

**QUANTITATIVE GENETIC AND GENOMIC MODELING OF FEED EFFICIENCY IN  
DAIRY CATTLE**

By

Yongfang Lu

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Animal Science — Doctor of Philosophy

2016

## ABSTRACT

### QUANTITATIVE GENETIC AND GENOMIC MODELING OF FEED EFFICIENCY IN DAIRY CATTLE

By

Yongfang Lu

Residual feed efficiency (**rFE**) is an increasingly important novel trait for maintaining the economic and environmental sustainability of dairy cattle production. Given that rFE is a derived trait based on the key energy source trait, dry matter intake (**DMI**), and key energy sink traits like milk energy (**MILKE**) and metabolic body weight (**MBW**), a more thorough quantitative genetic study of rFE would provide important baseline knowledge for dairy cattle management. A dairy consortium dataset focused on eventually developing genetic evaluations for rFE in US Holsteins, has been collected on DMI, MILKE, and MBW records from nearly 7000 dairy cattle in four countries. My dissertation was designed to address some motivating quantitative genetic research questions that could be addressed from the analyses of this dataset.

In Chapter 2, I reassessed the merit of using residual feed intake (**RFI**), defined as the estimated residual from regressing DMI on energy sink traits, as a measure of rFE, given that it is most commonly used in dairy cattle breeding research. We proposed a multiple trait (**MT**) modeling strategy involving all of the component traits, demonstrating that the use of the Cholesky Decomposition (**CD**) on the genetic and residual variance-covariance matrices lead to a more potentially flexible quantitative genetic approach to modeling rFE compared to RFI. The advantages of this MT approach were confirmed by simulation when the genetic versus residual relationships between energy sink and source traits were rather divergent. However, there appeared to be no meaningful differences in quantitative genetic inferences when applied to the dairy consortium dataset. Similar conclusions on non-distinctions between the two models for

genome wide association (**GWA**) analyses were also drawn in Chapter 4 although the MT GWA analyses shed further light that quantitative genetic inferences on DMI are distinctly independent of those on rFE.

I also investigated heterogeneous genetic relationships across environments for rFE with two broadly different approaches. In Chapter 3, I extended the MT model to discover substantial heterogeneity in genetic and residual partial efficiencies and variance components defining rFE as functions of environmental and management factors. Subsequently, I inferred upon genotype by environment interaction at the genomic level across environment conditions in Chapter 5, determining that some genomic regions are sensitive to environmental covariates such as average production and temperature.

## ACKNOWLEDGMENTS

This dissertation could not have been completed without the assistance and support from many professors, colleagues, and my family. I would like to acknowledge people who have made great contributions to my PhD research and my life.

First, I want to express my sincere appreciation to my advisor, Dr. Robert Tempelman, who has offered unconditional support and understanding over the years. Thank you for inspiring and thoughtful guidance through my graduate training. You have continually and convincingly conveyed a spirit of adventure in regard to research and scholarship.

I would like to thank my committee members, Dr. Mike VandeHaar, Dr. Juan P. Steibel, Dr. Nora M. Bello, and Dr. Andrew O. Finley for their encouragement, insightful comments, inspiring discussions, and useful suggestions. I appreciate your patience and understanding over the years. Special thanks to Dr. Mike VandeHaar, who has made every effort to create feed efficiency consortium dataset. My whole dissertation was based on the viability of this dataset.

In addition, thanks to my colleagues and department of Animal Science. They have supported me all the time. I am very grateful for the friendship of all of the members of the animal genetic and breeding research group. I have worked closely and puzzled over many research and personal problems with them.

Finally, many thanks to my parents and dear sister. They have let me pursue PhD degree freely and encouraged me as always. Also, to my husband, thank you for taking care of me and bringing so much happiness to my life.

## TABLE OF CONTENTS

LIST OF TABLES .....	viii
LIST OF FIGURES .....	x
Chapter1: Introduction .....	1
1.1 Data Resources .....	3
1.2 Research Issues to Resolve with Genetics of Residual Feed Efficiency.....	4
1.3 Statistical Approaches .....	7
1.4 Specific Objectives.....	7
Chapter2: An Alternative Approach to Modeling Genetic Merit of Feed Efficiency in Dairy Cattle.....	9
2.1 Abstract .....	9
2.2 Introduction .....	9
2.3 Materials and Methods .....	12
2.3.1 Proposed MT Model.....	12
2.3.2 Functional Similarities between RFI and Proposed MT Model .....	15
2.3.3 Simulation Study .....	17
2.3.4 Application to Dairy Consortium Data.....	19
2.4 Results .....	23
2.4.1 Simulation Study .....	23
2.4.2 Application to Dairy Consortium Data.....	28
2.4.2.1 Variance Component Estimates .....	28
2.4.2.2 Partial Efficiencies/Regressions.....	29
2.4.2.3 Correlation between EBV .....	30
2.4.2.4 Cross-validation Study.....	31
2.5 Discussion .....	32
2.6 Conclusions .....	36
Chapter3: Modeling Genetic and Non-Genetic Variation of Feed Efficiency and its Partial Relationships between Component Traits as a Function of Management and Environmental Factors.....	37
3.1 Abstract .....	37

3.2 Introduction .....	37
3.3 Materials and Methods .....	40
3.3.1 The Multiple Trait Model .....	40
3.3.2 Application to Dairy Consortium Data.....	47
3.3.2.1 Data Source and Editing .....	47
3.3.2.2 Model Comparison.....	49
3.3.2.3 Cross-validation Assessment .....	50
3.3.2.4 Marginal Means .....	51
3.3.3 Simulation Study .....	51
3.3.4 MCMC Implementation .....	52
3.4 Results .....	53
3.4.1 Analysis of Dairy Consortium Data .....	53
3.4.2 Inferences on Factors Influencing PE.....	53
3.4.3 Inferences on Variance Components and Heritability of rFE .....	60
3.4.4 Cross-validation.....	64
3.4.5 Simulation Study Assessment .....	65
3.5 Discussion .....	66
3.6 Conclusions .....	71
Chapter4: Genome Wide Association Analyses based on Alternative Strategies for Modeling	
Feed Efficiency .....	72
4.1 Abstract .....	72
4.2 Introduction .....	73
4.3 Materials and Methods .....	75
4.3.1 The MT Model.....	75
4.3.2 Single SNP Associations .....	78
4.3.3 Window-Based Associations.....	79
4.3.4 Application to Dairy Consortium Data.....	80
4.3.4.1 Data Description .....	80
4.3.4.2 Real Data Analyses .....	83
4.4 Results and Discussion.....	84
4.4.1 Genetic Parameter Estimates .....	84
4.4.2 Single SNP Associations .....	85
4.4.3 Window-Based Associations.....	88

4.4.4 General Discussion .....	93
4.5 Conclusions .....	96
Chapter5: Modeling genotype by environment interaction over multiple covariates for genome-wide association analyses of residual feed efficiency in dairy cattle.....	98
5.1 Abstract .....	98
5.2 Introduction .....	99
5.3 Materials and Methods .....	101
5.3.1 Conventional Genomic Prediction Model for Residual Feed Intake.....	101
5.3.2 Reaction Norm Extension Involving Multiple Covariates .....	102
5.3.3 Backsolving for SNP-specific Effects .....	104
5.3.4 Genome Wide Associations on SNP-specific Overall Effects and Reaction Norms	104
5.3.5 Simulation Study .....	108
5.3.6 Application to RFI Data .....	109
5.3.7 Cross-validation Assessment.....	111
5.4 Results .....	111
5.4.1 Simulation Study .....	111
5.4.2 Real Data Analyses.....	114
5.4.3 Genome-wide Association Analyses for RFI .....	116
5.5 Discussion .....	119
5.6 Conclusions .....	123
Chapter6: Conclusions and Future Work.....	124
6.1 Conclusions .....	124
6.2 Implications of Dissertation for Genetic Improvement and Control of Feed Efficiency.	127
6.3 Future Work .....	128
APPENDICES .....	131
APPENDIX A: Chapter 1 .....	132
APPENDIX B: Chapter 2.....	136
APPENDIX C: Chapter 3.....	151
APPENDIX D: Chapter 4.....	158
REFERENCES .....	169

## LIST OF TABLES

Table 2.1: Ranges of values on seven key parameters for testing proposed multivariate model versus RFI model under a response surface design simulation study.....	18
Table 2.2: Frequency of number of cows and lactations by station in dairy consortium study....	20
Table 2.3: Estimates of variance components, heritabilities, and repeatabilities of rFE from dairy consortium data analyses based on proposed multivariate approach versus RFI model. ....	29
Table 2.4: Estimates of partial efficiencies based on proposed multivariate model in dairy consortium data analysis .....	30
Table 2.5: Cross Validation Prediction Accuracies for each of 5 different validation datasets (fold) based on proposed (MT) versus RFI analyses of dairy consortium data .....	31
Table 3.1: Distribution of cows by region. ....	48
Table 3.2: Posterior inferences characterizing heterogeneity of genetic and residual partial efficiencies of $DMI^1$ on $MILKE^2$ .....	55
Table 3.3: Posterior inferences characterizing heterogeneity of genetic and residual partial efficiencies of $DMI^1$ on $MBW^2$ .....	56
Table 3.4: Posterior inferences characterizing heterogeneity of genetic and residual variation of $DMI^1 MILKE^2,MBW^3$ .....	62
Table 3.5: Inferences on station and parity specific heritabilities for $DMI^1 MILKE^2,MBW^3$ .....	64
Table 3.6: Cross validation prediction accuracies for each of 5 different validation datasets (fold) for homogeneous and heterogeneous models. ....	65
Table 3.7: Summary on 95% highest posterior density interval (HPD) and coverage probabilities across 20 replicated datasets in simulation study .....	66
Table 4.1: Brief characterization of dataset by stations <sup>1</sup> .....	82
Table 4.2: Estimated variance components and heritability for residual feed efficiency and its component traits.....	84

Table 4.3: Residual and genetic covariances between the three component traits .....	85
Table 4.4: Physical location <sup>1</sup> (in Mb) of significant regions detected for each trait .....	89
Table 4.5: Symbol, name, and involved biological functions of genes overlapped with 2 significant regions identified for residual feed efficiency in Holstein cows .....	91
Table 5.1: Brief characterization of dataset by stations .....	107
Table 5.2: Areas under receiver operating characteristic curve (AUC) between the classic GBLUP and the reaction norm (RN) model at different false positive rates.....	112
Table 5.3: Estimated variance components and likelihood ratio test for residual feed intake based on the reaction norm (RN) models with various breakpoints on temperature (TP).....	114
Table 5.4: Estimated variance components and likelihood ratio test for dry matter intake based on the reaction norm (RN) models with various breakpoints on temperature (TP).....	115
Table 5.5: Symbol, name, and biological process of genes have reaction norm (RN) effects with temperature breaks at 65F (TP65) and contemporary group milk production (AP) for the residual feed intake .....	118
Table 5.6: Cross Validation Prediction Accuracies for each of 5 different validation datasets (fold) under the regular GBLUP and the reaction norm (RN) model. ....	119
Table A.1: True values for each of seven key parameters in 162 run response surface design simulation study. ....	132
Table B.1: Posterior inferences characterizing heterogeneity of genetic and residual partial efficiencies of MBW <sup>1</sup> on MILKE <sup>2</sup> .....	147
Table B.2: Marginal Mean and Variance Component Inferences of Genetic and Residual Variance Components of MILKE <sup>1</sup> .....	148
Table B.3: Marginal Mean and Variance Component Inferences of Genetic and Residual Variance Components of MBW <sup>1</sup>   MILKE <sup>2</sup> .....	150

## LIST OF FIGURES

Figure 2.1: Estimated (with standard error bars) mean residual ( $\hat{\phi}_{3j}^{(e)}$  : grey bars) and genetic ( $\hat{\phi}_{3j}^{(g)}$  : white bars) partial efficiencies under proposed multivariate model and estimated partial regression coefficients ( $\hat{b}_j$  : black bars) under RFI model for DMI on (a) MILKE ( $j=1$ ) as a function of couplets of values for [ $\phi_{31}^{(u)} = \phi_{31}^{(e)}$ ] and for DMI on (b) MBW ( $j=2$ ) as a function of couplets of values for [ $\phi_{32}^{(u)} = \phi_{32}^{(e)}$ ] in response surface design simulation study..... 24

Figure 2.2: Estimated (with standard error bars) mean residual ( $\hat{\phi}_{3j}^{(e)}$  : grey bars) and genetic ( $\hat{\phi}_{3j}^{(g)}$  : white bars) partial efficiencies under proposed multivariate model and estimated partial regression coefficients ( $\hat{b}_j$  : black bars) under RFI model for DMI on (a) MILKE ( $j=1$ ) as a function of couplets of values for [ $\phi_{31}^{(u)} \neq \phi_{31}^{(e)}$ ] and for DMI on (b) MBW ( $j=2$ ) as a function of couplets of values for [ $\phi_{32}^{(u)} \neq \phi_{32}^{(e)}$ ] in response surface design simulation study..... 25

Figure 2.3: Estimated (with standard error bars) mean parameters under RFI (white bars) vs proposed multivariate (black bars) models for (a) genetic variances  $\hat{\sigma}_{u3|1,2}^2$  versus  $\hat{\sigma}_{uRFI}^2$ , (b) residual variances  $\hat{\sigma}_{e3|1,2}^2$  versus  $\hat{\sigma}_{eRFI}^2$  and (c) heritabilities  $\hat{h}_{3|1,2}^2$  versus  $\hat{h}_{RFI}^2$  as a function of equivalent partial efficiencies ( $\phi_{3.}^{(u)} = \phi_{3.}^{(e)}$  defined as  $\phi_{31}^{(u)} = \phi_{31}^{(e)}$  and  $\phi_{32}^{(u)} = \phi_{32}^{(e)}$ ) or nonequivalent partial efficiencies ( $\phi_{3.}^{(u)} \neq \phi_{3.}^{(e)}$  defined as  $\phi_{31}^{(u)} \neq \phi_{31}^{(e)}$  and  $\phi_{32}^{(u)} \neq \phi_{32}^{(e)}$ ) in response surface design simulation study. Any bars not sharing the same letter within a pair are different from each other ( $P<0.05$ ). ..... 26

Figure 2.4: Differences in accuracies of genetic merit prediction between proposed multivariate and RFI models versus  $\phi_{31}^{(e)}$  for  $\phi_{31}^{(u)} = 0.2(\bigcirc-\bigcirc-\bigcirc)$ ,  $0.3(\triangle-\triangle-\triangle)$ ,  $0.4(\blacktriangle-\blacktriangle-\blacktriangle)$ ,  $0.5(\bullet-\bullet-\bullet)$ , and  $0.6(\square-\square-\square)$  in response surface design simulation study..... 27

Figure 2.5: Difference in accuracies of prediction of genetic merit between proposed multivariate and RFI models versus  $\phi_{32}^{(e)}$  for  $\phi_{32}^{(u)}$  being  $0.0500(\bigcirc-\bigcirc-\bigcirc)$ ,  $0.0875(\triangle-\triangle-\triangle)$ ,  $0.1250(\blacktriangle-\blacktriangle-\blacktriangle)$ ,  $0.1625(\bullet-\bullet-\bullet)$ , and  $0.2000(\square-\square-\square)$  in response surface design simulation study. .... 28

Figure 2.6: Scatterplot of EBV of rFE based on proposed multivariate model ( $\hat{\mathbf{u}}_{3|1,2}$ ) versus RFI ( $\hat{\mathbf{u}}_{RFI}$ ) model of dairy consortium data. Reference line of slope 1 and intercept 0 superimposed ..... 31

Figure 3.1: Posterior means of station-specific genetic ( $\hat{\phi}_{(31)}^u$ and $\hat{\phi}_{(32)}^u$ ) and residual ( $\hat{\phi}_{(31)}^e$ and $\hat{\phi}_{(32)}^e$ ) partial efficiencies versus estimated partial regression coefficients ( $\hat{b}_1$ and $\hat{b}_2$ ) previously reported by Tempelman et al., 2015) for DMI <sup>1</sup> on MILKE <sup>2</sup> (Panel a) and MBW <sup>3</sup> (Panel a).....	59
Figure 4.1: Comparison of BLUP of SNP effects for RFI versus DMI conditional on MILKE and MBW. Line of slope 1 and intercept 0 superimposed. ....	86
Figure 4.2: Comparison of single SNP inferences ( $-\log_{10}(p)$ ) for DMI conditional on MILKE and MBW versus those for RFI. Line of slope 1 and intercept 0 superimposed. ....	86
Figure 4.3: The Manhattan plots of single SNP for DMI conditional on MILKE and MBW(a) and for RFI (b). Bonferroni correction threshold superimposed. ....	87
Figure 4.4: The Manhattan plots of single SNP on MILKE (a), MBW(b) , and DMI(c). Bonferroni correction threshold superimposed.....	88
Figure 4.5: The Manhattan plots of 1Mb windows for DMI conditional on MILKE and MBW(a) and for RFI (b). Bonferroni correction threshold superimposed. ....	90
Figure 4.6: The Manhattan plot of 1-Mb windows on MILKE (a), MBW(b) , and DMI(c). Bonferroni correction threshold superimposed.....	92
Figure 5.2: Receiver operating characteristic (ROC) curves for random intercept for 1Mb non-overlapping windows based on the reaction norm (RN) model and the classic GBLUP averaged across 10 replicates. Line of slope 1 and intercept 0 superimposed. ....	112
Figure 5.3: Receiver operating characteristic (ROC) curves for random slopes on first to fourth order polynomial terms of AP for 1Mb non-overlapping windows based on the reaction norm (RN) model and the classic GBLUP averaged across 10 replicates. Line of slope 1 and intercept 0 superimposed. ....	113
Figure 5.4: Comparison on $-\log_{10}(p)$ of SNP (a) and 1Mb window (b) for random intercept between the GBLUP and the reaction norm (RN) model. Line of slope 1 and intercept 0 superimposed. ....	116
Figure C.1: Scatterplot of $-\log_{10}(p)$ for SNPs (a) and windows (b) based on the homogeneous partial regression RFI versus heterogeneous partial regression RFI models. Line of slope 1 and intercept 0 superimposed. ....	156

Figure C.2: Manhattan plots for windows on RFI based on all cows (a) and genotyped cows (b). Bonferroni correction threshold superimposed.....	157
Figure D.1: Manhattan plot for SNP-specific random intercept based on the regular GBLUP (a) and the RN model (b).....	163
Figure D.2: Manhattan plot for Window-specific random intercept based on the regular GBLUP (a) and the RN model (b) .....	164
Figure D.3: Manhattan plot for SNP-specific random slope on the linear (a) and quadratic (b) terms of AP .....	164
Figure D.4: Manhattan plot for SNP-specific random slope on the linear (a) and quadratic (b) terms of TP65.....	165
Figure D.5: Manhattan plot for SNP-specific random slope on the linear (a) and quadratic (b) terms of RH (1) and interaction between TP65 and RH (2) .....	166
Figure D.6: Manhattan plot for Window-specific random on the linear (a) and quadratic (b) terms of AP .....	167
Figure D.7: Manhattan plot for Window-specific random on the linear (a) and quadratic (b) terms of TP65.....	167
Figure D.8: Manhattan plot for Window-specific random slope on the linear (a) and quadratic (b) terms of RH (1) and interaction between TP65 and RH (2) .....	168

## **Chapter1: Introduction**

According to the FAO, global food production must increase by 70% to support an anticipated population of 9.1 billion by 2050 ([http://www.fao.org/fileadmin/templates/wsfs/docs/expert\\_paper/How\\_to\\_Feed\\_the\\_World\\_in\\_2050.pdf](http://www.fao.org/fileadmin/templates/wsfs/docs/expert_paper/How_to_Feed_the_World_in_2050.pdf)). Milk is a complete source of nutrients and milk production must increase even further, roughly 80%, by 2050 (Steinfeld et al., 2006). However, increased milk production is associated with higher feed consumption and concomitant greater greenhouse gas emissions. Feed cost accounts for 50-80% of total cost for dairy production (Connor et al., 2013, Sova et al., 2013). Furthermore, for every kg of milk produced, there is an average 1.5 kg of carbon dioxide equivalent of greenhouse gas emissions (GHG) (Hagemann et al., 2012) though Capper et al. (2009) indicated that improving milk production will reduce GHG for every unit milk production. Thus, selection on production traits without considering dry matter intake (DMI) may threaten the economic and environmental sustainability of dairy cattle production. Hence, there have been strong arguments put forward to improve feed efficiency (FE) in dairy cattle (Hayes et al., 2013, Pryce et al., 2015, VandeHaar et al., 2016).

Overall feed efficiency has very broad definitions that generally can be classified into two types of traits: ratio based traits such as feed conversion rate, and residual based traits such as residual feed intake (RFI) (Berry and Crowley, 2013). Overall feed efficiency, defined as the ratio of feed intake captured in milk products, has been doubled for the past 50 years. This increase is mostly due to the dilution of maintenance; i.e., feeding cows with much more feed than the maintenance requirements to promote the proportion of feed utilized for milk production

(VandeHaar and St-Pierre, 2006, VandeHaar et al., 2016). Yet, VandeHaar et al. (2016) suggested that dilution of maintenance effect diminishes due to the decreased feed digestibility as cows consume higher levels of feed, and therefore more attention move toward residual based FE traits that define and enhance digestive and metabolic efficiency of feed. Thus, the overall feed efficiency has two major components, 1) feed digestive and metabolic efficiency and 2) dilution of maintenance (i.e. less feed being directed towards body maintenance). RFI and other residual based measures that evaluate variation in digestive and metabolic efficiency among animals that are independent of the dilution of maintenance are not complete measures of overall feed efficiency given they do not favor small cows with lower maintenance requirement. In this dissertation, we will refer to residual feed efficiency (rFE) as the feed digestive and metabolic efficiency that are defined by RFI or similar measures that do not necessarily equate with overall feed efficiency and economic efficiency. Note that selection for small body weight and hence a smaller maintenance requirement can be easily pursued in dairy breeding (Pryce et al., 2015).

Whole genome prediction (WGP), based on the use of high density single nucleotide polymorphism (SNP) markers (Meuwissen et al., 2001) has recently played a prominent role in accelerating the genetic improvement of dairy cattle since it was implemented in 2009 (Garcia-Ruiz et al., 2016, Taylor et al., 2016). That is, higher accuracies of genetic merit prediction and larger genetic gains are achieved by using genomic information compared to traditional selection based on using only pedigree information. Genomic selection (GS) based on the WGP has proven to be an efficient way to improve economically important traits in dairy cattle. With the implementation of GS, genetic merit on milk yield has increased by 8% over the past 8 years as

reported by the Council on Dairy Cattle Breeding (<https://www.cdcb.us/eval/summary/trend.cfm?>). Genetic merit has also improved for reproductive traits as well (Garcia-Ruiz et al., 2016). It has also been noted that GS is particularly useful selection for lowly heritable traits (Calus et al., 2013).

Multiple studies have reported the estimated heritability of rFE to be between 0.08 and 0.35 depending on various measures and populations. Hence, genetic variation for rFE appears to exist and should be considered in genetic selection programs (Vallimont et al., 2010, 2011, Pryce et al., 2014, Tempelman et al., 2015, VandeHaar et al., 2016). In particular, the use of GS for selecting and breeding cows with higher rFE appears to be promising in dairy cattle (Pryce et al., 2015, VandeHaar et al., 2016). Furthermore, the use of genome wide association analyses (GWA) could be also helpful to detect important genomic regions influencing rFE.

## **1.1 Data Resources**

Inference on genetic merit depends on the availability of phenotypic records from animals or their relatives. Similarly, the use of WGP and GWA relies upon a sufficiently large number of genotyped animals in order to obtain reliable and meaningful results. Thanks to the existence of national Dairy Herd Improvement (DHI) programs, milk yields and their components (i.e. fat and protein yields) on dairy cows have been regularly collected for decades. Furthermore, there are now over 1.4 million US dairy cattle that are genotyped ([https://www.cdcb.us/News/08\\_11\\_16\\_CDCB\\_Livability\\_1\\_1\\_clean\\_002.pdf](https://www.cdcb.us/News/08_11_16_CDCB_Livability_1_1_clean_002.pdf)). However, DMI is currently not a routinely recorded trait as individual DMI records requires substantial

investments in facilities and labor. Hence, the lack of DMI data is a major obstacle to the genetic improvement in rFE.

Combining data from different institutions is being increasingly used to assemble large enough datasets for genetic inferences (de Haas et al., 2012, Berry et al., 2014, Andersson et al., 2015, Hardie et al., 2015). Currently, there are several research consortia that involve the collection of DMI records from dairy cattle. The dairy feed efficiency project led by Michigan State University and funded by USDA-NIFA (United States Department of Agriculture National Institute of Food and Agriculture) has led to the creation of a database including records on DMI, milk yield (MY), milk components, body weight (BW), and body condition score (BCS) on nearly 7,000 cows from 16 different research stations in U.S., Canada, Dutch, and the United Kingdom (Tempelman et al., 2015); this is the dataset that has been used in this particular dissertation. The global DMI (gDMI) project led by Wageningen University also combined DMI data from multiple countries but with a primary objective of predicting genetic merit and selecting for DMI rather than for rFE per se (Veerkamp et al., 2013). There is yet another project involving dairy researchers from Canada, Australia, and New Zealand to perform genomic studies on residual feed intake (RFI). Some data has been shared across these three projects.

## **1.2 Research Issues to Resolve with Genetics of Residual Feed Efficiency**

The availability of the USDA-NIFA consortium dataset greatly facilitates genetic studies on feed efficiency and its component traits. With this dataset, I hope to answer some questions that are critical to genetic and genomic selection on rFE based on a relatively large population.

One question is how should we evaluate rFE in dairy cattle? There are a few different measures for rFE in dairy cattle, including feed conversion ratio, gross efficiency, RFI, and net energy efficiency (Koch et al., 1963, Kennedy et al., 1993, Berry and Crowley, 2013, Berry and Pryce, 2014). Among them, the currently most popular is RFI defined as the difference between actual DMI and that predicted based on energy sink requirements for production and maintenance (Koch et al., 1963). These predictions assumed that the partial relationships relating DMI to these energy sink traits are the same at genetic and non-genetic levels. Furthermore, the use of RFI requires that records on all the component traits used to compute RFI are available on each animal. Currently, the gDMI research group focuses on the analysis of DMI rather than RFI, in selecting for FE (Veerkamp et al., 2013). There have been reservations in using RFI as a “trait” for selection since it is not directly observable. A natural question is under what conditions would selection for DMI and RFI be equivalent?

As indicated previously, assembling data from various research stations is necessary to build a sufficiently large dataset for inference; nevertheless, there may be issues in integrating data from rather disparate sources (Banos et al., 2012). Data quality can be partly addressed by invoking careful edits and filters. However, there may be heterogeneity at various levels which may somewhat contaminate inferences involving combined datasets if these issues are not appropriately handled. Heterogeneity in residual variances has been widely recognized and considered in breeding for many economic traits in dairy cattle (Lopez-Romero et al., 2003, Ronnegard et al., 2010, Ronnegard et al., 2013, Ou et al., 2016) as well as heterogeneity at the genetic level (Martin-Collado et al., 2015, Shirali et al., 2015, Tempelman et al., 2015). For

example, Tempelman et al. (2015) found that genetic variances for RFI varied greatly across countries. Also, de Haas et al. (2012) also reported low genetic correlations among DMI from different countries, suggesting DMI may be genetically heterogeneous across countries. Each of those two studies implied that heterogeneity was influenced by research station or even country. However, residual and genetic heterogeneity of rFE may be caused by various other factors including weather, location, parities, rations, farm management, etc (Martin-Collado et al., 2015). Inferring such heterogeneity in rFE could provide important base knowledge for fine-tuning management across production systems.

Finally, GWA analyses of rFE require further research. There have been smaller scale GWA studies conducted for the rFE in dairy cattle populations (Bolormaa et al., 2011, Yao et al., 2013). Although Veerkamp et al. (2012) performed GWA analysis on DMI, no GWA study in dairy cattle has directly compared differences in GWA inferences between rFE and DMI to my knowledge. This is somewhat intriguing as there is some controversy as to whether to include DMI or rFE as a key trait for genetic improvement programs in dairy cattle (Berry and Pryce, 2014).

With rFE, there are other considerations involving GWA inferences. That is, are allelic substitution effects in a certain portion of the genome of a consistent magnitude across environments or are these effects environmentally sensitive? To address these concerns, GWA analyses should attempt to model genotype by environment (GxE) interaction in which the allelic substitution effects for various genotypes are modeled as functions of potentially important environmental covariates. This phenomenon is believed to be fairly common for complex traits

within a wide variety of species including dairy cattle (Valdar et al., 2006, Dempfle et al., 2008, Goddard and Hayes, 2009). Information on GxE could be particularly exploited in dairy cattle breeding because of advanced reproductive technologies (e.g. artificial insemination, embryo transfer, synchronized breedings) that facilitate the ready exchange of germplasm across environments.

By developing methodology and corresponding data analyses to answer these questions, I hope to develop some base knowledge on the complexities of genetic and environmental modeling of rFE that could be incorporated in future dairy cattle management schemes.

### **1.3 Statistical Approaches**

In this dissertation, I proposed an alternative strategy for modeling rFE based on a multivariate (MT) model involving the components traits of rFE, namely DMI, milk energy (MILKE), and metabolic body weight (MBW). One can readily think of rFE as being DMI adjusted for (or conditional on) MILKE and MBW. The Cholesky decomposition (Pourahmadi, 2007) on the genetic and residual variance-covariance matrices in the MT facilitates such a characterization at both genetic and non-genetic levels and has been previously used to characterize the relationship between production and reproduction traits in dairy cattle (Bello et al., 2012). Exploring partial relationships between the component traits of rFE at genetic and non-genetic levels could allow a richness of insight not otherwise allowed with RFI modeling.

### **1.4 Specific Objectives**

This dissertation aims to provide a new perspective to evaluate and perform genetic analysis on rFE using the USDA-NIFA funded dairy consortium dataset. The first objective of this

dissertation is to demonstrate how the Cholesky decomposition of genetic and residual variance-covariance matrices in an MT model on the key components traits of rFE leads to a potentially more elegant characterization of rFE compared to anything else previously provided. A second objective is to explore potential heterogeneity of partial efficiencies and variance components of rFE at both genetic and non-genetic levels as functions of various environment and management factors. A third objective was to assess whether or not there might be any advantages to using the MT over the RFI model for GWA and particularly to assess how the associations may differ between the various component traits of rFE and rFE itself. My final objective was to investigate the impact of GxE through reaction norm modeling and to assess its impact on GWA and resulting inferences. All objectives involve the development of methodologies and/or models, followed by a simple simulation study assessment and an application to the USDA-NIFA consortium dataset.

## **Chapter2: An Alternative Approach to Modeling Genetic Merit of Feed Efficiency in Dairy Cattle**

### **2.1 Abstract**

Genetic improvement of residual feed efficiency (rFE) in dairy cattle requires greater attention given increasingly important resource constraint issues. A widely accepted yet occasionally contested measure of rFE in dairy cattle is residual feed intake (RFI). The use of RFI is limiting for a number of reasons, including interpretation, differences in recording frequencies between the various component traits that define RFI, and potential differences in genetic versus non-genetic relationships between the component traits. Hence, analyses focusing on dry matter intake (DMI) as the response are often preferred. We propose an alternative multiple-trait (MT) modeling strategy that exploits the Cholesky decomposition to provide a potentially more robust measure of rFE. We assessed both approaches by simulation as well as by application to 26,383 weekly records from 50 to 200 days in milk on 2,470 cows from a dairy rFE consortium study involving 7 institutions. Although the proposed MT model fared better than the RFI model when simulated genetic and non-genetic associations between DMI and component traits were substantially different from each other, there were no meaningful differences in predictive performance between the two models when applied to the consortium data.

### **2.2 Introduction**

Residual feed efficiency (rFE) based on the efficient conversion of feed nutrients into saleable milk directly affects the profitability of dairy production. Thus, improving rFE in cattle is important to maximizing dairy production on limited inputs, especially as constraints on feed production become increasingly relevant. In addition, improving rFE is also of environmental

importance because more nutrients are directed into milk production with less nutrient loss in manure and methane excreted as rFE increases (Richardson and Herd, 2004).

A popular measure of rFE is residual feed intake (RFI) which is defined as the difference between actual feed intake and that predicted based on requirements for production and maintenance or the so-called “energy sinks” (Koch et al., 1963). The typical statistical modeling strategy for RFI includes two stages (Tempelman et al., 2015). For reasons that will be explained later, we will refer to DMI as Trait #3 such that  $y_{i3}$  represents the DMI record for animal  $i$ . In the first stage, DMI, as the energy source, is typically specified as a linear function of various energy sinks plus other fixed or random (non-animal) effects that potentially influence DMI; i.e.

$$y_{i3} = \mathbf{x}_i' \boldsymbol{\beta}_3^* + b_1 \text{MILKE}_i + b_2 \text{MBW}_i + b_3 \Delta \text{BW}_i + \text{RFI}_i . \quad [2.1]$$

Here  $\boldsymbol{\beta}_3^*$  represents a vector of various fixed effects connected to  $y_{i3}$  via known incidence row vector  $\mathbf{x}_i'$ . Key energy sinks include milk energy (MILKE), metabolic BW (MBW) being defined as BW raised to the  $3/4$  power and BW change ( $\Delta \text{BW}$ ) all indexed by animal or record  $i$  in Equation [2.1]. Note then that  $b_1$ ,  $b_2$ , and  $b_3$  are partial regression coefficients of DMI on energy sinks MILKE, MBW, and  $\Delta \text{BW}$ , respectively. Now  $\text{RFI}_i$  is merely the estimated residual from Model [2.1], thereby representing the RFI “record” for subject  $i$ ; this variable is further typically specified as the response variable in a second stage variance components model:

$$\text{RFI}_i = \mu^* + \mathbf{z}_i' \mathbf{u}_{\text{RFI}} + \text{other potential random animal effects} + e_{i,\text{RFI}} \quad [2.2]$$

In Equation [2.2],  $\mu^*$  is the overall mean, whereas  $\mathbf{u}_{RFI}$  is the animal genetic merit for RFI and connected to  $RFI_i$  via known incidence row vector  $\mathbf{z}_i'$ . Furthermore,  $\mathbf{u}_{RFI} = \{u_{RFI,i}\} \sim \text{MVN}(\mathbf{0}, \mathbf{A}\sigma_{u_{RFI}}^2)$  for  $\mathbf{A}$  being the numerator (Henderson, 1976) or genomic relationship matrix (VanRaden, 2008) or a hybrid of the two (Aguilar et al., 2010), whereas other potential random animal effects may include permanent environmental effects if there is more than one record per animal. Finally  $e_{iRFI}$  is a corresponding residual such that  $e_{iRFI} \sim \text{NIID}(0, \sigma_{e_{RFI}}^2)$ .

Conceptually, Equations [2.1] and [2.2] could be combined together as one model as shown in Equation [2.3], thereby reinforcing that RFI is really just an adjusted measure of DMI; i.e.

$$y_{i3} = \mu + \mathbf{x}_i' \boldsymbol{\beta}_3^* + b_1 \text{MILKE} + b_2 \text{MBW} + b_3 \Delta \text{BW} + \mathbf{z}_i' \mathbf{u}_{RFI} + \text{other potential random animal effects} + e_{iRFI} \quad [2.3]$$

As argued by Tempelman et al. (2015), analyses based on a single-stage model [2.3] may be desirable if effects in Equation [2.1] are not orthogonal to effects in Equation [1.2].

Given these limitations of RFI but, nevertheless, an inherent desire to meaningfully characterize rFE beyond a selection index involving DMI, we propose an alternative parameterization for multiple trait (MT) modeling of DMI jointly with key energy sink traits. Our objective is to demonstrate how this parameterization not only leads to a potentially more elegant characterization of rFE compared to RFI, but furthermore reconciles approaches focused on the analyses of RFI versus DMI (Berry and Pryce, 2014).

## 2.3 Materials and Methods

### 2.3.1 Proposed MT Model

Our proposed strategy for characterizing rFE is based on the square root free or modified Cholesky decomposition (CD) (Pourahmadi et al., 2007) which our group has previously adapted for the joint analysis of milk production and reproduction data albeit in a non-genetic context (Bello et al., 2010). We apply this decomposition on each variance-covariance matrix (e.g., genetic and residual) partition in a MT model analysis on DMI and two key energy sink traits (i.e., MILKE and MBW). As with RFI modeling, we prefer to keep  $\Delta BW$  as a covariate for DMI even in this proposed approach, in part due to its seemingly very low heritability (<1%) and greater relative variability. Statistically, the order for the three traits is rather arbitrary; however, for modeling rFE, it is necessary to specify DMI as the last trait in the sequence as noted later.

We write the MT linear mixed model for MILKE, MBW, and DMI in order as Traits 1, 2, and 3, respectively; extensions to any other number of rFE component traits are relatively straightforward, provided again that DMI is specified last. We momentarily assume one record per animal such that the model can be written in a classical quantitative genetics framework as:

$$\mathbf{y}_{.j} = \mathbf{X}_j \boldsymbol{\beta}_j + \mathbf{Z}_j \mathbf{u}_{.j} + \mathbf{e}_{.j}; j=1,2,3 \quad [2.4]$$

Here  $\mathbf{y}_{.j} = [y_{1j} \ y_{2j} \ \cdots \ y_{n,j}]'$  is the vector of  $n \times 1$  responses for Trait  $j$  whereas  $\boldsymbol{\beta}_j$  denotes the vector of fixed effects for Trait  $j$  connected to  $\mathbf{y}_{.j}$  by corresponding known incidence matrix  $\mathbf{X}_j$ , whereas  $\mathbf{u}_{.j} = [u_{1j} \ u_{2j} \ \cdots \ u_{q,j}]'$  denotes the  $q \times 1$  vector of additive genetic effects for Trait  $j$  connected to  $\mathbf{y}_{.j}$  by corresponding known incidence matrix  $\mathbf{Z}_j$ . Finally,  $\mathbf{e}_{.j} = [e_{1j} \ e_{2j} \ \cdots \ e_{n,j}]'$  is the  $n \times 1$  vector of residuals for Trait  $j$ . Further extensions are

possible to handle other cases, for example, if there are repeated measurements per trait per animal or even missing records for some traits on some animals.

Let  $\mathbf{u}$  denote the vector of additive genetic merit for all three traits jointly; i.e.  $\mathbf{u} = [\mathbf{u}_{\cdot 1} \quad \mathbf{u}_{\cdot 2} \quad \mathbf{u}_{\cdot 3}]'$  where elements are ordered by animals within traits. Suppose that we reorder  $\mathbf{u}$  by traits within animals as  $\mathbf{u} = [\mathbf{u}_{\cdot 1} \quad \mathbf{u}_{\cdot 2} \quad \mathbf{u}_{\cdot 3} \quad \cdots \quad \mathbf{u}_{\cdot q}]'$  with  $\mathbf{u}_{\cdot i} = [u_{i1} \quad u_{i2} \quad u_{i3}]'$  representing the genetic merit for the three traits on animal  $i$ . We specify the typical animal breeding MT modeling assumptions using  $\text{var}(\mathbf{u}_{\cdot i}) = \mathbf{G} \forall i$  where  $\mathbf{G}$  is the 3 x 3 genetic variance covariance matrix for the three traits with diagonal element  $j$ , ( $j=1,2,3$ ) being the genetic variance  $\sigma_{u_j}^2$  for Trait  $j$  and off-diagonals  $\sigma_{u_j u_{j'}}$  defining the genetic covariance between Traits  $j$  and  $j'$  ( $j'=1,2,3; j' \neq j$ ) across the  $q > n$  animals. Then,  $\text{var}(\mathbf{u}) = \mathbf{A} \otimes \mathbf{G}$  for  $\mathbf{A}$  being the relationship matrix, as previously noted, and  $\otimes$  denoting the Kronecker product with  $q$  also counting potential ancestors without phenotypic records. Similarly, writing  $\mathbf{e}_i = [e_{i1} \quad e_{i2} \quad e_{i3}]'$  as the 3 x 1 vector of residuals for animal  $i$ , we assume  $\text{var}(\mathbf{e}_i) = \mathbf{R} \forall i$  where  $\mathbf{R}$  is a 3 x 3 residual variance covariance matrix with residual variances ( $\sigma_{e_j}^2$ ) along the diagonal and residual covariances  $\sigma_{e_j e_{j'}}$  in the off-diagonals.

A CD on  $\mathbf{G}$  merely involves the following reparameterization of additive genetic effects for the three traits in a sequential manner:

$$u_{i1} = u_{i1} \tag{2.5a}$$

$$u_{i2} = \phi_{21}^{(u)} u_{i1} + u_{i2|1} \tag{2.5b}$$

$$u_{i3} = \phi_{31}^{(u)} u_{i1} + \phi_{32}^{(u)} u_{i2} + u_{i3|1,2} \tag{2.5c}$$

or jointly together as

$$\mathbf{u}_{i.} = \mathbf{T}_u \mathbf{u}_{i.} + \mathbf{u}_{i.}^* \quad [2.6]$$

Here  $\mathbf{T}_u$  is a 3 x 3 lower triangular matrix with zeroes on the diagonals and non-zero elements  $\phi_{21}^{(u)}$ ,  $\phi_{31}^{(u)}$ , and  $\phi_{32}^{(u)}$  with row-column addresses in  $\mathbf{T}_u$  indicated by their corresponding subscripts. These parameters are typically referred to as generalized autoregressive parameters by Pourahmadi et al.(2007); however, we prefer to refer to them as partial efficiencies at the genetic level in our rFE application. For example, we characterize  $\phi_{31}^{(u)}$  as the genetic partial efficiency of DMI on MILKE. Furthermore,  $\mathbf{u}_{i.}^* = [u_{i1} \ u_{i2|1} \ u_{i3|1,2}]'$  for animal  $i$  are specified to be normally and independently distributed effects having diagonal (co)variance matrix  $\Delta_u = \text{diag}(\sigma_{u_1}^2, \sigma_{u_{2|1}}^2, \sigma_{u_{3|1,2}}^2)$ . Here elements of  $\Delta_u$  were labeled by Pourahmadi et al. (2007) as innovation variance components but we prefer to refer to them as sequentially conditional variance components (SCVC) for Trait 1, Trait 2 given 1 (Trait 2|1) and Trait 3 given 1 and 2 (Trait 3|1,2), respectively. Similarly, the elements of  $\mathbf{u}_{i.}^*$  could be referred to as sequentially conditional genetic effects, consistent with our conditional subscripting notations as indicated above for elements of  $\mathbf{u}_{i.}^*$  and  $\Delta_u$ . It can be then demonstrated that  $\text{var}(\mathbf{u}_{i.}) = \mathbf{G} = (\mathbf{I} - \mathbf{T}_u)^{-1} \Delta_u (\mathbf{I} - \mathbf{T}_u)^{-1}$ , such that

$$\mathbf{G} = \begin{bmatrix} \sigma_{u_1}^2 & \phi_{21}^{(u)} \sigma_{u_1}^2 & (\phi_{31}^{(u)} + \phi_{21}^{(u)} \phi_{32}^{(u)}) \sigma_{u_1}^2 \\ \phi_{21}^{(u)} \sigma_{u_1}^2 & (\phi_{21}^{(u)})^2 \sigma_{u_1}^2 + \sigma_{u_{2|1}}^2 & \phi_{21}^{(u)} (\phi_{31}^{(u)} + \phi_{21}^{(u)} \phi_{32}^{(u)}) \sigma_{u_1}^2 + \phi_{32}^{(u)} \sigma_{u_{2|1}}^2 \\ (\phi_{31}^{(u)} + \phi_{21}^{(u)} \phi_{32}^{(u)}) \sigma_{u_1}^2 & \phi_{21}^{(u)} (\phi_{31}^{(u)} + \phi_{21}^{(u)} \phi_{32}^{(u)}) \sigma_{u_1}^2 + \phi_{32}^{(u)} \sigma_{u_{2|1}}^2 & (\phi_{31}^{(u)} + \phi_{21}^{(u)} \phi_{32}^{(u)})^2 \sigma_{u_1}^2 + (\phi_{32}^{(u)})^2 \sigma_{u_{2|1}}^2 + \sigma_{u_{3|1,2}}^2 \end{bmatrix} \quad [2.7]$$

Now with DMI specified as Trait 3 in this MT sequence,  $\sigma_{u_{3|1,2}}^2$  is the genetic variance for DMI conditional on genetic effects of MILKE and MBW. As a corollary,  $u_{i3|1,2}$  represents the genetic merit of DMI conditional on the genetic merit of MILKE and of MBW for animal  $i$  with  $\mathbf{u}_{i3|1,2} = \{u_{i3|1,2}\} \sim N(\mathbf{0}, \mathbf{A} \sigma_{u_{3|1,2}}^2)$ . It is this term that we propose as an alternative measure for

genetic merit of rFE as a counterpart to  $\mathbf{u}_{RFI}$  in Equations [2.2] or [2.3]. Negative values for  $u_{i3|1,2}$  indicates the animal  $i$  has favorably above average genetic merit for rFE with  $\sigma_{u_{3|1,2}}^2$  defining the genetic variance of our proposed measure of rFE. Of course, the magnitude of the partial efficiencies  $\phi_{21}^{(u)}$ ,  $\phi_{31}^{(u)}$ , and  $\phi_{32}^{(u)}$  factor into these relationships as well; i.e. together, these coefficients partly determine genetic covariance between each pair of the three traits; e.g., from Equation [2.7],  $\sigma_{u_1 u_2} = \phi_{21}^{(u)} \sigma_{u_1}^2$ .

Similarly, the residuals for the three traits on animal  $i$  can be reparameterized using the CD accordingly as

$$e_{i1} = e_{i1} \quad [2.8a]$$

$$e_{i2} = \phi_{21}^{(e)} e_{i1} + e_{i2|1} \quad [2.8b]$$

$$e_{i3} = \phi_{31}^{(e)} e_{i1} + \phi_{32}^{(e)} e_{i2} + e_{i3|1,2} \quad [2.8c]$$

For example,  $e_{i3|1,2}$  is the residual for DMI conditional on the MILKE and MBW; with  $\sigma_{e_{3|1,2}}^2 = \text{var}(e_{i3|1,2})$  being the residual variance for our proposed rFE measure. Also,  $\phi_{21}^{(e)}$ ,  $\phi_{31}^{(e)}$ , and  $\phi_{32}^{(e)}$  represent the partial efficiencies between pairs of the three traits at the residual level. In essence, the heritability measure for rFE based on our proposed approach with a classical animal model is  $h_{3|1,2}^2 = \sigma_{u_{3|1,2}}^2 / (\sigma_{u_{3|1,2}}^2 + \sigma_{e_{3|1,2}}^2)$ .

### 2.3.2 Functional Similarities between RFI and Proposed MT Model

Suppose the same fixed effect factors are specified for each of the three traits; in other words,  $\mathbf{x}_i'$  is used as the same incidence row vector connecting  $y_{ij}$  to  $\beta_j, j=1,2,3, \forall i$ . Furthermore, let us momentarily assume just one record per animal such that there are no other animal effects to model other than additive genetic effects. In Equation [2.9] below, we actually re-derive

Equation [2.3] for the one-stage RFI approach but incorporate the relationships provided in Equations [2.5c] and [2.8c]. Furthermore, let us assume momentarily that partial efficiencies as they involve DMI are identical at the genetic and residual levels; i.e.  $\phi_{31} = \phi_{31}^{(u)} = \phi_{31}^{(e)}$  and  $\phi_{32} = \phi_{32}^{(u)} = \phi_{32}^{(e)}$ . Then

$$\begin{aligned}
y_{i3} &= \mathbf{x}_i' \boldsymbol{\beta}_3 + b_3 \Delta BW + u_{i3} + e_{i3} \\
&= \mathbf{x}_i' \boldsymbol{\beta}_3 + b_3 \Delta BW + \left( \phi_{31}^{(u)} u_{i1} + \phi_{32}^{(u)} u_{i2} + u_{i3||,2} \right) + \left( \phi_{31}^{(e)} e_{i1} + \phi_{32}^{(e)} e_{i2} + e_{i3||,2} \right) \\
&= \mathbf{x}_i' \boldsymbol{\beta}_3 + b_3 \Delta BW + \phi_{31} (u_{i1} + e_{i1}) + \phi_{32} (u_{i2} + e_{i2}) + (u_{i3||,2} + e_{i3||,2}) \\
&= (\mathbf{x}_i' \boldsymbol{\beta}_3 - \phi_{31} \mathbf{x}_i' \boldsymbol{\beta}_1 - \phi_{32} \mathbf{x}_i' \boldsymbol{\beta}_2) + b_3 \Delta BW + \phi_{31} (\mathbf{x}_i' \boldsymbol{\beta}_1 + u_{i1} + e_{i1}) \\
&\quad + \phi_{32} (\mathbf{x}_i' \boldsymbol{\beta}_2 + u_{i2} + e_{i2}) + u_{i3||,2} + e_{i3||,2} \\
&= \mathbf{x}_i' (\boldsymbol{\beta}_3 - \phi_{31} \boldsymbol{\beta}_1 - \phi_{32} \boldsymbol{\beta}_2) + b_3 \Delta BW + \phi_{31} y_{i1} + \phi_{32} y_{i2} + u_{i3||,2} + e_{i3||,2}
\end{aligned} \tag{2.9}$$

In other words, it is analytically demonstrated comparing Equation [2.9] with Equation [2.3] that the RFI model leads to the same quantitative genetic inferences for genetic merit prediction and estimation of heritability on rFE as our proposed MT model, provided that  $\phi_{31} = \phi_{31}^{(u)} = \phi_{31}^{(e)}$  and  $\phi_{32} = \phi_{32}^{(u)} = \phi_{32}^{(e)}$ . Thus, if the partial efficiencies relating DMI to the energy sink traits are identical at both the genetic and residual levels, then our proposed MT approach is synonymous with RFI modeling. If so, it could then be analytically demonstrated that  $\boldsymbol{\beta}_3^* = \boldsymbol{\beta}_3 - \phi_{31} \boldsymbol{\beta}_1 - \phi_{32} \boldsymbol{\beta}_2$ ,  $b_1 = \phi_{31}$ , and  $b_2 = \phi_{32}$ , referring back to Equation [2.3], such that  $\mathbf{u}_{RFI} = \mathbf{u}_{3||,2}$  and  $h_{RFI}^2 = h_{3||,2}^2$ . That is, the classical RFI modeling strategy specifies the partial relationships between DMI and MILKE or MBW at one phenotypic level only (e.g.  $b_1$  or  $b_2$ ) whereas the proposed MT approach defines these relationships separately at the genetic ( $\phi_{31}^{(u)}$  and  $\phi_{32}^{(u)}$ ) from the residual level ( $\phi_{31}^{(e)}$  and  $\phi_{32}^{(e)}$ ) or at any other levels (e.g. permanent environment) as demonstrated later.

### 2.3.3 Simulation Study

We further explored the relationships between these two approaches (i.e. MT vs RFI) based on a simulation study invoking a response surface design. MILKE, MBW, and DMI (i.e.  $y_1$ ,  $y_2$  and  $y_3$ ) were simulated based on a MT mixed animal model with genetic and residual variance-covariance matrices  $\mathbf{G}$  and  $\mathbf{R}$  based on various specifications for  $\phi_{21}^{(u)}$ ,  $\phi_{21}^{(e)}$ ,  $\phi_{31}^{(u)}$ ,  $\phi_{31}^{(e)}$ ,  $\phi_{32}^{(u)}$ ,  $\phi_{32}^{(e)}$ , and the heritability of Trait 3|1,2, i.e.,  $h_{3|1,2}^2$ . An effect due to  $\Delta BW$  was not simulated. The heritabilities for Trait 1 (i.e., MILKE) and Trait 2|1 (i.e., MBW| MILKE) were consistently set to 0.32 and 0.23 based on estimates derived from a larger study (Tempelman et al., 2015). The same applied for phenotypic variances for Trait1, Trait2|1, and Trait 3|1,2. The sample size and pedigree structure for this simulation study was based on a subsequently described dataset except that only one record per trait was simulated per animal. A central composite response surface design designed for uniform precision based on using SAS ADX software (SAS Institute Inc., Cary, NC) was adopted whereby specifications for 7 different parameters, i.e.,  $\phi_{21}^{(u)}$ ,  $\phi_{21}^{(e)}$ ,  $\phi_{31}^{(u)}$ ,  $\phi_{31}^{(e)}$ ,  $\phi_{32}^{(u)}$ ,  $\phi_{32}^{(e)}$  and  $h_{3|1,2}^2$  (see Table 2.1 and Appendix Table A.1) were varied, resulting in 162 independent replicated datasets with 21 replications at the center point. The chosen ranges for each parameter, as provided in Table 2.1, were based on the variability across research stations from a larger study (Tempelman et al., 2015). Note that all but 13 of the runs in Appendix Table A.1 had parameter values specified within the ranges provided in Table 2.1; the remaining runs pertained to axial points which are characteristically specified in optimally constructed response surface designs (Kutner et al., 2005).

Table 2.1: Ranges of values on seven key parameters for testing proposed multivariate model versus RFI model under a response surface design simulation study.

Parameter	Range of Values <sup>1</sup>
$h_{3 12}^2$	[0.10, 0.50]
$\phi_{21}^{(u)}$	[0.10, 0.50]
$\phi_{21}^{(e)}$	[0.10, 0.50]
$\phi_{31}^{(u)}$	[0.20, 0.60]
$\phi_{31}^{(e)}$	[0.20, 0.60]
$\phi_{32}^{(u)}$	[0.05, 0.20]
$\phi_{32}^{(e)}$	[0.05, 0.20]

<sup>1</sup> Simulated values for each of 162 runs are provided in Appendix Table A.1.

For both sets of analyses, variance components were estimated by REML using the software ASREML (Gilmour et al., 2009) based on the one-stage RFI model in [2.3] and the proposed three-trait MT model. The BLUP of  $u_{iRFI}$ ,  $\hat{u}_{iRFI}$ , was determined to be the EBV conditionally on the corresponding REML estimates of  $\sigma_{u_{RFI}}^2$  and  $\sigma_{e_{RFI}}^2$  in the RFI analysis. Similarly, the BLUP of  $u_{i3|1,2}$ ,  $\hat{u}_{i3|1,2}$ , using the proposed MT approach was based on the corresponding REML estimates of variance-covariance matrices **G** and **R** and subsequent CD-based transformations. Comparisons were made between the two models for estimates of partial efficiencies (MT model) versus partial regressions (RFI model) as well as for estimates of variance components and heritabilities. Model performance was based on accuracy of genetic prediction defined as the correlation between  $\hat{u}_{iRFI}$  and  $u_{i3|1,2}$  for the RFI analysis and the correlation between  $\hat{u}_{i3|1,2}$  and  $u_{i3|1,2}$  for the proposed MT approach. Second order response surface model analyses were used to infer upon differences in accuracies between the two models as a function of the 7 parameters.

### 2.3.4 Application to Dairy Consortium Data

Data source and editing: DMI, milk yields (**MY**), BW, all measured in kg, as well as fat, protein, and lactose components of milk were collected from Holstein cows at 7 different North American research stations with most of the data being a subset of data from a larger international consortium Tempelman et al. (2015). These research stations were University of Alberta (**AB**), Iowa State University (**ISU**), Michigan State University (**MSU**), the University of Florida (**UF**), the University of Wisconsin-Madison (**UW**), the USDA Dairy Forage Research Center (**USDFRC**) in Madison, WI, and the USDA Animal Genomics and Improvement Laboratory (**AGIL**) in Beltsville, MD. For AB, all the above mentioned traits were collected on a monthly basis except DMI; more detailed description of that data can be found in Berry et al. (2014) and Manafiazar et al. (2013). Data editing procedures and conversion to weekly records on DMI, MILKE, MBW, and  $\Delta BW$  were conducted as according to those outlined in Tempelman et al. (2015) with a data summary provided in Table 2.2. This included restricting the data from 50 to 200 DIM; i.e., at a period of lactation that is somewhat post-peak such that cows are not likely to be in a negative energy balance. Pedigrees involving a total of 8,721 animals going back 3 generations were procured as well.

Table 2.2: Frequency of number of cows and lactations by station in dairy consortium study.

Station <sup>1</sup>	Number of cows	Number of weekly records	Number of lactations
AB	236	1450	360
UF	205	1664	220
ISU	758	8784	795
MSU	163	1807	192
USDFRC	245	2550	259
UW	575	6511	652
AGIL	288	3617	448
Total	2470	26383	2926

<sup>1</sup>AB= University of Alberta, UF = University of Florida, ISU = Iowa State University, MSU = Michigan State University, USDFRC = USDA Forage Research Center, UW = University of Wisconsin-Madison, AGIL=USDA Animal Genomics Improvement Laboratory

Statistical model and data analysis: All analyses were conducted using either the proposed MT approach or the single-step RFI model. However, given that there were multiple weekly records per lactation with some cows, in turn, having multiple lactations, it was necessary to additionally model two sets of permanent environmental effects. That is, the linear mixed model for DMI using the one-stage RFI approach in Equation [2.3] and for each of the three traits using the proposed MT approach in Equation [2.4] were as follows:

$$\mathbf{y}_{.j} = \mathbf{X}_{.j}\boldsymbol{\beta}_j + \mathbf{Z}_{.j}\mathbf{u}_{.j} + \mathbf{W}_{w,j}\mathbf{p}_{w,j} + \mathbf{W}_{b,j}\mathbf{p}_{b,j} + \mathbf{e}_{.j}; j=1,2,3 \quad [2.10]$$

Equation [2.10] then slightly extends Equation [2.4] (and also Equation [2.3] for RFI analyses) in that two more sets of random effects were fitted. Here  $\boldsymbol{\beta}_j$  include the fixed effects of linear effect of  $\Delta BW$ , research station, parity, rations, and fourth order polynomial function of DIM for trait  $j$ . The vector  $\mathbf{u}_{.j}$  represents the genetic effects for Trait  $j$  as indicated previously. Furthermore,  $\mathbf{p}_{w,j} = [p_{w1j} \ p_{w2j} \ \dots \ p_{wq,j}]$  represents the vector of *within-lactation* permanent environmental effects for Trait  $j$ , identifying common environmental effects consistently associated with an animal within lactation such that  $\mathbf{W}_{w,j}$  is the corresponding incidence matrix.

