

**DELVAN ALVES DA SILVA**

**GENETIC EVALUATION AND GENOTYPE BY ENVIRONMENT  
INTERACTION USING TEST-DAY MODELS FOR PORTUGUESE AND  
BRAZILIAN HOLSTEIN CATTLE**

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

Adviser: Paulo Sávio Lopes

Co-advisers: Júlio Gil Vale Carvalheira  
Fabyano Fonseca e Silva  
Cláudio Napolis Costa

**VIÇOSA - MINAS GERAIS  
2019**

**Ficha catalográfica preparada pela Biblioteca Central da Universidade  
Federal de Viçosa - Câmpus Viçosa**

T

S586g  
2019

Silva, Delvan Alves da, 1990-

Genetic evaluation and genotype by environment  
interaction using test-day models for Portuguese and Brazilian  
Holstein cattle / Delvan Alves da Silva. – Viçosa, MG, 2019.  
91 f. : il. ; 29 cm.

Texto em inglês.

Orientador: Paulo Sávio Lopes.

Tese (doutorado) - Universidade Federal de Viçosa.

Inclui bibliografia.

1. Bovinos de leite. 2. Autocorrelação. 3. Interação  
genótipo-ambiente. 4. Genética. 5. Lactação. 6. Análise de  
regressão. I. Universidade Federal de Viçosa. Departamento de  
Zootecnia. Programa de Pós-Graduação em Zootecnia. II. Título.

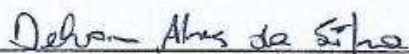
CDD 22. ed. 636.2142

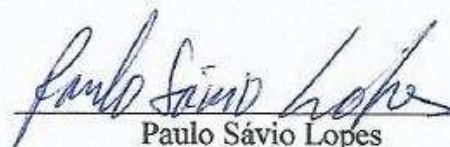
**DELVAN ALVES DA SILVA**

**GENETIC EVALUATION AND GENOTYPE BY ENVIRONMENT  
INTERACTION USING TEST-DAY MODELS FOR PORTUGUESE AND  
BRAZILIAN HOLSTEIN CATTLE**

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

APPROVED: July 16<sup>th</sup>, 2019.

  
\_\_\_\_\_  
Delvan Alves da Silva  
Author

  
\_\_\_\_\_  
Paulo Sávio Lopes  
Adviser

*I dedicate this work to my parents  
David and Geralda,  
my sister Dalva,  
my brothers Denival and João Paulo  
and my wife Alessandra.*

## ACKNOWLEDGMENTS

I am grateful to Universidade Federal de Viçosa, especially the Animal Science Graduate Program and the Department of Animal Science, for the opportunity of carrying out the course;

The Universidade do Porto: Research Center in Biodiversity and Genetic Resources (CIBIO-InBio), and the Institute of Biomedical Sciences Abel Salazar (ICBAS) for the opportunity, theoretical and technical support;

The Embrapa Gado de Leite for the opportunity, theoretical and technical support;

The National Council of Technological and Scientific Development (CNPq, Brazil), Coordination for the Improvement of Higher Education Personnel (CAPES, Brazil), and Portuguese National Funding Agency for Science, Research and Technology (FCT, Portugal) for financial support;

The Portuguese Dairy Cattle Breeders Association and Brazilian Holstein Cattle Breeders Association for providing the data and to have become this work possible;

I would like to express my sincere gratitude to my advisor Prof. Paulo Sávio Lopes for the continuous support of my Ph.D. study and related research, for his supervision, patience, motivation, friendship, advice, and opportunities given to me. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D. study.

My sincere thanks also to my co-advisor Prof. Júlio Carvalheira, who provided me an opportunity to join their team and who has access to the server and research facilities. Without their precious support, it would not be possible to conduct this research. Thanks for his supervision, patience, motivation, friendship, advice, and opportunities given to me.

My sincere thanks also to my co-advisor Dr. Cláudio Napolis Costa, for all teachings, suggestions, support, friendship, and opportunities given to me. Without his precious support, it would not be possible to conduct this research.

I would like to thank my co-advisor Prof. Fabyano Fonseca e Silva for the advices and support during my Ph.D. Thank you for the supervision, patience, motivation, opportunities and friendship. Without his precious support, it would not be possible to conduct this research. I have learned a lot from you.

I would like to thank my thesis committee: Prof. Sebastião Martins Filho, and Dr. Rodrigo Reis Mota, for their comments, encouragement, and contributions to this thesis;

I also thank Prof. Renata Veroneze for her contributions, support, discussions, and friendship;

I would also like to thank all friends, professors and staff of the CIBIO/UP, ICBAS/UP and DZO/UFV;

A special thank you to all friends (new and old) from “salinha”, GDMA, and LABTEC for the good conversations, discussions, barbecues and for the games, especially Teo, Darlene, Hugo, Hinayah, Ivan, Ândrea, Sirlene, Lais, Nadson, Giovanni, Vinicius, Edinho, Erica, Eula, Amanda, Matheus, Talita, Natalia, Pedro, Tulio, Leticia, Layla, Aricia, Ingrid, Daniele, Margareth, Susana, Karine, Walmir, Haniel, and Pamela. I also thank you my old friends from UFMT/Sinop Denise, Nayanne, and Patricia for the old and good times.

Last but not least, I would like to thank my family: my parents (David and Geralda), my brothers (Denival and João Paulo), my sister (Dalva), my grandfathers (João Rosa and José) and grandmothers (Dindinha and Maria), D. Joana, Rosimary, Mario, Júnior, Sabrina, Victor and Janete for all support and encouragement. Thank you very much for everything. Finally, I will never forget all the support and encouragement of my beloved wife and partner. Thank you so much for everything. I love you very much!

## **BIOGRAPHY**

DELVAN ALVES DA SILVA, son of David dos Santos Silva and Geralda Alves da Silva, was born on October 29<sup>th</sup>, 1990 in Juara - MT, Brazil.

In March 2009, he started his undergraduate studies in Animal Science at the Universidade Federal de Mato Grosso (UFMT), Sinop - MT, Brazil. Since the beginning, he was involved in several research and extension activities in the animal breeding and genetics areas under the supervision of Dr. Cláudio Vieira de Araújo. In July 2013, Delvan concluded his undergrad, with an honorable award from the UFMT for his excellent academic performance.

In August 2013, he started his graduate program at Animal Science Graduate Program by Universidade Federal de Viçosa (UFV, Viçosa – MG, Brazil), under supervision of Dr. Paulo Sávio Lopes to obtain his degree of Magister Scientiae in Animal Science. He presented his dissertation in July 2015.

In August 2015, started his Ph.D at Animal Science Graduate Program by UFV, under supervision of Dr. Paulo Sávio Lopes to obtain his degree of Doctor Scientiae in Animal Science. He had the opportunity to work in a research project of international cooperation (CAPES/FCT, n° 99999.008462/2014-03) between Brazil (UFV and Embrapa Gado de Leite) and Portugal (Universidade do Porto). As part of the project, he had an internship in the Embrapa Gado de Leite (Juiz de Fora - MG, Brazil), under supervision of Dr. Cláudio Nápolis Costa and in Universidade do Porto (Porto, Portugal), under supervision of Dr. Júlio Carvalheira. The results of the project development are summarized in this thesis. His thesis will be defended in July 2019.

## ABSTRACT

SILVA, Delvan Alves, D.Sc., Universidade Federal de Viçosa, July, 2019. **Genetic evaluation and genotype by environment interaction using test-day models for Portuguese and Brazilian Holstein cattle.** Adviser: Paulo Sávio Lopes. Co-advisers: Júlio Gil Vale Carvalheira, Fabyano Fonseca e Silva and Cláudio Napolis Costa.

The validation of test-day (TD) models (autoregressive, AR and random regression, RR) in Brazilian Holstein cattle via genetic evaluation is of great importance since there is an interest to change traditional by TD models for multiple lactations. In addition, the investigation of the presence of genotype by environment interaction (G x E) between Brazil and Portugal populations may be important for breeding strategies as long as breeders would be allowed to invest or not in foreign genotypes. To do so, the traits used in this study were milk, fat and protein yields, and somatic cell score (SCS), using the first three lactations. The data was provided by Portuguese and Brazilian Holstein Cattle Breeders Associations and recorded between 1994 and 2016. First, TD milk, fat and protein yields, and SCS were used to study the effect of fixed (HTDF) vs. random (HTDR) contemporary groups (herd-test-date) with or without unknown parent groups (UPG) by using AR. The recovery of information performed on the HTDR models increased the lower reliabilities from 0.50 to 0.75, 0.54 to 0.66, 0.64 to 0.71 and 0.25 to 0.67 for milk, fat and protein yields, and for SCS, respectively (for bulls with 10 or more daughters). The differences in annual genetic gains between models (HTDR vs HTDF) were for sires (cows) of 30.66 (38.59) kg, 1.18 (1.35) kg, 1.26 (1.22) kg and -0.001 (-0.03) scores for milk, fat and protein yields and SCS, respectively. The contemporary groups as a random effect in the AR model seemed to be more relevant than just the UPG effect itself. However, the combination of both may provide higher annual genetic gains. In general, the HTDR model with UPG was the procedure that best fit and should be used for genetic evaluations and genetic trend analysis of longitudinal traits in Brazilian Holstein cattle. Subsequently, TD milk yield and SCS records were fitted to AR and RR models with the objective to compare their efficiency in the Brazilian genetic evaluations. Milk yield (SCS) heritabilities were similar between both models and ranged from 0.17 to 0.23 (0.11 to 0.17). The rank correlation between estimated breeding values (EBV) obtained from AR and RR models were 0.96 and 0.94 for milk yield and 0.97 and 0.95 for SCS, respectively, for

bulls (with 10 or more daughters) and cows. Annual genetic gains for bulls (cows) obtained using AR model were 46.11 (49.50) kg for milk yield and -0.019 (-0.025) score for SCS. Using RR models this gain was 47.70 (55.56) kg for milk yield and -0.022 (-0.028) score for SCS. In general, the AR models were more efficient and, given the lower number of parameters to estimate and its suitability to fit data from small herds, these models are more parsimonious and should be used in genetic evaluations of Holstein cattle in Brazil. Finally, a two-step genetic evaluation was used to verify G x E. In step 1, we performed a within-country (Portugal and Brazil separately) genetic evaluation by using AR models. Similar heritability estimates for SCS were observed for both countries, whereas for milk yield the heritability were 0.31 for Portugal and 0.23 for Brazil. The rank correlation between EBV of common bulls were 0.75 for milk yield and 0.62 for SCS. In step 2, the precorrected phenotypes from Portugal and Brazil were considered two distinct traits in bi-trait reaction norm models. For milk yield, the genetic correlation between environmental gradient (HTD levels) within countries were higher than 0.92 for Portugal and 0.98 for Brazil. For SCS, the genetic correlation between HTD levels ranged from 0.64 to 0.99 for Portugal and from 0.79 to 0.99 for Brazil. The average of genetic correlation estimates between HTD levels from Portugal and HTD levels from Brazil was 0.73 for milk yield and 0.57 for SCS. In conclusion, our results indicated the presence of G x E in Holstein cattle in different production systems and climatic conditions. The low genetic correlation in the Portuguese population indicated the presence of G x E between extreme environmental gradient for SCS.

Keywords: Autocorrelation. Dairy cattle. Environment gradient. Genetic group. Multiple lactations. Random regression.

## RESUMO

SILVA, Delvan Alves, D.Sc., Universidade Federal de Viçosa, julho de 2019. **Avaliação genética e interação genótipo x ambiente usando modelos test-day para bovinos Holstein em Portugal e Brasil.** Orientador: Paulo Sávio Lopes. Coorientadores: Júlio Gil Vale Carvalheira, Fabyano Fonseca e Silva e Cláudio Napolis Costa.

Avaliações genéticas utilizando modelos test-day (TD) (como autorregressivo, AR e regressão aleatória, RR) em bovinos da raça Holandesa são necessárias para avaliar a aplicação desses modelos nesta população, uma vez que há interesse na mudança de modelos tradicionais para modelos TD usando múltiplas lactações. Da mesma forma, a avaliação da interação genótipo x ambiente (G x E) entre Brasil e Portugal pode ser importante para estratégias de melhoramento genético e para os produtores que investem em genótipos estrangeiros. Foram utilizados neste estudo as três primeiras lactações para produção leite, gordura e proteína, e escore de células somáticas (SCS). Os dados foram fornecidos pelas Associações Portuguesa e Brasileira de Criadores de Bovinos da Raça Holandesa e coletados entre 1994 e 2016. Primeiramente, os registros de produção de leite, gordura e proteína, e SCS foram utilizados para estudar o efeito de grupos de contemporâneos como fixo (HTDF) ou aleatório (HTDR) e com ou sem grupos de pais fantasmas (UPG) usando o modelo AR. A recuperação das informações usando modelos HTDR aumentou as confiabilidades mínimas de 0,50 para 0,75, 0,54 para 0,66, 0,64 para 0,71 e 0,25 para 0,67 para produção de leite, gordura e proteína, e SCS, respectivamente (para touros com 10 ou mais filhas). As diferenças nos ganhos genéticos anuais entre os modelos (HTDR vs. HTDF) para touros (vacas) foram de 30,66 (38,59) kg, 1,18 (1,35) kg, 1,26 (1,22) kg e -0,001 (-0,03) escores para produção de leite, gordura e proteína e SCS, respectivamente. Os grupos de contemporâneos como efeito aleatório no modelo AR são mais relevantes do que apenas considerar o efeito UPG. No entanto, a combinação de ambos pode proporcionar maiores ganhos genéticos anuais. Em geral, o modelo HTDR com UPG foi o que melhor se ajustou a essas características e deve ser o modelo de escolha para avaliações genéticas e análise de tendências genéticas de características longitudinais em bovinos da raça Holandesa. Posteriormente, os registros de produção de leite e SCS foram ajustados usando os modelos AR e RR, com o objetivo de comparar a eficiência destes modelos na

avaliação genética nacional. As herdabilidades para produção de leite (SCS) foram similares entre os dois modelos e variaram de 0,17 a 0,23 (0,11 a 0,17). As correlações de rank entre os valores genéticos estimados (EBV) obtidos com os modelos AR e RR foram de 0,96 e 0,94 para produção de leite e 0,97 e 0,95 para SCS, respectivamente, para touros com 10 ou mais filhas e vacas. Ganhos genéticos anuais para touros (vacas) obtidos com os modelos AR foram 46,11 (49,50) kg para produção de leite e -0,019 (-0,025) escores para SCS. Ao usar os modelos RR, esses ganhos foram de 47,70 (55,56) kg para produção de leite e -0,022 (-0,028) escores para SCS. Em geral, os modelos AR foram mais eficientes e, dado o menor número de parâmetros para estimar e sua adequação aos dados de pequenos rebanhos, esses modelos são mais parcimoniosos e devem ser escolhidos para avaliações genéticas de bovinos Holstein no Brasil. Finalmente, a avaliação genética em dois passos foi usada para verificar a presença ou não de G x E. No passo 1, foi realizada avaliação genética dentro de país (Portugal e Brasil separadamente) usando modelos AR. Foram observadas estimativas de herdabilidade semelhantes para SCS em ambos os países, enquanto que para a produção de leite a herdabilidade foi de 0,31 para Portugal e de 0,23 para o Brasil. A correlação de rank entre os EBVs dos touros comuns aos dois países foi de 0,75 para produção de leite e de 0,62 para SCS. No passo 2, os fenótipos corrigidos de Portugal e do Brasil foram considerados como duas características distintas usando modelos de normas de reação bi-característicos. Para produção de leite, a correlação genética entre os gradientes ambientais (níveis de HTD) dentro dos países foi superior a 0,92 para Portugal e a 0,98 para o Brasil. Para SCS, a correlação genética entre os níveis de HTD variou de 0,64 a 0,99 para Portugal e de 0,79 a 0,99 para o Brasil. A média das estimativas de correlação genética entre os níveis de HTD de Portugal com os níveis de HTD do Brasil foi de 0,73 para a produção de leite e de 0,57 para a SCS. Concluiu-se que há interação genótipo x ambiente para bovinos da raça Holandesa considerando diferentes sistemas de produção e condições climáticas. A baixa correlação genética na população portuguesa indicou a presença de G x E entre gradiente ambiental extremo para SCS.

Palavras-chave: Autocorrelação. Bovinos de leite. Gradiente ambiental. Grupos genéticos. Múltiplas lactações. Regressão aleatória.

## TABLE OF CONTENTS

<b>GENERAL INTRODUCTION .....</b>	<b>13</b>
References .....	17
<b>CHAPTER 1 .....</b>	<b>19</b>
<b>Unknown parent and contemporary groups for genetic evaluation of Brazilian Holstein using autoregressive test-day models .....</b>	<b>19</b>
1.1. Abstract .....	20
1.2. Introduction .....	21
1.3. Materials and methods .....	23
1.4. Results and Discussion .....	27
1.5. Conclusion .....	35
1.6. Acknowledgments .....	35
1.7. References .....	36
1.8. Supplementary material .....	39
<b>CHAPTER 2 .....</b>	<b>40</b>
<b>Autoregressive and random regression test-day models for multiple lactations in genetic evaluation of Brazilian Holstein cattle .....</b>	<b>40</b>
2.1. Abstract .....	41
2.2. Introduction .....	41
2.3. Material and methods .....	43
2.4. Results .....	50
2.5. Discussion .....	55
2.6. Conclusion .....	58
2.7. Acknowledgements .....	59
2.8. References .....	59
2.9. Supplementary material .....	63

<b>CHAPTER 3 .....</b>	<b>71</b>
<b>Genotype by environment interaction for Portuguese and Brazilian Holstein populations using autoregressive and multi-trait reaction norms test-day models .....</b>	<b>71</b>
3.1. Abstract .....	72
3.2. Introduction .....	73
3.3. Material and methods .....	75
3.4. Results and Discussion.....	79
3.5. Conclusion.....	87
3.6. Acknowledgements .....	87
3.7. References .....	87
<b>GENERAL CONCLUSION.....</b>	<b>91</b>

## GENERAL INTRODUCTION

In dairy cattle, the most important milk traits (e.g., milk yield and milk solids) are longitudinal traits measured over time. The genetic evaluation for these traits may be performed by using traditional models that consider yields aggregated (e.g. 305-d lactation records) or test-day (TD) models that consider the direct use of TD records for milk traits. The genetic evaluation of dairy cattle using TD models has been investigated by several research groups, and some countries have already implemented a routine genetic evaluation of large commercial dairy populations (e.g. Schaeffer et al., 2000; Carvalheira et al., 2002; Vasconcelos et al., 2004; Muir et al., 2007; Interbull, 2018).

Records of daily milk yield may be depicted as repeated observations of a single trait within one lactation. The main issue with TD models is to account for the covariance structure of the repeated records (Carvalheira et al., 2002; Sawalha et al., 2005a; b). Several approaches using TD records were proposed, such as multi-trait, simple repeatability, autoregressive (AR), and random regression (RR) models (Jensen, 2001; Carvalheira et al., 2002; Schaeffer, 2004). There are several advantages of using TD models compared with the traditional use of 305-d lactation records. Among these advantages are more precise adjustment for temporary environmental effects on the TD, avoidance of the use of extended records for culled cows and for records in progress, the possibility of genetic evaluation for persistency of lactation, and increase in reliability of estimates of genetic prediction (Schaeffer et al., 2000; Jensen, 2001; Carvalheira et al., 2002)..

Multiple-trait TD model is a model in which records at each day in milk are considered as different traits. These traits are then analyzed by performing multi-trait

analyses. For this reason, the disadvantage of this model is that many traits must be defined, leading to difficulties such fixed and random effects estimation (Jensen, 2001). According to Sawalha et al. (2005b), and the structure to account for (co)variation among random correlated effects of repeated TD records that should have fewer parameters than the multiple trait approach to avoid over-parameterization. On the other hand, the assumption of simple repeatability animal model of constant environmental and genetic correlations among different TD records may not be realistic.

