PEDRO VITAL BRASIL RAMOS

GENOMIC ANALYSIS FOR FEED EFFICIENCY TRAITS IN BEEF CATTLE AND ALGORITHMS TO APPROXIMATE ACCURACIES FOR SINGLE-STEP GENOMIC BEST LINEAR UNBIASED PREDICTOR

Thesis submitted to the Animal Science Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Doctor Scientiae*.

Adviser: Renata Veroneze

VIÇOSA - MINAS GERAIS
2023

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T
Ramos, Pedro Vital Brasil, 1994-
R175g Genomic analysis for feed efficiency traits in beef cattle and
2023 algorithms to approximate accuracies for single-step genomic best linear unbiased predictor / Pedro Vital Brasil Ramos. – Viçosa, MG, 2023.
1 tese eletrônica (71 f.): il. (algumas color.).

Texto em inglês.
Orientador: Renata Veroneze.
Tese (doutorado) - Universidade Federal de Viçosa, Departamento de Zootecnia, 2023.
Inclui bibliografia.
DOI: https://doi.org/10.47328/ufvbbt.2023.721
Modo de acesso: World Wide Web.

1. Bovinos de corte - Alimentação e rações. 2. Bovinos de corte - Genética. 3. Análise de regressão. 4. Medidas físicas. I. Veroneze, Renata, 1984-. II. Universidade Federal de Viçosa. Departamento de Zootecnia. Programa de Pós-Graduação em Zootecnia. III. Título.

CDD 22. ed. 636.2130855

Bibliotecário(a) responsável: Alice Regina Pinto Pires CRB-6/2523

PEDRO VITAL BRASIL RAMOS

GENOMIC ANALYSIS FOR FEED EFFICIENCY TRAITS IN BEEF CATTLE AND ALGORITHMS TO APPROXIMATE ACCURACIES FOR SINGLE-STEP GENOMIC BEST LINEAR UNBIASED PREDICTOR

Thesis submitted to the Animal Science Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Doctor Scientiae*.

.

APPROVED: October 6, 2023.

Assent:

Documento assinado digitalmente
PEDRO VITAL BRASIL RAMOS
Data: 08/12/2023 13:21:17-0300
Verifique em https://validar.iti.gov.br

Pedro Vital Brasil Ramos
Author

Documento assinado digitalmente
RENATA VERONEZE
Data: 08/12/2023 13:38:51-0300
Verifique em https://validar.iti.gov.br

Renata Veroneze
Adviser

I dedicate this work to
Fabyano Fonseca e Silva and
Luiz Otávio Campos da Silva

ACKNOWLEDGEMENTS

Firstly, I would like to express my gratitude to the Universidade Federal de Viçosa and the Graduate Program in Animal Science for providing me the opportunity to pursue this course. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001. I extend my appreciation to the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brazil) for granting the scholarship. Additionally, I would like to thank the University of Georgia for their support during the period of my sandwich Ph.D. program.

I am deeply thankful to Fabyano Fonseca e Silva and Luiz Otávio Campos da Silva for their friendship, teachings, and all their contributions during my academic and professional journey. I will be forever grateful for the impact they've had on my life.

I express my sincere appreciation to my advisor, Renata Veroneze, and my co-advisor, Gilberto Romeiro de Oliveira Menezes, for their unwavering support, especially during the challenging moments. Their contributions were instrumental in completing my Ph.D. and exploring new professional opportunities.

I would like to express my gratitude to Daniela Lourenco for believing in my potential and providing me with the opportunity to join the University of Georgia's research group. Furthermore, I am thankful for her guidance and support, which played a pivotal role in my professional development.

I extend my heartfelt thanks to Andre Garcia and Delvan Alves da Silva for their friendship and support during my Ph.D.

I am profoundly grateful to my mother, Adriana Vital Brasil, and my father, Roberto do Carmo Ramos, for their unwavering support throughout every step of my journey, dedicating their efforts to my success. Their presence makes me resilient in pursuing my dreams. I also appreciate my brother, Matheus Vital Brasil Ramos, for his friendship and invaluable advice, which undoubtedly contributed to my personal and professional growth.

Special thanks to my girlfriend, Luisa Crivelli Alvarenga, for her companionship. Your friendship, encouragement, and love provide me with the strength and determination to pursue my goals.

ABSTRACT

RAMOS, Pedro Vital Brasil, D.Sc., Universidade Federal de Viçosa, October, 2023. **Genomic analysis for feed efficiency traits in beef cattle and algorithms to approximate accuracies for single-step genomic best linear unbiased predictor.** Adviser: Renata Veroneze.

Feed efficiency plays a major role in the overall profitability and sustainability of the beef cattle industry, as it is directly related to the reduction of the animals' input requirements. Traditionally, the average daily feed intake and weight gain are used to calculate feed efficiency traits. The phenotypes are collected in feedlot systems with electronic feed bunks, requiring high financial costs, which limits the data collection. Therefore, new methodologies capable of predicting breeding values with high accuracy for these traits become crucial for selecting efficient animals. In this sense, feed efficiency traits can be analyzed longitudinally using random regression models (RRM), which allow fitting random genetic and environmental effects over time by considering the covariance pattern between the daily records. RRM can be a feasible alternative to evaluate genetic parameters as a function of the evaluation period and increase the accuracy of estimated breeding values (EBV) for feed efficiency. The accuracy of EBV is calculated to help make selection decisions, representing the correlation between the true breeding value and EBV. The theoretical accuracy can be calculated based on the prediction error variances obtained from the diagonal of the inverse of the left-hand side (LHS) of the mixed model equations (MME). However, inverting the LHS is not computationally feasible for an extensive system of equations, especially if genomic information is available. Thus, different algorithms to approximate accuracies have been proposed. Therefore, the first objective of this study was to propose genomic evaluations for dry matter intake (DMI), body weight gain (BWG), residual feed intake (RFI), and residual weight gain (RWG) data collected during an 84-day feedlot test period via random regression models and evaluate the genetic parameters behavior for feed efficiency traits and their implication for new selection strategies. Genetic parameters and genomic breeding values (GEBV) were estimated by random regression models under ssGBLUP for Nellore cattle using orthogonal Legendre polynomials and B-spline. The random regression model using linear B-splines proved a feasible alternative for the genomic evaluation of feed efficiency. Heritability estimates of DMI, RFI, BWG, and RWG indicate enough additive genetic variance to achieve a moderate response to selection. Based on genetic correlations and ranking comparisons between the test days, a new selection strategy can be adopted by reducing the performance test to 56 days for DMI and RFI selection and 77 days for BWG and RWG selection. The second objective of this study was to compare the approximated accuracies from two algorithms implemented in the BLUPF90 suite of programs and compare the approximated accuracies from the two algorithms against the exact accuracy based on the inversion of the LHS of MME. Algorithm 1 approximates accuracies based on the diagonal of the genomic relationship matrix $(\mathbf{G})$. In turn, Algorithm 2 uses block sparse inversion of $\mathbf{G}^{-1}$. The Data were provided by the American Angus Association and included three datasets of growth, carcass, and marbling traits. For the genomic evaluations, a multi-trait model was applied to the datasets. To ensure the feasibility of inverting the LHS of the MME, a subset of data under single-trait models was used to compare approximated and exact accuracies. Accuracies from Algorithm 2 presented a higher correlation with the exact accuracies than from Algorithm 1. Additionally, Algorithm 2's accuracies were, in general, closer to the exact accuracies according to the mean square error and demonstrated similar changes in behavior when adding new genotyped animals to the analysis. In summary, the random regression model is a feasible alternative for the genomic evaluation of feed efficiency in Nellore cattle and the Algorithm 2 is more suitable for approximating accuracies of GEBV and should be used for routine ssGBLUP evaluations.

Keywords: Feed intake; Longitudinal data; B-splines; Accuracy approximation; Genomic evaluation.

RESUMO

RAMOS, Pedro Vital Brasil, D.Sc., Universidade Federal de Viçosa, outubro de 2023. **Análise genômica para características de eficiência alimentar em bovinos de corte e algoritmos para aproximação de acurácia para melhor preditor linear não viesado genômico de passo único**. Orientadora: Renata Veroneze.

A eficiência alimentar desempenha um papel fundamental na lucratividade e na sustentabilidade da indústria de gado de corte, pois está diretamente relacionada à redução dos requisitos de insumos dos animais. Tradicionalmente, a média diária de consumo alimentar e ganho de peso são usados para calcular as características de eficiência alimentar. Os fenótipos são coletados em sistemas de confinamento com cochos de alimentação eletrônicos, o que demanda um alto custo financeiro e limita a coleta de dados. Portanto, novas metodologias capazes de predizer com alta acurácia os valores genéticos para essas características se tornam cruciais para selecionar animais eficientes. Nesse sentido, as características de eficiência alimentar podem ser analisadas longitudinalmente através de modelos de regressão aleatória (MRA), que permitem ajustar efeitos genéticos e ambientais aleatórios ao longo do tempo, considerando o padrão de covariância entre os registros diários. Assim, o MRA representa uma alternativa viável para avaliar os parâmetros genéticos em função do tempo e aumentar a acurácia das estimativas de valores genéticos (EBV) para eficiência alimentar. A acurácia dos EBV é calculada para auxiliar nas decisões de seleção, representando a correlação entre o valor genético verdadeiro e o estimado. A acurácia teórica pode ser calculada com base na variância do erro de predição obtidas através da diagonal da inversa da matriz de coeficientes das equações do modelo misto (EMM). No entanto, inverter a matriz de coeficientes não é computacionalmente viável para um sistema extenso de equações, especialmente se informações genômicas estiverem disponíveis. Desta forma, diferentes algoritmos foram propostos para aproximar acurácias. O primeiro objetivo deste estudo foi propor avaliações genômicas para dados de consumo de matéria seca (CMS), ganho de peso corporal (GPC), consumo alimentar residual (CAR) e ganho de peso residual (GPR) coletados durante um período de teste em confinamento de 84 dias por meio de modelos de regressão aleatória e avaliar o comportamento dos parâmetros genéticos destas características e suas implicações para novas estratégias de seleção. Parâmetros genéticos e valores genéticos genômicos (GEBV) foram estimados por meio de modelos de regressão aleatória via ssGBLUP para bovinos Nelore, utilizando polinômios ortogonais de Legendre e B-splines. O modelo de regressão aleatória usando B-splines lineares mostrou-se a alternativa mais viável para a avaliação genômica de eficiência alimentar, de acordo com a qualidade de ajuste. As estimativas de herdabilidade de CMS, CAR, GPC e GPR indicam variância genética aditiva suficiente para moderada resposta a seleção. Com base nas correlações genéticas e nas comparações de ranking entre os dias de teste, uma nova estratégia de seleção pode ser adotada, reduzindo o período de teste de desempenho para 56 dias para a seleção de CMS e CAR e 77 dias para a seleção de GPC e GPR. O segundo objetivo deste estudo foi comparar as acurácias aproximadas de dois algoritmos implementados no pacote de programas BLUPF90 e compará-las com a acurácia exata obtida com base na inversa da matriz de coeficientes da EMM. O Algoritmo 1 aproxima as acurácias com base na diagonal da matriz de parentesco genômico ($\mathbf{G}$). Por sua vez, o Algoritmo 2 utiliza a inversão esparsa por blocos de $\mathbf{G}^{-\mathbf{1}}$. Os dados foram fornecidos pela Associação Americana do Angus e incluíram de dados de características de crescimento, carcaça e marmoreio. Para as avaliações genômicas, um modelo multicaracterístico foi utilizado para os três conjuntos de dados. Para inversão da matriz de coeficientes da EMM, um modelo uni característico foi aplicado a um subconjunto de dados para comparar as acurácias aproximadas e exatas. As acurácias obtidas com o Algoritmo 2 apresentaram uma correlação superior com as acurácias exatas em comparação com as do Algoritmo 1. Além disso, as acurácias do Algoritmo 2 foram, no geral, mais próximas das acurácias exatas, conforme evidenciado pelo erro médio quadrático, e demonstraram comportamento semelhantes quando novos animais genotipados foram adicionados à análise. Em resumo, o modelo de regressão aleatória mostrou-se uma alternativa promissora para avaliação genômica de eficiência alimentar, e o Algoritmo 2 mostrou-se mais adequado para a aproximação de acurácias, portanto, é o mais indicado para avaliações genéticas de rotina

Palavras-chave: Ingestão alimentar; Dados longitudinais; B-splines; Aproximação de acurácia; Avaliação genômica.

SUMMARY

CHAPTER 1

1.1 General Introduction

The profitability of beef cattle enterprises is significantly impacted by the cost of providing feed, accounting for as much as 75% of their direct expenses (Nielsen et al., 2013). As a result, there is a growing focus on enhancing feed efficiency to increase beef production systems' economic viability and environmental sustainability.

At the animal level, various alternative definitions of feed efficiency exist, each with different applications (Berry and Crowley, 2013). Traditionally, the feed conversion ratio (i.e., feed: gain) or its mathematical inverse, feed conversion efficiency (i.e., gain: feed), was widely used. However, both traits are associated with the animal's growth rate, leading to increased body weight (BW) and a greater demand for nutrients in selected animals (Herd & Bishop, 2000). More recently, residual feed intake (RFI), has become the preferred measurement (Savietto et al., 2014). The RFI is defined as the feed intake expressed as the deviations from the expected feed intake based on metabolic weight and weight gain (Koch et al., 1963) and must be phenotypically independent of both traits. In addition, Koch et al. (1963) also proposed residual weight gain (RWG) as an alternative measure of identifying inter-animal variation in feed efficiency among animals. Using a principle similar to RFI, Koch et al. (1963) defined RWG as the residuals from the regression of ADG on feed intake and metabolic weight. Unlike RFI, where negative values indicate more efficient animals, positive RWG values (i.e., animals growing faster than expected) are deemed more efficient.

The main obstacles to widespread adoption of feed efficiency in cattle breeding programs are the high cost and technical difficulty associated with measuring the trait (Nielsen et al., 2013). The phenotypes are collected in feedlot systems with electronic feed bunks, requiring high financial costs, which limits the data collection. In this context, the advent of genomic selection approaches allows to accommodate information from single nucleotide polymorphisms (SNPs), increasing selection accuracy and accelerating genetic improvement (Berry et al., 2016). The use of the single-step genomic BLUP (ssGBLUP) methodology allows the evaluation of genotyped and non-genotyped animals simultaneously, by replacing the pedigree-based relationship matrix with a realized relationship matrix, which combines pedigree and genomic relationship (Lourenco et al. 2020). In this way, ssGBLUP results in more accurate breeding values compared to the traditional evaluation method (based on pedigree), especially for traits that are difficult to measure, such as efficiency traits (Van Eenennaam et al., 2014).

Traditionally, the average daily feed intake and weight gain are used to calculate feed efficiency traits. However, new methodologies capable of predicting breeding values with high accuracy for these traits become crucial for selecting efficient animals. Feed efficiency traits can be longitudinally analyzed using random regression models (RRM). RRM allow fitting random genetic and environmental effects over time. In addition, the RRM can consider the covariance pattern between the daily records (Oliveira et al., 2019). Thus, RRM provide insights into the temporal variation of biological processes and the physiological implications underlying the traits (Strucken et al., 2015). Therefore, RRM generate valuable information that can be leveraged in breeding programs.

According to Schaeffer (2004), an RRM should contain a set of fixed curves describing the phenotypic relationship among the various groups of individuals, and a set of random regressions, which should include those for the additive genetic and the permanent environmental effects. Consequently, the genetic and permanent environmental correlation between phenotypes at different times can be different from 1. Thus, RRM can be a feasible alternative to increase the accuracy of breeding values for feed efficiency and evaluate genetic parameters as a function of the evaluation period.

Random regression models (RRM) are commonly implemented using orthogonal Legendre polynomials (LP). Alternatively, other functions, such as splines or segmented polynomials, can also be employed (Huisman et al., 2002). B-splines (BSP) represent a specific type of spline function, and as noted by Rice and Wu (2001), they are well-suited for modeling random effects in mixed model analysis and efficient in estimating covariance functions. RRM based on Legendre polynomials typically require fewer coefficients compared to models using B-spline functions (Meyer, 2005), which can facilitate the use of more parsimonious models. However, RRM based on Legendre polynomials often lead to inflated variances at the extremes of the lactation curve, as demonstrated by Misztal (2006). In the context of feed efficiency traits in beef cattle, studies assessing the goodness of fit of RRM are scarce in the literature. Such studies are essential to identify the most suitable functions for modeling these traits.

Currently, it is possible to adapt RRM to include information from single nucleotide polymorphisms (SNP) through the single-step genomic BLUP (ssGBLUP), which results in higher accuracy of breeding values compared to the traditional evaluation method based on pedigree (Oliveira et al., 2019).

Genomic selection can significantly improve the rate of genetic gain in animal breeding programs (Meuwissen et. Al. 2016). To assist selection decisions in practical breeding programs, the accuracy of estimated genomic breeding values (GEBV) is provided, representing the correlation between the true and the estimated breeding value (Wilmink & Dommerholt, 1985). Theoretical accuracy can be calculated based on prediction error variances (PEV) obtained from the diagonal of the inverse of the left-hand side of the mixed model equations (MME - Henderson, 1984). For large MME, inverting the LHS matrix is not computationally feasible; thus, approximated accuracies are employed. Several methods to approximate accuracies from the pedigree-based best linear unbiased prediction (PBLUP) have been proposed and implemented (Misztal & Wiggans, 1988; Harris and Johnson, 1998; Liu et al., 2004; Tier and Meyer, 2004) by exploiting the sparse structure of the animal model (Quaas, 1976). These methods involve accounting for non-genetic effects in the LHS matrix to estimate the effective record contributions (ERC) for the animals. Subsequently, the ERC is combined with the pedigree-based relationship matrix information to calculate approximated accuracies.

However, the inclusion of genomic information introduces complexity, compromising the sparse structure of the MME and making it more challenging to calculate PEV. In scenarios where all animals are genotyped, genomic predictions can be calculated using the genomic BLUP (GBLUP) and single-nucleotide polymorphism BLUP (SNPBLUP) models (Liu et al. 2014). In GBLUP, the size of the MME increases with the number of genotyped animals, whereas SNPBLUP maintains a constant set of equations size with respect to genotyped animals but changes with the number of SNPs. For analyses involving both genotyped and non-genotyped animals through the single-step GBLUP (ssGBLUP; Aguilar et al., 2010), the calculation of PEV by inverting the MME becomes prohibitively expensive for large datasets, primarily due to the presence of dense matrix blocks within the MME, representing the genomic relationships.

Misztal et al. (2013) introduced two algorithms to approximate accuracies of estimated GEBV from ssGBLUP, building upon the traditional pedigree-based algorithm outlined in Misztal and Wiggans (1988). The first algorithm obtained genomic accuracies as a function of the inverse of the sum of a diagonal matrix of ERC and the inverse of the genomic relationship matrix. However, this inverse is not feasible for many genotyped animals; therefore, the second algorithm considers only the diagonal of the inverses of the genomic and pedigree relationship matrices for genotyped animals. In this case, the resulting matrix is diagonal, which has a straightforward inverse for any data size, thus is feasible to obtain the inverses of the genomic and pedigree relationship matrices for genotyped animals. However, the two algorithms provided overestimated accuracies with real datasets. To address this issue, another algorithm was developed. Tsuruta et al., (2016) developed an algorithm that calculates the genomic contribution based on the diagonal of the genomic relationship matrix, which is then combined with the animals' record and pedigree contributions following Misztal and Wiggans (1988). This approach provides less computational cost when working with large populations and has been implemented in the BLUPF90 suite of programs (Misztal et al., 2014a), specifically in the ACCF90GS software.

More recently, Bermann et al. (2022) presented an algorithm for calculating the PEV in the GBLUP model using block sparse inversion of $\mathbf{G}^{-1}$computed by the algorithm of proven and young animals (APY; Misztal et al. 2014b). The authors demonstrated that the described algorithm could be expanded to the ssGBLUP model to approximate accuracies using the ERC procedure, as shown by Liu et al. (2017). Since this approach proved feasible for calculating PEV for GEBV in ssGBLUP for large populations with computational efficiency, it has been implemented in the BLUPF90 suite of programs, in the ACCF90GS2 software.

The objectives of this thesis can be categorized into two primary areas of focus. First, our goals were to: (1) investigate the feasibility of genomic evaluations for dry matter intake, body weight gain, residual feed intake, and residual weight gain via RRM; (2) compare the goodness of fit of RRM using orthogonal Legendre polynomials and B-spline functions; (3) evaluate the behavior of genetic parameters for feed efficiency traits and their implication for selection strategies. Second, our objectives were: 1) compare approximated accuracies from the algorithms presented by Tsuruta et al. (2016) and Bermann et al. (2022); 2) compare the approximated accuracies from the two algorithms mentioned above against the exact accuracy based on the direct inversion of the LHS of MME; 3) Evaluate the impact of adding genotyped animals with and without phenotypes on the exact and approximated accuracies of GEBV.

