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.

MULTIVARIATE ESTIMATION OF VARIANCE AND COVARIANCE COMPONENTS USING RESTRICTED MAXIMUM LIKELIHOOD, IN DAIRY CATTLE.

A thesis presented in partial fulfilment
of the requirements for the degree
of Doctor of Philosophy
in
Animal Science
at
Massey University

CARLOS FRANCISCO SOSA FERREYRA

1992

AKNOWLEDGEMENTS

I am especially indebted to my Chief Supervisor Professor Hugh T. Blair, for his kind and helpful advice. The time and effort that Professor Blair put into supervising my work were beyond the call of duty. My wife and I will always treasure our friendship with Dr Hugh and Alana Blair.

I would like to express my gratitude to Professor Robert D. Anderson for his guidance in my learning of mathematical statistics and estimation of variance components. He has made himself available at all times to provide friendly help. My family and I will always remember the lovely Christmas dinners we enjoyed in the company of Professor Anderson and his family.

Also I am indebted to Professor Alexander L. Rae for his suggestions during the early stages of my doctoral studies, and for the enlightening conversations we had.

I especially thank Dr Dorian J. Garrick for suggesting algorithms and computing strategies for my research, and for many conversations that helped me to understand REML theory. Dr Garrick has provided a friendly learning experience and the oportunity to further my knowledge in applied aspects of animal evaluation.

Many thanks to Professor Stuart N. McCutcheon for his concern on the welfare of the overseas students and their families, and for doing everything at his reach to help me and my family.

I also want to thank Dr Hugo Varela Alvarez for his important suggestions that improved the code of the computer programs for my research.

Special thanks to my fellow graduate students, in particular Alison Darragh, Meing Jiao Shi, Russel Knutson, Laura Jacobson, Chris Butts, Chis Roberts, Mariusz Skorupsky, Tatsushi Inoue and Raymundo Rangel, and to the secretaries of the Department of Animal Science, who provided a friendly environment that I'll always remember.

I would like to thank the New Zealand Ministry of External Relations and Trade for the scholarship that allowed me to pursue higher education at Massey University, and to the

Instituto Nacional de Investigaciones Forestales, Agricolas y Pecuarias de Mexico for providing additional funding that allowed my family to join me in New Zealand.

My appreciation goes to Gisela Ahlborn and to Dr Brian Wickham of the Livestock Improvement Corporation for providing the data for my research.

My parents Ana Maria Ferreyra and Gregorio Sosa showed me the path to follow and gave all the support and encourangement that I could need. To them my deep gratitude and admiration.

All that I have done would be impossible without the help, support and patience of my wife Alba Patricia Rodriguez, who has been at my side in the difficult times and shared the happy times with me. To her I give my deepest gratitude. The same patience was shown by my stepchildren Alba and Nelson, who along with their mother, shared with me the rewarding experience of living in New Zealand.

ABSTRACT

The multivariate estimation of sire additive and residual variances and covariances by Restricted Maximum Likelihood (REML) is addressed. Particular emphasis is given to its application to dairy cattle data when all traits are explained by the same model and no observations are missing. Special attention is given to the analysis of new traits being included in a sire evaluation programme, for which a model has to be developed and no previous estimates of the population parameters exist.

Results obtained by using either the multivariate Method 3 of Henderson, multivariate REML excluding the Numerator Relationship Matrix (NRM) or by multivariate REML including the NRM were compared. When a large number of traits were fitted simultaneously the variance-covariance matrix estimated by Method 3 was negative-definite (outside the allowable parameter space). REML estimates obtained while ignoring the NRM were biased. The number and sequence of traits fitted in the analysis affected the estimates at convergence. A canonical transformation of the variance-covariance matrix was undertaken to simplify the computation by means of an Expectation Maximisation (EM) algorithm.

Approaches to choosing initial values for their use in iterative methods were compared via their values at convergence and the number of iterations required to converge.

To further simplify the use of multivariate REML, three transformations of the Mixed Model Equations (MME) were integrated: the absorption of proven sire effects taken as fixed, a triangular factorisation of the NRM, and the singular value decomposition of the coefficient matrix in the MME. One statistical algorithm (EM) and one mathematical algorithm (Scoring

type) were developed to iteratively solve the REML equations on the transformed scale, such that the transformed coefficient matrix of the MME did not need to be inverted at each iteration and the required quantities to build the REML equations were obtained through vector operations.