The most commonly used TD model is probably RR models (Schaeffer, 2004). This model allows us to adjust environmental effects from TD records and, to model the lactation curve shape leading to more accurate genetic evaluations than using the cumulative production in 305 days of lactation. However, RR models require the estimation of large number of parameters and may not be adequate at early or late stages of the lactation (Schaeffer, 2004; Meyer, 2005; Sawalha et al., 2005b).

The AR models may be an alternative approach for modeling TD records. Harville (1979) proposed the use of an autoregressive process to model covariance structures for random effects of repeated measures in animal breeding. In addition, Carvalheira et al. (2002) proposed the autoregressive TD model for multiple lactations. Daily yields are assumed as the same trait within and across lactations under AR. Furthermore, the animal permanent environmental effects are modeled as long- (correlations among records across lactations) and short-term (correlations among records within lactation) effects taking into account the non-genetic autocorrelations due to the cows' repeated performance. In this sense, AR explains random correlated effects of repeated TD records with fewer parameters compared to multi-trait or random regression models (Carvalheira et al., 2002; Sawalha et al., 2005b; a).

Autoregressive and random regression TD models are used by several members of Interbull, such as Canada, Germany, Italy, Netherlands, New Zealand, Portugal, and others in their national dairy cattle genetic evaluations (Interbull, 2018). However, in Brazil, genetic evaluations for the production traits in Holstein cattle use traditional models for 305-d adjusted lactation records. Also, there is no genetic evaluation for somatic cell score (Costa et al., 2014). Therefore, genetic evaluations by using TD models in Brazilian Holstein cattle are required to evaluate the application of these models in this population.

In addition, there is no international genetic evaluation for Brazilian Holstein cattle. This is an important issue considering that 94% of total semen has been imported, mainly from Northern American and some European countries (ASBIA, 2017). Thus, evaluation of the genotype by environment interaction (G x E) between Brazil and other countries with international genetic evaluation (e.g. an Interbull member) may be important for breeding strategies and breeders investing in foreign genotype. For this reason, an international research cooperation between Brazil and Portugal in the genetic evaluation for Holstein cattle was agreed (CAPES/FCT, nº 99999.008462/2014-03).

In Portugal and Brazil, the development of Holstein cattle has been based on continuous importation of semen from Northern American and European countries. Portugal and Brazil have been developing their national genetic evaluation considering their environmental conditions and production systems. The diversification of environmental conditions leads to the different expression of the genetic potential according to environment level. For this reason, the G x E is defined to exist if the difference between genotypes depends on the environment in which they are measured. This includes two different situations: the genotypes ranking can change

among environments, or they can retain the same ranking, but differences can be larger in one environment than in another (Kolmodin et al., 2002; Hayes et al., 2016). Therefore, genotypes that are chosen from distinct regions in which they would be introduced may not consummate breeders' expectations. Given these differences, knowledge of G x E may be relevant for breeders that invest in foreign genotypes (Silva et al., 2014).

The G x E is commonly analyzed as a multiple-trait situation in which the trait measured in the different environments are treated as different correlated traits. Then, the genetic correlation among traits indicates the degree of reranking of animals among the environments (Hayes et al., 2016). Another approach for measuring G x E is by using reaction norm models (Kolmodin et al., 2002; Streit et al., 2012). Reaction norm models are interesting because they use covariance functions to describe the individual's genetic merit as a function of an environmental gradient. The random regression coefficients estimated by these models allow the genetic differences between individuals as well as genetic correlations among different environmental classes (Silva et al., 2014).

Therefore, the general objectives of this thesis were to perform genetic evaluations for Brazilian Holstein cattle by using multiple lactations TD models (AR and RR) and to investigate the presence of genotype by environment interaction within- and across-country for Portuguese and Brazilian Holstein cattle populations.

In addition, the specific objectives were to: i) evaluate the influence of contemporary groups assumed as fixed or random effects and the impact of unknown parent groups inclusion on the genetic evaluations of milk, fat and protein yields and somatic cell scores of Brazilian Holstein cattle, by using an autoregressive repeatability animal model; ii) compare autoregressive and random regression models

for multiple lactations test-day records of milk yield and somatic cell score in the Brazilian Holstein cattle; iii) evaluate genotype by environment interaction for milk yield and somatic cell score from Holstein cattle between Portugal and Brazil by using a two-step approach: EBV prediction within-country using an autoregressive test-day model and genetic correlation estimates between environmental gradients within- and across-country using a multi-trait reaction norm model.

## References

- ASBIA. 2017. Index - Associação Brasileira de Insiminação Artificial 2017. Accessed April 10, 2019. [http://www.asbia.org.br/wp-content/uploads/2018/10/INDEX-ASBIA-2017\\_completo.pdf](http://www.asbia.org.br/wp-content/uploads/2018/10/INDEX-ASBIA-2017_completo.pdf).
- Carvalho, J., E.J. Pollak, R.L. Quaas, and R.W. Blake. 2002. An autoregressive repeatability animal model for test-day records in multiple lactations. *J. Dairy Sci.* 85:2040–2045. doi:10.3168/jds.S0022-0302(02)74281-1.
- Costa, C.N., Cobuci, J.A., Santos, G.G., Thompson, G., Carvalho, J., 2014. Genetic parameters of test day records in Brazilian Holstein cattle using an autoregressive multiple lactation animal model, in: 10th World Congress of Genetics Applied to Livestock Production. Vancouver, Canadá, pp. 5–7.
- Harville, D.A. 1979. Recursive estimation using mixed linear models with autoregressive random effects. Proc. Conf. R.C. Henderson, ed. Cornell Univ., Ithaca, NY.
- Hayes, B.J., H.D. Daetwyler, and M.E. Goddard. 2016. Models for genome x environment interaction: examples in livestock. *Crop Sci.* 56:2251–2259. doi:10.2135/cropsci2015.07.0451.
- Interbull. 2018. National Genetic Evaluation Forms Provided by Countries. Accessed June 19, 2018. <http://www.interbull.org/ib/geforms>.
- Jensen, J. 2001. Genetic evaluation of dairy cattle using test-day models. *J. Dairy Sci.* 84:2803–2812. doi:10.3168/jds.S0022-0302(01)74736-4.
- Kolmodin, R., E. Strandberg, P. Madsen, J. Jensen, and H. Jorjani. 2002. Genotype by Environment Interaction in Nordic Dairy Cattle Studied Using Reaction Norms. *Acta Agric. Scand.* 52:11–24. doi:10.1080/09064700252806380.

- Meyer, K. 2005. Estimates of genetic covariance functions for growth of Angus cattle. *J. Anim. Breed. Genet.* 122:73–85. doi:10.1111/j.1439-0388.2005.00503.x.
- Muir, B.L., G. Kistemaker, J. Jamrozik, and F. Canavesi. 2007. Genetic parameters for a multiple-trait multiple-lactation random regression test-day model in Italian Holsteins. *J. Dairy Sci.* 90:1564–1574. doi:10.3168/jds.S0022-0302(07)71642-9.
- Sawalha, R.M., J.F. Keown, S.D. Kachman, and L.D. Van Vleck. 2005a. Genetic evaluation of dairy cattle with test-day models with autoregressive covariance structures and with a 305-d model. *J. Dairy Sci.* 88:3346–3353. doi:10.3168/jds.S0022-0302(05)73018-6.
- Sawalha, R.M., J.F. Keown, S.D. Kachman, and L.D. Van Vleck. 2005b. Evaluation of autoregressive covariance structures for test-day records of Holstein cows: estimates of parameters. *J. Dairy Sci.* 88:2632–2642. doi:10.3168/jds.S0022-0302(05)72940-4.
- Schaeffer, L.R. 2004. Application of random regression models in animal breeding. *Livest. Prod. Sci.* 86:35–45. doi:10.1016/S0301-6226(03)00151-9.
- Schaeffer, L.R., J. Jamrozik, G.J. Kistemaker, and B.J. Van Doormaal. 2000. Experience with a test-day model. *J. Dairy Sci.* 83:1135–1144. doi:10.3168/jds.S0022-0302(00)74979-4.
- Silva, F.F., H.A. Mulder, E.F. Knol, M.S. Lopes, S.E.F. Guimaraes, P.S. Lopes, P.K. Mathur, J.M.S. Viana, and J.W.M. Bastiaansen. 2014. Sire evaluation for total number born in pigs using a genomic reaction norms approach. *J. Anim. Sci.* 92:3825–3834. doi:10.2527/jas.2013-6486.
- Streit, M., F. Reinhardt, G. Thaller, and J. Bennewitz. 2012. Reaction norms and genotype-by-environment interaction in the German Holstein dairy cattle. *J. Anim. Breed. Genet.* 129:380–389. doi:10.1111/j.1439-0388.2012.00999.x.
- Vasconcelos, J., A. Martins, M.F. Petim-Batista, J. Colaço, R.W. Blake, and J. Carnevalheira. 2004. Prediction of daily and lactation yields of milk, fat, and protein using an autoregressive repeatability test day model. *J. Dairy Sci.* 87:2591–8. doi:10.3168/jds.S0022-0302(04)73384-6.

## CHAPTER 1

### **Unknown parent and contemporary groups for genetic evaluation of Brazilian Holstein using autoregressive test-day models**

Delvan Alves da Silva<sup>1</sup>, Cláudio Napolis Costa<sup>2</sup>, Alessandra Alves da Silva<sup>1</sup>,  
Fabyano Fonseca e Silva<sup>1</sup>, Paulo Sávio Lopes<sup>1</sup>, Glaucyana Gouvêa dos Santos<sup>2</sup>,  
Gertrude Thompson<sup>3,4</sup> and Júlio Carneiro<sup>3,4</sup>

<sup>1</sup> Departamento de Zootecnia, Universidade Federal de Viçosa, 36570-000, Viçosa, Brazil

<sup>2</sup> Embrapa Gado de Leite, 36038-330, Juiz de Fora, Brazil

<sup>3</sup> Research Center in Biodiversity and Genetic Resources (CIBIO-InBio), University of Porto, 4485-661, Vairão, Portugal

<sup>4</sup> Institute of Biomedical Sciences Abel Salazar (ICBAS), University of Porto, 4050-313, Porto, Portugal

## 1.1. Abstract

Test-day milk, fat and protein yields and somatic cell score (SCS) of Brazilian Holstein cattle were used to study the effect of fixed (HTDF) vs random (HTDR) contemporary groups (herd-test-date) with or without unknown parent groups (UPG) using an autoregressive test-day animal model for genetic evaluations. Therefore, four models were used in this study to evaluate these effects: HTDF with or without UPG and HTDR with or without UPG. A total of 4,142,740 test-day records from the first three lactations were used in this study. The data was provided by the Brazilian Holstein Cattle Breeders Association and obtained from 1994 to 2016. UPG were defined by geographic region, sex and birth year of the animals with missing parents resulting in a total of 133 unknown sire groups and 153 unknown dam groups. The additive genetic variances estimated with HTDR decreased in all traits (except for SCS), but also did the estimated phenotypic variances, leaving the heritabilities almost the same in both models and ranging from 0.13 for fat (3<sup>rd</sup> lactation) fitted with HTDR to 0.24 for milk (1<sup>st</sup> lactation) fitted with HTDF. The recovery of information performed on the HTDR models increased the lower reliabilities from 0.50 to 0.75, 0.54 to 0.66, 0.64 to 0.71 and 0.25 to 0.67 for milk, fat and protein yields, and for SCS, respectively (considering only bulls with 10 or more daughters). The difference in annual genetic gains between models (HTDR vs HTDF) was for sires (cows) of 30.66 (38.59) kg, 1.18 (1.35) kg, 1.26 (1.22) kg and -0.001 (-0.03) scores for milk, fat and protein yields and SCS, respectively. The contemporary groups as a random effect in the AR model is more relevant than considering just the UPG effect, but the combination of both may provide higher annual genetic gains. The rank correlations between HTDF without UPG and the HTDR with UPG models were higher for cows (from 0.77 to 0.85) than for bulls (from 0.73 to 0.82). The HTDR with or without UPG, were the models with the best

results indicating that there was no significant bias in the genetic trends using Interbull's methods 1 and 2. In general, the HTDR model with UPG was the procedure that best fitted these traits and should be the model of choice for genetic evaluations and genetic trend analysis of longitudinal traits in Brazilian Holstein cattle.

**Keywords:** autoregressive model; dairy cattle; genetic group; genetic trend

## 1.2. Introduction

Test-day (**TD**) models have been used by several countries for genetic evaluation of longitudinal traits in dairy cattle. These models provide more accurate breeding value estimates of cows and bulls compared to cumulative 305-d lactation models. An alternative to the multiple traits random regressions approach for modeling TD records is the autoregressive (**AR**) model for multiple lactations proposed by Carvalheira et al. (2002). Under this model, daily yields are assumed as the same trait within and across lactations. Furthermore, the animal permanent environmental effects are modeled as long-term (correlations between records across lactations) and short-term (correlations between records within lactation) effects taking into account the non-genetic autocorrelations due to the cows' repeated performance.

In a genetic evaluation, the minimum number of records within each contemporary group (**CG**) is a critical factor related to bias in the prediction of breeding values. CG may be fitted as fixed or random effects, such that this decision will define the balance between accuracy and bias. Using the CG as a random effect may recover information across groups (Strabel et al., 2005; Ugarte et al., 1992; Van Vleck, 1987; Visscher and Goddard, 1993), especially when the number of records in some CG is small. However, it will introduce bias in the breeding values predictions

if there are non-random associations of animals with effects considered to be fixed (Ugarte et al., 1992; Vasconcelos et al., 2008; Visscher and Goddard, 1993). If these non-random associations can be considered negligible, assuming CG as a random effect will contribute to recover information that otherwise would be lost due to data editing.

In Brazil, the majority of dairy herds supervised by the official milk recording system are characterized by their small size, which implies the existence of a high frequency of CG (levels of herd-test-dates, **HTD**) with less than 3 records. Excluding these farms from the genetic evaluations may cause bias in the prediction of the bulls' breeding values due to the loss of information on their daughters, with negative consequences on the genetic progress (Vasconcelos et al., 2008).

The genetic composition of the Brazilian Holstein population is characterized by the continuous importation of semen from international proven bulls, the importation of cows from Latin American countries and by cows originated from commercial Holstein herds, previously not registered in the official identification system. This scenario contributes for a large number of animals of unknown ancestry. Assuming that they all come from a base population with the same genetic level may not be valid. Under these circumstances, including unknown parent groups (UPG) in the genetic modeling could help in eliminating or minimizing bias from the predicted EBVs.

Over the years, the importance of the genetic groups in the genetic evaluations has been described by several authors (Alfonso and Estany, 1999; Theron et al., 2002; Phocas and Laloë, 2004; Fikse, 2009; Misztal et al., 2013; Tsuruta et al., 2014). There is evidence that the inclusion of UPG in the model decreases bias in the genetic trend (Theron et al., 2002). On the other hand, Phocas and Laloë, (2004) reported that the

inclusion of UPG may lead to wrong ranking of animals and suboptimal selection decision. These apparent contradictory studies reinforce the need to assess the inclusion of UPG in national genetic evaluation schemes, particularly in heterogeneous structured populations, such as the Brazilian Holstein cattle.

This study was aimed at evaluate the influence of contemporary groups (HTD) assumed as fixed (HTDF) or random (HTDR) effects and the impact of UPG inclusion on the genetic evaluations of milk, fat and protein yields and somatic cell scores (SCS) of Brazilian Holstein cattle, using an autoregressive repeatability animal model.

### **1.3. Materials and methods**

#### Data and Pedigree file

A total of 4,142,740 test-day records of milk, fat and protein yields and somatic cell count (SCC) from the first three lactations of Brazilian Holsteins cattle collected from 1994 to 2016 were analyzed. The SCC measurements were log transformed to SCS using the formula  $SCS = \log_2 (SCC / 100) + 3$ . The data were provided by the Brazilian Holstein Cattle Breeders Association.

The dataset was edited to eliminate the following records of cows: i) with improper identification; ii) without herd codes; iii) with less than two TD records per lactation; iv) with unknown birth or calving dates; and v) farms with less than three cows. Test-day records corresponded to the range of 5 to 305 d in milk and time interval between consecutive TD was approximately 30 d. We assumed age at calving between 18 and 45, 28 and 58, 38 and 78 mo for the first, second and third lactations, respectively. Cow ages at calving were grouped into 16 classes: class 1, with cows less than 20 mo at calving; classes 2 to 15 were formed for every 4-mo interval from 20 to 75 mo; and class 16 for cows over 75 mo old at calving. Days in milk (DIM) were

grouped into 30 classes with 10 d intervals. To remove outliers, TD records over  $\pm 3$  phenotypic standard deviations within DIM, age class and parity, were deleted.

Two datasets were created according to the two scenarios under study. In the first, the data were edited assuming HTD as a fixed effect in the AR model and, in this case, herds with HTD levels with less than 3 observations were removed. In the second, HTD was considered a random effect and no restrictions were applied to the number of records per HTD level. Number of records and descriptive statistics for all traits in each scenario are in Table 1.

The pedigree file consisted of 7,912 bulls and 345,769 cows. A total of 409 bulls and 98,475 cows had at least one unknown parent (Table 2). Animals with unknown parents originated from 20 countries and were organized in 5 regions according to their geographic location, e.g., Europe, North America, South America (except Brazil), Oceania, and Brazil. UPG were arranged by region, sex and year of birth resulting in a total of 133 unknown sire groups and 153 unknown dam groups, ensuring a minimum of 10 observations per group.

#### Autoregressive test-day model

The genetic evaluation was performed using the AR test-day model (Carvalho et al. 1998, 2002). Four scenarios were analyzed: 1) HTDF w/ UPG, 2) HTDF w/o UPG, 3) HTDR w/ UPG and 4) HTDR w/o UPG.

In matrix notation, the AR model for scenarios 1 and 3 can be written as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{Q}\mathbf{g} + \mathbf{Z}\mathbf{a} + \mathbf{M}\mathbf{p} + \mathbf{N}\mathbf{t} + \mathbf{r}, \quad (1)$$

where:  $\mathbf{y}$  is the vector of TD records;  $\mathbf{X}\boldsymbol{\beta} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{X}_3\boldsymbol{\beta}_3$ , where  $\boldsymbol{\beta}_1$  the vector of the fixed or random effect of HTD;  $\boldsymbol{\beta}_2$  the vector of the fixed effect of age class at calving nested within herd; and  $\boldsymbol{\beta}_3$  the vector of the fixed effect of DIM nested within

herd and lactation order;  $\mathbf{Q}$  is the matrix that relates animals to their UPG;  $\mathbf{g}$  is the vector of UPG effect;  $\mathbf{a}$  is the vector of random additive genetic effect;  $\mathbf{p}$  is the vector of long-term environmental (LTE) random effect;  $\mathbf{t}$  is the vector of short-term environmental (STE) random effect; and  $\mathbf{r}$  is the vector of random residual effects fitted with a heterogeneous covariance structure across lactations. The  $\mathbf{X}_i$ ,  $\mathbf{Z}$ ,  $\mathbf{M}$  and  $\mathbf{N}$  are the incidence matrices relating observations to fixed and random effects. To fit the UPG effect, we used the same methodology proposed by Westell et al. (1988) with the transformed equations described by Quaas and Pollak (1981) and Quaas (1988). In this case,  $\mathbf{a}^*$  is a vector of random additive genetic effects, including UPG effects ( $\mathbf{g}$ ) and additive genetic effects ( $\mathbf{a}$ ):  $\mathbf{a}^* = \mathbf{Qg} + \mathbf{a}$ . The models without UPG (scenarios 2 and 4) has the same definitions, except the terms that include UPG information. A first order autoregressive covariance structure was assumed for HTD (when random), LTE and STE effects. More details on the AR model may be found in Carvalheira et al. (1998, 2002).