Similarly,  $\mathbf{p}_{b,j} = [p_{b1j} \ p_{b2j} \ \dots \ p_{bq,j}]$  represents the vector of *between-lactation* permanent environmental effects for Trait  $j$  identifying common environmental effects consistently associated with an animal across lactations, with  $\mathbf{W}_{b,j}$  being the corresponding incidence matrix. Both sets of permanent environmental effects were also considered in Connor et al. (2013) and Tempelman et al. (2015). Suppose that for the MT analyses, we specify  $\mathbf{p}_w = [\mathbf{p}'_{w,1} \ \mathbf{p}'_{w,2} \ \mathbf{p}'_{w,3}]'$  and  $\mathbf{p}_b = [\mathbf{p}'_{b,1} \ \mathbf{p}'_{b,2} \ \mathbf{p}'_{b,3}]'$  across all traits. As we did earlier for  $\mathbf{u}$  and  $\mathbf{e}$ , we could reorder elements of  $\mathbf{p}_w$  and  $\mathbf{p}_b$  by traits within animals; i.e. rewrite  $\mathbf{p}_w = [\mathbf{p}'_{w1.} \ \mathbf{p}'_{w2.} \ \dots \ \mathbf{p}'_{wm.}]'$  and  $\mathbf{p}_b = [\mathbf{p}'_{b1.} \ \mathbf{p}'_{b2.} \ \dots \ \mathbf{p}'_{bm.}]'$ , where  $\mathbf{p}_{w,i.} = [\mathbf{p}'_{w,i1} \ \mathbf{p}'_{w,i2} \ \mathbf{p}'_{w,i3}]'$  and  $\mathbf{p}_{b,i.} = [\mathbf{p}'_{b,i1} \ \mathbf{p}'_{b,i2} \ \mathbf{p}'_{b,i3}]'$  are the within-lactation and between-lactation permanent environmental effects, respectively, for the three traits on animal  $i$ . We then assume  $\text{var}(\mathbf{p}_{w,i.}) = \mathbf{P}_w \ \forall i$  and  $\text{var}(\mathbf{p}_{b,i.}) = \mathbf{P}_b \ \forall i$ ; i.e.,  $\mathbf{P}_w$  and  $\mathbf{P}_b$  are 3 x 3 covariance variance matrices for within-lactation and between-lactation permanent environmental effects, respectively. With this reordering for  $\mathbf{p}_w$  and  $\mathbf{p}_b$ , we specify independence for these effects across animals; i.e.  $\text{var}(\mathbf{p}_w) = \mathbf{I} \otimes \mathbf{P}_w$  and  $\text{var}(\mathbf{p}_b) = \mathbf{I} \otimes \mathbf{P}_b$ . We also adapt the same CD-based reparameterization for elements of  $\mathbf{p}_w$  and  $\mathbf{p}_b$  as we did for  $\mathbf{u}$  in Equations [5abc] and for  $\mathbf{e}$  in Equations [2.8abc].

Thus, the SCVC for DMI|MILKE&MBW are estimated on four different levels, i.e., conditional genetic ( $\sigma^2_{u_{3|1,2}}$ ), conditional within-lactation permanent environment ( $\sigma^2_{p_{w,3|1,2}}$ ), conditional between-lactation permanent environment ( $\sigma^2_{p_{b,3|1,2}}$ ), and conditional residual variance ( $\sigma^2_{e_{3|1,2}}$ ) for our proposed rFE variable. In other words, the corresponding heritability  $h^2_{3|1,2}$  is simply the ratio of  $\sigma^2_{u_{3|1,2}}$  and  $\sigma^2_{p_{3|1,2}} = \sigma^2_{u_{3|1,2}} + \sigma^2_{p_{b,3|1,2}} + \sigma^2_{p_{w,3|1,2}} + \sigma^2_{e_{3|1,2}}$ , whereas the corresponding repeatability  $r^2_{3|1,2}$  is the ratio of  $\sigma^2_{u_{3|1,2}} + \sigma^2_{p_{b,3|1,2}} + \sigma^2_{p_{w,3|1,2}}$  and  $\sigma^2_{p_{3|1,2}}$ . Similar representations of

heritabilities ( $h_{RFI}^2$ ) and repeatabilities ( $r_{RFI}^2$ ) for the RFI modeling approach were determined by extending Equation [3] as well, with further details provided in Tempelman et al. (2015).

Estimates of variance components, heritabilities, and partial efficiencies at both genetic (i.e.,  $\mathbf{u}$ ) and non-genetic (i.e.,  $\mathbf{p}_w$ ,  $\mathbf{p}_b$ , and  $\mathbf{e}$ ) levels were obtained by ASREML based on using the “ANTE” specification option for variance-covariance matrices (Gilmour et al., 2009), which directly provides estimates of SCVC and partial efficiencies and their approximate standard errors at the various levels. Furthermore, the correlation between the two models for EBV was also assessed.

Cross-validation assessment: In order to test the robustness of our cross-validation tests with respect to different strategies of splitting the data, a hybrid data splitting strategy was developed for five-fold cross validation with application to the dairy consortium data. The entire dataset was initially split into two equal non-overlapping sets (Split A and Split B) of records on completely different animals. In Split A, the data were further randomly divided across animals into five nearly equally sized subsplits, i.e. each subsplit included all the records from a randomly chosen 20% of the animals within Split A. In Split B, the data were also randomly divided into five subsplits except that roughly 20% of the records from each animal was represented within each subsplit. One validation dataset was then based on randomly assigning one of the 5 subsplits from Split A with one of the 5 subsplits from Split B, such that the remaining data constituted the training data for that particular fold as part of a 5 fold cross-validation. This cross-validation thereby allowed a simultaneous assessment of the two competing models on both across and within animal cross-validation predictive abilities. The

performance metric used was the cross-validation predictive accuracy defined as the correlation between predicted DMI in the validation dataset based on estimates derived from the analyses of the training dataset.

## 2.4 Results

### 2.4.1 Simulation Study

Under simulated scenarios where partial efficiencies were equivalent at both genetic and residual levels, (i.e.,  $\phi_{31}^{(u)} = \phi_{31}^{(e)}$  and  $\phi_{32}^{(u)} = \phi_{32}^{(e)}$ ), estimates of partial efficiencies were virtually identical to estimates of partial regression coefficients ( $b_1$  and  $b_2$ ) under the RFI analysis (Figure 2.1) as anticipated from Equation [2.9]. Conversely, under scenarios where partial efficiencies were not the same at the genetic and the corresponding residual levels, (i.e.,  $\phi_{31}^{(u)} \neq \phi_{31}^{(e)}$  and  $\phi_{32}^{(u)} \neq \phi_{32}^{(e)}$ ), partial regression estimates based on the RFI model were intermediate to the two corresponding partial efficiency estimates derived from the MT model; i.e.  $\hat{b}_1$  was intermediate to  $\hat{\phi}_{31}^{(e)}$  and  $\hat{\phi}_{31}^{(u)}$  whereas  $\hat{b}_2$  was intermediate to  $\hat{\phi}_{32}^{(e)}$  and  $\hat{\phi}_{32}^{(u)}$  (Figure 2.2). At any rate, whether or not genetic and residual partial efficiencies were equivalent (Figure 2.1) or non-equivalent (Figure 2.2), one can observe that the MT model provided unbiased estimates of these partial efficiencies.

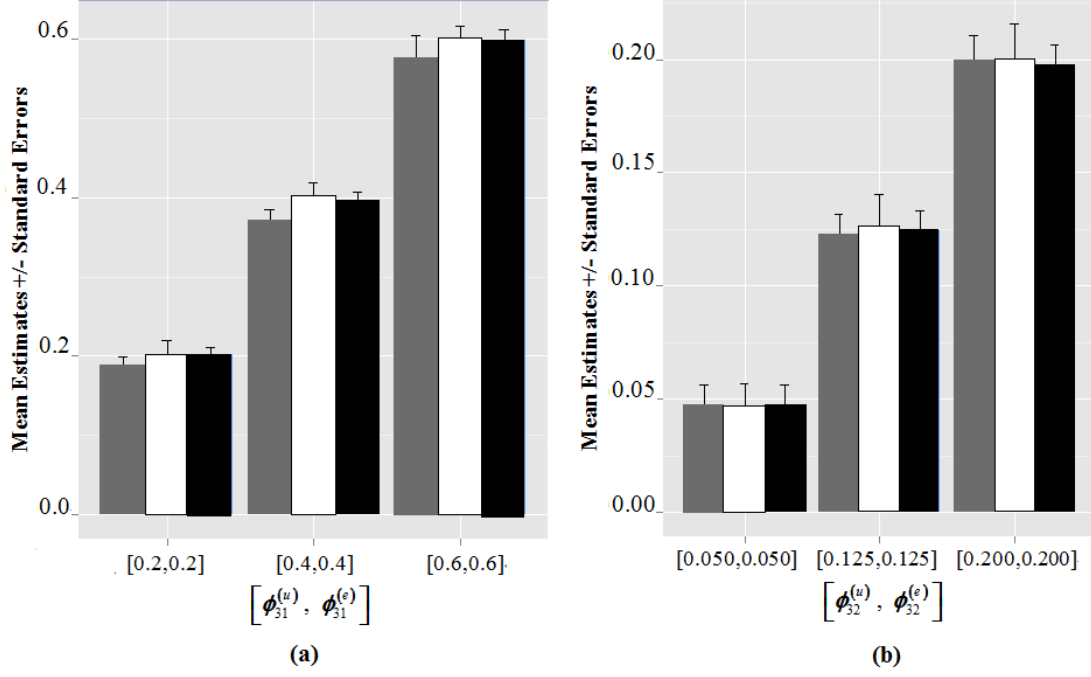


Figure 2.1: Estimated (with standard error bars) mean residual ( $\hat{\phi}_{3j}^{(e)}$  : grey bars) and genetic ( $\hat{\phi}_{3j}^{(g)}$  : white bars) partial efficiencies under proposed multivariate model and estimated partial regression coefficients ( $\hat{b}_j$  : black bars) under RFI model for DMI on (a) MILKE ( $j=1$ ) as a function of couplets of values for  $[\phi_{31}^{(u)} = \phi_{31}^{(e)}]$  and for DMI on (b) MBW ( $j=2$ ) as a function of couplets of values for  $[\phi_{32}^{(u)} = \phi_{32}^{(e)}]$  in response surface design simulation study.

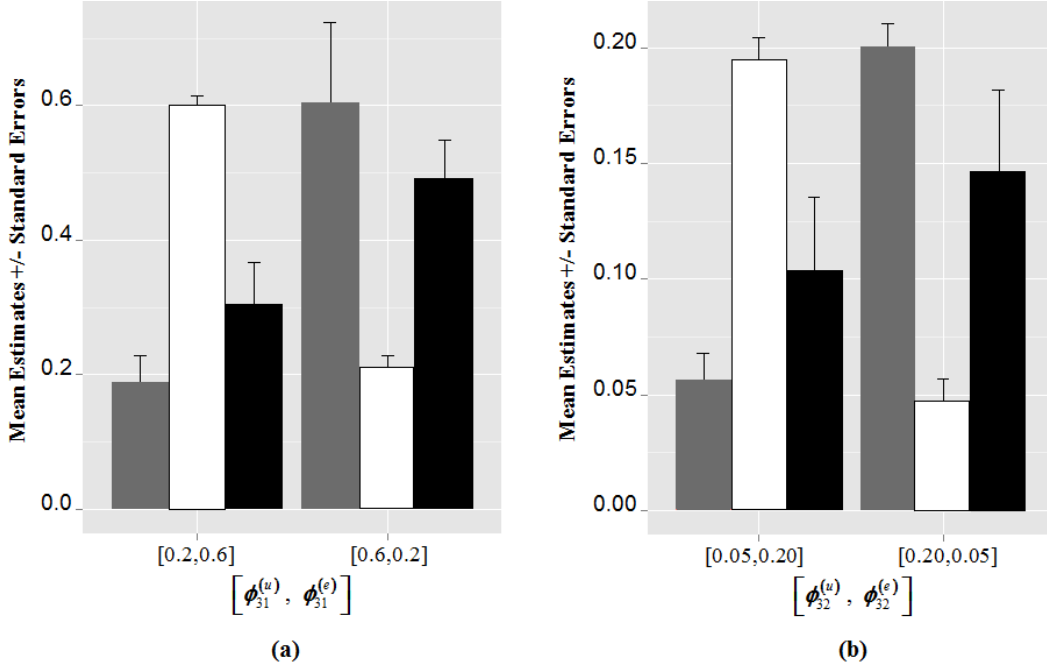


Figure 2.2: Estimated (with standard error bars) mean residual ( $\hat{\phi}_{3j}^{(e)}$  : grey bars) and genetic ( $\hat{\phi}_{3j}^{(g)}$  : white bars) partial efficiencies under proposed multivariate model and estimated partial regression coefficients ( $\hat{b}_j$  : black bars) under RFI model for DMI on (a) MILKE ( $j=1$ ) as a function of couplets of values for  $[\phi_{31}^{(u)} \neq \phi_{31}^{(e)}]$  and for DMI on (b) MBW ( $j=2$ ) as a function of couplets of values for  $[\phi_{32}^{(u)} \neq \phi_{32}^{(e)}]$  in response surface design simulation study.

Under equivalent partial efficiencies at the genetic and residual levels, estimated variance components, and heritabilities were virtually identical for the two modeling approaches, again as anticipated; i.e.,  $\hat{\sigma}_{uRFI}^2 \approx \hat{\sigma}_{u3|1,2}^2$  and  $\hat{\sigma}_{eRFI}^2 \approx \hat{\sigma}_{e3|1,2}^2$  such that  $\hat{h}_{RFI}^2 \approx \hat{h}_{3|1,2}^2$  (Figure 2.3). Conversely, under non-equivalent partial efficiencies, we observed that  $\sigma_{uRFI}^2 \neq \sigma_{u3|1,2}^2$  ( $P < 0.05$ ), whereas no significant differences were observed between  $\sigma_{eRFI}^2$  and  $\sigma_{e3|1,2}^2$  as well as between  $h_{RFI}^2$  and  $h_{3|1,2}^2$ .

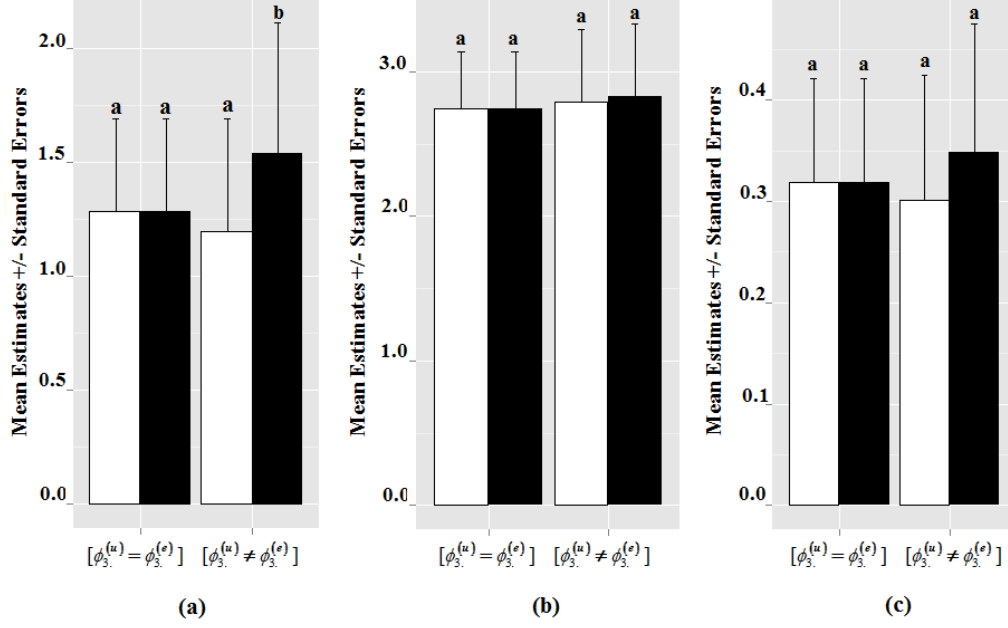


Figure 2.3: Estimated (with standard error bars) mean parameters under RFI (white bars) vs proposed multivariate (black bars) models for (a) genetic variances  $\hat{\sigma}_{u3|1,2}^2$  versus  $\hat{\sigma}_{uRFI}^2$ , (b) residual variances  $\hat{\sigma}_{e3|1,2}^2$  versus  $\hat{\sigma}_{eRFI}^2$  and (c) heritabilities  $\hat{h}_{3|1,2}^2$  versus  $\hat{h}_{RFI}^2$  as a function of equivalent partial efficiencies ( $\phi_3^{(u)} = \phi_3^{(e)}$  defined as  $\phi_{31}^{(u)} = \phi_{31}^{(e)}$  and  $\phi_{32}^{(u)} = \phi_{32}^{(e)}$ ) or nonequivalent partial efficiencies ( $\phi_3^{(u)} \neq \phi_3^{(e)}$  defined as  $\phi_{31}^{(u)} \neq \phi_{31}^{(e)}$  and  $\phi_{32}^{(u)} \neq \phi_{32}^{(e)}$ ) in response surface design simulation. Bars not sharing the same letter within a pair are different from each other ( $P < 0.05$ ).

Differences in accuracies of genetic merit prediction between the proposed MT and classical RFI approaches were determined for their prediction of  $\mathbf{u}_{3|1,2}$ . For the situation where the corresponding partial efficiencies were the same at both the genetic and residual levels, i.e.,  $\phi_{31}^{(u)} = \phi_{31}^{(e)}$  and  $\phi_{32}^{(u)} = \phi_{32}^{(e)}$ , differences in the accuracy of genetic prediction were close to 0 or even slightly higher for the classical RFI model (Figures 2.4 and 2.5). Again, this result was somewhat anticipated given Equation [2.9] because the classical RFI approach is not only mathematically equivalent but also is more parsimonious compared to the proposed MT approach in those specific cases. Nevertheless, when genetic partial efficiencies were substantially different relative to the corresponding residual partial efficiencies, i.e.,  $\phi_{31}^{(u)} \neq \phi_{31}^{(e)}$

and  $\phi_{32}^{(u)} \neq \phi_{32}^{(e)}$ , our proposed MT model fared substantially better (Figures 2.4 and 2.5). In particular, the accuracy of genetic merit prediction was as much as 4 percentage points greater for the proposed approach when  $\phi_{31}^{(u)}$  and  $\phi_{31}^{(e)}$  were most divergent from each other (Figure 2.4). However, the difference for accuracy of genetic merit prediction between the two approaches was far less pronounced across various combinations of  $\phi_{32}^{(u)}$  and  $\phi_{32}^{(e)}$  as shown in Figure 2.5, likely in part due to their smaller range of investigated values relative to  $\phi_{31}^{(u)}$  and  $\phi_{31}^{(e)}$ .

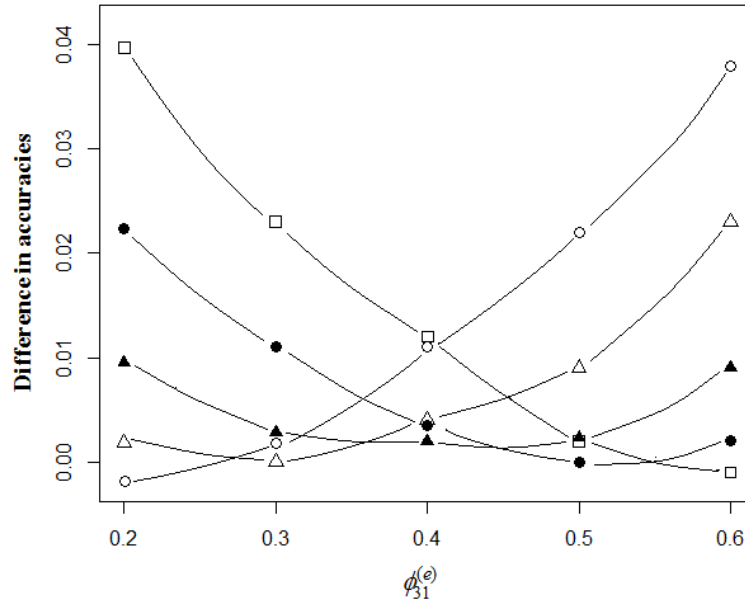


Figure 2.4: Differences in accuracies of genetic merit prediction between proposed multivariate and RFI models versus  $\phi_{31}^{(e)}$  for  $\phi_{31}^{(u)} = 0.2$ (○-○-○),  $0.3$ (△-△-△),  $0.4$  (▲-▲-▲),  $0.5$ (●-●-●), and  $0.6$ (□-□-□) in response surface design simulation study.

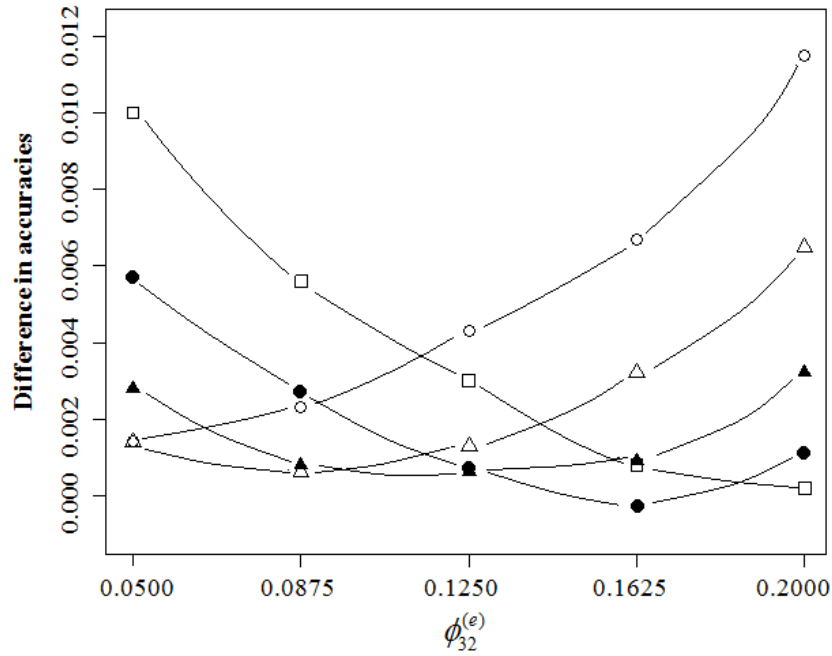


Figure 2.5: Difference in accuracies of prediction of genetic merit between proposed multivariate and RFI models versus  $\phi_{32}^{(e)}$  for  $\phi_{32}^{(u)}$  being 0.0500(○-○-○), 0.0875(△-△-△), 0.1250 (▲-▲-▲), 0.1625(●-●-●), and 0.2000(□-□-□) in response surface design simulation study.

## 2.4.2 Application to Dairy Consortium Data

### 2.4.2.1 Variance Component Estimates

Estimates of variance components, heritabilities, and repeatabilities using each of the two modeling approaches on the dairy consortium data are provided in Table 2.3. Heritability estimates were very similar; the estimate ( $\pm$  standard error) of  $h_{3|,2}^2$  using the proposed MT model was  $0.14 \pm 0.03$ , whereas the estimate of  $h_{RFI}^2$  using the one-stage RFI model was  $0.16 \pm 0.03$ . Repeatability estimates were also very similar (i.e.,  $0.64 \pm 0.01$  versus  $0.65 \pm 0.01$ ) between the two modeling approaches.

Table 2.3: Estimates of variance components, heritabilities, and repeatabilities of rFE from dairy consortium data analyses based on proposed multivariate approach versus RFI model.

	Multivariate Model		RFI Model	
Parameter Category	Parameter	Estimate( $\pm$ Standard Error)	Parameter	Estimate ( $\pm$ Standard Error)
Genetic	$\sigma^2_{u_{3 ,2}}$	0.61 $\pm$ 0.15	$\sigma^2_{u_{RFI}}$	0.70 $\pm$ 0.14
Within-lactation permanent Environment	$\sigma^2_{p_{w,3 ,2}}$	1.77 $\pm$ 0.12	$\sigma^2_{p_{w,RFI}}$	1.80 $\pm$ 0.12
Between-lactation permanent environment	$\sigma^2_{p_{b,3 ,2}}$	0.33 $\pm$ 0.18	$\sigma^2_{p_{b,RFI}}$	0.39 $\pm$ 0.16
Residual	$\sigma^2_{e_{3 ,2}}$	1.57 $\pm$ 0.02	$\sigma^2_{e_{RFI}}$	1.57 $\pm$ 0.02
Heritability	$h^2_{3 ,2}$	0.14 $\pm$ 0.03	$h^2_{RFI}$	0.16 $\pm$ 0.03
Repeatability	$r^2_{3 ,2}$	0.64 $\pm$ 0.01	$r^2_{RFI}$	0.65 $\pm$ 0.01

#### 2.4.2.2 Partial Efficiencies/Regressions

Estimated partial efficiencies based on the proposed MT approach are reported in Table 2.4; key estimates pertain to  $\phi_{31}^{(x)}$  or the partial efficiencies of DMI on MILKE and to  $\phi_{32}^{(x)}$  or the partial efficiencies of DMI on MBW for  $x = u, p_w, p_b$ , and  $e$ . For example, the  $\phi_{31}^{(u)}$  coefficient defines the change in the genetic merit for DMI for every unit change in the genetic merit of MILKE. Thus, these partial efficiencies partly characterize rFE at various levels in that the smaller their values, the more efficient the conversion of feed to MILKE or MBW at the various levels (i.e. genetic effects, within/across-permanent environmental effects or residual effects).

Estimates of  $\phi_{31}^{(x)}$  ranged from 0.32 to 0.52 whereas estimates of  $\phi_{32}^{(x)}$  ranged from 0.04 to 0.16.

Under the RFI analysis, the corresponding partial regression coefficient estimates were  $0.38 \pm 0.005$  for  $b_1$  and  $0.12 \pm 0.004$  for  $b_2$  which were well within the range of the corresponding partial efficiency estimates from the proposed MT analysis.

Table 2.4: Estimates of partial efficiencies based on proposed multivariate model in dairy consortium data analysis

Level <sup>2</sup>	Partial Efficiency Estimates ( $\pm$ Standard Errors)		
	$\phi_{21}^{(\cdot)}$ <sup>1</sup>	$\phi_{31}^{(\cdot)}$	$\phi_{32}^{(\cdot)}$
$\phi_{jk}^{(u)}$	$-0.21 \pm 0.32$	$0.52 \pm 0.08$	$0.14 \pm 0.02$
$\phi_{jk}^{(p_w)}$	$0.37 \pm 0.09$	$0.32 \pm 0.09$	$0.11 \pm 0.01$
$\phi_{jk}^{(p_b)}$	$0.76 \pm 0.41$	$0.45 \pm 0.01$	$0.04 \pm 0.01$
$\phi_{jk}^{(e)}$	$0.13 \pm 0.01$	$0.38 \pm 0.22$	$0.16 \pm 0.09$

<sup>1</sup>Estimated partial efficiencies ( $\pm$  standard errors) of MBW on MILKE ( $\phi_{21}^{(\cdot)}$ ), MBW on MILKE ( $\phi_{31}^{(\cdot)}$ ) and of DMI on MBW ( $\phi_{32}^{(\cdot)}$ ) at the <sup>2</sup>genetic ( $\phi_{jk}^{(u)}$ ), within-lactation permanent environment ( $\phi_{jk}^{(p_w)}$ ) between-lactation permanent environment ( $\phi_{jk}^{(p_b)}$ ) and residual ( $\phi_{jk}^{(e)}$ ) levels.

### 2.4.2.3 Correlation between EBV

The estimated correlation between EBV of rFE for cows having their own records based on the two different modeling approaches was  $cor(u_{3|1,2}, u_{RFI}) = 0.93$ . A scatterplot of the two sets of EBV is provided in Figure 2.6 further demonstrates the high level of agreement between the two different sets of EBV although it is apparent that  $u_{3|1,2}$  demonstrates a slightly greater level of shrinkage relative to  $u_{RFI}$  likely in part to the fact that the estimate of  $\sigma_{u_{3|1,2}}^2$  is slightly smaller than that for  $\sigma_{u_{RFI}}^2$  as provided earlier in Table 2.3.

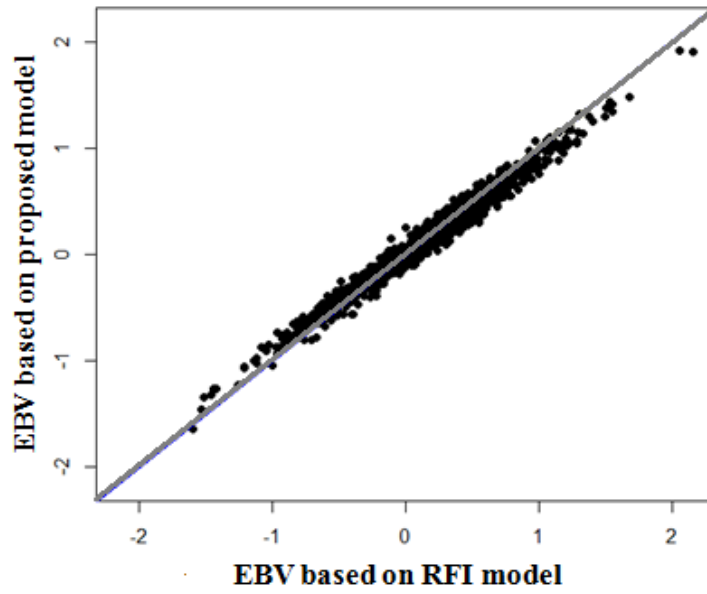


Figure 2.6: Scatterplot of EBV of rFE based on proposed multivariate model ( $\hat{\mathbf{u}}_{3|1,2}$ ) versus RFI ( $\hat{\mathbf{u}}_{RFI}$ ) model of dairy consortium data. Reference line of slope 1 and intercept 0 superimposed

#### 2.4.2.4 Cross-validation Study

The accuracies for the 5-fold cross validation were provided in Table 2.5. There was no evidence of a significant difference in predictive accuracy between the two modeling approaches, whether based on within animal or across animal splits of the data, or even combined.

Table 2.5: Cross Validation Prediction Accuracies for each of 5 different validation datasets (fold) based on proposed (MT) versus RFI analyses of dairy consortium data

	Across animals		Within animals		Combined	
Fold	MT	RFI	MT	RFI	MT	RFI
1	0.85	0.85	0.94	0.94	0.90	0.90
2	0.81	0.81	0.94	0.94	0.88	0.88
3	0.82	0.84	0.94	0.94	0.88	0.89
4	0.86	0.86	0.94	0.94	0.90	0.90
5	0.79	0.80	0.94	0.93	0.84	0.84
Mean	0.83	0.83	0.94	0.94	0.88	0.88

## 2.5 Discussion

We have proposed a modeling strategy for rFE that we believe helps to address the controversy between those researchers that prefer to focus on the analysis of DMI versus those that prefer analyzing RFI for quantitative genetic inference on rFE, recognizing that DMI is easier to understand, even though DMI is not a direct measure of rFE (Berry and Pryce, 2014). In particular, we demonstrated that MT inference involving DMI and the energy sink traits such as MILKE and MBW is equivalent to single trait inference on RFI under one special condition: that the partial efficiencies relating DMI to the energy sink traits are the same at genetic and non-genetic (i.e. permanent environment or residual) levels. However, we would quickly add that even Kennedy et al. (1993) had earlier suggested that selection on RFI is equivalent to multiple trait selection on its components. In fact, we would suggest that our proposed MT approach more closely aligns with the “genotypic regression” approach introduced by Kennedy et al. (1993) which is far more challenging to implement for real data analyses compared to the “phenotypic regression” approach used far more commonly for RFI modeling as also adapted in this paper. More recently, another group (Strathe et al., 2014) also presented a strategy for rFE modeling in pigs which is even more closely aligned to our proposed MT strategy.

We do believe our proposed MT model facilitates a more comprehensive investigation into the relationship between DMI and the key energy sinks (MILKE and MBW) at different levels; i.e., genetic, permanent environmental, and residual levels whereas the classical RFI approach assumes that such relationships are constant across all such levels. Indeed, our simulation results intuitively proved that two modeling approaches have similar accuracy of genetic merit

prediction when partial efficiencies are equivalent at the various levels (i.e.,  $\phi_{..}^{(u)} = \phi_{..}^{(e)}$ ); in fact, it could be effectively argued that RFI modeling would then be preferred since there would be fewer parameters to estimate. Nevertheless, we also demonstrated by simulation that there can be greater accuracies in genetic merit prediction using the proposed MT approach when these partial efficiencies are rather different from each other between the genetic and non-genetic levels (i.e.,  $\phi_{31}^{(u)} \neq \phi_{31}^{(e)}$ ).

Our empirical analysis of a subset of a larger dairy consortium study (Tempelman et al., 2015) suggested no real apparent advantage of our proposed approach relative to classical RFI modeling. This result again may be due to the fact that the partial efficiencies may not have demonstrated enough heterogeneity across the various genetic and non-genetic levels to warrant a greater cross-validation prediction accuracy for the proposed approach. However, we believe the proposed MT approach warrants future consideration and research for several reasons.

Firstly, a MT approach facilitates the incorporation of all data on animals that might have missing records on DMI or some of the energy sink components of rFE; conversely a RFI modeling strategy would require deletion of all data from any time period missing a record on any rFE component trait. Hence, an even greater accuracy of selection could be demonstrated based on our approach if such data were incorporated. Furthermore, the issue of ‘missingness’ is actually more subtle for rFE modeling because all component traits are generally not measured with the same level of frequency. For instance, we summarized records on all of the traits to a weekly basis just as we did previously, even though recording frequencies did not only differ among traits within each research station, they also differed among research stations for the same

traits(Tempelman et al., 2015). Our proposed approach thereby provides an opportunity to better formally account for these recording frequency differences.

Secondly, the proposed MT approach is not any more difficult to implement than any standard MT analysis. Variance-covariance component matrices can be readily estimated using REML in mixed model software followed by use of the CD to estimate the corresponding partial efficiencies. In fact, as previously noted, ASREML software (Gilmour et al., 2009) has the “ANTE” option that facilitates the direct estimation of partial efficiencies and SCVC, including their estimated standard errors using the average information matrix. With  $\hat{\mathbf{u}}_i$  denoting a standard MT-based EBV for MILKE, MBW, and DMI, the estimated genetic partial efficiencies ( $\hat{\mathbf{T}}_u$ ) can be used in a rearrangement of Equation [2.6] to backsolve  $\hat{\mathbf{u}}_i^* = (\mathbf{I} - \hat{\mathbf{T}}_u)\hat{\mathbf{u}}_i$  for the BLUP of  $\hat{u}_{i,3|1,2}$  being the last element of  $\hat{\mathbf{u}}_i^*$ , and which we propose as our genetic merit for rFE.

Note that a MT analysis is an important prerequisite for forming a selection index on the corresponding traits. Suppose that the relative  $\nu_1$ ,  $\nu_2$ , and  $\nu_3$  are known for MILKE, MBW, and DMI, respectively, such that a portion of the selection index, at least as it pertains to just these 3 traits, is  $\nu_1 u_{i,1} + \nu_2 u_{i,2} + \nu_3 u_{i,3}$  for animal  $i$  based on a MT analysis. Noting that a selection index would involve DMI or rFE, but not both, the selection index for animal  $i$  could be re-expressed as a function of its EBV for MILKE, MBW, and our proposed measure of rFE as  $(\nu_1 + \nu_3 \phi_{31}^{(u)})u_{i,1} + (\nu_2 + \nu_3 \phi_{32}^{(u)})u_{i,2} + \nu_3 u_{i,3|1,2}$  using Equation [2.5c]. So, for example, if  $\nu_1 > 0$ ,  $0 < \phi_{31}^{(u)} < 1$  (as based on our results),  $\nu_3 < 0$ , and  $\nu_1 > |\nu_3|$  as expected, then the ratio of the economic weight  $(\nu_1 + \nu_3 \phi_{31}^{(u)})$  of MILKE relative to that  $(\nu_3)$  on rFE should be smaller compared

with the same ratio involving the weight ( $v_1$ ) on MILKE relative to that ( $v_3$ ) on DMI in a selection index. In other words, rFE would be weighted more highly (in absolute value) relative to MILKE as compared with DMI relative to MILKE in a selection index. Given the previously noted similarities between our rFE measure and RFI, a similar comparison could be drawn between the economic weights of DMI and RFI relative to MILKE in alternative expressions of the same selection index.

Finally, our approach allows us to better formally investigate any potential heterogeneity in these partial efficiencies at the various levels across, say, research stations or even diets; this would involve further extending work by Bello et al. (2010) who similarly investigated heterogeneities in the associations between milk production and reproduction at the contemporary group and cow levels.

We failed to include  $\Delta BW$  as a fourth trait in our proposed model given its very high variability and low heritability, even though it is conceptually straightforward to do so. Hence, we treated BW as a covariate in both our proposed MT approach and the RFI model such that both sets of analysis on the dairy data implied a common partial efficiency or regression of DMI on  $\Delta BW$  at all genetic and non-genetic levels. Better approaches to deal with  $\Delta BW$  within the context of our proposed model warrant future research. We also did not report the effect of  $\phi_{21}^{(x)}$ ; i.e. the partial efficiency of MBW on MILKE at level  $x = u$  or  $e$  in our MT vs RFI modeling comparisons because we observed that neither the magnitude nor the range of the discrepancies between  $\phi_{21}^{(u)}$  and  $\phi_{21}^{(e)}$  had any impact (results not reported).

It has been shown that higher accuracies and larger genetic gains can be achieved by using genomic selection for rFE traits (deHaas et al., 2014). Even though we used pedigree-based BLUP in this paper, it is conceptually rather easy to substitute the numerator relationships matrix with a genomic relationship matrix or even a H relationship matrix that facilitates the use of genotyped and non-genotyped animals (Aguilar et al., 2010). We believe genomic extensions of our proposed model will be promising and may facilitate further discoveries as to which genes are important for rFE.

## **2.6 Conclusions**

In this study, an alternative MT model strategy based on a modified Cholesky decomposition for modeling rFE in dairy cattle was proposed as an alternative to the conventional RFI modeling approach. Both strategies were further studied and compared using simulation and application to a subset of dairy research consortium data. We demonstrated the equivalence between proposed and RFI modeling strategies when partial efficiencies between DMI with any of the energy sink traits were equivalent at all levels whereas the proposed MT approach had up to 4 percentage points greater accuracy in genetic merit prediction when there were large discrepancies between these partial relationships at genetic versus non-genetic levels. Application to the dairy research consortium data did not suggest any significant difference in cross-validation predictive accuracy between the two approaches. Nevertheless, we believe our proposed approach warrants further consideration in future research, especially when data on some of the rFE component traits are missing.

## **Chapter3: Modeling Genetic and Non-Genetic Variation of Feed Efficiency and its Partial Relationships between Component Traits as a Function of Management and Environmental Factors**

### **3.1 Abstract**

Residual feed efficiency (rFE), characterized as the fraction of feed nutrients converted into milk or meat, is of increasing economic importance in the dairy industry. We conjecture that rFE is a complex trait whose variation and relationships or partial efficiencies (PE) involving the conversion of DMI to milk energy and metabolic body weight may be highly heterogeneous across environments or management scenarios. In this study, a hierarchical Bayesian multivariate mixed model was proposed to jointly infer upon such heterogeneity at both genetic and non-genetic levels on PE and variance components (VC). The heterogeneity was modeled by embedding mixed effects specifications on PE and VC in addition to those directly specified on the component traits. We validated the model by simulation and applied it to a joint analysis of a dairy rFE consortium dataset with 5,088 Holstein cows from 13 research stations in Canada, the Netherlands, the United Kingdom, and the United States. Although no differences were detected among research stations for PE at the genetic level, there was some evidence of heterogeneity in residual PE. Furthermore, substantial heterogeneity in VC across stations, parities, and ration were observed with heritability estimates of rFE ranging from 0.16 to 0.46 across stations.

### **3.2 Introduction**

Residual feed efficiency (rFE) is becoming more important for the economic and environmental sustainability of dairy production and increases as a greater proportion of feed nutrients are directed towards milk production (Connor, 2015). A commonly used measure of

rFE is residual feed intake (RFI) which is defined as the difference between actual and predicted DMI. That is, RFI “responses” are typically derived as the estimated residuals from a linear model analysis whereby partial regression relationships are specified between DMI and energy sink covariates such as milk energy (MILKE) and metabolic body weight (MBW), for example. Given that there has been some reluctance to directly incorporate RFI in breeding goals because RFI can be a difficult concept to understand and explain and there are possible genetic antagonisms with other performance traits (Berry and Pryce, 2014), some investigators have focused their attention on DMI as the key phenotype for rFE analyses (Berry et al., 2014, de Haas et al., 2015).

Recently, Lu et al. (2015) proposed a multiple trait (MT) mixed model analysis of DMI with MILKE and MBW that further resolves whether RFI or DMI should be considered as the key response variable for rFE. They demonstrated that Cholesky decompositions performed on each of the estimated (e.g., by REML) 3 x 3 genetic (**G**) and residual (**R**) variance-covariance matrices among the 3 key traits (MILKE, MBW, and DMI) lead to a parameterization whereby the estimated partial regression relationships between DMI and MILKE, and between DMI and MBW are essentially partitioned into genetic and residual components. Additionally, the Cholesky decomposition leads to a determination of EBV for rFE that is identical to that based on a classical RFI analysis under the special case whereby the partial regression relationships relating DMI to MILKE and to MBW are specified to be identical at both genetic and residual levels. As a corollary, Lu et al. (2015) demonstrated by simulation that the greater the discrepancy between the partial relationships at genetic and non-genetic levels, the greater the EBV accuracy for their proposed MT approach relative to a classical RFI analysis. Furthermore, the MT approach facilitates the incorporation of data on cows that might be missing, even

selectively, on any of the 3 key phenotypes that would otherwise be discarded in a RFI analysis (Pollak et al., 1984).

A hierarchical Bayesian extension of a 2-trait MT model was developed earlier in a non-genetic context by (Bello et al., 2010) who later inferred heterogeneous partial regression relationships between calving intervals and milk yield at both herd and cow levels as a function of environmental and herd management factors for Michigan dairy herds. We surmise that the genetic and residual partial relationships between DMI and MILKE, or between DMI and MBW could also be modeled as a linked multifactorial function of various factors including parities, research stations, and rations, for example. Indeed, based on station-specific RFI analyses, Tempelman et al. (2015) determined that estimated partial regression coefficients of DMI on MILKE and on MBW were highly heterogeneous across research stations in several countries. Furthermore, adaptation of the hierarchical Bayesian approach as proposed by (Bello et al., 2010) would also infer the degree of heterogeneity in heritabilities of rFE across different management conditions.

The 2 primary objectives of this study were 1) to identify potential management or environmental factors that might impact genetic and residual partial regression relationships (i.e., PE) between DMI and MILKE, and between DMI and MBW, and 2) to infer whether there is evidence of heterogeneity of genetic and residual variances for rFE across these same management or environmental factors.

### 3.3 Materials and Methods

#### 3.3.1 The Multiple Trait Model

Our study closely combines developments provided in Lu et al. (2015) with those provided in (Bello et al., 2010). The 3 key component traits of rFE are numbered as follows in the MT model: 1) MILKE, 2) MBW, and 3) DMI. We write this MT model as follows:

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{u} + \mathbf{e}_i. \quad [3.1]$$

Here  $\mathbf{y}_i = [y_{i1} \ y_{i2} \ y_{i3}]'$  is the vector of responses for the 3 traits on record or animal  $i$ ,  $i = 1, 2, 3, \dots, n$ . Furthermore,  $\boldsymbol{\beta} = [\boldsymbol{\beta}_1' \ \boldsymbol{\beta}_2' \ \boldsymbol{\beta}_3']'$  is the vector of fixed effects connected to  $\mathbf{y}_i$  by known incidence matrix  $\mathbf{X}_i = (\mathbf{I}_3 \otimes \mathbf{x}_i')$  such that  $\boldsymbol{\beta}_j$  denotes the subvector of fixed effects for trait  $j$ ,  $j = 1, 2, 3$ . Note that  $\otimes$  denotes the Kronecker product such that we assume the same fixed effects incidence row vector  $\mathbf{x}_i'$  for each of the 3 traits for ease of presentation, although further generalization is possible. Similarly,  $\mathbf{u} = [\mathbf{u}_{\cdot 1}' \ \mathbf{u}_{\cdot 2}' \ \mathbf{u}_{\cdot 3}']'$  is the vector for animal genetic effects connected to  $\mathbf{y}_i$  by known incidence matrix  $\mathbf{Z}_i = (\mathbf{I}_3 \otimes \mathbf{z}_i')$  such that  $\mathbf{u}_{\cdot j} = \{u_{ij}\}_{i=1}^n$  denotes the subvector of random genetic effects on trait  $j$  for all  $n$  animals. Hence, we also assume the same random effects incidence row vector  $\mathbf{z}_i'$  for each of the 3 traits, although again further generalization is possible. To further simplify presentation, we specifically focus on the situation where there is 1 record per animal, and genetic merit is explicitly modeled only for animals having records, although extensions to the more common situation where genetic evaluations are also desired on animals without any records or multiple records is readily apparent. Finally,  $\mathbf{e}_i = [e_{i1} \ e_{i2} \ e_{i3}]'$  denotes the sub-vector of residuals for the 3 traits on animal  $i$ . Now  $\mathbf{e}_i$  is assumed to be independently multivariate normal across animals; i.e.,  $\mathbf{e}_i \sim MVN(\mathbf{0}, \mathbf{R}_i)$  where

$\mathbf{R}_i$  denotes the 3 x 3 residual (co)variance matrix among 3 traits specific for animal  $i$ .

Now  $\mathbf{u}$  can be alternatively reordered by traits within animals; i.e.,  $\mathbf{u} = [\mathbf{u}_{.1} \quad \mathbf{u}_{.2} \quad \cdots \quad \mathbf{u}_{.n}]$  with  $\mathbf{u}_{.i} = \{u_{ij}\}_{j=1}^3$  denoting the vector of random genetic effects for the three traits on animal  $i$ . Because of potential correlations between animal effects due to, for example, the numerator relationship matrix  $\mathbf{A}$ , the breeding values  $\mathbf{u}_i$  and  $\mathbf{u}_{i'}$  for animals  $i$  and  $i'$  ( $i \neq i'$ ) are not necessarily independent of each other. To flexibly allow for such correlation as well as for subject-specific genetic trait variances and covariances as developed later, we invoke a Cholesky decomposition on  $\mathbf{A}$ , writing  $\mathbf{A} = \mathbf{C}\mathbf{C}'$ . For clarity of presentation, we specify the dimension of  $\mathbf{A}$  as  $n \times n$  such that it only pertains to relationships between animals with records; nevertheless, its construction should be suitably based on all available ancestor information. In other words, we define  $\mathbf{A}$  as a subset of a larger numerator relationship matrix that involves both animals with records and their ancestors. Nevertheless, if genetic evaluations are also desired on ancestors, the dimensions of both  $\mathbf{A}$  and  $\mathbf{z}_i' \forall i$  can be augmented accordingly. We define  $\mathbf{u}_{.j}^*$  using  $\mathbf{u}_{.j} = \mathbf{C}\mathbf{u}_{.j}^*$  such that  $\mathbf{u}^* = [\mathbf{u}_{.1}^* \quad \mathbf{u}_{.2}^* \quad \mathbf{u}_{.3}^*]'$  is ordered by animals within traits in the same manner as  $\mathbf{u} = [\mathbf{u}_{.1}' \quad \mathbf{u}_{.2}' \quad \mathbf{u}_{.3}']'$ . Alternatively, if elements of  $\mathbf{u}^*$  are ordered by elements traits within animals, i.e.,  $\mathbf{u}^* = [\mathbf{u}_{1.}^* \quad \mathbf{u}_{2.}^* \quad \cdots \quad \mathbf{u}_{n.}^*]$ , then we specify  $\mathbf{u}_{i.}^* = \{u_{ij}^*\}_{j=1}^3 \overset{\text{independent}}{\sim} MVN(\mathbf{0}, \mathbf{G}_i)$ , for the 3 traits on animal  $i$ . That is,  $\mathbf{G}_i$  denotes the 3 x 3 genetic (co)variance matrix among the 3 traits being unique to the environmental or management circumstances peculiar to animal  $i$ . It is important to further clarify, using results from Kennedy et al. (1988), that  $\mathbf{u}_{i.}^*$  actually defines the genetic merit on the three traits for animal  $i$  only if it is a base population animal whereas  $\mathbf{u}_{i.}^*$  defines the Mendelian sampling effects on the three traits for animal  $i$  if it is a not a base

population animal. We thereby rewrite Equation [3.1] as

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i (\mathbf{I}_3 \otimes \mathbf{C}) \mathbf{u}^* + \mathbf{e}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i^* \mathbf{u}^* + \mathbf{e}_i \quad [3.2]$$

with  $\mathbf{Z}_i^* = \mathbf{Z}_i (\mathbf{I}_3 \otimes \mathbf{C}) = \mathbf{I}_3 \otimes \mathbf{z}^*_{i\cdot}$  such that  $\mathbf{z}^*_{i\cdot} = \mathbf{z}'_i \mathbf{C}$ , and all other terms defined as before.

Mirroring what was presented previously, albeit in a non-genetic context (Bello et al., 2012) and further extending Lu et al. (2015), we write the Cholesky decomposition for  $\mathbf{G}_i$  as  $\mathbf{G}_i = (\mathbf{I}_3 - \mathbf{T}_i^u)^{-1} \boldsymbol{\Delta}_i^u (\mathbf{I}_3 - \mathbf{T}_i^u)^{-1'}$ , where  $\mathbf{I}_3$  is an identity matrix of order 3,  $\boldsymbol{\Delta}_i^u = \text{diag}(\sigma_{u_{(1)i}}^2, \sigma_{u_{(21)i}}^2, \sigma_{u_{(31,2)i}}^2)$  is a diagonal matrix of sequentially conditional variance components (VC) and  $\mathbf{T}_i^u = \begin{bmatrix} 0 & 0 & 0 \\ \phi_{(21)i}^u & 0 & 0 \\ \phi_{(31)i}^u & \phi_{(32)i}^u & 0 \end{bmatrix}$  is a lower triangular matrix such that  $\boldsymbol{\Delta}_i^u$  and  $\mathbf{T}_i^u$  are modeled as being potentially unique to the environmental or management circumstances pertaining to animal  $i$ . Now  $\boldsymbol{\Delta}_i^u = \text{var}(\boldsymbol{\delta}_i)$  where  $\boldsymbol{\delta}_i = [u_{i(1)}^* \ u_{i(21)}^* \ u_{i(31,2)}^*]'$  is based on the specification  $\mathbf{u}_i^* = \mathbf{T}_i^u \mathbf{u}_i^* + \boldsymbol{\delta}_i$  whose elements can be written recursively Lu et al. (2015) as

$$u_{i1}^* = u_{i(1)}^* \quad [3.3a]$$

$$u_{i2}^* = \phi_{(21)i}^u u_{i(1)}^* + u_{i(21)}^* \quad [3.3b]$$

$$u_{i3}^* = \phi_{(31)i}^u u_{i1}^* + \phi_{(32)i}^u u_{i2}^* + u_{i(31,2)}^* \quad [3.3c]$$

For instance,  $u_{i(21)}^*$  is used to represent the genetic effect of MBW conditional on MILKE for animal  $i$  whereas  $u_{i(31,2)}^*$  pertains to the genetic effect of DMI conditional on MILKE and MBW and which we have previously proposed as the genetic effect for rFE for animal  $i$  (Lu et al., 2015). Similarly, we define  $\sigma_{u_{(31,2)i}}^2$  is the Mendelian sampling genetic variance of rFE for animal  $i$ . referred to  $\phi_{(21)i}^u$ ,  $\phi_{(31)i}^u$ , and  $\phi_{(32)i}^u$  as partial efficiencies (PE) at the genetic level. We further extend their work using Bello et al. (2010) to allow for the possibility that these genetic PE are potentially heterogeneous across cows as manifested by the additional subscript  $i$ .

Lu et al. (2015) likewise demonstrated  $\mathbf{R}_i = (\mathbf{I}_3 - \mathbf{T}_i^e)^{-1} \mathbf{\Delta}_i^e (\mathbf{I}_3 - \mathbf{T}_i^e)^{-1}$ , where  $\mathbf{\Delta}_i^e = \text{diag}(\sigma_{e_{(1)i}}^2, \sigma_{e_{(2)i}}^2, \sigma_{e_{(31,2)i}}^2)$  is a diagonal matrix and  $\mathbf{T}_i^e = \begin{bmatrix} 0 & 0 & 0 \\ \phi_{(21)i}^e & 0 & 0 \\ \phi_{(31)i}^e & \phi_{(32)i}^e & 0 \end{bmatrix}$  is a lower triangular matrix except that here both  $\mathbf{\Delta}_i^e$  and  $\mathbf{T}_i^e$  are potentially unique to the environmental circumstances peculiar to animal  $i$  following Bello et al. (2010). Now  $\mathbf{\Delta}_i^e$  is the covariance matrix of  $\mathbf{\varepsilon}_i = [e_{(1)i} \ e_{(2)i} \ e_{(31,2)i}]'$  as defined within  $\mathbf{e}_i = \mathbf{T}_i^e \mathbf{e}_i + \mathbf{\varepsilon}_i$  whose elements can be demonstrated to be written recursively, using Lu et al. (2015), as

$$e_{i1} = e_{i(1)} \quad [3.4a]$$

$$e_{i2} = \phi_{(21)i}^e e_{i1} + e_{i(2)} \quad [3.4b]$$

$$e_{i3} = \phi_{(31)i}^e e_{i1} + \phi_{(32)i}^e e_{i2} + e_{i(31,2)} \quad [3.4c]$$

Here we refer to  $\phi_{(21)i}^e$ ,  $\phi_{(31)i}^e$ , and  $\phi_{(32)i}^e$  as PE at the residual level, again being unique to animal  $i$ .

Equations [3.3c] and [3.4c] are particularly key for defining rFE in that they specify the key relationships between the energy source (DMI) and energy sink traits (MILKE and MBW) at both the genetic (i.e.,  $\phi_{(31)i}^u$ , and  $\phi_{(32)i}^u$ ) and residual ( $\phi_{(31)i}^e$ , and  $\phi_{(32)i}^e$ ) levels. For example,  $\phi_{(31)i}^u$  defines the change in  $u_{i3}$  for every unit change in  $u_{i1}$  holding constant  $u_{i2}$  whereas  $\phi_{(32)i}^u$  defines the change in  $u_{i3}$  for every unit change in  $u_{i2}$  holding constant  $u_{i1}$  on animal  $i$ . Furthermore,  $\phi_{(31)i}^e$  and  $\phi_{(32)i}^e$  are similarly interpreted but at the residual or non-genetic level with respect to animal  $i$ . In essence, heterogeneous PE at the genetic or residual levels indirectly imply heterogeneous genetic and residual correlations between traits across animals as determined by different environments and management factors.

As in Bello et al. (2010), we specify a structural mixed effects model on each of these PE. For example, at the genetic level, we write

$$\phi_{(r)i}^u = \mathbf{x}_{(r)i}^u{}' \boldsymbol{\gamma}_{(r)}^u + \mathbf{z}_{(r)i}^u{}' \mathbf{m}_{(r)}^u \quad [3.5]$$

with  $r = 21, 31$ , or  $32$  denoting indices representing PE relationships between MBW and MILKE ( $r = 21$ ), between DMI and MILKE ( $r = 31$ ) and between DMI and MBW ( $r = 32$ ) as per Equations [3.3b] and [3.3c]. Here  $\boldsymbol{\gamma}_{(r)}^u$  represents a  $p_{u(r)} \times 1$  vector of unknown “fixed effects” connected to  $\phi_{(r)i}^u$  by known incidence row vector  $\mathbf{x}_{(r)i}^u{}'$  whereas  $\mathbf{m}_{(r)}^u \sim N(\mathbf{0}, \mathbf{I}\sigma_{m_u(r)}^2)$  represents a  $q_{u(r)} \times 1$  vector of unknown “random effects” connected to  $\phi_{(r)i}^u$  by known incidence row vector  $\mathbf{z}_{(r)i}^u{}'$ . In this context, “fixed effects” pertain to those effects characterized by either non-informative or vaguely informative prior distributions that do not presume exchangeability amongst the elements of  $\boldsymbol{\gamma}_{(r)}^u$ . Strategically, “random effects” specifications are often best utilized for factors characterized by a large number of subclasses, with each subclass characterized by relatively limited data (Bello et al., 2010).