Traits other than Production (TOP) from New Zealand Holstein-Friesian dairy cows were analysed (4 management and 13 conformation characteristics), each trait scored using a linear scale from 1 to 9, with extreme values corresponding to extreme phenotypes. Mixed model methodology was used for the analysis of TOP as no significant departure from normality was observed. To model the TOP, the fixed effects of herd, inspector, age, stage of lactation (linear and quadratic) and breed of dam were tested for significance. Only the effects of inspector and herd were significant for all traits, with breed of dam significantly affecting adaptability to milking, shed temperament and stature.

Estimates of phenotypic means and standard deviations, and heritabilities for TOP were: adaptability to milking 5.4 ± 1.7, 0.20; shed temperament 5.5 ± 1.6, 0.12; milking speed 5.7 ± 1.5, 0.11; farmer's overall opinion 5.7 ± 1.7,.14; stature 5.1 ± 1.0, 0.14; weight 4.4 ± 1.0, 0.37; capacity 5.3 ± 1.0, 0.40; rump angle 5.4 ± 0.7, 0.16; rump width 5.2 ± 0.7, 0.08; legs 5.2 ± 0.6, 0.34; udder support 5.3 ± 1.0, 0.63; fore udder 4.9 ± 1.1, 0.48; rear udder 4.9 ± 1.0, 0.33; front teat placement 4.2 ± 0.7, 0.22; rear teat placement 5.2 ± 0.8, 0.22; udder overall 4.8 ± 1.1, 0.42; and dairy conformation 5.3 ± 1.1, 0.32.

Large positive phenotypic correlations among management traits were obtained, while the correlations of these traits with type were small and positive when significant. Large and positive correlations among udder traits were found. All traits related to size were positively correlated amongst themselves. Most of the traits were positively correlated with dairy conformation.

Estimated genetic correlations for stature and weight with other conformation traits were generally negative. With the exception of udder support, all udder traits were positively correlated amongst themselves. Dairy conformation was positively correlated with most

traits, except with stature, rump angle, legs, rear udder and udder overall.

The estimates obtained in this study shold be used in the evaluation of Holstein-Friesian sires and cows for TOP in New Zealand.

TABLE OF CONTENTS

Abstract.....i
List of Tables.....vii
List of Figures.....ix
General Introduction.....x

CHAPTER I:
MULTITRAIT ANALYSIS OF LINEAR SCORES OF MANAGEMENT AND CONFORMATION CHARACTERISTICS IN NEW ZEALAND HOLSTEIN-FRIESIAN HEIFERS FOR THE ESTIMATION OF THEIR COMPONENTS OF VARIANCE.....1

1. INTRODUCTION.....2
   1.1 Conformation and Dairy Cattle Breeding.....2
   1.2 Dairy Cattle Breeding and Traits Other than Production in New Zealand.....8
   1.3 Genetic and Environmental Factors Affecting Conformation.....9
   1.4 Relationship Among Management and Conformation Traits.....13
   1.5 Analysis of Type and Management Data.....18
2. OBJECTIVE.....21
3. MATERIALS AND METHODS.....21
4. RESULTS AND DISCUSSION.....26
   4.1 Means, Standard Deviations and Coefficients of Variation.....26
   4.2 Factors Affecting TOP.....30
   4.3 Phenotypic Correlations.....34
   4.4 Mixed Models and the Estimation of Variances and Covariances.....35
5. BIBLIOGRAPHY.....42

CHAPTER II:
MULTIVARIATE ESTIMATION OF VARIANCE AND COVARIANCE COMPONENTS OF TRAITS OTHER THAN PRODUCTION, IN NEW ZEALAND HOLSTEIN-FRIESIAN HEIFERS.....47

1. INTRODUCTION.....48
   1.1 Multitrait Selection and Estmation of Covariance Components.....48

1.2 Appropriateness of Model and REML....50
1.3 Multitrait Estimation of Variance Components....53
1.4 The Performance of Multitrait vs Single Trait Analysis....57
1.5 Selection Intensity and Realationships of Dairy Cattle in New Zealand....58
2. OBJECTIVE....59
3. MATERIALS AND METHODS....59
4. RESULTS AND DISCUSSION....63
4.1 Relationships....63
4.2 Variance and Covariance Components for Management Traits....64
4.2.1 Additive Relationships Between Young Sires Ignored....64
4.2.2 Additive Relationships Considered....64
4.3 Variance and Covariance Components for Linear Conformation....69
4.3.1 Relationships Ignored....69
4.3.2 Relationships Included....69
4.4 The Effect of Initial Values on an EM Algrithm....77
4.4.1 The Effect of Initial Values on Estimates....77
4.4.2 Effect on the Convergence Rate of Algorithms....79
4.5 Comparison of Estimates with Findings in Other Studies....80
4.5.1 Management Traits....80
4.5.2 Conformation Traits....81
4.6 The use of Estimated Parameters....83
5. CONCLUSIONS....85
6. BIBLIOGRAPHY....88