For the estimation of variance components (VC) and autocorrelation coefficients, six random samples of the data were created. Each sample consisted of data from twenty herds representing all regions of Brazil. The VC and autocorrelation coefficients were estimated using DFREML procedures (Boldman and Van Vleck, 1991; Smith and Graser, 1986). The convergence criterion was established when the variance of the simplex was less than  $10^{-8}$ . The occurrence of local maxima was checked by five consecutive cold starts without significant change in the log-likelihood (up to four decimal places). According to Theron et al. (2002), the inclusion of UPG in the model was not relevant for the estimation of VC. Therefore, in this study, only the models without UPG were used for estimation of those parameters.

## Genetic trend validation

Interbull's methods 1 and 2 described by Boichard et al. (1995) were used to validate the genetic trends. In the method 1, genetic trends are compared using the bull's genetic merit obtained using all 3 lactations versus just the first lactation. In this case, two regression models were fitted:

$$Y_1 = b_{01} + b_1 B_{\text{Year}} + e_i,$$

$$Y_3 = b_{03} + b_3 B_{\text{Year}} + e_i,$$

where  $Y_1$  and  $Y_3$  are the bull's EBV and  $b_1$  and  $b_3$  are the regression coefficients (the subscripts represent the number of lactations used on the EBV prediction);  $B_{\text{Year}}$  is the bull birth year. According to Interbull (2016), the  $|b_3 - b_1|$  should be less than  $0.02\sigma_a$  ( $\sigma_a$  = additive genetic standard deviation) and if  $b_1 < b_3$ , there is overestimation (the opposite if  $b_1 > b_3$ ).

In the method 2, daughter yield deviations – DYD (Liu et al., 2004) were computed for each bull by calving year of the daughters to determine whether these deviations remain stable over time (Boichard et al., 1995). The following regression model was fitted to these DYDs:

$$Y_{ij} = \text{Bull}_i + b_j + e_{ij},$$

where,  $Y_{ij}$  is the DYD considering daughters of the  $i^{\text{th}}$  bull that calved in the  $j^{\text{th}}$  year;  $j=0$  for the first year when at least 10 daughters of a bull calved for the first time;  $\text{Bull}_i$  is the effect of the  $i^{\text{th}}$  bull. In this case, the regression coefficient ( $b$ ) should be less than  $0.01\sigma_a$  and not significantly different from zero. If  $b > 0$  there is overestimation or the opposite, if  $b < 0$  (Boichard et al., 1995; Interbull, 2016).

#### 1.4. Results and Discussion

Assuming the HTD effect as random permitted recover information that otherwise would not contribute for the evaluations due to data editing. Strabel et al. (2005) reported that small HTD class is a consequence of the small herds in the population, which results in a large proportion of cows without or only with few contemporaries. In the present study, the loss of information corresponded to approximately 3% of the records and more than 25% of the herds. As expected, most of these herds were small and their exclusion had no impact on the levels of production (Table 1).

Animals with at least one unknown parent corresponded to almost 28% of the population (Table 2). The majority of these animals were females and the historical explanation refers to an open herd-book policy to cows from commercial Holstein herds, not yet included in the official identification system of the breeders association. There is also an important number of cows imported from Latin American countries with only partial pedigree information for official registration.

Variance components, autocorrelations and heritabilities ( $h^2$ ) estimates for milk, fat and protein yields and SCS are shown in Table 3. The relative magnitude of the VC and autocorrelations for all traits in this study are consistent with the literature (e.g., Carvalheira et al., 1998; Vasconcelos et al., 2008; Costa et al., 2014). The additive genetic variances estimated with random HTD decreased in all traits (except for SCS), but also did the estimated phenotypic variances, leaving the heritabilities almost the same in both models and ranging from 0.24 for milk in the first lactation fitted with HTDF to 0.13 for fat in the third lactation fitted with HTDR. The  $h^2$  estimates for all traits were moderate to low and decreased from the first to the third lactation (Table 3).

Working with an AR model with fixed HTD on a data sub-set of Holstein Brazilian cattle (only large and well-structured herds), Costa et al. (2014) reported higher  $h^2$  estimates (0.23 to 0.35, 0.23 to 0.38, 0.23 to 0.35 and 0.19 to 0.20 for milk, fat and protein yields and SCS, respectively) due to lower estimates of residual variances. On the other hand, the  $h^2$  estimates for the second and third lactations were similar to those reported by Vasconcelos et al. (2008) for the same traits using an AR model with fixed HTD in Portuguese Holstein. The first lactation  $h^2$  however, was higher than the ones obtained in the present study because of a corresponding lower residual variance.

The loss of information due to data editing in the HTDF models had a strong negative impact on the magnitude of the accuracies of EBV, especially on sires in the lower range of reliabilities (Figure 1). On average, 48% of bulls that had 10 or more daughters in the full data set (used for the HTDR analysis) lost between 1 and 10 (6% of bulls lost more than 10) daughters due to the data edition for the HTDF analysis. The recovery of information performed by the HTDR models increased the lower reliabilities from a minimum of 0.50 to 0.75, 0.54 to 0.66, 0.64 to 0.71 and 0.25 to 0.67 for milk, fat and protein yields, and for SCS, respectively (considering only bulls with 10 or more daughters). Strabel et al. (2005), also found that assuming HTD as a random effect may be a strategy to improve the genetic evaluation by reducing the residual and prediction error variances, thus increasing EBV reliabilities.

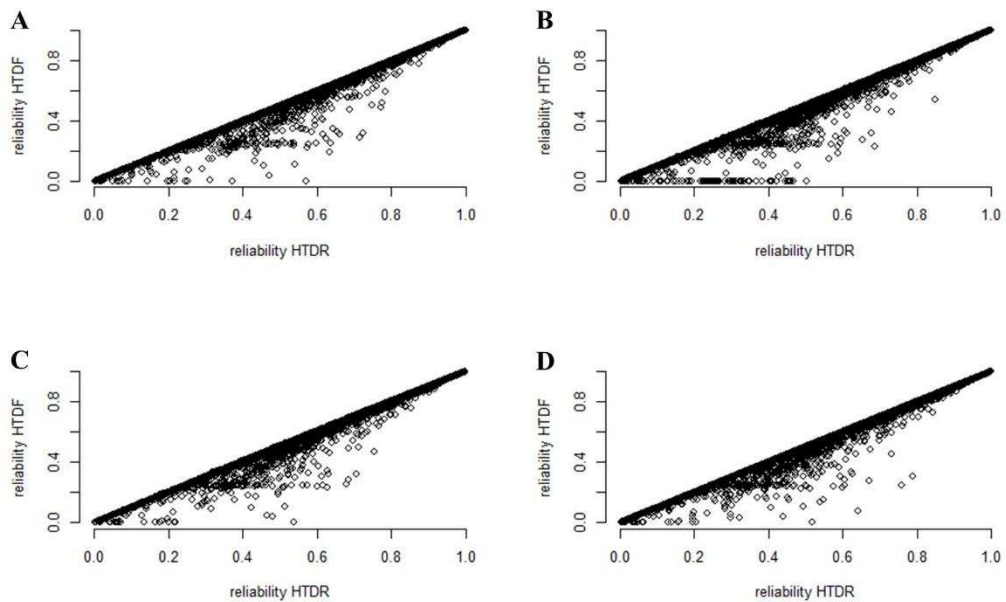
**Table 1.** Description of dataset and effects in the model using contemporary groups as a fixed (HTDF) or random (HTDR) effect for milk, fat and protein yields and somatic cell score (SCS).

Item <sup>a</sup>	HTDF				HTDR			
	Milk	Fat	Protein	SCS	Milk	Fat	Protein	SCS
TD records	3,247,029	2,863,077	2,766,534	2,668,776	3,333,752	2,951,521	2,842,279	2,741,591
TD records, lactation 1	1,528,069	1,338,094	1,287,159	1,237,563	1,577,965	1,387,845	1,328,894	1,277,589
TD records, lactation 2	1,041,944	922,817	893,465	865,264	1,065,493	947,653	915,080	886,362
TD records, lactation 3	677,016	602,166	585,910	565,949	690,294	616,023	598,305	577,640
Cows	210,556	201,824	191,851	186,851	217,197	210,023	199,050	194,143
Cows, lactation 1	193,425	182,008	172,126	167,131	200,068	189,971	178,759	173,779
Cows, lactation 2	131,592	124,796	118,559	115,531	134,632	128,794	121,976	119,006
Cows, lactation 3	86,708	81,996	78,267	76,508	88,291	84,094	80,177	78,294
Herds	897	828	785	755	1,204	1,152	1,073	1,029
HTD classes	85,595	76,758	71,250	68,445	116,733	107,948	97,389	93,457
HLDIM classes	62,579	55,737	52,766	50,639	70,102	63,387	59,563	57,183
HAGE classes	11,295	10,376	9,791	9,395	13,380	12,615	11,757	11,294
Mean (SD)	28.27 (8.35)	0.95 (0.31)	0.89 (0.25)	3.08 (2.16)	28.17 (8.36)	0.94 (0.31)	0.88 (0.25)	3.09 (2.16)
Mean (SD), lactation 1	26.32 (7.01)	0.88 (0.27)	0.82 (0.22)	2.63 (2.01)	26.23 (7.03)	0.88 (0.27)	0.82 (0.22)	2.65 (2.01)
Mean (SD), lactation 2	29.66 (8.74)	0.99 (0.32)	0.93 (0.26)	3.24 (2.18)	29.57 (8.75)	0.99 (0.32)	0.93 (0.26)	3.25 (2.19)
Mean (SD), lactation 3	30.52 (9.45)	1.02 (0.34)	0.95 (0.27)	3.80 (2.23)	30.44 (9.46)	1.02 (0.34)	0.95 (0.28)	3.80 (2.23)

<sup>a</sup>TD: test-day; HTD: herd-test-day; HLDIM: days in milk class within herd and lactation; HAGE: age class within herd; SD: standard deviation

**Table 2.** Animals with unknown parents in pedigree file.

Animals	Number	% <u>within</u> sex
Cows with unknown:		
Sire	179	0.05
Dam	6,840	1.98
Sire and Dam	91,456	26.45
Bulls with unknown:		
Sire	30	0.38
Dam	28	0.35
Sire and Dam	351	4.44



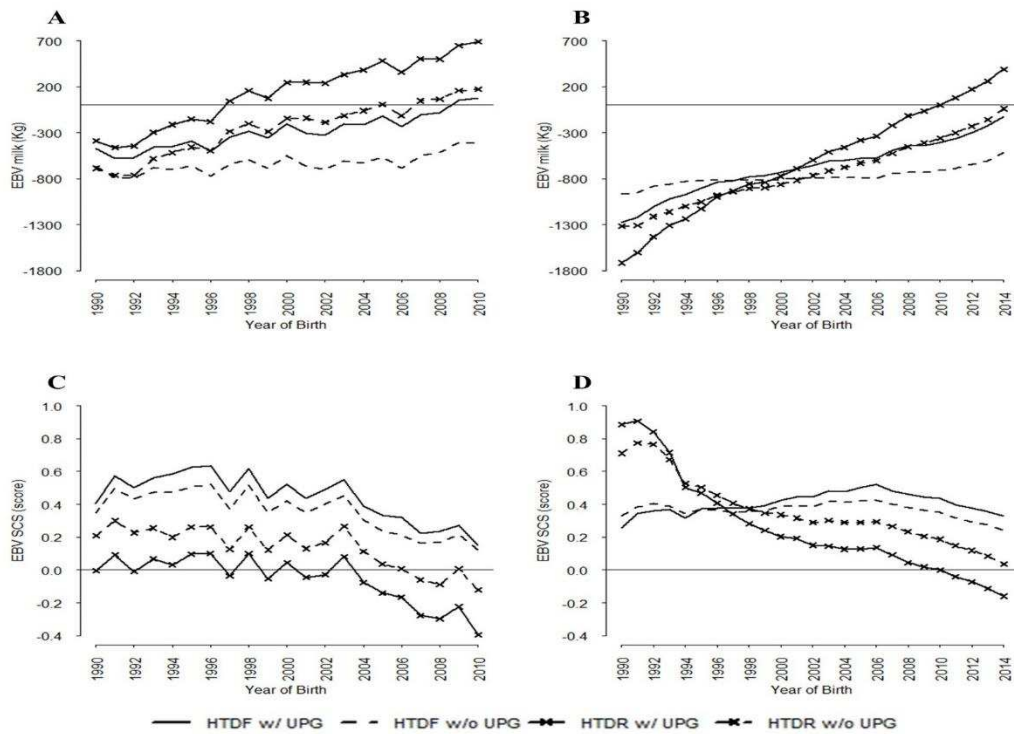
**Figure 1.** Reliability of the breeding values predicted for milk (A), fat (B) and protein (C) yield and somatic cell score (SCS-D) for bulls from genetic evaluations with contemporary groups as a fixed (HTDF) or random (HTDR) effect.

**Table 3.** Estimates<sup>1</sup> of variance components, autocorrelations, and heritability for milk, fat and protein yields and somatic cell score (SCS) considering contemporary groups as a fixed (HTDF) or random (HTDR) effects in AR model.

Estimates	HTDF				HTDR			
	Milk (SE)	Fat (SE)	Protein (SE)	SCS (SE)	Milk (SE)	Fat (SE)	Protein (SE)	SCS (SE)
$\hat{\sigma}_a^2$	10.32 (0.113)	0.013 (0.0002)	0.008 (0.0001)	0.57 (0.013)	8.83 (0.313)	0.012 (0.0005)	0.008 (0.0002)	0.66 (0.017)
$\hat{\sigma}_{e1}^2$	11.74 (0.100)	0.031 (0.0007)	0.010 (0.0002)	1.12 (0.005)	8.35 (0.632)	0.028 (0.0013)	0.011 (0.0005)	1.33 (0.038)
$\hat{\sigma}_{e2}^2$	14.49 (0.215)	0.042 (0.0011)	0.013 (0.0002)	1.21 (0.005)	9.89 (0.799)	0.039 (0.0022)	0.014 (0.0007)	1.34 (0.044)
$\hat{\sigma}_{e3}^2$	14.96 (0.225)	0.046 (0.0014)	0.013 (0.0003)	1.15 (0.006)	11.64 (0.891)	0.046 (0.0023)	0.016 (0.0009)	1.40 (0.040)
$\hat{\sigma}_p^{2*}$	≈0.00 (≈0.00)	≈0.00 (≈0.00)	≈0.00 (≈0.00)	≈0.00 (≈0.00)	≈0.00 (≈0.00)	≈0.00 (≈0.00)	≈0.00 (≈0.00)	≈0.00 (≈0.00)
$\hat{\sigma}_{t1}^2$	21.31 (0.321)	0.018 (0.0002)	0.017 (0.0003)	1.45 (0.007)	17.32 (0.459)	0.018 (0.0009)	0.016 (0.0004)	1.64 (0.015)
$\hat{\sigma}_{t2}^2$	29.38 (0.578)	0.029 (0.0005)	0.025 (0.0006)	1.84 (0.014)	20.56 (0.672)	0.026 (0.0014)	0.020 (0.0009)	2.02 (0.056)
$\hat{\sigma}_{t3}^2$	36.26 (0.729)	0.035 (0.0004)	0.030 (0.0007)	2.35 (0.042)	26.68 (1.190)	0.033 (0.0017)	0.025 (0.0009)	2.35 (0.089)
$\hat{\sigma}_c^2$	–	–	–	–	3.71 (0.128)	0.007 (0.0002)	0.004 (0.0003)	0.21 (0.010)
$\hat{\rho}_p^*$	≈0.00 (≈0.00)	≈0.00 (≈0.00)	≈0.00 (≈0.00)	≈0.00 (≈0.00)	≈0.00 (≈0.00)	≈0.00 (≈0.00)	≈0.00 (≈0.00)	≈0.00 (≈0.00)
$\hat{\rho}_{t1}$	0.69 (0.002)	0.71 (0.0017)	0.67 (0.0029)	0.84 (0.001)	0.66 (0.004)	0.67 (0.0090)	0.66 (0.0021)	0.84 (0.006)
$\hat{\rho}_{t2}$	0.78 (0.002)	0.80 (0.0018)	0.77 (0.0020)	0.83 (0.001)	0.75 (0.004)	0.79 (0.0035)	0.75 (0.0037)	0.82 (0.005)
$\hat{\rho}_{t3}$	0.76 (0.001)	0.79 (0.0025)	0.74 (0.0022)	0.85 (0.003)	0.76 (0.004)	0.80 (0.0026)	0.76 (0.0040)	0.84 (0.003)
$\hat{\rho}_c$	–	–	–	–	0.67 (0.008)	0.62 (0.0149)	0.66 (0.0155)	0.58 (0.009)
$\hat{h}_{L1}^2$	0.24 (0.004)	0.21 (0.0051)	0.22 (0.0030)	0.18 (0.003)	0.23 (0.007)	0.19 (0.0034)	0.20 (0.0070)	0.17 (0.005)
$\hat{h}_{L2}^2$	0.19 (0.004)	0.16 (0.0045)	0.17 (0.0027)	0.16 (0.003)	0.21 (0.007)	0.15 (0.0035)	0.16 (0.0068)	0.16 (0.005)
$\hat{h}_{L3}^2$	0.17 (0.003)	0.14 (0.0038)	0.15 (0.0023)	0.14 (0.004)	0.17 (0.007)	0.13 (0.0027)	0.14 (0.0060)	0.14 (0.004)

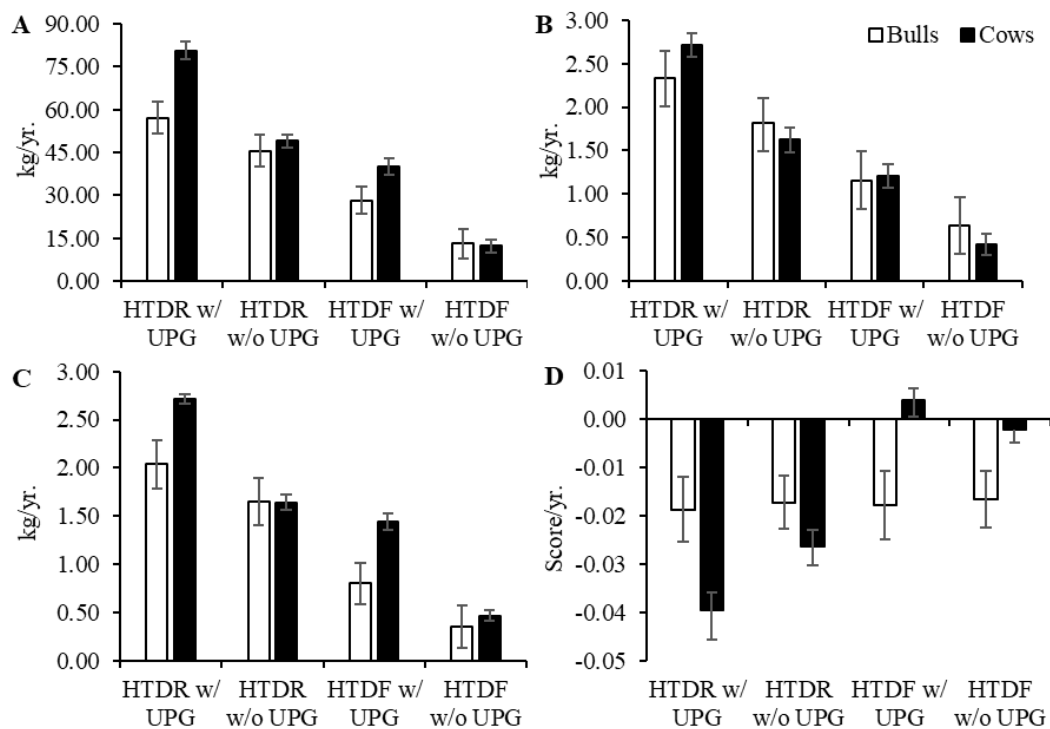
<sup>1</sup>  $\hat{\sigma}_a^2$ : additive genetic variance (kg<sup>2</sup> for milk, fat and protein yields and score units<sup>2</sup> for SCS);  $\hat{\sigma}_p^2$ : LTE variance;  $\hat{\sigma}_{ti}^2$ : STE variance;  $\hat{\sigma}_c^2$ : herd-test-day (HTD) variance;  $\hat{\sigma}_{ei}^2$ : residual variance;  $\hat{\rho}_p$ ,  $\hat{\rho}_{ti}$  and  $\hat{\rho}_c$ : autocorrelations for LTE, STE and HTD;  $\hat{h}_{Li}^2$ : heritability, where  $i=1, 2$  and  $3$  correspond to first, second and third lactations, respectively. SE: standard error. \* Values <0.0001(<0.00001)

The effect of assuming a fixed or random HTD, with or without UPG, on the genetic progress of the Brazilian Holstein population, is illustrated in Figure 2 for milk yield and SCS (the shape of the curves for fat and protein yields are similar to the one of milk). For comparison, all EBVs were deviated from the mean EBV of cows born in 2010 (genetic reference base) evaluated with the HTDR w/ UPG model. A much steeper genetic trend and higher annual genetic gains were obtained using the HTDR when compared with the HTDF models, regardless of including or not UPG. Similar results were also obtained by other authors when considering the CG as a random effect in the model (Mayers et al., 2002; Oikawa and Sato, 1997; Strabel and Szwaczkowski, 1999). The difference in annual genetic gains between models (HTDR vs HTDF) was for sires (cows) of 30.6 (38.59) kg, 1.18 (1.35) kg, 1.26 (1.22) kg and -0.001 (-0.03) scores for milk, fat and protein yields and SCS, respectively (Figure 3). Within each model, the effect of the UPG was also important, increasing the annual genetic gains for all traits. For bulls (cows), the mean annual genetic gains for milk, fat and protein yields and SCS were 13.35 (29.74) kg, 0.51 (0.93) kg, 0.42 (1.02) kg and -0.001 (-0.004) scores, respectively, higher in models w/ UPG than in models w/o UPG (Figure 3). Similar results were found by Theron et al. (2002) in South African Holstein cattle. These results show that the HTD as a random effect in the AR model was more relevant than considering just the UPG effect, but the combination of both may promote even higher annual genetic gains.



