GENOMIC SELECTION FOR TRAITS OF ECONOMIC IMPORTANCE IN SHEEP

A thesis presented in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy
in
Animal Breeding and Genetics

at
Massey University, Manawatū,
New Zealand

Alfredo Andrés Lepori Honeyman

2014

Supervisors:
Professor Nicolas Lopez-Villalobos
Professor Hugh T. Blair

To my beloved wife, Loreto

ABSTRACT

The main objective of this thesis was to analyse the inclusion of genomic information of production traits into a multitrait sheep breeding programme evaluated for 20 years using deterministic and stochastic simulation models. The breeding objective was to reduce faecal egg score (FES), decrease yearling weight (YW) and increase 160 days lamb carcass weight (CW). The selection criteria included 160 days live weight (instead of CW) plus YW and FES. The first study developed a stochastic model selecting animals based on their individual breeding values estimated using best linear unbiased predictor (BLUP) procedure with a multitrait animal model. The model was validated using a deterministic multitrait selection index; obtaining similar prediction responses for breeding objective and selection criteria traits. The second study deterministically evaluated the inclusion of genomic information explaining different proportions of CW and YW genetic variances into a selection index. Under the same selection scheme a selection index having only genomic information obtained lower accuracies and genetic gains compared to a selection index considering phenotypic information. If shorter generation intervals are implemented, a selection index including phenotypic and genomic information explaining low proportions of the trait's genetic variance could achieve higher genetic and economic gains. The third study evaluated genetic responses of a stochastically modelled breeding flock selecting ewes based on BLUP estimated breeding values and selecting rams based on genomic breeding values (GBV) for CW. The fourth study evaluated accuracy of prediction of CW GBV using the same simulated model. Carcass weight GBVs were calculated in a validation population using single nucleotide polymorphism (SNP) effects from a training population. The further apart the genetic relationship between these two populations, lower the GBV accuracy. Resultant accuracies depended on the proportion of total genetic variance explained by genomic information and also the variance accounted by each SNP, therefore an appropriate GBV estimating method has to be chosen to achieve accuracies as high as possible. Stochastic models proved to be more versatile for managing data, also showing variation of the genetic responses. In contrast, deterministic models were faster regarding computer processing times. The study proved that a breeding programme combining GBV and BLUP estimated breeding values can increase genetic responses by selecting animals at early stages of life.

DECLARATION

This thesis contains no material that has been accepted for a degree or diploma by the University or any other institution. To the best of my knowledge no material previously published or written by another person has been used, except where due acknowledgement has been made in text.

Alfredo Lepori

19 March 2014.

ACKNOWLEDGEMENTS

I would like to sincerely thank my supervisors, Professor Nicolás Lopez-Villalobos and Professor Hugh Blair, for all their assistance, guidance and counselling over the course of my doctoral studies at Massey University.

I also would like to thank Debbie Hill, Postgraduate and Research Administrator for the Institute of Veterinary, Animal and Biomedical Sciences, Massey University, for her cooperation and support in the attainment of the goal of this study.

I want to express my sincere gratitude to the Christian community of Emmanuel congregational church, especially to the Toulon family, Hugh and Nola Neilson, and Bob and Helen Abblet. Their love and support was extremely valuable making my family and I felt at home.

I am extremely grateful to the Adeyinka family, Folusho, Femi, Mojo, Omolayo and Ayokunle. Thank you for accepting me as part of your family; you have become part of mine.

My gratitude is also extended to Andrés and Carolina Reidel, José and Elizabeth Solis and also Alex and Vicky Grinberg for their friendly support and encouragement while completing this study.

I would like to thank all my fellow post-graduate students, for providing a friendly environment at the Institute of Veterinary, Animal and Biomedical Sciences, Massey University. Especially, to the Modelling and Breeding Club for being a constant source of knowledge, discussion and good conversation. I would like acknowledge Dr. José García Muñiz, for his valuable comments regarding this study.

Special thanks must go to my parents-in-law, Milagro, Carla and Juan Carlos whose encouragement and blessings were extremely appreciated.

Finally, I would like to give my deep gratitude to my wife, Loreto, and my beloved children, Martina, Sofia and Agustín. Their encouragement, understanding, patience and support made this thesis possible.

CONTENTS

LIST OF TABLES

Table **Title** **Page**

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

160W	Live weight at 160 days
BLUE	Best linear unbiased estimate
BLUP	Best linear unbiased prediction
CW	Lamb carcass weight
DNA	Deoxyribonucleic acid
EBV	Estimated breeding value
EV	Economic Value
FES	Faecal egg score at 160 days
GBLUP	Genomic best linear unbiased prediction
GBV	Genomic breeding value
MAS	Marker assisted selection
PI	Phenotypic based multi trait selection index
REW	Relative economic weight
SNP	Single nucleotide polymorphism
TBV	True breeding value
WW	Weaning weight
YW	Live weight at 1 year of age (yearling weight)

CHAPTER 1

General introduction

New Zealand is recognised worldwide for livestock production, specifically dairy, beef and sheep production. New Zealand sheep industry strongly relies in lamb meat exports (Beef + Lamb New Zealand 2013); therefore the amount of lamb carcass weight produced is a trait that is very desirable to be improved as is directly related to the farms income. As carcass weight is a trait that can not be measured in living animals, a positive correlated trait like live weight at 160 days can be used, in order to select animals as breeders improving carcass weight (Bennett et al. 1991). Other traits of economic importance are, faecal egg score related to the animal's parasite load and adult weight associated with maintenance costs (Huisman et al. 2008). These traits affect negatively farm profit by increasing farm costs, therefore it is desired to minimise or reduce the genetic gain of these traits in order to improve the profit of the farms (Amer 2000).

Well designed breeding programmes achieve a breeding goal by identifying and choosing for breeding, the most suitable animals for the production system (Harris et al. 1984). Genetic improvement programmes can be designed by the implementation of a proper selection index that maximises the genetic gain of the traits included in the breeding objective, which are traits identified as targets, and have an impact on the production system (Blair & Garrick 2007). It can be said that New Zealand has been successful in the implementation of breeding programmes, as an example, average lamb carcass weight has increased from 16.9 kg in 2006 to 18.2 kg on average in 2010 (Beef + Lamb New Zealand 2012b).

Classical genetic improvement programmes rely on quantitative genetics to select individuals as breeders using phenotypic records from the selection candidates and/or their relatives (Ruane & Sonnino 2007). Genetic evaluations have been very important in New Zealand, helping to improve livestock production (Crawford 2003). These evaluations are statistical techniques that generate predictions of an animal genetic merit, allowing the ranking of animals in a breeding population for replacement. Initially, genetic evaluations were conducted using best linear predictor procedures, but when computers with enough processing capacity were available, best linear unbiased prediction (BLUP) procedures were commercially implemented (Blair & Garrick 2007). BLUP procedures use phenotypic records and genealogical data to generate a numerical representation of an animal genetic merit named estimated

breeding values, and with these values it is possible construct an animal genetic ranking (Henderson 1975).

The development of technologies that enables the manipulation of genetic material at DNA level, have allowed the discovery of the molecular source of variation for animal production and diseases traits (Crawford 2003). Meuwissen et al. (2001) presented the methodology for genomic selection, using information from thousands of single nucleotide polymorphism genotypes allowing the estimation of a trait's genomic breeding value, for animals that may not have phenotypic records of their own or from their relatives. The expectation of incorporating genomic information into animal production selection programmes is that, the use of DNA information will help to improve the rate of genetic gain compared with programmes using just phenotypic information (Meuwissen et al. 2001).

In sheep production, studies have been developed suggesting that the use of genomic selection in breeding programmes will be beneficial to the genetic gains of the evaluated production system. The researches involve evaluations of genetic gains in selection indices for dual purpose animals, meat (as terminal sires) or wool related traits (Pickering et al. 2013; Swan & Brown 2013; Van der Werf 2009), and assessment of accuracy of genomic breeding values for production traits (Sise & Amer 2009; Slack-Smith et al. 2010). The conclusion of these studies showed that the use of genomic selection can produce higher genetic gains for traits of economic importance, but none of them included genomic information in the same way.

None of the reviewed studies have shown the long term variation markers when using genomic selection in a breeding programme, nor have accuracies been evaluated based on the true breeding value of the genotyped trait.

The objective of this thesis was to analyse how, the long-term genetic responses of economic important production traits are affected, when genomic information of one of those traits is included into a multitrait sheep breeding programme evaluated over 20 years. Subsequently, analyse the long-term behaviour of genomic breeding values accuracies using the true breeding values of the genotyped trait, with genomic breeding values estimated under two different methodologies. Finally, a comparison of breeding programmes that include genomic information and breeding programmes based only

on phenotypic records was required, to allow readers to make an informed conclusion on the differences between them.

To achieve the objectives of this thesis, deterministic and stochastic simulation models were developed. Simulated models were created with a breeding objective that decreases faecal egg score (to improve parasite resistance), decrease yearling weight (to reduce maintenance costs) and to increase carcass weight of the lambs (increase production income).

CHAPTER 2

A review of genomic selection for production traits

2.1. Introduction

2.1.1. Current selection

Animal breeding is a technology involving application of the knowledge of genetics and other disciplines (e.g. economics, farm management and reproduction) to improve animals (Garrick & Snell 2005). Differences between individuals are the raw materials on which breeders work; it is known that the variation between individuals is due to differences in their genotypes and their environmental experiences. A third component of the difference comes from joint effects between genotype and environment which cannot yet fairly be attributed to one or the other, or non-additive combination effects (Lush 1943). Quantitative genetics is concerned with the inheritance of those differences between individuals that are by degree rather than of kind, in other words quantitative rather than qualitative differences (Falconer & Mackay 1996). To decide which animals are the best ones to produce the next generation, the breeder has to choose which animals have the best combination of several traits to achieve the highest genetic improvement (in some cases stated as the breeding objective).

In 1942, Hazel and Lush stated that multi-trait selection using a selection index is more effective than other methods ways of selection, because it achieves the maximum genetic improvement per unit of time and effort expended (Hazel & Lush 1942). The aggregate value of an animal is the addition of its various genotypes (assuming different but overlapping genotypes for each economic trait), and each genotype should be incorporated into the index according to the relative economic influence of that trait (Hazel 1943). Having this in mind, the net genetic improvement obtained by selecting a group of animals is the accumulation of the genetic gains made for all the traits which have economic importance.

Animal industries around the world require the identification of animals of high genetic merit for traits of economic relevance. The breeding objective consists of two components, estimates of genetic merit for the traits to be improved and their economic values. An estimate of the genetic merit is required for each trait included in the breeding objective. This estimated genetic merit is typically called the breeding value, estimated breeding value or expected progeny difference (Blair & Garrick 2007). Selection for economically important quantitative traits in animals and plants is

traditionally based on phenotypic records of an individual and/or its relatives (Meuwissen et al. 2001). To enable the estimation of genetic merit, it is assumed that a large number of genes, each with small individual effects (the infinitesimal model) contribute to the expression of the phenotype and by their interaction the additive effects can be elucidated. The prediction or estimation of breeding values for animals using phenotypic and genealogical information is nowadays commonly calculated by best linear unbiased prediction (BLUP) (Garrick & Snell 2005; Meuwissen et al. 2001; Montaldo & Meza-Herrera 1998).

2.1.2. Opportunities for selection schemes

Selection schemes are based in the principle of optimising the annual genetic gain, which according to the equation presented by Rendel and Robertson in 1950 (Garrick & Snell 2005; Lopez-Villalobos & Garrick 2005; Meuwissen 2003) shows 4 elements that can be controlled to change the rate of genetic gain,

$$\Delta G_T = \left(\frac{i \times r_{TI}}{L} \right) \times \sigma_g$$

where:

ΔG_T is the annual genetic gain of the objective.
i is the intensity of selection.
r_{TI} is the correlation between the true breeding value (TBV) and the estimated breeding value also known as accuracy of selection.
L is the length of the generation interval for the selected population.
σ_g is the genetic standard deviation of aggregate genotype of all the animals in the population.

Therefore any new technology or system that enables the optimisation of the genetic gain by changing any component of this equation will be of high commercial interest, but as stated by Ruane & Sonnino (2007) it needs to be assessed if implementing these new technologies into breeding programmes are cost-effective. Molecular or DNA technologies, could offer the possibility to include measurements with reliable accuracy of evaluation at younger ages compared with the current selection schemes (Garrick & Snell 2005). This evaluation at a younger age could be as early as

embryonic stages (Georges & Massey 1991) or even could be done at a sexual cellular level (Haley & Visscher 1998). Improved genetic gain would also be possible when the traits under a conventional selection programme have low accuracy, such as traits with low heritability or traits with few recordings (Meuwissen 2003); when traits are not available at the time of selection (carcass traits) or when traits are sex-limited (Haley & Visscher 1998; Meuwissen 2003).

2.1.3. Molecular genetic markers

A simple definition of genetic markers can be, an observable genetically-controlled variation that follows a Mendelian pattern of inheritance (Williams 2005). Molecular techniques allow for the detection of these variations or polymorphisms existing among individuals in a population for certain DNA regions (Montaldo & Meza-Herrera 1998). Genetic markers can take a number of forms depending on which molecular technique is used to detect and produce them, hence the choice of using any of them depends on the goal of study and the variability of the information needed (Vignal et al. 2002).

Vignal et al. (2002) presented the following examples of molecular markers that according to the authors are the main ones:

- Restriction fragment length polymorphisms (RFLP)
- Polymerase chain reaction - Restriction fragment length polymorphisms (PCR-RFLP)
- Randomly amplified polymorphic DNA (RAPD)
- Amplification fragment length polymorphisms (AFLP)
- Single stranded conformation polymorphism (SSCP)
- Microsatellite
- Single nucleotide polymorphism (SNP).

The differences between them involve basically the type of information that they can provide, and the requirements and characteristic of the techniques used for obtaining each marker.

2.1.3.1. RFLP

Restriction fragment length polymorphism (RFLP) are considered to be the first DNA-based molecular markers (Edwards & McCouch 2007). RFLPs can be described as species-specific sequences of genomic variations detected as differences in DNA fragment lengths after a process called Southern blot hybridisation. This is achieved by treating whole genomic DNA material is treated with a restriction enzyme followed by separation by gel electrophoresis (Mueller & Wolfenbarger 1999; Southern 1975; Zehner et al. 1998). The polymorphisms identified with RFLPs can be due to single base changes generating loss or gain of restriction sites, or may be from insertion or deletion between restriction sites, RFLPs are highly locus-specific molecular markers (Edwards & McCouch 2007; Mueller & Wolfenbarger 1999).

2.1.3.2. PCR-RFLP

Isolation of enough DNA material for RFLPs and its analysis using Southern blot methodology is considered technically demanding and also time consuming (Beckmann 1988). For this reason another method which combines the Polymerase chain reaction (PCR) and RFLP was implemented.

The PCR technique was first proposed by Mullis et al. (1986). One PCR reaction cycle can be described as a three-step process: denaturation, annealing and extension. A PCR cycle duplicates a target DNA sequence by using two flanking primers (short synthetically generated nucleotides sequences), that bind to each of the target DNA strands in opposite directions, serving as a staring point for the synthesis of the DNA. By the successive repetition of a cycle the so called "chain reaction" amplification is accomplished, obtaining in very short time 2^N times the amount of the target DNA sequence, with N being the number of the performed cycles (Huang 2014; Mueller & Wolfenbarger 1999; Mullis et al. 1986).

In PCR-RFLP, the PCR process is used to rapidly amplify only the DNA regions flanked by the two primers in the annealing stage, after which the PCR products are submitted to RFLP analysis to identify polymorphisms. Therefore, a larger quantity of genomic samples can be evaluated in a shorter time (Edwards & McCouch 2007; González Andrade 2010; Higuchi et al. 1999; Zehner et al. 1998).

2.1.3.3. RAPD

Random amplified polymorphic DNA (RAPD) was first described in 1990 by Williams et al. (1990) and Welsh & McClelland (1990). This procedure utilises low stringency PCR amplification with a random DNA sequence as a single primer that generates an array of unknown DNA fragments that is specific for each DNA strain, and are used to generate a genetic profile (Wang et al. 1993; Welsh & McClelland 1990; Williams et al. 1990). One issue is the low potential of exactly reproducing RAPD because the methodology depends highly on the PCR conditions (Vignal et al. 2002). This methodology generates dominant markers (Mueller & Wolfenbarger 1999; Vignal et al. 2002; Williams et al. 1990), but homologous alleles can sometimes be identified with the help of pedigrees (Mueller & Wolfenbarger 1999).

2.1.3.4. AFLP

This technique was proposed by Vos et al. (1995), based on a selective PCR amplification of restriction fragments from a total digest of genomic DNA. They structured the methodology as a three step procedure:

1. Restriction of the DNA and ligation of oligonucleotide adapters.
2. Selective amplification of sets of restriction fragments.
3. Gel analysis of the amplified fragments.

AFLPs are easy to use in the laboratory and compared with RAPDs have better reproducibility. But on the other hand, the process is more technically demanding because AFLP patterns are more complex as multiple loci are screened at the same time (Edwards & McCouch 2007; Mueller & Wolfenbarger 1999; Zhang et al. 2014). The main drawback of this method is the difficulty in detecting homologous markers, because it primarily generates dominant rather than co-dominant markers (Mueller & Wolfenbarger 1999; Vignal et al. 2002).

This methodology is still largely used for genotyping, phylogenetic analysis or the identification of species which feature large complex genomes, therefore new tools are designed for scoring, studying or managing ALFP data (Zhang et al. 2014).

2.1.3.5. SSCP

The single-strand conformational polymorphism (SSCP) method is a genotyping technology which was first proposed by Orita et al. (1989a). The authors developed this methodology as a fast, simple, powerful and sensitive tool for detecting specific DNA sequence changes, even on a single-base scale (Fukuoka et al. 1994; Orita et al. 1989b). This methodology allows segments of DNA that have been amplified with specific primers and PCR to be quickly examined, and to identify in a single strand of DNA any sequence variation (Humphries et al. 1997; Yao et al. 1996). This is done without the involvement of restriction enzyme digestion, blotting, or hybridization to probes (Orita et al. 1989b).

SSCP patterns can be significantly affected for example by the temperature used for the electrophoresis process, the glycerol concentrations in gel, the buffer concentration and conductivity. Also, the optimal conditions for detection of the polymorphisms are affected by the length of the analysed nucleotides sequences (Fukuoka et al. 1994; Humphries et al. 1997).

2.1.3.6. Microsatellite

Microsatellites, also known as simple sequence length polymorphisms (SSLP) or simple sequence repeats (SSR), consist of short nucleotide sequences (di-, tri- or tetra-nucleotide patterns) repeated in tandem several times (Edwards & McCouch 2007; Mueller & Wolfenbarger 1999). One advantage is that these markers are co-dominant; therefore the heterozygous form can be differentiated from the homozygous form. Another benefit is that the results are highly reproducible and are abundant in eukaryotic genomes (Edwards & McCouch 2007).

A drawback of this method is the need of developing species-specific primers, which requires a considerable degree of molecular skills (e.g. cloning and sequencing) and also patience and time as the procedure may take several months (Mueller & Wolfenbarger 1999). It is very unusual that primers developed for one species could be used further than just the closest relatives. Therefore, for every new species analysed new microsatellite primers need to be developed (Edwards & McCouch 2007; Mueller & Wolfenbarger 1999).

2.1.3.7. SNP

SNP is the acronym for single nucleotide polymorphism (Montaldo & Meza-Herrera 1998; Vignal et al. 2002). These are stable point mutations that constitute the most common type of genetic variation (Gilles et al. 1999), and they are characterised by the variation of a nucleotide at a single base. Because of their widespread nature throughout the genome, they are considered as potentially valuable genetic markers (Garrick & Snell 2005; Gilles et al. 1999). There are at most four alleles, for the four bases, A, T, C and G, at any one position in the genome but most often there are only two alleles (Garrick & Snell 2005; Vignal et al. 2002). Most commonly, only two alleles occur, firstly, because the probability of two independent base changes occurring at a single position is very low (Vignal et al. 2002). Secondly because of a bias towards transition mutations (purine-purine (A↔G) or pyrimidine-pyrimidine (C↔T)) over transversion mutations (purine-pyrimidine or pyrimidine-purine (A↔C, A↔T, G↔C, G↔T)), leading to a prevalence of two SNP types. This diallelic nature makes individual SNPs intrinsically less informative than other markers which show more variation, but a large number of SNPs compensate for their low variability (Haley & Visscher 1998). This is due to how the SNPs are organised in the chromosome, SNPs that are near each other tend to be inherited together so regions of close related markers can be identified. These regions of linked variants are known as haplotypes (The International HapMap Consortium 2008).

SNPs, of all the molecular markers are considered to be the best choice (Edwards & McCouch 2007).SNPs provide the most marker density, have a very low mutation rate and allow to make inferences across independent datasets (Nielsen 2000). Also, in some situations SNPs are the only way for finding markers very close or within a gene of significant importance, or can provide the means to detect functional genetic elements (Edwards & McCouch 2007; Meuwissen et al. 2013). In addition this technology is becoming more cost-effective for most of the major livestock species. Using SNP-chip genotyping technologies panels of approximately 50,000 or even over 700,000 genome-wide SNPs are available (Pérez-Rodríguez et al. 2013; VanRaden et al. 2011).

2.2. Selection schemes using genomic information

It has been stated that most traits of economic interest behave as if there were many loci influencing their expression and therefore their variation. This assumption leads to the idea of polygenic inheritance and the infinitesimal model (Garrick & Snell 2005; Hayes & Goddard 2001). In reality, a small number of genes seem to affect a large proportion of the variation in phenotypic characteristics; accompanied by a large number of genes each having small effects (Garrick & Snell 2005; Georges & Massey 1991; Hayes & Goddard 2001). The inclusion of genomic information into selection programmes to improve animal production rely on the expectation that information at the DNA level will lead to faster genetic gain than the one achieved based only on phenotypic information (Meuwissen et al. 2001). Therefore accurate appraisal about the implications of introducing genomic data into a breeding scheme concerning the impact that this inclusion will have in the phenotypic response, is needed before developing an accurate breeding programme.

Lande & Thompson (1990) presented a deterministic simulation model that included molecular genetic information using selection index theory. This was further developed by Dekkers (2007) where he presented equations to include marker information as a correlated trait (with heritability equal to 1). Several assumptions were made to develop the model, highlighting the multivariate normality of the marker information, which the author states that it can be considered "approximately valid" if markers breeding values are based on a large number of markers allowing to be represented as a polygenic trait. Another consideration is that selection index theory does not account for changes in the genetic variance due to changes in gene (allele) frequencies (Bulmer effect). This issue was solved using a selection index software package that adjusts for the Bulmer effect. Based on Dekker's (2007) publication, Janssen-Tapken et al. (2010) compared different selection strategies for improved productivity and marker assisted selection (MAS) for disease traits; Togashi & Lin (2010) analysed different selection methods for genetic improvement of net merit for two traits with the inclusion of marker information; Pryce et al. (2010) presented a deterministic model using the four pathways of selection of Rendel & Robertson (1950) while also accounting for the rate of inbreeding per generation. In sheep production Sise & Amer (2009) presented a deterministic approach using selection

index theory to predict the response to genomic selection in dual purpose sheep flocks. Besides the methodology used the results of these studies suggest that a large amount of information is required to a estimate genomic breeding values (GBV). Better genetic gains may be obtained because genomic information explains a larger amount of the phenotypic variance, and/or a reduction of the generation interval.

Even though deterministic models have the advantage of being less demanding in computing time than stochastic models, most of the work evaluating the impact of information on individual genes or markers has been done using stochastic simulations. This is because deterministic prediction of the selection response by selection index theory needs multivariate normality, which does not occur when only a limited number genes are used in the selection simulation, and also because selection index predictions ignore Bulmer effects (Dekkers 2007).

Stochastic simulation models using BLUP with the inclusion of genetic markers have been compared with conventional BLUP to assess the effectiveness of MAS (Fernando & Grossman 1989; Zhang & Smith 1992), even though it was never commercially applied. The great change in the perception of MAS was when Meuwissen et al. (2001) published their work in which they presented the methodology for genomic selection, a term that was first introduced by Haley and Visscher (1998) (Meuwissen 2007). The methodology is a form of MAS that uses SNPs as a dense marker map combined as "haplotypes", which represents the total genetic variation of the trait under study, and phenotypes in a training population to estimate the effect of each haplotype in the breeding objective traits. This allows the estimation of breeding values for animals that have no phenotypic record of their own and no progeny (Meuwissen et al. 2001). These predictions of animal genetic merit became known as genomic estimated breeding values (GBV). The statistical methods used by Meuwissen et al. (2001) will be explained in the next section.

Since the appearance of genomic selection, several methodologies have arisen to estimate the GBVs and also for analysing the accuracy of different methods in different scenarios (Amer & Payne 2009; Hayes et al. 2006; Hayes & Goddard 2010; Luan et al. 2009; Pérez-Rodríguez et al. 2013; Schaeffer 2006; Toosi et al. 2010; Weigel et al. 2010). Regarding sheep, studies have evaluated the accuracy of genomic selection for production traits (Sise & Amer 2009; Slack-Smith et al. 2010), genetic

gains in selection indices for dual purpose animals or wool related traits (Pickering et al. 2013; Swan & Brown 2013), or for the detection of faulty genes by identifying animals carrying the undesirable SNP allele (Sise et al. 2008).

In general, the literature concludes that the inclusion of genomic data into selection schemes produces an increase in the rate of genetic gain for the traits or index under simulation. All studies concur that large datasets are required to achieve high accuracies but there is no agreement on a unique methodology for estimating the GBVs, with small differences shown by different statistical analyses.

2.3. Prediction of genomic values

Genetic improvement programmes try to optimise the genetic merit of the population studied. This merit can be a linear or nonlinear combination of genetic values of the traits under selection, and cannot be directly observed. This requires that the genetic values must be inferred from data (Gianola 2000). The novel approach proposed by Meuwissen et al. (2001) using dense marker maps, analysed data sets that included thousands of variables (SNPs) to estimate the allelic effects of each SNP, for a finite animal population. Their intention was to utilise genomic selection procedures to explain all genetic variation by genetic markers and thereby, acquire a higher accuracy of selection in situations, where the accuracy of selection of "non genomic procedures" was low, like traits with low heritability, measured late-in-life, or after-slaughter. Also their proposal of genomic selection was to fit all markers (whether or not they were statistically significant) and to estimate all gene effects simultaneously.

Meuwissen et al. (2001) showed three procedures for analysing genomic data and therefore inferring genomic breeding value estimates or GBVs. These methods were Least squares, BLUP and Bayesian estimation (BayesA and BayesB). Numerical examples will be presented to illustrate how genomic information is incorporated into the mathematical procedures to obtain GBVs.

2.3.1. Least squares procedure

The least squares procedure in statistics is used to predict the expected value of a vector of observations, being $E(\mathbf{y}) = \mathbf{Xb}$, leading to the equation:

$$\mathbf{X'X\hat{b} = X'y}$$

where:

X is an incidence matrix associating individuals and their fixed effects,

y is a vector of the known observations,

$\hat{\mathbf{b}}$ is a vector with the regression coefficients for fixed effects.

Using this statistical procedure to estimate the solution for a very large number of effects (variables) is not possible, because there will not be enough degrees of freedom to fit all the effects at the same time (Meuwissen et al. 2001; Meuwissen 2003; Searle 2006). However, least squares can be used to test all the genomic variables included, by analysing them one by one for their statistical significance (stepwise), and therefore enabling the inclusion only of the ones that improve the fit of the model (Meuwissen et al. 2001; Meuwissen 2003). The major issue that arises from the use of least squares is that a bias problem occurs by setting the effects of the non-significant genes to zero and giving full effect to the ones that are statistically significant. Therefore, this procedure performs poorly because it greatly overestimates some variable effects (the ones that do have a statistically significant effect) and underestimates others (the ones that are non-significant but with some effect) (Meuwissen et al. 2001).

2.3.1.1. Numerical example illustrating the use of the least squares procedure to estimate genomic breeding value

Two sheep populations of 20 animals were simulated, a training population (Table A2.1) and a predicted population (Table A2.2), each one of them with randomly generated weaning weights (WW) with a mean value of 28 kg and a variance of 4 kg^2, also a sequence of 40 genetic markers (SNPs) per animal was simulated, associating them to the weaning weight of each animal, these were generated assuming that the SNPs explain a 100% of the genetic variance for the trait, with frequencies of 0.5 for heterozygous and 0.25 for each homozygous allele and that they have a gamma distribution with a shape parameter of 0.4 and a scale set to 1.66 (Calus & Veerkamp 2007; Hayes & Goddard 2001; Meuwissen et al. 2001; Toosi et al. 2010) this is to represent a small number of genes affecting a large variation proportion, and large number of genes affecting a small proportion.

To see which SNPs predict most accurately the WW of each animal presented in Table A2.1, a stepwise regression (forward-backward) was done using the step function of the data analysis language R (Ihaka & Gentleman 1996), which showed that SNPs 1, 2, 24, 4, 13, 5, 31, 33, 14 and 3 were the ones that most accurately predicted the WW for the 20 animals in the training population (R^2 0.92), leaving aside the information provided by the rest of SNPs. The SNPs identified as significant were analysed using a least squares procedure to obtain the regression coefficients (*b* values) of the 10 selected SNPs. This *b* values, also known as best linear predictors, were utilised to estimate the GBVs of the predicted population as seen in Figure 2.1, obtaining an accuracy of prediction of 0.38.

Results presented in Figure 2.1 shows large differences between weaning weights TBVs and GBVs, being 3.83 kg the largest prediction difference (animal 1,500) and 0.18 kg the lowest prediction difference (animal 2,065).

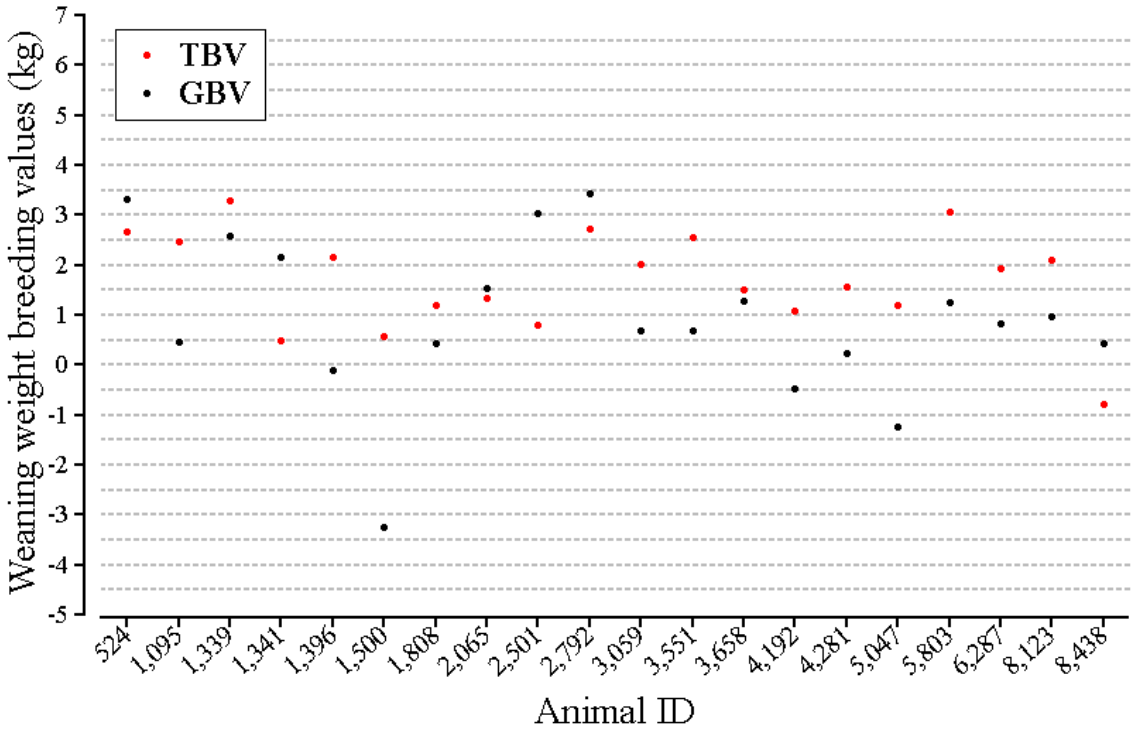

Figure 2.1. True breeding values (TBV) and genomic estimated breeding values (GBV) for simulated sheep weaning weights using least squares.

2.3.2. Best linear unbiased prediction procedure

This statistical procedure is used to estimate the fixed effects and breeding values (random effects) (Falconer & Mackay 1996), and assumes genetic and phenotypic variances are known. In reality the genetic variance (and also the residual variance) might not be known, so they can be estimated using "restricted maximum likelihood" procedure, better known as REML (Falconer & Mackay 1996; Gianola 2000), which accounts for the loss of degrees of freedom incurred in estimating fixed effects (Gianola 2000; Meuwissen et al. 2001).

The SNP information or allelic effects are fitted in genomic BLUP as random effects; therefore this procedure is also called random regression-BLUP (RR-BLUP) (Habier et al. 2007; Moser et al. 2009) or SNPBLUP (Meuwissen et al. 2013). This methodology requires an estimate of the SNP variance. However this procedure assumes that every locus will get the same variance and considers a very small impact of the allelic effects on the related trait (Meuwissen et al. 2001; Meuwissen 2003). Another consideration of interest is that the gene or SNP effects are assumed to be additive (which is appropriate for the prediction of breeding values), but the reality is that some degree of dominance will probably occur in practice (Meuwissen et al. 2001), therefore the dominance effect of some genes is going to be masked by the average effect of the total gene population.

2.3.2.1. SNPBLUP and GBLUP

Another BLUP model which includes genomic information is GBLUP (Hayes et al. 2009b; Hayes et al. 2009a). This model can be defined as an adaptation of the SNPBLUP method. Instead of fitting the SNP information as random effects, in GBLUP the markers information is included using a genomic relationship matrix G, which is derived from them (Daetwyler et al. 2010; VanRaden 2008; Visscher et al. 2006).

SNPBLUP and GBLUP models are equivalent, consequently the genomic breeding values estimates yield from each of the models are going to be the same (Goddard 2009; Habier et al. 2007; Meuwissen et al. 2013).

2.3.2.2. Numerical example illustrating the use of the best linear unbiased predictor procedure to estimate genomic breeding values

Two examples were made to show the use of BLUP with the same two sheep populations of 20 animals simulated in point 2.4.1.1.; two simulated markers, SNP1 and SNP2 have known variances explaining 40 and 30% respectively of the WW total genetic variance. To obtain the genomic information predicting values, the 40 simulated SNPs were included as random variables with the mixed model equation:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \lambda\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix}$$

where:

X is a known incidence matrix relating animals with their fixed effects,
Z is an incidence matrix relating the recorded animal and their random effects,
y is a vector of the known observations (phenotypic records),
b and **a** are regression coefficients for fixed and random effects respectively (BLUE and BLUP values),
I is an identity matrix of the same order as the **Z'Z** matrix.
λ (also known as alpha value) is the value obtained when the residual variance is divided by the random effect variance (for each of the random variables used in the model) $\lambda = \frac{\sigma_e^2}{\sigma_a^2}$.

The first of the SNPBLUP examples (1-alpha) was based on Meuwissen et al. (2001), where each SNP variance was assumed to be the WW total genetic variance (1.6 kg^2) divided by the number of SNPs, this brings an individual variance value of 0.04 kg^2. The residual variance with a value of 2.4 kg^2 was obtained considering the total genetic variance and the phenotypic variance utilised (4 kg^2) in the stochastic simulation. The SNPs variance and the residual variance are needed to obtain the alpha value used in the mixed model equations. The estimates for the random SNPs effects obtained from the training population were multiplied with the SNPs effects values of the predicted population in order to get their GBVs (Figure 2.2). The accuracy of prediction obtained was 0.47.

The second example (2-alpha) differs from 1-alpha in that, as the variances of the SNP1 and SNP2 were known (representing a large proportion of the total genetic variance), these variances were used to estimate the alpha values for each respective known random effect (SNP). The rest of the SNPs with unknown variances were assumed to have a variance value obtained as the difference of the known SNPs variances and the total genetic variance. With this information the BLUP values obtained from the training population were used as for the previous example with the purpose of acquiring the GBVs (Figure 2.2); this resulted in an accuracy of prediction of 0.45.

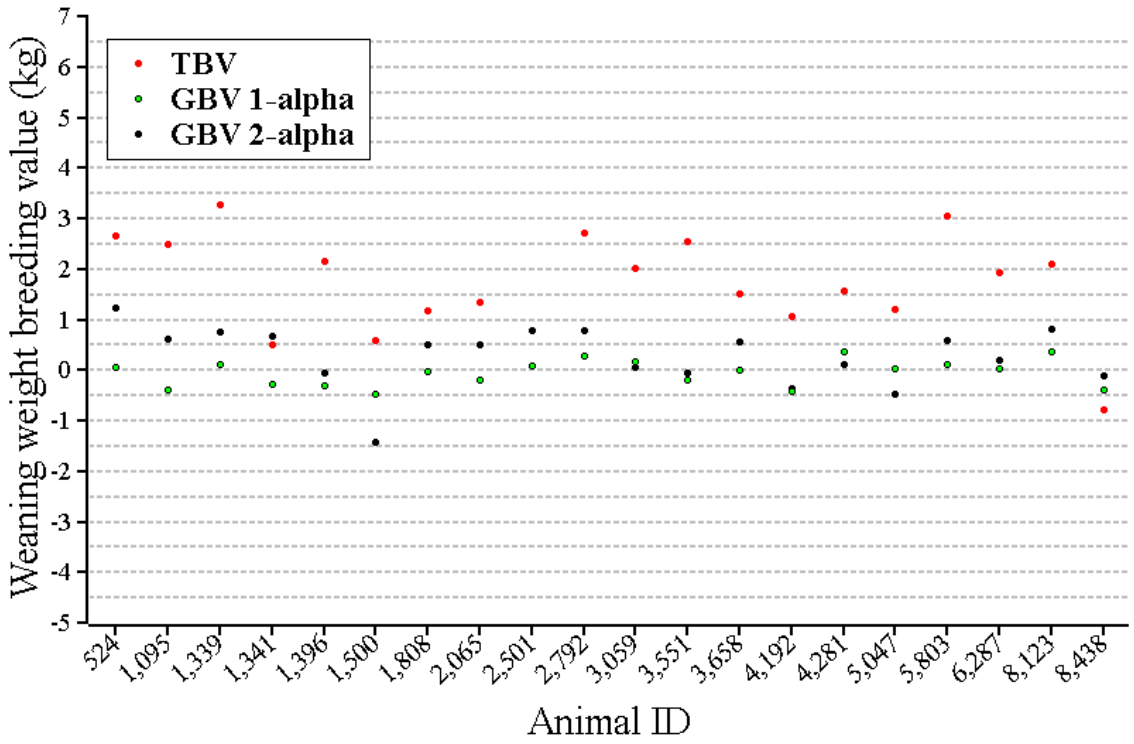

Figure 2.2. True breeding values (TBV) and genomic estimated breeding values (GBVs) for simulated sheep weaning weights using two different BLUP approaches (1-alpha and 2-alpha).

