dots, i$ where i is the number of possibilities of usage according to the breeding scheme, e.g. breeding sows, production sows). These sex-use-classes (C_{su}) can be divided into age-classes ($a = 1, \dots, j$ where j is the number of age classes according to the breeding scheme). Thus, the total number of sex-use-age-classes (C_{sua}) within the population depends on the number of different C_{su} 's and the number of age-classes existing at the same time within these C_{su} 's. The number of age-classes within the C_{su} 's can be different for every C_{su} and is a function of the breeding scheme. Animals enter the youngest C_{sua} when they are born. As only one round of selection and the following spread of genes of the selected animals within the population over time is considered, the selected individuals are called founders (\tilde{F}). If more than one animal is selected as an \tilde{F} , the selection process can take place in different C_{su} 's (e.g. breeding boars, breeding sows). The origin of the time-scale is the time of birth of the selected \tilde{F}_{sua} . To distinguish between groups of \tilde{F} 's selected from different C_{su} 's (and therefore, different C_{sua} 's) the abbreviation is extended to \tilde{F}_{sua} . Every \tilde{F}_{sua} is the origin of one or many selection paths xy , where x is the selected group (\tilde{F}_{sua} itself) and y is the target group (e.g. boars to sows, $x=\textit{boars}$, $y=\textit{sows}$). If a \tilde{F}_{sua} founds a certain xy , it is denoted as $\tilde{F}_{sua.xy}$. Every $\tilde{F}_{sua.xy}$ can consist of one or many individuals selected from the same C_{sua} .

According to the gene flow method (Hill, 1974; McClintock & Cunningham, 1974), a matrix P can be derived mirroring the population structure mentioned above and its reproductive system. P consists of reproduction rows and aging rows. Reproduction rows contain the probabilities that a new born C_{sua} inherits genes from the C_{sua} 's already existing. Aging rows contain the probability that a C_{sua} existing from time point $t - 1$ to t inherits genes from itself by aging, which is one. Thus, all rows of P must sum up to one. Mathematically, P is a Markov chain transmission matrix. A matrix Q containing only aging probabilities and a matrix E_{xy} containing only reproduction probabilities can

be derived from P for every selection path xy . Thus, E_{xy} contains only the reproduction row of a certain selection path (e.g. boars to boars) and all the other cells are zero. All rows of E_{xy} sum up to zero except the selection path specific reproduction row which sums up to 0.5.

The vector n_{xy_0} reflects the distribution of genes of the $\tilde{F}_{sua_{xy}}$ over C_{sua} 's of the population at time zero. Therefore, it contains only zeros except a one at the position of the C_{sua} from which the $\tilde{F}_{sua_{xy}}$ were selected. The vector m_{xy_0} is a null vector. The vector m_{xy_t} reflects the proportion of the genome of the $\tilde{F}_{sua_{xy}}$ on the genome of the C_{sua} 's after t time units due to the selection path xy . It is calculated by the following equations:

$$n_{xy_t} = Q \times n_{xy_{t-1}} \quad (1.2)$$

$$m_{xy_t} = E_{xy} \times n_{xy_{t-1}} + P \times m_{xy_{t-1}}. \quad (1.3)$$

The realisation vector r_k represents the proportion of every C_{sua} on the total realisation of trait k . These proportions are the numbers of trait realising animals in a certain C_{sua} divided by the total number of trait realising animals. Thus, r_k has the same dimension as m_{xy_t} and its elements sum up to one. The realisation $R_{xy_{kt}}$ of the trait k at a certain time point t due to the selection path xy is calculated according to:

$$R_{xy_{kt}} = m'_{xy_t} \times r_k. \quad (1.4)$$

Thus, $R_{xy_{kt}}$ is the weighted sum of the genome proportions of the $\tilde{F}_{sua_{xy}}$ in the trait realising C_{sua} 's having arrived there due to selection path xy . The weights are the corresponding values in the r_k vector. The genetic net present value g_{xy_k} as with interest

rate p capitalised realisation of trait k due to the selection path xy is calculated as

$$g_{xy_k} = \sum_{t_0}^{t_{max}} R_{xy_{kt}} \times \frac{1}{(1+p)^t}, \quad (1.5)$$

where p is the interest rate and t_{max} is the planning horizon. These g_{xy_k} 's are the correction factors for the economic weights, which are necessary to calculate an overall selection index out of the breeding values of k traits.

To evaluate the importance of different \tilde{F}_{sua} 's for the trait realisation of the population, the genetic net present values of all selection paths being founded by the same \tilde{F}_{sua} have to be summarised. This can be made by using m_{x_t} vectors, which are calculated as follows:

$$m_{x_t} = \sum_y m_{xy_t}. \quad (1.6)$$

Thus, m_{x_t} vectors are a cumulation of the m_{xy_t} vectors of all selection paths xy being founded by the same \tilde{F}_{sua} (e.g. all selection paths being founded by primary selected breeding boars) The realisation of trait k at time point t calculated with m_{x_t} vectors is denoted by $R_{x_{kt}}$. The genetic net present value calculated with m_{x_t} vectors is denoted by g_{x_k} .

1.2.2 Gametic Gene Flow Method

Every individual inherits genes from its sire and dam. As the C_{sua} 's are assemblies of individuals of the same sex, use and age, C_{sua} 's can be divided into two sex-use-age-gamete-classes (C_{suag}); one class for all paternally inherited genes (the male C_{suag} where $g = 1$) and one class for all maternally inherited genes (the female C_{suag} where $g = 2$). Even the genome of the whole population at a certain time point t can be divided into

two gamete classes (C_g); the gamete class of male sex as the sire derived part ($C_g, g = 1$) of the population genome, and the gamete class of female sex as the dam derived part ($C_g, g = 2$) of the population genome. Thus, genes of a certain $\tilde{F}_{sua.xy}$ can occur in C_1 and C_2 due to selection path xy after t time units.

In each case 50% of the genes of an individual are sire-of-origin and 50% are dam-of-origin. The paternally derived genes are a mixture of the genes of the sire's sire and the sire's mother. Thus, the genes of the C_{sua1} of a new born C_{sua} are a mixture of the genes of its paternal grandparents. The same applies to the C_{sua2} of a new born C_{sua} .

Taking these extensions into account, a transmission matrix P^* can be derived on the gametic level. Thus, P^* is of double dimension as P . Every C_{sua} is divided into two C_{sua_g} 's. P^* has twice the number of reproduction rows as P . The values within the reproduction rows represent the transmission probabilities of the genome of the C_{sua1} and C_{sua2} of already existing C_{sua} 's to the C_{sua1} and C_{sua2} of a new born C_{sua} . All rows of P^* must sum up to one. The matrices Q^* and E_{xy}^* are derived from P^* as explained for the non-gametic model.

A vector $n_{xy_0}^*$ of double dimension as n_{xy_0} is set up containing only zeros except a one at each of both C_{sua1} and C_{sua2} positions of the C_{sua} from which the founders were selected. The vectors $m_{xy_0}^*$ and $m_{xy_t}^*$ are of double dimension as their counterparts in the non-gametic model, but fulfill the same functions. $m_{xy_t}^*$ vectors are calculated as follows:

$$n_{xy_t}^* = Q^* \times n_{xy_{t-1}}^*, \quad (1.7)$$

$$m_{xy_t}^* = E_{xy}^* \times n_{xy_{t-1}}^* + P^* \times m_{xy_{t-1}}^*. \quad (1.8)$$

In contrast to the non-gametic model, an $m_{xy_t}^*$ vector represents the proportion of the genome of a certain founder group $\tilde{F}_{sua_{xy}}$ on the genome of the sex-use-age-gamete-classes C_{sua_g} after t time units due to the selection path xy . Thus, it allows a statement about the probability that a certain gene taken from the $\tilde{F}_{sua_{xy}}$ gene pool has arrived in a C_{sua} and is inherited from its direct male or female ancestors.

The extension of the gene flow method also affects the realisation vector r_k . Instead of one realisation vector r_k , two vectors r_{kg} are used, one for every C_g . This is a prerequisite for calculating the realisation of trait k out of the genome proportions of a $\tilde{F}_{sua_{xy}}$ in the paternally (C_1) and maternally (C_2) derived gamete class due to selection path xy separately. The dimension of the r_{kg} vectors are twice the dimension of r_k . The values within the r_{kg} vectors are derived mainly in the same way as explained for the non-gametic model. Additionally, they are weighted with 0.5. Thus, the elements of r_{kg} sum up to 0.5. The vector r_{k1} for calculating the weighted sum of genome proportions of $\tilde{F}_{sua_{xy}}$ in the C_{sua1} 's contains zeros at the positions where the $m_{xy_t}^*$ vectors contain the values of the C_{sua2} 's. The same applies to r_{k2} being used to calculate the weighted sum of genome proportions of $\tilde{F}_{sua_{xy}}$ in the C_{sua2} 's. It contains zeros at positions where the $m_{xy_t}^*$ vectors contain the values of the C_{sua1} 's.

The realisation $R_{xy_{ktg}}$ of the trait k due to the genome proportions of $\tilde{F}_{sua_{xy}}$ having arrived at C_g via selection path xy at a certain time point t is calculated as:

$$R_{xy_{ktg}} = m_{xy_t}^{*'} \times r_{kg}. \quad (1.9)$$

Thus, the $R_{xy_{kt}}$ of the non-gametic model is divided into two $R_{xy_{ktg}}$'s according to the sex of the direct parents of which the trait realising animals inherited the genome proportions of $\tilde{F}_{sua_{xy}}$.

The genetic net present value $g_{xy_{kg}}$ as with interest rate p capitalised $R_{xy_{ktg}}$ values is

derived by the formula

$$g_{xy_{kg}} = \sum_{t_0}^{t_{max}} R_{xy_{ktg}} \times \frac{1}{(1+p)^t}, \quad (1.10)$$

where t_{max} is the planning horizon and p is the interest rate. These $g_{xy_{kg}}$'s are the weighting coefficients for the breeding values of trait k of the individuals as a sire and a dam. They are necessary to calculate a combined breeding value for trait k . The sum of these $g_{xy_{kg}}$'s is the correction factor for the economic weight of trait k in the selection index.

The evaluation of the relative importance of a certain \tilde{F}_{sua} for the trait realisation of the population can be calculated as explained above. Thus, all $m_{xy_t}^*$ vectors of a certain \tilde{F}_{sua} have to be combined to $m_{x_t}^*$ vectors as in (1.6). The realisations of trait k at time point t calculated with $m_{x_t}^*$ vectors are denoted by $R_{x_{ktg}}$. The genetic net present values calculated with $m_{x_t}^*$ vectors are denoted by $g_{x_{kg}}$.

As we assume an additive infinitesimal model, the trait realisation depends on the additive effects of numerous genes and its alleles. Thus, the increase in trait realisation within a certain C_{sua} at time point t is the weighted sum of the genome proportions of $\tilde{F}_{sua_{xy}}$ within the C_{sua1} and C_{sua2} of this C_{sua} . This implies that the gametic model can also be used for non-gametic calculations just by the addition of both r_{kg} vectors and the weighted sum of the C_{sua1} and C_{sua2} position of all C_{sua} positions of the $m_{xy_t}^*$ vectors. When doing so, the results of the gametic model are equal to those calculated with the non-gametic model. Therefore, the gametic model is a generalisation of the non-gametic model.

1.2.3 Inbreeding

As the cumulated genome proportions of a group of founder individuals \tilde{F}_{sua} in the C_{sua} 's due to calculation of $m_{x_t}^*$ vectors are known, a vector $\Delta F_{t_{\tilde{F}_{sua}}}$ representing the average increase of inbreeding coefficients of every C_{sua} at time unit t can be calculated. The values of $\Delta F_{t_{\tilde{F}_{sua}}}$ reflect the increase in inbreeding at time point t due to the selection of a certain \tilde{F}_{sua} . If all individuals within a certain group of founders \tilde{F}_{sua} are not related, are not inbred and have the same chance of reproduction, $\Delta F_{t_{\tilde{F}_{sua}}}$ is calculated by

$$\Delta F_{t_{\tilde{F}_{sua}}} = m_{x_{t_1}}^* \cdot m_{x_{t_2}}^* \times \frac{1}{2} \times \frac{1}{N_{\tilde{F}_{sua}}}, \quad (1.11)$$

where $m_{x_{t_1}}^*$ is a vector containing only the C_{sua1} values and $m_{x_{t_2}}^*$ is a vector containing only the C_{sua2} values of the $m_{x_t}^*$ vector in time unit t . Thus, the dimension of $m_{x_{t_1}}^*$ and $m_{x_{t_2}}^*$ is half the dimension of the $m_{x_t}^*$ vector. $N_{\tilde{F}_{sua}}$ is the number of individuals of the \tilde{F}_{sua} . Note that $m_{x_{t_1}}^* \cdot m_{x_{t_2}}^*$ is the Hadamard product of the mentioned vectors. If more than one \tilde{F}_{sua} has been selected initially (e.g. from new born breeding boars and breeding sows), the vectors $\Delta F_{t_{\tilde{F}_{sua}}}$ must be calculated for every \tilde{F}_{sua} . A generalisation for calculating $\Delta F_{t_{\tilde{F}_{sua}}}$ accounting for unequal breeding probabilities of individuals of a \tilde{F}_{sua} is given in the Appendix 1. The overall increase in inbreeding due to one round of selection at time point t is

$$\Delta F_t = \sum_{\tilde{F}_{sua}} \Delta F_{t_{\tilde{F}_{sua}}}. \quad (1.12)$$

Thus, ΔF_t is just a cumulation of the $\Delta F_{t_{\tilde{F}_{sua}}}$ representing the increase in inbreeding due to selection of the different founder groups.

If P^* is derived in a way that generations are discrete (being born, breeding use and slaughtering of an animal is within one time unit), the results of this algorithm become

equivalent to the formula of Wright (1931)

$$\Delta F = \frac{1}{8N_m} + \frac{1}{8N_f}, \quad (1.13)$$

where ΔF is the increase in inbreeding from one generation to the next depending on the number of selected males, N_m , and females, N_f . In contrast to the formula of Wright (1931), the gametic gene flow method allows to account for complex breeding situations and gives an insight into the fluctuation of inbreeding over time. A more general application of the gene flow method to estimate the increase in inbreeding is given by Johnson (1977). It is useful to model the development of the kinship coefficient between all groups of animals within a population due to multiple cycles of selection, but does not allow for quantifying the contribution of a certain group of animals to the overall increase in inbreeding if the planning horizon exceeds the live-span of this group.

1.2.4 Application to a Breeding Programme

For visualising the extensions of the matrix P^* , the $m_{x_t}^*$ vectors and the r_{kg} vector, and for calculating the effect on $g_{xy_{kg}}$'s, we considered a hypothetical pig breeding programme of the following structure: A breeding stage for producing breeding boars and breeding sows to reproduce and produce boars for the production stage. A production stage where production sows, mated with breeding boars, reproduce and produce piglets for fattening, and a fattening stage where piglets of both sexes are fattened. All animals are pure-bred. The discrete time unit is six months. The first offspring at breeding, production and fattening stage is born when boars and breeding and production sows enter the third time period. Boars are in the fourth time period when their last offspring is born. Breeding and production sows are in the fifth time period when their last offspring is born. The survival rate of boars and sows after maturity is 75%. The traits

of interest are litter size and backfat thickness. Litter size exemplifies a reproductive trait with repeated realisation by females. On the contrary, backfat thickness exemplifies a carcass trait realised only once at slaughter. The realisation of litter size occurs in females at breeding (10%) and production stage (90%), whereas backfat thickness is realised only by fattening animals. The length of the planning horizon is either 5, 10, 15 and 30 time units. The interest rate is 5% per year and, therefore, roughly 2.5% per time unit. Based on the given parameters, we derived P^* with respect to gametic inheritance as displayed in table 1.1. Given that backfat thickness is realised at fattening stage, the r_{kg} vectors for this trait contain only zeros except at C_{sua1} and C_{sua2} position of fattening stage. Trait realisations due to genes in C_{sua1} results in a zero at the C_{sua2} position, whereas trait realisations due to genes in C_{sua2} results in a zero at the C_{sua1} position. As reproduction traits are characterised by multiple realisations (e.g. litter numbers, lactation numbers) the r_{kg} values are derived with respect to realisation probabilities of reproduction cycles. The realisation vectors derived with respect to trait and C_{sua} are shown in table 1.2.

Table 1.1: Transmission matrix P^* .

C_{su}	boars										breeding sows										production sows										fatten- ing									
	1		2		3		4		1		2		3		4		5		1		2		3		4		5		♂	♀										
C_{sua}	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀		
1	0	0	0	0	0.29	0.29	0.21	0.21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.22	0.22	0.16	0.16	0.12	0.12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
5	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
breed- ing sows	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀		
1	0	0	0	0	0.29	0.29	0.21	0.21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
pro- duct- ion sows	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀		
1	0	0	0	0	0.29	0.29	0.21	0.21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
fatten- ing	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀		
1	0	0	0	0	0.29	0.29	0.21	0.21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

C_{su} : sex-use-classes, C_{sua} : sex-use-age-classes, C_{sua} : sex-use-age-gamete-classes

Table 1.2: Vector $n_{xy_0}^*$ for the sire derived selection pathways and trait/gamete class dependent realisation vectors r_{kg} .

	boars					breeding sows					production sows					fattening	
	1	2	3	4	1	2	3	4	5	1	2	3	4	5	1	1	
vector	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	
$n_{xy_0}^*$	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
backfat thickness,	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
r_{k2}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
backfat thickness,	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
r_{k1}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
litter size,	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
r_{k2}	0	0	0	0	0	0	0	0	0	0	0	0	0.043	0	0.032	0	
litter size,	0	0	0	0	0	0	0	0	0	0	0	0	0.043	0	0.032	0	
r_{k1}	0	0	0	0	0	0	0	0	0	0	0	0	0.024	0	0	0	
litter size,	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
r_{k1}	0	0	0	0	0	0	0	0	0	0	0	0	0.389	0	0.292	0	
litter size,	0	0	0	0	0	0	0	0	0	0	0	0	0.292	0	0.219	0	
r_{k1}	0	0	0	0	0	0	0	0	0	0	0	0	0.219	0	0	0	

r_{k1} : realisation vector of the male (paternal) gamete class, r_{k2} : realisation vector of the female (maternal) gamete class

1.3 Results

Table 1.3 displays the vectors $m_{x_0}^*$ to $m_{x_{20}}^*$ for the cumulated sire derived selection paths. Every vector reflects the distribution of the genes of the selected founding sire group over all C_{sua_g} 's via all sire derived selection paths after t time units. The values in these vectors mirror the proportion of the founding sire group genome on the genome of the C_{sua_g} 's of a certain C_{sua} . Thus, the average of C_{sua1} and C_{sua2} of a certain C_{sua} is the overall proportion of the founder group genome on the genome of a certain C_{sua} in t . Therefore, the results are equivalent to the results of the non-gametic model. Up to time unit five, due to the specificity of the breeding programme, all genes of the male founder animals occur in C_{sua1} 's. At time unit six, 16.3% of the genome of new born piglets in every C_{su} are descended from the founder group genome, but 43% of this genome occur in C_{sua2} 's and 57% in C_{sua1} 's. In later time units, the proportion of the founder genome in C_{sua2} 's increases up to 50% and in certain combinations of C_{sua} 's and time units to more than 50% (e.g. C_{sua} "Boars 1" in time unit eight). After 20 time units, the differences between the values in C_{sua1} 's and C_{sua2} 's are approximately zero. Thus, the probabilities of inheriting parts of the founder genome from a sire or a dam are almost equal. As expected from the non-gametic model, these proportions are equal if calculation is carried out on infinite number of time units (not shown).

Table 1.3: Vectors $m_{x_0}^*$ to $m_{x_{20}}^*$ for the cumulated sire derived selection paths.

	C_{su}	C_{stua}	C_{stuaq}	time units																				
				0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
boars	1	♂	0	0	0	0.571	0.429	0	0.163	0.245	0.092	0.082	0.184	0.158	0.105	0.135	0.159	0.135	0.126	0.143	0.144	0.134	0.136	
		♀	0	0	0	0	0	0	0.124	0.185	0.139	0.114	0.14	0.155	0.132	0.13	0.144	0.141	0.134	0.137	0.141	0.138	0.136	
		♂	0	0	0	0	0.571	0.429	0	0.163	0.245	0.092	0.082	0.184	0.158	0.105	0.135	0.159	0.135	0.126	0.143	0.144	0.134	
		♀	0	0	0	0	0	0	0	0.124	0.185	0.139	0.114	0.14	0.155	0.132	0.13	0.144	0.141	0.134	0.137	0.141	0.138	
breeding sows	2	♂	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
		♀	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		♂	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		♀	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
pro-duction sows	3	♂	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
		♀	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		♂	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		♀	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
fattening	4	♂	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
		♀	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		♂	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		♀	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
fattening	5	♂	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
		♀	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		♂	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		♀	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
fattening	1	♂	0	0	0	0	0	0	0.571	0.429	0	0.163	0.245	0.092	0.082	0.184	0.158	0.105	0.135	0.159	0.135	0.126	0.143	0.136
		♀	0	0	0	0	0	0	0	0	0	0.124	0.185	0.139	0.114	0.14	0.155	0.132	0.13	0.144	0.141	0.138	0.136	
		♂	0	0	0	0	0	0	0	0	0	0.163	0.245	0.092	0.082	0.184	0.158	0.105	0.135	0.159	0.135	0.126	0.143	0.134
		♀	0	0	0	0	0	0	0	0	0	0.124	0.185	0.139	0.114	0.14	0.155	0.132	0.13	0.144	0.141	0.138	0.136	
fattening	2	♂	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
		♀	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
		♂	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
		♀	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
fattening	3	♂	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
		♀	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
		♂	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
		♀	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
fattening	4	♂	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
		♀	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
		♂	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
		♀	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
fattening	5	♂	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
		♀	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
		♂	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
		♀	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

C_{su} : sex-use-classes, C_{stua} : sex-use-age-classes, C_{stuaq} : sex-use-age-gamete-classes, length of one time unit: 6 months

Figures 1.1(a) and 1.1(b) illustrate the trait realisation values $R_{x_{ktg}}$ due to the cumulated sire derived selection paths subject to trait and planning horizon. As expected from the breeding scheme and, therefore, possible gene flow, the fluctuations of $R_{x_{ktg}}$ values within the first 10 time units are enormous. Up to time unit five, backfat thickness is realised only due to genes in gamete class C_1 , starting in time unit three, reaching a maximum in time unit four and decreasing to zero in time unit six. After time unit six, $R_{x_{kt1}}$ and $R_{x_{kt2}}$ values of backfat thickness are almost equal, but a stable level is not reached before time unit 20. The $R_{x_{ktg}}$ values of litter size show a similar pattern as backfat thickness, but the time lag until realising the trait due to genes in gamete class C_2 holds on until time unit eight. Litter size realisation due to genes in C_1 starts in time unit five, increases sharply until time unit seven and decreases to the maximum level of $R_{x_{kt2}}$. In later stages of realisation, there are almost no differences between $R_{x_{kt1}}$ and $R_{x_{kt2}}$ of litter size. This results from the fact that the proportions of the founder group genome inherited by a sire or a dam are equal for every C_{sua} in later time units (exemplified by the last column of table 1.3).

1.3.1 Genetic Net Present Values

Table 1.4 shows the genetic net present values $g_{xy_{kg}}$ of backfat thickness classified by selection paths for different planning horizons. The $g_{xy_{k1}}$'s and $g_{xy_{k2}}$'s could be calculated for cumulative selection paths, where the first one was always higher than the second one. However, generally $g_{xy_{kg}}$'s of these selection paths increase with time.

The picture is different for non-cumulative selection paths. Due to the breeding scheme, there are no $g_{xy_{k2}}$'s for the sire-fattening pig selection path independent of planning horizon. The constant $g_{xy_{k1}}$ values of this selection path only result from using the selected founding boars as sires for fattening pigs during their third and fourth time period of life. The sire-production dam path allows no gene flow of paternally inherited

genes to the fattening stage. Thus, $g_{xy_{k_1}}$'s of this selection path must always be zero (see figures 1.2(a) and 1.2(b)).

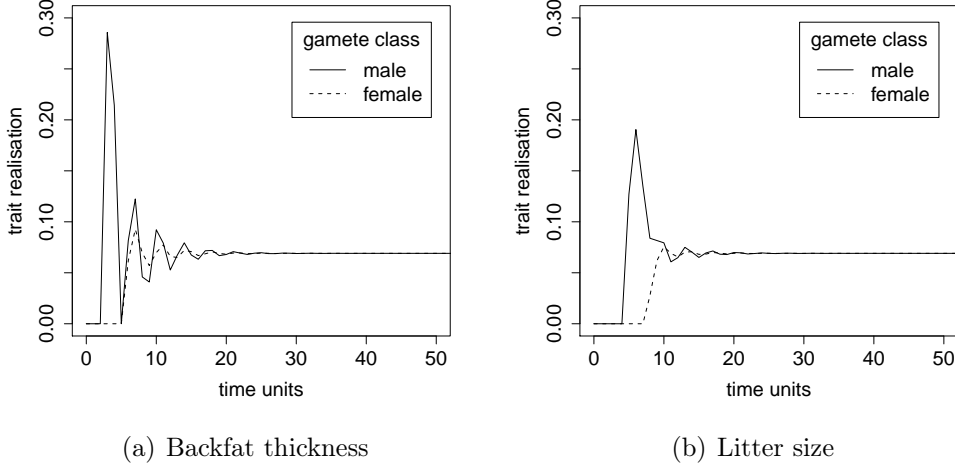
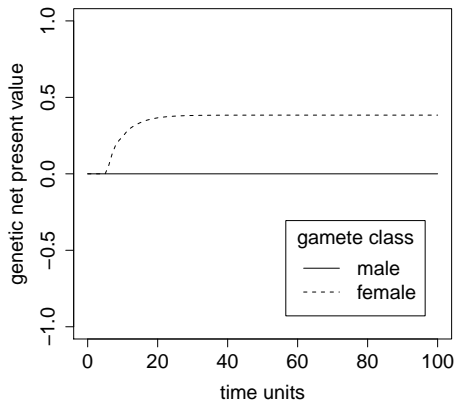


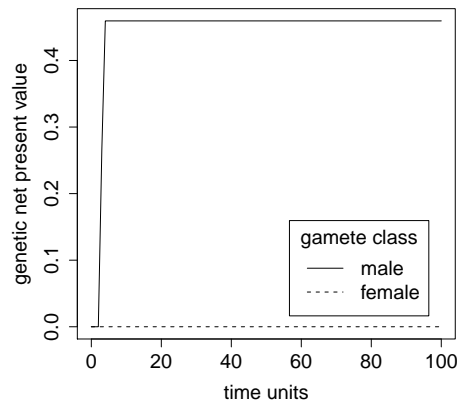
Figure 1.1: Trait realisation subject to planning horizon and gamete class for the cumulated sire derived selection paths.

Figures 1.2(c) and 1.2(d) show plots of the $g_{x_{k_g}}$'s of backfat thickness for the cumulated sire derived selection paths and cumulated breeding dam derived selection paths. The founder genome proportions in C_1 are more important for both because they arrive earlier at trait realising animals. The maternally inherited founder genome arrives at trait realising animals with a certain time lag. This creates a sustained difference in genetic net present values even if the planning horizon is unlimited (results not shown). As the $g_{x_{k_g}}$'s increase asymptotically in time, the difference between both can be regarded as fixed if the planning horizon is long enough. Considering backfat thickness, this fixed difference between the $g_{x_{k_1}}$ and the $g_{x_{k_2}}$ is about 20% for both cumulated selection paths.

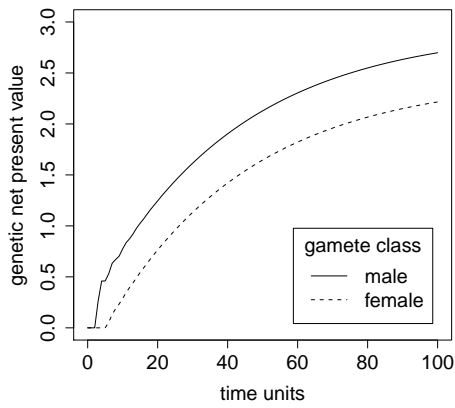
Comparing cumulated selection paths, asymptotical genetic net present values of backfat thickness due to cumulated breeding dam derived selection paths are always lower than due to the cumulated sire derived selection paths. This is explained by specificities of the breeding programme (a breeding dam is not producing fattening pigs).



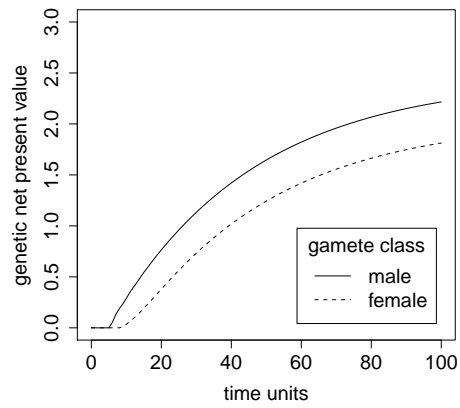
(a) Backfat thickness - sire to production dam



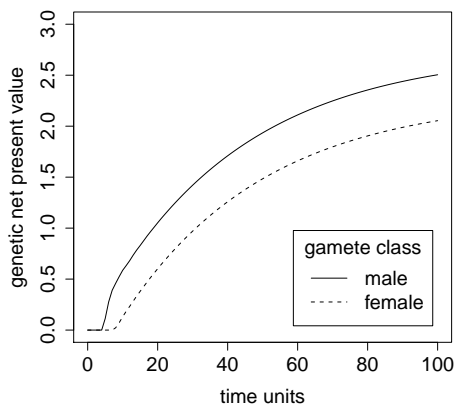
(b) Backfat thickness - sire to fattening pig



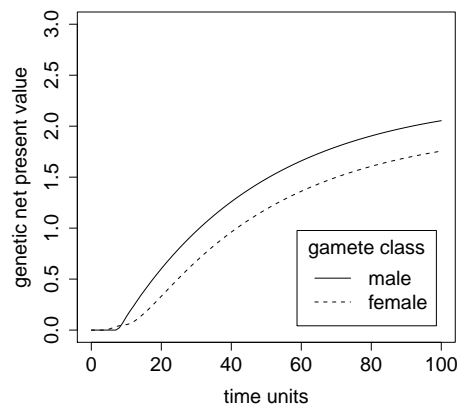
(c) Backfat thickness - cumulated sire derived paths



(d) Backfat thickness - cumulated breeding dam derived paths



(e) Litter size - cumulated sire derived paths



(f) Litter size - cumulated breeding dam derived paths

Figure 1.2: Genetic net present values subject to trait, selection paths, planning horizon and gamete class.

Table 1.5 shows the $g_{xy_{kg}}$'s of litter size classified by selection paths for different planning horizons. As litter size is realised in reproductive females, $g_{xy_{kg}}$'s of the selection path sire-fattening pig are always zero independent of the planning horizon. The $g_{x_{kg}}$'s are generally lower compared to backfat thickness. This results from the time lag in realisation and its lower fluctuations within the first time units because the long term trait realisation of litter size is equal to backfat thickness (see figures 1.1(a) and 1.1(b)). Illustrated by figures 1.2(e) and 1.2(f), the $g_{x_{kg}}$'s out of C_1 are always higher than out of C_2 except for the cumulated breeding dam derived selection paths in the first time units. The shape of the curves and the differences between the cumulated selection paths are almost the same as of backfat thickness.

Table 1.4: Genetic net present value of backfat thickness subject to selection path, planning horizon and gamete class (C_g).

