THE EFFECT OF LINKAGE AND GENETIC GROUPING ON THE ACCURACY OF ACROSS-FLOCK GENETIC EVALUATION IN AUSTRALIAN MERINO SHEEP

BY

Mohammad Khusro

A thesis submitted for the degree of Doctor of Philosophy of the University of New England, Armidale, Australia

MAY 2007

Animal Genetics and Breeding Unit
# Table of Contents

CONTENTS .............................................................................................................. i

Preface ................................................................................................................... v

Acknowledgements ................................................................. vi

Abbreviations .............................................................................................. viii

List of Tables ............................................................................................... x

List of Figures .............................................................................................. xiii

Abstract ............................................................................................................ xvi

Chapter 1. General Introduction ................................................................. 1

Chapter 2. Literature Review ........................................................................ 7

2.1 The concept of genetic grouping ................................................................. 7

2.2 Theoretical aspects of genetic grouping ....................................................... 10

2.2.1 Criteria for defining genetic groups in a model .................................. 10

2.2.2 Sire model with genetic grouping ......................................................... 12

2.2.3 Animal model with genetic grouping ..................................................... 13

2.2.4 Fixed versus random genetic group effects ........................................... 17

2.2.5 Estimation of breeding values with maternal genetic groups ............... 22

2.2.5.1 Differential genetic grouping for direct additive and maternal genetic
effects ........................................................................................................... 23

2.2.6 Incorporating genetic groups in a multi-trait animal model evaluation..... 25

2.3 Considerations regarding estimation of genetic group effects .................... 28

2.3.1 Connectedness in across-flock evaluation ............................................ 28

2.3.1.1 Measures of connectedness ........................................................... 29

2.3.1.2 Factors affecting the accuracy of across-flock genetic evaluation ...... 37
2.3.2 Estimability of genetic group effects using mixed model methodology ........................................... 39
2.4 Application of genetic grouping ......................................................................................................... 41
  2.4.1 Effect of genetic grouping on estimates of variance components ........................................... 41
  2.4.2 Effect of genetic grouping on the accuracy of genetic evaluation ........................................ 45
  2.4.3 Inclusion of genetic groups in models for estimating QTL effects ...................................... 53
  2.4.4 Effect of genetic grouping and deregression on international evaluation ....................... 55
  2.4.5 Impact of genetic grouping on the accuracy of across-flock Merino sire evaluation in Australia ............................................................ 59
2.5 Alternative Bayesian approaches to genetic grouping for evaluation with uncertain parentage ....................................................................................................................... 69
2.6 Characteristics of industry data used for genetic evaluation of Merino sheep in Australia ........................................................................................................................................ 73
  2.6.1 OVIS software package ........................................................................................................... 73
  2.6.2 Sheep Genetics ....................................................................................................................... 74
2.7 Clustering methods and algorithms .................................................................................................. 76
  2.7.1 Application of cluster analysis in genetic evaluation ......................................................... 80
2.8 Conclusions ....................................................................................................................................... 82

Chapter 3. The effect of linkage and genetic grouping on the partitioning of genetic and environmental differences between flocks ................................................. 88
3.1 Introduction ..................................................................................................................................... 88
3.2 Materials and Methods ................................................................................................................... 89
3.3 Results and Discussion .................................................................................................................... 95
3.4 Conclusions and Implications ......................................................................................................... 119

Chapter 4. The effect of flock size on the accuracy of estimated genetic group effects and genetic group variance ............................................................... 124
4.1 Introduction ..................................................................................................................................... 124
4.2 Materials and Methods ................................................................................................................... 126
  4.2.1 Genetic analysis ....................................................................................................................... 128
Chapter 5. The effect of cluster size on the accuracy of clustering mechanism in single-trait genetic evaluation

5.1 Introduction .................................................................................................................. 148
5.2 Materials and Methods .............................................................................................. 149
  5.2.1 Data simulation ........................................................................................................ 149
  5.2.2 Genetic analysis I (Non-clustering) ....................................................................... 150
  5.2.3 Clustering flocks .................................................................................................... 150
  5.2.4 Genetic analysis II (Clustering) ............................................................................ 152
5.3 Results and Discussion .............................................................................................. 153
5.4 Conclusions and Implications .................................................................................... 164