1.2 References

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. Dairy Sci. 93:743–752. doi: 10.3168/ jds .2009 -2730

Bermann, M., D. A. L. Lourenco, and I. Misztal. 2022. Efficient approximation of reliabilities for single-step genomic best linear unbiased predictor models with the algorithm for proven and young. J. Anim. Sci. 100:1–7. doi:10.1093/jas/skab353

Berry, D. P., Garcia, J. F., & Garrick, D. J. 2016. Development and implementation of genomic predictions in beef cattle. Anim. Front., 6(1), 32–38. doi:10.2527/af.2016-0005

Berry, D. P., & Crowley, J. J. 2013. CELL BIOLOGY SYMPOSIUM: Genetics of feed efficiency in dairy and beef cattle1. J. Anim. Sci, 91(4), 1594–1613. doi:10.2527/jas.2012-5862

Boichard, D., & Lee, A. J. 1992. Approximate Accuracy of Genetic Evaluation Under a Single-Trait Animal Model. J. Dairy Sci., 75(3), 868–877. doi:10.3168/jds.s0022-0302(92)77827-8

Crowley, J. J., McGee, M., Kenny, D. A., Crews, D. H., Evans, R. D., & Berry, D. P. 2010. Phenotypic and genetic parameters for different measures of feed efficiency in different breeds of Irish performance-tested beef bulls. J. Anim. Sci, 88(3), 885–894. doi:10.2527/jas.2009-1852

Durunna, O. N., Mujibi, F. D. N., Goonewardene, L., Okine, E. K., Basarab, J. A., Wang, Z., & Moore, S. S. 2011. Feed efficiency differences and reranking in beef steers fed grower and finisher diets1. J. Anim. Sci, 89(1), 158–167. doi:10.2527/jas.2009-2514

Harris, B., & Johnson, D. 1998. Approximate Reliability of Genetic Evaluations Under an Animal Model. J. Dairy Sci., 81(10), 2723–2728. doi:10.3168/jds.s0022-0302(98)75829-1

Henderson CR. Applications of linear models in animal breeding. Guelph: University of Guelph; 1984

Herd, R.M., and Bishop, S.C. 2000. Genetic variation in residual feed intake and its association with other production traits in British Hereford cattle. Livest. Prod. Sci., 63(2), 111–119. https://doi.org/10.1016/s0301-6226(99)00122-0

Huisman, A. E., Veerkamp, R. F., & van Arendonk, J. A. M. 2002. Genetic parameters for various random regression models to describe the weight data of pigs. J. Anim. Sci, 80(3), 575–582. https://doi.org/10.2527/2002.803575x

Koch, R. M., Swiger, L. A., Chambers, D., Gregory, K. E. 1963. Efficiency of Feed Use in Beef Cattle. J. Anim. Sci, 22(2), 486–494. https://doi.org/10.2527/jas1963.222486x

Liu, Z., P. M. VanRaden, M. H. Lidauer, M. P. Calus, H. Benhajali, H. Jorjani, and V. Ducrocq. 2017. Approximating genomic reliabilities for national genomic evaluation. Interbull Bull. 51:75–85.

Liu, Z., M. E. Goddard, F. Reinhardt, and R. Reents. 2014. A singlestep genomic model with direct estimation of marker effects. J. Dairy Sci. 97:5833–5850. doi:10 .3168/ jds .2014 -7924

Liu, Z., Reinhardt, F., Bünger, A., & Reents, R. 2004. Derivation and Calculation of Approximate Reliabilities and Daughter Yield-Deviations of a Random Regression Test-Day Model for Genetic Evaluation of Dairy Cattle. J. Dairy Sci, 87(6), 1896–1907. doi:10.3168/jds.s0022-0302(04)73348-2

Lourenco, D., Legarra, A., Tsuruta, S., Masuda, Y., Aguilar, I., Misztal, I. 2020. Single-Step Genomic Evaluations from Theory to Practice: Using SNP Chips and Sequence Data in BLUPF90. Genes, 11(7), 790. https://doi.org/10.3390/genes11070790

Meuwissen, T., Hayes, B., & Goddard, M. 2016. Genomic selection: A paradigm shift in animal breeding. Anim. Front., 6(1), 6–14. doi:10.2527/af.2016-0002

Meyer, K. 2005. Random regression analyses using B-splines to model growth of Australian Angus cattle. Genet. Sel., 37(6), 473. doi:10.1186/1297-9686-37-6-473

Meyer, K. 1989. Approximate accuracy of genetic evaluation under an animal model. Livest. Prod. Sci., 21(2), 87–100. doi:10.1016/0301-6226(89)90041-9

Misztal, I., S. Tsuruta, D.A.L. Lourenco, I. Aguilar, A. Legarra, and Z. Vitezica. 2014a. Manual for BLUPF90 family of programs. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all1.pdf. (Accessed 5 October 2022.)

Misztal, I., A. Legarra, and I. Aguilar. 2014b. Using recursion to compute the inverse of the genomic relationship matrix. J. Dairy Sci. 97:3943–3952. doi:10.3168/jds.2013-7752

Misztal, I., S. Tsuruta, I. Aguilar, A. Legarra, P. M. VanRaden, and T. J. Lawlor. 2013. Methods to approximate reliabilities in single-step genomic evaluation. J. Dairy Sci. 96:647–654. doi: 10 .3168/ jds .2012 -5656.

Misztal, I. 2006. Properties of random regression models using linear splines. J. Anim. Breed. Genet., 123(2), 74–80. doi:10.1111/j.1439-0388.2006.00582.x

Misztal, I., & Wiggans, G. R. 1988. Approximation of Prediction Error Variance in Large-Scale Animal Models. J. Dairy Sci., 71, 27–32. doi:10.1016/s0022-0302(88)79976-2

Nielsen, M. K., MacNeil, M. D., Dekkers, J. C. M., Crews, D. H., Rathje, T. A., Enns, R. M., & Weaber, R. L. 2013. Review: Life-cycle, total-industry genetic improvement of feed efficiency in beef cattle: Blueprint for the Beef Improvement Federation11The development of this commentary was supported by the Beef Improvement Federation. The Professional Animal Scientist, 29(6), 559–565. doi:10.15232/s1080-7446(15)30285-0

Nkrumah, J. D., Keisler, D. H., Crews, D. H., Basarab, J. A., Wang, Z., Li, C., … Moore, S. S. 2007. Genetic and phenotypic relationships of serum leptin concentration with performance, efficiency of gain, and carcass merit of feedlot cattle1. J. Anim. Sci., 85(9), 2147–2155. doi:10.2527/jas.2006-764

Oliveira, H. R., Brito, L. F., Lourenco, D. A. L., Silva, F. F., Jamrozik, J., Schaeffer, L. R., Schenkel, F. S. 2019. Invited review: Advances and applications of random regression models: From quantitative genetics to genomics. J. Dairy Sci., 102, 7664-7683. https://doi.org/10.3168/jds.2019-16265

Quaas, R. 1976. Computing the diagonal elements and inverse of a large numerator relationship matrix. Biometrics. 32(4):949–953. doi:10.2307/2529279

Rice, J. A., Wu, C. O. 2001. Nonparametric Mixed Effects Models for Unequally Sampled Noisy Curves. Biometrics, 57(1), 253–259. https://doi.org/10.1111/j.0006-341x.2001.00253.x

Savietto, D., Berry, D. P., & Friggens, N. C. 2014. Towards an improved estimation of the biological components of residual feed intake in growing cattle1. J. Anim. Sci., 92(2), 467–476. doi:10.2527/jas.2013-6894

Strucken, E. M., Laurenson, Y. C. S. M., & Brockmann, G. A. 2015. Go with the flow-biology and genetics of the lactation cycle. Front. Genet., 6. doi:10.3389/fgene.2015.00118

Tier, B., & Meyer, K. 2004. Approximating prediction error covariances among additive genetic effects within animals in multiple-trait and random regression models. J. Anim. Breed. Genet., 121(2), 77–89. doi:10.1111/j.1439-0388.2003.00444.x

Tsuruta, S., D.A.L. Lourenco, Y. Masuda, D.W. Moser, and I. Misztal. 2016. Practical approximation of accuracy in genomic breeding values for a large number of genotyped animals. In: Joint Annual Meeting of American Society of Animal Science, Salt Lake City. https://doi.org/10.2527/jam2016-0337

Van Eenennaam, A.L., Weigel, K.A., Young, A.E., Cleveland, M.A., Dekkers, J.C.M. 2014. Applied Animal Genomics: Results from the Field. Annual Review of Animal Biosciences, 2(1), 105–139. https://doi.org/10.1146/annurev-animal-022513-114119

Wilmink J.B., Dommerholt J. 1985. Approximate Reliability of Best linear Unbiased Prediction in models with and without relationships, J. Dairy Sci. 68 (1985) 946–952. doi: 10.3168/jds.S0022-0302(85)80913-9

CHAPTER 2

GENOMIC ANALYSIS OF FEED EFFICIENCY TRAITS IN BEEF CATTLE USING RANDOM REGRESSION MODELS

2.1 Abstract

Feed efficiency plays a major role in the overall profitability and sustainability of the beef cattle industry, as it is directly related to the reduction of the animal demand for input and methane emissions. Traditionally, the average daily feed intake and weight gain are used to calculate feed efficiency traits. However, feed efficiency traits can be analyzed longitudinally using random regression models (RRM), which allow fitting random genetic and environmental effects over time by considering the covariance pattern between the daily records. Therefore, the objectives of this study were to: (1) propose genomic evaluations for dry matter intake (DMI), body weight gain (BWG), residual feed intake (RFI), and residual weight gain (RWG) data collected during an 84-day feedlot test period via random regression models; (2) compare the goodness of fit of RRM using Legendre polynomials and B-spline functions; (3) evaluate the genetic parameters behavior for feed efficiency traits and their implication for new selection strategies. Genetic parameters and genomic breeding values (GEBV) were estimated by random regression models under ssGBLUP for Nellore cattle using orthogonal Legendre polynomials and B-spline. Models were compared based on the deviance information criterion (DIC). The ranking of the average GEBV of each test week and the overall GEBV average was compared by the percentage of individuals in common and the Spearman correlation coefficient (top 1%, 5%, 10%, and 100%). The highest goodness of fit was obtained with linear B-Spline function considering heterogeneous residual variance. The heritability estimates across the test period for DMI, BWG, RFI, and RWG ranged from 0.06 to 0.21, 0.11 to 0.30, 0.03 to 0.26, and 0.07 to 0.27, respectively. DMI and RFI presented within-trait genetic correlations ranging from low to high magnitude across different performance test-day. In contrast, BWG and RWG presented negative genetic correlations between the first three weeks and the other days of performance tests. DMI and RFI presented a high-ranking similarity between the GEBV average of week eight and the overall GEBV average, with Spearman correlations and percentages of individuals selected in common ranging from 0.90 to 1.00. The highest Spearman correlations (ranging from 0.94 to 0.98) and percentages of individuals selected in common (ranging from 0.94 to 0.95) of BWG and RWG were observed at week 11. In conclusion, the random regression model using linear B-splines is a feasible alternative for the genomic evaluation of feed efficiency. Heritability estimates of DMI, RFI, BWG, and RWG indicate enough additive genetic variance to achieve a moderate response to selection. A new selection strategy can be adopted by reducing the performance test to 56 days for DMI and RFI selection and 77 days for BWG and RWG selection.

Keywords: feed intake, model comparison, longitudinal data, genomic selection, B-splines, Legendre polynomials.

2.2 Introduction

In beef cattle systems, the feed costs represent the largest effective operational cost (Souza et al., 2021). Therefore, the profitability and sustainability of livestock activities rely on feed efficiency. In addition, feed efficiency is also related to methane emissions (Manzanilla-Pech et al., 2022), playing a role in the environmental impact of beef cattle production. In this context, selection for feed efficiency is an important strategy in animal breeding programs to reduce production costs by lowering animal demand for inputs (Tizioto et al., 2016) and reducing methane emissions. Traditionally, feed efficiency is measured by gross feed efficiency (intake over weight gain) and conversion (weight gain over intake). Both traits are associated with the animal's growth rate, leading to increased body weight (BW) and a greater demand for nutrients in selected animals (Herd & Bishop, 2000). In this sense, the recent focus of beef cattle breeding programs has been to use residual feed intake (RFI) as a selection criterion for feed efficiency. The RFI is defined as the feed intake expressed as the deviations from the expected feed intake based on metabolic weight and weight gain (Koch et al., 1963) and must be phenotypically independent of both traits. In addition, residual weight gain (RWG) has also been proposed as a feed efficiency trait. The RWG is the weight gain expressed as deviations from the expected weight gain based on feed intake and metabolic weight (Koch et al., 1963).

Although feed efficiency has great economic relevance, phenotypes are collected in feedlot systems with electronic feed bunks, requiring high financial costs, which limits the data collection. Therefore, new methodologies capable of predicting breeding values with high accuracy for these traits become crucial for selecting efficient animals. Feed efficiency traits can be analyzed longitudinally using random regression models (Wang et al., 2019). Unlike the traditional model, in which the average daily feed intake and weight gain are used, random regression models (RRM) allow fitting random genetic and environmental effects over time. In addition, the RRM can consider the covariance pattern between the daily records (Oliveira et al., 2019). Thus, RRM can be a feasible alternative to increase the accuracy of breeding values for feed efficiency and evaluate genetic parameters as a function of the evaluation period.

Random regression models are usually applied with orthogonal Legendre polynomials (LP). Other functions, such as splines or segmented polynomials, can also be adopted (Huisman et al., 2002). A specific type of spline function is B-splines (BSP), and according to Rice and Wu (2001), B-splines are suitable for modeling random effects in mixed model analysis and efficient in estimating covariance functions. Currently, it is possible to adapt RRM to include information from single nucleotide polymorphisms (SNP) through the single-step genomic BLUP (ssGBLUP), which results in more accurate breeding values compared to the traditional evaluation method based on pedigree (Oliveira et al., 2019). However, studies that assess the goodness of fit of RRM for feed efficiency traits in beef cattle are scarce in the literature.

In this context, we aimed to: (1) investigate the feasibility of genomic evaluations for dry matter intake, body weight gain, residual feed intake, and residual weight gain via RRM; (2) compare the goodness of fit of RRM using orthogonal Legendre polynomials and B-spline functions; (3) evaluate the behavior of genetic parameters for feed efficiency traits and their implication for selection strategies.

2.2 Materials and methods

The datasets were provided by EMBRAPA-GENEPLUS beef cattle breeding program (Campo Grande, Mato Grosso do Sul, Brazil). They included phenotypic records for 3,138 Nellore young bulls born between 2014 and 2019, raised under extensive system. The animals underwent an adaptation period of 21 days to the diet and station, after which they were subjected to an 84-day feed efficiency test with ad libitum access to diet and water. The tests were performed in two performance test stations (Embrapa’s Farm and Genética Aditiva’s farm), where the dry matter intake and daily weight data were automatically obtained using electronic feed bunks (Intergado® system, Intergado Ltda., Contagem, Minas Gerais, Brazil).

2.2.1 Phenotypic data

Dry matter intake (DMI) was defined as the daily amount of individually consumed feed automatically recorded by the electronic system. The body weight gain (BWG) was calculated as the difference between the average daily body weights of consecutive test days.

Data consistency analysis was performed using the R software (R Core Team, 2019). Contemporary groups (CG) were determined based on sex, herd, year, and season (1 = January to March, 2 = April to June, 3 = July to September, and 4 = October to December) of birth. Animals belonging to CG with less than three records were excluded. Records of animals with phenotypic data above or below three standard deviations from the mean of each test day within CG were also removed. The complete pedigree contained 10,106,839 animals. Descriptive statistics of the data are presented in Table 1.

Table 1. Number of Animals (Animals), Number of records (N), Number of Performance Tests (Tests), Number of Contemporary Groups (CG), means, maximum and minimum values, and standard deviations (SD) obtained for dry matter intake (DMI) and body weight gain (BWG).

Trait	Animals	N	Tests	CG	Mean	Maximum	Minimum	SD
DMI	2920	168,052	18	306	11.37	45.22	0.05	4.53
BWG	2696	153,977	18	274	1.34	6.00	-2.60	0.79

DMI= dry matter intake; BWG= body weight gain.

2.2.2 Genotipic data

A total of 4,675 animals were genotyped using one of the 777K, 35K, 30K, 27K, and 26K Illumina SNP panels. The genotype quality control was performed separately for each SNP density, using the snpStats v1.16.0 package (Clayton, 2014) and the preGSf90 software (Misztal et al., 2014). The following criteria for SNP exclusion were assumed: genotype calling score < 15%; sample and SNP Call Rate < 95%; minor allele frequency (MAF) < 5%; sample heterozygosity ± three standard deviations; p-value of the chi-squared test for Hardy–Weinberg equilibrium lower than 10^{-6}.

Pedigree consistency analysis was evaluated through the seekparentf90 software (Aguilar, 2014). Relationships between parent and progeny with Mendelian inconsistencies above 1% were excluded. The imputation of all genotypes to the 777k panel was performed using the FImpute v.2.2 software (Sargolzaei, Chesnais, & Schenkel, 2014). Imputation accuracy was estimated via cross-validation analysis based on the random exclusion of genotyped animals and was higher than 97%, as previously reported by Carvalheiro et al.

(2014). After the imputation process, markers with MAF <5% were excluded. The genomic quality control resulted in a final dataset with 4,551 genotyped animals and 122,447 SNPs.

2.2.3 Statistical analyses

Variance components, genetic parameters, and genomic estimated breeding values (GEBV) for each day were obtained by fitting a random regression model using the GIBBS3f90 software (Misztal et al., 2014) under ssGBLUP. The additive genetic and permanent environment effects were modeled using orthogonal Legendre polynomials of order one to four and linear B-splines segmented polynomials, with homogeneous and heterogeneous residual variance. Different residual variances every two weeks of the performance test were evaluated in models with heteroscedasticity. The B-splines models assumed 7 knots to divide the performance test duration time (84 days) into equally spaced intervals of 2 weeks. The most suitable model was chosen after comparing the goodness of fit through the deviance information criterion (DIC). Genetic correlations between the different days of the performance test were estimated using the genetic (co)variances of the traits on each test day.

The Bayesian random regression model can be represented in matrix notation as:

$$\mathbf{y=Xb+Za+Wp+e} \qquad [1]$$

where $\mathbf{y}$ is the vector of observations (body weight gain and dry matter intake), $\mathbf{y|b,a,p,G_a, R_p,}\sigma_e^2 \sim \mathbf{N\ (Xb+Za+Wp, I}\sigma_e^2\mathbf{)}$, in which σ_e^2 is the residual variance; **b** is the vector of systematic effects (CG, and performance test group), assuming a uniform prior distribution; **a** is the vector of regression coefficients for the additive genetic effect, $\mathbf{a|\Sigma}_a\mathbf{, H} \sim \boldsymbol{N}\ (\mathbf{0}, \mathbf{\Sigma}_a \otimes \mathbf{H})$, in which **H** is a hybrid relationship matrix that includes pedigree and genomic information, and $\mathbf{\Sigma}_a$ is the additive genetic covariance matrix; **p** is the vector of regression coefficients for the random permanent environmental effect, assuming $\mathbf{p|\Sigma}_p \sim \boldsymbol{N}\ (\mathbf{0}, \mathbf{\Sigma}_p \otimes \mathbf{I})$, which **I** is an identity matrix and $\mathbf{\Sigma}_p$ is a matrix of permanent environmental variance. The incidence matrices for **b**, **a,** and **p,** are **X**, **Z**, and **W**, respectively.

According to Aguillar et al. (2010), the **H** inverse matrix is given by:

$$\mathbf{H}^{-1}=\mathbf{A}^{-1}+\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{G}^{-1}-\mathbf{A}_{22}^{-1}\end{bmatrix} \qquad [2]$$

where $\mathbf{A}_{22}^{-1}$ is the inverse of the numerator relationship matrix for genotyped animals; $\mathbf{G}^{-1}$ is the inverse of the genomic relationship matrix (**G**) constructed based on the type I described by VanRaden (2008).

For the genetic parameters estimation of RFI, the model assumed **y** as the vector of daily dry matter intake observations, and **b** included the systematic effects of daily metabolic body weight ($\mathrm{DBW}^{0.75}$), daily body weight gain, CG, and performance test group. For RWG, the model assumed **y** as the vector of daily body weight gain observations and the systematic effects of daily dry matter intake, daily metabolic weight, CG, and performance test group.

A total of 500,000 MCMC iterations were used to infer all evaluated traits. The burn-in and thinning were given by 300,000 and ten iterations, respectively. The POSTGIBBSF90 software (Misztal et al., 2014) was used for posterior inference, and the convergence was evaluated through the Geweke test (Geweke, 1992).

In order to identify the most efficient individuals during the entire test period, the animals were ranked through the overall average GEBV ($\mathrm{GEBV_A}$). In addition, the average GEBV of each test week were estimated, assuming the $\mathrm{GEBV_x}$ (where X is the corresponding week number) as the average of the GEBV from the beginning of the test to the last day of the corresponding week. The ranking based on $\mathrm{GEBV_x}$ was evaluated by the percentage of individuals in common and the Spearman correlation coefficient (top 1%, 5%, 10%, and 100%). The rank comparison analysis was computed considering only animals with genotypic and phenotypic information.

2.3 Results

The DIC for the different models are in Table 2. The DIC values decreased as the order of Legendre polynomials (LP) increased, indicating that models using LP of higher order obtained a better goodness of fit. Additionally, models considering heterogeneous residual variance also presented a lower DIC value. The highest goodness of fit was obtained with linear B-Spline function considering heterogeneous residual variance, in which the DIC values for DMI, BWG, RFI, and RWG were 735,075.5, 217,52.2, 632,078.4, and 149,652.0, respectively (Table 2). Thus, we applied a random regression model using B-splines segmented polynomials with six classes of residual variance for genetic parameters and breeding values predictions.

Table 2. Deviance Information Criterion (DIC) for different models with Legendre polynomials and B-splines segmented polynomials.

