Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

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

ABSTRACT i

DECLARATION iii

ACKNOWLEDGEMENTS v

CONTENTS vii

LIST OF TABLES xi

LIST OF FIGURES xvii

LIST OF ABBREVIATIONS AND ACRONYMS xxi

CHAPTER 1

General introduction 1

CHAPTER 2

A review of genomic selection for production traits 7

2.1. Introduction 9

2.1.1. Current selection 9

2.1.2. Opportunities for selection schemes 10

2.1.3. Molecular genetic markers 11

2.1.3.1. RFLP 12

2.1.3.2. PCR-RFLP 12

2.1.3.3. RAPD 13

2.1.3.4. AFLP 13

2.1.3.5. SSCP 14

2.1.3.6. Microsatellite 14

2.1.3.7. SNP 15

2.2. Selection schemes using genomic information 16

2.3. Prediction of genomic values 18

2.3.1. Least squares procedure 18

2.3.1.1. Numerical example illustrating the use of the least squares procedure to estimate genomic breeding value 19

2.3.2. Best linear unbiased prediction procedure 21

2.3.2.1. SNPBLUP and GBLUP 21

2.3.2.2. Numerical example illustrating the use of the best linear unbiased predictor procedure to estimate genomic breeding values 22
2.3.3. Bayesian Procedures 24
2.3.3.1. The bayesian alphabet 25
2.3.4. Statistical learning methods for estimation of genomic values 27
2.3.5. Multi-step vs. Single-step approaches for genomic evaluations 30
2.4. Concluding observations 30
2.5. Appendix to chapter 2 33

CHAPTER 3

Deterministic and stochastic simulation of a breeding programme for a nucleus flock based on individual selection 37
3.1. Abstract 39
3.2. Introduction 39
3.3. Materials and methods 42
3.3.1. Deterministic model 42
3.3.1.1. Breeding objective 43
3.3.1.2. Selection index 44
3.3.1.3. Breeding scheme 45
3.3.1.4. Correlated responses for individual traits 46
3.3.1.5. Contribution of traits in the selection index 47
3.3.2. Stochastic simulation 47
3.3.2.1. Flock structure 49
3.3.2.2. Genetic evaluation 50
3.3.2.3. Data generation structure 52
3.4. Results 54
3.4.1. Deterministic simulation 54
3.4.2. Stochastic simulation 55
3.5. Discussion 57

CHAPTER 4

Deterministic simulation analysis using selection index theory for sheep genomic selection 61
4.1. Abstract 63

4.2. Introduction 63

4.3. Materials and methods 65

4.3.1. Genomic selection index conceptualisation 66

4.3.2. Selection scenarios 69

4.3.2.1. Scenario 1: Different selection indices including genomic information 70

4.3.2.2. Scenario 2: Using genomic information to change age at selection 74

4.4. Results 77

4.4.1. Scenario 1 (Typical flock age structure) 77

4.4.2. Scenario 2 (Using ram and ewe lambs as parents) 82

4.5. Discussion 84

4.6. Conclusions 87

4.7. Appendix to chapter 4 88

CHAPTER 5

Stochastic simulation model for sheep breeding schemes using genomic selection and multitrait total merit index 95

5.1. Abstract 97

5.2. Introduction 97

5.3. Materials and methods 100

5.3.1. Flock structure 101

5.3.2. Breeding scheme 102

5.3.2.1. Best linear unbiased prediction selection 103

5.3.2.2. Genomic best linear unbiased prediction selection 105

5.3.2.3. Accuracy of genomic breeding values 106

5.3.2.4. Hardy-Weinberg equilibrium 107

5.3.3. Data generation 107

5.3.3.1. Phenotypic information 107

5.3.3.2. SNP generation 108

5.3.4. Data generation structure 110

5.4. Results 112

5.5. Discussion 116

CHAPTER 6

Accuracy of prediction of genomic breeding values for lamb carcass weight using simulation 119

6.1. Abstract 121

6.2. Introduction 121

6.3. Materials and methods 122

6.3.1. Estimation of genomic breeding values 123

6.3.2. Accuracy of GBV 125

6.4. Results 125

6.5. Discussion 128

6.6. Conclusions 131

6.7. Appendix to chapter 6 132

CHAPTER 7

General Discussion 135

7.1. Introduction 137

7.2. Effect of including genomic selection in a sheep breeding programme 137

7.3. Limitations and considerations of the simulation programmes 139

7.4. Possible new applications of the model 141

7.4.1. Economical analysis of breeding programmes 141

7.4.2. Accuracies of different statistical methods to estimate 142

7.4.3. Number of SNPs or animals needed to maximise genetic gain 142

7.4.4. Analysis of other breeds and other traits 144

7.5. Conclusions 144

7.5.1. Simulation techniques for genetic evaluation. 144

7.5.2. Analysis of genomic selection 145

7.5.3. Genomic breeding values accuracies 145

Addendum 147

References 151

LIST OF TABLES

Table **Title** **Page**

Table A2.1. Simulated single nucleotide polymorphisms associated with weaning weight (WW) for a training population of 20 animals. 33

Table A2.2. Simulated single nucleotide polimorphisms associated with weaning weight (WW) for a predicted population of 20 animals. 34

Table A2.3. Genomic breeding values (GBV) for weaning weight estimated using least squares methods. 35

Table A2.4. Genomic breeding values (GBV) for weaning weight is sheep estimated using best linear unbiased prediction procedures. 36

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. 44

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. 44

Table 3.3. Simulated ewes and rams pathways of selection. 45

Table 3.4. Age structure for ewes and rams over two years old, of a stochastically simulated sheep breeding flock. 50

Table 3.5. Birth rank percentages for lambs born in a stochastically simulated sheep breeding flock. 50

Table 3.6. Regression coefficients (b-values) contribution of traits in the selection index and correlated responses for deterministically simulated traits[1]. 54

Table 3.7. Animal genetic responses for deterministic and stochastic (mean ($\overline{x}_{\Delta G}$) and standard error (se)) simulations. 56

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. 69

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. 70

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. 73

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. 74

Table 4.5. Age structure for ewes over two years, old of a deterministically simulated sheep nucleus. 74

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 (% σ_G^2). 75

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. 76

Table 4.8. Age structures of the two nucleus flocks. 76

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. 83

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. 88

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. 88

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. 88

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. 89

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. 89

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. 90

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. 91

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. 92

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. 93

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. 93

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. 101

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. 101

Table 5.3. Age structure of ewes over two years old, of a stochastically simulated sheep nucleus flock. 102

Table 5.4. Birth rank percentages for lambs born in a stochastically simulated sheep breeding flock. 102

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). 115

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. 132

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. 132

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. 133

LIST OF FIGURES

Table	Title	Page
Figure 2.1.	True breeding values (TBV) and genomic estimated breeding values (GBV) for simulated sheep weaning weights using least squares.	20
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).	23
Figure 3.1.	Simulated sheep nucleus breeding scheme structure.	45
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).	53
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.	55
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.	57
Figure 4.1.	Two pathways of selection breeding scheme structure for the simulated sheep nucleus.	74

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. 78

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. 79

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}). 80

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}). 82

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. 84

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). 103

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). 110

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). 111

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. 113

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. 114

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. 115

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). 126

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. 127

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. 128

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). 138

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)