Chapter 6. Multiple-trait animal model with differential genetic grouping ........ 167
6.1 Introduction .................................................................................................................. 167
6.2 The animal model and mixed model equations ....................................................... 168
  6.2.1 Differential genetic grouping via Q-matrix .............................................................. 169
  6.2.2 Differential genetic grouping via W-matrix ............................................................ 170
6.3 Numerical example ..................................................................................................... 173
  6.3.1 Q-matrix approach ............................................................................................... 173
  6.3.2 W-matrix approach .............................................................................................. 178
6.4 Discussion .................................................................................................................... 181
6.5 Conclusions and Implications .................................................................................... 182

Chapter 7. Conventional versus differential genetic grouping in across-flock multiple-trait evaluation .............................................................. 184
7.1 Introduction .................................................................................................................. 184
7.2 Materials and Methods .............................................................................................. 185
  7.2.1 Data ...................................................................................................................... 185
7.2.2 Data analysis ........................................................................................................... 187
7.3 Results and Discussion ............................................................................................. 192
  7.3.1 Complete dataset .................................................................................................. 192
  7.3.1.1 Regression analysis using complete dataset ..................................................... 196
  7.3.2 Reduced dataset .................................................................................................. 199
  7.3.2.1 Regression analysis using reduced dataset ...................................................... 203
7.4 Conclusions and Implications .................................................................................... 211

Chapter 8. General Discussion ....................................................................................... 215

Chapter 9. General Conclusions .................................................................................... 225

BIBLIOGRAPHY ............................................................................................................... 228

Appendix. Standardization of estimated genetic group effects ...................................... 244
Preface

I certify that the substance of this thesis has not already been submitted for any degree and is not currently being submitted for any other degree or qualification.

I certify that any help received in preparing this thesis, and all sources used, have been acknowledged in this thesis.

(Mohammad Khusro)
Acknowledgements

First, my sincere thanks to Dr. Daniel Brown (Principal supervisor) and Dr. Hans Graser (Co-supervisor) for offering me the postgraduate scholarship from Meat & Livestock Australia (MLA) without which I could have never even imagined pursuing a doctoral program at an Australian university. I will always remain indebted to Dr. Graser and Dr. Brown for providing me casual employment at AGBU on several occasions to support myself financially. I gratefully acknowledge MLA for providing funding and support for this study. I really appreciate and admire the principles and discipline of Dr. Brown and Dr. Graser that have taught me so eloquently how to accomplish tasks and meet deadlines independently. I would also extend my heartfelt thanks to Dr. Graser for being so considerate, generous and understanding for providing me extra scholarship for making modifications in this thesis.

My whole-hearted thanks and gratitude to Dr. Brown who was always willing to help and guide especially in those initial stages when I was struggling with “FORTRAN” and “GAWK”. I am greatly indebted to him for his endless support, invaluable advice and encouragement. The idea for the work accomplished in Chapter 6 was stimulated by Dr. Brown. Also, my sincere thanks to Dr. Graser for making me understand the basics of mixed model equations and guiding so eloquently through all stages of the research undertaken in Chapter 6. Thanks are also due to Dr. Bruce Tier, Dr. Abe Huisman and Dr. Ron Crump for their enormous help in making me understand some of the theoretical aspects of quantitative genetics and programming in FORTRAN. I am grateful to Dr. Gilbert Jeyaruban for his valuable discussions on statistical tests used in Chapter 4 and for the time he had so kindly spent in going through this thesis, and also for providing valuable feedback and critical comments.

I thank all the staff at AGBU especially Marlene Youman for their kind assistance that has added to the completion of this thesis, and my colleagues, Alex Woolaston and Rob Jones for sharing those laughing moments especially those focusing on the intake of caffeine in coke and coffee to stay awake during post-lunch sessions. I would also like
to thank the three anonymous examiners for their valuable comments and feedback which has enhanced the overall quality and standard of this thesis.