The results presented in Figure 2.2 do not show large differences of the animals predicted GBVs for the two presented methods. The biggest GBV difference of 3.2 kg was obtained by animal 1,339 analysed with 1-alpha and the lowest GBV difference of 0,008 kg was achieved by animal 2,501 analysed with 2-alpha method. Even though the accuracy of the 1-alpha method was slightly higher, as shown in Figure 2.2 the

differences between the animals TBVs and GBVs in average were lower for the 2-alpha BLUP method.

2.3.3. Bayesian Procedures

The premise of Bayesian analysis is the Bayes' theorem, which is based on defining two probability values:

The likelihood probability, which is in a general sense, the probability that is derived from the observed data.
The prior probability, which is an estimate of how likely this set of observations and the associated prior probability, is to occur in the population.

The combination of these two probabilities is an estimation of the probability that an event will occur (known as the joint probability) (Hair et al. 2006). In general Bayesian estimation (BayesA and BayesB) is similar to BLUP (Meuwissen et al. 2001; Meuwissen 2003). The difference between these approaches is that the variance of the allelic effect is assumed individually (for every gene or SNP included in the model), which is estimated by using a prior distribution for the variance of gene *i* (V_{ai}) assumed as:

$V_{ai=0}$ with probability π, and

$V_{ai} \sim \chi^{-2}(v, S)$ with probability (1- π).

Where p depends on the gene mutation rate, and $\chi^{-2}(v, S)$ represents the inverse – Chi squared distribution with v degrees of freedom and scale parameter S. These two parameters (v and S) depend on the distribution of the mutational effects, and in practice they need to be estimated (Meuwissen et al. 2001; Meuwissen 2003).

This methodology of analysis wasn't chosen in the experimental part of this thesis to estimate the simulated GBVs mainly because is more computational demanding and secondly because the level of accuracy shown in the literature doesn't differ much compared to the BLUP methodology including genomic information.

2.3.3.1. The bayesian alphabet

Since Meuwissen et al. (2001) presented the Bayes hierarchical models (BayesA and BayesB) several other bayesian models have been proposed due to discrepancy on the mathematical validity of the models (e.g. assumptions of the SNPs effects distributions). Therefore BayesA and BayesB are considered the starting point for what is now known as the bayesian alphabet.

2.3.3.1.1. BayesA and BayesB

As stated in point 2.3.3, the variance of the allelic effect of the SNP included is assumed individually, this is because considering the SNP effects to have a normal distribution with an invariable variance may not be accurate. According to Meuwissen et al. (2001) the BayesA model, assumed that the variance of SNPs effects had a scaled inverted chi-square distribution, allowing some SNPs to have larger effects than they do under an assumption of normality. However the prior density does not have a peak when the variance of a genomic segment equals 0. This, according to Gianola et al. (2009), actually is the kernel of the density of the t-distribution, which is the *de facto* prior assigned to a SNPs effects in Meuwissen et al. (2001).

Meuwissen et al. (2001) highlights that the difference between their two proposed bayes models, is that in the BayesB model, some SNPs (with a probability of π) have no effect on the evaluated trait, and another proportion of SNPs (with a probability of $1-\pi$) have an effect drawn from a t-distribution. Therefore BayesB can be reduced to BayesA by having $\pi = 0$ (Gianola et al. 2009).

Gianola et al. (2009) state that, in a Bayesian learning context BayesB is wrongly formulated based on the assumption that an *a priori* variance value equal to zero implies absence of an effect of the SNP on the trait. In a Bayesian sense, a parameter having an *a priori* variance value of zero, does not inevitably indicate that it will obtain the value of 0, in fact it could be any value, but known with certainty. Besides the previous statements, BayesB methodology has been largely used and as shown by the Meuwissen et al. (2001) delivers higher accuracies of prediction compared with other analysis methodologies.

2.3.3.1.2. Bayes SSVS or BayesC

The Stochastic Search Variable Selection (SSVS) methodology developed by George & McCulloch (1993) (Verbyla et al. 2009) also called Bayes C (Verbyla et al. 2010) is a methodology is similar to BayesB. It differs by allowing a constant dimensionality to be maintained across all models while enabling the SNPs in the predictive set to change. In other words, a single effect variance common to all SNPs is used instead of locus-specific variances. With this change, the influence of the scaling parameter is reduced (Habier et al. 2011). The major advantage of this method is that it can be implemented using the Gibbs sampler instead of the more computationally demanding algorithms such as the reverse jump algorithm.

2.3.3.1.3. BayesD

This is another methodology which can be described to be a modification of BayesB. This procedure according to Habier et al. (2011) differs from BayesB in that the scale parameter of the inverse chi-square prior for locus-specific variances is treated as an unknown having its own prior.

2.3.3.1.4. BayesCπ and BayesDπ

The purpose of Habier et al. (2011) in developing these two bayesian methods was to address for the drawbacks exposed by Gianola et al. (2009), of the models presented by Meuwissen et al. (2001). BayesCπ and BayesDπ models are similar to BayesC and BayesD models respectively in terms of how the SNP effect variances are simulated leading to different strategies for the inclusion of SNPs in the model. The difference between the respective models being that the proportion π was considered as unknown with a prior distributed uniformly (0.1).

Regarding accuracy of GBVs, the authors have acknowledged that none of the models (BayesA, BayesB, BayesCπ and BayesDπ) outperformed the others. But in terms of computing time BayesCπ and BayesDπ are the models which took less time to run.

2.3.3.1.5. Bayesian Lasso

Bayesian least absolute shrinkage and selection operator (Bayesian Lasso) proposed by Park & Casella (2008) (de los Campos et al. 2009), combines the subset selection (i.e., variable selection) with the shrinkage produced by the standard Bayesian regression, but it does not accommodate pedigree information or regression on (co)variates other than the markers for which a different shrinkage approach may be desired.

2.3.4. Statistical learning methods for estimation of genomic values

Statistical learning or machine learning methods are terminologies referring to a group of methods that emerged due to the need of obtaining information from an immense amount of data. They were developed to optimise the predictive performance based on the automatic discovery of patterns in a dataset through the use of computer algorithms from a training dataset (Bishop 2006). Methods such as least squares procedures, BLUP and the entire bayesian alphabet can be classified as machine learning approaches.

To obtain different approaches to analyse and obtain GBVs with higher accuracies than the ones presented by Meuwissen et al. (2001) other statistical methods such as machine learning algorithms have been proposed for genomic selection (Long et al. 2007). An example is the double hierarchical generalized linear model (DHGLM) developed by Lee & Nelder (2006) which according to Shen et al. (2011) uses a likelihood framework to allow for the estimation of marker-specific variances and therefore GBVs without the need for a prior distribution.

Two other methods also utilised to estimate the marker effects, are partial least squares regression (PLSR) (Moser et al. 2009; Solberg et al. 2007; Solberg et al. 2009; Sölkner et al. 2007) and principal components regression (PCR) (Solberg et al. 2007; Solberg et al. 2009). These are considered dimension reduction methods, and provide another approach to deal with having a greater number of predictors than records (Solberg et al. 2009). PLSR obtained similar accuracies compared with other methodologies analysed by the authors (Bayes regression and RR-BLUP), and PCR obtained slightly lower accuracies. Regarding computational requirements, PLSR and PCR methods needed less computing time.

Other nonparametric approximation methods for obtaining additive genetic values are Kernel regression (KR) and the semi-parametric function of reproducing kernel Hilbert spaces regression (RKHS) procedures. These methods can be embedded into standard mixed-effects linear models without introducing non-linearity (Gianola et al. 2006), thereby allowing the identification of the multiple and complex interactions between many loci at different chromosomes (Gianola & van Kaam 2008).

Moser et al. (2009) and Long et al. (2011) also proposed another method which can be classified as a specific learning algorithm of RKHS. For this method, the nonparametric support vector regression (SVR) proposed by Vapnik (1998), simultaneously minimises an objective function which includes both model complexity and the error on the training data.

The Elastic-Net (EN) algorithm (Croiseau et al. 2011) has been used to estimate each of the available SNP effects, when there is a larger number of SNPs than animal records. This methodology originally proposed by Zou & Hastie (2005) is defined as a penalised regression approach (Croiseau et al. 2011; Waldron et al. 2011). The approach encourages a grouping effect whereby highly correlated prediction variables are likely to be conjointly included or discarded from the model. This leads to a reduction of SNPs included in the prediction equation with only a minor effect on the quality of prediction (Croiseau et al. 2011).

Artificial neural networks, or neural networks (NN) have been proposed for the prediction of complex genotypes using genomic information (Gianola et al. 2011; Okut et al. 2011; Okut et al. 2013). The concept of NN is to find a mathematical representation of the information processing in biological systems, in particular the central nervous system (Bishop 2006). NN is a machine learning procedure that operates as a universal approximator of several complex input functions. These functions called neurons operate in parallel and are arranged into layers which converge into a single output (Gianola et al. 2010; Gianola et al. 2011; Okut et al. 2011). The methodology can be considered as a non-linear regression model, with parameters tuned by Markov Chain Monte Carlo methods or by Bayes theory. Therefore it can capture non-linear relationships between predictors and responses (Gianola et al. 2010; Gianola et al. 2011), and in this way, learn complex functional forms as well as pattern recognition due to their adaptive nature (Okut et al. 2013). The

main challenge is that the methodology is very computational demanding and hence time consuming, especially with large datasets (Pérez-Rodríguez et al. 2013).

There are some other special cases of learning algorithms that are combinations of several machine learning models. These combined models, occasionally labelled as committees, are implemented in an attempt to obtain better performances (higher accuracies or less computing times) than the ones achieved by using an isolated model (Bishop 2006). Bagging is one example of a committee method, the name being an abbreviation for bootstrap aggregating (Bishop 2006; Flach 2012). Bagging is a method developed by Breiman (1996) (Bishop 2006; Gianola et al. 2010; Gonzalez-Recio & Forni 2011). As the name states, the method obtains an aggregated predictor value from generating different predictor values by bootstrapping different random samples or subsets of the original learning dataset and using them as new learning sets (Breiman 1996; Flach 2012).

A committee method used for genome-wide prediction which exploits the bagging method with another machine learning procedure, called tree models, is random forest (RF) (Flach 2012). This algorithm was developed by Breiman (2001) and employed in a genomic breeding value context for the first time by Gonzalez-Recio & Forni (2011). RF constructs several decision trees models on bootstrapped subsets of the dataset (which includes genomic information), averaging each generated predictor to make final predictions. This methodology reduces the prediction error by a factor of the number of tree models evaluated.

Gonzáles-Recio et al. (2010) utilised a further committee method called boosting for genomic selection. Specifically they evaluated the performance of the L_2-Boosting algorithm, showing that the method can be used for regression in high-dimensional problems, and can evince complexity brought by covariates, such as SNPs. The authors stated that L_2-Boosting could be a viable method for genomic selection, but to improve the performance of the procedure, aspects such as the choice of weak learner, stopping criterion, step-size parameter and programming strategy have to be evaluated.

2.3.5. Multi-step vs. Single-step approaches for genomic evaluations

Methodologies for genomic evaluations are executed as a series of steps (Misztal et al. 2009), hence the term multi-step. In theory, it consists of two phases, a training phase for the estimation of the genotypic effects and an application phase where the GBVs of the selection individuals are estimated based on their own genotypic information and the effects from the training phase (Garrick et al. 2009). Some of the methods mentioned could involve traditional BV estimations using an animal model, prediction of genotypic effects (usually in mixed models as random effects), the inclusion of new input data derived from the observations (e.g. offspring averages as data, deregressed evaluations, removing parent average effects), genomic selection, and maybe developing a selection index that merges genomic and polygenic data (Garrick et al. 2009; Misztal et al. 2009; VanRaden 2008). For the inclusion of either all or some of these options the multi-step procedure is said to be a blended methodology, but as reviewed there is not a single way to approach the analysis.

According to Misztal et al. (2009) the multi-step procedure has some disadvantages such as the requirements for certain parameters needed for the estimation of genomic effects and GBVs (e.g. prior variances and weights), and also loss of accuracy and biases attributable to selection. Therefore, they developed a single-step methodology which simultaneously includes phenotypic, genomic and relationship information (Harris et al. 2013). The method modifies the pedigree-based numerator relationship matrix with the inclusion of the genomic relationship matrix proposed by VanRaden (2008) (Harris et al. 2013; Misztal et al. 2009; Swan et al. 2012). This new methodology allows the analysis in one unique procedure of all available individuals, regardless of whether they have been genotyped or just have phenotypic information.

2.4. Concluding observations

The selection of young bulls by artificial insemination companies using genomic selection has increased genetic gains mainly by shortening the generation intervals (Wiggans et al. 2011). But even though more than a decade has passed since Meuwissen (2001) proposed the use of genomic selection, there is still no agreement in

a unique statistical procedure for estimating GBVs, as new methodologies are implemented (Pérez-Rodríguez et al. 2013).

The methodologies for validating genomic selection have become a very important topic as more countries are interested in utilising this new technology (Olson et al. 2011). The main issue that has arisen concerns the realised accuracy of the GBVs which are low compared with progeny tests. It has been very difficult to increase the accuracy to levels that may be considered acceptable, for example 75% for yield traits in dairy cow populations (Wiggans et al. 2011), and 90% for beef related traits (Boerner & Johnston 2013). One of the ways that this lower accuracy has been managed is by using very large training populations that have both genotypic and phenotypic information (Saatchi et al. 2013). Since the number of animals within breeds is sometimes not enough to obtain desired accuracy levels, multi-breed methods have been proposed to increase accuracy (MacLeod et al. 2013; Moghaddar et al. 2013; Saatchi & Garrick 2013).

Another way that has been proposed to increase reliabilities is the use of quantitative trait nucleotides (Ron & Weller 2007; Weller & Ron 2011); these are specific polymorphisms that have a direct effect over a phenotypic characteristic. High density chips (over 700,000 SNPs) are another way that has been proposed for increasing the estimation accuracies, but for the moment this is less cost effective than using a lower density chip with more animals (VanRaden et al. 2011).

As technologies develop and costs decrease, these new molecular and quantitative genetic tools will become a possible alternative in animal selection programmes. Different analysis techniques, such as the ones presented in this chapter, are designed in an effort to integrate the information provided by molecular genetics, and by using quantitative procedures to estimate a trait's genomically estimated breeding value of genotyped animals in a selection breeding programme. The accuracies of the breeding values obtained depend not only on the analysis method utilised but also on the amount of information used to estimate the GBVs and the distribution of the genotypes. Therefore the results obtained in the presented examples must be considered as examples only, because they consist of just a few animals.

This review suggests that the power of genomic selection currently does not rely on the expectation of acquiring better realised accuracies (although there is a lot of effort to overcome this issue) in the prediction of GBVs, but rather on the possibility of selecting breeding animals at younger ages. Because of the limited evidence for increase genetic gain provided by the empirical studies, further simulation studies are needed in order to assess the long-term effects of including genomic information about production traits in multitrait sheep breeding programmes.

2.5. Appendix to chapter 2

Table A2.1. Simulated single nucleotide polymorphisms associated with weaning weight (WW) for a training population of 20 animals.

ID	WW	SNP																																						
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
2239	27.72	0	-1	-1	0	-1	1	0	0	0	-1	0	0	0	-1	0	0	0	0	0	1	0	-1	1	0	1	0	0	0	1	0	0	1	1	0	1	0	0	1	0	1
5678	30.38	1	0	1	1	-1	1	0	-1	0	1	1	1	0	-1	0	1	1	0	-1	0	1	0	0	1	-1	0	1	0	-1	0	1	0	0	-1	0	-1	0	1	1	0
7269	30.70	0	1	0	0	1	0	0	0	1	0	0	-1	-1	0	1	0	0	-1	-1	0	1	-1	0	1	0	0	1	0	0	0	0	1	-1	0	0	-1	0	-1	-1	-1
6867	29.37	0	-1	0	1	1	0	-1	1	1	-1	0	-1	1	1	0	0	-1	0	0	-1	1	-1	0	1	-1	0	0	-1	0	1	0	1	-1	0	1	-1	-1	0	0	0
760	29.46	-1	1	1	-1	0	0	-1	0	1	-1	-1	-1	-1	0	0	-1	-1	0	0	0	1	0	1	1	0	0	1	0	0	1	0	1	1	-1	0	1	0	1	1	0
3439	28.18	-1	-1	1	0	0	1	0	1	0	0	1	1	1	-1	0	-1	1	0	1	0	-1	0	1	1	-1	0	1	1	0	-1	0	0	-1	-1	-1	-1	1	0	-1	1
1347	31.05	0	0	0	-1	-1	0	0	0	1	1	1	-1	1	0	0	1	0	-1	1	1	1	0	-1	1	-1	0	0	1	-1	0	1	0	1	0	-1	0	0	0	-1	-1
3916	28.48	-1	1	0	-1	0	1	1	0	0	0	0	1	0	-1	0	1	0	0	0	0	0	-1	1	0	1	-1	0	-1	-1	-1	0	1	0	0	0	0	1	0	0	-1
6412	32.26	1	1	-1	0	1	-1	-1	0	0	-1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	-1	0	1	-1	1	0	1	-1	1	0	0	-1	0	1	0	0
8579	27.48	-1	0	1	-1	0	0	0	0	1	-1	0	1	1	-1	0	0	1	0	1	1	0	1	0	0	0	-1	1	0	-1	0	1	1	0	0	0	-1	0	0	0	0
755	27.32	0	-1	1	1	0	0	0	1	-1	-1	1	0	1	0	-1	-1	0	1	0	-1	0	1	1	0	-1	-1	0	1	-1	0	0	0	-1	-1	0	-1	0	1	0	0
5777	28.54	0	-1	0	0	1	1	1	0	0	0	0	-1	0	1	1	0	0	0	0	1	1	0	0	0	0	1	0	1	0	0	1	1	1	0	1	0	0	0	1	-1
8347	28.63	1	-1	-1	1	0	1	0	0	-1	0	0	-1	-1	-1	0	0	1	0	-1	1	-1	1	-1	-1	1	0	1	0	-1	1	0	0	0	0	1	-1	1	0	0	-1
5204	29.98	0	-1	1	0	1	1	0	1	0	-1	0	0	0	1	1	1	-1	-1	-1	1	1	-1	1	1	1	0	1	0	0	0	0	-1	-1	0	0	0	1	0	0	-1
4752	28.28	-1	0	0	0	0	0	1	1	-1	0	0	1	0	1	-1	-1	1	1	0	0	0	0	0	1	0	-1	1	1	0	0	0	0	1	1	0	1	0	0	1	1
7768	29.13	-1	1	-1	-1	0	-1	1	-1	0	1	0	0	1	0	0	0	0	0	0	1	-1	0	1	-1	0	0	0	0	0	-1	0	1	-1	1	-1	0	0	-1	0	-1
1375	30.61	1	-1	0	0	0	0	1	-1	0	0	0	0	1	0	0	1	0	1	-1	-1	1	1	1	-1	1	0	0	0	-1	1	0	1	-1	1	-1	1	0	0	-1	1
916	32.43	1	1	-1	1	1	0	-1	1	-1	0	0	-1	0	0	-1	0	1	1	1	0	1	0	-1	1	-1	1	1	1	0	1	1	-1	-1	-1	-1	-1	1	-1	-1	0
8041	29.27	0	0	1	0	0	1	1	-1	-1	-1	0	1	0	0	1	-1	0	-1	0	-1	0	-1	-1	-1	0	1	1	0	0	1	1	-1	0	-1	0	1	0	-1	0	1
5458	26.75	-1	-1	1	1	0	-1	1	-1	-1	0	0	0	1	-1	-1	1	0	-1	1	-1	-1	-1	0	-1	0	-1	0	1	-1	-1	-1	0	-1	-1	1	-1	1	0	1	-1

Table A2.2. Simulated single nucleotide polimorphisms associated with weaning weight (WW) for a predicted population of 20 animals.

ID	WW	SNP																																						
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
6287	29,95	1	-1	0	1	0	-1	0	0	-1	1	1	0	1	1	1	0	1	0	1	1	1	0	-1	0	1	1	-1	-1	1	1	0	0	1	0	-1	-1	0	0	0	0
4192	29,08	-1	1	0	0	0	1	1	1	-1	0	-1	0	1	-1	1	-1	1	1	0	1	0	1	1	1	0	-1	1	-1	1	0	0	0	0	0	0	0	1	1	-1	0
3551	30,55	0	0	0	-1	0	0	0	-1	-1	-1	1	0	0	1	0	0	-1	0	-1	-1	0	0	0	0	1	0	1	0	1	1	1	0	0	1	1	1	0	0	1	1
3658	29,52	0	1	0	0	-1	-1	0	0	-1	-1	0	1	0	0	-1	1	-1	0	0	1	-1	0	-1	1	1	0	1	0	-1	-1	0	0	0	0	-1	0	-1	0	0	1
3059	30,03	0	0	0	0	0	0	0	0	0	1	-1	0	1	0	0	0	1	1	0	0	1	0	-1	0	0	1	0	0	1	0	-1	0	0	1	-1	-1	1	0	-1	1
4281	29,57	0	0	0	0	-1	0	1	-1	-1	-1	1	1	-1	0	-1	0	0	1	0	-1	1	0	1	1	1	0	-1	0	-1	0	0	-1	-1	-1	-1	0	1	1	-1	0
1341	28,50	1	0	1	-1	0	-1	0	0	-1	0	-1	0	-1	0	0	1	1	0	0	-1	0	0	0	0	-1	1	0	1	1	-1	0	1	1	0	-1	0	0	0	1	-1
1808	29,19	0	1	1	0	0	1	0	-1	-1	-1	0	0	0	0	0	-1	0	0	-1	0	1	1	-1	0	0	0	0	-1	-1	0	1	1	1	1	0	0	1	1	0	-1
2065	29,35	0	1	-1	0	0	-1	0	-1	0	-1	0	-1	0	1	-1	-1	0	-1	1	0	0	-1	0	1	-1	-1	0	0	0	-1	0	1	1	1	0	-1	1	0	0	0
8438	27,22	0	0	-1	-1	0	1	-1	0	0	0	0	0	-1	1	0	0	0	0	0	-1	-1	-1	-1	-1	0	-1	1	0	0	0	-1	0	0	0	1	0	0	0	0	0
8123	30,10	1	0	0	1	1	0	-1	0	-1	1	1	0	1	0	0	1	-1	0	0	0	-1	1	0	-1	-1	-1	1	0	0	0	0	-1	1	1	-1	0	1	1	0	0
5047	29,20	0	-1	0	0	-1	0	1	1	1	1	1	0	0	1	-1	0	-1	1	0	0	1	0	-1	-1	1	1	0	-1	0	-1	0	1	0	0	0	0	0	0	1	0
5803	31,06	0	1	0	-1	0	-1	1	1	1	1	-1	-1	-1	0	1	0	-1	-1	0	0	1	0	0	0	0	1	0	0	1	1	0	0	0	0	-1	0	1	0	0	0
2792	30,72	1	0	1	-1	-1	0	0	0	0	-1	1	0	0	0	-1	1	0	-1	0	0	1	1	0	1	1	0	0	-1	1	0	-1	1	0	0	0	1	0	-1	0	0
1095	30,48	1	0	-1	1	0	-1	0	0	1	-1	1	0	-1	-1	-1	0	-1	0	1	0	0	-1	0	0	0	-1	1	1	0	0	0	0	0	-1	1	0	-1	-1	-1	0
1500	28,59	-1	-1	-1	1	1	0	0	-1	0	-1	0	0	0	-1	-1	-1	0	0	0	1	0	1	1	0	1	0	0	1	1	-1	-1	1	-1	0	0	0	0	-1	1
2501	28,80	1	0	-1	-1	1	0	1	-1	0	-1	1	-1	-1	0	0	-1	0	1	-1	0	0	-1	1	1	0	-1	-1	0	0	-1	0	0	-1	-1	-1	0	0	0	1	1
524	30,67	1	1	-1	-1	-1	1	-1	0	0	-1	-1	0	0	0	0	0	0	0	-1	0	-1	1	0	0	0	1	-1	0	-1	0	0	0	0	-1	-1	0	-1	1	0	1
1396	30,15	0	0	1	1	0	0	0	0	0	0	0	1	1	-1	0	1	-1	1	1	0	0	1	0	1	0	0	-1	0	0	-1	0	0	0	0	0	-1	1	-1	1	0
1339	31,29	1	0	0	0	0	0	-1	1	-1	0	0	-1	1	1	-1	1	0	1	1	0	1	-1	-1	0	0	0	0	-1	-1	-1	0	-1	0	-1	0	1	1	1	1	1

Table A2.3. Genomic breeding values (GBV) for weaning weight estimated using least squares methods.

Animal ID	GBV
6287	0.83
4192	-0.46
3551	0.69
3658	1.30
3059	0.70
4281	0.22
1341	2.16
1808	0.43
2065	1.53
8438	0.44
8123	0.97
5047	-1.22
5803	1.25
2792	3.44
1095	0.46
1500	-3.25
2501	3.05
524	3.32
1396	-0.09
1339	2.59

Table A2.4. Genomic breeding values (GBV) for weaning weight is sheep estimated using best linear unbiased prediction procedures.

Animal ID	True breeding values	GBV (1-alpha)[1]	GBVs (2-alpha)[2]
6287	1.95	0.04	0.21
4192	1.08	-0.41	-0.34
3551	2.55	-0.18	-0.04
3658	1.52	0.01	0.56
3059	2.03	0.18	0.07
4281	1.57	0.37	0.12
1341	0.50	-0.27	0.68
1808	1.19	-0.02	0.51
2065	1.35	-0.19	0.52
8438	-0.78	-0.39	-0.10
8123	2.10	0.37	0.83
5047	1.20	0.04	-0.48
5803	3.06	0.13	0.60
2792	2.72	0.29	0.80
1095	2.48	-0.38	0.63
1500	0.59	-0.48	-1.43
2501	0.80	0.08	0.79
524	2.67	0.07	1.24
1396	2.15	-0.29	-0.05
1339	3.29	0.11	0.77

[1] Equal variance for all SNPs
[2] Known variances for SNP1 and SNP2

CHAPTER 3

Deterministic and stochastic simulation of a breeding programme for a nucleus flock based on individual selection

3.1. Abstract

The use of an adequate selection index in a breeding unit to identify the individuals with best genetics will improve the breeding objective traits. New Zealand has had very strong capacity of quantitative genetics, which has enabled the development of successful sheep breeding programmes. The study developed a stochastic model that simulated a sheep breeding flock selecting their breeding animals using best linear unbiased predictor methodology. The system had the breeding objective of improving parasite resistance by decreasing faecal eggs score; decreasing yearling weight to reduce maintenance feed costs in adult sheep and to increase 160 days lamb's carcass weight representing an increase of the system income. A selection index theory deterministic model was utilised to validate the stochastic model. Very similar genetic responses values were obtained for the analysed traits between simulation models (e.g. 0.113 kg and 0.099 ± 0.04 kg for carcass weight deterministic and stochastic, respectively) also appreciable on all the genetic trends of the traits; therefore the stochastic model has proven to be adequate to simulate the proposed system. The models developed in this study enable the construction base for new models to investigate different breeding options such as the inclusion of DNA information to improve production traits.

3.2. Introduction

An animal breeding programme can be defined as a series of organised steps designed to obtain the farm system's breeding goal (Harris et al. 1984). This can be achieved by the identification of the genetically fittest individuals of a breeding unit using an appropriate selection index, to improve the traits in the breeding objective. This enables the implementation of custom designed selection schemes for the breeding population. Subsequently, the implementation of an appropriate mating system through dissemination of the improved genetic material to commercial farmers is required. The programme is not complete until the whole process is evaluated (economically, technically and genetically) allowing identification of any weak points in the system; any necessary improvements should then be made if required.

Some concepts that require attention from the breeding programme definition are breeding goal, breeding objective and selection index.

The breeding goal is a high level description of the desired outcome as a result of the breeding management of the population, for example, increased profit per unit of feed. More recently, breeders have also considered aspects like ethics and other social aspects of human and animal welfare and wellbeing (Groen 2000).

The breeding or selection objective is a statement of all the traits that are to be modified via selection. The relative (economic) importance of these different traits are needed to apportion selection pressure between the traits to achieve the desired breeding goal (Van der Werf 2000).

The selection index is the sum of weighted selection criteria (the traits measured) to predict the animal's breeding value. Animals are selected based on their index (Cameron 1997).

According to Hazel and Lush (1942), selection for an index which gives proper weight to each trait is more effective than other ways of selection to achieve maximum genetic progress in the breeding objective. The selection index theory equations developed by Hazel (1943) defined the aggregate breeding value of an animal as the sum of its various genotypes weighted by the relative economic influence of each trait.

The New Zealand sheep industry breeding goal is to improve the genetic merit of the animals thereby improving farm profit. The success of implementing proper breeding programmes can be seen by reviewing the countries sheep statistics. From 1996 to 2011 the total number of New Zealand breeding ewes 2 years and older exposed to ram declined by 38.75% (Statistics New Zealand 2012), but the number of exported lambs (1998-2011) only declined by 19.6% (Beef + Lamb New Zealand 2012a). This means that more lambs are exported by means of less breeding females. In addition, lamb carcass weight increased from an average of 16.9 kg in 2006 to 18.2 kg on average in 2010 (Beef + Lamb New Zealand 2012b).

New Zealand has been fortunate in having a strong quantitative animal breeding capability. Initial genetic evaluations were based on Best Linear Prediction procedure (Blair & Garrick 2007), but as computing capacity allowed, Best Linear Unbiased

Prediction (BLUP) was introduced commercially to a subset of ram breeders in 1985 (Garrick 1991). BLUP maximises the use of phenotypic and genealogical data to obtain estimated breeding values (EBVs), thereby enabling an animal genetic ranking (Henderson 1963, 1975; Meuwissen et al. 2001). BLUP enables the subjects to be compared between different environments (e.g. flocks, years, etc.) providing animals are genetically linked.

The New Zealand sheep industry is comprised of several distinct breeding populations (typically breeds), and each of these populations will make genetic progress at different rates. There are four factors that control the rate of genetic gain towards the breeding goal:

1. Selection differential
2. Accuracy of selection
3. Genetic variance of the population
4. Generation interval

These four factors are interdependent, so rather than seeking to either maximise or minimise each factor, it is required to consider them jointly to optimise the rate of genetic change (Blair & Garrick 2007). In a real life situation, the best scenario to optimize the genetic gain of a production system is not always possible, especially because of difficulties related to farm management decisions. For example, the costs of obtaining some data at a specific moment could make the sampling unviable, affecting as a result the selection strategies of the animals.

Simulation techniques (or data modelling) are a very useful tool that can be described as: “A hypothetical or stylized representation that attempts to predict aspects of the behaviour of some system by the adoption of a language, using a special set of symbols, letters, numerals, etc., creating an approximate (mathematical) model of it, imitating the internal processes and not merely the results of the thing being simulated” (Reingruber & Gregory 1994). Stochastic or random simulation is when the values of the parameters needed to develop the model are generated randomly using predefined distributions, in order to minimise bias (Moore & McCabe 1990). On the other hand, deterministic simulation can be described as an algorithm in which the correct next step depends only on the current state, or in other words states that for a

given data or variable input the output obtained will always be the same (Pomar et al. 1991).

The objectives of this study were firstly to develop a stochastic simulation model for a multitrait sheep breeding scheme including four correlated traits under selection, to assess the averages and variances of the genetic responses. And secondly, to validate the stochastic simulation model against a deterministic model using selection index theory.

3.3. Materials and methods

Two simulation models were developed in this study, a deterministic and a stochastic model. Four traits were simulated based on the phenotypic and genetic parameters (Tables 3.1 and 3.2) presented by Bennett *et al.* (1991), Huisman & Brown (2008) and Huisman *et al*. (2008). The traits considered were live weight at 160 days (160W), faecal egg score at 160 days (FES), live weight at 1 year of age (YW) and lamb carcass weight at 160 days (CW); 160 days as selected age for CW, 160W and FES was to simulate an average age for the animals to be slaughtered.

Special attention has to taken when incorporating values from different studies, as the included variables when arranged as a (co)variance matrix could result in a non positive definite matrix causing some inconvenient in posterior analysis that may use this (co)variance matrix.

Both simulation models used the animal's own trait performance with one measurement per animal. CW was included in the study to include a trait that cannot be directly measured in the live animal. Using the correlation between CW and the measured traits makes it possible to generate EBVs for the unobserved trait (Cameron 1997).

3.3.1. Deterministic model

The deterministic model was developed using the selection index theory software SIP (Wagenaar et al. 1995). The selection index equations in matrix form are:

$$\mathbf{Pb} = \mathbf{Ga}$$

where:

P is a phenotypic variance-covariance matrix for selection criteria traits

b is a vector of selection index weights for phenotypic information

G is a genetic variance-covariance matrix between selection criterion and breeding objective traits

a is a vector of economic weights for the breeding objective traits.

3.3.1.1. Breeding objective

The breeding objective proposed for the present study included FES, YW and CW as the traits to be economically important for the production system. The breeding objective of decreasing FES (to improve parasite resistance), to decrease YW (to decrease maintenance feed costs) and to increase CW of the lambs (increase production income), is expected to improve farm profit.

The breeding objective also included an economic component to apportion the selection pressure between the traits in the objective. The economic component can be considered as:

- Economic value (EV), is the amount of profit change per unit of improvement in a trait keeping all other traits constant (Van Arendonk 1991).
- Relative economic weights (REW), can be stated to be values or weighting factors assigned with a relative importance depending on an expected proportion of change (Hazel 1943).

For the present study, the EV for each of the traits was estimated using the trait's genetic standard deviation (σ_G) and REWs, in order that an absolute value of the EV of a trait multiplied by the σ_G of the same trait represented the desired relative economic weight absolute value (AEW) (Table 3.1). The positive and negative values of the EV and REW represent the purpose of increasing (positive) or decrease (negative) the genetic value of the traits related with each EV.

Table 3.1. Genetic standard deviations (σ_G), economic values (EV), relative economic weights (REW) and relative economic weights absolute values (AEW) of the traits[1] included in the breeding objective.

Trait	σ_G	EV(¢)	REW	AEW
FES	0.78	-10.72	-30%	30%
YW (kg)	3.30	-1.70	-20%	20%
CW (kg)	0.83	16.93	50%	50%
Total				100%

[1]FES= faecal egg score defined as cubic root of number of eggs per gram [(eggs/g)$^{1/3}$], YW= yearling live weight and CW = carcass weight at 160 days of age.

3.3.1.2. Selection index

The selection index consisted of three traits 160W, FES and YW. 160W was the primary predictor of CW due to the high genetic correlation between these two traits.

The genetic and phenotypic parameters utilised as input data in the SIP software are shown in Table 3.2.

Table 3.2. Phenotypic standard deviations (σ_P) heritabilities (on the diagonal), phenotypic (above the diagonal) and genetic (below the diagonal) correlations of traits included in a selection index for sheep genetic improvement.

			Correlations			
Trait	Unit	σ_P	160W	CW	FES	YW
160W	kg	4.524	**0.54**	0.94	-0.01	0.65
CW	kg	1.766	0.92	**0.22**	-0.01	0.61
FES[1]	score	1.483	0.34	0.31	**0.28**	0.1
YW	kg	5.216	0.76	0.69	0.13	**0.40**

[1]Defined as Cube root of number of eggs per gram [(eggs/g)$^{1/3}$]

The selection index equations were:

$$\underset{\mathbf{P}}{\begin{bmatrix} 20.47 & -0.06711 & 15.34 \\ -0.06711 & 2.2 & 0.7737 \\ 15.34 & 0.7737 & 27.21 \end{bmatrix}} \underset{\mathbf{b}}{\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}} = \underset{\mathbf{G}}{\begin{bmatrix} 11.05 & 0.8872 & 8.336 & 2.534 \\ 0.8872 & 0.616 & 0.3366 & 0.2034 \\ 8.336 & 0.3366 & 10.88 & 1.911 \end{bmatrix}} \underset{\mathbf{a}}{\begin{bmatrix} 0 \\ -10.72 \\ -1.7 \\ 16.93 \end{bmatrix}}$$

3.3.1.3. Breeding scheme

The breeding scheme was based on a sheep breeding nucleus structure that consisted of 300 ewes, the size of the New Zealand average performance recorded flock (Garrick et al. 2000). Figure 3.1 shows the reproductive and replacement parameters for the nucleus flock.