We similarly specify a structural mixed effects model for the residual PE  $\phi_{(21)i}^e$ ,  $\phi_{(31)i}^e$ , and  $\phi_{(32)i}^e$  specific to observation  $i$

$$\phi_{(r)i}^e = \mathbf{x}_{(r)i}^e{}' \boldsymbol{\gamma}_{(r)}^e + \mathbf{z}_{(r)i}^e{}' \mathbf{m}_{(r)}^e \quad [3.6]$$

with  $r = 21, 31$ , or  $32$  as before as per Equations [3.4b] and [3.4c]. Here,  $\boldsymbol{\gamma}_{(r)}^e$  denotes a  $p_{e(r)} \times 1$  vector of fixed effects connected to  $\phi_{(r)i}^e$  by known incidence row vector  $\mathbf{x}_{(r)i}^e{}'$  whereas  $\mathbf{m}_{(r)}^e \sim N(\mathbf{0}, \mathbf{I}\sigma_{m_e(r)}^2)$  represent a  $q_{e(r)} \times 1$  vector of random effects connected to  $\phi_{(r)i}^e$  by  $\mathbf{z}_{(r)i}^e{}'$  where terms are similarly interpreted as with genetic PE.

Following Bello et al. (2010), each VC specific to animal  $i$  in  $\Delta_i^u = \text{diag}(\sigma_{u_{(1)i}}^2, \sigma_{u_{(2)1i}}^2, \sigma_{u_{(3)1,2i}}^2)$  is also specified as a function of environmental or management factors in a structural mixed effects model:

$$\log(\sigma_{u(s)i}^2) = \mathbf{x}_{(s)i}^u{}' \log(\boldsymbol{\tau}_{(s)}^u) + \mathbf{z}_{(s)i}^u{}' \log(\mathbf{v}_{(s)}^u) \quad [3.7]$$

where  $s = 1, 2|1$ , and  $3|1,2$  is used to denote, respectively, the genetic variance for MBW, the genetic variance for MBW conditional on MILKE, and the genetic variance for DMI conditional on MILKE and MBW. Here  $\boldsymbol{\tau}_{(s)}^u$  is a  $p_{u(s)} \times 1$  vector of fixed effects connected to  $\sigma_{u(s)i}^2$  by known incidence row vector  $\mathbf{x}_{(s)i}^u{}'$  whereas  $\mathbf{v}_{(s)}^u$  is a  $q_{u(s)} \times 1$  vector of random effects connected to  $\sigma_{u(s)i}^2$  by known incidence row vector  $\mathbf{z}_{(s)i}^u{}'$ . Independent inverted Gamma priors  $IG(\eta_{(s)}^u, \eta_{(s)}^u - 1)$  are assigned to each element of  $\mathbf{v}_{(s)}^u = \{v_{(s)d}^u\}_{d=1}^{q_{u(s)}}$  such that  $E(v_{(s)d}^u) = 1$  with coefficient of variation (CV) of  $CV(\mathbf{v}_{(s)}^u) = \frac{1}{\sqrt{\eta_{(s)}^u - 2}}$  as previously demonstrated by Kizilkaya and Tempelman (2005) and Bello et al. (2010). That is,  $CV(\mathbf{v}_{(s)}^u)$  characterizes the standard deviation of subclass-specific variances expressed relative to the average subclass variance.

Finally, the VC in  $\Delta_i^e = \text{diag}(\sigma_{e(1)i}^2, \sigma_{e(2|1)i}^2, \sigma_{e(3|1,2)i}^2)$  are similarly modeled with a mixed effects specification:

$$\log(\sigma_{e(s)i}^2) = \mathbf{x}_{(s)i}^e{}' \log(\boldsymbol{\tau}_{(s)}^e) + \mathbf{z}_{(s)i}^e{}' \log(\mathbf{v}_{(s)}^e) \quad [3.8]$$

where  $s = 1, 2|1$ , and  $3|1,2$  is used to denote, respectively, the residual variance for MBW, the residual variance for MBW conditional on MILKE, and the residual variance for DMI conditional on MILKE and MBW. Here  $\boldsymbol{\tau}_{(s)}^e$  denotes a  $p_{e(s)} \times 1$  vector of fixed effects connected to  $\sigma_{e(s)i}^2$  by known incidence row vector  $\mathbf{x}_{(s)i}^e{}'$  whereas  $\mathbf{v}_{(s)}^e$  represents a  $q_{e(s)} \times 1$  vector of

random effects connected to  $\sigma_{e(s)i}^2$  by known incidence row vector  $\mathbf{z}_{(s)i}^e$  . Here, elements of

$\mathbf{v}_{(s)}^{e'} = \left\{ v_{(s)l}^e \right\}_{l=1}^{q_{e(s)}}$  are assumed to be random independent draws from  $IG(\eta_{(s)}^e, \eta_{(s)}^e - 1)$  such that

$$E\left(v_{(s)l}^e\right) = 1 \text{ and } CV\left(\mathbf{v}_{(s)}^e\right) = \frac{1}{\sqrt{\eta_{(s)}^e - 2}}.$$

Following Bello et al. (2010) and Kizilkaya and Tempelman (2005), vaguely informative priors are specified on  $\eta_{(s)}^e$  and  $\eta_{(s)}^u$  as follows:

$$\eta_{(s)}^u \sim p\left(\eta_{(s)}^u\right) = \frac{1}{\left(1 + \eta_{(s)}^u\right)^2} \quad [3.9]$$

$$\eta_{(s)}^e \sim p\left(\eta_{(s)}^e\right) = \frac{1}{\left(1 + \eta_{(s)}^e\right)^2} \quad [3.10]$$

for  $s = 1, 2|1$ , and  $3|1,2$ .

For all analyses in this paper, flat unbounded priors are specified on  $\beta_1, \beta_2, \beta_3, \gamma_{(21)}^u, \gamma_{(31)}^u, \gamma_{(32)}^u, \gamma_{(21)}^e, \gamma_{(31)}^e, \gamma_{(32)}^e, \tau_{(1)}^u, \tau_{(2|1)}^u, \tau_{(3|1,2)}^u, \tau_{(1)}^e, \tau_{(2|1)}^e, \tau_{(3|1,2)}^e, \sigma_{m_u(21)}^2, \sigma_{m_u(31)}^2, \sigma_{m_u(32)}^2, \sigma_{m_e(21)}^2, \sigma_{m_e(31)}^2$  and  $\sigma_{m_e(32)}^2$ , as also adapted by Bello et al. (2010), although vaguely or even rather informative priors could be specified as well. Due to these flat prior specifications, it is imperative to invoke identifiability restrictions on  $\gamma_{(21)}^u, \gamma_{(31)}^u, \gamma_{(32)}^u, \gamma_{(21)}^e, \gamma_{(31)}^e, \gamma_{(32)}^e, \tau_{(1)}^u, \tau_{(2|1)}^u, \tau_{(3|1,2)}^u, \tau_{(1)}^e, \tau_{(2|1)}^e$ , and  $\tau_{(3|1,2)}^e$  similar to those commonly invoked for fixed classification factors (i.e.,  $\beta_1, \beta_2$ , and  $\beta_3$ ) in classical linear model specifications. In that regard, we utilize the corner parameterization (Clayton, 1996, Kizilkaya and Tempelman, 2005) also referred to as the set-to-zero restriction (Milliken and Johnson., 2009) and as also used by Bello et al. (2010) whereby an overall

intercept is specified and the effect corresponding to one arbitrarily chosen level of each fixed effects factor is ignored or “zeroed out”.

Full conditional densities (FCD) of all unknown (hyper)parameters as required for conducting Markov Chain Monte Carlo (MCMC) Bayesian analysis are provided in the Appendix B.

### **3.3.2 Application to Dairy Consortium Data**

#### **3.3.2.1 Data Source and Editing**

A dataset of MILKE, DMI, and MBW on 5,088 Holstein cows was collected on 13 research stations and studies from the United Kingdom, the Netherlands, Canada, and the United States (US); data sources and editing procedures used for this study are extensively characterized elsewhere (Berry et al., 2014, Lu et al., 2015, Tempelman et al., 2015, Manzanilla-Pech et al., 2016). Phenotypes on all 3 traits were further condensed into 42-d records by taking the average of the first 6 weekly records on all 3 traits between 50 and 200 DIM. If data from multiple lactations were available on individual cows, only the earlier lactation was chosen. One research station was in Canada, being the University of Alberta (AB). Six stations were in the US, specifically, Iowa State University (ISU), Michigan State University (MSU), the University of Florida (UF), the University of Wisconsin-Madison (UW), the USDA Dairy Forage Research Center (USDFRC) in Madison, WI, and the USDA Beltsville Agricultural Research Center (BARC) in Beltsville, MD. Four research stations were in the Netherlands including an experimental herd ‘t Gen (TGEN) in Lelystad, the Nij Bosma Zathe (NBZ) herd located near Leeuwarden, a third study (ZOM) based on the work by Zom et al. (2012) ; and a compilation of studies (NLN) based on data collected from various nutritional experiments, all of which were

previously characterized in detail by Tempelman et al. (2015). The remaining 2 herds were from the United Kingdom, the Langhill (LAN) farm near Edinburgh from 1992 to 2001 and from the Scottish Agricultural College (SAC) Dairy Research Centre based at Crichton Royal Farm near Dumfries with data collection from 2003 to 2011; again, these data are also described extensively elsewhere (Tempelman et al., 2015). A summary of the number of cows by research station/study is provided in Table 3.1.

Table 3.1: Distribution of cows by region.

Region	Station	No. of cows with phenotypes
Canada	AB	236
United States	UF	175
	ISU	742
	MSU	158
	UW	237
	USDFRC	565
	BARC	286
Netherlands	TGEN	569
	NBZ	99
	ZOM	661
	NLN	329
United Kingdom	LAN	590
	SAC	441
Overall		5,088

<sup>1</sup>AB = University of Alberta; UF = University of Florida; ISU = Iowa State University; MSU = Michigan State University; UW = University of Wisconsin–Madison; USDFRC = USDA Dairy Forages Research Center; BARC = USDA Beltsville Agricultural Research Center; TGEN = ‘t Gen experimental herd, Lelystad, the Netherlands; NBZ= Nij Bosma Zathe, Leeuwarden, the Netherlands; ZOM = data based on the work of Zom et al. (2012); NLN = compilation of studies previously characterized by Tempelman et al. (2015); LAN = Langhill farm, Edinburgh, UK; SAC = Scottish Agricultural College.

The specification for the full hierarchical Bayesian model included the fixed effects of research station and parity (primiparous versus multiparous) for each of  $\tau_{(1)}^u, \tau_{(2|1)}^u, \tau_{(3|1,2)}^u, \tau_{(1)}^e, \tau_{(2|1)}^e$ , and  $\tau_{(3|1,2)}^e$ . Research station, parity, and the linear and quadratic effects of DIM were specified in  $\gamma_{(21)}^u, \gamma_{(31)}^u, \gamma_{(32)}^u, \gamma_{(21)}^e, \gamma_{(31)}^e$ , and  $\gamma_{(32)}^e$ . Finally, research station, parity, ration with station, and up to 4<sup>th</sup> order polynomial effects of DIM were specified in  $\beta_1, \beta_2$ , and  $\beta_3$ . Ration within station was specified in all cases as the single random effects factor for each of  $\mathbf{m}_{(r)}^u, \mathbf{m}_{(r)}^e, \mathbf{v}_{(s)}^u$ , and  $\mathbf{v}_{(s)}^e$ . Given that there were a total of 271 rations across the 13 research stations, it was more important to specify these ration effects as random at the deepest levels of the model hierarchy (i.e., for PE and VC) in order to facilitate efficient borrowing of information across rations. Conversely, treating ration effects as fixed in  $\beta_1, \beta_2$ , and  $\beta_3$  was deemed to be feasible because these effects are specified closer to the data ( $\mathbf{y}$ ) in the model hierarchy. To account for changes in energy balance, the average daily change in BW ( $\Delta\text{BW}$ ) was fitted as a covariate in  $\beta_1, \beta_2$ , and  $\beta_3$  as well.

### 3.3.2.2 Model Comparison

Model comparison and selection in hierarchical Bayesian inference is often formally based on the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002) as well as on the relative sizes of the posterior  $z$ -scores; i.e., the ratio of the posterior means to their posterior standard deviations (Gelman et al., 2012). The absolute values of these posterior  $z$ -scores can be used in Wald-like tests to ascertain, for example, whether or not they exceed 1.96 in absolute value for a Type I error rate of 5%, assuming the marginal posterior density of the corresponding parameter is reasonably symmetric. The DIC has previously been used in a stepwise manner to infer

evidence for fixed and random sources of heterogeneity on VC and PE within a 2-trait model by Bello et al. (2012). However adapting their stepwise process would be far more onerous for our 3-trait model given the need for an assessment of which fixed or random effects factors to keep for each of 12 different mixed model subsets (i.e., 3 separate models for each of Equations [3.5], [3.6], [3.7], and [3.8]) beyond the classical mixed model specification in Equation [3.1]. Thus, we only compared 5 models starting at one extreme from a purely homoscedastic MT model with homogeneous PE and VC at all levels ( $M_0$ ) as developed and used previously in Lu et al. (2015) to the proposed fully heteroscedastic MT model with heterogeneous PE and VC, based on all structural mixed effects model specifications as detailed in this paper ( $M_4$ ). The remaining models evaluated were  $M_1$ ) fixed effects (i.e., without random effects) modeling on both genetic and residual PE with homogeneous genetic and residual VC,  $M_2$ ) fixed effects modeling on genetic and residual PE and on genetic and residual VC and  $M_3$ ) full mixed effects modeling on VC with only fixed effects modeling on PE.

### **3.3.2.3 Cross-validation Assessment**

To assess the relative importance of modeling heterogeneous genetic and residual VC and PE, we assessed the cross-validation performance of our proposed model ( $M_4$ ) relative to the more conventional MT mixed model ( $M_0$ ) of Lu et al. (2015). More specifically, a 5-fold cross validation study was conducted by merely randomly partitioning the entire dataset into 5 equally sized subsets such that for each of the folds, 4 subsets were chosen as training data with the remaining subset chosen as validation data. The performance of the 2 models were compared with respect to cross-validation predictive accuracy defined as the correlation between true DMI

and prediction in the validation datasets based on estimates derived from the analyses of the training datasets.

### 3.3.2.4 Marginal Means

All reported estimates were based on the full model ( $M_4$ ), determined to be the best fitting of those evaluated. To facilitate interpretation, however, estimates of all fixed effects were re-expressed to a “marginal mean” or “least-squares means” basis as also popularized by SAS software (Littell et al., 2002, Milliken and Johnson., 2009); further illustration in the context of our MT model is presented by Bello et al. (2010) and Bello et al. (2012). For example, linear combinations of  $\gamma_{(21)}^u, \gamma_{(31)}^u, \gamma_{(32)}^u, \gamma_{(21)}^e, \gamma_{(31)}^e$ , and  $\gamma_{(32)}^e$  were used to estimate marginal genetic and residual PE relationships for the various research stations, averaged across levels of the remaining factor(s) (i.e., parity) and at the midpoint covariate values (i.e., DIM = 125 d). Marginal mean linear combinations of  $\tau_{(1)}^u, \tau_{(21)}^u, \tau_{(31,2)}^u, \tau_{(1)}^e, \tau_{(21)}^e$ , and  $\tau_{(31,2)}^e$  were exponentiated to the VC scale in a manner similar to that described by Bello et al. (2010). Pairwise comparisons between marginal posterior means were conducted between levels of each fixed effects factor (i.e., stations and parity) using 2-tailed Bayesian  $P$ -values in the manner described previously by Bello et al. (2010).

### 3.3.3 Simulation Study

We realize that our proposed Bayesian MT model on 3 traits is massively hierarchical in nature since it entails a total of 15 different mixed model specifications, including 3 within the basic first stage specification for  $\mathbf{y}_i$  in Equation [3.1], one for each of  $\phi_{(21)i}^u, \phi_{(31)i}^u$ , and  $\phi_{(32)i}^u$  in Equation [3.5], one for each of  $\phi_{(21)i}^e, \phi_{(31)i}^e$ , and  $\phi_{(32)i}^e$  in Equation [3.6], one for each of  $\sigma_{u(1)i}^2$ ,

$\sigma_{u(2|1)i}^2$ , and  $\sigma_{u(3|1,2)i}^2$  in Equation [3.7], and one for each of  $\sigma_{e(1)i}^2$ ,  $\sigma_{e(2|1)i}^2$ , and  $\sigma_{e(3|1,2)i}^2$  in Equation [3.8]. Hence, similar to Bello et al. (2010), we used a simulation study to validate the proposed model and Bayesian inference strategy as developed elsewhere.

Twenty replicated datasets were simulated based on only the subset of the data and pedigree deriving from the US research stations to facilitate computational tractability. An overall intercept and station effects were generated as fixed effects for PE ( $\gamma_{(21)}^u, \gamma_{(31)}^u, \gamma_{(32)}^u, \gamma_{(21)}^e, \gamma_{(31)}^e$ , and  $\gamma_{(32)}^e$ ) and for the VCs ( $\tau_{(1)}^u, \tau_{(2|1)}^u, \tau_{(3|1,2)}^u, \tau_{(1)}^e, \tau_{(2|1)}^e$ , and  $\tau_{(3|1,2)}^e$ ), whereas ration within station effects were generated as random effects with variance components  $\sigma_{m_u(21)}^2, \sigma_{m_u(31)}^2, \sigma_{m_u(32)}^2$  for the genetic PE and  $\sigma_{m_e(21)}^2, \sigma_{m_e(31)}^2$ , and  $\sigma_{m_e(32)}^2$  for the residual PE and with hyperparameters  $\eta_{(1)}^u, \eta_{(2|1)}^u$ , and  $\eta_{(3|1,2)}^u$  for the genetic VC and  $\eta_{(1)}^e, \eta_{(2|1)}^e$ , and  $\eta_{(3|1,2)}^e$  for the residual VC. True values for these (hyper)parameters were roughly determined as estimates from the analysis of subset of data deriving from the US. Coverage probabilities of the 95% highest posterior density (HPD) interval for each parameter or hyper-parameter were based on how many times the HPD included the true values over the 20 replicated datasets.

### 3.3.4 MCMC Implementation

For all analyses involving either the dairy or simulated data, trace plots, autocorrelation, and effective sample size (ESS) for samples of all unknown parameters drawn from the posterior density using MCMC were monitored (Brooks and Gelman, 1998) using the R package Coda (Plummer et al., 2006). The number of burnin cycles were 30,000 for the simulation study and 20,000 for the dairy data analysis. After burnin, every tenth cycle was saved from 100,000

MCMC cycles such that posterior inference was based on 10,000 cycles. This sampling strategy was sufficient to ensure that  $ESS > 100$  for all hyperparameters.

### **3.4 Results**

#### **3.4.1 Analysis of Dairy Consortium Data**

As indicated previously, only a small subset of all possible models was considered, the simplest model ( $M_0$ ) being equivalent to a classical homogeneous (co)variance matrix multiple-trait modeling specification as in Lu et al. (2015). DIC values were expressed relative to  $M_0$  (i.e.  $DIC = 0$  for  $M_0$ ) which was determined to be the worst fitting model. DIC values became progressively smaller (i.e., better fitting) for increasingly more complex models starting with  $M_1$ ) fixed effects but no random effects modeling on both genetic and residual PE with no structural modeling at all for VC ( $DIC = -1,673$ );  $M_2$ ) only fixed effects modeling on both genetic and residual PE and VC ( $DIC = -4,305$ );  $M_3$ ) full mixed effects modeling on VC with fixed effects modeling only on PE ( $DIC = -4,633$ ); and finally  $M_4$ ) full mixed effects modeling on both VC and PE ( $DIC = -5,400$ ), the best fitting of all models and upon which all subsequent inferences on the dairy data analysis are based. Clearly then, it was important to structurally fit genetic and residual sources of heterogeneous PE and VE per Equations [3.5], [3.6], [3.7], and [3.8], recognizing that DIC differences exceeding 7 are generally considered to represent significant difference in model fit (Spiegelhalter et al., 2002).

#### **3.4.2 Inferences on Factors Influencing PE**

The posterior means (PMEAN) and posterior standard deviations (PSD), 95% Highest Posterior Density (HPD) intervals, and ESS for all PE related inferences on marginal means of

station, parity, linear, and quadratic effects of DIM at both genetic ( $\gamma_{(21)}^u$ ,  $\gamma_{(31)}^u$ , and  $\gamma_{(32)}^u$ ) and residual levels ( $\gamma_{(21)}^e$ ,  $\gamma_{(31)}^e$ , and  $\gamma_{(32)}^e$ ) as well as variance components for random effects at the genetic ( $\sigma_{m_u(21)}^2$ ,  $\sigma_{m_u(31)}^2$ , and  $\sigma_{m_u(32)}^2$ ) and residual ( $\sigma_{m_e(21)}^2$ ,  $\sigma_{m_e(31)}^2$ , and  $\sigma_{m_e(32)}^2$ ) levels are summarized in Appendix Table B.1 for MBW|MILKE, in Table 3.2 for DMI|MILKE, and in Table 3.3 for DMI|MBW, respectively.

Table 3.2: Posterior inferences characterizing heterogeneity of genetic and residual partial efficiencies of DMI<sup>1</sup> on MILKE<sup>2</sup>

		PMEAN (PSD) <sup>3</sup>	95% HPD <sup>4</sup>	ESS <sup>5</sup>	PMEAN (PSD) <sup>3</sup>	95% HPD <sup>4</sup>	ESS <sup>5</sup>
Factor	Levels	Genetic Partial Efficiencies ( $\phi_{(31)}^u$ in kg/Mcal)			Residual Partial Efficiencies ( $\phi_{(31)}^e$ in kg/Mcal)		
Stations <sup>6</sup>	AB	0.47(0.18)	[0.10,0.82]	939	0.43(0.13)	[0.16,0.66]	670
	BARC	0.46(0.10)	[0.27,0.65]	2,547	0.27(0.11)	[0.06,0.48]	1,622
	ISU	0.42(0.09)	[0.24,0.59]	887	0.36(0.05)	[0.26,0.47]	1,143
	MSU	0.46(0.15)	[0.16,0.74]	1,232	0.40(0.06)	[0.28,0.50]	1,439
	UF	0.36(0.18)	[0.02,0.71]	736	0.35(0.09)	[0.17,0.54]	649
	USDFRC	0.49(0.11)	[0.27,0.70]	491	0.37(0.06)	[0.25,0.49]	471
	UW	0.32(0.12)	[0.08,0.57]	907	0.31(0.06)	[0.18,0.43]	1,100
	NBZ	0.44(0.07)	[0.30,0.60]	1,458	0.49(0.13)	[0.22,0.75]	1,551
	NLN	0.32(0.13)	[0.07,0.59]	1,016	0.29(0.05)	[0.20,0.38]	1,405
	TGEN	0.40(0.13)	[0.15,0.65]	1,059	0.32(0.10)	[0.11,0.51]	878
	ZOM	0.32(0.06)	[0.21,0.43]	766	0.28(0.04)	[0.21,0.35]	1,301
	LAN	0.28(0.10)	[0.09,0.48]	2,445	0.31(0.06)	[0.20,0.42]	4,176
	SAC	0.30(0.14)	[0.02,0.59]	1,966	0.29(0.06)	[0.18,0.40]	3,776
DIM	Linear	0.03(0.08)	[-0.12,0.19]	9,637	0.06(0.10)	[-0.14,0.25]	8,913
	Quadratic	-0.01(0.11)	[-0.21,0.21]	9,514	0.02(0.13)	[-0.20,0.29]	6,713
Parity	Primiparous	0.35(0.05)	[0.25,0.44]	597	0.34(0.03)	[0.27,0.41]	456
	Multiparous	0.43(0.07)	[0.29,0.58]	591	0.35(0.03)	[0.28,0.40]	840
Ration(Station)	Variance Component (kg/Mcal) <sup>2</sup>						
	Genetic ( $\sigma_{m_u(31)}^2$ )	5.2e-03(5.2e-03)	[3.4e-06,1.5e-2]	198			
	Residual ( $\sigma_{m_e(31)}^2$ )				3.6e-03(2.7e-03)	[1.6e-05,8.7e-03]	263

<sup>1</sup> Dry matter intake (in Kg)

<sup>2</sup> Milk energy (in Mcal)

<sup>3</sup> Posterior (marginal) mean ( $\pm$  posterior standard deviation)

<sup>4</sup> 95% Highest posterior density interval

<sup>5</sup> Effective sample size

<sup>6</sup> Station levels characterized in body of paper.

Table 3.3: Posterior inferences characterizing heterogeneity of genetic and residual partial efficiencies of DMI<sup>1</sup> on MBW<sup>2</sup>

		PMEAN (PSD) <sup>3</sup>	95% HPD <sup>5</sup>	ESS <sup>6</sup>	PMEAN (PSD) <sup>3,4</sup>	95% HPD <sup>5</sup>	ESS <sup>6</sup>
Factor	Levels	Genetic Partial Efficiencies( $\phi_{(32)}^u$ in Kg/ Kg <sup>0.75</sup> )			Residual Partial Efficiencies( $\phi_{(32)}^e$ in Kg/ Kg <sup>0.75</sup> )		
<i>Stations</i> <sup>7</sup>	<i>AB</i>	0.21 (0.08)	[0.05,0.37]	399	0.04 <sup>ab</sup> (0.12)	[-0.21,0.28]	345
	<i>BARC</i>	0.11 (0.07)	[-0.03,0.25]	1,358	0.08 <sup>ab</sup> (0.05)	[-0.01,0.17]	1,780
	<i>ISU</i>	0.13 (0.04)	[0.05,0.20]	795	0.04 <sup>ab</sup> (0.04)	[-0.05,0.12]	823
	<i>MSU</i>	0.10 (0.05)	[0.01,0.18]	745	0.08 <sup>ab</sup> (0.05)	[-0.04,0.19]	606
	<i>UF</i>	0.10 (0.10)	[-0.06,0.32]	127	0.18 <sup>ab</sup> (0.12)	[-0.06,0.38]	114
	<i>USDFRC</i>	0.11 (0.04)	[0.04,0.19]	385	0.13 <sup>a</sup> (0.04)	[0.04,0.21]	330
	<i>UW</i>	0.13 (0.04)	[0.04,0.22]	711	0.10 <sup>ab</sup> (0.05)	[0.00,0.20]	675
	<i>NBZ</i>	0.08 (0.03)	[0.02,0.14]	1,283	0.04 <sup>ab</sup> (0.07)	[-0.10,0.18]	1,059
	<i>NLN</i>	0.10 (0.04)	[0.02,0.19]	596	0.06 <sup>ab</sup> (0.03)	[0.00,0.11]	802
	<i>TGEN</i>	0.09 (0.06)	[-0.03,0.21]	740	0.07 <sup>ab</sup> (0.05)	[-0.03,0.17]	677
	<i>ZOM</i>	0.07 (0.03)	[0.02,0.13]	1,031	0.05 <sup>b</sup> (0.02)	[0.00,0.10]	1,143
	<i>LAN</i>	0.10 (0.03)	[0.03,0.16]	1,814	0.05 <sup>ab</sup> (0.04)	[-0.03,0.12]	2,350
	<i>SAC</i>	0.09 (0.05)	[-0.01,0.20]	1,029	0.02 <sup>ab</sup> (0.04)	[-0.07,0.10]	1,126
<i>DIM</i>	<i>Linear</i>	-0.07 (0.15)	[-0.35,0.22]	8,772	-0.04(0.14)	[-0.31,0.25]	8,626
	<i>Quadratic</i>	0.07 (0.16)	[-0.25,0.39]	6,556	0.12(0.16)	[-0.18,0.42]	7,899
<i>Parity</i>	<i>Primiparous</i>	0.11 (0.02)	[0.08,0.15]	490	0.08(0.02)	[0.03,0.12]	370
	<i>Multiparous</i>	0.10 (0.02)	[0.06,0.15]	514	0.07(0.02)	[0.02,0.11]	442
<i>Ration(Station)</i>	Variance Component ( (kg/kg <sup>0.75</sup> ) <sup>2</sup> )						
	<i>Genetic</i> ( $\sigma_{m_{(32)}}^2$ )	6.0e-04(6.0e-4)	[2.9e-06,1.8e-03]	188			
	<i>Residual</i> ( $\sigma_{m_{(32)}}^2$ )				7.0e-04(7.1e-04)	[1.4e-06,2.1e-03]	214

<sup>1</sup> Dry matter intake (in Kg)

<sup>2</sup> Metabolic body weight (in Kg)

<sup>3</sup> Posterior (marginal) mean ( $\pm$  posterior standard deviation)

<sup>4</sup> Estimates not sharing the same letters within factors are statistically different ( $P < 0.05$ ).

<sup>5</sup> 95% Highest posterior density interval

<sup>6</sup> Effective sample size

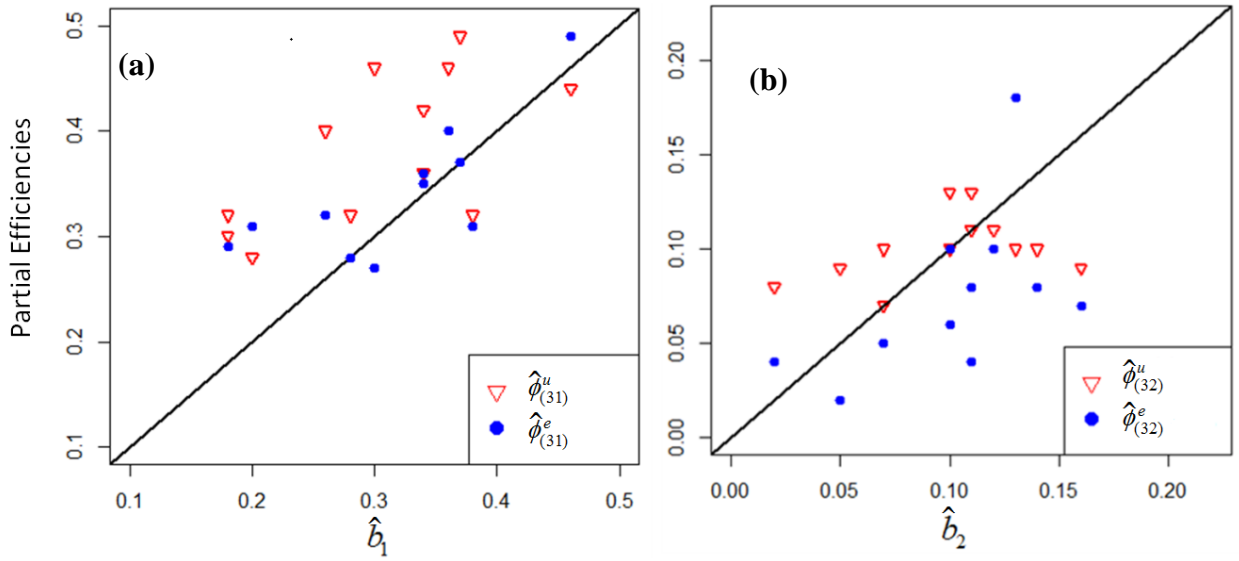
<sup>7</sup> Station levels characterized in body of paper.

Although we do not particularly highlight results regarding MBW|MILKE in this paper, it is intriguing to note from Supplementary Table B.1 that the genetic PE between the 2 traits were not generally different from 0 for any one research station nor for either parity because 0 fell within the 95% HPD in all cases. Conversely, residual relationships were universally positive. Taken together and using Equation [3.7] in (Lu et al., 2015), these results imply a universally negligible genetic correlation but a positive residual correlation between MBW and MILKE. At any rate, there was also no formal evidence of any heterogeneity in these relationships between research stations or parities based on pairwise comparisons between the corresponding marginal means. Also, there was no evidence of heterogeneous genetic variances but strong evidence of heterogeneous residual variances across stations for MILKE (Supplementary Table B.2) whereas there was no evidence of any heterogeneous genetic nor residual variances for MBW|MILKE across stations (Supplementary Table B.3)

Because our focus was on rFE, we were particularly interested in PE relationships involving DMI. As a basis for reference, the overall mean PE at the genetic and residual levels were respectively determined to be  $\bar{\phi}_{(r)}^u = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_{(r)i}^u ' \hat{\gamma}_{(r)}^u + \mathbf{z}_{(r)i}^u ' \hat{\mathbf{m}}_{(r)}^u)$  and  $\bar{\phi}_{(r)}^e = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_{(r)i}^e ' \hat{\gamma}_{(r)}^e + \mathbf{z}_{(r)i}^e ' \hat{\mathbf{m}}_{(r)}^e)$  where  $\hat{\gamma}_{(r)}^u$ ,  $\hat{\mathbf{m}}_{(r)}^u$ ,  $\hat{\gamma}_{(r)}^e$ , and  $\hat{\mathbf{m}}_{(r)}^e$  denote the PMEAN of the corresponding parameters for  $r = 31$  or  $32$ ; i.e., PE of DMI on MILKE and of DIM on MBW, respectively. Specifically, the PMEAN( $\pm$  PSD) of  $\bar{\phi}_{(31)}^u$  was  $0.38(\pm 0.10)$  kg/Mcal whereas the PMEAN ( $\pm$  PSD) of  $\bar{\phi}_{(32)}^u$  was  $0.10(\pm 0.05)$  kg/kg<sup>0.75</sup> implying then, for example, the genetic merit of DMI is estimated to increase by 0.38 kg on average for every Mcal increase in the genetic merit of MILKE holding constant MBW. At the residual level, the PMEAN( $\pm$  PSD) of

$\bar{\phi}_{(31)}^e$  was 0.33 ( $\pm 0.06$ ) kg/Mcal whereas the PMEAN( $\pm$  PSD) of  $\bar{\phi}_{(32)}^e$  was 0.07( $\pm 0.04$ ) kg/kg<sup>0.75</sup> such that, for example, as the residual or temporary environment effect of MBW increases by 1 kg<sup>0.75</sup>, the residual effect of DMI is estimated to increase by 0.07 kg holding constant MILKE. These overall genetic and residual PE estimates are in reasonable agreement with averages of station specific partial regression coefficients relating DMI to MILKE and to MBW within a classical RFI analysis on the same data reported previously by Tempelman et al. (2015) who, in turn, demonstrated these estimates to be consistent with estimates reported by NRC (2001).

A scatterplot of the station-specific marginal means for genetic and residual PE versus the corresponding partial regression coefficients relating DMI to MILKE and to MBW from the classical RFI analyses in Tempelman et al. (2015) are provided in Figure 3.1, noting that estimates from Station AB are not included in Figure 3.1 since that station was not considered in Tempelman et al. (2015). Correlations between station-specific partial regression coefficients relating DMI to MILKE, reported by Tempelman et al. (2015), with corresponding genetic and residual PE estimates were 0.66 and 0.76, respectively. Correlations between station-specific partial regression coefficients relating DMI to MBW, reported by Tempelman et al. (2015), with corresponding genetic and residual PE estimates were 0.39 and 0.56, respectively. Although the agreement between the 2 analyses was generally good, it should be clearly seen that the genetic PE were generally larger than residual PE, particularly for DMI on MILKE (Figure 3.1a ). Furthermore, there appeared to be less heterogeneity between stations in the genetic PE relative to the residual PE for DMI on MBW (Figure 3.1b).



<sup>1</sup> Dry matter intake (in Kg)

<sup>2</sup> Milk energy (in Mcal)

<sup>3</sup> Metabolic body weight (in Kg)

Figure 3.1: Posterior means of station-specific genetic ( $\hat{\phi}_{(31)}^u$  and  $\hat{\phi}_{(32)}^u$ ) and residual ( $\hat{\phi}_{(31)}^e$  and  $\hat{\phi}_{(32)}^e$ ) partial efficiencies versus estimated partial regression coefficients ( $\hat{b}_1$  and  $\hat{b}_2$ ) previously reported by Tempelman et al., 2015) for DMI<sup>1</sup> on MILKE<sup>2</sup> (Panel a) and MBW<sup>3</sup> (Panel a).

A list of these same station-specific marginal mean inferences for the genetic and residual PE are provided in Table 3.2 for DMI on MILKE and in Table 3.3 for DMI on MBW. Inference summaries are also provided for primiparous versus multiparous animals as well as for linear and quadratic coefficients of DIM on these PE. Although the estimated marginal mean genetic PE between DMI and MILKE ranged widely from 0.28 to 0.49 kg/Mcal across stations, there was no formal evidence of any difference ( $P > 0.05$ ) between any pair of stations or between primiparous versus multiparous cows. Variation due to ration within station effects on genetic PE between DMI and MILKE seemed substantial; based on the reported point estimate

$\sqrt{\hat{\sigma}_{m_u(31)}^2} = 0.07$  kg/Mcal such that one might anticipate roughly a range of  $4 \times (0.07) = 0.28$

kg/Mcal difference on genetic PE between the most extreme ration effects, assuming normality and adjusting for station and DIM effects. As with the genetic PE, any differences in the marginal mean residual PE of DMI to MILKE did not appear to be important even though again point estimates varied widely from 0.27 kg/Mcal to 0.49 kg/Mcal. Furthermore, the estimate  $\sqrt{\hat{\sigma}_{m_e(31)}^2} = 0.06$  kg/Mcal further suggested that ration within station could be a substantial source of variability for the residual PE relationship between DMI and MILKE. Linear and quadratic effects of DIM also did not seem to be important for either residual or genetic PE; in other words, stage of lactation did not appear to be important for modeling PE between DMI and MILKE.

The posterior summary of station and parity level effects on the genetic and residual PE of DMI to MBW are provided in Table 3.3. There appeared to be no significant differences between stations for the genetic PE although point estimates ranged from 0.07 kg/kg<sup>0.75</sup> to 0.21 kg/kg<sup>0.75</sup>. Furthermore, heterogeneity in genetic PE of DMI to MBW due to the rations appeared to be relatively substantial ( $\sqrt{\hat{\sigma}_{m_u(32)}^2} = 0.024$  kg/kg<sup>0.75</sup>). However, at the residual level, there was formal evidence of at least one difference for PE of DMI on MBW between ZOM ( $0.05 \pm 0.02$  kg/kg<sup>0.75</sup>) and USDFRC ( $0.13 \pm 0.04$  kg/kg<sup>0.75</sup>) even though point estimates  $\pm$  standard errors ranged more widely from  $0.02 \pm 0.04$  kg/kg<sup>0.75</sup> (SAC) to  $0.18 \pm 0.12$  kg/kg<sup>0.75</sup> (UF). The magnitude of variation in the residual PE of DMI on MBW due to the ration was again rather substantial ( $\sqrt{\hat{\sigma}_{m_e(32)}^2} = 0.026$  kg/kg<sup>0.75</sup>).

### 3.4.3 Inferences on Variance Components and Heritability of rFE

Supplementary Tables C.2 and C.3, and Table 3.4 summarize marginal inferences on fixed and random effects influencing VC for MILKE, MBW|MILKE, and DMI|MILKE,MBW,

respectively. We focus in particular on DMI|MILKE,MBW; i.e., our proposed rFE trait. The overall genetic variance as based on the P<sub>MEAN</sub> ( $\pm$  PSD) of  $\bar{\sigma}_{u3|12}^2 = \frac{1}{n} \sum_{i=1}^n \exp\left(\mathbf{x}_i^{u'} \log\left(\hat{\boldsymbol{\tau}}_{(3|12)}^u\right) + \mathbf{z}_i^{u'} \log\left(\hat{\mathbf{v}}_{(3|12)}^u\right)\right)$  was  $0.65 (\pm 0.19) \text{ kg}^2$ , whereas the overall residual variance was based on the P<sub>MEAN</sub> ( $\pm$  PSD) of  $\bar{\sigma}_{e3|1,2}^2 = \frac{1}{n} \sum_{i=1}^n \exp\left(\mathbf{x}_i^{e'} \log\left(\hat{\boldsymbol{\tau}}_{3|1,2}^e\right) + \mathbf{z}_i^{e'} \log\left(\hat{\mathbf{v}}_{3|1,2}^e\right)\right)$  being  $1.86(\pm 0.45) \text{ kg}^2$ , such that the P<sub>MEAN</sub> ( $\pm$  PSD) of overall heritability was  $0.26 (\pm 0.08)$ . Evidence for heterogeneity in the genetic variance for rFE was considerable (Table 3.4). The NBZ station appeared to have the smallest estimated genetic variance ( $0.47 \pm 0.10 \text{ kg}^2$ ) whereas the largest genetic variance was estimated at UF ( $1.28 \pm 0.41 \text{ kg}^2$ ) among all stations. No significant differences in genetic variance in rFE were found between primiparous and multiparous cows. Furthermore, genetic variances were demonstrated to be highly heterogeneous between rations within stations with the 95% HPD for the CV on genetic variance between rations being between 0.26 and 0.52.

Table 3.4: Posterior inferences characterizing heterogeneity of genetic and residual variation of DMI<sup>1</sup>|MILKE<sup>2</sup>,MBW<sup>3</sup>

Factor	Levels	PMEAN (PSD) <sup>4,5</sup>	95% HPD <sup>6</sup>	ESS <sup>7</sup>	PMEAN (PSD) <sup>4,5</sup>	95% HPD <sup>6</sup>	ESS <sup>7</sup>
		Genetic variance (Mcal <sup>2</sup> )			Residual variance (Mcal <sup>2</sup> )		
Stations <sup>8</sup>	AB	1.18 <sup>ab</sup> (0.46)	[0.44,2.09]	620	2.41 <sup>a</sup> (1.11)	[0.69,4.60]	579
	BARC	0.68 <sup>abcd</sup> (0.23)	[0.32,1.14]	1,846	2.17 <sup>ab</sup> (1.06)	[0.71,4.30]	920
	ISU	0.81 <sup>abcd</sup> (0.21)	[0.43,1.23]	865	2.15 <sup>a</sup> (0.89)	[0.86,3.94]	412
	MSU	0.69 <sup>abcd</sup> (0.18)	[0.37,1.05]	985	1.32 <sup>ab</sup> (0.46)	[0.55,2.21]	904
	UF	1.24 <sup>ab</sup> (0.41)	[0.52,2.04]	693	1.49 <sup>ab</sup> (0.62)	[0.55,2.71]	1,047
	USDFRC	0.98 <sup>b</sup> (0.20)	[0.60,1.38]	630	1.88 <sup>a</sup> (0.57)	[0.86,3.01]	734
	UW	0.77 <sup>abcd</sup> (0.20)	[0.41,1.17]	663	1.38 <sup>ab</sup> (0.45)	[0.61,2.27]	792
	NBZ	0.47 <sup>c</sup> (0.10)	[0.28,0.67]	260	1.56 <sup>ab</sup> (0.66)	[0.53,2.84]	1,431
	NLN	0.57 <sup>acd</sup> (0.12)	[0.34,0.80]	711	0.85 <sup>b</sup> (0.24)	[0.43,1.34]	686
	TGEN	0.61 <sup>abcd</sup> (0.20)	[0.29,1.00]	968	2.28 <sup>ab</sup> (1.15)	[0.67,4.59]	418
	ZOM	0.76 <sup>abd</sup> (0.10)	[0.56,0.96]	312	1.7 <sup>a</sup> (0.29)	[1.19,2.26]	231
	LAN	0.51 <sup>cd</sup> (0.13)	[0.28,0.78]	1,423	3.01 <sup>a</sup> (1.25)	[1.08,5.38]	380
	SAC	0.78 <sup>abcd</sup> (0.23)	[0.39,1.26]	1,448	2.58 <sup>a</sup> (1.20)	[0.79,4.95]	622
Parity	<u>Primiparous</u>	0.63(0.07)	[0.48,0.78]	438	1.01 <sup>x</sup> (0.18)	[0.67,1.37]	290
	Multiparous	0.83(0.13)	[0.59,1.09]	222	2.91 <sup>y</sup> (0.47)	[2.06,3.86]	286
Ration(Station)	CV <sup>9</sup>						
	Genetic	0.37(0.08)	[0.26,0.52]	600			
	Residual				0.89(0.56)	[0.47,1.57]	2,542

<sup>1</sup> Dry matter intake (in Kg)

<sup>2</sup> Milk energy (in Mcal)

<sup>3</sup> Metabolic body weight (in Kg)

<sup>4</sup> Posterior (marginal) mean ( $\pm$  posterior standard deviation)

<sup>5</sup> Estimates not sharing the same letters within factors are statistically different ( $P < 0.05$ ).

<sup>6</sup> 95% Highest posterior density interval

<sup>7</sup> Effective sample size

<sup>8</sup> Station levels characterized in body of paper.

<sup>9</sup> Coefficient of variation of ration within station-specific genetic and residual variance components

The magnitude of heterogeneity on the residual variance of rFE across research stations was also quite considerable (Table 3.4). The marginal P<sub>MEAN</sub> ( $\pm$  PSD) for stations ranged from a low of 0.85 ( $\pm$  0.24) kg<sup>2</sup> at NLN to 3.01 ( $\pm$  1.25) kg<sup>2</sup> at LAN. Furthermore, primiparous cows were inferred to have a substantially lower residual variance (1.01  $\pm$  0.18 kg<sup>2</sup>) compared to multiparous cows (2.91  $\pm$  0.47 kg<sup>2</sup>). Also, there appeared to be considerable evidence for heterogeneity of residual variances across rations as the 95% HPD the CV on genetic variance between rations was defined as between 0.47 to 1.57.

Table 3.5 summarizes inferences on heritabilities of rFE across different stations and parity classes. Evidence in favor of heterogeneity of heritability was substantial with the smallest heritability estimated at LAN (0.16  $\pm$  0.06) at the largest estimates at UF (0.46  $\pm$  0.13) and NLN (0.41  $\pm$  0.09). Furthermore, the estimated heritability (0.39  $\pm$  0.06) for primiparous cows was almost twice that (0.22  $\pm$  0.04) for multiparous cows with the difference being significant ( $P < 0.05$ ).

Table 3.5: Inferences on station and parity specific heritabilities for DMI<sup>1</sup>|MILKE<sup>2</sup>,MBW<sup>3</sup>

		PMEAN (PSD) <sup>4,5</sup>	95% HPD <sup>6</sup>	ESS <sup>7</sup>
Factor	Levels			
<i>Station</i> <sup>8</sup>	<i>AB</i>	0.35 <sup>ab</sup> (0.14)	[0.10,0.62]	668
	<i>BARC</i>	0.26 <sup>ab</sup> (0.10)	[0.08,0.47]	1,003
	<i>ISU</i>	0.29 <sup>ab</sup> (0.10)	[0.12,0.50]	539
	<i>MSU</i>	0.35 <sup>ab</sup> (0.10)	[0.16,0.56]	652
	<i>UF</i>	0.46 <sup>a</sup> (0.13)	[0.22,0.71]	922
	<i>USDFRC</i>	0.36 <sup>ab</sup> (0.10)	[0.18,0.55]	460
	<i>UW</i>	0.37 <sup>ab</sup> (0.10)	[0.18,0.57]	701
	<i>NBZ</i>	0.25 <sup>ab</sup> (0.09)	[0.09,0.42]	670
	<i>NLN</i>	0.41 <sup>a</sup> (0.09)	[0.22,0.57]	540
	<i>TGEN</i>	0.24 <sup>ab</sup> (0.10)	[0.07,0.44]	554
	<i>ZOM</i>	0.31 <sup>a</sup> (0.05)	[0.21,0.41]	254
	<i>LAN</i>	0.16 <sup>b</sup> (0.06)	[0.05,0.29]	667
	<i>SAC</i>	0.25 <sup>ab</sup> (0.10)	[0.08,0.45]	804
	<i>Primiparous</i>	0.39 <sup>x</sup> (0.06)	[0.27,0.50]	342
<i>Parity</i>	<i>Multiparous</i>	0.22 <sup>y</sup> (0.04)	[0.14,0.31]	208

<sup>1</sup> Dry matter intake (in Kg)<sup>2</sup> Milk energy (in Mcal)<sup>3</sup> Metabolic body weight (in Kg)<sup>4</sup> Posterior mean ( $\pm$  posterior standard deviation)<sup>5</sup> Estimates not sharing the same letters are statistically different ( $P < 0.05$ ).<sup>6</sup> 95% Highest posterior density interval<sup>7</sup> Effective sample size<sup>8</sup> Station levels characterized in body of paper.

### 3.4.4 Cross-validation

The accuracies for the 5-fold cross validation on the dairy consortium data are provided in Table 3.6. There was strong evidence for differences in model fit between the homogeneous ( $M_0$ ) and fully heterogeneous ( $M_4$ ) models. The accuracy for predicting DMI was consistently higher (~4%) using the heterogeneous model compared to that of the homogeneous model, which

further implies that substantial heterogeneity of some nature (i.e., on genetic or residual PE, or on genetic or residual VC) exists in this consortium dataset.

Table 3.6: Cross validation prediction accuracies for each of 5 different validation datasets (fold) for homogeneous and heterogeneous models.

Fold	Combined	
	Homo (M <sub>0</sub> )	Heter (M <sub>4</sub> )
1	0.75	0.78
2	0.78	0.82
3	0.80	0.84
4	0.77	0.80
5	0.82	0.85
Mean	0.78	0.82

### 3.4.5 Simulation Study Assessment

Coverage frequencies, ranging from 16/20 to 20/20 for the 95% HPD of key hyperparameters across the 20 replicates in the simulation study, are listed in Table 3.7 and were as nearly expected (i.e., 95%). Summing across the last column of Table 3.7, the overall coverage frequency across all parameters was 458/470 or very close to the expected 95%. The posterior means of  $\gamma''_{(r)}$  were more variable and their 95% HPD thus was wider than those of  $\gamma^e_{(r)}$ , as there is typically greater uncertainty for parameters characterizing distribution of random effects as opposed to those for residuals, as reported in an earlier study (Bello et al., 2010).

Table 3.7: Summary on 95% highest posterior density interval (HPD) and coverage probabilities across 20 replicated datasets in simulation study

		True value	Range of 95% HPD lower limits	Range of 95% HPD upper Limits	Coverage probability
Genetic PE	$\gamma_{(21)}^u$	0.35	[-0.236,-0.182]	[0.298,0.369]	17/20
	$\gamma_{(31)}^u$	0.40	[-0.019,0.028]	[0.419,0.455]	20/20
	$\gamma_{(32)}^u$	0.20	[-0.034,-0.018]	[0.188,0.286]	19/20
	$\sigma_{m,(21)}^2$	0.10	[0.068,0.072]	[0.143,0.148]	20/20
	$\sigma_{m,(31)}^2$	0.10	[0.067,0.068]	[0.130,0.133]	20/20
	$\sigma_{m,(32)}^2$	0.10	[0.065,0.101]	[0.124,0.129]	19/20
Genetic VC	$\tau_{(1)}^u$	4	[1.288,1.637]	[3.932,4.302]	17/20
	$\tau_{(21)}^u$	8	[6.413,6.753]	[12.537,13.632]	20/20
	$\tau_{(31,2)}^u$	1.5	[0.859,1.948]	[2.056,2.496]	18/20
	$\eta_{(1)}^u$	8	[2.002,8.046]	[21.108,27.842]	19/20
	$\eta_{(21)}^u$	8	[3.268,4.693]	[12.525,15.652]	20/20
	$\eta_{(31,2)}^u$	8	[2.000,2.962]	[15.473,24.727]	20/20
Residual PE	$\gamma_{21}^e$	0.30	[0.043,0.084]	[0.430,0.472]	20/20
	$\gamma_{31}^e$	0.40	[0.176,0.195]	[0.399,0.494]	19/20
	$\gamma_{32}^e$	0.30	[0.174,0.192]	[0.402,0.425]	20/20
	$\sigma_{m,(21)}^2$	0.10	[0.068,0.071]	[0.140,0.144]	20/20
	$\sigma_{m,(31)}^2$	0.10	[0.084,0.109]	[0.197,0.202]	18/20
	$\sigma_{m,(32)}^2$	0.10	[0.071,0.073]	[0.133,0.137]	20/20
Residual VC	$\tau_{(1)}^e$	8	[6.080,6.326]	[10.153,10.484]	20/20
	$\tau_{(21)}^e$	32	[27.336,27.755]	[33.788,34.472]	20/20
	$\tau_{(31,2)}^e$	3	[2.281,3.616]	[5.376,7.849]	18/20
	$\eta_{(1)}^e$	6	[2.000,3.709]	[20.687,49.602]	20/20
	$\eta_{(21)}^e$	8	[5.769,16.163]	[20.891,53.724]	18/20
	$\eta_{(31,2)}^e$	8	[2.306,19.901]	[37.749,51.161]	16/20

### 3.5 Discussion

We have previously proposed a MT model analysis involving MILKE, MBW, and DMI using the Cholesky decomposition to provide a statistically more elegant approach for characterizing the genetic merit and heritability of rFE relative to classical RFI approaches (Lu et al., 2015). Note that the proposed MT model is more closely related than the classical specification of RFI to the genotypic regression model proposed by Kennedy et al. (1993); so the developments we

have presented here somewhat represent a heterogeneous yet more elegant extension of their genotypic regression model. As we had also indicated in Lu et al. (2015), it certainly would have been ideal to include  $\Delta BW$  as a fourth trait given that, like MILKE and MBW, it is considered to be an important energy sink for lactating dairy cattle (Berry and Crowley, 2013). But, for reasons already outlined in Lu et al. (2015), we modeled it as a covariate to account for its effect on rFE in our MT analyses. We further augment this work here by modeling heterogeneity in the genetic and residual (co)variances, albeit indirectly, between MILKE, MBW, and DMI, by specifying structural mixed effects models on each key parameter defined by the square root free Cholesky decomposition on the (co)variance matrices, similar to Bello et al. (2010). Based on formal model choice criteria (DIC), we determined that specifying such heterogeneity was important for the analysis of an international dairy feed efficiency consortium dataset.

Due to computing constraints, we did not consider every possible submodel given that there are a total of 15 different structural model specifications, 3 per each of Equations [3.1], [3.5], [3.6], [3.7], and [3.8]. We might further conclude, nevertheless, that based on our analyses, specifying this heterogeneity was ultimately most important for VC and somewhat less so for PE, particularly for PE at the genetic level. The latter is not particularly too surprising since it is typically more challenging to detect heterogeneity in genetic (co)variances relative to residual (co)variances. It was somewhat surprising to us that we found no evidence of differences between parity classes for genetic or residual PE for DMI on MBW or for DMI on MILKE given that the metabolic relationships between these traits have been reported to differ between primiparous and multiparous cows (Wathes et al., 2007).

There was mild evidence of some heterogeneity in residual PE for DMI on MBW between stations with the only significant difference involving two research stations. Any differences in ration-specific or station specific PE could be partly attributed to differences, for example, in management strategies, ration compositions, measurement error, or ambient temperature; e.g., whether or not the animal is a thermo-neutral environment (NRC, 1981). Furthermore, we demonstrated that genetic PE were generally larger than residual PE implying that the strength of the relationships between DMI with MILKE and MBW is stronger at the genetic level than at the residual level. In a classical RFI analysis, the residual and genetic PE are constrained to be identical as previously noted by Lu et al. (2015); nevertheless estimates of these 2 sets of PE tracked very well with the corresponding partial regression coefficients from the station-specific RFI-based analyses conducted by Tempelman et al. (2015) as demonstrated earlier in Figure 3.1. Note that any heterogeneity in the PE, particularly at the genetic level, has implications for accurately estimating breeding values on rFE, because these PE determine the key relationship between the genetic merit of DMI and that of rFE using Equation [3.3c] for example.

We concluded that the genetic PE of MBW|MILKE did not differ from 0 for any station, implying then no evidence of genetic association between MILKE and MBW. Our results are then consistent with near zero genetic correlation estimates between these 2 traits as reported elsewhere (Berry et al., 2003, Spurlock et al., 2012).

As previously indicated, there was much stronger evidence of a difference in genetic and residual variances and, consequently, in heritabilities for rFE between research stations, adding further evidence that rFE is indeed a complex trait as also indicated by Tempelman et al. (2015)

who previously determined country specific differences for heritabilities of RFI based on analyses of weekly data. The average heritability estimates reported in Table 3.6 appeared to be somewhat larger here than in Lu et al. (2015) and in Tempelman et al. (2015). Any differences between the current study with those 2 previous studies may be partly due to the fact that 42 d data were used here instead of weekly data, such that the effect of  $\Delta BW$  was better characterized by relatively less measurement error in the current study. However, it could also reflect the fact that the modeled heterogeneity in genetic and residual PE in Equations [3.3c] and [3.4c] leads to more accurate characterizations of genetic and residual effects of rFE (i.e.,  $u_{i(3|1,2)}$  and  $e_{i(3|1,2)}$ , respectively), thereby resulting in a more reliable estimates of the heritabilities of rFE. Our analyses furthermore indicated that the heritability of rFE in primiparous cows was nearly double that of multiparous cows due to the large difference between the parity classes for residual variance. Again, these results may partly reflect the differences in energy dynamics as it relates to growth demands in primiparous cows relative to multiparous cows. Since selection indices involving rFE inherently depend upon the correct specification of genetic/residual variances and covariances between the key traits of MILKE, MBW, and DMI (Lu et al., 2015), it appears that differential selection index weightings may need to reflect this heterogeneity across stations and parities as well. It is also important to note that between-station heterogeneity may partly reflect differences in time periods for data collection (see Table 1 in Tempelman et al., 2015), noting that higher heritability estimates tended to be associated with more recent data.

Although our genetic analysis is based on the use of pedigree information and not on genomic marker information to specific animal relationships, it is conceptually easy to extend the

proposed multivariate model for genomic evaluations simply by replacing the  $\mathbf{A}$  with the genomic or realized relationship matrix  $\mathbf{G}$  or even a hybrid matrix  $\mathbf{H}$  based on data from both genotyped and non-genotyped animals as with single step GBLUP (Aguilar et al., 2010). We focused on a pedigree-based analysis in this paper for one primary reason;  $\mathbf{A}$  is a relatively sparse matrix with 83.6% of its entries being 0, thereby allowing us to effectively utilize sparse matrix software (i.e., “SparseM” package in R) to help improve computational efficiency for sampling elements of  $\mathbf{u}$  whereas  $\mathbf{G}$  is typically a dense matrix. We also determined (not reported) inconsequential differences in EBV between using  $\mathbf{A}$  versus  $\mathbf{G}$  based on computationally tractable MT analyses assuming homogeneous PE and VC specifications (Lu et al., 2015). Nevertheless, we do believe it will be important to further optimize our software code to facilitate the incorporation of genomic information in future analyses.