**Figure 2.** Genetic trend for milk yield and somatic cell score (SCS) for bulls (A and C) and cows (B and D), respectively, born between 1990 and 2014 from genetic evaluations with contemporary groups as a fixed (HTDF) or random (HTDR) effect and with (w/) or without (w/o) unknown parent groups (UPG).

The greatest impact on the EBV re-ranking of animals was between the HTDF without UPG and the HTDR with UPG models (Supplementary Table S1). As expected, the loss of information here represented by the exclusion of small farms, was more important for bulls (rank correlations varied from 0.73 in protein to 0.82 in SCS) than for cows (rank correlations varied from 0.77 in protein to 0.85 in SCS). Vasconcelos et al. (2005) also reported that the exclusion of the small herds in the Portuguese Holstein genetic evaluation may have influenced the selection of bulls and possibly reducing the rate of the genetic progress.



**Figure 3.** Annual genetic gain for milk (A), fat (B) and protein (C) yield and somatic cell score (SCS-D) for bulls and cows born between 1990 and 2014 from evaluations with contemporary groups as a fixed (HTDF) or random (HTDR) effect and, with (w/) or without (w/o) unknown parent groups (UPG).

Bias on genetic trends was tested by Interbull's methods 1 and 2 for all traits and validated (no significant bias) the results for the HTDR model with or without UPG (Table 4). On the other hand, the analysis by method 2 on the genetic trends obtained with the HTDF model (especially without UPG), indicated that there was a significant bias, suggesting an overestimation of the EBVs for fat and protein yields and SCS (Table 4). In general, the HTDR model with UPG was the procedure that best fitted these traits and should be considered the model of choice for genetic evaluations and genetic trend analysis of longitudinal traits in Brazilian Holstein cattle.

**Table 4.** The absolute values of regression coefficients (b) for validation Method 2 for milk, fat and protein yields and somatic cell score (SCS).

	HTDF			HTDR		
	$0.01\sigma_a$	b  w/ UPG	b  w/o UPG	$0.01\sigma_a$	b  w/ UPG	b  w/o UPG
Milk	9.797	0.339	7.718	9.062	5.528	1.074
Fat	0.353	0.239	0.569**	0.340	0.281	0.054
Protein	0.265	0.238*	0.474***	0.264	0.205	0.027
SCS	0.008	0.010	0.010	0.008	0.001	0.003

$\sigma_a$  = genetic additive standard deviation; \* ( $P < 0.05$ ), \*\*( $P < 0.01$ ) and \*\*\*( $P < 0.001$ ); genetic evaluations with contemporary groups as a fixed (HTDF) or random (HTDR) effect and, with (w/) or without (w/o) unknown parent groups (UPG).

## 1.5. Conclusion

The amount of information recovered by using random CG and the correction of the base population with the inclusion of UPG showed a strong and positive effect on the genetic evaluations of longitudinal traits in Brazilian Holstein cattle. In general, the HTDR model with UPG was the one with the best results, maximizing the use of the available data, with the highest EBV reliabilities, unbiased genetic trends and maximizing the annual genetic gains of the Brazilian Holstein dairy cattle.

## 1.6. Acknowledgments

The authors acknowledge Brazilian Holstein Cattle Breeders Association (ABCBRH) for providing data for this study. This study was partially financed by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior and Fundação para Ciência e Tecnologia (CAPES/FCT, nº 99999.008462/2014-03 and 88887.125450/2016-00), and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq 465377/2014-9 - PROGRAMA INCT and CNPq 142467/2015-4).

## 1.7. References

- Alfonso, L., Estany, J., 1999. An expression of mixed animal model equations to account for different means and variances in the base population. *Genet. Sel. Evol.* 31, 105–113. doi:10.1186/1297-9686-31-2-105.
- Boichard, D., Bonaiti, B., Barbat, A., Mattalia, S., 1995. Three Methods to Validate the Estimation of Genetic Trend for Dairy Cattle. *J. Dairy Sci.* 78, 431–437.
- Boldman, K.G., Van Vleck, L.D., 1991. Derivative-free restricted maximum likelihood estimation in animal models with a sparse matrix solver. *J. Dairy Sci.* 74, 4337–4343. doi:10.3168/jds.S0022-0302(91)78629-3.
- Carvalho, J., Pollak, E.J., Quaas, R.L., Blake, R.W., 2002. An autoregressive repeatability animal model for test-day records in multiple lactations. *J. Dairy Sci.* 85, 2040–2045. doi:10.3168/jds.S0022-0302(02)74281-1.
- Carvalho, J.G., Blake, R.W., Pollak, E.J., Quaas, R.L., Duran-Castro, C. V., 1998. Application of an autoregressive process to estimate genetic parameters and breeding values for daily milk yield in a tropical herd of Lucerna cattle and in United States Holstein herds. *J. Dairy Sci.* 81, 2738–2751. doi:10.3168/jds.S0022-0302(98)75831-X.
- Costa, C.N., Cobuci, J.A., Santos, G.G., Thompson, G., Carvalho, J., 2014. Genetic parameters of test day records in Brazilian Holstein cattle using an autoregressive multiple lactation animal model, in: 10th World Congress of Genetics Applied to Livestock Production. Vancouver, Canadá, pp. 5–7.
- Fikse, F., 2009. Fuzzy classification of phantom parent groups in an animal model. *Genet. Sel. Evol.* 41, 1–8. doi:10.1186/1297-9686-41-42.
- Interbull, 2016. Interbull CoP - APPENDIX III - Interbull trend validation procedure. Accessed October 12, 2017. [http://www.interbull.org/ib/cop\\_appendix2](http://www.interbull.org/ib/cop_appendix2).
- 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, 1896–1907. doi:10.3168/jds.S0022-0302(04)73348-2.
- Mayers, P., Stoll, J., Reents, R., Gengler, N., 2002. Alternative modeling of fixed effects in test-day models to increase their usefulness for management decisions. *Interbull Bull.* 29, 128–132.

- Misztal, I., Vitezica, Z.G., Legarra, A., Aguilar, I., Swan, A.A., 2013. Unknown-parent groups in single-step genomic evaluation. *J. Anim. Breed. Genet.* 130, 252–258. doi:10.1111/jbg.12025.
- Oikawa, T., Sato, K., 1997. Treating small herds as fixed or random in an animal model. *J. Anim. Breed. Genet.* 114, 177–183. doi:10.1111/j.1439-0388.1997.tb00503.x.
- Phocas, F., Laloë, D., 2004. Should genetic groups be fitted in BLUP evaluation? Practical answer for the French AI beef sire evaluation. *Genet. Sel. Evol.* 36, 325–345. doi:10.1186/1297-9686-36-3-325.
- Quaas, R.L., 1988. Additive Genetic Model with Groups and Relationships. *J. Dairy Sci.* 71, 1338–1345. doi:10.1016/S0022-0302(88)79986-5.
- Quaas, R.L., Pollak, E.J., 1981. Modified Equations for Sire Models with Groups. *J. Dairy Sci.* 64, 1868–1872. doi:10.3168/jds.S0022-0302(81)82778-6.
- Smith, S.P., Graser, H.U., 1986. Estimating variance components in a class of mixed models by restricted maximum likelihood. *J. Dairy Sci.* 69, 1156–1165. doi:10.3168/jds.S0022-0302(86)80516-1.
- Strabel, T., Szwaczkowski, T., 1999. The use of test day models with small size of contemporary groups. *J. Anim. Breed. Genet.* 116, 379–386.
- Strabel, T., Szyda, J., Ptak, E., Jamrozik, J., 2005. Comparison of random regression test-day models for Polish Black and White cattle. *J. Dairy Sci.* 88, 3688–3699. doi:10.3168/jds.S0022-0302(05)73055-1.
- Theron, H.E., Kanfer, F.H.J., Rautenbach, L., 2002. The effect of phantom parent groups on genetic trend estimation. *S. Afr. J. Anim. Sci.* 32, 130–135.
- Tsuruta, S., Misztal, I., Lourenco, D.A.L., Lawlor, T.J., 2014. Assigning unknown parent groups to reduce bias in genomic evaluations of final score in US Holsteins. *J. Dairy Sci.* 97, 5814–5821. doi:10.3168/jds.2013-7821.
- Ugarte, E., Alenda, R., Carabaño, M.J., 1992. Fixed or random contemporary groups in genetic evaluations. *J. Dairy Sci.* 75, 269–278. doi:10.3168/jds.S0022-0302(92)77762-5.
- Van Vleck, L.D., 1987. Contemporary Groups for Genetic Evaluations. *J. Dairy Sci.* 70, 2456–2464.
- Vasconcelos, J., Martins, A., Ferreira, A., Carvalheira, J., 2005. Effects of the exclusion of small herds from the genetic evaluations of dairy cattle in Portugal. *Rev. Port. Zootec.* 2, 105–117.

- Vasconcelos, J., Santos, F., Bagnato, A., Carvalheira, J., 2008. Effects of clustering herds with small-sized contemporary groups in dairy cattle genetic evaluations. *J. Dairy Sci.* 91, 377–384. doi:10.3168/jds.2007-0202.
- Visscher, P.M., Goddard, M.E., 1993. Fixed and random contemporary groups. *J Dairy Sci.* 76, 1444–1454. doi:10.3168/jds.S0022-0302(93)77475-5.
- Westell, R.A., Quaas, R.L., Van Vleck, L.D., 1988. Genetic groups in an animal model. *J. Dairy Sci.* 71, 1310–1318.

## 1.8. Supplementary material

**Table S1.** Spearman correlation estimates between breeding values predicted for milk, fat and protein yields and somatic cell score (SCS) from evaluations with contemporary groups as a fixed (HTDF) or random (HTDR) effect and with (w/) or without (w/o) unknown parent groups (UPG).

Traits <sup>1</sup>	Models	HTDF w/ UPG	HTDF w/o UPG	HTDR w/ UPG	HTDR w/o UPG
Milk	HTDF w/ UPG	-	0.96	0.90	0.91
	HTDF w/o UPG	0.95	-	0.77	0.79
	HTDR w/ UPG	0.94	0.79	-	0.99
	HTDR w/o UPG	0.96	0.88	0.97	-
Fat	HTDF w/ UPG	-	0.97	0.90	0.92
	HTDF w/o UPG	0.97	-	0.79	0.82
	HTDR w/ UPG	0.92	0.81	-	0.99
	HTDR w/o UPG	0.96	0.89	0.97	-
Protein	HTDF w/ UPG	-	0.95	0.89	0.91
	HTDF w/o UPG	0.93	-	0.73	0.76
	HTDR w/ UPG	0.93	0.77	-	0.99
	HTDR w/o UPG	0.96	0.86	0.97	-
SCS	HTDF w/ UPG	-	0.99	0.82	0.84
	HTDF w/o UPG	0.99	-	0.82	0.84
	HTDR w/ UPG	0.80	0.85	-	0.99
	HTDR w/o UPG	0.86	0.90	0.99	-

<sup>1</sup> Bulls with more than 10 daughters (above the diagonal) and cows with records (below the diagonal).

## CHAPTER 2

### **Autoregressive and random regression test-day models for multiple lactations in genetic evaluation of Brazilian Holstein cattle**

Delvan Alves da Silva<sup>1</sup>, Claudio Napolis Costa<sup>2</sup>, Alessandra Alves da Silva<sup>1</sup>, Hugo Teixeira Silva<sup>1</sup>, Paulo Savio Lopes<sup>1</sup>, Fabyano Fonseca e Silva<sup>1</sup>, Renata Veroneze<sup>1</sup>, Gertrude Thompson<sup>3,4</sup>, Ignacio Aguilar<sup>5</sup>, Julio Carvalheira<sup>3,4</sup>

<sup>1</sup> Department of Animal Science, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil

<sup>2</sup> Embrapa Dairy Cattle, Juiz de Fora, Minas Gerais, Brazil

<sup>3</sup> Research Center in Biodiversity and Genetic Resources (CIBIO-InBio), University of Porto, Vairao, Porto, Portugal

<sup>4</sup> Institute of Biomedical Sciences Abel Salazar (ICBAS), University of Porto, Porto, Portugal.

<sup>5</sup> Instituto Nacional de Investigacion Agropecuaria, Montevideo, Uruguay

## 2.1. Abstract

Test-day records from the first three lactations of Brazilian Holstein cattle were fitted to Autoregressive (AR) and Random regression (RR) models with the objective to compare their efficiency in national genetic evaluations. The data comprised 4,142,740 records of milk yield (MY) and somatic cell score (SCS), registered between 1994 and 2016 from 2,322 herds. Although heritabilities were similar between models and traits, MY and SCS additive genetic variances estimates using AR models were, respectively, 7.0 and 22.2% higher than those obtained from RR. Residual variances on the other hand, were lower in both traits when estimated via AR. The rank correlation between estimated breeding values (EBV) obtained from AR and RR models were 0.96 and 0.94 for MY and 0.97 and 0.95 for SCS, respectively, for bulls with 10 or more daughters, and cows. Annual genetic gains for bulls (cows) obtained using AR model were 46.11 (49.50) kg for MY and -0.019 (-0.025) score for SCS. By using RR models these gain were 47.70 (55.56) kg for MY and -0.022 (-0.028) score for SCS. AR Akaike information criterion values were lower than those from RR models for both traits. In general, AR models were more efficient and, given the lower number of parameters to estimate and its suitability to fit data from small herds, these models should be chosen for genetic evaluations of the Holstein cattle in Brazil.

**Keywords:** autoregression, random regression, dairy cattle, Legendre polynomials

## 2.2. Introduction

The most commonly used test-day (**TD**) model for the analysis of multiple lactations is probably the multi-trait random regression (**RR**) model (Schaeffer, 2004). Several members of Interbull, such as Canada (Schaeffer et al., 2000), Germany,

Austria, Luxembourg (Liu et al., 2000), Italy (Muir et al., 2007), Netherlands (Roos et al., 2001), New Zealand (Harris et al., 2006) and others have been implemented RR models in their national dairy cattle genetic evaluations (Interbull, 2018).

This model allow us to adjust environmental effects from TD records and, to model the lactation curve shape leading to more accurate genetic evaluations than using the cumulative production in 305 days of lactation (Schaeffer et al., 2000). RR models are commonly fitted by using Legendre polynomials (LP) as sub-models, which implies in a high number of parameters to be estimated, as demonstrated by Schaeffer (2004). When few records per animal are available or even in populations with a high frequency of small herds, parameter estimation and artifacts issues may appear, especially at the lactation extremes (Meyer, 2005; Misztal, Strabel, Jamrozik, Mäntysaari, & Meuwissen, 2000; Nobre et al., 2003). To minimize these effects, a minimum of five TD records per animal has been adopted in several studies (Bohmanova et al., 2008; Konstantinov et al., 2015; Oliveira et al., 2017). Nevertheless, this strategy implies a loss of records due to editing the data.

In Portugal, which is also an Interbull member, the autoregressive test-day (**AR**) model for multiple lactations proposed by Carvalheira, Pollak, Quaas, & Blake (2002) has been routinely used for national genetic evaluations of dairy cattle. The AR model assumes additive genetic correlation of unity and partitions the animal's permanent environment into two terms (long - between lactations and short - within lactations effects) fitted with first-order autoregressive (co)variances structures. Thus, it takes into account the non-genetic correlations due to the repeated performance of cows. In this sense, AR explains random correlated effects of repeated TD records with fewer parameters compared to multi-trait or random regression models (Carvalheira et al., 2002; Sawalha et al., 2005a; b; Silva et al., 2019).

Brazil is not a member of Interbull, but it is attempting to become one (Silva et al., 2019). The country strategy envisages the change from traditional lactation models to TD models for multiple lactations in its national genetic evaluation of Holstein cattle. Although both RR and AR models present remarkable advantages (Bohmanova et al., 2008; Carvalheira et al., 2002; Costa, Carvalheira, Cobuci, Freitas, & Thompson, 2009; Costa et al., 2008; Muir et al., 2007; Oliveira et al., 2017; Vasconcelos et al., 2004), there are no reports that compare them simultaneously under a national genetic evaluation viewpoint. Therefore, we aimed to compare AR and RR models for multiple lactations test-day records of milk yield and somatic cell score in the Brazilian Holstein cattle.

### **2.3. Material and methods**

#### **Data**

The data used in this study were provided by the Brazilian Holstein Cattle Breeders Association. A total of 4,142,740 TD records of milk yield (**MY**) and somatic cell score (**SCS**) from the first three lactations of Brazilian Holstein cattle recorded between 1994 and 2016 from 2,322 herds, comprised the complete data set (**D16**). A reduced data set (**D11**) for both traits, with records truncated from 2011, were also used to evaluate prediction bias for both models.