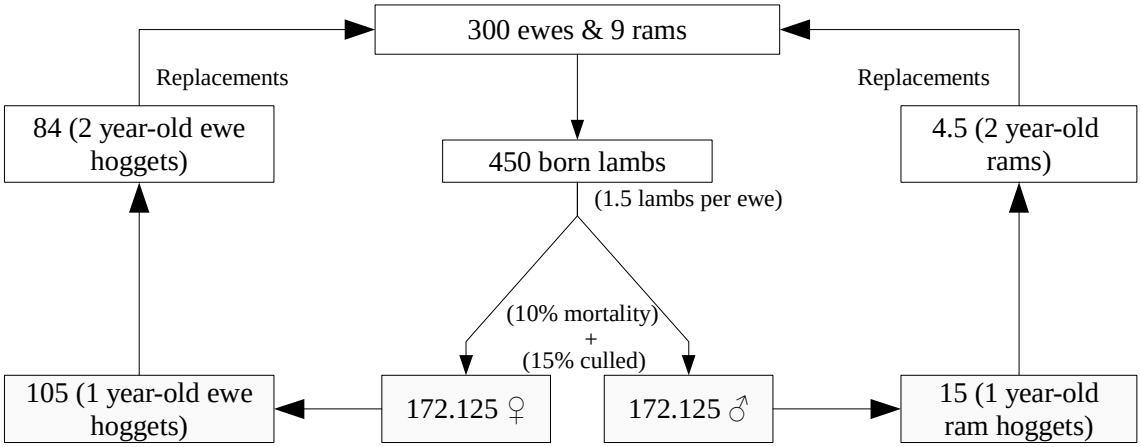

Figure 3.1. Simulated sheep nucleus breeding scheme structure.

In order to assess the annual genetic progress in the selection index traits, a simulated breeding flock was developed. The information of the virtual flock (Table 3.3) was simulated based on the structure presented in Figure 3.1. One year-old ewe and ram hoggets represent 35 and 5% respectively of the 300 breeding ewes. The age structure for males and females older than 1 year-old is presented in Table 3.4. Using selection pathways theory, the flock information was arranged into two categories, ewes to breed ewes, and rams to breed rams, based in the pathways model presented by Lopez-Villalobos & Garrick (2005). The generation interval was estimated using the same age structure utilised for the stochastic model (Table 3.4).

Table 3.3. Simulated ewes and rams pathways of selection.

Selection Pathway	Population Size	Number selected	Proportion selected	Generation Interval
Ewes	105	84	0.8	3.4
Rams	15	4.5	0.3	2.5

With the selection index information and the pathways model, annual genetic gain ($\Delta G / yr$) of the simulated population was estimated using Rendel & Robertson (1950) selection pathways theory formula:

$$\Delta G / yr = \frac{\left(\bar{i}_d \times r_d\right) + \left(\bar{i}_s \times r_s\right)}{L_d + L_s} \times \sigma_G$$

Where:

subscripts *d* and *s* represents dams and sires pathways respectively;

$\bar{i}$ represents the intensity of selection of the corresponding pathway,

r their accuracy of selection,

L their generation interval and

σ_G is the standard deviation of the true genetic value.

3.3.1.4. Correlated responses for individual traits

To asses individual responses of the breeding objective traits to selection based on the selection index, annual correlated responses (CR_j/yr) for each trait were estimated using Turner (1959) equation,

$$CR_j/yr = \frac{i_I \left(\mathbf{b'G_j} \Big/ \sqrt{\mathbf{b'Pb}} \right)}{L}$$

Where:

CR_j is the correlated response of trait *j*,

$\mathbf{G_j}$ is the $\mathbf{j}^{th}$ column of matrix **G** of the breeding objective,

b is the regression coefficient vector of the selection criteria,

P is the phenotypic (co)variance matrix

L is the generation interval of the selected population and

i_I is the selection intensity of the index.

As the breeding scheme involved two pathways with different intensities of selection, the correlated responses of each trait were estimated as the sum of both equations for each pathway.

$$CR_j/yr = \frac{i_{IE}\left(\mathbf{b'G_j}/\sqrt{\mathbf{b'Pb}}\right) + i_{IR}\left(\mathbf{b'G_j}/\sqrt{\mathbf{b'Pb}}\right)}{L_E + L_R}$$

being i_{IE} and L_E the intensity of selection and generation interval for the ewes pathway, and respectively i_{IR} and L_R being for the rams pathway.

3.3.1.5. Contribution of traits in the selection index

Cameron (1997), presented a way to asses how much contribution over the selection objective a trait included in the selection index has. This contribution is considered to be the proportional reduction to the accuracy of the selection index if the trait was excluded from the selection index. The equation to estimate this contribution is

$$\frac{r^{*}_{IH}}{r_{IH}} = \mathbf{1} - \sqrt{\frac{b_j^2}{\mathbf{b'Pb}P_{jj}^{-1}}}$$

where

r^{*}_{IH} is the accuracy of the selection index with the j trait omitted from the selection index,

P_{jj}^{-1} is the jth diagonal element of the inverse **P** matrix and

b_j^2 is the square of the jth element of the regression coefficients vector.

3.3.2. Stochastic simulation

The stochastic model was developed using Base SAS, SAS/IML and SAS 9.3 Macro language (SAS Institute Inc. 2011). The procedures used in developing the genetic simulation model were based on the studies of Analla et al. (1995), Cameron (1997), Falconer and Mackay (1996) and Dzama et al. (2001). The simulation represents a flock under a breeding programme, the time under evaluation is 20 years and the simulation process was replicated 100 times in order obtain a measure of the variation of response to selection. The genetic and phenotypic parameters utilised were the same as those used in the deterministic simulation. The effects of dominance and epistasis

were not modelled. The four traits considered in the deterministic simulation (LW160, FEC, YWT, CW) were modelled as:

$$y_{gij} = \mu + M_g + F_i + G_j + e_{gij}$$

where, y_{gij} is the phenotypic value of any of the traits being simulated

μ is the mean of the population for the trait

M_g is the effect of year g

F_i is the effect of flock i

G_j is the additive genetic effect of animal j, and

e_{gij} is an environmental effect.

The year effect was assumed to account for 5% of the phenotypic (co)variance matrix. The inclusion of the flock effect (F) was to account for a flock difference per replicate, and it was assumed to account for 10% of the phenotypic (co)variance matrix. The matrices containing the information of the traits genetic, environmental, flock and year effects, were created by the product of a randomly generated normal distribution matrix, having as many columns as traits simulated and as many rows as the number of randomly generated animals, with the lower triangular matrix **D** of the Cholesky decomposition (Nejati-Javaremi et al. 2007) of the (co)variance matrix of each effect.

In matrix notation the generation model for each effect can be expressed as:

$$\underset{\text{nx4}}{\left[\mathbf{V_1} \quad \mathbf{V_2} \quad \mathbf{V_3} \quad \mathbf{V_4}\right]} = \underset{\text{nx4}}{\left[\mathbf{\Phi_1} \quad \mathbf{\Phi_2} \quad \mathbf{\Phi_3} \quad \mathbf{\Phi_4}\right]} \underset{\text{4x4}}{\begin{bmatrix} D_{1,1} & 0 & 0 & 0 \\ D_{2,1} & D_{2,2} & 0 & 0 \\ D_{3,1} & D_{3,2} & D_{3,3} & 0 \\ D_{4,1} & D_{4,2} & D_{4,3} & D_{4,4} \end{bmatrix}}$$

Where:

$\mathbf{V_i}$ is a random multivariate normal deviate vector for the trait i, $\mathbf{\Phi_i}$ represents a random number vector taken from a normal distribution with mean 0 and variance 1 for the i^{th} trait and the matrix containing $D_{i,j}$ results from the transpose of Cholesky

decomposition of the (co)variance matrix of the effect **V** (environmental, flock or year effect) for the traits under selection.

The true genetic or breeding value ($\mathbf{TBV_i}$) of any animal *i* of the base population was generated using the method described in the previous paragraph. The $\mathbf{TBV_i}$ of any animal *i* in the following generations was estimated assuming that the value of an individual was equal to the average value of the parents genetic value ($\mathbf{TBV_s}$ and $\mathbf{TBV_d}$) plus a deviation due to Mendelian sampling (m_i) (Mrode 2006).

For the present multitrait simulation the model utilised to estimate the TBV of each lamb considered a Mendelian sampling in the absence of inbreeding (Kemp et al. 1986), then

$$\mathbf{TBV_i} = 0.5\left(\mathbf{TBV_s} + \mathbf{TBV_d}\right) + \left(\phi\,\mathbf{D}\right)0.5$$

where **D** is the lower triangular matrix of the Cholesky decomposition of the genetic (co)variance matrix, $\mathbf{TBV_i}$, $\mathbf{TBV_s}$ and $\mathbf{TBV_d}$ are row vectors of TBVs for animal *i*, the sire and the dam of animal *i* respectively, and ϕ is a random number vector taken from a normal distribution (N(0,1)).

The phenotypic values of all the simulated traits for each available animal were obtained as the sum of the TBVs, environmental effects, year and flock effects.

3.3.2.1. Flock structure

The base population simulated consisted of a flock with an average of 300 ewes with a standard deviation of 25 animals (SD=25). The number of rams corresponded to 3% of the ewes. The number of hoggets (between 1 and 2 year old) represented 35% and 5% of the ewe population females and males respectively. Table 3.4 shows the age structure of the flock for ewes and rams older than 2 years. Animals older than the ages shown in Table 3.4 were culled. Also 15% of the lambs were culled due to reproductive and phenotypic defects.

Table 3.4. Age structure for ewes and rams over two years old, of a stochastically simulated sheep breeding flock.

Age in years	2	3	4	5
Ewes (%)	28	26	24	22
Rams (%)	50	50		

The sex of lambs was assumed as 50% males and 50% females at birth and the lambing proportion used was 1.5 lambs per ewe. Table 3.5 shows the birth rank probability assumed for the lambing.

Table 3.5. Birth rank percentages for lambs born in a stochastically simulated sheep breeding flock.

N^{o} of lambs	1	2	3
Lambs (%)	60	30	10

For the death simulation process, 10% of the born lambs die before reaching the first year of life. Ewe's death percentage is 5% for the 1 year old and 2 % for older ewes. 1% of the rams die.

3.3.2.2. Genetic evaluation

Estimated breeding values were generated via multitrait analysis using the package AIREML (Johnson & Thompson 1995). This software used the average information matrix as second derivatives in a quasi-Newton procedure. The multitrait mixed model (Henderson & Quaas 1976) presented in matrix form was:

$$\begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \mathbf{Y}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{X}_3 \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Z}_3 \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \mathbf{e}_3 \end{bmatrix}$$

where:

$\mathbf{Y_i}$ is a vector that represents the observed phenotypic records of trait i,

$\mathbf{X_i}$ is a incidence matrix of fixed effects associated with trait i,

$\mathbf{b_i}$ is a vector of regression coefficients of the fixed effects,

$\mathbf{Z_i}$ is a relationship matrix of the recorded animals with random effects (breeding values),

$\mathbf{a_i}$ is the vector of breeding values for trait i,

$\mathbf{e_i}$ is a vector of the residual (error) of trait i,

The matrices of genetic (co)variance (G) and residual (co)variance (R) are represented as:

$$\mathbf{var}\begin{bmatrix}\mathbf{a_1}\\ \mathbf{a_2}\\ \mathbf{a_3}\end{bmatrix} = \begin{bmatrix}\mathbf{A}\sigma_{a_{11}} & \mathbf{A}\sigma_{a_{12}} & \mathbf{A}\sigma_{a_{13}}\\ \mathbf{A}\sigma_{a_{12}} & \mathbf{A}\sigma_{a_{22}} & \mathbf{A}\sigma_{a_{23}}\\ \mathbf{A}\sigma_{a_{13}} & \mathbf{A}\sigma_{a_{23}} & \mathbf{A}\sigma_{a_{33}}\end{bmatrix} = \mathbf{G} \text{ and}$$

$$\mathbf{var}\begin{bmatrix}\mathbf{e_1}\\ \mathbf{e_2}\\ \mathbf{e_3}\end{bmatrix} = \begin{bmatrix}\mathbf{I}\sigma_{e_{11}} & \mathbf{I}\sigma_{e_{12}} & \mathbf{I}\sigma_{e_{13}}\\ \mathbf{I}\sigma_{e_{12}} & \mathbf{I}\sigma_{e_{22}} & \mathbf{I}\sigma_{e_{23}}\\ \mathbf{I}\sigma_{e_{13}} & \mathbf{I}\sigma_{e_{23}} & \mathbf{I}\sigma_{e_{33}}\end{bmatrix} = \mathbf{R}$$

Where:

$\mathbf{I}_i$ is an identity matrix of the order nxn where n is the number of measured animals,

A is the numerator relationship matrix between animals and,

$\sigma_{a_{ij}}$ with $\sigma_{e_{ij}}$ are their corresponding covariances between traits *i* and *j*.

To enable comparison between the deterministic and stochastic simulations the relationship matrix was assumed to be an identity matrix, i.e., the EBVs for each animal were obtained based on its own performance but accounting for the genetic correlations between the traits.

3.3.2.2.1. Selection Index

A total merit index (IDX) (Hazel 1943) was created using the EBVs each weighted by a regression coefficient (*b* value) obtained from the deterministic simulation using the

SIP software. The purpose of generating this IDX was to enable the ranking of each animal based on the relative importance of the traits on which the evaluation was based.

$$IDX_i = (b_1 \times EBV_{i160W}) + (b_2 \times EBV_{i\,FES}) + (b_3 \times EBV_{i\,YW})$$

3.3.2.3. Data generation structure

The modelling process was built in subroutines modules (M1 to M9) using SAS Macro language. Each subroutine generated specific information that contributed to the creation of a 20 year sheep flock database. The model was replicated 100 times.

- M1 developed a base population using the age structure presented in the assumptions. The phenotypic record for each animal in the base generation was obtained as the sum of the trait's true genetic value, environmental value, flock effect and a year effect.
- M2 generated EBV's for each animal for the traits of interest (without considering genealogical information). From this point the replication process started, the purpose being to simulate a 20 year period of reproduction, death, culling and selection routines.
- M3 implemented the reproduction process for which a random mating within flock was done.
- M4 simulated the losses from death, in which a percentage of females and males based on their age were randomly categorised as dead.
- M5 assigned the year effect, to the phenotypic value of each animal.
- M6 obtained the phenotypic information for all live animals about to be 1 year old.
- M7 was the first subroutine to cull the animals (males and females) based on two aspects; the age structure presented in the assumptions and their IDX (selection of the ranked animals). This subroutine also culled 15% of the lambs younger than one year, because of reproductive and phenotypic defects. The aim of this subroutine was to preserve the age structure of the flock considering all the available animals (alive and not culled) per age category. If there were more sheep than needed, the animals that had lower IDX per age were culled.
- M8 ensured that all the live animals became one year older.

- M9 was the second culling subroutine, which culled all animals (males and females) that exceeded the maximum age of the age structure tables.

In the final step subroutine, M2 was run again in order to genetically evaluate the animals that were 1 year old. A flowchart of the subroutines is shown in Figure 3.2.

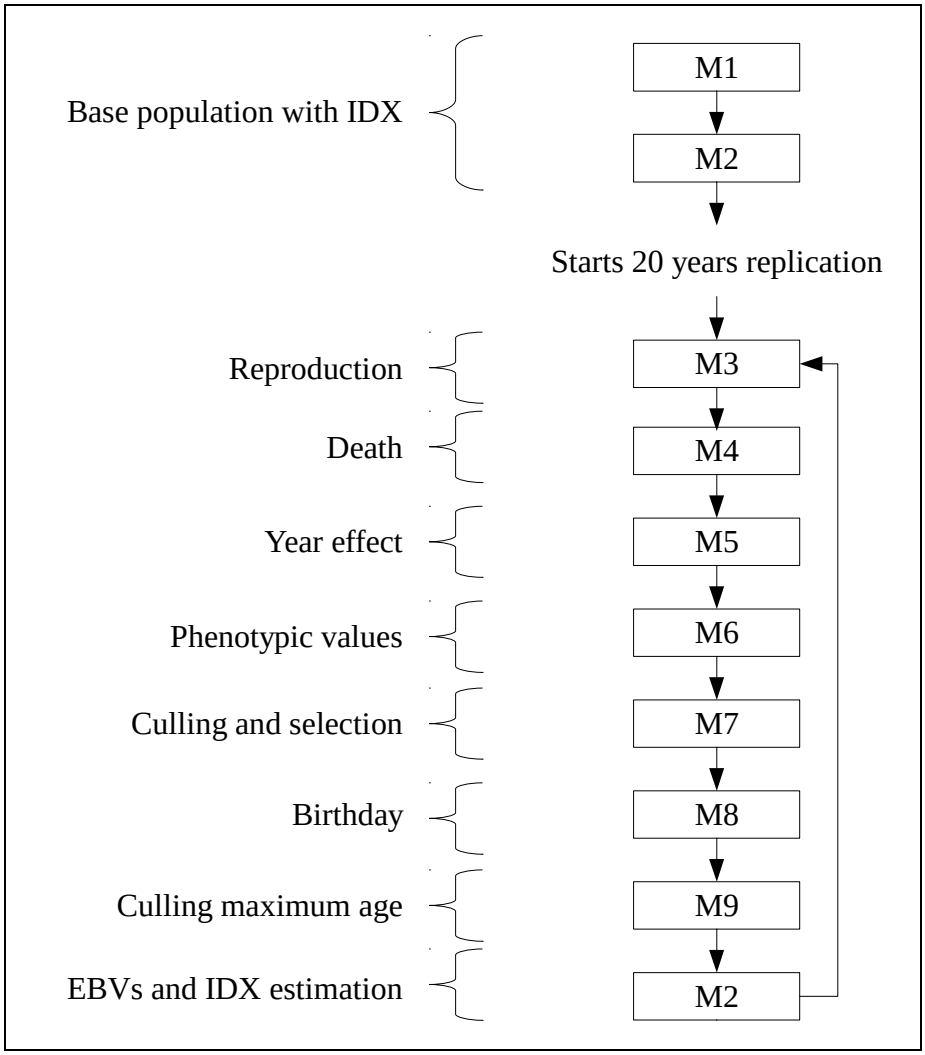

Figure 3.2. Subroutine structure of a stochastically simulated 20 years sheep breeding programme, selecting the breeding stock with a multitrait total merit index (IDX), built using estimated breeding values (EBVs).

3.4. Results

3.4.1. Deterministic simulation

Relative economic weights (Table 3.1), and genetic and phenotypic parameters (Table 3.2) were used to estimate the regression coefficients (b-values) to weight the selection criteria traits in the selection index. The index regression coefficients, the relative contribution of the trait in the selection criteria (%), and the correlated responses of the traits are shown in Table 3.6.

The economic value of response was ¢4.97, the index genetic standard deviation was 11.29 and the accuracy of predicted genetic merit was 0.44. The annual genetic gain obtained, considering the two pathways presented in Table 3.3, was of ¢1.27 per year.

Table 3.6. Regression coefficients (b-values) contribution of traits in the selection index and correlated responses for deterministically simulated traits[1].

Trait	b-value	Contribution of trait in selection index (%)	Correlated response[2]
160W (kg)	1.069	32.06	1.79 i
FES	-1.6	12.15	-0.0197 i
YW (kg)	-0.1809	1.006	1.29 i
CW (kg)			0.4104 i

[1]FES= faecal egg score defined as cubic root of number of eggs per gram [(eggs/g)$^{1/3}$], YW= yearling live weight and CW = carcass weight at 160 days.
[2] i represents the intensity of selection.

Genetic responses to selection for the four evaluated traits are presented in Figure 3.3. The annual genetic gains are shown in Table 3.7.

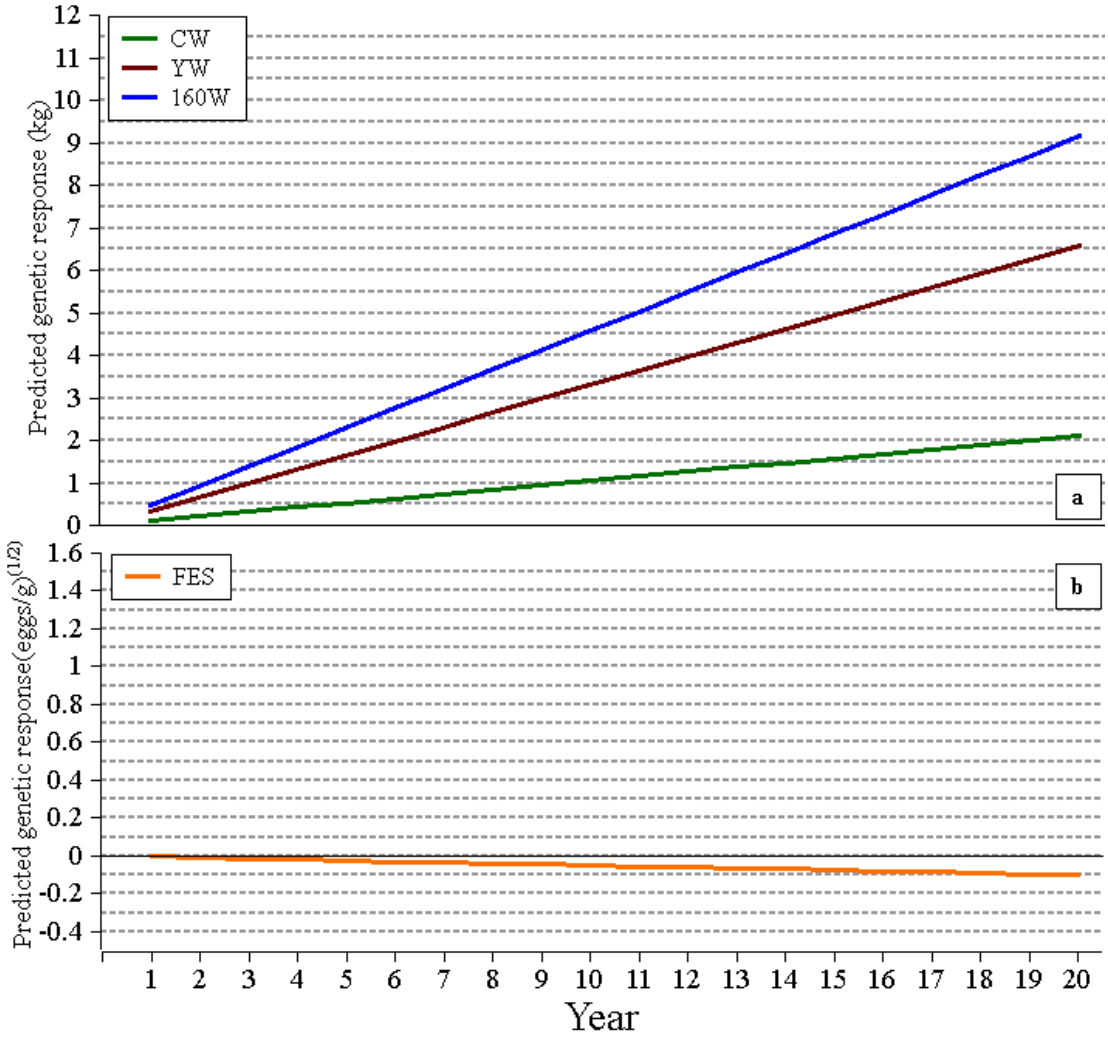

Figure 3.3. Annual predicted genetic responses for (a) live weight at 160 days (160W), yearling weight (YW) and carcass weight (CW) and (b) faecal egg score (FES), for a deterministically simulated sheep breeding programme.

3.4.2. Stochastic simulation

Figure 3.4 shows the mean true breeding values (TBV) for trait per year and per replicate, and the trend lines as an average TBVs for all the replicates. The annual average genetic gain for each trait (correlated responses) measured were calculated as, the slope of the regression of the mean TBV per year for the twenty simulated years for the 100 simulated replicates (Table 3.7).

Table 3.7. Animal genetic responses for deterministic and stochastic (mean ($\bar{x}_{\Delta G}$) and standard error (se)) simulations.

Trait[1]	Deterministic	Stochastic $\bar{x}_{\Delta G}$ ± se
160W (kg)	0.458	0.486 ± 0.007
FES	-0.005	0.022 ± 0.001
YW (kg)	0.330	0.358 ± 0.006
CW (kg)	0.105	0.106 ± 0.001

[1]FES= faecal egg score defined as cubic root of number of eggs per gram [$(eggs/g)^{1/3}$], YW= yearling live weight and CW = carcass weight at 160 days.

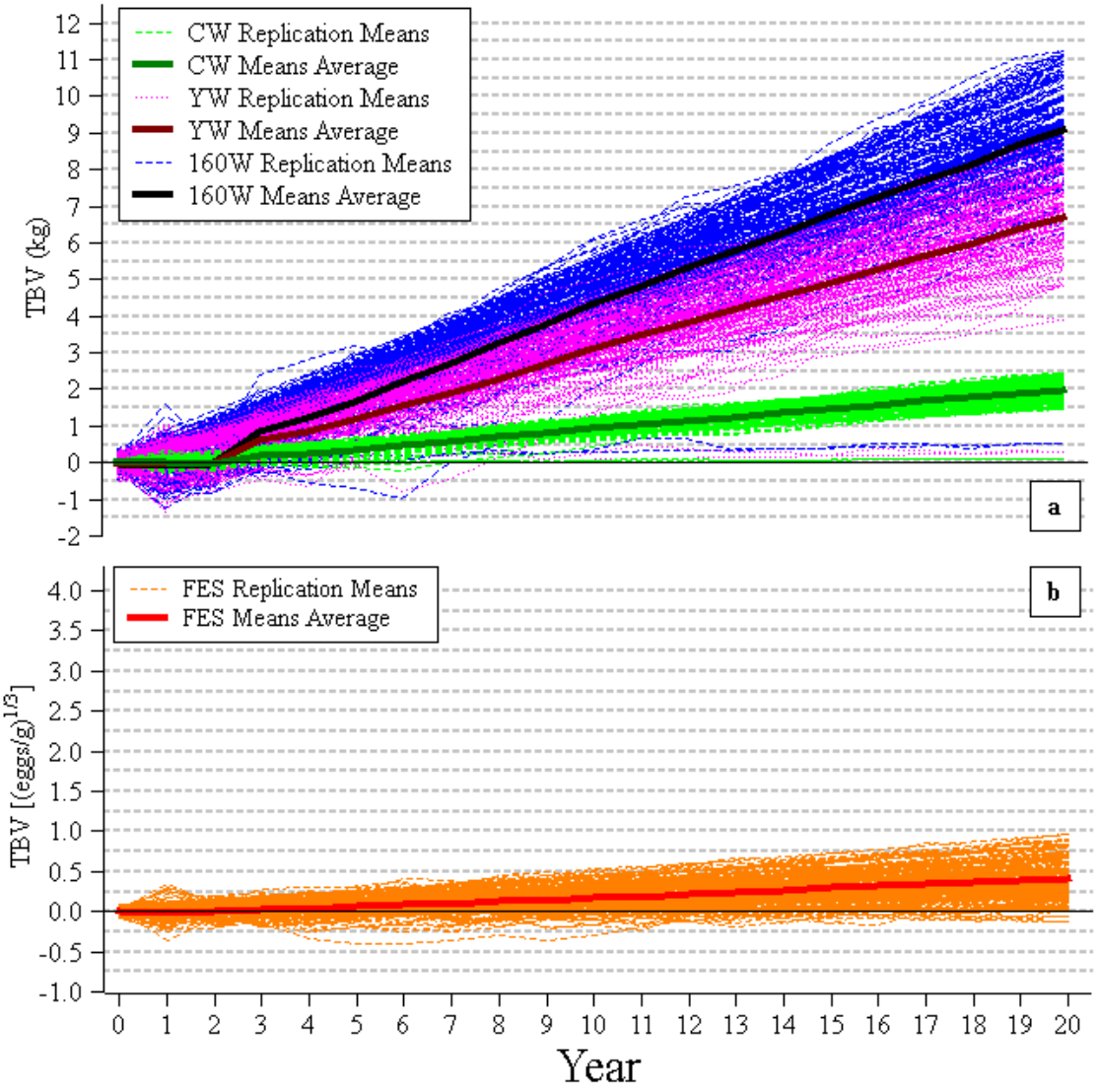

Figure 3.4. True breeding value (TBV) trends of individual replicates (dotted lines) and replicates averages (continuous lines) for (a) live weight at 160 days (160W), yearling weight (YW) and carcass weight (CW) and (b) faecal egg score (FES), for a stochastically simulated sheep breeding programme.

3.5. Discussion

For any simulation model an accurate knowledge of the variables to be included as input data it is required, especially regarding the mean and variance of the traits involved. Based on the work done to set up the present research (Table 3.2) it is suggested that, to obtain adequate simulated phenotypic values, the ratio of the phenotypic standard deviation to the mean or coefficient of variation (CV) for each

trait should be less than 0.2, otherwise phenotypic values generated with a stochastic simulation could turn out to be negative which is highly improvable for real measured data.

The main purpose of this study was to develop a stochastic model that simulated the genetic behaviour of a breeding flock with animals selected based on individual records of three different traits using a multitrait selection index. One of the objective traits (CW) was not directly measured. Genetic trends from a deterministic model were used to validate the results of the stochastic model, which has proven to be adequate to simulate the breeding programme, because it obtains similar genetic results compared with the deterministic simulation (Table 3.7). Important is to highlight that the shape (straight lines) of the genetic response tendencies of the deterministic simulation (Figure 3.3) represent perfectly what someone can expect from a deterministic model which accounts for the same outcome per year (Pomar et al. 1991). Comparing the deterministic genetic trends and average genetic trends of the stochastic model, all simulated traits behaved in a similar way. This can be explained because the values of the animal genetic responses (input data for Figures 3.3 and 3.4), differ very little between them (Table 3.7) especially if the standard error values are considered.

The variation of the stochastic simulation can be appreciated in Figure 3.4 showing a high variability of 160W and YW, and for FES and CW the represented deviations from the mean shows to be in a lower amount, therefore the input data used for the stochastic model (SE in Table 3.7) its very well represented, this fluctuation of values is one of the main points of comparison between a stochastic and a deterministic model.

For the stochastic results (Figure 3.4) a wide range genetic means within year were obtained. During the first three years (0-2) no genetic gain occurred. This was an adjustment period for the flock being selected. The base population had not been under selection for the proposed breeding objective, and therefore base animals had average TBVs of zero. An increase in average genetic merit occurs after culling those animals in the base population and the inclusion of genetically selected animals into the programme.

YW becomes an important trait because bigger animals require more nutritional input (maintenance cost) than a smaller ones (Nicol & Brookes 2007), therefore if the system management is not changed (by adjusting the farm feed input), having bigger animals will lead to a reduction in the farm stocking rate and conjointly less lambs born per hectare ergo less income. In the other hand Figures 3.3-a and 3.4-a showed an increase in CW genetic values, having heavier carcasses weights per lamb increases the income per unit sold.

One of the targets proposed by the breeding objective was to decrease or lower YW (to limit maintenance cost of the flock); this was done by assigning it a negative relative economic weight. However Figures 3.3-a and 3.4-a both show significant genetic trends for this trait. This previous situation can be explained by two aspects:

- As first point it is needed to be highlighted the fact that CW and YW have a considerable high genetic correlation (near to 0.7), hence as CW received a large positive economic value, the response of YW will tend to be in the same direction of CW, therefore it will be very hard to control the behaviour of the trait.
- The second and maybe more important than the first point is that the economic weights used were not the best to maintain or reduce the yearling weight of the flock. The economic value assigned to YW has proven to be modestly negative and in order to achieve lower or negative genetic gains; a lower economic value is required by YW. Therefore, new input data (economic weights) are needed to be simulated and evaluated in order to reach a better outcome.

FES impacts the production system by affecting the appropriate weight gain of the animals, increasing the drenching costs and also in extreme cases could lead to death loses (Vagenas et al. 2007). Figures 3.3-b and 3.4-b showed a very low genetic change (close to 0) for FES tendencies, even considering that the genetic correlation between FES and CW had a positive value, hence the negative restriction weight utilized (estimated on the trait's genetic standard deviation) showed to be effective to achieve the objective of limiting the traits effect.

Even though the stochastic model of this study has proven to be adequate to simulate the breeding programme, bio-economic farm models are needed (Van Arendonk 1991; Wolfová et al. 2009) to analyse the economical implication of having heavier CW and

YW with the corresponding maintenance of FES (Figures 3.3-b and 3.4-b), and as a result adjust the economic information of the present breeding objective to enhance the simulated breeding programme for the proposed breeding goal (Harris et al. 1984).

The present study has developed two models (one deterministic and one stochastic) simulating a breeding programme for a sheep breeding flock, providing the groundwork to develop simulations that could explore several other scenarios such as, the possible advantages of including genomic information into a breeding programme to improve production characteristics (Amer 2011).

As shown by the results obtained from this study; compared to a deterministic simulation a stochastic model allows a more flexible way of data control, permitting for example changing age structures or the inclusion of more variables in an easiest way. But as stated by Dekkers (2007), one of the advantages of a deterministic model over a stochastic one is that the deterministic model is less demanding regarding computer processing time (once all the variables to be included are modelled), therefore a fast overview of different scenarios including genomic information can be evaluated to optimize selection strategies.

CHAPTER 4

Deterministic simulation analysis using selection index theory for sheep genomic selection

4.1. Abstract

For animal breeding, molecular technologies could provide information with reliable accuracy of evaluation at younger ages. There is also the opportunity to enhance the development of breeding schemes by using genetic information associated with traits that are either not measurable at the time of the animal selection, or are sex-limited traits. Based on deterministic selection index models, the present study estimated the accuracy of selection and predicted the genetic gain using different selection indices that included either phenotypic and genomic information or indices with only marker information. Two SNP genotypes were included that explained four different percentages (1, 10, 30 and 50%) of one simulated trait genetic variance. The breeding objective for the simulation included decreased faecal egg score (to improve parasite resistance); reduced yearling weight (to decrease adult sheep feed maintenance costs) and to increased 160 day lamb carcass weight (to increase the farm's income). Results from the different simulated scenarios showed that an increase of genetic gain and accuracy of prediction was seen when genomic information was included together with phenotypic information. Compared to a selection index using only phenotypic information, a selection index that uses genomic information without phenotypic information can result in lower selection accuracies and genetic gains. The results of this study suggest that the use of genomic information in a selection scheme could increase genetic gains with lower accuracies of selection, only if a reduction of the generation interval was also achieved.

4.2. Introduction

The inclusion of genomic information into breeding schemes to improve animal production, relies on the expectation that information at the DNA level will lead to faster genetic gain compared to that which could be achieved using only phenotypic information (Meuwissen et al. 2001). For that reason, accurate appraisal about the implications of introducing genomic data into a breeding scheme is needed before its development. Meuwissen et al. (2001) demonstrated the methodology for genomic selection, a term that according to Meuwissen (2007) was first introduced in 1998 by Haley & Visscher (1998). Genomic selection uses information from thousands of single nucleotide polymorphism (SNP) genotypes combined as “haplotypes”, to allow

the estimation of genomic breeding values for animals that may have no phenotypic records of their own (Meuwissen et al. 2001).

Hazel (1943) defined the aggregate genetic value of an animal as the sum of its various genotypes (assuming a distinct genotype for each economic trait) weighted by the relative economic influence of that trait. Hazel and Lush (1942) stated that a multi-trait selection using a selection index is more effective than other types of selection, because it achieves the maximum genetic improvement per unit of time. A deterministic simulation model was presented by Lande & Thompson (1990) which included molecular genetic information using selection index theory. This model was further extended by Dekkers (2007) who presented equations to include marker information as a correlated trait (with heritability equal to 1). It was assumed that the marker information (combined genetic effects of those markers) showed multivariate normality to develop the model, which according to Lande & Thompson (1990) does not occur when genotypes representing only a small number of genes are used because a large proportion of the trait variance is associated with the gene or the marker. The methodology developed by Dekkers (2007) does not account for changes in the genetic variance due to changes in gene (allele) frequencies, the so-called Bulmer effect (Bulmer 1971). Several studies have used Dekker's (2007) methodology to simulate the impact of including genomic selection into breeding schemes. Janssen-Tapken et al. (2010) compared different selection strategies in beef cattle using genetic markers for diseases traits. Togashi & Lin (2010) analysed different selection methods for genetic improvement of net merit for two traits with the inclusion of marker information. Pryce & Goddard (2010) presented a deterministic model to simulate genomic selection in dairy cattle using the four pathways of selection of Rendel & Robertson (1950) while accounting for the rate of inbreeding per generation. In sheep production, Sise & Amer (2009) presented a deterministic approach using selection index theory to predict the response to genomic selection in dual purpose sheep flocks. They showed that when SNP markers were included as part of the selection index, the increase in annual genetic response was due to the use of young rams.

One of the contributions that molecular technologies can provide for genetic improvement might be the inclusion of measurements with reliable accuracy at younger ages (Garrick & Snell 2005). This could be as early as the embryonic stage

(Georges & Massey 1991) or even at the gamete level (Haley & Visscher 1998). Other situations in which these techniques could enhance the breeding schemes may be when traits are not available at the time of selection (Haley & Visscher 1998; Meuwissen 2003), or when traits are sex-limited (Haley & Visscher 1998).

In the New Zealand sheep industry ewes can give birth to their first lamb when they are one-year-old (Kenyon et al. 2008). This creates an opportunity to reduce the generation interval if one-year old ewes can be used as mothers of the following generation. However, it has to be considered that one-year-old ewes and ram hoggets have a lower reproductive performance compared with older animals in the flock (Kenyon et al. 2007).

The objectives of this study were:

- For one breeding scheme, design a deterministic selection index model to include genomic information. Based on this deterministic model, estimate the accuracy of selection and the genetic gain for different selection indices that include phenotypic and genomic information conjointly and also indices with only marker information.
- By using different breeding schemes, evaluate the impact on the accuracy of selection and the genetic gain in selection indices having phenotypic and marker information and also indices with only marker information.