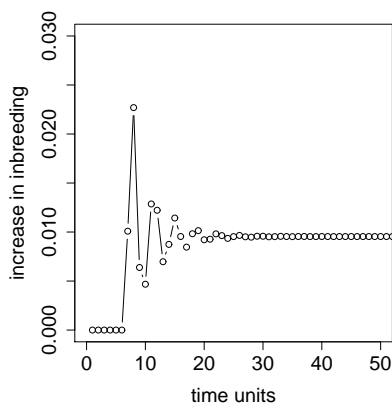
selection path	planning horizon in time units									
	10		15		20		25		30	
	gamete class (C_g)									
	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂
sire-sire	0.05	0.27	0.17	0.39	0.28	0.5	0.38	0.6	0.47	0.69
sire-breeding dam	0	0.05	0.04	0.17	0.12	0.28	0.2	0.38	0.28	0.47
sire-production dam	0.24	0	0.33	0	0.37	0	0.38	0	0.38	0
sire-fattening pig	0	0.46	0	0.46	0	0.46	0	0.46	0	0.46
breeding dam -breeding dam	0	0.03	0.03	0.16	0.11	0.27	0.19	0.37	0.28	0.46
breeding dam-sire	0.03	0.25	0.16	0.38	0.27	0.49	0.37	0.59	0.46	0.68
cumulated sire derived	0.29	0.78	0.54	1.03	0.76	1.25	0.96	1.44	1.13	1.62
cumulated breeding dam derived	0.03	0.29	0.19	0.54	0.38	0.76	0.56	0.96	0.73	1.13

Table 1.5: Genetic net present value of litter size subject to selection path, planning horizon and gamete class (C_g).

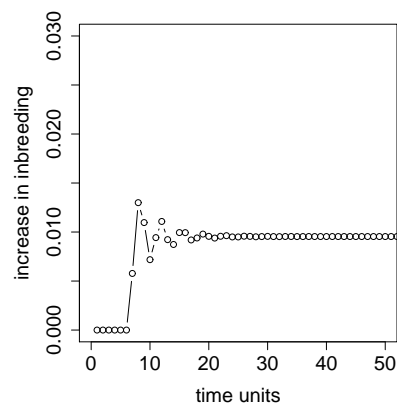
selection path	planning horizon in time units									
	10		15		20		25		30	
	gamete class (C_g)									
	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂
sire-sire	0	0.15	0.1	0.3	0.21	0.41	0.31	0.51	0.39	0.6
sire-breeding dam	0.01	0.04	0.03	0.14	0.1	0.25	0.18	0.35	0.26	0.43
sire-production dam	0.12	0.39	0.25	0.39	0.3	0.39	0.31	0.39	0.32	0.39
breeding dam-breeding dam	0.05	0	0.07	0.09	0.13	0.2	0.21	0.3	0.29	0.38
breeding dam-sire	0	0.13	0.09	0.29	0.2	0.4	0.3	0.5	0.38	0.59
cumulated sire derived	0.13	0.58	0.38	0.83	0.6	1.05	0.8	1.25	0.97	1.42
cumulated breeding dam derived	0.05	0.13	0.16	0.38	0.33	0.6	0.51	0.8	0.67	0.97

1.3.2 Inbreeding

Figures 1.3(a) and 1.3(b) illustrate the development of the inbreeding coefficient of new born boars due to founding boars and founding breeding sows when only one animal was initially selected. As expected from the genome distribution reflected by the $m_{x_t}^*$ vectors, the inbreeding reaches a maximum in time unit eight and then is stabilised at roughly 1%. The long term increase of the inbreeding coefficient induced by selecting founding boars is equal to the increase induced by selecting founding breeding sows. The short-term increase within the first 10 time units often exceeds the long term increase considerably. The peak caused by founding boars exceeds the long term increase by more than 100%, whereas the peak caused by founding breeding sows exceeds the long term increase by about 20%.



(a) Primary selected boars



(b) Primary selected breeding sows

Figure 1.3: Inbreeding coefficient subject to planning horizon for new born boars due to different primary selected animals.

1.4 Discussion

The results clearly show that the gene flow via male ancestors is much faster than via female ancestors independent of the sex of the primary selected animal group. This leads to a higher genetic net present value for trait realisation out of paternal inherited genome proportions. Additionally, the maximum increase of the inbreeding coefficient caused by selecting males exceeds its long term increase much more than when it is caused by selecting females.

1.4.1 Trait Realisation and Genetic Net Present Values

In general, our suggestions about extending the gene flow method (Hill, 1974) offer the possibility of gaining knowledge about the sex-dependent origin of the genes of primary selected animals when they occur in a certain C_{sua} . If we assume an imprinting affected trait and an imperfect correlation between the breeding values of an individual as sire and as dam, changes in the economic weights for these breeding values change the ranking of the individuals in an overall breeding value.

The results from our hypothetical breeding scheme clarify that the breeding values of an individual as a sire or a dam can not always be equally weighted. In cross-breeding schemes, it is quite clear which breeding value is of interest. Depending on the breeding goal of the used breed, only breeding values as a sire (in the sire line) or a dam (in the dam line) are relevant for ranking animals. In pure-breeding schemes with discrete generations, males and females of one generation equally contribute to the next generation. In this case, breeding values of individuals as a sire or a dam have to be equally weighted. Apart from that, in complex nucleus breeding schemes the gene flow of primary selected animals to trait realising C_{sua} 's depends heavily on the special construction of the scheme. The complexity of such breeding schemes can lead to

counter-intuitive results. This is exemplified by our hypothetical breeding scheme. The spread of the genes of the selected breeding dams via their sons is much faster than via their daughters. Thus, the breeding values of the breeding dams as sires (the genes of these dams reach trait realising animals via a male progeny) should be weighted higher than their breeding values as dams. Additionally, sires selected in the nucleus are used for producing trait realising C_{sua} 's and its female precursors. Due to this construction, the males have breeding values as a sire and a dam which are not realised at the same time. As almost every breeding scheme has its own specificities, computations are necessary to derive the exact values of realisation for both breeding values.

Independent of the special construction of the scheme, how the weights should be chosen is also a question of the length of the planning horizon. From a theoretical point of view breeding programmes are of long duration. Therefore, variations from the long term realisation during the first time units are of little interest, and weights for both breeding values can be equal. In practice, breeding programmes have to fulfil short term economic demands as well. Thus, correct estimation of trait realisation and, therefore, different weights for both breeding values might clarify also the short-term perspective of such programmes. However, if correct weighting factors are almost equal and the genetic correlation between the breeding value as a sire and the breeding value as a dam is high, then equal weighting may lead to very much the same selection response as correct weighting.

An increase in the interest rate decreases the difference between the genetic net present values. Thus, the decision how to weight both breeding values must also account for the future development of the interest rate.

For the deterministic prediction of genetic gain, even for new developments like genomic breeding value estimation, the selection index theory (Dekkers, 2007; Villanueva *et al.*, 1993) and the selection path model (Rendel & Robertson, 1950) are still the methods

of choice. Under genomic imprinting it can be shown that the genetic gain due to such a programme depends on the correlation between the genetic effect as sire and as dam, the proportion of the additive genetic variance explained by each genetic effect and the economic weighting factors. For a detailed outline see the Appendix B.

1.4.2 Inbreeding

The deterministic prediction of inbreeding for breeding schemes relies still on the formula of Wright (1931) or upon prediction of the long term contribution of an ancestor (Woolliams & Bijma, 2000). The later copes with overlapping generations, different selective advantages of breeding animals and deviations from Poisson-distributed family size. However, these predictions give no impression about short-term fluctuations of inbreeding. A third approach merging gene flow method and kinship coefficient was developed by Johnson (1977). It allows to assess the increase of the average kinship coefficient within and between selection groups even in complex breeding schemes and for multiple cycles of selection with the limitation that the ratio of the number of animals within the selection groups and the planning horizon must be sufficiently large.

As now the probability that a C_{sua} receives the same genes from its sire and dam is calculable, fluctuations of inbreeding coefficients can be modeled for every C_{sua} and the whole planning horizon without being limited to certain number of individuals or number of time units. The results of our calculations show clearly that increase in inbreeding coefficients of special combinations of C_{sua} 's and time unit can exceed average increase inbreeding coefficients by far. Thus, a possible inbreeding depression might decrease the real trait realisation especially in early time units.

In contrast to Johnson (1977), our approach models only one cycle of selection. Merging both might be a way to overcome the limitations inherent in both approaches.

In conclusion, relative weighting factors for parent-specific additive genetic effects can be derived appropriately with a gametic version of the discounted gene flow method. Moreover, gametic gene flow may also be useful for predicting short-term fluctuations of inbreeding coefficients.

1.5 Summary

Findings within the last fifteen years emphasise the possible role of genomic imprinting for trait expression in livestock species. In genetic evaluation, genomically imprinted traits can be treated by models with two different breeding values per animal; one accounts for the paternal and the other for the maternal expression pattern. Relative weighting factors for these breeding values were derived by a generalised version of the discounted gene flow method, which was extended to a gametic level to account for parent-of-origin effects. The gametic approach proved also useful for calculating the expected increase in inbreeding induced by one round of selection and its dynamics over time. The gametic gene flow method was applied to a hypothetical pig breeding programme. Relative weighting factors were higher for the paternally inherited genetic effect even in female selection paths, but depend on the breeding scheme heavily. The maximum medium-term increase in inbreeding due to selection exceeded the long term increase in a range of 20 % to 100 %.

1.6 Acknowledgement

This work was financed by by the German Federal Ministry for Education and Research (BMBF), project "FUGATO-Brain".

1.7 Bibliography

- Balcerzak, K. M., Freeman, A. E. & Willham, R. L. (1989), Selection for the direct and maternal genetic effects for dystocia in holsteins. *J Dairy Sci* **72**(5), 1273–1279.
- Cockett, N. E., Jackson, S. P., Snowden, G. D., Shay, T. L., Berghmans, S., Beever, J. E., Carpenter, C. & Georges, M. (1999), The callipyge phenomenon: evidence for unusual genetic inheritance. *J Anim Sci* **77 Suppl 2**, 221–227.
- de Koning, D., Harlizius, B., Rattink, A., Groenen, M., Brascamp, E. & van Arendonk, J. (2001a), Detection and characterization of quantitative trait loci for meat quality traits in pigs. *J Anim Sci* **79**(11), 2812–2819, URL [PM:11768109](#).
- de Koning, D., Rattink, A., Harlizius, B., Groenen, M., Brascamp, E. & van Arendonk, J. (2001b), Detection and characterization of quantitative trait loci for growth and reproduction traits in pigs. *Livest Prod Sci* **72**(3), 185–198, URL [ISI:000172280000001](#).
- de Koning, D., Rattink, A., Harlizius, B., van Arendonk, J., Brascamp, E. & Groenen, M. (2000), Genome-wide scan for body composition in pigs reveals important role of imprinting. *Proc Natl Acad Sci U S A* **97**(14), 7947–7950, URL [PM:10859367](#).
- de Vries, A. G., Kerr, R., Tier, B., Long, T. & Meuwissen, T. (1994), Gametic imprinting effects on rate and composition of pig growth. *Theor Appl Genet* **88**, 1037–1042.
- Dekkers, J. C. M. (2007), Prediction of response to marker-assisted and genomic selection using selection index theory. *J Anim Breed Genet* **124**(6), 331–341, URL <http://dx.doi.org/10.1111/j.1439-0388.2007.00701.x>.
- Engelland, T. & Tier, B. (2002), Genetic variances due to imprinted genes in cattle. *J Anim Breed Genet* **119**, 154–165, URL <http://dx.doi.org/10.1046/j.1439-0388.2002.00323.x>.

- Freking, B. A., Murphy, S. K., Wylie, A. A., Rhodes, S. J., Keele, J. W., Leymaster, K. A., Jirtle, R. L. & Smith, T. P. L. (2002), Identification of the single base change causing the callipyge muscle hypertrophy phenotype, the only known example of polar overdominance in mammals. *Genome Res* **12**(10), 1496–1506, URL <http://dx.doi.org/10.1101/gr.571002>.
- Groen, A. F. (1989), Economic values in cattle breeding. i. influences of production circumstances in situations without output limitations. *Livest Prod Scie* **22**, 1–16.
- Haig, D. (2004), Genomic imprinting and kinship: how good is the evidence? *Annu Rev Genet* **38**, 553–585, URL [PM:15568986](https://pubmed.ncbi.nlm.nih.gov/15568986/).
- Hill, W. G. (1974), Prediction and evaluation of response to selection with overlapping generations. *Anim Prod* **18**, 117–139.
- Johnson, D. L. (1977), Inbreeding in populations with overlapping generations. *Genetics* **87**(3), 581–591.
- McClintock, A. & Cunningham, E. (1974), Selection in dual-purpose cattle populations: Defining the breeding objective. *Anim Prod* **18**, 237–247.
- Neugebauer, N., Räder, I., Schild, H. J., Zimmer, D. & Reinsch, N. (2010), Evidence for parent-of-origin effects on genetic variability of beef traits. *J Anim Sci* **88**(2), 523–532, URL <http://dx.doi.org/10.2527/jas.2009-2026>.
- Reik, W., Collick, A., Norris, M. L., Barton, S. C. & Surani, M. A. (1987), Genomic imprinting determines methylation of parental alleles in transgenic mice. *Nature* **328**(6127), 248–251, URL <http://dx.doi.org/10.1038/328248a0>.
- Reinsch, N. & Guiard, V. (2005), Mixed mendelian and imprinted inheritance, unpublished.

-
- Rendel, J. M. & Robertson, A. (1950), Estimation of genetic gain in milk yield by selection in a closed herd of dairy cattle. *J Genet* **50**, 1–8.
- Sapienza, C., Peterson, A. C., Rossant, J. & Balling, R. (1987), Degree of methylation of transgenes is dependent on gamete of origin. *Nature* **328**(6127), 251–254, URL <http://dx.doi.org/10.1038/328251a0>.
- Stella, A., Stalder, K., Saxton, A. & Boettcher, P. (2003), Estimation of variances for gametic effects on litter size in yorkshire and landrace swine. *J Anim Sci* **81**(9), 2171–2178, URL [PM:12968691](https://pubmed.ncbi.nlm.nih.gov/12968691/).
- Tuiskula-Haavisto, M., de Koning, D.-J., Honkatukia, M., Schulman, N. F., Mäki-Tanila, A. & Vilkki, J. (2004), Quantitative trait loci with parent-of-origin effects in chicken. *Genet Res* **84**(1), 57–66.
- Van Laere, A.-S., Nguyen, M., Braunschweig, M., Nezer, C., Collette, C., Moreau, L., Archibald, A. L., Haley, C. S., Buys, N., Tally, M., Andersson, G., Georges, M. & Andersson, L. (2003), A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature* **425**(6960), 832–836, URL <http://dx.doi.org/10.1038/nature02064>.
- Villanueva, B., Wray, N. & Thompson, R. (1993), Prediction of asymptotic rates of response from selection on multiple traits using univariate and multivariate best linear unbiased predictors. *Anim Prod* **57**, 1–13.
- Whitehead, J. (2004), *Genomic imprinting in development and evolution*. Ph.D. thesis, Uppsala University, Faculty of Science and Technology.
- Wilkins, J. F. & Haig, D. (2003), What good is genomic imprinting: the function of parent-specific gene expression. *Nat Rev Genet* **4**(5), 359–368, URL <http://dx.doi.org/10.1038/nrg1062>.

Woolliams, J. A. & Bijma, P. (2000), Predicting rates of inbreeding in populations undergoing selection. *Genetics* **154**(4), 1851–1864.

Wright, S. (1931), Evolution in mendelian populations. *Genetics* **16**(2), 97–159.

1.8 Appendix 1

As the genes in the C_{sua1} and C_{sua2} of a C_{sua} can be from different animals of \tilde{F}_{sua} , one has to assess the probability that these genes are identical by descent. Let x be a row vector of the dimension $N_{\tilde{F}_{sua}}$ containing the reproduction probabilities of the individuals of a certain \tilde{F}_{sua} . D is a diagonal matrix of the dimension $N_{\tilde{F}_{sua}} \times N_{\tilde{F}_{sua}}$ containing the kinship coefficients of the founder individuals with itself. Assuming that the founders are not related, the overall probability that two alleles taken at random from the gene pool of \tilde{F}_{sua} are identical by descent (p) is calculated as follows:

$$p = tr((x'x) \cdot D). \quad (1.14)$$

Note that $(x'x) \cdot D$ is the Hadamard product of the matrices $(x'x)$ and D . Following these derivations, the average increase of the inbreeding coefficient $\Delta F_{t_{\tilde{F}_{sua}}}$ of a certain C_{sua} due to the selection of \tilde{F}_{sua} is

$$\Delta F_{t_{\tilde{F}_{sua}}} = m_{x_{t_1}}^* \cdot m_{x_{t_2}}^* \times p, \quad (1.15)$$

where $m_{x_{t_1}}^*$ is a vector containing only the C_{sua1} values and $m_{x_{t_2}}^*$ is a vector containing only the C_{sua2} values of the $m_{x_t}^*$ vector in time unit t . Thus, their dimensions are half the dimension of the $m_{x_t}^*$ vector.

1.9 Appendix 2

The genetic gain (ΔH) can be approximated applying selection index theory by

$$\begin{aligned}
 \Delta H &= i r_{TI} \sigma_T \\
 &= i \sigma_I \\
 &= i \sqrt{b' P b} \\
 &= i \sqrt{(P^{-1} G w)' P P^{-1} G w} \\
 &= i \sqrt{w' G' P^{-1}' P P^{-1} G w} \\
 &= i \sqrt{w' G' P^{-1}' G w}
 \end{aligned} \tag{1.16}$$

where i is the selection intensity, r_{TI} is the correlation between the Index I and the aggregated genotype T , σ_T is the standard deviation of T , σ_I is the standard deviation of I , b is a vector of regression coefficients regressing the index value of a selection candidate on the phenotypic measurements of the informants, P is the covariance matrix of all phenotypic informants, G is the genotypic covariance matrix between the selection candidate and all the informants and w is a vector with economic weights. In a breeding programme selecting only on one trait, G is a vector and w is a scalar and, therefore, dispensable. Now assume a breeding programme selecting only on one trait and accounting for genomic imprinting. Selection candidates are supposed to be not inbred. The column dimension of G and the row dimension of w extend to two.

As outlined in the introduction, the methylation pattern of the DNA in the gametes forming a progeny depends on the sex of the parents. Thus, the Co-ancestry between the selection candidate and all its relatives can be divided into two categories. For simplicity, it is assumed that the selection candidate doesn't have an own performance.

The category (1) is made from the sire, all direct ancestors of the sire (e.g. grand-sire, grand-dam), all indirect relatives of the selection candidate which are connected to

it's pedigree by a final male gene transmission (e.g. daughters of a half-brother) and all descendants of the selection candidate having a male as it's direct ancestor (e.g. progeny if the candidate is a male, progeny of a son).

The category (2) is made from the dam, all direct ancestors of the dam (e.g. grandsire, grand-dam), all indirect relatives of the selection candidate which are connected to it's pedigree by a final female gene transmission (e.g. daughters of a half-sister) and all descendants of the selection candidate having a dam as it's direct ancestor (e.g. progeny if the candidate is a female, progeny of a daughter).

The covariance between the candidates breeding value as a sire and all individuals belonging to category (1) is the variance of the additive genetic effect when acting as a sire times the numerator relationship, whereas the covariance between the candidate's breeding value as a sire and all individuals in category (2) is the covariance between the additive genetic effects when acting as a sire and when acting as a dam times the numerator relationship.

Now imagine that the row order of G is the following: the first part is occupied by category (1) and the second by category (2). The column order of G is as follows: the first column is used for the breeding value as a sire and the second for the breeding value as a dam. Due to categorizing the relatives and the candidate, G can be partitioned into four sub-matrices G_{11}, G_{12}, G_{21} and G_{22} (1.17).

$$\begin{aligned}
 G_{11} &= n_{i_1} \sigma_{a_m}^2, & G_{11} &= n_{i_1} c_m^2 \sigma_a^2, \\
 G_{12} &= n_{i_1} \sigma_{a_m a_f}, & G_{12} &= n_{i_1} c_m c_f \sigma_a^2, \\
 G_{21} &= n_{i_2} \sigma_{a_f}^2, & G_{21} &= n_{i_2} c_f^2 \sigma_a^2, \\
 G_{22} &= n_{i_2} \sigma_{a_f a_m}, & G_{22} &= n_{i_2} c_m c_f \sigma_a^2.
 \end{aligned}
 \tag{1.17}
 \tag{1.18}$$

In (1.17) n_{i_1} and n_{i_2} are the numerator relationships between the individual and the informant i in category 1 and category 2 respectively, $\sigma_{a_m}^2$ is the additive genetic variance

due to acting as a sire, $\sigma_{a_f}^2$ is the additive genetic variance due to acting as a dam and σ_{ama_f} is the covariance between both. $\sigma_{a_m}^2$, $\sigma_{a_f}^2$ and σ_{ama_f} can be expressed in terms of σ_a^2 by applying correction factors such that $c_m^2\sigma_a^2$ is equal to $\sigma_{a_m}^2$, $c_f^2\sigma_a^2$ is equal to $\sigma_{a_f}^2$ and $c_m c_f \sigma_a^2$ equals σ_{ama_f} . Thus, (1.17) changes to (1.18).

Now, G can be partitioned into a matrix N with dimension equal to G , a matrix C and a matrix [1] of dimension 2×2 , and a scalar σ_a^2 . N and C contain the values:

$$\begin{array}{ll}
 N_{11} = n_{i_1}, & C_{11} = c_m, \\
 N_{12} = 0, & C_{12} = 0, \\
 N_{21} = 0, & C_{21} = 0, \\
 N_{22} = n_{i_2}, & C_{22} = c_f,
 \end{array} \tag{1.19} \tag{1.20}$$

and [1] contains only ones. Thus, G can now be written as:

$$G = NC[1]C\sigma_a^2 \tag{1.21}$$

The values in w are derived from gametic gene flow method as described above. Assuming that the correction factor for the breeding value as a sire (w_s) is higher than the factor for the breeding value as a dam (w_d), w can be transformed into a vector a relating all values in w to w_s . Thus, at the first position of a is a one and at the second position is a coefficient below one. Now, the whole formula for calculating ΔH can be expressed as:

$$\begin{aligned}
 \Delta H &= i\sqrt{w_s a' (NC[1]C\sigma_a^2)' P^{-1} NC[1]C\sigma_a^2 a w_s} \\
 &= i\sigma_a^2 w_s \sqrt{a' (NC[1]C)' P^{-1} NC[1]C a}
 \end{aligned} \tag{1.22}$$

The above formula clearly shows that, in contrast to the non imprinting accounting model, ΔH depends additionally on the economic weights of both breeding values, the amount of σ_a^2 ascribed to each and the correlation between both. As ΔH is overestimated

if both breeding values are weighted equally with the factor w_m , it is recommended to use the correct weighting factors.

2 Decorrelated Selection Indices versus Optimum Selection Indices in Optimising Multistage Dairy Cattle Breeding Schemes regarding Genomic Selection

Decorrelated Selection Indices versus Optimum Selection Indices in Optimising Multistage Dairy Cattle Breeding Schemes regarding Genomic Selection

Vinzent Börner & Norbert Reinsch
Leibniz Institute for the Biology of Farm Animals, FBN
Research Unit Genetics and Biometry
18196 Dummerstorf, Germany

Abstract

Decorrelated selection indices have been used for the optimisation of multistage breeding schemes. However, they underestimate the genetic gain to an unforeseeable extent and have not been applied to breeding schemes with different generation intervals and selection intensities in each selection path. As the occurrence of genomic selection makes the re-evaluation of breeding programs with special regard to multistage selection necessary, a grid search was applied in order to identify optimum breeding plans maximising the genetic gain per year in a multistage multipath dairy cattle breeding program, in which the accuracy of genomically estimated breeding values and its cost per individual were varied over an semi-continuous parameter space and the total breeding costs were restricted. Results were obtained from applying decorrelated indices as well as from optimal selection indices. In comparison to the results obtained by using the optimal indices, decorrelated indices underestimated the genetic gain up to 40% and the proportion of individuals undergoing genomic selection was different. Additionally, the interaction of selection paths was counter-intuitive and difficult to forecast when using decorrelated indices. Independent of the applied method, genomic selection replaced traditional progeny testing when maximising the genetic gain per year as long as the accuracy of genomically estimated breeding values was ≥ 0.45 .

2.1 Introduction

Today, genomic selection (GS) offers breeders the possibility to reduce costs, decrease the generation interval (Schaeffer, 2006) and possibly avoid inbreeding (Daetwyler *et al.*, 2007). GS is based on predicting breeding values from observable individual genotypes as soon as the genotype is formed. These genotypes consists of large number of DNA markers in form of single nucleotide polymorphism (SNP) being in linkage disequilibrium to quantitative trait loci coding for economical important traits.

In dairy cattle, the accuracies (r) of breeding values for milk production traits estimated on this basis are reported to be as high as 0.75 (Hayes *et al.*, 2009b). Although the accuracy of the genomically estimated breeding values (r_{GEBV}) for this trait is below those from progeny testing, the tremendously decreased generation interval (GI) may lead to a much higher genetic gain per year (ΔG_a) (Meuwissen *et al.*, 2001; Schaeffer, 2006). In a detailed calculation, Schaeffer (2006) summarized the potential effects of GS on dairy cattle breeding schemes assuming a r_{GEBV} of 0.75 and costs for GS per animal (C_{GEBV}) of 500 USD in a one-stage selection approach. Since r_{GEBV} and C_{GEBV} may change, research work concerning optimum breeding schemes for different combinations of this parameters and possible multistage selection approaches are still of interest.

Beside the pure substitution of progeny testing by GS with a given single accuracy, as a variety of SNP-chips for different costs yielding breeding values of different accuracies is available now, it is also feasible to use special SNP-chips in each selection path or more than one chip in a multistage preselection approach. Additionally, the advantage of GS might also be combined with traditional progeny testing schemes as currently applied in certain dairy cattle breeding programs (Spelman *et al.*, 2010). Thus, beside the already answered questions about the applicability of GS, the answer to the question how a breeding scheme should be structured has become tremendously complex because of

the given possibility of combining different information sources according to their costs and correlations to the breeding goal, resulting in a variety of one-, two- or multistage breeding schemes for each selection path to choose from.

It is a question of economical optimisation to select the breeding scheme maximising a defined utility function. As in multistage breeding schemes the information about the selection candidates gathered at all previous stages are combined with those gathered at the current stage, the selection indices of successive stages are correlated. Thus, one of the major challenges for optimisation research in multistage breeding schemes is the necessity of using computational sophisticated multiple integration techniques to derive the selection intensities of the selection stages with regard to non normally distributed breeding values. In an application of such methods on multistage dairy cattle breeding schemes Ducrocq & Colleau (1989) faced the problem that the convergence of such algorithms is difficult to achieve if the correlation between stages is high. Additionally, the computational time was seen as unacceptable if the number of selection stages becomes too high (Saxton, 1989).

A possible solution to circumvent these problems is the decorrelation of the stage selection indices, proposed by Xu & Muir (1991, 1992). Then, stage selection indices can be assumed to be normally distributed allowing for an easy calculation of the selection intensities and the use of maximisation techniques like multidimensional Newton method. Xu & Muir (1992) amounted the loss in genetic gain due to the decorrelation in a two stage breeding scheme up to 10% compared to the exact calculation solving the integral. This loss was justified by the possibility of implementing an unlimited number of stages which is otherwise not applicable.

However, to our knowledge decorrelated indices have been used for model calculations of breeding schemes for poultry (Hicks *et al.*, 1998; Xu & Muir, 1992), beef cattle (Xu *et al.*, 1995) and marker assisted selection (Xie & Xu, 1998), where in all cases for the

modelled selection paths of a particular breeding plan identical accuracies of breeding values, selection intensities and generation intervals were assumed. Thus, the applicability of the decorrelated indices to complex breeding schemes regarding different selection paths and its interaction due to effects of the selection strategy in one path on the accuracy of breeding values of the other paths, different selection intensities and breeding value accuracies in each path, the interaction between the generation interval and the number and character of the selection stages and, finally, the opportunity to split the investments between selection path, has not been investigated in detail.

Since numerical integration techniques developed more recently by Genz (1992) allow for a fast and stable calculation of the exact selection intensities making the implementation of many stages and the comparison of a huge number of different breeding schemes in an optimisation process feasible, the aim of our work was to compare the results of breeding scheme optimisations when the approach of Xu & Muir (1992) (decorrelated indices) or numerical integration (optimal indices) was used for deriving selection intensities, breeding value accuracies and the genetic gain. Both methods were applied to a dairy cattle breeding scheme regarding the complexity given above, and including the possibility of using GS as an optional selection stage in a way that it can be used for preselection or as the only information source. Optimisation was over semi-continuous parameter space of r_{GEBV} and its costs per genotyped individual, and economical resources were restricted. Therefore, the results furthermore provide insights into the sensitivity of dairy cattle breeding plans to a variation in these parameters.

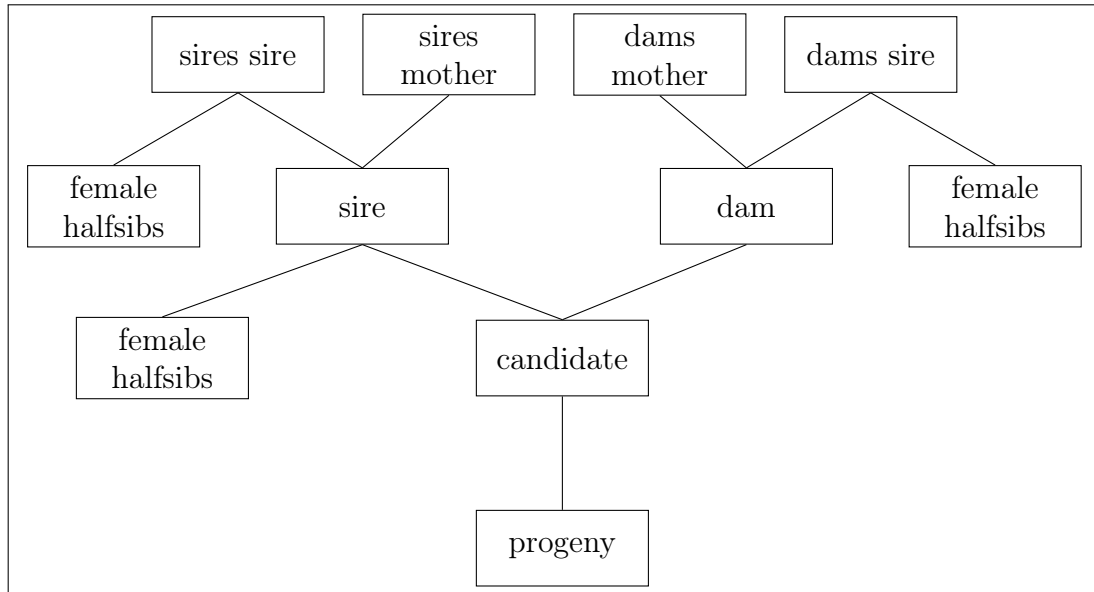


Figure 2.1: Standard pedigree used to derive breeding values in all selection paths

2.2 Methods

2.2.1 Construction of Selection Indices and the Implementation of GEBV

For calculating the reliability of breeding values selection index methodology was used based on the pedigree given in figure 2.1.

Two different methods were used to derive the regression coefficients for information sources, which then were used to calculate variances, covariances and accuracies of estimated breeding values.

For the optimal selection indices (*OSI*), these regression coefficients were constructed fulfilling the following constraint:

$$\sigma_{(i,j),T} = w' \times G_{i,j} \times b_{i,j} \rightarrow \max \quad (2.1)$$

where w is the vector of economic weights of the traits in the breeding goal, $\sigma_{(i,j),T}$ is the

covariance between the estimated and true breeding value, $b_{i,j}$ is the vector of regression coefficients for all information sources being available, and $G_{i,j}$ is the genetic-phenotypic covariance matrix at selection stage i of selection path j .

