

THE UNIVERSITY OF ADELAIDE

Faculty of Science

School of Agriculture, Food and Wine

**Incorporating Pedigree Information
into the Analysis of
Agricultural Genetic Trials**

Helena Oakey

Doctor of Philosophy

May 2008

Contents

- 1 Introduction** **1**
 - 1.1 A new approach to the analysis of agricultural genetic trials 13

- 2 Measures of Relatedness** **18**
 - 2.1 Genes, alleles, genotypes and genetic effects 19
 - 2.2 Identity Modes 23
 - 2.3 Coefficient of Coancestry 26
 - 2.4 Coefficient of Inbreeding 27
 - 2.5 Special Case of the coefficient of Coancestry 28
 - 2.6 The genetic variance and covariance under inbreeding and Mendelian sampling 29
 - 2.6.1 Genetic Variance 29
 - 2.6.2 Genetic Covariance 34
 - 2.7 Full Variance-Covariance matrix 36
 - 2.8 Additive Relationship Matrix 37

CONTENTS

2.8.1	Adjustment for self-fertilization	38
2.8.2	The coefficient of parentage matrix-adjustment for self-fertilization .	41
2.9	Dominance relationship matrix	42
2.9.1	Gamete allocation	44
2.9.2	Forming ancestral gamete pairs	44
2.9.3	Determining the dominance relationship between gamete pairs . . .	48
2.9.4	Diagonal elements of $\mathbf{M3}$	50
2.9.5	Adjustment for Self-fertilization $\mathbf{M3}$	51
2.9.6	Updating the rules (Section 2.9.3) that determine the dominance relationship between gamete pairs	53
2.9.7	Updating the rules (Sections 2.9.1 and 2.9.2) that form the ancestral gamete pairs	55
2.10	Special Case: The dominance relationship matrix under no inbreeding . . .	59
2.11	A new method for calculating the dominance relationship matrix under no inbreeding	60
2.11.1	Gamete Allocation	62
2.11.2	The probability of the inheritance of gametes	62
2.11.3	Calculating dominance relationships	64
2.12	Inverse of the Relationship Matrices	67
2.12.1	Inverse of the Additive Relationship Matrix	67

3	Modern approaches for the analysis of field trials	71
3.1	<i>Standard</i> Statistical Model	72
3.1.1	Models for the non-genetic effects	73
3.1.2	Models for the genetic line means	75
3.2	Extending the <i>Standard</i> Statistical model	80
3.3	Fitting the dominance genetic effect d	83
3.3.1	Determination of the family pedigree	94
3.3.2	Forming gamete pairs	94
3.3.3	Determining the dominance relationship between gamete pairs	94
3.3.4	The dominance genetic effect assuming no inbreeding	96
3.3.5	Determination of the family pedigree	96
3.3.6	Gamete allocation and the probability of gamete inheritance	96
3.3.7	Calculating between and within dominance relationships	97
3.4	Estimation and Fitting	98
3.5	Selection indices	100
3.6	Heritability generalized	102
4	Analysis of Wheat Breeding Trials	114
4.1	Trial details	115
4.2	Single Site Analysis	118
4.2.1	Statistical Model	120

4.2.2	Analysis	122
4.3	Multi-site analysis	130
4.3.1	Statistical Model	130
4.3.2	Analysis	131
5	Analysis of Sugarcane Breeding Trials	144
5.1	Trial Details	145
5.2	Statistical Model	147
5.3	Analysis	149
5.4	Comparison of the results with the analysis presented by Oakey et al. (2007)	163
6	Model performance under simulation	165
6.1	Method	166
6.1.1	Data Models	167
6.2	Analysis Models	171
6.2.1	Indicators of the Performance of the Analysis Models	172
6.3	Results	174
6.3.1	REML estimation of variance components	174
6.3.2	Bias of REML estimation	175
6.3.3	Performance of Analysis Models	177
6.3.4	Total Genetic Effect	177

CONTENTS

6.3.5	Additive Genetic Effect	178
6.3.6	Partially-replicated design verses replicated design	179
6.3.7	Conclusion	179
7	Discussion and Conclusions	186
	Appendix A - Functions written in R code	198
A.1	Creating the additive relationship matrix with adjustment for inbreeding .	198
A.2	Simulation code to generate data models	203
A.2.1	R code to Run simulations	206
	Appendix B - ASReml code	221
B.1	ASReml code for fitting the <i>Extended</i> model in the wheat example (single site)	221
B.2	ASReml code for the final MET <i>Extended</i> model in the wheat example . .	226
B.3	ASReml code for the final MET <i>Extended</i> model in the sugarcane example	231
B.4	ASReml code for fitting the Analysis models	236

List of Tables

2.1 Summary of the mutually exhaustive and exclusive events that cover the possible likeness and non likeness of the alleles α_{jY} and α_{jZ} of individual j and alleles α_{kU} and α_{kV} of individual k respectively at locus l 25