4.3. Materials and methods

The present work is a theoretical study, developed to illustrate the methodology used to incorporate genomic information into a selection index. It was not a purpose of the study to simulate the New Zealand sheep industry in detail or to present an optimised breeding scheme design.

The traits included the breeding objective were carcass weight at 160 days (CW), yearling live weight (YW) and faecal egg score at 160 days (FES). The intent was to:

- Increase CW, because it is an income trait only measurable after slaughter (Meuwissen 2003),

- decrease YW, which is a trait measured later in life and affects maintenance costs (Nicol & Brookes 2007) and,
- decrease FES, because it is a health and welfare trait that affects the profit of sheep industry (Vagenas et al. 2007).

4.3.1. Genomic selection index conceptualisation

Following Dekkers (2007), the phenotype *(P)* of a trait was represented as:

$$P = G + E$$

with *G* partitioned as

$$G = Q + R,$$

where,

G represents the additive genetic value of the trait under study,
Q is the genetic effects correlated with SNP genotypes,
R is the residual genetic effects independent of the markers,
E is the random environmental effects.

The linear form of the selection index methodology suggested by Lande & Thompson (1990) can be represented as:

$$I_i = b_Q Q_i + b_P P_i$$

which represents the selection index criteria shaped by the SNP genotype information for the individual i (Q_i), the individual's phenotypic information (P_i), and the regression coefficients for SNP genetic and phenotypic information b_Q and b_P respectively.

The selection index equations in matrix form are:

$$\mathbf{Pb} = \mathbf{Ga}$$

where:

P is a variance-covariance matrix with traits in the selection index,

b is a vector of index weighting factors for SNP genotypes and phenotypic information,
G is a variance-covariance matrix associating traits of the selection index and the breeding objective,
a is a vector of economic weights for the breeding objective traits.

SNP genotypic information included in the selection index was simulated based on the models presented by Lande & Thompson (1990), Cameron (1997) and Dekkers (2007). Correlations between markers (SNP genotypes) and the traits were derived using the path coefficient theory diagram presented by Dekkers (2007), assuming the accuracies of $\hat{Q}$ predictor of Q for every SNP genotype $r_{\hat{Q}} = 1$, or equivalently $\hat{Q} = Q$. Therefore, the genetic correlation between any trait and a SNP genotype trait can be estimated as:

$$r_{G_j Q_i} = q_i r_{G_{ij}}$$

where:

q_i is the square root of the proportion of the genetic variance of the trait *i* associated with the SNP genotype $q^2 = \sigma^2_Q/\sigma^2_G$ and

$r_{G_{ij}}$ is the genetic correlation between the traits *i* and *j*.

For the estimation of the phenotypic correlation between any trait and the SNP genotype for trait i, the equation utilised was:

$$r_{P_j Q_i} = h_j q_i r_{G_{ij}}$$

where:

h_j is the square root of the heritability of the correlated trait *j*.

The heritability of the SNP genotype for trait *i* was assumed to be 1, meaning that there was no difference between genetic and phenotypic variances of the SNP genotype. Covariances were obtained using standard quantitative genetics theory for estimating covariances between correlated traits (Dekkers 2007; Falconer & Mackay 1996).

The accuracy of the selection index (r_{TI}), also known as the correlation between the selection criteria and the breeding objective, can be estimated in matrix notation as:

$$r_{TI} = \sqrt{(\mathbf{b'Pb})/(\mathbf{a'Ca})}$$

where:

C is a variance-covariance matrix of the traits included in the breeding objective.

This accuracy of selection was used for estimating the annual genetic gain of the simulated population, based on the pathways model presented by Lopez-Villalobos & Garrick (2005) using Rendel & Robertson (1950) selection pathways theory formula:

$$\Delta G / yr = \frac{\sum (\bar{i}_i \times r_i)}{\sum L_i} \times \sigma_G$$

where:

$\bar{i}_i$ represents the intensity of selection for the pathway *i*,

r_i the accuracy of selection of animals for path *i*,

L_i the generation interval for the pathway *i*, and

σ_G is the standard deviation of the true genetic value or in matrix notation $\sqrt{\mathbf{a'Ca}}$.

In order to calculate genetic responses in individual traits in the selection index, annual correlated responses (R_j/yr) for each trait were estimated using the Turner (1959) equation:

$$R_j/yr = \frac{i_E\left(\mathbf{b'G_j} \Big/ \sqrt{\mathbf{b'Pb}}\right) + i_R\left(\mathbf{b'G_j} \Big/ \sqrt{\mathbf{b'Pb}}\right)}{L_E + L_R}$$

where:

$\mathbf{G_j}$ is the jth column of matrix **G** of the breeding objective,

b is the regression coefficient vector of the selection criteria,

P is the phenotypic (co)variance matrix

L_E and L_R are the generation intervals of ewes and rams pathways, respectively and i_E and i_R are the selection intensity of ewes and rams pathways, respectively.

4.3.2. Selection scenarios

The present study considered scenarios, one focusing on each of the stated study objectives. All simulations in both scenarios were developed using the same breeding objective (*H*) represented as:

$$H = a_{FES} BV_{FES} + a_{CW} BV_{CW} + a_{YW} BV_{YW}$$

Where:

H represents the desired breeding objective,

a_i an economic value of trait *i* and

BV_i is the breeding value of trait *i*.

The absolute values of the breeding objective economic values (*a*) multiplied by the trait's genetic standard deviation (σ_G) corresponds to the desired relative economic weight absolute value (AEW), which were needed to obtain relative economic weights (REW) (Table 4.1). Considering the intention of decreased YW (decreased maintenance costs), lower FES (decreased health costs) and increased CW (increased income), positive or negative EV and REW were assigned to the traits.

Table 4.1. Genetic standard deviations (σ_G), economic values (*a*), relative economic weights (REW) and relative economic weights absolute values (AEW) of the traits[1] included in the breeding objective.

Trait	σ_G	*a*(¢)	REW	AEW
FES	0.78	-10.72	-30%	30
YW (kg)	3.30	-1.70	-20%	20
CW (kg)	0.83	16.93	50%	50
Total				100%

[1]FES= faecal egg score defined as cubic root of number of eggs per gram [(eggs/g)$^{1/3}$], YW= yearling live weight and CW = carcass weight at 160 days.

4.3.2.1. Scenario 1: Different selection indices including genomic information

The purpose of developing this scenario was to evaluate the effect of including genomic information into a selection index. Several selection indices were created to account for different proportions of genetic variance explained by the included markers. All indices in this scenario were simulated under the same breeding scheme which was defined as a set of organised steps that included reproductive, management and economic and productive decisions conducted to select the animals predicted to have the best genetic merit.

The phenotypic and genetic parameters for the quantitative traits required to obtain the selection index weighting factors for a sheep population (Table 4.2) were obtained from Bennett et al. (1991) and Huisman & Brown and Huisman et al. (2008). The traits included in the selection index were live weight at 160 days (160W), FES and YW, and the traits included in the breeding objective were CW, FES and YW.

Genetic and phenotypic (**G** and **P**) and residual (**P-G**) covariance matrices were tested to be positive definite. The residual covariance matrix was not positive definite, therefore a bending process was utilised to make it positive definite and then the **P** matrix was recalculated.

Table 4.2. Phenotypic standard deviation (σ_P), heritability (diagonal), phenotypic (above diagonal) and genetic (below diagonal) correlations of traits[1] considered in selection index and breeding objective for sheep genetic improvement.

			Correlations			
Trait	Unit	σ_P	160W	CW	FES	YW
160W	Kg	4.524	**0.54**	0.94	-0.01	0.65
CW	Kg	1.766	0.92	**0.22**	-0.01	0.61
FES	Score	1.483	0.34	0.31	**0.28**	0.1
YW	Kg	5.216	0.76	0.69	0.13	**0.40**

[1]160W= live weight at 160 days, CW = carcass weight at 160 days, FES= faecal egg score defined as cubic root of number of eggs per gram [(eggs/g)1/3] and YW= yearling live weight.

Table 4.3 describes all the simulated selection indices which can be summarised as follows:

- A phenotypic-based (PI) multi-trait selection index, containing three traits (160W, FES and YW) as selection index, with YW and FES also being included in *H*.
- Eight selection indices that adds one SNP genotype (M_i) to PI as selection index. The SNP genotypes included as part of *H*, the first one (M_{CW}) linked to CW and the second (M_{YW}) associated to YW. Each explained four different percentages of the genetic variance of the respective traits (1, 10, 30 and 50%).
- Sixteen selection indices that add two SNP genotypes ($M_{CW \text{ and }} M_{YW}$) to PI as selection index. Each SNP genotype was correlated to all traits in *H* and represented different genetic variance proportions of the traits.
- Sixteen selection index using only $M_{CW \text{ and }} M_{YW}$ as selection criteria. Each SNP genotype was correlated to all *H* traits and represented different genetic variance proportions for each trait. No phenotypic information was included in these selection indices.

The selection index equations in matrix form used for PI that added 2 correlated SNP genotypes with the *H* traits were:

$$\begin{bmatrix} \sigma^2_{P_{160W}} & \sigma_{P_{160W,FES}} & \sigma_{P_{160W,YW}} & \sigma_{G_{160W,M_{CW}}} & \sigma_{G_{160W,M_{YW}}} \\ \sigma_{P_{FES,160W}} & \sigma^2_{P_{FES}} & \sigma_{P_{FES,YW}} & \sigma_{G_{FES,M_{CW}}} & \sigma_{G_{FES,M_{YW}}} \\ \sigma_{P_{YW,160W}} & \sigma_{P_{YW,FES}} & \sigma^2_{P_{YW}} & \sigma_{G_{YW,M_{CW}}} & \sigma^2_{M_{YW}} \\ \sigma_{G_{M_{CW},160W}} & \sigma_{G_{M_{CW},FES}} & \sigma_{G_{M_{CW},YW}} & \sigma^2_{M_{CW}} & \sigma_{G_{M_{CW},M_{YW}}} \\ \sigma_{G_{M_{YW},160W}} & \sigma_{G_{M_{YW},FES}} & \sigma^2_{M_{YW}} & \sigma_{G_{M_{YW},M_{CW}}} & \sigma^2_{M_{YW}} \end{bmatrix} \begin{bmatrix} b_{160W} \\ b_{FES} \\ b_{YW} \\ b_{M_{CW}} \\ b_{M_{YW}} \end{bmatrix} =$$

$$\begin{bmatrix} \sigma_{G_{160W,FES}} & \sigma_{G_{160W,YW}} & \sigma_{G_{160W,CW}} \\ \sigma^2_{G_{FES}} & \sigma_{G_{FES,YW}} & \sigma_{G_{FES,CW}} \\ \sigma_{G_{YW,FES}} & \sigma^2_{G_{YW}} & \sigma_{G_{YW,CW}} \\ \sigma_{G_{M_{CW},FES}} & \sigma_{G_{M_{CW},YW}} & \sigma_{G_{M_{CW},CW}} \\ \sigma_{G_{M_{YW},FES}} & \sigma_{G_{M_{YW},YW}} & \sigma_{G_{M_{YW},CW}} \end{bmatrix} \begin{bmatrix} a_{FES} \\ a_{CW} \\ a_{YW} \end{bmatrix}$$

where:

$\sigma^2_{P_i}$, $\sigma_{P_{i,j}}$, $\sigma^2_{G_i}$ and $\sigma_{G_{i,j}}$ are the phenotypic and genetic variances and covariances of the traits i and j, respectively,

b_i and a_i are the regression coefficients and the economic values for the trait i, respectively.
The G matrix has an order of 5x3.

Two selection pathways were simulated (Figure 4.1 and Table 4.4). The first pathway (Ewes) considers 300 breeding ewes (Garrick et al. 2000) reproduced by natural mating, with 84 ewes being replaced every year and lambing ages ranging between 2 and 5 years (Table 4.5). The second pathway (Rams) consisted of 4.5 rams selected from 202.5 male lambs; once selected these animals were used for two mating seasons. All animals (rams and ewes) were first mated at 1.5 years old having their first offspring at 2 years old. The average number of lambs born per ewe was assumed at 1.5, with a mortality of 10% during the first year. The total number of rams in the flock was assumed to be 3% of the total number of ewes.

Table 4.3. Selection indices including SNP genotypes (M_{CW} and M_{YW}) explaining different percentages (1, 10, 30, 50%) of the correlated trait genetic variance.

Selection Index	Percentage of genetic variance	
	M_{CW}	M_{YW}
PI[1]		
PI+M_{CW}1	1	
PI+M_{CW}10	10	
PI+M_{CW}30	30	
PI+M_{CW}50	50	
PI+M_{YW}1		1
PI+M_{YW}10		10
PI+M_{YW}30		30
PI+M_{YW}50		50
PI+M_{CW}1-M_{YW}1	1	1
PI+M_{CW}1-M_{YW}10	1	10
PI+M_{CW}1-M_{YW}30	1	30
PI+M_{CW}1-M_{YW}50	1	50
PI+M_{CW}10-M_{YW}1	10	1
PI+M_{CW}10-M_{YW}10	10	10
PI+M_{CW}10-M_{YW}30	10	30
PI+M_{CW}10-M_{YW}50	10	50
PI+M_{CW}30-M_{YW}1	30	1
PI+M_{CW}30-M_{YW}10	30	10
PI+M_{CW}30-M_{YW}30	30	30
PI+M_{CW}30-M_{YW}50	30	50
PI+M_{CW}50-M_{YW}1	50	1
PI+M_{CW}50-M_{YW}10	50	10
PI+M_{CW}50-M_{YW}30	50	30
PI+M_{CW}50-M_{YW}50	50	50
M_{CW}1-M_{YW}1	1	1
M_{CW}1-M_{YW}10	1	10
M_{CW}1-M_{YW}30	1	30
M_{CW}1-M_{YW}50	1	50
M_{CW}10-M_{YW}1	10	1
M_{CW}10-M_{YW}10	10	10
M_{CW}10-M_{YW}30	10	30
M_{CW}10-M_{YW}50	10	50
M_{CW}30-M_{YW}1	30	1
M_{CW}30-M_{YW}10	30	10
M_{CW}30-M_{YW}30	30	30
M_{CW}30-M_{YW}50	30	50
M_{CW}50-M_{YW}1	50	1
M_{CW}50-M_{YW}10	50	10
M_{CW}50-M_{YW}30	50	30
M_{CW}50-M_{YW}50	50	50

[1]PI= phenotypic information selection index including 160W+FES+YW as traits, 160W= live weight at 160 days, CW = carcass weight at 160 days, FES= faecal egg score and YW= yearling live weight.

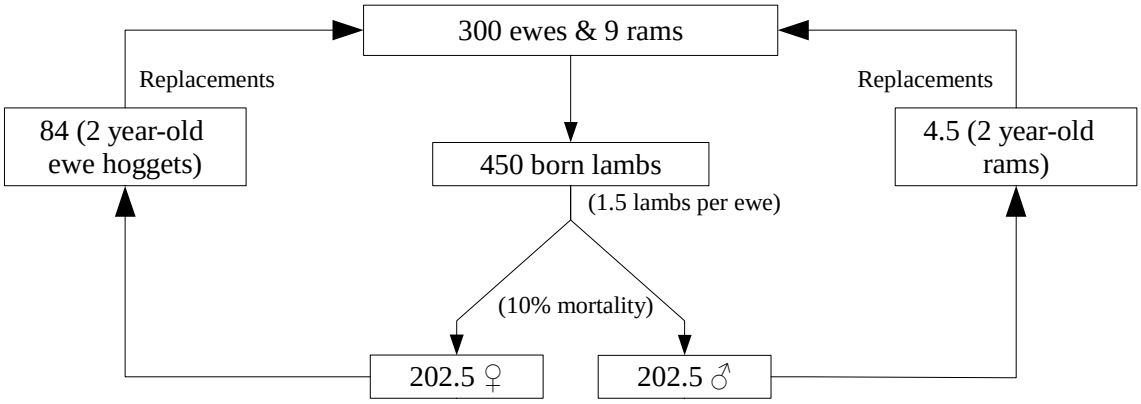

Figure 4.1. Two pathways of selection breeding scheme structure for the simulated sheep nucleus.

Table 4.4. Population parameters to simulate two pathways of selection (ewes and rams pathways), for a selection index with rams selected at 1 year old and ewes lambing from 2 to 5 years old.

Selection Pathway	Population Size	Number selected	Proportion selected	Generation Interval
Ewes	202.5	84	0.4148	3.4
Rams	202.5	4.5	0.0222	2.5

Table 4.5. Age structure for ewes over two years, old of a deterministically simulated sheep nucleus.

Age in years	2	3	4	5	**Total**
Ewes %	28%	26%	24%	22%	**100%**

4.3.2.2. Scenario 2: Using genomic information to change age at selection

The same SNP genotypes considered in scenario one (M_{CW} and M_{YW}) were included as components of the selection index. Each SNP genotype explained only 1% or 10% of the genetic variance of their correlated trait (Table 4.6). The inclusion of SNP genotypes as part of the selection criteria was depended on the age at sampling of the males. Therefore, three different periods of data recording or selection were simulated (Table 4.7):

- Males of 1 year old, using only phenotypic data were recorded.

- Males at 160 days old, in which 160W and FES were recorded and M_{YW} was genotyped, both were included as selection criteria.
- Males sampled at birth for M_{CW} and M_{YW} as traits included in the selection criteria.

The differences of the selection indices for scenario 2 compared with scenario 1 were, that in scenario 2 the SNP genotypes were not included together with the correlated trait in the selection index but instead they substituted for them (Table 4.6). This reduced the age of the animals at selection. The breeding scheme SS1 (Table 4.7) was the same used for scenario 1. Selection using genomic information was carried through the males of the population, to reduce the number of samples needed to be taken from the whole population. Table 4.8 shows the ewe flock age structure.

Table 4.6. Selection indices using phenotypic data (160W, FES and YW) and SNP genotypes (M_{CW} and M_{YW}) explaining different percentages of the genetic variance of the trait ($\%\sigma_G^2$).

Selection Index	Trait				
	160W	FES	YW	M_{CW} ($\%\sigma_G^2$)	M_{YW} ($\%\sigma_G^2$)
PI	✓	✓	✓		
$M_{YW}1$	✓	✓			1
$M_{cW}1M_{YW}1$				1	1
$M_{YW}10$	✓	✓			10
$M_{cW}10M_{YW}10$				10	10

[1]160W= live weight at 160 days, FES= faecal egg score, YW= yearling live weight, M_{CW}= SNP genotype for CW and M_{YW}= SNP genotype for YW.

The scenario considered a breeding nucleus of 300 lambing ewes (Garrick et al. 2000) simulated as base population. The ewes were bred by natural mating, and had two age structures (Tables 4.7 and 4.8) implemented as:

- Ages ranging from 2 to 5 years old at lambing (E2), and with a prolificacy of 150%.
- Ages ranging from 1 to 4 years old at lambing (E1), with 136% lambs per ewe due to lower prolificacy of the ewe lambs (Kenyon et al. 2008; Mulvaney et al. 2010; Notter 2000).

For males having their last trait recorded at one year of age a ram to ewe ratio of 1:33.3 was used. A ratio of 1:20 rams per ewe was used when animals were genotyped for selection at ages younger than one year, and were 1 year old when offspring were born (Kenyon et al. 2007).

Based on the previous information two selection pathways were simulated, one pathway for ewes and another pathway for rams. Differences of population sizes and number of animal selected (Table 4.4 and Appendices Table A4.1 to Table A4.5) were due to different prolificacy rates and ram to ewe ratios used.

Table 4.7. Simulated breeding schemes and selection indices for the genetic improvement of a 300-ewe flock size using two different ewes age structure considering the inclusion of SNP genotypes (M_{CW} and M_{YW}) explaining different percentage of the genetic variance of their correlated traits according to the males selection age.

Breeding scheme	Selection Index[1]	Males selection age	Number of years mated	Ewes age range at lambing	Ewe prolificacy (%)	Ram/ewe ratio
SS1	PI	1 year	2	2 to 5 years	150	1:33.3
SS2(a)	$M_{YW}1$	160 days	1		150	1:20
SS2(b)	$M_{YW}10$	160 days	1		150	1:20
SS2(c)	$M_{cW}1M_{YW}1$	At birth	1		150	1:20
SS2(d)	$M_{cW}10M_{YW}10$	At birth	1		150	1:20
SS3	PI	1 year	2	1 to 4 years	136	1:33.3
SS4(a)	$M_{YW}1$	160 days	1		136	1:20
SS4(b)	$M_{YW}10$	160 days	1		136	1:20
SS4(c)	$M_{cW}1M_{YW}1$	At birth	1		136	1:20
SS4(d)	$M_{cW}10M_{YW}10$	At birth	1		136	1:20

[1] Refer to table 4.6

Table 4.8. Age structures of the two nucleus flocks.

Ewes age structure	Ewes%			
	28%	26%	24%	22%
E2	2	3	4	5
E1	1	2	3	4

4.4. Results

4.4.1. Scenario 1 (Typical flock age structure)

The predicted rate of genetic gain in the breeding objective using a selection index based on only phenotypic information (PI) was 2.796 ¢/year with an accuracy of 0.440 (Figure 4.2, Table A4.6). The rate of genetic gain was also 2.796 ¢/year for PI+M_{YW} and it slightly increased to 2.810 ¢/year for PI+M_{CW} when they were incorporated in the selection index explaining 1% of the genetic variance. When only one of the SNP genotypes (M_{CW} or M_{YW}) was incorporated to PI, the genetic gains and accuracies obtained by PI+M_{CW} were higher than PI-M_{YW} when comparing the same proportion of genetic variance explained.

Using the same breeding scheme in all scenarios, accuracies of selection indices using only marker information (Figure 4.3, Table A4.7) were lower, compared to those obtained with selection indices that conjointly had phenotypic and marker information (Figure 4.2) when considering the same proportion of variance explained by M_{CW} and M_{YW}. The highest accuracy in Figure 4.2 was 0.569 for PI+M_{CW}50 with M_{YW} included at any proportion of YW genetic variance (0, 1, 10, 30 and 50%). These combinations generated a predicted genetic gain of 3.62 ¢/y. The highest accuracy in Figure 4.3 was 0.469 for M_{CW}50-M_{YW}50 giving a genetic gain of 2.982 ¢/y.

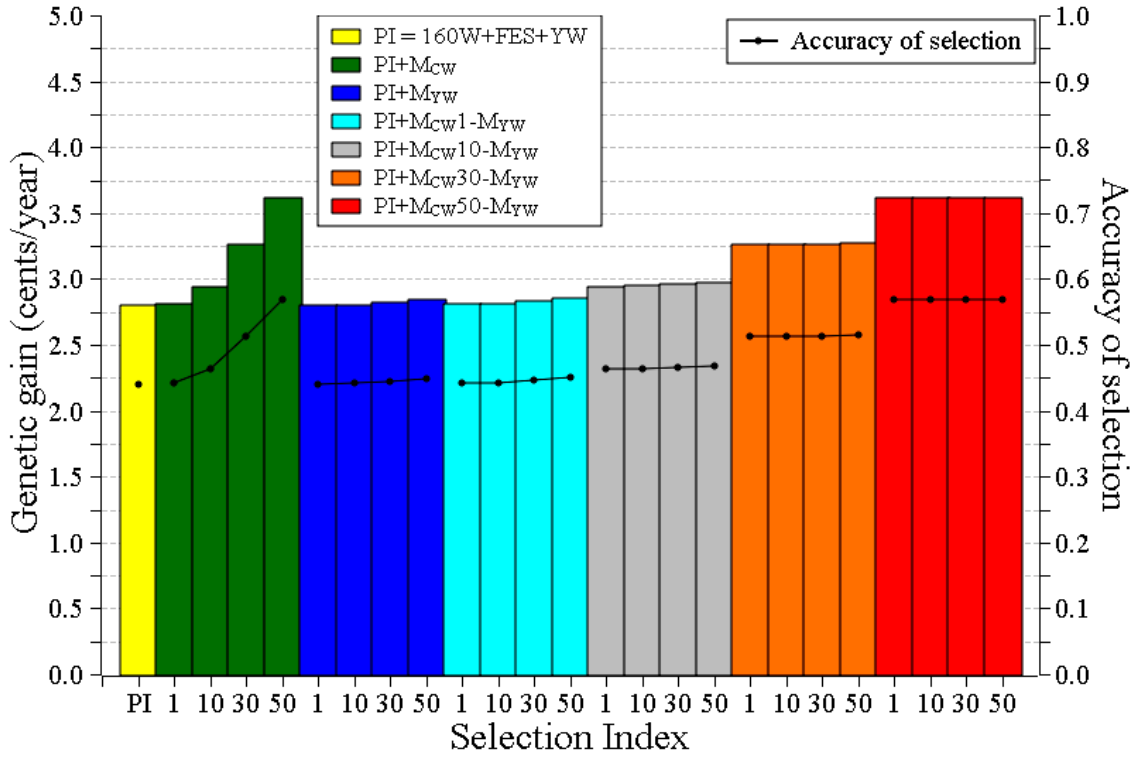

Figure 4.2. Genetic gains (shown in bars) and accuracies of selection (scattered lines) for different selection indices including phenotypic information (PI) and SNP genotypes (M_{CW} and M_{YW}) explaining different proportions (1, 10, 30 and 50%) of the total genetic variance of carcass weight (CW) and yearling live weight (YW) evaluated in a sheep nucleus.

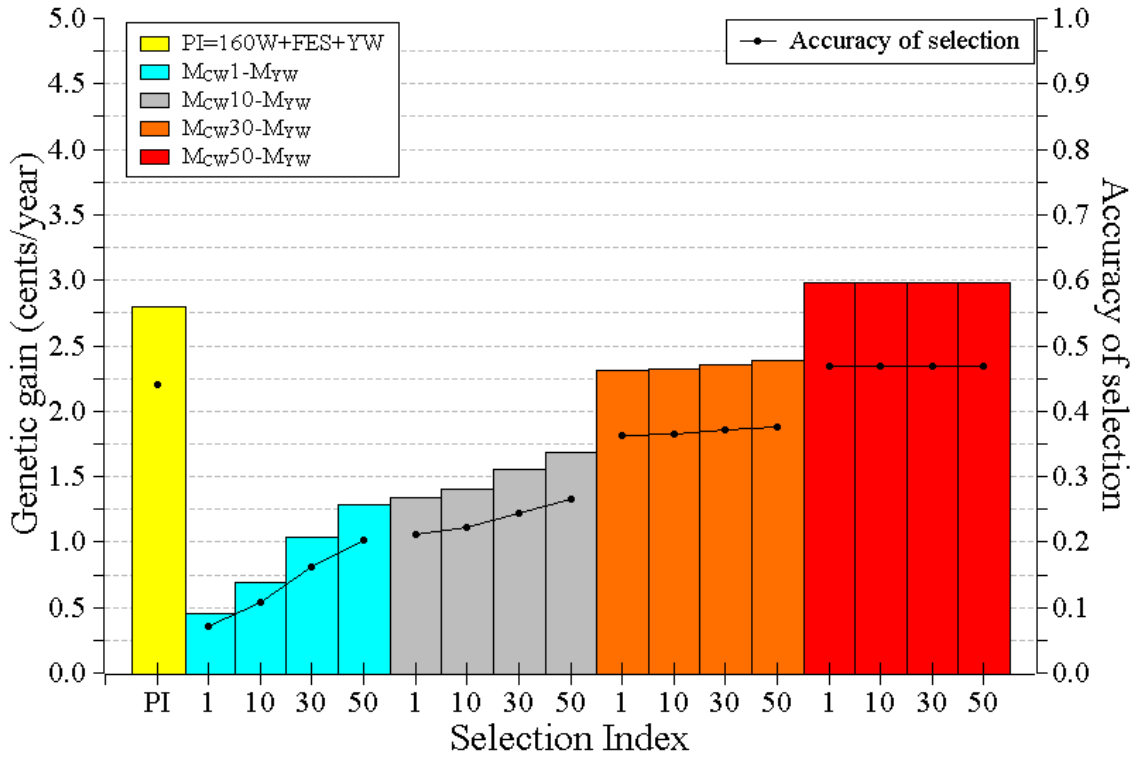

Figure 4.3. Genetic gains (shown in bars) and accuracies of selection (scattered lines) for different selection indices including only phenotypic information (PI), or having only SNP genotypes (M_{CW} and M_{YW}) explaining different proportions (1, 10, 30 and 50%) of the total genetic variance of carcass weight (CW) and yearling live weight (YW) evaluated in a sheep nucleus.

Annual predicted correlated genetic responses for 160W, FES, YW and CW are presented in Figure 4.4 (from Table A4.8). The highest gain for each of the included traits (all positive values), occurred when PI+M_{CW}-M_{YW} accounted for 50% of both CW σ_G^2 and YW σ_G^2. For FES the same gain was obtained using the PI+M_{CW}-M_{YW} index when M_{CW} accounted for 50% of CW σ_G^2 and M_{YW} accounted for 10 and 30% of YW σ_G^2 (Figure 4.4-b). For CW the highest genetic response was obtained using the PI+M_{CW}-M_{YW} index when M_{CW} accounted for 50% of CW σ_G^2 and M_{YW} accounting for 30% of YW σ_G^2 (Figure 4.4-d). The lowest gains for correlated traits was when PI was applied with no marker information included in the selection index. The same values were obtained for FES and CW in PI-M_{YW} when M_{YW} accounted for 1% of YW σ_G^2.

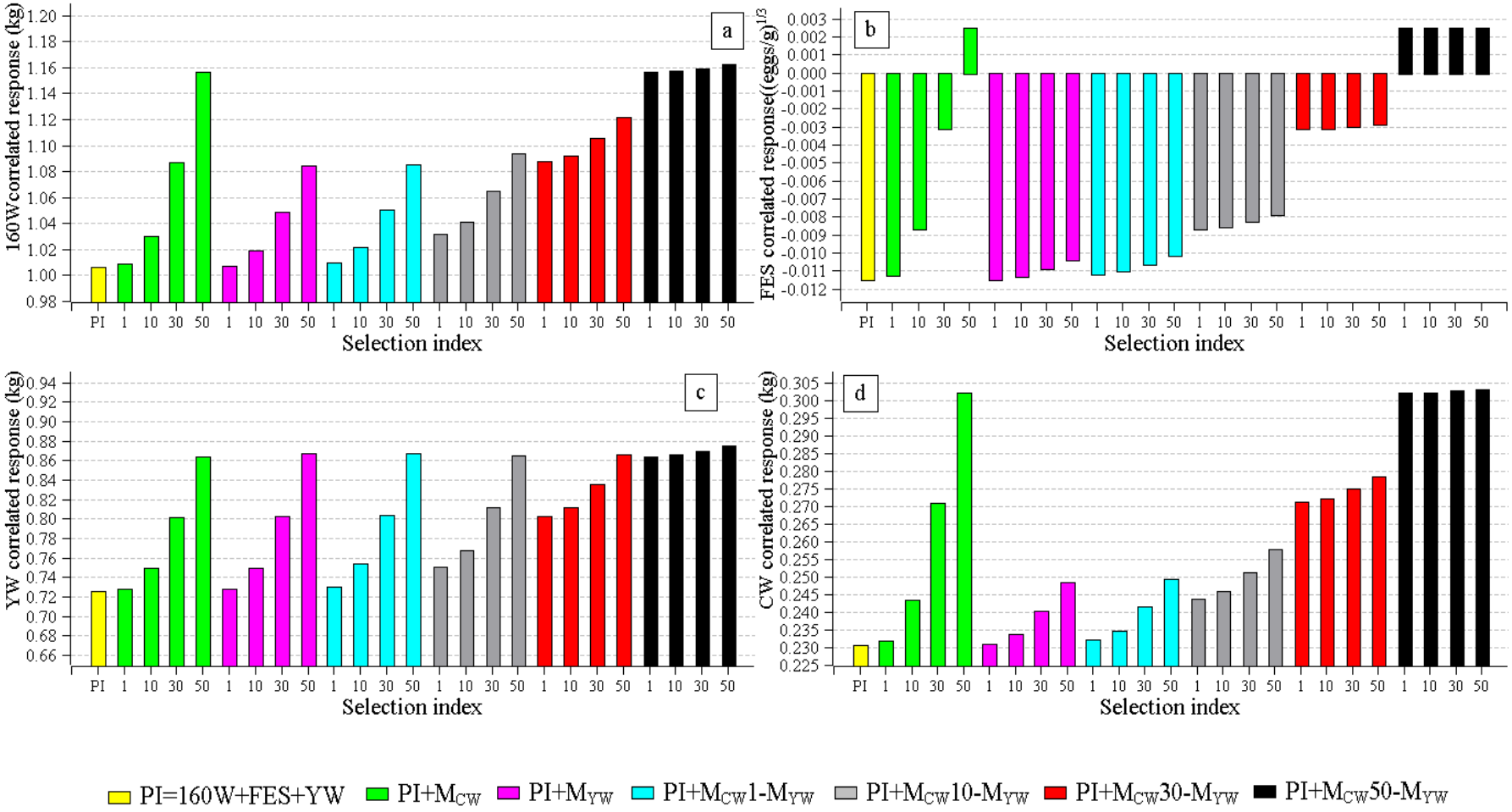

Figure 4.4. Genetic responses per year of live weight at 160 days (160W) (a), faecal egg score (FES) (b) , yearling live weight (YW) (c) and carcass weight (CW) (d) for different selection indices (colours) including phenotypic information (PI) and different proportions of SNP genotypes (M_{CW} and M_{YW}).

Figure 4.5 and Table A4.9 show correlated responses for 160W, FES, YW and CW when selecting on SNP genotypes only. The lowest gain values for each of the included traits were obtained when M_{CW}-M_{YW} accounts for 1% of both CW σ_G^2 and YW σ_G^2. The highest annual predicted genetic value for 160W and CW was achieved by M_{CW}50-M_{YW}50. The highest gain value for YW was achieved in M_{CW}-M_{YW} with M_{CW} accounting for 1% of CW σ_G^2 and M_{YW} accounting for 50% of YW σ_G^2. For FES (Figure 4.5b) the highest gain value (being undesirable as the lowest is better) was obtained in all the percentages of M_{YW} (1, 10, 30 and 50% of YW σ_G^2) when M_{CW} accounted for 50% of CW σ_G^2.

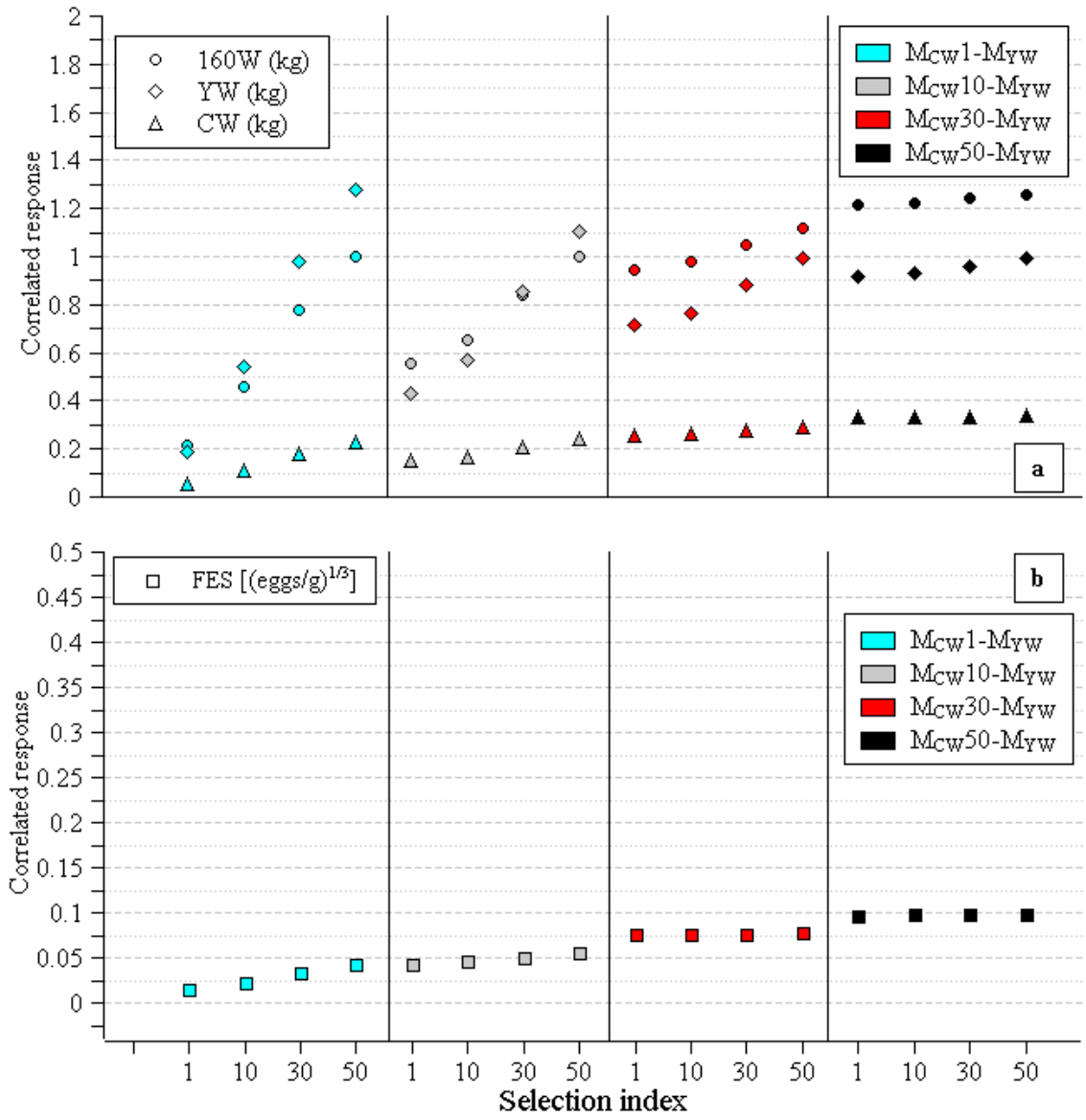

Figure 4.5. Genetic responses per year of (a) live weight at 160 days (160W), yearling weight (YW) and carcass weight (CW) and (b) faecal egg score (FES) (shapes), for different selection indices (colours) having only SNP genotypes (M_{CW} and M_{YW}).