The decorrelated selection indices (*DSI*) were constructed maximising $\sigma_{(i,j),T}$ too (Xu & Muir, 1992), but regarding the marginal constraint that the covariances of the considered stage to all previous stages equals zero:

$$\begin{aligned}\sigma_{(i,j),(i-1,j)} &= b'_{i,j} \times P_{(i,j),(i-1,j)} \times B_{i-1,j} = 0 \quad (i > 1) \\ \sigma_{i,j}^2 &= b'_{i,j} \times P_{(i,j),(i,j)} \times b_{i,j} = 1 \quad (i > 1)\end{aligned}\tag{2.2}$$

where $P_{(i,j),(i,j)}$ is the phenotypic covariance matrix and $\sigma_{i,j}^2$ is the variance of the breeding values at stage i of path j , $\sigma_{(i,j),(i-1,j)}$ is a vector of covariances, $P_{(i,j),(i-1,j)}$ is the phenotypic covariance matrix between the stage i and all previous stages and $B_{i-1,j}$ is a matrix of regression coefficients of all stages being previous to i within path j . The second marginal condition assures that a solution exists.

Following the approach of Dekkers (2007), the GEBV was included in the selection index as a trait with a heritability of one and a genetic correlation determined by its accuracy.

2.2.2 Genetic Gain

The genetic gain of each path was calculated differently according to the number of stages being used. For single-stage breeding schemes and the first stage of multistage selection schemes, independently of the used algorithm, the genetic gain was derived from

$$\Delta G_{i,j} = z_{i,j} \sigma_{i,j},\tag{2.3}$$

where $\Delta G_{i,j}$ is the genetic gain, $z_{i,j}$ is the selection intensity and $\sigma_{i,j}$ is the standard deviation of the estimated breeding value of stage i for $i=1$ of path j . The selection intensity was derived from one-dimensional integration of a standard normal distribution according to:

$$z_{i,j} = \frac{\exp(-0.5u_{i,j}^2)}{\sqrt{2\Pi}} \frac{1}{(1 - \int_{-\infty}^{u_{i,j}} \exp(-0.5x^2)dx)}, \quad (2.4)$$

where $u_{i,j}$ is the truncation point of stage i within path j .

In the case of multistage selection schemes, the genetic gain of the stages following the first one are derived differently according to using the decorrelated indices or the optimal indices. When using the optimal indices, the following formula was applied:

$$\begin{aligned} \Delta G_{i,j} &= z_{i,j}\sigma_a \\ \Delta G_j &= z_{n,j}\sigma_a \end{aligned} \quad (2.5)$$

where $\Delta G_{i,j}$ is the genetic gain at stage i of path j , $z_{i,j}$ is the selection intensity after selecting at stage i of path j and σ_a is the standard deviation of the breeding goal. ΔG_j is the overall genetic gain of path j and $z_{n,j}$ is the selection intensity after selecting at the last stage of path j . Selection intensities were derived using the moment generating function of a truncated multi normal distribution of Tallis (1961) including all used selection stages and the breeding goal, where the truncation points were calculated using integration algorithms of Genz (1992) and maximisation techniques of Brent (1972).

Using the decorrelated index approach of Xu & Muir (1992) , the selection gain of multistage breeding schemes was calculated avoiding multiple integration according to

$$\begin{aligned} \Delta G_{i,j} &= z_{i,j}\sigma_{i,j} \\ \Delta G_j &= \sum_{i=1}^{n_j} \Delta G_{i,j}, \end{aligned} \quad (2.6)$$

where $\Delta G_{i,j}$ is the genetic gain, $z_{i,j}$ is the selection intensity and $\sigma_{i,j}$ is the standard deviation of estimated breeding values of stage i within path j . ΔG_j is the genetic gain and n_j is the number of stages of path j . Due to the zero covariance between the stage indices, $z_{i,j}$ could be calculated by one-dimensional integration as in formula 2.4.

The genetic gain per year for the whole selection scheme was calculated according to Rendel & Robertson (1950):

$$\Delta G_a = \frac{\sum_{i=1}^n \Delta G_j}{\sum_{i=1}^n GI_j}, \quad (2.7)$$

where ΔG_a is the genetic gain per year, ΔG_j is the genetic gain and GI_j is generation interval of path j , and n is the number of selection paths.

2.2.3 Breeding Program

For modeling a dairy cattle breeding program in a population of 100.000 dairy cows, the structure of a cooperative holstein dairy cattle breeding organization was adopted. Bull dams are selected as heifers from all available heifers and are assumed to be used as bull dams only once. Male selection candidates are produced by contract matings to the bull dams. For the purpose of comparing methods only one trait (milk trait) with a heritability of 0.25 is in the breeding goal. Selection can take place on the basis of performance data and GEBVs.

The biological, technical and economical parameters of the program are given in table 2.1. All parameters of the breeding program expressible in terms of probabilities and not given in table 2.1 are assumed to be zero or one (e.g. success of first insemination, survival rate of calves etc.).

The selection paths are “sire to sire”(SS), “sire to dam”(SD), “dam to sire”(DS) and “dam to dam” (DD). As in practice almost no selection takes place within the path DD,

Table 2.1: biological, economical and technical parameter of the breeding program

parameter	unit	value
h^2 milk trait		0.25
phenotypic standard deviation	kg	700
accuracy GEBV ¹		0.3 - 0.9
age at first calving	month	26
time between calvings	month	12
length of lactation	month	10
maturity of test bulls	month	14
number of daughters per sire	head	99 ³
inseminations with test bulls	%	20
average age of bull calves at purchase	month	6
price of bull calves	€	4000
husbandry costs of bull calves until maturity	€/day	5
husbandry costs of test bulls	€/day	4.5
cost for genomic selection ²	€	25 - 400
population of cows	head	100.000
demand for cow sires	head/year	10
demand for bull sires	head/year	5
initial male selection candidates	head/year	500
demand for bull dams (contract matings)	head/year	1000
compensation payments for test bull matings	€/test bull	3000
maximum breeding costs	€/year	719050

¹: GEBV = genomically estimated breeding values, ²: per genotyped individual, includes DNA isolation, genotyping and calculation of breeding values but not blood sample recovery, ³: minimum number, recalculated according to the number of bulls entering the progeny testing stage

a selection intensity of zero is assumed for this path. Each selection path is structured in stages. Each stage is characterized by a certain amount of information about the selection candidate becoming available at this stage. The selection stages available in the path SS and SD are a) selection on performance data and GEBVs of ancestors and half sibs (hereafter termed “pedigree information”), b) selection on the candidates GEBV and c) selection on progeny performance data. The selection stages available in the path DS are a) selection on performance data and GEBVs of ancestors and half sibs (hereafter termed “pedigree information”), and b) selection on the candidates GEBV. The selection in the path DD is only possible on performance data of ancestors, half sibs and own performance. Table 2.2 summarizes the amount of information at each stage within a certain path.

The number of selection candidates of the path SD before selection determine via reproduction parameters the demand for finally selected bull dams. The initial number of selection candidates in the path DS is determined by the cow population size, the calving interval and the sex ratio (here 1:1). The demand for finally selected cow sires and bull sires is appointed by the breeding organization. The initial candidates of the path SS before selection are the same individuals as the finally selected candidates of the path SD. Thus, the finally selected individuals of the path SS are a subset of those finally selected in the path SD.

For selecting male calves, the tissue sample for genotyping the selection candidate is taken by a veterinarian on farm at birth and has to be paid by the farmer. DNA-isolation and SNP-genotyping is carried out by a central laboratory, and is on the expenses of the breeding organization. Selection candidates are kept on farm until an age of six months. Farmers husbanding male candidates being selected at pedigree stage and being slaughtered after genotyping receive a compensation payment from the breeding organization. In the case of genotyping female calves as potential bull dams the division

Table 2.2: Maximum amount of information at each selection stage according to the selection path

Informat	Information	selection path								
		SS			SD			DS		DD
		stages			stages			stages		stages
		Ped	GBV	PT	Ped	GBV	PT	Ped	GBV	Ped
sires sire	SNP	x			x			x		x
sires dam	milk yield	x			x			x		x
sires dam	SNP	x			x			x		x
dams sire	SNP	x			x			x		x
dams dam	SNP	x			x			x		x
dams dam	milk yield	x			x			x		x
sire	SNP	x			x			x		x
dam	SNP	x			x			x		x
dam	milk yield	x			x			x		x
female halfsibs of the dam	milk yield	x			x			x		x
female halfsibs of the sire	milk yield	x			x			x		x
female halfsibs of the candidate	milk yield	x			x			x		x
candidate	milk yield									x
candidate	SNP		x			x			x	
female progeny of the candidate	milk yield			x			x			
generation interval in months		23	23	71	23	23	61	26	26	48

selection path: SS = “sire to sire”, SD = “sire to dam”, DS = “dam to sire”, DD = “dam to dam”, selection stages: Ped = selection is based on pedigree data (in the path “dam to dam” this stages includes the performance of the selection candidate), GBV = selection is based on genomically breeding values, PT = selection is based on progeny performance data, generation interval: months between the birth of the selection candidate and the birth of its reproductive progeny if the selection in this path stops at that stage

of the costs is the same, but no compensations are payed.

2.2.3.1 Breeding Costs

Costs being independent of selection strategies (e.g. performance recording of females, calculating pedigree index, marketing, semen processing) are not regarded. To allow for reduced costs for labor and infrastructure due to different selection schemes, all expenses are derived from some invariant basic cost given in table 2.1 via the cost function of NamKoong (1970) or some versions of this function.

Costs were divided into five categories: cost for GEBVs of selection candidates, compensation payments to breeders for husbanding male selection candidates, costs for purchasing bull calves, costs for husbanding bull calves until maturity and costs for husbanding bull calves until progeny performance. The first and second kind of costs occur only in the case of GS. The third and fourth kind of costs are variable concerning its absolute amount due to different selection strategies. The last kind of cost is only part of the calculation if progeny testing takes place. As mentioned above, the path SS is a derivative of the path SD, and the selection candidates in the path SS are a subset of those in the path SD. Thus, to avoid double counting, cost for purchasing bull calves and husbandry costs for bull calves until maturity were assigned to path SD.

Costs for GEBVs of selection candidates were calculated by:

$$C_i^{genomic} = C_{GEBV} \prod_{k=1}^{j-1} p_{ik} n_i, \quad (2.8)$$

where $C_i^{genomic}$ are the total costs for GS in path i , C_{GEBV} are the costs for genotyping one individual independent of the path, p_{ik} is the probability that an individual is selected at stage k , j is the stage within path i at which GS takes place and n_i is the number of initial selection candidates of path i . C_{GEBV} includes all costs for estimating GEBVs of

one individual except taking the blood samples of selection candidates.

Since compensation payments to farmers were payed only for non-selected genotyped male calves, this cost were derived from:

$$C_i^{comp} = C^{comp} \prod_{k=1}^{j-1} p_{ik}(1 - p_{ij})n_i, \quad (2.9)$$

where C_i^{comp} are the total compensation payments in path i , C^{comp} are the general compensation payments payed for one non selected genotyped bull calf, p_{ik} is the probability that an individual is selected at stage k , j is the stage of path i at which GS takes place, and n_i is the number of initial selection candidates of path i .

Costs for purchasing bull calves were expressed as follows:

$$C_{SD}^{bull} = C^{bull} \prod_{k=1}^j p_{(SD)k} n_{SD}, \quad (2.10)$$

where C_{SD}^{bull} are the costs for purchasing bull calves in the path SD, C^{bull} is the general price for one bull calf, j is the last stage of path SD before the bull calf is being bought, $p_{(SD)k}$ is the probability that an individual is selected at stage k of path SD, and n_{SD} is the number of initial selection candidates of the path SD.

Husbandry Costs for bulls until maturity were calculated according to:

$$C_{SD}^{hus_{calf}} = C^{hus_{calf}/day} (t_{mat} - t_j) \prod_{k=1}^j p_{(SD)k} n_{SD}, \quad (2.11)$$

where $C_{SD}^{hus_{calf}}$ are the husbandry costs for bull calves in the path SD, $C^{hus_{calf}/day}$ are the general husbandry costs per day for one bull calf until maturity, j is the last stage of path SD before the bull calf is being bought, $p_{(SD)k}$ is the probability that an individual is selected at stage k of path SD, t_{mat} is the age of the bull in days at maturity, t_j is the age of the bull calf in days at the last stage before it is bought and n_{SD} is the number

of initial selection candidates of the path SD.

Costs for progeny testing of bulls are mainly the husbandry cost of the time period between maturity and the availability of progeny performance data, and the compensation payments for test daughters. Thus, the cost of progeny testing are calculated according to:

$$C_i^{prog} = (C^{hus_{bull}/day}(t_{prog} - t_{mat}) + C^{daug}) \prod_{k=1}^j p_{ik} n_i, \quad (2.12)$$

where C_{SD}^{prog} are the progeny testing costs for bulls of the path i , $C^{hus_{bull}/day}$ are the general husbandry costs per day for one bull after maturity, C^{daug} are the compensation payments per test bull for test matings, j is the last stage of path i before the progeny testing stage, p_{ik} is the probability that an individual is selected at stage k of path i , t_{mat} is the age of the bull in days at maturity, t_{prog} is the age of the bull in days when progeny performance data are available, and n_i is the number of initial selection candidates of the path SD. Regarding that progeny testing can take place only in one male derived selection path (as the path SS is derived from the path SD), the index i in the formula above is disposable.

The overall costs for the path SS were then:

$$C_{SS}^{total} = C_{SS}^{prog} + C_{SS}^{comp} + C_{SS}^{genomic} \quad (2.13)$$

and the overall costs for the path SD:

$$C_{SD}^{total} = C_{SD}^{prog} + C_{SD}^{comp} + C_{SD}^{genomic} + C_{SD}^{bull} + C_{SD}^{hus_{calf}} \quad (2.14)$$

According to the system of original and derived paths selection stages being used in the path SD cannot be part of the path SS. Thus, costs for estimating GEBVs and compensation payments can only occur in the path SS or SD. The same applies to the

costs for progeny testing. Additionally, cost for GS or progeny testing only occur, if the selection stage where these informations are raised is part of the selection scheme.

The overall costs of the path DS consisted only of the cost for genomic selection:

$$C_{DS}^{total} = C_{DS}^{genomic} \quad (2.15)$$

The overall costs for the whole selection scheme were the sum over all selection paths. The maximum breeding costs (C_{max}) were calculated assuming a progeny testing selection scheme with the same cost structure.

$$C_{max} = C^{hus_{catf}/day}(t_{mat} - t_j) \prod_{k=1}^j p_{(SD)k} n_{SD} + (C^{hus_{bull}/day}(t_{prog} - t_{mat}) + C^{daug}) \prod_{k=1}^j p_{ik} n_i, \quad (2.16)$$

2.2.4 Parameter Variation

Parameters being varied during the calculation process were C_{GEBV} and r_{GEBV} . r_{GEBV} was varied between 0.3 and 0.9 in steps of 0.025 and of C_{GEBV} between 20 € and 400 € in steps of 10 €, resulting in 975 combinations of r_{GEBV} and C_{GEBV} .

2.2.5 Maximisation

2.2.5.1 Combination of Stages

To find the best breeding strategy for a given combination r_{GEBV} and C_{GEBV} , the number of stages in each path and, thus, the stage at which the selection process stops, was variable.

For each path, an initial number of selection candidates and a final number of selected individual for reproduction was fixed. The product of the proportion of the selected

individuals at each stage had to fulfill the equation:

$$\frac{n_{i_{final}}}{n_{i_{initial}}} = \prod_{j=1}^{m_i} p_{ij}, \quad (2.17)$$

where $n_{i_{final}}$ is the final number of selected individuals being actually used for reproduction, $n_{i_{initial}}$ is the initial number of selection candidates before any selection has taken place, m_i is the number of used selection stages and p_{ij} is the proportion of individuals being selected at stage j of path i on the number of individual having been tested at stage j .

For finding the best possible breeding scheme we applied a grid search with the following parameters. The percentage of selected individuals at each stage was varied between one and 100 % in steps of 2.5 %. To fulfill equation 2.17 the value of the last stage having a percentage of selected individuals $\leq 99\%$ was then erased and calculated as dependent variable. Only combinations of stages having a value calculated for the last stages ≤ 1 were seen as valid to be used in further calculation steps.

The valid stage combinations of all paths were totally cross classified over paths, and each of that combinations was seen as a possible breeding scheme. The goal function to be maximised was the genetic gain per year. The only constraint was to keep the breeding costs below maximum costs being calculated such that no GS takes place in all paths and in the path SD a conventional progeny performance breeding scheme with 50 test bulls per year is used. Breeding costs were derived for every of the valid schemes, but stages within such combinations having a proportion of selected individual > 0.99 were excluded. For all schemes fulfilling the cost constraint ΔG_a was calculated, and the scheme which maximised this parameter was evaluated as the best solution for the given set of input parameters.

Regarding the system of original and derived paths for SD and SS, the optimisation

process consisted of more than one calculation cycle. In the first cycle, the used selection stages in the paths SD and SS were equal. If the final result maximising the goal function contained a selection stage of the path SD with a proportion selected ≥ 0.99 and being located at the end of the selection process, and not between two stages with such a proportion < 0.99 , then the calculation was restarted excluding this stage from the path SD but still using it for the path SS. This variability allowed for calculating breeding schemes such that cow sires are selected on GEBVs and bull sires additionally on progeny performance data.

For each combination of r_{GEBV} and C_{GEBV} a grid of about 60000 breeding schemes was searched for optimisation. The total amount of evaluated breeding plans was 58 millions for each method.

2.2.5.2 Dynamic Pedigree Information Content

Since every selected candidate becomes an ancestor of its progeny, selected candidates of every selection path occur as ancestors in the pedigree of each selection candidate independent of the path. Furthermore, the exclusion of stages during the maximisation process changes the amount of information gathered about the selection candidate (e.g. GEBV and/or progeny performance for males). If a changed selection strategy in one path alters the amount of information about the selection candidate, the information content of the pedigree being used for selection candidates in other path might be altered too. This information content is altered additionally if a changed selection strategy of a certain path is changing its generation interval. In this case, it has to be regarded that at the time point of selecting the candidate some possible informants don't have performance data (e.g. daughters of the sire). Thus, if the selection strategy in the path SD and SS is changed from progeny performance testing to GS, the pedigrees of all selection paths are altered twice. Daughters of the candidates sire are excluded from

the pedigrees, whereas the GEBVs of the sires are included.

The main impact of the described dynamics of the pedigree information content is a variable accuracy of the pedigree derived breeding values. The maximum amount of data at each selection stage with respect to the selection path is shown in table 2.2. This amount was adapted according to the used stages in each path.

2.2.6 Software

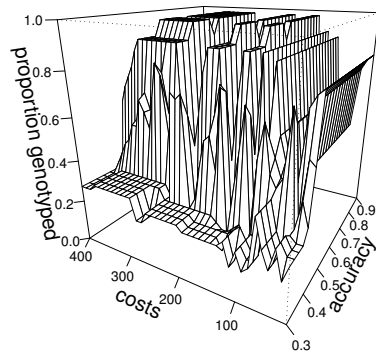
All calculations were carried out with a FORTRAN 90 program written by the first author. For calculating the selection intensity of an optimal index used in a multistage breeding program, FORTRAN routines of Genz (1992) and Brent (1972) were used.

2.3 Results

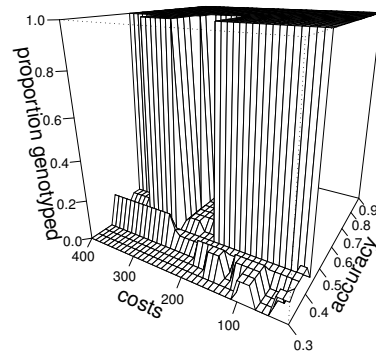
2.3.1 Comparison of Methods for Calculating the Genetic Gain

In the following, parameters and results of the breeding programs maximising the genetic gain per year and fulfilling the cost constraint are compared between *OSI* and *DSI*. Figures 2.2 and 2.3 represent the proportion of initial selection candidates of path SD being genotyped or progeny tested respectively. Both methods have in common that a r_{GEBV} of 0.4 was seen as the upper limit for applying progeny testing in this path. But two major differences between the both calculation methods were obvious.

Firstly, for an r_{GEBV} between 0.3 and 0.4 and for all values of C_{GEBV} , giving 195 possible parameter combinations, *DSI* identified 120 parameter combinations where only pedigree data and progeny performance are used, whereas *OSI* found no such parameter combinations. Additionally, *DSI* found the inclusion of all three selection stages as the highest yielding selection strategy in 52 of 195 parameter combinations (see table 2.3).

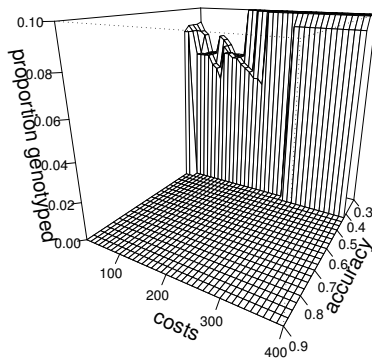


(a) optimal index

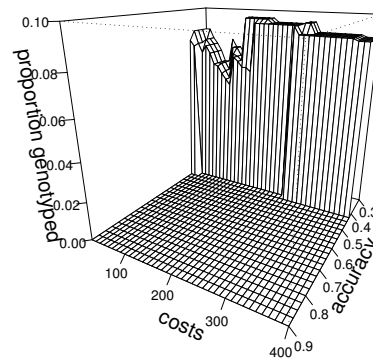


(b) decorrelated index

Figure 2.2: Parameters of the breeding schemes maximising the genetic gain per year as a function of the genotyping costs per individual, the accuracy of the genomically estimated breeding value and the used calculation method: proportion of genotyped selection candidates of the path “sire-dam”



(a) decorrelated index



(b) optimal index

Figure 2.3: Parameters of the breeding schemes maximising the genetic gain per year as a function of the genotyping costs per individual, the accuracy of the genomically estimated breeding value and the used calculation method: proportion of progeny tested initial selection candidates of the path “sire-dam”

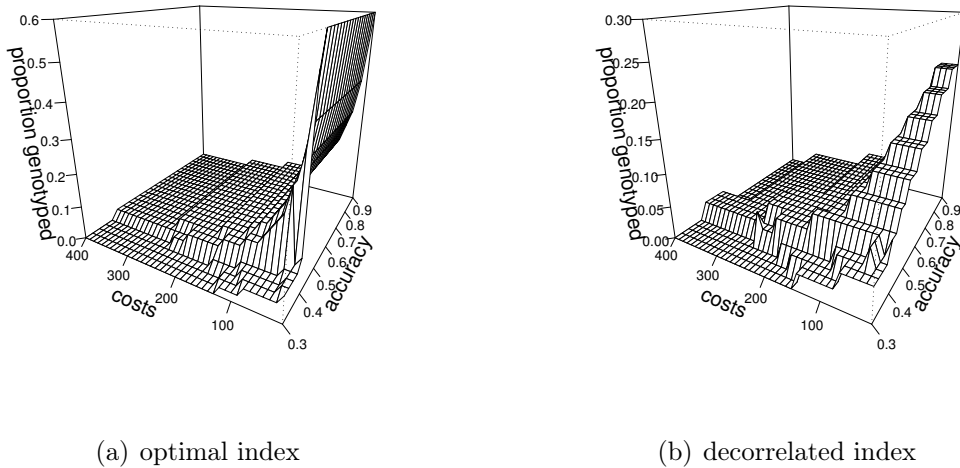


Figure 2.4: Parameters of the breeding schemes maximising the genetic gain per year as a function of the genotyping costs per individual, the accuracy of the genomically estimated breeding value and the used calculation method: proportion of genotyped selection candidates of the path “dam-sire”

On the contrary, this was the case in 173 parameter combinations if *OSI* was applied.

Secondly, in the r_{GEBV} range between 0.45 and 0.9, and across the whole range of C_{GEBV} , *DSI* found in 680 of 741 possible parameter combinations that genotyping all 500 candidates as the sole basis of selection is the best strategy for the path SD (see table 2.3). On the contrary, *OSI* never uses GS as the sole basis of selection independent of C_{GEBV} and r_{GEBV} , and the maximum proportion of genotyped male candidates was 0.975. Beside this major differences, the fluctuations in the proportion of genotyped initial selection candidates were much higher when using *OSI* compared to using *DSI*. Additionally, in 971 of 975 parameter combinations, the proportion of genotyped cow sires differed between both methods. In 680 cases, *DSI* found a higher optimum proportion compared to *OSI*, and in 291 cases it was vice versa. The maximum difference was found at C_{GEBV} of 260 € in a r_{GEBV} range between 0.55 and 0.9, where this proportion was one for *DSI*, compared to 0.075 for *OSI*.

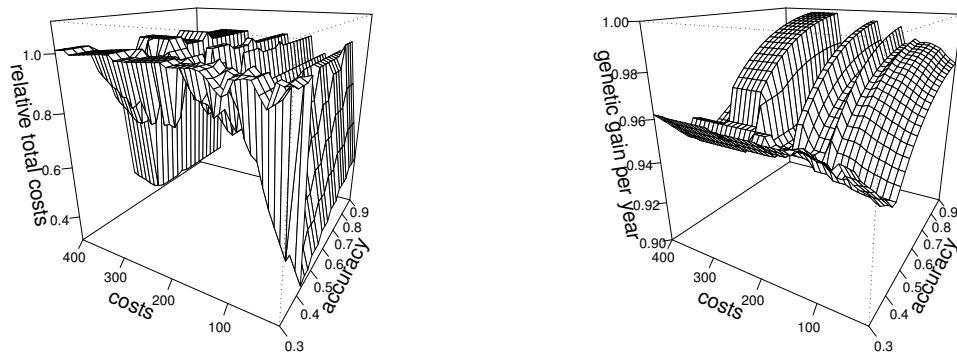
Table 2.3: Distribution of the different selection strategies according to the selection path being applied on and the accuracy of genomically estimated breeding values

path	r_{GEBV}^1	n ²	n_{Pd}^3		n_{Pd+PT}^4		n_{Pd+GS}^5		$n_{Pd+GS+PT}^6$		n_{GS}^7	
			a ⁸	b ⁹	a	b	a	b	a	b	a	b
SD ¹⁰	0.3-0.4	195	0	0	120	0	23	22	52	173	0	0
SS ¹¹	0.3-0.45	273	0	0	120	0	30	25	123	248	0	0
SD	0.45-0.9	741	0	0	0	0	61	741	0	0	680	0
DS ¹²	0.425-0.5	156	0	0	0	0	156	156	0	0	0	0
DS	0.3-0.9	975	221	121	0	0	754	854	0	0	0	0

1:parameter space for the accuracy of genomically estimated breeding values, 2:number of possible breeding schemes within the given parameter space, 3:number of schemes with selection on pedigree data, 4:number of schemes with selection on pedigree data and progeny performance, 5:number of schemes with selection on pedigree data and genomically estimated breeding values, 6:number of schemes with selection on pedigree data, genomically estimated breeding values and progeny performance, 7:number of schemes with selection on genomically estimated breeding values, 8:decorrelated indices are used, 9:optimal indices are used, 10:selection path “sire-dam”, 11:selection path “sire-sire”, 12:selection path “dam-sire”, the parameter space for the cost of genomic selection per genotyped individual is 20-400 €

The results for the path SS were very similar to those for SD (see table 2.3), but the upper limit for r_{GEBV} using progeny performance testing in this path was 0.45, independent of the applied method. Thus, in a range of r_{GEBV} between 0.425 and 0.45, potential cow sires were selected applying pedigree data and GS or GS solely, whereas the potential bull sires were selected from the cow sires on the basis of their progeny performance data. Outside of this range, bull sires were always selected from cow sires without gathering any additional information.

The proportion of the initial 50000 female selection candidates in the path DS being selected on the basis of pedigree data and GEBV depended on the calculation method used is given in figure 2.4. Both methods have in common that the proposed proportion of genotyped bull dams never reach the levels of cow sires because it is impossible due to the cost limitations imposed. In 535 of 975 possible parameter combinations, the proportion to be genotyped examined by *DSI* was different from that examined by *OSI*. In detail, in 13 of the 535 cases the proportion suggested by *DSI* was higher,



(a) relative breeding costs per year

(b) relative genetic gain per year (kg milk)

Figure 2.5: Comparison of parameters of the breeding schemes maximising the genetic gain per year as a function of the genotyping costs per individual, the accuracy of the genomically estimated breeding value and the used calculation method: decorrelated index relative to the optimal index

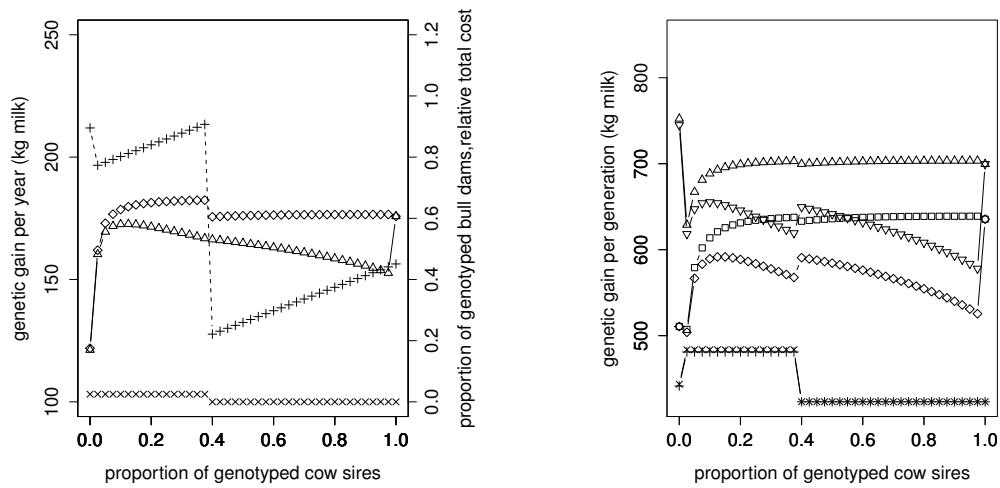
and in all other 522 cases lower compared to that suggested by *OSI*. When *OSI* was applied, only at an r_{GEBV} between 0.3 and 0.4 and above a C_{GEBV} of 120 € bull dams were selected solely on the basis of their pedigree data, summing up to 121 out of 975 parameter combinations. On the contrary, when using *DSI* at almost all r_{GEBV} values GS was excluded as a possible selection procedure for certain values of C_{GEBV} as long as they were above 150 €. Only at r_{GEBV} values between 0.425 and 0.5, GS was applied independently of C_{GEBV} . Thus, in 221 of 975 parameter combinations *DSI* excluded GS from the selection process in the path DS (see table 2.3). Another difference between both methods was the maximum proportion of genotyped bull dams, which was 0.6 when *OSI* was applied, C_{GEBV} equaled 20 € and r_{GEBV} was between 0.45 and 0.9, and 0.225 when *DSI* was used at r_{GEBV} levels from 0.825 to 0.9 and C_{GEBV} levels from 20 to 50 €.

The costs of the breeding schemes suggested by *DSI* exceeded those of *OSI* in 291 of 975 parameter combinations, but only to a maximum amount of 3.6 %, whereas in 680 cases

they were up to 70% lower compared to those out of *OSI* (see figure 2.5(a)). Except for one-stage selection procedures, *DSI* always suggested breeding schemes yielding less genetic gain per year than actually possible, with a maximum reduction of 5.5% and a mean of 2% (see figure 2.5(b)). If all optimisation results obtained applying *DSI* were recalculated using *OSI*, the underestimation due to using *DSI* was up to 5.5% for the summarized genetic gain per year, and up to 7% for the genetic gain per generation of the selection path SS (results not shown). For the reverse recalculation, the summarized genetic gain per year was underestimated due to *DSI* up to 29%, and the genetic gain per generation of selection path SS up to 40% (results not shown).