Improper identification or herd codes, herds with less than three recorded cows, lactations with less than five TD records, as well as animals with unknown birth or calving dates and without first lactation records, were removed from data set. TD records varied from 5 to 305 days in milk (**DIM**), with time intervals between consecutive TD of approximately 30 days. Only records with age at calving between 18 and 45, 28 and 58, 38 and 78 months (**mo**) for the first, second and third lactations,

respectively, were used in this study. Age at calving was then grouped into 16 classes: class 1, < 20 mo; classes 2 to 15 were formed for every 4 mo interval from 20 to 75 mo; and class 16 for cows over 75 mo old at calving. DIM was grouped into 30 classes with 10 days intervals. TD records with more or less than three phenotypic standard deviations within DIM, age class and parity, were deleted. The number of records and descriptive statistics for MY and SCS in each data set (D16 or D11) are shown in Table 1.

**Table 1.** Description of the complete (D16) and reduced (D11) data sets

Item <sup>†</sup>	MY		SCS <sup>‡</sup>	
	D16	D11	D16	D11
TD records	2,976,510	2,019,062	2,330,142	1,507,970
TD records, lactation 1	1,518,134	1,016,633	1,192,638	757,105
TD records, lactation 2	920,453	625,562	714,653	465,497
TD records, lactation 3	537,923	376,867	422,851	285,368
Cows, lactation 1	177,140	116,607	142,750	88,856
Cows, lactation 2	106,355	71,114	83,137	52,946
Cows, lactation 3	62,566	43,149	49,204	32,495
TD records per cow	16.8	17.31	16.32	16.97
TD records per cow, lactation 1	8.57	8.71	8.35	8.52
TD records per cow, lactation 2	8.65	8.79	8.60	8.79
TD records per cow, lactation 3	8.60	8.73	8.59	8.78
Average (SD), all lactations	28.27 (8.28)	27.33 (8.11)	3.01 (2.13)	3.17 (2.11)
Average (SD), lactation 1	26.26 (6.99)	25.27 (6.77)	2.64 (2.00)	2.79 (2.00)
Average (SD), lactation 2	29.95 (8.67)	28.96 (8.48)	3.19 (2.16)	3.37 (2.12)
Average (SD), lactation 3	31.04 (9.41)	30.19 (9.23)	3.73 (2.22)	3.86 (3.15)

<sup>†</sup>TD: test-day; SD: standard deviation; Averages in kg for milk yield (MY) and score for somatic cell score (SCS). <sup>‡</sup>SCS =  $\log_2(\text{SCC} / 100) + 3$  and SCC: somatic cell count.

### Statistical models

The genetic evaluation for multiple lactations was performed by using AR (Carvalho, Blake, Pollak, Quaas, & Duran-Castro, 1998; Carvalho et al., 2002) and RR (4<sup>th</sup> order LP) test-day models (Schaeffer et al., 2000; Costa et al., 2008).

Autoregressive test-day model (AR)

In matrix notation, the AR model may be described as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{H}\mathbf{c} + \mathbf{M}\mathbf{p} + \mathbf{Q}\mathbf{t} + \mathbf{e},$$

wherein,  $\mathbf{y}$  is the vector of TD records;  $\boldsymbol{\beta}$  is the vector of fixed effects (age at calving class nested within herd, and days in milk class nested within herd and lactation);  $\mathbf{a}$  is the vector of random additive genetic effects;  $\mathbf{c}$  is the vector of random contemporary group (herd-test-day, **HTD**) effects;  $\mathbf{p}$  is the random long-term environmental effects (**LTE**);  $\mathbf{t}$  is the random short-term environmental effects (**STE**); and  $\mathbf{e}$  is the vector of random residual effects fitted with heterogeneous covariance structure across lactations. The  $\mathbf{X}$ ,  $\mathbf{Z}$ ,  $\mathbf{H}$ ,  $\mathbf{M}$  and  $\mathbf{Q}$  are incidence matrices relating observations to fixed and random effects. A first order autoregressive covariance structure was assumed for HTD (within herds), LTE (between lactations) and STE (between TD, within lactations) effects.

For AR model, in which  $L = 1, 2$  or  $3$  correspond to the first, second and third lactations, respectively, we have:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{a} \\ \mathbf{c} \\ \mathbf{p} \\ \mathbf{t} \\ \mathbf{e} \end{bmatrix} \sim N \left( \begin{bmatrix} \mathbf{X}\boldsymbol{\beta} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \begin{bmatrix} \mathbf{V} & \mathbf{ZG} & \mathbf{HW} & \mathbf{MJ} & \mathbf{QS} & \mathbf{R} \\ & \mathbf{G} & \emptyset & \emptyset & \emptyset & \emptyset \\ & & \mathbf{W} & \emptyset & \emptyset & \emptyset \\ & & & \mathbf{J} & \emptyset & \emptyset \\ & & & & \mathbf{S} & \emptyset \\ & \text{Sim.} & & & & \mathbf{R} \end{bmatrix} \right)$$

where,

the phenotypic (co)variance is  $\mathbf{V} = \mathbf{ZGZ}' + \mathbf{HWH}' + \mathbf{MJM}' + \sum_{L=1}^3 (\mathbf{Q}_L \mathbf{S}_L \mathbf{Q}'_L) + \mathbf{R}$ ;

with  $\mathbf{G} = \mathbf{A}\sigma_a^2$ , where  $\mathbf{A}$  represents the pedigree-based additive genetic relationship matrix and  $\sigma_a^2$  the additive genetic variance.  $\mathbf{R}$  is the residual variance matrix and the covariance matrices  $\mathbf{W}$ ,  $\mathbf{J}$  and  $\mathbf{S}$  are based on first order autoregressive structures for

HTD, LTE, STE, respectively, as follow (for simplicity, let's assume that each herd has three HTD levels and cows have three TD records in each lactation):

$$\mathbf{W} = \sigma_c^2 \begin{bmatrix} 1 & \rho_c & \rho_c^2 \\ & 1 & \rho_c \\ & & 1 \end{bmatrix} \otimes \mathbf{I}_q;$$

$$\mathbf{J} = \sigma_p^2 \begin{bmatrix} 1 & \rho_p & \rho_p^2 \\ & 1 & \rho_p \\ & & 1 \end{bmatrix} \otimes \mathbf{I}_m;$$

$$\mathbf{S} = \sigma_{tL}^2 \begin{bmatrix} 1 & \rho_{tL} & \rho_{tL}^2 \\ & 1 & \rho_{tL} \\ & & 1 \end{bmatrix} \otimes \mathbf{I}_{mL};$$

$$\mathbf{R} = \mathbf{I}_{nL} \sigma_{eL}^2,$$

where,  $\mathbf{I}$  is the identity matrix;  $q$  is the number of herds;  $m$  is the number of cows with records;  $mL$  is the number of cows with records within the  $L^{\text{th}}$  lactation;  $nL$  is the number of records in each lactation;  $\sigma_c^2$  is HTD variance component,  $\rho_c$  is HTD autocorrelation coefficient,  $\sigma_p^2$  is LTE variance component,  $\rho_p$  is LTE autocorrelation coefficient,  $\sigma_{tL}^2$  is STE variance component,  $\rho_{tL}$  is STE autocorrelation coefficient and  $\sigma_{eL}^2$  are the residual variance component for the  $L^{\text{th}}$  lactation. More details about AR model can be found in (Carvalho et al., 1998; Carvalho et al., 2002; Silva et al., 2019).

For the estimation of variance components and autocorrelation coefficients, five random samples of the data were created for each trait. Each sample consisted of data from 40 (25) herds for MY (SCS) representing all Brazilian regions with records. These samples were used to obtain estimates of variance components and autocorrelations coefficients by using DFREML methodology (Smith and Graser, 1986). Likelihood functions were maximized by the multivariate simplex algorithm (Nelder and Mead, 1965). The convergence criterion was defined as  $10^{-8}$ . The local maxima was checked by running five consecutive cold starts without significant

changes in the log-likelihood (up to four decimals places). This process was repeated for each data sub-set. Reliabilities (**REL**) for the predicted breeding values (**EBV**), were computed as  $\left[REL = 1 - \frac{PEV}{\sigma_a^2}\right]$ , where PEV is the prediction error variance obtained from the diagonal of the inverse of the coefficient matrix.

#### Random regression test-day model (RR)

The fixed and random regressions were fitted using a 4<sup>th</sup> order LP (five covariables). In matrix notation, this model can be described as follows:

$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{H}\mathbf{c} + \mathbf{M}\mathbf{pe} + \mathbf{e}$ , so that:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \emptyset & \emptyset \\ \emptyset & \mathbf{X}_2 & \emptyset \\ \emptyset & \emptyset & \mathbf{X}_3 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \boldsymbol{\beta}_3 \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_1 & \emptyset & \emptyset \\ \emptyset & \mathbf{Z}_2 & \emptyset \\ \emptyset & \emptyset & \mathbf{Z}_3 \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{bmatrix} + \begin{bmatrix} \mathbf{H}_1 & \emptyset & \emptyset \\ \emptyset & \mathbf{H}_2 & \emptyset \\ \emptyset & \emptyset & \mathbf{H}_3 \end{bmatrix} \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \mathbf{c}_3 \end{bmatrix} + \begin{bmatrix} \mathbf{M}_1 & \emptyset & \emptyset \\ \emptyset & \mathbf{M}_2 & \emptyset \\ \emptyset & \emptyset & \mathbf{M}_3 \end{bmatrix} \begin{bmatrix} \mathbf{pe}_1 \\ \mathbf{pe}_2 \\ \mathbf{pe}_3 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \mathbf{e}_3 \end{bmatrix},$$

where,  $\mathbf{y}$  is the vector of TD records, represented in each lactation by  $\mathbf{y}_i$ ;  $\boldsymbol{\beta}$  is the vector of fixed effects (age class at calving nested within herd, and fixed regression coefficients for days in milk nested within herd), represented in each lactation by  $\boldsymbol{\beta}_i$ ;  $\mathbf{a}$  is the vector of random regression coefficients for additive genetic effect, represented in each lactation by  $\mathbf{a}_i$ ;  $\mathbf{c}$  is the vector of random HTD effects, represented in each lactation by  $\mathbf{c}_i$ ;  $\mathbf{pe}$  is the vector of random regression coefficients for permanent environmental effect, represented in each lactation by  $\mathbf{pe}_i$ ; and  $\mathbf{e}$  is the vector of random residual effects, represented in each lactation by  $\mathbf{e}_i$ . It was assumed homogeneity of variances within each lactation. The  $\mathbf{X}_i$ ,  $\mathbf{Z}_i$ ,  $\mathbf{H}_i$ , and  $\mathbf{M}_i$  are incidence matrices relating observations to fixed and random effects in lactation  $i$ , in which  $i = 1, 2$ , and  $3$ , corresponding to first, second and third lactations, respectively. We made the following assumptions:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{a} \\ \mathbf{c} \\ \mathbf{pe} \\ \mathbf{e} \end{bmatrix} \sim \text{NMV} \left( \begin{bmatrix} \mathbf{X}\boldsymbol{\beta} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \begin{bmatrix} \mathbf{V} & \mathbf{ZG} & \mathbf{HW} & \mathbf{MP} & \mathbf{R} \\ & \mathbf{G} & \emptyset & \emptyset & \emptyset \\ & & \mathbf{W} & \emptyset & \emptyset \\ & & & \mathbf{P} & \emptyset \\ \text{Sim.} & & & & \mathbf{R} \end{bmatrix} \right),$$

where:  $\mathbf{V} = \mathbf{ZGZ}' + \mathbf{HWH}' + \mathbf{MPM}' + \mathbf{R}$ ;  $\mathbf{G} = \mathbf{G}_0 \otimes \mathbf{A}$ ,  $\mathbf{G}_0$  is the 15x15 covariance matrix of additive genetic regression coefficients;  $\mathbf{W} = \mathbf{W}_0 \otimes \mathbf{I}_{nc}$ , where  $\mathbf{W}_0$  is a diagonal matrix containing the variances of HTD and  $\mathbf{I}_{nc}$  is the identity matrix, in which  $nc$  is the number of HTD;  $\mathbf{P} = \mathbf{P}_0 \otimes \mathbf{I}_n$ , wherein  $\mathbf{P}_0$  is the 15x15 covariance matrix of the permanent environmental regression coefficients and  $\mathbf{I}_n$  is the identity matrix, in which  $n$  is the number of animals with records;  $\mathbf{R} = \mathbf{R}_0 \otimes \mathbf{I}_N$ , where  $\mathbf{R}_0$  is a diagonal matrix of residual variances,  $\mathbf{I}_N$  is the identity matrix, in which  $N$  is the total number of animals. The HTD and residual covariances between traits were assumed as zero.

The vector of estimated breeding values ( $\mathbf{EBV}_{ij}$ ) for each lactation of each animal  $i$  at test-day  $j$  was obtained as:

$$\mathbf{EBV}_{ij} = \mathbf{K}\mathbf{a}_i,$$

where,  $\mathbf{a}_i$  is the vector of estimated breeding values for the regression coefficients of each animal  $i$ , and  $\mathbf{K}$  is the 301x5 matrix of LP (ranged 5-305 DIM). The EBV for accumulated 5 to 305 DIM were obtained as the sum of all EBVs per animal in each lactation.

The same samples used for AR model were also adopted to estimate the variance components for the RR using the AIREMLF90 software (Misztal et al., 2015). The convergence criterion was defined as  $10^{-8}$ . The covariances matrices for additive genetic ( $\mathbf{G}_1$ ) and permanent environmental ( $\mathbf{P}_1$ ) effects between and within lactations were estimated between DIM as follows:

$$\mathbf{G}_1 = \mathbf{K}_1\mathbf{G}\mathbf{K}'_1,$$

$$P_1 = K_1 P K_1',$$

where,  $G$  and  $P$  were previously defined; and  $K_1 = I_3 \otimes K$ .

The approximate EBV prediction error variances (PEVs) per animal using the RR model were estimated according to Tier & Meyer (2004). The REL by animal was obtained as  $(1 - K T_i K' / K G K')$ , where  $T_i$  is the approximate prediction error covariances for breeding values predicted to regression coefficients.

### Model comparisons

The AR and RR models were compared based on estimates of variance components, genetic parameters, REL, rank correlation, genetic trends, annual genetic gains and Akaike information criterion (AIC) obtained by using the complete data sets. Genetic trends were estimated by regressing the animal's EBV on birth year, after they were adjusted to the genetic base-year defined as the mean of EBVs of cows born in 2010.

AIC was defined as:  $AIC = -2\log(L) + 2p$ , where  $L$  is the Likelihood function, and  $p$  is the number of estimated parameters ( $p = 14$  or  $p = 256$  for the AR or the RR model, respectively).

The predictive ability of the AR or RR models was based on the methodology proposed by Mäntysaari, Liu, & VanRaden (2010). The comparisons were based in the results obtained by each model, where young bulls (208 for MY and 197 for SCS) had no daughter information in D11, but more than 19 daughters in the D16. To validate the genetic prediction, EBVs predicted in the D16 from the AR or RR models were regressed on the parent average (PA) predicted in D11 for both models. This regression model was as follows:

$$y_i = b_0 + b_1 x_i + e_i,$$

where  $y_i$  is the bull EBV from D16 (AR and RR),  $x_i$  is the bull PA from D11 (AR and RR),  $b_0$  and  $b_1$  are the intercept and regression coefficients,  $e_i$  is the residual error. The  $b_1$  was used as an indicator of bias in the genetic prediction with an expected value of one. The coefficient of determination ( $R^2$ ) was used to assess the validation reliability of the evaluation models. Confidence intervals (CI) were estimated for  $b_1$  and  $R^2$  using nonparametric bootstrap. The CI were calculated by using `boot` and `boot.ci` functions of the R packages with 10,000 bootstrap samples.

## 2.4. Results

### Variance components and genetic parameters estimates

Variance components and heritability estimates ( $h^2$ ) for AR and RR models for MY and SCS are shown in Table 2. For MY and SCS, the genetic variances ( $\sigma_a^2$ ) were 7.5 and 28.6% higher when estimated with AR models, respectively. For MY (SCS), the residual variances were 14.2 (12.2), 28.1 (15.9) and 28.5% (16.0%) lower for first, second and third lactation, when estimated with AR models, respectively. Similar trends were observed for the phenotypic variance estimates in both models, and between STE (AR) and permanent environmental (RR) variances. The  $h^2$  estimates were moderate to low and decreased from the first to the third lactation for both models, except for SCS by using RR model.

**Table 2.** Variance components and heritabilities with respective standard-errors (SE) for milk yield (MY) and somatic cell score (SCS) estimated by the autoregressive (AR) and random regression (RR) models

Parameters	AR <sup>†</sup>		RR <sup>‡</sup>	
	MY ± SE	SCS ± SE	MY ± SE	SCS ± SE
$\sigma_a^2$	8.89 ± 0.268	0.63 ± 0.009	8.27 ± 0.559	0.49 ± 0.015
$\sigma_{a1}^2$	-	-	7.28 ± 0.411	0.41 ± 0.010
$\sigma_{a2}^2$	-	-	9.13 ± 0.808	0.58 ± 0.024
$\sigma_{a3}^2$	-	-	9.60 ± 0.602	0.61 ± 0.023
$\sigma_p^{2§}$	≈0.00 ± ≈0.000	≈0.00 ± ≈0.000	-	-
$\sigma_{t1}^2$	18.47 ± 0.270	1.63 ± 0.022	21.35 ± 0.437	1.64 ± 0.022
$\sigma_{t2}^2$	22.39 ± 0.392	1.95 ± 0.017	24.41 ± 0.690	1.89 ± 0.020
$\sigma_{t3}^2$	29.57 ± 0.928	2.34 ± 0.027	31.59 ± 0.867	2.31 ± 0.025
$\sigma_c^2$	3.18 ± 0.101	0.18 ± 0.005	1.56 ± 0.053	0.09 ± 0.001
$\sigma_{c1}^2$	-	-	1.68 ± 0.054	0.12 ± 0.002
$\sigma_{c2}^2$	-	-	1.47 ± 0.056	0.07 ± 0.001
$\sigma_{c3}^2$	-	-	1.41 ± 0.075	0.05 ± 0.002
$\sigma_{e1}^2$	7.88 ± 0.209	1.22 ± 0.018	9.18 ± 0.114	1.39 ± 0.016
$\sigma_{e2}^2$	9.60 ± 0.236	1.22 ± 0.029	13.36 ± 0.250	1.45 ± 0.024
$\sigma_{e3}^2$	11.09 ± 0.246	1.26 ± 0.023	15.50 ± 0.239	1.50 ± 0.020
$\sigma_{p1}^2$	38.42 ± 0.579	3.66 ± 0.030	39.48 ± 0.513	3.56 ± 0.030
$\sigma_{p2}^2$	44.05 ± 0.801	3.98 ± 0.029	48.37 ± 0.891	3.99 ± 0.029
$\sigma_{p3}^2$	52.73 ± 1.210	4.41 ± 0.027	58.10 ± 1.335	4.47 ± 0.022
$h^2$	0.21 ± 0.003	0.16 ± 0.002	0.19 ± 0.012	0.13 ± 0.004
$h_1^2$	0.23 ± 0.004	0.17 ± 0.003	0.21 ± 0.012	0.11 ± 0.003
$h_2^2$	0.20 ± 0.003	0.16 ± 0.003	0.19 ± 0.015	0.14 ± 0.006
$h_3^2$	0.17 ± 0.003	0.14 ± 0.002	0.17 ± 0.008	0.14 ± 0.005

<sup>†</sup> $\sigma_a^2$ : additive genetic variance (kg<sup>2</sup> for MY and score units<sup>2</sup> for SCS);  $\sigma_p^2$ : long-term environmental variance;  $\sigma_{ti}^2$ : short-term environmental variance;  $\sigma_c^2$ : herd-test-day variance;  $\sigma_{ei}^2$ : residual variance;  $\sigma_{pi}^2$ : phenotypic variance;  $h^2$ : Average heritability weighted by the number of records in each sample;  $h_i^2$ : heritabilities; where  $i=1, 2$  and  $3$  correspond to first, second and third lactations, respectively. <sup>§</sup>values for  $\sigma_p^2 < 0.0001 \pm < 0.00001$ . <sup>‡</sup> $\sigma_a^2$ : Average additive genetic variance weighted by the number of records in each sample (kg<sup>2</sup> for MY and score units<sup>2</sup> for SCS);  $\sigma_{ai}^2$ : additive genetic variance;  $\sigma_{ti}^2$ : permanent environmental variance;  $\sigma_{ci}^2$ : Average herd-test-day variance weighted by the number of records in each sample;  $\sigma_{ei}^2$ : herd-test-day variance;  $h^2$ : Average heritability weighted by the number of records in each sample;  $h_i^2$ : heritabilities; where  $i=1, 2$  and  $3$  correspond to first, second and third lactations, respectively.