4.4.2. Scenario 2 (Using ram and ewe lambs as parents)

The accuracies of selection between the same schemes letters (a,b,c,d) shown in Table 4.9 are the same because the sources of information are the same. The highest accuracy and genetic gain were obtained by using selection index $M_{YW}10$ in breeding schemes SS2 and SS4. When the female generation interval was reduced higher genetic gains were achieved compared to their respective selection indices when ewe first lambed at 2 years old (Table 4.9).

Table 4.9. Genetic gains in the breeding objective and accuracies of selection with ewes ages from 2 to 5 years old (SS1 and SS2) and from 1 to 4 years (SS3 and SS4), for different breeding schemes and selection indices with M_{CW} and M_{YW} explaining 1% and 10% of the genetic variance for CW and YW.

Breeding scheme	Selection index	Genetic gain (¢/year)	Accuracy of selection
SS1 [1]	PI	2.796	0.4399
SS2(a) [2]	$M_{YW}1$	3.198	0.4400
SS2(b) [2]	$M_{YW}10$	3.206	0.4412
SS2(c) [2]	$M_{cW}1$-$M_{YW}1$	0.520	0.0715
SS2(d) [2]	$M_{cW}10$-$M_{YW}10$	1.610	0.2216
SS3 [1]	PI	3.162	0.4399
SS4(a) [2]	$M_{YW}1$	3.806	0.4400
SS4(b) [2]	$M_{YW}10$	3.816	0.4412
SS4(c) [2]	$M_{cW}1$-$M_{YW}1$	0.619	0.0715
SS4(d) [2]	$M_{cW}10$-$M_{YW}10$	1.916	0.2216

[1] Ram generation interval (L) = 2.5 years; [2] Ram L = 1 year

Figure 4.6 and Table A4.10 shows the genetic responses per year for 160W, FES, YW and CW, using different breeding schemes and selection indices. It can be seen that a selection indices with lower female generation interval (SS3 and SS4) achieved greater genetic gain in all traits compared with the correspondent selection indices using SS1 and SS2 breeding schemes. The use of PI achieved higher values than $M_{cW}1$-$M_{YW}1$ and $M_{cW}10$-$M_{YW}10$ (except FES) but lower than $M_{YW}1$ and $M_{YW}10$ (except CW SS2 and 160W SS2 and SS4).

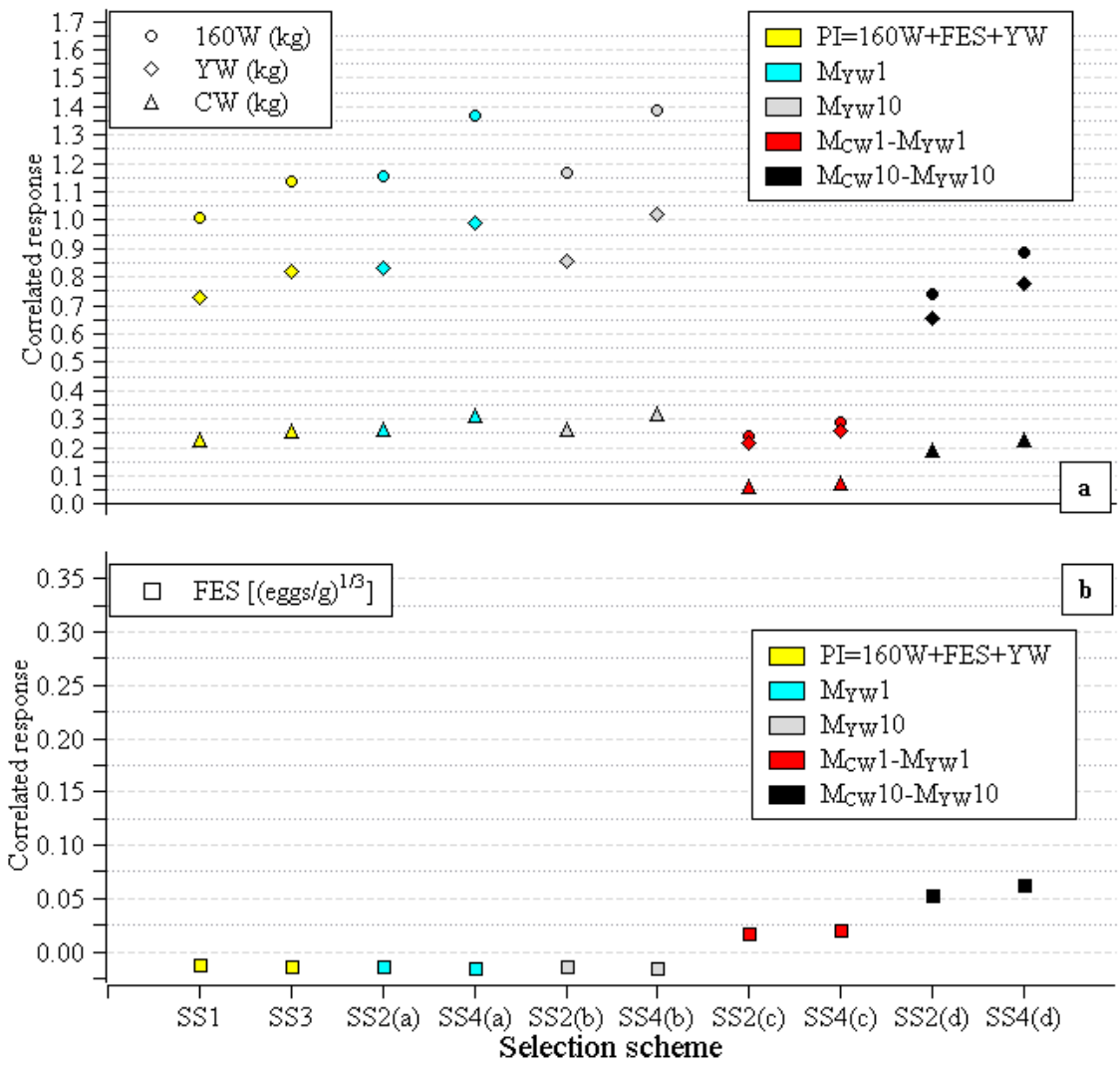

Figure 4.6. Correlated responses for (a) live weight at 160 days (160W), yearling weight (YW) and carcass weight (CW) and (b) faecal egg score (FES) (shapes), using different selection index (colours) within four different breeding schemes (SS1-SS4(d) from Table 4.7) with M_{CW} and M_{YW} explaining 1% and 10% of CW and YW total genetic variance.

4.5. Discussion

Results from scenario 1 suggest that the use of genomic information in combination with phenotypic information can increase the genetic gain compared with a traditional selection index without genomic information. As presented in Figure 4.2, PI+M_{CW} obtained higher genetic gains compared with PI+M_{YW} for all the different proportions of genetic variance explained by the SNP genotypes. This is due to the higher REW assigned to CW compared to YW (50 v/s -20% respectively). This big difference

between the REWs was selected in order to limit the genetic gain of YW in favour of CW due to the higher heritability that YW has (0.4) compared with CW (0.22), expressing that a higher heritability represents a higher proportion of the phenotype is determine by the trait's genotype (Falconer & Mackay 1996).

The use of M_{CW} and M_{YW} in combination with PI (PI+M_{CW}-M_{CW} Figure 4.2) increased genetic gains and accuracies of selection compared to PI. Different levels of genetic gains and accuracies of prediction can be appreciated depending on the proportion of genetic variance explained by each SNP genotype utilised as stated by Hayes et al. (2010b). This was shown in Figures 4.2 and 4.3, where much higher rates of genetic gains were obtained when the genomic information included into the selection index explained higher percentages of the total genetic variance (30% and 50%).

In relation to the accuracies of selection, and considering that for scenario one the same breeding scheme was used for all selection indices, the results obtained by indices without phenotypic information (Figure 4.3) were lower compared to the results obtained using selection indices having phenotypic and genomic information together (Figure 4.2). These results are in agreement with Janssen-Tapken et al. (2010).

One of the advantages of using genomic information in selection indices is the possibility of selecting animals at younger ages because the genetic merit of the animals can be estimated immediately after birth rather than waiting for the animal to produce phenotypic records. A consequence of this early selection is that the generation interval can be decreased and subsequently genetic gain will increase, providing accuracy of selection is maintained (Georges & Massey 1991; Goddard et al. 2010; Haley & Visscher 1998; Meuwissen 2003; Schaeffer 2006; Spelman et al. 2012). Supporting evidence is provided in Table 4.9; the only difference between these selection indices being the age structures utilised in the respective breeding schemes. The results show that a breeding scheme that decreases the generation interval of the breeding population can increase the rate of genetic gain in the nucleus flock. Moreover, it is apparent that when SNP genotypes are included conjointly with phenotypic information, higher genetic gains can be achieved in association with the reduction of the generation interval. This occurs even having selection indices with higher proportions of the traits genetic variances explained by the SNP genotype.

Even though in scenario 2 $M_{YW}10$ only described 10% of YW σ^2_G, and it's correlated response was not the highest of all the simulated indices the genetic gain obtained by SS4(b) (Table 4.9) was the highest outcome of all the study. This demonstrates that reducing the generation interval is critical to obtain higher rates of genetic improvement.

The economic value applied in the breeding objective for FES achieved the proposed objective of minimising the traits genetic change in the population when traits correlated to FES were included in the selection index (Figures 4.4, 4.5-b and 4.6-b). When no phenotypic information was included in the selection index, correlated responses in FES became positive but with values close to 0. This can be explained because the model was simulated considering the respective covariances between SNP genotypes and FES which were very low (the highest being 0.17 for $\sigma_{G_{M_{YW}},FES}$, with M_{YW} representing 50% of YW genetic variance).

For YW correlated responses, when the correlated trait phenotypic data is included jointly with M_{YW} using the same breeding scheme (Figure 4.4-c), the values obtained were all higher than when using PI alone. Likewise, when no phenotypic information was used (Figure 4.5-a and selection indices (c) and (d) on Figure 4.6-a) values obtained with M_{CW} and M_{YW} representing 1% and 10% of CW σ^2_G and YW σ^2_G were all lower than the gain obtained under PI. As a result, the objective of lowering YW genetic responses was very difficult to achieve with the traits included in the model. This is because of the high positive genetic correlation between YW and CW (0.7) and YW and 160W (0.76).

Correlated responses in CW were all positive (Figures 4.4, 4.5-a and 4.6-a) with some being close to the response obtained with PI, such as, PI-M_{YW} representing 1% of YW σ^2_G (Figure 4.4). The correlated responses for YW and CW shown in Figures 4.4 and 4.5-a reflect the high positive correlation between the two traits. Thus even though opposite genetic gains are desired (positive in CW; negative in YW), the chosen REWs did not enable this to occur.

This study provides a theoretical framework to illustrate the use of genomic information with selection index theory. But the use of more complex statistical and

genetic procedures are needed to simulate more precisely the application of this new technology. For example, the implementation of economic analysis taking into account costs and benefits of the application of molecular techniques are needed (Ruane & Sonnino 2007; Sonnino et al. 2007). Also, it must be noted that the change in genetic variance with time also known as Bulmer effect (Bulmer 1971; Dekkers 2007; Falconer & Mackay 1996), was not accounted for in the present study. Therefore, the predicted genetic gains are likely to be overestimates of the true genetic gain when the Bulmer effect is accounted for (Bulmer 1971). This problem can be addressed using stochastic simulation models considering several years, and is the subject of chapter 5.

4.6. Conclusions

The results obtained in this study illustrate the potential effects of genomic selection on the rate of genetic gain in a nucleus flock. Reviewing PI outcomes of scenario one, an increase of genetic gain and accuracy can be seen when genomic information is included together with phenotypic information. As expected, as the proportion of the trait explained by each SNP genotype increases, so it does the rate of gain in the objective. Considering that several traits can be included in a selection index, special care has to be taken when including correlated traits with high heritabilities in order to assign appropriate economic values to achieve a desired genetic response.

Under the same breeding scheme, the results from the present study suggest that the use of genomic information without phenotypic information can result in lower selection accuracies and lower genetic gains compared to a selection index using only phenotypic information. Higher genetic gains can be expected using just genotypic information (and higher accuracies) if the genotypic information accounts for a large proportion of the objective trait's genetic variance. Based on the results obtained for a selection index using phenotypic and genomic information together, the inclusion of a SNP genotype that explains only a low proportion of the correlated trait's genetic variance is not the best alternative to achieve greater genetic and economic gains, unless a reduction of the generation interval is achieved. The best genetic responses for the individual traits did not always lead to the best genetic gain in the objective and therefore the best economic outcome.

To assess more accurately the effect of including genomic information into a sheep breeding programme, stochastic models have to be implemented. This will give the flexibility to control different aspects such as environmental effects, population structure, farm management and the number of traits included. In addition, selection index predictions ignore changes in genetic variances changes (Dekkers 2007). Thus, the use of stochastic models will better demonstrate long term effects of genomic selection on selection accuracies and genetic gains in a breeding population.

4.7. Appendix to chapter 4

Table A4.1. Population parameters to simulate two pathways of selection (ewes and rams pathways), for a selection index with rams selected at 1 year old and ewes ages from 1 to 4 years old.

Selection Pathway	Population Size	Number selected	Proportion selected	Generation Interval
Ewes	183.6	84	0.4575	2.54
Rams	183.6	4.5	0.0245	2.50

Table A4.2. Population parameters to simulate two pathways of selection (ewes and rams pathways), for a selection index with rams selected at 160 days old and ewes ages from 2 to 5 years old.

Selection Pathway	Population Size	Number selected	Proportion selected	Generation Interval
Ewes	202.5	84	0.4148	3.40
Rams	202.5	15	0.0741	1.00

Table A4.3. Population parameters to simulate two pathways of selection (ewes and rams pathways), for a selection index with rams selected at 160 days old and ewes ages from 1 to 4 years old.

Selection Pathway	Population Size	Number selected	Proportion selected	Generation Interval
Ewes	183.6	84	0.4575	2.54
Rams	183.6	15	0.0817	1.00

Table A4.4. Population parameters to simulate two pathways of selection (ewes and rams pathways), for a selection index with rams selected at birth and ewes ages from 2 to 5 years old.

Selection Pathway	Population Size	Number selected	Proportion selected	Generation Interval
Ewes	202.5	84	0.4148	3.40
Rams	202.5	15	0.0741	1.00

Table A4.5. Population parameters to simulate two pathways of selection (ewes and rams pathways), for a selection index with rams selected at birth and ewes ages from 1 to 4 years old.

Selection Pathway	Population Size	Number selected	Proportion selected	Generation Interval
Ewes	183.6	84	0.4575	2.54
Rams	183.6	15	0.0817	1.00

Table A4.6. Genetic gains and accuracies of selection for selection indices including phenotypic information (PI) and SNP genotypes (M_{CW} and M_{YW}) explaining different proportions (1%,10%,30% and 50%) of the total genetic variance of carcass weight (CW) and yearling weight (YW) respectively.

Selection index	Genetic gain (¢/y)	Accuracy of selection
PI	2.796	0.440
PI+$M_{CW}1$	2.810	0.442
PI+$M_{CW}10$	2.942	0.463
PI+$M_{CW}30$	3.258	0.513
PI+$M_{CW}50$	3.617	0.569
PI+$M_{YW}1$	2.796	0.440
PI+$M_{YW}10$	2.804	0.441
PI+$M_{YW}30$	2.822	0.444
PI+$M_{YW}50$	2.844	0.448
PI+$M_{CW}1$-$M_{YW}1$	2.811	0.442
PI+$M_{CW}1$-$M_{YW}10$	2.812	0.442
PI+$M_{CW}1$-$M_{YW}30$	2.836	0.446
PI+$M_{CW}1$-$M_{YW}50$	2.857	0.450
PI+$M_{CW}10$-$M_{YW}1$	2.942	0.463
PI+$M_{CW}10$-$M_{YW}10$	2.947	0.464
PI+$M_{CW}10$-$M_{YW}30$	2.960	0.466
PI+$M_{CW}10$-$M_{YW}50$	2.976	0.468
PI+$M_{CW}30$-$M_{YW}1$	3.259	0.513
PI+$M_{CW}30$-$M_{YW}10$	3.260	0.513
PI+$M_{CW}30$-$M_{YW}30$	3.265	0.514
PI+$M_{CW}30$-$M_{YW}50$	3.270	0.515
PI+$M_{CW}50$-$M_{YW}1$	3.617	0.569
PI+$M_{CW}50$-$M_{YW}10$	3.617	0.569
PI+$M_{CW}50$-$M_{YW}30$	3.617	0.569
PI+$M_{CW}50$-$M_{YW}50$	3.618	0.569

Table A4.7. Genetic gains and accuracies for a selection index including phenotypic information (PI) and selection indices having only SNP genotypes (M_{CW} and M_{YW}) explaining different proportions (1%, 10%, 30% and 50%) of the total genetic variance of carcass weight (CW) and yearling weight (YW) respectively.

Selection index	Genetic gain (¢/y)	Accuracy of selection
PI[1]	2.796	0.440
$M_{CW}1$-$M_{YW}1$	0.454	0.072
$M_{CW}1$-$M_{YW}10$	0.687	0.108
$M_{CW}1$-$M_{YW}30$	1.032	0.162
$M_{CW}1$-$M_{YW}50$	1.287	0.203
$M_{CW}10$-$M_{YW}1$	1.338	0.211
$M_{CW}10$-$M_{YW}10$	1.408	0.222
$M_{CW}10$-$M_{YW}30$	1.554	0.244
$M_{CW}10$-$M_{YW}50$	1.689	0.266
$M_{CW}30$-$M_{YW}1$	2.306	0.363
$M_{CW}30$-$M_{YW}10$	2.321	0.365
$M_{CW}30$-$M_{YW}30$	2.355	0.371
$M_{CW}30$-$M_{YW}50$	2.391	0.376
$M_{CW}50$-$M_{YW}1$	2.975	0.468
$M_{CW}50$-$M_{YW}10$	2.976	0.468
$M_{CW}50$-$M_{YW}30$	2.979	0.469
$M_{CW}50$-$M_{YW}50$	2.982	0.469

[1]PI= phenotypic information selection index including 160W+FES+YW as traits, 160W= live weight at 160 days, CW = carcass weight at 160 days, FES= faecal egg score and YW= yearling live weight.

Table A4.8. Correlated responses year of live weight at 160 days (160W), faecal egg score (FES), yearling weight (YW) and carcass weight (CW), for selection indices including phenotypic information (PI) and SNP genotypes (M_{CW} and M_{YW}) explaining different proportions (1%,10%,30% and 50%) of the total genetic variance of CW and YW respectively.

Selection index	Correlation response			
	160W	FES	YW	CW
PI	1.006	-0.012	0.726	0.231
PI+$M_{CW}1$	1.009	-0.011	0.728	0.232
PI+$M_{CW}10$	1.030	-0.009	0.749	0.243
PI+$M_{CW}30$	1.087	-0.003	0.802	0.271
PI+$M_{CW}50$	1.156	0.002	0.864	0.302
PI+$M_{YW}1$	1.008	-0.012	0.728	0.231
PI+$M_{YW}10$	1.019	-0.011	0.749	0.234
PI+$M_{YW}30$	1.049	-0.011	0.803	0.240
PI+$M_{YW}50$	1.084	-0.010	0.867	0.249
PI+$M_{CW}1$-$M_{YW}1$	1.010	-0.011	0.730	0.232
PI+$M_{CW}1$-$M_{YW}10$	1.022	-0.011	0.754	0.235
PI+$M_{CW}1$-$M_{YW}30$	1.050	-0.011	0.804	0.241
PI+$M_{CW}1$-$M_{YW}50$	1.085	-0.010	0.867	0.249
PI+$M_{CW}10$-$M_{YW}1$	1.031	-0.009	0.751	0.244
PI+$M_{CW}10$-$M_{YW}10$	1.041	-0.009	0.768	0.246
PI+$M_{CW}10$-$M_{YW}30$	1.065	-0.008	0.812	0.251
PI+$M_{CW}10$-$M_{YW}50$	1.094	-0.008	0.865	0.258
PI+$M_{CW}30$-$M_{YW}1$	1.087	-0.003	0.803	0.271
PI+$M_{CW}30$-$M_{YW}10$	1.092	-0.003	0.812	0.272
PI+$M_{CW}30$-$M_{YW}30$	1.105	-0.003	0.836	0.275
PI+$M_{CW}30$-$M_{YW}50$	1.121	-0.003	0.866	0.278
PI+$M_{CW}50$-$M_{YW}1$	1.156	0.002	0.864	0.302
PI+$M_{CW}50$-$M_{YW}10$	1.157	0.003	0.866	0.302
PI+$M_{CW}50$-$M_{YW}30$	1.159	0.003	0.870	0.303
PI+$M_{CW}50$-$M_{YW}50$	1.162	0.003	0.875	0.303

Table A4.9. Correlated responses year of live weight at 160 days (160W), faecal egg score (FES), yearling weight (YW) and carcass weight (CW), for selection indices having only SNP genotypes (M_{CW} and M_{YW}) explaining different proportions (1%,10%,30% and 50%) of the total genetic variance of CW and YW respectively.

Selection index	Correlation response			
	160W	FES	YW	CW
$M_{CW}1$-$M_{YW}1$	0.213	0.015	0.190	0.055
$M_{CW}1$-$M_{YW}10$	0.458	0.023	0.542	0.109
$M_{CW}1$-$M_{YW}30$	0.776	0.034	0.978	0.181
$M_{CW}1$-$M_{YW}50$	1.000	0.042	1.279	0.231
$M_{CW}10$-$M_{YW}1$	0.555	0.044	0.427	0.150
$M_{CW}10$-$M_{YW}10$	0.650	0.046	0.571	0.170
$M_{CW}10$-$M_{YW}30$	0.837	0.051	0.855	0.210
$M_{CW}10$-$M_{YW}50$	1.003	0.056	1.103	0.246
$M_{CW}30$-$M_{YW}1$	0.947	0.076	0.717	0.256
$M_{CW}30$-$M_{YW}10$	0.977	0.076	0.766	0.262
$M_{CW}30$-$M_{YW}30$	1.045	0.077	0.879	0.276
$M_{CW}30$-$M_{YW}50$	1.117	0.079	0.995	0.291
$M_{CW}50$-$M_{YW}1$	1.218	0.098	0.920	0.330
$M_{CW}50$-$M_{YW}10$	1.225	0.098	0.931	0.331
$M_{CW}50$-$M_{YW}30$	1.241	0.098	0.959	0.334
$M_{CW}50$-$M_{YW}50$	1.258	0.098	0.990	0.338

Table A4.10. Correlated responses year of live weight at 160 days (160W), faecal egg score (FES), yearling weight (YW) and carcass weight (CW), for different breeding schemes and selection indices with M_{CW} and M_{YW} explaining 1% and 10% of CW and YW total genetic variance.

Breeding scheme	Selection index	Correlation response			
		160W	FES	YW	CW
SS1	PI	1.006	-0.012	0.726	0.231
SS2	$M_{YW}1$	1.152	-0.013	0.833	0.264
SS2	$M_{YW}10$	1.165	-0.013	0.857	0.267
SS2	$M_{cW}1$-$M_{YW}1$	0.243	0.017	0.217	0.063
SS2	$M_{cW}10$-$M_{YW}10$	0.743	0.053	0.653	0.194
SS3	PI	1.138	-0.013	0.821	0.261
SS4	$M_{YW}1$	1.371	-0.016	0.991	0.314
SS4	$M_{YW}10$	1.387	-0.015	1.020	0.318
SS4	$M_{cW}1$-$M_{YW}1$	0.290	0.020	0.259	0.075
SS4	$M_{cW}10$-$M_{YW}10$	0.884	0.063	0.778	0.231

CHAPTER 5

Stochastic simulation model for sheep breeding schemes using genomic selection and multitrait total merit index

5.1. Abstract

New DNA technologies allow the identification of genetic markers related to production traits, providing information to predict genomic breeding values for those animal traits at early age. The inclusion of this information into breeding schemes could allow the selection of breeding stock in early stages of their life. The present study developed a stochastic model that simulated a sheep breeding flock in which females were selected based on breeding values estimated using best linear unbiased predictor methodology and males were selected using only genomic breeding values for carcass weight estimated with a genomic best linear unbiased predictor methodology. The breeding objective of the simulated breeding population was to reduce the parasite load by decreasing faecal egg score; to reduce maintenance feed costs in ewes by decreasing yearling weight and to improve the income of the system by augmenting 160 day lamb carcass weight. The genomic information included to estimate the genomic breeding values represented 20% of the carcass weight genetic variance, a trait that can not be measured in living animals. Results of this study showed an increasing accuracy for genomic selection as phenotypic information is added to the training population. Also, proportionally higher genetic gains were obtained for carcass weight compared with the other simulated traits. This proved that the use of genomic selection combined with a multitrait selection index could be a valid option for increasing the genetic gains of traits recorded expressed after the animals have been selected for breeding.

5.2. Introduction

The sheep industry is comprised of several distinct breeding populations (typically breeds), and each of these populations will be progressing at various rates of genetic improvement. There are four factors that control the rate of towards the breeding goal:

- Intensity of selection
- Selection accuracy
- Genetic standard deviation
- Generation interval

These four factors that control the rate of genetic change are interdependent, so rather than seeking to either maximise or minimise each factor, it is necessary to consider them jointly to optimise the rate of genetic change (Blair & Garrick 2007).

The breeding goal of most sheep industries around the world is to identify animals of high genetic merit in traits that improves the profit of the production system. Selection for economically important quantitative traits in sheep is traditionally based on phenotypic records of the individual or relatives. Estimated breeding values, based on this phenotypic data, are commonly calculated by best linear unbiased prediction (BLUP), using phenotypic records (Meuwissen et al. 2001). With respect to New Zealand, initial sheep genetic evaluations were based on best linear prediction procedures, but as computing capacity allowed, BLUP was introduced (Blair & Garrick 2007).

The development of useful breeding schemes is not an easy task, with several processes needing to be considered. Harris et al. (1984), proposed a nine step systematic approach to build a comprehensive animal breeding scheme, which was used to develop a computer simulation model to iteratively analyse different breeding plans for broiler chickens.

Computer stochastic simulations can be described as mathematical routines developed with randomly generated parameters using predefined distributions, imitating the internal processes of a system (Moore & McCabe 1990; Reingruber & Gregory 1994). This technique has been widely used to evaluate breeding schemes in different species like, goats (Analla et al. 1995), swine (Pomar et al. 1991), dairy sheep (Smulders et al. 2007) or dairy cattle (Sörensen et al. 1999).

One or several genes producing an effect of any measurable level on a quantitative trait are called quantitative trait loci (QTL) (Hayes & Goddard 2001), and if the measurable trait is of economic importance the genes are called economic trait loci (Garrick & Snell 2005). Georges and Massey (1991), stated that the phenotypic expression of a trait in an animal is due the combination between the environmental effects and the effect of several “polygenes” known as Quantitative Trait Loci. Therefore as quantitative traits are usually affected by many genes the benefit from selection using

genetic markers is limited by the proportion of the genetic variance explained by the QTL (Meuwissen et al. 2001).

A single nucleotide polymorphism (SNP) can be defined as a genetic marker that is characterised by the variation in a nucleotide at a single base (Garrick & Snell 2005). The availability of thousands of SNPs spread across the genome gives the opportunity to include genome-wide marker information to predict total breeding values for the species under study and allows genomic selection (Calus et al. 2008; Meuwissen et al. 2001).

The term genomic selection was first proposed by Hillel et al. (1990), but the mathematical procedure was first described by Meuwissen et al. (2001). In genomic selection, breeding values are predicted using a large number of marker haplotypes across the entire genome (Calus et al. 2008). The theory behind this statement is, that some markers very near to a QTL could be combined into an haplotype. Therefore chromosome segments containing the same haplotypes, possibly are going to be identical by descent (IBD) and consequently have the same QTL allele (Meuwissen et al. 2001).

The tendency of some alleles at two different loci to be inherited together, generally because they are physically very close, is known as linkage disequilibrium (LD) (Falconer & Mackay 1996; Ruane & Sonnino 2007). This non-random association between alleles is expressed based on the amount of recombination between two loci during the gametic recombination phase (Falconer & Mackay 1996).

Considering all the previous statements, it is possible to state that genomic selection is a method that allows the estimation of genomic breeding values (GBV) without phenotypic information on the animals under selection. The genetic merit would then be calculated by summing up the values of each and every chromosomal segment (Garrick & Snell 2005). As a result, breeding schemes can be adapted to select breeding stock based on a ranking of the animal's GBVs, this could be done in early stages of their life (for example at birth), achieving an increase in the trait's genetic gain providing the breeding scheme's generation interval is reduced (Pryce et al. 2010).

Based on Meuwissen et al. (2001), several studies have been developed to evaluate different methods to incorporate genomic information to obtain GBVs. But regarding

deterministic or stochastic simulation studies analysing production traits in sheep breeding schemes, the author found very few studies (Pickering et al. 2013; Swan & Brown 2013).

Therefore taking into account all the studies utilising simulation techniques, the hypothesis of including genomic information into a simulated multitrait sheep breeding programme seems to be an interesting option for evaluating the long term genetic response of the traits under selection.

The objectives of this study were:

- To develop a stochastic simulation model for a sheep flock using a multitrait BLUP selection index conjointly with SNP genotypes associated with one production trait.
- To evaluate the implication of using these two selection procedures (SNPBLUP and phenotypic BLUP) together on a sheep flock.

5.3. Materials and methods

A stochastic model was programmed using Base SAS, SAS/IML and SAS 9.3 Macro language (SAS Institute Inc. 2011). The simulation represented a flock under a breeding programme, evaluated for the period of 20 years. This simulation process was replicated 100 times in order to obtain measurements of the variation of responses to selection.

The simulated traits (Table 5.1) were the same traits simulated in Chapter 3 (Table 3.2), being live weight at 160 days (160W), faecal egg score at 160 days (FES), live weight at 1 year or yearling weight (YW) and the lamb carcass weight (CW). Phenotypic and genetic parameters for these traits were obtained from Bennett et al. (1991), Huisman & Brown (2008) and Huisman et al. (2008). The breeding objective was to decrease FES (to improve parasite resistance), decrease YW (to decrease maintenance costs) and to increase CW of the lambs (increase production income).

The simulation model considered two procedures to select the animals as part of the breeding flock:

A BLUP selection to choose the flock's ewes.

A genomic selection (using SNPBLUP) to select males (new born) to be used at seven month of age as breeding rams.

For the SNPBLUP selection, the simulated SNP genotype was associated with CW, its inclusion was determined because CW is a trait that can only be recorded in dead animals (therefore it is too late to select measured animals as breeders), and also because it was the most economically important trait for income based on the economic weights utilised in Table 5.2.

Table 5.1. Phenotypic standard deviations (σ_P) heritabilities (on the diagonal), phenotypic (above the diagonal) and genetic (below the diagonal) correlations of traits[1] included in a selection index for sheep genetic improvement.

			Correlations			
Trait	Unit	σ_P	160W	CW	FES	YW
160W	kg	4.524	**0.54**	0.94	-0.01	0.65
CW	kg	1.766	0.92	**0.22**	-0.0094	0.611
FES[1]	score	1.483	0.34	0.3128	**0.28**	0.1
YW	kg	5.216	0.76	0.6992	0.13	**0.40**

[1]160W= live weight at 160 days, CW = carcass weight at 160 days, FES= faecal egg score defined as cubic root of number of eggs per gram [(eggs/g)$^{1/3}$] and YW= yearling live weight.

Table 5.2. Genetic standard deviations (σ_G), economic values (EV) and relative economic weights (REW) of the traits[1] included in the breeding objective.

Trait	σ_G	EV($)	REW
FES	0.78	-10.72	-30%
YW (kg)	3.30	-1.70	-20%
CW (kg)	0.83	16.93	50%

[1]FES= faecal egg score defined as cubic root of number of eggs per gram [(eggs/g)$^{1/3}$], YW= yearling live weight and CW = carcass weight at 160 days.

5.3.1. Flock structure

The base population simulated represented the size of an average New Zealand performance recorded flock (Garrick et al. 2000). It consisted of a 300 ewes flock as average and a standard deviation of 25 animals. The number of rams utilised the first 2 years corresponded to 3% of the ewes. Later, when lamb rams were used, the percentage of breeding males was 5% of the breeding females to account for worst

reproductive performance due difficulty in detecting and/or mounting females in oestrus, or that young rams mated less frequently than older rams (Kenyon et al. 2007). The number of ewe hoggets (between 1 and 2 year old) represented 31% of the ewes between 2 and 5 year old population. Table 5.3 shows the age structure of the flock for ewes older than 2 years. Ewes older than 5 years old and males older than 1 year old were culled.

Table 5.3. Age structure of ewes over two years old, of a stochastically simulated sheep nucleus flock.

Age in years	2	3	4	5
Ewes (%)	28%	26%	24%	22%

The sex of the born lambs was assumed as 50% males and 50% females and the lambing proportion was 1.5 lambs per ewe. Table 5.4 shows the birth rank probability assumed for the lambing.

Table 5.4. Birth rank percentages for lambs born in a stochastically simulated sheep breeding flock.

Number of lambs	1	2	3
Lambs (%)	60	30	10

For the death simulation process, it was assumed that 10% of the born lambs die before reaching the first year of life. Ewe's death percentage was 5% for 1 year old ewes and 2 % for older ewes.

5.3.2. Breeding scheme

Figure 5.1 shows the reproductive and replacement parameters assumed in the simulated flock. Female selection was based on BLUP breeding values at one year old (lambing at 2 years) and males were selected at birth using only SNPBLUP breeding values to breed at 7 months of age. A higher number of rams was selected (compared with the stochastic model in chapter 3) to account for a possible lower reproductive performance.

The purpose of selecting the males of the population based only on their GBVs was based on the idea presented by Meuwissen et al. (2001) that animals could be selected without any phenotypic record. Also it was taken into consideration that the decision of adopting this new technology by the farmers will be not only to improve the genetic gain of their production system but also if it could ease their working load by reducing the amount of data recording.

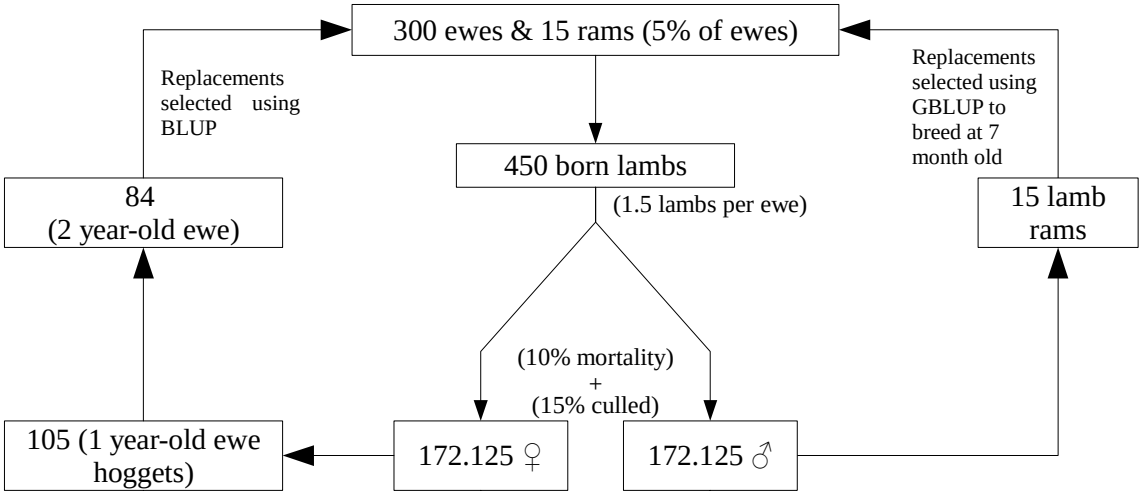

Figure 5.1. Breeding scheme structure for a simulated sheep breeding nucleus, selecting the females with BLUP breeding values and the males with genomic breeding values (SNPBLUP).