One potential extension of our proposed model is to directly model genetic effects as random effects on genetic and residual PE and VC, much like what we did with ration effects. In other words, one could specify Equations [3.5], [3.6], [3.7], and [3.8] as a function of polygenic or even SNP-specific effects (i.e., in genomic) models. In fact, with respect to residual variances as per Equation [3.8], related work has already been conducted by Sorensen and Waagepetersen (2003) for polygenic effects and by Yang et al. (2011) for genomic marker effects. With respect to PE, somewhat indirectly related work has been conducted in beef cattle by Savietto et al. (2014) and in poultry (Aggrey and Rekaya, 2013, Rekaya and Aggrey, 2015) whereby animal specific random regressions are specified on partial regression relationships between DMI and MBW and average daily gain as based on a classical RFI model. At any rate, it seems likely that

many more records would be needed than would be required here to infer upon genetic heterogeneity of this nature at these higher levels of the Bayesian model hierarchy. Our simulation study, nevertheless, indicates that it is possible to infer upon the nature of the heterogeneity specified in our proposed model.

### **3.6 Conclusions**

We have conducted a hierarchical Bayesian analysis to infer upon potential heterogeneity on genetic and residual components of DMI conversion to MILKE and DMI to MBW as well as in genetic and residual VC of rFE across research stations. We detected no evidence of heterogeneity in genetic PE with slight evidence of heterogeneity in the residual PE of DMI to MBW across research stations. Yet evidence for genetic and residual heteroscedasticity was substantial across stations and rations within stations levels for rFE. The estimated heritability of rFE ranged from 0.16 to 0.46 across stations, thereby implying that rFE is a more complex trait than what is currently considered in most quantitative genetic analyses. Heterogeneous relationships across environments should be taken into consideration in the genetic characterization and management of rFE in dairy cattle.

## **Chapter4: Genome Wide Association Analyses based on Alternative Strategies for Modeling Feed Efficiency**

### **4.1 Abstract**

Genome wide association (**GWA**) of residual feed efficiency (**rFE**) using genome wide association (**GWA**) analysis could help target important genomic regions influencing rFE. Data provided by an international dairy rFE research consortium consisted of phenotypic records on dry matter intakes (**DMI**), milk energy (**MILKE**), and metabolic body weight (**MBW**) on 6937 cows from 16 stations in four counties. Of these 6937 cows, 4916 had genotypes on 57,347 single nucleotide polymorphism (**SNP**) markers. We adapted our previously proposed multiple trait (**MT**) approach for modeling rFE to conduct a GWA study and compared it to a GWA analysis based on the more classical residual feed intake (**RFI**) model. Both models were based on a single-step genomic BLUP procedure that allows the use of phenotypes from both genotyped and non-genotyped cows. Estimated effects for single SNP markers were small and not statistically important for rFE. However, upon further refining this analysis to develop joint tests on SNP markers within non-overlapping 1-Mb windows, significant associations were detected between rFE with a region on each of BTA 12 and 26 under both MT and RFI models. There was no overlap between detected genomic regions for rFE and genomic regions influencing its component traits (i.e. DMI, MILKE, and MBW). This confirms that rFE and its component traits are different such that genomic regions that control DMI are not necessarily synonymous with genomic regions influencing rFE.

## 4.2 Introduction

Residual feed efficiency (**rFE**), defined as the efficiency of converting feed nutrients into milk and body tissue, has gained considerable attention in recent years, partly due to increasing constraints on arable land as well as mounting concerns regarding environmental pollution (Richardson and Herd, 2004). Most notably, rFE has been included in the national selection index for beef and dairy cattle breeding in Australia in 2014 and 2015, respectively (Gonzalez-Recio et al., 2014, Byrne et al., 2016) with some other countries considering the same (Hayes et al., 2013). Whole genome prediction (WGP) based on the use of single nucleotide polymorphism (SNP) markers (Meuwissen et al., 2001) is a particularly promising strategy to improve rFE in the long term (Hayes et al., 2013, Pryce et al., 2015, VandeHaar et al., 2016). Nevertheless, it is equally important to identify potential causal variants or quantitative trait loci (QTL) affecting rFE in order to fine-tune physiological strategies for further improving rFE (Herd and Arthur, 2009).

Lu et al. (2015) proposed a model that elegantly derived rFE based on a simple reparameterization of a standard multiple trait (MT) analysis of its key component traits, namely, dry matter intake (DMI), milk energy (MILKE) and metabolic body weight (MBW). The MT model recognizes that the partial regressions relating DMI to MILKE and MBW may be different at genetic and non-genetic levels whereas residual feed intake (RFI), a classical measure of rFE, is based on specifying one universal set of partial relationships at the phenotypic level. Although Lu et al. (2015) used pedigree information to determine the numerator relationship matrix (**A**) between animals, it is certainly possible to substitute **A** with genomic relationships (**G**) constructed from SNP markers or even a hybrid relationship matrix (**H**)

combining genotyped and non-genotyped animals having phenotypes as in the single step (SSBLUP) approach (Aguilar et al., 2010). One important advantage of MT over RFI analyses is that it allows the use of data on animals with missing records on any of its component traits; subsequently, MT analyses mitigates the impact of selection bias when records on at least one trait are rarely missing such as, for example, MILKE (Pollak et al., 1984). Furthermore, the MT model is potentially useful for identifying potential pleiotropic SNP markers associated with multiple traits (Banerjee et al., 2008). Thus, MT extensions to GWA based on the model of Lu et al. (2015) could help further elucidate how candidate regions for rFE differ, if at all, from those of its component traits.

Increasingly, more GWA studies have been based on mixed model association methods that simultaneously fit all markers to infer upon any one particular marker of interest (Yang et al., 2014). A particularly popular strategy is based on pursuing generalized least squares (GLS) inference upon the SNP marker of interest (i.e., treating it as a fixed effect) while treating all other SNP markers as random. This strategy is a feature of the EMMAX software (Kang et al., 2010) with an efficient implementation provided in Gualdron Duarte et al. (2014). In GWA analyses that simultaneously fit all SNP markers, any given marker may exhibit only a weak association with a closely linked QTL, in part because of the multicollinearity issues incurred with linkage disequilibrium (LD) with neighboring markers and the large Bonferroni multiple comparisons penalty incurred with testing many different SNP in high LD with each other (Johnson et al., 2010). Hence there has been some momentum to using GWA inference based on joint tests on SNP markers within various partitions or windows containing groups of SNP

markers in order to increase sensitivity and specificity of GWA (Pahl and Schafer, 2010, Moser et al., 2015, Goddard et al., 2016).

In this study, we compare GWA analysis on rFE based on using the classical RFI model with the MT model proposed by Lu et al. (2015) on an international rFE consortium dataset involving phenotypes on 6937 cows of which 4916 cows are genotyped. Our objectives were to 1) infer upon the genetic architecture for rFE and its components traits; 2) compare GWA inferences for rFE between base on the MT model and the RFI model; 3) demonstrate the utility of window-based inference for GWA.

### 4.3 Materials and Methods

#### 4.3.1 The MT Model

We propose extending our previously proposed MT model (Lu et al., 2015) to conduct GWA analyses on rFE as well as on its component traits, DMI, MILKE, and MBW. We reintroduce the MT model in Equation [4.1] for the three component traits arranged in order as  $i=1$ ) MILKE,  $i=2$ ) MBW, and  $i=3$ ) DMI

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{u}_1 + \mathbf{e}_1 \\ \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{u}_2 + \mathbf{e}_2 \\ \mathbf{X}_3\boldsymbol{\beta}_3 + \mathbf{u}_3 + \mathbf{e}_3 \end{bmatrix}. \quad [4.1]$$

Now  $\mathbf{y}_i = [y_{1i} \ y_{2i} \ \dots \ y_{qi}]'$  is the vector of responses for all  $n$  cows on Trait  $i = 1, 2$ , and 3.

Here MILKE and MBW are energy sink traits whereas DMI is the energy source trait that all jointly determine rFE. Although  $\Delta BW$  is also considered to be an important third energy sink trait, we specified it as a covariate for the reasons outlined previously by Lu et al. (2015). For each trait,  $\boldsymbol{\beta}_i$  is a vector of fixed effects that included a linear regression on  $\Delta BW$ , station, first

through fourth order polynomial regressions on days in milk (DIM), parity class (i.e. primiparous versus multiparous), ration, the interaction between first through fourth polynomial terms on DIM and parity class for trait  $i$ . All covariates and dummy variables for these effects are defined by the known incidence matrix  $\mathbf{X}_i$ . Furthermore,  $\mathbf{u}_i = [u_{i1} \ u_{i2} \ \dots \ u_{iq}]$  and  $\mathbf{e}_i = [e_{i1} \ e_{i2} \ \dots \ e_{iq}]$  represent vectors of animal genomic effects and residual effects, respectively, for trait  $i$ . The vector of residuals for all three traits were assumed to be random draws from a multivariate normal distribution, i.e.,  $[\mathbf{e}'_1 \ \mathbf{e}'_2 \ \mathbf{e}'_3]' \sim MVN(\mathbf{0}, \mathbf{I} \otimes \Sigma_e)$ , for  $\Sigma_e$  being the 3 x 3 residual (co)variance matrix among the traits. Similarly, the vector of animal genomic effects for all three traits were assumed to be multivariate normally distributed  $[\mathbf{u}'_1 \ \mathbf{u}'_2 \ \mathbf{u}'_3]' \sim MVN(\mathbf{0}, \Sigma_u \otimes \mathbf{H})$  for  $\mathbf{H}$  being the genetic relationship matrix constructed between genotyped and non-genotyped cows and  $\Sigma_u$  being the 3x3 genetic (co)variance matrix among the three traits within any animal. Specifically, we define the genetic variance for trait  $i$  as  $\sigma_{u_i}^2$ , being diagonal element  $i,i$  of  $\Sigma_u$  and the genetic covariance between traits  $i$  and  $i'$  as  $\sigma_{u_i, u_{i'}}$  or element  $i,i'$  of  $\Sigma_u$ . We similarly define  $\sigma_{e_i}^2$  and  $\sigma_{e_i, e_{i'}}$  as components of  $\Sigma_e$ . Here  $\mathbf{H}$  was constructed based on hybridizing the numerator relationship matrix  $\mathbf{A}$  between all animals and the genomic relationship matrix  $\mathbf{G} = \mathbf{Z}\mathbf{Z}'$  between the genotyped animals for  $\mathbf{Z}$  being the centered and standardized genotype matrix (Aguilar et al., 2010).

Based on the square root free Cholesky decomposition (**CD**) on  $\Sigma_u$ , one can express the relationship between the genetic effects for the three traits in a recursive manner (Lu et al., 2015):

$$u_{1i} = u_{1i} \tag{4.2a}$$

$$u_{2i} = \phi_{21}^{(u)} u_{1i} + u_{2|i} \quad [4.2b]$$

$$u_{3i} = \phi_{31}^{(u)} u_{1i} + \phi_{32}^{(u)} u_{2i} + u_{3|1,2i} \quad [4.2c]$$

Here  $u_{2|i}$  denotes the genetic merit for subject  $i$  on MBW conditional on MILKE.  $u_{3|1,2i}$  denotes the genetic merit for subject  $i$  on DMI conditional on MILKE and MBW (i.e., DMI|MILKE,MBW), which Lu et al. (2015) have previously proposed as an alternative expression of genetic merit for rFE. Furthermore,  $\phi_{21}^{(u)}$ ,  $\phi_{31}^{(u)}$ , and  $\phi_{32}^{(u)}$  are partial efficiencies or relationships at the genetic level and derived from the CD on  $\Sigma_u$ , that are used to specify these recursive relationships in Equations [4.2a]-[4.2c]. Finally,  $\mathbf{u}_1 = \{u_{1i}\}_{i=1}^n \sim N(\mathbf{0}, \mathbf{H}\sigma_{u_1}^2)$ ,  $\mathbf{u}_{2|1} = \{u_{2|i}\}_{i=1}^n \sim N(\mathbf{0}, \mathbf{H}\sigma_{u_{2|1}}^2)$ , and  $\mathbf{u}_{3|1,2} = \{u_{3|1,2i}\}_{i=1}^n \sim N(\mathbf{0}, \mathbf{H}\sigma_{u_{3|1,2}}^2)$  with  $\mathbf{u}_1$ ,  $\mathbf{u}_{2|1}$ , and  $\mathbf{u}_{3|1,2}$  defined to be distributionally independent from each other. Note that  $\sigma_{u_{3|1,2}}^2$  is the genetic variance for the rFE measure proposed by Lu et al. (2015). Lu et al. (2015) also demonstrated how relationships similar to those in Equations [4.2a], [4.2b], and [4.2c] can be specified at the residual level such that  $\phi_{21}^{(e)}$ ,  $\phi_{31}^{(e)}$ , and  $\phi_{32}^{(e)}$  are analogously defined as residual partial efficiencies based on the CD on  $\Sigma_e$ .

A classical RFI mode (Lu et al., 2015) can be written as a regression of DMI on MILKE and MBW with random genetic ( $\mathbf{u}_{RFI}$ ) and residual ( $\mathbf{e}_{RFI}$ ) effects

$$\mathbf{y}_3 = b_1 \mathbf{y}_1 + b_2 \mathbf{y}_2 + \mathbf{X}_{RFI} \boldsymbol{\beta}_{RFI} + \mathbf{u}_{RFI} + \mathbf{e}_{RFI} \quad [4.3]$$

Here  $\boldsymbol{\beta}_{RFI}$  denotes the vector of fixed effects for RFI with corresponding known incidence matrix  $\mathbf{X}_{RFI}$ . The fixed effects fitted in  $\boldsymbol{\beta}_{RFI}$  were the same as those fitted in  $\boldsymbol{\beta}_i$  from the MT model. Additionally,  $b_1$  and  $b_2$  are regression coefficients of DMI on MILKE and DMI on MBW, respectively, and define partial efficiencies at a composite phenotypic level. It is also assumed

that  $\mathbf{u}_{RFI} \sim N(\mathbf{0}, \mathbf{H}\sigma_{u_{RFI}}^2)$  and  $\mathbf{e}_{RFI} \sim N(\mathbf{0}, \mathbf{I}\sigma_{e_{RFI}}^2)$ . Lu et al. (2015) demonstrated that when partial efficiencies are identical at the genetic and residual levels (i.e.  $\phi_{31}^{(e)} = \phi_{31}^{(u)}$  and  $\phi_{32}^{(e)} = \phi_{32}^{(u)}$ ), then inferences under the two competing models (RFI versus MT) should be essentially identical.

### 4.3.2 Single SNP Associations

All (co)variance components ( $\sigma$ ) can be estimated using REML for both the MT ( $\sigma = [\Sigma_u \quad \Sigma_e]$ ) and the classical RFI models ( $\sigma = [\sigma_{u_{RFI}}^2 \quad \sigma_{e_{RFI}}^2]$ ). Conditional on these REML estimates ( $\hat{\sigma}$ ), Henderson's mixed model equations can be used to directly solve for BLUP of the animal genetic effects for each of the three component traits (i.e.  $\mathbf{u}_1, \mathbf{u}_2$ , and  $\mathbf{u}_3$ ) in the MT model and for RFI (i.e.,  $\mathbf{u}_{RFI}$ ) in the classical RFI model. Furthermore, as demonstrated by Lu et al. (2015),  $\mathbf{u}_1, \mathbf{u}_2$ , and  $\mathbf{u}_3$  and the CD on  $\Sigma_u$  can be used to derive the BLUP of  $\mathbf{u}_{3|1,2}$  (i.e.,  $\mathbf{u}_{3|1,2}$ ) based on Equation [4.2c].

We partition  $\mathbf{u}_i = [\mathbf{u}_{(non)i}' \quad \mathbf{u}_{(gen)i}']$  into effects for non-genotyped ( $\mathbf{u}_{(non)i}$ ) and genotyped ( $\mathbf{u}_{(gen)i}$ ) animals and do so similarly for  $\hat{\mathbf{u}}_i$ , the BLUP of  $\mathbf{u}_i$ . As demonstrated previously by Wang et al. (2012), the BLUP  $\hat{\mathbf{g}}_i$  of SNP effects for trait  $i$  can be backsolved from  $\hat{\mathbf{u}}_i$  using Equation [4.4]

$$\hat{\mathbf{g}}_i = \mathbf{Z}'\mathbf{G}^{-1}\hat{\mathbf{u}}_{(gen)i} . \quad [4.4]$$

Here  $\hat{\mathbf{g}}_i = [\hat{g}_{i1} \quad \hat{g}_{i2} \quad \dots \quad \hat{g}_{im}]'$  denote the BLUP of the SNP effects for the corresponding traits MILKE ( $i=1$ ), MBW ( $i=2$ ), DMI ( $i=3$ ), DMI|MILKE,MBW ( $i=3|1,2$ ) and RFI ( $i=RFI$ ). Borrowing from Gualdron Duarte et al. (2014), it can be readily demonstrated that  $\mathbf{V}_{\hat{\mathbf{g}}_i} = \text{var}(\hat{\mathbf{g}}_i) = \mathbf{Z}'\mathbf{G}^{-1}(\mathbf{G}\sigma_{u_i}^2 - \mathbf{C}_{(gen)}^{\mathbf{u}_i, \mathbf{u}_i})\mathbf{G}^{-1}\mathbf{Z}$  where  $\mathbf{C}_{(gen)}^{\mathbf{u}_i, \mathbf{u}_i}$  is the Bayesian posterior (co)variance matrix for  $\mathbf{u}_{(gen)i}$  conditional on  $\sigma = \hat{\sigma}$ , or equivalently, the prediction error (co)variance matrix

for  $\hat{\mathbf{u}}_{(gen)i}$  conditional on  $\boldsymbol{\sigma} = \hat{\boldsymbol{\sigma}}$  based on classical linear mixed models theory (Searle et al., 1992), i.e.,  $\mathbf{C}_{(gen)}^{\mathbf{u}_i} = \text{var}(\mathbf{u}_{(gen)i} | \mathbf{y}, \boldsymbol{\sigma} = \hat{\boldsymbol{\sigma}})$ , or  $\mathbf{C}_{(gen)}^{\mathbf{u}_i} = \text{var}(\hat{\mathbf{u}}_{(gen)i} - \mathbf{u}_{(gen)i} | \boldsymbol{\sigma} = \hat{\boldsymbol{\sigma}})$ , respectively. Further details are provided in Appendix B. A test statistic used for GWA on SNP  $k$  for trait  $i$  can be written as in Equation [4.5]

$$z_{ik} = \frac{\hat{g}_{ik}}{\text{sep}(\hat{g}_{ik})} \quad [4.5]$$

Here, the denominator in Equation [4.5] is the square root of the  $k,k$  element in  $\mathbf{V}_{\hat{\mathbf{g}}_i}$ .

Gualdron Duarte et al. (2014) and Bernal Rubio et al. (2016) demonstrated that this test statistic is equivalent to conducting GLS inference on  $g_{ik}$  while momentarily treating all other elements of  $\mathbf{g}_i$  as random, just as with EMMAX (Kang et al., 2010). That is, under the null hypothesis  $H_0: g_{ik} = 0$ , it can be demonstrated that  $z_{ik} \sim N(0,1)$ .

#### 4.3.3 Window-Based Associations

Recognizing the potential issues with single SNP associations, adjusting for collinear SNP in high LD as described earlier, we also conducted window-based inferences. That is, we partitioned the genome into non-overlapping 1-Mb regions, similar to what has been done for GWA analyses in other cattle populations (Guo et al., 2012, Hawken et al., 2012, Alpay et al., 2014). This resulted in a total of  $R=2,673$  windows. Suppose that  $\mathbf{Z}$  can be partitioned accordingly into these  $R$  windows as  $\mathbf{Z} = [\mathbf{Z}_1 \ \mathbf{Z}_2 \ \cdots \ \mathbf{Z}_R]$  with  $\mathbf{Z}_r$  having  $n_r$  columns, implying then that window  $r$  contains  $n_r$  SNP markers. Suppose that  $\mathbf{g}_i$  is similarly partitioned; i.e.  $\mathbf{g}_i = [\mathbf{g}_{i1}' \ \mathbf{g}_{i2}' \ \cdots \ \mathbf{g}_{iR}']'$  such that  $\mathbf{g}_{ir}$  is of dimension  $n_r \times 1$ . A joint test for all  $n_r$  markers can be readily determined by computing the test statistic in Equation [4.6]

$$\chi_r^2 = \hat{\mathbf{g}}_{ir}' \mathbf{V}_{\hat{\mathbf{g}}_i(r)}^{-1} \hat{\mathbf{g}}_{ir} \quad [4.6]$$

where  $\mathbf{V}_{\hat{\mathbf{g}}_{ir}} = \text{var}(\hat{\mathbf{g}}_{ir}) = \mathbf{Z}_r' \mathbf{G}^{-1} \left( \mathbf{G} \sigma_{u_i}^2 - \mathbf{C}_{(gen)}^{u_i u_i} \right) \mathbf{G}^{-1} \mathbf{Z}_r$ . Similar to Equation [4.5], the test statistic in Equation [4.6] allows for GLS inference on  $\mathbf{g}_{ir}$ , treating all other SNP effects as random. That is, extending Bernal Rubio et al. (2016), under the null hypothesis  $H_0: \mathbf{g}_{ir} = 0$ , it can be readily demonstrated that  $\chi_r^2$  is distributed as a chi-square random variable with  $n_r$  degrees of freedom.

### 4.3.4 Application to Dairy Consortium Data

#### 4.3.4.1 Data Description

The dataset was collected from 6937 Holstein cows involving 16 research stations or major studies from Scotland, the Netherlands, Canada, and the United States. All records on DMI, MILKE, MBW, and RFI were based on the first 28d after 50 DIM, based on the data edits described in Tempelman et al. (2015) and Lu et al. (2015). For cows having multiple lactations in the dataset, only records from the earliest lactation were used. Nine of the research stations were from the United States including Iowa State University (**ISU**) at Ames, IA, Michigan State University (**MSU**) at East Lansing, MI, the University of Florida (**UF**) at Gainesville, FL, the University of Wisconsin-Madison (**UW**) at Madison, WI, the United States Dairy Forages Research Center (**USDFRC**) at Madison, WI, the USDA Animal Genomics and Improvement Laboratory (**AGIL**) at Beltsville, MD, Virginia Polytechnic Institute and State University (**VT**) at Blacksburg, VA, the Dairy Research Facility at Miner Institute (**MIN**) at Chazy, NY, and the Purina Animal Nutrition Center (**PANC**) at Gray Summit, MO. Four stations/studies were from the Netherlands including an experimental herd (**TGEN**) in Lelystad previously described by Veerkamp et al. (2000) and Banos et al. (2012); the Nij Bosma Zathe (**NBZ**) herd located near

Leeuwarden and also previously described by Banos et al. (2012) ; and a compilation of studies (**NLN**) based on data collected from various nutritional experiments as described in Tempelman et al. (2015). Two stations derived from the UK: the Langhill (**LAN**) farm near Edinburgh from 1992 to 2001 upon which the herd was relocated to the Scottish Agricultural College (**SAC**) Dairy Research Centre based at Crichton Royal Farm near Dumfries with data collection from 2003 to 2011. Finally, the last station was the University of Alberta (**UA**) in Edmonton, Canada. A complete breakdown of number of animals with phenotypes and genotypes and timeline of data collection for each station is provided in Table 4.1.

Table 4.1: Brief characterization of dataset by stations<sup>1</sup>

Area	Station	Number of Phenotypes	Number of Genotypes	Dates of Data Collection
Canada	UA	237	220	09/2007-06/2012
	UF	507	377	02/2009-07/2015
	ISU	952	930	05/2008-09/2015
	MSU	271	264	02/2011-01/2015
	MIN	57	51	02/2013-03/2013
United States	UW	816	780	12/2007-07/2015
	PANC	144	18	08/2009-12/2011
	USDFRC	407	347	10/2009-04/2015
	USDA AGIL	518	488	11/2007-09/2015
	VT	90	54	02/2011-10/2013
NL	TGEN	586	461	11/1991-04/1998
	NBZ	100	91	07/2003-11/2004
	ZOM	661	0	11/1991-01/2001
	NLN	555	385	03/2003-03/2014
UK	LAN	590	157	12/1990-03/2001
	SAC	446	293	09/2003-11/2010
Overall		6937	4916	

<sup>1</sup> UA = University of Alberta; UF = University of Florida; ISU = Iowa State University; MSU = Michigan State University; MIN= Dairy Research Facility at Miner Institute; UW = University of Wisconsin–Madison; PANC= Purina Animal Nutrition Center;USDFRC = USDA Dairy Forages Research Center; USDA AGIL = USDA Beltsville Agricultural Research Center; VT= Virginia Polytechnic Institute and State University; TGEN = ‘t Gen experimental herd, Lelystad, the Netherlands; NBZ= Nij Bosma Zathe, Leeuwarden, the Netherlands; ZOM = data based on the work of Zom et al. (2012); NLN = compilation of studies previously characterized by Tempelman et al. (2015); LAN = Langhill farm,Edinburgh, UK; SAC = Scottish Agricultural College.

Genomic characterization was based on Illumina® BovineSNP50 Genotyping BeadChip, which features 54,609 informative SNP probes uniformly spanning the entire bovine genome. Genotypes were obtained on 4,917 of the 6,937 cows and pre-processed for parent–progeny conflicts checking by AGIL. These genotypes were imputed to a final set of 61,013 SNPs, including an additional 15,818 SNPs selected for the magnitude of their effect on economically important traits evaluated in the US (Wiggans et al., 2016). After editing and filtering, 57,347 SNPs were saved for the GWA analyses. The markers on the X chromosome were further partitioned into pseudoautosomal and X-specific components.

#### **4.3.4.2 Real Data Analyses**

GWA analyses based on the MT and RFI models described above were applied to this rFE dairy consortium dataset. Variance components under both models were estimated using REML. Furthermore, BLUP and the corresponding SNP and window-based association tests, as previously described, for  $\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3$ , and  $\mathbf{g}_{3|1,2}$  under the MT model and for  $\mathbf{g}_{RFI}$  under the RFI model were determined as well. GWA analysis was provided for each component trait in the MT analysis. Bonferroni corrections based on a genome wide error rate of 5% were applied with the correction being based on the number (57,347) of SNP markers for single SNP associations and the number (2,673) of windows for window-based associations. Potential candidate genes located within windows deemed to have an association with the various traits were determined using the Bioconductor package biomaRt based on the UMD3.1 bovine assembly from Ensembl.

## 4.4 Results and Discussion

### 4.4.1 Genetic Parameter Estimates

The estimated genetic and residual variances as well as the estimated heritabilities for two definitions of rFE and its component traits are provided in Table 4.2. Estimated residual and genetic covariances among the three component traits of the MT model are reported in Table 4.3. Residual variance estimates were nearly identical for RFI and DMI|MILKE,MBW such that the respective heritabilities were  $0.18 \pm 0.04$  and  $0.17 \pm 0.03$ , consistent with those based on using only pedigree and not genotype information as previously reported by our group (Lu et al., 2015). The estimated genetic partial efficiencies relating DMI to MILKE and DMI to MBW were  $\phi_{31}^{(u)} = 0.42 \pm 0.06$  and  $\phi_{32}^{(u)} = 0.11 \pm 0.02$ , respectively, whereas the estimated residual partial efficiencies relating DMI to MILKE and DMI to MBW were  $\phi_{31}^{(e)} = 0.36 \pm 0.07$  and  $\phi_{32}^{(e)} = 0.10 \pm 0.02$ , again similar to what we previously reported (Lu et al., 2015).

Table 4.2: Estimated variance components and heritability for residual feed efficiency and its component traits.

Trait	Genetic variance	Residual variance	heritability
MILKE <sup>1</sup>	2.84±0.31	10.12±0.28	0.22±0.05
MBW <sup>2</sup>	27.44±2.09	32.75±1.23	0.46±0.07
DMI <sup>3</sup>	1.19±0.12	3.87±0.11	0.23±0.05
RFI <sup>4</sup>	0.51±0.07	2.39±0.06	0.18±0.04
DMI MILKE,MBW <sup>5</sup>	0.49±0.06	2.39±0.05	0.17±0.03

<sup>1</sup> Milk energy

<sup>2</sup> Metabolic body weight

<sup>3</sup> Dry matter intake

<sup>4</sup> Residual feed intake

<sup>5</sup> Dry matter intake conditional on milk energy and metabolic body weight

Table 4.3: Residual and genetic covariances between the three component traits

Covariances $\pm$ SE		Trait $(i,i')^1$		
		(1,2)	(1,3)	(2,3)
Level	Residual	4.42 $\pm$ 0.36	3.52 $\pm$ 0.13	4.30 $\pm$ 0.24
	Genetic	0.17 $\pm$ 0.48	1.14 $\pm$ 0.14	2.67 $\pm$ 0.33

<sup>1</sup>  $i=1$ ) MILKE (Milk energy),  $i=2$ ) MBW (metabolic body weight), and  $i=3$ ) DMI (Dry matter intake)

#### 4.4.2 Single SNP Associations

We compared GLS inferences of SNP effects for rFE between the two models (i.e.  $\mathbf{g}_{RFI}$  vs.  $\mathbf{g}_{3|1,2}$ ) based on the methods we described above. We noted that  $\hat{\mathbf{g}}_{3|1,2}$  computed were slightly more shrunk to zero as opposed to that for  $\hat{\mathbf{g}}_{RFI}$  (Figure 4.1). This larger spread for  $\hat{\mathbf{g}}_{RFI}$  was likely caused by the slightly larger REML estimates of  $\sigma^2_{u_{RFI}}$  relative to  $\sigma^2_{u_{3|1,2}}$ . Nevertheless, they were still relatively close to and highly correlated with each other, as were diagonals of  $\mathbf{V}_{\hat{\mathbf{g}}_{RFI}}$  and  $\mathbf{V}_{\hat{\mathbf{g}}_{3|1,2}}$ , such that negative  $\log_{10}$  P-values (  $-\log_{10}(p)$  ) for GWA inference on RFI and for DMI|MILKE,MBW were in relatively good agreement (Figure 4.2).

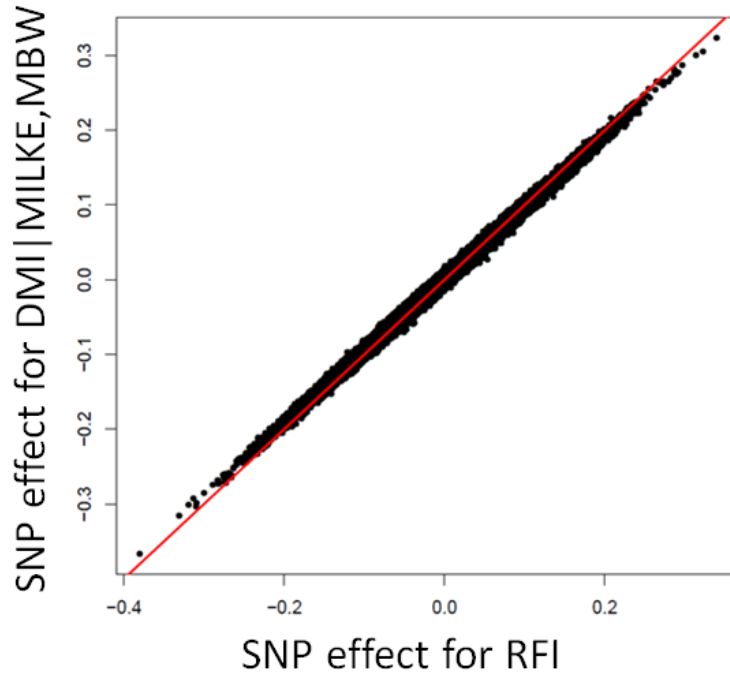


Figure 4.1: Comparison of BLUP of SNP effects for RFI versus DMI conditional on MILKE and MBW. Line of slope 1 and intercept 0 superimposed.

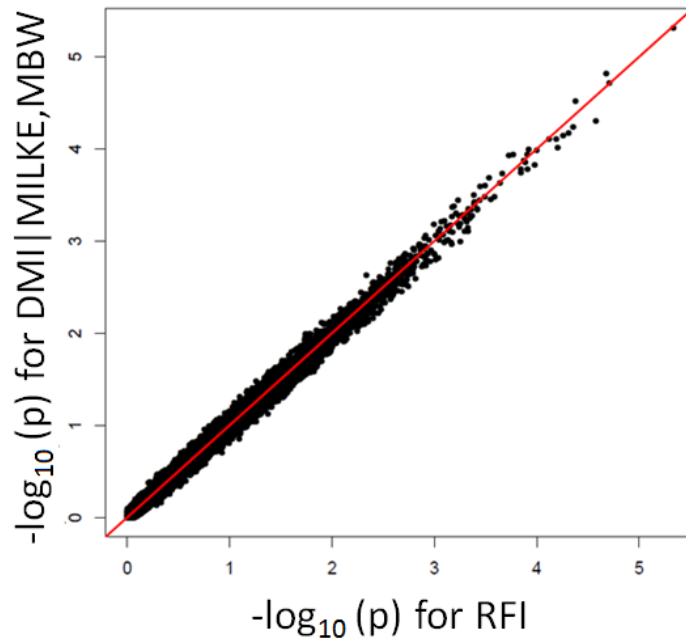


Figure 4.2: Comparison of single SNP inferences ( $-\log_{10}(p)$ ) for DMI conditional on MILKE and MBW versus those for RFI. Line of slope 1 and intercept 0 superimposed.

Manhattan plots for single SNP GWA inference for both rFE traits (RFI vs DMI|MILKE,MBW) are provided in Figure 4.3. There were no significant SNP associations detected for either DMI|MILKE,MBW (Figure 4.3a) nor for RFI (Figure 4.3b). For the component traits of rFE, the only significant associations were discovered for MBW (Figure 4.4b), namely two SNP markers (BovineHD0500030398 and Hapmap60480-ss46526970) located at 105,804,923bp and 105,870,613bp on BTA5 and 2 markers (BovineHD1800016753 and ARS-BFGL-NGS-109285) located at 57,516,245bp and 57,589,121bp on BTA18. However, no significant associations were determined for MILKE (Figure 4.4a) nor for DMI (Figure 4.4c).

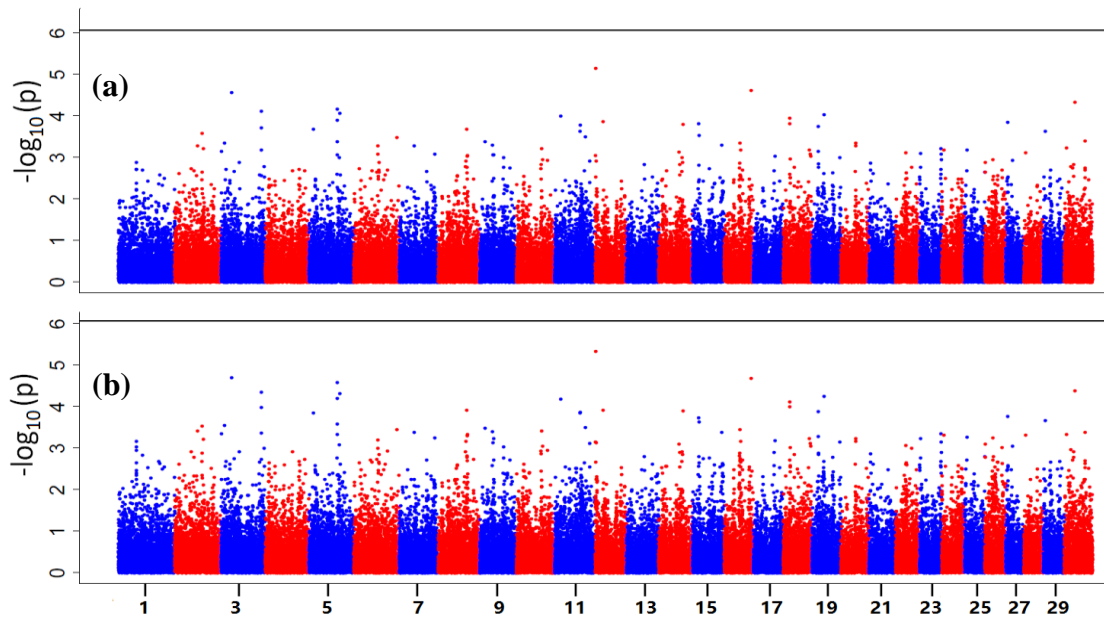


Figure 4.3: The Manhattan plots of single SNP for DMI conditional on MILKE and MBW(a) and for RFI (b). Bonferroni correction threshold superimposed.

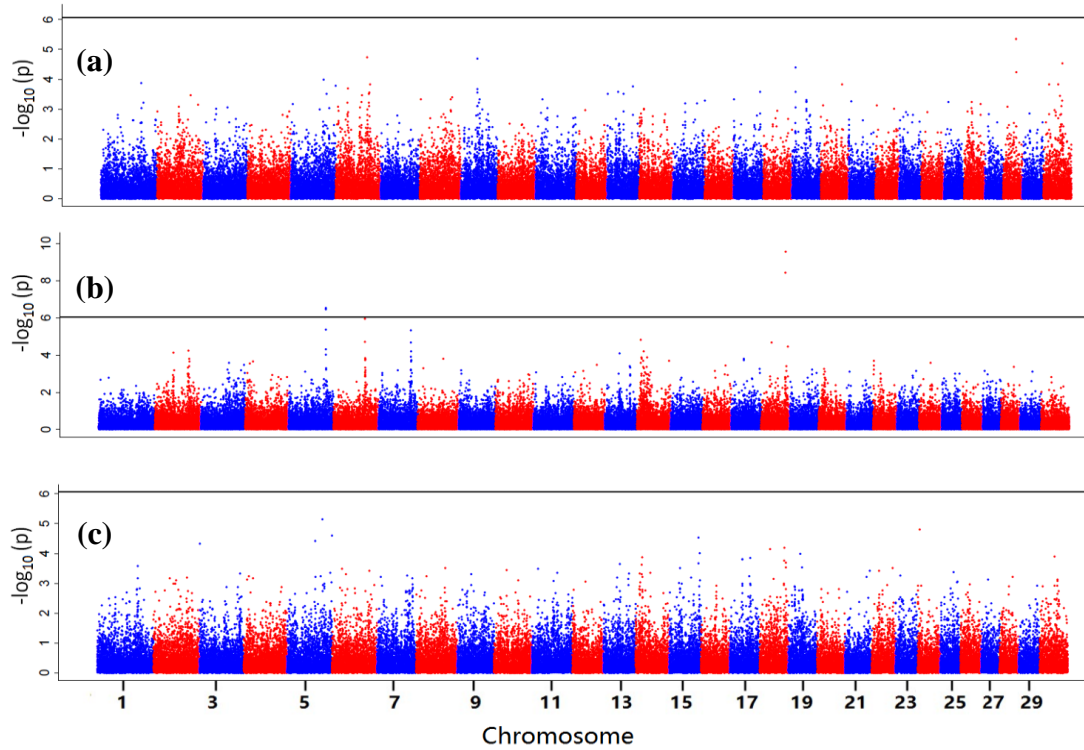


Figure 4.4: The Manhattan plots of single SNP on MILKE (a), MBW(b) , and DMI(c).  
Bonferroni correction threshold superimposed.

#### 4.4.3 Window-Based Associations

The window-based Manhattan plots for RFI, and DMI|MILKE,MBW are provided in Figure 4.5. As further detailed in Table 4.4, the same 2 regions were found to be important for both RFI and DMI|MILKE,MBW, being on chromosome BTA12 centered at 2.50 Mb (**BTA12\_2.50**) and on BTA 26 centered at 46.50Mb (**BTA26\_46.50**). BTA26\_46.50 was previously confirmed as an important region for rFE in Nelore cattle (Olivieri et al., 2014). Furthermore, Sherman et al. (2009) and Nkrumah et al. (2007) reported associations on BTA26 at 47 Mb for RFI in beef cattle, which were in good agreement with the association that we found within BTA26\_46.50.

Table 4.4: Physical location<sup>1</sup> (in Mb) of significant regions detected for each trait

Chr	DMI <sup>3</sup>	MBW <sup>4</sup>	MILKE <sup>5</sup>	RFI <sup>6</sup>	DMI MILKE,MBW <sup>7</sup>
5	106.50				
12				2.50	2.50
14		7.50,9.50, 25.50			
18	58.50,59.50				
20			14.50		
26				46.50	46.50
X-specific <sup>2</sup>			102.50		

<sup>1</sup>Location of a window defined as center of that window

<sup>2</sup> X-specific section on X chromosome

<sup>3</sup> Dry matter intake

<sup>4</sup> Metabolic body weight

<sup>5</sup> Milk energy

<sup>6</sup> Residual feed intake

<sup>7</sup> Dry matter intake conditional on milk energy and metabolic body weight

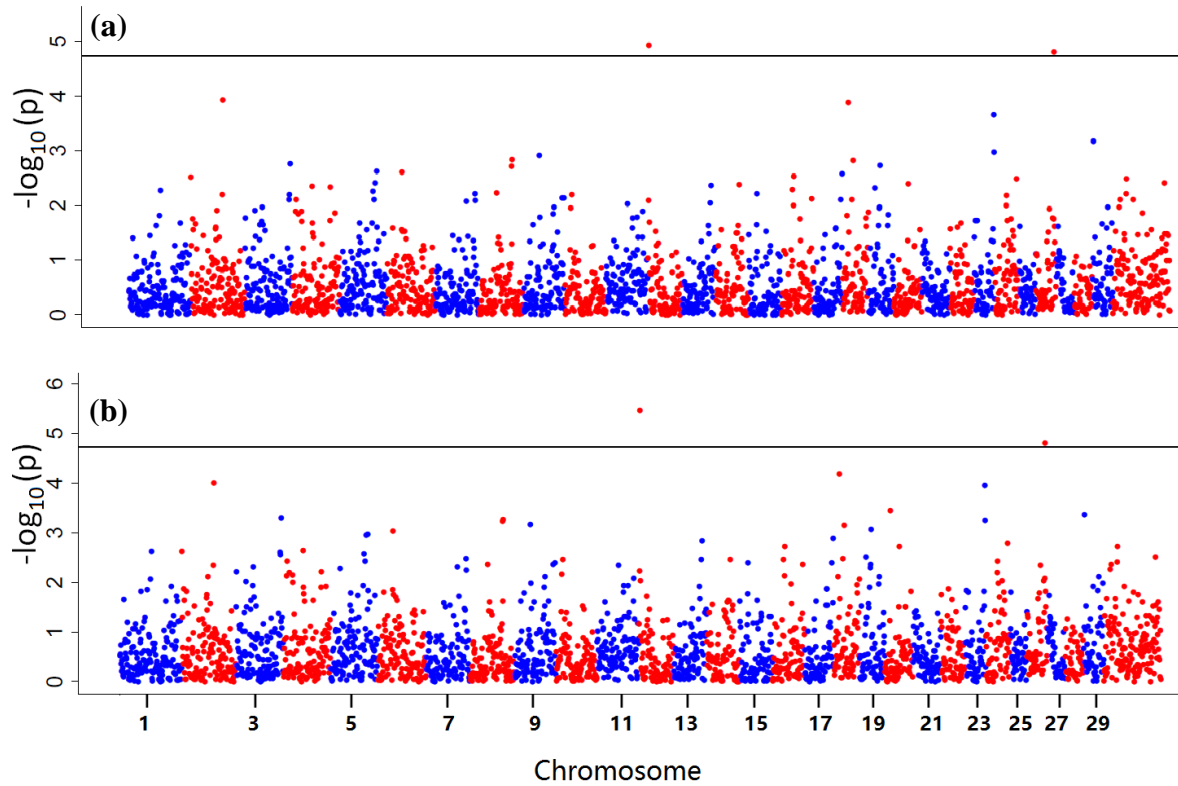


Figure 4.5: The Manhattan plots of 1MB windows for DMI conditional on MILKE and MBW(a) and for RFI (b). Bonferroni correction threshold superimposed.

We found 4 genes in the bovine genome assembly that physically overlapped with those two windows associated with rFE; these genes and a summary of their biological processes are listed in Table 4.5. These genes include disintegrin and metallopeptidase domain 12 (*ADAM12*, 45,848,827-46,238,138 bp). *ADAM12* regulates broad biological processes, including modulation of cell morphological changes, satellite cell activation, and ectodomain shedding during signaling of muscle and fat development (Cao et al., 2003, Kawaguchi et al., 2003). The *ADAM12* gene is also identified as a regulator for transforming growth factor- $\beta$ 1, which is involved in the differentiation of human adipose tissue-derived mesenchymal stem cells into smooth muscle cells (Kim et al., 2012). Furthermore, *ADAM12* regulates myogenesis and

adipogenesis in beef cattle (Coles et al., 2014) and was reported as a candidate gene for marbling score in Korean cattle (Ryu and Lee, 2016). *ADAM12* was also suggested as a candidate gene for rFE in Nelore cattle (Olivieri et al., 2014, Olivieri et al., 2016). It is possible that *ADMA12* gene affects rFE by directly regulating muscle development and fatty acid utilization. Thus, *ADMA12* can be a promising candidate gene for rFE

Table 4.5: Symbol, name, and involved biological functions of genes overlapped with 2 significant regions identified for residual feed efficiency in Holstein cows

Symbol	Name	Functions involved
<i>RPL3</i>	Ribosomal protein L3	poly(A) RNA binding and structural constituent of ribosome
<i>DIAPH3</i>	Protein diaphanous homolog 3	binding and Rho GTPase binding
<i>ADAM12</i>	A Disintegrin And Metalloproteinase Domain 12	SH3 domain binding and metalloproteinase activity
<i>DOCK1</i>	Dedicator Of Cytokinesis 1	phagocytosis and cell migration

The window-based Manhattan plots for the component traits of rFE are provided in Figure 4.6. There were 2, 3, and 3 regions found to be associated with MILKE (Figure 4.6a), MBW (Figure 4.6b), and DMI (Figure 4.6c), respectively. The three detected regions for DMI were centered at 106.50 Mb on BTA5, at 58.50 Mb and at 59.50Mb on BTA18 (see Table 4.4). The detected window on BTA5 is between 2 SNPs (at 100,826,813 bp and 118,501,191 bp) detected for DMI by Veerkamp et al. (2012). The three identified regions associated with MBW were all on BTA14, centered at 7.50 Mb, 9.50 Mb, and 25.50Mb. Lee et al. (2013) detected 6 SNPs at 24-25 MB on BTA14 for carcass weight in Korean Hanwoo. Lu et al. (2013) reported 3 SNP associations at HAPMAP29891-BTC-007427 ( 8,485,892 bp) , UA-IFASA-6233

(10,090,807), and BTB-01532239(24,437,778 bp) in beef cattle that fall within the 3 windows detected for MBW in our study. Saatchi et al. (2014a) also reported a region at 24 Mb on BTA14 that was associated with MBW in Angus and Simmental  $\times$  Angus populations. Saatchi et al. (2014b) later detected a genomic region at 24 Mb on BTA14 associated with BW across multiple beef cattle species. The 2 important regions for MILKE were centered at 14.50 Mb on BTA20 and at 102.50Mb on the X chromosome. The former window was previously reported to be associated with milk fat percentage and milk yield in dairy cattle (Plante et al., 2001, Waters et al., 2011).

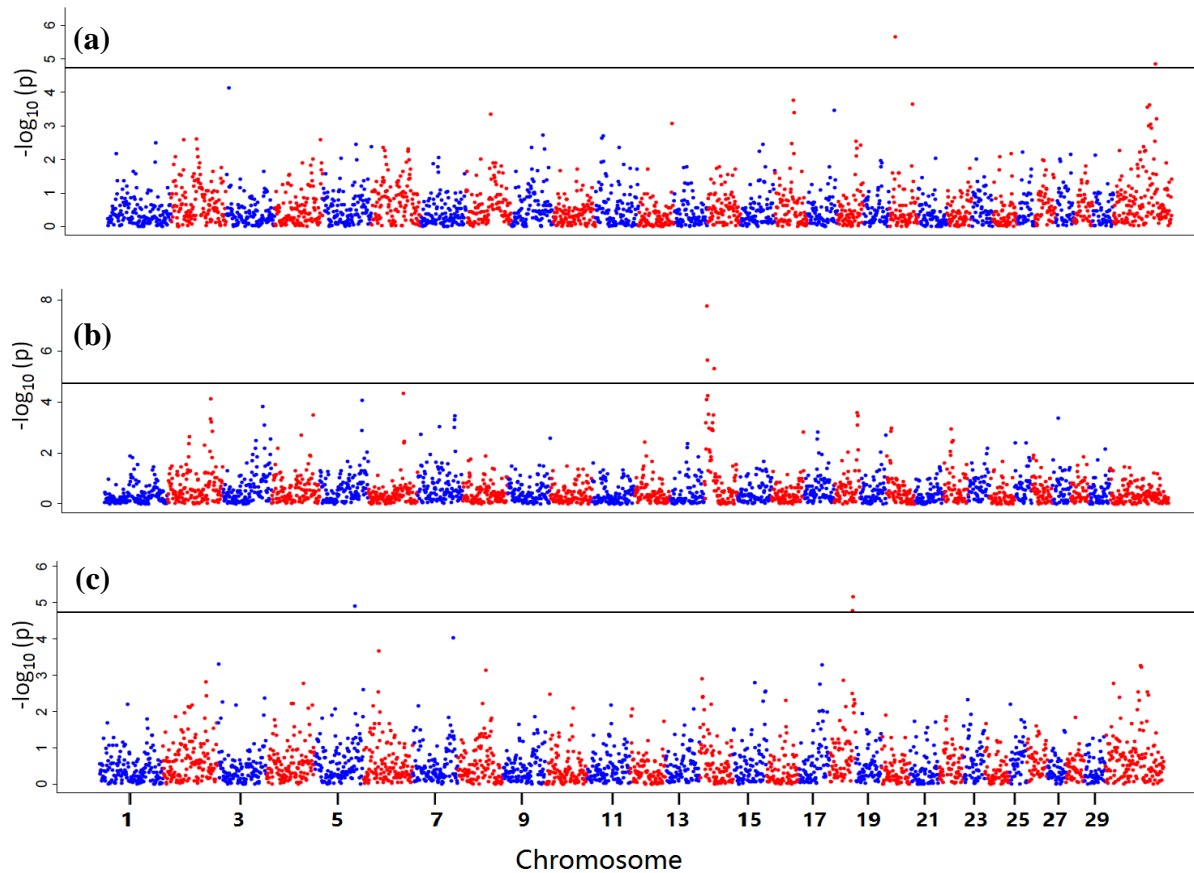


Figure 4.6: The Manhattan plot of 1-Mb windows on MILKE (a), MBW(b) , and DMI(c). Bonferroni correction threshold superimposed.

#### 4.4.4 General Discussion

GWA analyses have been increasingly used to infer upon the genetic architecture of traits given the increasing availability of high density SNP marker panels. Our primary purpose, however, was not to just conduct GWA analyses on rFE and its component traits, DMI, MILKE, and MBW, but rather to demonstrate how the genomic architecture may differ substantially between traits that define rFE. We particularly focus on rFE(closely related to RFI) and DMI because of the current controversy on where to focus selection for improving efficiency of feed utilization (Berry and Pryce, 2014).

Among genomic regions determined to be associated with the various traits using the window-based test, there was no regions in common between any of the four traits. This indicates that the genetic architecture of rFE, whether specified as RFI or as DMI|MILKE,MBW, is distinctly different from its component traits, including DMI. This result may at first seem counterintuitive since Kennedy et al. (1993) and Lu et al. (2015) demonstrated that a selection index involving DMI, MILKE, and MBW should be equivalent to a selection index on rFE, MILK, and MBW. However, because of multivariate normal assumptions on elements of  $\mathbf{u}_1$ ,  $\mathbf{u}_2$ , and  $\mathbf{u}_3$ , it can be readily demonstrated that Equations [4.2a], [4.2b], and [4.2c] can be similarly expressed at the level of SNP effects. For example, by analogy with Equation [4.2c], it can be determined that  $g_{3j} = \phi_{31}^{(u)} g_{1j} + \phi_{32}^{(u)} g_{2j} + g_{3|1,2j}$  such that  $g_{3|1,2j}$  is independent of  $g_{3j}$ , and also of  $g_{1j}$  and  $g_{2j}$ . Hence, GWA inferences on DMI|MILKE,MBW should be independent of inferences on DMI, MILKE or MBW simply by construction. We previously discovered quantitative genetic inferences on RFI were nearly identical to those for DMI|MILKE,MBW (Lu et al., 2015). Hence,

a orthogonal relationship between RFI with DMI, MILKE, or MBW similar to that between DMI|MILKE,MBW and these three component traits should nearly hold as well. In beef cattle studies, for example, no overlap was also determined between detected genomic regions for rFE and BW (Rolf et al., 2012, Saatchi et al., 2014a).

The relationship  $g_{3j} = \phi_{31}^{(u)} g_{1j} + \phi_{32}^{(u)} g_{2j} + g_{3|1,2j}$  suggests that GWA inferences on DMI should be somewhat driven by MILKE and MBW. As an illustration consider the 2 most significant associations, both on BTA 18, for DMI as we discovered earlier. The corresponding  $P$ -values for MILKE and MBW in those same regions were 0.05/0.008 and 0.0003/0.003. Hence, selection against DMI in those regions would have detrimental consequences for MILKE and MBW. By contrast, the 2 detected associations for DMI|MILKE,MBW, as well as for RFI, had corresponding  $P$ -values for MILKE/MBW of 0.38/0.26 and 0.32/0.10, thereby partly confirming the genetic independence between rFE and its component traits. From the viewpoint of identifying candidate genes for maximizing efficiency, it seems that one should target attention on inferences involving  $g_{3|1,2}$  rather than  $g_1$ .

We previously noted that a general movement to more window-based inference as opposed to single SNP inference in GWA studies. Indeed, we did not detect any single SNP associations for either measure of rFE, nor for DMI or MILKE. It is possible that these traits are very complex in the sense that they are controlled by many genes each with small effects. Although we did identify two genomic 1MB regions associated with rFE based on the window test, the proportion of the genetic variance jointly accounted by those two regions was rather small (~1.1%) using the approach suggested by Fernando and Garrick (2013). Similar findings and conclusions on

non-detection of genomic regions for RFI in pigs and in beef cattle have been discovered (Lu et al., 2013, Do et al., 2014, Santana et al., 2014, Serão et al., 2016) . Herd and Arthur (2009) further noted that the physiological basis for feed energy utilization is a complicated process involving digestion, metabolism, activity, and thermoregulation, thereby suggesting a plethora of genes. Similarly, MILKE is a composite of milk, fat, protein, and lactose yield (Tempelman et al., 2015).

GWA inferences on RFI versus DMI|MILKE,MBW were virtually identical, whether based on single SNP and window tests suggesting it may be rather inconsequential whether one chooses the MT versus RFI model for GWA. These similarities may be somewhat anticipated given the similarities in partial efficiency estimates between the genetic and residual levels for both DMI on MILKE ( $\hat{\phi}_{31}^{(u)} \approx \hat{\phi}_{31}^{(e)}$ ) and DMI on MBW ( $\hat{\phi}_{32}^{(u)} \approx \hat{\phi}_{32}^{(e)}$ ) as we have noted previously (Lu et al., 2015). However, we have recently noted that each of the partial efficiencies  $\phi_{31}^{(u)}$ ,  $\phi_{31}^{(e)}$ ,  $\phi_{32}^{(u)}$ , and  $\phi_{32}^{(e)}$  may be heterogeneous across environments (Lu et al., 2017a) being similar to our previous findings that the partial regression parameters  $b_1$  and  $b_2$  in the RFI model of Equation [4.3] are heterogeneous across research stations (Tempelman et al., 2015). We briefly investigated whether the specification of heterogeneous station the partial regression parameters  $b_1$  and  $b_2$  to derive RFI in a manner similar to that pursued by Tempelman et al. (2015) might impact GWA inferences based on the RFI model. Figure C.1 in Appendix A provides the scatterplot of  $-\log_{10}(p)$  for SNPs and window-based on the homogeneous partial regression RFI versus heterogeneous partial regression RFI models. The correlation between  $-\log_{10}(p)$  values for SNPs and window-based on either model was 0.87 and 0.79. Admittedly, the detected regions between

the homogeneous and heterogeneous partial regression RFI models were not identical based on a formal Bonferroni test; nevertheless, both analyses models overlapped on 66% of the windows for  $P < 0.01$ . Accommodating heterogeneous partial efficiencies across research stations in a MT GWA analysis would be rather laborious using Monte Carlo Markov Chain approaches as pursued by Lu et al. (2017a). Nevertheless, these developments and others are worth pursuing in hierarchical Bayesian extensions for the MT model. For example, GBLUP assumes constant genetic variability across the genome whereas heavy tailed and/or sparse prior specifications have been known to exhibit better GWA properties (Moser et al., 2015). Similarly, one might perceive that  $\phi_{31}^{(u)}$ ,  $\phi_{31}^{(e)}$ ,  $\phi_{32}^{(u)}$ , and  $\phi_{32}^{(e)}$  (i.e., pleiotropic associations) are also heterogeneous across the genome as well.

It is also important to recognize that the single-step methods proposed by Wang et al. (2012) were vital to this GWA study given that 2021 of the 6937 cows with phenotypes were missing genotypes. For instance, when GWA analysis was based only using only phenotypes on genotyped cows, no significant associations were determined using window-based inference for RFI (Figure C.2bTable). Nevertheless the windows with the two most significant associations were the same in either case.