The arrival of Aamna and Afeefa, my lovely daughters, were very special and memorable moments. They have been a constant source of joy and happiness all these years. My heartfelt thanks to Farzana, my beautiful, loving and understanding wife, who along with Aamna and Afeefa, sacrificed a lot. I will always and forever remain grateful to my loving and supporting parents, Dr. Farreed Ali Shamsi and Mrs. Naseem Fatma Fareed, who sacrificed so much all these years for my coming to Australia. I wholeheartedly thank my parents for being so patient all these years and for remembering me in their daily prayers along with my brother, Mohammad Bilal and sisters, Dr. Fatima Tabassum and Asra Fareed. I dedicate this thesis to my loving parents.

Finally, all praise is due to Almighty God for giving me strength, courage, patience, and perseverance all these years without which I would not have completed this thesis.
Abbreviations

**ABV:** Australian breeding value

**AI:** Artificial insemination

**AM:** Animal model

**ANRM:** Average numerator relationship matrix

**ASBV:** Australian sheep breeding value

**AWI:** Australian Wool Innovation Limited

**BLUP:** Best linear unbiased prediction

**BLUP-A:** Best linear unbiased prediction without genetic groups

**BLUP-G:** Best linear unbiased prediction with genetic groups

**BWT:** Body weight

**CD:** Coefficient of determination

**CG:** Contemporary group

**CTSE:** Central Test Sire Evaluation

**CR:** Connectedness rating

**EBV:** Estimated breeding value

**FBV:** Flock breeding value

**FD:** Fibre diameter

**GFW:** Greasy fleece weight

**GG:** Genetic group

**GLt:** Genetic links

**h²:** Heritability

**HIER:** Hierarchical
MACE: Multiple-trait across country evaluation
MAS: Marker-assisted selection
MLA: Meat and Livestock Australia
MME: Mixed model equations
MSE: Mean square error
NRM: Numerator relationship matrix
PEC: Prediction error correlation
PEV: Prediction error variance
PEVD: Prediction error variance of difference
QTL: Quantitative trait loci
$R^2$: Coefficient of determination
RAM: Reduced animal model
REML: Restricted maximum likelihood
SED: Standard error of difference
SD: Standard deviation
SE: Standard error
TBV: True breeding value
WEC: Worm egg count
List of Tables

Table 2.1. Comparison of national genetic evaluation results (X) with results from MACE applied to only mastitis in Sweden (I₁), with MACE results using genetic correlations set to zero (I₂) and with MACE results using estimated genetic correlations (I₃)∗........................................................................................................57

Table 2.2. The average PEV of EBV differences between sires (genetic variance units) for the medium wool schemes (1989-2000) with and without elite link schemes in 1996 and 2000 for no genetic groups, sire x year groups and bloodline groups. The average PEV of bloodline mean differences and the variance proportion compared with no elite link schemes (in brackets) are also presented∗........................................61

Table 2.3. Estimates for sire and sire x management group (sire.mg) components of variance (estimated as a fraction of residual variance) for models without and with genetic group effects∗ ........................................................................................................63

Table 2.4. Estimates of variances for sires, sire x dam source and error from models with varying numbers of genetic groups∗ ........................................................................................................64

Table 2.5. Effect of using sire variances from the full model on BLUP solutions from the incomplete models. Results are expressed as standard deviations (SD) among sires solutions and the regressions (and correlations, r) of sire solutions from the incomplete models versus the full model∗ ........................................................................66

Table 2.6. Estimates for additive genetic (σ²_a), genetic group (σ²_g), and residual (σ²_e) variance components, σ²_g/σ²_e (p²_g) and the ratio for Bayes factor (RBF) for yearling WEC (standard errors in parentheses) using different grouping strategies∗ ......68

Table 3.1. Differences between the two scenarios..................................................................................................................93

Table 3.2. Difference between the flock mean EBVs (micron) of WNsires at varying levels of Xsire and WFsire linkage for different models (Scenario I).................104

Table 3.3. Genetic group solutions (±SE; micron) at varying levels of Xsire and WFsire linkage (Scenario II)..................................................................................................................105

Table 3.4. Estimated flock effects expressed as deviations from Flock 1 (±SE; micron) at varying levels of Xsire linkage for different models (Scenario II)...............112