CHAPTER III:

COMPARISON OF TWO ALGORITHMS FOR THE ESTIMATION OF VARIANCE AND COVARIANCE COMPONENTS OF TRAITS WITH EQUAL DESIGN MATRICES....94

1. INTRODUCTION....95
1.1 The Multivariate Mixed Model....96
1.2 Estimation by REML....100
1.3 General Properties of the Described Methods....106
1.4 Simplification of Computations for Multivariate Models....107
1.5 Inclusion of the Numerator Relationship Matrix....113
1.6 Including Records of Daughters of Selected Sires....116
1.7 Objective....117
2. MATERIALS AND METHODS....117
3. RESULTS AND DISCUSSION....119

4. CONCLUSIONS.....126
5. BIBLIOGRAPHY.....127

APPENDIX I:
CLASSIFICATION TYPE IN DAIRY COWS.....132

APPENDIX II:
CANONICAL TRANSFORMATION OF THE VARIANCE-COVARIANCE MATRIX.....136

APPENDIX III:
MULTIVARIATE ESTIMATES OF GENETIC PARAMETERS FOR TRAITS OTHER THAN PRODUCTION IN NEW ZEALAND DAIRY COWS.....140

LIST OF TABLES

CHAPTER I:

Table I. Scales and fixed effects in various analyses in linear type and management traits....11

Table II. Heritabilities for linear type and management traits for Holstein-Friesian cows in the literature and their mean....12

Table III. Ranges for phenotypic correlations between linear type traits reported in the literature....16

Table IV. Ranges for genetic correlations between linear type traits reported in the literature....19

Table V. List of traits, abbreviations and phenotypic values corresponding to extreme scores....23

Table VI. Distribution of TOP records by herd and sire....28

Table VII. Unadjusted means, standard deviations and coefficients of variation for traits other than production in New Zealand Holstein-Friesian heifers, and from the study by Meyer et al.(1987)....29

Table VIII. Percentage of variation explained by the linear model, and phenotypic correlations between management and conformation traits in New Zealand Holstein-Friesian heifers....36

Table IX. Phenotypic correlations between conformation traits in New Zealand Holstein-Friesian heifers....37

Table X. Residual and sire variance components and heritabilities for management traits estimated by multivariate Method 3 and REML in New Zealand Holstein-Friesian heifers....40

Table XI. Genetic correlations for management traits in New Zealand Holstein-Friesian heifers, computed from multivariate Method 3 and REML sire covariance components....40

Table XII. Heritabilities and genetic correlations for linear type traits in New Zealand Holstein-Friesian heifers....41

CHAPTER II:

Table I. REML estimates of heritabilities and genetic correlations for linear management traits, from bivariate, trivariate and multivariate analysis, using a model including the NRM.....68

Table II. Multitrait REML estimates of residual and sire variance and covariance components by assuming sires are unrelated.....71

Table III. Multitrait REML estimates of residual and sire variance and covariance components by including the additive relationship between sires.....72

Table IV. Heritabilities and genetic correlations for linear type traits obtained when the additive relationship matrix is ignored and included in the analysis.....74

CHAPTER III:

Table I. Sets of initial values utilised in the multivariate REML estimation of variance components for stature, udder support and dairy conformation of New Zealand Ayrshire cattle.....123

Table II. REML estimates of variances, covariances, heritabilities and genetic correlations for stature, udder support, and dairy conformation, in New Zealand Ayrshire cattle.....123

APPENDICES:

Appendix I.

Table A1. Ranges of heritabilities for descriptive type and management traits for three breeds of dairy cows.....134

Appendix III.

Table 1. Phenotypic means and standard deviations, heritabilities, genetic and phenotypic correlations for Traits Other Than Production in New Zealand dairy heifers.....143

LIST OF FIGURES

CHAPTER II.