5.3.2.1. Best linear unbiased prediction selection

The females of the flock were selected over 1 year old (once the phenotypic information was recorded) using BLUP breeding values. Estimated breeding values (EBV) were generated via multitrait analysis using the package AIREML (Johnson & Thompson 1995). This software uses the average information matrix as second derivatives in a quasi-Newton procedure. The analysis used a multitrait mixed model considering the inclusion of pedigree information (Henderson & Quaas 1976). For the present simulation, analysing three traits (160W, FES and YW) the model can be written as:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{X}_3 \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Z}_3 \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \mathbf{e}_3 \end{bmatrix}$$

where:

$\mathbf{y_i}$ is a vector that represents the observed phenotypic records of trait i,

$\mathbf{X_i}$ is a incidence matrix of fixed effects (mean, year, flock) associated with trait i,

$\mathbf{b_i}$ is a vector of fixed effects for trait i,

$\mathbf{Z_i}$ is an incidence matrix relating animals with records to animals in the pedigree,

$\mathbf{a_i}$ is the vector of random animal effects for trait i,

$\mathbf{e_i}$ is a vector of the random residual effects of trait i,

The matrices of genetic (co)variance (**G**) and residual (co)variance (**R**) are represented as:

$$\mathbf{var}\begin{bmatrix}\mathbf{a_1}\\ \mathbf{a_2}\\ \mathbf{a_3}\end{bmatrix} = \begin{bmatrix}\mathbf{A}\sigma^2_{a_1} & \mathbf{A}\sigma_{a_{12}} & \mathbf{A}\sigma_{a_{13}}\\ \mathbf{A}\sigma_{a_{12}} & \mathbf{A}\sigma^2_{a_2} & \mathbf{A}\sigma_{a_{23}}\\ \mathbf{A}\sigma_{a_{13}} & \mathbf{A}\sigma_{a_{23}} & \mathbf{A}\sigma^2_{a_3}\end{bmatrix} = \mathbf{G} \text{ and}$$

$$\mathbf{var}\begin{bmatrix}\mathbf{e_1}\\ \mathbf{e_2}\\ \mathbf{e_3}\end{bmatrix} = \begin{bmatrix}\mathbf{I}\sigma^2_{e_1} & \mathbf{I}\sigma_{e_{12}} & \mathbf{I}\sigma_{e_{13}}\\ \mathbf{I}\sigma_{e_{12}} & \mathbf{I}\sigma^2_{e_2} & \mathbf{I}\sigma_{e_{23}}\\ \mathbf{I}\sigma_{e_{13}} & \mathbf{I}\sigma_{e_{23}} & \mathbf{I}\sigma^2_{e_3}\end{bmatrix} = \mathbf{R}$$

where:

I is an identity matrix of the order nxn where n is the number of measured animals,

A is the relationship matrix between animals (pedigree information),

$\sigma^2_{a_i}$ and $\sigma^2_{e_i}$ are the genetic and residual effects variances for trait *i*, and

$\sigma_{a_{ij}}$ with $\sigma_{e_{ij}}$ are their corresponding covariances between traits *i* and *j*.

After obtaining the EBV of each trait, a total merit index (IDX) for each animal (Hazel 1943) was generated, multiplying these EBVs with a regression coefficient (*b* value) as weighting values which were utilised in the stochastic simulation of chapter 3, to rank each animal based on the relative importance of the traits for which the evaluation was based.

$$IDX = (b_1 \times EBV_{160W}) + (b_2 \times EBV_{FES}) + (b_3 \times EBV_{YW})$$

5.3.2.2. Genomic best linear unbiased prediction selection

All simulated males were genotyped when they were born. The purpose was to have their genetic evaluation to select these animals as early in their life as possible, to allow a reduction of generation interval and sale of unwanted rams. The genomic information model represented 100 SNPs distributed within the same chromosome (Reich et al. 2001), for each simulated animal. For the present study, it was assumed that the whole SNP genotype represented 20% of CW genetic variance ($\sigma^2_{a_{CW}}$) and that one SNP (SNP 15) controlled 10% of the whole SNP genotype (2% of $\sigma^2_{a_{CW}}$). Therefore the individual SNP variance was assumed to be the $\sigma^2_{a_{CW}}$ multiplied by 20% (representing the amount accounted by the SNP genotype) divided by the number of simulated SNPs (100) giving an individual SNP variance of 0.00137 kg^2.

It was assumed a biallelic loci form for the simulated SNP genotype. There were two possible alleles for each SNP (1 and 2), and three possible genotypes: -1, 0 and 1 associated with a homozygous with a negative effect, heterozygous with a neutral effect and a homozygous with a positive effect, respectively.

The SNP information or allelic effects were fitted in a SNPBLUP as random effects (Meuwissen et al. 2001; Meuwissen 2003). It was assumed that every SNP had the same proportion of the total genetic variance considering therefore the same impact of each allelic effect on the related trait. The random effects were estimated using the mixed model equation:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \lambda\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix}$$

where:

X is a column vector of ones relating the animals and the trait mean (fixed effect),
Z is an incidence matrix, relating all the animals with each allelic form (-1, 0, 1) of the included SNPs (random effects),
y is a vector of the known CW phenotypic records,
b and **a** are vectors of fixed effects and random SNP effects respectively (BLUE and BLUP values),

I is an identity matrix of the same order as the **Z'Z** matrix.

λ is a value obtained when the residual variance is divided by each individual SNP effect variance.

The GBV estimation considered the effects covering the whole SNP genotype. Genomic selection used a subset of the total population called the training population, the requirement to be included as part of the training population was having phenotypic and genomic information at the moment of estimating GBVs. The information was analysed in order to estimate the regression coefficients for the random effects (each allelic form of each SNP). These estimates for the random effects were multiplied with the SNP genotypes of the predicted population (all animals with genomic information, with or without phenotypic records), and then all these estimates of each SNP were summed to obtain the GBV (Luan et al. 2009). Therefore, SNP effects were assumed to have an additive effect but not a dominance effect. The mathematical procedure previously described allowed estimation of GBV for all animals with genomic information.

5.3.2.3. Accuracy of genomic breeding values

The accuracy of GBV ($r_{TBV,GBV}$) was estimated as the correlation between the true breeding value and the predicted genomic breeding value shown as:

$$r_{TBV,GBV} = \frac{\sigma_{TBV,GBV}}{\sigma_{TBV}\sigma_{GBV}}$$

where, $\sigma_{TBV,GBV}$ is the covariance between TBV and GBV.

σ_{TBV} is the standard deviation of the true breeding values and

σ_{GBV} is the standard deviation of the estimated genomic breeding values.

The information utilised to evaluate the annual accuracies consisted of the training population (animals with phenotypic and genomic information) and animals with only genomic information updated annually.

5.3.2.4. Hardy-Weinberg equilibrium

To assess if SNP15 had any changes along the simulated selection period in order to see if that specific allelic form in the population had been under selection (Hardy-Weinberg equilibrium), a chi-square Pearson test χ^2 was determined (Falconer & Mackay 1996). The analysis was done measuring the allelic genotypic values at year 0 and at year 20 using the equation:

$$\chi^2 = \sum \frac{(O_{SNP} - E_{SNP})^2}{E_{SNP}}$$

where O_{SNP} is the observed allelic frequency at a certain period and E_{SNP} is the expected allelic frequency in the same period.

To assess the significance of this test, one degree of freedom is used (number of genotypes - number of alleles). For significance level of 5% and one degree of freedom the limit value utilised to accept or reject the null hypothesis is 3.84. Any result of χ^2 below this value, the null hypothesis that the population is in Hardy–Weinberg equilibrium is accepted.

5.3.3. Data generation

5.3.3.1. Phenotypic information

The phenotypic values of all the simulated traits for each available animal were obtained as the sum of the true breeding values (TBV), environmental effects, year and flock effects modelled as:

$$y_{gijklm} = \mu + M_g + F_i + G_j + e_{gij}$$

where, y_{gijklm} is the phenotypic value of any of the traits being simulated

μ is the mean of the population for the trait,

M_g is the effect of year *g*,

F_i is the flock effect *i*,

G_m is the TBV effect of animal *k*, and

e_{ijklm} is an environmental effect.

The year effect was assumed to have a variation of 5% of the phenotypic (co)variance matrix. The inclusion of the flock effect (*F*) was to account for a flock difference per replicate, and it was assumed to have a variation of 10% of the phenotypic (co)variance matrix. Genetic, environmental, flock and year effects matrices were created by the product of a randomly generated normal distribution matrix with the lower triangular matrix **D** of the Cholesky decomposition (Nejati-Javaremi et al. 2007) of the (co)variance matrix of each effect.

The genetic effect matrix was created with a randomly generated normal distribution matrix times the lower triangular matrix **D** of the Cholesky decomposition (Nejati-Javaremi et al. 2007) of a modified genetic (co)variance matrix including 160W, FES, YW and CW. The modification consisted in replacing the CW genetic variance ($\sigma^2_{a_{CW}}$) with a value representing 80% of $\sigma^2_{a_{CW}}$ as the SNP genotype represented 20% of $\sigma^2_{a_{CW}}$.

Then, based on Dekkers (2007) the true breeding value for CW was simulated as

$$G = Q + R$$

where,
G represents the additive genetic value of the trait under study,
Q is the genetic effects correlated with SNP genotypes,
R is the residual genetic effects independent of the markers (polygenic effect),

5.3.3.2. SNP generation

5.3.3.2.1. Base population SNP generation

In order to create the SNP genotype for the base population, two alleles for each SNP were generated (1 and 2) each of them representing a frequency (*p* and *q*) of 0.5. Based on Falconer & Mackay (1996), the additive amount contributed by each homozygous for each SNP considering no dominance was generated as:

$$a_i = \sqrt{\frac{V_{A_i}}{2p_iq_i}}$$

where,

a_i is the deviation from the mean (additive effect) of the positive homozygous of the SNP *i*.
V_{Ai} is the additive variance of the SNP *i*.
p_i and q_i are the allelic frequencies for the SNP *i*.

The additive effect for the negative homozygous of the SNP *i* was estimated as the product of -1 times a_i. The heterozygous form was simulated with an additive value of 0.

5.3.3.2.2. Offspring SNP generation

After pairing males and females based on their GBV and their IDX respectively, and according on Table 5.4 information, by using uniform random generation procedure, the number of offspring for each ewe was simulated. Once the number of lambs per ewe and sire are known, the inherited SNP genotype of each lamb was generated using one strand of each parent recombined genomic information, emulating the information carried by one gamete. All the SNPs were assumed to be allocated in one chromosome (Reich et al. 2001) therefore presenting LD between them; it was considered that the SNP in the 15th position (SNP15) was the "anchor" SNP. Recombination was assumed to be a crossing over recombination as an exchange of single strands between two participating chromatids (Andersen & Sekelsky 2010; Bernstein et al. 2011).

Figure 5.2 shows how the SNP genotype of each lamb is formed by inheriting one randomly recombined allele strand (gamete) from its sire and the other from its dam. The LD was simulated as the probability of a SNP to be represented as a specific allelic form in one locus of an allele strand. Therefore the higher the simulated value, the higher the probability of a neighbouring allele of the anchor SNP (randomly selected allele of SNP15) to be inherited together. The simulated probabilities ranged from 0.575, representing that the alleles have a probability of 42.5% to be recombined by the opposite allele, and 1 indicating that one of the alleles had 100% chance to be the chosen allele, and consequently the opposite allele has no possibility to be represented (no recombination). SNP15 as the anchor SNP had always a value of 1.

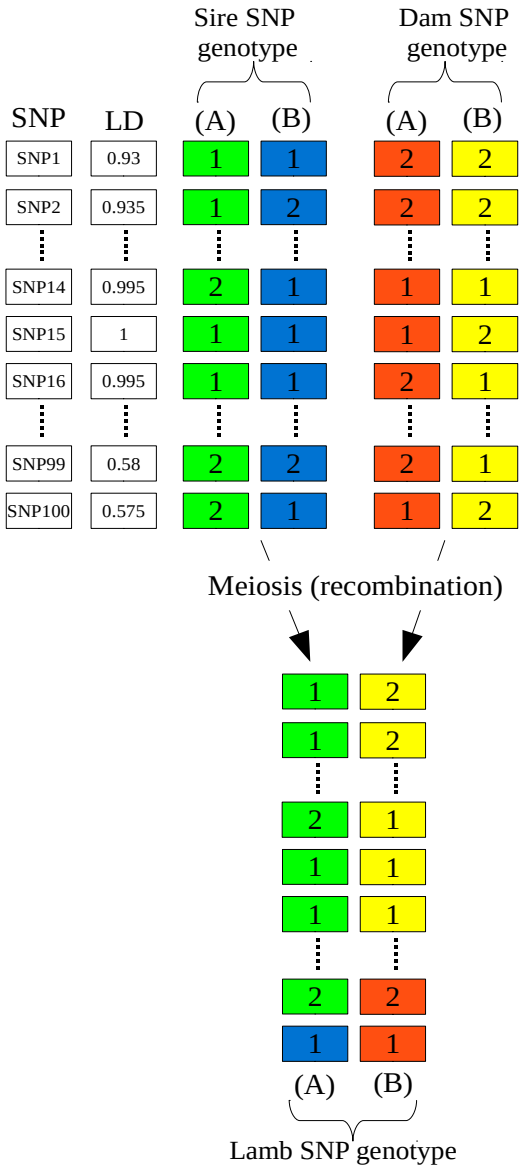

Figure 5.2. Schematic presentation of an inherited lamb SNP genotype formed by recombined parental randomly selected SNP genotypes allelic strands (A or B), based on simulated linkage disequilibrium (LD).

5.3.4. Data generation structure

The modelling process was built in subroutine modules using SAS Macro language. Each subroutine generated specific information that contributed to the creation of a database for a sheep flock under a 20 year selection programme. The model was replicated 100 times. A flowchart of the subroutines is shown in Figure 5.3.

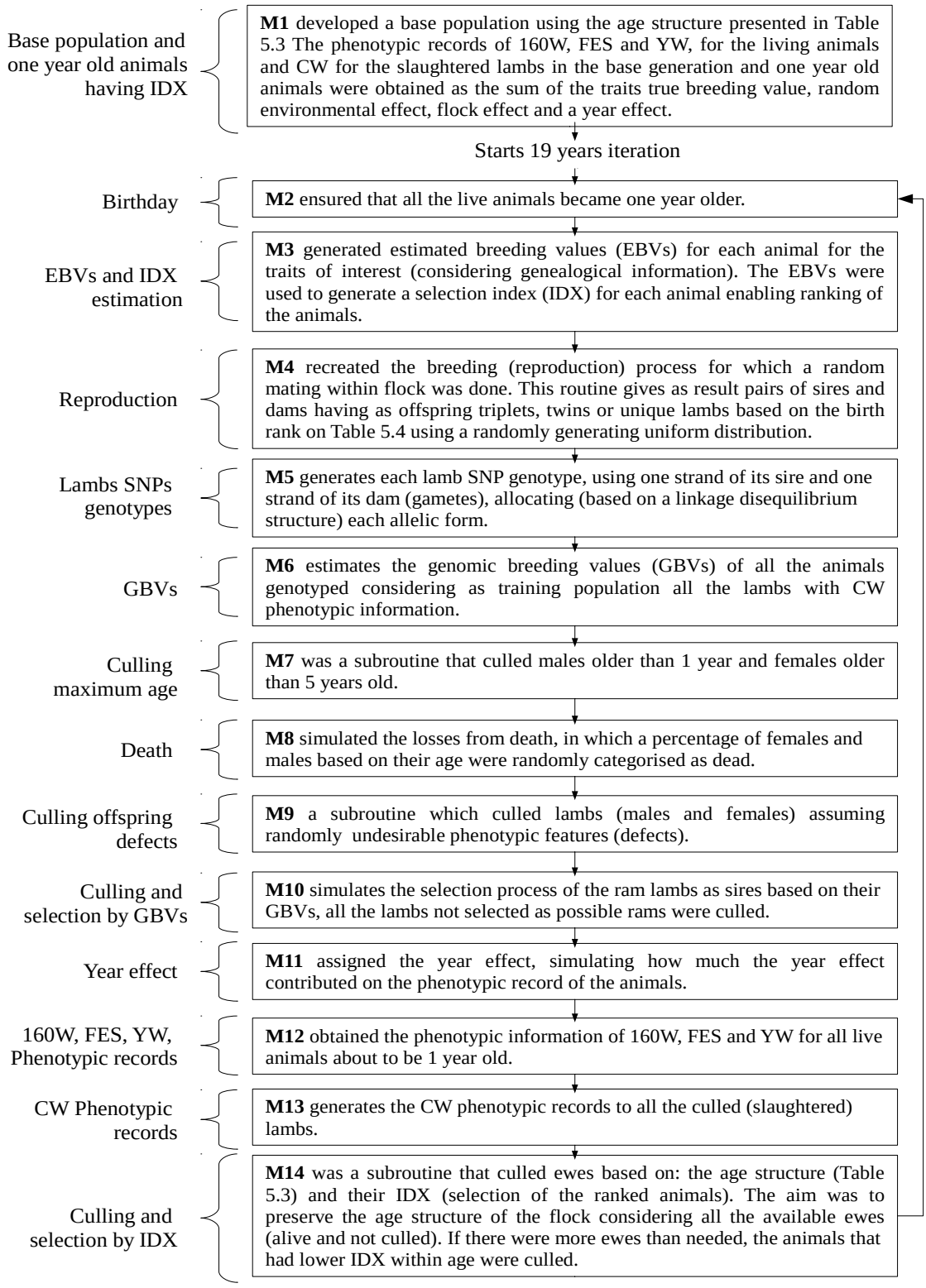

Figure 5.3. Stochastic simulation subroutine modules (**M1-M14**) of a 20 years sheep breeding programme, selecting the rams using carcass weight (CW) genomic breeding values (GBV) and ewes with a total merit index (IDX), built using estimated breeding values (EBV) of live weight at 160 days (160W), faecal egg score (FES) and yearling weight (YW).

5.4. Results

Figure 5.4 shows the mean of TBV for each trait per year and per replication, and the trend lines of average TBVs for all the replications. High genetic variance (σ_G^2) is observed for CW, 160W and YW, while for FES the variance is much lower. CW shows the highest average genetic trend of all the simulated traits over all the evaluated years. CW presents an increasing curvilinear trend, reaching at year 20 a TBV of 10.44 kg. Regarding the TBVs trends for 160W and YW the shape of the tendency lines appear to be more linear showing less gain than CW over the years. Their high values at year 20 were 5.66 and 4.33 kg respectively. FES showed least change during the analysed time period. The trend line of its TBVs was very close to 0 during the first 8 years after which it presented a slight increase, reaching the value of 0.27 $(eggs/g)^{1/3}$ at year 20.

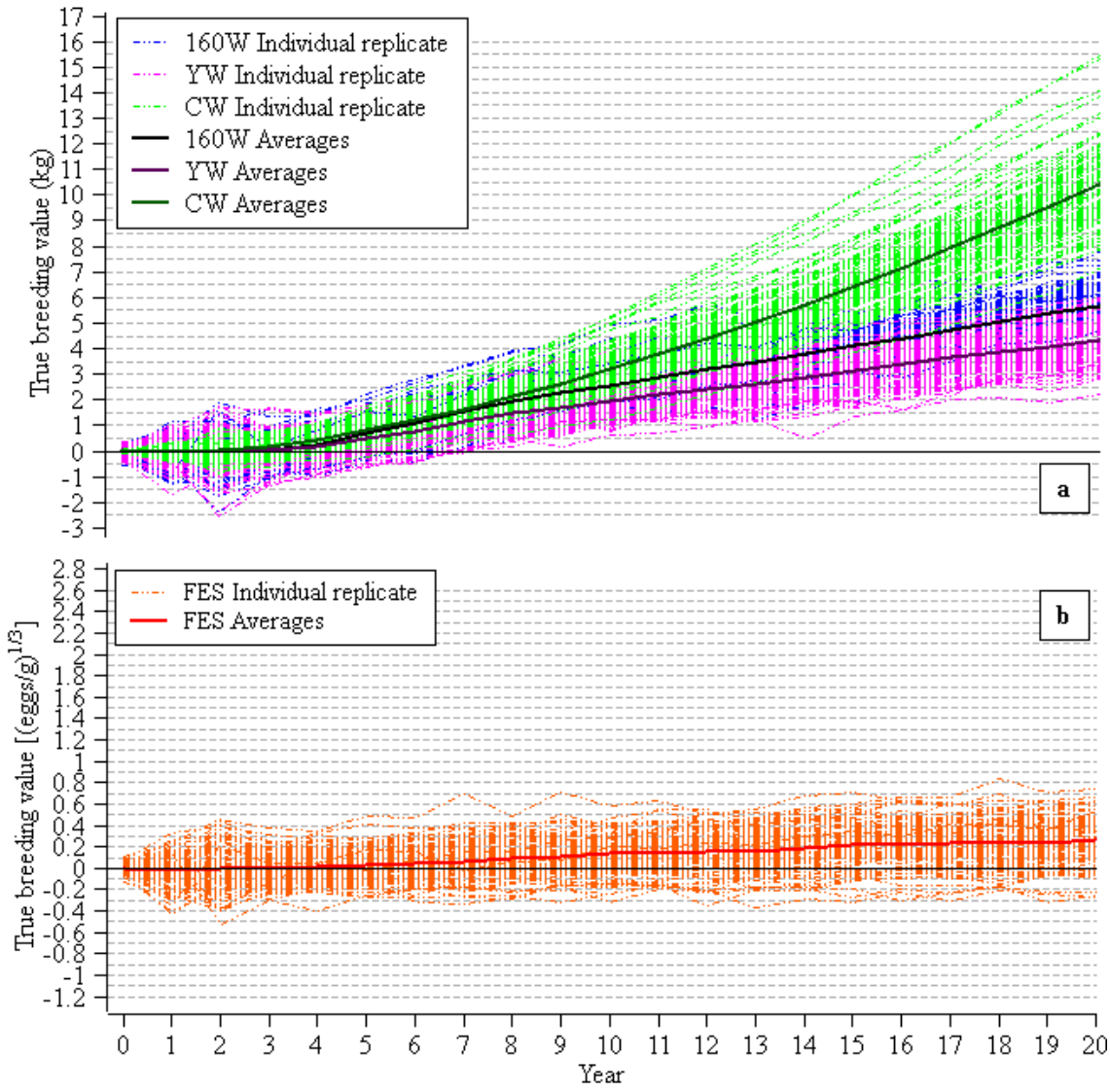

Figure 5.4. Trends of true breeding values of individual replicates (dotted lines) and replicates averages (continuous lines) for (a) live weight at 160 days (160W), yearling weight (YW) and carcass weight (CW) and (b) faecal egg score (FES), in a sheep breeding flock in a twenty years simulated breeding programme using genomic selection.

Figure 5.5 shows accuracies of prediction of the GBVs when using the CW phenotypic information provided by the training population updated on a yearly basis. Big dispersion of the accuracies can be appreciated during the first years of using genomic selection, ranging between 6 and 37% at year 2 and from 37 to 73% at year 6. Accuracies over 90% were achieved after 20 years of simulation.

Figure 5.6 presents the frequency change of the allelic forms of SNP15 across the 20 simulated years. The allelic form that was simulated to increase CW (SNP15 positive homozygous) shows an increase frequency over time.

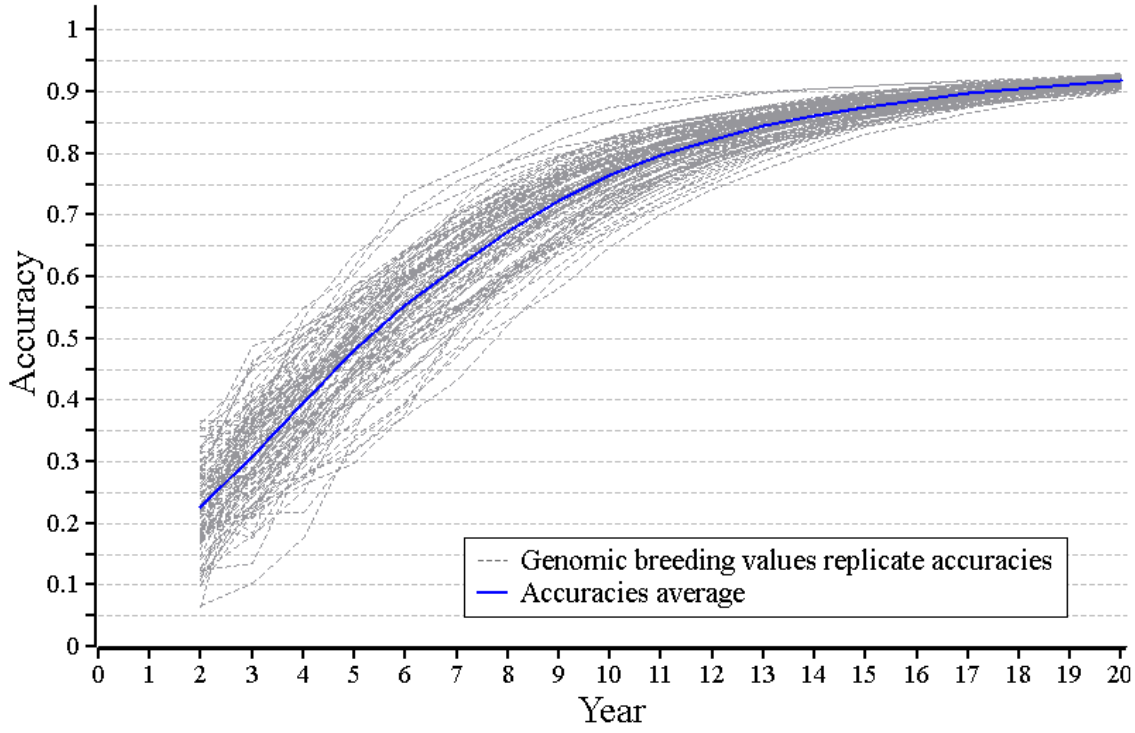

Figure 5.5. Trends of genomic breeding values accuracies for carcass weight (CW) using a yearly updated training population. Individual replicates are shown in grey dotted lines and averages of the replicates are shown in the continuous blue line.

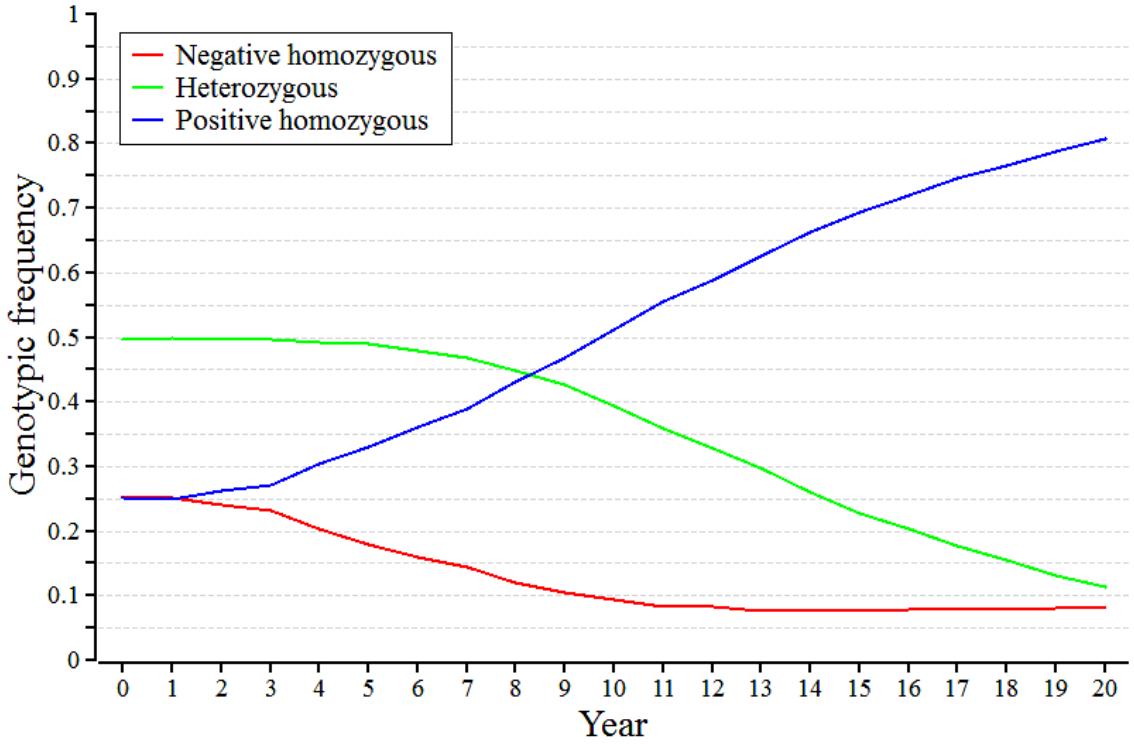

Figure 5.6. Long term response genotypic frequency averages for a non-recombinant anchor SNP (SNP15) for a simulated sheep population under 20 years breeding programme using genomic selection for carcass weight.

Table 5.5 shows frequencies of the 3 genotypes between the first simulated year (year 0) and the last simulated year (year 20). The positive homozygous (AA) presents a higher value, 0.806 at year 20 compared with 0.252 obtained at year 0. The other two allelic forms reduced their frequencies over the simulated 20 years. A high chi squared value of 26.821 was obtained considering the genotypic frequencies at year 20.

Table 5.5. Average of genotypic frequencies for positive homozygous (1), heterozygous (0) and negative homozygous (-1); and chi squared values (χ^2) for the simulated sheep flock at year 0 and year 20 (n = 100 replicates).

	Year 0			Year 20		
Allelic form	1	0	-1	1	0	-1
Genotypic frequency	0.252	0.496	0.252	0.806	0.114	0.08
χ^2		0.006			26.821	

5.5. Discussion

The aim of this study was to develop a stochastic model that simulated a breeding flock with ewes selected using a multitrait BLUP of three different traits, and rams selected based on GBV for CW estimated using SNPBLUP. The combined breeding scheme was chosen to simulate a scenario in which not all new-born animals were genotyped due to the costs that these procedures involve, a fact that a sheep breeder will absolutely consider at the moment to decide the implementation of this new technique.

The flexibility of adding or modifying information to nourish the stochastic simulation model was a key aspect compared with the deterministic model presented in chapter 4. Even though a stochastic simulation is considerably more demanding in terms of the required computer memory capacity and computing time as stated by Dekkers (2007). Also the results presented by the developed stochastic model (Figures 5.4 and 5.5) had a mean and a variance around the mean, not just an average linear trend such as the one produced by the deterministic model developed in Chapter 4.

The genetic trend of individual replicates obtained by the simulation (Figure 5.4) showed a large variation of within year records, presenting a wide range of TBV results due to the randomness of the simulation procedure. But it is important to highlight that besides the differences of the obtained values, all the TBVs follow the same tendency along the 20 years of evaluations.

Figure 5.4 also presented very low (practically 0) average genetic gain for all the traits during the first years. This is caused by an adjustment period of the population to the selection process, similar to the results obtained in the stochastic simulation of Chapter 3.

After the initial years of low TBV variation, a curvilinear trend line demonstrating the average true breeding value for CW showed an increasing genetic gain for the trait (Figure 5.4-a). This increasing of values was higher on a yearly basis, differing from the genetic trends presented by 160W, YW and FES that showed a linear growth with proportionally lower genetic gains. This can be explained because for the simulated population, two different methods were used to select the breeding animals. Firstly, the

males were selected using only genomic selection directly associated with CW, and second the females were selected based on an IDX, developed with the selection objective of also increasing CW as a correlated trait. The fact that CW had an heritability slightly over 0.2 (considered medium) and that the generation intervals were reduced (Schaeffer 2006) also helped to achieve the genetic gains of the selected traits.

The effect of using genomic selection having 20% of the genetic variance of CW explained by the whole SNP genotype, including one SNP (SNP15) that accounts for 2% of the same trait genetic variance, achieved, after 20 years of selection, a genetic gain five times higher, compared to the use of only IDX (stochastic simulation in Chapter 3). Regarding the genetic gains average results after 20 years of selection of the other two breeding objective traits, YWT and FES presented values of 4.33 kg and 0.27 $(eggs/g)^{1/3}$ using genomic selection, representing 65% and 67% respectively of the true breeding values obtained using only IDX. Then it can be said that, in the present simulated breeding programme because of the use of genomic selection for CW, the genetic gains for FES and YW were limited (Dekkers & Van der Werf 2007), accomplishing in a more suitable way the proposed selection objective of increasing CW.

Another point to highlight as part of the results of the simulation is the effect that the inclusion of SNP information had on the accuracy of prediction of the GBVs for CW. It can be appreciated in Figure 5.5 that average accuracies over 90% were achieved after 20 years of simulation. But to obtain these results the training population had to be updated in a yearly basis, also considering that as shown in the subroutine predicting the GBVs (M6 in Figure 5.3) the estimates for the SNPs effects were updated yearly as new animals with phenotypic and genomic information are incorporated adding more input data to the training population database. The annual replicate accuracies of prediction trends (grey dotted lines) showed a great variability between simulated replicates, especially during the first five years using genomic selection, but as the evaluated years and the accuracies increased, the variation between replicates becomes smaller.

Figure 5.6 presents the increase of the genotypic frequency of the positive homozygous of SNP15 (the one that was simulated to increase CW), reducing the amount of

heterozygous and negative homozygous frequencies close to 10% for both of them. This result illustrates the reduction of genotypic variation for this marker in the population after 20 years of selection. In a similar way this change of genotypic frequency after 20 years is reflected in Table 5.5, which showed that the simulated population was not in Hardy-Weinberg equilibrium as demonstrated by the χ^2 value of 26.821 (3.84 being the limit value for 5% of significance level). This loss of equilibrium clearly can be stated to be caused by a selection process (Falconer & Mackay 1996), as the simulation did not consider mutation of the simulated SNP neither migration.

Another point to consider regarding this increasing in the frequency of the favourable genotype for SNP15 is that the methodology used for genomic selection (SNPBLUP) has proven to be effective to achieve the desired objective of increased CW genetic values, even taking into account that in the utilised mixed model equation all the included random variables (SNPs) were analysed with the same individual SNP variance.

The present study proved that using genomic selection in combination with a multitrait selection index is an alternative option to increase genetic gains of traits recorded when animals are selected before the phenotype is available (e.g. lambs being slaughtered), as stated by Meuwissen and Goddard (1996), and also when early selection can be done (Dekkers & Van der Werf 2007; Schaeffer 2006), thereby reducing generation intervals. Results obtained in this work, contribute to the statement that genomic selection can be commercially useful if, the cost of implementing schemes using genomic information are lowered (Amer 2011), or markets gain interest in products that could be enhanced using genomic selection (Sumner & Davison 2006) justifying the implementation of this technology. Accuracies achieved by the model can reach very high levels (over 90%) but only considering that the training population and therefore the regression estimates for predicting the GBVs were updated and accumulated in a yearly basis. Using the simulation structure developed in the present model other situations or scenarios considering genomic selection for a sheep breeding programme could be evaluated.

CHAPTER 6

Accuracy of prediction of genomic breeding values for lamb carcass weight using simulation

6.1. Abstract

The use of genomic selection in breeding programme expects that the use of DNA information will lead to faster genetic gains. For a successful application of genomic selection an important point to consider is the accuracy of prediction of the breeding values obtained from the genomic information. This study analysed the accuracy of prediction for only newborn males either with their genomic breeding values being re-estimated on a yearly basis or alternatively using regression coefficients from a static training population (not updated on a yearly basis). The accuracy of prediction was estimated as the correlation between the true breeding values and the genomic breeding values (GBV) of a stochastically simulated sheep breeding flock. GBV accuracies of prediction were found using regression coefficients of a static training population to estimate the lambs genomic breeding values. These were compared with the accuracy tendency obtained when SNP random effects solutions were re-estimated yearly. The stochastic simulation showed a large accuracy variance within years caused by the males of the population being the only genomically selected animals, leaving to randomness the genomic contribution of the ewes. To obtain the highest possible accuracy of prediction, the most adequate statistical analysis method has to be chosen to predict GBVs. This is because the accuracy level of the predicted GBVs depends on the amount of variance that the genomic information represents and the variance distribution of the analysed SNPs.

6.2. Introduction

The use of genomic selection enables the design of novel breeding schemes (Pryce et al. 2010), and has become a very important field of investigation in recent years (Goddard 2012; Hayes & Goddard 2010; Pérez-Rodríguez et al. 2013; Van Eenennaam et al. 2014). The decision of implementing genomic selection as part of a breeding programme relies on the conviction that the use of DNA information will accomplish a faster genetic gain compared to a breeding programme based just on phenotypic information (Meuwissen et al. 2001).

A very important subject to consider for the successful application of genomic selection, is the accuracy of prediction (Luan et al. 2009). This should refer to how

accurate the estimated genomic breeding values (GBVs) are in predicting the true genetic value of an animal (TBVs) (Calus 2010; Cameron 1997). In real life, as the TBVs are not known, the accuracy estimation is done by correlating GBVs and phenotypic information (Harris et al. 2013; Hayes et al. 2010a; Hayes et al. 2010b; Luan et al. 2009; Saatchi et al. 2013), leading to the use of many different evaluation models. The purpose behind the development of these different statistical models to obtain GBVs, responds to the need to increase the GBVs accuracy of estimation for the selection of breeding animals, crops or forages (Pérez-Rodríguez et al. 2013; Wimmer et al. 2013).

Simulation techniques (or data modelling) are a very useful tool that, by imitating the internal processes of a system, attempts to predict certain aspects of it (Reingruber & Gregory 1994). Based on the previous information, a stochastic simulation model representing a sheep breeding flock selecting the breeding rams with genomic selection was developed in Chapter 5. This model provided information such as TBVs, phenotypic records and GBVs, which were needed to analyse some aspects of the long-term genetic response of traits for which the population was selected.

The objective of this study was to evaluate some aspects regarding how the accuracy of selection performs when:

The estimation of GBVs is done using the SNP random effect solutions obtained from a static training population (not updated on a yearly basis).
The accuracy of prediction is evaluated using only newborn males TBVs and GBVs.

6.3. Materials and methods

The study analysed a stochastically developed sheep breeding flock under a genomic selection breeding programme over 20 years (chapter 5). In order to obtain measurements of the variation of responses to selection, the simulation process was replicated 100 times. The resulting database (the same database obtained in chapter 5) contained genetic (true breeding values), environmental (e.g. year effect, dam age and flock effect), genomic (a 100 single nucleotide polymorphism sequence) and phenotypic information. The model considered the inclusion of four correlated simulated traits: live weight at 160 days (160W), faecal egg score at 160 days (FES),

yearling weight (YW) and lamb carcass weight (CW). Phenotypic and genetic parameters for these traits were obtained from Bennett et al. (1991), Huisman & Brown (2008) and Huisman et al. (2008). Phenotypes (*P*) for LW160, FEC and YWT were modelled as Falconer & Mackay (1996):

$$P = G + E$$

G represents the genetic value of the trait under study,
E is the random environmental effects affecting the trait.