Figure 2.6 represents optimum breeding plan results using *DSI* if for a given proportion of genotyped bull calves all other parameters are chosen such that the genetic gain per time unit was maximised. Then, the results were recalculated with the same set of input parameters using *OSI*. Figure 2.6(a) clearly shows, that the optimum genetic gain per year as a function of the proportion of genotyped bull calves has two peaks (at 0.1 and 1) if *DSI* was used, whereas the application of *OSI* yielded a plateau for this result as long as other parameters (e.g. proportion of genotyped bull dams) remained unchanged. Additionally the exclusion of bull dams from genomic selection results in a decreased genetic gain per year when optimal indices were used, whereas the shape of the curve of the genetic gain per year was not changing if decorrelated indices were used. Moreover, when applying *DSI*, an increase in the proportion of genotyped male selection candidates above 0.4 leads to an exclusion of bull dams from GS (see figure 2.6(b)). This exclusion results in a decreased genetic gain per generation in the path DS and increased it in the path SD. As the overall genetic gain per year was roughly unaffected (see figure 2.6(a)), the loss in genetic gain per year due to the exclusion of GS in the path DS was compensated by an increase in the path SS and SD.

Another remarkable fact is the sharp increase in genetic gain per year resulting from



(a) Genetic gain per year calculated with decorrelated indices (\triangle) and calculated with optimal indices (\diamond), relative total breeding costs (+) and proportion of genotyped bull dams (\times).

(b) Genetic gain per generation of the paths "sire-sire" (\triangle) "sire-dam" (\square) and "dam-sire" (+) calculated with optimal indices, and of the paths "sire-sire" (∇), "sire-dam" (\diamond) and "dam-sire" (\times) calculated with decorrelated indices.

Figure 2.6: Calculation results as a function of the number of genotyped cow sires if at a given proportion of genotyped cow sires the breeding scheme maximising the genetic gain per year using the decorrelated index and regarding maximum breeding costs is used. Results of the optimal index are the recalculated results of the decorrelated index. accuracy of the genomically estimated breeding values: 0.75, cost for genomic selection: 400 € per genotyped individual.

Table 2.4: development of the accuracy of stage breeding values as a function of the implementation of genomic selection and the used selection indices, accuracy of $\text{GEBV}^2 = 0.75$

selection path	used selection stage		accuracy of stage breeding values			
			optimal indices		decorrelated indices	
	Pedigree ¹	GEBV	pedigree index	pedigree index + GEBV	pedigree index	GEBV
sire - dam	yes	no	0.177	-	0.177	-
dam - sire	yes	no	0.347	-	0.347	-
sire - dam	yes	no	0.427	-	0.427	-
dam - sire	yes	yes	0.379	0.766	0.379	0.606
sire - dam	yes	yes	0.441	0.754	0.441	0.612
dam - sire	yes	no	0.499	-	0.499	-
sire - dam	yes	yes	0.537	0.754	0.537	0.530
dam - sire	yes	yes	0.50	0.773	0.50	0.589

1: for the amount of information in the pedigree index see table 3.2, 2: genomically estimated breeding value

DSI if all bull calves are genotyped, whereas the results from using *OSI* were slightly reduced when the preselection at the pedigree stage was abandoned (see figure 2.6(b)). The plot of the breeding costs in figure 2.6(a) shows a strong decrease when bull dams were excluded from GS. Thus, almost independently of the C_{GEBV} level, genotyping bull dams was the main source of breeding costs in our simulation study. The comparison of the plots of *DSI* and *OSI* clearly shows that for a multi-stage selection scheme the application of decorrelated indices only yield in an underestimation in a range of up to 6% if the proportion of genotyped males was between 2.5 and 25% of the initial selection candidates. Above this range, the genetic gain per year was underestimated up to 14% across all paths, but even up to 18% in the path SS.

Table 2.4 summarizes the accuracy of pedigree derived breeding values as a function of the implementation of genomic selection and the used selection indices. When using

OSI, the accuracy of the selection indices out of pedigree information for male calves were always increased due to the implementation of GS because then, as the GEBV was treated as a trait, genotyped individuals within the pedigree had an additional information. Additionally, due to the implementation of GS in the path SD, the accuracy of the selection index at the GS stage of path DS was further increased. In a sharp contrast to this results, the application of *DSI* leads to completely different results. Although the accuracy at the pedigree stage increased too when implementing GS, the one at the GS stage not even reached the accuracy of the information source itself. Moreover, the accuracy at the GS stage was further reduced if, in addition to GS in the path SD, GS was implemented in the path DS too (and vice versa). Finally, applying GS in the paths SD and DS led to an accuracy of the selection index at the GS selection stage of the path SD being slightly below the one at its pedigree stage.

Taking together these results with special regard to those given in figure 2.6, one can conclude that both methods tend to put more emphasis on implementing GS in the male selection path because of cost limitations. Additionally, due to the strong peaks of the genetic gain as a function of the proportion of genotyped males when using *DSI*, this algorithm seems to be more invariable in reducing the proportion of genotyped males in order to allow for genotyping bull dams when economical resources are limited. Thus, *OSI* can reduce the proportion of genotyped males without losing genetic gain as long as the increasing cost for GS not requesting a decision about where to implement GS. Additionally, as in all paths the variation steps for the proportion selected at each stage were 2.5%, a reduction of the proportion genotyped in the path DS in order to keep the cost limitation is setting free enough economical resources to immediately increase the proportion of genotyped male calves. This behaviour is mirrored in the sharp declines and increases of the proportion of genotyped males given in figure 2.2(a), and explains why the results when using *DSI* differ from those of *OSI* not only because of an

underestimated genetic gain, but also in the recommendation concerning the structure of path-specific breeding strategies.

2.3.2 Genetic Gain of Optimal Indices

The genetic gain per year when using *OSI* for deriving the accuracy of selection indices is given in figure 2.7. The genetic gain per year increases with an increasing r_{GEBV} , but is almost independent of C_{GEBV} . The highest genetic gain per year of 236.94 kg or 0.67 genetic standard deviations was achieved at a parameter combination of $r_{GEBV}=0.9$ and $C_{GEBV}=20$ €, whereas the lowest genetic gain per year of 113.09 or 0.32 genetic standard deviations was achieved at a parameter combination of $r_{GEBV}=0.3$ and C_{GEBV} between 380 and 400 €. The relative contribution of the different selection paths to the summarized genetic gain of the breeding scheme was between 0.46 and 0.34 for the path SS, between 0.36 and 0.29 for the path SD, and between 0.35 and 0.19 for the path DS. In 21 of 975 parameter combinations, the relative contribution of the path DS exceeded that of the path SS, but to not more than 0.01. The relative contribution of the path SD was exceeded by that of the path DS in 202 of 975 cases, wherein the maximum exceeding was 0.04. Thus, in far the most cases of parameter combinations, the main contribution to the selection gain came from the path SS.

2.4 Discussion

2.4.1 Comparison of the results for the decorrelated and optimal index

The results of this simulation study amount the underestimation of the genetic gain per year using decorrelated indices up to 5.5% for the whole breeding scheme, and up to 6%, for certain selection paths. This is within the range given by Xu & Muir (1992). As in our simulation study the goal function was almost a plateau for the most possible

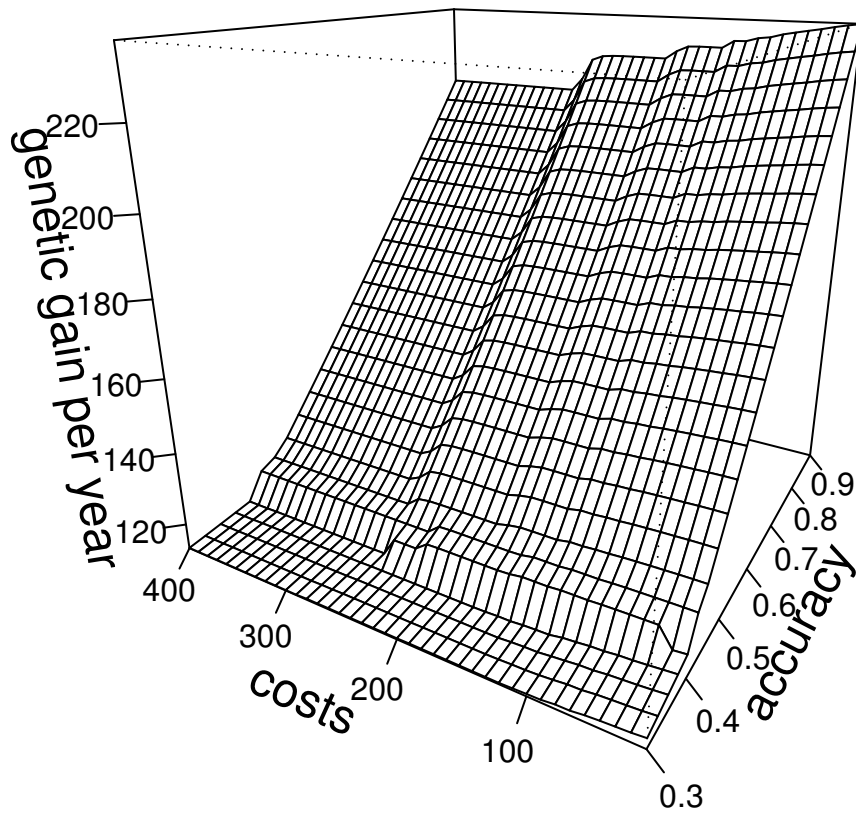


Figure 2.7: Calculation results as a function of the genotyping costs per individual and the accuracy of genomically estimated breeding values if optimal indices were used: genetic gain per year

values of proportion genotyped, the range of underestimation results from the fact that the proportion genotyped yielding the highest genetic gain per year when applying *DSI* was within the range in which a variation of this proportion has only a minor effect on the genetic gain per year. This outcome might be by chance or a result of the breeding scheme structure. Thus, a reliable prediction of the extent of this underestimation is not possible.

The possible interaction of paths and stages due to an altered pedigree information content as a function of path specific selection procedures was not regarded in previous publications applying decorrelated indices. As our results clearly show, the consideration of these interactions uncovered the side effect of reducing the accuracy of the stages following the pedigree stage further, if due to a change selection strategy in another path the information content of the pedigree is increased. Thus, the completely counterintuitive situation arose where excluding genomic selection in path DS increases the accuracy of GEBV and the genetic gain in path SS and SD. Furthermore, the fact that the proposed breeding schemes were more cost efficient when applying *DSI* compared to using *OSI* showed another difficult to handle and to forecast feature of decorrelated indices. Due to the structure of the simulated breeding program, overall breeding costs were tremendously affected by the implementation of GS in the path DS, whereas the use of GS in the path SS and SD could never increase the breeding costs to more than 324.175 € (which is less than 50% of the maximum breeding costs), even if C_{GEBV} was 400 €. Excluding GS from the path DS decreases the breeding costs immediately, even when the proportion genotyped in the path SD is kept at a high level. The fact that *DSI* proposed more cost efficient breeding programs compared to *OSI* results from keeping the proportion of genotyped bull dams at a low level, or excluding GS completely from the path DS. This results from the following complex interaction. As given in table 2.4, including information sources can lead to decreased accuracies at some selection stages

in certain paths. An additional selection stage is only being worth regarded when the additional genetic gain out of this stage overcompensates the loss due to decorrelation in all other stages. This includes also the complete exchange of stages if the accuracy of the included stage is higher than that of the excluded, and measuring the information at the new stage without preselection is not overdrawing the cost limitation, because then the accuracy of the new stage is not decreased by decorrelating the included and the excluded stage. Given that for an r_{GEBV} of 0.75 the shape of the genetic gain curve as a function of the proportion of genotyped bull dams for the path DS (not shown) is the same as presented for the path SD in figure 2.6(b), without cost limitation *DSI* would have found breeding schemes genotyping all selection candidates in all possible paths, because these maximise the genetic gain per year, whereas *OSI* would still have combined pedigree information and GEBV. Thus, as GS was beside progeny testing the only cost producing selection information in this simulation, in such a scenario *DSI* would always select the most expensive breeding scheme. If overall breeding costs are limited, the number of genotyped bull dams might be too low to generate a sufficient additional genetic gain to compensate the loss in all other paths. Then, GS is excluded from the path DS, leading in turn to a tremendous cost reduction. If the additional gain is sufficient, but cost limitations do not allow for genotyping the optimal amount of dams (e.g. 100%), then the proportion is genotyped where the genetic gain as a function of this proportion has a local maximum, or around this point (exemplified by figure 2.6(b) for the path SD). Where this local maximum is located is difficult to predict since the genetic gain is a nonlinear function of the selection intensities of stages, selection intensities are not independent of each other (one can be described as a function of all others) and its optimal composition depends on the correlation between breeding values and its correlation to the breeding goal.

Thus, given the facts above, one can say that the behaviour of the algorithm applying

DSI has its intrinsic logic, but is actually confusing for decision makers and makes the results difficult to interpret. Additionally, the results out of *DSI* underestimate the genetic gain for sure, but it is impossible to forecast if they are close to the optimum breeding strategy or not. Furthermore, the application of *DSI* leads to a strong peak of the goal function suggesting that the proportion of genotyped individuals can only be varied in a very narrow range without losing genetic gain, whereas the application of *OSI* clearly shows that genetic gain is relatively insensitive for a change of this parameter. As in this simulation the breeding scheme and the breeding goal were of low complexity, applying *DSI* to more complex breeding schemes including a variety of selection paths and stages, the interactions out of this method might be untraceable and leading to illogical advises.

One of the advantages of decorrelated indices mentioned by Xu & Muir (1992) is the applicability of maximisation methods using first and second derivatives of the goal function. As the cost function of NamKoong (1970) is not continuous if the selection intensities of certain selection stages converge to zero within the maximisation process, and the generation interval of a selection path cannot be expressed as a function of the proportion selected of each selection stage, the convergence of such maximisation algorithms is uncertain. Thus, grid search or heuristic approaches are still the methods of choice when goal functions are difficult or impossible to differentiate or not continuous. Since limitations due to cpu-time has been almost overcome because of developments of efficient hardware and fast algorithms (Genz, 1992), the exact calculation of optimal indices in combination with the above mentioned methods is maybe the better alternative.

2.4.2 Results using the optimal index

On the contrary to a genetic gain per year of 0.46 genetic standard deviations published by Schaeffer (2006), we found for a given r_{GEBV} of 0.75 a genetic gain per year between 0.53 and 0.57 genetic standard deviation. This difference may result from differences in the underlying selection intensities in the path SS and SD, which were higher in our calculations. Additionally, we did not find the path DS to be the main source of genetic gain, which may be based in the assumption of Schaeffer (2006) to have GEBV's available for all potential bull dams. Since in our calculations genotyping large numbers of bull dams was the major source of breeding costs, maximising genetic gain per year by including genomic selection in the path DS did not show the proposed great potentials in cost reduction.

Some breeding organisations still rely on using genomic selection as a preselection stage followed by some progeny testings (Spelman *et al.*, 2010). As the continuation of progeny testing in combination with genomic selection was found to be economical only at an $r_{GEBV} \leq 0.4$ in the path SD and ≤ 0.45 in the path SS, and r_{GEBV} of 0.7 can be achieved in practical breeding programs (Hayes *et al.*, 2009a), there might be no alternative to replacing conventional progeny testing by genomic selection in order to maximise the genetic gain per year.

2.5 Conclusions

The application of decorrelated indices to multistage dairy cattle breeding schemes including genomic selection in an optimisation approach regarding the strong interaction between selection path may lead to results being not only counterintuitive but also difficult to interpret yielding in improper decisions. Since the fast and stable calculation of selection intensities in multistage breeding programs is possible even for highly correlated

selection indices and small proportions of selected individuals, the optimal selection index is the method of choice for the deterministic optimisation of breeding scheme using selection index methodology.

2.6 Acknowledgement

This work was financed by the German Federal Ministry for Education and Research (BMBF), project "FUGATO-Brain".

2.7 Bibliography

- Brent, R. P. (1972), *Algorithms for Minimisation without Derivatives*. Prentice Hall.
- Daetwyler, H. D., Villanueva, B., Bijma, P. & Woolliams, J. A. (2007), Inbreeding in genome-wide selection. *J Anim Breed Genet* **124**(6), 369–376, URL <http://dx.doi.org/10.1111/j.1439-0388.2007.00693.x>.
- Dekkers, J. C. M. (2007), Prediction of response to marker-assisted and genomic selection using selection index theory. *J Anim Breed Genet* **124**(6), 331–341, URL <http://dx.doi.org/10.1111/j.1439-0388.2007.00701.x>.
- Ducrocq, V. & Colleau, J. (1989), Optimum truncation points for independent culling level selection on a multivariate normal distribution, with an application to dairy cattle selection. *Genet Sel Evol* **21**, 185–191.
- Genz, A. (1992), Numerical computation of multivariate normal probabilities. *J Comput Graph Stat* **1**, 141 – 149.
- Hayes, B. J., Bowman, P. J., Chamberlain, A. C., Verbyla, K. & Goddard, M. E. (2009a), Accuracy of genomic breeding values in multi-breed dairy cattle populations. *Genet Sel Evol* **41**, 51, URL <http://dx.doi.org/10.1186/1297-9686-41-51>.

-
- Hayes, B. J., Bowman, P. J., Chamberlain, A. J. & Goddard, M. E. (2009b), Invited review: Genomic selection in dairy cattle: progress and challenges. *J Dairy Sci* **92**(2), 433–443, URL <http://dx.doi.org/10.3168/jds.2008-1646>.
- Hicks, C., Muir, W. & Stick, D. (1998), Selection index updating for maximizing rate of annual genetic gain in laying hens. *Poult Sci* **77**(1), 1–7, URL [PM:9469744](https://pubmed.ncbi.nlm.nih.gov/9469744/).
- Meuwissen, T. H., Hayes, B. J. & Goddard, M. E. (2001), Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**(4), 1819–1829.
- NamKoong, G. (1970), Optimum allocation of selection intensity in two stages of truncation selection. *Biometrics* **26**, 465 – 476.
- Rendel, J. M. & Robertson, A. (1950), Estimation of genetic gain in milk yield by selection in a closed herd of dairy cattle. *J Genet* **50**, 1–8.
- Saxton, A. (1989), Indcull version 3.0: Independent culling for two or more traits. *J Hered* **80** (2), 166–167.
- Schaeffer, L. R. (2006), Strategy for applying genome-wide selection in dairy cattle. *J Anim Breed Genet* **123**(4), 218–223, URL <http://dx.doi.org/10.1111/j.1439-0388.2006.00595.x>.
- Spelman, R., Arias, J., Keehan, M., Obolonkin, V., Winkelman, A., Johnson, D. & Harris, B. (2010), Applicatino of genomic selection in the new zealand dairy cattle industriy. In *Proceedings of the nineth World Congress on Genetics Applied to Livestock Production, Leipzig, Germany, August 1-6, 2010*.
- Tallis, G. M. (1961), The moment generating function of the truncated multi-normal distribution. *J R Statist Soc B* **23**, 223 – 229.

- Xie, C. & Xu, S. (1998), Efficiency of multistage marker-assisted selection in the improvement of multiple quantitative traits. *Heredity* **80** (Pt 4), 489–498, URL [PM:9618913](#).
- Xu, S., Martin, T. G. & Muir, W. M. (1995), Multistage selection for maximum economic return with an application to beef cattle breeding. *J Anim Sci* **73**(3), 699–710, URL [PM:7608002](#).
- Xu, S. & Muir, W. M. (1991), Multistage selection for genetic gain by orthogonal transformation. *Genetics* **129**(3), 963–974, URL [PM:1752432](#).
- Xu, S. & Muir, W. M. (1992), Selection index updating. *Theor Appl Genet* **83**(4), 451–458, URL <http://dx.doi.org/10.1007/BF00226533>.

3 Optimisation of Genomic Selection Dairy Cattle Breeding Schemes regarding different SNP-Chips and Imputing

Optimisation of Genomic Selection Dairy Cattle Breeding Schemes regarding different SNP-Chips and Imputing

Vinzent Börner & Friedrich Teuscher & Norbert Reinsch
Leibniz Institute for the Biology of Farm Animals, FBN
Research Unit Genetics and Biometry
18196 Dummerstorf, Germany

Abstract

The availability of different SNP chips and the development of imputation algorithms allow for multistage dairy cattle breeding schemes applying various genomic selection strategies on the basis of SNP chips which yield genomically estimated breeding values (*GEBV*) with different accuracies for different costs. Since the costs and accuracies of these *GEBVs* can be regarded as uncertain, and the interaction of path specific selection strategies linked via restricted economical resources has not been examined, a grid search was applied in order to identify optimum breeding plans maximising the genetic gain per year in a multistage dairy cattle breeding program allowing for the use a high density (hd) and a low density (ld) SNP chip, and varying the costs and accuracies of *GEBVs* estimated from these chips in a semi-continuous range. Furthermore, the overall breeding costs were limited and the genetic gain was predicted deterministically. The results clearly show that the use of ld chips is mainly indicated for the selection of bull dams, but they also put emphasis on the non-linearity of the genetic gain as a function of the selection intensity because increasing the number of genotyped bull dams above a certain limit due to abandoning hd genotyping was not seen as economical. Furthermore, the genetic gain was much more sensible to changes in the costs of *GEBVs* from ld chips compared to hd chips, and changes of costs and accuracies of *GEBVs* from ld chips put the most pressure on the breeding scheme structure in order to maintain a high genetic gain per year.

3.1 Introduction

Genomic selection (*GS*) of farm animals is revolutionising animal breeding in theory and practice since the beginning of the century (Meuwissen *et al.*, 2001). Highly dense maps of genetic markers being in linkage disequilibrium to genes or quantitative trait loci coding for economical important traits allow to derive genomically estimated breeding values (*GEBV*). The expectations on selecting individuals on the basis of these *GEBV*s concern three major frontiers of animal breeding: the accuracy of breeding values for traits with a low heritability (Calus & Veerkamp, 2007; Muir, 2007), the control of inbreeding (Daetwyler *et al.*, 2007) and the generation interval (Schaeffer, 2006). Since in dairy cattle breeding the length of the generation interval is the major obstacle for increasing the genetic gain per year, the attention in this sector of animal breeding are focused mainly on the latter promise of *GS*. Abandoning progeny testing for potential breeding bulls and own performance tests for potential breeding cows leads to such a substantial decrease of the generation interval, that although the accuracies of these *GEBV*s (r_{GEBV}) for milk production is still below those out of conventional progeny testing schemes (Hayes *et al.*, 2009), there is an increase in genetic gain per year (Schaeffer, 2006). In a detailed calculation, Schaeffer (2006) summarized the potential effects of *GS* on dairy cattle breeding schemes assuming an r_{GEBV} of 0.75, costs for *GS* per animal (C_{GEBV}) of 500 USD and a one-stage selection approach where a highly accurate *GEBV* was available for a large number of potential bull dams. Especially the last assumption might not be realistic since genotyping large proportions of the cow population with SNP chips allowing for such an r_{GEBV} may overdraw the economical resources of breeding organisations. Furthermore, if overall breeding costs are limited and C_{GEBV} are high, *GS* of bull dams will barely reach a dimension yielding a substantial surplus in genetic gain being generated in this selection path justifying the financial efforts (Börner & Reinsch, 2010).

As today SNP chips of different densities are available, economical limitations may be overcome due to the usage of cheap low density SNP chips in order to genomically evaluate large proportions of the cow population. Suppose the SNPs of such chips are a subset of those being used for high density panels, the accuracy of breeding values from low density genotyping ($r_{GEBV,LD}$) depends on the special construction of such subset. The markers in the subset can be selected such that they represent those having the highest effect on the phenotype when their effects were estimated on the basis of high density genotyping, or such that they are evenly spaced across the genome (Habier *et al.*, 2009). If marker values are re-estimated for such subsets, r_{GEBV} based on “evenly spaced” SNPs is somewhat less than that using “highest effect” SNPs (Weigel *et al.*, 2010). On the contrary, subsets of evenly spaced markers can be used for imputing the high density SNP genotype of an individual being genotyped with a low density SNP chip (Habier *et al.*, 2009) which is possible across traits and breeds. Imputation algorithms can rely only on population based haplotype information (Howie *et al.*, 2009), or merge population and parent haplotype information (VanRaden, 2010). Marker effects estimated from high density SNP genotypes can then be used to calculate the $GEBV$ for the imputed genotypes. Using a population based imputation algorithm, Weigel *et al.* (2010) found a reduction of r_{GEBV} for milk yield, protein percentage and daughter pregnancy rate of Jersey dairy cattle due to imputation between zero and 45% depending on the size of the SNP subset, whereupon between 0.9% (366) and 6.9% (2942) of the SNPs being used for the high density chip were used for the low density chip. In a study on simulated data VanRaden (2010) reported a reduction of r_{GEBV} of only 4% if 5,000 out of 50,000 markers are used for the subset, and 8% for 500 out of 50,000. The underlying r_{GEBV} estimated from 50,000 marker was 0.91 and an imputation algorithm combining pedigree and population information was used.

However, the decision room of breeding organisations concerning the detailed struc-

ture of breeding schemes has become tremendously complex since the actually realised r_{GEBV} of breeding values being estimated from high and low density SNP genotypes may change in practical breeding programs, and the future development of genotyping costs is still unclear. Additionally, GS and the availability of different SNP chips allows for a variety of one-, two- or multistage breeding schemes in every single selection path, and economical resources are not selection path specific and limited. As the breeding scheme construction affects breeding value accuracies, selection intensities, generation intervals and, therefore, the genetic gain per year, answers to questions concerning whether, where, when and to what costs a certain information correlated with the breeding goal should be gathered is of paramount interest. Moreover, decision must be taken under a uncertainty about the future development of important parameters such as costs and accuracies of genomically estimated breeding values.

The aim of this study was to elucidate the outlined decision room in order to find optimum dairy cattle breeding schemes regarding genomic selection and the possibility of using different SNP chips under the constraint of limited economical resources. Additionally, the sensitivity of the genetic gain and the robustness of the structure breeding schemes to changes in costs and accuracies of genomically estimated breeding values could be examined. Therefore, multi-stage selection schemes were evaluated using a grid search and varying the cost ($C_{GEBV,HD}$) and accuracy ($r_{GEBV,HD}$) of breeding values ($GEBV_{HD}$) out of high density and low density SNP chips ($C_{GEBV,LD}$, $r_{GEBV,LD}$, $GEBV_{LD}$). The outcomes are providing both, suggestions for the optimal structure of a dairy cattle breeding scheme for every given combination of r_{GEBV} and C_{GEBV} , and answers to questions concerning the sensitivity of the genetic gain per year if external parameters changing or selection schemes are constructed improper.

3.2 Methods

3.2.1 Construction of Selection Indices and the Implementation of GEBV

For calculating the reliability of breeding values selection index methodology was used based on the pedigree given in figure 3.1.

Coefficients for regressing the genotype of the selection candidate on phenotypic measurements, being used to calculate variances, covariances and accuracies of estimated breeding values, were constructed such that

$$\sigma_{(i,j),T} = w' \times G_{i,j} \times b_{i,j} \rightarrow \max, \quad (3.1)$$

where $\sigma_{(i,j),T}$ is the covariance between the estimated and true breeding value, w is the vector of economic weights of the traits in the breeding goal, $b_{i,j}$ is the vector of regression coefficients for all information sources being available, and $G_{i,j}$ is the genetic-phenotypic covariance matrix at selection stage $j = 1, \dots, m$ of selection path $i = 1, \dots, n$.

Following the approach of Dekkers (2007), the $GEBV_{HD}$ and $GEBV_{LD}$ were included in the selection index as a trait with a heritability of one and a genetic correlation determined by its accuracy. The correlation between $GEBV_{HD}$ and $GEBV_{LD}$ was calculated as follows:

$$r_{HD:LD} = \frac{\sigma_{LD}}{\sigma_{HD}}, \quad (3.2)$$

where $r_{HD:LD}$ is the correlation between $GEBV_{HD}$ and $GEBV_{LD}$, σ_{LD} is the standard deviation of $GEBV_{LD}$ and σ_{HD} is the standard deviation of $GEBV_{HD}$. A detailed derivation of this formula is given in 3.8 (Appendix).

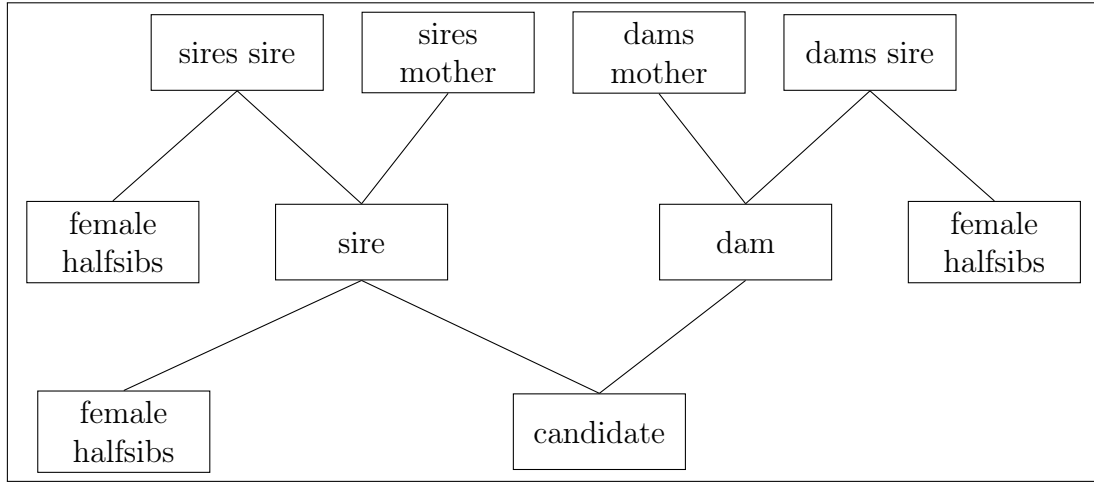


Figure 3.1: Standard pedigree used to derive breeding values in all selection paths

3.2.2 Genetic Gain

The genetic gain of each path was calculated according to the number of stages being used. For single-stage breeding schemes and the first stage of multistage selection schemes the genetic gain was derived from

$$\Delta G_{i,1} = z_{i,1}\sigma_{i,1}, \quad (3.3)$$

where $\Delta G_{i,1}$ is the genetic gain, $z_{i,1}$ is the selection intensity and $\sigma_{i,1}$ is the standard deviation of the estimated breeding value at the first stage of path i . The selection intensity was derived from one-dimensional integration of a standard normal distribution according to:

$$z_{i,1} = \frac{\exp(-0.5u_{i,1}^2)}{\sqrt{2\pi}} \frac{1}{(1 - \int_{-\infty}^{u_{i,1}} \exp(-0.5x^2)dx)}, \quad (3.4)$$

where $u_{i,1}$ is the truncation point at the first stage within path j .

In the case of multistage selection schemes, the genetic gain of the stages following the

first one were derived applying the following formula:

$$\begin{aligned}\Delta G_{i,j} &= z_{i,j}\sigma_a \\ \Delta G_i &= z_{i,m}\sigma_a\end{aligned}\tag{3.5}$$

where $\Delta G_{i,j}$ is the genetic gain at stage j of path i , $z_{i,j}$ is the selection intensity at stage j of path i and σ_a is the standard deviation of the aggregated genotype. ΔG_i is the overall genetic gain of path i and $z_{m,i}$ is the selection intensity after selecting at the last stage of path i . Selection intensities were derived using the moment generating function of a truncated multi normal distribution of Tallis (1961) including all used selection stages and the aggregated genotype, where the truncation points were calculated using integration algorithms of Genz (1992) and maximisation techniques of Brent (1972).

The genetic gain per year for the whole selection scheme was calculated according to Rendel & Robertson (1950):

$$\Delta G_a = \frac{\sum_{i=1}^n \Delta G_i}{\sum_{i=1}^n GI_i},\tag{3.6}$$

where ΔG_a is the genetic gain per year, ΔG_i is the genetic gain and GI_i is generation interval of path $i = 1, \dots, n$.