## 4.5 Conclusions

A GWA study on rFE and its components traits DMI, MBW, and MILKE in dairy cattle was conducted based on a MT model involving the component traits and the RFI model. The same 2 genomic regions on BTA12 and BTA26 were concluded to be associated with rFE in both models. The BTA26 region included a putative candidate gene ADAM12 that possibly regulates

muscle development and fatty acid utilization. The MT model facilitated comparisons of important SNPs or regions between rFE (i.e., RFI or DMI|MILKE,MBW) and its component traits (i.e. MILKE,MBW, and DMI). GWA efforts focusing primarily on DMI will not be useful for targeting genomic regions important for rFE.

## **Chapter5: Modeling genotype by environment interaction over multiple covariates for genome-wide association analyses of residual feed efficiency in dairy cattle**

### **5.1 Abstract**

Residual Feed efficiency (**rFE**) vitally impacts the environmental sustainability of the dairy industry. Genome-wide association (**GWA**) analyses based on high-density genomic marker panels are useful to narrow down genomic markers or regions and, eventually the identification of potential candidate genes associated with traits of interest such as rFE. However, most such studies infer an overall effect for each marker across environments, without allowing for the fact that allelic substitution effects at each marker may be sensitive to various environmental covariates. This form of genotype by environment (**GxE**) interaction is plausible for complex traits like rFE, particularly for data collected over a wide range of climatic and production systems. However, as the number of plausible environmental covariates increase, such inferences can be too intractable and computationally infeasible. We tested a reaction norms (**RN**) model extension to an conventional genomic model (**GBLUP**) whereby some simplifying assumptions in the RN model allow one to tractably solve for a cumulative RN effect across multiple covariates followed by equally tractable back-solving for RN specific marker effects. Based on simulation embedded in the genotypes and design structure of our consortium data, we determined that the RN model had superior GWA properties to GBLUP at false positive rate from 0.025 to 0.15 and was able to detect RN effects. We also applied both models to the dairy rFE consortium data analyses, with environmental covariates being linear and quadratic terms on average production (**AP**), relative humidity (**RH**), temperature (**TP**), and the interaction between RH and TP. In a cross-validation study, the proposed RN model yielded 5% greater prediction

accuracy compared to the regular GBLUP analysis. Gene *ELOVL4* and *ATP5E* were suggested as the candidate genes that have RN effects with linear term on AP and quadratic term on TP65, respectively. Our study indicated that GxE is an important genetic determinant for rFE.

## 5.2 Introduction

Residual feed efficiency (**rFE**) or dry matter intake (**DMI**) is being increasingly considered in dairy cattle selection programs. The availability of high-density single nucleotide polymorphism (**SNP**) marker panels has not only facilitated whole genomic prediction (**WGP**) for rFE or DMI but it also allows the use of genome-wide association (**GWA**) analyses to infer important genomic regions, and, potentially, any resident candidate genes or quantitative trait loci (**QTL**) that may influence these traits.

Most GWA studies, however, focus on estimating the overall allelic substitution effect of any SNP marker across environments, assuming that any QTL in linkage disequilibrium (**LD**) with that marker has a constant effect across environments. Conversely, heterogeneity in these effects across environment, also known as genotype by environment (**GxE**) interaction, could be important for rFE in dairy cattle as is true with other complex traits. Advanced cattle reproductive technologies, such as artificial insemination and embryo transfer, facilitate the exchange of germplasm across many different environments such that inference and management of GxE would seem more important in dairy cattle production than in many other livestock species where such technologies are not as widely used.

GxE modeling within a WGP or GWA context can lead to an extremely parameterized model if environment-specific SNP effects need to be modeled. A parsimonious strategy to model GxE is based on the use of environmental covariates. These models are often referred to as reaction

norm (**RN**) models and are based on modeling genetic or SNP effects as a function of a single covariate, often specified as the mean response for that particular environment or herd (Kolmodin, 2003). This covariate is often regarded to be a proxy for level or quality of herd management. In other words, GxE is modeled by specifying allelic substitution effects to be a linear function of a covariate (Lillehammer et al., 2009). If these SNP specific linear functions are large and important, the SNP marker is said to be environmentally sensitive. These models are being increasingly used to model GxE in WGP models (Lillehammer et al., 2009, Lopez-Cruz et al., 2015) since the dimensionality of the genomic part of the model is only doubled with a linear function on one covariate (Lillehammer et al., 2007a). A RN model is also an example of a random regression (Lillehammer et al., 2007b) or random coefficients model as the overall (i.e. average) and GxE effects for each SNP marker are modeled using random intercept and random slope specifications.

There may be any number of potential covariates that could be responsible for generating GxE. However, increasing the number of such covariates, including second or higher order polynomials and interactions between covariates, may unduly increase model dimensionality as well. Recently, (Jarquin et al., 2014) proposed the use of covariance functions for modeling GxE as a function of a large number of environmental covariates. We adapt their approach and demonstrate how to extend it further to conduct GWA analyses for covariate specific components of GxE. Specifically, our objectives were to infer upon GxE for rFE and, specifically, to determine which genomic regions may be environmentally sensitive to covariate-specific values.

## 5.3 Materials and Methods

### 5.3.1 Conventional Genomic Prediction Model for Residual Feed Intake

A commonly used response to characterize rFE is residual feed intake (**RFI**) (Tempelman et al., 2015). Based on definitions provided in (Berry and Crowley, 2013) and Kennedy et al. (1993), RFI is typically defined as DMI adjusted linearly for various energy sink covariates such as milk energy (**MILKE**), metabolic body weight (**MBW**), and change in body weight (**ΔBW**) over the course of the period from which DMI was recorded. Additionally specifying other effects that influence RFI, the statistical model can be specified as in Equation [5.1].

$$\text{DMI}_i = \mathbf{x}_i^{*'} \boldsymbol{\beta}^* + b_1 \text{MILKE}_i + b_2 \text{MBW}_i + b_3 \Delta \text{BW}_i + \mathbf{z}_i' \mathbf{g}_0 + e_i; i = 1, 2, \dots, n \quad [5.1]$$

Here, the subscript  $i$  on DMI, MILKE, MBW, and  $\Delta \text{BW}$  indicates that records on those specific traits belong to animal  $i$ , and  $b_1$ ,  $b_2$ , and  $b_3$  are partial regression coefficients of DMI on ME, MBW, and  $\Delta \text{BW}$  respectively. Furthermore,  $\boldsymbol{\beta}^*$  is a  $p \times 1$  vector of fixed effects with corresponding known row incidence matrix  $\mathbf{x}_i^{*'}$  whereas  $\mathbf{g}_0 = [g_{01} \ g_{02} \ \dots \ g_{0m}]$  is a vector of random SNP marker effects (i.e. random intercept) on RFI for  $\mathbf{z}_i' = [z_{i1} \ z_{i2} \ \dots \ z_{im}]$  being a known row incidence vector of genotypes on  $m$  markers for animal  $i$ . Even though genotypes are typically coded as the number of copies (0,1, or 2) of a reference allele,  $\mathbf{z}_i' = [z_{i1} \ z_{i2} \ \dots \ z_{im}]$  was standardized in the way implemented in VanRaden (2008). We assume throughout that  $\mathbf{g}_0 \sim N(\mathbf{0}, \mathbf{I}\sigma_g^2)$  and the residual  $e_i \sim \text{NIID}(0, \sigma_e^2) \forall i$ . Note that RFI response for animal  $i$ , can be written as  $\text{RFI}_i = \text{DMI}_i - b_1 \text{MILKE}_i + b_2 \text{MBW}_i + b_3 \Delta \text{BW}_i$  where typically  $b_1$ ,  $b_2$ , and  $b_3$  are estimated in a first stage model to derive  $\text{RFI}_i$ . In turn,  $\text{RFI}_i$  is written as a function of all

remaining effects in a second stage model; however, we prefer to fit all such effects simultaneously in Equation [5.1] as we've argued previously in Lu et al. (2015).

### 5.3.2 Reaction Norm Extension Involving Multiple Covariates

To model GxE using RN specifications for each of  $c$  covariates, we extend Equation [5.1] to Equation [5.2].

$$DMI_i = \mathbf{x}_i^* \boldsymbol{\beta}^* + b_1 MILKE_i + b_2 MBW_i + b_3 \Delta BW_i + \mathbf{z}_i' \mathbf{g}_0 + \sum_{k=1}^c d_{ik} \mathbf{z}_i' \mathbf{g}_k + e_i . \quad [5.2]$$

Additionally to what was already specified in Equation [5.1],  $\mathbf{g}_k = [g_{k1} \ g_{k2} \ \dots \ g_{km}]$  are the marker specific RN (i.e. random slope) effects on environmental covariate  $k$ ,  $k=1,2,\dots,c$  with  $d_{ik}$  being the value for environmental covariate  $k$  on animal  $i$ . For computational convenience, as we illustrate later, we invoke the distributional assumption identical to that used by Jarquin et al. (2014) whereby  $\mathbf{g}_k \sim N(\mathbf{0}, \mathbf{I}\sigma_b^2) \forall k$ ; in other words, the variance component is specified to be same (i.e.,  $\sigma_b^2$ ) for all covariate-specific slopes. With that strong assumption, it then seems particularly important that each covariate being standardized to a similar scale; e.g.; null mean and unit variance. Furthermore, standardizing all covariates to a null mean (i.e.,  $\bar{d}_{.k} = \frac{1}{n} \sum_{i=1}^c d_{ik} = 0 \forall k$ ) and variance of 1 allows the interpretation of  $\mathbf{g}_0$  as SNP effects for an average or typical environment.

In situations where  $m \gg n$ , it may be prudent to re-write Equation [5.2] as a direct function of animal effects rather than as SNP effects in the spirit of genomic BLUP (**GBLUP**) (VanRaden, 2008). That is, we write  $u_{0i} = \mathbf{z}_i' \mathbf{g}_0$  as the overall genetic merit for animal  $i$  and  $u_{ki} = \mathbf{z}_i' \mathbf{g}_k$  as the RN genetic merit for environment  $k$  on animal  $i$ . Subsequently, we sum all of the RN components  $u_i = [u_{1i} \ u_{2i} \ \dots \ u_{ci}]$  into one term  $v_i = \sum_{k=1}^c d_{ik} u_{ki}$  such that Equation [5.2] can be

rewritten as Equation [5.3]

$$\text{DMI}_i = \mathbf{x}_i^* \boldsymbol{\beta}^* + b_1 \text{MILKE}_i + b_2 \text{MBW}_i + b_3 \Delta \text{BW}_i + u_{0i} + v_i + e_i. \quad [5.3]$$

Write  $\mathbf{u}_0 = [u_1 \ u_2 \ \dots \ u_n]'$  and  $\mathbf{v} = [v_1 \ v_2 \ \dots \ v_n]'$  as vectors of animal specific effects of overall and cumulative (across covariates) RN genetic merit, respectively, with  $\mathbf{Z} = \{\mathbf{z}_i\}_{i=1}^n$  being the  $n \times m$  matrix of all genotypes. Note then that this implies  $\mathbf{u}_0 \sim N(0, \mathbf{G}\sigma_g^2)$  where  $\mathbf{G} = \mathbf{Z}\mathbf{Z}'$  is the genomic relationship matrix among animals (VanRaden, 2008). Also,  $\mathbf{v} = \{v_i\}_{i=1}^n \sim N(0, \mathbf{Z}\mathbf{Z}' \# \mathbf{D}\mathbf{D}'\sigma_b^2)$  where  $\mathbf{D} = [\mathbf{d}_1 \ \mathbf{d}_2 \ \dots \ \mathbf{d}_c]$  is of dimension  $n \times c$  with column  $k$  being  $\mathbf{d}_k = [d_{1k} \ d_{2k} \ \dots \ d_{nk}]'$ , an  $n \times 1$  vector of values on covariate  $k$ . Furthermore,  $\#$  denotes the Hadamard product. This covariance structure  $\mathbf{Z}\mathbf{Z}' \# \mathbf{D}\mathbf{D}'\sigma_b^2$  is identical to that specified in Jarquin et al. (2014) and can also be rewritten as  $\sum_{k=1}^c \text{diag}(\mathbf{d}_k) \mathbf{Z}\mathbf{Z}' \text{diag}(\mathbf{d}_k) \sigma_b^2$  where  $\text{diag}(\mathbf{d}_k)$  denotes a  $n \times n$  diagonal matrix with elements of  $\mathbf{d}_k$  along the diagonal.

To simplify notation further, we write  $\mathbf{x}_i' = [\mathbf{x}_i^* \ \text{MILKE}_i \ \text{MBW}_i \ \Delta \text{BW}_i]$  and  $\boldsymbol{\beta} = [\boldsymbol{\beta}^* \ b_1 \ b_2 \ b_3]$ . Then with  $\mathbf{X} = \{\mathbf{x}_i'\}$  the mixed model equations used to directly solve for the generalized least squares (GLS) of  $\boldsymbol{\beta}$  (i.e.,  $\hat{\boldsymbol{\beta}}$ ), and for the BLUP of  $\mathbf{u}_0$  and  $\mathbf{v}$  (i.e.,  $\hat{\mathbf{u}}_0$  and  $\hat{\mathbf{v}}$ ), respectively, is in Equation [5.4].

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}' & \mathbf{X}' \\ \mathbf{X} & \mathbf{I} + (\mathbf{Z}\mathbf{Z}'\sigma_g^2)^{-1}\sigma_e^2 & \mathbf{I} \\ \mathbf{X} & \mathbf{I} & \mathbf{I} + (\mathbf{Z}\mathbf{Z}'\# \mathbf{D}\mathbf{D}'\sigma_b^2)^{-1}\sigma_e^2 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}}_0 \\ \hat{\mathbf{v}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{y} \\ \mathbf{y} \end{bmatrix} \quad [5.4]$$

Note that  $\sigma_b^2$  and  $\sigma_e^2$  can either be specified by the user or readily estimated using REML.

### 5.3.3 Backsolving for SNP-specific Effects

Using the theory of backsolving for BLUP of one set of random effects as linear function of another set within an equivalent linear mixed model (Henderson, 1977),  $\hat{\mathbf{u}}_k$  being the BLUP of  $\mathbf{u}_k$  or the vector of animal effects for RN on covariate  $k$  can be determined from  $\hat{\mathbf{v}}$  in Equation [5.4] using Equation [5.5].

$$\begin{aligned}\hat{\mathbf{u}}_k &= \text{cov}(\mathbf{u}_k, \mathbf{v}')(\text{var}(\mathbf{v}))^{-1} \hat{\mathbf{v}} \\ &= (\text{diag}(\mathbf{d}_k) \sigma_b^2)(\mathbf{ZZ}' \# \mathbf{DD}' \sigma_b^2)^{-1} \hat{\mathbf{v}}. \\ &= \text{diag}(\mathbf{d}_k)(\mathbf{ZZ}' \# \mathbf{DD}')^{-1} \hat{\mathbf{v}}\end{aligned}\quad [5.5]$$

Similarly, one could use the same theory to solve for  $\hat{\mathbf{g}}_k = [\hat{g}_{k1} \quad \hat{g}_{k2} \quad \hat{g}_{km}]' = \text{BLUP}(\mathbf{g}_k)$  or the vector of SNP RN effects specific to covariate  $k$ :

$$\begin{aligned}\hat{\mathbf{g}}_k &= \text{cov}(\mathbf{g}_k, \mathbf{v}')(\text{var}(\mathbf{v}))^{-1} \hat{\mathbf{v}} \\ &= (\mathbf{Z}' \text{diag}(\mathbf{d}_k) \sigma_b^2)(\mathbf{ZZ}' \# \mathbf{DD}' \sigma_b^2)^{-1} \hat{\mathbf{v}} \\ &= \mathbf{Z}' \text{diag}(\mathbf{d}_k)(\mathbf{ZZ}' \# \mathbf{DD}')^{-1} \hat{\mathbf{v}}\end{aligned}\quad [5.6]$$

### 5.3.4 Genome Wide Associations on SNP-specific Overall Effects and Reaction Norms

BLUP estimates (i.e.,  $\hat{\mathbf{g}}_k$ ) by themselves do not directly indicate as to which chromosomal regions might be directly important for overall genetic or environment-specific RN effects. A common GWA inference strategy is to use the EMMAX procedure (Kang et al., 2010) whereby the marker of interest is tested using GLS, treating all other markers as random, or equivalently modeling correlation between animals based on a genomic relationship matrix constructed from all SNP markers. Recent developments described by Gualdron Duarte et al. (2014) and adopted previously by our group lead to the following GLS test statistic for testing overall genetic effects ( $k=0$ ) and the RN effect for SNP  $j$  on covariate  $k$ :

$$z_{kj} = \frac{g_{kj}}{\sqrt{\text{var}(g_{kj})}} . \quad [5.7]$$

Underneath the null hypothesis  $H_0: g_{kj} = 0$ , it can be demonstrated that  $z_{kj} \sim N(0,1)$ . Using mixed model theory,  $\text{var}(g_{kj}) = \text{var}(g_{kj}) - \text{var}(g_{kj} - g_{kj})$  and can be determined as element  $k,k$  of the expression in Equation [5.8].

$$\begin{aligned} \text{var}(\hat{\mathbf{g}}_k) &= \text{var}\left(\mathbf{Z}' \text{diag}(\mathbf{d}_k)(\mathbf{ZZ}'\mathbf{DD}')^{-1} \hat{\mathbf{v}}\right) = \\ &= \mathbf{Z}' \text{diag}(\mathbf{d}_k)(\mathbf{ZZ}'\mathbf{DD}')^{-1} \text{var}(\hat{\mathbf{v}})(\mathbf{ZZ}'\mathbf{DD}')^{-1} \text{diag}(\mathbf{d}_k) \mathbf{Z} \quad . \quad [5.8] \\ &= \mathbf{Z}' \text{diag}(\mathbf{d}_k)(\mathbf{ZZ}'\mathbf{DD}')^{-1} (\mathbf{ZZ}'\mathbf{DD}'\sigma_b^2 - \mathbf{C}_{bb}\sigma_e^2)(\mathbf{ZZ}'\mathbf{DD}')^{-1} \text{diag}(\mathbf{d}_k) \mathbf{Z} \end{aligned}$$

Here  $\mathbf{C}_{bb}$  is portion of the inverse of the coefficient matrix in Equation [5.4] that corresponds to the lower right portion corresponding to cumulative RN effects; i.e.

$$\begin{bmatrix} \mathbf{C}_{\beta\beta} & \mathbf{C}_{\beta g} & \mathbf{C}_{\beta b} \\ \mathbf{C}_{g\beta} & \mathbf{C}_{gg} & \mathbf{C}_{gb} \\ \mathbf{C}_{b\beta} & \mathbf{C}_{bg} & \mathbf{C}_{bb} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}' & \mathbf{X}' \\ \mathbf{X} & \mathbf{I} + (\mathbf{ZZ}'\sigma_g^2)^{-1} \sigma_e^2 & \mathbf{I} \\ \mathbf{X} & \mathbf{I} & \mathbf{I} + (\mathbf{ZZ}'\mathbf{DD}'\sigma_b^2)^{-1} \sigma_e^2 \end{bmatrix}^{-1} \sigma_e^2 \quad [5.9]$$

More detailed derivations can be found in Appendix D.

For a GWA analysis that jointly fits all SNP markers, any given marker may exhibit only a weak association with a closely linked QTL, in part because of the high LD that exists with neighboring markers (Lu et al., 2017b). Hence, we also considered genomic based windows associations based on partitioning columns of  $\mathbf{Z}$ , and hence elements of  $\mathbf{g}_0, \mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_c$ , into  $R$  distinct components or windows with each window containing a certain number of SNP markers and/or based on a genomic length. In this study, we partitioned the genome into non-overlapping 1megabase (**MB**) regions resulting in a total of  $R = 2,673$  windows. For joint testing  $H_0: \mathbf{g}_{kr} = 0$

pertaining to window  $r$  for SNP effects in  $\mathbf{g}_k$ ,  $k=0,1,2,...c$ , Lu et al. (2017b) extended the test from Gualdron Duarte et al. (2014) to the test statistic as in Equation [5.10]

$$\chi_{kr}^2 = \mathbf{g}_{kr}' \left( \text{var}(\mathbf{g}_{kr}) \right)^{-1} \mathbf{g}_{kr} \quad [5.10]$$

where

$$\text{var}(\hat{\mathbf{g}}_{kr}) = \mathbf{Z}_r' \text{var}(\hat{\mathbf{g}}_k) \mathbf{Z}_r. \quad [5.11]$$

Under  $H_0$ :  $\mathbf{g}_{kr} = 0$ , the test statistic in Equation [5.10] is chi-square distributed with degrees of freedom  $n_r$ , being the number of SNP markers in window  $r$ .

Data on all key phenotypes were collected on 3,463 Holstein cows from 8 different dairy research stations including University of Alberta (**AB**) in Edmonton, Canada, Iowa State University (**ISU**) in Ames, IA, Michigan State University (**MSU**) in East Lansing, MI, University of Florida (**UF**) in Gainesville, FL, University of Wisconsin-Madison (**UW**) in Madison, WI, United States Dairy Forages Research Center (**USDFRC**) in Madison, WI, Virginia Polytechnic Institute and State University (**VT**) in Blacksburg, VA, and the USDA Animal Genomics and Improvement Laboratory (**AGIL**) in Beltsville, MD. Data editing procedures were outlined in Tempelman et al. (2015) and based on the first 28 d after 50 DIM as in Lu et al. (2017b) and Manzanilla-Pech et al. (2016). The distribution of cows in each station can be found in Table 5.1. Daily weather data including relative humidity (**RH**) and temperature (**TP**) was retrieved using the R package weatherData for all the test dates within each station. These weather records were averaged across the 28 days of data on each animal.

Table 5.1: Brief characterization of dataset by stations

Country	Station <sup>1</sup>	No. of animals
United State	UF	362
	ISU	930
	MSU	264
	UW	780
	PANC	18
	USDFRC	347
	USDA AGIL	488
	VT	54
Canada	AB	220
Overall		3463

<sup>1</sup> AB = University of Alberta; UF = University of Florida; ISU = Iowa State University; MSU = Michigan State University; MIN= Dairy Research Facility at Miner Institute; UW = University of Wisconsin–Madison; PANC= Purina Animal Nutrition Center; USDFRC = USDA Dairy Forages Research Center; USDA AGIL = USDA Animal Genomics and Improvement Laboratory; VT= Virginia Polytechnic Institute and State University;

Genomic characterization was based on Illumina® BovineSNP50 Genotyping BeadChip, which features 54,609 informative SNP probes uniformly spread over the bovine genome. Raw Genotypes were pre-processed by the USDA’s Animal Improvement Programs Laboratory (AIPL) where parent-progeny conflicts were checked. Genotypes were then increased to a final set of 61,013 SNPs, including additional 15,818 selected SNPs from bovine HD chips based on the magnitude of their effect on traits evaluated in the U.S. (Wiggans et al., 2016). Further editing and filtering were implemented to rule out SNPs with MAF <0.05 and in perfect LD which resulted in 57,347 SNPs being left. The SNPs on X chromosome were classified into two groups: pseudo-autosomal and X-specific SNPs.

### 5.3.5 Simulation Study

To partly validate our proposed method and model, a simulation study was conducted based on actual genotype and data structure within our dairy rFE consortium dataset. Environmental covariates were first, second, third and fourth order polynomial terms on average contemporary group production (**AP**) for the contemporary group defined as a group of cows in 28-day test period within the same station. First to fourth polynomial terms of AP were standardized to mean of 0 and variance of 1. Thirty of the 57,347 markers were randomly chosen to be QTL. With columns of **Z** also standardized to mean 0 and variance of 1, random intercepts (i.e., QTL subset of **g<sub>o</sub>**) was simulated from independent Gaussian distributions with mean of 0 and variance of 1000 such that variance of overall genetic merit **u<sub>o</sub>** was roughly 0.50. RN effects on first to fourth order polynomial terms of AP (i.e.,  $c = 4$ ) were simulated from independent Gaussian distributions with the same variance of 200 and mean of 0. Phenotypes were based on simulating breeding values on all such QTL effects plus a residual generated from a Gaussian distribution with variance  $\sigma_e^2 = 2.0$  so that the heritability under the average environmental covariate value was roughly 0.20. The heritability of rFE ranged from 0.20 to 0.46 across the values of AP. Ten replicated datasets were simulated in this manner and then analyzed based on the proposed RN model and a conventional GBLUP analyses. Variance components were estimated using REML in all cases with the R package regress.

Window-based GWA tests based on non-overlapping 1MB genomic regions and the chi-square test presented by Lu et al. (2017b) were conducted for random intercepts (**g<sub>o</sub>**) and slopes (**g<sub>1</sub>**, **g<sub>2</sub>**, **g<sub>3</sub>**, and **g<sub>4</sub>**) on each replicate using Equation [5.10]. A true positive was claimed if the

corresponding window including that QTL was declared to be significant based on a Bonferroni test. Conversely, a false positive was determined if a window was deemed to be statistically significant but did not include a QTL. The rates of true and false positives were used to compare receiver operating characteristic (**ROC**) curves averaged across the 10 replicates between the proposed RN model and conventional GBLUP. The area under the ROC curve (**AUC**) compared between two models at various false positive rates (**FPR**) (Wray et al., 2010) based on a nonparametric Wilcoxon Signed-Rank test, blocking on data replicate.

### **5.3.6 Application to RFI Data**

The proposed RN model was applied to the joint dairy rFE dataset described above. For modeling RN on milk production, it has been determined that a “broken stick” relationship might be best suited to model the relationship between milk yield and TP, recognizing that heat stress and hence potential GxE effects due to TP might not occur until beyond a certain breakpoint. This breakpoint has often deemed to be around 72°F in various studies involving milk yield (Bohmanova et al., 2007, Aguilar et al., 2009, Ludovico et al., 2015). Below this breakpoint up until cold stress, which is not likely a major consideration with intensively managed research facilities in this study, animals are believed to be in a thermoneutral environment by which there should be little, if any, regulatory changes in metabolic heat production or evaporative heat loss .

We decided to infer upon this breakpoint via a series of 8 RN models/analyses whereby 6 of the models were labeled TP55, TP60, TP65, TP70, TP75, TP80, for which TP breakpoints on RN specifications ranged from 55 to 80 °F in 5 °F increments, respectively. The other two models were TPN where no breakpoint was specified and TPG as a conventional GBLUP model.

Specifically, environmental covariates used for SNP-specific RN effects for each of 8 RN models included linear and quadratic terms on corresponding TP (i.e., TPN, TP55, TP60, TP65, TP70, TP75, or TP80), AP, RH, and crossproduct between corresponding TP and RH. Thus, a total of  $c=8$  environmental covariates were considered for any RN analyses on RFI.

The vector of fixed effects  $\beta^*$  in Equation [5.3] for the RN models included the effects of stations, rations within stations, first through fourth order polynomial effects of DIM, parity effects, the interaction between first through fourth order polynomial effects of DIM and parity effects, and fixed environmental covariates. To facilitate a likelihood ratio test (or equivalently information criteria like assessments based on REML), the same fixed environmental covariates were specified in  $\beta^*$  for all 8 candidate models. These fixed environmental covariates included linear and quadratic effects on RH and AP and 6th order Bspline functions on each of TP and TP\*RH, the linear by linear crossproduct of TP and RH.

All variance components were estimated by REML. SNP-specific RN effects for each environmental covariate were back-solved using Equation [5.6] as previously noted and single SNP hypothesis tests were conducted. We subsequently investigated inferences for non-overlapping 1MB windows based on the chi-square test in Lu et al. (2017b) as shown in Equation [5.10]. Association between regions and the trait of interest were declared if the test was significant based on a Bonferroni adjustment for the number of SNP markers for single SNP tests or number of windows for window-based tests. Potential candidate genes located within windows deemed to have RN with the various environmental covariates were determined and

their gene function information were retrieved using the Bioconductor package biomaRt based on the UMD3.1 bovine assembly from Ensembl.

As a comparison between RFI and DMI, we also fitted the proposed RN model on DMI. The difference between the analyses on RFI and DMI was simply that the covariates ME, MBW, and  $\Delta$ BW were not fitted in RN model in Equations [5.2-5.3]; i.e.,  $\beta = \beta^*$ . All other effects were the same as those considered for RFI.

### 5.3.7 Cross-validation Assessment

In order to compare the performances between the best fitting RN model (based on the chosen breakpoint for TP) and the regular GBLUP model for rFE, a 5-fold cross-validation was performed. DMI prediction accuracy determined as the correlation between true DMI and predicted DMI in the validation dataset based on estimates derived from the analyses of the training dataset was computed and compared to determine the performances of two models.

## 5.4 Results

### 5.4.1 Simulation Study

ROC curves for the random intercepts or overall SNP effects ( $\mathbf{g}_o$ ) based on the RN model and the conventional GBLUP were plotted for 1MB windows (Figure 5.1). Generally, the ROC curve for the RN model straddled the ROC curve for the GBLUP model although the ROC curve for the RN model tended to exceed that for the GBLUP model for false positive rate (**FPR**) rates between 0.025 and 0.30. To formally test if difference was significant, we calculated the AUC for FPR at 0.01, 0.025, 0.05, 0.15, 0.30, 0.40, and 0.50 for the regular GBLUP and the RN model. The nonparametric Wilcoxon Signed-Rank test on AUC implied that the RN model significantly

increased QTL detection power at FPR ranging from 0.025 to 0.15 (see Table 5.2); nevertheless differences were small.

Table 5.2: Areas under receiver operating characteristic curve (AUC) between the classic GBLUP and the reaction norm (RN) model at different false positive rates.

	False positive rates						
	.01	.025	.05	.15	.30	.40	.50
RN	.15±.04	.58±.13	1.49±.3	6.48±1.2	15.8±2.4	22.9±3.2	30.6±3.7
GBLUP	.14±.04	.53±.12	1.41±.3	6.16±1.3	15.3±2.7	22.6±3.4	30.3±4.1
$p^1$	.221	.008	0.042	0.041	0.056	0.262	0.68

<sup>1</sup> Based on the nonparametric Wilcoxon Signed-Rank test on the AUC

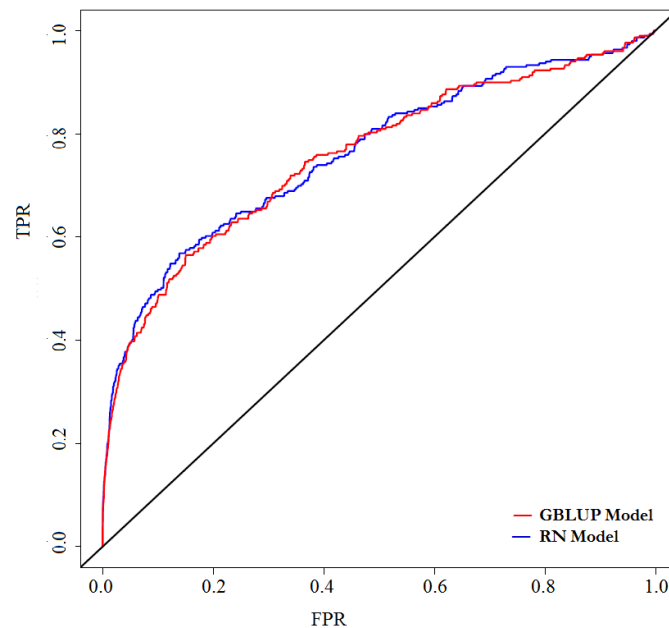


Figure 5.1: Receiver operating characteristic (ROC) curves for random intercept for 1Mb non-overlapping windows based on the reaction norm (RN) model and the classic GBLUP averaged across 10 replicates. Line of slope 1 and intercept 0 superimposed.

The ROC curve for RN inferences on first to fourth order polynomial terms on AP are also provided in Figure 5.2. Generally, the ROC curves for the RN effects on all four polynomial

terms of AP were above the reference line with slope 1 and intercept 0. This reference line is expected if QTL windows for RN effects are simply guessed. The nonparametric Wilcoxon Signed-Rank test on AUC at FPR ranging from 0.01 to 0.50 confirmed that RN effects on all four polynomial terms of AP were significantly different from reference line ( $P < 0.01$ ), indicating that inferences on windows containing RN effects are at least superior to random guessing. Furthermore, there seems to be no evidence of differences between AUC for ROC curves for any of the four RN polynomial specifications. Furthermore, AUC were generally much smaller for RN inferences (Figure 5.2) compared to random intercept or overall genetic effects (Figure 5.1).

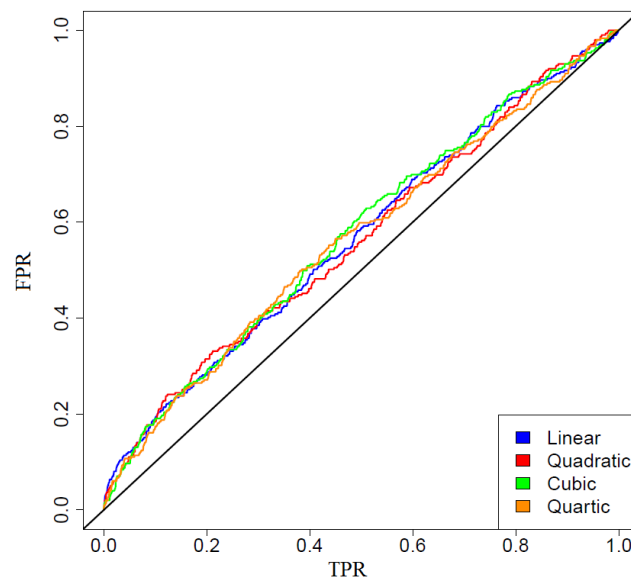


Figure 5.2: Receiver operating characteristic (ROC) curves for random slopes on first to fourth order polynomial terms of AP for 1Mb non-overlapping windows based on the reaction norm (RN) model and the classic GBLUP averaged across 10 replicates. Line of slope 1 and intercept 0 superimposed.

### 5.4.2 Real Data Analyses

To determine an appropriate breakpoint for TP on RFI and DMI, variance components and LRT statistics, i.e. -2 times the difference in log REML likelihood between full models in TPN and TP55-TP80 and reduced model in TPG, were calculated and summarized in Tables 3 and 4. Compared to the regular GBLUP model in TPG, the 7 tested RN models involving various breakpoints on TP all significantly fitted better than GBLUP for both RFI (Table 5.3), and DMI (Table 5.4), with seemingly greater statistical significance of RN on RFI compared to DMI based on the relative magnitude of the LRT.

Table 5.3: Estimated variance components and likelihood ratio test for residual feed intake based on the reaction norm (RN) models with various breakpoints on temperature (TP).

Models	$\sigma_u^2$	$\sigma_b^2$	$\sigma_e^2$	LRT Statistics <sup>1</sup>
TPN	.46±.08	.043±.010	2.31±.10	27.60
TP55	.46±.08	.041±.010	2.32±.10	24.60
TP60	.45±.08	.051±.010	2.25±.10	33.87
TP65	.45±.08	.061±.010	2.18±.10	50.91
TP70	.48±.08	.030±.008	2.39±.10	20.55
TP75	.48±.08	.027±.006	2.42±.09	24.09
TP80	.49±.08	.023±.006	2.46±.08	21.22
TPG	.49±.08		2.61±.09	

<sup>1</sup> Likelihood ratio test statistics, relative to TPG model

Table 5.4: Estimated variance components and likelihood ratio test for dry matter intake based on the reaction norm (RN) models with various breakpoints on temperature (TP).

Models	$\sigma_u^2$	$\sigma_b^2$	$\sigma_e^2$	LRT Statistics <sup>1</sup>
TPN	1.40±.18	.051±.016	4.15±.18	12.51
TP55	1.41±.18	.045±.017	4.18±.19	8.44
TP60	1.40±.18	.050±.016	4.15±.19	10.48
TP65	1.41±.18	.047±.015	4.17±.18	11.50
TP70	1.42±.18	.034±.014	4.27±.18	7.52
TP75	1.42±.18	.024±.010	4.34±.17	6.39
TP80	1.44±.18	.021±.008	4.36±.16	7.93
TPG	1.42±.18		4.52±.16	0

<sup>1</sup> Likelihood ratio test statistics, relative to TPG model

Model TP65 was the best fitting model for RFI having the largest LRT statistic relative to GBLUP suggesting a broken stick RN relationship for TP at a breakpoint of 65°F. This thereby suggests that GxE as a function of TP is important above 65°F. For the best fitting RN model for RFI (i.e. TP65),  $\hat{\sigma}_b^2 = 0.06 \pm 0.01$ . Given that genotypes are standardized in the manner described previously, the diagonals of  $\mathbf{ZZ}'$  are relatively close to 1 (0.913-1.162), implying that the diagonals of  $\mathbf{ZZ}'\mathbf{DD}'$  are relatively close to the diagonals of  $\mathbf{DD}'$  which are simply the sum of the squared environmental covariates for any one animal; i.e., element  $i$  of  $\mathbf{DD}'$  is  $\sum_{k=1}^c d_{ik}^2$ . So for a genomically non-inbred animal (i.e. corresponding diagonal element of  $\mathbf{ZZ}'$  is close to 1),

the heritability based on this model is defined to be  $h^2 = \frac{\sigma_g^2 + \sum_{k=1}^c d_{ik}^2 \sigma_b^2}{\sigma_g^2 + \sum_{k=1}^c d_{ik}^2 \sigma_b^2 + \sigma_e^2}$ . This model further

constrains the heritability to be lowest in an average environment (i.e.,  $d_{ik} = 0 \forall k$  such that

$\sum_{k=1}^c d_{ik}^2 \sigma_b^2 = 0$  and subsequently  $h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$ ). Based on the estimates provided in Table 5.3,

this implies that the estimated heritability for rFE  $\hat{h}^2 = \frac{0.45}{0.45 + 2.18} = 0.17$  at an average

environment, consistent with what we have reported previously (Lu et al., 2017b). The highest heritability determined for RFI is  $\hat{h}^2 = \frac{0.45 + 106.04 * 0.061}{0.45 + 106.04 * 0.061 + 2.18} = 0.76$  for  $\sum_{k=1}^8 d_{ik}^2 = 106.04$  at the most extreme environment.

### 5.4.3 Genome-wide Association Analyses for RFI

*Random Intercept:* Genome-wide association analyses for RFI using both the single SNP and 1MB window-based inference strategies were conducted under the best fitting RN model (i.e., TP65) as well as GBLUP (i.e., TPG). Firstly, there was no evidence of an association for overall genetic effects ( $\mathbf{g}_o$ ) based on neither single SNP (Figure D.1 in Appendix D) nor window-based inferences (Figure D.2 in Appendix D) whether using the conventional GBLUP model or the TP65model. Furthermore, we noted that there was little change in  $-\log_{10}(\text{P-values})$  between either model for either of the two types (single SNP versus window-based) of inferences on random intercept  $\mathbf{g}_o$  (Figure 5.3).

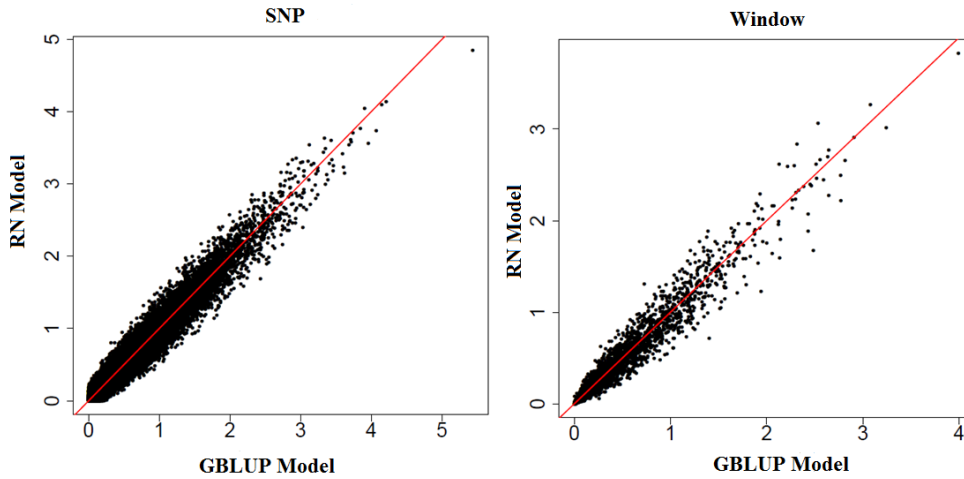


Figure 5.3: Comparison on  $-\log_{10}(p)$  of SNP (a) and 1Mb window (b) for random intercept between the GBLUP and the reaction norm (RN) model. Line of slope 1 and intercept 0 superimposed.

We also inferred SNP-specific and window-specific associations for RN on each environmental covariate or polynomial or interaction thereof. Manhattan plots for SNP-specific RN effects (i.e. random slopes) on linear and quadratic terms of TP65, RH, AP, and interaction between TP65 and RH were provided (see Figure D.3-D.4 in Appendix D). For SNP specific associations, there was only 1 detected SNP (Hapmap58910-rs29012613, 20,171,743 bp) on BTA 9 pertaining to the linear RN on AP (see Figure D.3).

Manhattan plots for window-specific RN effects on various environmental covariates are also provided (see Figures D.6-D.8 in Appendix D). A region at 20.5 Mb on BTA9 was found to have a significant association for linear RN effect of AP whereas a region at 82.5 Mb on BTA5 was associated with a quadratic RN effect of AP (see Figure D.6a and Figure D.6b). Two windows on BTA13 centered at 49.5 Mb and 58.5 Mb, were associated with the quadratic RN effect of TP65 (Figure D.7b) whereas no such region was detected to be associated with the linear RN term of TP65 (Figure D.7a). Furthermore, there was no detected important region having RN association for linear and quadratic terms of RH nor interaction between TP65 and RH (see Figure D.8). A few candidate genes in the regions that identified having substantial RN effects with corresponding environmental covariates were retrieved and are listed in Table 5.5.

Table 5.5: Symbol, name, and biological process of genes have reaction norm (RN) effects with temperature breaks at 65F (TP65) and contemporary group milk production (AP) for the residual feed intake

Environmental covariate	Symbol	Name	Function involved
Linear on AP	<i>SH3BGRL2</i>	SH3 Domain Binding Glutamate Rich Protein Like 2	protein disulfide oxidoreductase activity
	<i>ELOVL4</i>	Bos taurus ELOVL fatty acid elongase 4	the biosynthesis of fatty acids
Quadratic on AP	<i>KLHL42</i>	Kelch Like Family Member 42	ubiquitin-protein transferase activity
Quadratic on TP65	<i>CRLS1</i>	Bos taurus cardiolipin synthase 1	synthesis of cardiolipin, critical for mitochondrial function
	<i>ATP5E</i>	Bos taurus ATP synthase, H <sup>+</sup> transporting, mitochondrial F1 complex, epsilon subunit	synthase catalyzes ATP synthesis

#### 5.4.4 Cross-validation Accuracy

We computed the accuracy of DMI prediction for both the regular GBLUP model and the best fitting RN model. Prediction accuracy for DMI defined as the correlation between true and estimated DMI in validation dataset for each folder was listed in Table 5.6. Prediction accuracy for DMI in the simplified RN model was generally 5% higher compared to that based on the regular GBLUP ( $P<0.01$ ).

Table 5.6: Cross Validation Prediction Accuracies for each of 5 different validation datasets (fold) under the regular GBLUP and the reaction norm (RN) model.

Fold	Models	
	GBLUP (TPG)	RN (TP65)
1	0.77	0.81
2	0.76	0.80
3	0.74	0.79
4	0.73	0.79
5	0.80	0.84
Mean	0.76	0.81

## 5.5 Discussion

In this study, we considered the high dimensional RN model proposed by Jarquin et al. (2014) to investigate GxE for RFI over linear and quadratic terms on TP, RH, AP, and interaction between TP and RH. These covariates are commonly considered in GxE studies in plant and animal breeding (Hayes et al., 2016).

We did not discover any significant SNP for overall genetic effect (i.e. random intercept). Similarly, there have been other studies failing to detect any genomic associations with rFE suggesting that it is a rather complex trait (Lu et al., 2013, Do et al., 2014, Santana et al., 2014, Serão et al., 2016). We also detected no significant associations for overall genetic effects based on window-based associations. This result was unlike what we have determined previously (Lu et al., 2017b). However, that previous work entailed the use of single-step GBLUP techniques (Aguilar et al., 2010) which permit the incorporation of data on non-genotyped cows. Furthermore, that previous analysis also included cows from countries other than U.S. and Canada; we were not able to obtain weather information on data from research stations in the United Kingdom and in the Netherlands.

We also noticed that inferences on overall or random intercept effects whether based on single SNP or window-based associations differed very little between conventional GBLUP and the RN models. This might imply that whether or not one fits RN effects does not weaken the ability to detect QTL effects averaged across environments.

Noting that cross validation indicated that our RN model yielded higher prediction accuracies for DMI compared to the regular GBLUP, GWA inferences lead to the detection of four regions having significant RN effects with environmental covariates AP or TP65. These results suggest these genomic regions are environmentally sensitive to these covariates for their effect on rFE. Among genes identified in those regions, the two most promising candidates seemed to be gene *ELOVL* Fatty Acid Elongase 4(*ELOVL4*, 19,842,903-19,881,703 bp) and ATP synthase subunit epsilon (*ATP5E*, 57,866,138-57,869,829bp). *ELOVL4* has been related to the biosynthesis of fatty acids, while gene *ATP5E* controls mitochondrial ATP synthesis (Mayr et al., 2010). Fatty acids are important in feed digestion and metabolism and control body weight and insulin sensitivity (Canfora et al., 2015), whereas ATP plays important roles in intracellular energy transfer and involves in many biological processes including growth and maintenance (Maruyama, 1991, Tort et al., 2011).

The RN model proposed by Jarquin et al. (2014) provided an efficient approach to investigate GxE under multiple environmental factors for WGP and GWA studies. It is much more computationally feasible compared to the classical RN model, thus widely used in plant and human genomic research (Crossa et al., 2016, De Coninck et al., 2016, Vazquez et al., 2016). Note that if all those random RN slopes are random draws from an identically and independent

normal distribution and zero correlations are specified between random intercept and RN slopes, then Jarquin's model would be equivalent to a conventional RN model.

Note that the common variance for all RN slopes is a rather strong assumption. The natural follow-up question is what are the implications for GWA inferences if, as expected, variance components are rather heterogeneous across different RN covariates, including different polynomials or interactions? We speculate that the implications may be not so serious at first glance, although admittedly the seriousness of this issue will depend greatly on the degree of heterogeneity in variance components between RN terms. Firstly, as indicated earlier, it is important to invoke standardizations on covariates so that scale dependence is alleviated, and this is what we did in this study. Secondly, our GWA inferences are based on test statistics which are equivalent to GLS test statistics. That is, these tests specify the single SNP effect or joint SNP effects in a window fixed effects in hypothesis testing. Hence the mis-specification of variance components should only misalign somewhat the specification of genomic-based kinship covariances between individuals in a manner similar to that used in EMMAX, noting that one would estimate an "average" variance component (i.e.,  $\sigma_b^2$ ) with the Jarquin's model.

One possible strategy to allow for covariate or RN specific variance components is to fit each covariate separately in a series of single-covariate RN analysis and estimate the covariate-specific RN variance component  $\sigma_{bk}^2$  for covariate  $k$  for each of  $c$  separate single covariate RN analyses. Upon collating the corresponding estimates  $\hat{\sigma}_{bk}^2$ ,  $k=1,2,3,\dots,c$  across these analyses, one might then jointly fit all covariate-specific RN effects in the Jarquin's model fixing  $\text{var}(\mathbf{v}) = \sum_{k=1}^c \text{diag}(\mathbf{d}_k) \mathbf{Z} \mathbf{Z}' \text{diag}(\mathbf{d}_k) \hat{\sigma}_{b,k}^2$ . Admittedly, this might be considered a rather ad hoc

solution. A potentially more elegant yet computationally demanding method could be based on Bayesian variable selection. Bayesian variable selection approach automatically allows difference variances on those individual RN terms and performs shrinkage on those less important RN terms so that no preselection on the environmental covariates is needed. Some of this work has been already been done in the context of single RN covariate specifications (Yang et al., 2015). For high dimensional environmental covariate modeling, it seems rather challenging to estimate unique variance components and/or covariances between all possible RN terms in somewhat finite datasets. One conceivable strategy might be to extend the model with augmented variables such that there is a borrowing of information across RN specific variance and covariance components using hierarchical Bayesian modeling (Bello et al., 2012).

Another problem with the Jarquin's model relates to the fact that no covariances are specified between any of  $\mathbf{g}_0, \mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_c$ . With doing so, this invokes the dubious assumption that the total genetic variance  $\sigma_g^2 + \sum_{k=1}^c d_{ik}^2 \sigma_b^2$  and hence, heritability, is lowest at the point where  $\sum_{k=1}^c d_{ik}^2$  is smallest. However, it is not necessarily logical to presume that this should even be so. Perhaps one strategy to help circumvent this issue is to determine the environments (and covariate values thereof) where heritability is estimated to be lowest in a series single covariate RN analyses and specify these values as the reference values for which the standardized covariates are expressed as zeroes in those cases. It also seems possible to extend Jarquin's model to specify a common covariance between intercept and every single slope and fit that to solve the mixed model equation of Equation [5.4] but that may be as dubious as an assumption as specifying zero covariance between intercept and every single slope.

Finally, we believe Jarquin's RN procedure could be readily extended to single-step GBLUP analyses to include information on non-genotyped animals as well. Furthermore, the analyses in this chapter were based on the RFI phenotype whereas we have highlighted a multiple trait approach to modeling rFE in this dissertation (Lu et al., 2015). The extension of Jarquin's model to the multiple trait approach is possible although that might potentially entail specifying RN on all of the component traits which could be somewhat laborious.

## **5.6 Conclusions**

We extended the high-dimensional environmental covariate RN model from Jarquin et al. (2014) to investigate which genomic regions might be contributing to environmentally sensitive to various covariates for rFE. We were able to demonstrate using a simple simulation study that this model is able to detect associations for various orders of polynomial RN effects although the impact of detection of causal variants for overall genetic effects was very slight relative to a conventional GBLUP model. We conducted GWA analyses on a dairy feed consortium study using inferences based on both single SNP marker and 1 MB window associations. We detected no association for overall genetic effects. However, two windows on BTA9 and BTA5 were found to have significant associations with the linear and quadratic RN on AP, respectively, and two genomic regions on BTA 13 were associated with the quadratic RN effects of TP65. Cross-validation confirmed that GxE plays an important role for influencing rFE. However, this model was based on some simplifying assumptions that should be extended with future methodological research.

## Chapter6: Conclusions and Future Work

### 6.1 Conclusions

The primary focus of this dissertation has been to develop meaningful methodological extensions, whether based on either classical mixed models theory or on hierarchical Bayesian inference, in order to develop a much stronger statistical framework for the quantitative genetic modeling of residual feed efficiency (**rFE**) in dairy cattle than what has been considered in previous research. These extensions are timely since there are simultaneous global research efforts in collecting dry matter intake (DMI) data for the purpose of improving the long-term economic and environmental sustainability of dairy cattle production. One such project is the USDA-NIFA project entitled “Genomic selection and herd management for improved efficiency of the dairy industry” (USDA-NIFA Grant # 2011-68004-30340) which has funded this effort and provided the data for this dissertation. This data consists of DMI, body weights (BW) and milk production and its components collected between 50 and 200 days in milk (DIM) on nearly 7,000 cows from 16 research stations in 4 countries.

In Chapter 2, we developed a genetic measure of rFE based on a multiple trait (MT) analyses of DMI, metabolic body weight ( $MBW = BW^{0.75}$ ), and milk energy (MILKE) based on a linear function of milk component yields. The currently most popular “response” variable for rFE is residual feed intake (RFI) which is based on the estimated residual from fitting DMI as a linear function of MILKE and MBW as well as change in body weight throughout the recording period. The RFI is then typically modeled as a function of genetic effects in a second stage model although we have made the case that RFI inferences can be based on a single-step model (Lu et al., 2015). We demonstrated in Chapter 2 how the RFI modeling strategy assumes that the partial relationships between DMI and MILKE and MBW are the same at both genetic and non-

genetic levels. Conversely, our proposed MT model, invoking the use of the Cholesky decomposition on the genetic and residual variance-covariance matrices, allows these partial relationships to be specified differently for genetic and residual levels. The potential impact of large discrepancies between genetic and residual partial relationships was demonstrated by simulation such that the MT model can lead up to 4% greater accuracy of genetic prediction. However, there appeared to be no appreciable difference in fit between the MT and RFI model when applied to the USDA-NIFA feed efficiency project data. Heritability estimates ( $\pm$  standard error) were also similar for rFE measures under both models, being  $0.14 \pm 0.03$  under the MT model and  $0.16 \pm 0.03$  using the RFI model. Nevertheless, the MT approach is promising in the sense that it could handle missing data; this is a non-trivial consideration for future genetic evaluations of rFE knowing that the number of records on DMI will generally be far less than MILKE, for example. This should then provide a more promising approach to provide genetic merit on rFE for all US dairy cows than the current ad hoc strategy consider in the Total Performance Index (TPI) by Holstein Association USA which bases rFE on a linear formula involving milk produced and DMI ([http://www.holsteinusa.com/genetic\\_evaluations/ss\\_tpi\\_formula.html](http://www.holsteinusa.com/genetic_evaluations/ss_tpi_formula.html)). In fact, we also demonstrated in Chapter 2 that once economic weights (i.e. revenues and/or costs) of DMI, MBW, and MILKE are known, it is fairly straightforward to determine the relative economic weightings of rFE to MILKE for example.

Given that Tempelman et al. (2015) discovered that the partial relationships relating DMI to MILKE and MBW are heterogeneous across research stations in RFI modeling, we extended the MT model in Chapter 3 using hierarchical Bayesian analyses to allow for multifactorial modeling of these partial relationships at both the genetic and residual level as a function of

various management and environmental factors, including station, parity, DIM, and ration. Data on all components of feed efficiency (i.e. DMI, MILKE, MBW, and  $\square$ BW) were based on first 28 days after 50 DIM for each cow within the USDA-NIFA study. Results clearly indicated that genetic and residual variation in rFE is highly heterogeneous across stations, parity, or diet such that the heritability for rFE ranged widely from 0.16 to 0.46. We also found that heritability for rFE in primiparous cows was nearly twice of that in multiparous cows. Although there was some evidence of heterogeneity across stations for partial relationships between DMI and MILKE and DMI and MBW at the residual level, there was no such evidence at the genetic level. At any rate, modeling these different layers of heterogeneity contributed to a 4% higher prediction accuracy on DMI compared to the standard specification from Chapter 2.

We then investigated a comparison of the MT versus RFI approaches for genome wide association (GWA) analyses in Chapter 4. For both approaches, we adopted the single-step genomic BLUP approaches developed by Aguilar et al. (2010) using the Illumina® BovineSNP50 Genotyping BeadChip panel and 28-day records from the USDA-NIFA FE dataset. Inferences were conducted based on single SNP tests and 1 MB window-based joint tests on multiple SNP markers, the latter used to mitigate the impact of multicollinearity due to linkage disequilibrium. For either modeling approach, no single SNP associations with rFE were detected whereas the window-based approach led to the detection of 2 genomic regions within BTA12 and BTA26 significantly associated with rFE. Nevertheless, the MT approach provided important insights that GWA inferences on DMI and on rFE are not synonymous as we provided evidence that the two most significant GWA associations for DMI were partially driven by mild pleiotropic associations with MILKE and MBW whereas GWA inferences on rFE are completely different from the three other traits.

It is possible that any number of environmental covariates could be drivers for genotype by environment (GxE) interactions for economically important dairy traits such as rFE. These sorts of GxE inferences are typically based on reaction norm (RN) models whereby the allelic substitution effect for a particular SNP marker is modeled to be a linear function of a covariate (or polynomial or interaction thereof). Hence, it might be useful to determine specifically which SNP markers and/or regions are associated with these particular GxE covariate drivers so that we can identify genomic regions that are environmentally sensitive. We adapted the model proposed by Jarquin et al. (2014) which allows a large number of such covariate driven RN terms to be fitted simultaneously based on some simplifying assumptions; namely that the RN variance is the same for each covariate and that there is no covariance between the overall average SNP effect and its RN.

After validating GWA and various levels of environmental sensitivity inferences from this model using a simple simulation study, we applied it to the North American subset of the USDA-NIFA consortium dataset. Environmental covariates considered were various terms involving temperature (TP), relative humidity (RH), and average contemporary group milk production (AP). In the GWA studies on GxE effect, 4 regions were found that were environmentally sensitive to various polynomials in TP or AP. The importance of modeling GxE as a function of RN on these covariates was further established as the GxE RN model yielded a 5% higher cross validation prediction accuracy on DMI compared to the conventional GBLUP model.

## **6.2 Implications of Dissertation for Genetic Improvement and Control of Feed Efficiency**

We confirmed that rFE is a heritable trait such that it is possible to improve it by genetic selection. Furthermore, our proposed MT model offers an innovative way to incorporate rFE in the dairy breeding plan. We proposed to simultaneously select for DMI, MBW, and MILKE with

recognition that multiple-trait selection on DMI, MBW, and MILKE is the same as multiple-trait selection on rFE, MBW, and MILKE (Kennedy et al., 1993).

We subsequently determined that the nature of the genetic and non-genetic relationships among the component traits of rFE is highly heterogeneous and subject to different environmental and management conditions. Furthermore, within a GxE context, we again confirmed the importance of herd management impacting allelic substitution effects, and hence heterogeneity of genetic variance. The importance of management practices on influencing genetic variability, and hence estimated genetic merit, for various dairy production traits has been made in other studies as well (Bello et al., 2012, Martin-Collado et al., 2015).