Phenotypes for CW was modelled based on Dekkers (2007):

$P = G + E$ with *G* simulated as $G = Q + R$, where,

Q is the sum of the single nucleotide polymorphism (SNP genotypes) effects on the specific trait,
R is the residual genetic effects independent of the SNP effects.

In the present study, the estimated *Q* or $\hat{Q}$ is referred as genomic breeding values (GBVs).

6.3.1. Estimation of genomic breeding values

In order to obtain new GBVs, two phase process was undergone to the information provided by the same simulated population of chapter 5, therefore, the pedigree structure of the population was the same (same parents for the same offspring), even though other animals might have better GBVs to be selected as breeders. Phase one used a subset of the total population available by year, called the training population, consisting of all the animals that at the moment of estimating GBVs had been slaughtered (when new lambs are born), therefore having the CW phenotypic information recorded and a sequence of 100 SNPs as genomic information associated with CW. The second phase used all the animals that had genomic information with or without phenotypic records, named as the predicted population.

The information provided for each animal in the training population by the 100 simulated SNPs was fitted in a single-trait genomic BLUP (SNPBLUP) as random

effects (Meuwissen et al. 2001; Meuwissen 2003), every SNP was assumed to account for the same proportion of the total genetic variance of CW. To this, end random effect values were estimated using the mixed model equation:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \lambda\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix}$$

where:

X is a column vector formed only by ones relating each animal with the trait mean (fixed effect),
Z is an incidence matrix, relating each animal with an individual allelic form (-1, 0, 1) of the present SNPs (random effects),
y is a vector of the known CW phenotypic records,
b is the regression coefficient for fixed effects (BLUE),
a is the vector of estimate solutions for random effects (SNP),
I is an identity matrix of the same order as the Z'Z matrix and
λ is a value obtained when the residual variance is divided by each individual SNP effect variance.

When the training population was analysed, the estimates of the random effects were acquired (one for each allelic form of each SNP). Following, as the second step, the obtained random effect estimates were multiplied with an incidence matrix that related each animal of the predicted population to their SNP genotype allelic structure. The resulting products of each animal SNP sequence were summed within individual to obtain their CW GBV (Luan et al. 2009). Therefore, SNP effects were assumed to have an additive effect with no dominance or epistatic effects. The mathematical procedure previously described allowed estimation of GBV for all animals with SNP information.

In the present study two different GBVs data sets were evaluated. Firstly, the GBVs generated on chapter 5 which were estimated using regression coefficients for random effects updated on a yearly base. The second consisted of new GBVs for animals born from year 10 until year 20, using for each lamb the same SNP genotypes of the previous data set. These GBVs were obtained using the solution values for random SNP effects estimated using as the training population every animal with genomic and

phenotypic information only until year 9; therefore not updating the information to estimate the random SNP effects on a yearly basis.

6.3.2. Accuracy of GBV

The accuracy of GBV (r_{TG}) also known as the correlation between the true breeding value and the predicted genomic breeding value (Cameron 1997) was estimated as:

$$r_{TG} = \frac{\sigma_{TBV,GBV}}{\sigma_{TBV}\sigma_{GBV}}$$

Where $\sigma_{TBV,GBV}$ is the covariance between TBV and GBV,

σ_{TBV} is the standard deviation of the true breeding values and

σ_{GBV} is the standard deviation of the estimated genomic breeding values.

The purpose of using this statistic method was to evaluate how precisely the TBVs of the simulated sheep population are predicted by the SNP effects. Considering how these GBVs were obtained (as explained in point 6.3.1.) accuracies were estimated taking into account these two different GBVs data sets. Consequently two ways of evaluating the accuracies were used. The first way considered the estimation of accuracies of the GBVs for the training population plus the male lambs without phenotypic information, accumulating the amount of animals with GBVs on a yearly basis. The second considered the estimation of the accuracies for only the newly evaluated male lambs (males without phenotypic information) year by year, not considering the information provided by animals born in previous years.

6.4. Results

Figure 6.1 shows the accuracy of prediction trend of CW GBVs for newborn male lambs (red trend line). These GBVs are from the simulated population in Chapter 5 which were obtained using an accumulated yearly updated training population. An initial accuracy value of 0.28 was obtained in year 2, from this point on; the accuracy values constantly increased reaching a value of 0.77 at year 10. After year 10 the rate of increase in accuracy values declines, reaching an accuracy of to 0.91 at year 20. Figure 6.1 also shows the accuracy of CW GBVs for each of the 100 replicates (grey

dotted lines) using the solution of the SNP random effects obtained using the information of the training population to year 9 (not yearly updated) to estimate the GBVs of animals from year 10 onwards, therefore no accuracies were estimated before year 10. The figure shows a great dispersion of individual replicate values on year 10, having a standard deviation of 0.045 (Table A6.1), this dispersion gradually reduces during the following years, reaching at year 20 a standard deviation value of 0.013 (Table A6.1). The blue trend line included in Figure 6.1 represents the average by year of all the replicates trend lines presenting values at year 20 of 0.89.

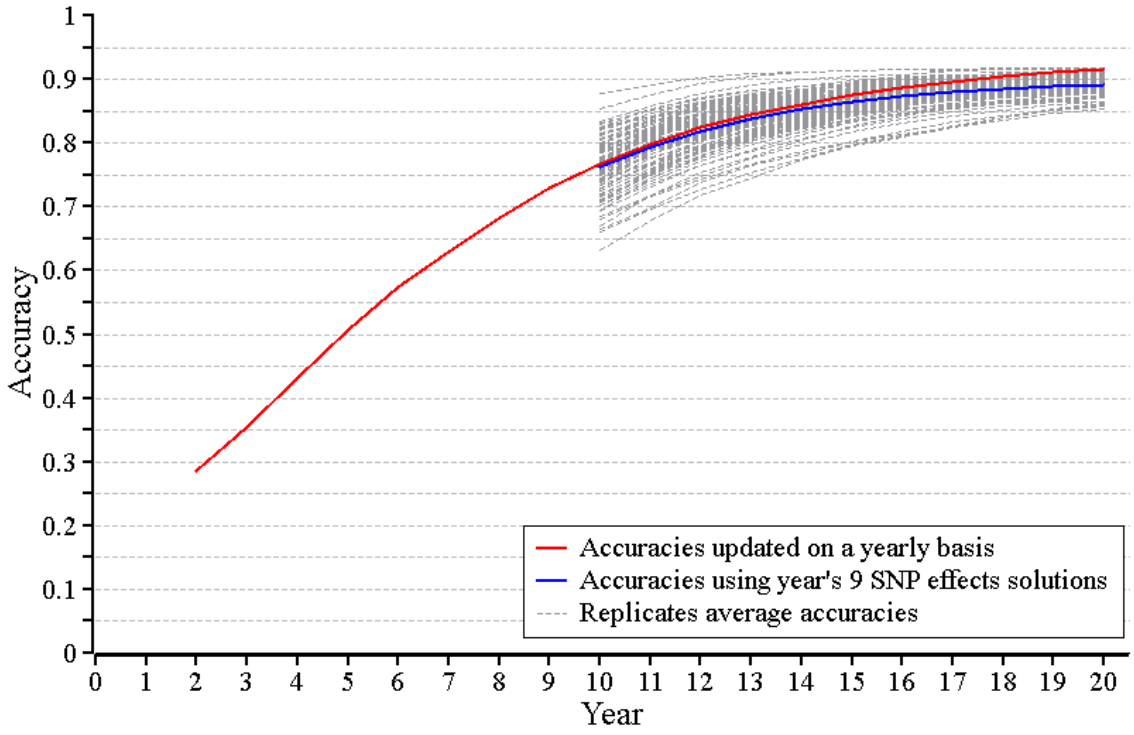

Figure 6.1. Accuracy of lamb carcass weight predicted genomic breeding values, estimated using random SNP effects solutions of a yearly accumulated and updated training population starting at year 1 (red trend line), and within-year individual replicate for genomic breeding values estimated using year's 9 SNP effects solutions (grey dotted trend lines) with replicate averages accuracies mean (blue trend line).

Figure 6.2 shows trend lines of the CW GBVs accuracies of prediction for new-born male lambs. Random SNP effects to obtain GBVs were estimated using a yearly updated training population. The grey dotted trend lines represent the within-year individual replicate accuracy averages for all the newborn lambs. An evident dispersion of the accuracies can be seen all along the simulated years, presenting the

higher differences of replicate variation between years two and seven, with standard deviation values of 0.083 and 0.078 respectively (Table A6.2), after which those differences were slightly diminished reaching a standard deviation value of 0.053 (Table A6.2) at year 20. The blue trend line presented in Figure 6.2 represents the year averages for the replicate accuracy averages trend lines. It shows low accuracy values in the early years (being 0.12 the lowest accuracy value at year two) but steadily increased until year nine reaching a value of 0.45; after which the trend line presents a plateau state obtaining at year twenty an average accuracy value of 0.42.

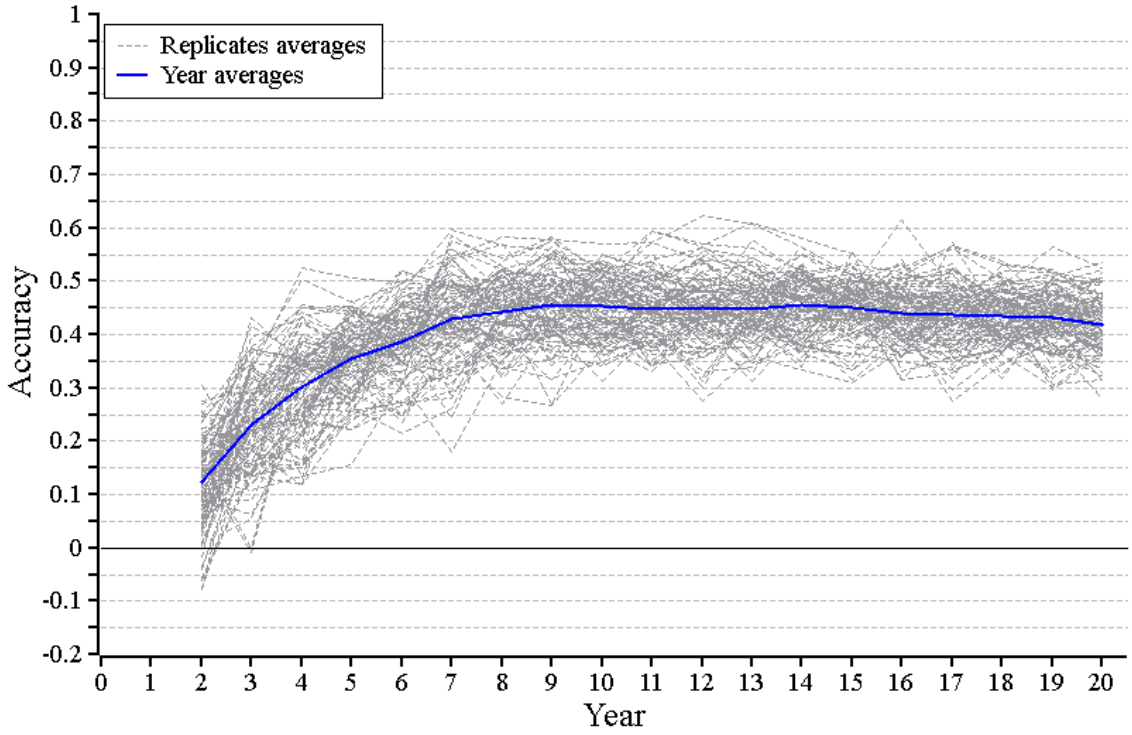

Figure 6.2. Individual replicate accuracies of prediction (grey dotted trend lines) and mean accuracy value from 100 replicates (blue trend line), for carcass weight genomic breeding values of newborn male lambs estimated using random SNP effects obtained from a yearly updated training population.

The information presented in Figure 6.3 shows accuracies of prediction of CW GBVs for newborn male lambs. The GBVs utilised for the accuracy estimation were obtained using year nine random SNP solution effects, therefore, same as for Figure 6.1, before year 10 no accuracies were estimated. Figure 6.3 shows a large dispersion of accuracies between individual replicates (grey dotted trend lines) along the evaluated years, presenting the lowest difference at year ten with a standard deviation value of

0.058 (Table A6.3), after which an irregular descending trend can be appreciated. The average of the replicates (blue trend line) shows the highest mean accuracy value of 0.45 at year ten with a clear decreasing tendency all through the following years reaching the lowest accuracy value (0.35) at year twenty.

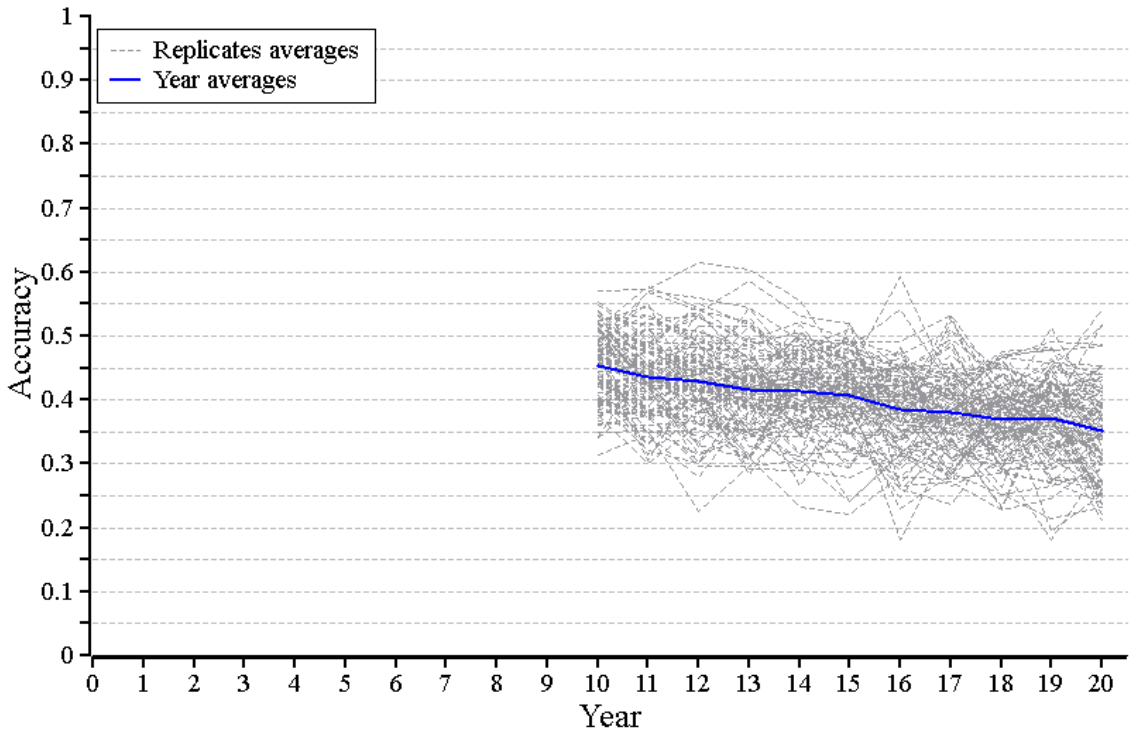

Figure 6.3. Individual replicate accuracies of prediction (grey dotted trend lines) and mean accuracy value from 100 replicates (blue trend line), for carcass weight genomic breeding values of newborn male lambs estimated using only year's 9 random SNP effects solutions.

6.5. Discussion

Studies evaluating the accuracy of genomic selection in sheep populations were not found in the literature; therefore studies involving other agricultural enterprises (poultry, crops, dairy and beef cattle) were used in the present discussion.

As stated by Reingruber & Gregory (1994), the use of simulation techniques allowed the prediction of certain aspects of a specific system. In this study, the simulation model developed in Chapter 5 representing a sheep breeding flock with breeding rams selected using genomic selection, delivered information such as TBVs (not known in

real life), GBVs and a sequence of SNP genotypes for each simulated animal. This information allowed the estimation of accuracy of prediction of GBVs for CW which relied on the correlation between TBVs and GBVs (Calus 2010; Cameron 1997).

The study showed that when GBVs were estimated using the solution of random SNP effects obtained from a static training population, in this case fixed at the ninth year (Figure 6.3), a steady decreasing tendency of the accuracies was observed in the following years. If these results are compared with the plateau tendency shown from year 10 by the average accuracies of prediction presented in Figure 6.2, it can be stated that higher accuracies can be achieved when the GBVs are re-estimated using the random SNP effects solutions of a yearly updated training population. It is important to highlight that the animals and therefore the genomic information utilised to estimate carcass weight GBVs were exactly the same for the two comparisons. Therefore the only difference between these two studies was the way that random SNP effects solutions were utilised. Based on the previous comparison, it may be inferred that, the more distant the genetic relationship between the training population and the predicted animals, the lower the accuracy of prediction between GBVs and TBVs. This previous statement, concurs with Saatchi et al. (2013), who estimated accuracies for Hereford beef cattle, comparing four nationally evaluated training populations (U.S., Canada, Argentina and Uruguay), using two different Bayesian methods to estimate the GBVs. Also Habier et al. (2010) evaluating GBV in German Holstein cattle and Van Eenennaam et al. (2014) analysing GBV accuracies of production and quality traits in layer chickens arrived at the same conclusion, stating that in order to provide a good accuracy of prediction, SNP effects should be re-estimated including the most recent phenotypic data from relatives.

The reduction in the accuracy of prediction of GBVs can have significant consequences, especially if the breeding animals are selected only on their genomic information (without recording any phenotypic data). The long-term response of selecting animals based on GBVs estimated using random SNP effects solutions from an old training population, will incur a reduction of the GBVs accuracies. As a result, if breeding animals of lesser genetic value are selected, lower genetic gain of the population for the trait under selection can be expected.

The wide range of the GBVs accuracies of prediction presented by the individual simulated replicates (grey trend lines Figures 6.1-6.3), can be explained by two causes. The first cause may be due to the fact that for the simulated breeding flock, the genomic selection was carried out only through the males of the population without considering the females. This approach left to randomness the ewe's contribution of genomic information for the flock. The second cause can be linked also with a randomness issue, but this time related to crossing over during meiosis which occurs during the simulated reproduction process (Andersen & Sekelsky 2010). Considering these causes, the randomness effect occurred in the breeding population through utilisation of animals with unknown SNP genotypes into the breeding scheme; this deficiency could be minimised with the inclusion of the female's genomic information.

The high levels of GBVs accuracy of prediction shown in Figure 6.1 for all replicates is likely to be a result of the recurrent inclusion of animals with more phenotypic and genomic information. As the number of years progress, the number of animals in the database with both phenotypic and genomic information increases, in contrast to the steady number of new animals included having only genomic information; therefore by having more information on a yearly basis, higher accumulated accuracy levels are obtained. The difference in the GBV accuracy of prediction between the red trend line and the blue trend line of Figure 6.1, can be explained in the same way that arises in Figure 6.3. The accuracy reduction when GBVs were obtained using only the SNP effects solutions of a fixed-year (year 9) training population (blue trend line, less information), against the use of yearly updated training population to predict GBVs (red trend line, more information).

When the accuracy levels are measured only on the predicted animals (Figure 6.2), low accuracy levels can be appreciated at the beginning of the simulated time period. The accuracies in the following years increase as the number of animals in the training population also increase. Thus more phenotypic and genomic information is included, in order to re-estimate the random SNP effects on a yearly basis to obtain GBVs for the desired trait (Goddard 2012). After year nine of simulation, the accuracy levels reached a plateau state, showing that even when more information is incorporated on a yearly base to estimate GBVs this is very little gain in accuracy. This suggests that, without changing the methodology for predicting the GBVs the accuracy level was

mediated by the genetic information that nourishes the simulation model. As stated by Hayes et al. (2010a) the accuracy of the predicted GBVs depends on the amount of the trait's variance that the analysed genomic information is accounting for. This also confirms the results of the deterministic simulation generated in Chapter 4, showing higher accuracies when more of the total genetic variance is explained by the included genomic information.

The use of different statistical procedures to analyse the genomic information in order to predict GBVs will result in different accuracy outcomes. But it is not obvious if the resulting accuracy level will be higher or lower. As presented by Habier et al. (2011), Wimmer et al. (2013) or Wolc et al. (2011), the most accurate analysis method has to be chosen individually for each trait separately. This is because the accuracy of the method depends on the variance distribution of the analysed genomic information. For example as stated by Hayes et al. (2010b) SNPBLUP could be equally, or even most accurate than other analysis methods (eg. Bayesian methods) especially if a large number of SNP effects account for a small amount of variance (Hayes et al. 2010a).

6.6. Conclusions

Based on the analysis of the data obtained from a stochastic simulation model developed in Chapter 5, it can be expected that, when the genetic relationship between the training population and the predicted animals is more distant, the level of accuracy of prediction for the GBVs will be reduced. This suggests that SNP random effect solutions required to estimate GBVs should be re-estimated on a yearly basis using both, accumulated phenotypic and SNP information.

The analysis showed a large variation of accuracies of prediction for CW GBVs within year. This was caused because the genomic selected animals were only the ram lambs, leaving to randomness the contribution of genomic information provided by the ewes. In relation to the accuracy level of the predicted GBVs, it can be stated that it depends on the amount of variance that the genomic information explains and also, the contribution that each analysed SNP makes to the total genetic variance. Therefore the most adequate analysis method has to be chosen to predict GBVs with the highest possible accuracy of prediction.

6.7. Appendix to chapter 6

Table A6.1. Within-year standard deviations for 100 replicates carcass weight predicted genomic breeding values accuracies, estimated using year's 9 SNP effects solutions.

Year	Accuracy of selection standard deviation
10	0.045
11	0.040
12	0.035
13	0.030
14	0.027
15	0.023
16	0.020
17	0.018
18	0.016
19	0.014
20	0.013

Table A6.2. Individual replicate accuracies of prediction standard deviations, for carcass weight genomic breeding values of newborn male lambs estimated using random SNP effects obtained from a yearly updated training population.

Year	Accuracy of selection standard deviation
2	0.083
3	0.088
4	0.089
5	0.065
6	0.070
7	0.078
8	0.063
9	0.067
10	0.058
11	0.062
12	0.070
13	0.066
14	0.062
15	0.062
16	0.068
17	0.064
18	0.059
19	0.065
20	0.072

Table A6.3. Individual replicate accuracies of prediction standard deviations, for carcass weight genomic breeding values of newborn male lambs estimated using only year's 9 random SNP effects solutions.

Year	Accuracy of selection standard deviation
10	0.058
11	0.062
12	0.070
13	0.066
14	0.062
15	0.062
16	0.068
17	0.064
18	0.059
19	0.065
20	0.072

CHAPTER 7

General Discussion

7.1. Introduction

Animal breeding is the controlled propagation of domestic animals in order to improve desirable traits in future generations and involves the use of knowledge from several branches of science, including genetics statistics, reproduction, computer science and economics (Garrick & Snell 2005). Molecular genetics is becoming a real alternative to be employed in different animal breeding programmes, due to reduction in the cost of DNA sequencing and the possibility of imputing marker sequences from a group of animals having a complete genomic sequence into related individuals having incomplete genomic sequences (Goddard 2012). The inclusion of DNA information into selection programmes to improve animal production, rely on the prospect that the use of this information for genomic selection will produce a faster genetic gain than the one achieved with only phenotypic information (Meuwissen et al. 2001).

The studies presented in this thesis followed a structured sequence, in which the final objective was to develop a stochastic simulation model for sheep breeding which included genomic and phenotypic information in order to analyse the genetic responses for traits of economic importance in sheep production. The topics considered in this thesis were, the development of a deterministic and a stochastic model to simulate a sheep breeding scheme with four correlated traits under selection (Chapter 3). The development of deterministic multitrait selection index models considering the inclusion of genomic information to evaluate accuracy of selection and genetic gains (Chapter 4). The development of a stochastic model that simulated a sheep breeding flock selecting ewe replacements using a multitrait best linear unbiased predictor methodology (BLUP) and the rams with genomic selection (Chapter 5). Finally the analysis of prediction accuracies of the genomic breeding values estimated using a yearly updated training population or with the information provided by a static training population (Chapter 6).

7.2. Effect of including genomic selection in a sheep breeding programme

The effectiveness of including genomic information into a sheep breeding programme is shown in Figure 7.1. The compared genetic responses were obtained from a breeding flock selecting all animals based on their individual breeding values estimated with

BLUP (Chapter 3) and from another breeding flock selecting ewes using BLUP and rams with genomic selection (Chapter 5).

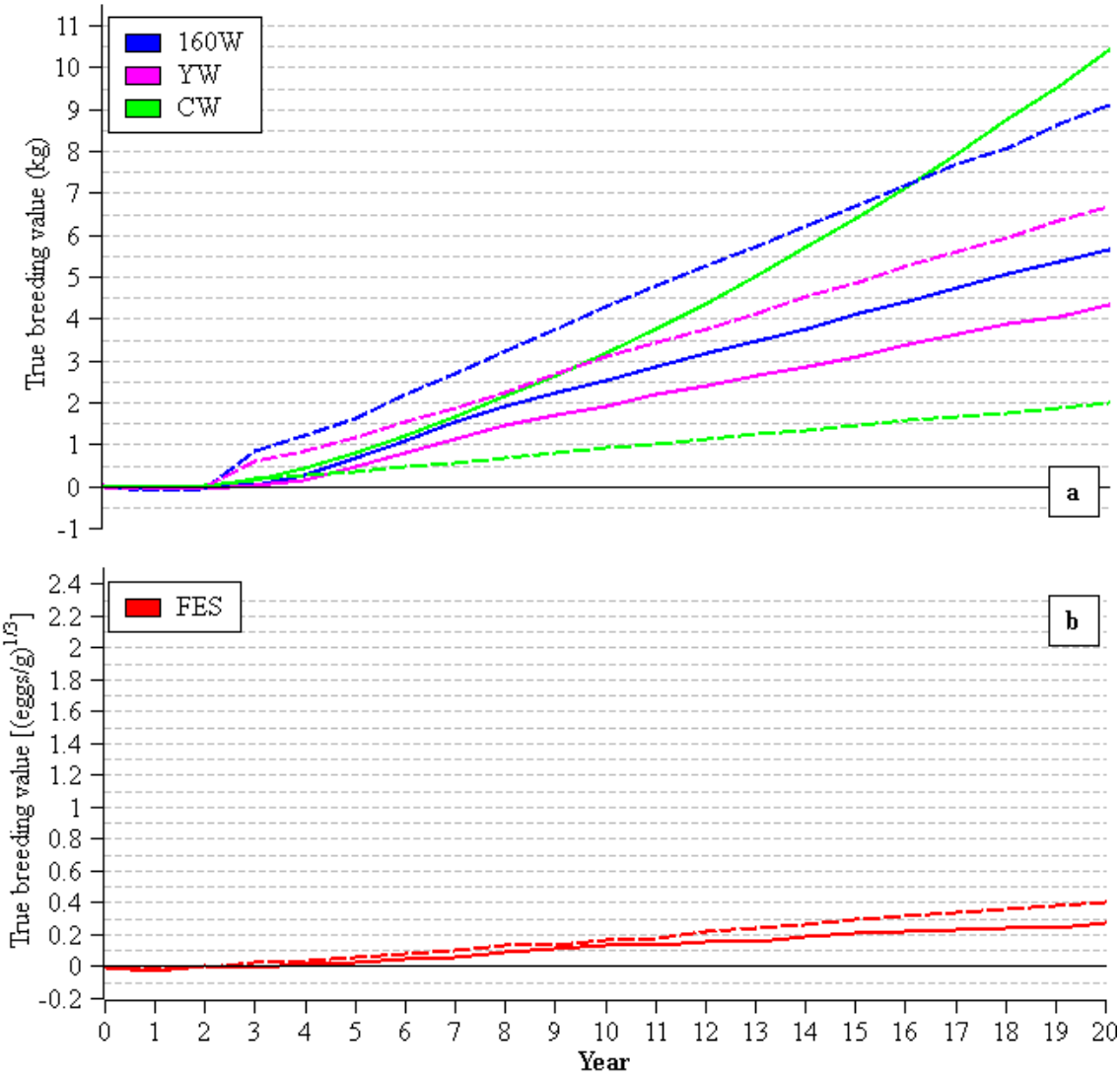

Figure 7.1. Changes in true breeding values for (a) live weight at 160 days (160W), yearling weight (YW) and carcass weight (CW) and for (b) faecal egg score (FES), of a sheep breeding flock with a breeding programme using genomic selection (continuous lines) and a breeding programme without genomic information (dotted lines).

The breeding programme utilising genomic information shows a higher genetic response for carcass weight (CW) compared with the genetic gain obtained by a traditional breeding programme without the inclusion of genomic information. These results suggest that for an objective trait that cannot be measured directly, genomic

selection may enable greater rates of genetic gain than a selection based on the use of a correlated trait. The degree of benefit will depend on the accuracy of selection and reduction in generation interval achieved in the genomic system, relative to the genetic correlation between the objective trait and the indicator trait in the traditional selection system.

The use of the proposed genomic selection in a breeding programme also resulted in lower rates for yearling weight (YW) and faecal egg score (FES), compared with the genetic gains obtained by the simulated breeding programme which did not use genomic selection. After 20 years of evaluation, the genomic selection programme obtained 2.37 kg and 0.14 (eggs/g) $^{1/3}$ lower for YW and FES, respectively. These outcomes represent a better way to achieve the breeding objective of reducing the maintenance costs of adult ewes by lowering YW and by decreasing health problems and costs related to parasite loads by reducing the amount of FES.

One of the main differences between the two simulated breeding schemes is the age at which the rams were utilised as breeders (ewe age structure was the same for both simulations). In the breeding scheme that didn't use genomic selection (Chapter 3) the rams were used as breeders for two years giving a generation interval of 2.5 years. This is compared to the breeding scheme that included genomic selection (Chapter 5), whereby rams were utilised only once being one year old when their lambs were born. The previous point illustrates that the reduction of the flock's generation interval was due to the males of the flock. Considering the genetic responses of the analysed traits the simulation model demonstrated that genomic selection is a viable technology to successfully select new-born animals as breeders.

7.3. Limitations and considerations of the simulation programmes

There are several key factors that could lead to the success or failure of any simulation programme. The importance of the biological information nourishing the model is critical and was discussed earlier. However, some non-biological aspects that restricted the number of selection scenarios simulated in the present thesis were features regarding the programming language and the software utilised to develop the simulation models. The deterministic models of this thesis were developed using

Microsoft Excel (2003). The implementation of selection index equations and calculations of selection intensities, accounting for the age structure of the ewes and rams, required the estimation of asymptotic genetic gains demanding knowledge of quantitative genetics and the design of breeding programmes. Once the deterministic model was implemented, a number of scenarios could be simulated and computer time was not a limiting factor.

Using selection index theory, the variance of the selection index or the variance of the breeding objective can be estimated. Also, as shown by Falconer & Mackay (1996) and Cameron (1997), the individual genetic gain for each trait included in the analysis (correlated response) can be estimated. However, the author of this thesis did not found a formula to calculate the variance of individual traits genetic responses using selection index theory.

The stochastic simulation models were developed using SAS 9.3 (2011), that provided statistical, data management and programming capabilities, allowing the construction of recursive and random routines (as presented in Chapters 3 and 5).

Stochastic models provide many advantages over deterministic models. Changes in genetic gain can be determined at any time point compared to the asymptotic genetic gains which only presents a linear trend. In addition, and compared with deterministic models; in stochastic models besides estimating the variances of genetic gains for the index and breeding objective, the genetic response variances for individual traits can be estimated. An estimation of the variance of genetic response enables the determination of economic risk which is very important for those making investment decisions and whose livelihood depend on the productivity and profitability of the livestock enterprise (Conington et al. 2004; Dekkers & Shook 1990).

However, some limitations have to be considered when developing stochastic models including genomic selection:

Becoming competent in the computer language can became a very time demanding process (especially if there is no previous programming knowledge).

Simulations produce very large datasets, therefore a large amount of computer memory is required for running the simulation processes. In the present thesis, a 2.3 GHz quad-core computer with 8-gigabyte of RAM memory was used. The developed genomic

selection stochastic simulation, produced a 2-gigabyte database and took on average 20 hours of actual computing time for the entire simulation model using macros of SAS 9.3 (2011).

7.4. Possible new applications of the model

The use of sub modules (macros) to structure the developed stochastic simulation programme proved to be a very versatile tool. This technique of modelling, by adding or modifying small segments of the computational routine, allowed the implementation of several models in order to study or analyse the performance of a simulated breeding programme under different scenarios or any other question regarding genomic information such as the:

- Economic impact of the breeding programme
- Accuracies of different statistical methods to estimate GBVs
- Number of SNPs or animals needed to maximise genetic gain,
- Analysis of other breeds and traits

7.4.1. Economical analysis of breeding programmes

Economic analysis is an important step in considering in the efficacy of breeding programmes (Harris et al. 1984; Lopez-Villalobos & Garrick 2005). An economic analysis can help to understand the reasons for success/failures of animal breeding programmes. This improved understanding should allow the development of better techniques to enhance livestock genetics (Amer 2011).

The use of bioeconomic farm models is a useful methodology to assess the economic effects of modifying biological characteristics or production related traits (Van Arendonk 1991), and also to create economically rational breeding programmes (Thomasen et al. 2014; Wolfová et al. 2009). The present thesis developeded a genomic selection approach to include in a breeding programme, and evaluated the long-term genetic response to selection of different traits in a sheep breeding flock. However, an economic analysis is required to evaluate the financial effect of CW, YW and FES genetic changes on farm profitability.

Few studies were found by the author that analysed the economic impact of genomic selection where traits were constrained using an antagonistic genetic correlation (Pickering et al. 2013). The genomic selection model developed here could be incorporated with complex simulation models such as whole farm models. Examples are those developed by Baudracco et al. (2013) for dairy grazing systems and by Pickering et al. (2013) for a dual-purpose sheep population, to appraise how the change of genetic values for the selected traits will affect economic aspects of the sheep industry like costs of production (by including genotyping and evaluation costs) and income of the production system.

7.4.2. Accuracies of different statistical methods to estimate

It was discussed (Chapters 2 and 6) that different statistical methods can be used to estimate GBVs, for example SNPBLUP utilised in Chapter 5 or BayesB and BayesC methods utilised by Saatchi et al. (2013). Each of these procedures for estimating GBVs, deliver different accuracies of prediction depending on the contribution in the genetic variance of each SNP effect included (Habier et al. 2011; Hayes et al. 2010a).

The stochastic genomic selection model developed in this thesis utilised SNPBLUP to estimate GBVs, this was achieved by statistically fitting the information provided by the SNPs as random effects (Meuwissen et al. 2001). Nevertheless, the model has the power to utilise any other statistical procedure that predicts GBVs. Future research using other methods; exploring the genetic responses or the accuracy impact of predicting GBVs for a sheep breeding flock could be developed. In addition, studies comparing the outcomes of several different statistical estimation techniques could be analysed. These experiments could be implemented by either modifying the existing module or developing a new module, to perform the genomic breeding value estimation (M6 in Figure 5.3) and incorporating it as a new subroutine of the genomic selection stochastic model.

7.4.3. Number of SNPs or animals needed to maximise genetic gain

As found by Calus (2010), the accuracy of the genomic estimations depends on the number of individuals from the reference population, number of analysed markers and the proportion of genetic variance affecting the trait. The stochastic simulation model

developed in this thesis could be used in future studies to predict the number of individuals needed to accurately detect certain number of SNPs or to evaluate the number of SNPs required for a determined number of animals to achieve a certain accuracy level for any desired trait of a sheep breeding flock.

To evaluate the number of individuals needed to achieve a certain accuracy level using a specific statistical procedure to estimate GBVs, the heritability of the trait for which the GBVs are being estimated must be known. In addition, it is necessary to assume a fixed number of SNP, which are controlling a known percentage of the total genetic variance of that trait. Once the previous information is determined, the simulation model could be modified to deliver the answer of the desired accuracy. This modification is just for the first subroutine M1 in Figure 5.3. Then the model should be adjusted to use the desired statistical methodology to estimate GBVs considering a fixed number of animals. Once the routine runs, accuracies of prediction will be estimated as the correlation between the true breeding values and GBVS (Calus 2010; Cameron 1997) of the simulated animals (which is the same methodology as utilised in Chapters 5 and 6). The previous routine (population generation, GBV estimation and accuracy prediction) has to be run for a number of iterations, possibly one hundred times as was performed in the present thesis, to obtain an average and the deviation values of the accuracies of prediction for that number of evaluated animals. After this process has finished the routine has to be rerun simulating a population with additionally evaluated animals to assess the change in the accuracies of prediction when more individuals are incorporated.

The model could also be used to identify how many SNPs are required to achieve a desired level of predicted accuracy with a specific number of animals in a sheep breeding flock. This simulation model would be very similar to the one proposed previously for evaluating the number of individuals needed to obtain a certain level of accuracy. The major difference is that the number of SNPs will be variable, increasing in every replicate run, instead of changing the number of animals. This model has to consider the contribution of each SNP to the total genetic variance, because as stated in 7.3.2, the achieved accuracy depends on the statistical procedure chosen to estimate the GBVs.

7.4.4. Analysis of other breeds and other traits

In this thesis, the stochastic simulation model was implemented to simulate a single breed sheep breeding flock. Therefore, to nourish the simulation model all the input information (physiological aspects, traits parameters and management decisions) was carefully chosen to represent a long-term response of this unique animal population.

The phenotypic and genetic mean values and variances of the traits included in the stochastic model could be modified to account for other traits to be analysed. The simulation model could also be extended to include more traits to be analysed, either in the selection index or in the breeding objective. These modifications, combined with changes in physiological and management, could be made to represent other sheep breeds or production system. For example, this modification could be used to simultaneously evaluate the genetic responses of multiple sheep breeds breeding programmes. It could also enable, as presented by Hayes et al. (2009a) in dairy cattle, Saatchi & Garrick (2013) for beef cattle or Moghaddar et al. (2013) for sheep, the comparison of predicting GBVs accuracies using genomic information from multiple breeds.