Table 4.1. Estimated genetic group variance (±SE; micron²) for different flock sizes at varying levels of linkage (26 genetic groups)..................................................132
Table 4.2. Variance of the difference (micron²) between simulated and estimated genetic group effects for different flock sizes at varying levels of linkage (26 genetic groups).................................................................................................................................................................................. 132
Table 4.3. Correlation between TBVs and EBVs for different flock sizes at varying levels of linkage (26 genetic groups) .................................................................................................................................................................................................................................................. 133
Table 4.4. Estimated genetic group variance (±SE; micron²) for different flock sizes at varying levels of linkage (51 genetic groups) .................................................................................................................................................................................................................. 136
Table 4.5. Variance of difference (micron²) between simulated and estimated genetic group effects for different flock sizes at varying levels of linkage (51 genetic groups) .................................................................................................................................................................................................................................................. 136
Table 4.6. Correlation between TBVs and EBVs for different flock sizes at varying levels of linkage (51 genetic groups) .................................................................................................................................................................................................................................................. 137
Table 4.7. Difference in the variance (of difference between simulated and estimated genetic group effects; micron²) for 26 versus 51 genetic groups for varying flock sizes and levels of linkage* .................................................................................................................................................................................................................................................. 138
Table 4.8. Difference in the variance (of difference; micron²) averaged over different levels of linkage for 51 genetic groups (above diagonal) and 26 genetic groups (below diagonal) for varying combinations of any two flock sizes¹,² ............................................... 138
Table 4.9. Difference in the variance (of difference; micron²) averaged over different flock sizes for 51 genetic groups (above diagonal) and 26 genetic groups (below diagonal) for varying combinations of two levels of linkage¹,² .................................................. 139
Table 4.10. Difference in correlation (26 versus 51 genetic groups) between TBVs and EBVs for varying flock sizes and levels of linkage* .................................................................................................................................................................................................................................................. 140
Table 4.11. Difference in correlation between TBVs and EBVs averaged over different levels of linkage for 51 genetic groups (above diagonal) and 26 genetic groups (below diagonal) for varying combinations of two flock sizes¹,² .................................................. 140
Table 4.12. Difference in correlation between TBVs and EBVs averaged over different flock sizes for 51 genetic groups (above diagonal) and 26 genetic groups (below diagonal) for varying combinations of two levels of linkage¹,² .................................................. 141
Table 6.1. Hypothetical dataset for illustrating differential genetic grouping.................. 173
Table 6.2. Pedigree illustrating Scenario 2 of differential genetic grouping.................. 176
Table 7.1. The numbers of progeny derived from common link sires for different combinations of flocks (values on diagonal are the total number of animals in each flock).......................................................... 186
Table 7.2. Summary statistics of Sheep Genetics data used for estimating breeding values for BWT and FD using conventional and differential genetic grouping ..... 187
Table 7.3. Summary statistics of CTSE data used for estimating regression coefficients (transformed as per Sheep Genetics evaluations) ......................................................... 191
Table 7.4. Cluster allocation using conventional and differential genetic grouping (complete dataset) ........................................................................................................ 194
Table 7.5. Regression coefficients (±SE) derived using all sires for BWT and FD using conventional and differential clustering and without clustering (complete dataset) ........................................................................................................ 196
Table 7.6. Regression coefficients (±SE) for BWT and FD estimated using different threshold of Sheep Genetics data for conventional (Con) and differential (Diff) clustering (complete dataset)........................................................................................................ 198
Table 7.7. Regression coefficients (±SE) derived using all sires for BWT and FD for conventional and differential clustering and without clustering (reduced dataset). 204
Table 7.8. Allocation of flocks to different clusters using conventional and differential genetic grouping (reduced dataset) ................................................................. 204
Table 7.9. Regression coefficients (±SE) for BWT and FD estimated using different threshold of Sheep Genetics data for conventional (Con) and differential (Diff) clustering (reduced dataset)........................................................................................................ 207
List of Figures