The autocorrelations for LTE, STE, and HTD obtained with AR model for MY and SCS are shown in Table 3. The additive genetic ( $r_g$ ) and permanent environmental

( $r_{pe}$ ) correlations obtained with RR models, for both traits, are shown in Table 4. The LTE autocorrelations were close to zero (AR) whereas the permanent environmental correlations among lactations using the RR model ranged from 0.10 to 0.30.

Genetic correlation among TD within or among lactations for AR are assumed as unity. On the other hand, RR genetic correlations between TD within lactation ranged from 0.82 to 0.92, whereas between lactations ranged from 0.68 to 0.86, for both traits. Lower genetic correlation values were observed between first and third lactation, for both traits (Table 4).

**Table 3.** Autocorrelation parameters and respective standard-errors (SE) for milk yield (MY) and somatic cell score (SCS) estimated by the autoregressive model (AR)

Parameters <sup>†</sup>	MY $\pm$ SE	SCS $\pm$ SE
$\rho_p^{\ddagger}$	$\approx 0.00 \pm \approx 0.000$	$\approx 0.00 \pm \approx 0.000$
$\rho_{t1}$	$0.68 \pm 0.007$	$0.85 \pm 0.003$
$\rho_{t2}$	$0.76 \pm 0.006$	$0.82 \pm 0.002$
$\rho_{t3}$	$0.76 \pm 0.002$	$0.84 \pm 0.001$
$\rho_c$	$0.68 \pm 0.011$	$0.61 \pm 0.005$

<sup>†</sup> $\rho_p$ ,  $\rho_{ti}$  and  $\rho_c$ : autocorrelations for LTE, STE and HTD effects; where  $i=1, 2$  and  $3$  correspond to first, second and third lactations, respectively. <sup>‡</sup>values for  $\hat{\rho}_p < 0.0001 \pm < 0.00001$ .

**Table 4.** Averages of additive genetic ( $r_g$ ) and permanent environment ( $r_{pe}$ ) correlations between test-days (TD) within and between lactations for milk yield (MY) and somatic cell score (SCS) estimated by the random regression model (RR)

TD of the lactation	MY		SCS	
	$r_g$ (SD)	$r_{pe}$ (SD)	$r_g$ (SD)	$r_{pe}$ (SD)
1 with 1	0.82(0.210)	0.60(0.370)	0.90(0.113)	0.75(0.209)
1 with 2	0.77(0.141)	0.26(0.149)	0.76(0.091)	0.15(0.063)
1 with 3	0.68(0.127)	0.15(0.069)	0.71(0.079)	0.10(0.048)
2 with 2	0.90(0.086)	0.70(0.273)	0.92(0.083)	0.73(0.232)
2 with 3	0.79(0.088)	0.30(0.108)	0.86(0.075)	0.28(0.074)
3 with 3	0.85(0.132)	0.69(0.281)	0.90(0.093)	0.76(0.204)

SD: Standard deviation.

## EBV reliability, rank correlation and genetic trends

The greatest increase in REL by using AR model was observed for bulls with daughters. For MY, bulls without daughters, had a mean (standard deviation - **SD**) REL of 0.27 (0.14) with the AR, whereas with the RR this estimate was 0.18 (0.11). For bulls with 1 to 9 daughters, the mean (SD) for AR was of 0.58 (0.14), compared with 0.35 (0.10) for RR. For bulls with 10 or more daughters, the mean (SD) for AR was 0.92 (0.06), compared with 0.73 (0.13) for RR. Similar results were observed for SCS.

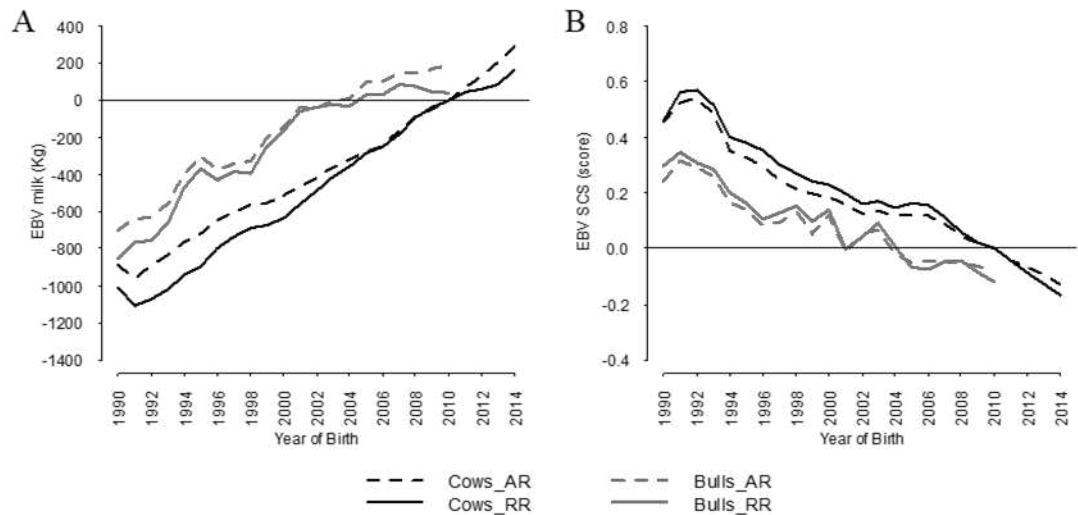
EBV rank correlation coefficients between models (AR vs RR) for bulls with 10 or more daughters in at least 5 herds, were 0.96 and 0.97 for MY and SCS, respectively. For cows with records, rank correlation coefficients were 0.94 and 0.95, respectively. There was a drop in rank correlation for the top 100 bulls and cows. For the top 100 bulls with 10 or more daughters, rank correlation estimates was 0.59 and 0.67 for MY and SCS, respectively. For the top 100 cows with records, the rank correlations were 0.33 and 0.54, respectively.

MY and SCS genetic trends for bulls with 10 or more daughters and cows with records, born between 1990 and 2014, are presented Figure 1. Both models showed a similar pattern for the genetic progress. The annual genetic gains average for the AR model, for bulls (cows) were of 46.11 (49.50) kg for MY and -0.019 (-0.025) score for SCS. The RR model (average over lactations) showed values of 47.70 (55.56) kg for MY and -0.022 (-0.028) score for SCS.

## Model comparisons

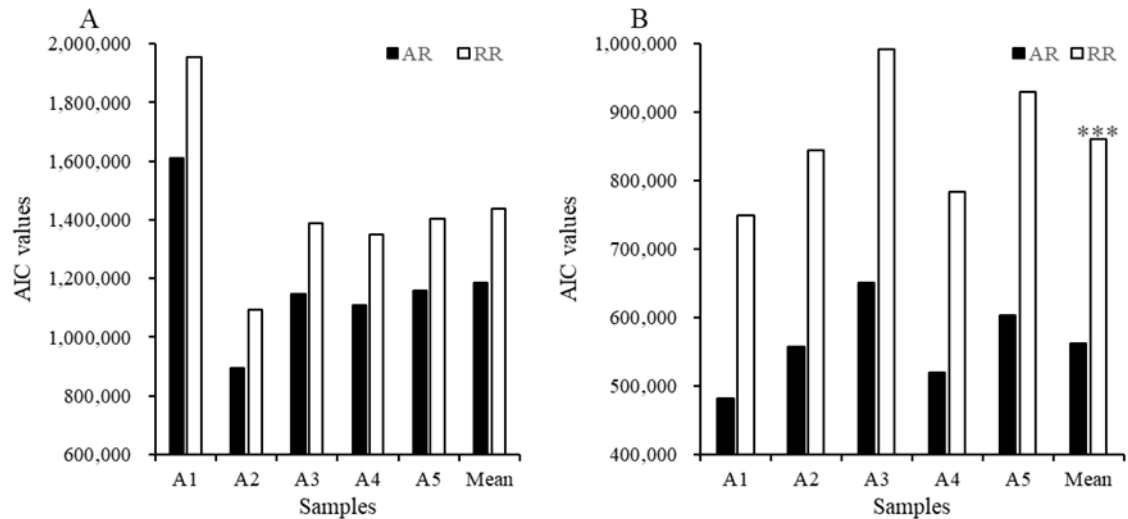
The AIC obtained with both models for the five samples used in the estimation of the variance components and their general mean are shown in Figure 2. In general,

AIC values were lower for the AR (21.5% lower for MY and 52.9% lower ( $P < 0.05$ ) for SCS).



**Figure 1.** Genetic trends for milk yield (A) and somatic cell score (SCS - B) for bulls and cows born between 1990 and 2014 from evaluations using autoregressive (AR) and random regression (RR) models (base year = 2010).

The  $b_1$  regression coefficients and  $R^2$  validation reliabilities with 95% bootstrap CI (10,000 bootstrap resampling) obtained using the model validation are shown in Table 5. The  $R^2$  difference between models were small. For both models, the  $b_1$  values for MY were within  $\pm 15\%$  of the optimal value ( $b_1 = 1$ ), whereas for SCS, the  $b_1$  values were outside of the acceptable range for both models (Tsuruta, Misztal, Aguilar, & Lawlor, 2011).



**Figure 2.** Akaike information criterion (AIC) values from five samples (A1-A5) used in the estimation of the variance components by the autoregressive (AR) and random regression (RR) models.

**Table 5.** Values of regression coefficients ( $b_1$ ), validation reliabilities ( $R^2$ ), and their 95% bootstrap confidence intervals (CI) from the validation test model for milk yield (MY) and somatic cell scores (SCS) from evaluations using autoregressive (AR) and random regression (RR) models

Models	$b_1$ (CI)	$R^2$ (CI)
MY (kg)		
AR	0.85 (0.58 - 1.09)	0.17 (0.09 - 0.27)
RR	0.87 (0.61 - 1.15)	0.17 (0.09 - 0.27)
SCS (score)		
AR	0.56 (0.35 - 0.76)	0.14 (0.05 - 0.23)
RR	0.63 (0.44 - 0.80)	0.20 (0.11 - 0.30)

## 2.5. Discussion

In our study, lactations with less than 5 TD records were eliminated and all cows had to have records in the first lactation as previously mentioned. Several studies have used these criteria to editing data considering the RR model (Bohmanova et al., 2008; Konstantinov et al., 2015; Oliveira et al., 2017). A lack of convergence can occur

because of numerical problems associated with a smaller number of TD per cow, per lactation in the RR model when using LP (Misztal, 2006; Robbins, Misztal, & Bertrand, 2005). In addition, the RR models fitted with LP with few records per cow may also be a problem for estimations at the extremes of the lactation curve (Meyer, 2005; Misztal et al., 2000; Nobre et al., 2003). This problem may be reduced using autoregressive processes (Misztal et al., 2000).

On the other hand, other studies (Carvalho et al., 1998; Vasconcelos, Santos, Bagnato, & Carvalho, 2008; Silva et al., 2019) have shown that the AR model could be fitted using a minimum of 2 TD records per cow, per lactation. This may represent an advantage for the AR models in situations such in Brazil, in which the number of small herds with a relatively high frequency of cows having few records, may be important to avoid loss of information due to data editing.

The genetic correlation between TD may differ from unit meaning that the expression at each DIM may have different additive genetic variance. The RR allow modeling changes in variance along a continuous scale, which is a necessary condition to the assumption that the trait genetic control may change over time. On the other hand, the AR model as a repeatability model, assumes the same genetic variance for all TD within or between the lactations (Carvalho et al., 2002). In the present study, the changes in  $\sigma_a^2$  between lactations (RR model) for MY were relatively small and, on average, both models provided similar estimates and heritabilities (Table 2). However, for SCS, the RR average  $\sigma_a^2$  was relatively smaller than the estimate from the AR and is reflected in a smaller heritability. One reason for this occurrence may be due to the changes in the data structure (number of herds, animals and records) by considering the lactations as different traits in the RR model.

The long-term (LTE autocorrelation or  $r_{pe}$  between lactations) environmental correlations were low in both models. Most of the permanent environmental correlations were captured by the short-term (between TD within lactations – STE autocorrelations and  $r_{pe}$  within lactations) environmental effect (Table 3 and 4). These results are supported by previous studies also using the AR model (Costa, Cobuci, Santos, Thompson, & Carvalheira, 2014; Vasconcelos et al., 2004, 2008). Muir et al. (2007) reported lower  $r_{pe}$  values within lactation for RR model, ranging from 0.36 to 0.48 for MY and from 0.26 to 0.44 for SCS. The average genetic correlations between TD for both traits using the RR model were high and positive (ranging from 0.68 to 0.92) in all lactations (Table 4). Muir et al. (2007) reported similar genetic correlations for MY (range 0.70 - 0.86), but for SCS these values were smaller (range 0.43 - 0.52).

In the AR model approach, the TD records of all lactations are considered simultaneously, which leads to an increase in the number of records per bull-daughter. On the other hands, TD records were considered by lactation by using RR models. The higher the number of observations and relatives information of bulls in the relationship matrix, greater REL estimates are (Carvalheira et al., 2002; Schaeffer et al., 2000). According to Tier & Meyer (2004), the reliabilities of regression coefficients and the quality of their approximation using RR models depend upon the quantity and quality of the available data. These authors reported that when few observations per animals are available, lower REL estimation for higher order coefficients are observed and vice-versa. This is supported by our lower REL estimates by using RR model, in which a maximum of 10 TD per animal were available for REL estimation each day (ranging from 5 to 305 DIM), which may explain the lower reliabilities obtained using the RR model.

In general, no relevant EBV re-ranking were observed between AR and RR. Genetic trends and annual genetic gains were also similar for both traits and models (Figure 1). In absolute terms however, annual genetic gains from RR for both traits were higher than for AR in either bulls and cows (ranging from 3.5 to 15.8%).

The AR model best fit to the data for both traits due to its smaller AIC values (Figure 2). The AR model is more parsimonious considering the lower number of parameters to be estimated (14 vs. 256 for AR and RR, respectively). Both models provided unbiased estimations of the genetic trend for MY indicating a good predictive ability (Table 5). The opposite happened with SCS where the regression coefficient as a measure of bias was significantly less than one ( $P < 0.05$ ). Meuwissen & Pool, (2001), reported similar findings when comparing AR and RR models.

The performance of AR and RR models were compared by using MY and SCS TD records from the first three lactations of Brazilian Holstein cows. The RR model requires higher number of information per animal if compared with the AR model. The EBV reliabilities estimated with the AR model were higher on average than those estimated by the RR model. Although the models differ in their structures, similar genetic trends were observed, but the RR model provided the highest annual genetic gains. The AR model provided a better fitting (AIC values) for the data, but both models revealed similar results for predictive ability.

## **2.6. Conclusion**

Both models performed well and may be used for parameter estimation and genetic evaluations of milk yield and somatic cell score in multiple-lactation of the Holstein cattle in Brazil. In general, the AR models were more efficient and, given the lower number of parameters to estimate and its suitability to fit data from small herds

(especially relevant to avoid loss of information due to data editing), these models are more parsimonious and should be chosen for genetic evaluations of longitudinal traits of the Holstein cattle in Brazil.

## **2.7. Acknowledgements**

The authors acknowledge the Brazilian Holstein Cattle Breeders Association (ABCBRH) for providing data for this study. This study was partially financed by Coordination for the Improvement of Higher Education Personnel and Portuguese National Funding Agency for Science, Research and Technology (CAPES/FCT, n° 99999.008462/2014-03 and 88887.125450/2016-00), and National Council of Technological and Scientific Development (CNPq 465377/2014-9 - PROGRAMA INCT and CNPq 142467/2015-4).

## **2.8. References**

- Bohmanova, J., F. Miglior, J. Jamrozik, I. Misztal, and P.G. Sullivan. 2008. Comparison of random regression models with Legendre Polynomials and Linear Splines for production traits and somatic cell score of Canadian Holstein cows. *J. Dairy Sci.* 91:3627–3638. doi:10.3168/jds.2007-0945.
- Carvalho, J., E.J. Pollak, R.L. Quaas, and R.W. Blake. 2002. An autoregressive repeatability animal model for test-day records in multiple lactations. *J. Dairy Sci.* 85:2040–2045. doi:10.3168/jds.S0022-0302(02)74281-1.
- Carvalho, J.G., R.W. Blake, E.J. Pollak, R.L. Quaas, and C. V. Duran-Castro. 1998. Application of an autoregressive process to estimate genetic parameters and breeding values for daily milk yield in a tropical herd of Lucerna cattle and in United States Holstein herds. *J. Dairy Sci.* 81:2738–2751. doi:10.3168/jds.S0022-0302(98)75831-X.
- Costa, C.N., J. Carvalho, J.A. Cobuci, A.F. Freitas, and G. Thompson. 2009. Estimation of genetic parameters of test day fat and protein yields in Brazilian

- Holstein cattle using an autoregressive multiple lactation animal model. *South African J. Anim. Sci.* 39:165–168.
- Costa, C.N., J.A. Cobuci, G.G. Santos, G. Thompson, and J. Carvalheira. 2014. Genetic parameters of test day records in Brazilian Holstein cattle using an autoregressive multiple lactation animal model. Pages 5–7 in 10th World Congress of Genetics Applied to Livestock Production, Vancouver, Canadá.
- Costa, C.N., C.M.R. Melo, I.U. Packer, A.F. Freitas, N.M. Teixeira, and J.A. Cobuci. 2008. Genetic parameters for test day milk yield of first lactation Holstein cows estimated by random regression using Legendre polynomials. *Rev. Bras. Zootec.* 37:602–608. doi:10.1590/S1516-35982008000400003.
- Harris, B.L., A.M. Winkelman, D.L. Johnson, and W.A. Montgomerie. 2006. Development of a national production test-day model for New Zealand. *Interbull Bull.* 35:27–30.
- Interbull. 2018. National Genetic Evaluation Forms Provided by Countries. Accessed June 19, 2018. <http://www.interbull.org/ib/geforms>.
- Konstantinov, K. V, G.J. Nieuwhof, and T.P. Hancock. 2015. Implementation of multiple traits multi lactation random regression test day model for production traits in Australia. *Interbull Bull.* 49.
- Liu, Z., F. Reinhardt, and R. Reents. 2000. Estimating parameters of a random regression test day model for first three lactation milk production traits using the covariance function approach. *Interbull Bull.* 25:74–80.
- Mäntysaari, E.A., Z. Liu, and P. VanRaden. 2010. Interbull validation test for genomic evaluations. *Interbull Bull.* 17–22.
- Meuwissen, T.H.E., and M.H. Pool. 2001. Autoregressive versus random regression test-day models for prediction of milk yields 7:172–178.
- Meyer, K. 2005. Estimates of genetic covariance functions for growth of Angus cattle. *J. Anim. Breed. Genet.* 122:73–85. doi:10.1111/j.1439-0388.2005.00503.x.
- Misztal, I. 2006. Properties of random regression models using linear splines. *J. Anim. Breed. Genet.* 123:74–80. doi:10.1111/j.1439-0388.2006.00582.x.
- Misztal, I., T. Strabel, J. Jamrozik, E.A. Mäntysaari, and T.H.E. Meuwissen. 2000. Strategies for estimating the parameters needed for different test-day models. *J. Dairy Sci.* 83:1125–1134. doi:10.3168/jds.S0022-0302(00)74978-2.