Trait	Residual Variance Classes	LEG1	LEG2	LEG3	LEG4	B-Spline1
DMI	1	758101.2	756511.5	754714.0	754079.0	741326.6
	6	749517.1	747292.7	745749.4	744859.5	735075.5
BWG	1	122574.6	78652.2	55898.6	38507.8	25589.3
	6	101773.6	52526.6	30117.9	24457.8	21752.2
RFI	1	1007562.3	998456.5	897568.2	889657.6	636029.8
	6	985621.1	923568.7	876598.2	897854.4	632078.4
RWG	1	214589.5	204589.3	194321.2	175895.4	150543.6
	6	201548.6	197845.7	196324.3	165428.2	149652.0

DMI= dry matter intake; BWG= body weight gain; RFI=residual feed intake; RWG= residual weight gain; LEGx = Legendre polynomial, in which x = 1, 2, 3, and 4 is the order for the additive genetic and permanent environmental regressions; B-Spline1 = linear B-spline functions.

The maximum and minimum values of posterior means and their 95% highest probability density intervals (HPD95) for genetic parameters obtained for DMI, BWG, RFI, and RWG from days 1 to 84 of the performance test are presented in Table 3. The heritability was low to moderate for all traits across the performance test. The lowest heritability posterior mean was observed for RFI (0.03), whereas the highest was for BWG (0.30). According to the HPD intervals, all parameters differed from zero and no difference was observed in the highest heritability of the four traits. The lowest heritability estimates of DMI, RFI, and RWG were similar. In contrast, the lowest heritabilities for DMI and RFI differed from those of BWG.

Table 3. Posterior means and 95% highest density probability for heritability (h_a^2), additive genetic variance (σ_a^2), permanent environment variance (σ_{pe}^2), and residual variance (σ_e^2). Minimum and maximum estimates across the performance test days (1 to 84 days) for DMI, BWG, RFI, and RWG.

Parameters		DMI	BWG	RFI	RWG
h_a^2	Min	0.06 (0.04 - 0.08)	0.11 (0.09 - 0.13)	0.03 (0.02 - 0.05)	0.07 (0.03 - 0.11)
	Max	0.21 (0.15 - 0.26)	0.30 (0.24 - 0.35)	0.26 (0.21 - 0.32)	0.27 (0.18 - 0.41)
σ_a^2	Min	0.58 (0.41 - 0.78)	0.02 (0.01 - 0.034)	0.31 (0.17 - 0.50)	0.02 (0.01 - 0.04)
	Max	1.57 (1.14 - 2.03)	0.51 (0.29 - 0.73)	2.17 (1.67 - 2.75)	0.45 (0.28 - 0.71)
σ_{pe}^2	Min	1.41 (1.08 - 1.86)	0.09 (0.08 - 0.10)	0.80 (0.55 - 1.09)	0.08 (0.06 - 0.09)
	Max	2.97 (2.51 - 3.40)	1.40 (0.53 - 2.28)	1.78 (1.42 - 2.14)	0.95 (0.77 - 1.09)
σ_e^2	Min	3.53 (3.47 - 3.58)	0.01 (0.01- 0.01)	3.59 (3.50 - 3.68)	0.01 (0.01 - 0.01)
	Max	6.96 (6.86 - 7.07)	0.08 (0.08 - 0.09)	7.55 (7.46 - 7.64)	0.26 (0.16 - 0.36)

DMI= Dry matter intake; BWG= Body weight gain; RFI= Residual feed intake; RWG= Residual weight gain.

Figure 1 shows the posterior means and their 95% highest probability density intervals (HPD95) of heritability obtained for BWG, RWG, DMI, and RFI along the 84 days of the performance test.

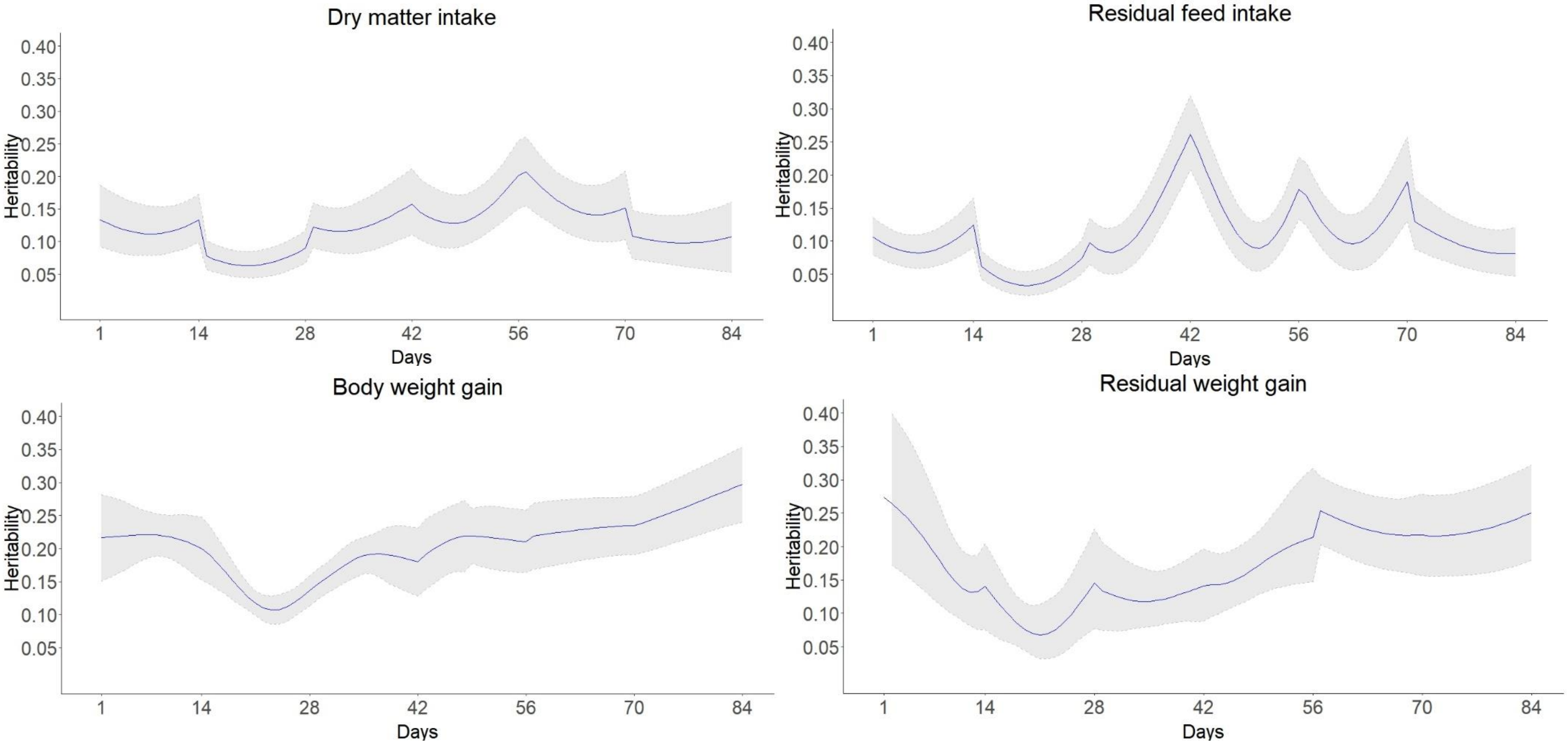


Figure 1. Posterior means and 95% highest probability density intervals (HPD95) for heritability over test-days, for dry matter intake, residual feed intake, body weight gain and residual weight gain in Nellore cattle.

The posterior means of heritability of the four traits presented dynamic changes and different behavior throughout the performance test period (Figure 1). However, based on the HPD95, the heritability estimates for all four traits at the beginning and end of the performance test did not differ significantly. The heat maps of the genetic correlations between different days of performance tests for BWG, RWG, DMI, and RFI are in Figure 2.

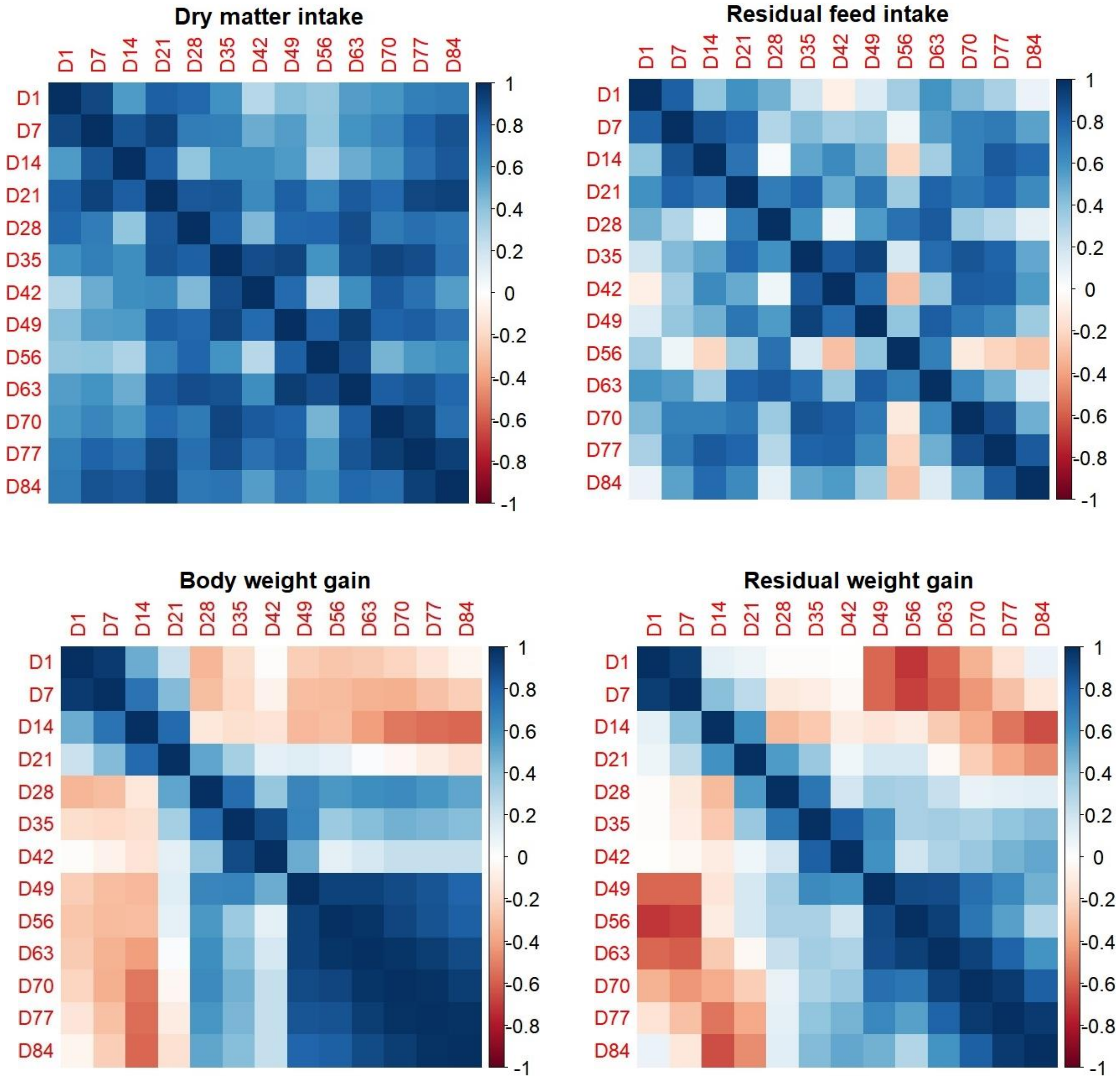


Figure 2. Heat map of genetic correlations between the performance test-days (Dx, with x ranging from 1 to 84 days) for dry matter intake, residual feed intake, body weight gain, and residual weight gain in Nellore cattle.

Dry matter intake presented only positive genetic correlations between performance test days (Figure 2). In contrast, genetic correlations for residual feed intake were negative between a few test days. Body weight gain and residual weight gain showed more significant changes in the genetic correlations between the test days. Both traits presented negative genetic correlations between the days of the first three weeks and the other days of performance tests.

Spearman correlations between the $GEBV_A$ and the average $GEBV_X$ of test weeks, considering 1%, 5%, 10%, and 100% percentiles of the best animals, are presented in Table 4. Dry matter intake and residual feed intake presented a high Spearman correlation when considering 100% of evaluated animals with phenotypic information. From week eight onwards, dry matter intake presented values above 90% in 1%, 5%, and 10% percentiles. Residual feed intake, in turn, presented Spearman correlations greater than 90% from week seven onwards. On the other hand, body weight gain and residual weight gain generally presented low, negative Spearman correlations.

Table 4. Spearman correlations for the four evaluated traits, based on the general average GEBV ($GEBV_A$) and the average GEBV of other weeks ($GEBV_X$) of the performance test, considering 1%, 5%, 10% and 100% percentiles of the best animals.

	Spearman Correlation*															
Weeks of test	Dry matter intake				Residual feed intake				Body weight gain				Residual weight gain			
	1%	5%	10%	100%	1%	5%	10%	100%	1%	5%	10%	100%	1%	5%	10%	100%
$GEBV_1$ and $GEBV_A$	0.21	0.42	0.54	0.79	0.50	0.36	0.34	0.50	-0.29	0.02	0.05	-0.17	-0.38	-0.35	-0.20	-0.28
$GEBV_2$ and $GEBV_A$	0.07	0.48	0.54	0.79	0.55	0.35	0.30	0.55	-0.15	0.00	0.04	-0.24	-0.54	-0.33	-0.19	-0.34
$GEBV_3$ and $GEBV_A$	0.10	0.54	0.58	0.82	0.60	0.41	0.37	0.60	0.04	-0.02	0.04	-0.23	-0.61	-0.27	-0.16	-0.33
$GEBV_4$ and $GEBV_A$	0.30	0.63	0.72	0.90	0.72	0.60	0.60	0.72	0.12	-0.01	0.06	-0.11	-0.54	-0.22	-0.11	-0.24
$GEBV_5$ and $GEBV_A$	0.51	0.72	0.80	0.95	0.83	0.76	0.78	0.83	0.17	-0.03	0.07	0.07	-0.48	-0.20	-0.04	-0.06
$GEBV_6$ and $GEBV_A$	0.62	0.79	0.84	0.97	0.91	0.90	0.90	0.91	0.00	-0.10	0.03	0.24	-0.31	-0.20	0.01	0.15
$GEBV_7$ and $GEBV_A$	0.69	0.83	0.88	0.98	0.92	0.90	0.91	0.92	0.18	-0.03	0.10	0.49	-0.10	-0.10	0.12	0.39
$GEBV_8$ and $GEBV_A$	0.95	0.95	0.98	1.00	0.99	0.99	0.99	0.99	0.75	0.37	0.43	0.78	0.29	0.24	0.39	0.67
$GEBV_9$ and $GEBV_A$	0.95	0.95	0.99	1.00	0.89	0.92	0.93	0.89	0.94	0.66	0.68	0.92	0.61	0.63	0.68	0.87
$GEBV_{10}$ and $GEBV_A$	0.96	0.97	0.99	1.00	0.91	0.93	0.94	0.91	0.97	0.81	0.84	0.98	0.88	0.85	0.84	0.96
$GEBV_{11}$ and $GEBV_A$	0.99	0.99	1.00	1.00	0.97	0.98	0.98	0.97	0.98	0.94	0.96	1.00	0.94	0.94	0.95	0.99

$GEBV_X$= The average GEBV from the first to the last test-day of the week (X); $GEBV_A$ = The average GEBV of the entire period of the test.

*p<.001.

The Percentage of commonly selected individuals for BWG, DMI, RWG, and RFI between GEBV_A and GEBV_X, assuming 1%, 5%, and 10% percentiles of the best animals, are presented in Table 5.

Table 5. Percentage of commonly selected individuals for the four evaluated traits, based on GEBV_A and GEBV_X of the performance test, considering 1%, 5% and 10% percentiles of the best animals.

	Percentage of commonly selected individuals											
Weeks of test	Dry matter intake			Residual feed intake			Body weight gain			Residual weight gain		
	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
GEBV_1 and GEBV_A	48	53	57	30	35	39	0	7	10	0	4	7
GEBV_2 and GEBV_A	48	52	61	41	39	36	0	05	06	0	04	06
GEBV_3 and GEBV_A	52	56	66	44	43	44	0	03	07	0	04	06
GEBV_4 and GEBV_A	69	71	75	56	61	61	0	05	13	04	04	08
GEBV_5 and GEBV_A	86	80	79	70	72	77	0	13	19	04	08	14
GEBV_6 and GEBV_A	86	81	86	85	83	87	0	16	23	04	13	19
GEBV_7 and GEBV_A	90	85	91	85	88	90	0	25	33	07	23	31
GEBV_8 and GEBV_A	93	93	97	100	99	99	26	50	59	30	47	52
GEBV_9 and GEBV_A	100	97	97	89	90	91	63	73	75	59	70	72
GEBV_{10} and GEBV_A	100	97	98	89	93	93	74	81	85	74	84	83
GEBV_{11} and GEBV_A	100	99	98	93	96	96	85	93	94	85	92	92

GEBV_X = the average GEBV from the first test-day to the last test-day of the corresponding week (X); GEBV_A = the average GEBV of the entire period of the test.

DMI and RFI presented a higher percentage of commonly selected individuals between the overall average GEBV and the average GEBV of the performance test weeks. The percentages of commonly selected individuals in weeks 7, 8, 9, 10, and 11 were above 80% for the three percentiles (Table 5). In contrast, body weight gain and residual weight gain presented, in general, a low percentage of commonly selected individuals.

2.4 Discussion

2.4.1 Models Comparison

Variance components and GEBV for feed efficiency traits in a Nellore population were estimated using random regression models. The additive genetic and permanent environment effects were modeled using orthogonal LP and BSP with homogeneous and heterogeneous residual variance. The DIC values of RRM decreased as the order of LP for the animal effect increased. These results corroborate with Wang et al. (2022), who reported that models using LP of higher order were also better for genomic prediction of feed efficiency in pigs. The models that assumed heterogeneous residual variance also showed better goodness of fit, suggesting considerable changes in the residual variance over the feedlot period.

The model based on B-spline functions with seven knots and six classes of residual variance was the best based on the DIC. According to Brito et al. (2018), major multicollinearity issues are expected when using Legendre polynomials of higher order, and the dependency problems of regression parameter estimates across the curve are reduced when B-splines are used. In addition, RRM using B-splines is less susceptible to erratic estimates and inflated variances at the extreme of the curves, which is often observed when using Legendre polynomials (Meyer, 2005). Most beef cattle studies using RRM based on spline functions analyzed growth traits (Oliveira et al., 2022), but none have reported using such a model for feed efficiency traits. Spline functions are a feasible alternative since they present similar estimated breeding values as Legendre polynomial functions. Therefore, due to the better goodness of fit, the random regression model using B-splines segmented polynomials was chosen for genetic parameters and GEBV estimation for the evaluated traits.

2.4.2 Heritabilities

The results of the present study indicate that genetic parameters may change over the test days (Figure 1). The heritability estimates for the four traits change dynamically during the performance test. In addition, DMI, RFI, BWG, and RWG presented enough genetic variance to respond to genetic selection.

Posterior means of heritability for RFI were low to moderate, ranging from 0.03 to 0.26 (Table 3) across the performance test. At the beginning of the test, RFI presented lower heritabilities. The low heritability until week four indicates that, at the beginning of the performance test, RFI was significantly influenced by the higher permanent environmental effects (Supl 2). The highest heritability estimates were observed in weeks 6 and 7, with a maximum value of 0.26 on day 42 (Figure 1), influenced by a high increase in the genetic variance (Supl 1). These results agreed with Polizel et al. (2018), de Oliveira et al. (2022), Robinson et al. (2004), and Ceacero et al. (2016), who reported heritability estimates ranging from 0.14 to 0.28 for RFI in beef cattle, using traditional animal models. In contrast, Wang et al. (2022) related higher heritability estimates, ranging from 0.11 to 0.48, in a pig population using a random regression model based on Legendre polynomials.

Heritabilities of DMI and RFI showed similar trends (Figure 1). The posterior means of heritability for DMI ranged from 0.06 to 0.21 and presented lower values at the beginning of the test (Figure 1). Contributing to this was the relatively higher permanent environmental effects (Supl 2) in the first two weeks and mainly the high residual variance in the third and fourth weeks (6.96). In contrast, Torres-Vásquez et al. (2019) obtained higher estimates of heritability between 0.32 and 0.40 for DMI in Angus cattle, using a random regression model based on Legendre polynomials. The high permanent environmental and residual variance of DMI and RFI, particularly at the start of the performance test, suggest that the environment significantly impacts feeding behavior, even in controlled feedlots. In turn, Haskell et al. (2019) showed that feeding behavior, such as the length of feeder visits and dominance, affected RFI and DMI, respectively, in a beef cattle population. Therefore, these results can be attributed to the animals' adaptation, as they were previously raised in pasture systems and were not used to the feedlot system and electronic feed bunks.

Heritability for BWG and RWG showed similar trends (Figure 1), ranging from 0.11 to 0.3 and 0.07 to 0.27, respectively (Table 3). The lower heritability estimates were observed for both traits in weeks 3 and 4. The decrease in genetic variance (ranging from 0.023 to 0.066 and 0.023 to 0.061 for RWG and BWG, respectively) can explain the low heritability in these two weeks. Therefore, the selection based on the GEBV for these weeks would result in a lower genetic gain. The other weeks showed little fluctuations in the posterior means of heritability. The heritabilities for BWG were lower than those reported by Torres-Vázquez et al. (2019), who obtained estimates of around 0.4 in early- and 0.5 in mid-test, considering a test period of 391 days. Genetic parameters for RWG predicted by random regression models were not found in the literature.