Figure 3.1. Diagrammatic illustration of the use of sires with different origin........... 90
Figure 3.2. Solutions for genetic group effects at different degrees of Xsire linkage (Scenario I)........................................................................................................... 96
Figure 3.3. Solutions for flock environmental effects for different models and different degrees of Xsire linkage (Scenario I)........................................................................ 96
Figure 3.4. Across-flock correlation between the TBVs and the EBVs at varying levels of Xsire and WFsire linkage (0 to 50%) for different models (Scenario I)........ 99
Figure 3.5. The variance of EBVs across all animals at varying levels of Xsire and WFsire linkage (0 to 50%) for different models (Scenario I).............................. 99
Figure 3.6. Within-flock (genetic group 1) correlation between the TBVs and the EBVs of all animals at varying levels of Xsire and WFsire linkage for different models (Scenario I)..................................................................................................................... 101
Figure 3.7. True breeding values versus estimated breeding values at 10% Xsire linkage and without genetic groups for a single replicate (Scenario I; dark and light colored points represent flocks 1 and 2, respectively)..................................................... 102
Figure 3.8. True breeding values versus estimated breeding values at 10% Xsire linkage and with genetic groups for a single replicate (Scenario I; dark and light colored points represent flocks 1 and 2, respectively)..................................................... 102
Figure 3.9. Difference in the means of genetic groups 1 and 2 and their associated SEDs at varying levels of Xsire and WFsire linkage (Scenario II)......................... 103
Figure 3.10. Across-genetic group correlation between TBVs and EBVs at varying levels of Xsire and WFsire linkage for different models (Scenario II)................. 108
Figure 3.11. Correlation between TBVs and EBVs within genetic group 1 at varying levels of Xsire and WFsire linkage for different models (Scenario II).............. 114
Figure 3.12. Difference in the mean EBVs of WNsires by flock at varying levels of Xsire linkage and without genetic groups (Scenario II)...................................... 115
Figure 3.13. Difference in the mean EBVs of WNsires by flock at varying levels of Xsire linkage and with genetic groups (Scenario II)........................................ 116
Figure 3.14. Standard errors of estimated genetic group effects (flocks 1 or 2) at varying levels of Xsire linkage and heritability (Scenario II)................................. 117
Figure 3.15. Average standard errors of estimated genetic group effects at varying levels
of Xsire linkage and progeny per sire (Scenario II) .................................................. 119
Figure 4.1. Coefficient of determination between simulated and estimated genetic group
effects for different flock sizes at varying levels of linkage (26 genetic groups) ... 131
Figure 4.2. Coefficient of determination between simulated and estimated genetic group
effects for different flock sizes at varying levels of linkage (51 genetic groups) ... 135
Figure 4.3. Simulated and estimated genetic group variance at different levels of
progeny per link sire (51 genetic groups) ................................................................. 142
Figure 4.4. Standard errors of estimated genetic group effects for varying flock size (for
single replicate at 50 progeny each from two Xsires) ........................................ 142
Figure 4.5. Coefficient of determination between simulated and estimated genetic group
effects for small, medium and large flocks at varying levels of progeny per link sire
........................................................................................................................................ 143
Figure 4.6. Standard errors of genetic group solutions for small, medium and large
flocks at varying levels of progeny per link sire .................................................... 144
Figure 5.1. Flow diagram of the methodology used for the estimation of genetic group
and cluster effects........................................................................................................ 153
Figure 5.2. Estimated between-cluster variance for different cluster sizes at varying
levels of progeny per link sire ($\sigma_{gg}^2 = 1.5$ micron$^2$) ..................................... 154
Figure 5.3. Standard errors of estimated between-cluster variance for different cluster
sizes at varying levels of progeny per link sire ....................................................... 156
Figure 5.4. Average linkage (averaged over numbers of flocks in each category) in each
category of small, medium and large flocks at varying levels of progeny per link sire
........................................................................................................................................ 157
Figure 5.5. Correlation between simulated flock mean TBVs and EBVs for varying
cluster sizes at different levels of progeny per link sire for small flocks ............ 158
Figure 5.6. Standard errors of estimated cluster effects for different cluster sizes for
small, medium and large flocks (25 progeny per link sire) ................................. 159
Figure 5.7. Standard errors of estimated cluster effects for small, medium and large
flocks for 10 and 25 clusters at varying levels of progeny per link sire .......... 160
Figure 5.8. Across-flock correlation between simulated flock mean TBVs and EBVs for
varying cluster sizes at different levels of progeny per link sire ..................... 162
Figure 5.9. Across-flock correlation between simulated TBVs and EBVs with estimated cluster effects for varying cluster sizes at different levels of progeny per link sire 163