- Misztal, I., S. Tsuruta, D. Lourenco, I. Aguilar, A. Legarra, and Z. Vitezica. 2015. Manual for BLUPF90 Family of Programs. Accessed October 10, 2017. [http://nce.ads.uga.edu/wiki/doku.php?id=application\\_programs](http://nce.ads.uga.edu/wiki/doku.php?id=application_programs).
- Muir, B.L., G. Kistemaker, J. Jamrozik, and F. Canavesi. 2007. Genetic parameters for a multiple-trait multiple-lactation random regression test-day model in Italian Holsteins. *J. Dairy Sci.* 90:1564–1574. doi:10.3168/jds.S0022-0302(07)71642-9.
- Nelder, J.A., and R. Mead. 1965. A Simplex method for function minimization. *Comput. J.* 7:308–313. doi:10.1093/comjnl/7.4.308.
- Nobre, P.R.C., I. Misztal, S. Tsuruta, J.K. Bertrand, L.O.C. Silva, and P.S. Lopes. 2003. Analyses of growth curves of Nellore cattle by multiple-trait and random regression models. *J. Anim. Sci.* 81:918–926. doi:10.2527/2003.814927x.
- Oliveira, H.R., F.F. Silva, M.V.G.B. Silva, O.H.G.B.D. Siqueira, M.A. Machado, J.C.C. Panetto, L.S. Glória, and L.F. Brito. 2017. Bayesian models combining Legendre and B-spline polynomials for genetic analysis of multiple lactations in Gyr cattle. *Livest. Sci.* 201:78–84. doi:10.1016/j.livsci.2017.05.007.
- Robbins, K.R., I. Misztal, and J.K. Bertrand. 2005. A practical longitudinal model for evaluating growth in Gelbvieh cattle. *J. Anim. Sci.* 83:29–33. doi:10.2527/2005.83129x.
- Roos, A.P.W., A.G.F. Harbers, and G. Jong. 2001. Random regression test-day model in The Netherlands. *Interbull Bull.* 27:155 – 158.
- Sawalha, R.M., J.F. Keown, S.D. Kachman, and L.D. Van Vleck. 2005a. Evaluation of autoregressive covariance structures for test-day records of Holstein cows: estimates of parameters. *J. Dairy Sci.* 88:2632–2642. doi:10.3168/jds.S0022-0302(05)72940-4.
- Sawalha, R.M., J.F. Keown, S.D. Kachman, and L.D. Van Vleck. 2005b. Genetic evaluation of dairy cattle with test-day models with autoregressive covariance structures and with a 305-d model. *J. Dairy Sci.* 88:3346–3353. doi:10.3168/jds.S0022-0302(05)73018-6.
- Schaeffer, L.R. 2004. Application of random regression models in animal breeding. *Livest. Prod. Sci.* 86:35–45. doi:10.1016/S0301-6226(03)00151-9.
- Schaeffer, L.R., J. Jamrozik, G.J. Kistemaker, and B.J. Van Doormaal. 2000. Experience with a test-day model. *J. Dairy Sci.* 83:1135–1144. doi:10.3168/jds.S0022-0302(00)74979-4.

- Silva, D.A., C.N. Costa, A.A. Silva, F.F. Silva, P.S. Lopes, G.G. dos Santos, G. Thompson, and J. Carvalheira. 2019. Unknown parent and contemporary groups for genetic evaluation of Brazilian Holstein using autoregressive test-day models. *Livest. Sci.* 220:1–7. doi:10.1016/j.livsci.2018.11.022.
- Smith, S.P., and H.U. Graser. 1986. Estimating variance components in a class of mixed models by restricted maximum likelihood. *J. Dairy Sci.* 69:1156–1165. doi:10.3168/jds.S0022-0302(86)80516-1.
- Tier, B.B., and K. Meyer. 2004. Approximating prediction error covariances among additive genetic effects within animals in multiple-trait and random regression models. *J. Anim. Breed. Genet.* 121:77–89.
- Tsuruta, S., I. Misztal, I. Aguilar, and T.J. Lawlor. 2011. Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. *J. Dairy Sci.* 94:4198–4204. doi:10.3168/jds.2011-4256.
- Vasconcelos, J., A. Martins, M.F. Petim-Batista, J. Colaço, R.W. Blake, and J. Carvalheira. 2004. Prediction of daily and lactation yields of milk, fat, and protein using an autoregressive repeatability test day model. *J. Dairy Sci.* 87:2591–8. doi:10.3168/jds.S0022-0302(04)73384-6.
- Vasconcelos, J., F. Santos, A. Bagnato, and J. Carvalheira. 2008. Effects of clustering herds with small-sized contemporary groups in dairy cattle genetic evaluations. *J. Dairy Sci.* 91:377–384. doi:10.3168/jds.2007-0202.

## 2.9. Supplementary material

Programming codes in R to access EBV by animal using Random Regression model

```
# EBV by animal - Random Regression
#
# 18-04-2018
# Delvan Silva
# -----
#Pakages
library(lattice)

#setwd("C:\\Analises Doutorado Portugal_CAPES\\Analise de
dados\\Analises_Br\\Avaliacao\\random
regression\\Leite\\prediction")
#
=====
#Polinomios de Legendre
x=seq(5,305) #especificando meses de 1 a 10
x1=as.matrix((-1+2*(x-min(x))/(max(x)-min(x)))
M=cbind(as.matrix(rep(1,nrow(x1))),x1,x1^2,x1^3,x1^4)
library(orthopolynom)
n=4
b=legendre.polynomials(n, normalized=T)
gama0=polynomial.coefficients(b)
aux=rbind(gama0[[1]],gama0[[2]],gama0[[3]],gama0[[4]],gama0[[5]])
aux[upper.tri(aux)] <- 0
gamat=aux
gama=t(gamat)

K=M%*%gama #é a variável independente original transformada
dim(K)
#
=====
#####lendo coef geneticos aditivos#####

solution=read.table("solutions",skip=1)
head(solution)
dim(solution)

#EBV in lactation 1 by days in milk
solution1 = solution[solution[,1]==1,][,]
head(solution1)
dim(solution1)

#Polinomios de Legendre for lactation 1
coef0=solution1[solution1[,2]==8,][,3:4]
coef1=solution1[solution1[,2]==9,][,3:4]
coef2=solution1[solution1[,2]==10,][,3:4]
coef3=solution1[solution1[,2]==11,][,3:4]
coef4=solution1[solution1[,2]==12,][,3:4]

Nanimal=nrow(coef0)
Nanimal

VG_anil=NULL
for(i in 1:Nanimal)
{
```

```

VG_ani1[i] <- list((coef0[i,2]*K[,1])+(coef1[i,2]*K[,2])+
(coef2[i,2]*K[,3])+(coef3[i,2]*K[,4])+(coef4[i,2]*K[,5]))
}

ebv1_dim=data.frame(cbind( sort(rep(seq(1,Nanimal),301)),
rep(seq(5,305),Nanimal), unlist(VG_ani1)[1:(Nanimal*301)])) ##EBV
lactation 1
colnames(ebv1_dim) = c('id','dim','ebv1')
head(ebv1_dim)
dim(ebv1_dim)
class(ebv1_dim$ebv)

ebv1 = cbind(unique(ebv1_dim$id),
tapply(ebv1_dim$ebv1,ebv1_dim$id,FUN=sum))
colnames(ebv1) = c('id','ebv1_305')
head(ebv1)
dim(ebv1)

#EBV lact2
solution2 = solution[solution[,1]==2,][,]
head(solution2)
dim(solution2)

#Polinomios de Legendre for lactation 2
coef0=solution2[solution2[,2]==8,][,3:4]
coef1=solution2[solution2[,2]==9,][,3:4]
coef2=solution2[solution2[,2]==10,][,3:4]
coef3=solution2[solution2[,2]==11,][,3:4]
coef4=solution2[solution2[,2]==12,][,3:4]

Nanimal=nrow(coef0)
Nanimal

VG_ani2=NULL
for(i in 1:Nanimal)
{
VG_ani2[i] <- list((coef0[i,2]*K[,1])+(coef1[i,2]*K[,2])+
(coef2[i,2]*K[,3])+(coef3[i,2]*K[,4])+(coef4[i,2]*K[,5]))
}

ebv2_dim=data.frame(cbind( sort(rep(seq(1,Nanimal),301)),
rep(seq(5,305),Nanimal), unlist(VG_ani2)[1:(Nanimal*301)])) ##EBV
lactation 1
colnames(ebv2_dim) = c('id','dim','ebv2')
head(ebv2_dim)
dim(ebv2_dim)
class(ebv2_dim$ebv2)

ebv2 = cbind(unique(ebv2_dim$id),
tapply(ebv2_dim$ebv2,ebv2_dim$id,FUN=sum))
colnames(ebv2) = c('id','ebv2_305')
head(ebv2)
dim(ebv2)

#EBV lact3
solution3 = solution[solution[,1]==3,][,]
head(solution3)
dim(solution3)

#Polinomios de Legendre for lactation 3

```

```

coef0=solution3[solution3[,2]==8,][,3:4]
coef1=solution3[solution3[,2]==9,][,3:4]
coef2=solution3[solution3[,2]==10,][,3:4]
coef3=solution3[solution3[,2]==11,][,3:4]
coef4=solution3[solution3[,2]==12,][,3:4]

Nanimal=nrow(coef0)
Nanimal

VG_ani3=NULL
for(i in 1:Nanimal)
{
  VG_ani3[i] <- list((coef0[i,2]*K[,1])+(coef1[i,2]*K[,2])+
  (coef2[i,2]*K[,3])+(coef3[i,2]*K[,4])+(coef4[i,2]*K[,5]))
}

ebv3_dim=data.frame(cbind( sort(rep(seq(1,Nanimal),301)),
rep(seq(5,305),Nanimal), unlist(VG_ani3)[1:(Nanimal*301)])) ##EBV
lactation 1
colnames(ebv3_dim) = c('id','dim','ebv3')
head(ebv3_dim)
dim(ebv3_dim)
class(ebv3_dim$ebv3)

ebv3 = cbind(unique(ebv3_dim$id),
tapply(ebv3_dim$ebv3,ebv3_dim$id,FUN=sum))
colnames(ebv3) = c('id','ebv3_305')
head(ebv3)
dim(ebv3)

ebv_dim = cbind(ebv1_dim,ebv2_dim[,3],ebv3_dim[,3])
colnames(ebv_dim) = c('id','dim','ebv1','ebv2','ebv3')
head(ebv_dim)
dim(ebv_dim)

ebv_305 = cbind(ebv1,ebv2[,2],ebv3[,2])
colnames(ebv_305) = c('id','ebv1_305','ebv2_305','ebv3_305')
head(ebv_305)
tail(ebv_305)
dim(ebv_305)
class(ebv_305)

ebv_305 = as.data.frame(ebv_305)

#Voltar codigos originais
ped = read.table('./renum/renadd25.ped',sep=' ')
code_orig = ped[,c('V1','V10')]
colnames(code_orig) = c('id','id_orig')
head(code_orig)

idx = ebv_305$id %in% code_orig$id
table(idx)

ebv_305 = merge(ebv_305,code_orig, by='id')
ebv_305 = subset(ebv_305, select =
c('id','id_orig','ebv1_305','ebv2_305','ebv3_305'))
head(ebv_305)
dim(ebv_305)

#Exporta arquivos
write.table(ebv_305,file='./bv_animal_rr.txt',sep=' ',

```



```

b=legendre.polynomials(n, normalized=T)
gama0=polynomial.coefficients(b)
aux=rbind(gama0[[1]],gama0[[2]],gama0[[3]],gama0[[4]],gama0[[5]])
aux[upper.tri(aux)] <- 0
gamat=aux
gama=t(gamat)

K=M%*%gama #é a variável independente original transformada
dim(K)
#
=====
#Pedigree
ped = read.table('./renum/renadd25.ped',sep=' ')
code_orig = ped[,c('V1','V2','V3','V10')]
colnames(code_orig) = c('id','pai','mae','id_orig')
head(code_orig)
#
=====
#Variancias genéticas
vg = read.table('./GM.txt')
vg = as.matrix(vg)
Vg1=diag(K%*%(q1%*%vg%*%t(q1))%*%t(K))
Vg2=diag(K%*%(q2%*%vg%*%t(q2))%*%t(K))
Vg3=diag(K%*%(q3%*%vg%*%t(q3))%*%t(K))

#lendo arquivo pev_pec_bf90
pev = read.table('./pev_pec_accrr')
dim(pev)
pev = as.matrix(pev)

#PEV by animals
Nanimal=nrow(pev)
Nanimal

pev1=NULL
pev2=NULL
pev3=NULL
for(i in 1:Nanimal)
{
  pev1[i] <- list(s1 =
diag(K%*(q1%*(symMatrix(pev[i,2:121],15, byrow = T,upper =
T))%*%t(q1))%*%t(K))
  pev2[i] <- list(s1 =
diag(K%*(q2%*(symMatrix(pev[i,2:121],15, byrow = T,upper =
T))%*%t(q2))%*%t(K))
  pev3[i] <- list(s1 =
diag(K%*(q3%*(symMatrix(pev[i,2:121],15, byrow = T,upper =
T))%*%t(q3))%*%t(K))
}

#PEV para lactacao 1
pev_l1=data.frame(cbind( sort(rep(seq(1,Nanimal),301)),
rep(seq(5,305),Nanimal), unlist(pev1)[1:(Nanimal*301)],
rep(Vg1,Nanimal)))
colnames(pev_l1) = c('id','dim','pev','vg')
pev_l1 = pev_l1[pev_l1$id %in% code_orig$pai,]
head(pev_l1)
dim(pev_l1)
class(pev_l1$pev)

#PEV para lactacao 2

```

```

pev_12=data.frame(cbind( sort(rep(seq(1,Nanimal),301)),
rep(seq(5,305),Nanimal), unlist(pev2)[1:(Nanimal*301)],
rep(Vg2,Nanimal)))
colnames(pev_12) = c('id','dim','pev','vg')
pev_12 = pev_12[pev_12$id %in% code_orig$pai,]
head(pev_12)
dim(pev_12)
class(pev_12$pev)

#PEV para lactacao 3
pev_13=data.frame(cbind( sort(rep(seq(1,Nanimal),301)),
rep(seq(5,305),Nanimal), unlist(pev3)[1:(Nanimal*301)],
rep(Vg3,Nanimal)))
colnames(pev_13) = c('id','dim','pev','vg')
pev_13 = pev_13[pev_13$id %in% code_orig$pai,]
head(pev_13)
dim(pev_13)
class(pev_13$pev)

#reliability por DIM
pev_11$rel = 1-(pev_11$pev/pev_11$vg)

pev_12$rel = 1-(pev_12$pev/pev_12$vg)

pev_13$rel = 1-(pev_13$pev/pev_13$vg)

#Voltar codigos originais
head(code_orig)

idx = acc_a$id %in% code_orig$id
table(idx)

pev_11 = merge(pev_11,code_orig, by='id',all.x=T,all.y=F)
pev_11 = subset(pev_11, select =
c('id','id_orig','dim','pev','vg','rel'))
pev_12 = merge(pev_12,code_orig, by='id',all.x=T,all.y=F)
pev_12 = subset(pev_12, select =
c('id','id_orig','dim','pev','vg','rel'))
pev_13 = merge(pev_13,code_orig, by='id',all.x=T,all.y=F)
pev_13 = subset(pev_13, select =
c('id','id_orig','dim','pev','vg','rel'))

#Exporta arquivos pev por dim
write.table(pev_11,file='./pev_11.txt',sep=' ',
col.names=F,row.names=F,quote=F)
write.table(pev_12,file='./pev_12.txt',sep=' ',
col.names=F,row.names=F,quote=F)
write.table(pev_13,file='./pev_13.txt',sep=' ',
col.names=F,row.names=F,quote=F)

#acc_dim =
cbind(pev_11[,c(1,2,5,6)],pev_12[,c(5,6)],pev_13[,c(5,6)])
#colnames(acc_dim) = c('id','dim','rel1','rel2','rel3')
#head(acc_dim)
#dim(acc_dim)

#pev por animal (media)
pev1_a = tapply(pev_11$pev,pev_11$id,FUN=mean)
pev1_a = as.data.frame(as.numeric(pev1_a))
sd pev1_a = tapply(pev_11$pev,pev_11$id,FUN=sd)

```

```

sd_peg1_a = as.data.frame(as.numeric(sd_peg1_a))
peg1_a = cbind(unique(peg_11$id),peg1_a,sd_peg1_a)
colnames(peg1_a) = c('id','peg1','sdpeg1')
head(peg1_a)
dim(peg1_a)

peg2_a = tapply(peg_12$peg,peg_12$id,FUN=mean)
peg2_a = as.data.frame(as.numeric(peg2_a))
sd_peg2_a = tapply(peg_12$peg,peg_12$id,FUN=sd)
sd_peg2_a = as.data.frame(as.numeric(sd_peg2_a))
peg2_a = cbind(unique(peg_12$id),peg2_a,sd_peg2_a)
colnames(peg2_a) = c('id','peg2','sdpeg2')
head(peg2_a)
dim(peg2_a)

peg3_a = tapply(peg_13$peg,peg_13$id,FUN=mean)
peg3_a = as.data.frame(as.numeric(peg3_a))
sd_peg3_a = tapply(peg_13$peg,peg_13$id,FUN=sd)
sd_peg3_a = as.data.frame(as.numeric(sd_peg3_a))
peg3_a = cbind(unique(peg_13$id),peg3_a,sd_peg3_a)
colnames(peg3_a) = c('id','peg3','sdpeg3')
head(peg3_a)
dim(peg3_a)

#reliability por animal (media)
rel_11a = tapply(peg_11$rel,peg_11$id,FUN=mean)
rel_11a = as.data.frame(as.numeric(rel_11a))
sd_rel_11a = tapply(peg_11$rel,peg_11$id,FUN=sd)
sd_rel_11a = as.data.frame(as.numeric(sd_rel_11a))
rel_11a = cbind(unique(peg_11$id),rel_11a,sd_rel_11a)
colnames(rel_11a) = c('id','rel1','sdrel1')
head(rel_11a)
dim(rel_11a)

rel_12a = tapply(peg_12$rel,peg_12$id,FUN=mean)
rel_12a = as.data.frame(as.numeric(rel_12a))
sd_rel_12a = tapply(peg_12$rel,peg_12$id,FUN=sd)
sd_rel_12a = as.data.frame(as.numeric(sd_rel_12a))
rel_12a = cbind(unique(peg_12$id),rel_12a,sd_rel_12a)
colnames(rel_12a) = c('id','rel2','sdrel2')
head(rel_12a)
dim(rel_12a)

rel_13a = tapply(peg_13$rel,peg_13$id,FUN=mean)
rel_13a = as.data.frame(as.numeric(rel_13a))
sd_rel_13a = tapply(peg_13$rel,peg_13$id,FUN=sd)
sd_rel_13a = as.data.frame(as.numeric(sd_rel_13a))
rel_13a = cbind(unique(peg_13$id),rel_13a,sd_rel_13a)
colnames(rel_13a) = c('id','rel3','sdrel3')
head(rel_13a)
dim(rel_13a)

#conferencias
table(rel_12a$id == peg2_a$id)

acc_a =
cbind(rel_11a,rel_12a[,2:3],rel_13a[,2:3],peg1_a[,2:3],peg2_a[,2:3
],peg3_a[,2:3])
head(acc_a)

acc_a = merge(acc_a,code_orig, by='id',all.x=T,all.y=F)