3.2.3 Breeding Program

For modeling a dairy cattle breeding program in a population of 100.000 dairy cows, the structure of a cooperative holstein dairy cattle breeding organization was adopted. Bull dams are selected as heifers from all available heifers and are assumed to be used as bull dams only once. Male selection candidates are produced by contract matings to the bull dams. For the sake of simplicity only one trait (milk trait) with a heritability of 0.25 is in the breeding goal. Selection can take place on the basis of pedigree data, GEV_{LD}

and $GEBV_{HD}$.

The biological, technical and economical parameters of the program are given in table 3.1. All parameters of the breeding program expressible in terms of probabilities and not given in table 3.1 are assumed to be zero or one (e.g. success of first insemination, survival rate of calves etc.).

The selection paths are “sire to sire”(SS), “sire to dam”(SD), “dam to sire”(DS) and “dam to dam” (DD). As in practice almost no selection takes place within the path DD, a selection intensity of zero is assumed for this path. Each selection path is structured in stages. Each stage is characterized by a certain amount of information about the selection candidate becoming available at this stage. The selection stages available in the path SS, SD and DS are a) selection on performance data and possible $GEBV$ s of ancestors and half sibs (hereafter termed “pedigree information”), b) selection on the candidates $GEBV_{LD}$ and c) selection on the candidates $GEBV_{HD}$. The selection in the path DD is only possible on performance data of ancestors, half sibs and own performance. Table 3.2 summarizes the amount of information at each stage within a certain path.

The number of selection candidates of the path SD before selection determines via reproduction parameters the demand for finally selected bull dams. The initial number of selection candidates in the path DS is determined by the cow population size, the calving interval and the sex ratio (here 1:1). The demand for finally selected cow sires and bull sires is appointed by the breeding organization. The initial candidates of the path SS before selection are the same individuals as the finally selected candidates of the path SD. Thus, the finally selected individuals of the path SS are a subset of those finally selected in the path SD.

For selecting male calves, the tissue sample for genotyping the selection candidate

Table 3.1: biological, economical and technical parameter of the breeding program

parameter	unit	value
h^2 milk trait		0.25
phenotypic standard deviation	kg	700
$r_{GEBV,HD}$		0.7 - 0.9
$r_{GEBV,LD}$		0.4 - 0.65
age at first calving	month	26
time between calvings	month	12
length of lactation	month	10
maturity of test bulls	month	14
average age of bull calves at purchase	month	6
price of bull calves	€	4000
husbandry costs of bull calves until maturity	€/day	5
$C_{GEBV,HD}$	€	150 - 250
$C_{GEBV,LD}$	€	20 - 100
population of cows	head	100.000
demand for cow sires	head/year	10
demand for bull sires	head/year	5
initial male selection candidates	head/year	500
demand for bull dams (contract matings)	head/year	1000
compensation payments for husbanding male calves until $GEBV_{LD}$ estimation	€/calf	150
compensation payments for husbanding male calves until $GEBV_{HD}$ estimation	€/calf	150
maximum breeding costs	€/year	719050

$GEBV_{HD}$ = selection is based on genomic breeding values estimated with high density SNP chips, $GEBV_{LD}$ = selection is based on genomic breeding values estimated with low density SNP chips, $r_{GEBV,HD}$ = accuracy of $GEBV_{HD}$, $r_{GEBV,LD}$ = accuracy of $GEBV_{LD}$, $C_{GEBV,HD}$ = costs for estimating genomic breeding values out of using high density SNP chips, per genotyped individual, includes DNA isolation, genotyping and calculation of breeding values but not blood sample recovery, $C_{GEBV,LD}$ = costs for estimating genomic breeding values out of using low density SNP chips, per genotyped individual, includes DNA isolation, genotyping and calculation of breeding values but not blood sample recovery,

Table 3.2: Maximum amount of information at each selection stage according to the selection path

Informat	Information	selection path			
		SS & SD & DS			DD
		stages			
		Ped	$GEBV_{LD}$	$GEBV_{HD}$	Ped
sires sire	$GEBV_{HD}$	x			x
sires sire	$GEBV_{LD}$	x			x
sires dam	milk yield	x			x
sires dam	$GEBV_{LD}$	x			x
sires dam	$GEBV_{HD}$	x			x
dams sire	$GEBV_{LD}$	x			x
dams sire	$GEBV_{HD}$	x			x
dams dam	$GEBV_{LD}$	x			x
dams dam	$GEBV_{HD}$	x			x
dams dam	milk yield	x			x
sire	$GEBV_{LD}$	x			x
sire	$GEBV_{HD}$	x			x
dam	$GEBV_{LD}$	x			x
dam	$GEBV_{HD}$	x			x
dam	milk yield	x			x
halfsibs of the dam	milk yield	x			x
halfsibs of the sire	milk yield	x			x
halfsibs of the candidate	milk yield	x			x
candidate	milk yield				x
candidate	$GEBV_{LD}$		x		
candidate	$GEBV_{HD}$			x	
generation interval in months for males		23	23	23	-
generation interval in months for bull dams		26	26	26	-
generation interval in months for cow dams		-	-	-	48

selection path: SS = "sire to sire", SD = "sire to dam", DS = "dam to sire", DD = "dam to dam", selection stages: Ped = selection is based on pedigree data (in the path "dam to dam" this stages includes the performance of the selection candidate), $GEBV_{HD}$: selection is based on genomic breeding values estimated with high density SNP chips, $GEBV_{LD}$: selection is based on genomic breeding values estimated with low density SNP chips, generation interval: months between the birth of the selection candidate and the birth of its reproductive progeny if the selection in this path stops at that stage

is taken by a veterinarian on farm at birth and has to be payed by the farmer. DNA-isolation and SNP-genotyping is carried out by a central laboratory, and is on the expense of the breeding organization. Selection candidates are kept on farm until an age of six months. Farmers husbanding male candidates being selected at pedigree stage and being slaughtered after the first (low density) or second (high density) genotyping receive a compensation payment from the breeding organization to cover the husbanding costs for each genotyping process. In the case of genotyping female calves as potential bull dams the division of the costs is the same, but no compensations are payed.

3.2.4 Breeding Costs

Costs being independent of selection strategies (e.g. performance recording of females, calculating pedigree index, marketing, semen processing) are not regarded. To allow for reduced costs for labor and infrastructure due to different selection schemes, all expenses are derived from some invariant basic cost given in table 3.1 via the cost function of NamKoong (1970) or adopted versions of this function.

Operational Costs Costs were divided into four categories: cost for $GEBV_{LD}$ and $GEBV_{HD}$ of selection candidates, compensation payments to breeders for husbanding male selection candidates, costs for purchasing bull calves and costs for husbanding bull calves until maturity. The first and second kind of costs occur only in the case of GS . The third and fourth kind of costs are variable concerning its absolute amount due to different selection strategies. As mentioned above, the path SS is a derivative of the path SD , and the selection candidates in the path SS are a subset of those in the path SD . Thus, to avoid double counting, cost for purchasing bull calves and husbandry costs for bull calves until maturity were assigned to path SD .

Costs for GS_{LD} were calculated by:

$$C_{(GS,LD)i} = C_{GEBV,LD} \prod_{k=1}^{j-1} p_{ik} n_i, \quad (3.7)$$

where $C_{(GS,LD)i}$ are the total costs for GS on the basis of $GEBV_{LD}$ in path i , $C_{GEBV,LD}$ are the costs for genotyping one individual independent of the path with a low density SNP chip, p_{ik} is the probability that an individual is selected at stage k , j is the stage within path i at which GS takes place and n_i is the number of initial selection candidates of path i . $C_{GEBV,LD}$ includes all costs for estimating $GEBV_{LD}$ of one individual except taking the blood samples of selection candidates. The same formula was applied to calculate the costs for $GEBV_{HD}$ with changes in abbreviation to $C_{GEBV,HD}$ and $C_{(GS,HD)i}$.

Compensation payments to farmers were paid only for non-selected genotyped male calves. This cost were derived for low density genotyped individuals from:

$$C_{(Comp,LD)i} = C_{Comp,LD} \prod_{k=1}^{j-1} p_{ik} (1 - p_{ij}) n_i, \quad (3.8)$$

where $C_{(Comp,LD)i}$ are the total compensation payments in path i , $C_{Comp,LD}$ are the general compensation payments paid for one non selected genotyped bull calf, p_{ik} is the probability that an individual is selected at stage k , j is the stage of path i at which GS takes place, and n_i is the number of initial selection candidates of path i . The same formula was applied for compensation payments out of high density genotyping with changes in abbreviations to $C_{(Comp,HD)i}$ and $C_{Comp,HD}$.

Costs for purchasing bull calves were expressed as follows:

$$C_{calf,SD} = C_{calf} \prod_{k=1}^j p_{(SD)k} n_{SD}, \quad (3.9)$$

where $C_{calf,SD}$ are the costs for purchasing bull calves in the path SD , C_{calf} is the general price for one bull calf, j is the last stage of path SD before the bull calf is being bought, $p_{(SD)k}$ is the probability that an individual is selected at stage k of path SD , and n_{SD} is the number of initial selection candidates of the path SD .

Husbandry Costs for bulls until maturity were calculated according to:

$$C_{(hus,calf)SD} = C_{(hus,calf)/day}(t_{mat} - t_j) \prod_{k=1}^j p_{(SD)k} n_{SD}, \quad (3.10)$$

where $C_{(hus,calf)SD}$ are the husbandry costs for bull calves in the path SD , $C_{(hus,calf)/day}$ are the general husbandry costs per day for one bull calf until maturity, j is the last stage of path SD before the bull calf is being bought, $p_{(SD)k}$ is the probability that an individual is selected at stage k of path SD , t_{mat} is the age of the bull in days at maturity, t_j is the age of the bull calf in days at the last stage before it is bought and n_{SD} is the number of initial selection candidates of the path SD .

The overall costs for the path SS were then:

$$C_{total,SS} = C_{(Comp,HD)SS} + C_{(GS,HD)SS} + C_{(Comp,LD)SS} + C_{(GS,LD)SS} \quad (3.11)$$

and the overall costs for the path SD

$$C_{total,SS} = C_{(Comp,HD)SD} + C_{(GS,HD)SD} + C_{(Comp,LD)SD} + C_{(GS,LD)SD} \\ + C_{(hus,calf)SD} + C_{calf,SD} \quad (3.12)$$

According to the system of original and derived paths selection stages being used in the path SD could not be part of the path SS . Thus, costs for estimating a certain $GEBV$ and the following compensation payments could only be accounted for in the path SS or SD . Additionally, cost for GS_{HD} or GS_{LD} only accrued, if the selection

stage where these informations are raised was part of the selection scheme.

The overall costs of the path DS were:

$$C_{total,DS} = C_{(GS,LD)DS} + C_{(GS,HD)DS} \quad (3.13)$$

The summarised costs for the whole selection scheme were the sum over all selection paths.

Maximum Costs The maximum breeding costs (C_{max}) being used as a constraint during the maximisation process were calculated assuming a selection scheme based on progeny performance testing 50 test bulls per year with following cost structure:

$$C_{max} = C_{(hus,calf)SD} + C_{calf,SD} + (C^{husbull/day}(t_{prog} - t_{mat}) + C^{daug}) \prod_{k=1}^j p_{ik} n_i, \quad (3.14)$$

where t_{prog} is the age of the male when daughter lactation records are available and C^{daug} are the insemination compensation payments for one test bull payed by the breeding organisation for one test bull, p_{ik} is the probability that an individual is selected at stage k , j is the last stage within path i before progeny testing takes place and n_i is the number of initial selection candidates of path i .

3.2.5 Maximisation

For finding the best possible breeding scheme accounting for the variability of r_{GEBV} and C_{GEBV} , we applied a double layer grid search with the following parameters.

Parameter Variation In the first layer of the grid, the parameters being varied during the calculation process were $r_{GEBV,HD}$, $r_{GEBV,LD}$, $C_{GEBV,HD}$ and $C_{GEBV,LD}$. $r_{GEBV,HD}$

was varied between 0.7 and 0.9 in steps of 0.05, $r_{GEBV,LD}$ between 0.4 and 0.65 with the aforementioned step width, $C_{GEBV,HD}$ between 150 € and 250 € and $C_{GEBV,LD}$ between 20 € and 100 € in steps of 20 €, resulting in 900 combination out cross classifying all four mentioned parameters.

Combination of Stages In the second layer of the grid search, the number of stages in each path and, thus, the stage at which the selection process stops, was varied in order to find the best breeding strategy for a given combination of $r_{GEBV,HD}$, $r_{GEBV,LD}$, $C_{GEBV,HD}$ and $C_{GEBV,LD}$.

For each path, an initial number of selection candidates and a final number of selected individual for reproduction was fixed. The product of the proportion of the selected individuals at each stage had to fulfill the equation:

$$\frac{n_{i_final}}{n_{i_initial}} = \prod_{j=1}^{m_i} p_{ij}, \quad (3.15)$$

where n_{i_final} is the final number of selected individuals being actually used for reproduction, $n_{i_initial}$ is the initial number of selection candidates before any selection has taken place, m_i is the number of used selection stages and p_{ij} is the proportion of individuals being selected at stage j of path i from the number of individual having been tested at stage j .

The percentage of selected individuals at each stage was varied between one and 100% in steps of 5%. These proportion are denoted in the following as PG_{HD} for the proportion selected at $GEBV_{HD}$ stage and PG_{LD} for the proportion selected at the $GEBV_{LD}$ stage. Distinction concerning the selection path are made due to extending the abbreviation with the already mentioned abbreviations for the selection path (e.g. $PG_{HD,SD}$, $PG_{HD,DS}$, $PG_{LD,SD}$, $PG_{LD,DS}$). To fulfill equation 3.15 the value of the last stage having

a percentage of selected individuals $\leq 99\%$ was then erased and calculated as dependent variable. Only if the value calculated for this stage was ≤ 1 the breeding scheme was seen as valid to be used in further calculation steps. The valid stage combinations of all paths were totally cross classified over paths, and each of that combinations was seen as a possible breeding scheme.

The goal function to be maximised was the genetic gain per year. The only constraint was to keep the breeding costs below maximum costs being calculated such that no *GS* takes place in all paths and in the path *SD* a conventional progeny performance breeding scheme with 50 test bulls per year is used. Breeding costs were derived for each of the valid schemes, but stages within such combinations having a proportion of selected individual >0.99 were excluded. ΔG_a was calculated for all schemes fulfilling the cost constraint, and the scheme maximising this parameter was evaluated as the best solution for a given set of input parameters.

Regarding the system of original and derived paths for *SD* and *SS*, the optimisation process was carried out iteratively. In the first cycle, the used selection stages in the paths *SD* and *SS* were equal. If the final result maximising the goal function contained a selection stage of the path *SD* with a proportion selected ≥ 0.99 , and being located at the end of the selection process and not between two stages with such a proportion <0.99 , then the calculation was restarted excluding this stage from the path *SD* but still using it in the path *SS*. This variability allowed for calculating breeding schemes such that cow sires are selected on $GEBV_{LD}$ and bull sires additionally on $GEBV_{HD}$.

Across all 900 possible combinations of input parameters, the total amount of evaluated breeding plans in terms of overall breeding costs was 146 million. 6.7 million of these fulfilled the cost constraint and were compared concerning the genetic gain per year.

3.2.6 Dynamic Pedigree Information Content

Since every selected candidate becomes an ancestor of its progeny, selected candidates of every selection path occur as ancestors in the pedigree of each selection candidate independent of the path. Furthermore, the exclusion of stages during the maximisation process changes the amount of information gathered about the selection candidate (e.g. $GEBV_{HD}$ and/or $GEBV_{LD}$). If a changed selection strategy in one path alters the amount of information about the selection candidate, the information content of the pedigree being used for selection candidates in other path might be altered too. This information content is altered additionally if a changed selection strategy of a certain path is changing its generation interval. In this case, it has to be regarded that at the time point of selecting the candidate some possible informants don't have performance data (e.g. daughters of the sire). Thus, if the selection strategy in the path SD and SS is changed from GS_{HD} to GS_{LD} , the pedigrees of all selection paths are altered.

The main impact of the described dynamics of the pedigree information content is a variable accuracy of the pedigree derived breeding values. The maximum amount of data at each selection stage with respect to the selection path is shown in table 3.2. This amount was adapted according to the used stages in each path.

3.2.7 Breeding Scheme Similarity Indicator

In order to analyse the effect of variations of C_{GEBV} and r_{GEBV} on the structure of breeding schemes in terms of changes of the proportion selected at certain selection stages within a selection path without a variety of diagrams and tables, we used a correlation like similarity measurement to reflect these changes. It was constructed as follows: Let x_{ij} be a vector of the selection path j within breeding scheme i with a dimension of the possibly, but not necessarily used, selection stages within that path

except pedigree stage. The values within x_{ij} are the genotyped proportions of the initial selection candidates at GS_{HD} and GS_{LD} stage within j , where not used stages are set to zero. Furthermore, let $x_i = \{x_{i1}, \dots, x_{in}\}$ be a vector out of concatenating all x_{ij} vectors of the breeding scheme i , and \bar{x}_i the mean of x_i . The similarity between the structure of breeding schemes i and z , S_{iz} , was then calculated as follows:

$$S_{iz} = \frac{\frac{x'_i x_z}{n} - \bar{x}_i \bar{x}_z}{\sqrt{\left(\frac{x'_i x_i}{n} - \bar{x}_i^2\right)\left(\frac{x'_z x_z}{n} - \bar{x}_z^2\right)}} \quad (3.16)$$

where n is the dimension of x_i and x_z . The possible values of S_{iz} range from minus one to one, where a value of one indicates a absolute similarity between two breeding schemes concerning its proportion selected at each possible selection stage.

3.2.8 Software

All calculations were carried out with a FORTRAN 90 program written by the first author. For calculating the selection intensity in a multistage breeding program, FORTRAN routines of Genz (1992) and Brent (1972) were used.

3.3 Results

3.3.1 Optimisation Results

Similarities among optimum breeding schemes independent of costs ($C_{GEBV,HD}$, $C_{GEBV,LD}$) and accuracies ($r_{GEBV,HD}$, $r_{GEBV,LD}$) of genomically estimated breeding values could be found for the following results. In general no additional information was gathered to select bull sires from the finally selected cow sires. Thus, breeding schemes selecting cow sires on the basis of $GEBV_{LD}$ and bull sires on $GEBV_{HD}$ were not found to maximise

the genetic gain per year. Since selecting bull dams only on pedigree information was never an option in optimum breeding schemes, one kind of GS was always applied to that path leading in turn to the fact that the highest proportion of the overall breeding costs was always generated due to genotyping bull dams. The contributions of the different selection paths to the overall genetic gain were always in the following order: $\Delta G_{SS} > \Delta G_{SD} > \Delta G_{DS}$.

Analysing the breeding schemes maximising ΔG_a of each input parameter combination, table 3.3 summarizes the results of the ΔG_a , ΔG_{SS} , ΔG_{SD} , ΔG_{DS} , absolute and relative total breeding costs, and the proportion of genotyped selection candidates in the path SS ($PG_{HD,SD}, PG_{LD,SD}$) and DS ($PG_{HD,DS}, PG_{LD,DS}$) in terms of mean, maximum and minimum. Furthermore, the concrete values for a reference scenario with a given parameter combination of $r_{GEBV,HD}=0.75$, $r_{GEBV,LD}=0.6$, $C_{GEBV,HD}=210$ € and $C_{GEBV,LD}=100$ € are given in the last column of table 3.3. The calculated ΔG_a varied between 223 and 161 kg milk, which equals 0.64 and 0.46 genetic standard deviations, respectively. The genetic gain was mainly generated due to the selection of sires, whereas the contribution of bull dams never exceeded 31 %. Independent of C_{GEBV} and r_{GEBV} , GS_{HD} was always used to select cow sires, whereas the inclusion of an additional preselection on GS_{LD} was affected by the concrete parameter combinations. On the contrary, for the path DS breeding schemes were found suggesting selection on both, GS_{LD} and GS_{HD} , or excluding one of these. The highest proportion of male selection candidates being high density genotyped (≥ 0.95) were found across the whole range of accuracies for $GEBV_{HD}$, but were accompanied by a minimum difference in accuracy of 0.15 to $GEBV_{LD}$ and the abandon of low density genotyping in the path SD . On the contrary, this proportion was lowest when $r_{GEBV,HD}$ was equal to 0.70 and $r_{GEBV,LD}$ equal to 0.65. Since this proportion varied between 0.285 and 0.903 when $PG_{LD,SD}$ was at its maximum of 0.95, a high number of low density genotyped male selection candidates

gives no hint to the extent of the successive selection intensity at the $GEBV_{HD}$ stage.

The initial number of selection candidates in the path DS was 100 times higher than in the path SD leading to a much smaller number of genotyped individuals in the path $DS(PG_{LD,DS}, PG_{HD,DS})$ compared to those of SD due to cost limitations. The levels of $PG_{LD,DS}$ and $PG_{HD,DS}$ were much more affected by cost of the corresponding $GEBVs$ than in path SD . Since for selecting bull dams GS_{HD} was mostly combined with GS_{LD} , a maximum number of 2650 high density genotyped heifers could only be obtained if the cost for both $GEBVs$ remained at its lowest possible value (150 and 20 €), and not more than 10,000 individuals were pregenotyped with a low density SNP chip. An increase of $C_{GEBV,LD}$ lead to abandon the selection of bull dams on $GEBV_{LD}$ if the costs exceed a certain level, where this level decreases with an increasing difference between the accuracies of $GEBVs$. For example even if $C_{GEBV,HD}$ is equal to 250 €, bull dams should be solely selected on $GEBV_{HD}$ given a $r_{GEBV,HD}$ of 0.9 if the accuracy of $GEBV_{LD}$ is ≤ 0.5 . If at the above given costs the accuracy of $GEBV_{HD}$ drops to 0.75, GS_{LD} was not seen to be useful for selecting bull dams as long as $r_{GEBV,LD}$ remained at 0.4. Highest levels of $PG_{LD,DS}$ of 0.55, meaning low density genotyping 27,500 heifers, were only possible at lowest genotyping cost for $GEBV_{LD}$ in combination with excluding GS_{HD} from the path DS , where this exclusion and, therefore, the selection of bull dams only on $GEBV_{LD}$ required a minimum accuracy of $GEBV_{LD}$ of 0.55.

Minimum breeding costs regarding the goal function could be obtained when males and females were selected only on $GEBV_{HD}$, its costs were lowest and almost all male selection candidates ($PG_{HD,SD} \geq 0.95$) but only 5% of the female selection candidates were genotyped. The precondition for excluding GS_{LD} even from the path DS was a minimum difference in r_{GEBV} of 0.3 to 0.35 dependent on $r_{GEBV,HD}$. Such an extreme composition of a breeding scheme occurred especially, when the marginal costs of genotyping 5% more females already led to an overdraw of the cost limitation and a shift to

low density genotyping of females triggered a decreased genetic gain per year.

The maximisation of the genetic gain per year for the reference scenario suggested genotyping only a minor proportion of the potential bull dams with a low density SNP chip, an exclusion of GS_{HD} from the path DS and the selection of potential bull sires on the basis of their $GEBV_{LD}$ and $GEBV_{HD}$. In absolute numbers, the breeding scheme maximising the genetic gain per year for the given parameter combination suggests genotyping the best 475 out of 500 male selection candidates according to their pedigree index with the low density chip and subsequently genotyping 238 of these 475 with a high density chip. On the contrary, only 5,000 out of 50,000 heifers should be used for selection on $GEBV_{LD}$. Breeding cost of almost 100% of the maximum breeding cost were generated mainly due to genotyping females (70%). The 30% resulting from male calve selection can be divided into cost for estimating $GEBV_{LD}$, $GEBV_{HD}$ and purchasing cost (each 22%), and expenditures for compensation payments to farmers (31%). The genetic gain was generated mainly in the paths SS and SD , whereas the path DS contributed only 25%.

3.3.2 Sensitivity of Maximisation Results to Changes in Accuracy and Cost

Figure 3.2 reflect the distribution of ΔG_{SD} and ΔG_{DS} and figure 3.3 the distribution of ΔG_a as functions of cost and accuracies of $GEBV$ such that all but one parameter of $r_{GEBV,HD}$, $r_{GEBV,LD}$, $C_{GEBV,LD}$ and $C_{GEBV,HD}$ are varied across all possible values and the parameter given on the x-axis is fixed. Thus, the distribution of for example ΔG_a is the conditional distribution given the value of the x-axis. If these distributions are broad and show no trend across all possible values of the x-axis parameter, the outcome of the optimisation is not sensitive to a variation of this input parameter.

General trends in calculation results due to input parameter variation were only found

Table 3.3: Parameters of results across accuracies and costs of genomic estimated breeding values

	\bar{x}	max	min	reference ¹
ΔG_a (%)	191.21 (55 %)	223.97 (64 %)	161.38 (46 %)	178.50 (51 %)
ΔG_{SS} (%)	745.21 (39 %)	839.83 (41 %)	647.30 (36 %)	701.52 (39 %)
ΔG_{SD} (%)	675.21 (35 %)	762.83 (37 %)	582.35 (33 %)	637.18 (36 %)
ΔG_{DS} (%)	491.70 (26 %)	639.65 (31 %)	373.02 (21 %)	446.27 (25 %)
$PG_{HD,SD}$	0.46	1	0.06	0.475
$PG_{HD,DS}$	0.03	0.05	0	0
$PG_{LD,SD}$	0.56	0.95	0	0.95
$PG_{LD,DS}$	0.1	0.55	0	0.10
absolute and relative total breeding costs	705,091 (98 %)	719,050 (100 %)	566,675 (79 %)	717,800 (99 %)

ΔG_a : genetic gain per year and as proportion of the additive genetic standard deviation, $PG_{HD,SD}$: proportion of the initial selection in the path “sire-dam” being genotyped with a high density SNP chip, $PG_{LD,SD}$: proportion of the initial selection in the path “sire-dam” being genotyped with a low density SNP chip, $PG_{HD,DS}$: proportion of the initial selection in the path “dam-sire” being genotyped with a high density SNP chip, $PG_{LD,DS}$: proportion of the initial selection in the path “dam-sire” being genotyped with a low density SNP chip, ΔG_{SS} : genetic gain per generation in the path “sire-sire” and as proportion of the total genetic gain, ΔG_{SD} : genetic gain per generation in the path “sire-dam” and as proportion of the total genetic gain, ΔG_{DS} : genetic gain per generation in the path “dam-sire” and as proportion of the total genetic gain, 1: calculation results for a parameter combination of $r_{GEBV,HD}=0.75$, $r_{GEBV,LD}=0.6$, $C_{GEBV,HD}=210$ € and $C_{GEBV,LD}=100$ €.

for ΔG_a and ΔG_{DS} as functions of r_{GEBV} and C_{GEBV} (see figures 3.2(c), 3.2(d), 3.3), and ΔG_{SD} as a function of $r_{GEBV,HD}$ (see figure 3.2(b)). In detail, increasing costs for $C_{GEBV,LD}$ decreased ΔG_{DS} much faster compared to $C_{GEBV,HD}$. The same effect could be observed less distinctive for ΔG_a . On the contrary, r_{GEBV} had an increasing effect on the genetic gain. ΔG_{SD} was strong positively affected by a higher $r_{GEBV,HD}$ but unaffected by a more accurate $GEBV_{LD}$. Furthermore, the effect of the accuracies of both $GEBV$ on ΔG_{DS} was positive to a similar extend. In result, ΔG_a was positively affected too, but it was much more sensitive to an increasing $r_{GEBV,HD}$ than to an increasing $r_{GEBV,LD}$. Although no effect of $C_{GEBV,LD}$, $C_{GEBV,HD}$ and $r_{GEBV,LD}$ on the genetic gain per generation of the path SD could be found, the figures 3.2(a) and 3.2(b) show a clear clustering of calculation results. These clusters emphasis the importance of the accuracy of $GEBV_{HD}$ for the genetic gain in this path since the breeding scheme being grouped in one cluster have this accuracy in common.

Reorganising the breeding scheme in terms of adjusting the extent genotyped individuals in each selection path regarding the cost constraint is a strategy to decrease the changes in calculation outcomes in terms of ΔG_a , ΔG_{SD} and ΔG_{DS} induced by variations of costs and accuracies of $GEBV$. The magnitude of this reorganisation process is reflected by a similarity of breeding schemes structures before and after reorganisation. Figure 3.4 visualises the development of the mean of this similarity between optimum breeding schemes differing only in one input parameter. These similarities were calculated between a optimum basic scheme characterised by a input parameter combination and optimum schemes differing from this basic scheme only in one single input parameter (the basic parameter). Thus, in figure 3.4(a) the input parameter allowed to be different to the basic scheme is given at the abscissa, and the right ordinate of each sub-diagram shows the value of the basic input parameter being used in the basic scheme. For example the upper diagram of figure 3.4(a) shows the similarity between a basic breeding

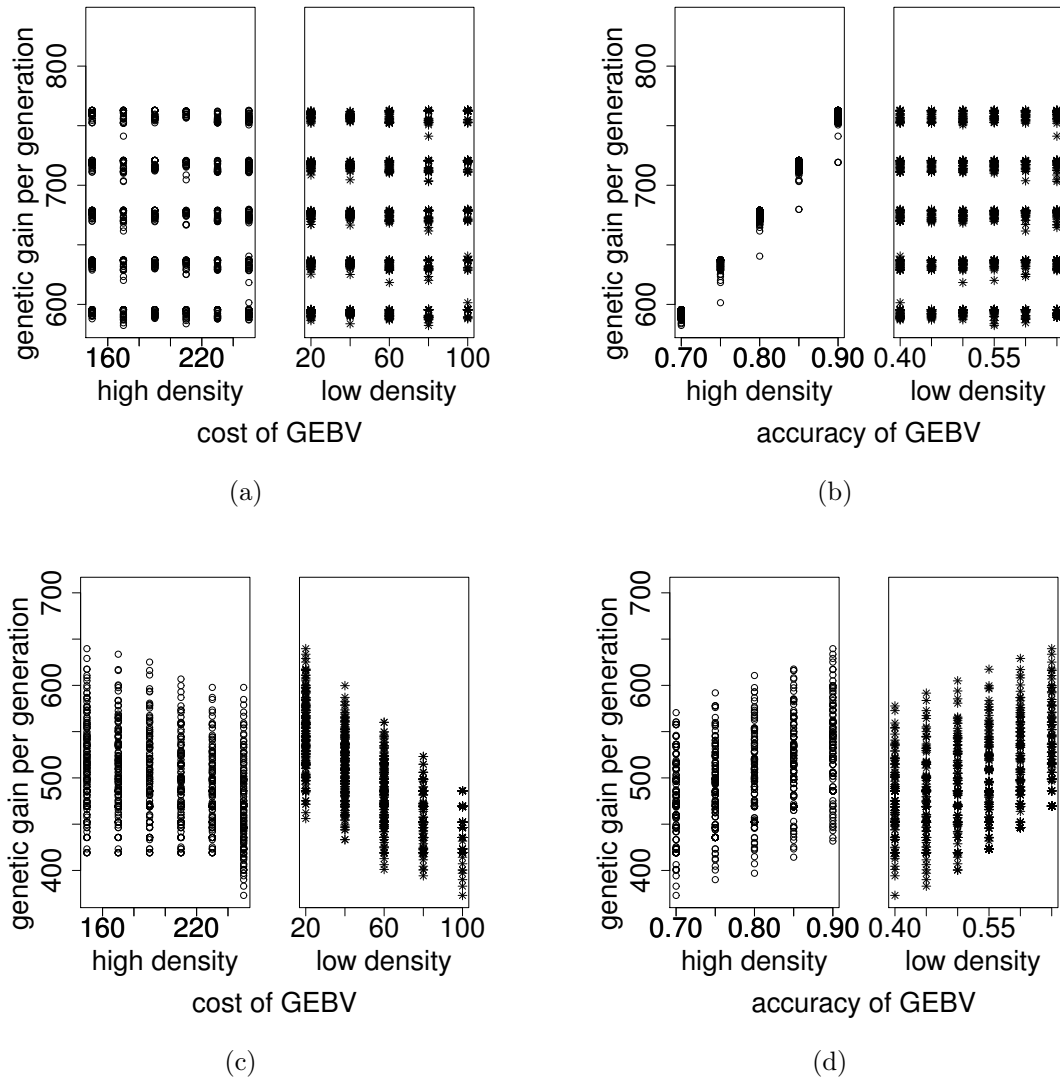


Figure 3.2: The genetic gain per generation as a function of costs (3.2(a) and 3.2(c)) and accuracies (3.2(b) and 3.2(d)) of genomically estimated breeding values. Only those breeding schemes are plotted maximising the genetic gain per year for a given combination of cost and accuracy of the genomically estimated breeding values from high density and low density SNP chips. Figure 3.2(a) and 3.2(b): results for the path “sire-dam”. Figure 3.2(c) and 3.2(d): results for the path “dam-sire”.