Figure 5.10. Within-flock correlation between simulated TBVs and EBVs from genetic groups model and different clustering models at varying levels of progeny per sire .............................................................................................................................................. 163

Figure 7.1. Flow diagram depicting the methodology used for the comparison of EBVs from conventional and differential genetic grouping strategies for BWT and FD . 189

Figure 7.2. Scatter plot of estimated genetic group solutions (without standardization) for BWT and FD (numbers in bold and italics represent different genetic groups; complete dataset) ........................................................................................................................................ 194

Figure 7.3. Scatter plot of sire EBVs for BWT from conventional and differential genetic grouping (complete dataset) ........................................................................................................................................ 195

Figure 7.4. Scatter plot of sire EBVs for FD from conventional and differential genetic grouping (complete dataset) ........................................................................................................................................ 195

Figure 7.5. Scatter plot of estimated genetic group effects for BWT using complete and reduced datasets (bigger dots represent flocks from which BWT data were deleted) ........................................................................................................................................ 200

Figure 7.6. Scatter plot of estimated genetic group effects for FD using complete and reduced datasets ........................................................................................................................................ 201

Figure 7.7. Scatter plot of sire EBVs for BWT from conventional and differential genetic grouping (reduced dataset) ........................................................................................................................................ 202

Figure 7.8. Scatter plot of sire EBVs for FD from conventional and differential genetic grouping (reduced dataset) ........................................................................................................................................ 203

Figure 7.9. Scatter plot of sire EBVs for BWT from conventional and differential clustering models (sires having progeny in their own flocks only; complete dataset) ........................................................................................................................................ 209

Figure 7.10. Scatter plot of sire EBVs for BWT from conventional and differential clustering models (sires having progeny in one flock only; complete dataset) ...... 209

Figure 7.11. Scatter plot of sire EBVs for BWT from conventional and differential clustering models (sires having progeny in more than one flock; complete dataset) ........................................................................................................................................ 210
Abstract

The aim of this research was to characterize the significance of linkage and genetic grouping, and to develop better and alternative models of genetic grouping for across-flock evaluations in the Australian Merino sheep industry. Most livestock industry datasets used for genetic evaluation have missing pedigree and performance data in varying amounts. The implicit assumption underlying an animal model evaluation that all base animals have similar genetic merit with a common variance will not hold true in a majority of cases. Not accounting for previous selection on base animals will bias estimated breeding values (EBVs). The use of outside sires having different population means and with either incomplete or no pedigree information is a common practice which may affect the mean breeding value of animals in a flock. To account for the differences in the pedigree and data recording, and differences in the genetic means of base populations, genetic groups are generally included in the evaluation procedure. In Merino evaluations by Sheep Genetics in Australia, genetic groups are included in the model to account for differences in the genetic means of base populations of different flocks.

The basic theory and numerous aspects associated with genetic grouping from published literature were reviewed in Chapter 2. Genetic groups can be treated as either fixed or random effects depending on the data structure, flock size, heritability and genetic links across flocks. Treating genetic groups as random effects avoids potential singularity in the coefficient matrix of the mixed model equations leading to the estimation of pair-wise differences among different groups. Several criteria have been proposed by different researchers for defining genetic groups in the evaluation schemes. These include year of birth, flock or herd of origin, selection path, sex and generation interval. However, none of these criteria has universal acceptance and not all are appropriate for all circumstances. Formation of genetic groups on the basis of source of origin (e.g., flock) should be considered provided differences between groups are large. For traits influenced by maternal effects and having differences in selection history and data recording, differential genetic groups can be included in the model for direct additive and maternal genetic effects. Genetic groups can be included in the multi-trait animal model evaluation via Q-matrix and W-matrix approaches using mixed model
equations (MME). The $\mathbf{Q}$-matrix is a design matrix which describes the proportion of genes contributed by each genetic group to the animals of interest. The $\mathbf{W}$-matrix is similar to the conventional numerator relationship matrix ($\mathbf{A}$) except that it has genetic groups included in it as extra animals. Both approaches ($\mathbf{Q}$-matrix and $\mathbf{W}$-matrix) provide identical solutions.