Our GWA results suggested that rFE is likely controlled by many genes with small effects making it rather challenging to target specific genomic areas for special attention. Although we identified two regions having significant associations with rFE based on the window-based test, it appeared that those two regions only contribute to a very small proportion of the total genetic variance (1.1%). Thus, I personally believe it may not be beneficial to conduct further functional experiments to locate candidate genes within these two regions. We also noted that GxE and management driven heterogeneities play an important role in genetic regulation of rFE, thereby adding to the genetic complexity of this trait.

### **6.3 Future Work**

This dissertation has addressed several important issues with various integrated chapters or studies although it might be apparent that there may be a need to jointly infer some of the various phenomena considered in this dissertation. For example, the heterogeneous variance and partial regression modeling at the genetic and residual levels in Chapter 3 should not be considered really distinctly different from the GxE modeling considered in Chapter 5 since GxE generates

heterogeneous genetic variances across environmental covariates. However, the fact that both models are already highly parameterized may require some judicious approaches to model development; for example, perhaps heterogeneous genetic variance and partial relationship modeling might be specified only at the level of the overall (i.e. random intercept) genetic effects in the RN based GxE inferences. Furthermore, we note that others (Aggrey and Rekaya, 2013, Saatchi et al., 2014a) have additionally sought to determine whether or not there is genetic variability in the partial relationships of DMI with MILKE and with MBW. Again, it may be difficult to accommodate many more extensions without increasing dataset sizes.

Except for Chapter 3, the bulk of this work has been based on the implementation of classical mixed models theory; i.e., BLUP based on REML (Chapters 2, 4, and 5). We recognize that Bayesian regression methods, especially those based on heavy-tailed or variable selection specifications (de los Campos et al., 2013) may have advantages over the BLUP assumption of normally independently and identically distributed SNP effects. Implementation of these procedures using Markov Chain Monte Carlo methods can be computationally onerous and plagued by numerical instabilities; e.g. poor mixing, especially in high dimensional models (Tempelman, 2015). However, if conducted conscientiously, these Bayesian regression procedures can augment prediction accuracies as well as accuracies of GWA inferences (Moser et al., 2015). For example, although Bayesian regression strategies have already been proposed for single covariate RN models (Yang et al., 2015), they conceptually could be extended further to select covariates for higher dimensional GxE analyses as we conducted in Chapter 5. Bayesian inference may be particularly useful to extend the distributional assumptions of the Jarquin's model by allowing RN specific variances by borrowing information across covariates with a prior distribution.

It is also important to recognize that the concept of an RFI-like rFE measure as I have investigated in this dissertation is not necessarily synonymous with overall feed efficiency or economic efficiency. Selection index for dairy breeding should include component traits in order to favor the dilution of maintenance and decrease feed intake. In an addition, fixed overhead costs (e.g. stall or pen space) are not adequately represented when penalizing high producing cows that have a much higher DMI consumption than anticipated (i.e. highly positive RFI or smaller feed saved). With increasingly better ways to quantify all input costs, including feed, for the individual cow, it is necessary to integrate the economic and genetic efficiency of dairy production, and promote novel economic evaluations of rFE that are not just based on minimization of feed costs. Future work on the genetics-economics interface of rFE is then vitally important. This should include more optimal herd management, including grouping strategies, to maximize milk output relative to feed and other input costs of which the genetic potential of RFI-like measures is only a small component.

## **APPENDICES**

## APPENDIX A: Chapter 1

### Supplementary Figures and Tables

Table A.1: True values for each of seven key parameters in 162 run response surface design simulation study.

Run	$h^2_{31,2}$	$\phi_{21}^{(\varepsilon)}$	$\phi_{21}^{(u)}$	$\phi_{31}^{(\varepsilon)}$	$\phi_{31}^{(u)}$	$\phi_{32}^{(u)}$	$\phi_{32}^{(\varepsilon)}$
1	0.1	0.1	0.1	0.2	0.2	0.05	0.05
2	0.1	0.1	0.1	0.2	0.2	0.05	0.2
3	0.1	0.1	0.1	0.2	0.2	0.2	0.05
4	0.1	0.1	0.1	0.2	0.2	0.2	0.2
5	0.1	0.1	0.1	0.2	0.6	0.05	0.05
6	0.1	0.1	0.1	0.2	0.6	0.05	0.2
7	0.1	0.1	0.1	0.2	0.6	0.2	0.05
8	0.1	0.1	0.1	0.2	0.6	0.2	0.2
9	0.1	0.1	0.1	0.6	0.2	0.05	0.05
10	0.1	0.1	0.1	0.6	0.2	0.05	0.2
11	0.1	0.1	0.1	0.6	0.2	0.2	0.05
12	0.1	0.1	0.1	0.6	0.2	0.2	0.2
13	0.1	0.1	0.1	0.6	0.6	0.05	0.05
14	0.1	0.1	0.1	0.6	0.6	0.05	0.2
15	0.1	0.1	0.1	0.6	0.6	0.2	0.05
16	0.1	0.1	0.1	0.6	0.6	0.2	0.2
17	0.1	0.1	0.5	0.2	0.2	0.05	0.05
18	0.1	0.1	0.5	0.2	0.2	0.05	0.2
19	0.1	0.1	0.5	0.2	0.2	0.2	0.05
20	0.1	0.1	0.5	0.2	0.2	0.2	0.2
21	0.1	0.1	0.5	0.2	0.6	0.05	0.05
22	0.1	0.1	0.5	0.2	0.6	0.05	0.2
23	0.1	0.1	0.5	0.2	0.6	0.2	0.05
24	0.1	0.1	0.5	0.2	0.6	0.2	0.2
25	0.1	0.1	0.5	0.6	0.2	0.05	0.05
26	0.1	0.1	0.5	0.6	0.2	0.05	0.2
27	0.1	0.1	0.5	0.6	0.2	0.2	0.05
28	0.1	0.1	0.5	0.6	0.2	0.2	0.2
29	0.1	0.1	0.5	0.6	0.6	0.05	0.05
30	0.1	0.1	0.5	0.6	0.6	0.05	0.2
31	0.1	0.1	0.5	0.6	0.6	0.2	0.05
32	0.1	0.1	0.5	0.6	0.6	0.2	0.2
33	0.1	0.5	0.1	0.2	0.2	0.05	0.05
34	0.1	0.5	0.1	0.2	0.2	0.05	0.2
35	0.1	0.5	0.1	0.2	0.2	0.2	0.05

Table A.1: (cont'd)

36	0.1	0.5	0.1	0.2	0.2	0.2	0.2
37	0.1	0.5	0.1	0.2	0.6	0.05	0.05
38	0.1	0.5	0.1	0.2	0.6	0.05	0.2
39	0.1	0.5	0.1	0.2	0.6	0.2	0.05
40	0.1	0.5	0.1	0.2	0.6	0.2	0.2
41	0.1	0.5	0.1	0.6	0.2	0.05	0.05
42	0.1	0.5	0.1	0.6	0.2	0.05	0.2
43	0.1	0.5	0.1	0.6	0.2	0.2	0.05
44	0.1	0.5	0.1	0.6	0.2	0.2	0.2
45	0.1	0.5	0.1	0.6	0.6	0.05	0.05
46	0.1	0.5	0.1	0.6	0.6	0.05	0.2
47	0.1	0.5	0.1	0.6	0.6	0.2	0.05
48	0.1	0.5	0.1	0.6	0.6	0.2	0.2
49	0.1	0.5	0.5	0.2	0.2	0.05	0.05
50	0.1	0.5	0.5	0.2	0.2	0.05	0.2
51	0.1	0.5	0.5	0.2	0.2	0.2	0.05
52	0.1	0.5	0.5	0.2	0.2	0.2	0.2
53	0.1	0.5	0.5	0.2	0.6	0.05	0.05
54	0.1	0.5	0.5	0.2	0.6	0.05	0.2
55	0.1	0.5	0.5	0.2	0.6	0.2	0.05
56	0.1	0.5	0.5	0.2	0.6	0.2	0.2
57	0.1	0.5	0.5	0.6	0.2	0.05	0.05
58	0.1	0.5	0.5	0.6	0.2	0.05	0.2
59	0.1	0.5	0.5	0.6	0.2	0.2	0.05
60	0.1	0.5	0.5	0.6	0.2	0.2	0.2
61	0.1	0.5	0.5	0.6	0.6	0.05	0.05
62	0.1	0.5	0.5	0.6	0.6	0.05	0.2
63	0.1	0.5	0.5	0.6	0.6	0.2	0.05
64	0.1	0.5	0.5	0.6	0.6	0.2	0.2
65	0.5	0.1	0.1	0.2	0.2	0.05	0.05
66	0.5	0.1	0.1	0.2	0.2	0.05	0.2
67	0.5	0.1	0.1	0.2	0.2	0.2	0.05
68	0.5	0.1	0.1	0.2	0.2	0.2	0.2
69	0.5	0.1	0.1	0.2	0.6	0.05	0.05
70	0.5	0.1	0.1	0.2	0.6	0.05	0.2
71	0.5	0.1	0.1	0.2	0.6	0.2	0.05
72	0.5	0.1	0.1	0.2	0.6	0.2	0.2
73	0.5	0.1	0.1	0.6	0.2	0.05	0.05
74	0.5	0.1	0.1	0.6	0.2	0.05	0.2
75	0.5	0.1	0.1	0.6	0.2	0.2	0.05
76	0.5	0.1	0.1	0.6	0.2	0.2	0.2
77	0.5	0.1	0.1	0.6	0.6	0.05	0.05
78	0.5	0.1	0.1	0.6	0.6	0.05	0.2
79	0.5	0.1	0.1	0.6	0.6	0.2	0.05
80	0.5	0.1	0.1	0.6	0.6	0.2	0.2
81	0.5	0.1	0.5	0.2	0.2	0.05	0.05

Table A.1: (cont'd)

82	0.5	0.1	0.5	0.2	0.2	0.05	0.2
83	0.5	0.1	0.5	0.2	0.2	0.2	0.05
84	0.5	0.1	0.5	0.2	0.2	0.2	0.2
85	0.5	0.1	0.5	0.2	0.6	0.05	0.05
86	0.5	0.1	0.5	0.2	0.6	0.05	0.2
87	0.5	0.1	0.5	0.2	0.6	0.2	0.05
88	0.5	0.1	0.5	0.2	0.6	0.2	0.2
89	0.5	0.1	0.5	0.6	0.2	0.05	0.05
90	0.5	0.1	0.5	0.6	0.2	0.05	0.2
91	0.5	0.1	0.5	0.6	0.2	0.2	0.05
92	0.5	0.1	0.5	0.6	0.2	0.2	0.2
93	0.5	0.1	0.5	0.6	0.6	0.05	0.05
94	0.5	0.1	0.5	0.6	0.6	0.05	0.2
95	0.5	0.1	0.5	0.6	0.6	0.2	0.05
96	0.5	0.1	0.5	0.6	0.6	0.2	0.2
97	0.5	0.5	0.1	0.2	0.2	0.05	0.05
98	0.5	0.5	0.1	0.2	0.2	0.05	0.2
99	0.5	0.5	0.1	0.2	0.2	0.2	0.05
100	0.5	0.5	0.1	0.2	0.2	0.2	0.2
101	0.5	0.5	0.1	0.2	0.6	0.05	0.05
102	0.5	0.5	0.1	0.2	0.6	0.05	0.2
103	0.5	0.5	0.1	0.2	0.6	0.2	0.05
104	0.5	0.5	0.1	0.2	0.6	0.2	0.2
105	0.5	0.5	0.1	0.6	0.2	0.05	0.05
106	0.5	0.5	0.1	0.6	0.2	0.05	0.2
107	0.5	0.5	0.1	0.6	0.2	0.2	0.05
108	0.5	0.5	0.1	0.6	0.2	0.2	0.2
109	0.5	0.5	0.1	0.6	0.6	0.05	0.05
110	0.5	0.5	0.1	0.6	0.6	0.05	0.2
111	0.5	0.5	0.1	0.6	0.6	0.2	0.05
112	0.5	0.5	0.1	0.6	0.6	0.2	0.2
113	0.5	0.5	0.5	0.2	0.2	0.05	0.05
114	0.5	0.5	0.5	0.2	0.2	0.05	0.2
115	0.5	0.5	0.5	0.2	0.2	0.2	0.05
116	0.5	0.5	0.5	0.2	0.2	0.2	0.2
117	0.5	0.5	0.5	0.2	0.6	0.05	0.05
118	0.5	0.5	0.5	0.2	0.6	0.05	0.2
119	0.5	0.5	0.5	0.2	0.6	0.2	0.05
120	0.5	0.5	0.5	0.2	0.6	0.2	0.2
121	0.5	0.5	0.5	0.6	0.2	0.05	0.05
122	0.5	0.5	0.5	0.6	0.2	0.05	0.2
123	0.5	0.5	0.5	0.6	0.2	0.2	0.05
124	0.5	0.5	0.5	0.6	0.2	0.2	0.2
125	0.5	0.5	0.5	0.6	0.6	0.05	0.05
126	0.5	0.5	0.5	0.6	0.6	0.05	0.2
127	0.5	0.5	0.5	0.6	0.6	0.2	0.05

Table A.1: (cont'd)

128	0.5	0.5	0.5	0.6	0.6	0.2	0.2
129	0.973	0.3	0.3	0.4	0.4	0.125	0.125
130	0.3	-0.373	0.3	0.4	0.4	0.125	0.125
131	0.3	0.973	0.3	0.4	0.4	0.125	0.125
132	0.3	0.3	-0.373	0.4	0.4	0.125	0.125
133	0.3	0.3	0.973	0.4	0.4	0.125	0.125
134	0.3	0.3	0.3	-0.273	0.4	0.125	0.125
135	0.3	0.3	0.3	1.073	0.4	0.125	0.125
136	0.3	0.3	0.3	0.4	-0.273	0.125	0.125
137	0.3	0.3	0.3	0.4	1.073	0.125	0.125
138	0.3	0.3	0.3	0.4	0.4	-0.128	0.125
139	0.3	0.3	0.3	0.4	0.4	0.378	0.125
140	0.3	0.3	0.3	0.4	0.4	0.125	-0.128
141	0.3	0.3	0.3	0.4	0.4	0.125	0.378
142	0.3	0.3	0.3	0.4	0.4	0.125	0.125
143	0.3	0.3	0.3	0.4	0.4	0.125	0.125
144	0.3	0.3	0.3	0.4	0.4	0.125	0.125
145	0.3	0.3	0.3	0.4	0.4	0.125	0.125
146	0.3	0.3	0.3	0.4	0.4	0.125	0.125
147	0.3	0.3	0.3	0.4	0.4	0.125	0.125
148	0.3	0.3	0.3	0.4	0.4	0.125	0.125
149	0.3	0.3	0.3	0.4	0.4	0.125	0.125
150	0.3	0.3	0.3	0.4	0.4	0.125	0.125
151	0.3	0.3	0.3	0.4	0.4	0.125	0.125
152	0.3	0.3	0.3	0.4	0.4	0.125	0.125
153	0.3	0.3	0.3	0.4	0.4	0.125	0.125
154	0.3	0.3	0.3	0.4	0.4	0.125	0.125
155	0.3	0.3	0.3	0.4	0.4	0.125	0.125
156	0.3	0.3	0.3	0.4	0.4	0.125	0.125
157	0.3	0.3	0.3	0.4	0.4	0.125	0.125
158	0.3	0.3	0.3	0.4	0.4	0.125	0.125
159	0.3	0.3	0.3	0.4	0.4	0.125	0.125
160	0.3	0.3	0.3	0.4	0.4	0.125	0.125
161	0.3	0.3	0.3	0.4	0.4	0.125	0.125
162	0.3	0.3	0.3	0.4	0.4	0.125	0.125

## APPENDIX B: Chapter 2

### Derivation of Full Conditional Densities for MCMC implementation

Much of the following developments closely follow the Appendix in Bello et al. (2010) noting that their developments were based on a 2 trait rather than a 3 trait system.

#### FCD of $\beta$ and $u$ :

We write the entire dataset on all subjects as  $\mathbf{y} = [\mathbf{y}_1' \ \mathbf{y}_2' \ \cdots \ \mathbf{y}_n']'$ . Furthermore, write the overall fixed and random effects design matrices as  $\mathbf{X}_T = [\mathbf{X}_1' \ \mathbf{X}_2' \ \cdots \ \mathbf{X}_n']'$  and  $\mathbf{Z}_T^* = [\mathbf{Z}_1^{*'} \ \mathbf{Z}_2^{*'} \ \cdots \ \mathbf{Z}_n^{*'}]'$ , respectively. We also define  $\Sigma_e = \text{var}(\mathbf{e}) = \left( \bigoplus_{i=1}^n \mathbf{R}_i \right)$ , noting that  $\mathbf{e} = [\mathbf{e}_1' \ \mathbf{e}_2' \ \mathbf{e}_3' \ \dots \ \mathbf{e}_n']'$  is ordered by traits within animals and  $\bigoplus$  denotes the direct sum operator (Searle, 1982). Hence, it should then be readily noted that  $\Sigma_e^{-1} = \bigoplus_{i=1}^n \mathbf{R}_i^{-1}$ . We similarly define  $\text{var}\left(\begin{bmatrix} \mathbf{u}_1^* & \mathbf{u}_2^* & \dots & \mathbf{u}_n^* \end{bmatrix}\right) = \bigoplus_{i=1}^n \mathbf{G}_i$  with inverse  $\bigoplus_{i=1}^n \mathbf{G}_i^{-1}$  noting that  $\Sigma_g^{-1} = \left( \text{var}\left(\begin{bmatrix} \mathbf{u}_{.1}^* & \mathbf{u}_{.2}^* & \mathbf{u}_{.3}^* \end{bmatrix}\right) \right)^{-1}$  can be readily determined by merely rearranging elements of  $\bigoplus_{i=1}^n \mathbf{G}_i^{-1}$  by animals within traits rather than by traits within animals.

One might specify subjective multivariate normal prior densities on fixed location parameters for each trait:  $\beta_j | \beta_j^0, \mathbf{V}_j^{(\beta)} \sim N\left(\beta_j^0, \mathbf{V}_j^{(\beta)}\right)$ , with hyperparameters  $\beta_j^0$  and  $\mathbf{V}_j^{(\beta)}$  being specified as known. Bounded uniform priors are also commonly considered, i.e.,  $\mathbf{V}_j^{(\beta)-1} = \mathbf{0}$

(Sorensen and Gianola, 2002) as, typically, enough data is available to infer upon elements of  $\beta_i$  with any reasonable noninformative prior distribution in large field studies (Gelman, 2006).

Using (Sorensen and Gianola, 2002) that the joint FCD of  $\theta = [\beta' \quad \mathbf{u}^*]'$  is multivariate Gaussian:

$$\theta \sim N \left( \left( \mathbf{W}' \Sigma_e^{-1} \mathbf{W} + \Sigma_\theta^{-1} \right)^{-1} \left( \mathbf{W}' \Sigma_e^{-1} \mathbf{y} + \Sigma_\theta^{-1} \theta_0 \right), \left( \mathbf{W}' \Sigma_e^{-1} \mathbf{W} + \Sigma_\theta^{-1} \right)^{-1} \right) \quad [\text{B1}]$$

for  $\mathbf{W} = [\mathbf{X}_T \quad \mathbf{Z}_T^*]$ ,  $\Sigma_\theta = \text{diag}(\mathbf{V}_1^{(\beta)} \quad \mathbf{V}_2^{(\beta)} \quad \mathbf{V}_3^{(\beta)} \quad \Sigma_g)$ , and  $\theta_0 = [\beta_1^0 \quad \beta_2^0 \quad \beta_3^0 \quad \mathbf{0}_{3q \times 1}]'$ . There are a number of different alternative strategies for sampling from elements of  $\theta$ , including single site or univariate Gibbs updates and block sampling strategies that exploit the sparsity (i.e., high frequency of zero elements) in  $\left( \mathbf{W}' \Sigma_e^{-1} \mathbf{W} + \Sigma_\theta^{-1} \right)^{-1}$  as we invoked in this study.

Note then that draws of breeding values for the three traits,  $\mathbf{u}_{.1}$ ,  $\mathbf{u}_{.2}$ , and  $\mathbf{u}_{.3}$  can readily be determined as  $\mathbf{C}^{-1} \mathbf{u}_{.1}^*$ ,  $\mathbf{C}^{-1} \mathbf{u}_{.2}^*$  and  $\mathbf{C}^{-1} \mathbf{u}_{.3}^*$ , respectively, for  $\mathbf{A} = \mathbf{C}\mathbf{C}'$ , i.e.  $\mathbf{C}$  being the Cholesky decomposition of the relationship matrix  $\mathbf{A}$ . Furthermore, draws of  $\mathbf{u}_{2|1}^*$  can then simply be determined as a vector with elements  $\left\{ u_{i2}^* - \phi_{(21)i}^u u_{i1}^* \right\}_{i=1}^n$  whereas draws of  $\mathbf{e}_{2|1}$  can be determined as a vector with elements  $\left\{ e_{i2} - \phi_{(21)i}^e e_{i1} \right\}_{i=1}^n$  noting that  $e_{i1} = y_{i1} - \mathbf{x}_{i1}' \beta_1 - \mathbf{z}_{i1}' \mathbf{u}_1^*$ . Similarly, draws of  $\mathbf{u}_{3|1,2}^*$  can be determined as  $\left\{ u_{i3}^* - \phi_{(31)i}^u u_{i1}^* - \phi_{(32)i}^u u_{i2}^* \right\}_{i=1}^n$  whereas draws of  $\mathbf{e}_{3|1,2}$  are determined as  $\left\{ e_{i3} - \phi_{(31)i}^e e_{i1} - \phi_{(32)i}^e e_{i2} \right\}_{i=1}^n$ .

**FCD of  $\gamma_{(21)}^u$ ,  $\mathbf{m}_{(21)}^u$ ,  $\gamma_{(21)}^e$  and  $\mathbf{m}_{(21)}^e$ .**

Note from Equations [2.2], [2.3b] and [2.4b] that one could rewrite the model for Trait # 2 (i.e. MBW) as follows:

$$y_{i2} = \mathbf{x}_i' \boldsymbol{\beta}_2 + \mathbf{z}_i^* \mathbf{u}_2^* + e_{i2} = \mathbf{x}_i' \boldsymbol{\beta}_2 + \mathbf{z}_i^* \mathbf{U}_1^* \boldsymbol{\phi}_{(21)}'' + \mathbf{z}_i^* \mathbf{u}_{2||}^* + e_{i1} \phi_{(21)i}^e + e_{i2||} \quad [\text{B2}]$$

where  $\boldsymbol{\phi}_{(21)}'' = \left\{ \phi_{(21)i}'' \right\}_{i=1}^n$  and  $\mathbf{U}_1^* = \text{diag}(\mathbf{u}_1^*)$  and all other terms defined as before. We rearrange

Equation [B2] to be a linear model in  $\boldsymbol{\phi}_{(21)}''$  and hence of  $\boldsymbol{\gamma}_{(21)}''$  and  $\mathbf{m}_{(21)}''$  using Equation [2.5]:

$$\begin{aligned} y_{i21}'' &= y_{i2} - \mathbf{x}_i' \boldsymbol{\beta}_2 - \mathbf{z}_i^* \mathbf{u}_{2||}^* - e_{i1} \phi_{(21)i}^e = \mathbf{z}_i^* \mathbf{U}_1^* \boldsymbol{\phi}_{(21)}'' + e_{i2||} \\ &= \mathbf{z}_i^* \mathbf{U}_1^* \mathbf{X}_{(21)}'' \boldsymbol{\gamma}_{(21)}'' + \mathbf{z}_i^* \mathbf{U}_1^* \mathbf{Z}_{(21)}'' \mathbf{m}_{(21)}'' + e_{i2||} \end{aligned} \quad [\text{B3}]$$

It can then be readily demonstrated using Equation [B5] that the joint FCD of  $\boldsymbol{\gamma}_{(21)}''$  and  $\mathbf{m}_{(21)}''$  is

multivariate normal just like Equation [B1] except one substitutes  $\begin{bmatrix} \mathbf{Z}^* \mathbf{U}_1^* \mathbf{X}_{(21)}'' & \mathbf{Z}^* \mathbf{U}_1^* \mathbf{Z}_{(21)}'' \end{bmatrix}$  for  $\mathbf{W}$ ,

$\text{diag}(\sigma_{e_{(21)i}}^2)$  for  $\boldsymbol{\Sigma}_e$ ,  $\text{diag}(\mathbf{V}_{\boldsymbol{\gamma}_{(21)}}'', \mathbf{I} \sigma_{m_u(21)}^2)$  for  $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ ,  $\left( \boldsymbol{\mu}_{\boldsymbol{\gamma}_{(21)}}^{(u)} \quad \mathbf{0}_{q_u \times 1} \right)'$  for  $\boldsymbol{\theta}_0$ , and  $\left\{ y_{i21}'' \right\}_{i=1}^n$  for  $\mathbf{y}$

noting that  $\mathbf{Z}^* = \begin{bmatrix} \mathbf{z}_1^* & \mathbf{z}_2^* & \cdots & \mathbf{z}_n^* \end{bmatrix}'$ ,  $\mathbf{X}_{(21)}'' = \begin{bmatrix} \mathbf{x}_{(21)1}'' & \mathbf{x}_{(21)2}'' & \cdots & \mathbf{x}_{(21)n}'' \end{bmatrix}'$  and

$\mathbf{Z}_{(21)}'' = \begin{bmatrix} \mathbf{z}_{(21)1}'' & \mathbf{z}_{(21)2}'' & \cdots & \mathbf{z}_{(21)n}'' \end{bmatrix}'$ . Similarly, one could rearrange [B2] to be a linear model in

$\boldsymbol{\phi}_{(21)}^e = \left\{ \phi_{(21)i}^e \right\}_{i=1}^n$  and hence of  $\boldsymbol{\gamma}_{(21)}^e$  and  $\mathbf{m}_{(21)}^e$  using Equation [2.6]:

$$\begin{aligned} y_{i21}^e &= y_{i2} - \mathbf{x}_i' \boldsymbol{\beta}_2 - \mathbf{z}_i^* \mathbf{U}_1^* \boldsymbol{\phi}_{(21)}'' - \mathbf{z}_i^* \mathbf{u}_{2||}^* = e_{i1} \phi_{(21)i}^e + e_{i2||} \\ &= e_{i1} \mathbf{x}_{(21)i}^e \boldsymbol{\gamma}_{(21)}^e + e_{i1} \mathbf{z}_{(21)i}^e \mathbf{m}_{(21)}^e + e_{i2||} \end{aligned} \quad [\text{B4}]$$

such that the joint FCD of  $\gamma_{(21)}^e$  and  $\mathbf{m}_{(21)}^e$  is multivariate normal just like in [B1] except one

substitutes  $\begin{bmatrix} \mathbf{E}_1 \mathbf{X}_{(21)}^e & \mathbf{E}_1 \mathbf{Z}_{(21)}^e \end{bmatrix}'$  for  $\mathbf{W}$ ,  $\text{diag}(\sigma_{e_{(21)i}}^2)$  for  $\Sigma_e$ ,  $\text{diag}(\mathbf{V}_{\gamma_{(21)}}^e, \mathbf{I}\sigma_{m_u(21)}^2)$  for  $\Sigma_\theta$ ,

$(\mu_{\gamma_{(21)}}^{(e)} \quad \mathbf{0}_{q_u \times 1})'$  for  $\theta_0$ , and  $\{y_{i21}^e\}_{i=1}^n$  for  $\mathbf{y}$  noting that  $\mathbf{E}_1 = \text{diag}\{\sigma_{e_{i1}}^2\}_{i=1}^n$ ,

$\mathbf{X}_{(21)}^e = \begin{bmatrix} \mathbf{x}_{(21)1}^e & \mathbf{x}_{(21)2}^e & \cdots & \mathbf{x}_{(21)n}^e \end{bmatrix}'$  and  $\mathbf{Z}_{(21)}^e = \begin{bmatrix} \mathbf{z}_{(21)1}^e & \mathbf{z}_{(21)2}^e & \cdots & \mathbf{z}_{(21)n}^e \end{bmatrix}'$ .

**FCD of  $\gamma_{(31)}^u$ ,  $\mathbf{m}_{(31)}^u$ ,  $\gamma_{(31)}^e$ ,  $\mathbf{m}_{(31)}^e$ ,  $\gamma_{(32)}^u$ ,  $\mathbf{m}_{(32)}^u$ ,  $\gamma_{(32)}^e$ , and  $\mathbf{m}_{(32)}^e$**

Note from Equations [2.2], [2.3c] and [2.4c] that one could rewrite the model for Trait #3 (i.e. DMI) as follows:

$$y_{i3} = \mathbf{x}_i' \boldsymbol{\beta}_3 + \mathbf{z}_i^* \mathbf{u}_3^* + e_{i3} = \mathbf{x}_i' \boldsymbol{\beta}_3 + \mathbf{z}_i^* \mathbf{U}_1^* \boldsymbol{\phi}_{(31)}^u + \mathbf{z}_i^* \mathbf{U}_2^* \boldsymbol{\phi}_{(32)}^u + \mathbf{z}_i^* \mathbf{u}_{3||,2}^* + e_{i1} \phi_{(31)i}^e + e_{i2} \phi_{(33)i}^e + e_{i3||,2}$$

[B5]

where  $\boldsymbol{\phi}_{(31)}^u = \{\phi_{(31)i}^u\}_{i=1}^n$ ,  $\boldsymbol{\phi}_{(32)}^u = \{\phi_{(32)i}^u\}_{i=1}^n$ ,  $\mathbf{U}_2^* = \text{diag}(\mathbf{u}_2^*)$ , and all other terms defined as before. We

rearrange [B5] to be a linear model in  $\boldsymbol{\phi}_{(31)}^u$  and hence of  $\gamma_{(31)}^u$  and  $\mathbf{m}_{(31)}^u$  using Equation [2.5]:

$$\begin{aligned} y_{i31}^u &= y_{i3} - \mathbf{x}_i' \boldsymbol{\beta}_3 - \mathbf{z}_i^* \mathbf{U}_2^* \boldsymbol{\phi}_{(32)}^u - \mathbf{z}_i^* \mathbf{u}_{3||,2}^* - e_{i1} \phi_{(31)i}^e - e_{i2} \phi_{(33)i}^e \\ &= \mathbf{z}_i^* \mathbf{U}_1^* \boldsymbol{\phi}_{(31)}^u + e_{i3||,2} \\ &= \mathbf{z}_i^* \mathbf{X}_{(31)}^u \mathbf{U}_1^* \gamma_{(31)}^u + \mathbf{z}_i^* \mathbf{Z}_{(31)}^u \mathbf{U}_1^* \mathbf{m}_{(31)}^u + e_{i3||,2} \end{aligned}$$

[B6]

It can then be readily demonstrated that the joint FCD of  $\gamma_{(31)}^u$  and  $\mathbf{m}_{(31)}^u$  is multivariate normal just

like in [B1] except one substitutes

$$\begin{bmatrix} \mathbf{Z}^* \mathbf{U}_1^* \mathbf{X}_{(31)}^u & \mathbf{Z}^* \mathbf{U}_1^* \mathbf{Z}_{(31)}^u \end{bmatrix} \text{ for } \mathbf{W}, \text{diag}(\sigma_{e_{(3|1,2)i}}^2) \text{ for } \Sigma_e, \text{diag}(\mathbf{V}_{\gamma(31)}^u, \mathbf{I} \sigma_{m_u(31)}^2) \text{ for } \Sigma_\theta, (\boldsymbol{\mu}_{\gamma(31)}^u \quad \mathbf{0}_{q_u \times 1})' \text{ for } \theta_0, \text{ and } \{y_{i31}^u\}_{i=1}^n \text{ for } \mathbf{y}. \text{ Similarly, one could rearrange [B5] to be a linear model in } \boldsymbol{\phi}_{(31)}^e = \{\phi_{(31)i}^e\}_{i=1}^n$$

and hence of  $\gamma_{(31)}^e$  and  $\mathbf{m}_{(31)}^e$  using Equation [2.6]:

$$\begin{aligned} y_{i31}^e &= y_{i3} - \mathbf{x}_i' \boldsymbol{\beta}_3 - \mathbf{z}_i^* \mathbf{U}_1^* \boldsymbol{\phi}_{(31)}^u - \mathbf{z}_i^* \mathbf{U}_2^* \boldsymbol{\phi}_{(32)}^u - \mathbf{z}_i^* \mathbf{u}_{3|1,2}^* - e_{i2} \phi_{(32)i}^e \\ &= e_{i1} \phi_{(31)i}^e + e_{i3|1,2} \\ &= e_{i1} \mathbf{x}_{(31)i}^e \boldsymbol{\gamma}_{(31)}^e + e_{i1} \mathbf{z}_{(31)i}^e \mathbf{m}_{(31)}^e + e_{i3|1,2} \end{aligned} \quad [\text{B7}]$$

such that the joint FCD of  $\gamma_{(31)}^e$  and  $\mathbf{m}_{(31)}^e$  is multivariate normal just like in [B1] except one

substitutes  $\begin{bmatrix} \mathbf{E}_1 \mathbf{X}_{(31)}^e & \mathbf{E}_1 \mathbf{Z}_{(31)}^e \end{bmatrix}'$  for  $\mathbf{W}$ ,  $\text{diag}(\sigma_{e_{(3|1,2)i}}^2)$  for  $\Sigma_e$ ,  $\text{diag}(\mathbf{V}_{\gamma(31)}^e, \mathbf{I} \sigma_{m_e(31)}^2)$  for  $\Sigma_\theta$ ,

$(\boldsymbol{\mu}_{\gamma(31)}^{(e)} \quad \mathbf{0}_{q_e \times 1})'$  for  $\theta_0$ , and  $\{y_{i31}^e\}_{i=1}^n$  for  $\mathbf{y}$  noting that  $\mathbf{X}_{(31)}^e = \begin{bmatrix} \mathbf{x}_{(31)1}^e & \mathbf{x}_{(31)2}^e & \cdots & \mathbf{x}_{(31)n}^e \end{bmatrix}'$  and

$$\mathbf{Z}_{(31)}^e = \begin{bmatrix} \mathbf{z}_{(31)1}^e & \mathbf{z}_{(31)2}^e & \cdots & \mathbf{z}_{(31)n}^e \end{bmatrix}'.$$

We rewrite [B5] to be a linear model in  $\boldsymbol{\phi}_{(32)}^u$  and hence of  $\gamma_{(32)}^u$  and  $\mathbf{m}_{(32)}^u$  using Equation [2.5]:

$$\begin{aligned} y_{i32}^u &= y_{i3} - \mathbf{x}_i' \boldsymbol{\beta}_3 - \mathbf{z}_i^* \mathbf{U}_1^* \boldsymbol{\phi}_{(31)}^u - \mathbf{z}_i^* \mathbf{u}_{3|1,2}^* - e_{i1} \phi_{(31)i}^e - e_{i2} \phi_{(33)i}^e \\ &= \mathbf{z}_i^* \mathbf{U}_2^* \boldsymbol{\phi}_{(32)}^u + e_{i3|1,2} \\ &= \mathbf{z}_i^* \mathbf{X}_{(32)}^u \mathbf{U}_2^* \boldsymbol{\gamma}_{(32)}^u + \mathbf{z}_i^* \mathbf{Z}_{(32)}^u \mathbf{U}_2^* \mathbf{m}_{(32)}^u + e_{i3|1,2} \end{aligned} \quad [\text{B8}]$$

It can then be readily demonstrated that the joint FCD of  $\gamma_{(32)}^u$  and  $\mathbf{m}_{(32)}^u$  is multivariate normal

just like in [B1] except one substitutes

$$\begin{bmatrix} \mathbf{Z}^* \mathbf{U}_2^* \mathbf{X}_{(32)}^u & \mathbf{Z}^* \mathbf{U}_2^* \mathbf{Z}_{(32)}^u \end{bmatrix} \text{ for } \mathbf{W}, \text{diag}\left(\sigma_{e_{(31,2)i}}^2\right) \text{ for } \Sigma_e, \text{diag}\left(\mathbf{V}_{\gamma(32)}^u, \mathbf{I}\sigma_{m_u(32)}^2\right) \text{ for } \Sigma_\theta, \left(\boldsymbol{\mu}_{\gamma(32)}^u \quad \mathbf{0}_{q_u \times 1}\right)'$$

for  $\theta_0$ , and  $\{y_{i32}^u\}_{i=1}^n$  for  $\mathbf{y}$ .

Finally, one could also rearrange [B5] to be a linear model in  $\boldsymbol{\phi}_{(32)}^e = \{\phi_{(32)i}^e\}_{i=1}^n$  and hence of  $\gamma_{(32)}^e$  and

$\mathbf{m}_{(32)}^e$  using Equation [2.6]:

$$\begin{aligned} y_{i32}^e &= y_{i3} - \mathbf{x}_i' \boldsymbol{\beta}_3 - \mathbf{z}_i^* \mathbf{U}_1^* \boldsymbol{\phi}_{(31)}^u - \mathbf{z}_i^* \mathbf{U}_2^* \boldsymbol{\phi}_{(32)}^u - \mathbf{z}_i^* \mathbf{u}_{31,2}^* - e_{i1} \phi_{(31)i}^e \\ &= e_{i2} \phi_{(32)i}^e + e_{i31,2} \\ &= e_{i2} \mathbf{x}_{(32)i}^e \gamma_{(32)}^e + e_{i2} \mathbf{z}_{(32)i}^e \mathbf{m}_{(32)}^e + e_{i31,2} \end{aligned} \quad [\text{B9}]$$

such that the joint FCD of  $\gamma_{(32)}^e$  and  $\mathbf{m}_{(32)}^e$  is multivariate normal just like in [B1] except one

substitutes  $\begin{bmatrix} \mathbf{E}_2 \mathbf{X}_{(32)}^e & \mathbf{E}_2 \mathbf{Z}_{(32)}^e \end{bmatrix}'$  for  $\mathbf{W}$ ,  $\text{diag}\left(\sigma_{e_{(31,2)i}}^2\right)$  for  $\Sigma_e$ ,  $\mathbf{E}_2 = \text{diag}\left\{\sigma_{e_{i2}}^2\right\}_{i=1}^n \text{diag}\left(\mathbf{V}_{\gamma(32)}^e, \mathbf{I}\sigma_{m_e(32)}^2\right)$

for  $\Sigma_\theta$ ,  $\left(\boldsymbol{\mu}_{\gamma(32)}^{(e)} \quad \mathbf{0}_{q_e \times 1}\right)'$  for  $\theta_0$ , and  $\{y_{i32}^e\}_{i=1}^n$  for  $\mathbf{y}$  noting that  $\mathbf{X}_{(32)}^e = \begin{bmatrix} \mathbf{x}_{(32)1}^e & \mathbf{x}_{(32)2}^e & \cdots & \mathbf{x}_{(32)n}^e \end{bmatrix}'$ ,

and  $\mathbf{Z}_{(32)}^e = \begin{bmatrix} \mathbf{z}_{(32)1}^e & \mathbf{z}_{(32)2}^e & \cdots & \mathbf{z}_{(32)n}^e \end{bmatrix}'$ .

### FCD of $\sigma_{m_u(r)}^2$ for $r=21,31,32$

The FCD for  $\sigma_{m_u(r)}^2$  for  $r=21,31,32$  can be determined as follows:

$$\begin{aligned} p\left(\sigma_{m_u(r)}^2 | \text{ELSE}, \mathbf{y}\right) &\propto p\left(\mathbf{m}_{(r)}^u | \sigma_{m_u(r)}^2\right) * p\left(\sigma_{m_u(r)}^2\right) = \\ &\propto \left(\sigma_{m_u(r)}^2\right)^{-\frac{q_u}{2}} \exp\left(\frac{\mathbf{m}_{(r)}^u \mathbf{m}_{(r)}^u}{-2\sigma_{m_u(r)}^2}\right) p\left(\sigma_{m_u(r)}^2\right) \end{aligned}$$

Note that if  $p\left(\sigma_{m_u(r)}^2\right) \propto 1$  as adapted in this paper, then the FCD of  $\sigma_{m_u(r)}^2$  is Scaled Inverted Chi-

Square with parameters  $q_u - 2$  and  $\frac{\mathbf{m}_{(r)}^u \cdot \mathbf{m}_{(r)}^u}{q_u - 2}$ . If  $p\left(\sigma_{m_u(r)}^2\right)$  is a proper scaled inverted chi-square density, then it could be readily shown that  $p\left(\sigma_{m_u(r)}^2 | ELSE, \mathbf{y}\right)$  is scaled inverted chi-square distributed as well.

### FCD of $\sigma_{m_e(r)}^2$ where $r=21,31,32$

The FCD for  $\sigma_{m_e(r)}^2$  for  $r=21,31,32$  can be determined as follows:

$$\begin{aligned} p\left(\sigma_{m_e(r)}^2 | ELSE, \mathbf{y}\right) &\propto p\left(\mathbf{m}_{(r)}^e | \sigma_{m_e(r)}^2\right) p\left(\sigma_{m_e(r)}^2\right) \\ &\propto \left(\sigma_{m_e(r)}^2\right)^{-\frac{q_e}{2}} \exp\left(\frac{\mathbf{m}_{(r)}^e \cdot \mathbf{m}_{(r)}^e}{-2\sigma_{m_e(r)}^2}\right) \end{aligned}$$

Note that if  $p\left(\sigma_{m_e(r)}^2\right) \propto 1$  as adapted in this paper, then the FCD of  $\sigma_{m_e(r)}^2$  is Scaled Inverted Chi-

Square with parameters  $q_e - 2$  and  $\frac{\mathbf{m}_{(r)}^e \cdot \mathbf{m}_{(r)}^e}{q_e - 2}$ . If  $p\left(\sigma_{m_e(r)}^2\right)$  is a proper scaled inverted chi-square density, then it could be readily shown that  $p\left(\sigma_{m_e(r)}^2 | ELSE, \mathbf{y}\right)$  is scaled inverted chi-square density as well.

### FCD of $\tau_{(s)}^u$ and $\mathbf{v}_{(s)}^u$ for $s = 1, 2|1$ , and $3|1,2$ .

Equation [2.7] can be rewritten as  $\sigma_{u(s)i}^2 = \prod_{c=1}^{P_{u(s)}} \left(\tau_{(s)c}^u\right)^{X_{(s)i,c}^u} \prod_{d=1}^{q_{u(s)}} \left(v_{(s)d}^u\right)^{Z_{(s)i,d}^u}$  noting that

$$\mathbf{x}_{(s)i}^u = \begin{bmatrix} x_{(s)i,1}^u & x_{(s)i,2}^u & x_{(s)i,P_{u(s)}}^u \end{bmatrix} \text{ and } \mathbf{z}_{(s)i}^u = \begin{bmatrix} z_{(s)i,1}^u & z_{(s)i,2}^u & z_{(s)i,q_{u(s)}}^u \end{bmatrix}. \text{ Given an "random}$$

effects” prior of  $p(v_{(s)d}^u | \eta_{(s)}^u) \sim \text{IG}(\eta_{(s)}^u, \eta_{(s)}^u - 1) \quad \forall d$  and noting that elements of  $\mathbf{z}_{(s)i}^u$  are typically

either 0 or 1, then the FCD of can be written as:

$$\begin{aligned}
 p(v_{(s)d}^u | ELSE, \mathbf{y}) &\propto \left( \prod_{i=1}^n \left( \prod_{c=1}^{P_{u(s)}} (\tau_{(s)c}^u)^{x_{(s)i,c}^u} \prod_{d=1}^{Q_{u(s)}} (v_{(s)d}^u)^{z_{(s)i,d}^u} \right)^{\frac{1}{2}} \right) \\
 &\exp \left( -\frac{1}{2v_{(s)d}^u} \sum_{i=1}^n I(z_{(s)i,d}^u = 1) \frac{(u_{i(s)}^*)^2}{\left( \prod_{c=1}^{P_{u(s)}} (\tau_{(s)c}^u)^{x_{(s)i,c}^u} \prod_{d'=1, d' \neq d}^{Q_{u(s)}} (v_{(s)d'}^u)^{z_{(s)i,d'}^u} \right)} \right) v_{(s)d}^u^{-(\eta_{(s)}^u + 1)} \exp \left( -\frac{\eta_{(s)}^u - 1}{v_{(s)d}^u} \right) \\
 &\propto v_{(s)d}^u \left[ -\left( \frac{n_{(s)d}^u}{2} + \eta_{(s)}^u + 1 \right) \right] \exp \left( -\frac{1}{2v_{(s)d}^u} \sum_{i=1}^n I(z_{(s)i,d}^u = 1) \frac{(u_{i(s)}^*)^2}{\left( \prod_{c=1}^{P_{u(s)}} (\tau_{(s)c}^u)^{x_{(s)i,c}^u} \prod_{d'=1, d' \neq d}^{Q_{u(s)}} (v_{(s)d'}^u)^{z_{(s)i,d'}^u} \right)} - \frac{\eta_{(s)}^u - 1}{v_{(s)d}^u} \right)
 \end{aligned}$$

In other words, the FCD of  $v_{(s)d}^u$  can be demonstrated to be Inverse Gamma with parameters

$$\left( \frac{n_{(s)d}^u}{2} + \eta_{(s)}^u \right) \text{ and } \frac{1}{2} \sum_{i=1}^n I(z_{(s)i,d}^u = 1) \frac{(u_{i(s)}^*)^2}{\left( \prod_{c=1}^{P_{u(s)}} (\tau_{(s)c}^u)^{x_{(s)i,c}^u} \prod_{d'=1, d' \neq d}^{Q_{u(s)}} (v_{(s)d'}^u)^{z_{(s)i,d'}^u} \right)} + \eta_{(s)}^u - 1. \text{ Note that } I(.) \text{ refers}$$

to the indicator function which is equal to 1 if the condition defined therein is true; otherwise  $I(.) =$

0; furthermore,  $n_{(s)d}^u = \sum_{i=1}^n I(z_{(s)i,d}^u = 1)$ ; i.e., the number of animals that are directly connected to

$v_{(s)d}^u$ . Using similar developments, it can be also demonstrated that the FCD for  $\tau_{(s)c}^u$  with a flat

prior ( $p(\tau_{(s)c}^u) \propto 1$ ) is also inverse Gamma distributed with parameters:  $\frac{n_{(s)c}^u}{2} - 1$  and

$$\frac{1}{2} \sum_{i=1}^n I(x_{(s),i,c}^u = 1) \frac{(u_{i(s)}^*)^2}{\left( \prod_{c'=1, c' \neq c}^{p_{u(s)}} (\tau_{(s)c'}^u)^{x_{(s),i,c}^u} \prod_{d=1}^{q_{u(s)}} (v_{(s)d}^u)^{z_{(s),i,d}^u} \right)} \text{ where } n_{(s)c}^u = \sum_{i=1}^n I(x_{(s),i,c}^u = 1); \text{ i.e., the number of}$$

animals that are directly connected to  $\tau_{(s)c}^u$ .

**FCD for  $\tau_{(s)}^e$  and  $v_{(s)}^e$  for  $s = 1, 2|1$ , and  $3|1, 2$ .**

Equation [8] can be rewritten as  $\sigma_{e(s)i}^2 = \prod_{c=1}^{p_{e(s)}} (\tau_{(s)c}^e)^{x_{(s),i,c}^e} \prod_{d=1}^{q_{e(s)}} (v_{(s)d}^e)^{z_{(s),i,d}^e}$  noting that

$$\mathbf{x}_{(s)i}^e = \begin{bmatrix} x_{(s),i,1}^e & x_{(s),i,2}^e & x_{(s),i,p_{e(s)}}^e \end{bmatrix} \text{ and } \mathbf{z}_{(s)i}^e = \begin{bmatrix} z_{(s),i,1}^e & z_{(s),i,2}^e & z_{(s),i,q_{e(s)}}^e \end{bmatrix}. \text{ Given a "random effects"}$$

prior of  $p(v_{(e)d}^u | \eta_{(e)}^u) \sim \text{IG}(\eta_{(e)}^u, \eta_{(e)}^u - 1) \forall d$  and noting that elements of  $\mathbf{z}_{(s)i}^e$  are typically either 0

or 1, then the FCD of can be written as:

$$\begin{aligned} p(v_{(s)d}^e | ELSE, \mathbf{y}) &\propto \left( \prod_{i=1}^n \left( \prod_{c=1}^{p_{e(s)}} (\tau_{(s)c}^e)^{x_{(s),i,c}^e} \prod_{d=1}^{q_{e(s)}} (v_{(s)d}^e)^{z_{(s),i,d}^e} \right)^{-\frac{1}{2}} \right) \\ &\exp \left( -\frac{1}{2v_{(s)d}^e} \sum_{i=1}^n I(z_{(s),i,d}^e = 1) \frac{(e_{i(s)})^2}{\left( \prod_{c=1}^{p_{e(s)}} (\tau_{(s)c}^e)^{x_{(s),i,c}^e} \prod_{d'=1, d' \neq d}^{q_{e(s)}} (v_{(s)d'}^e)^{z_{(s),i,d'}^e} \right)} \right) v_{(s)d}^e^{-\eta_{(s)}^e - 1} \exp \left( -\frac{\eta_{(s)}^e - 1}{v_{(s)d}^e} \right) \\ &\propto v_{(s)d}^e^{\left[ \frac{-\eta_{(s)}^e}{2} - (\eta_{(s)}^e + 1) \right]} \exp \left( -\frac{1}{2v_{(s)d}^e} \sum_{i=1}^n I(z_{(s),i,d}^e = 1) \frac{(e_{i(s)})^2}{\left( \prod_{c=1}^{p_{e(s)}} (\tau_{(s)c}^e)^{x_{(s),i,c}^e} \prod_{d'=1, d' \neq d}^{q_{e(s)}} (v_{(s)d'}^e)^{z_{(s),i,d'}^e} \right)} - \frac{\eta_{(s)}^e - 1}{v_{(s)d}^e} \right) \end{aligned}$$

which can be demonstrated to be Inverse Gamma with parameters  $\left( \frac{n_{(s)d}^e}{2} + \eta_{(s)}^e \right)$  and

$$\frac{1}{2} \sum_{i=1}^n I(z_{(s),i,d}^e = 1) \frac{(e_{i(s)})^2}{\left( \prod_{c=1}^{P_{e(s)}} (\tau_{(s)c}^e)^{x_{(s),i,c}^e} \prod_{d'=1, d' \neq d}^{Q_{e(s)}} (v_{(s)d'}^e)^{z_{(s),i,d'}^e} \right)} + \eta_{(s)}^e. \text{ Note that } I(.) \text{ refers to the indicator}$$

function which is equal to 1 if the condition defined therein is true; otherwise  $I(.) = 0$ ; furthermore,

$n_{(s)d}^e = \sum_{i=1}^n I(z_{(s),i,d}^e = 1)$ . Using similar developments, it can be also demonstrated that the FCD for

$\tau_{(s)c}^e$  is also inverse Gamma distributed with parameters  $\frac{n_{(s)c}^e}{2} + \eta_{(s)}^e$  and

$$\frac{1}{2} \sum_{i=1}^n I(x_{(s),i,c}^e = 1) \frac{(e_{i(s)})^2}{\left( \prod_{c'=1, c' \neq c}^{P_{e(s)}} (\tau_{(s)c'}^e)^{x_{(s),i,c'}^e} \prod_{d=1}^{Q_{e(s)}} (v_{(s)d}^e)^{z_{(s),i,d}^e} \right)} + \eta_{(s)}^e$$

### FCD of $\eta_{(s)}^u$

$$\begin{aligned} P(\eta_{(s)}^u | ELSE) &\propto \prod_{d=1}^{P_{u(s)}} p(v_{(s)d}^u | \eta_{(s)}^u) * p(\eta_{(s)}^u) \\ &\propto \frac{(\eta_{(s)}^u - 1)^{\eta_{(s)}^u P_{u(s)}}}{\Gamma(\eta_{(s)}^u)^{P_{u(s)}}} \left( \prod_{d=1}^{P_{u(s)}} v_{(s)d}^u^{-\left(\eta_{(s)}^u + 1\right)} \right) \exp \left( -\left(\eta_{(s)}^u - 1\right) \sum_{d=1}^{P_{u(s)}} \frac{1}{v_{(s)d}^u} \right) (1 + \eta_{(s)}^u)^{-2} \end{aligned}$$

Since this FCD is not recognizable, a Metropolis Hastings step is used here; further details are provided in Bello et al. (2012)

### FCD of $\eta_{(s)}^e$

$$\begin{aligned}
P\left(\eta_{(s)}^e | ELSE\right) &\propto \prod_{d=1}^{P_{e(s)}} p\left(v_{(s)d}^e | \eta_{(s)}^e\right) * p\left(\eta_{(s)}^e\right) \\
&\propto \frac{\left(\eta_{(s)}^e - 1\right)^{\eta_{(s)}^e P_{e(s)}}}{\Gamma\left(\eta_{(s)}^e\right)^{P_{e(s)}}} \left( \prod_{d=1}^{P_{e(s)}} v_{(s)d}^e^{-\left(\eta_{(s)}^e + 1\right)} \right) \exp\left(-\left(\eta_{(s)}^e - 1\right) \sum_{d=1}^{P_{e(s)}} \frac{1}{v_{(s)d}^e}\right) \left(1 + \eta_{(s)}^e\right)^{-2}
\end{aligned}$$

Since this FCD is not recognizeable, a Metropolis Hastings step is also used here

### Supplementary Figures and Tables

Table B.1: Posterior inferences characterizing heterogeneity of genetic and residual partial efficiencies of MBW<sup>1</sup> on MILKE<sup>2</sup>

		PMEAN (PSD) <sup>3</sup>	95% HPD <sup>4</sup>	ESS <sup>5</sup>	PMEAN (PSD) <sup>3</sup>	95% HPD <sup>4</sup>	ESS <sup>5</sup>
Factor	Levels	Genetic Partial Efficiencies ( $\phi_{(21)}^u$ in kg/Mcal)			Residual Partial Efficiencies ( $\phi_{(21)}^e$ in kg/Mcal)		
<i>Stations</i> <sup>6</sup>	<i>AB</i>	0.22(0.22)	[-0.23,0.63]	1417.61	0.62(0.17)	[0.29,0.96]	2027.56
	<i>BARC</i>	0.03(0.18)	[-0.32,0.38]	1485.44	0.56(0.15)	[0.27,0.85]	1327.47
	<i>ISU</i>	0.05(0.20)	[-0.35,0.44]	1370.28	0.48(0.15)	[0.19,0.78]	1720.10
	<i>MSU</i>	0.11(0.20)	[-0.28,0.50]	1678.18	0.41(0.16)	[0.10,0.73]	2840.62
	<i>UF</i>	0.10(0.21)	[-0.31,0.49]	2153.18	0.47(0.17)	[0.13,0.80]	2118.45
	<i>USDFRC</i>	0.02(0.21)	[-0.40,0.42]	1963.80	0.54(0.16)	[0.24,0.86]	1251.10
	<i>UW</i>	0.10(0.21)	[-0.31,0.53]	2292.54	0.40(0.17)	[0.07,0.74]	2194.10
	<i>NBZ</i>	0.25(0.17)	[-0.09,0.58]	994.46	0.51(0.16)	[0.20,0.83]	2464.04
	<i>NLN</i>	0.11(0.20)	[-0.29,0.52]	1757.27	0.47(0.16)	[0.18,0.80]	2166.99
	<i>TGEN</i>	0.17(0.20)	[-0.20,0.57]	1495.81	0.41(0.15)	[0.12,0.70]	1437.23
	<i>ZOM</i>	0.11(0.13)	[-0.16,0.37]	1030.10	0.47(0.09)	[0.29,0.65]	1041.08
	<i>LAN</i>	0.09(0.21)	[-0.32,0.51]	1786.41	0.33(0.14)	[0.05,0.61]	2885.50
	<i>SAC</i>	0.14(0.21)	[-0.26,0.54]	1533.41	0.45(0.13)	[0.19,0.70]	2510.04
<i>DIM</i>	<i>Linear</i>	-4.3e-03(0.02)	[-0.04,0.03]	8905.04	-0.01(0.02)	[-0.06,0.03]	8767.77
	<i>Quadratic</i>	2.8e-03(0.02)	[-0.04,0.04]	8447.56	0.01(0.03)	[-0.05,0.07]	8932.12
<i>Parity</i>	<i>Primiparous</i>	0.09(0.16)	[-0.22,0.42]	797.32	0.50(0.13)	[0.25,0.75]	765.76
	<i>Multiparous</i>	0.14(0.15)	[-0.16,0.43]	1398.08	0.44(0.08)	[0.29,0.61]	2580.82
<i>Ration(Station)</i>	<i>Variance component</i> ( $(\text{kg}^{0.75})^2 / \text{Mcal}^2$ )						
	<i>Genetic</i> ( $\sigma_{m_u(21)}^2$ )	0.09(0.11)	[3.1e-04,0.32]	134.97			
	<i>Residual</i> ( $\sigma_{m_e(21)}^2$ )				0.10(0.06)	[2.5e-03,0.22]	394.00

<sup>1</sup> Metabolic body weight (in Kg)

<sup>2</sup> Milk energy (in Mcal)

<sup>3</sup> Posterior mean ( $\pm$  posterior standard deviation)

<sup>4</sup> 95% Highest posterior density interval

<sup>5</sup> Effective sample size

Table B.1: (cont'd)

<sup>6</sup> Station levels characterized in body of manuscript.