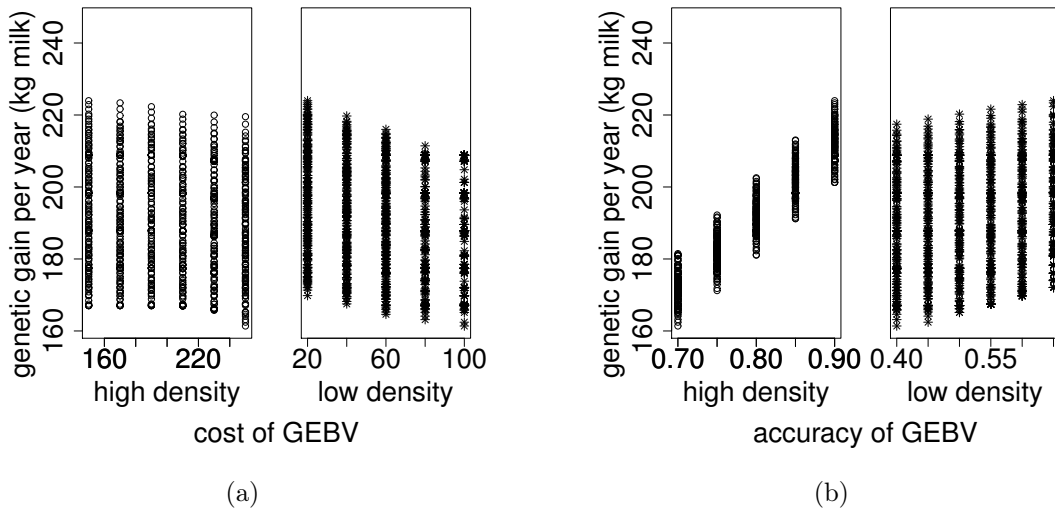


Figure 3.3: The genetic gain per year as a function of costs (3.3(a)) and accuracies (3.3(b)) of genomically estimated breeding values. Only those breeding schemes are plotted maximising the genetic gain per year for a given combination of cost and accuracy of the genomically estimated breeding values from high density and low density SNP chips

scheme characterised by $C_{GEBV,LD}$ of 20 € and all schemes using the same input values as the basic scheme for $r_{GEBV,HD}$, $r_{GEBV,LD}$, and $C_{GEBV,HD}$, but differ from the basic scheme in $C_{GEBV,LD}$. The slope of the mean curve can be seen as indicators for the sensitivity of breeding scheme structures to a change of a single input parameter value. In general, no differences could be found regarding the similarity between breeding scheme as a function of the basic parameter. Thus, transforming the graphs of each sub-figure of figure 3.4 into a similarity matrix with the column and row dimension of the number of graphs, it would be symmetrical. Breeding scheme structures are relatively insensitive to a change of $C_{GEBV,LD}$ up to 40 €, but have to be reorganised if this cost level is exceeded. Roughly the same result is found for $r_{GEBV,LD}$, but breeding scheme structures react to any increase in the accuracy of $GEBV_{LD}$ with a reorganisation. On the contrary, and counter-intuitive, breeding scheme structures remain relatively unchanged if the costs of $GEBV_{HD}$ increase, and seem to be more robust against an increase of

$r_{GEBV,HD}$ compared to $r_{GEBV,LD}$.

3.3.3 Sensitivity of Maximisation Results to Changes of Breeding Scheme Structures

As mentioned before, the sensitivity of breeding schemes structures to a shift in input parameter values is induced by a need to reorganise this structure in order to maximise the genetic gain per year even if input parameters change. Beside the risk of changing input parameters there is also a risk of strategical management mistakes in terms of choosing initially a already improper breeding scheme structure not maximising the genetic gain. Thus, the sensitivity of the genetic gain per year to mis-structuring breeding schemes given a certain input parameter set might be of interest. As in our model calculation PG_{HD} is directly affected by PG_{LD} , this sensitivity can be described more precisely as the sensitivity of the goal function concerning PG_{LD} , and a sensitivity concerning PG_{HD} within a given PG_{LD} . Figure 3.5 reflects this sensitivity in terms of the genetic gain per year and generation in the path SD for the reference scenario. Breeding schemes whose genetic gain is plotted in figures 3.5(a) and 3.5(b) were chosen from all breeding schemes calculated for the reference scenario such that for a given $PG_{LD,SD}$ the breeding scheme is selected maximising the genetic gain per year. Thus, only the number of male calves to be genotyped with a low density SNP chip was fixed, but $PG_{HD,SD}$, $PG_{LD,DS}$ and $PG_{HD,DS}$ could still be optimised. The shape of the curves suggest that up to a total genotyping of selection candidates an increase of $PG_{LD,SD}$ can generate always an additional genetic gain, but with a range between 0.35 and 1.00 the surplus will not exceed 2%. Additionally, since a $PG_{LD,SD}$ of zero indicates genotyping all male selection candidates with high density SNP chips, the sole usage of $GEBV_{HD}$ as genomic selection basis might generate the same genetic gain.

The sensitivity of ΔG_a to the proportion of high density genotyped initial male selec-

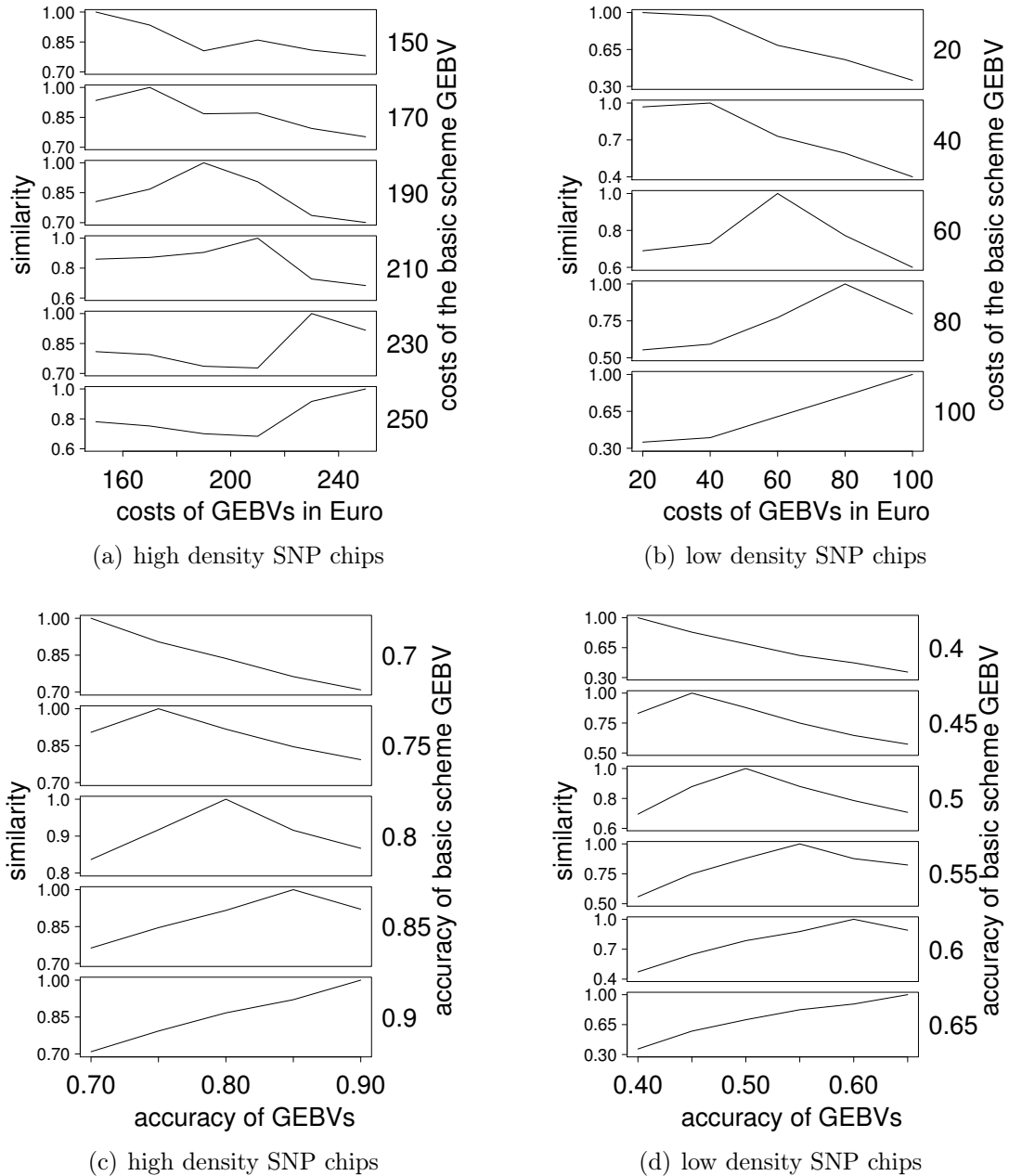


Figure 3.4: The mean of the similarity (correlation) between breeding schemes structures. The similarity graph given in each sub-sub-diagram was calculated only between a breeding scheme having a value for the abscissa parameter (the basic parameter) given at the right ordinate (the basic scheme), and breeding schemes differing in their input parameter combination from the basic scheme only in the basic parameter, where the actual value of the basic parameter for this schemes is given at the abscissa. The basic parameters are the cost for genomically estimated breeding values out of high density chips (3.4(a)) and the resulting accuracy (3.4(c)), as well as the cost for genomically estimated breeding values out of low density chips (3.4(b)) and the resulting accuracy (3.4(d)).

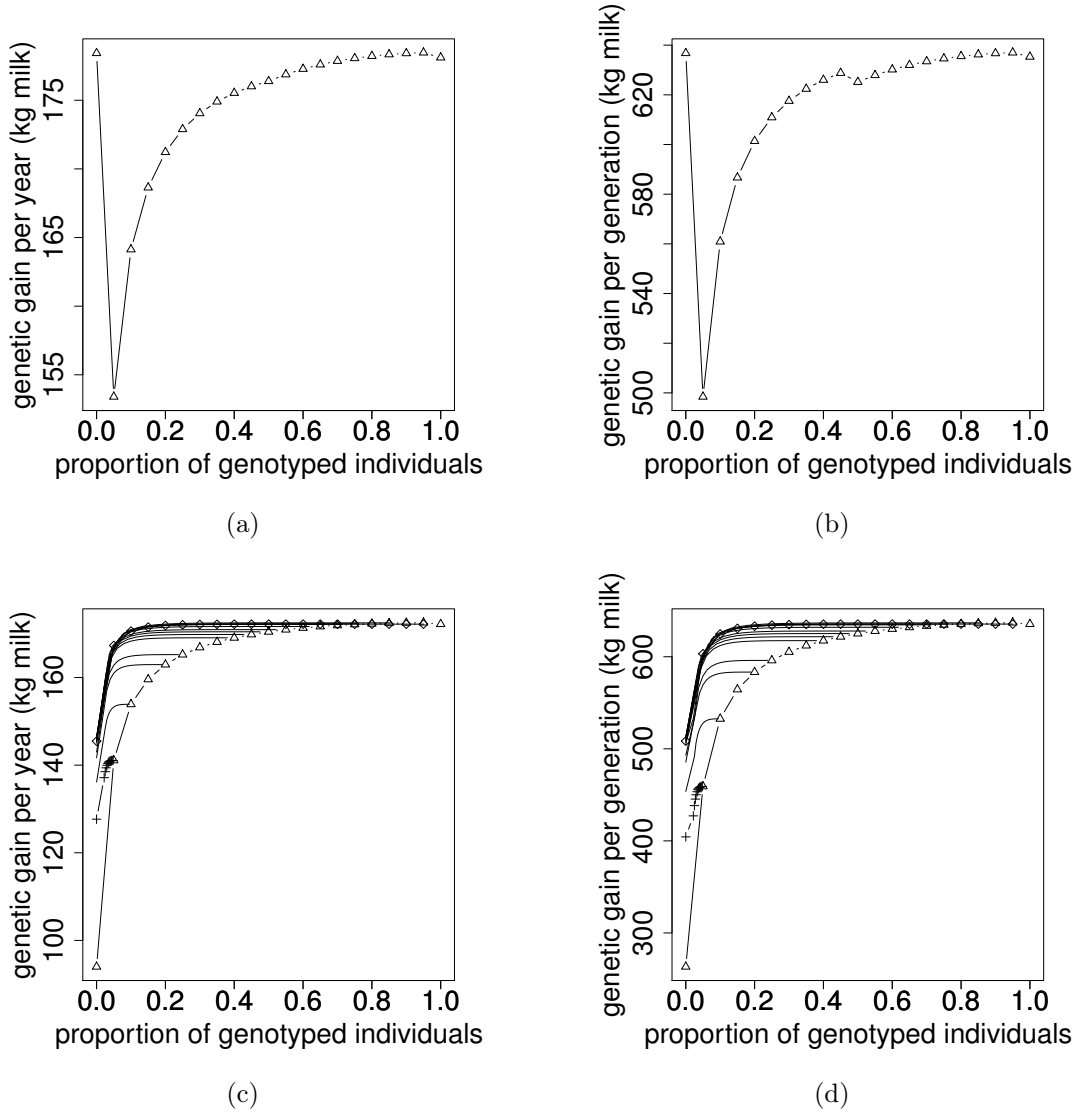


Figure 3.5: Genetic gain per generation and year of breeding schemes calculated using an input parameter set equal to the reference scenario. Figures 3.5(a) and 3.5(b) show the outcomes of breeding schemes maximising ΔG_a and ΔG_{SD} for a given $PG_{LD,SD}$. Thus, selection intensities of all other selection stages within paths could be chosen such that ΔG_a is maximised, where $PG_{LD,SD}$ was the abscissa parameter. Figures 3.5(c) and 3.5(d) show the outcomes of breeding schemes maximising ΔG_a and ΔG_{SD} for a given $PG_{HD,SD}$ within a given $PG_{LD,SD}$. On the contrary to 3.5(a) and 3.5(b), the selection intensity of the high density SNP chip in path SD could not be chosen in a above mentioned manner because it was the abscissa parameter, and $PG_{LD,DS}$ was equal to 0.05 and $PG_{HD,DS}$ equal to zero. One graph in 3.5(c) and 3.5(d) shows the results for the variation of $PG_{HD,SD}$ within a give $PG_{LD,SD}$, where only graphs with an underlying $PG_{LD,SD}$ of zero (Δ), 0.05 ($+$) and 0.95 (\diamond) are marked.

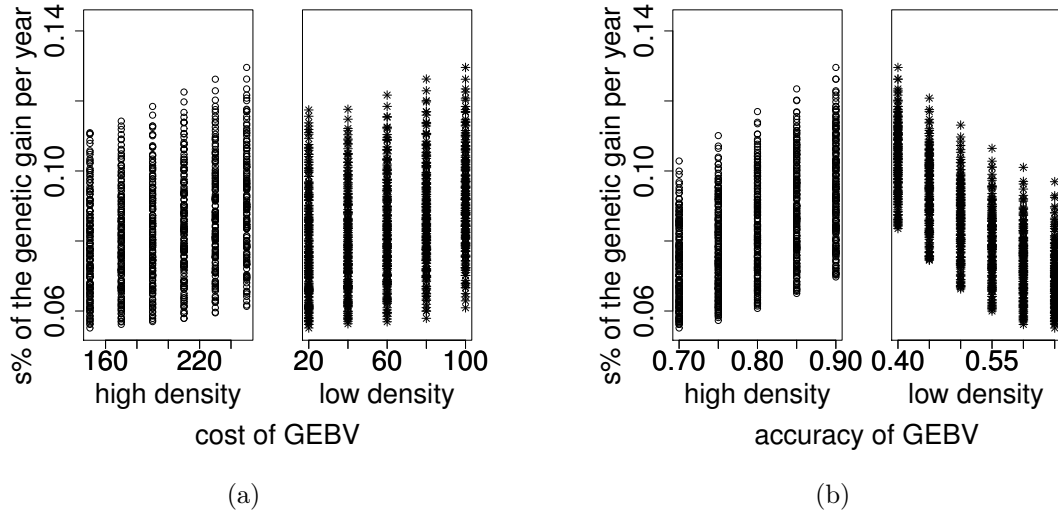


Figure 3.6: The variation coefficient (s%) of ΔG_a of all breeding schemes calculated for a given combination of input parameters as a function of the variation the costs (3.6(a)) and accuracies (3.6(b)) of genomically estimated breeding values

tion candidates given a $PG_{LD,DS}$ of 0.05 and $PG_{HD,DS}$ of zero is illustrated by figures 3.5(c) and 3.5(d). Each curve represents a certain $PG_{LD,SD}$ value. Curves for a $PG_{LD,SD}$ between 0.1 and 0.9 are not marked. As the parameter $PG_{LD,SD}$ is fixed for each curve, $PG_{HD,SD}$ can be chosen conditional the marginal constraint given in formula 3.15. Thus, the range of possible values for $PG_{HD,SD}$ is a function of $PG_{LD,SD}$ leading to short curves if this parameter is small. A maximum genetic gain per year of 172.53 kg is achievable if no low density genotyping is applied, and a GEV_{HD} is estimated for 95% of the initial male selection candidates. If a loss of 2% (-3.45 kg) in the genetic gain is accepted, this proportion can be decreased to 40%, and to 25% if this loss can reach up to 5% (-8.62 kg). The curve representing an underlining $PG_{LD,SD}$ of zero can be regarded as the most sensitive for a shift of $PG_{HD,SD}$. As the above mentioned loss in genetic gain per year assumes this curve, this loss is less when calculated for curves representing a underlying $PG_{LD,SD}$ exceeding zero.

To visualise a possible interaction of risks based on the uncertainty of future devel-

opments of input parameters and risks out of resource miss-allocation, the variation coefficient of the genetic gain per year of all calculated breeding schemes within each input parameter combination is plotted in figure 3.6 as a function of the varied parameters. More accurately estimated $GEBV_{HD}$ as well as higher $C_{GEBV,HD}$ and $C_{GEBV,LD}$ increased the risk of miss-allocation, whereas more accurately estimated $GEBV_{LD}$ decrease it. Relating this findings to figures 3.5(b) and 3.5(d) it can be determined that the spread of the curves within the diagrams becomes less dense if C_{GEBV} or $r_{GEBV,HD}$ is increased.

3.4 Discussion

Several authors found using low density SNP chips in conjunction with imputation algorithms to be a strategy for applying genomic selection to species where females have low reproductive capacities (Habier *et al.*, 2009; VanRaden, 2010). In dairy cattle, this might be especially the case for selecting bull dams, since the relatively low number of yearly recruited cow and bull sires, and the advantages out of the shortened generation interval justify even high expenses for genotyping individuals. One of the major difficulties for the application of *GS* in the path *DS* is the necessity of a large number of dams for producing male selection candidates making a high selection intensity almost impossible if *GS* cost per individual are high and economical resources are limited. This might be also the case when using embryo transfer techniques in order to decrease this number because its costs may overdraw the economical resources as well as increasing the number of genotyped individuals. Since cheap low density SNP chips allow for genotyping large proportion of potential bull dams, a *GEBV* with a sufficient accuracy may generate a higher genetic gain compared to using expensive high density SNP panels. This assumption is supported by the outcomes of this study, clearly showing that the selection path “dam-sire” is the major application field of low density genotyping.

The results of our calculations highlighting the non-linearity of the genetic gain as a function of the selection intensity, and, therefore, the proportion selected because we found in more than a half of all parameter combination a three stage selection of bull dams on the basis of pedigree information, $GEBV_{LD}$ and $GEBV_{HD}$ the most rewarding strategy in terms of genetic gain per year. Thus, if a sufficiently accurate pedigree pre-selection takes place, an extend of GS_{LD} above a certain limit will generate less genetic gain than using an additional GS_{HD} stage with lowest selection intensity. Additionally, due to such multi stage selection procedures a given genetic gain per generation in a certain selection path can be achieved at less expenses, or even a higher genetic gain per year at the same expenses due to allocating investment resources regarding the marginal benefit of an additional selection intensity. This is exemplified by the usage of GS_{LD} for selecting male calves in combination with GS_{HD} , being only a result of the possible cost reductions in the path SD , and not of a higher genetic gain due to including another information source in the selection process. The available economic resources could then be used to increase the GS_{LD} selection intensity or implementing an additional GS_{HD} stage in path DS leading to an increased ΔG_{DS} and a higher ΔG_a . These outcomes are in a sharp contrast to ideas of genotyping even whole sub-populations in order to identify individuals with highest $GEBVs$ neglecting the non-linearity of the genetic gain as a function of the selection intensity (König & Swalve, 2009).

The fact that the inclusion of GS_{LD} in the path SD and, therefore, the availability of an additional information source correlated with the breeding goal was not leading to a higher ΔG_{SD} results from the covariance structure between $GEBV_{HD}$ and $GEBV_{LD}$. On the contrary to multistage selection processes were the covariances between the information gathered at different stages are smaller than its variances, and an exclusion of information sources will decrease the maximum achievable genetic gain, the genetic gain from a multistage selection process including only GS_{LD} and GS_{HD} will never

exceed the one generated by a single stage selection procedure using only GS_{HD} because the covariance between $GEBV_{HD}$ and $GEBV_{LD}$ is equal to the variance of the latter.

In this study, we found even for low $r_{GEBV,HD}$ of 0.7 and $r_{GEBV,LD}$ of 0.4 a genetic gain per year of 0.47 genetic standard deviations. This is still higher compared to other published results assuming higher accuracies of $GEBVs$ (Lillehammer *et al.*, 2010; Schaeffer, 2006), and might be induced by the very high selection intensities for bull sires and cow sires assumed in our calculations and the assumptions of a minimum $r_{GEBV,HD}$ of 0.7. Without regarding cost limitations, Schaeffer (2006) found the selection path DS to become the most important assuming to have available $GEBVs$ for the whole cow population being as accurate as those estimated for selecting male calves, and predicted a large decrease in breeding costs. Under restricted cost, however, this might be not possible due to limiting the number of genotyped potential bull dams. As in our calculations genotyping all potential bull dams was not possible due to cost limitations even with very cheap low density SNP chips, the genetic gain generated in the path DS was never exceeding the one generated in the paths SS and SD . Furthermore, as the need to genotype large proportions of potential bull dams was found to be most cost producing, the application of GS in the path DS might lead to breeding schemes being as expensive as progeny testing schemes as long as the genetic gain per year is the maximisation criterion.

The above mentioned problem of genomic selection to generate a sufficient genetic gain in the path DS because of low selection intensities when economic resources are limited is the major force behind the sensitivity of breeding schemes concerning the genetic gain per year, genetic gain per generation and the breeding scheme structure to changes in r_{GEBV} and C_{GEBV} . In order to maximise ΔG_a under restricted overall costs and an increasing C_{GEBV} the proportion of genotyped individuals can be reduced. As shown, ΔG_{SD} as a function of $PG_{HD,SD}$ and $PG_{HD,SD}$ within $PG_{LD,SD}$ is a plateau

above a certain selection intensity, and, therefore, an increasing selection intensity has a diminishing marginal utility in terms of the genetic gain. Thus, in those paths where the goal function has such a shape the selection intensity at the genomic selection stages can be reduced without a general loss in genetic gain if it is already above a certain limit. ΔG_{SD} is, therefore, almost not affected by increasing C_{GEBV} because a almost similar outcome can be achieved with less expenditures. On the contrary, the selection intensity in the path DS has almost levels where the marginal benefit of an increasing proportion of genotyped individuals in terms of genetic gain is still high. Thus, a reduction of this proportion induced by increasing genotyping costs will lead immediately to a decrease in ΔG_{DS} . As the ΔG_a is the sum over the genetic gains per generation from different selection paths weighted with the generation interval (Rendel & Robertson, 1950), the effect of increasing C_{GEBV} on it was not as high as on ΔG_{DS} but higher than on ΔG_{SD} . The sensitivity of ΔG_{SD} , ΔG_{DS} and ΔG_a to changes of the $GEBV$ accuracy is influenced by extent of using different SNP chips in each path. ΔG_{SD} is generated mainly due to applying GS_{HD} and, therefore, the accuracy of $GEBV_{LD}$ has no effects, whereas an increasing $r_{GEBV,HD}$ clearly leads to a higher genetic gain per generation. Since in many parameter combinations $GEBV_{HD}$ and $GEBV_{LD}$ were used to select females, increases in accuracies of both breeding values effected ΔG_{DS} positively. Comparing ΔG_{SD} and ΔG_{DS} as functions of $r_{GEBV,HD}$, ΔG_{SD} is much more sensitive to a change in $r_{GEBV,HD}$ compared to ΔG_{DS} , which is a result of a higher selection intensity in the path SD . Trends of ΔG_a as a function of r_{GEBV} are within the trends found for ΔG_{SD} and ΔG_{DS} .

A direct result from the given interactions between genetic gain, accuracies of breeding values and cost of genomic selection is the sensitivity of breeding scheme structures to changes of all four variable input parameters. The highest pressure on breeding scheme structure was induced due to changes of $r_{GEBV,LD}$ and $C_{GEBV,LD}$, whereas the need of reorganisation in order to maximise the genetic gain per year was not that high if

$r_{GEBV,HD}$ and $C_{GEBV,HD}$ changed. This is caused due to the facts that low density SNP chips are mainly used in the path DS , a reduction of the selection intensity in this path has almost immediate effects on the genetic gain and the high number of genotyped individuals will easily increase the breeding costs above the given limitation if C_{GEBV} increase. Additionally, as the variation step for the proportion selected (5 %) in the path DS is equal to genotyping 2500 more or less individuals, a decrease in the number of genotyped females of 5 % at a $C_{GEBV,LD}$ of 20 € sets free economical resources allowing for high density genotyping 40 % of the male selection candidates even at a $C_{GEBV,HD}$ of 250 €. Thus, increasing $GEBV_{LD}$ costs will lead to a decrease in $PG_{LD,DS}$, and by the same way to an increase in $PG_{HD,SD}$. On the contrary, higher cost for GS_{HD} will induce only a slightly reduction of $PG_{HD,SD}$ and no other changes. The same context is responsible for the sensitivity of breeding scheme structures to changes in r_{GEBV} .

Breeders might regard higher accuracies of genomically estimated breeding values positively. But our results also show that the risk of obtaining an suboptimal ΔG_a compared to other competitors is increased because its distribution enlarges with an enhanced $r_{GEBV,HD}$. Thus, the uncompetitiveness of suboptimal breeding scheme is increased the more genetic gain is generated due to using highly accurate $GEBVs$. On the contrary, higher accuracies of $GEBV_{LD}$ decrease this risk because the outcomes of the three main groups of selection strategies, two-stage including $GEBV_{LD}$, two-stage including $GEBV_{HD}$ and three-stage including $GEBV_{LD}$ and $GEBV_{HD}$, are becoming more similar. The increasing effect of higher cost for $GEBV_{LD}$ and $GEBV_{HD}$ results from excluding more breeding schemes where high proportions of the selection candidates are genotyped because of the cost constraint. Once a sufficient selection intensity has been reached, the additional genetic gain due to an additional proportion selected is rather small. Thus, breeding schemes yielding an almost equal genetic gain at high costs are excluded. Since the genetic gain of the remaining breeding schemes react very sensible to

changing the proportion selected, the distribution of possible outcomes in terms of ΔG_a is enlarged but remaining breeding schemes yield a higher benefit per invested capital unit on average.

3.5 Conclusions

Genomic selection using low density SNP chips is useful for an extensive genotyping of potential bull dams in dairy cattle breeding programs. This holds even if accuracies of genomically estimated breeding values are low. Since pedigree information will be available for all selection candidates, the unlimited extend of genotyping this candidates is not economical even if financial resources are unlimited. The limitation of economical resources leads to a strong interaction between the selection strategies of different selection paths. Thus, changes in accuracy and cost of genomically estimated breeding values will not only have an effect on the on the genetic gain generated in the path where the affected chip is used. Furthermore, breeding organisations have to reveal the structure of their selection schemes, reallocate financial resources between selection paths and adjust selection intensities in order to maintain competitiveness due to achieving the maximum genetic gain per year. Additionally, the non-linearity of the genetic gain as a function of the selection intensity offers the possibility of achieving an advantage in competition due to generating a given genetic gain at lowest costs.

3.6 Acknowledgement

This work was financed by by the German Federal Ministry for Education and Research (BMBF), project "FUGATO-Brain".

3.7 Bibliography

- Brent, R. P. (1972), *Algorithms for Minimisation without Derivatives*. Prentice Hall.
- Börner, V. & Reinsch, N. (2010), Decorrelated selection indices versus optimum selection indices in optimising multistage dairy cattle breeding schemes regarding genomic selection. *personal communications* .
- Calus, M. P. L. & Veerkamp, R. F. (2007), Accuracy of breeding values when using and ignoring the polygenic effect in genomic breeding value estimation with a marker density of one snp per cm. *J Anim Breed Genet* **124**(6), 362–368, URL <http://dx.doi.org/10.1111/j.1439-0388.2007.00691.x>.
- Daetwyler, H. D., Villanueva, B., Bijma, P. & Woolliams, J. A. (2007), Inbreeding in genome-wide selection. *J Anim Breed Genet* **124**(6), 369–376, URL <http://dx.doi.org/10.1111/j.1439-0388.2007.00693.x>.
- Dekkers, J. C. M. (2007), Prediction of response to marker-assisted and genomic selection using selection index theory. *J Anim Breed Genet* **124**(6), 331–341, URL <http://dx.doi.org/10.1111/j.1439-0388.2007.00701.x>.
- Genz, A. (1992), Numerical computation of multivariate normal probabilities. *J Comput Graph Stat* **1**, 141 – 149.
- Habier, D., Fernando, R. L. & Dekkers, J. C. M. (2009), Genomic selection using low-density marker panels. *Genetics* **182**(1), 343–353, URL <http://dx.doi.org/10.1534/genetics.108.100289>.
- Hayes, B. J., Bowman, P. J., Chamberlain, A. J. & Goddard, M. E. (2009), Invited review: Genomic selection in dairy cattle: progress and challenges. *J Dairy Sci* **92**(2), 433–443, URL <http://dx.doi.org/10.3168/jds.2008-1646>.