2.2 Summary of the $E(g_{jl}|I_x)$ of individual j 30

2.3 Summary of the $E(g_{jl}g_{kl}|I_x)$ of individual j and k respectively 35

2.4 Pedigree of Example 43

2.5 Gamete Allocation to Pedigree of Example (Table 2.4) 44

2.6 Gamete Inheritance 45

2.7 All possible Gamete Pairs 47

2.8 Pedigree and Gamete Allocation 56

2.9 All possible Gamete Pairs 57

2.10 Inheritance of Base Gamete 62

2.11 Table of probabilities for the base gamete 64

3.1 Summary of the variance models for \mathbf{G}_e 79

LIST OF TABLES

3.2	Family Pedigree of Example (Table 2.4)	87
3.3	Gamete Allocation to Family Pedigree of Table 3.2	87
4.1	Details of the wheat example trials ^a	116
4.2	Tests of significance for improvement in the prediction of yield (kg/ha) resulting from the <i>Standard</i> verses <i>Extended</i> model and the average prediction error variance of the total genetic effect (\mathbf{g}_t) for the <i>Standard</i> and the <i>Extended</i> model.	123
4.3	Environmental terms fitted in the <i>Extended</i> model of the analysis of yield for each of the trials.	124
4.4	The Total or overall genetic variance of yield (kg/ha) for lines with pedigree information ($\sigma_{g_t}^2$) and lines without pedigree ($\sigma_{h_t}^2$) at each of the trials from the <i>Standard</i> and <i>Extended</i> models and broad (H^2) and narrow (h^2) sense heritability ^b	125
4.5	The correlations between the E-BLUPs of \mathbf{g}_t from the <i>Standard</i> model and the E-BLUPs $\mathbf{g}_t = \mathbf{a}_t + \mathbf{i}_t$ and \mathbf{a}_t respectively from <i>Extended</i> model	127
4.6	Summary of models fitted showing the structure of the trial genetic variance matrices \mathbf{G}_a , \mathbf{G}_i and \mathbf{G}_h for each of the genetic line effects \mathbf{a} , \mathbf{i} and \mathbf{h} respectively.	133

LIST OF TABLES

4.7 REML estimate of the components of the additive and epistatic genetic variance matrices^a for yield (kg/ha) at each trials, in the final *Extended* model (Model 8, Table 4.6) 136

4.8 Summary of the REML estimates of the total genetic variance and percent additive and epistatic variance in yield (t/ha) for lines with pedigree information at the final model (Model 8, Table 4.6). 138

5.1 Summary of the design layout and other details of the sugar example subtrials. 146

5.2 Non-genetic terms (excluding blocking terms^b) used in the MET analysis of the sugar example. 150

5.3 Summary of models fitted showing the structure of the trial genetic variance matrix for each of the genetic components. 152

5.4 REML estimate of the components of the additive, dominance and residual non-additive genetic variance matrices^a for CCS% at each trial in the final model (Model 11, Table 5.3) 154

5.5 Summary of the REML estimates of the total genetic variance and percent additive, dominance and epistatic variance in CCS for the final model (Model 11, Table 5.3, page 152) 156

6.1 Summary of the data models showing the additive variance as a percentage of the total genetic variance and the genetic variance as a percentage of the total variance 170

6.2 Summary of the three analysis models for the random vector \mathbf{g} the genetic effect of lines 171

6.3 Summary of y and x used in the calculation of the mean square error of prediction (Eqn. 6.2.3) and the relative response to selection (Eqn. 6.2.4) . 173

6.4 Summary of the proportion of REML estimates where either σ_a^2 or σ_i^2 were zero and thus not present in the *Extended* model. 174

6.5 Summary of the true and estimated variance components $\sigma_a^2, \sigma_i^2, \sigma_g^2$ and the percentage of genetic variance under the *Extended* models for the 9 data models (Table 6.1) in each of the partially replicated and replicated designs. 181

6.6 Summary of the ^amean square error of prediction for the total genetic effect^b under *Extended* analysis model in the partially replicated and replicated designs for the nine data models (Table 6.1). 182

6.7 Summary of the ^arelative response for the total genetic effect^b under the *Extended* model in the partially replicated and replicated designs for the nine data models (Table 6.1) 183

6.8 Summary of the ^amean square error of prediction for the additive genetic effect under the *Extended* model in the partially replicated and replicated designs for the nine data models (Table 6.1). 184

LIST OF TABLES

6.9 Summary of the ^arelative response for the additive genetic effect under the *Extended* model in the partially replicated and replicated designs for the nine data models (Table 6.1). 185