Figure 1. Estimate of sire variance component for shed temperament, from two initial values.....66

Figure 2. Estimate of residual variance component for shed temperament, from two initial values.....67

CHAPTER I.

Figure 1. Rate of convergence of estimate of sire variance component for STA of New Zealand Ayrshire cows, by Method of Scoring and EM algorithms.....124

Figure 2. Rate of convergence of estimate of sire covariance component between STA and US of Ayrshire, by a Method of Scoring and an EM algorithms.....125

GENERAL INTRODUCTION

When undertaking the task of finding estimates of heritabilities and correlations that are to be used in animal selection programmes, both theoretical and practical aspects have to be considered. These deal mainly with the properties of the sample, the model chosen to describe the sampling procedure, the method of computation of the estimates, and the properties of the estimates.

The biology of animals and its interaction with farming practices will impact on the statistical properties of the measurements made on them. Domestic animal species differ in their reproductive physiology and in the length of their life-cycle; that is, they reproduce at different rates and have different generation intervals. The modification of the reproductive rate through artificial breeding, and embryo cloning and transfer may enhance or eliminate some of the natural differences. Some traits are restricted to animals of one sex (eg. ovulation rate), to a given age (eg. weaning weight) or condition (eg. days in calf), these factors affect the kind of selection programme that can be implemented. They also affect

Ideally, all available information should be used in the estimation of population parameters but, this is seldom the case as all records are not always available or validated. Unbalanced size of fixed and random classes is a common finding in field data, and not uncommon in experimental data, imposing restrictions on the statistical methodology that can be used for the study. The methods available frequently involve difficult and tedious computations.

The interpretation of the estimates derived will be limited by the assumptions that were made during the development of the statistical model and by the distribution of the measurements, the underlying traits and unexplained variation.

Often when new traits are considered in a selection programme no previous estimates of the population parameters exist, and sometimes the distribution properties that these traits follow may also be unknown. Thus basic descriptive statistics and tests of goodness of fit are required before preceeding to the modelling of the traits. If several traits are included simultaneously, a multivariate approach is required.

In the present work these problems are addressed with particular emphasis on traits observed in dairy cows. The case studied involves the development of a model by testing fixed and random effects, as well as some covariables, before a method to estimate variance components is chosen. The selection of statistical and computational methods is studied, and practical and theoretical aspects are reviewed.

This thesis consists of three self-contained chapters that deal in sequence with the descriptive study of a new set of traits in a selection programme, their modelling and, the statistical and computational aspects of estimation of their components of variance and covariance, under a multivariate model. A set of management and conformation traits of New Zealand dairy cows is used as an example throughout the study.

In Chapter I an historical review of the measurement, use and evaluation of traits other than production in dairy cows is undertaken. A descriptive study of a set of such traits in New Zealand dairy cows is undertaken and an appropiate model for the multivariate estimation of

their variance and covariance components is developed. Restricted Maximum Likelihood (REML) estimation of parameters for traits with equal design matrices under a multivariate model is applied, and the resulting estimates are compared with those obtained by a method of fitting constants.

Further development of a multivariate model is described in Chapter II, in which the importance of including the Numerator Relationship Matrix (NRM), when different from the identity matrix, is shown. A computational strategy for the application of an Expectation Maximisation (EM) algorithm, after a transformation of the vector of observations, is described. The effect of initial values on the rate of convergence of an EM algorithm is studied. The estimates obtained are interpreted and compared with those previously reported in the literature, and a set of recommendations given for the use of the estimates.

Although the desirable properties of REML make it the method of estimation of choice, the number of computations required, and its relative lack of availability in commercial statistical packages, have been limiting factors for its application. The use of a canonical transformation on multivariate models, where traits are affected by the same fixed and random effects, considerably simplifies the computations required by reducing the size of the coefficient matrix of the mixed model equations that is to be inverted. However, the use of slow converging algorithms, that require the inversion of the reduced coefficient matrix at each round of iteration, still represents a large computational effort. Faster algorithms have the disadvantage that an even greater computational effort is needed at each round of iteration. In Chapter III an algorithm that further simplifies the computations is developed, such that the use of EM or Method of Scoring-like algorithms are simpler and faster. This exploits the properties of the singular value decomposition of the coefficient matrix, which allows the inclusion of the NRM. The intention of this section is to make the use of REML more appealing to researchers in the field of animal breeding.