Table B.2: Marginal Mean and Variance Component Inferences of Genetic and Residual Variance Components of MILKE<sup>1</sup>

		PMEAN(PSD) <sup>2</sup>	95%HPD <sup>4</sup>	ESS <sup>5</sup>	PMEAN(PSD) <sup>2,3</sup>	95%HPD <sup>4</sup>	ESS <sup>5</sup>
Factor	Levels	<i>Genetic variance</i> (Mcal <sup>2</sup> )			<i>Residual variance</i> (Mcal <sup>2</sup> )		
<i>Stations</i> <sup>6</sup>	<i>AB</i>	3.13(1.20)	[1.19,5.55]	676.96	5.1 <sup>abc</sup> (1.33)	[2.65,7.77]	1350.53
	<i>AGIL</i>	6.72(2.84)	[2.07,12.40]	868.98	5.81 <sup>abc</sup> (1.96)	[2.51,9.68]	1714.74
	<i>ISU</i>	7.31(2.79)	[2.38,12.92]	342.90	9.54 <sup>ab</sup> (2.46)	[5.06,14.53]	619.60
	<i>MSU</i>	4.14(2.05)	[1.17,8.17]	628.56	9.26 <sup>a</sup> (1.95)	[5.74,13.19]	1673.32
	<i>UF</i>	4.89(2.34)	[1.43,9.69]	647.11	6.99 <sup>abc</sup> (1.62)	[3.88,10.14]	1228.41
	<i>USDFRC</i>	6.26(2.30)	[2.16,10.69]	348.36	7.08 <sup>abc</sup> (1.25)	[4.55,9.75]	406.75
	<i>UW</i>	6.60(3.44)	[1.25,13.37]	357.71	7.70 <sup>abc</sup> (2.07)	[3.75,11.68]	654.00
	<i>NBZ</i>	3.92(1.10)	[1.87,6.03]	162.34	4.06 <sup>c</sup> (1.37)	[1.75,6.79]	1989.52
	<i>NLN</i>	3.33(1.38)	[1.21,6.14]	647.24	5.31 <sup>bc</sup> (0.94)	[3.52,7.19]	1470.02
	<i>TGEN</i>	4.17(1.73)	[1.34,7.55]	527.75	5.47 <sup>abc</sup> (1.79)	[2.43,9.00]	817.54
	<i>ZOM</i>	4.09(0.83)	[2.63,5.77]	146.86	5.50 <sup>bc</sup> (0.64)	[4.26,6.77]	544.24
	<i>LAN</i>	3.64(1.33)	[1.47,6.23]	555.44	7.30 <sup>abc</sup> (1.70)	[4.28,10.68]	1359.89
	<i>SAC</i>	3.34(1.36)	[1.25,5.96]	493.17	9.01 <sup>abc</sup> (2.45)	[4.58,13.71]	1600.81
<i>Parity</i>	<i>Primiparous</i>	4.90(1.00)	[2.95,6.82]	208.45	4.10 <sup>x</sup> (0.55)	[3.05,5.17]	324.21
	<i>Multiparous</i>	3.81(1.00)	[1.95,5.80]	214.50	10.01 <sup>y</sup> (0.96)	[8.09,11.89]	637.15
<i>Ration(Station)</i>	<i>CV</i> <sup>7</sup>						
	<i>Genetic</i>	0.50(0.38)	[0.26,0.92]	296.11			
	<i>Residual</i>				0.41(0.09)	[0.27,0.59]	909.23

<sup>1</sup> Milk energy (in Mcal)

<sup>2</sup> Posterior mean ( $\pm$  posterior standard deviation)

Table B.2: (cont'd)

<sup>3</sup> Estimates not sharing the same letters within factors are statistically different ( $P < 0.05$ ).

<sup>4</sup> 95% Highest posterior density interval

<sup>5</sup> Effective sample size

<sup>6</sup> Station levels characterized in body of paper.

<sup>7</sup> Coefficient of variation of ration within station specific genetic and residual variance components.

Table B.3: Marginal Mean and Variance Component Inferences of Genetic and Residual Variance Components of MBW<sup>1</sup> | MILKE<sup>2</sup>

		PMEAN (PSD) <sup>3</sup>	95% HPD <sup>4</sup>	ESS <sup>5</sup>	PMEAN (PSD) <sup>3,4</sup>	95% HPD <sup>5</sup>	ESS <sup>6</sup>
Factor	Levels	<b>Genetic variance</b> ((Kg <sup>0.75</sup> ) <sup>2</sup> )			<b>Residual variance</b> ((Kg <sup>0.75</sup> ) <sup>2</sup> )		
<i>Stations</i> <sup>7</sup>	<i>AB</i>	26.73(7.92)	[13.18,43.05]	532.48	18.77(6.06)	[8.43,30.96]	926.47
	<i>AGIL</i>	23.13(7.72)	[10.25,38.84]	1046.34	28.68(8.84)	[13.27,46.10]	2023.31
	<i>ISU</i>	34.78(8.79)	[18.37,52.12]	917.23	23.00(5.78)	[12.34,34.28]	953.39
	<i>MSU</i>	47.47(18.48)	[14.47,81.86]	305.45	30.01(12.67)	[10.35,55.56]	286.01
	<i>UF</i>	48.77(16.24)	[18.94,80.02]	566.48	26.96(10.67)	[9.39,48.22]	527.74
	<i>USDFRC</i>	34.63(9.78)	[17.36,54.59]	182.86	25.66(6.67)	[12.97,38.85]	286.84
	<i>UW</i>	38.53(12.99)	[14.07,62.86]	331.67	25.15(9.12)	[9.52,42.94]	394.20
	<i>NBZ</i>	23.02(5.62)	[13.34,34.13]	228.00	19.51(6.67)	[7.92,32.42]	1183.86
	<i>NLN</i>	32.95(12.43)	[11.59,57.01]	371.09	31.35(8.91)	[14.41,48.13]	621.35
	<i>TGEN</i>	24.72(8.53)	[10.16,41.37]	805.66	25.01(8.08)	[10.96,41.06]	766.92
	<i>ZOM</i>	29.31(5.17)	[19.43,39.20]	165.22	23.58(3.19)	[17.61,29.85]	243.28
	<i>LAN</i>	28.08(7.31)	[15.80,43.53]	966.57	24.73(6.07)	[13.61,36.51]	1324.91
	<i>SAC</i>	32.27(10.76)	[13.85,53.20]	1049.48	34.39(9.79)	[16.30,53.26]	1194.97
<i>Parity</i>	<i>Primiparous</i>	27.26(3.96)	[19.46,34.74]	257.33	19.78 <sup>x</sup> (2.85)	[14.29,25.35]	360.83
	<i>Multiparous</i>	34.57(7.01)	[20.74,47.89]	160.59	30.71 <sup>y</sup> (5.26)	[20.65,40.87]	142.90
<i>Ration(Station)</i>	<i>CV</i> <sup>8</sup>						
	<i>Genetic</i>	0.39(0.05)	[0.34,0.48]	108.33			
	<i>Residual</i>				0.32(0.05)	[0.23,0.42]	1204.37

<sup>1</sup> Metabolic body weight (in Kg)<sup>2</sup> Milk energy (in Mcal)<sup>3</sup> Posterior mean ( $\pm$  posterior standard deviation)<sup>4</sup> Estimates not sharing the same letters within factors are statistically different ( $P < 0.05$ ).<sup>5</sup> 95% Highest posterior density interval<sup>6</sup> Effective sample size<sup>7</sup> Station levels characterized in body of paper.<sup>8</sup> Coefficient of variation of ration within station specific genetic and residual variance components.

## APPENDIX C: Chapter 3

### Derivations

Let's start with a simple univariate animal model:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}' \\ \mathbf{X} & \mathbf{I} + (\mathbf{ZZ}')^{-1}\lambda \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{y} \end{bmatrix} \text{ where } \lambda = \frac{\sigma_e^2}{\sigma_u^2} \text{ and } \begin{bmatrix} \mathbf{C}^{\beta\beta} & \mathbf{C}^{\beta u} \\ \mathbf{C}^{u\beta} & \mathbf{C}^{uu} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}' \\ \mathbf{X} & \mathbf{I} + \mathbf{ZZ}'\lambda \end{bmatrix}^{-1}$$

And one may also absorb the fixed effects too:

$$\left[ \mathbf{W} + (\mathbf{ZZ}')^{-1}\lambda \right] [\hat{\mathbf{g}}] = [\mathbf{W}\mathbf{y}] \text{ where } \mathbf{W} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \text{ such that } \mathbf{C}^{uu} = \left[ \mathbf{W} + (\mathbf{ZZ}')^{-1}\lambda \right]^{-1}.$$

$$\begin{aligned} \mathbf{C}^{gg} &= \mathbf{Z}'\mathbf{G}^{-1}\mathbf{C}^{uu}\mathbf{G}^{-1}\mathbf{Z}'\sigma_u^2 \\ \text{Then } \mathbf{C}^{gg} \text{ part can be computed as } &= \mathbf{Z}'\mathbf{G}^{-1} \left[ \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + (\mathbf{ZZ}')^{-1}\lambda \right]^{-1} \mathbf{G}^{-1}\mathbf{Z}'\sigma_u^2 \end{aligned}$$

And we already know that  $\mathbf{g} = \mathbf{Z}'\mathbf{G}^{-1}\mathbf{u}$  and

$$z_j = \frac{\hat{g}_j}{\text{sqrt}\left((\mathbf{Z}'\mathbf{G}^{-1}\mathbf{Z})_j \sigma_u^2 - C_j^{gg}\right)} \text{ where } (\mathbf{Z}'\mathbf{G}\mathbf{Z})_j \text{ and } C_j^{gg} \text{ are the } j,j \text{ diagonal element of } (\mathbf{Z}'\mathbf{G}\mathbf{Z})$$

and  $C_{jj}^{gg}$  is the prediction error variance of  $\hat{g}_j$  being the GBLUP estimate for SNP  $j$ .

$$\text{Consider the MT (multiple-trait) model on MILKE, MBW, and DMI } \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{X}\boldsymbol{\beta}_1 + \mathbf{u}_1 + \mathbf{e}_1 \\ \mathbf{X}\boldsymbol{\beta}_2 + \mathbf{u}_2 + \mathbf{e}_2 \\ \mathbf{X}\boldsymbol{\beta}_3 + \mathbf{u}_3 + \mathbf{e}_3 \end{bmatrix},$$

and the BV estimates and variance genetic and residual covariance matrixes between three traits are

$$\Sigma_g = \text{var} \begin{bmatrix} u_{1i} \\ u_{2i} \\ u_{3i} \end{bmatrix} = \text{var}(\mathbf{u}_i) = \begin{bmatrix} \sigma_{g_1}^2 & \sigma_{g_{12}} & \sigma_{g_{13}} \\ \sigma_{g_{12}} & \sigma_{g_2}^2 & \sigma_{g_{23}} \\ \sigma_{g_{13}} & \sigma_{g_{23}} & \sigma_{g_3}^2 \end{bmatrix}; i = 1, 2, \dots, q \text{ It just happens that}$$

$$\Sigma_g = \text{var} \begin{bmatrix} g_{1i} \\ g_{2i} \\ g_{3i} \end{bmatrix} = \text{var}(\mathbf{u}_i) = \begin{bmatrix} \sigma_{g_1}^2 & \sigma_{g_{12}} & \sigma_{g_{13}} \\ \sigma_{g_{12}} & \sigma_{g_2}^2 & \sigma_{g_{23}} \\ \sigma_{g_{13}} & \sigma_{g_{23}} & \sigma_{g_3}^2 \end{bmatrix}; i = 1, 2, \dots, m. \text{ Furthermore,}$$

$$\mathbf{R} = \text{var} \begin{bmatrix} e_{1j} \\ e_{2j} \\ e_{3j} \end{bmatrix} = \text{var}(\mathbf{e}_j) = \begin{bmatrix} \sigma_{e_1}^2 & \sigma_{e_{12}} & \sigma_{e_{13}} \\ \sigma_{e_{12}} & \sigma_{e_2}^2 & \sigma_{e_{23}} \\ \sigma_{e_{13}} & \sigma_{e_{23}} & \sigma_{e_3}^2 \end{bmatrix}; j = 1, 2, \dots, n.$$

$$\text{Let's write: } \Sigma_g^{-1} = \begin{bmatrix} g^{11} & g^{12} & g^{13} \\ g^{21} & g^{22} & g^{23} \\ g^{31} & g^{23} & g^{33} \end{bmatrix} \text{ and } \Sigma_e^{-1} = \begin{bmatrix} r^{11} & r^{12} & r^{13} \\ r^{21} & r^{22} & r^{23} \\ r^{31} & r^{23} & r^{33} \end{bmatrix}$$

The mixed model equation for MT model is:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X}r^{11} & \mathbf{X}'\mathbf{I}r^{11} & \mathbf{X}'\mathbf{X}r^{12} & \mathbf{X}'\mathbf{I}r^{12} & \mathbf{X}'\mathbf{X}r^{13} & \mathbf{X}'\mathbf{I}r^{13} \\ \mathbf{I}r^{11} & \mathbf{I}r^{11} + \mathbf{H}^{-1}g^{11} & \mathbf{I}r^{12} & \mathbf{I}r^{12} + \mathbf{H}^{-1}g^{12} & \mathbf{I}r^{13} & \mathbf{I}r^{13} + \mathbf{H}^{-1}g^{13} \\ \mathbf{X}'\mathbf{X}r^{12} & \mathbf{X}'\mathbf{I}r^{12} & \mathbf{X}'\mathbf{X}r^{22} & \mathbf{X}'\mathbf{I}r^{22} & \mathbf{X}'\mathbf{X}r^{23} & \mathbf{X}'\mathbf{I}r^{23} \\ \mathbf{I}r^{12} & \mathbf{I}r^{12} + \mathbf{H}^{-1}g^{12} & \mathbf{I}r^{22} & \mathbf{I}r^{22} + \mathbf{H}^{-1}g^{21} & \mathbf{I}r^{23} & \mathbf{I}r^{23} + \mathbf{H}^{-1}g^{23} \\ \mathbf{X}'\mathbf{X}r^{13} & \mathbf{X}'\mathbf{I}r^{13} & \mathbf{X}'\mathbf{X}r^{23} & \mathbf{X}'\mathbf{I}r^{23} & \mathbf{X}'\mathbf{X}r^{33} & \mathbf{X}'\mathbf{I}r^{33} \\ \mathbf{I}r^{13} & \mathbf{I}r^{13} + \mathbf{H}^{-1}g^{13} & \mathbf{I}r^{23} & \mathbf{I}r^{23} + \mathbf{H}^{-1}g^{23} & \mathbf{I}r^{33} & \mathbf{I}r^{33} + \mathbf{H}^{-1}g^{33} \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\mathbf{u}}_1 \\ \hat{\beta}_2 \\ \hat{\mathbf{u}}_2 \\ \hat{\beta}_3 \\ \hat{\mathbf{u}}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y}_1 \\ \mathbf{y}_1 \\ \mathbf{X}'\mathbf{y}_2 \\ \mathbf{y}_2 \\ \mathbf{X}'\mathbf{y}_3 \\ \mathbf{y}_3 \end{bmatrix}$$

such that the inverse of the mixed model equations can be written as follows:

$$\begin{aligned}
& \begin{bmatrix} \mathbf{X}'\mathbf{X}_r^{11} & \mathbf{X}'\mathbf{I}_r^{11} & \mathbf{X}'\mathbf{X}_r^{12} & \mathbf{X}'\mathbf{I}_r^{12} & \mathbf{X}'\mathbf{X}_r^{13} & \mathbf{X}'\mathbf{I}_r^{13} \\ \mathbf{I}_r^{11} + \mathbf{H}^{-1}\mathbf{g}^{11} & \mathbf{I}_r^{12} + \mathbf{H}^{-1}\mathbf{g}^{12} & \mathbf{I}_r^{13} + \mathbf{H}^{-1}\mathbf{g}^{13} \\ \mathbf{X}'\mathbf{X}_r^{12} & \mathbf{X}'\mathbf{I}_r^{12} & \mathbf{X}'\mathbf{X}_r^{22} & \mathbf{X}'\mathbf{I}_r^{22} & \mathbf{X}'\mathbf{X}_r^{23} & \mathbf{X}'\mathbf{I}_r^{23} \\ \mathbf{I}_r^{12} + \mathbf{H}^{-1}\mathbf{g}^{12} & \mathbf{I}_r^{22} + \mathbf{H}^{-1}\mathbf{g}^{21} & \mathbf{I}_r^{23} + \mathbf{H}^{-1}\mathbf{g}^{23} \\ \mathbf{X}'\mathbf{X}_r^{13} & \mathbf{X}'\mathbf{I}_r^{13} & \mathbf{X}'\mathbf{X}_r^{23} & \mathbf{X}'\mathbf{I}_r^{23} & \mathbf{X}'\mathbf{X}_r^{33} & \mathbf{X}'\mathbf{I}_r^{33} \\ \mathbf{I}_r^{13} + \mathbf{H}^{-1}\mathbf{g}^{13} & \mathbf{I}_r^{23} + \mathbf{H}^{-1}\mathbf{g}^{23} & \mathbf{I}_r^{33} + \mathbf{H}^{-1}\mathbf{g}^{33} \end{bmatrix}^{-1} \\
& = \begin{bmatrix} \mathbf{C}^{\beta_1\beta_1} & \mathbf{C}^{\beta_1u_1} & \mathbf{C}^{\beta_1\beta_2} & \mathbf{C}^{\beta_1u_2} & \mathbf{C}^{\beta_1\beta_3} & \mathbf{C}^{\beta_1u_3} \\ \mathbf{C}^{u_1\beta_1} & \mathbf{C}^{u_1u_1} & \mathbf{C}^{u_1\beta_2} & \mathbf{C}^{u_1u_2} & \mathbf{C}^{u_1\beta_3} & \mathbf{C}^{u_1u_3} \\ \mathbf{C}^{\beta_2\beta_1} & \mathbf{C}^{\beta_2u_1} & \mathbf{C}^{\beta_2\beta_2} & \mathbf{C}^{\beta_2u_2} & \mathbf{C}^{\beta_2\beta_3} & \mathbf{C}^{\beta_2u_3} \\ \mathbf{C}^{u_2\beta_1} & \mathbf{C}^{u_2u_1} & \mathbf{C}^{u_2\beta_2} & \mathbf{C}^{u_2u_2} & \mathbf{C}^{u_2\beta_3} & \mathbf{C}^{u_2u_3} \\ \mathbf{C}^{\beta_3\beta_1} & \mathbf{C}^{\beta_3u_1} & \mathbf{C}^{\beta_3\beta_2} & \mathbf{C}^{\beta_3u_2} & \mathbf{C}^{\beta_3\beta_3} & \mathbf{C}^{\beta_3u_3} \\ \mathbf{C}^{u_3\beta_1} & \mathbf{C}^{u_3u_1} & \mathbf{C}^{u_3\beta_2} & \mathbf{C}^{u_3u_2} & \mathbf{C}^{u_3\beta_3} & \mathbf{C}^{u_3u_3} \end{bmatrix}
\end{aligned}$$

Of course, the key part of that inverse pertains to the random effects for three traits which we illustrate below:

$$\mathbf{C}^{uu} = \begin{bmatrix} \mathbf{C}^{u_1u_1} & \mathbf{C}^{u_1u_2} & \mathbf{C}^{u_1u_3} \\ \mathbf{C}^{u_2u_1} & \mathbf{C}^{u_2u_2} & \mathbf{C}^{u_2u_3} \\ \mathbf{C}^{u_3u_1} & \mathbf{C}^{u_3u_2} & \mathbf{C}^{u_3u_3} \end{bmatrix}$$

Suppose one wants to backsolve for solutions of  $\mathbf{g}$  using breeding value estimates from genotyped animals. Refer to these genotypes as  $\mathbf{Z}$  and the corresponding breeding values from the genotyped animals as  $[\mathbf{u}_1^* \ ' \ \mathbf{u}_2^* \ ' \ \mathbf{u}_3^* \ ']$  on the three traits. Then, it can be established that:

$$\begin{aligned}
\begin{bmatrix} \hat{\mathbf{g}}_1 \\ \hat{\mathbf{g}}_2 \\ \hat{\mathbf{g}}_3 \end{bmatrix} &= \text{cov} \left( \begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \\ \mathbf{g}_3 \end{bmatrix}, [\mathbf{u}_1^* \ ' \ \mathbf{u}_2^* \ ' \ \mathbf{u}_3^* \ '] \right) \left( \text{var} \begin{bmatrix} \mathbf{u}_1^* \\ \mathbf{u}_2^* \\ \mathbf{u}_3^* \end{bmatrix} \right)^{-1} \begin{bmatrix} \hat{\mathbf{u}}_1^* \\ \hat{\mathbf{u}}_2^* \\ \hat{\mathbf{u}}_3^* \end{bmatrix} \\
&= \text{cov} \left( \begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \\ \mathbf{g}_3 \end{bmatrix}, [\mathbf{u}_1^* \ ' \ \mathbf{u}_2^* \ ' \ \mathbf{u}_3^* \ '] \right) \left( \text{var} \begin{bmatrix} \mathbf{Z}\mathbf{g}_1 \\ \mathbf{Z}\mathbf{g}_2 \\ \mathbf{Z}\mathbf{g}_3 \end{bmatrix} \right)^{-1} \begin{bmatrix} \hat{\mathbf{u}}_1^* \\ \hat{\mathbf{u}}_2^* \\ \hat{\mathbf{u}}_3^* \end{bmatrix} \\
&= \begin{bmatrix} \text{cov}(\mathbf{g}_1, \mathbf{g}_1' \mathbf{Z}') & \text{cov}(\mathbf{g}_1, \mathbf{g}_2' \mathbf{Z}') & \text{cov}(\mathbf{g}_1, \mathbf{g}_3' \mathbf{Z}') \\ \text{cov}(\mathbf{g}_2, \mathbf{g}_1' \mathbf{Z}') & \text{cov}(\mathbf{g}_2, \mathbf{g}_2' \mathbf{Z}') & \text{cov}(\mathbf{g}_2, \mathbf{g}_3' \mathbf{Z}') \\ \text{cov}(\mathbf{g}_3, \mathbf{g}_1' \mathbf{Z}') & \text{cov}(\mathbf{g}_3, \mathbf{g}_2' \mathbf{Z}') & \text{cov}(\mathbf{g}_3, \mathbf{g}_3' \mathbf{Z}') \end{bmatrix} \left( (\mathbf{Z}\mathbf{Z}')^{-1} \otimes \Sigma_u^{-1} \right)^{-1} \begin{bmatrix} \hat{\mathbf{u}}_1^* \\ \hat{\mathbf{u}}_2^* \\ \hat{\mathbf{u}}_3^* \end{bmatrix} \\
&= \begin{bmatrix} \sigma_{g_1}^2 \mathbf{Z}' & \sigma_{g_{12}} \mathbf{Z}' & \sigma_{g_{13}} \mathbf{Z}' \\ \sigma_{g_{12}} \mathbf{Z}' & \sigma_{g_2}^2 \mathbf{Z}' & \sigma_{g_{23}} \mathbf{Z}' \\ \sigma_{g_{13}} \mathbf{Z}' & \sigma_{g_{23}} \mathbf{Z}' & \sigma_{g_3}^2 \mathbf{Z}' \end{bmatrix} \begin{bmatrix} (\mathbf{Z}\mathbf{Z}')^{-1} g^{11} & (\mathbf{Z}\mathbf{Z}')^{-1} g^{12} & (\mathbf{Z}\mathbf{Z}')^{-1} g^{13} \\ (\mathbf{Z}\mathbf{Z}')^{-1} g^{12} & (\mathbf{Z}\mathbf{Z}')^{-1} g^{22} & (\mathbf{Z}\mathbf{Z}')^{-1} g^{23} \\ (\mathbf{Z}\mathbf{Z}')^{-1} g^{13} & (\mathbf{Z}\mathbf{Z}')^{-1} g^{23} & (\mathbf{Z}\mathbf{Z}')^{-1} g^{33} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{u}}_1^* \\ \hat{\mathbf{u}}_2^* \\ \hat{\mathbf{u}}_3^* \end{bmatrix} \\
&= (\Sigma_u \otimes \mathbf{Z}') (\Sigma_u^{-1} \otimes (\mathbf{Z}\mathbf{Z}')^{-1}) \begin{bmatrix} \hat{\mathbf{u}}_1^* \\ \hat{\mathbf{u}}_2^* \\ \hat{\mathbf{u}}_3^* \end{bmatrix} = (\mathbf{I}_3 \otimes \mathbf{Z}' (\mathbf{Z}\mathbf{Z}')^{-1}) \begin{bmatrix} \hat{\mathbf{u}}_1^* \\ \hat{\mathbf{u}}_2^* \\ \hat{\mathbf{u}}_3^* \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{Z}' (\mathbf{Z}\mathbf{Z}')^{-1} \hat{\mathbf{u}}_1^* \\ \mathbf{Z}' (\mathbf{Z}\mathbf{Z}')^{-1} \hat{\mathbf{u}}_2^* \\ \mathbf{Z}' (\mathbf{Z}\mathbf{Z}')^{-1} \hat{\mathbf{u}}_3^* \end{bmatrix} = \begin{bmatrix} \mathbf{Z}' & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Z}' \end{bmatrix} \begin{bmatrix} (\mathbf{Z}\mathbf{Z}')^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (\mathbf{Z}\mathbf{Z}')^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & (\mathbf{Z}\mathbf{Z}')^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{u}}_1^* \\ \hat{\mathbf{u}}_2^* \\ \hat{\mathbf{u}}_3^* \end{bmatrix} \\
&= (\mathbf{I}_3 \otimes \mathbf{Z}') (\mathbf{I}_3 \otimes (\mathbf{Z}\mathbf{Z}')^{-1}) \begin{bmatrix} \hat{\mathbf{u}}_1^* \\ \hat{\mathbf{u}}_2^* \\ \hat{\mathbf{u}}_3^* \end{bmatrix}
\end{aligned}$$

such that

$$\begin{aligned}
\text{var} \begin{bmatrix} \hat{\mathbf{g}}_1 \\ \hat{\mathbf{g}}_2 \\ \hat{\mathbf{g}}_3 \end{bmatrix} &= (\mathbf{I}_3 \otimes \mathbf{Z}') (\mathbf{I}_3 \otimes (\mathbf{Z}\mathbf{Z}')^{-1}) \text{var} \begin{bmatrix} \hat{\mathbf{u}}_1^* \\ \hat{\mathbf{u}}_2^* \\ \hat{\mathbf{u}}_3^* \end{bmatrix} (\mathbf{I}_3 \otimes (\mathbf{Z}\mathbf{Z}')^{-1}) (\mathbf{I}_3 \otimes \mathbf{Z}) \\
&= (\mathbf{I}_3 \otimes \mathbf{Z}') (\mathbf{I}_3 \otimes (\mathbf{Z}\mathbf{Z}')^{-1}) (\Sigma_u \otimes (\mathbf{Z}\mathbf{Z}') - \mathbf{C}_{uu}) (\mathbf{I}_3 \otimes (\mathbf{Z}\mathbf{Z}')^{-1}) (\mathbf{I}_3 \otimes \mathbf{Z}) \\
&= (\Sigma_u \otimes \mathbf{Z}' (\mathbf{Z}\mathbf{Z}')^{-1} (\mathbf{Z}\mathbf{Z}') (\mathbf{Z}\mathbf{Z}')^{-1} \mathbf{Z}) - (\mathbf{I}_3 \otimes \mathbf{Z}' (\mathbf{Z}\mathbf{Z}')^{-1}) \mathbf{C}_{uu} (\mathbf{I}_3 \otimes (\mathbf{Z}\mathbf{Z}')^{-1} \mathbf{Z}) \\
&= (\Sigma_u \otimes \mathbf{Z}' (\mathbf{Z}\mathbf{Z}')^{-1} \mathbf{Z}) - (\mathbf{I}_3 \otimes \mathbf{Z}' (\mathbf{Z}\mathbf{Z}')^{-1}) \mathbf{C}_{uu} (\mathbf{I}_3 \otimes (\mathbf{Z}\mathbf{Z}')^{-1} \mathbf{Z})
\end{aligned}$$

Let's refer to this expression as  $\Sigma_{\hat{\mathbf{g}}} = \text{var} \begin{bmatrix} \hat{\mathbf{g}}_1 \\ \hat{\mathbf{g}}_2 \\ \hat{\mathbf{g}}_3 \end{bmatrix} = \begin{bmatrix} \Sigma_{\hat{\mathbf{g}}_{11}} & \Sigma_{\hat{\mathbf{g}}_{12}} & \Sigma_{\hat{\mathbf{g}}_{13}} \\ \Sigma_{\hat{\mathbf{g}}_{21}} & \Sigma_{\hat{\mathbf{g}}_{22}} & \Sigma_{\hat{\mathbf{g}}_{23}} \\ \Sigma_{\hat{\mathbf{g}}_{31}} & \Sigma_{\hat{\mathbf{g}}_{32}} & \Sigma_{\hat{\mathbf{g}}_{33}} \end{bmatrix}$ . So the so-called “fixed

effects” test for any SNP effect in  $\begin{bmatrix} \hat{\mathbf{g}}_1 \\ \hat{\mathbf{g}}_2 \\ \hat{\mathbf{g}}_3 \end{bmatrix}$  is based on dividing an element in that vector by the

square root of the corresponding diagonal element in the big expression for  $\Sigma_{\hat{\mathbf{g}}}$  above. In an

another word, test statistics for SNP  $k$  on trait  $i$  is  $z_{ik} = \frac{\hat{g}_{ik}}{\text{sep}(\hat{g}_{ik})}$ , where  $\hat{g}_{ik}$  is the SNP estimates

on trait  $i$  and the denominator is the square root of the  $k,k$  element in  $\text{var}(\hat{\mathbf{g}}_i)$ . It can be proved

that  $z_{ik} \sim N(0,1)$ .

Now our proposed rFE trait based on Cholesky decomposition is:

$\hat{\mathbf{g}}_{3|1.2} = \begin{bmatrix} -\phi_{31}^{(u)} \mathbf{I}_m & -\phi_{32}^{(u)} \mathbf{I}_m & \mathbf{I}_m \end{bmatrix} \begin{bmatrix} \hat{\mathbf{g}}_1 \\ \hat{\mathbf{g}}_2 \\ \hat{\mathbf{g}}_3 \end{bmatrix}$ . The corresponding variance covariance matrix of can be

then determined as follows:

$$\begin{aligned} \text{var}(\hat{\mathbf{g}}_{3|1.2}) &= \begin{bmatrix} -\phi_{31}^{(u)} \mathbf{I}_m & -\phi_{32}^{(u)} \mathbf{I}_m & \mathbf{I}_m \end{bmatrix} \text{var} \begin{bmatrix} \hat{\mathbf{g}}_1 \\ \hat{\mathbf{g}}_2 \\ \hat{\mathbf{g}}_3 \end{bmatrix} \begin{bmatrix} -\phi_{31}^{(u)} \mathbf{I}_m \\ -\phi_{32}^{(u)} \mathbf{I}_m \\ \mathbf{I}_m \end{bmatrix} \\ &= \begin{bmatrix} -\phi_{31}^{(u)} \mathbf{I}_m & -\phi_{32}^{(u)} \mathbf{I}_m & \mathbf{I}_m \end{bmatrix} \Sigma_{\hat{\mathbf{g}}} \begin{bmatrix} -\phi_{31}^{(u)} \mathbf{I}_m \\ -\phi_{32}^{(u)} \mathbf{I}_m \\ \mathbf{I}_m \end{bmatrix} \end{aligned}$$

For the element  $i$  of  $\hat{\mathbf{g}}_{3|1.2}$ , its variance then really turns out to be :

$$\Sigma_{\hat{\mathbf{g}}_{31,2}} = \Sigma_{\hat{\mathbf{g}}_{33}} + \left(\phi_{31}^{(u)}\right)^2 \Sigma_{\hat{\mathbf{g}}_{11}} + \left(\phi_{32}^{(u)}\right)^2 \Sigma_{\hat{\mathbf{g}}_{22}} - 2\phi_{31}^{(u)}\Sigma_{\hat{\mathbf{g}}_{31}} - 2\phi_{32}^{(u)}\Sigma_{\hat{\mathbf{g}}_{32}} + 2\phi_{31}^{(u)}\phi_{32}^{(u)}\Sigma_{\hat{\mathbf{g}}_{12}}$$

Hence the fixed effects test is based on the square root of this expression as the standard error.

Now, for the window-base test on window  $r$  covering SNPs  $j$  to  $k$  on trait  $i$ , the chi-square test statistics for that window will be :

$$\chi_{ir}^2 = \mathbf{g}_{ir} \left( \mathbf{Z}_r \Sigma_{\mathbf{g}_i} \mathbf{Z}_r \right)^{-1} \mathbf{g}_{ir}$$

where  $\mathbf{g}_{ir}$  and  $\mathbf{Z}_r$  are the corresponding parts in  $\mathbf{g}_i$  and  $\mathbf{Z}$  for SNPs  $j$  to  $k$ .  $\chi_{ri}^2$  statistics will be compared to a Chi-square distribution with  $df$  being number of SNPs in window  $r$ .

### Supplementary Figures and Tables

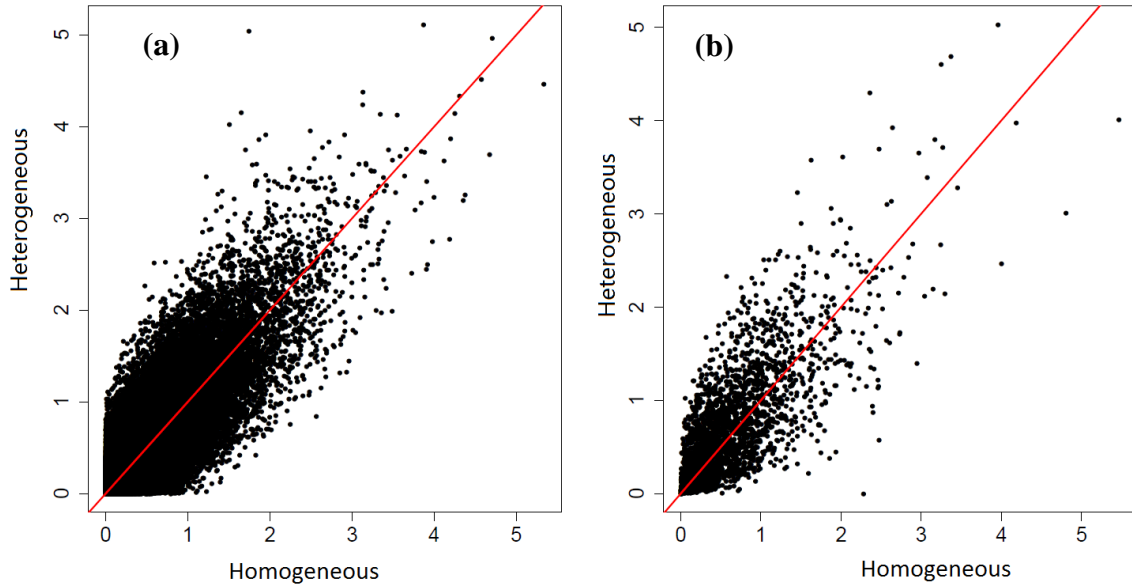


Figure C.1: Scatterplot of  $-\log_{10}(p)$  for SNPs (a) and windows (b) based on the homogeneous partial regression RFI versus heterogeneous partial regression RFI models. Line of slope 1 and intercept 0 superimposed.

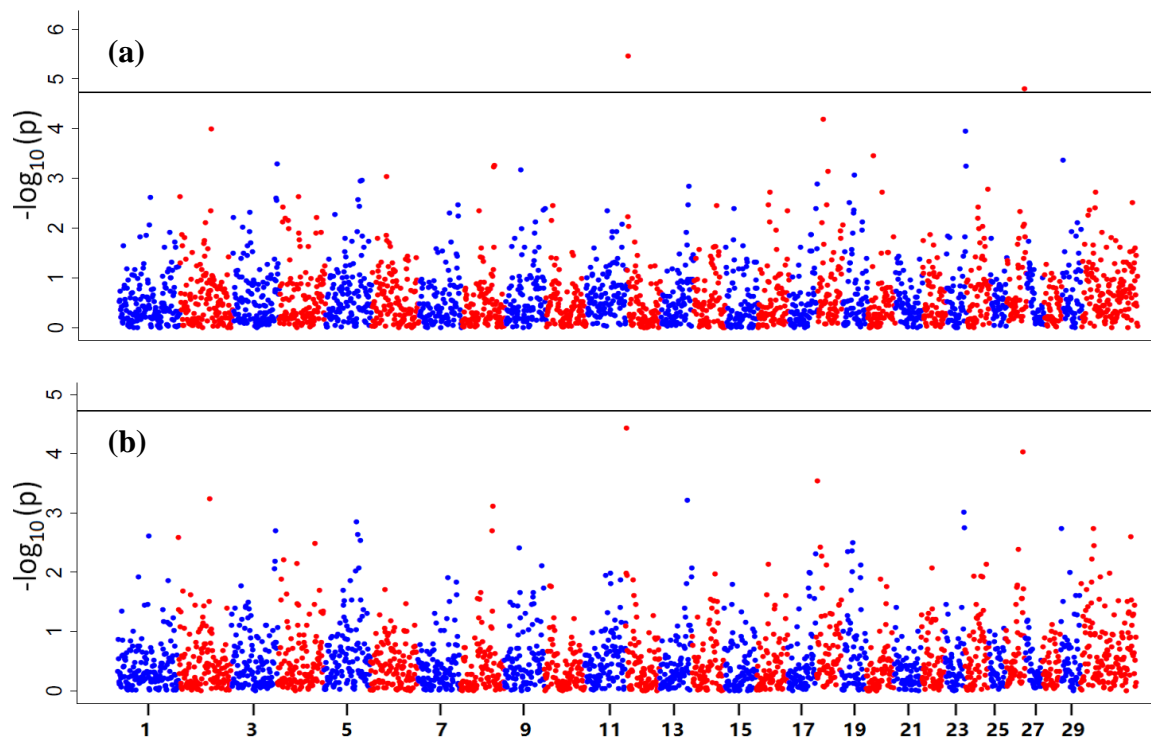


Figure C.2: Manhattan plots for windows on RFI based on all cows (a) and genotyped cows (b). Bonferroni correction threshold superimposed.

## APPENDIX D: Chapter 4

### Derivations

Assuming environmental covariates were fitted as fixed effects, a general genomic prediction model with GxE effect can be written like this:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{g}_0 + \mathbf{D}_1\mathbf{Z}\mathbf{g}_{b1} + \mathbf{D}_2\mathbf{Z}\mathbf{g}_{b2} + \mathbf{D}_3\mathbf{Z}\mathbf{g}_{b3} + \dots + \mathbf{D}_c\mathbf{Z}\mathbf{g}_{bc} + \mathbf{e}$$

for  $\mathbf{g}_0$  and  $\{\mathbf{g}_{bi}\}_{i=1}^c$  being vectors for SNP-specific random intercept and slope, respectively.

Suppose that

$$\begin{bmatrix} \mathbf{g}_0 \\ \mathbf{g}_{b1} \\ \mathbf{g}_{b2} \\ \mathbf{g}_{bc} \end{bmatrix} \sim N \left( \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{I}\sigma_{g0}^2 & \mathbf{I}\sigma_{g0b1} & \mathbf{I}\sigma_{g0b2} & \mathbf{I}\sigma_{g0bc} \\ & \mathbf{I}\sigma_{b1}^2 & \mathbf{I}\sigma_{b1b2} & \mathbf{I}\sigma_{b1bc} \\ & & \mathbf{I}\sigma_{b2}^2 & \mathbf{I}\sigma_{b2bc} \\ & & & \mathbf{I}\sigma_{bc}^2 \end{bmatrix} \right).$$

Now, we rewrite  $\mathbf{u}_0 = \mathbf{Z}\mathbf{g}_0$  and  $\mathbf{v} = \sum_{k=1}^c \mathbf{D}_k\mathbf{Z}\mathbf{g}_{bk}$  as being the genomic breeding values for

intercept and slope such that  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}_0 + \mathbf{v} + \mathbf{e}$ .

To simplify things, let's assume that all of the covariances are zero and all of the reaction norm variances are the same; i.e.,  $\sigma_{bk}^2 = \sigma_b^2 \forall k$  for  $k=1,2,3,\dots,c$ , such that

$$\begin{bmatrix} \mathbf{u}_0 \\ \mathbf{v} \end{bmatrix} \sim N \left( \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{Z}\mathbf{Z}'\sigma_{g0}^2 & \mathbf{0} \\ \mathbf{0} & \sigma_b^2 \sum_{k=1}^c \mathbf{D}_k\mathbf{Z}\mathbf{Z}'\mathbf{D}_k \end{bmatrix} \right)$$

where  $\sum_{i=1}^c \mathbf{D}_i \mathbf{Z} \mathbf{Z}' \mathbf{D}_i$  can be also written as  $\mathbf{Z} \mathbf{Z}' \# \mathbf{D}^* \mathbf{D}^{*'}$  such that

$\mathbf{D}^* = [\text{vecdiag}(\mathbf{D}_1) \text{ vecdiag}(\mathbf{D}_2) \text{ vecdiag}(\mathbf{D}_3) \cdots \text{vecdiag}(\mathbf{D}_c)]$  is a  $n \times c$  matrix of covariates for  $n$  being number of records and  $c$  being number of environmental covariates fitted.

The mixed model equation for above model can be written as

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}' & \mathbf{X} \\ \mathbf{X} & \mathbf{I} + (\mathbf{Z}\mathbf{Z}'\sigma_{g0}^2)^{-1}\sigma_e^2 & \mathbf{I} \\ \mathbf{X} & \mathbf{I} & \mathbf{I} + \left(\sum_{k=1}^c \mathbf{D}_k \mathbf{Z} \mathbf{Z}' \mathbf{D}_k \sigma_b^2\right)^{-1}\sigma_e^2 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}}_0 \\ \hat{\mathbf{v}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{y} \\ \mathbf{y} \end{bmatrix}.$$

Assume

$$\begin{bmatrix} \mathbf{C}_{00} & \mathbf{C}_{01} & \mathbf{C}_{02} \\ \mathbf{C}_{10} & \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{20} & \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}' & \mathbf{X} \\ \mathbf{X} & \mathbf{I} + (\mathbf{Z}\mathbf{Z}'\sigma_{g0}^2)^{-1}\sigma_e^2 & \mathbf{I} \\ \mathbf{X} & \mathbf{I} & \mathbf{I} + \left(\sum_{k=1}^c \mathbf{D}_k \mathbf{Z} \mathbf{Z}' \mathbf{D}_k \sigma_b^2\right)^{-1}\sigma_e^2 \end{bmatrix}^{-1}$$

To backsolve for SNP-specific random intercept and slopes, it pertains to any of the covariates use the following:

$$\begin{aligned}
\begin{bmatrix} \hat{\mathbf{g}}_0 \\ \hat{\mathbf{g}}_{b1} \\ \hat{\mathbf{g}}_{b2} \\ \hat{\mathbf{g}}_{bc} \end{bmatrix} &= \text{COV} \left( \begin{bmatrix} \mathbf{g}_0 \\ \mathbf{g}_{b1} \\ \mathbf{g}_{b2} \\ \mathbf{g}_{bc} \end{bmatrix}, [\mathbf{u}' \quad \mathbf{v}'] \right) \left( \text{var} \begin{bmatrix} \mathbf{u}_0 \\ \mathbf{v} \end{bmatrix} \right)^{-1} \begin{bmatrix} \hat{\mathbf{u}}_0 \\ \hat{\mathbf{v}} \end{bmatrix} \\
&= \text{COV} \left( \begin{bmatrix} \mathbf{g}_0 \\ \mathbf{g}_{b1} \\ \mathbf{g}_{b2} \\ \mathbf{g}_{bc} \end{bmatrix}, \begin{bmatrix} \mathbf{g}_0' \mathbf{Z}' & \sum_{k=1}^c \mathbf{g}_{bk}' \mathbf{Z}' \mathbf{D}_k \end{bmatrix} \right) \left( \text{var} \begin{bmatrix} \mathbf{Z} \mathbf{g}_0 \\ \sum_{k=1}^c \mathbf{D}_k \mathbf{Z} \mathbf{g}_{bk} \end{bmatrix} \right)^{-1} \begin{bmatrix} \hat{\mathbf{u}}_0 \\ \hat{\mathbf{v}} \end{bmatrix} \\
&= \begin{bmatrix} \text{COV}(\mathbf{g}_0, \mathbf{g}_0' \mathbf{Z}') & \mathbf{0} \\ \mathbf{0} & \text{COV}(\mathbf{g}_{b1}, \mathbf{g}_{b1}' \mathbf{Z}' \mathbf{D}_1) \\ \mathbf{0} & \text{COV}(\mathbf{g}_{b2}, \mathbf{g}_{b2}' \mathbf{Z}' \mathbf{D}_2) \\ \mathbf{0} & \text{COV}(\mathbf{g}_{bc}, \mathbf{g}_{bc}' \mathbf{Z}' \mathbf{D}_2) \end{bmatrix} \begin{bmatrix} \mathbf{Z} \mathbf{Z}' \sigma_{g_0}^2 & \mathbf{0} \\ \mathbf{0} & \sigma_b^2 \sum_{k=1}^c \mathbf{D}_k \mathbf{Z} \mathbf{Z}' \mathbf{D}_k \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mathbf{u}}_0 \\ \hat{\mathbf{v}} \end{bmatrix} \\
&= \begin{bmatrix} \sigma_{g_0}^2 \mathbf{Z}' & \mathbf{0} \\ \mathbf{0} & \sigma_v^2 \mathbf{Z}' \mathbf{D}_1 \\ \mathbf{0} & \sigma_v^2 \mathbf{Z}' \mathbf{D}_2 \\ \mathbf{0} & \sigma_v^2 \mathbf{Z}' \mathbf{D}_c \end{bmatrix} \begin{bmatrix} \sigma_{g_0}^{-2} (\mathbf{Z} \mathbf{Z}')^{-1} \hat{\mathbf{u}}_0 \\ \left( \sigma_b^2 \sum_{k=1}^c \mathbf{D}_k \mathbf{Z} \mathbf{Z}' \mathbf{D}_k \right)^{-1} \hat{\mathbf{v}} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{Z}' (\mathbf{Z} \mathbf{Z}')^{-1} \hat{\mathbf{u}}_0 \\ \mathbf{Z}' \mathbf{D}_1 \left( \sum_{i=1}^c \mathbf{D}_i \mathbf{Z} \mathbf{Z}' \mathbf{D}_i \right)^{-1} \hat{\mathbf{v}} \\ \mathbf{Z}' \mathbf{D}_2 \left( \sum_{i=1}^c \mathbf{D}_i \mathbf{Z} \mathbf{Z}' \mathbf{D}_i \right)^{-1} \hat{\mathbf{v}} \\ \mathbf{Z}' \mathbf{D}_c \left( \sum_{i=1}^c \mathbf{D}_i \mathbf{Z} \mathbf{Z}' \mathbf{D}_i \right)^{-1} \hat{\mathbf{v}} \end{bmatrix} = \begin{bmatrix} \mathbf{Z}' (\mathbf{Z} \mathbf{Z}')^{-1} \hat{\mathbf{u}}_0 \\ \mathbf{Z}' \mathbf{D}_1 \left( \sum_{k=1}^c \mathbf{Z} \mathbf{Z}' \# \mathbf{D}^* \mathbf{D}^* \right)^{-1} \hat{\mathbf{v}} \\ \mathbf{Z}' \mathbf{D}_2 \left( \sum_{k=1}^c \mathbf{Z} \mathbf{Z}' \# \mathbf{D}^* \mathbf{D}^* \right)^{-1} \hat{\mathbf{v}} \\ \mathbf{Z}' \mathbf{D}_c \left( \sum_{k=1}^c \mathbf{Z} \mathbf{Z}' \# \mathbf{D}^* \mathbf{D}^* \right)^{-1} \hat{\mathbf{v}} \end{bmatrix}
\end{aligned}$$

$$\text{var}(\hat{\mathbf{g}}_0) = \text{var}(\mathbf{Z}' (\mathbf{Z} \mathbf{Z}')^{-1} \hat{\mathbf{u}}_0) = \mathbf{Z}' (\mathbf{Z} \mathbf{Z}')^{-1} \text{var}(\hat{\mathbf{u}}_0) (\mathbf{Z} \mathbf{Z}')^{-1} \mathbf{Z}$$

$$\mathbf{Z}' (\mathbf{Z} \mathbf{Z}')^{-1} (\mathbf{Z} \mathbf{Z}' \sigma_{g_0}^2 - \mathbf{C}_{11} \sigma_e^2) (\mathbf{Z} \mathbf{Z}')^{-1} \mathbf{Z} \sigma_e^2$$

and

$$\begin{aligned}
\text{var}(\hat{\mathbf{g}}_{bk}) &= \text{var}\left(\mathbf{Z}'\mathbf{D}_k\left(\sum_{k=1}^c \mathbf{Z}\mathbf{Z}'\#\mathbf{D}^*\mathbf{D}^{*'}\right)^{-1}\hat{\mathbf{v}}\right) \\
&= \mathbf{Z}'\mathbf{D}_k\left(\sum_{k=1}^c \mathbf{Z}\mathbf{Z}'\#\mathbf{D}^*\mathbf{D}^{*'}\right)^{-1}\text{var}(\hat{\mathbf{v}})\left(\sum_{k=1}^c \mathbf{Z}\mathbf{Z}'\#\mathbf{D}^*\mathbf{D}^{*'}\right)^{-1}\mathbf{D}_k\mathbf{Z} \\
&= \mathbf{Z}'\mathbf{D}_k\left(\sum_{k=1}^c \mathbf{Z}\mathbf{Z}'\#\mathbf{D}^*\mathbf{D}^{*'}\right)^{-1}\left(\sum_{k=1}^c \mathbf{Z}\mathbf{Z}'\#\mathbf{D}^*\mathbf{D}^{*'}\sigma_b^2 - \mathbf{C}_{22}\sigma_e^2\right)\left(\sum_{k=1}^c \mathbf{Z}\mathbf{Z}'\#\mathbf{D}^*\mathbf{D}^{*'}\right)^{-1}\mathbf{D}_k\mathbf{Z}
\end{aligned}$$

### SNP based GWA test

Then, the GWA tests on the SNP-specific random intercept  $\hat{g}_{0i}$  and slope  $\hat{g}_{bki}$  on environmental covariate  $k$  for SNP  $i$  can be written as

$$z_{0i} = \frac{\hat{g}_{0i}}{\sqrt{\text{var}(\hat{g}_{0i})}} \text{ and } z_{bki} = \frac{\hat{g}_{bki}}{\sqrt{\text{var}(\hat{g}_{bki})}}.$$

Z scores will be compared to a standardized normal distribution to obtain p values for each SNP.

### Window-based GWA test

We also considered window-based associations test with partitioning columns of  $\mathbf{Z}$ , and hence elements of  $\mathbf{g}_0, \mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_c$ , into  $R$  distinct 1MB windows. Then, for the window  $r$  with genotype of  $\mathbf{Z}_r$  for  $r_n$  SNPs covered in this window, GWA tests on random intercept vector  $\hat{\mathbf{g}}_{0r}$  and slope vector  $\hat{\mathbf{g}}_{bkr}$  on environmental covariate  $k$  can be written as

$$\begin{aligned}
\chi_{0r}^2 &= \hat{\mathbf{g}}_{0r}' \text{var}(\hat{\mathbf{g}}_{0r})^{-1} \hat{\mathbf{g}}_{0r} \quad \text{and} \quad \chi_{bkr}^2 = \hat{\mathbf{g}}_{bkr}' \text{var}(\hat{\mathbf{g}}_{bkr})^{-1} \hat{\mathbf{g}}_{bkr}, \quad \text{where} \quad \text{var}(\hat{\mathbf{g}}_{0r}) = \mathbf{Z}_r' \text{var}(\hat{\mathbf{g}}_0) \mathbf{Z}_r \text{ and} \\
\text{var}(\hat{\mathbf{g}}_{bkr}) &= \mathbf{Z}_r' \text{var}(\hat{\mathbf{g}}_{bk}) \mathbf{Z}_r
\end{aligned}$$

To obtain p values for a window, Chi-square statistics will be compared to a Chi-square distribution with degree freedom being  $r_n$ .

### A simple approach to solve for the MME model

The simplest way to solve for the solutions to the MME would be as follows:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}' \\ \mathbf{X} & \mathbf{I} + \mathbf{M}^{-1}\sigma_e^2 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{m}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{y} \end{bmatrix}$$

Where  $\mathbf{M} = \mathbf{ZZ}'\sigma_{g0}^2 + \mathbf{ZZ}'\mathbf{D}^*\mathbf{D}^*\sigma_b^2$  and  $\mathbf{m} = \mathbf{u} + \mathbf{v}$ .