The present study shows that genetic parameter estimates for feed efficiency traits and weight gain fluctuate over time. Thus, depending on the selection strategy, it may impact the potential response to selection for feed efficiency. After all, the selection based on different periods of the feedlot test will result in different genetic gains due to the heritability difference across the test days. In addition, the four evaluated traits presented changes in the residual variance and the permanent environmental effects across the feedlot test period. Therefore, the environmental components considerably influence the feed efficiency traits, which can be associated with the animal adaptation to the management practices and the feedlot structure.

2.4.3 Genetic Correlations

The (co)variance components for each day of the performance test were obtained from the regression coefficients, and the genetic correlations between different test days were estimated (Figure 2). Within-trait correlations were generally strongest between adjacent days and weakened as the interval between test days increased. Dry matter intake was the unique trait that presented only positive genetic correlations ranging from low to high magnitude across different performance test-day (Figure 2). These indicate that the selection for DMI, based on any test day, would increase DMI during the entire performance test period. Additionally, DMI on day 84 presented an average genetic correlation of 0.76 (SD = 0.11) with the other test days, which suggests that DMI can be well predicted from early days. In general, RFI presented smaller genetic correlations between test days when compared with DMI. Residual feed intake showed a few negative genetic correlations between some test days. These results suggest that different allelic variability may govern RFI at different stages of the performance test.

Body weight gain and RWG showed a similar genetic correlation pattern. Both traits showed negative correlations between the first three weeks' test days and the remaining weeks of the performance test. This suggests a genetic antagonism between both traits from the beginning to the end of the test. Thus, the selection decisions taken at the beginning of the test would have an unfavorable impact on the final test performance. This change in the genetic correlations after the third week may be related to the animal's adaptation to the feedlot. The adaptation period of 21 days assumed for the performance test follows Parra et al. 2018, which stated that 21 days is enough for Nellore's nutritional adaption in a feedlot. Additionally, the authors reported that this period could be reduced to 14 days with no significant differences. However, other intrinsic components can influence the animal's feeding behavior. According to von Keyserlingk and Weary (2010), the timing and frequency of feed delivery, the feed bunk structure, and space allowance are important features that affect the feeding behavior of ruminants. Neave et al. (2018), in turn, related that the propensity of individuals to explore their feeding environment, the reactivity of individuals in response to management stressors, dominance status, and degree of sociability can affect growth and productivity in ruminants.

Therefore, genetic correlations of feed efficiency traits between performance test days can vary, suggesting that different genes regulate these traits depending on the test period. As a result, selecting animals based on different test periods may lead to different genetic responses. Thus, to ensure the effective selection of efficient animals throughout the efficiency test, it is crucial to design a selection strategy that considers these differences across days.

2.4.4 GEBVs ranking of animals.

The overall average of GEBV ($\mathrm{GEBV_A}$) across the entire period and $\mathrm{GEBV_X}$ of each week were estimated and compared for the four traits. The comparison between the $\mathrm{GEBV_X}$ ranking of different weeks and the $\mathrm{GEBV_A}$ allowed us to assess the similarity of the selected candidates in different percentiles and weeks of the performance test. The ranking was evaluated by the percentage of individuals in common (top 1%, 5%, and 10% - Table 4) and the Spearman correlation coefficient (top 1%, 5%, 10%, and 100%- Table 5). In general, Spearman correlations and the percentage of commonly selected individuals between $\mathrm{GEBV_X}$ and $\mathrm{GEBV_A}$ increased over the weeks (Table 4). Dry matter intake showed Spearman correlations ranging from 0.79 to 1.0 according to the $\mathrm{GEBV_X}$ of the entire evaluated population (top 100%). However, correlations within the 1%, 5% and 10% percentiles ranged from 0.07 to 0.99, 0.42 to 0.99, and 0.54 to 1.00, respectively. This suggests greater differences in the ranking of animals according to lower percentiles of selected animals. Spearman correlations (Table 4) and the percentage of individuals selected in common (Table 5) from week eight onwards were greater than 90% for the four percentiles (1%, 5%, 10%, and 100%). Thus, selection based on $\mathrm{GEBV_A}$, $\mathrm{GEBV_8}, \mathrm{GEBV_9}, \mathrm{GEBV_{10}}, \mathrm{GEBV_{11}}$ would lead to the selection of a very similar set of individuals.

Therefore, the high similarity between the ranking obtained for the 84 days and shorter test periods may suggest the selection of the same candidates through a shorter test period. Shortening the performance test period may improve the feed efficiency selection response by increasing the number of cattle recorded in feedlots with fixed capacities (Englishby et al., 2018). Additionally, efficiency phenotypes are measured with electronic feed bunks, which are costly. In this way, performance tests with shorter periods of recording individual feed intake would have greater economic viability, increasing the farmer's interest in collecting feed efficiency phenotypic data.

Residual feed intake presented a moderate to high Spearman correlation considering the 100% percentile (Table 4). However, the Spearman correlations variation within the 5% and 10% percentiles were higher (0.35 to 0.99 and 0.30 to 0.99, respectively). Spearman correlations at percentiles 1%, 5%, and 10%, from week six onwards, were greater than 90%, while the percentages of individuals selected in common were higher than 80% (Table 5). Thus, using $\mathrm{GEBV_6}$ as a selection criterion would lead to the selection of very similar individuals compared to the selection based on $\mathrm{GEBV_A}$. This result corroborates with Englishby et al. (2018), who used a traditional multiple-trait model and found that shortening the test duration from 83 days to 42 days would not impact selection for RFI. In turn, Archer et al. (1997) reported that shortening the test duration to 35 days (five weeks) would not impact the selection for feed efficiency. However, the present study shows that selection based on $\mathrm{GEBV_5}$ may result in the re-ranking of animals. Thus, choosing $\mathrm{GEBV_5}$ as a selection criterion would result in a great difference in selected animals compared to the selection using $\mathrm{GEBV_A}$. In addition, $\mathrm{GEBV_8}$ showed the highest-ranking similarity with $\mathrm{GEBV_A}$, with Spearman correlations of 0.99 and percentages of individuals selected in common ranging from 0.99 to 1.00. This suggests that selection based on the GEBV average of week eight would result in the selection of the same animals as in $\mathrm{GEBV_A}$.

Body weight gain and RWG presented greater differences in the ranking of animals during the test weeks compared to the other traits. The Spearman correlation between the average GEBV of the first five test weeks and the $\mathrm{GEBV_A}$, ranged from -0.24 to 0.07 and -0.34 to -0.06 for BWG and RWG, respectively (Table 4). Thus, selecting for these traits based on the beginning of the test would result in a considerable difference in selected animals. Different from DMI and RFI, the results for BWG and RWG indicate that greatly reducing the test period is not possible. The highest Spearman correlations (ranging from 0.94 to 0.98) and percentages of individuals selected in common (ranging from 0.94 to 0.95) were observed at week 11. Therefore, $\mathrm{GEBV_{11}}$ seems to be the best option for the indirect selection of $\mathrm{GEBV_A}$ to improve BWG and RWG.

In conclusion, the random regression model is a feasible alternative for the genomic evaluation of feed efficiency in Nellore cattle. Random regression models using linear B-splines with six segments (seven knots) and six classes of residual variances is recommended for DMI, RFI, BWG, and RWG evaluations, according to better goodness of fit. Genetic parameter estimates for feed efficiency traits and weight gain fluctuated over the performance test. Heritability estimates indicate enough additive genetic variance to achieve a moderate response to selection. Based on genetic correlations and ranking comparisons, it is possible to change the selection strategy by reducing the performance test period from 84 to 56 days since it would result in selecting almost the same animals for RFI and DMI. However, for weight gain traits (BWG and RWG), a larger test period of 77 days would be adequate.

2.5 References

Archer, J.A., Arthur, P.F., Herd, R.M., Parnell, P.F., Pitchford, W.S. (1997). Optimum post weaning test for measurement of growth rate, feed intake, and feed efficiency in British cattle. *Journal of Animal Science*, 75, 2024–2032. https://doi.org/ 10.2527/1997.7582024x

Aguilar, I. (2014). SeekParentF90. http://nce.ads.uga.edu/wiki/doku.php?id=readme.seekparentf90 [Accessed March 15, 2022].

Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., & Lawlor, T. J. (2010). Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score1. *Journal of Dairy Science*, 93(2), 743–752. https://doi.org/10.3168/jds.2009-2730

Brito, L. F., Gomes da Silva, F., Rojas de Oliveira, H., Souza, N., Caetano, G., Costa, E. V., Menezes, G.R.O., Melo, A.L.P., Rodrigues, M.T., Torres, R.A. de Almeida Torres, R. (2017). Modelling lactation curves of dairy goats by fitting random regression models using Legendre polynomials or B-splines. *Canadian Journal of Animal Science*, 98, 73-83. https://doi.org/10.1139/cjas-2017-0019

Ceacero, T. M., Mercadante, M. E. Z., Cyrillo, J. N. dos S. G., Canesin, R. C., Bonilha, S. F. M., & de Albuquerque, L. G. (2016). Phenotypic and Genetic Correlations of Feed Efficiency Traits with Growth and Carcass Traits in Nellore Cattle Selected for Postweaning Weight. *PLOS ONE*, 11: e0161366. https://doi.org/10.1371/journal.pone.0161366

Carvalheiro, R., Boison, S. A., Neves, H. H. R., Sargolzaei, M., Schenkel, F. S., Utsunomiya, Y. T., … Garcia, J. F. (2014). Accuracy of genotype imputation in Nelore cattle. *Genetics Selection Evolution*, 46, 69. https://doi.org/10.1186/s12711-014-0069-1

Clayton, D. (2014). SNPSTATS: SnpMatrix and XSnpMatrix Classes and Methods. R package, Bioconductor. https://bioconductor.org/packages/release/bioc/html/snpStats.html [Accessed March 20, 2022].

de Oliveira, M.H.V., Malheiros, J.M., Ospina, A.M.T. Dominguez-Castaño, P., Benfica, L.F., Correia, E.C.S., et al. (2022). Analysis of residual feed intake in Nellore bulls of different ages, rib eye area, and backfat thickness. *Tropical Animal Health and Production*, 54, 295. https://doi.org/10.1007/s11250-022-03285-6

Englishby, T. M., Banos, G., Moore, K. L., Coffey, M. P., Evans, R. D., Berry, D. P. (2016). Genetic analysis of carcass traits in beef cattle using random regression models. *Journal of Animal Science*, 94(4), 1354–1364. https://doi.org/10.2527/jas.2015-0246

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (with discussion). In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. Smith (Eds.), Bayesian statistics, 4 (pp. 169–193). Oxford, UK: Oxford University Press

Haskell, M. J., Rooke, J. A., Roehe, R., Turner, S. P., Hyslop, J. J., Waterhouse, A., Duthie, C.A. (2018). Relationships between feeding behaviour, activity, dominance, and feed efficiency in finishing beef steers. *Applied Animal Behaviour Science*. https://doi.org/10.1016/j.applanim.2018.10.012

Herd, R.M., and Bishop, S.C. (2000). Genetic variation in residual feed intake and its association with other production traits in British Hereford cattle. Livestock Production Science, 63(2), 111–119. https://doi.org/10.1016/s0301-6226(99)00122-0

Huisman, A. E., Veerkamp, R. F., & van Arendonk, J. A. M. (2002). Genetic parameters for various random regression models to describe the weight data of pigs. *Journal of Animal Science*, 80(3), 575–582. https://doi.org/10.2527/2002.803575x

Koch, R. M., Swiger, L. A., Chambers, D., Gregory, K. E. (1963). Efficiency of Feed Use in Beef Cattle. *Journal of Animal Science*, 22(2), 486–494. https://doi.org/10.2527/jas1963.222486x

Lancaster, P. A., Carstens, G. E., Crews, D. H., Welsh, T. H., Forbes, T. D. A., Forrest, D. W. Rouquette, F. M. (2009). Phenotypic and genetic relationships of residual feed intake with performance and ultrasound carcass traits in Brangus heifers1. *Journal of Animal Science*, 87(12), 3887–3896. https://doi.org/10.2527/jas.2009-2041

Lourenco, D., Legarra, A., Tsuruta, S., Masuda, Y., Aguilar, I., Misztal, I. (2020). Single-Step Genomic Evaluations from Theory to Practice: Using SNP Chips and Sequence Data in BLUPF90. *Genes*, 11(7), 790. https://doi.org/10.3390/genes11070790

Manzanilla-Pech, C.I.V., Stephansen, R.B., Difford, G.F., Løvendahl, P., Lassen, J. (2022). Selecting for feed efficient cows will help to reduce methane gas emissions. *Frontiers in Genetics*, 13, 885932. https://doi.org/10.3389/fgene.2022.885932.

Meyer, K. (2005). Random regression analyses using B-splines to model growth of Australian Angus cattle. *Genetics Selection Evolution*, 37: 1. https://doi.org/10.1186/1297-9686-37-6-473

Misztal, I., S. Tsuruta, D.A.L. Lourenco, I. Aguilar, A. Legarra, and Z. Vitezica. 2014. Manual for BLUPF90 family of programs. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all1.pdf. (Accessed 5 October 2022.)

Novomestky, F. (2013). Orthopolynom: Collection of functions for orthogonal and orthonormal polynomials, R package version 1.0-5. https://CRAN.R-project.org/package=orthopolynom [Accessed April 10, 2022].

Neave, H.W., Costa, J.H.C., Weary, D.M., von Keyserlingk, M.A.G. (2018). Personality is associated with feeding behavior and performance in dairy calves. *Journal of Dairy Science*, 101, 7437–7449. https://doi.org/10.3168/jds.2017-14248

Oliveira, H. R., Brito, L. F., Lourenco, D. A. L., Silva, F. F., Jamrozik, J., Schaeffer, L. R., Schenkel, F. S. (2019). Invited review: Advances and applications of random regression models: From quantitative genetics to genomics. *Journal of Dairy Science*, 102, 7664-7683. https://doi.org/10.3168/jds.2019-16265

Parra, F. S., Ronchesel, J. R., Martins, C. L., Perdigão, A., Pereira, M. C. S., Millen, D. D., Arrigoni, M. D. B. (2019). Nellore bulls in Brazilian feedlots can be safely adapted to high-concentrate diets using 14-day restriction and step-up protocols. *Animal Production Science*, 59, 1858-1867. https://doi.org/10.1071/an18207

Polizel, G. H. G., Grigoletto, L., Carvalho, M. E., Rossi Junior, P., Ferraz, J. B. S., Santana, M. H. (2018). Genetic correlations and heritability estimates for dry matter intake, weight gain and feed efficiency of Nellore cattle in feedlot. *Livestock Science*, 214, 209–210. https://doi.org/10.1016/j.livsci.2018.06.013

R Core Team (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. https:// www.R-project.org/ [Accessed April 5, 2022].

Rice, J. A., Wu, C. O. (2001). Nonparametric Mixed Effects Models for Unequally Sampled Noisy Curves. *Biometrics*, 57(1), 253–259. https://doi.org/10.1111/j.0006-341x.2001.00253.x

Robinson, D.L. and Oddy, V. H. (2004). Genetic parameters for feed efficiency, fatness, muscle area and feeding behaviour of feedlot finished beef cattle. *Livestock Production Science*, 90(2-3), 255–270. https://doi.org/10.1016/j.livprodsci.2004.06.

Sargolzaei, M., Chesnais, J. P., Schenkel, F. S. (2014). A new approach for efficient genotype imputation using information from relatives. *BMC Genomics*, 15(1), 478. https://doi.org/10.1186/1471 -2164-15-478

Souza, F.M., Lopes, F.B., Rosa, G.J.M., Magnabosco, C.U. (2021). Economic values of reproductive, growth, feed efficiency and carcass traits in Nellore cattle. *Journal of Animal Breeding and Genetics*, 139, 170–180. https://doi.org/10.1111/jbg.12652.

Speidel, S. E., Enns, R. M., Crews Jr, D.H. (2010). Genetic analysis of longitudinal data in beef cattle: A review. *Genetics and Molecular Research*, 9, 19–33. https://doi.org/10.4238/vol9-1gmr675

Tizioto, P. C., Coutinho, L. L., Oliveira, P. S. N., Cesar, A. S. M., Diniz, W. J. S., Lima, et al. (2016). Gene expression differences in Longissimus muscle of Nelore steers genetically divergent for residual feed intake. *Scientific Reports*, 6:39493. https://doi.org/10.1038/srep39493

Van Eenennaam, A.L., Weigel, K.A., Young, A.E., Cleveland, M.A., Dekkers, J.C.M. (2014). Applied Animal Genomics: Results from the Field. *Annual Review of Animal Biosciences*, 2(1), 105–139. https://doi.org/10.1146/annurev-animal-022513-114119

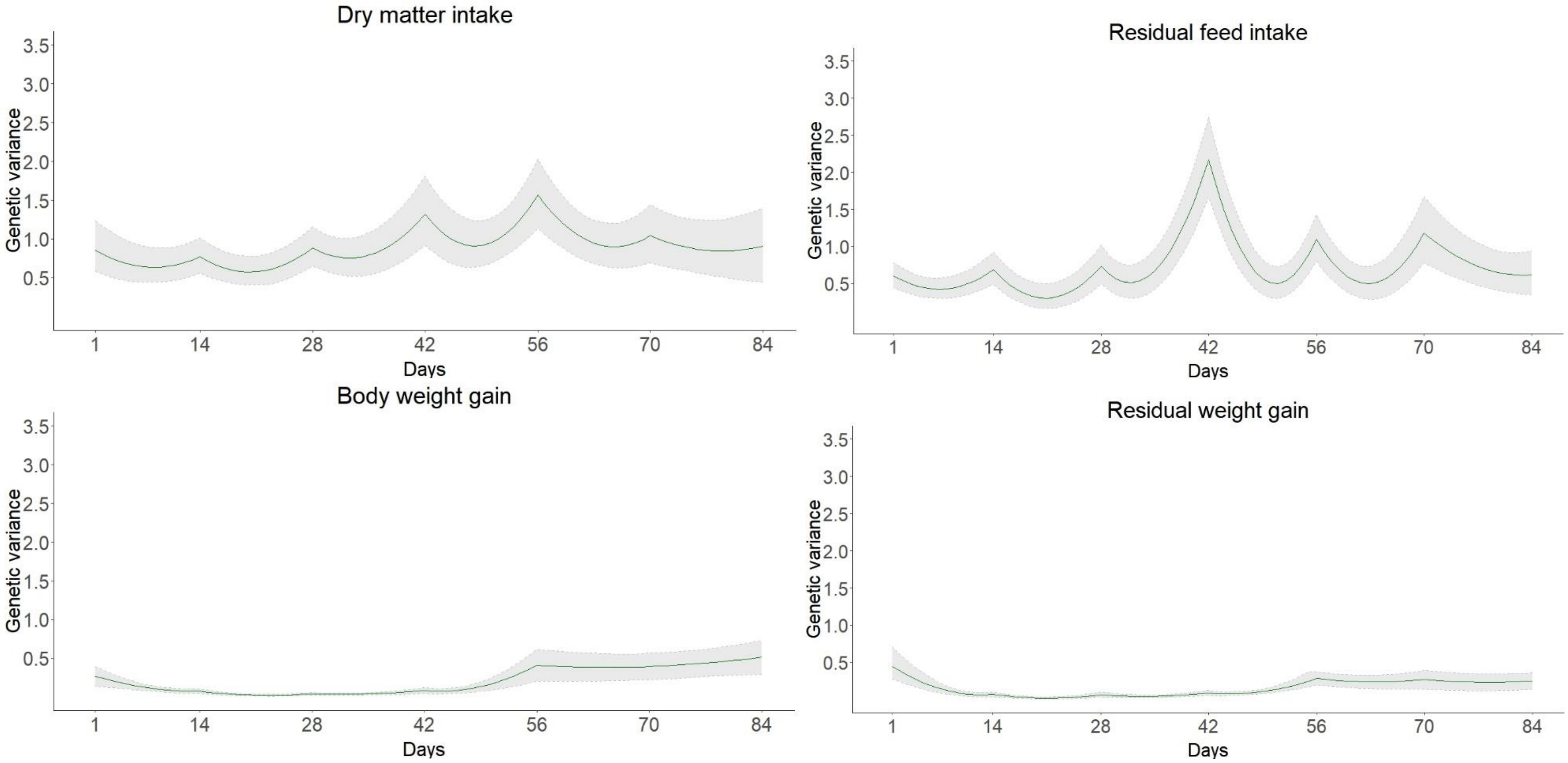
Wang, Y., Diao, C., Kang, H., Hao, W., Mrode, R., Chen, J., Liu, J., Zhou, L. (2021) A Random Regression Model Based on a Single-Step Method for Improving the Genomic Prediction Accuracy of Residual Feed Intake in Pigs. *Frontiers in Genetics*, 12: 769849. doi:10.3389/fgene.2021.769849.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, 91(11), 4414–4423. https://doi.org/10.3168/jds.2007-0980