Adequate genetic links between flocks and accounting for genetic groups in the evaluation model are a prerequisite for an unbiased across-flock evaluation in the Australian Merino sheep industry. Besides linkage, proportion of progeny derived from common parents, and genetic grouping, data structure is also equally important for an accurate estimation of genetic group and environmental effects in across-flock evaluations. The estimability of genetic and environmental differences between flocks is dependent on the groups having animals represented across different levels of other systematic fixed effects included in the model such as flock or herd, year, or season. In the absence of linkage (or connectedness) between flocks, genetic group and environmental effects will be confounded. Different measures of connectedness have been proposed in the literature. However, there is no measure of connectedness that has universal acceptance. Most of these measures are very demanding computationally and not suitable for large datasets.

Chapter 3 used simulation to focus on the issue of disentangling genetic and environmental effects on the phenotype in across-flock evaluations. The results obtained in this chapter have shown that the accurate partitioning of genetic and environmental differences between flocks can be achieved when at least 25% of the total progeny in each flock is generated from link sires and genetic groups are included in the evaluation model. The link sires used for establishing linkage should have adequate numbers of progeny records to accurately estimate their genetic group effects. However, the actual proportion of linkage required for disentangling genetic and environmental differences between flocks could vary depending on the magnitude of genetic differences between flocks. The accurate estimation of genetic group effects is also dependent on the heritability of the trait. For traits with low, moderate or high heritability, genetic group effects can be estimated when at least 25% linkage exists between flocks. The proportion of linkage required for the estimation of genetic group effects remains the
same irrespective of the level of heritability. However, the standard errors of estimated genetic group solutions are slightly lower (more accuracy) when the heritability is low and vice-versa. The impact of heritability on the accuracy of estimated genetic group solutions decreases when higher levels of genetic links (more than 50%) exist between flocks.

At the same degree of linkage, the average standard errors of estimated genetic group effects are marginally higher when more progeny are generated from fewer link sires compared to fewer progeny from more sires. The accuracy of estimated genetic group effects is determined to a greater extent by the total number of progeny generated from link sires than by the number of progeny per link sire. To derive genetic group solutions with an accuracy of 70%, the SED associated with the estimated difference in the means of two genetic groups needs to be no greater than one-half of the genetic standard deviation. If the level of accuracy desired in the estimated genetic group solutions is 80%, then the SED must be less than one-third of the genetic standard deviation.

The effect of flock and genetic group size (number of animals) on the accuracy of estimation of genetic group effects and genetic group variance was investigated using simulated data in Chapter 4. The accuracy of estimated genetic group effects and genetic group variance is dependent on the size of genetic groups and the proportion of linkage between genetic groups. The smaller the size of genetic groups, the lower is the accuracy of estimation and vice versa. For larger flocks, genetic group effects and genetic group variance are estimable with higher accuracy than for smaller flocks. The size of genetic groups formed should be sufficiently large for obtaining unbiased across-flock EBVs. It would be erroneous to make prior assumptions regarding distributions and covariance structure of genetic groups. The actual genetic group variance is expected to depend on the selection history, which for most livestock breeding datasets cannot be traced back to the unselected base population. Alternatively, genetic group variance could be estimated from industry data without making prior assumptions. For the data structure simulated in this study and at the same level of linkage, generating a fixed number of progeny (per link sire) from different numbers of link sires or varying
numbers of progeny (per link sire) from constant number of link sires does not alter the accuracy of prediction of genetic group effects and across-flock EBVs.

Chapter 5 examined the possibility of clustering flocks on the basis of their estimated genetic group effects so as to reduce the number of genetic groups included in the model. Genetic group effects were estimated from simulated data. If the size (number of animals) of a large number of flocks is too small to be treated as separate genetic groups, it is recommended to cluster such flocks on the basis of their estimated genetic group solutions. Clustering leads to the estimation of a smaller number of genetic group effects and hence improves their accuracy. The accuracy of prediction (correlation between true and estimated breeding values) after clustering is not significantly different from that when flocks are not clustered. For a given dataset, different numbers of clusters could be formed based on the estimated genetic group effects. The accuracy of clustering is dependent on the number of cluster effects that needs to be estimated, the number of genetic groups and the magnitude of their associated effects, and the degree of linkage between genetic groups. Clustering provides a solution to the problem of reduced precision of estimating genetic group effects for smaller flocks in across-flock evaluations. The technique of clustering flocks based on their estimated genetic group effects could be implemented in Sheep Genetics evaluations. Further research should focus on developing clustering algorithms that would take into consideration issues regarding the level of linkage between flocks and therefore the estimability of cluster effects prior to clustering flocks in across-flock evaluations.