7.5. Conclusions

The main conclusions of this thesis can be summarised under three topics,

- Simulation techniques applied to genetic evaluations.
- Genomic selection analysis
- Accuracies of estimating genomic breeding values

7.5.1. Simulation techniques for genetic evaluation.

All studies in this thesis demonstrated that the use of simulation techniques is a very useful tool that enabled a prospective view for a production system. Regardless of the type of simulation utilised (deterministic or stochastic), models can be used to assess the long-term responses to selection of a breeding population. It was shown that, compared with a deterministic simulation, a stochastic model allows the variance of genetic gain to be estimated for the selection index, breeding objective and individual traits. On the other hand, once a deterministic model was fully developed, it allowed a

faster review of various scenarios (including genomic selection), because it was less demanding regarding computer processing time.

7.5.2. Analysis of genomic selection

The use of genomic information in a breeding programme can increase the rate of genetic gain for traits that are difficult to measure, e.g., carcass weight in this study. The power of genomic selection relies on the possibility of selecting breeding animals at younger stages of life, achieving higher genetic gains due to a reduction of the generation interval of the breeding population.

7.5.3. Genomic breeding values accuracies

Considering the methods of estimating genomic breeding values, accuracies of genomic selection are lower than the accuracies obtained with progeny testing. The accuracies of genomic breeding values depend on the amount of information used to estimate them and the contribution to the total genetic variance of each included SNP. Therefore the most adequate analysis method must be chosen to predict GBVs with the highest possible accuracy of prediction. An important point shown in this thesis is that, accuracies of GBVs were reduced as the genetic distance between the training population and the population where GBVs are estimated increased. Therefore, it is recommended to re-estimate SNPs effects on a yearly basis.

Addendum

Since the mathematical work of this thesis was originally completed (at the end of 2011), several advances in the field of study have arisen. For example, parameters of the simulated traits were gathered from different publication sources rather than New Zealand parameter estimates (e.g. Pickering et al. (2012) or Pickering et al.(2013)). Other advances that occurred while, or after the present thesis was being developed are some topic discussed in Chapter 2 like, the estimation of GBVs using a genomic relationship matrix in GBLUP or utilizing some of the machine learning methodologies described.

Another point that might be worthy to mention is that, in the genomic simulation developed in this thesis the inheritance of the SNP genotypes or linked locations, was set up assuming an anchor SNP (with it's alleles having an equal possibility to be sampled), and the rest of the alleles at other SNPs of the SNP genotype, were sampled assuming a linkage disequilibrium model, rather than using genetic distances and making inheritance dependent on the result for the adjacent marker.

Considering all this statement, the author of this thesis acknowledge that the results of this study might have delivered different results, like different accuracies of prediction for the estimated genomic breeding values for the simulated population (possibly higher). Other result that might have a different outcome due to a difference in the simulation procedure utilised, might be the genetic trend of CW for the genomic selection chapter (Figures 5.4-a and 7.1). Assumptions of a genetic relationship between the SNP panel utilized and all the traits in the breeding objective (and not only CW), might have restricted the genetic response of the traits under selection showing a more believable genetic gain for CW (compared with the extreme curvilinear genetic gain shown).

References

Amer P 2011. The economics of successes and failures in animal breeding. Journal of Animal Breeding and Genetics 128: 327-328.

Amer PR 2000. Trait economic weights for genetic improvement with SIL. Proceedings of the New Zealand Society of Animal Production 60: 189-191.

Amer PR, Payne GM 2009. Combining estimates of SNP effects when they are subpopulation specific. Proceedings of the Association for the Advancement of Animal Breeding and Genetics 18: 216-219.

Analla M, Sanchez-Palma A, Munoz-Serrano A, Serradilla JM 1995. Simulation analysis with BLUP methodology of different data-structures in goat selection schemes in spain. Small Ruminant Research 17: 51-55.

Andersen SL, Sekelsky J 2010. Meiotic versus mitotic recombination: Two different routes for double-strand break repair. BioEssays 32: 1058-1066.

Baudracco J, Lopez-Villalobos N, Holmes CW, Comeron EA, Macdonald KA, Barry TN 2013. E-dairy: A dynamic and stochastic whole-farm model that predicts biophysical and economic performance of grazing dairy systems. Animal : an international journal of animal bioscience 7: 870-878.

Beckmann JS 1988. Oligonucleotide polymorphisms: A new tool for genomic genetics. Nat Biotech 6: 1061-1064.

Beef + Lamb New Zealand 2012a. Annual slaughter statistics for New Zealand lamb, dating back to 1998. http://www.beeflambnz.com/information/on-farm-data-and-industry-production/industry-production-trends/ [accessed 13 November 2012].

Beef + Lamb New Zealand 2012b. Sheep and beef new season outlook 2012-13. http://www.beeflambnz.com/information/on-farm-data-and-industry-production/industry-production-trends/ [accessed 13 November 2012].

Beef + Lamb New Zealand 2013. Beef + lamb new zealand's annual report on the financial year from 1 october 2012 to 30 september 2013. http://www.beeflambnz.com/data/events-archive/agm/ [accessed 29 December 2013].

Bennett GL, Johnson DL, Kirton AH, Carter AH 1991. Genetic and environmental-effects on carcass characteristics of southdown x romney lambs. 2. Genetic and phenotypic variation. Journal of Animal Science 69: 1864-1874.

Bernstein H, Bernstein C, Michod RE 2011. Chapter19. Meiosis as an evolutionary adaptation for DNA repair. In: DNA Repair. Dr. Inna Kruman (ed.) InTech Open Publisher. doi: 10.5772/25117.

Bishop CM 2006. Pattern recognition and machine learning. Springer. 738 p.

Blair HT, Garrick DJ 2007. Application of new technologies in sheep breeding. New Zealand Journal of Agricultural Research 50: 89-102.

Boerner V, Johnston DJ 2013. Accuracy of Igenity direct genomic values in Australian Angus. Proceedings of the Association for the Advancement of Animal Breeding and Genetics No. 20. p 211-214, Napier, New Zealand.

Breiman L 1996. Bagging predictors. Machine Learning 24: 123-140.

Breiman L 2001. Random forests. Machine Learning 45: 5-32.

Bulmer MG 1971. The effect of selection on genetic variability. The American Naturalist 105: 201-211.

Calus MP, Meuwissen TH, de Roos APW, Veerkamp RF 2008. Accuracy of genomic selection using different methods to define haplotypes. Genetics 178: 553-561.

Calus MPL 2010. Genomic breeding value prediction: Methods and procedures. Animal : an international journal of animal bioscience 4: 157-164.

Calus MPL, Veerkamp RF 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. Journal of Animal Breeding and Genetics 124: 362-368.

Cameron ND 1997. Selection indices and prediction of genetic merit in animal breeding, CAB International, Wallingford, Oxfordshire OX10 8DE, UK. 203 p.

Conington J, Bishop SC, Waterhouse A, Simm G 2004. A bioeconomic approach to derive economic values for pasture-based sheep genetic improvement programs. Journal of Animal Science 82: 1290-1304.

Crawford AM 2003. The use of genetic modification technologies in the discovery of genes affecting production traits and disease resistance in animals. New Zealand Veterinary Journal 51: 52-57.

Croiseau P, Legarra A, Guillaume F, Fritz S, Baur A, Colombani C, Robert-Granié C, Boichard D, Ducrocq V 2011. Fine tuning genomic evaluations in dairy cattle through SNP pre-selection with the elastic-net algorithm. Genetics Research 93: 409-417.

Daetwyler HD, Hickey JM, Henshall JM, Dominik S, Gredler B, van der Werf JHJ, Hayes BJ 2010. Accuracy of estimated genomic breeding values for wool and meat traits in a multi-breed sheep population. Animal Production Science 50: 1004-1010.

de los Campos G, Naya H, Gianola D, Crossa J, Legarra A, Manfredi E, Weigel K, Cotes JM 2009. Predicting quantitative traits with regression models for dense molecular markers and pedigree. Genetics 182: 375-385.

Dekkers JCM 2007. Prediction of response to marker-assisted and genomic selection using selection index theory. Journal of Animal Breeding and Genetics 124: 331-341.

Dekkers JCM, Shook GE 1990. Economic evaluation of alternative breeding programs for commercial artificial insemination firms. Journal of Dairy Science 73: 1902-1919.

Dekkers JCM, Van der Werf JHJ 2007. Strategies, limitations and opportunities for marker-assisted selection in livestock. In: Guimarães EP, Ruane J, Scherf BD, Sonnino A, Dargie JD (eds.) Marker-assisted selection: Current status and future perspectives in crops, livestock, forestry and fish. Pg. 167-184. FAO, Rome.

Dzama K, Walter JP, Ruvuna F, Sanders JO, Chimonyo M 2001. Index selection of beef cattle for growth and milk production using computer simulation modelling. South African Journal of Animal Science 31: 65-75.

Edwards JD, McCouch SR 2007. Molecular markers for use in plant molecular breeding and germplasm evaluation. In: Guimarães EP, Ruane J, Scherf BD, Sonnino A, Dargie JD (eds.) Marker-assisted selection: Current status and future perspectives in crops, livestock, forestry and fish. Pg. 29-49. FAO, Rome.

Falconer DS, Mackay TFC 1996. Introduction to quantitative genetics. Longman Group Ltd, Harlow, Essex, England. 438 p.

Fernando RL, Grossman M 1989. Marker assisted selection using best linear unbiased prediction. Genetics Selection Evolution 21: 467-477.

Flach P 2012. Machine learning: The art and science of algorithms that make sense of data. Cambridge University Press. 409 p.

Fukuoka S, Inoue T, Miyao A, Monna L, Zhong HS, Sasaki T, Minobe Y 1994. Mapping of sequence-tagged sites in rice by single strand conformation polymorphism. DNA Research 1: 271-277.

Garrick D, Taylor J, Fernando R 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. Genetics Selection Evolution 41: 55.

Garrick DJ 1991. Best linear unbiased prediction for across-flock/year breeding values. Proceedings of the New Zealand Society of Animal Production 51: 411-416.

Garrick DJ, Blair HT, Clarke JN 2000. Sheep industry structure and genetic improvement. Proceedings of the New Zealand Society of Animal Production 60: 175-179.

Garrick DJ, Snell RG 2005. Emerging technologies for identifying superior dairy cows in New Zealand. New Zealand Veterinary Journal 53: 390-399.

George E, McCulloch R 1993. Variable selection via gibbs sampling. Journal of the American Statistical Association 88: 881 - 889.

Georges M, Massey JM 1991. Velogenetics, or the synergistic use of marker assisted selection and germ-line manipulation. Theriogenology 35: 151-159.

Gianola D 2000. Statistics in animal breeding. Journal of the American Statistical Association 95: 296-299.

Gianola D, de los Campos G, Gonzáles-Recio O, Long N, Okut H, Rosa GJM, Weigel KA, Wu X-L 2010. Statistical learning methods for genome-based analysis of quantitative traits. Proceedings of the 9th World Congress on Genetics Applied to Livestock Production, Liepzig, Germany 1-6 August 2010. CD-ROM.

Gianola D, de los Campos G, Hill WG, Manfredi E, Fernando R 2009. Additive genetic variability and the bayesian alphabet. Genetics 183: 347-363.

Gianola D, Fernando RL, Stella A 2006. Genomic-assisted prediction of genetic value with semiparametric procedures. Genetics 173: 1761-1776.

Gianola D, Okut H, Weigel KA, Rosa GJM 2011. Predicting complex quantitative traits with bayesian neural networks: A case study with jersey cows and wheat. BMC Genetics 12: 87-100.

Gianola D, van Kaam JBCHM 2008. Reproducing kernel hilbert spaces regression methods for genomic assisted prediction of quantitative traits. Genetics 178: 2289-2303.

Gilles PN, Wu DJ, Foster CB, Dillon PJ, Chanock SJ 1999. Single nucleotide polymorphic discrimination by an electronic dot blot assay on semiconductor microchips. Nature Biotechnology 17: 365-370.

Goddard ME 2009. Genomic selection: Prediction of accuracy and maximisation of long term response. Genetica 136: 245-257.

Goddard ME 2012. Uses of genomics in livestock agriculture. Animal Production Science 52: 73-77.

Goddard ME, Hayes BJ, Meuwissen THE 2010. Genomic selection in farm animal species - lessons learnt and future perspectives. Proceedings of the 9th World

Congress on Genetics Applied to Livestock Production, Liepzig, Germany 1-6 August 2010. CD-ROM.

Gonzáles-Recio O, Weigel KA, Gianola D, Naya H, Rosa GJM 2010. L_2-boosting algorithm applied to high-dimensional problems in genomic selection. Genetics Research 92: 227-237.

Gonzalez-Recio O, Forni S 2011. Genome-wide prediction of discrete traits using bayesian regressions and machine learning. Genetics Selection Evolution 43: 7.

González Andrade F 2010. Forensic genetics research progress. Nova Science Publishers Inc., New York. 335 p.

Groen AF 2000. Breeding goal definition. In: Galal S, Boyazoglu J, Hammond K (eds.) Workshop on developing breeding strategies for lower input animal production environments. Pg. 25-104. ICAR Technical series N° 3, Bella (Italy).

Habier D, Fernando R, Kizilkaya K, Garrick D 2011. Extension of the bayesian alphabet for genomic selection. BMC Bioinformatics 12: 186.

Habier D, Fernando RL, Dekkers JCM 2007. The impact of genetic relationship information on genome-assisted breeding values. Genetics 177: 2389-2397.

Habier D, Tetens J, Seefried F-R, Lichtner P, Thaller G 2010. The impact of genetic relationship information on genomic breeding values in german holstein cattle. Genetics Selection Evolution 42: 5.

Hair JF, Black WC, Babin BJ, Anderson RE, Tatham RL 2006. Multivariate data analysis. 6th ed. Pearson Education, Upper Saddle River, New Jersey 07458, USA. 924 p.

Haley CS, Visscher PM 1998. Strategies to utilize marker-quantitative trait loci associations. Journal of Dairy Science 81: 85-97.

Harris BL, Winkelman AM, Johnson DL 2013. Impact of including a large number of female genotypes on genomic selection. Proceedings of the 2013 Interbull meeting, Nantes, France. 23 -25 August 2013.

Harris DL, Stewart TS, Arboleda CR 1984. Animal breeding programs: A systematic approach to their design, USDA-ARS Bull. AAT-NC-8. USDA-ARS, Peoria, IL.

Hayes BJ, Bowman PJ, Chamberlain AC, Verbyla K, Goddard ME 2009a. Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genetics Selection Evolution 41: 51.

Hayes BJ, Chamberlain AJ, Goddard ME 2006. Use of markers in linkage disequilibrium with QTL in breeding programs. Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, MG, Brasil. 13-18 August 2006.

Hayes BJ, Goddard ME 2001. The distribution of the effects of genes affecting quantitative traits in livestock. Genetics Selection Evolution 33: 209-229.

Hayes BJ, Goddard ME 2010. Genome-wide association and genomic selection in animal breeding. Genome / National Research Council Canada = Genome / Conseil national de recherches Canada 53: 876-883.

Hayes BJ, Pryce JE, Bowman PJ, Chamberlain AC, Goddard ME 2010a. Genetic architecture and accuracy of genomic prediction. Proceedings of the 9th World Congress on Genetics Applied to Livestock Production, Liepzig, Germany 1-6 August 2010. CD-ROM.

Hayes BJ, Pryce JE, Chamberlain AJ, Bowman PJ, Goddard ME 2010b. Genetic architecture of complex traits and accuracy of genomic prediction: Coat colour, milk-fat percentage, and type in holstein cattle as contrasting model traits. PLoS Genet. Vol. 6(9) doi: 10.1371/journal.pgen.1001139.

Hayes BJ, Visscher PM, Goddard ME 2009b. Increased accuracy of artificial selection by using the realized relationship matrix. Genetics Research 91: 47-60.

Hazel LN 1943. The genetic basis for constructing selection indexes. Genetics 28: 476-490.

Hazel LN, Lush JL 1942. The efficiency of three methods of selection. Journal of Heredity 33: 393-399.

Henderson CR 1963. Selection index and expected genetic advance. In: Hanson WD, Robinson HF (eds.) Statistical genetics and plant breeding. Pg. 141-163. National Academy of Sciences-National Research Council, Washington, D. C., U.S.A.

Henderson CR 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics 31: 423-447.

Henderson CR, Quaas RL 1976. Multiple trait evaluation using relatives' records. Journal of Animal Science 43: 1188-1197.

Higuchi M, Miyashita N, Awata T 1999. Rapid communication: A PCR-rflp in the coding region of the bovine tumor necrosis factor-alpha locus. Journal of Animal Science 77: 3400-3401.

Hillel J, Schaap T, Haberfeld A, Jeffreys AJ, Plotzky Y, Cahaner A, Lavi U 1990. DNA fingerprints applied to gene introgression in breeding programs. Genetics 124: 783-789.

Huang W 2014. Molecular genetics techniques and high throughput technologies. In: Khatib H (ed.) Molecular and quantitative animal genetics. Pg. 163-175. Wiley-Blackwell, Madison, WI, USA.

Huisman AE, Brown DJ 2008. Genetic parameters for bodyweight, wool, and disease resistance and reproduction traits in merino sheep. 2. Genetic relationships between bodyweight traits and other traits. Australian Journal of Experimental Agriculture 48: 1186-1193.

Huisman AE, Brown DJ, Ball AJ, Graser HU 2008. Genetic parameters for bodyweight, wool, and disease resistance and reproduction traits in merino sheep. 1. Description of traits, model comparison, variance components and their ratios. Australian Journal of Experimental Agriculture 48: 1177-1185.

Humphries SE, Gudnason V, Whittall R, Day INM, Techniques ftIFoCCSDCoMB 1997. Single-strand conformation polymorphism analysis with high throughput modifications, and its use in mutation detection in familial hypercholesterolemia. Clinical Chemistry 43: 427-435.

Ihaka R, Gentleman R 1996. R: A language for data analysis and graphics. Journal of Computational and Graphical Statistics 5: 299-314.

Janssen-Tapken U, Dekkers JCM, Kadarmideen HN 2010. Genetic analyses of cattle breeding schemes including genetic markers for trypanotolerance. Proceedings of the 9th World Congress on Genetics Applied to Livestock Production, Liepzig, Germany 1-6 August 2010. CD-ROM.

Johnson DL, Thompson R 1995. Restricted maximum likelihood estimation of variance components for univariate animal models using sparse matrix techniques and average information. Journal of Dairy Science 78: 449-456.

Kemp RA, Kennedy BW, Wilton JW 1986. The effect of positive assortative mating on genetic parameters in a simulated beef cattle population. TAG Theoretical and Applied Genetics 72: 76-79.

Kenyon PR, Morel PCH, Morris ST, West DM 2007. Effect of the age of rams on reproductive performance of ewe hoggets. New Zealand Veterinary Journal 55: 184-187.

Kenyon PR, Proctor L, Morel PCH, Morris ST, West DM 2008. The effect of breeding ewe lambs on subsequent two-year-old ewe performance. Livestock Science 115: 206-210.

Lande R, Thompson R 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124: 743-756.

Lee Y, Nelder JA 2006. Double hierarchical generalized linear models (with discussion). Journal of the Royal Statistical Society: Series C (Applied Statistics) 55: 139-185.

Long N, Gianola D, Rosa GJM, Weigel KA, Avendaño S 2007. Machine learning classification procedure for selecting snps in genomic selection: Application to early mortality in broilers. Journal of Animal Breeding and Genetics 124: 377-389.

Long N, Gianola D, Rosa GM, Weigel K 2011. Application of support vector regression to genome-assisted prediction of quantitative traits. Theoretical and Applied Genetics 123: 1065-1074.

Lopez-Villalobos N, Garrick D 2005. Methodology for the design and enhancement of genetic improvement programs illustrated in the context of the New Zealand dairy industry. Agrociencia IX: 553-568.

Luan T, Woolliams JA, Lien S, Kent M, Svendsen M, Meuwissen TH 2009. The accuracy of genomic selection in norwegian red cattle assessed by cross-validation. Genetics 183: 1119-1126.

Lush JL 1943. The genetic basis of variation. In: Animal breeding plans. Pg. 71-85. The Iowa State College Press.

MacLeod IM, Hayes BJ, Goddard ME 2013. Will sequence SNP data improve the accuracy of genomic prediction in the prescence of long term selection? Procedings of the Association for the Advancement of Animal Breeding and Genetics 20: 215-219.

Meuwissen T, Hayes B, Goddard M 2013. Accelerating improvement of livestock with genomic selection. Annual Review of Animal Biosciences 1: 221-237.

Meuwissen THE 2003. Genomic selection: The future of marker assisted selection and animal breeding. Proceedings of FAO international workshop "Marker assisted selection: A fast track to increase genetic gain in plant and animal breeding?", Turin, Italy 17-18 October 2003. Pg. 54-59.

Meuwissen THE 2007. Genomic selection: Marker assisted selection on a genome wide scale. Journal of Animal Breeding and Genetics 124: 321-322.

Meuwissen THE, Goddard ME 1996. The use of marker haplotypes in animal breeding schemes. Genetics Selection Evolution 28: 161-176.

Meuwissen THE, Hayes BJ, Goddard ME 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819-1829.

Misztal I, Legarra A, Aguilar I 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. Journal of Dairy Science 92: 4648-4655.

Moghaddar N, Swan AA, Van der Werf JHJ 2013. Accuracy of genomic prediction from multi-breed sheep reference population. Procedings of the Association for the Advancement of Animal Breeding and Genetics 20: 220-223.

Montaldo HH, Meza-Herrera CA 1998. Use of molecular markers and major genes in the genetic improvement of livestock. Electronic Journal of Biotechnology [online] 1: 83-89.

Moore DS, McCabe G 1990. Introduction to the practice of statistics. 1st ed. W.H.Freeman & Co Ltd, New York, USA. 790 p.

Moser G, Tier B, Crump RE, Khatkar MS, Raadsma HW 2009. A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers. Genetics Selection Evolution 41: 56.

Mrode R 2006. Linear models for the prediction of animal breeding values. 2nd ed. CABI, Oxfordshire, UK. 368 p.

Mueller UG, Wolfenbarger LL 1999. Aflp genotyping and fingerprinting. Trends in Ecology & Evolution 14: 389-394.

Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H 1986. Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. Cold Spring Harbor Symposia On Quantitative Biology 51 Pt 1: 263-273.

Mulvaney FJ, Morris ST, Kenyon PR, Morel PCH, West DM 2010. Effect of nutrition pre-breeding and during pregnancy on breeding performance of ewe lambs. Animal Production Science 50: 953-960.

Nejati-Javaremi A, Izadi F, Rahimi G, Moradi M 2007. Selection in fat-tailed sheep based on two traits of fat-tail and body weight versus single-trait total body weight. International Journal of Agriculture & Biology 9: 645–648.

Nicol AM, Brookes IM 2007. The metabolisable energy requirements of grazing livestock. In: Rattray PV, Brookes IM, Nicol AM (eds.) Pasture and supplements for grazing animals N^{o} 14. Pg. 151-172. New Zealand Society of Animal Production (Inc.), Hamilton, New Zealand.

Nielsen R 2000. Estimation of population parameters and recombination rates from single nucleotide polymorphisms. Genetics 154: 931-942.

Notter DR 2000. Effects of ewe age and season of lambing on prolificacy in US Targhee, Suffolk, and Polypay sheep. Small Ruminant Research 38: 1-7.

Okut H, Gianola D, Rosa GJM, Weigel KA 2011. Prediction of body mass index in mice using dense molecular markers and a regularized neural network. Genetics Research 93: 189-201.

Okut H, Wu X-L, Rosa G, Bauck S, Woodward B, Schnabel R, Taylor J, Gianola D 2013. Predicting expected progeny difference for marbling score in angus cattle using artificial neural networks and bayesian regression models. Genetics Selection Evolution 45: 34.

Olson KM, Vanraden PM, Tooker ME, Cooper TA 2011. Differences among methods to validate genomic evaluations for dairy cattle. Journal of Dairy Science 94: 2613-2620.

Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T 1989a. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. Proceedings of the National Academy of Sciences 86: 2766-2770.

Orita M, Suzuki Y, Sekiya T, Hayashi K 1989b. Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. Genomics 5: 874-879.

Park T, Casella G 2008. The bayesian lasso. Journal of the American Statistical Association 103: 681-686.

Pérez-Rodríguez P, Gianola D, Weigel KA, Rosa GJM, Crossa J 2013. Technical note: An R package for fitting bayesian regularized neural networks with applications in animal breeding. Journal of Animal Science 91: 3522-3531.

Pickering NK, Dodds KG, Auvray B, McEwan JC 2013. The impact of genomic selection on genetic gain in the New Zealand sheep dual purpose selection index. Proceedings of the Association for the Advancement of Animal Breeding and Genetics No. 20. p 175-178, Napier, New Zealand.

Pickering NK, Dodds KG, Blair HT, Hickson RE, Johnson PL, McEwan JC 2012. Genetic parameters for production traits in New Zealand dual-purpose sheep, with an emphasis on dagginess. Journal of Animal Science 90: 1411-1420.

Pomar C, Harris DL, Savoie P, Minvielle F 1991. Computer simulation model of swine production systems: I.I.I. A dynamic herd simulation model including reproduction. Journal of Animal Science 69: 2822-2836.

Pryce JE, Goddard ME, Hayes BJ 2010. Breeding schemes for dairy cows under genomic selection – what can we do? Proceedings of the 9th World Congress on Genetics Applied to Livestock Production, Liepzig, Germany 1-6 August 2010. CD-ROM.

Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES 2001. Linkage disequilibrium in the human genome. Nature 411: 199-204.

Reingruber MC, Gregory WW 1994. The data modeling handbook: A best-practice approach to building quality data models. Wiley-QED, New York, USA. 362 p.

Rendel JM, Robertson A 1950. Estimation of genetic gain in milk yield by selection in a closed herd of dairy cattle. Journal of Genetics 50: 1-8.

Ron M, Weller JI 2007. From QTL to qtn identification in livestock-winning by points rather than knock-out: A review. Animal genetics 38: 429-439.

Ruane J, Sonnino A 2007. Marker-assisted selection as a tool for genetic improvement of crops, livestock, forestry and fish in developing countries: An overview of the

issues. In: Guimarães EP, Ruane J, Scherf BD, Sonnino A, Dargie JD (eds.) Marker-assisted selection: Current status and future perspectives in crops, livestock, forestry and fish. Pg. 3-13. FAO, Rome.

Saatchi M, Garrick DJ 2013. Accuracies of genomic predictions in US beef cattle. Procedings of the Association for the Advancement of Animal Breeding and Genetics 20: 207-210.

Saatchi M, Ward J, Garrick DJ 2013. Accuracies of direct genomic breeding values in hereford beef cattle using national or international training populations. Journal of Animal Science 91: 1538-1551.

SAS Institute Inc. 2011. Sas 9.3. User's guide; windows version, Cary, NC. USA.

Schaeffer LR 2006. Strategy for applying genome-wide selection in dairy cattle. Journal of Animal Breeding and Genetics 123: 218-223.

Searle SR 2006. Matrix algebra useful for statistics. 1st ed. John Wiley & Sons, Inc, Hoboken, N.J. USA. xxii, 438 p.

Shen X, Ronnegard L, Carlborg O 2011. Hierarchical likelihood opens a new way of estimating genetic values using genome-wide dense marker maps. BMC Proceedings 5: S14.

Sise JA, Amer PR 2009. SNP predictors to accelerate the rate of genetic progress in sheep. Proceedings of the Association for the Advancement of Animal Breeding and Genetics 18: 200-223.

Sise JA, Auvray B, Dodds KG, Amer PR 2008. SNiP and cut: Quantifying the potential benefits of genomic selection tools for genetic fault elimination in sheep. Proceedings of the New Zealand Society of Animal Production 68: 33-36.

Slack-Smith A, Kinghorn BP, Hickey JM, Van der Werf JHJ 2010. The accuracy of genomic selection in predicting carcass traits in meat sheep. Proceedings of the 9th World Congress on Genetics Applied to Livestock Production, Liepzig, Germany. 1-6 August 2010. CD-ROM.

Smulders JP, Serrano M, Pérez-Guzmán MD, Jimenez MA, Uribe H, Jurado JJ 2007. Stochastic simulation of manchega sheep breed selection scheme. Impact of artificial insemination, progeny testing system and nucleus size on genetic progress and inbreeding. Livestock Science 106: 218-231.

Solberg T, Sonesson A, Woolliams J, Meuwissen T 2009. Reducing dimensionality for prediction of genome-wide breeding values. Genetics Selection Evolution 41: 29.

Solberg TR, Sonesson AK, Woolliams JA, Meuwissen THE 2007. Using partial least square regression (plsr) and principal component regression (PCR) in prediction of genome wide breeding values. Proceedings of the 58th Annual Conference of the European Association for Animal Production, Dublin, Ireland 26.-29 August 2007, http://www.eaap.org/Previous_Annual_Meetings/2007Dublin/Papers/S18_03_Solberg.pdf. [accessed 23 November 2009].

Sölkner J, Tier B, Crump R, Moser G, Thompson P, Raadsma HW 2007. A comparison of different regression methods for genomic-assisted prediction of genetic values in dairy cattle. Proceedings of the 58th Annual Conference of the European Association for Animal Production, Dublin, Ireland 26.-29 August 2007, http://www.eaap.org/Previous_Annual_Meetings/2007Dublin/Papers/S18_02_Solkner.pdf. [accessed 23 November 2009].

Sonnino A, Carena MJ, Guimarães EP, Baumung R, Pilling D, Rischkowsky B 2007. An assessment of the use of molecular markers in developing countries. In: Guimarães EP, Ruane J, Scherf BD, Sonnino A, Dargie JD (eds.) Marker-assisted selection: Current status and future perspectives in crops, livestock, forestry and fish. Pg. 16-26. FAO, Rome.

Sörensen MK, Berg P, Jensen J, Christensen LG 1999. Stochastic simulation of breeding schemes for total merit in dairy cattle. Interbull Bulletin 23: 183-191.

Southern EM 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. Journal of Molecular Biology 98: 503-517.

Spelman R, Arias J, Keehan M, Obolonkin V, Winkelman A, Johnson D, Harris B 2012. Application of genomic selection in the New Zealand dairy cattle industry.

Proceedings of the 38th ICAR Session, Cork, Ireland 28 May - 1 June 2012, http://www.icar.org/Cork_2012/Manuscripts/Published/Spelman.pdf [accessed 20 November 2012].

Spiegel MR, Heredero RH 1991. Estadística. 2nd ed. McGraw Hill, Guadalajara, Mexico. 556 p.

Statistics New Zealand 2012. Total New Zealand breeding ewes 2 tooth and over put to ram. http://www.stats.govt.nz/infoshare/Default.aspx [accessed 10 November 2012].

Sumner RMW, Davison RM 2006. Economic implications for growing wool from using genomic technologies. Proceedings of the New Zealand Society of Animal Production 66: 149-154.

Swan AA, Brown DJ 2013. The impact of measuring adult fleece traits with genomic selection on economic gain in merino selection indexes. Procedings of the Association for the Advancement of Animal Breeding and Genetics 20: 233-236.

Swan AA, Johnston DJ, Brown DJ, Tier B, Graser H-U 2012. Integration of genomic information into beef cattle and sheep genetic evaluations in australia. Animal Production Science 52: 126-132.

The International HapMap Consortium 2008. What is the hapmap? http://hapmap.ncbi.nlm.nih.gov/whatishapmap.html [accessed September 2008].

Thomasen JR, Egger-Danner C, Willam A, Guldbrandtsen B, Lund MS, Sørensen AC 2014. Genomic selection strategies in a small dairy cattle population evaluated for genetic gain and profit. Journal of Dairy Science 97: 458-470.

Togashi K, Lin CY 2010. Theoretical efficiency of multiple-trait quantitative trait loci-assisted selection. Journal of Animal Breeding and Genetics 127: 53-63.

Toosi A, Fernando RL, Dekkers JC 2010. Genomic selection in admixed and crossbred populations. Journal of Animal Science 88: 32-46.

Turner HN 1959. Ratios as criteria for selection in animal or plant breeding, with particular reference to efficiency of food conversion in sheep. Australian Journal of Agricultural Research 10: 565-580.

Vagenas D, Doeschl-Wilson A, Bishop SC, Kyriazakis I 2007. In silico exploration of the effects of host genotype and nutrition on the genetic parameters of lambs challenged with gastrointestinal parasites. International Journal of Parasitology 37: 1617-1630.

Van Arendonk JAM 1991. Use of profit equations to determine relative economic value of dairy cattle herd life and production from field data. Journal of Dairy Science 74: 1101-1107.

Van der Werf J 2000. Livestock staright-breeding system structures for the sustainable intensification of extensive grazing systems. In: Galal S, Boyazoglu J, Hammond K (eds.) Workshop on developing breeding strategies for lower input animal production environments. Pg. 105-177. ICAR Technical series N° 3, Bella (Italy).

Van der Werf J 2009. Potential benefit of genomic selection in sheep. Procedings of the Association for the Advancement of Animal Breeding and Genetics 18: 38-41.

Van Eenennaam AL, Weigel KA, Young AE, Cleveland MA, Dekkers JCM 2014. Applied animal genomics: Results from the field. Annual Review of Animal Biosciences 2: 105-139.

VanRaden PM 2008. Efficient methods to compute genomic predictions. Journal of Dairy Science 91: 4414-4423.

VanRaden PM, O'Connell JR, Wiggans GR, Weigel KA 2011. Genomic evaluations with many more genotypes. Genetics Selection Evolution 43: 10.

Vapnik VN 1998. Statistical learning theory. 1st. ed. John Wiley & Sons, INC., New York, USA. 736 p.

Verbyla K, Bowman P, Hayes B, Goddard M 2010. Sensitivity of genomic selection to using different prior distributions. BMC Proceedings 4: S5.

Verbyla K, Hayes B, Bowman P, Goddard M 2009. Accuracy of genomic selection using stochastic search variable selection in australian holstein friesian dairy cattle. Genetics Research 91: 307 - 311.

Vignal A, Milan D, SanCristobal M, Eggen A 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Genetics Selection Evolution 34: 275-305.

Visscher PM, Medland SE, Ferreira MAR, Morley KI, Zhu G, Cornes BK, Montgomery GW, Martin NG 2006. Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. PLoS Genet 2: e41.

Vos P, Hogers R, Bleeker M, Reijans M, Lee Tvd, Hornes M, Friters A, Pot J, Paleman J, Kuiper M, Zabeau M 1995. Aflp: A new technique for DNA fingerprinting. Nucleic Acids Research 23: 4407-4414.

Wagenaar D, Van Arendonk J, Kramer M 1995. Selection index program (SIP). Wageningen Agricultural University.

Waldron L, Pintilie M, Tsao M-S, Shepherd FA, Huttenhower C, Jurisica I 2011. Optimized application of penalized regression methods to diverse genomic data. Bioinformatics 27: 3399-3406.

Wang G, Whittam TS, Berg CM, Berg DE 1993. Rapd (arbitrary primer) PCR is more sensitive than multilocus enzyme electrophoresis for distinguishing related bacterial strains. Nucleic Acids Research 21: 5930-5933.

Weigel KA, de los Campos G, Vazquez AI, Van Tassell CP, Rosa GJM, Gianola D, O'Connell JR, VanRaden PM, Wiggans GR 2010. Genomic selection and its effects on dairy cattle breeding programs. Proceedings of the 9th World Congress on Genetics Applied to Livestock Production, Liepzig, Germany 1-6 August 2010. CD-ROM.

Weller JI, Ron M 2011. Invited review: Quantitative trait nucleotide determination in the era of genomic selection. Journal of Dairy Science 94: 1082-1090.

Welsh J, McClelland M 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Research 18: 7213-7218.

Wiggans GR, Vanraden PM, Cooper TA 2011. The genomic evaluation system in the united states: Past, present, future. Journal of Dairy Science 94: 3202-3211.

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research 18: 6531-6535.

Williams JL 2005. The use of marker-assisted selection in animal breeding and biotechnology. Revue Scientifique et Technique de l'OIE 24: 379-392.

Wimmer V, Lehermeier C, Albrecht T, Auinger H-J, Wang Y, Schön C-C 2013. Genome-wide prediction of traits with different genetic architecture through efficient variable selection. Genetics 195: 573-587.

Wolc A, Stricker C, Arango J, Settar P, Fulton J, O'Sullivan N, Preisinger R, Habier D, Fernando R, Garrick D, Lamont S, Dekkers J 2011. Breeding value prediction for production traits in layer chickens using pedigree or genomic relationships in a reduced animal model. Genetics Selection Evolution 43: 5.

Wolfová M, Wolf J, Krupová Z, Kica J 2009. Estimation of economic values for traits of dairy sheep: I. Model development. Journal of Dairy Science 92: 2183-2194.

Yao J, Aggrey SE, Zadworny D, Hayes JF, Kühnlein U 1996. Sequence variations in the bovine growth hormone gene characterized by single-strand conformation polymorphism (sscp) analysis and their association with milk production traits in holsteins. Genetics 144: 1809-1816.

Zehner R, Zimmermann S, Mebs D 1998. Rflp and sequence analysis of the cytochrome b gene of selected animals and man: Methodology and forensic application. International Journal Of Legal Medicine 111: 323-327.

Zhang W, Smith C 1992. Computer simulation of marker-assisted selection utilizing linkage disequilibrium. TAG Theoretical and Applied Genetics 83: 813-820.

Zhang Z, van Parijs FD, Xiao B 2014. The status of aflp in the genomics era and a pipeline for converting aflps into single-locus markers. Molecular Breeding 34: 1245-1260.

Zou H, Hastie T 2005. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67: 301-320.