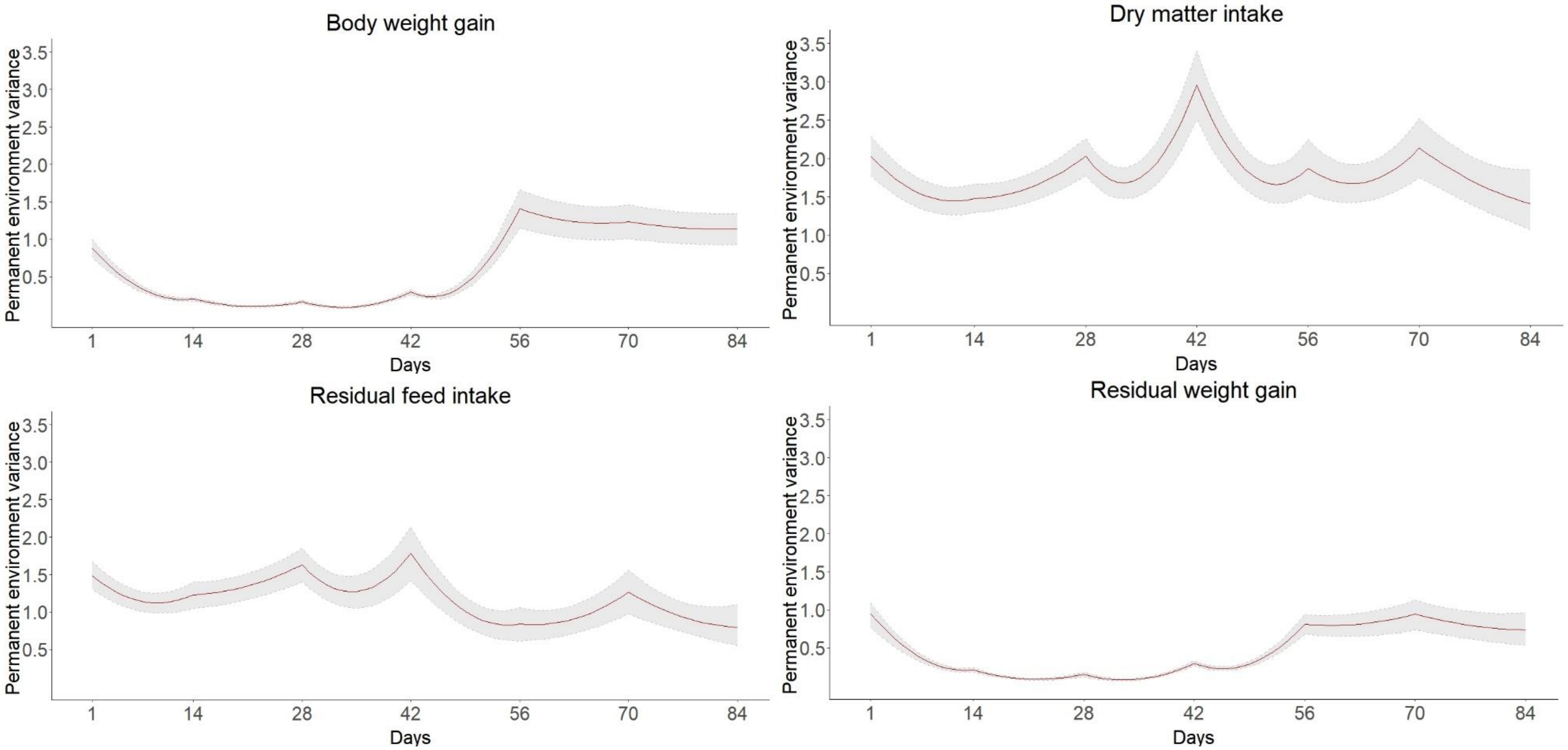
Torres-Vázquez, J.A., Duijvesteijn, N., van der Werf, J.H.J., Clark, S.A. (2019). Longitudinal analysis of body weight and average daily feed intake during the feedlot test period in Angus cattle. *Journal of Animal Breeding and Genetics*, 137(3):281-291. https://doi.org/10.1111/jbg.12439

Von Keyserlingk, M. A. G., & Weary, D. M. (2010). Review: Feeding behaviour of dairy cattle: Meaures and applications. *Canadian Journal of Animal Science*, 90(3), 303–309. https://doi.org/10.4141/cjas09127

2.6 Supplementary Material



Supl 1. Posterior means and 95% highest probability density intervals (HPD95) for genetic variance over test-days, for dry matter intake, residual feed intake, body weight gain and residual weight gain in Nellore cattle.



Supl 2. Posterior means and 95% highest probability density intervals (HPD95) for permanent environment variance over test-days, for dry matter intake, residual feed intake, body weight gain and residual weight gain in Nellore cattle.

CHAPTER 3

Comparing algorithms to approximate accuracies for single-step genomic best linear unbiased predictor

3.1 Abstract

The theoretical accuracy of breeding values can be calculated based on the prediction error variances obtained from the diagonal of the inverse of the left-hand side (LHS) of the mixed model equations (MME). However, inverting the LHS is not computationally feasible for an extensive system of equations, especially if genomic information is available. Thus, different algorithms to approximate accuracies have been proposed. The objectives of this study were to: 1) compare the approximated accuracies from two algorithms implemented in the BLUPF90 suite of programs; 2) compare the approximated accuracies from the two algorithms against the exact accuracy based on the inversion of the LHS of MME; 3) Evaluate the impact of adding genotyped animals with and without phenotypes on the exact and approximated accuracies. Algorithm 1 approximates accuracies based on the diagonal of the genomic relationship matrix ($\mathbf{G}$). In turn, Algorithm 2 uses block sparse inversion of $\mathbf{G^{-1}}$. The Data for this study were provided by the American Angus Association and included three datasets of growth, carcass, and marbling traits. The genotype files contained 1,235,930 animals, and the pedigree file contained 12,492,581 animals. For the genomic evaluations, a multi-trait model was applied to the datasets. To ensure the feasibility of inverting the LHS of the MME, a subset of data under single-trait models was used to compare approximated and exact accuracies. The correlations between exact accuracies and approximated accuracies from Algorithms 1 and 2 of genotyped animals ranged from 0.87 to 0.90 and 0.98 to 0.99, respectively. The intercept and slope of the regression of exact on approximated accuracies from Algorithm 2 ranged from 0.00 to 0.01 and 0.82 to 0.87, respectively. However, the intercept and the slope for Algorithm 1 ranged from -0.10 to 0.05 and 0.98 to 1.10, respectively. In more than 80% of the traits, Algorithm 2 exhibited smaller mean square error than Algorithm 1. The correlation between the approximated accuracies obtained from Algorithms 1 and 2 ranged from 0.56 to 0.74, 0.38 to 0.71, and 0.71 to 0.97 in the groups of genotyped animals, genotyped animals without phenotype, and proven genotyped sires, respectively. The approximated accuracy from Algorithm 2 presented closer behavior to the exact when including genotyped animals in the analysis. According to the results, Algorithm 2 is recommended for genetic evaluations since it proved more precise.

Key words: accuracy approximation, BIF accuracy, genomic evaluation, prediction error variance

3.2 Introduction

In animal breeding programs, the accuracy of estimated breeding values (EBV) is calculated to help make selection decisions, representing the correlation between the true breeding value and EBV (Wilmink & Dommerholt, 1985). The theoretical accuracy can be calculated based on the prediction error variances (PEV) obtained from the diagonal of the inverse of the left-hand side of the mixed model equations (MME - Henderson, 1984). However, inverting the coefficient matrix is not computationally feasible for an extensive system of equations. In this context, methods for approximating PEV have been proposed and implemented for pedigree-based evaluations (Misztal & Wiggans, 1988; Harris and Johnson, 1998; Liu et al., 2004; Tier and Meyer, 2004) by exploiting the sparse structure of the animal model (Quaas, 1976). These methods are based on calculating the animals' effective record contribution (ERC), accounting for non-genetics effects in the LHS of MME and calculating the accuracy approximation by combining the ERC with the pedigree-based relationship matrix.

However, when genomic information is included, the sparse structure of the MME is compromised, and obtaining PEV becomes more challenging. For evaluations where all animals are genotyped, genomic predictions can be calculated by the genomic BLUP (GBLUP) and the single-nucleotide polymorphism BLUP (SNPBLUP- Liu et al., 2014) models. In the GBLUP, the size of MME increases with the number of genotyped animals, whereas in SNPBLUP, the dimension of the set of equations remains constant with the number of genotyped animals but changes with the number of SNP. For the simultaneous analysis of genotyped and non-genotyped animals, the genomic and pedigree information can be combined in a single model through single-step GBLUP (ssGBLUP; Aguilar et al., 2010) and single-step SNP-BLUP (ssSNPBLUP; Liu et al., 2014). In this case, the calculation of PEV by inverting the MME is also prohibitively expensive for large datasets since it presents dense matrix blocks in the MME, represented by genomic relationships.

Misztal et al. (2013) introduced an algorithm to approximate accuracies of estimated genomic breeding values (GEBV) from ssGBLUP, building upon the traditional pedigree-based algorithm outlined in Misztal and Wiggans (1988). They proposed two algorithms that initially approximate the accuracy by disregarding the genomic information. Subsequently, they converted these accuracies into ERC for the genotyped animals. Then, they obtained genomic accuracies as a function of the inverse of the sum of a diagonal matrix of ERC and the inverse of the genomic relationship matrix, removing the double-counting of pedigree relationships for genotyped animals. However, this inverse is not feasible for datasets with many genotyped animals; therefore, the second algorithm considers only the diagonal elements of the inverses of the genomic and pedigree relationship matrices for genotyped animals. The resulting matrix is diagonal, which has a straightforward inverse for any data size, given that the inverses of the genomic and pedigree relationship matrices for genotyped animals can be obtained.

Nonetheless, the two algorithms usually provided overestimated accuracies with real datasets. To address this issue, another algorithm was developed such that the genomic contribution is based on the diagonal of the genomic relationship matrix (Tsuruta et al., 2016), which is then combined with the animals' record and pedigree contributions following Misztal and Wiggans (1988). This approach offers computational advantages when working with large populations and has been implemented in the BLUPF90 suite of programs (Misztal et al., 2014a), specifically in the ACCF90GS software, and has been adopted by the industry.

Liu et al., (2017) proposed a method based on SNPBLUP for approximating the accuracy of GEBV, accounting for the residual polygenic effect for genotyped animals. For non-genotyped animals, genomic information was propagated via ERC. In turn, Edel et al. (2019) introduced a single-step SNPBLUP approach that imputes the genotypes of non-genotyped animals based on the observed genotypes with a simplified pedigree. More recently, Bermann et al. (2022) presented an algorithm for calculating the PEV in the GBLUP model using block sparse inversion of $\mathbf{G}^{-1}$computed by the algorithm of proven and young animals (APY; Misztal et al. 2014b). Furthermore, the authors demonstrated that the described algorithm could be expanded to the ssGBLUP model to approximate accuracies using the ERC procedure, as shown by Liu et al. (2017). Since this approach proved feasible for calculating PEV for GEBV in ssGBLUP for large populations with computational efficiency, it has been implemented in the BLUPF90 suite of programs, in the ACCF90GS2 software.

Approximating contributions from genomic information to the accuracy of GEBV is more complex than the contributions from pedigree and phenotypes. Often, the lack of precision leads to overestimated accuracies, and despite constant improvement, no algorithm has proven to be completely free of bias. Bermann et al. (2022) compared accuracies from their algorithm against the ones from the inverse of the LHS of MME and obtained correlations as strong as 0.98 with minimum bias. However, the performance and precision of their algorithm have not been tested against other methods currently used for ssGBLUP routine genomic evaluations. In this context, the objectives of this study were to: 1) compare approximated accuracies from the algorithms presented by Tsuruta et al. (2016) and Bermann et al. (2022); 2) compare the approximated accuracies from the two algorithms mentioned above against the exact accuracy based on the direct inversion of the LHS of MME; 3) Evaluate the impact of adding genotyped animals with and without phenotypes on the exact and approximated accuracies of GEBV. The latter can help verify if both algorithms correctly account for the information on genotyped animals whether they have phenotypes or not.

3.3 Material and Methods

3.3.1 Data and analyses

The American Angus Association (Saint Joseph, MO) provided the datasets for this study. DATA_GT included phenotypes for growth traits: birth weight (BW), weaning weight (WW), and post-weaning gain (PWG). DATA_CT included phenotypes for carcass traits: rib eye area (RIB), carcass weight (CWT), weight at ultrasound time (USWT), ultrasound rib eye area of bulls (URIB_B), ultrasound rib eye area of cows and steers (URIB_CS), fat thickness (FAT), ultrasound fat of bulls (UFAT_B), and ultrasound fat of cows and steers (UFAT_CS). Finally, DATA_MT included marbling traits: marbling (MARB), ultrasound intramuscular fat of bulls (UIMFB), ultrasound intramuscular fat of cows (UIMFC), and ultrasound intramuscular fat of steers (UIMFS). The genotype files contained 1,235,930 animals genotyped for 39,707 SNPs, and the complete pedigree file contained 12,492,581 animals. A subset of each dataset (DATA_sub) was created to compute the accuracy based on the inverse of the LHS of MME, which we will refer to as “exact accuracy” from now on. The number of animals with records, in the pedigree, and with genotypes for both the datasets and the subset data are presented in Table 1.

Table 1. Number of animals in the pedigree (Nº pedigree), number of animals with records (Nº records), and number of genotyped animals (Nº genotyped), mean, and standard deviations (SD) for growth and carcass traits in each dataset.

	Trait[1]	Nº pedigree	Nº records	Nº genotyped	Mean	SD
DATA_GT[2]	BW	12,538,121	9,771,567	1,235,930	80.44	9.85
	WW	12,538,121	10,135,518	1,235,930	595.95	99.66
	PWG	12,538,121	4,951,956	1,235,930	363.56	148.18
DATA_CT[2]	CWT	5,672,541	132,248	1,235,930	792.33	96.82
	FAT	5,672,541	132,240	1,235,930	0.57	0.18
	RIB	5,672,541	132,248	1,235,930	12.70	1.45
	UFAT_B	5,672,541	1,367,115	1,235,930	0.28	0.10
	UFAT_CS	5,672,541	897,188	1,235,930	0.27	0.11
	URIB_B	5,672,541	1,365,090	1,235,930	12,64	1.81
	URIB_CS	5,672,541	895,451	1,235,930	9.85	1.71
	USWT	5,672,541	2,649,080	1,235,930	1018.39	183.01
DATA_MT[2]	MARB	4,362,918	132,248	1,235,930	6.26	1.21
	UIMFB	4,362,918	1,363,830	1,235,930	3.84	1.11
	UIMFC	4,362,918	876,746	1,235,930	4.89	1.36
	UIMFS	4,362,918	14,048	1,235,930	5.02	1.36
DATA_SUB[2]	BW	177,644	31,277	30,000	80.39	8.88
	WW	223,341	35,194	30,000	632.30	96.83
	PWG	534,327	101,837	50,000	393.00	149.44
	CWT	485,315	132,248	50,000	792.30	96.82
	FAT	485,826	132,240	50,000	0.57	0.18
	RIB	485,711	132,248	50,000	12.70	1.45
	UFAT_B	474,832	125,708	50,000	0.27	0.11
	UFAT_CS	401,143	117,531	50,000	0.26	0.11
	URIB_B	473,844	125,685	50,000	12.85	1.86
	URIB_CS	400,577	117,520	50,000	10.01	1.80
	USWT	417,089	86,660	50,000	1051.40	190.86
	MARB	485,086	132,248	50,000	6.26	1.21
	UIMFB	475,203	125,702	50,000	3.82	1.43
	UIMFC	398,516	116,222	50,000	5.25	1.63
	UIMFS	267,708	14,048	50,000	5.02	1.36

[1] BW, birth weight; WW, weaning weight ; PWG, post-weaning gain ; CWT, carcass weight; FAT, fat thickness; RIB, rib eye area; URIB_B, ultrasound rib eye area of bulls; USWT, weight at ultrasound time; URIB_CS, ultrasound rib eye area of cows and steers; UFAT_B, ultrasound fat of bulls; UFAT_CS, ultrasound fat of cows and steers; MARB, marbling; UIMFB, ultrasound intramuscular fat of bulls; UIMFC, ultrasound intramuscular fat of cows; UIMFS, ultrasound intramuscular fat of steers.

[2] DATA_GT, Data of growth traits used for multi-trait analysis; DATA_CT, Data of carcass traits used for multi-trait analysis, DATA_MT, Data of marbling traits used for multi-trait analysis; DATA_SUB, subset data used for single-trait analyses and exact accuracy computation.

For the genomic evaluations, a multi-trait model was applied to DATA_GT, DATA_CT, and DATA_MT. Single-trait models were used when dealing with DATA_sub to make sure the inverse of the LHS of MME was feasible. The matrix notation of the general statistical model for single- and multi-trait models is as follows:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{Wm} + \mathbf{Smpe} + \mathbf{e}$$

where **y** is the vector of observations, **b** is the vector of fixed effects (contemporary group); **u** is the vector of random additive genetic effects; **m** is the vector of random maternal additive genetic effects (assumed for birth and weaning weight); **mpe** is the vector of random maternal permanent environmental effects (assumed only for weaning weight); **e** is the vector of random residuals; **X**, **Z**, **W** and **S** are the incidence matrices for the effects in **b**, **u**, **m**, and **mpe** respectively. The genomic evaluations were implemented using ssGBLUP, with the inverse of a relationship matrix that combines pedigree and genomic information ($\mathbf{H}^{-1}$), according to Aguilar et al. (2010):

$$\mathbf{H}^{-1}=\mathbf{A}^{-1}+\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{G}^{-1}-\mathbf{A}_{22}^{-1}\end{bmatrix} \qquad [1]$$

where $\mathbf{A}_{22}^{-1}$ is the inverse of the numerator relationship matrix for genotyped animals; $\mathbf{G}^{-1}$ is the inverse of the genomic relationship matrix ($\mathbf{G}$) constructed based on the type I described by VanRaden (2008). Because of the large number of genotyped animals in DATA_GT, DATA_CT, and DATA_MT, the algorithm for proven and young (APY; Misztal et al. 2014b) was applied to obtain a sparse representation of $\mathbf{G}^{-1}$ without having to invert $\mathbf{G}$. Therefore, $\mathbf{G}^{-1}$ in [1] was replaced by the $\mathbf{G}_{\mathrm{APY}}^{-1}$ constructed as below:

$$\mathbf{G}_{\mathrm{APY}}^{-1}=\begin{bmatrix}\mathbf{I} & -\mathbf{P}_{\mathrm{cn}}\\ 0 & \mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{G}_{\mathrm{cc}}^{-1} & 0\\ 0 & \mathbf{M}_{\mathrm{nn}}^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{I} & 0\\ -\mathbf{P}_{\mathrm{nc}} & \mathbf{I}\end{bmatrix}=\begin{bmatrix}\mathbf{G}^{\mathrm{cc}} & \mathbf{G}^{\mathrm{cn}}\\ \mathbf{G}^{\mathrm{nc}} & \mathbf{M}_{\mathrm{nn}}^{-1}\end{bmatrix} \qquad [2]$$

where the subscript c and n represent the core and noncore animals, respectively; $\mathbf{P}_{\mathrm{nc}} = \mathbf{G}_{\mathrm{nc}}\mathbf{G}_{\mathrm{cc}}^{-1}$; $\mathbf{P}_{\mathrm{cn}} = \mathbf{G}_{\mathrm{cc}}^{-1}\mathbf{G}_{\mathrm{cn}}$; and $\mathbf{M}_{\mathrm{nn}} = \mathrm{diag}\,(\mathbf{G}_{\mathrm{nn}} - \mathbf{G}_{\mathrm{nc}}\mathbf{G}_{\mathrm{cc}}^{-1}\mathbf{G}_{\mathrm{cn}})$.

The algorithm behind APY is based on the number of genotyped core animals representing the independent chromosome segments in the genome (Misztal, 2016). In this population, 22,000 genotyped animals were randomly selected as core based on the number of largest eigenvalues explaining 98% of the variance in the spectrum of the genomic relationship matrix (Pocrnic et al., 2016). The number of core animals was the same across all three full datasets, as well as the number of SNPs.

After the genomic evaluations, BIF (Beef Improvement Federation) accuracies were calculated as:

$$BIF_{accuracy} = 1 - \sqrt{\frac{PEV}{\sigma_u^2}} \qquad [3]$$

where σ_u^2 is the additive genetic variance. For DATA_sub, $BIF_{accuracy}$ was computed from PEV obtained based on the inverse of the LHS of MME.

The approximated accuracies tested in this study were proposed by Tsuruta et al. (2016) and Bermann et al. (2022) and implemented in ACCF90GS and ACCF90GS2 software from the BLUPF90 family (Misztal et al., 2014a). From now on, the algorithm proposed by Tsuruta et al. (2016) will be referred to as Algorithm 1, and the one presented by Bermann et al. (2022) will be the Algorithm 2. Both algorithms are detailed below.

3.3.2 Algorithm 1

Algorithm 1 was published as a conference abstract, so the detailed equations had never been shown before. This Algorithm approximates the genomic contribution based on the diagonal of the genomic relationship matrix, which is added to the phenotype and pedigree contributions obtained according to Misztal and Wiggans (1988). Based on phenotypes and pedigree, the reliability of animal i (REL$_i$) can be approximated as $1 - [\frac{\alpha}{\alpha + d_i}]$, where α is the ratio of residual to animal genetic variance and d$_i$ is the information in units of the effective number of records of the animal i. The information can be calculated by the inversion of the LHS of the mixed model equation as $LHS_{uu}^{ii} = 1 / (\alpha + d_i)$, where uu represents the block of the LHS for the animal effect and *ii* represents the diagonal element of the animal i. Thus, d_i can be portioned as $d_i^r + d_i^p$ when there is no genomic information, where d_i^r is the contribution from phenotypes d_i^p is the contribution from the pedigree.