The feasibility of allocating genetic groups differently for each trait (differential genetic grouping) using MME in a multiple-trait animal model evaluation was explored in Chapter 6. Differential genetic grouping can easily be implemented in the MME following certain modifications. Two different approaches ($Q$-matrix and $W$-matrix) of differential genetic grouping have been proposed for multi-trait evaluations. In the first approach ($Q$-matrix), differential genetic groups can be included as additional fixed effects in the model. The latter approach ($W$-matrix) includes genetic groups and possible phantom animals as extra animals in the numerator relationship matrix resulting in an augmented matrix $W$. The rules described can be used to modify $W^{-1}$.
$G_o^{1}$ block in the MME to account for differential genetic groups where $G_o$ is the genetic covariance matrix. Both methods provide identical solutions as the models are equivalent. This study considered only fictional data on two traits and two genetic groups for differential genetic grouping. The rules described for assigning groups differentially can be modified accordingly to accommodate more than two traits and two genetic groups. Differential genetic grouping can be implemented in OVIS for Sheep Genetics evaluations. It is recommended that the routine evaluations in OVIS should consider fitting genetic groups differentially for traits that are characterized by significant differences in selection history and data recording, for example, fleece weight and worm egg count (WEC).

Chapter 7 examined the variation in multi-trait EBVs obtained from a subset of Sheep Genetics data using conventional (same genetic group for all traits) and differential (different genetic groups for different traits) genetic grouping. The subset of data was chosen so as to include a large number of sires whose progeny performance was recorded in the CTSE (Central Test Sire Evaluation) database. Flocks were clustered on the basis of their estimated genetic group effects for body weight (BWT) and fibre diameter (FD). The sire EBVs derived using conventional and differential strategies of genetic grouping were validated by regressing sire EBVs on progeny records from CTSE. The results obtained show that when flocks are clustered using conventional strategy, the clustering mechanism is significantly influenced by the trait (BWT) with a large variance. This could lead to less than optimal clustering for the less variable trait(s). Therefore, in a multi-trait across-flock evaluation involving traits with large differences in their variance, estimated genetic group effects need to be standardized prior to clustering flocks. When complete data were used, the EBVs derived through conventional and differential genetic grouping were very similar for both traits. However, when incomplete (data on BWT deleted from some flocks) dataset was used, the accuracy of estimated genetic group effects for body weight was reduced which in turn affected the efficiency of clustering mechanism. Therefore, the results of conventional and differential clustering were different. The differences in these two strategies will get more evident depending on the amount of data available on different traits and the differences in their mean and variance. The level of similarity desired
between flocks for each trait(s) with regard to their estimated genetic group solutions will impact the pattern of clusters formed for that trait(s) which in turn may lead to differences between conventional and differential genetic grouping.

Several aspects characteristic of field data were not included in the experimental analyses undertaken in this thesis. These commonly include multiple and overlapping generations and differential grouping criteria for additive genetic and maternal genetic effects. The latter would depend on the amount of missing dam pedigree and possible differences in the number of records available on dams. Differential grouping may also be required for some traits characterized by differences in genetic trend for direct and maternal genetic effects. Further research could investigate how genetic grouping affects the accuracy of evaluation procedure when these factors are taken into consideration. All the experimental chapters except Chapters 6 and 7 used simulated data to examine the effect of linkage established through common sires used across all flocks and genetic grouping on the accuracy of across-flock evaluations. The minimum linkage required for accurately estimating the genetic group effects may vary if different sires are used for creating genetic links across flocks.

The ongoing improvement in the models of genetic evaluation will lead to further developments of current methods used for accounting for missing data and pedigree information. The recent advancements in molecular genetics techniques (for example, DNA finger printing and gene mapping) and simultaneous reduction in the costs associated with their widespread use in ascertaining parentage in different livestock industries, may in future reduce the need for grouping animals. However, some animals will always need to be grouped as base populations have varying means for traits of interest and pedigree can only be traced back for a few generations in farm animals.