- Howie, B., Donnelly, P. & Marchini, J. (2009), A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**(6), e1000529, URL <http://dx.doi.org/10.1371/journal.pgen.1000529>.
- König, S. & Swalve, H. H. (2009), Application of selection index calculations to determine selection strategies in genomic breeding programs. *J Dairy Sci* **92**(10), 5292–5303, URL <http://dx.doi.org/10.3168/jds.2009-2232>.
- Lillehammer, M., Meuwissen, T. & Sonesson, A. (2010), Effects of alternative genomic selection breeding schemes on genetic gain in dairy cattle. In *Proceedings of the Ninth World Congress on Genetics Applied to Livestock Production*.
- Lynch, M. & Walsh, B. (1998), *Genetics and Analysis of Quantitative Traits*. Sinauer Associates Inc., U.S.
- Meuwissen, T. H., Hayes, B. J. & Goddard, M. E. (2001), Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**(4), 1819–1829.
- Muir, W. M. (2007), Comparison of genomic and traditional blup-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. *J Anim Breed Genet* **124**(6), 342–355, URL <http://dx.doi.org/10.1111/j.1439-0388.2007.00700.x>.
- NamKoong, G. (1970), Optimum allocation of selection intensity in two stages of truncation selection. *Biometrics* **26**, 465 – 476.
- Rendel, J. M. & Robertson, A. (1950), Estimation of genetic gain in milk yield by selection in a closed herd of dairy cattle. *J Genet* **50**, 1–8.
- Schaeffer, L. R. (2006), Strategy for applying genome-wide selection in dairy cattle. *J Anim Breed Genet* **123**(4), 218–223, URL <http://dx.doi.org/10.1111/j.1439-0388.2006.00595.x>.

Tallis, G. M. (1961), The moment generating function of the truncated multi-normal distribution. *J R Statist Soc B* **23**, 223 – 229.

VanRaden, P. (2010), Genomic evaluations with many more genotypes and phenotypes. In *Proceedings of the Ninth World Congress on Genetics Applied to Livestock Production*, hallo, p. 149.

Weigel, K., de los Campos, G., Vazquez, A., Rosa, C. V. T. G., Gianola, D., O’Connell, J., VanRaden, P. & Wiggans, G. (2010), Genomic selection and its effects on dairy cattle breeding programs. In *Proc. 9th WGCALP*.

3.8 Appendix

According to path coefficient methodology (Lynch & Walsh, 1998) and the approach of Dekkers (2007) a certain phenotype P can be decomposed as follows:

$$\begin{aligned}
 P &= G + E \\
 G &= Q + R \\
 Q &= \hat{Q}_i + e_i \\
 \hat{Q}_i &= \hat{Q}_j + e_j \\
 P &= E + R + e_i + \hat{Q}_j + e_j
 \end{aligned}
 \tag{3.17}$$

where G is the additive genetic effect, E is the rest effect, Q is the true marker breeding value, R is the polygenic effect not explainable by the markers, \hat{Q}_i is estimated marker breeding value using a high density SNP chip, e_i is the error due to estimating the marker effects, \hat{Q}_j is the estimated marker breeding value using a low density SNP and imputation algorithms chip and e_j is the imputation error.

Assuming a zero covariance between \hat{Q}_j and e_j , the correlation between \hat{Q}_i and \hat{Q}_j is

$$r_{(\hat{Q}_i\hat{Q}_j)} = \frac{\sigma_{(\hat{Q}_j+e_j,\hat{Q}_j)}}{\sigma_{(\hat{Q}_j+e_j)}\sigma_{(\hat{Q}_j)}} = \frac{\sigma_{(\hat{Q}_j)}^2}{\sigma_{(\hat{Q}_i)}\sigma_{(\hat{Q}_j)}} = \frac{\sigma_{(\hat{Q}_j)}}{\sigma_{(\hat{Q}_i)}}. \quad (3.18)$$

General Discussion

Gametic Gene Flow Method

In part one of this thesis the discounted gene flow method (Hill, 1974; McClintock & Cunningham, 1974) was extended to traits with imprinted or mixed Mendelian and imprinted inheritance. The flow of genes of founder individuals is traced across generations and tiers of the population with special regard to the probability that an individual has inherited this genes from its sire or dam by considering parent-specific parts of the genome (gametes) rather than genotypes. The summarised standard discounted expressions from maternally and paternally inherited genes can be used as weighting coefficients for breeding values as sire and as dam. The general idea of the extension was the definition of each tier within the population as a pool of genes, and dividing this pool into two subclasses, where one subclass contains all genes being maternally inherited and the other all genes being paternally inherited. The Markovian transmission matrix contains the transmission probabilities between tiers and subclasses due to reproduction and ageing.

The concept of extending the gene flow method due to dividing the tiers of a population on the genome level has been taken up by Ytournal & Simianer (2010) by defining subclasses for each possible haplotype and tracing the flow this haplotypes over generations in order to predict how selection measures affect gene frequencies over time. Both

related extension contribute to the further development of deterministic approaches for optimising animal breeding plans.

Modelling Genomic Selection in Animal Breeding Plans

Since genomically estimated breeding values (*GEBV*) are the sum over all estimated marker effects, stochastic breeding scheme simulations allow to regard genomic selection (*GS*) by modelling the QTL and marker genome of each individual within a population (Lillehammer *et al.*, 2010). This detailed approach might be useful for the prediction of the genetic gain over many generations from *GS* because the decay of the linkage disequilibrium between markers and QTLs can be regarded. However, the applicability is rather limited because of the computation time concerning the number of individuals, simulated marker and QTLs, and the number of generations. Furthermore, assumptions are necessary e.g. about the number of trait affecting QTLs, the size of QTL effects and crossing over probabilities. A possible simplification is the simulation of *GEBVs* for each individual by drawing from a certain distribution. This procedure requires assumptions about the reliability of such *GEBVs* if they estimate the aggregate genotype (Buch *et al.*, 2010; Winkelman & Spelman, 2010). If the *GEBVs* are trait-specific, additional assumptions are necessary concerning the correlation between trait *GEBVs*, and between the true breeding value and the *GEBV* of a trait. Then, by taking appropriate assumptions, the decay of linkage disequilibrium can be considered, but the computational demands are likely to limit the number of breeding schemes to be compared. Deterministic models of breeding schemes allow for a large number of breeding schemes in order to search for an optimum because distributional parameters allow to circumvent the simulation of individuals. However, the same problem as for the stochastic simulation occurs in deterministic calculations: the knowledge on correlations between trait specific *GEBVs* and true breeding values.

In part two and three of this thesis, an approach of Dekkers (2007) was used to include *GEBVs* in breeding calculations by modelling the individual *GEBV* as a trait with a heritability of one and a correlation to the breeding goal equal to the *GEBV* accuracy. In order to model the simultaneous application of low- and high-density SNP chips, this approach was extended in part three by assuming that the *GEBVs* estimated from high density genotyping can be decomposed into a *GEBV* obtained by transforming a low density genotype into a high density genotype of a somewhat reduced accuracy via imputation algorithms (Howie *et al.*, 2009; VanRaden, 2010), which implies a certain amount of imputation error. In case of a single trait - or the aggregate genotype - the correlation between two *GEBVs* for the same trait (from different kinds of SNP chips) can be derived from the correlation between the *GEBVs* and the true breeding value. In case of two or multiple traits, the variances of true *GBVs* (i.e. *GEBVs* without prediction error) has also to be known. The oblivion of these variances (e.g. König & Swalve (2009)), which is equal to capture the complete additive genetic variance by markers, leads to an underestimation of the correlation between *GEBVs* of traits, an overestimation of the accuracy of breeding values derived from index methodology including *GEBVs* as traits and, as a consequence, the genetic gain.

Decorrelated Selection Indices in Multistage Selection Schemes

Xu & Muir (1992) pointed out that according to Saxton (1989) the application of multi-dimensional integration techniques to multistage selection schemes for the derivation of the selection intensity of the aggregate genotype leads to an exponential increase of computational time with an increasing number of stages. Furthermore, Ducrocq & Colleau (1989) reported the stability of integration results of multivariate normal distributions to be dependent on the correlation structure between variates. Thus, regarding computer capacities and available algorithms at that time, the application of decorrelated

selection indices (Xu & Muir, 1992) for optimising simple multistage breeding schemes has been developed and applied, despite of suboptimal results. At the same time Genz (1992) presented a method for integrating multivariate standard normal distributions yielding, to the authors knowledge, fast and stable integration results independent of the correlation structure of variates. Although both methods were developed almost 20 years ago, the integration algorithm of Genz (1992) has, to the authors knowledge, not been extensively used in the animal breeding literature. This algorithm together with zero-finding procedures of Brent (1972) is well suited for calculating selection intensities in multistage breeding schemes (Mi & Utz, 2008). In conclusion, there is no justification for using decorrelated selection indices any longer, and, even for a large number of possible selection stages, exact integration methods should be used instead.

Applicability of Maximisation Techniques to Multistage Selection Schemes

In part 2 and 3, a grid search was applied in order to find optimum breeding plans. This grid search was structured in two layers, one for the stepwise variation of input parameters (cost and accuracies of *GEBVs*), and the other for the stepwise variation of selection intensities at successive stages in each selection path. The latter has the strongest effect on the grid density because the number of possible selection variants in each selection path depends on the number of possible selection stages and the step width used for selection intensities. As each additional selection path and stage increases the number of dimensions of the global optimisation problem, the number of breeding schemes to be evaluated increases dramatically even if the number of stages within paths is still small. A possibility to circumvent the evaluation of such a large number of selection schemes is the application of the multidimensional Newton method (Ducrocq & Colleau, 1989; Xu & Muir, 1992). The application of such maximisation technique requires an explicit and continuous goal function, and its first and second derivation

to be continuous. The genetic gain per generation of a selection path can be modelled as a continuous function of the stage selection intensities. The generation interval is, however, non-continuous because in multistage breeding schemes the generation interval depends on the number of stages. Thus, if the genetic gain per year is the goal function to be maximised and all selection stages are not mandatory, the multidimensional Newton method can not be used in a global maximisation approach since the goal function is non-continuous.

Instead of maximising the genetic gain per year, the genetic gain per invested monetary unit can also be used as a goal function (Xu & Muir, 1992). As breeding costs due to trait measurements can be attributed to certain selection stages and depend on the number of measured individuals, trait measurement cost at stage n can be modelled as linear function of number of selected individuals at stage $n-1$, and, therefore, as a function of the selection intensity at stage $n-1$ (NamKoong, 1970). Thus, the costs and the genetic gain are continuous functions of the stage selection intensities allowing for the application of the multidimensional Newton method. However, since the cost function calculates the breeding costs of a certain selection stage on the basis of its number of initial selection candidates, breeding costs at this stage will still be positive even if the selection intensity at this stage is zero indicating that this selection stage is not used. As this is a wrong result, correct results can only be obtained by transforming the continuous cost function in a non-continuous one, and, therefore, the application of the multidimensional Newton method again becomes impossible.

Pedigree data are a major information source for pre-selecting individuals independent of the character of the following selection stages (e.g. genomic selection or progeny testing). The amount of information about an ancestor or sibling occurring within the pedigree is directly affected by the applied selection strategy if informants had been a selection candidates too. Thus, changing the selection strategy of a certain selection path

within a maximisation process by excluding selection stages will change the accuracy of breeding values estimated from pedigree data of this and all other selection paths. In such a case the goal function is again non-continuous and cannot be maximised via multidimensional Newton method. The same applies to a proper consideration of the Bulmer effect and the derivation of the selection intensity for small numbers of selection candidates via order statistics.

Summarising the given facts, the grid search is, to the best knowledge of the author, the method of choice for the global optimisation of breeding schemes when selection stages are optional because it allows for adjusting the generation interval, the calculation of the exact breeding costs, the consideration of the Bulmer effect and the correction of the selection intensities for order statistics.

The Role of Genomic Selection in Multistage Dairy Cattle Breeding Programmes

Several studies on dairy cattle breeding schemes were carried out in order to compare genomic selection and progeny performance as a possible information sources, founding genomic selection to be always superior to conventional progeny testing (Buch *et al.*, 2010; König *et al.*, 2009; König & Swalve, 2009; Schaeffer, 2006; Winkelman & Spelman, 2010). A possible disadvantage of these studies might be the assumption of a constant accuracy of *GEBVs* over generations. As this accuracy is negatively affected by a decay of linkage disequilibrium, SNP effects have to be continuously re-estimated in successive generations. Usually highly accurate estimated breeding values from progeny tested bulls are used as phenotypes for this purpose. Without progeny testing estimated breeding values from progeny performance are only available for those bulls being selected as sires on the basis of their *GEBVs*. Then, only such individuals will enter the validation and

estimation sample. This might lead to a biased re-estimation of SNP effects leading a decreased genetic gain in the future (Lillehammer *et al.*, 2010). Although a proper consideration of such effects may decrease the advantage of breeding schemes being solely based on genomic selection, the results given in part 2 of this thesis suggests that accuracies of genomically estimated breeding values below of 0.45 are necessary to justify a combination of progeny performance testing and genomic selection.

Biological and Technical Parameters of Dairy Cattle Breeding Program Optimisations

The optimisation of breeding schemes carried out in part two and three of this thesis was based on certain assumptions about biological and technical parameters.

Since the implementation of *GS* in dairy cattle breeding schemes will reduce costs especially by skipping progeny testing, whereas cost for semen storage, marketing etc. will roughly be unaffected, only such costs being changed by this new technology have to be regarded when deciding about its economical usefulness. The underlying cost in part two and three of this thesis where provided by the “Rinderzuchtverband Mecklenburg-Vorpommern”, and, therefore, reflect both a possible strategy to implement *GS* and the decision relevant costs concerned by this strategy. Thus, the outcomes of the optimisations must be interpreted regarding this assumptions. This is especially exemplified by the compensations payments to breeders keeping selection candidates because if *GEBVs* are available for low costs this payments make up the majority of genomic selection costs. This costs structure is an example for what might be expected in a cooperative breeding program.

Fixed biological parameters affect optimisations results as linear scaling factors not changing the ranking of breeding schemes in terms of the optimisation goal (e.g. success

of first insemination) as long as they are not interacting with variable input parameters or optimisations results. Thus, not regarding these constant parameters has no effect on the outcomes. However, a possible interaction of the breeding scheme optimisations and input parameters might be the reduced semen production of young bulls in breeding schemes without progeny testing leading to a higher number of necessary cow and bull sires (Winkelman & Spelman, 2010). As shown in part two and three, the selection intensity for cow sires is in most cases on such a level that a reduction of the number of selection candidates or the increase in selected individuals has only a very small effect on the genetic gains in the paths “sire-dam” and “sire-sire”. Thus, a fine-tuning of the effects of GS on the number of necessary cow sires would not have changed the ranking of breeding schemes in terms of genetic gain per year.

Summary

The goal of breeding activities in commercial livestock populations is the increase of the mean of the genetically based performance capacity concerning one or numerous traits being summarised the in aggregate genotype via weighting factors if a certain breeding scheme is applied. Given the breeding scheme, the extent of this increase depends on the accuracy of breeding value estimation which is a function of a) the amount of information available about the selection candidates and its correlation structure to the aggregate genotype, and b) the usefulness of the statistical model in order to regress the genotype of the selection candidate on these available information. Recent molecular-genetical findings concern both, the amount of available information as well as the statistical model. The latter is affected by a “genomic imprinting” called mechanism, leading to an alteration of a genes effect on the phenotype of offspring due to a sex specific DNA methylation during gametogenesis in parents. Genomic imprinting can be regarded in breeding value estimation due to the calculation of two breeding values for each individual, one if it acts as sire and the other if it acts as a dam. Weighting factors to summarise this breeding values can be derived by an extension of the gene flow method. This extension is developed in the first part of this thesis and allows for tracing the flow of genes of a certain founder or group of founders within a population across tiers (e.g. nucleus, multiplier, production) and generations with special regard to the sex of the direct parent of an individual carrying these genes. Thus, it allows to assess the

probability that a gene of a founder is inherited to its descendants via their direct sires or dams. The discounted and summarised trait realisations out of the genes inherited by the sire and the dam can be used as weighting coefficients for summarising the breeding values of an individual as a sire and a dam. The extended gene flow method is applied to a hypothetical pig breeding program showing that the weights for the breeding values as a dam and as a sire can differ according to the chosen breeding scheme and the planning horizon. Furthermore, it is shown that depending the breeding scheme the breeding value of a dam when acting as a sire might be weighted higher than when acting as a dam. Additionally, a possibility to predict the increase in inbreeding due to one round of selection inherent in the method is presented.

The above mentioned amount of available information about a selection candidate is affected by the discovery of hundreds of thousands DNA markers in form of single nucleotide polymorphisms (SNP marker) being in strong linkage disequilibrium with neighbouring trait affecting genes or quantitative trait loci (QTL). The sum over all estimated marker effects on the phenotype, the genomically estimated breeding value (GEBV), allow for the explanation of a certain proportion of the additive genetic variance dependent on the trait, and can be used as an additional information about the selection candidate for estimating breeding values.

The application of genomic selection (GS) as the selection on the basis of GEBVs may lead to multistage selection schemes especially in dairy cattle, using GS as a preselection stage in order to reduce the number of test bulls in breeding schemes using progeny testing, or to replace this information source. A major problem of multistage selection is to choose the combination of stages and selection intensities maximising the genetic gain. Approaches of optimisation research may be applied, but since the selection indices of successive stages are correlated, multidimensional integration for deriving the selection intensities at selection stages is necessary, which might be unstable and time consuming

according to the correlation structure and number of stages. The second part of this thesis compares the optimisation results of multistage breeding schemes regarding genomic selection, where two different approaches for deriving the selection intensity and the genetic gain are used and the accuracy and cost of GEBVs are varied. The first approach derives the stage dependent breeding values such that the correlation between stages is zero allowing for the calculation of the stage selection intensity via one dimensional integration and, therefore, a fast optimisation of breeding schemes containing even an unlimited number of selection stages. A disadvantage of this approach is a loss in variance of stage breeding values and the genetic gain. The second approach uses new developments for the integration of multivariate normal distributions and calculates an exact solution for the selection intensity and the genetic gain after a certain number of selection stages. The results clearly show that the integration algorithm is fast and stable enough to compare even a large number of possible breeding schemes. Furthermore, the loss in breeding value variance is unpredictable when using the decorrelated selection indices, and a proper consideration of the interaction between selection paths due to cost limitation and paths specific selection strategies will lead to illogical suggestions concerning the breeding scheme structure. As the accuracies and costs of GEBVs were varied in a certain range, the results also show that GS is competitive to conventional progeny testing in dairy cattle breeding even if the accuracy of GEBVs is decreased to 0.45.

GS will increase the breeding costs linear due to the number of genotyped individuals. Thus, genotyping large proportions of a population might lead to uneconomical breeding schemes. This is especially the case for bull dam selection in dairy cattle breeding because the cow population size is equal to the number of potential selection candidates. Additionally, the number of selected bull dams is dictated by the demand for potential sires. Therefore, decreasing the number of genotyped selection candidates in order to

fulfil economical limitations might lead to a very small selection intensity making the financial efforts difficult to justify concerning the genetic gain. A possible way out is the usage of inexpensive SNP chips containing only a minor number of SNPs for genotyping huge proportions of the selection candidates population and estimate less accurate GEBVs on this basis by using imputation algorithms. The third part of this thesis investigates multistage dairy cattle breeding schemes regarding the possibility of using a low density and high density SNP chip in each selection path. The costs of each chip and the accuracy of the subsequently estimated GEBVs were varied within a certain parameter space, where it was assured that the costs of the low density SNP chip and the subsequent accuracy of the GEBVs were always lower than those for the high density SNP chip. The results underline the potential of low density SNP chips for selecting bull dams from large cow populations, but also draw the attention to the non-linearity of the genetic gain as a function of the selection intensity. Thus, there exist combinations of cost and accuracies where it was found to be economical to limit the number of low density genotyped bull dams and include a further selection stage using high density SNP chips in that path. Furthermore, the results also show that the genetic gain is much more influenced by the cost and accuracy of the GEBV out of a high density chip, but the breeding scheme structure reacts more sensible to a change of this parameter concerning the low density chip.

Zusammenfassung

Züchtungsplanung beabsichtigt unter anderem die Vorhersage des genetisch bedingten Leistungszuwachses einer Population bezüglich eines oder mehrerer im aggregierten Genotyp zusammengefasster Merkmale bei Anwendung einer bestimmten Zuchtprogrammes. Der Umfang dieses Leistungszuwachses hängt bei gegebenem Zuchtprogramm im wesentlichen von der Genauigkeit der Zuchtwertschätzung ab, welche wiederum als Funktion der Informationsmenge über einen Selektionskandidaten, als auch der Eignung des statistische Modells beschrieben werden kann. Neuerer molekulargenetische Erkenntnisse betreffen sowohl das statistische Modell als auch die Informationsmenge. Mögliche Änderungen am statistischen Modell ergeben sich aus der Entdeckung der genomischen Prägung, einer auf DNA-Methylierung während der Gametogenese beruhenden Abschwächung der Genexpression im Nachkommen in Abhängigkeit von Geschlecht des vererbenden Elters. Genomisches Prägung kann durch Modellierung jeweils eines Zuchtwertes als Vater und als Mutter für jedes Individuum in der Zuchtwertschätzung berücksichtigt werden. Gewichte zur Zusammenfassung dieser Zuchtwerte in einem Gesamtzuchtwert können durch eine Erweiterung der Genflussmethode abgeleitet werden. Diese Erweiterung wird im ersten Teil dieser Dissertation entwickelt. Sie erlaubt die Verfolgung der Gene eines einzelnen oder einer Gruppe von Gründertieren über unterschiedliche Tiergruppen (z.B, männliche und weibliche Nucleustiere, Vermehrer, Schlachttiere) und Generationen hinweg, wobei zusätzlich zur Wahrscheinlichkeit, dass eine bestimmte Tier-

gruppe zu einer bestimmten Zeit Gene der Gründertiere erhalten hat, auch eine Aussage darüber getroffen werden kann, mit welcher Wahrscheinlichkeit diese Gene vom unmittelbaren Vater oder der unmittelbaren Mutter auf die Tiere der Subpopulation übertragen wurden. Die diskontierten und summierten Merkmalsrealisierungen aus maternal bzw. paternal vererbten Genen der Gründertiere können sodann als Gewichte zur Zusammenfassung der Zuchtwerte als Vater und als Mutter verwendet werden. Die Anwendung der erweiterten Genflussmethode auf ein hypothetisches Schweinezuchtprogramm zeigt, dass die Gewichte für die beiden o.g. Zuchtwerte in Abhängigkeit vom Zuchtplan und dem Planungshorizont differieren können, wobei selbst bei weiblichen Tieren der Zuchtwert als Vater unter Umständen höher zu gewichten ist als jener als Mutter. Weiterhin kann auf Grund der Eigenschaften der erweiterten Genflussmethode der Inzuchtanstieg durch eine Selektionsrunde abgeschätzt werden.

Die Änderung der Informationsmenge für einen Selektionskandidaten ergibt sich aus der Entdeckung von hunderttausenden DNA-Markern in Form von singulären Nukleotidpolymorphismen (SNP-Marker), welche sich jeweils im starken Kopplungsungleichgewicht mit benachbarten merkmalsbeeinflussenden Genen oder DNA-Abschnitten (QTL) befinden, und je nach Merkmal einen gewissen Anteil der additive-genetischen Varianz erklären. Genomische Zuchtwerte (GZW) als die Summe über alle SNP-Marker/-effekte können genutzt werden, um Individuen zu selektieren sobald deren Markergenotyp bekannt ist.

Die Anwendung genomischer Selektion (GS) als Selektion basierend auf GZWs könnte insbesondere in der Milchrindzucht zu Zuchtprogrammen mit mehr als zwei Selektionsstufen führen, wenn GZWs als Vorselektion für eine folgende Nachkommenschaftsprüfung verwendet werden. Die richtige Kombination von Selektionsstufen und Selektionsintensitäten zur Maximierung des Zuchtfortschrittes kann mit Hilfe von Maximierungsalgorithmen gefunden werden, wird jedoch durch die Notwendigkeit multipler numerischer

Integration zur exakten Berechnung der Selektionsintensität erschwert, da die Selektionsindices aufeinanderfolgender Stufen korreliert sind und die Stabilität der Lösung durch Integration sowie deren Rechenzeitbedarf von der Korrelationsstruktur und der Stufenzahl abhängig ist. Der zweite Teil dieser Arbeit beschäftigt sich daher mit einem Methodenvergleich zur Berechnung der Selektionsintensität und des Zuchtfortschrittes bei der Optimierung von mehrstufigen Zuchtprogrammen, welche genomische Selektion berücksichtigen und in denen die Kosten und Genauigkeiten genomischer Zuchtwerte variiert werden. Die erste Methode leitet die Stufenzuchtwerte derart ab, dass deren Korrelation zum Zuchtziel maximal ist, die Korrelation zwischen den Stufen jedoch null. Dies erlaubt die Berechnung der Selektionsintensität jeder Stufe mittels eindimensionaler Integration, und daher eine schnelle Optimierung von Zuchtprogrammen mit einer theoretisch unbegrenzten Stufenzahl. Der Nachteil dieser Methode ist eine verminderte Varianz der Stufenzuchtwerte ab der zweiten Stufen und somit ein verminderter Zuchtfortschritt. Die zweite Methode benutzt neue Entwicklungen zur numerischen Integration multivariater Normalverteilungen zur exakten Berechnung der Selektionsintensitäten der einzelnen Stufen und des Zuchtfortschrittes. Die Ergebnisse des Methodenvergleiches zeigen deutlich das der Integrationsalgorithmus schnell und stabil genug rechnet um selbst eine große Anzahl von Zuchtprogrammen zu vergleichen. Dagegen führt die Anwendung unkorrelierter Selektionsindices zu einem nicht vorhersehbaren Verlust von vorhergesagtem Zuchtfortschritt, und eine exakte Berücksichtigung der Interaktion zwischen den Selektionspfaden durch Kostenbudgetierung und pfadspezifischen Selektionsstrategien ergibt Optimierungsergebnisse die zwar die Logik des Algorithmus widerspiegeln, deren Annäherung an das wahre Optimum jedoch unbekannt ist. Da die Kosten und Genauigkeiten der GZW variiert wurden zeigen die Ergebnisse weiterhin, dass GS bezüglich des Zuchtfortschrittes pro Jahr mit konventionellen Milchviehzuchtprogrammen konkurrieren kann, selbst wenn die GZW-Genauigkeit auf 0.45 sinkt.

Durch GS steigen die Züchtungskosten linear mit der Anzahl genotypisierter Tiere an. Die Genotypisierung großer Teile einer Population kann somit zu unverhältnismäßig hohen Kosten und unökonomischen Zuchtprogrammen führen. Dies trifft insbesondere auf die Anwendung von GS zur Selektion von Bullenmüttern in Milchviehzuchtprogrammen zu, da die Anzahl potentieller Selektionskandidaten equivalent zur Größe der Kuhpopulation ist. Weiterhin problematisch ist der durch den Bedarf an männlichen Selektionskandidaten vorgegebenen Bedarf an Bullenmüttern, und die daraus resultierenden geringe Selektionsintensität der GS wenn die Zahl genotypisierter Bullenmütter vorhandenen Kostenlimitierungen angepasst wird. Der Umfang des Zuchtfortschrittes könnte dann in keinem Verhältnis zu den Kosten stehen. Ein möglicher Ausweg ist die Verwendung preisgünstiger SNP-Chips, welche eine deutlich verringerte Anzahl von Markern beinhalten, um damit große Teile der Kuhpopulation zu genotypisieren. Die damit zu schätzenden Zuchtwerte besitzen eine geringere Genauigkeit, die sich jedoch durch Anwendung von Zuweisungsalgorithmen (*engl.* imputing) verbessern lässt. Der dritte Teil dieser Arbeit untersucht mehrstufige Milchviehzuchtprogramme, welche die Möglichkeit berücksichtigen, unterschiedliche SNP Chips in jedem Selektionspfad zu verwenden bzw. diese zu kombinieren. Die Kosten und Genauigkeiten der GZWs die auf Basis dieser SNP Chips berechnet werden können, wurden semikontinuierlich variiert, wobei die Kosten bzw. Genauigkeiten der GZWs auf Basis des preisgünstigen Chips immer niedriger waren als jene auf Basis des kostenintensiven Chips. Die Ergebnisse dieser Untersuchung zeigen deutlich das Potenzial der genomischen Selektion auf Basis kostengünstiger SNP Chips für die Selektion von Bullenmüttern aus großen Kuhpopulationen, unterstreichen jedoch auch den nichtlinearen und asymptotischen Zusammenhang zwischen Selektionsintensität und Zuchtfortschritt. Es wurden daher auch Kombinationen aus Kosten und Genauigkeiten genomischer Zuchtwerte gefunden, bei denen auf eine weitere Ausdehnung der Genotypisierung von potentiellen Bullenmüttern mit einem kostengünstigen

SNP Chip zugunsten einer nachfolgenden Selektionsstufe und der zusätzlichen Anwendung des kostenintensiven SNP Chips verzichtet wurde um den Zuchtfortschritt zu maximieren. Weiterhin wurde der Zuchtfortschritt durch den Preis des kostenintensiven Chips und die mit ihm erzielte Genauigkeit der GZW wesentlich stärker beeinflusst als durch Änderung dieser Parameter für den preisgünstigen Chip. Im Gegensatz dazu war der Änderungsdruck auf die Struktur des Zuchtprogrammes durch Veränderungen der Kosten und erzielten Genauigkeiten der GZWs auf Basis des kleinen Chips am größten.

Bibliography

- Balcerzak, K. M., Freeman, A. E. & Willham, R. L. (1989), Selection for the direct and maternal genetic effects for dystocia in holsteins. *J Dairy Sci* **72**(5), 1273–1279.
- Brent, R. P. (1972), *Algorithms for Minimisation without Derivatives*. Prentice Hall.
- Buch, L. H., Sørensen, M. K., Lassen, J., Berg, P. & Sørensen, A. C. (2010), Dairy cattle breeding schemes with or without genomic selection and progeny testing. In *Proceedings of the nineth World Congress on Genetics Applied to Livestock Production, Leipzig, Germany, August 1-6, 2010*.
- Börner, V. & Reinsch, N. (2010), Decorrelated selection indices versus optimum selection indices in optimising multistage dairy cattle breeding schemes regarding genomic selection. *personal communications* .
- Calus, M. P. L. & Veerkamp, R. F. (2007), Accuracy of breeding values when using and ignoring the polygenic effect in genomic breeding value estimation with a marker density of one snp per cm. *J Anim Breed Genet* **124**(6), 362–368, URL <http://dx.doi.org/10.1111/j.1439-0388.2007.00691.x>.
- Cockett, N. E., Jackson, S. P., Snowden, G. D., Shay, T. L., Berghmans, S., Beever, J. E., Carpenter, C. & Georges, M. (1999), The callipyge phenomenon: evidence for unusual genetic inheritance. *J Anim Sci* **77 Suppl 2**, 221–227.

- Daetwyler, H. D., Villanueva, B., Bijma, P. & Woolliams, J. A. (2007), Inbreeding in genome-wide selection. *J Anim Breed Genet* **124**(6), 369–376, URL <http://dx.doi.org/10.1111/j.1439-0388.2007.00693.x>.
- de Koning, D., Harlizius, B., Rattink, A., Groenen, M., Brascamp, E. & van Arendonk, J. (2001a), Detection and characterization of quantitative trait loci for meat quality traits in pigs. *J Anim Sci* **79**(11), 2812–2819, URL [PM:11768109](https://pubmed.ncbi.nlm.nih.gov/11768109/).
- de Koning, D., Rattink, A., Harlizius, B., Groenen, M., Brascamp, E. & van Arendonk, J. (2001b), Detection and characterization of quantitative trait loci for growth and reproduction traits in pigs. *Livest Prod Sci* **72**(3), 185–198, URL [ISI:000172280000001](https://pubmed.ncbi.nlm.nih.gov/000172280000001/).
- de Koning, D., Rattink, A., Harlizius, B., van Arendonk, J., Brascamp, E. & Groenen, M. (2000), Genome-wide scan for body composition in pigs reveals important role of imprinting. *Proc Natl Acad Sci U S A* **97**(14), 7947–7950, URL [PM:10859367](https://pubmed.ncbi.nlm.nih.gov/10859367/).
- de Vries, A. G., Kerr, R., Tier, B., Long, T. & Meuwissen, T. (1994), Gametic imprinting effects on rate and composition of pig growth. *Theor Appl Genet* **88**, 1037–1042.
- Dekkers, J. C. M. (2007), Prediction of response to marker-assisted and genomic selection using selection index theory. *J Anim Breed Genet* **124**(6), 331–341, URL <http://dx.doi.org/10.1111/j.1439-0388.2007.00701.x>.
- Ducrocq, V. & Colleau, J. (1989), Optimum truncation points for independent culling level selection on a multivariate normal distribution, with an application to dairy cattle selection. *Genet Sel Evol* **21**, 185–191.
- Engelland, T. & Tier, B. (2002), Genetic variances due to imprinted genes in cattle. *J Anim Breed Genet* **119**, 154–165, URL <http://dx.doi.org/10.1046/j.1439-0388.2002.00323.x>.