Assuming the single-trait model:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e},$$

where **y** is a vector of phenotypes, **b** is the vector of fixed effects, **u** is the random additive animal effect, **X** and **Z** are incidence matrices of **b** and **u**, respectively, and **e** is the random residual effect. The LHS can be represented as:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{A}^{-1}\boldsymbol{\alpha} \end{bmatrix}$$

Where **A** is the numerator relationship matrix. So, the PEV of $\hat{u}_i$ is:

$$\mathrm{PEV}(\hat{\mathrm{u}}_\mathrm{i}) = \mathrm{LHS}_{\mathrm{uu}}^{\mathrm{ii}}\,\sigma_e^2$$

Where, $\mathrm{LHS}_{\mathrm{uu}}^{\mathrm{ii}}$ can be represented as:

$$\mathrm{LHS}_{\mathrm{uu}}^{\mathrm{ii}} = 1\,/(\alpha + d_i^r + d_i^p) \qquad [4]$$

If it is assumed that $\mathbf{D}^\mathrm{r}=\{\mathrm{d}_\mathrm{i}^\mathrm{r}\}$ and $\mathbf{D}^\mathrm{p}= \{\mathrm{d}_\mathrm{i}^\mathrm{p}\}$ are known. The equation [4] can be approximated as:

$$\mathrm{LHS}_{\mathrm{uu}}^{\mathrm{ii}} \approx \left[\, (\mathbf{D}_\mathbf{i}^\mathbf{r} + \mathbf{D}_\mathbf{i}^\mathbf{p} + \mathbf{I}\boldsymbol{\alpha})^{-1}\, \right]_{\mathrm{ii}}$$

3.3.2.1 Contribution due to records

Suppose that animals are unrelated. Then $\mathrm{LHS}_{\mathrm{uu}}^{\mathrm{ii}}$ becomes:

$$\mathrm{LHS}_{\mathrm{uu}}^{\mathrm{ii}} \approx [\, (\mathbf{D}_\mathbf{i}^\mathbf{r} + \mathbf{I}\boldsymbol{\alpha})^{-1}\,]_{\mathrm{ii}} \qquad [5]$$

Thus, the record contribution equals the effective number of records. If all effects are absorbed into animal effects (u) and animals are assumed unrelated, the equation [5] should be equal to:

$$[\mathbf{Z'MZ} + \mathbf{I}\alpha]^{-1}$$

where **M** is an absorption matrix.

3.3.2.2 Contributions due to relationships

The contributions from relationships can be obtained by an iterative formula according to Misztal and Wiggans (1988) as follows:

$$\begin{bmatrix} 1.5\alpha + d_s - d^r_{s_i} & 0.5\alpha & -\alpha \\ 0.5\alpha & 1.5\alpha + d_d - d^r_{d_i} & -\alpha \\ -\alpha & -\alpha & 2\alpha + d_i - d^r_{i_s} - d^r_{i_d} \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 1/(\alpha + d_s) & \cdots & \cdots \\ \cdots & 1/(\alpha + d_d) & \cdots \\ \cdots & \cdots & 1/(\alpha + d_i) \end{bmatrix}$$

Where d_i, d_s, and d_d are information from the animal *i* and its sire (s) and dam (d), respectively; $d^r_{s_i}$ and $d^r_{d_i}$ are the contributions to sire and dam information from records of animal i, respectively; and $d^r_{i_s}$ and $d^r_{i_d}$ are contributions to information for animal i from records of its sire and dam, respectively.

3.3.2.3 Genomic contribution

When genomic information is available as presented in [1], it is included in the diagonal elements of the inverse of the LHS for animal i: $\mathrm{LHS}^{\mathrm{ii}}_{\mathrm{uu}} = 1/(\alpha + d^r_i + d^p_i + d^g_i)$. If $\mathbf{D}^{\mathrm{r}}$ and $\mathbf{D}^{\mathrm{p}}$ are known, the LHS can be approximated for ssGBLUP models as proposed by Misztal et al. (2013):

$$\mathbf{LHS}^{\mathrm{ii}}_{\mathrm{uu}} \approx \left\{ \left[\mathbf{D}^{\mathbf{r}}_{\mathbf{i}} + \mathbf{D}^{\mathbf{p}}_{\mathbf{i}} + (\mathbf{I} + \mathbf{G}^{-1} - \mathbf{A}^{-1}_{22})\alpha \right]^{-1} \right\}_{\mathrm{ii}}$$

However, this approach has limitations due to the computational demand since it requires inverting both the genomic and pedigree-based relationship matrices. Therefore, in Algorithm 1, the genomic contribution is approximated based on the diagonal of the **G** as:

$$d^g_i = \alpha * \left[\overline{\mathrm{REL}} + (1 - \mathrm{g}_{\mathrm{ii}}) + z * \left(\overline{\mathrm{REL} - \mathrm{REL}_{\mathrm{pa}}}\right)\right] \qquad [6]$$

where $\overline{\mathrm{REL}}$ is the average reliability based on pedigree and phenotypes for genotyped animals; gii is the diagonal of **G** for animal *i*; z is an adjustment factor defined as $(N_g \times 1e^{-3})$, with N_g the number of genotyped animals; $\mathrm{REL}_{\mathrm{pa}}$ is the average reliability based on the pedigree information for genotyped animals with phenotypes. Thus, $\overline{\mathrm{REL} - \mathrm{REL}_{\mathrm{pa}}}$ can be defined as the contribution from phenotypes of genotyped animals. While the equation [6] is completely empirical and was obtained based on dairy and beef cattle datasets, it provides a good approximation for the genomic contribution and has been widely used in several species.

Additionally, the computing time is minimal and correlations between accuracies from Algorithm 1 and the exact accuracies are usually greater than 0.85 (Lourenco et al., 2018).

3.3.3 Algorithm 2

The Algorithm 2 was proposed by Bermann et al. (2022) and it calculates the PEV for the genotyped animals in the GBLUP model using a block sparse inversion of $\mathbf{G}_{\mathrm{APY}}^{-1}$ after adding effective record contributions as weights to its diagonal (see below). The method calculates the PEV in GBLUP as:

$$\mathrm{PEV} = \mathrm{diag}(\mathbf{D} + \mathbf{G}_{\mathrm{APY}}^{-1}\alpha)^{-1} \qquad [7]$$

where $\mathbf{D}$ can be represented as in VanRaden and Freeman (1985) and Misztal and Wiggans (1988):

$$\mathbf{D} \approx \mathbf{Z}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Z} \qquad [8]$$

where $\mathbf{X}$ and $\mathbf{Z}$ are design matrices for the fixed and random effects, respectively. Therefore, the algorithm does not require setting up the MME, but only approximating weights to be added to the diagonal elements of the $\mathbf{G}_{\mathrm{APY}}^{-1}$ matrix. The $\mathbf{G}_{\mathrm{APY}}^{-1}$ is created according to the block implementation in Masuda et al. (2016) and the equation [7] is obtained through an algorithm for calculating block sparse inverse, following Henderson and Searle (1981). The steps for calculating the PEV, according to Bermann et al. (2022) are:

1. Read $\mathbf{G}_{\mathrm{APY}}^{-1}$ from the disk and store it in memory.
2. Approximate $\mathbf{D}$ as in equation [8] and overwrite $\mathbf{G}_{\mathrm{APY}}^{-1}$ as $\mathbf{G}_{\mathrm{APY}}^{-1} = (\mathbf{D} + \mathbf{G}_{\mathrm{APY}}^{-1}\alpha)$.
3. Calculate $\mathbf{G}^{\mathrm{cn}^*} = \mathbf{G}^{\mathrm{cn}}\mathbf{M}_{\mathrm{nn}}$.
4. Overwrite $\mathbf{G}^{\mathrm{cc}}$ as $\mathbf{G}^{\mathrm{cc}}=\mathbf{G}^{\mathrm{cc}} - \mathbf{G}^{\mathrm{cn}}\left(\mathbf{G}^{\mathrm{cn}^*}\right)'$.
5. Directly invert $\mathbf{G}^{\mathrm{cc}}$.
6. Overwrite $\mathbf{G}^{\mathrm{cn}}$ as $\mathbf{G}^{\mathrm{cn}} = -\mathbf{G}^{\mathrm{cc}}\mathbf{G}^{\mathrm{cn}^*}$.
7. Overwrite $\mathbf{M}_{\mathrm{nn}}^{-1}$ as $\mathbf{M}_{\mathrm{nn}_i}^{-1} = \mathbf{M}_{\mathrm{nn}_i} - \left(\mathbf{G}_{.\mathrm{i}}^{\mathrm{cn}^*}\right)' \mathbf{G}_{.\mathrm{i}}^{\mathrm{cn}^*}$, where the subscript i refers to the i^{th}element of $\mathbf{M}_{\mathrm{nn}}^{-1}$ and $\mathbf{M}_{\mathrm{nn}}$, or the i^{th} row of $\mathbf{G}^{\mathrm{cn}^*}$and $\mathbf{G}^{\mathrm{cn}}$.
8. Obtain PEV from diag ($\mathbf{G}^{\mathrm{cc}}$) and $\mathbf{M}_{\mathrm{nn}}^{-1}$ for core and noncore animals, respectively.

The Algorithm also allows for replacing $\mathbf{G}_{\mathrm{APY}}^{-1}$ with the $\mathbf{G}^{-1}$ in [7], when the number of genotyped animals is less than 100k. In such a case, a dense inversion increases the computing cost with the number of genotyped animals. Genomic contributions are not propagated to non-genotyped animals based on ssGBLUP because the increase in accuracy for such animals is small (i.e., on average 0.02, Bermann et al. 2022) and they are less likely to be selected.

3.3.4 Algorithm comparison

Comparisons between exact and approximated accuracies included the intercept and slope of the regression of the exact accuracies on the approximated accuracies of genotyped animals obtained through Algorithms 1 and 2, as well as the correlation and the mean square error (MSE- predicted as the mean of the square of the difference) between them. The comparison between the algorithms was performed by calculating the intercept and slope of the regression of the approximated accuracies from Algorithm 1 on those obtained from Algorithm 2 and the correlation between them within three different groups of animals: 1= all genotyped animals, 2= genotyped animals born in 2022 without phenotypes; 3= genotyped sires with 100 or more offspring with phenotypes.

Additionally, we examined the changes in exact and approximated accuracies of PWG and URIB_B obtained through single-trait models to assess the impact of adding newly genotyped animals to the analyses. This assessment involved 40,000 genotyped animals, of which 25,000 lacked phenotypic data. Three scenarios were compared: A=inclusion of 10,000 genotyped animals with phenotypes; B= inclusion of 10,000 genotyped animals without phenotypes; C= no inclusion of animals. A descriptive analysis of exact and approximated accuracies was conducted for each scenario. We also compared the scenarios within exact and approximated accuracies, by the correlation, intercept, and slope of the regression of the accuracies obtained in Scenario A on those of Scenario B and the regression of the accuracies obtained in Scenario B on those of Scenario C.

3.4 Results and discussion

3.4.1 Comparing exact and approximated accuracies

Figures 1 and 2 present scatter plots and correlations, b0 and b1, for the exact and approximated accuracies of genotyped animals obtained through Algorithms 1 and 2 for growth traits. The correlations between exact accuracies and those approximations obtained from Algorithms 1 and 2 ranged from 0.87 to 0.90 and 0.99 to 0.99 respectively. Furthermore, the intercept and slope of the regression of exact accuracies on approximated accuracies from Algorithm 2 ranged from 0.00 to 0.01 and 0.82 to 0.87, respectively. However, the intercept and the slope for Algorithm 1 ranged from -0.10 to 0.05 and 0.98 to 1.10, respectively. When the intercept is close to 0 and the slope is lower than 1, it can be assumed that the predictions are overdispersed; if greater than 1, they are underdispersed. When the intercept deviates from 0 and is negative, there is overestimation; if positive, there is underestimation. Both algorithms presented some level of over or underdispersion (i.e., dispersion bias), with Algorithm 1 being more severely affected by under or overestimation (i.e., level bias). Algorithm 2 had a slight overestimation.

In Algorithm 1, one reason for level and dispersion biases was the failure to account for relationships among all genotyped animals, as only the diagonals of $\mathbf{G}$ were used. On the other hand, Algorithm 2 assumes the information of genotyped relatives based on the full $\mathbf{G}^{-1}$ (Berman et al. 2021). Additionally, bias was associated with the step to approximate accuracies based on pedigree and phenotypes (results not shown), which was also the primary source of bias in Algorithm 2. Extra sources of bias may be small or single-sire contemporary groups and missing data.

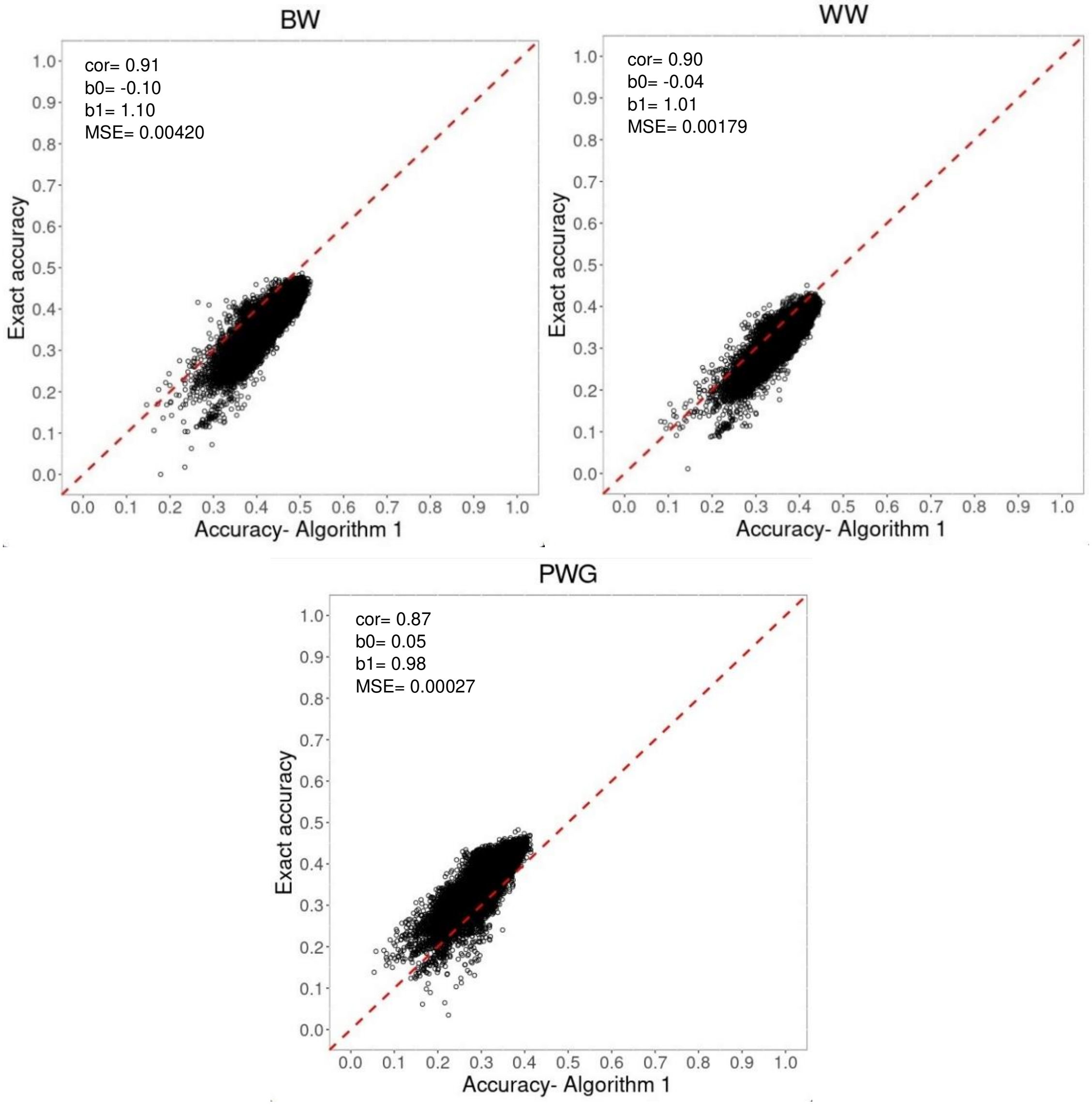


Figure 1. Scatter plots comparing exact against approximated accuracy obtained through Algorithm 1, for growth traits. BW, birth weight; WW, weaning weight; PWG, post-weaning gain.

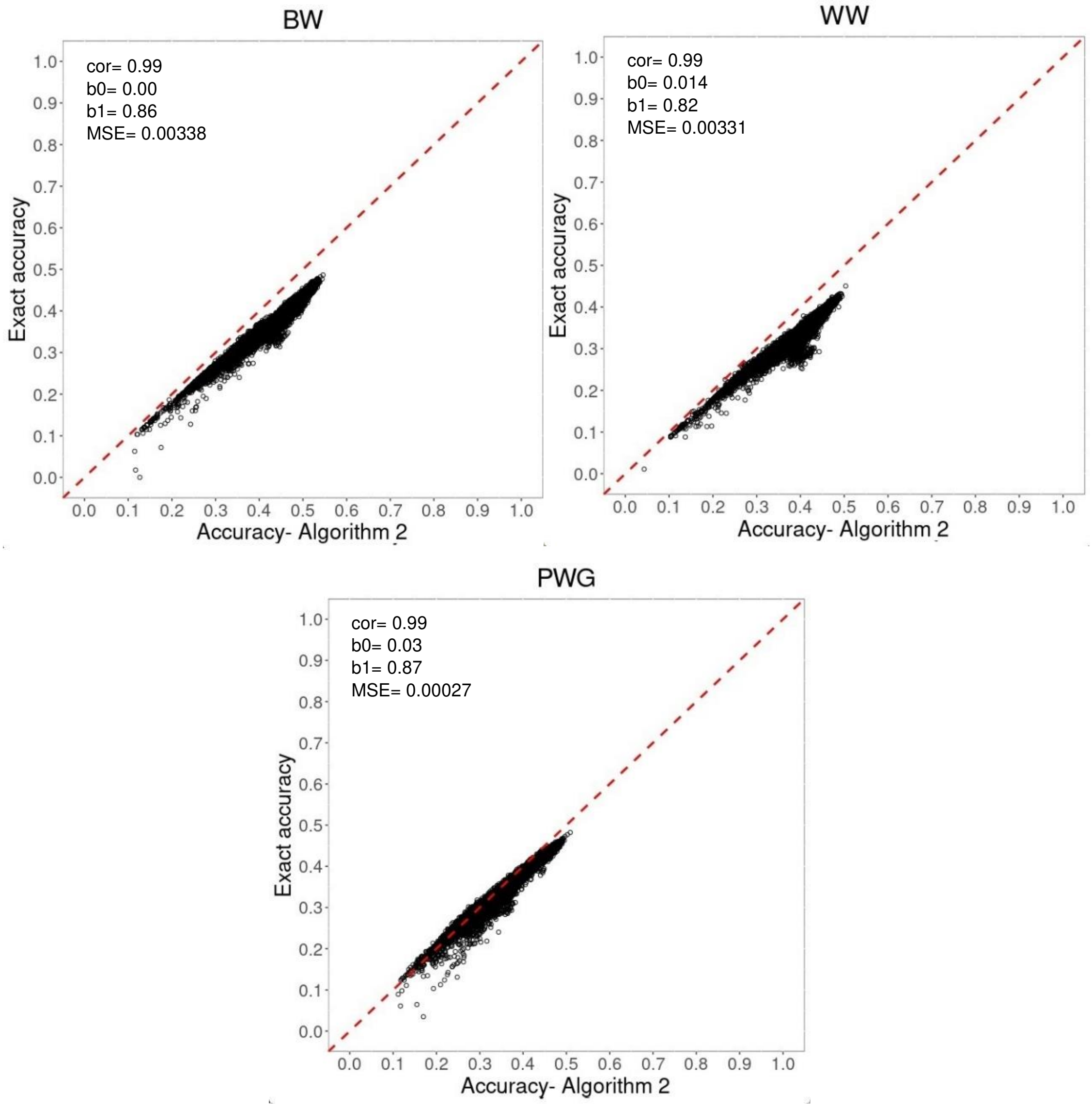


Figure 2. Scatter plots comparing exact against approximated accuracy obtained through Algorithm 2, for growth traits. BW, birth weight (direct effect); WW, weaning weight (direct effect); PWG, post-weaning gain.

Figures 3, 4, 5, and 6 present scatterplots of accuracies for carcass and marbling traits. Both groups of traits presented similar results. The accuracies obtained by Algorithm 2 also presented a higher correlation (ranging from 0.98 to 0.99) with the exact accuracy and a minor overestimation (slope ranging from 0.86 to 0.92) for all traits.

In addition, Algorithm 2 exhibited smaller MSE values than Algorithm 1 in more than 80% of the traits, indicating that Algorithm 2, in general, was also closer to the exact accuracies.

Therefore, Algorithm 2 seems more viable for approximating accuracies in ssGBLUP models. An alternative to further reduce overestimation and mean squared error (MSE), thereby enhancing the accuracy approximation to match exact values closely, entails scaling the approximated accuracies obtained through Algorithm 2. In this context, additional studies are required to determine the optimal adjustment factor for optimizing the results of Algorithm 2.