Now let the inverse of  $\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}' \\ \mathbf{X} & \mathbf{I} + \mathbf{M}^{-1}\sigma_e^2 \end{bmatrix}$  be written as

$$\begin{bmatrix} \mathbf{C}_{00} & \mathbf{C}_{0m} \\ \mathbf{C}_{m0} & \mathbf{C}_{mm} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}' \\ \mathbf{X} & \mathbf{I} + \mathbf{M}^{-1}\sigma_e^2 \end{bmatrix}^{-1} \text{ but what we really want is this inverse:}$$

$$\begin{bmatrix} \mathbf{C}_{00} & \mathbf{C}_{01} & \mathbf{C}_{02} \\ \mathbf{C}_{10} & \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{20} & \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}' & \mathbf{X} \\ \mathbf{X} & \mathbf{I} + (\mathbf{ZZ}'\sigma_{g0}^2)^{-1}\sigma_e^2 & \mathbf{I} \\ \mathbf{X} & \mathbf{I} & \mathbf{I} + \left( \sum_{k=1}^c \mathbf{D}_k \mathbf{ZZ}' \mathbf{D}_k \sigma_b^2 \right)^{-1} \sigma_e^2 \end{bmatrix}^{-1}$$

Let's write  $\mathbf{G}_1 = \mathbf{ZZ}'\sigma_{g0}^2$  and  $\mathbf{G}_2 = \mathbf{ZZ}'\mathbf{D}^*\mathbf{D}^*\sigma_b^2$

Based on Henderson (1985),  $\mathbf{C}_{jj} = \sigma_e^{-2}\mathbf{G}_j - \mathbf{G}_j\mathbf{M}^{-1}(\mathbf{M}\sigma_e^{-2} - \mathbf{C}_{mm})\mathbf{M}^{-1}\mathbf{G}_j$ ;  $j=1$  and  $2$

Then one could derive

$$\mathbf{C}_{11} = \sigma_e^{-2}\mathbf{G}_1 - \mathbf{G}_1\mathbf{M}^{-1}(\mathbf{M}\sigma_e^{-2} - \mathbf{C}_{mm})\mathbf{M}^{-1}\mathbf{G}_1 \text{ that pertains to the random intercept}$$

and

$$\mathbf{C}_{22} = \sigma_e^{-2}\mathbf{G}_2 - \mathbf{G}_2\mathbf{M}^{-1}(\mathbf{M}\sigma_e^{-2} - \mathbf{C}_{mm})\mathbf{M}^{-1}\mathbf{G}_2 \text{ that pertains to the random slope}$$

Along with

$$\hat{\mathbf{u}}_0 = \mathbf{G}_1\mathbf{M}^{-1}\hat{\mathbf{m}}$$

And

$$\hat{\mathbf{v}} = \mathbf{G}_2 \mathbf{M}^{-1} \hat{\mathbf{m}}$$

### Supplementary Figures and Tables

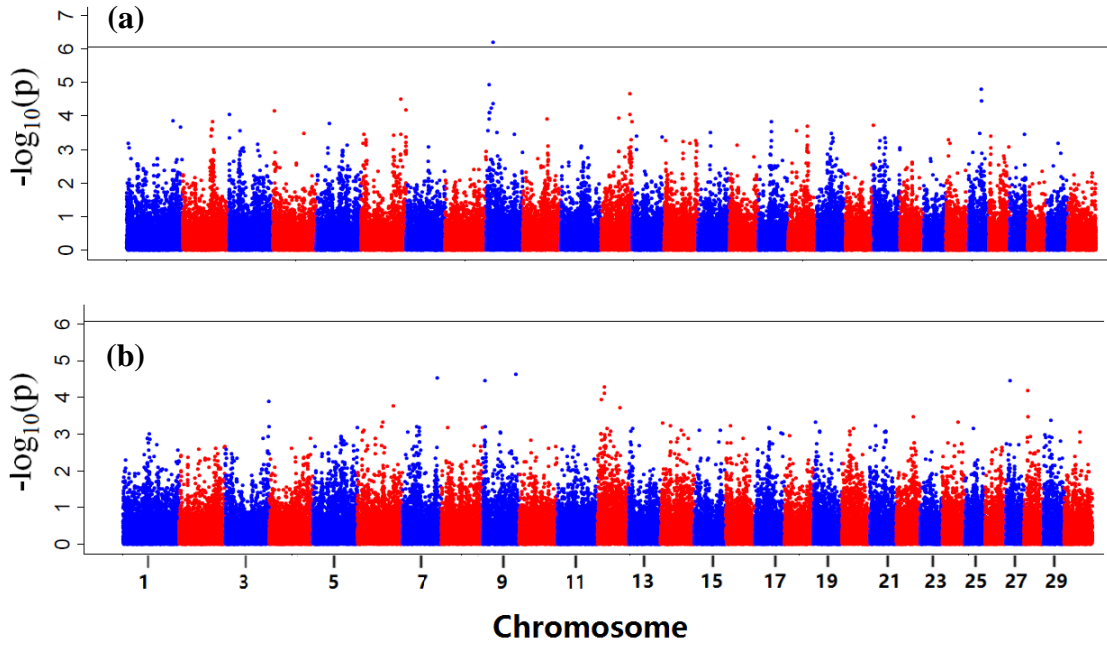


Figure D.1: Manhattan plot for SNP-specific random intercept based on the regular GBLUP (a) and the RN model (b)

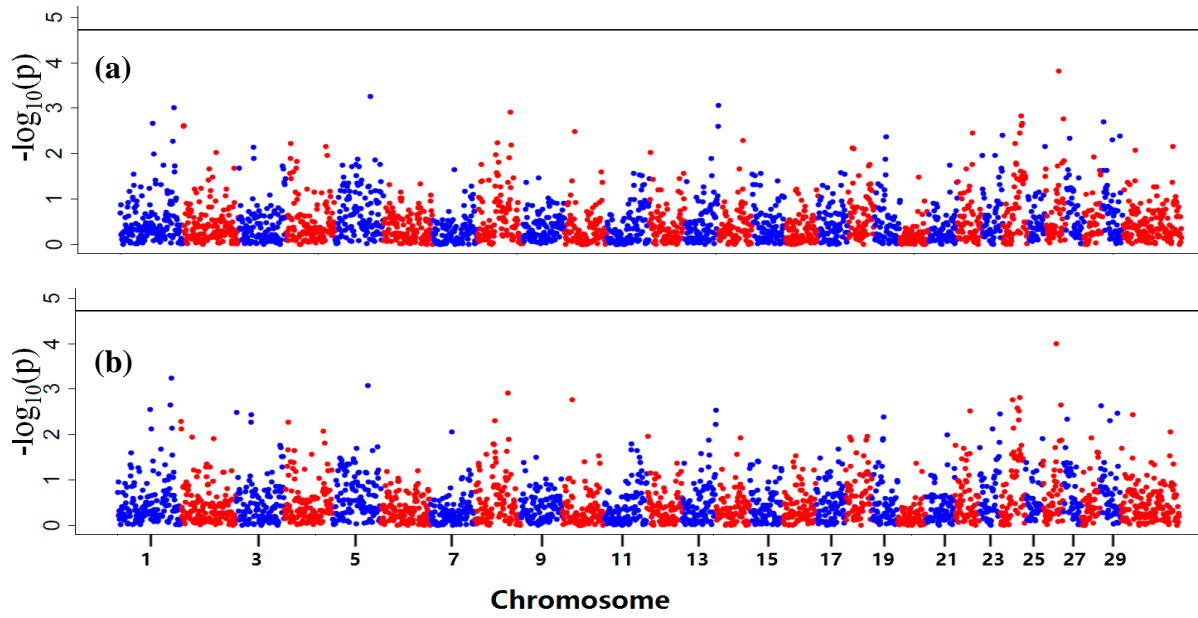


Figure D.2: Manhattan plot for Window-specific random intercept based on the regular GBLUP (a) and the RN model (b)

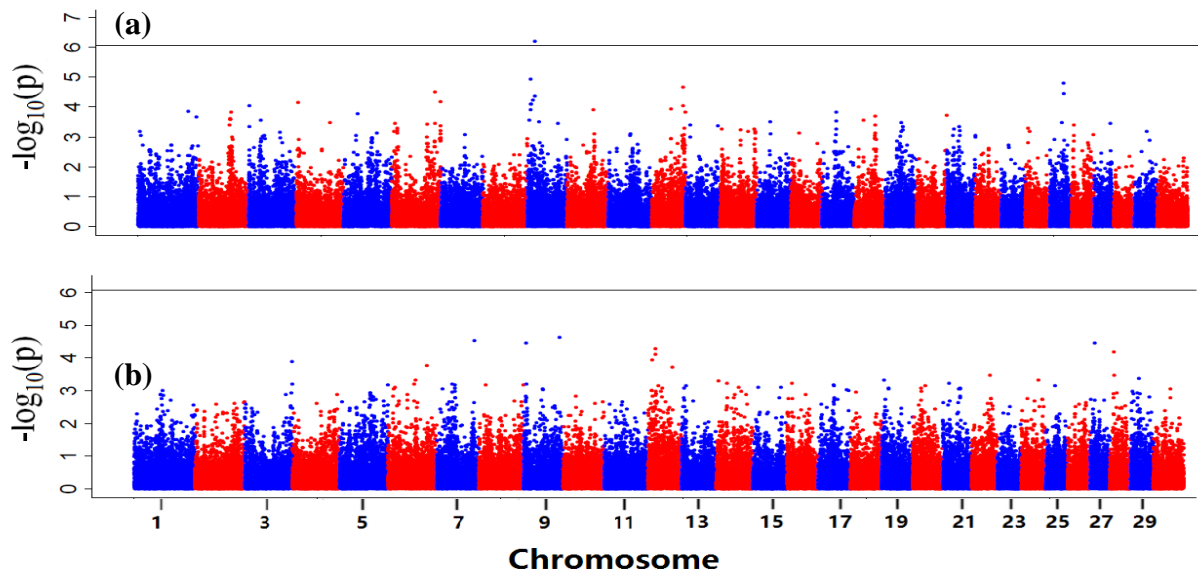


Figure D.3: Manhattan plot for SNP-specific random slope on the linear (a) and quadratic (b) terms of AP

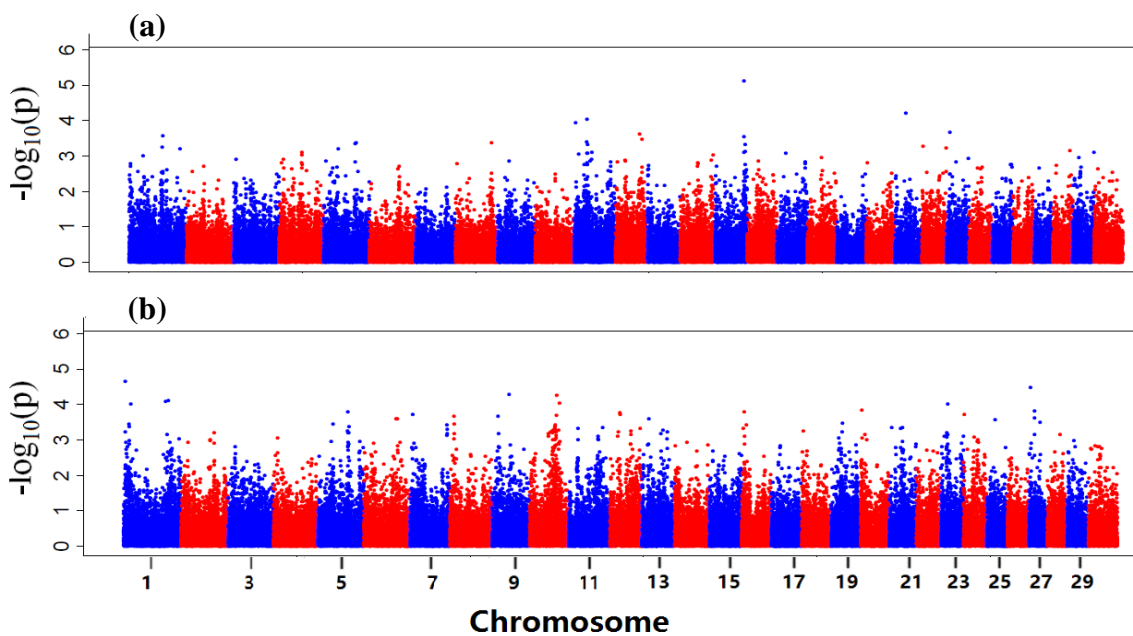


Figure D.4: Manhattan plot for SNP-specific random slope on the linear (a) and quadratic (b) terms of TP65

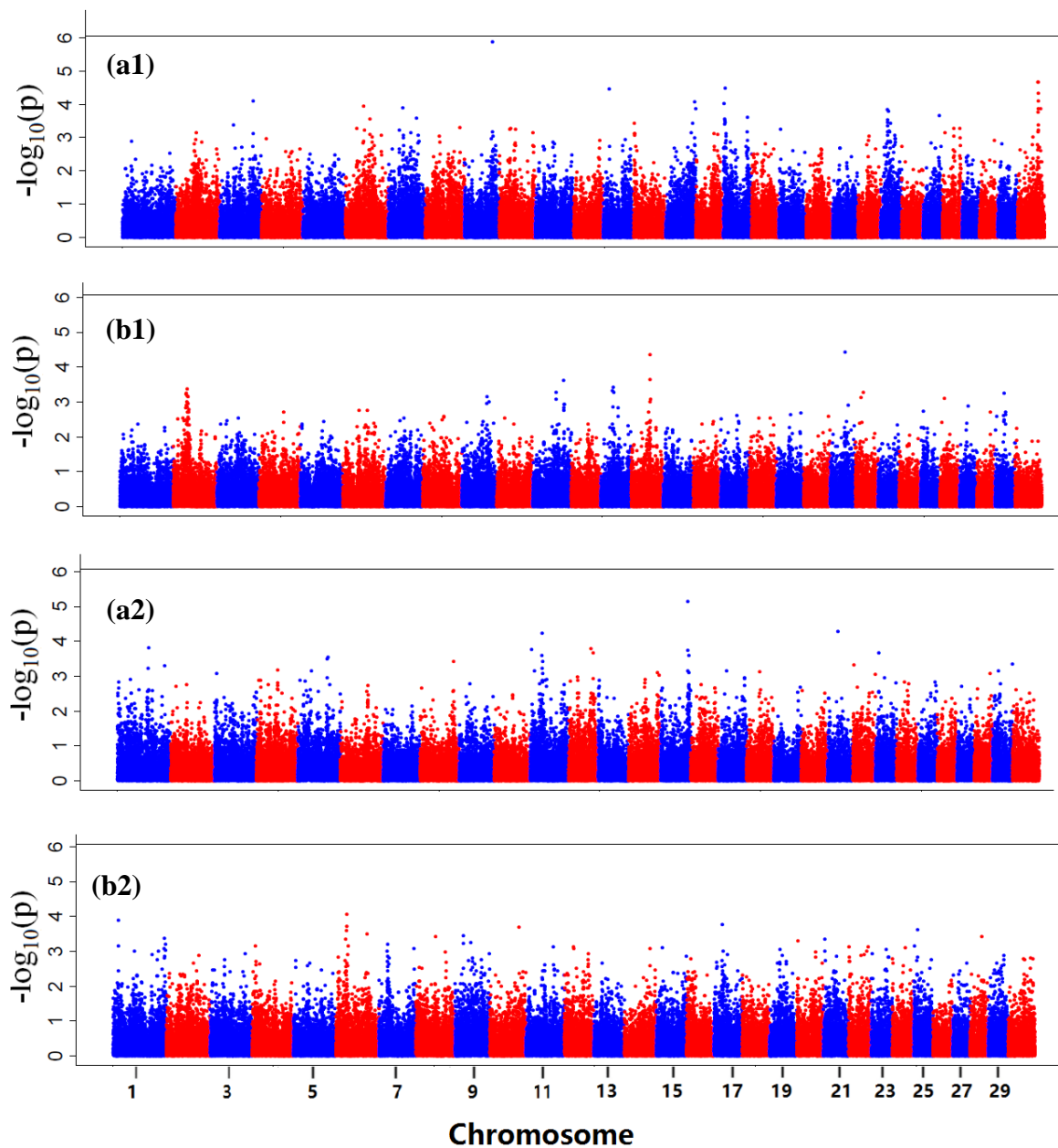


Figure D.5: Manhattan plot for SNP-specific random slope on the linear (a) and quadratic (b) terms of RH (1) and interaction between TP65 and RH (2)

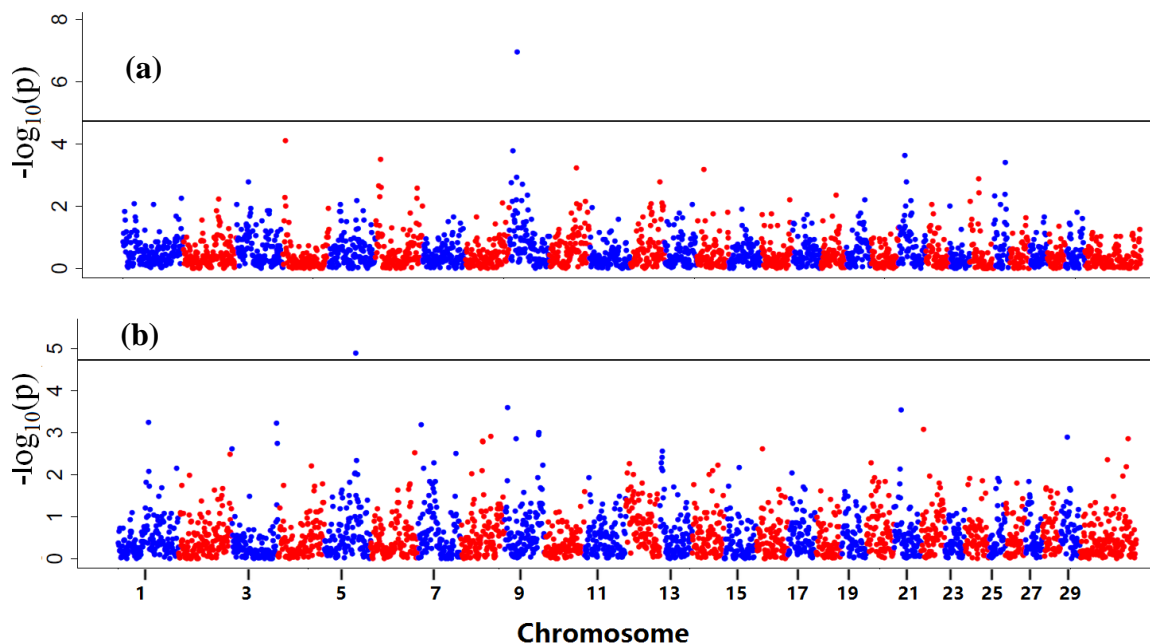


Figure D.6: Manhattan plot for Window-specific random on the linear (a) and quadratic (b) terms of AP

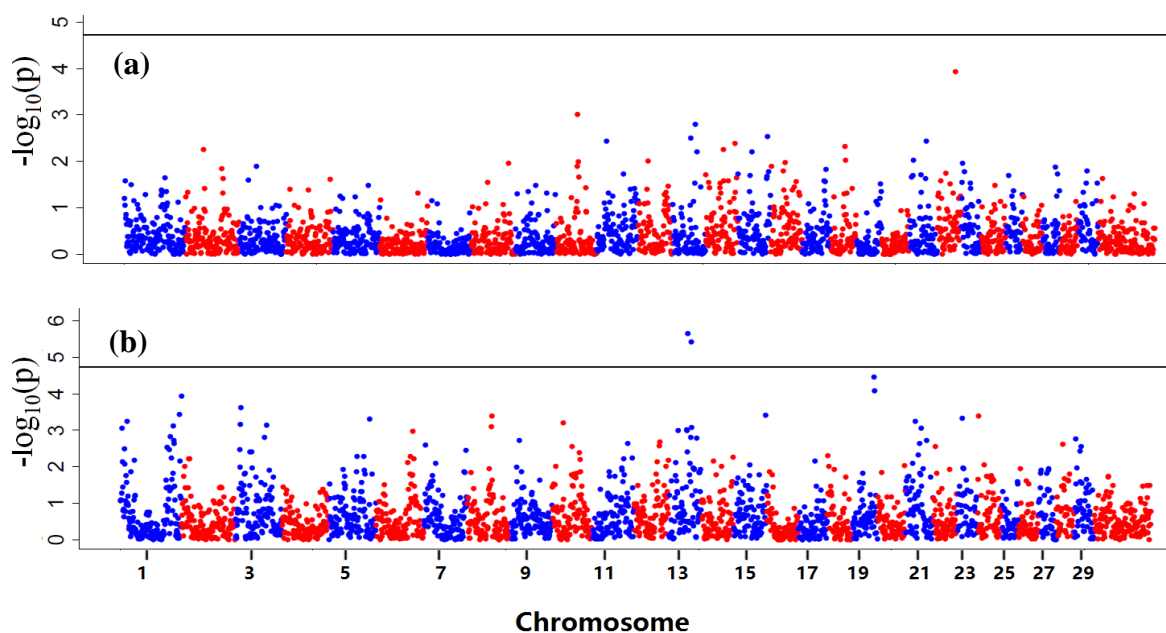


Figure D.7: Manhattan plot for Window-specific random on the linear (a) and quadratic (b) terms of TP65

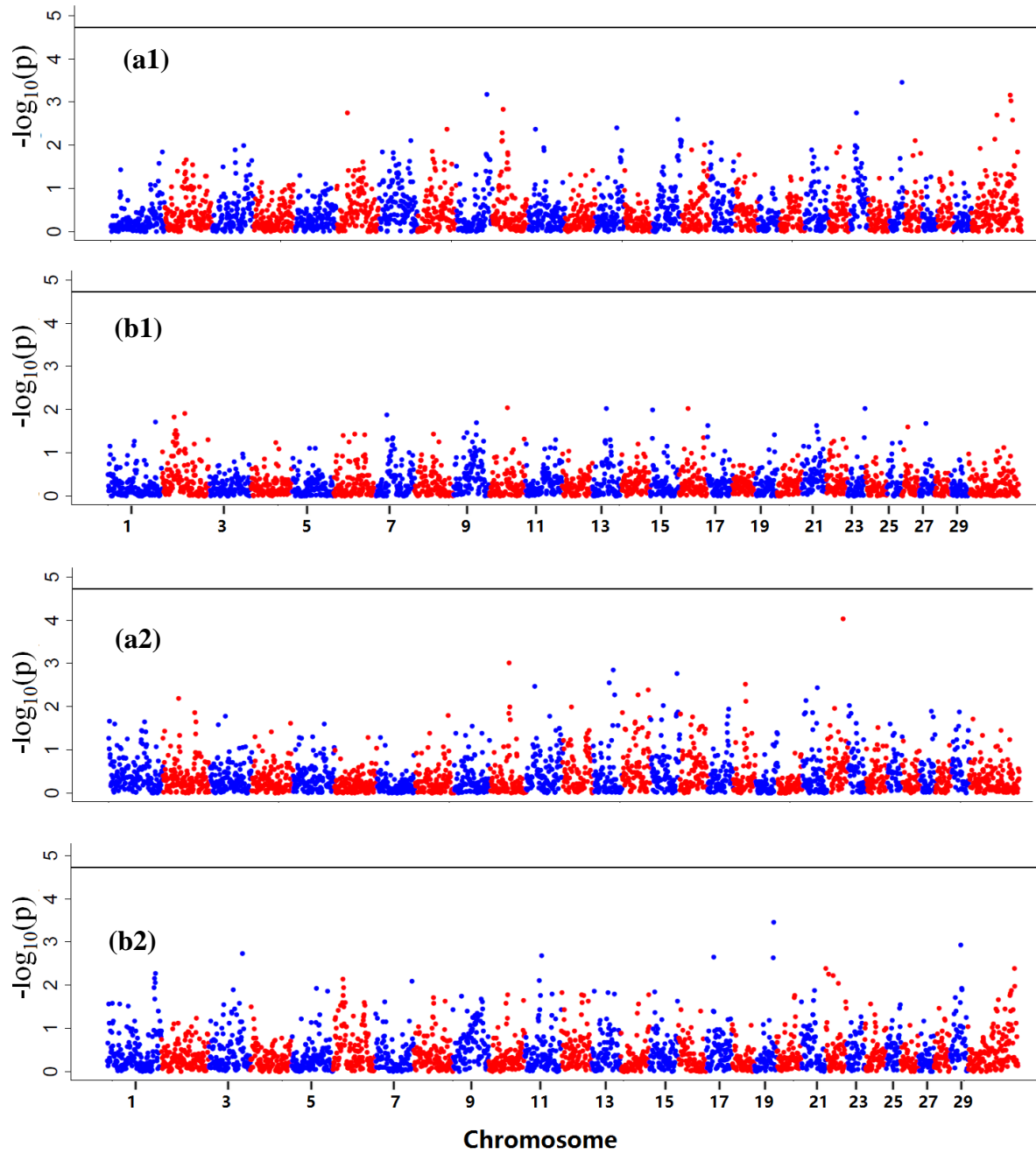


Figure D.8: Manhattan plot for Window-specific random slope on the linear (a) and quadratic (b) terms of RH (1) and interaction between TP65 and RH (2)

## REFERENCES

## REFERENCES

- Aggrey, S. E. and R. Rekaya. 2013. Dissection of Koch's residual feed intake: implications for selection. *Poult. Sci.* 92(10):2600-2605.
- Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Anim. Sci.* 93(2):743-752.
- Aguilar, I., I. Misztal, and S. Tsuruta. 2009. Genetic components of heat stress for dairy cattle with multiple lactations. *J. Anim. Sci.* 92(11):5702-5711.
- Alpay, F., Y. Zare, M. H. Kamalludin, X. Huang, X. Shi, G. E. Shook, M. T. Collins, and B. W. Kirkpatrick. 2014. Genome-wide association study of susceptibility to infection by *Mycobacterium avium* subspecies paratuberculosis in Holstein cattle. *PLoS ONE* 9(12):e111704.
- Andersson, L., A. L. Archibald, C. D. Bottema, R. Brauning, S. C. Burgess, D. W. Burt, E. Casas, H. H. Cheng, L. Clarke, C. Couldrey, B. P. Dalrymple, C. G. Elsik, S. Foissac, E. Giuffra, M. A. Groenen, B. J. Hayes, L. S. Huang, H. Khatib, J. W. Kijas, H. Kim, J. K. Lunney, F. M. McCarthy, J. C. McEwan, S. Moore, B. Nanduri, C. Notredame, Y. Palti, G. S. Plastow, J. M. Reecy, G. A. Rohrer, E. Sarropoulou, C. J. Schmidt, J. Silverstein, R. L. Tellam, M. Tixier-Boichard, G. Tosser-Klopp, C. K. Tuggle, J. Vilkki, S. N. White, S. H. Zhao, H. Zhou, and F. Consortium. 2015. Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. *Genome Biol.* 16:57.
- Banerjee, S., B. S. Yandell, and N. Yi. 2008. Bayesian quantitative trait loci mapping for multiple traits. *Genetics* 179:2275-2289.
- Banos, G., M. P. Coffey, R. F. Veerkamp, D. P. Berry, and E. Wall. 2012. Merging and characterising phenotypic data on conventional and rare traits from dairy cattle experimental resources in three countries. *Animal : an international journal of animal bioscience* 6(7):1040-1048.
- Bello, N. M., J. P. Steibel, R. J. Erskine, and R. J. Tempelman. 2012. Inferring Upon Heterogeneous Associations in Dairy Cattle Performance Using a Bivariate Hierarchical Model. *J AGRIC BIOL ENVIR S* 17(1):142-161.
- Bello, N. M., J. P. Steibel, and R. J. Tempelman. 2010. Hierarchical Bayesian modeling of random and residual variance-covariance matrices in bivariate mixed effects models. *Biom. J.* 52(3):297-313.
- Bernal Rubio, Y. L., J. L. Gualdrón Duarte, R. O. Bates, C. W. Ernst, D. Nonneman, G. A. Rohrer, A. King, S. D. Shackelford, T. L. Wheeler, R. J. Cantet, and J. P. Steibel. 2016. Meta-analysis of genome-wide association from genomic prediction models. *Anim Genet* 47:36-48.

- Berry, D. and J. Pryce. 2014. Feed Efficiency in Growing and Mature Animals. in Proc. 10th World Congress on Genetics Applied to Livestock Production. In Proc. 10th World Congress on Genetics Applied to Livestock Production. ASAS, August 17-22, 2014, Vancouver, CANADA.
- Berry, D. P., F. Buckley, P. Dillon, R. D. Evans, M. Rath, and R. F. Veerkamp. 2003. Genetic relationships among body condition score, body weight, milk yield, and fertility in dairy cows. *J. Dairy Sci.* 86:2193-2204.
- Berry, D. P., M. P. Coffey, J. E. Pryce, Y. de Haas, P. Lovendahl, N. Krattenmacher, J. J. Crowley, Z. Wang, D. Spurlock, K. Weigel, K. Macdonald, and R. F. Veerkamp. 2014. International genetic evaluations for feed intake in dairy cattle through the collation of data from multiple sources. *J. Anim. Sci.* 97(6):3894-3905.
- Berry, D. P. and J. J. Crowley. 2013. Cell Biology Symposium: genetics of feed efficiency in dairy and beef cattle. *J. Anim. Sci.* 91(4):1594-1613.
- Bohmanova, J., I. Misztal, and J. B. Cole. 2007. Temperature-humidity indices as indicators of milk production losses due to heat stress. *J. Anim. Sci.* 90(4):1947-1956.
- Bolormaa, S., B. J. Hayes, K. Savin, R. Hawken, W. Barendse, P. F. Arthur, R. M. Herd, and M. E. Goddard. 2011. Genome-wide association studies for feedlot and growth traits in cattle. *J. Anim. Sci.* 89(6):1684-1697.
- Brooks, S. P. and A. Gelman. 1998. General methods for monitoring convergence of iterative simulations. *J. Comput. Graph. Stat.* 7:434-455.
- Byrne, T. J., B. F. Santos, P. R. Amer, D. Martin-Collado, J. E. Pryce, and M. Axford. 2016. New breeding objectives and selection indices for the Australian dairy industry. *J. Dairy Sci.* 99:8146-8167.
- Calus, M. P. L., D. P. Berry, G. Banos, Y. d. Haas, and R. F. Veerkamp. 2013. Genomic selection: the option for new robustness traits? *Adv. Anim. Biosci.* 4(3):618-625.
- Canfora, E. E., J. W. Jocken, and E. E. Blaak. 2015. Short-chain fatty acids in control of body weight and insulin sensitivity. *Nat. Rev. Endocrinol.* 11:577-591.
- Cao, Y., Z. Zhao, J. Gruszczynska-Biegala, and A. Zolkiewska. 2003. Role of metalloprotease disintegrin ADAM 12 in determination of quiescent reserve cells during myogenic differentiation in vitro. *Mol. Cell Biol.* 23:6725-6738.
- Capper, J. L., R. A. Cady, and D. E. Bauman. 2009. The environmental impact of dairy production: 1944 compared with 2007. *Journal of animal science* 87(6):2160-2167.
- Clayton, D. G. 1996. Generalized linear mixed models. Pages 275-301 in Markov chain Monte Carlo in practice. Springer, New York, US.

Coles, C. A., J. Wadeson, M. I. Knight, L. M. Cafe, W. H. Johns, J. D. White, P. L. Greenwood, and M. B. McDonagh. 2014. A disintegrin and metalloprotease-12 is type I myofiber specific in *Bos taurus* and *Bos indicus* cattle. *J. Anim. Sci.* 92:1473-1483.

Connor, E. E. 2015. Invited review: improving feed efficiency in dairy production: challenges and possibilities. *Animal* 9:395-408.

Connor, E. E., J. L. Hutchison, H. D. Norman, K. M. Olson, C. P. Van Tassell, J. M. Leith, and R. L. Baldwin. 2013. Use of residual feed intake in Holsteins during early lactation shows potential to improve feed efficiency through genetic selection. *J. Anim. Sci.* 91(8):3978-3988.

Crossa, J., D. Jarquin, J. Franco, P. Perez-Rodriguez, J. Burgueno, C. Saint-Pierre, P. Vikram, C. Sansaloni, C. Petrolis, D. Akdemir, C. Sneller, M. Reynolds, M. Tattaris, T. Payne, C. Guzman, R. J. Pena, P. Wenzl, and S. Singh. 2016. Genomic prediction of gene bank wheat landraces. *G3 (Bethesda)* 6:1819-1834.

De Coninck, A., B. De Baets, D. Kourounis, F. Verbosio, O. Schenk, S. Maenhout, and J. Fostier. 2016. Needles: Toward large-scale genomic prediction with marker-by-environment interaction. *Genetics* 203:543-558.

de Haas, Y., M. P. Calus, R. F. Veerkamp, E. Wall, M. P. Coffey, H. D. Daetwyler, B. J. Hayes, and J. E. Pryce. 2012. Improved accuracy of genomic prediction for dry matter intake of dairy cattle from combined European and Australian data sets. *J. Anim. Sci.* 95(10):6103-6112.

de Haas, Y., J. E. Pryce, M. P. Calus, E. Wall, D. P. Berry, P. Lovendahl, N. Krattenmacher, F. Miglior, K. Weigel, D. Spurlock, K. A. Macdonald, B. Hulsege, and R. F. Veerkamp. 2015. Genomic prediction of dry matter intake in dairy cattle from an international data set consisting of research herds in Europe, North America, and Australasia. *J. Anim. Sci.* 98(9):6522-6534.

de los Campos, G., J. M. Hickey, R. Pong-Wong, H. D. Daetwyler, and M. P. L. Calus. 2013. Whole Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. *Genetics* 193:327-345.

deHaas, Y., J. E. Pryce, D. P. Berry, and R. F. Veerkamp. 2014. Genetic and genomic solutions to improve feed efficiency and reduce environmental impact of dairy cattle. 10th World Congress on Genetics Applied to Livestock Production. ASAS, August 17-22, 2014, Vancouver, CANADA.

Dempfle, A., A. Scherag, R. Hein, L. Beckmann, J. Chang-Claude, and H. Schafer. 2008. Gene-environment interactions for complex traits: definitions, methodological requirements and challenges. *Eur. J. Hum. Genet.* 16(10):1164-1172.

Do, D. N., A. B. Strathe, T. Ostensen, S. D. Pant, and H. N. Kadarmideen. 2014. Genome-wide association and pathway analysis of feed efficiency in pigs reveal candidate genes and pathways for residual feed intake. *Front Genet* 5:307.

- Fernando, R. L. and D. Garrick. 2013. Bayesian Methods Applied to GWAS, pp. 237-274 in *Genome-Wide Association Studies and Genomic Prediction*, edited by C. Gondro, J. van der Werf and B. Hayes. Humana Press, New York, U.S.
- Garcia-Ruiz, A., J. B. Cole, P. M. VanRaden, G. R. Wiggans, F. J. Ruiz-Lopez, and C. P. Van Tassell. 2016. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. *Proc. Natl. Acad. Sci. U.S.A.* 113(28):E3995-E4004.
- Gelman, A. 2006. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 1(3):515-533.
- Gelman, A., J. Hill, and M. Yajima. 2012. Why We (Usually) Don't Have to Worry About Multiple Comparisons. *Journal of Research on Educational Effectiveness* 5(2):189-211.
- Gilmour, A. R., B. J. Gogel, B. R. Cullis, and R. Thompson. 2009. *ASReml 3.0 User Guide*, 3.0 VSN International Ltd, Hemel Hempstead, HP1 1ES, UK [www.vsnl.co.uk](http://www.vsnl.co.uk).
- Goddard, M. E. and B. J. Hayes. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat. Rev. Genet.* 10(6):381-391.
- Goddard, M. E., K. E. Kemper, I. M. MacLeod, A. J. Chamberlain, and B. J. Hayes. 2016. Genetics of complex traits: prediction of phenotype, identification of causal polymorphisms and genetic architecture. *Proc Biol Sci.* 283(1835).
- Gonzalez-Recio, O., J. E. Pryce, M. Haile-Mariam, and B. J. Hayes. 2014. Incorporating heifer feed efficiency in the Australian selection index using genomic selection. *J. Dairy Sci.* 97:3883-3893.
- Gualdron Duarte, J. L., R. J. Cantet, R. O. Bates, C. W. Ernst, N. E. Raney, and J. P. Steibel. 2014. Rapid screening for phenotype-genotype associations by linear transformations of genomic evaluations. *BMC bioinformatics* 15:246.
- Guo, J. Z., H. Jorjani, and O. Carlborg. 2012. A genome-wide association study using international breeding-evaluation data identifies major loci affecting production traits and stature in the Brown Swiss cattle breed. *BMC Genet.* 13:82.
- Hagemann, M., A. Ndambi, T. Hemme, and U. Latacz-Lohmann. 2012. Contribution of milk production to global greenhouse gas emissions. An estimation based on typical farms. *Environ Sci Pollut Res Int* 19(2):390-402.
- Hardie, L. C., L. E. Armentano, R. D. Shaver, M. J. VandeHaar, D. M. Spurlock, C. Yao, S. J. Bertics, F. E. Contreras-Govea, and K. A. Weigel. 2015. Considerations when combining data from multiple nutrition experiments to estimate genetic parameters for feed efficiency. *J. Anim. Sci.* 98(4):2727-2737.
- Hawken, R. J., Y. D. Zhang, M. R. Fortes, E. Collis, W. C. Barris, N. J. Corbet, P. J. Williams, G. Fordyce, R. G. Holroyd, J. R. Walkley, W. Barendse, D. J. Johnston, K. C. Prayaga, B. Tier, A.

- Reverter, and S. A. Lehnert. 2012. Genome-wide association studies of female reproduction in tropically adapted beef cattle. *J. Anim. Sci.* 90(5):1398-1410.
- Hayes, B. J., H. D. Daetwyler, and M. E. Goddard. 2016. Models for Genome x Environment Interaction: Examples in Livestock. Vol. 56. *Crop Sci.*
- Hayes, B. J., H. A. Lewin, and M. E. Goddard. 2013. The future of livestock breeding: genomic selection for efficiency, reduced emissions intensity, and adaptation. *Trends Genet.* 29(4):206-214.
- Henderson, C. R. 1976. A Simple Method for Computing the Inverse of a Numerator Relationship Matrix Used in Prediction of Breeding Values. *Biometrics* 32(1):69-83.
- Henderson, C. R. 1977. Best linear unbiased prediction of breeding values not in model for records. 60:783-787.
- Henderson, C. R. 1985. Best linear unbiased prediction of nonadditive genetic merits in noninbred populations. *J. Anim. Sci.* 60:111-117.
- Herd, R. M. and P. F. Arthur. 2009. Physiological basis for residual feed intake. *J. Anim. Sci.* 87(14):E64-71.
- Jarquín, D., J. Crossa, X. Lacaze, P. Du Cheyron, J. Daucourt, J. Lorgeou, F. Piraux, L. Guerreiro, P. Perez, M. Calus, J. Burgueno, and G. de los Campos. 2014. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor. Appl. Genet.* 127(3):595-607.
- Johnson, R. C., G. W. Nelson, J. L. Troyer, J. A. Lautenberger, B. D. Kessing, C. A. Winkler, and S. J. O'Brien. 2010. Accounting for multiple comparisons in a genome-wide association study (GWAS). *BMC Genomics* 11:724.
- Kang, H. M., J. H. Sul, S. K. Service, N. A. Zaitlen, S. Y. Kong, N. B. Freimer, C. Sabatti, and E. Eskin. 2010. Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* 42:348-U110.
- Kawaguchi, N., C. Sundberg, M. Kveiborg, B. Moghadaszadeh, M. Asmar, N. Dietrich, C. K. Thodeti, F. C. Nielsen, P. Moller, A. M. Mercurio, R. Albrechtsen, and U. M. Wewer. 2003. ADAM12 induces actin cytoskeleton and extracellular matrix reorganization during early adipocyte differentiation by regulating beta1 integrin function. *J. Cell Sci.* 116:3893-3904.
- Kennedy, B. W., L. R. Schaeffer, and D. A. Sorensen. 1988. Genetic Properties of Animal Models. *Journal of Dairy Science* 71:17-26.
- Kennedy, B. W., J. H. J. Vanderwerf, and T. H. E. Meuwissen. 1993. Genetic and statistical properties of residual feed-intake. *J. Anim. Sci.* 71(12):3239-3250.
- Kim, Y. M., J. Kim, S. C. Heo, S. H. Shin, E. K. Do, D. S. Suh, K. H. Kim, M. S. Yoon, T. G. Lee, and J. H. Kim. 2012. Proteomic Identification of ADAM12 as a Regulator for TGF-beta 1-

Induced Differentiation of Human Mesenchymal Stem Cells to Smooth Muscle Cells. *Plos ONE* 7:e40820.

Kizilkaya, K. and R. J. Tempelman. 2005. A general approach to mixed effects modeling of residual variances in generalized linear mixed models. *Genet. Sel. Evol.* 37:31-56.

Koch, R. M., K. E. Gregory, D. Chambers, and L. A. Swiger. 1963. Efficiency of Feed Use in Beef Cattle. *J. Anim. Sci.* 22(2):486-&.

Kolmodin, R. 2003. Reaction norms for the study of genotype by environment interaction in animal breeding. *Sveriges lantbruksuniv., Acta Universitatis agriculturae Sueciae. Agraria*, 1401-6249.

Kutner, M. H., C. J. Nachtsheim, J. Neter, and W. Li. 2005. *Applied Linear Statistical Models*, 5th Edition. McGraw-Hill Irwin, New York.

Lee, S. H., B. H. Choi, D. Lim, C. Gondro, Y. M. Cho, C. G. Dang, A. Sharma, G. W. Jang, K. T. Lee, D. Yoon, H. K. Lee, S. H. Yeon, B. S. Yang, H. S. Kang, and S. K. Hong. 2013. Genome-wide association study identifies major loci for carcass weight on BTA14 in Hanwoo (Korean cattle). *PLoS ONE* 8:e74677.

Lillehammer, M., M. Arnyasi, S. Lien, H. G. Olsen, E. Sehested, J. Odegard, and T. H. Meuwissen. 2007a. A genome scan for quantitative trait locus by environment interactions for production traits. *J. Dairy Sci.* 90:3482-3489.

Lillehammer, M., B. J. Hayes, T. H. Meuwissen, and M. E. Goddard. 2009. Gene by environment interactions for production traits in Australian dairy cattle. *J. Dairy Sci.* 92:4008-4017.

Lillehammer, M., J. Odegard, and T. H. Meuwissen. 2007b. Random regression models for detection of gene by environment interaction. *Genet. Sel. Evol.* 39:105-121.

Littell, R. C., W. W. Stroup, R. J. Freund, and SAS Institute. 2002. *SAS for linear models*. 4th ed. SAS Institute ;Wiley, Cary, N.C. New York.

Lopez-Cruz, M., J. Crossa, D. Bonnett, S. Dreisigacker, J. Poland, J. L. Jannink, R. P. Singh, E. Autrique, and G. de los Campos. 2015. Increased prediction accuracy in wheat breeding trials using a marker x environment interaction genomic selection model. *G3 (Bethesda)* 5:569-582.

Lopez-Romero, P., R. Rekaya, and M. J. Carabano. 2003. Assessment of homogeneity vs. heterogeneity of residual variance in random regression test-day models in a Bayesian analysis. *J. Anim. Sci.* 86(10):3374-3385.

Lu, D., S. Miller, M. Sargolzaei, M. Kelly, G. Vander Voort, T. Caldwell, Z. Wang, G. Plastow, and S. Moore. 2013. Genome-wide association analyses for growth and feed efficiency traits in beef cattle. *J. Anim Sci.* 91(8):3612-3633.

- Lu, Y., M. J. Vandehaar, D. M. Spurlock, K. A. Weigel, L. E. Armentano, C. R. Staples, E. E. Connor, Z. Wang, N. M. Bello, and R. J. Tempelman. 2015. An alternative approach to modeling genetic merit of feed efficiency in dairy cattle. *J. Anim. Sci.* 98(9):6535-6551.
- Lu, Y., M. J. VandeHaar, D. M. Spurlock, K. A. WEigel, L. E. Armentano, C. R. Staples, E. E. Connor, Z. Wang, M. Coffey, R. F. Veerkamp, Y. de Haas, and R. J. Tempelman. 2017a. Modeling genetic and nongenetic variation of feed efficiency and its partial relationships btween component traits as a function of management and environmental factors. *Journal of Dairy Science* 100((in press)).
- Lu, Y., M. J. VandeHaar, D. M. Spurlock, K. A. Weigel, E. E. Connor, M. Coffey, R. F. Veerkamp, Y. D. Haas, C. R. Staples, Z. Wang, M. D. Hanigan, and R. J. Tempelman. 2017b. Genome wide association analyses based on alternative strategies for modeling feed efficiency. *Journal of Dairy Science* (in preparation).
- Ludovico, A., V. B. Maion, D. E. Bronkhorst, F. D. C. R. Grecco, L. F. C. D. Cunha, I. Y. Mizubuti, K. M. de Almeida, M. D. Ludovico, and E. H. W. de Santana. 2015. Losses in milk production and quality due to milk somatic cell count and heat stress of Holsteins cows in temperate climate. *Semin-Cienc Agrar* 36(5):3455-3470.
- Manafiazar, G., T. McFadden, L. Goonewardene, E. Okine, J. Basarab, P. Li, and Z. Wang. 2013. Prediction of residual feed intake for first-lactation dairy cows using orthogonal polynomial random regression. *J. Anim. Sci.* 96(12):7991-8001.
- Manzanilla-Pech, C. I., R. F. Veerkamp, R. J. Tempelman, M. L. van Pelt, K. A. Weigel, M. VandeHaar, T. J. Lawlor, D. M. Spurlock, L. E. Armentano, C. R. Staples, M. Hanigan, and Y. De Haas. 2016. Genetic parameters between feed-intake-related traits and conformation in 2 separate dairy populations--the Netherlands and United States. *J. Anim. Sci.* 99(1):443-457.
- Martin-Collado, D., T. J. Byrne, P. R. Amer, B. F. S. Santos, M. Axford, and J. E. Pryce. 2015. Analyzing the heterogeneity of farmers' preferences for improvements in dairy cow traits using farmer typologies. *J. Anim. Sci.* 98(6):4148-4161.
- Maruyama, K. 1991. The siscovery of Adenosine-Triphosphate and the establishment of its structure. *J. Hist. Biol.* 24:145-154.
- Mayr, J. A., V. Havlickova, F. Zimmermann, I. Magler, V. Kaplanova, P. Jesina, A. Pecinova, H. Nuskova, J. Koch, W. Sperl, and J. Houstek. 2010. Mitochondrial ATP synthase deficiency due to a mutation in the ATP5E gene for the F1 epsilon subunit. *Hum. Mol. Genet.* 19:3430-3439.
- Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819-1829.
- Milliken, G. A. and D. E. Johnson. 2009. *Analysis of Messy Data Volume 1: Designed Experiments*. CRC Press, Boca Raton,FL.

- Moser, G., S. H. Lee, B. J. Hayes, M. E. Goddard, N. R. Wray, and P. M. Visscher. 2015. Simultaneous Discovery, Estimation and Prediction Analysis of Complex Traits Using a Bayesian Mixture Model. *Plos Genet* 11(4):e1004969.
- Nkrumah, J. D., E. L. Sherman, C. Li, E. Marques, D. H. Crews, Jr., R. Bartusiak, B. Murdoch, Z. Wang, J. A. Basarab, and S. S. Moore. 2007. Primary genome scan to identify putative quantitative trait loci for feedlot growth rate, feed intake, and feed efficiency of beef cattle. *J. Anim. Sci.* 85:3170-3181.
- NRC. 1981. Effect of environment on nutrient requirements of domestic animals. Natl. Acad. Press, Washington, D.C.
- NRC. 2001. Nutrient requirements of dairy cattle. 7th rev. ed. Natl. Acad. Press, Washington, D.C.
- Olivieri, B. F., A. S. M. Cesar, M. L. do Nascimento, A. S. Chaves, P. C. Tizioto, R. R. Tullio, D. P. D. Lanna, A. N. Rosa, T. S. Sonstegard, G. B. Mourao, J. M. Reecy, D. J. Garrick, M. A. Mudadu, L. L. Coutinho, and L. C. A. Regitano. 2014. Identification of genomic regions associated with feed efficiency in Nelore cattle. *BMC Genet.* 15:100.
- Olivieri, B. F., M. E. Mercadante, J. N. Cyrillo, R. H. Branco, S. F. Bonilha, L. G. de Albuquerque, R. M. Silva, and F. Baldi. 2016. Genomic regions associated with feed efficiency indicator traits in an experimental Nelore cattle population. *PLoS ONE* 11(10):e0164390.
- Ou, Z. N., R. J. Tempelman, J. P. Steibel, C. W. Ernst, R. O. Bates, and N. M. Bello. 2016. Genomic Prediction Accounting for Residual Heteroskedasticity. *G3-Genes Genom Genet* 6(1):1-13.
- Pahl, R. and H. Schafer. 2010. PERMORY: an LD-exploiting permutation test algorithm for powerful genome-wide association testing. *Bioinformatics* 26:2093-2100.
- Plante, Y., J. P. Gibson, J. Nadesalingam, H. Mehrabani-Yeganeh, S. Lefebvre, G. Vandervoort, and G. B. Jansen. 2001. Detection of quantitative trait loci affecting milk production traits on 10 chromosomes in Holstein cattle. *J. Dairy Sci.* 84:1516-1524.
- Plummer, M., N. , K. C. Best, and K. Vines. 2006. CODA: convergence diagnosis and output analysis for MCMC. Vol. 6. R News.
- Pollak, E. J., J. Vanderwerf, and R. L. Quaas. 1984. Selection bias and multiple trait evaluation. 67:1590-1595.
- Pourahmadi, M. 2007. Cholesky decompositions and estimation of a covariance matrix: orthogonality of variance-correlation parameters. *Biometrika* 94(4):1006-1013.
- Pourahmadi, M., M. J. Daniels, and T. Park. 2007. Simultaneous modelling of the Cholesky decomposition of several covariance matrices. *J Multivariate Anal* 98(3):568-587.

Pryce, J. E., O. Gonzalez-Recio, G. Nieuwhof, W. J. Wales, M. P. Coffey, B. J. Hayes, and M. E. Goddard. 2015. Hot topic: Definition and implementation of a breeding value for feed efficiency in dairy cows. *J. Anim. Sci.* 98(10):7340-7350.

Pryce, J. E., W. J. Wales, Y. de Haas, R. F. Veerkamp, and B. J. Hayes. 2014. Genomic selection for feed efficiency in dairy cattle. *Animal : an international journal of animal bioscience* 8(1):1-10.

Rekaya, R. and S. E. Aggrey. 2015. Genetic properties of residual feed intakes for maintenance and growth and the implications of error measurement. *J. Anim Sci.* 93(3):944-948.

Richardson, E. C. and R. M. Herd. 2004. Biological basis for variation in residual feed intake in beef cattle. 2. Synthesis of results following divergent selection. *Aust. J. Exp. Agric.* 44(4-5):431-440.

Rolf, M. M., J. F. Taylor, R. D. Schnabel, S. D. McKay, M. C. McClure, S. L. Northcutt, M. S. Kerley, and R. L. Weaber. 2012. Genome-wide association analysis for feed efficiency in Angus cattle. *Anim. Genet.* 43:367-374.

Ronnegard, L., M. Felleki, F. Fikse, H. A. Mulder, and E. Strandberg. 2010. Genetic heterogeneity of residual variance - estimation of variance components using double hierarchical generalized linear models. *GENET. SEL. EVOL.* 42(8).

Ronnegard, L., M. Felleki, W. F. Fikse, H. A. Mulder, and E. Strandberg. 2013. Variance component and breeding value estimation for genetic heterogeneity of residual variance in Swedish Holstein dairy cattle. *J. Anim. Sci.* 96(4):2627-2636.

Ryu, J. and C. Lee. 2016. Genetic association of marbling score with intragenic nucleotide variants at selection signals of the bovine genome. *Animal : an international journal of animal bioscience* 10(4):566-570.

Saatchi, M., J. E. Beever, J. E. Decker, D. B. Faulkner, H. C. Freetly, S. L. Hansen, H. Yampara-Iquise, K. A. Johnson, S. D. Kachman, M. S. Kerley, J. Kim, D. D. Loy, E. Marques, H. L. Neibergs, E. J. Pollak, R. D. Schnabel, C. M. Seabury, D. W. Shike, W. M. Snelling, M. L. Spangler, R. L. Weaber, D. J. Garrick, and J. F. Taylor. 2014a. QTLs associated with dry matter intake, metabolic mid-test weight, growth and feed efficiency have little overlap across 4 beef cattle studies. *BMC genomics* 15:1004.

Saatchi, M., R. D. Schnabel, J. F. Taylor, and D. J. Garrick. 2014b. Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. *BMC genomics* 15:442.

Santana, M. H. A., Y. T. Utsunomiya, H. H. R. Neves, R. C. Gomes, J. F. Garcia, H. Fukumasu, S. L. Silva, G. A. Oliveira, P. A. Alexandre, P. R. Leme, R. A. Brassaloti, L. L. Coutinho, T. G. Lopes, F. V. Meirelles, J. P. Eler, and J. B. S. Ferraz. 2014. Genome-wide association analysis of feed intake and residual feed intake in Nellore cattle. *Bmc Genet* 15.

Saviotto, D., D. P. Berry, and N. C. Friggens. 2014. Towards an improved estimation of the biological components of residual feed intake in growing cattle. *J. Anim. Sci.* 92:467-476.

- Searle, S. R. 1982. Matrix Algebra Useful for Statistics. First ed. Wiley-Interscience.
- Searle, S. R., G. Casella, and C. E. McCulloch. 1992. Variance Components. John Wiley and Sons, New York.
- Serão, N. V., E. D. Mauch, S. K. Onteru, M. F. Rothschild, and J. C. M. Dekkers. 2016. Genome wide association study for residual feed intake and component traits of feed efficiency in pigs divergently selected for residual feed intake. 2016 Joint Annual Meeting. ASAS, July 19-23, 2016, Salt Lake City, UT.
- Sherman, E. L., J. D. Nkrumah, C. Li, R. Bartusiak, B. Murdoch, and S. S. Moore. 2009. Fine mapping quantitative trait loci for feed intake and feed efficiency in beef cattle. *J. Anim. Sci.* 87:37-45.
- Shirali, M., V. H. Nielsen, S. H. Moller, and J. Jensen. 2015. Longitudinal analysis of residual feed intake and BW in mink using random regression with heterogeneous residual variance. *Animal : an international journal of animal bioscience* 9(10):1597-1604.
- Sorensen, D. and D. Gianola. 2002. Likelihood, Bayesian and MCMC Methods in Quantitative Genetics. Statistics for Biology and Health. Springer Verlag, New York.
- Sorensen, D. and R. Waagepetersen. 2003. Normal linear models with genetically structured residual variance heterogeneity: a case study. *Genet. Res.* 82:207-222.
- Sova, A. D., S. J. LeBlanc, B. W. McBride, and T. J. DeVries. 2013. Associations between herd-level feeding management practices, feed sorting, and milk production in freestall dairy farms. *J. Anim. Sci.* 96(7):4759-4770.
- Spiegelhalter, D. J., N. G. Best, B. R. Carlin, and A. van der Linde. 2002. Bayesian measures of model complexity and fit. *J Roy Stat Soc B* 64:583-616.
- Spurlock, D. M., J. C. Dekkers, R. Fernando, D. A. Koltes, and A. Wolc. 2012. Genetic parameters for energy balance, feed efficiency, and related traits in Holstein cattle. *J. Dairy Sci.* 95:5393-5402.
- Steinfeld, H., P. Gerber, T. D. Wassenaar, V. Castel, M. R. M, and C. d. Haan. 2006. Livestock's long shadow: environmental issues and options. Roma.
- Strathe, A. B., T. Mark, B. Nielsen, D. N. Do, H. N. Kadarmideen, and J. Jensen. 2014. Deriving genomic breeding values for residual feed intake from covariance functions of random regression models. In Proc. 10th World Congress Genet. Appl. Livest. Prod. ASAS, August 17-22, 2014, Vancouver, BC, Canada.
- Taylor, J. F., K. H. Taylor, and J. E. Decker. 2016. Holsteins are the genomic selection poster cows. *Proc. Natl. Acad. Sci. U.S.A.* 113(28):7690-7692.

Tempelman, R. J. 2015. Statistical and Computational Challenges in Whole Genome Prediction and Genome-Wide Association Analyses for Plant and Animal Breeding. *Journal of Agricultural, Biological, and Environmental Statistics* 20(4):442-466.

Tempelman, R. J., D. M. Spurlock, M. Coffey, R. F. Veerkamp, L. E. Armentano, K. A. Weigel, Y. de Haas, C. R. Staples, E. E. Connor, Y. Lu, and M. J. Vandehaar. 2015. Heterogeneity in genetic and non-genetic variation and energy sink relationships for residual feed intake across research stations and countries. *J. Anim. Sci.* 98(3):(in press).

Tort, F., M. Del Toro, W. Lissens, J. Montoya, M. Fernandez-Burriel, A. Font, N. Bujan, A. Navarro-Sastre, E. Lopez-Gallardo, J. A. Arranz, E. Riudor, P. Briones, and A. Ribes. 2011. Screening for nuclear genetic defects in the ATP synthase-associated genes TMEM70, ATP12 and ATP5E in patients with 3-methylglutaconic aciduria. *Clin. Genet.* 80:297-300.

Valdar, W., L. C. Solberg, D. Gauguier, W. O. Cookson, J. N. P. Rawlins, R. Mott, and J. Flint. 2006. Genetic and environmental effects on complex traits in mice. *Genetics* 174(2):959-984.

Vallimont, J. E., C. D. Dechow, J. M. Daubert, M. W. Dekleva, J. W. Blum, C. M. Barlieb, W. Liu, G. A. Varga, A. J. Heinrichs, and C. R. Baumrucker. 2010. Genetic parameters of feed intake, production, body weight, body condition score, and selected type traits of Holstein cows in commercial tie-stall barns. *J. Anim. Sci.* 93(10):4892-4901.

Vallimont, J. E., C. D. Dechow, J. M. Daubert, M. W. Dekleva, J. W. Blum, C. M. Barlieb, W. Liu, G. A. Varga, A. J. Heinrichs, and C. R. Baumrucker. 2011. Heritability of gross feed efficiency and associations with yield, intake, residual intake, body weight, and body condition score in 11 commercial Pennsylvania tie stalls. *J. Anim. Sci.* 94(4):2108-2113.

VandeHaar, M. J., L. E. Armentano, K. Weigel, D. M. Spurlock, R. J. Tempelman, and R. Veerkamp. 2016. Harnessing the genetics of the modern dairy cow to continue improvements in feed efficiency. *J. Anim. Sci.* 99(6):4941-4954.

VandeHaar, M. J. and N. St-Pierre. 2006. Major advances in nutrition: relevance to the sustainability of the dairy industry. *J. Anim. Sci.* 89(4):1280-1291.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *J. Anim. Sci.* 91(11):4414-4423.

Vazquez, A. I., Y. Veturi, M. Behring, S. Shrestha, M. Kirst, M. F. R. Resende, and G. de los Campos. 2016. Increased proportion of variance explained and prediction accuracy of survival of breast cancer patients with use of whole-genome multiomic profiles. *Genetics* 203:1425-1438.

Veerkamp, R. F., J. E. P. , D. Spurlock, D. Berry, M. Coffey, P. Løvendahl, R. v. d. Linde, J. Bryant, F. Miglior, Z. Wang, M. Winters, N. Krattenmacher, N. Charfeddine, J. Pedersen, and Y. d. Haas. 2013. Selection on Feed Intake or Feed Efficiency: A Position Paper from gDMI Breeding Goal Discussions. *Interbull Bulletin* No. 47. Nantes, France, August 23 - 25, 2013.

Veerkamp, R. F., M. Coffey, D. Berry, Y. de Haas, E. Strandberg, H. Bovenhuis, M. Calus, and E. Wall. 2012. Genome-wide associations for feed utilisation complex in primiparous Holstein-

Friesian dairy cows from experimental research herds in four European countries. *Animal : an international journal of animal bioscience* 6(11):1738-1749.

Veerkamp, R. F., J. K. Oldenbroek, H. J. Van der Gaast, and J. H. Van der Werf. 2000. Genetic correlation between days until start of luteal activity and milk yield, energy balance, and live weights. *J. Anim. Sci.* 83(3):577-583.

Wang, H., I. Misztal, I. Aguilar, A. Legarra, and W. M. Muir. 2012. Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet Res.* 94:73-83.

Waters, S. M., M. S. McCabe, D. J. Howard, L. Giblin, D. A. Magee, D. E. MacHugh, and D. P. Berry. 2011. Associations between newly discovered polymorphisms in the *Bos taurus* growth hormone receptor gene and performance traits in Holstein-Friesian dairy cattle. *Anim. Genet.* 42:39-49.

Wathes, D. C., Z. Cheng, N. Bourne, V. J. Taylor, M. P. Coffey, and S. Brotherstone. 2007. Differences between primiparous and multiparous dairy cows in the inter-relationships between metabolic traits, milk yield and body condition score in the periparturient period. *Domest. Anim. Endocrin.* 33:203-225.

Wiggans, G. R., T. A. Cooper, P. M. VanRaden, C. P. Van Tassell, D. M. Bickhart, and T. S. Sonstegard. 2016. Increasing the number of single nucleotide polymorphisms used in genomic evaluation of dairy cattle. *J. Dairy Sci.* 99:4504-4511.

Wray, N. R., J. Yang, M. E. Goddard, and P. M. Visscher. 2010. The genetic interpretation of area under the ROC curve in genomic profiling. *PLoS Genet.* 6:e1000864.

Yang, J., N. A. Zaitlen, M. E. Goddard, P. M. Visscher, and A. L. Price. 2014. Advantages and pitfalls in the application of mixed-model association methods. *Nat Genet* 46:100-106.

Yang, W., C. Chen, J. P. Steibel, C. W. Ernst, R. O. Bates, L. Zhou, and R. J. Tempelman. 2015. A comparison of alternative random regression and reaction norm models for whole genome predictions. *J. Anim. Sci.* 93(6):2678-2692.

Yang, Y., O. F. Christensen, and D. Sorensen. 2011. Use of genomic models to study genetic control of environmental variance. *Genet. Res.* 93:125-138.

Yao, C., D. M. Spurlock, L. E. Armentano, C. D. Page, Jr., M. J. VandeHaar, D. M. Bickhart, and K. A. Weigel. 2013. Random Forests approach for identifying additive and epistatic single nucleotide polymorphisms associated with residual feed intake in dairy cattle. *J. Anim. Sci.* 96(10):6716-6729.

Zom, R. L. G., G. Andre, and A. M. van Vuuren. 2012. Development of a model for the prediction of feed intake by dairy cows: prediction of feed intake. *Livest Sci* 143:43-57.