-
- Freking, B. A., Murphy, S. K., Wylie, A. A., Rhodes, S. J., Keele, J. W., Leymaster, K. A., Jirtle, R. L. & Smith, T. P. L. (2002), Identification of the single base change causing the callipyge muscle hypertrophy phenotype, the only known example of polar overdominance in mammals. *Genome Res* **12**(10), 1496–1506, URL <http://dx.doi.org/10.1101/gr.571002>.
- Genz, A. (1992), Numerical computation of multivariate normal probabilities. *J Comput Graph Stat* **1**, 141 – 149.
- Groen, A. F. (1989), Economic values in cattle breeding. i. influences of production circumstances in situations without output limitations. *Livest Prod Sci* **22**, 1–16.
- Habier, D., Fernando, R. L. & Dekkers, J. C. M. (2009), Genomic selection using low-density marker panels. *Genetics* **182**(1), 343–353, URL <http://dx.doi.org/10.1534/genetics.108.100289>.
- Haig, D. (2004), Genomic imprinting and kinship: how good is the evidence? *Annu Rev Genet* **38**, 553–585, URL [PM:15568986](https://pubmed.ncbi.nlm.nih.gov/15568986/).
- Hayes, B. J., Bowman, P. J., Chamberlain, A. C., Verbyla, K. & Goddard, M. E. (2009a), Accuracy of genomic breeding values in multi-breed dairy cattle populations. *Genet Sel Evol* **41**, 51, URL <http://dx.doi.org/10.1186/1297-9686-41-51>.
- Hayes, B. J., Bowman, P. J., Chamberlain, A. J. & Goddard, M. E. (2009b), Invited review: Genomic selection in dairy cattle: progress and challenges. *J Dairy Sci* **92**(2), 433–443, URL <http://dx.doi.org/10.3168/jds.2008-1646>.
- Hicks, C., Muir, W. & Stick, D. (1998), Selection index updating for maximizing rate of annual genetic gain in laying hens. *Poult Sci* **77**(1), 1–7, URL [PM:9469744](https://pubmed.ncbi.nlm.nih.gov/9469744/).
- Hill, W. G. (1974), Prediction and evaluation of response to selection with overlapping generations. *Anim Prod* **18**, 117–139.

- Howie, B., Donnelly, P. & Marchini, J. (2009), A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**(6), e1000529, URL <http://dx.doi.org/10.1371/journal.pgen.1000529>.
- Johnson, D. L. (1977), Inbreeding in populations with overlapping generations. *Genetics* **87**(3), 581–591.
- König, S., Simianer, H. & Willam, A. (2009), Economic evaluation of genomic breeding programs. *J Dairy Sci* **92**(1), 382–391, URL <http://dx.doi.org/10.3168/jds.2008-1310>.
- König, S. & Swalve, H. H. (2009), Application of selection index calculations to determine selection strategies in genomic breeding programs. *J Dairy Sci* **92**(10), 5292–5303, URL <http://dx.doi.org/10.3168/jds.2009-2232>.
- Lillehammer, M., Meuwissen, T. & Sonesson, A. (2010), Effects of alternative genomic selection breeding schemes on genetic gain in dairy cattle. In *Proceedings of the Ninth World Congress on Genetics Applied to Livestock Production*.
- Lynch, M. & Walsh, B. (1998), *Genetics and Analysis of Quantitative Traits*. Sinauer Associates Inc., U.S.
- McClintock, A. & Cunningham, E. (1974), Selection in dual-purpose cattle populations: Defining the breeding objective. *Anim Prod* **18**, 237–247.
- Meuwissen, T. H., Hayes, B. J. & Goddard, M. E. (2001), Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**(4), 1819–1829.
- Mi, X. & Utz, F. (2008), *R-Package selection gain*. University of Hohenheim, mi@pz.uni-hohenheim.de, URL www.r-project.org/.

-
- Muir, W. M. (2007), Comparison of genomic and traditional blup-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. *J Anim Breed Genet* **124**(6), 342–355, URL <http://dx.doi.org/10.1111/j.1439-0388.2007.00700.x>.
- NamKoong, G. (1970), Optimum allocation of selection intensity in two stages of truncation selection. *Biometrics* **26**, 465 – 476.
- Neugebauer, N., Luther, H. & Reinsch, N. (2010a), Parent-of-origin effects cause genetic variation in pig performance traits. *Animal* **4**, 672–681.
- Neugebauer, N., Räder, I., Schild, H. J., Zimmer, D. & Reinsch, N. (2010b), Evidence for parent-of-origin effects on genetic variability of beef traits. *J Anim Sci* **88**(2), 523–532, URL <http://dx.doi.org/10.2527/jas.2009-2026>.
- Reik, W., Collick, A., Norris, M. L., Barton, S. C. & Surani, M. A. (1987), Genomic imprinting determines methylation of parental alleles in transgenic mice. *Nature* **328**(6127), 248–251, URL <http://dx.doi.org/10.1038/328248a0>.
- Reinsch, E., N. und Kalm (1995), Genfluß und relative Bedeutung maternaler, paternaler und direkter Effekte beim Milchrind. *Arch Anim Breed* **38**(4), 355–366.
- Reinsch, N. & Guiard, V. (2005), Mixed mendelian and imprinted inheritance, unpublished.
- Rendel, J. M. & Robertson, A. (1950), Estimation of genetic gain in milk yield by selection in a closed herd of dairy cattle. *J Genet* **50**, 1–8.
- Sapienza, C., Peterson, A. C., Rossant, J. & Balling, R. (1987), Degree of methylation of transgenes is dependent on gamete of origin. *Nature* **328**(6127), 251–254, URL <http://dx.doi.org/10.1038/328251a0>.

- Saxton, A. (1989), Indcull version 3.0: Independent culling for two or more traits. *J Hered* **80** (2), 166–167.
- Schaeffer, L. R. (2006), Strategy for applying genome-wide selection in dairy cattle. *J Anim Breed Genet* **123**(4), 218–223, URL <http://dx.doi.org/10.1111/j.1439-0388.2006.00595.x>.
- Spelman, R., Arias, J., Keehan, M., Obolonkin, V., Winkelman, A., Johnson, D. & Harris, B. (2010), Application of genomic selection in the new zealand dairy cattle industry. In *Proceedings of the ninth World Congress on Genetics Applied to Livestock Production, Leipzig, Germany, August 1-6, 2010*.
- Stella, A., Stalder, K., Saxton, A. & Boettcher, P. (2003), Estimation of variances for gametic effects on litter size in yorkshire and landrace swine. *J Anim Sci* **81**(9), 2171–2178, URL [PM:12968691](https://pubmed.ncbi.nlm.nih.gov/12968691/).
- Tallis, G. M. (1961), The moment generating function of the truncated multi-normal distribution. *J R Statist Soc B* **23**, 223 – 229.
- Tuiskula-Haavisto, M., de Koning, D.-J., Honkatukia, M., Schulman, N. F., Mäki-Tanila, A. & Vilkki, J. (2004), Quantitative trait loci with parent-of-origin effects in chicken. *Genet Res* **84**(1), 57–66.
- Van Laere, A.-S., Nguyen, M., Braunschweig, M., Nezer, C., Collette, C., Moreau, L., Archibald, A. L., Haley, C. S., Buys, N., Tally, M., Andersson, G., Georges, M. & Andersson, L. (2003), A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature* **425**(6960), 832–836, URL <http://dx.doi.org/10.1038/nature02064>.
- VanRaden, P. (2010), Genomic evaluations with many more genotypes and phenotypes.

-
- In *Proceedings of the Ninth World Congress on Genetics Applied to Livestock Production*, hallo, p. 149.
- Villanueva, B., Wray, N. & Thompson, R. (1993), Prediction of asymptotic rates of response from selection on multiple traits using univariate and multivariate best linear unbiased predictors. *Anim Prod* **57**, 1–13.
- Weigel, K., de los Campos, G., Vazquez, A., Rosa, C. V. T. G., Gianola, D., O’Connell, J., VanRaden, P. & Wiggans, G. (2010), Genomic selection and its effects on dairy cattle breeding programs. In *Proc. 9th WGCALP*.
- Whitehead, J. (2004), *Genomic imprinting in development and evolution*. Ph.D. thesis, Uppsala University, Faculty of Science and Technology.
- Wilkins, J. F. & Haig, D. (2003), What good is genomic imprinting: the function of parent-specific gene expression. *Nat Rev Genet* **4**(5), 359–368, URL <http://dx.doi.org/10.1038/nrg1062>.
- Winkelman, A. & Spelman, R. (2010), Response using genome-wide selection in dairy cattle breeding schemes. In *Proceedings of the Ninth World Congress on Genetics Applied to Livestock Production*.
- Woolliams, J. A. & Bijma, P. (2000), Predicting rates of inbreeding in populations undergoing selection. *Genetics* **154**(4), 1851–1864.
- Wright, S. (1931), Evolution in mendelian populations. *Genetics* **16**(2), 97–159.
- Xie, C. & Xu, S. (1998), Efficiency of multistage marker-assisted selection in the improvement of multiple quantitative traits. *Heredity* **80** (Pt 4), 489–498, URL [PM:9618913](https://doi.org/10.1046/j.1365-3113.1998.00489.x).
- Xu, S., Martin, T. G. & Muir, W. M. (1995), Multistage selection for maximum economic

- return with an application to beef cattle breeding. *J Anim Sci* **73**(3), 699–710, URL [PM:7608002](#).
- Xu, S. & Muir, W. M. (1991), Multistage selection for genetic gain by orthogonal transformation. *Genetics* **129**(3), 963–974, URL [PM:1752432](#).
- Xu, S. & Muir, W. M. (1992), Selection index updating. *Theor Appl Genet* **83**(4), 451–458, URL <http://dx.doi.org/10.1007/BF00226533>.
- Ytournal, F. & Simianer, H. (2010), Inclusion of externally computed genotype frequencies in the gene-flow method. In *Book of Abstracts of the 61st Annual Meeting of the European Association for Animal Production*.

Appendix

Fortran 95 routines were used for calculating the selection intensity of the finally selected individuals after n successive selection stages in terms of standard deviations of the aggregate genotype. In the following code the respective subroutines are integrated into a main program *Test* which sets the number of selection stages, creates a correlation matrix and a vector of the proportions of selected individuals at each selection stage, calls the subroutine *get_SelectionGain_sub* and writes out the results. Thus, the source code given here should compile without errors with an Intel Ifort 11.1 compiler yielding a working program. The main program *Test* has not been used for the model calculations of part two and three, where all subroutines had been called directly from the calculation program. The subroutine *mvtdst.f* of Genz (1992) can be found at <http://www.math.wsu.edu/faculty/genz/software/software.html> and *zero.f90* of Brent (1972) can be found at http://people.sc.fsu.edu/~jburkardt/f_src/brent/brent.html.

Module Data_Kind

Implicit None

*! saving the processor dependent variable of the kind of real and integer
! numbers*

Integer, **Parameter** :: IkXL=Selected_Int_Kind(12)

Integer, **Parameter** :: IkL=Selected_Int_Kind(9)

Integer, **Parameter** :: IkM=Selected_Int_Kind(6)

Integer, **Parameter** :: IkS=Selected_Int_Kind(3)

Integer(IkS), **Parameter** :: RkDbl=Selected_Real_Kind(15,100)

Integer(IkS), **Parameter** :: RkSgl=Selected_Real_Kind(6,37)

End Module Data_Kind

Module Selection_Gain

use Data_Kind

Implicit None

private *!make everything within the module only accessible within the module*

Type, **Private** :: Int_Var

Integer(IkM) :: ISNuVar *!< number of selection stages*

!< variables for the integration process of subroutine mvtdstpack.f

Real(RkDBL), **Allocatable**, **Dimension**(:) :: RVUpper *!< Vector of upper*

! integration limits

Real(RkDBL), **Allocatable**, **Dimension**(:) :: RVLower *!< Vector of lower*

! integration limits

Real(RkDBL), **Allocatable**, **Dimension**(:) :: RVDelta *!<*

Real(RkDBL), **Allocatable**, **Dimension**(:) :: RVCorr *!< Vector of*

! correlations built from the correlation matrix

Integer(IkM), **Allocatable**, **Dimension**(:) :: IVInfin

Real(RkDBL) :: Error, Val

Real(RkDBL) :: Integral

Integer(IkM) :: Inform

Real(RkDBL) :: RELEPS

Integer(IkM) :: Maxpts

contains

Procedure, Pass :: **allocate** => allocate_Integration_Variables_sub

End type Int_Var

!declare variables to be known across all subroutines within the module

!integration constants for the subroutine mvtldstpack.f

Integer(IkM), **Parameter** :: ISDeFr=-1

Real(RkDBL), **Parameter** :: ABSEPS=0.0000001

!derived data type to values of the multivariate integration process

Type(Int_Var) :: TSInt

!< let only the subroutine get_SelectionGain_sub accessible from outside

! the module

public :: get_SelectionGain_sub

contains

Subroutine get_SelectionGain_sub(&

&ISNuSt, RVPrSe, RMCorr, & *!In*

&RVGeGa) *!Out*

!< top level subroutine, communicates with the outer and inner environment

Implicit None

!< number of stages

Integer(Iks), **Intent**(In) :: ISNuSt

!< vector of proportion selected

Real(RkDBL), **Intent**(In), **Dimension**(ISNuSt) :: RVPrSe

!< correlation matrix of selection stages including the aggregated genotype

Real(RkDBL), **Intent**(In), **Dimension**(ISNuSt+1, ISNuSt+1) :: RMCorr

!< vector genetic gain in standard deviations

Real(RkDBL), **Intent**(Out), **Allocatable**, **Dimension**(:) :: RVGeGa

Integer(Iks) :: ISStat, c1 *!< others*

!< machine constant (lowest value to distinguish between two real numbers)

Real(RkDBL), **External** :: dlmach

!< vector of truncation points of selection stages

Real(RkDBL), **Allocatable**, **Dimension**(:) :: RVTrPo

!< vector of truncation points of selection stages and aggregated genotype

Real(RkDBL), **Allocatable**, **Dimension**(:) :: RVTrPo_new

```

! calculate truncation points of selection stages
call get_TruncationPoints_sub(&
    &ISNuSt ,RVPrSe ,RMCorr(2:ISNuSt+1,2:ISNuSt+1),& !In
    &RVTrPo) !Out

Allocate(RVTrPo_new(ISNuSt+1), RVGeGa(ISNuSt), Stat=ISStat)
If (ISStat/=0) Then
    write(* ,*) "allocation_error_in_get_SelectionGain_sub"
    Stop
End If

! copy the truncation points of selection stages and set the truncation
! point of the aggregated genotype to -infinity
RVTrPo_new(1)=-1000
RVTrPo_new(2:ISNuSt+1)=RVTrPo

! calculate the genetic gain of successive stages
Do c1=2,ISNuSt+1
    If (c1==2) Then
        ! one dimensional integration
        RVGeGa(c1-1)=dnorm(RVTrPo(c1-1))/RVPrSe(c1-1)
    Else
        call calc_Selection_Gain_sub(&
            &Int(c1 ,Iks) ,RMCorr(1:c1 ,1:c1) ,RVTrPo_new(1:c1)&
            &,product(RVPrSe(1:c1-1))& !In
            &,RVGeGa(c1-1)) !Out
    End If
End Do

Deallocate(RVTrPo_new,RVTrPo)
End Subroutine get_SelectionGain_sub

Subroutine get_TruncationPoints_sub(&
    &ISNuVar ,RVProb ,RMCorr,& !In
    &RVLoInBo) !Out

! < calculates successive the truncation points of stages due to
! iterative integration

```

```

use zerofinder

Implicit None
!number of variables of a multivariate normal distribution
Integer(Iks), Intent(In) :: ISNuVar
!vector of probabilities
Real(RkDBL), Intent(In), Dimension(ISNuVar) :: RVProb
! correlation matrix of the variables
Real(RkDBL), Intent(In), Dimension(ISNuVar,ISNuVar) :: RMCorr
! vector of lower integration bounds
Real(RkDBL), Intent(Out), Allocatable, Dimension(:) :: RVLoInBo
Integer(Iks) :: ISStat, c1 !others
!parameters for the zerofinder
Real(RkDBL), Parameter :: ax=100. !upper iteration value
Real(RkDBL), Parameter :: bx=-100. !lower iteration value
Real(RkDBL), Parameter :: tol=0.000001 !convergence criteria
Real(RkDBL) :: dlmach ! machine constant
Real(RkDBL) :: machep ! saves the machine constant
If(Allocated(RVLoInBo)) Then
    Deallocate(RVLoInBo)
End If
Allocate(RVLoInBo(ISNuVar), Stat=ISStat)
RVLoInBo=0.
call TSInt%allocate(ISNuVar)
!transform the correlation matrix into a correlation vector (since
! mvtdstpack.f can work only with correlation vectors)
call make_Correlation_Vector_sub(ISNuVar,RMCorr,TSInt%RVCCorr)
!get the machine constant for the zerofinder
machep=dlmach(4)
!set the vector of non-centrality parameter to zero
TSInt%RVDelta=0.
!set the vector IVInfin to one. This indicates to
! mvtdstpack.f that the upper integration bound is +infinity and the
! lower is found in RVLower
TSInt%IVInfin=1

```

```

TSInt%RVUpper=0.
TSInt%RVLower=0.
!calculate the lower integration bounds
Do c1=1,ISNuVar
    TSInt%ISNuVar=c1

    !calculate the integration area
    TSInt%Integral=Product(RVProb(1:TSInt%ISNuVar))

    !start the zerofinder of Richard Brent and get via iterative
    ! integration given the area and the upper bound the lower bound
    TSInt%RVlower(TSInt%ISNuVar)=zero(&
        &ax,& !< lower limit of the parameter space
        &bx,& !< upper limit of the parameter space
        &machep,& !< machine constant
        &tol,& !< tolerance
        &minimise& !< function to be minimised
        &)

    RVLoInBo(TSInt%ISNuVar)=TSInt%RVlower(TSInt%ISNuVar)
End Do
End Subroutine get_TruncationPoints_sub

Function minimise(a)
    !< function to be minimised due to the zero finder
    use Interface_Definitions
    use Data.Kind
    Implicit None
    Real(RkDBL), Intent(In) :: a !< function argument to be set by the zero
    ! finder
    Real(RkDBL) :: minimise !< function to be minimised
    Integer(Iks) :: c_1
    !< set the lower integration bound of the stage to be integrated
    TSInt%RVlower(TSInt%ISNuVar)=a
    TSInt%Inform=0

```

```

TSInt%RELEPS=0.00001
! calculate the integral
c_1=0
Do
  TSInt%Maxpts=Int(10000*(TSInt%ISNuVar+c_1),IkM)
  ! call the subroutine mvtdst of Alan Genz
  CALL MVIDST(TSInt%ISNuVar,ISDeFr, TSInt%RVLower(1:TSInt%ISNuVar),&
    & TSInt%RVUpper(1:TSInt%ISNuVar),TSInt%IVInfin(1:TSInt%ISNuVar)&
    & ,TSInt%RVCorr(1:TSInt%ISNuVar*(TSInt%ISNuVar-1)/2),TSInt&
    & %RVDelta(1:TSInt%ISNuVar),TSInt%MAXPTS,ABSEPS,TSInt%RELEPS,TSInt&
    & %ERror, TSInt%VAL,TSInt%Inform )
  ! calculate the difference between the calculated integral and the
  ! expected integral
  If(TSInt%Inform/=0) Then
    write(*,*) TSInt%Inform
    write(*,*) c_1
    write(*,*) "integration_error_when_searching_for_a_truncation_point"
    c_1=c_1+1
  Else
    write(*,*) TSInt%Maxpts
    exit
  End If
End Do
  minimise=(TSInt%val-TSInt%Integral)
End Function minimise

Subroutine calc_Selection_Gain_sub(&
  &ISDim,RMCorr,RVTrPo,RSPrSe,& !In
  &RSSeGa) !Out
  ! calculate the selection gain of each selection stage
  Implicit None
  ! number of stages + 1
  Integer(Iks), Intent(In) :: ISDim
  ! correlation matrix of the aggregated genotype and the selection stages
  Real(RkDBL), Intent(In), Dimension(ISDim,ISDim) :: RMCorr

```

```

Real(RkDBL) , Intent(In) , Dimension(ISDim) :: RVTrPo !< truncation points
Real(RkDBL) , Intent(In) :: RSPrSe !< proportion selected
Real(RkDBL) , Intent(Out) :: RSSeGa !< selection gain
Integer(Iks) :: ISc1 , ISc2 , ISc3 !< others
Real(RkDBL) , Dimension(ISDim,ISDim) :: RMCorr_new !< second correlation
! matrix
Real(RkDBL) , Dimension(ISDim,ISDim,ISDim) :: RMCorr_part !< 3 dimensional
! matrix of partial correlation coefficients

!< set up a matrix of conditional truncation points due to correcting the
! old truncation point from the effect of on variable
RMCorr_new=0.
Forall( ISc1=1:ISDim , ISc2=1:ISDim , ISc1/=ISc2)
  RMCorr_new( ISc1 , ISc2)=(RVTrPo( ISc2)–RMCorr( ISc1 , ISc2)*RVTrPo( ISc1 ))&
  &/sqrt(1–RMCorr( ISc1 , ISc2)**2)
End Forall

!set up matrix of partial correlation coefficients (conditional
! correlations) between all variables being corrected for on variable
RMCorr_part=1.
Forall( ISc1=1:ISDim , ISc2=1:ISDim , ISc3=1:ISDim , (( ISc1&
  &/=ISc3) .And.( ISc2/=ISc3) .And.( ISc2/=ISc1)))

  RMCorr_part( ISc1 , ISc2 , ISc3)=&
    &(&
      &RMCorr( ISc1 , ISc2)–RMCorr( ISc1 , ISc3)*RMCorr( ISc2&
        &, ISc3)&
      &)&
    &/&
    &(&
      &sqrt(1–RMCorr( ISc1 , ISc3)**2)&
      &*&
      &sqrt(1–RMCorr( ISc2 , ISc3)**2)&
      &)&
End Forall

```



```

! calculate the mean of the truncated multivariate normal distribution
RSSeGa=0.
Do ISc1=1,ISDim
  RSSeGa=RSSeGa+RMCorr(1,ISc1)*dnorm(RVTrPo(ISc1))*get_Integral(ISDim,ISc1&
    &,RMCorr_new,RMCorr_part)
End Do
RSSeGa=RSSeGa/RSPrSe
End Subroutine calc_Selection_Gain_sub

Function get_Integral(ISDim,ISPos,RMCorr_new,RMCorr_part)
! calculate the integral of the distribution of the aggregated genotype
Implicit None
Integer(Iks), Intent(In) :: ISDim
Integer(Iks), Intent(In) :: ISPos
Real(RkDBL), Intent(In), Dimension(ISDim,ISDim) :: RMCorr_new
Real(RkDBL), Intent(In), Dimension(ISDim,ISDim,ISDim) :: RMCorr_part
Real(RkDBL), Dimension(ISDim-1,ISDim-1) :: RMCorr
Real(RkDBL) :: get_Integral
Integer(Iks) :: c-1
call TSInt%allocate(Int(ISDim-1,Iks))
TSInt%RVUpper=0
!get the vector of lower integration bounds
TSInt%RVLower(1:ISPos-1)=RMCorr_new(ISPos,1:ISPos-1)
TSInt%RVLower(ISPos:ISDim-1)=RMCorr_new(ISPos,ISPos+1:ISDim)
!get the partial correlation matrix
RMCorr(1:ISPos-1,1:ISPos-1)=RMCorr_part(1:ISPos-1,1:ISPos-1,ISPos)
RMCorr(1:ISPos-1,ISPos:ISDim-1)=RMCorr_part(1:ISPos-1,ISPos+1:ISDim,ISPos)
RMCorr(ISPos:ISDim-1,1:ISPos-1)=RMCorr_part(ISPos+1:ISDim,1:ISPos-1,ISPos)
RMCorr(ISPos:ISDim-1,ISPos:ISDim-1)=RMCorr_part(ISPos+1:ISDim,ISPos&
  &+1:ISDim,ISPos)
!transform the correlation matrix into a correlation vector
call make_Correlation_Vector_sub(Int(ISDim-1,Iks),RMCorr,TSInt%RVCorr)
!set the integration parameter
TSInt%IVInfin=1

```

```

TSInt%IVInfin(1)=-1
TSInt%RVDelta=0
TSInt%val=0.
TSInt%Inform=0
TSInt%RELEPS=0.0000001
c_1=0
Do
  TSInt%Maxpts=Int(10000*((ISDim-1)+c_1),lkm)
  !get the integral, call the subroutine mvtdst of Alan Genz
  CALL MVT DST(ISDim-1,ISDeFr, TSInt%RVLOWer, TSInt%RVUPper, TSInt&
    &%dVINFIN, TSInt%RVCorr, TSInt%RVDelta,TSInt%MAXPTS,ABSEPS,&
    &TSInt%RELEPS, TSInt%ERRor, TSInt%VAL, TSInt%Inform)
  If (TSInt%Inform/=0) Then
    write (*,*) TSInt%Inform
    write (*,*) c_1
    write (*,*) "integration_error_when_searching_for_the_mean"
    c_1=c_1+1
  Else
    write (*,*) c_1, TSInt%Maxpts
  Exit
End If
End Do
  get_Integral=TSInt%Val
End Function get_Integral

Function dnorm(RSIn)
  !< calculates the density of a standard normal distribution
  Implicit None
  Real(RkDBL), Intent(In) :: RSIn
  Real(RkDBL) :: dnorm
  Real(RkDBL), Parameter :: Pi=3.14159265358979
  dnorm=Real(1.,RkDBL)/sqrt(Real(2.,RkDBL)*Pi)*exp(-(RSIn**2/Real(2.,RkDBL)))
End Function dnorm

Subroutine allocate_Integration_Variables_sub(this,ISDim)

```

```

Implicit None
Class(Int_Var), Intent(InOut) :: this
Integer(Iks), Intent(In) :: ISDim
Integer(Iks) :: ISStat
If(Allocated(this%RVUpper)) Then
    deallocate(this%RVUpper)
EndIf
If(Allocated(this%RVLower)) Then
    deallocate(this%RVLower)
EndIf
If(Allocated(this%RVDelta)) Then
    deallocate(this%RVDelta)
EndIf
If(Allocated(this%IVInfin)) Then
    deallocate(this%IVInfin)
EndIf
If(Allocated(this%RVCorr)) Then
    deallocate(this%RVCorr)
EndIf
Allocate(this%RVUpper(ISDim),&
    &this%RVLower(ISDim),&
    &this%IVInfin(ISDim),&
    &this%RVDelta(ISDim),&
    &this%RVCorr(ISDim*(ISDim-1)/2),&
    &Stat=ISStat)
If(ISStat/=0) Then
    write(*,*) "allocation_error"
    stop
End If
End Subroutine allocate_Integration_Variables_sub

Subroutine make_Correlation_Vector_sub(ISDim,RMCorr,RVCorr)
    !< transforms a correlation matrix into a correlation vector
Implicit None
Integer(Iks), Intent(In) :: ISDim

```

```

    Real(RkDBL), Intent(In), Dimension(ISDim,ISDim) :: RMCorr
    Real(RkDBL), Intent(InOut), Dimension(ISDim*(ISDim-1)/2) :: RVCorr
    Integer(Iks) :: c1,c2
    RVCorr=0.
    Forall(c1=1:ISDim,c2=1:ISDim,c1<c2)
        RVCorr(c1+((c2-1)*(c2-2))/2)=RMCorr(c1,c2)
    End Forall
End Subroutine make_Correlation_Vector_sub

end Module Selection_Gain

Module Interface_Definitions
    Interface
        SUBROUTINE MVIDST( N, NU, LOWER, UPPER, INFIN, CORREL, DELTA,MAXPTS,&
            & ABSEPS, RELEPS, ERROR, VALUE, INFORM )
        Use Data_Kind
        Implicit None
        Integer(IkM), Intent(In) :: N, NU
        Integer(IkM), Intent(In), Dimension(N) :: Infin
        Real(RkDBL), Intent(In), Dimension(N) :: Correl, Delta, Lower, Upper
        Integer(IkM), Intent(In) :: Maxpts
        Real(RkDBL), Intent(In) :: Abseps
        Real(RkDBL), Intent(Out) :: Releps, Error, Value
        Integer(IkM), Intent(Out) :: Inform
        End SUBROUTINE MVIDST
    End Interface
End Module Interface_Definitions

Program Test
    use Data_Kind
    use Selection_Gain
    Implicit None
    Integer, Parameter :: ISNuSt=10 !< number of selection stages
    Character(len=10):: time !< time
    Real(RkDBL), Allocatable, Dimension(:) :: RVPrSe !< vector with proportion
        selected

```

```

Real(RkDBL), Allocatable, Dimension(:) :: RVGeGa !< vector with genetic gain
Real(RkDBL), Allocatable, Dimension(:, :) :: RMCorrel !< correlation matrix,
! upper right: breeding goal, rest: selection stages
Integer(Iks) :: c1, c2, ISStat !< else
Allocate(RMCorrel(ISNuSt+1, ISNuSt+1), RVPrSe(ISNuSt)&
&, Stat=ISStat)
! set up correlation matrix
RMCorrel=1.
Forall(c1=1:ISNuSt+1, c2=1:ISNuSt+1, c1/=c2)
    RMCorrel(c1, c2)=0.95
End Forall
! set up vector with proportion selected at each stage
RVPrSe=0.5
! calculate the genetic gain after each selection stage in standard deviations
call get_SelectionGain_sub(&
    &ISNuSt, &
    &RVPrSe, &
    &RMCorrel, & !input
    &RVGeGa & ! output
    &)
write(* , *) "mean_of_the_truncated_multivariate_normal_distribution", RVGeGa
End Program

```


Danksagung

Bedanken möchte ich mich bei Prof. Norbert Reinsch für die Überlassung des Themas und die richtungweisenden Diskussionen.

Weiterhin möchte ich mich bei Dörte Wittenburg und Friedrich Teuscher für wertvolle mathematische Hinweise sowie Heiko Güldenpfenning vom Rinderzuchtverband Mecklenburg-Vorpommern für die Überlassung von Daten zur Züchtungsplanung bedanken.

Ein besonderer Dank gilt meinen Eltern und Annett Specht für die wichtige moralische Unterstützung während der vergangenen drei Jahre.

Lebenslauf

Zur Person

Name	Börner
Vorname	Vinzent
Geburtstag	01.01.1976
Geburtsort	Ueckermünde
Staatsangehörigkeit	Bundesrepublik Deutschland

Ausbildung

07.1994	Abitur, Gymnasium Eggesin
08.1994-08.1997	Ausbildung zum Bankkaufmann, Sparkasse Uecker-Randow
04.2002-09.2002	Chemiestudium, Freie Universität Berlin
10.2002-03.2005	Bachelorstudium Agrarwissenschaften, Humboldt-Universität Berlin
04.2005-09.2006	Masterstudium Nutztierwissenschaften, Humboldt-Universität Berlin