List of Figures

- 2.1 Example Pedigree. 43
- 4.1 The predicted (breeding value) yield (kg/ha) under the *Extended* model and the *Standard* model for lines with pedigree information. 128
- 4.2 The *additive* predicted (breeding value) yield (kg/ha) for the *Extended* model plotted against the predicted yield (kg/ha) of the *Standard* model for lines with pedigree information. 129
- 4.3 A bi-plot of the loadings of the first factor against the loadings of the second factor for the additive genetic line effect (\mathbf{a}). 137
- 4.4 The predicted *total* selection index of the *Standard* model (Model 4, Table 4.6) plotted against the predicted *total* selection index of yield (kg/ha) for the final model (Model 8, Table 4.6) 141
- 4.5 The predicted *total* selection index of the *Standard* model (Model 4, Table 4.6) plotted against the predicted *additive* genetic effects (breeding values) of yield (kg/ha) for the final model (Model 8, Table 4.6). 142

LIST OF FIGURES

4.6 The predicted *total* selection index of Model 1 (Table 4.6) plotted against the predicted *total* selection index of yield (kg/ha) for the final model (Model 8, Table 4.6). 143

5.1 A bi-plot of the loadings of the first factor against the loadings of the second factor for the additive genetic line effect \mathbf{a} 155

5.2 The predicted *dominance between family* selection index plotted against the predicted *dominance with family line* selection index of CCS for the final model (Model 11, Table 5.3). 159

5.3 The predicted *additive* selection index (breeding value index) plotted against the predicted *dominance* selection index of CCS for the final model (Model 11, Table 5.3). 160

5.4 The predicted *total* selection index of the *Standard* model (Model 2, Table 5.3) plotted against the predicted *total* selection index of CCS for the final model (Model 11, Table 5.3). 161

5.5 The predicted *total* selection index of the *Standard* model (Model 2, Table 5.3) plotted against the predicted *additive* genetic effects (breeding values) of CCS for the final model (Model 11, Table 5.3). 162

6.1 The additive relationship matrix used to simulate the data. 169

Abstract

This thesis presents a statistical approach which incorporates pedigree information in the form of relationship matrices into the analysis of standard agricultural genetic trials, where elite lines are tested. Allowing for the varying levels of inbreeding of the lines which occur in these types of trials, the approach involves the partitioning of the genetic effect of lines into additive genetic effects and non-additive genetic effects. The current methodology for creating relationship matrices is developed and in particular an approach to create the dominance matrix under full inbreeding in a more efficient manner is presented. A new method for creating the dominance matrix assuming no inbreeding is also presented.

The application of the approach to the single site analyses of wheat breeding trials is shown. The wheat lines evaluated in these trials are inbred lines so that the total genetic effect of each of the lines is partitioned into an additive genetic effect and an epistatic genetic effect. Multi-environment trial analysis is also explored through the application of the approach to a sugarcane breeding trial. The sugarcane lines are hybrids and therefore the total genetic effect of each hybrid is partitioned into an additive genetic effect, a heterozygous dominance genetic effect and a residual non-additive genetic effect. Finally, the approach for inbred lines is examined in a simulations study where the levels of heritability and the genetic variation as a proportion of total trial variation is explored in single site analyses.

Declaration

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being made available in all forms of media, now or hereafter known.

Acknowledgements

I would like to thank my supervisors, the three wise men, (in reverse alphabetical order) Ari Verbyla, Wayne Pitchford and Brian Cullis.

Thanks to Ari, for his statistical expertise and wisdom and his willingness to give this unstintingly – you have given me something to aspire to. Also thanks to him for his flexibility and understanding of my other job as a mother with two children. For his kindness and patience throughout the 10 years I have known him, it has been a pleasure working with you.

Thanks to Wayne for keeping our meetings on track. Your genetic expertise and experience in animal breeding was an asset that helped guide the research along the path it has finally taken. Wayne thanks for always having a smile and a positive slant even when things weren't going according to plan (which seemed to be often!).

Thanks to Brian (aka Brain) for his statistical expertise, suggestions and support with almost everything, but especially with getting the models fitted and the simulations. For his excellent critique of the PhD chapters and the papers and for managing to keep me on my toes at all times despite being in the next state.

Many thanks to Arthur Gilmour without his programming of the adaptation of the de Boer & Hoeschele (1993) method for creating the dominance matrices, the analysis of the sugarcane data would not have been possible. His quick replies to my ASReml queries throughout the duration of my PhD were also a great help.

My thanks to the Grains Research Development Council for providing the scholarship that made this PhD possible, I hope that the research present herewith has some practical benefits.

Finally, to my husband Shaun I owe my heartfelt gratitude for his understanding and encouragement throughout the trials and tribulations of the PhD journey. Without his steadfast support this journey could never have run the course to completion.

To my children Aberdeen and Jolyon, whom I adore, this PhD is dedicated to you both – may you always have the opportunity to follow your dreams.