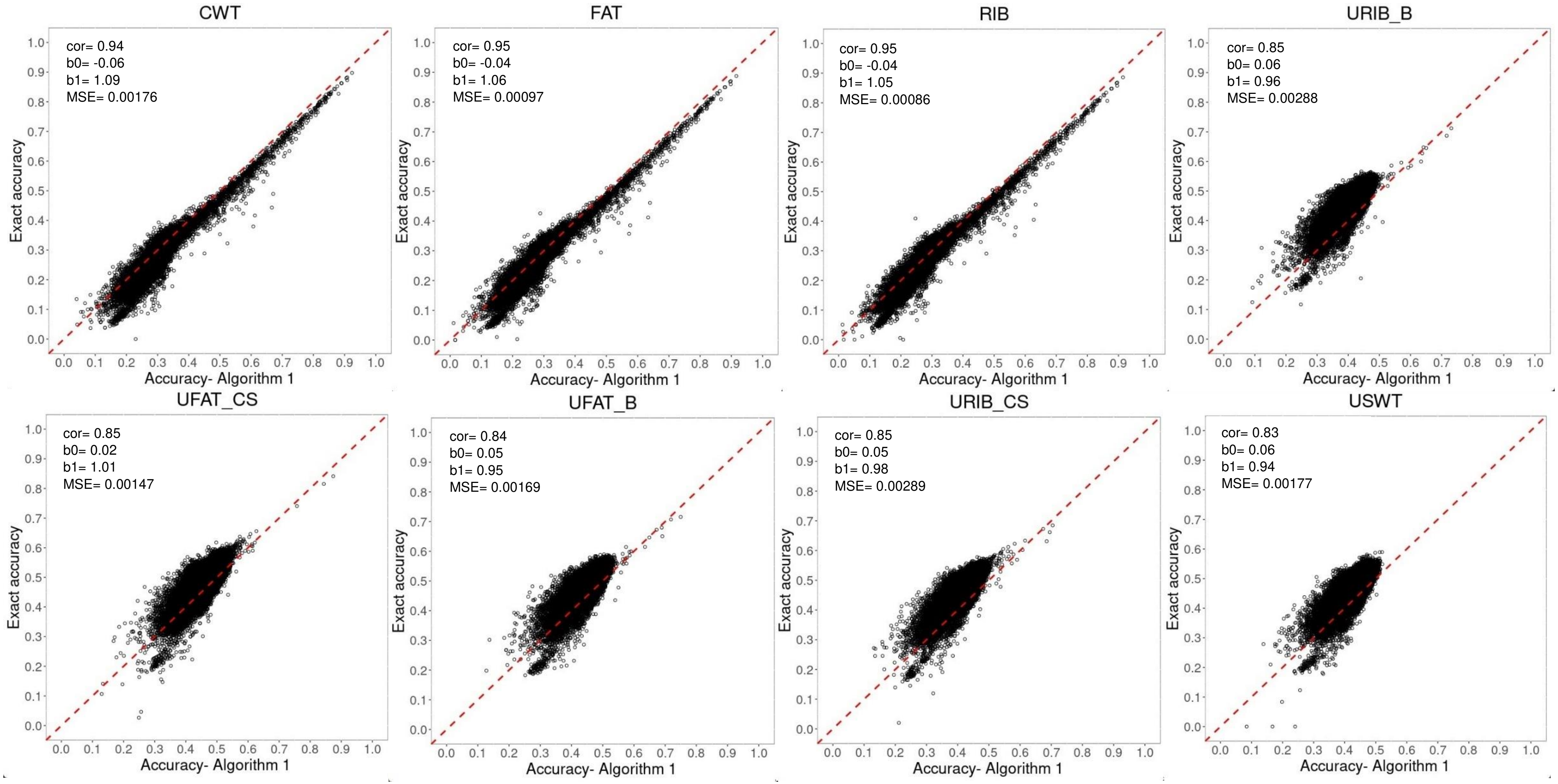


Figure 3. Scatter plots comparing exact against approximated accuracy obtained through Algorithm 1, for carcass traits. CWT, carcass weight; FAT, fat thickness; RIB, rib eye area; URIB_B, ultrasound rib eye area of bulls; USWT, weight at ultrasound time; URIB_CS, ultrasound rib eye area of cows and steers; UFAT_B, ultrasound fat of bulls; UFAT_CS, ultrasound fat of cows and steers.

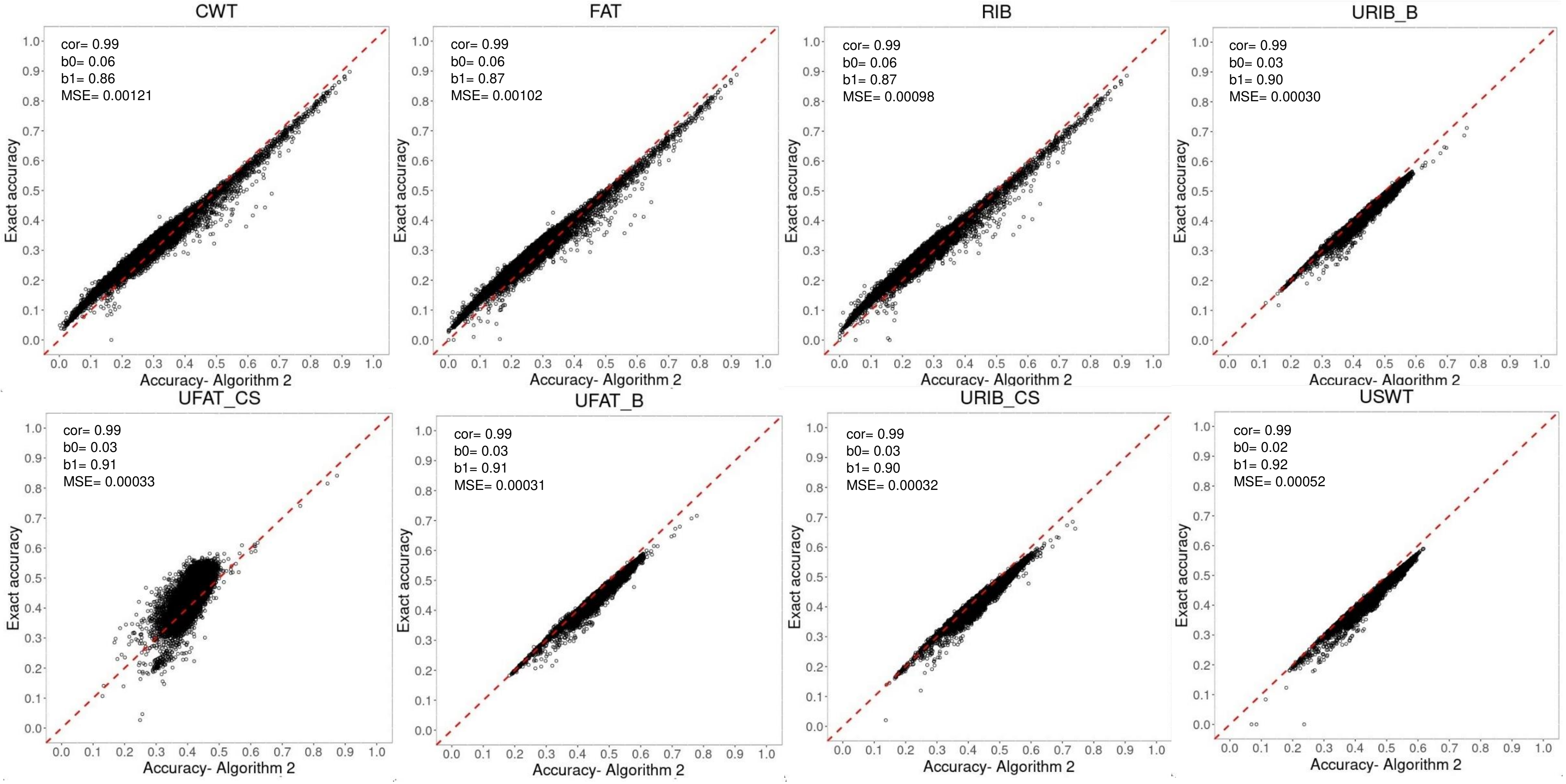


Figure 4. Scatter plots comparing exact against approximated accuracy obtained through Algorithm 2, for carcass traits. CWT, carcass weight; FAT, fat thickness; RIB, rib eye area; URIB_B, ultrasound rib eye area of bulls; USWT, weight at ultrasound time; URIB_CS, ultrasound rib eye area of cows and steers; UFAT_B, ultrasound fat of bulls; UFAT_CS, ultrasound fat of cows and steers.

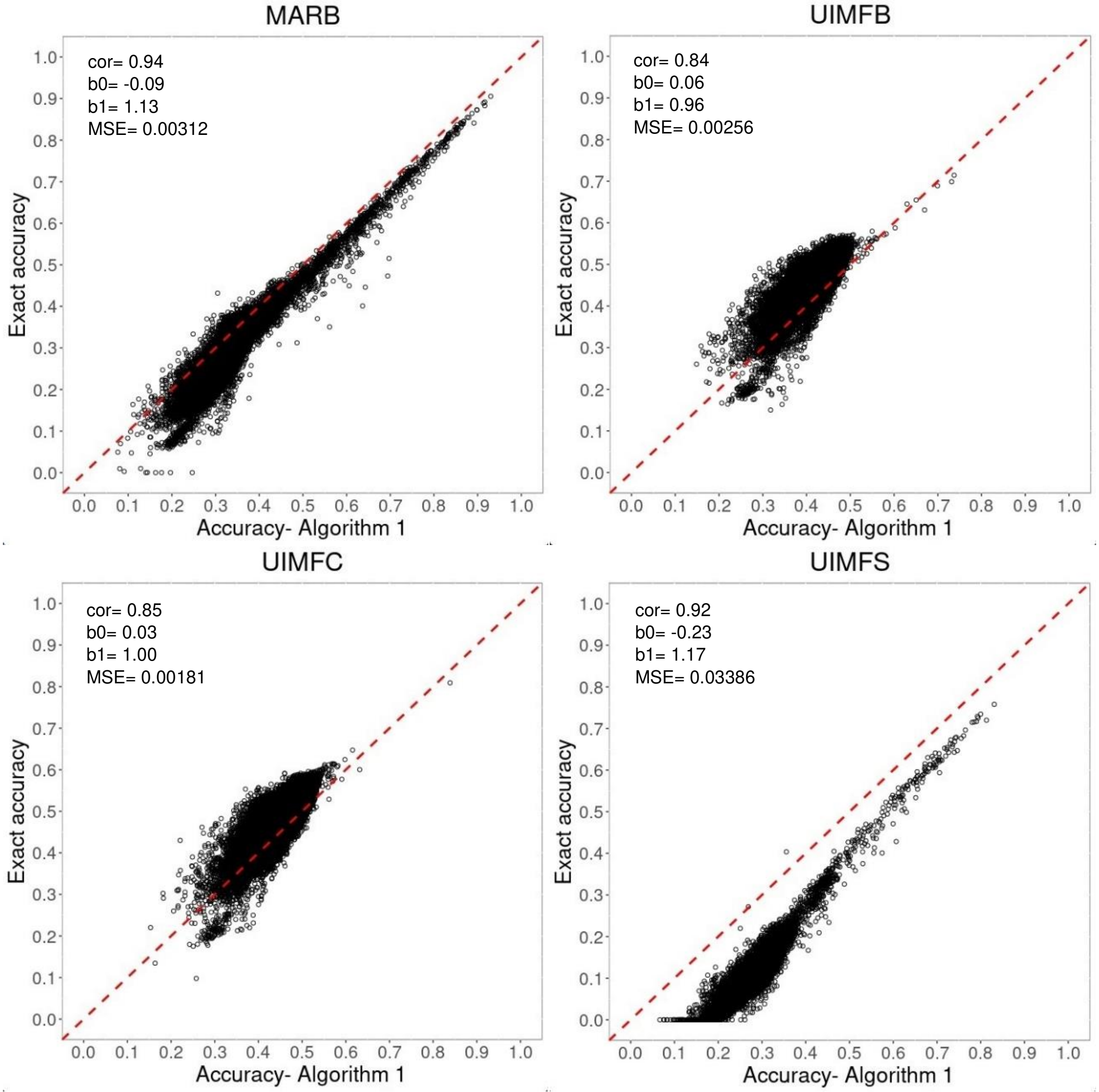


Figure 5. Scatter plots comparing exact against approximated accuracy obtained through Algorithm 1, for marbling traits. MARB, marbling; UIMFB, ultrasound intramuscular fat of bulls; UIMFC, ultrasound intramuscular fat of cows; UIMFS, ultrasound intramuscular fat of steers.

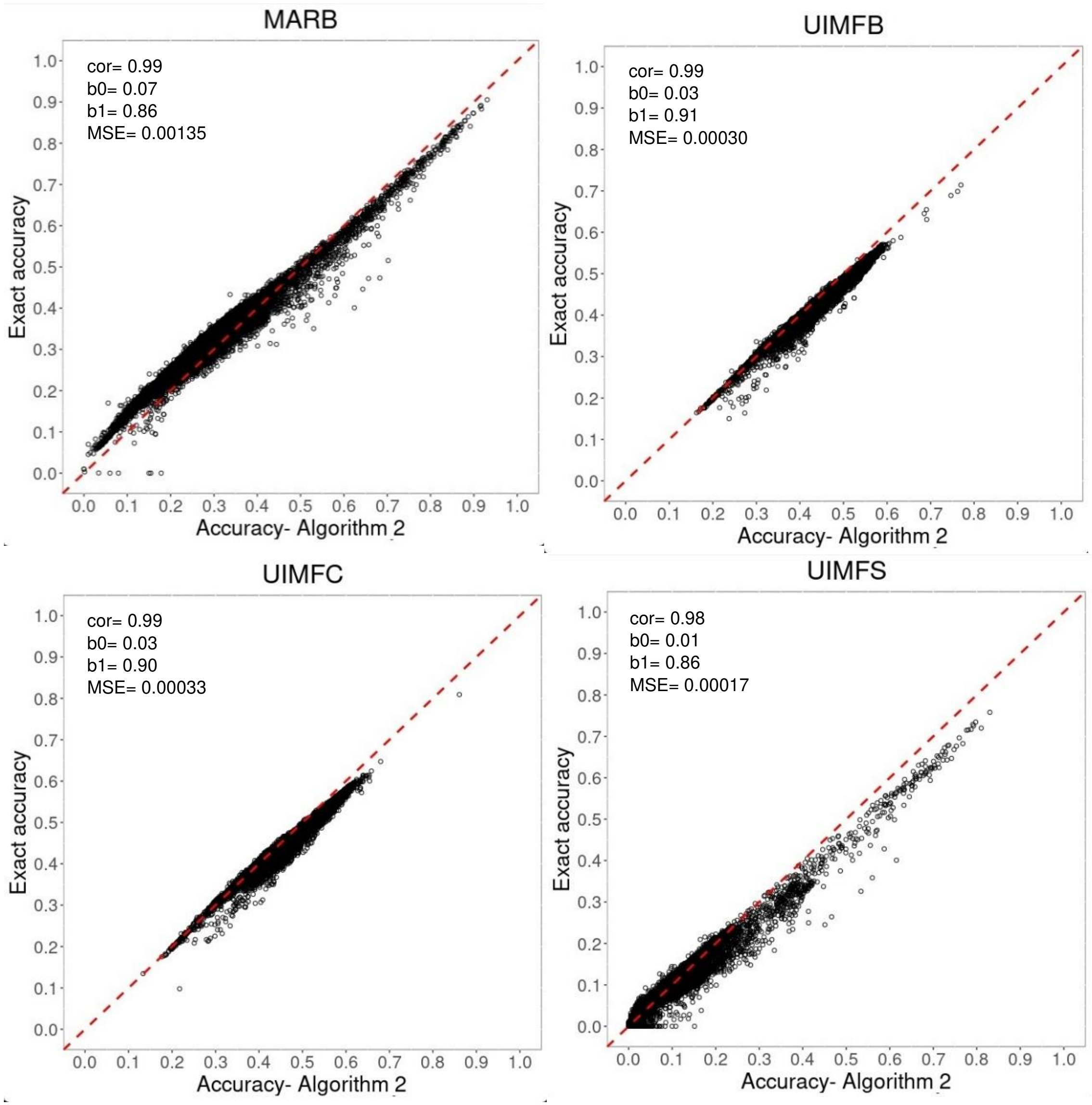


Figure 6. Scatter plots comparing different scenarios of exact and approximated accuracy obtained through Algorithm 2, for marbling traits. MARB, marbling; UIMFB, ultrasound intramuscular fat of bulls; UIMFC, ultrasound intramuscular fat of cows; UIMFS, ultrasound intramuscular fat of steers.

3.4.2 Comparing approximated accuracies from Algorithms 1 and 2

The comparison statistic between approximated accuracies obtained through Algorithms 1 and 2 for all genotyped animals (group 1), young genotyped animals (group 2), and proven genotyped sires (group 3) for growth traits (DATA_GT), carcass traits (DATA_CT), and marbling traits (DATA_MT) are in Table 2. For growth traits, the correlation between the approximated accuracies for genotyped animals ranged from 0.56 to 0.59. Furthermore, the slope and intercept of the regression of the accuracies for genotyped animals from Algorithm 1 on those from Algorithm 2 ranged from 0.54 to 0.58 and 0.04 to 0.18, respectively. These results indicate a considerable difference between the methods.

The differences between the algorithms are particularly pronounced when comparing within group 2, which exhibits the lowest correlations and slopes and the highest intercept values. This distinction can be attributed to the composition of group 2, which consists of young genotyped animals without phenotypes. Both algorithms compute accuracies as a function of records, pedigree, and genomic contribution (Misztal et al., 2013; Berman et al., 2022). Therefore, in group 2, where animals lack phenotypic information, the genomic contribution calculated differently in the algorithms significantly influences the mean of approximated accuracies. This difference between algorithms, especially for group 2, combined with stronger correlations between the exact accuracies and those from Algorithm 2, emphasizes that Algorithm 2 is more feasible than Algorithm 1. After all, group 2 represents a relevant category of animals in a selection program since it is composed of young selection candidates with limited phenotypic information and no offspring data available. In this context, having more precise approximated accuracies can help selecting the right animals among the selection candidates.

Table 2. Intercept (b0), slope (b1), and correlation (cor) between the approximated accuracies obtained through Algorithms 1 and 2 for each trait within different scenarios of genotyped animals.

		DATA_GT[2]			DATA_CT[2]								DATA_MT[2]			
Group[1]		BW	WW	PWG	RIB	CWT	USWT	URIB_B	URIB_CS	FAT	UFAT_B	UFAT_CS	MARB	UIMFB	UIMFC	UIMFS
1	b0	0.18	0.11	0.04	0.04	0.08	0.13	-0.02	0.02	0.01	0.00	0.00	0.04	-0.03	-0.03	0.08
	b1	0.54	0.55	0.58	0.55	0.52	0.59	0.76	0.70	0.55	0.72	0.72	0.54	0.74	0.75	0.49
	cor	0.56	0.55	0.59	0.74	0.71	0.63	0.72	0.71	0.74	0.69	0.71	0.74	0.71	0.72	0.69
2	b0	0.30	0.21	0.12	0.09	0.13	0.21	0.09	0.11	0.07	0.13	0.12	0.09	0.10	0.09	0.13
	b1	0.30	0.33	0.41	0.44	0.43	0.44	0.57	0.54	0.42	0.48	0.52	0.44	0.53	0.54	0.38
	cor	0.38	0.48	0.55	0.65	0.61	0.53	0.64	0.63	0.64	0.61	0.63	0.65	0.64	0.64	0.59
3	b0	-0.08	-0.14	-0.23	0.23	0.24	-0.17	-0.25	-0.23	0.18	-0.23	-0.23	0.22	-0.24	-0.25	-
	b1	1.07	1.14	1.23	0.68	0.69	1.17	1.25	1.23	0.75	1.23	1.23	0.71	1.24	1.25	-
	cor	0.97	0.97	0.96	0.69	0.71	0.96	0.95	0.94	0.77	0.94	0.93	0.72	0.95	0.94	-

BW, birth weight; WW, weaning weight; PWG, post-weaning gain; RIB, rib eye area; CWT, carcass weight; USWT, weight at ultrasound time; URIB_B, ultrasound rib eye area of bulls; URIB_CS, ultrasound rib eye area of cows and steers; FAT, fat thickness; UFAT_B, ultrasound fat of bulls; UFAT_CS, ultrasound fat of cows and steers; MARB, marbling; UIMFB, ultrasound intramuscular fat of bulls; UIMFC, ultrasound intramuscular fat of cows; UIMFS, ultrasound intramuscular fat of steers.

[1] 1, genotyped animals; 2, genotyped animals born in 2022 without phenotypes; 3, genotyped sires with 100 or more offspring with phenotypes.

[2] DATA_GT, Data of growth traits used for multi-trait analysis; DATA_CT, Data of carcass traits used for multi-trait analysis, DATA_MT, Data of marbling traits used for multi-trait analysis.

In contrast, in group 3, the correlation between the algorithms was close to one. This result was expected, as this group comprises sires with a minimum of 100 offspring with records. Consequently, in this case, offspring records have a more substantial impact on the accuracy than the genomic information. In both algorithms, the contributions due to phenotypes and pedigree are approximated using the method described by Misztal and Wiggans (1988); therefore, correlations are not one for this group because of the different approximation of genomic contributions between Algorithm 1 and 2.

For carcass traits (DATA_CT), the correlation between the approximated accuracies ranged from 0.63 to 0.74, 0.53 to 0.64, and 0.71 to 0.96 in groups 1, 2, and 3, respectively. The correlation patterns exhibited similar behavior across the different trait groups. The young animals (group 2) displayed lower correlation and slope values, whereas the proven sires (group 3) exhibited higher correlation and slopes. These results also suggest significant differences between the approximated accuracies obtained from the two algorithms for carcass traits. For marbling traits (DATA_MT), the correlation between the approximated accuracies ranged from

0.69 to 0.74, 0.59 to 0.65, and 0.72 to 0.95 in groups 1, 2, and 3, respectively. Due to the small number of animals (six sires) in group 3, the regression and correlation between algorithms could not be calculated for ultrasound intramuscular fat of steers. The results for marbling traits align with those observed for carcass traits in DATA_MT, indicating significant differences between the algorithms in the evaluated multi-trait carcass models. These results emphasize the importance of careful consideration when selecting the appropriate algorithm for accuracy approximation in the context of carcass, marbling and growth traits.

3.4.3 Impact of adding more genotyped animals on the average accuracy

To assess the influence of adding genotyped individuals on the average accuracy, we evaluated the changes in exact and approximated accuracies of PWG and URIB_B by comparing scenarios involving 40,000 genotyped, of which 25,000 lacked phenotypic data. This analysis also helped investigate whether both algorithms considered extra genomic information similarly. The extra information comprised 10,000 genotyped animals with and without phenotypes. Table 3 provides descriptive statistics of the exact and approximated accuracies calculated in different inclusion scenarios.

Across all scenarios, Algorithm 1 consistently yielded lower accuracies, while Algorithm 2 presented higher values. As expected, including genotyped animals with phenotypes had the most significant impact on increasing the exact and approximated accuracies. The increase in mean accuracy was more pronounced for Algorithm 2, which aligns closely with the exact accuracies. On the other hand, approximated accuracies from Algorithm 1 exhibited a comparatively lower increase due to including genotyped animals with phenotypes. This result can also be attributed to the fact that in Algorithm 1, the genomic contribution is not complete since it relies solely on the diagonals of **G** and does not consider information from relatives. In this context, it can be noticed that accuracy from Algorithm 2 presented closer behavior to the exact accuracies across different scenarios of including new genotyped animals in the analysis, meaning it can consider extra information more appropriately.

Table 3. Number of records (N), mean, maximum and minimum values, and standard deviations (SD) of exact accuracies (Acc) and the approximated GEBV accuracies of the 40,000 genotyped animals, and 25000 genotyped animals without phenotype, obtained through the Algorithm 1 (Algo_1) and 2 (Algo_2) for post-weaning gain (PWG) and ultrasound rib eye area of bulls (URIB_B).

			Mean			SD			Minimum			Maximum		
	Scenario[1]	N	Acc	Algo_1	Algo_2	Acc	Algo_1	Algo_2	Acc	Algo_1	Algo_2	Acc	Algo_1	Algo_2
PWG	A	40000	0.39	0.32	0.41	0.05	0.04	0.05	0.04	0.07	0.13	0.53	0.44	0.56
	B	40000	0.35	0.31	0.36	0.05	0.04	0.06	0.01	0.04	0.09	0.49	0.42	0.52
	C	40000	0.34	0.31	0.36	0.05	0.04	0.06	0.01	0.05	0.09	0.49	0.43	0.52
URIB_B	A	40000	0.44	0.39	0.46	0.05	0.05	0.06	0.12	0.09	0.12	0.71	0.73	0.76
	B	40000	0.40	0.37	0.41	0.06	0.05	0.06	0.09	0.07	0.07	0.70	0.72	0.74
	C	40000	0.40	0.37	0.41	0.06	0.05	0.06	0.08	0.07	0.07	0.69	0.72	0.74
PWG	A	25000	0.37	0.30	0.39	0.05	0.04	0.05	0.04	0.07	0.13	0.51	0.40	0.54
	B	25000	0.33	0.28	0.34	0.05	0.04	0.06	0.01	0.04	0.09	0.47	0.38	0.50
	C	25000	0.33	0.29	0.34	0.05	0.04	0.06	0.01	0.05	0.09	0.47	0.39	0.50
URIB_B	A	25000	0.42	0.37	0.43	0.05	0.03	0.05	0.12	0.09	0.12	0.71	0.73	0.76
	B	25000	0.38	0.34	0.38	0.05	0.03	0.06	0.09	0.07	0.07	0.70	0.72	0.74
	C	25000	0.38	0.35	0.38	0.05	0.03	0.06	0.08	0.07	0.07	0.69	0.72	0.74

PWG, post-weaning gain; URIB_B, ultrasound rib eye area of bulls

[1] A, inclusion of 10000 genotyped animals with phenotype; B, inclusion of 10000 genotyped animals without phenotype; C, no inclusion of new animals.

Figures 7 and 8 show scatter plots and comparative statistics of exact and approximated accuracies obtained through Algorithm 2 for PWG and URIB_B, respectively. These plots present the comparison between different scenarios of genotyped animal inclusion, considering the 40,000 genotyped animals.

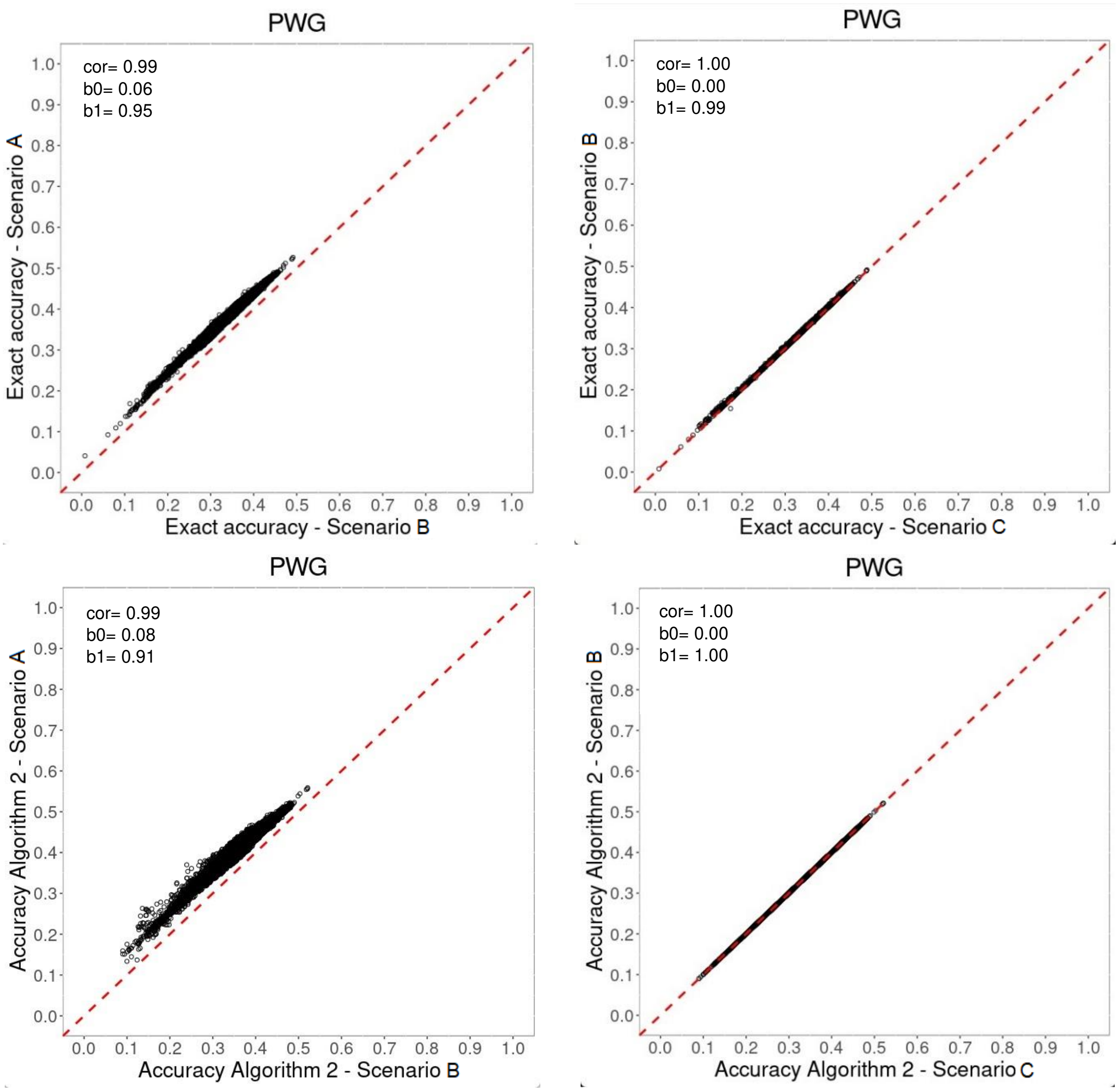


Figure 7. Scatter plots comparing different scenarios of exact and approximated accuracy obtained through Algorithm 2, for post-weaning gain (PWG). Scenario A, inclusion of 10000 genotyped animals with phenotype; Scenario B, inclusion of 10000 genotyped animals without phenotype; Scenario C, no inclusion of new animals.

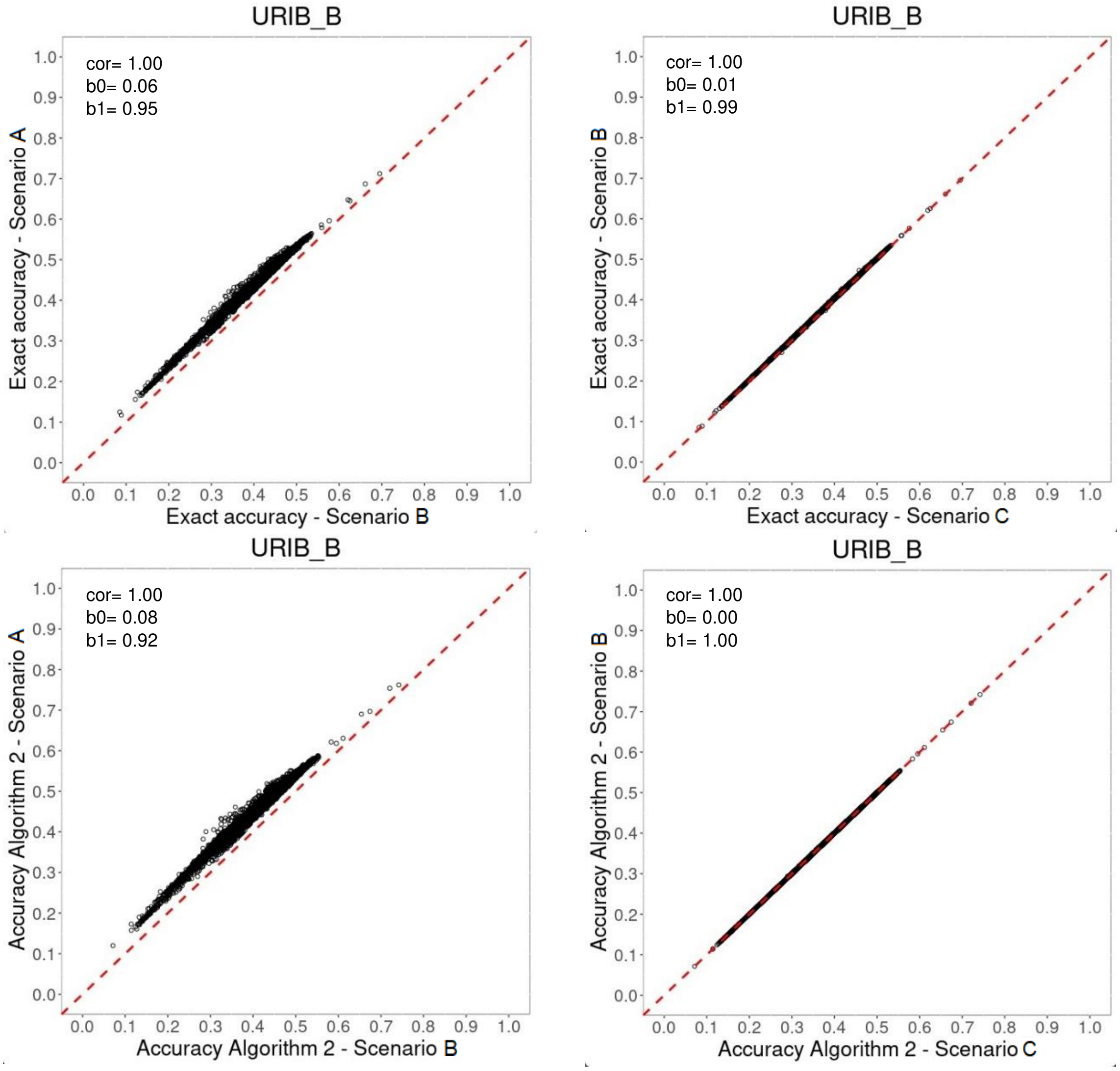


Figure 8. Scatter plots comparing different scenarios of exact and approximated accuracy obtained through Algorithm 2, for ultrasound rib eye area of bulls (URIB_B). Scenario A, inclusion of 10000 genotyped animals with phenotype; Scenario B, inclusion of 10000 genotyped animals without phenotype; Scenario C, no inclusion of new animals.

The regression slope and correlation between scenarios A and B for exact and approximated accuracies ranged from 0.91 to 0.95 and 0.99 to 1.00, respectively. These findings emphasize that including genotyped animals with phenotypes increases exact and approximated accuracies for both traits. Moreover, the increase in approximated accuracies closely mirrors the increase in exact accuracies. In contrast, including genotyped animals without phenotypes does not significantly increase exact or approximated accuracies for the analyzed traits. This result was expected because, although in ssGBLUP the new genomic relationships may modify old pedigree relationships, the effect of genotypes of young animals on the GEBV of other genotyped animals is likely to be minor, especially if the number of evaluated animals is very large (Misztal et al., 2014). However, changes in accuracies when including genotyped animals without phenotypes should be confirmed with larger numbers of genotyped animals.

3.5 Conclusion

Two algorithms for approximating accuracies from ssGBLUP were compared. Algorithm 1, which approximates accuracies based on the diagonal of **G**, exhibited significant differences in approximated accuracies compared to Algorithm 2, which utilizes block sparse inversion of $\mathbf{G}^{-1}$. Furthermore, accuracies from Algorithm 2 were closer to the exact accuracies and demonstrated similar changes in behavior when adding new genotyped animals in the analysis. Therefore, Algorithm 2 is more suitable for approximating the accuracies of GEBV and should be used for routine ssGBLUP evaluations.

3.6 Literature Cited

Aguilar, I., E. N. Fernandez, A. Blasco, O. Ravagnolo, and A. Legarra. 2020. Effects of ignoring inbreeding in model-based accuracy for BLUP and SSGBLUP. *J. Anim. Breed.* Genet. 137:356–364. https: //doi.org/10.1111/jbg.12470.

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. Dairy Sci.* 93:743–752. doi: 10.3168/ jds .2009 -2730

Bermann, M., D. A. L. Lourenco, and I. Misztal. 2022. Efficient approximation of reliabilities for single-step genomic best linear unbiased predictor models with the algorithm for proven and young. *J. Anim. Sci.* 100:1–7. doi:10.1093/jas/skab353

Edel, C., E. C. G. Pimentel, M. Erbe, R. Emmerling, and K.-U. Götz. 2019. Short communication: Calculating analytical reliabilities for single-step predictions. *J. Dairy Sci.* 102:3259–3265. doi: 10 .3168/ jds .2018 -15707.

Harris, B., and D. Johnson. 1998. Approximate reliability of genetic evaluations under an animal model. J. Dairy Sci. 81:2723–2728. doi: 10.3168/ jds .S0022 -0302(98)75829 -1.

Henderson CR. Applications of linear models in animal breeding. Guelph: University of Guelph; 1984

Henderson, H. V., and S. R. Searle. 1981. On deriving the inverse of a sum of matrices. SIAM Rev. 23:53–60. http://www.jstor.org/stable/2029838

Liu, Z., P. M. VanRaden, M. H. Lidauer, M. P. Calus, H. Benhajali, H. Jorjani, and V. Ducrocq. 2017. Approximating genomic reliabilities for national genomic evaluation. Interbull Bull. 51:75–85.

Liu, Z., M. E. Goddard, F. Reinhardt, and R. Reents. 2014. A singlestep genomic model with direct estimation of marker effects. *J. Dairy Sci.* 97:5833–5850. doi:10 .3168/ jds .2014 -7924

Liu, Z., F. Reinhardt, A. Bünger, and R. Reents. 2004. Derivation and calculation of approximate reliabilities and daughter yielddeviations of a random regression test-day model for genetic evaluation of dairy cattle. *J. Dairy Sci.* 87:1896–1907. doi: 10 .3168/ jds .S0022 - 0302(04)73348 -2.

Lourenco, D. A. L., S. Tsuruta, B. O. Fragomeni, Y. Masuda, I. Aguilar, A. Legarra, S. Miller, D. Moser, and I. Misztal. 2018. Single-step genomic BLUP for national beef cattle evaluation in US: from initial developments to final implementation. Proc. World. Cong. Appl. Livest. Prod. 11:495

Masuda, Y., I. Misztal, S. Tsuruta, A. Legarra, I. Aguilar, D. A. L. Lourenco, B. O. Fragomeni, and T. J. Lawlor. 2016. Implementation of genomic recursions in single-step genomic best linear unbiased predictor for US Holsteins with a large number of genotyped animals. J. Dairy Sci. 99:1968–1974. doi:10.3168/jds.2015-10540

Misztal I, 2016 Inexpensive computation of the inverse of the genomic relationship matrix in populations with small effective population size. Genetics 202: 401–409. https://doi.org/10.1534/genetics.115.182089

Misztal, I., S. Tsuruta, D.A.L. Lourenco, I. Aguilar, A. Legarra, and Z. Vitezica. 2014a. Manual for BLUPF90 family of programs. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all1.pdf. (Accessed 5 October 2022.)

Misztal, I., A. Legarra, and I. Aguilar. 2014b. Using recursion to compute the inverse of the genomic relationship matrix. *J. Dairy Sci.* 97:3943–3952. doi:10.3168/jds.2013-7752

Misztal, I., S. Tsuruta, I. Aguilar, A. Legarra, P. M. VanRaden, and T. J. Lawlor. 2013. Methods to approximate reliabilities in single-step genomic evaluation. *J. Dairy Sci*. 96:647–654. doi: 10 .3168/ jds .2012 -5656.

Misztal, I., and G. R. Wiggans. 1988. Approximation of prediction error variance in large-scale animal models. *J. Dairy Sci*. 71:27–32. doi: 10 .1016/ S0022 -0302(88)79976 -2.

Pocrnic I, Lourenco DAL, Masuda Y, Misztal I. Dimensionality of genomic information and performance of the Algorithm for Proven and Young for different livestock species. Genet Sel Evol. 2016;48:82.

Quaas, R. 1976. Computing the diagonal elements and inverse of a large numerator relationship matrix. Biometrics. 32(4):949–953. doi:10.2307/2529279

Strabel, T., I. Misztal, and J. K. Bertrand. 2001. Approximation of reliabilities for multiple-trait model with maternal effects. *J. Anim. Sci*. 79:833–839. doi:10.2527/2001.794833x

Tier, B., and K. Meyer. 2004. Approximating prediction error covariancesamong additive genetic effects within animals in multiple-trait and random regression models. *J. Anim. Breed. Genet*. 121:77–89. doi: 10 .1111/ j .1439 -0388 .2003 .00444 .x.

Tsuruta, S., D. A. L. Lourenco, Y. Masuda, T. J. Lawlor, and I. Misztal. 2021. Reducing computational cost of large-scale genomic evaluation by using indirect genomic prediction. *JDS Commun*. 2:356–360. doi: 10 .3168/ jdsc .2021 -0097

Tsuruta, S., D.A.L. Lourenco, Y. Masuda, D.W. Moser, and I. Misztal. 2016. Practical approximation of accuracy in genomic breeding values for a large number of genotyped animals. In: Joint Annual Meeting of American Society of Animal Science, Salt Lake City. https://doi.org/10.2527/jam2016-0337

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci*. 91(11), 4414–4423. doi: 10.3168/jds.2007-0980

VanRaden, P. M., and A. E. Freeman. 1985. Rapid method to obtain bounds on accuracies and prediction error variances in mixed models. *J. Dairy Sci*. 68:2123–2133. doi: 10 .3168/ jds .S0022-0302(85)81078 -X.

Wilmink J.B., Dommerholt J. 1985. Approximate Reliability of Best linear Unbiased Prediction in models with and without relationships, *J. Dairy Sci*. 68 (1985) 946–952. doi: 10.3168/jds.S0022-0302(85)80913-9

CHAPTER 4

4.1 General conclusions

The present study has contributed to the expansion of knowledge about feed efficiency traits, offering important information that can be considered in genetic evaluations for breeding programs. Additionally, it has provided valuable insights to the industry regarding algorithms for approximating accuracy, recommending the adoption of a new algorithm for routine genetic evaluations due to its demonstrated higher precision.

As discussed in Chapter 2, the random regression model is a viable approach for genomic evaluations of feed efficiency in Nellore cattle. The models using linear B-splines with seven knots and six classes of residual variances are recommended for evaluations of dry matter intake, residual feed intake, body weight gain, and residual weight gain according to better goodness of fit. Genetic parameter estimates for feed efficiency traits exhibited variation throughout the performance test. Heritability estimates indicated sufficient additive genetic variance, suggesting the potential for achieving a moderate response to selection. Based on genetic correlations and ranking comparisons, it is possible to adopt a new selection strategy by reducing the performance test period from 84 to 56 days since it would result in selecting almost the same animals for RFI and DMI. However, a more extended test period of 77 days appears more appropriate for body weight gain and residual weight gain.

In Chapter 3, we compared two algorithms implemented in the BLUPF90 suite of programs to approximate accuracies for the single-step Genomic Best Linear Unbiased Predictor (GEBV). The results hold significant value for breeding programs, indicating that Algorithm 2 (implemented in the ACCF90GS2 program) closely approximates the exact accuracies and exhibits similar behavior when new genotyped animals are added to the analysis. Therefore, Algorithm 2 has proven to be a more suitable choice for approximating the accuracies of GEBV, making it highly recommended for routine genetic evaluations.