```

```
acc_a = subset(acc_a, select =  
c('id','id_orig','rel1','sdrel1','rel2','sdrel2','rel3','sdrel3',  
'pev1','sdpev1','pev2','sdpev2','pev3','sdpev3'))  
head(acc_a)  
dim(acc_a)  
  
#Exporta arquivos  
write.table(acc_a,file='./acc_a_rr.txt',sep=' ',  
col.names=F,row.names=F,quote=F)
```

## CHAPTER 3

### **Genotype by environment interaction for Portuguese and Brazilian Holstein populations using autoregressive and multi-trait reaction norms test-day models**

D. A. Silva<sup>1</sup>, P. S. Lopes<sup>1</sup>, C. N. Costa<sup>2</sup>, A. A. Silva<sup>1</sup>, H. T. Silva<sup>1</sup>, F. F. Silva<sup>1</sup>, R.  
Veroneze<sup>1</sup>, G. Thompson<sup>3,4</sup> and J. Carneiro<sup>3,4</sup>

<sup>1</sup> Department of Animal Science, Universidade Federal de Viçosa, 36570-000,  
Viçosa, Brazil

<sup>2</sup> Embrapa Gado de Leite, 36038-330, Juiz de Fora, Brazil

<sup>3</sup> Research Center in Biodiversity and Genetic Resources (CIBIO-InBio), University  
of Porto, 4485-661, Vairão, Portugal

<sup>4</sup> Institute of Biomedical Sciences Abel Salazar (ICBAS), University of Porto, 4050-  
313, Porto, Portugal

### 3.1. Abstract

A two-step genetic evaluation was used to verify the presence of genotype by environment interaction (G x E). In step 1, we performed a within-country (Portugal and Brazil separately) genetic evaluation by using an autoregressive test-day (AR) model to obtain precorrected phenotypes, solutions for environmental effects (herd-test-day, HTD levels), genetic parameters, and rank correlations. In step 2, the precorrected phenotypes from Portugal and Brazil were considered as two distinct traits in a bi-trait reaction norm models to provide estimates of the intercept, regression coefficients and genetic correlation between environments in Portugal and Brazil. In step 1, genetic additive variance from Portugal was 14.1% higher from Brazil for milk yield (MY). For somatic cell score (SCS), the genetic additive variance for Portugal was 12.7% lower compared to Brazil. On the other hand, within lactation permanent environmental (STE) variance from Brazil for MY and SCS were 68.4 and 33.6% higher than Portugal, respectively. Although similar heritability estimates for SCS were observed for both countries, for MY the heritability was 0.31 for Portugal and 0.23 for Brazil. The rank correlation between predicted breeding values of the common bulls obtained from the within-country evaluations were 0.75 for MY and 0.62 for SCS. In step 2, for MY, the genetic correlation among HTD levels within countries were higher than 0.92 for Portugal and 0.98 for Brazil. For SCS, the genetic correlation among HTD levels ranged from 0.64 to 0.99 for Portugal and from 0.79 to 0.99 for Brazil. The average of genetic correlation estimates between Portugal and Brazil HTD levels was 0.73 for MY and 0.57 for SCS. Our results suggested genotype by environment interaction in Holstein cattle by considering different production systems and climatic conditions (temperate in Portugal and tropical in Brazil). The low genetic correlation in the Portuguese population indicated the presence of G x E among

extreme HTD levels for SCS. This study evidences the importance of testing genotypes under local environmental circumstances before imported semen investments.

**Keywords:** Dairy cattle, G x E, Environment gradient, Temperate climate, Tropical climate

### 3.2. Introduction

Reproductive biotechnologies, such as artificial insemination, has contributed to the extensive and global distribution of genetic material from animals with high productive potential. This process allows bulls to have daughters in multiple management levels and different climatic conditions. In addition, it may lead to different expressions of the genetic potential according to environment level (Kolmodin et al., 2002; Hayes et al., 2016) and, finally culminating in a reclassification of the animals (Zwald et al., 2003). Therefore, knowledge of genotype by environment interaction (G x E) may be relevant for breeders investing in foreign genotypes (Silva et al., 2014).

The G x E for milk traits within- and across-countries is routinely evaluated by using multi-trait or reaction norm models (e.g. Hammami et al., 2009a; Streit et al., 2012; Li et al., 2016). The reaction norm models are interesting because they use covariance functions to describe the individual's genetic merit as a function of an environmental gradient as well as to estimate of the genetic correlations over the environmental gradient(Silva et al., 2014).

Several studies have been reported the effect of G x E between pairs of countries with different production systems and climatic conditions in Holstein cattle. For example, Li et al. (2016) reported G x E (genetic correlation of 0.63) between

Brazil and European countries (Nordic and France) populations using multi-trait models. Similar results were reported by Hammami et al. (2009a; 2009b) when G x E was investigated between Luxembourg and Tunisian populations (genetic correlation ranged from 0.39 to 0.70). On the other hand, when G x E was investigated within-country, Streit et al. (2012) and Moreira et al. (2019) reported no G x E effect between environmental gradients. To our knowledge, no multi-trait reaction norm models for evaluating G x E between environmental gradients within- and across-country were performed for milk traits between the Portuguese and Brazilian Holstein populations.

The Portuguese and Brazilian Holstein populations are composed by continuous importation of genotypes. For example, in Brazil, around 94% of total semen has been imported, mainly from Northern American and European countries (ASBIA, 2017). A similar situation (about 97%) also occurs to Portugal. These two countries have distinct climatic conditions and management levels. Portugal, as in most European countries, is characterized by a temperate climate, which naturally favors the Holstein cattle performance. In addition, dairy farms may be considered as high-input production systems. On the other hand, climate conditions in Brazil is defined as a tropical climate, where the Holstein cattle production is influenced by many limitations, such as health stressors, variable feed quality and management (Costa et al., 2000).

These differences reinforce the importance of having reliable evaluation of G x E for imported genotypes in order to adjust the breeding strategies of these countries, if necessary. Therefore, our aim with this study was to evaluate G x E for milk yield (MY) and somatic cell scores (SCS) between Portuguese and Brazilian Holstein cattle populations by using a two-step approach: EBV prediction within-country using an autoregressive test-day model and genetic correlation estimation among

environmental gradients within- and across-country using a multi-trait reaction norm model.

### **3.3. Material and methods**

A two-step genetic evaluation was used to investigate the presence of G x E in Portuguese and Brazilian Holstein cattle (Calus et al., 2002; Kolmodin et al., 2002). In step 1, we performed a within-country (Portugal and Brazil separately) genetic evaluation by using autoregressive test-day (AR) model to estimate environment effects (herd-test-day, HTD levels), genetic parameters, and rank correlations. In step 2, genetic evaluation by using a bi-trait reaction norm model was performed to estimate the intercept and regression coefficients, as well as genetic correlation between Portugal and Brazil environments.

#### **Data**

In step 1, first lactation test-day records of MY and SCS from Portuguese and Brazilian Holstein cows, that calve between 1994 and 2017 were provided by Portuguese Dairy Cattle Breeders and Brazilian Holstein Cattle Breeders Associations. The data sets were edited according to predefined criteria for genetic analysis using AR models (Carvalheira et al., 2002; Silva et al., 2019).

In step 2, the two countries combined data sets for MY (SCS) were assembled using the daughters' information from 797 (737) bulls in common. The bulls were from different origins: USA 472 (441), Canada 185 (164), the Netherlands 85 (83), France 26 (21), Spain 13 (13), Italy 10 (9), and Germany, Belgium, Australia, and United Kingdom 6 (6). These bulls were required to have a minimum of 10 daughters in at

least five herds in each country. A general description of the data sets used in both steps is shown in Table 1.

**Table 1** Description of data sets (Portugal and Brazil) used for within-country evaluations using autoregressive model (step 1) and reaction norms analyses (step 2)

Item	Step 1		Step 2	
	Portugal	Brazil	Portugal	Brazil
Milk yield				
TD records	4,615,920	1,611,689	1,133,965	894,225
Cows	565,494	204,977	137,506	112,200
Bulls	22,824	5,842	797	797
Bulls > 10 daughters	12,486	4,636	797	797
Herds	3,997	1,635	3,037	1,500
Mean, kg (SD)	25.91(6.92)	26.15(7.06)	27.53(6.81)	26.88(7.00)
TD records per cow	8.16	7.86	8.25	7.97
Somatic cell score (SCS) <sup>1</sup>				
TD records	4,071,754	1,312,073	985,394	733,921
Cows	513,993	179,839	123,399	98,562
Bulls	21,710	5,163	737	737
Bulls > 10 daughters	11,671	3,939	737	737
Herds	3,516	1,538	2,857	1,384
Mean, score (SD)	2.62(1.73)	2.67(2.02)	2.52(1.70)	2.58(2.01)
TD records per cow	7.92	7.3	7.99	7.45

<sup>1</sup>SCS =  $\log_2(\text{SCC}/100) + 3$ , in which SCC is somatic cell count.

#### Statistical models for step 1

The AR model was fitted for each country and can be described as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{H}\mathbf{c} + \mathbf{Q}\mathbf{t} + \mathbf{e},$$

where  $\mathbf{y}$  is the vector of TD records;  $\boldsymbol{\beta}$  is the vector of fixed effects (age class at calving nested within herd, and days in milk class nested within herd and lactation);  $\mathbf{a}$  is the vector of random additive genetic effects;  $\mathbf{c}$  is the vector of random contemporary group (HTD) effects;  $\mathbf{t}$  is the random short-term environmental effects (STE); and  $\mathbf{e}$  is the vector of random residual. The  $\mathbf{X}$ ,  $\mathbf{Z}$ ,  $\mathbf{H}$  and  $\mathbf{Q}$  are incidence matrices relating

observations to fixed and random effects. A first order autoregressive covariance structure was assumed for HTD and STE effects. More details about AR model are found in Carvalheira et al. (1998, 2002) and Silva et al. (2019).

For the estimation of variance components (VC) and autocorrelation coefficients, five random samples of the data were created for both countries. Each sample consisted of data from 20 herds for MY representing all regions with records in Portugal and Brazil. These samples were used to obtain estimates of VC and autocorrelations coefficients using DFREML procedures (Smith and Graser, 1986). Likelihood functions were maximized by the multivariate simplex algorithm (Nelder and Mead, 1965). The convergence criterion was defined as  $10^{-8}$ . The occurrence of local maxima was checked by running five consecutive cold starts without significant changes in the log-likelihood (up to four decimal places).

The genetics parameters, rank correlation, and coincidence percentage were used to evaluate the G x E between Portugal and Brazil.

Jointly with solutions from HTD estimates ( $\hat{\mathbf{c}}$ ), the precorrected phenotypes  $\mathbf{y}^* = \mathbf{y} - (\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{H}\hat{\mathbf{c}} + \mathbf{Q}\hat{\mathbf{t}}) = \mathbf{Z}\hat{\mathbf{a}} + \hat{\mathbf{e}}$ , were used in the reaction norm model in step 2.

Statistical models for step 2

The precorrected phenotypes from Portugal and Brazil were considered two distinct but correlated traits. The bi-trait reaction norm model was fitted using Legendre polynomial. In matrix notation, this model can be described as follows:

$$\mathbf{y}^* = \mathbf{X}\mathbf{b} + \mathbf{Z}_1\mathbf{s} + \mathbf{Z}_2\mathbf{p} + \mathbf{e}$$

where  $\mathbf{y}^*$  is the vector of precorrected phenotypes;  $\mathbf{b}$  is a vector of fixed regression coefficients for HTD levels;  $\mathbf{s}$  is the vector of random regression coefficients for the sire additive genetic effect;  $\mathbf{p}$  is the vector of random regression coefficients for the

permanent environment effect; and  $\mathbf{e}$  is the vector of random residual effects. The  $\mathbf{X}$ ,  $\mathbf{Z}_1$  and  $\mathbf{Z}_2$  are incidence matrices relating observations to fixed and random effects. These matrices contain Legendre Polynomials corresponding to the HTD levels.

Expectations and covariance structure for random effects were defined as follows:

$$\mathbf{E}(\mathbf{y}^*) = \mathbf{X}\mathbf{b}, \mathbf{E}(\mathbf{s}) = \mathbf{0}, \mathbf{E}(\mathbf{p}) = \mathbf{0}, \mathbf{E}(\mathbf{e}) = \mathbf{0},$$

and

$$\mathbf{V}(\mathbf{y}^*) = \mathbf{Z}_1\mathbf{G}\mathbf{Z}_1' + \mathbf{Z}_2\mathbf{P}\mathbf{Z}_2' + \mathbf{R}, \mathbf{V}(\mathbf{s}) = \mathbf{G}, \mathbf{V}(\mathbf{p}) = \mathbf{P}, \mathbf{V}(\mathbf{e}) = \mathbf{R},$$

where  $\mathbf{G} = \mathbf{A} \otimes \mathbf{G}_0$ ,  $\mathbf{A}$  represents the pedigree-based additive genetic relationship and  $\mathbf{G}_0$  is the covariance matrix of the sire additive genetic regression coefficients; and  $\mathbf{P} = \mathbf{I} \otimes \mathbf{P}_0$ ,  $\mathbf{I}$  is the identity matrix and  $\mathbf{P}_0$  is the covariance matrix of the

permanent environment regression coefficients.  $\mathbf{R} = \begin{bmatrix} \mathbf{I}\sigma_{PT}^2 & 0 \\ 0 & \mathbf{I}\sigma_{BR}^2 \end{bmatrix}$ , where  $\sigma_{PT}^2$  and  $\sigma_{BR}^2$  are the residual variances in Portugal and Brazil, respectively.

Reaction norm random regression were solved by using REMLF90 software (Misztal et al., 2015). The appropriate order of the fixed and random regression models based on Akaike information criterion (AIC) and Akaike Weights (AW) were 3 for fixed regression, 3 and 1 for random regressions (sire additive genetic and permanent environment effects, respectively). AIC is defined as  $AIC = -2\log(L) + 2p$ , where  $L$  is the Likelihood function, and  $p$  is the number of estimated parameters and AW as  $AW_i = \exp\left(-\frac{\Delta_i}{2}\right) / \sum_{i=1}^2 \exp\left(-\frac{\Delta_i}{2}\right)$ , for  $i$  representing each model evaluated with different orders (Burnham and Anderson, 2004). AIC provides a qualitative comparison between models (smallest AIC is better) and  $AW_i$  represents the probability of model  $i$  to be better than the other model evaluated.

The predicted breeding values (EBV<sub>ij</sub>) for each country of each bull *i* at HTD level *j* was obtained as:  $EBV_{ij} = \mathbf{K}_L \mathbf{s}_i$ , where  $\mathbf{s}_i$  is the vector of predicted breeding values for the regression coefficients of each bull *i* and  $\mathbf{K}_L$  is the matrix of Legendre Polynomials corresponding to the HTD levels in each country. The additive genetic covariances matrix ( $\widehat{\mathbf{G}}$ ) between HTD level within- and across-country were given by  $\widehat{\mathbf{G}} = 4 * \mathbf{K}_1 \widehat{\mathbf{G}}_0 \mathbf{K}'_1$ , where  $\mathbf{K}_1 = \mathbf{I} \otimes \mathbf{K}_L$ .

The genetic correlation between HTD level within- and between-country was calculated from entries of the covariance matrix  $\widehat{\mathbf{G}}$  of the additive effects at each HTD level. These estimates were used as an indicator for G x E between HTD levels within- and across-country. The changes in EBV along of the HTD levels were also analyzed.

### 3.4. Results and Discussion

#### Step 1 – within-country evaluation

Variance components, autocorrelations and heritability estimates for MY obtained from within-country evaluations using the AR model are shown in Table 2. The relative magnitude of the VC and autocorrelations for MY and SCS are consistent with the literature (Vasconcelos et al., 2008; Silva et al., 2019). For MY, genetic additive variance from Portugal was 14.1% higher than corresponding estimates from Brazil. For SCS, genetic additive variance from Portugal was 12.7% lower compared to Brazil. On the other hand, STE variance from Brazil for MY and SCS were 68.4 and 33.6% higher than Portugal, respectively. Similar results were reported by Hammami et al. (2008), in which the Tunisian Holstein population presented lower genetic variance if compared to the Luxembourg population. The authors suggested that this may occurred due to difficulties encountered by daughters of superior bulls to express their genetic potential under harsh conditions. In addition, they associated the

largest environmental variances in the Tunisian population to poor management and feeding quality fluctuations during the year.

In this context, the lowest genetic variance for milk traits is expected in tropical conditions. This is because daughters of superior bulls (in favorable conditions) are unable to express their maximum genetic potential when in unfavorable production systems and/or harsh climatic conditions. In addition, the largest STE variances observed in Brazil could be due to the health stressors and climate conditions, which introduce additional variation that may be permanently associated with each cow (Costa et al., 2000).

MY heritability was higher in Portugal when compared to Brazil, but similar for SCS (Table 2). Hammami et al. (2008) found a lower heritability for MY in the Tunisian population (0.17) compared to Luxembourg (0.41). The authors concluded that these differences were caused by divergences in production levels, resulting from differences in climatic and management conditions.

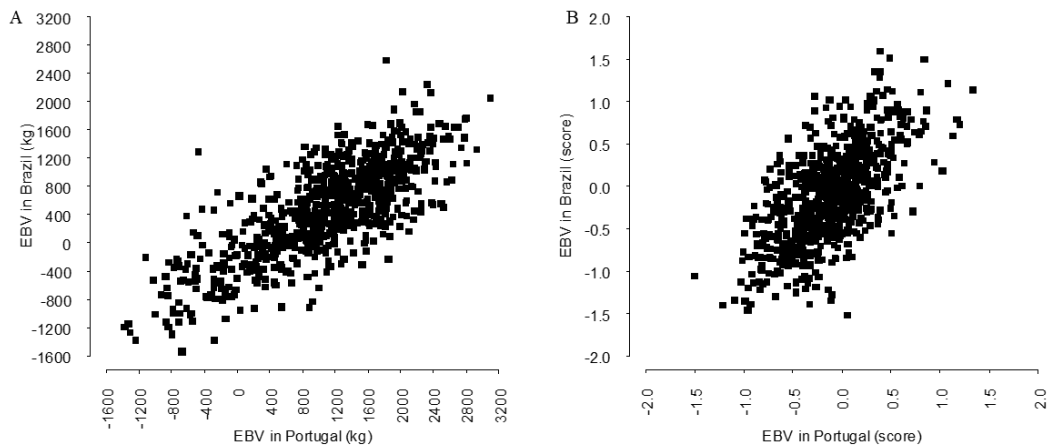
Figure 1 plots the predicted breeding values (EBV) for the common bulls obtained from within-country evaluations. The coefficient of regression (Hammami et al., 2008) was used as an indicator for the expected response on daughters in Brazil (tropical environment) when bulls are selected using the EBV from Portugal (temperate environment). The regression coefficients were 0.62 for MY and 0.85 for SCS, indicating that a higher response is expected in Brazil for SCS than MY. For milk yield, these results indicated which expression of genetic potential is more favored by Portuguese environmental conditions. On the other hand, for SCS, the environmental conditions of both countries seemed to have a similar effect on the phenotypic expression. This interpretation is in line with our results of heritability estimates for SCS, in which similar results were observed for both countries. Lower regression

coefficients have been reported in literature. Ojango and Pollott (2002) reported a regression coefficient of 0.32 when bulls EBV from Kenya were regressed on corresponding EBV from the United Kingdom, whereas Hammami et al. (2008) reported a regression coefficient of 0.18 when bulls EBV from Tunisia were regressed on corresponding EBV from Luxembourg.

**Table 2** Variance components, autocorrelation coefficients and heritability estimates with respective standard-errors (SE) estimated from within-country evaluations using autoregressive model (step 1) in Portugal and Brazil

Parameters <sup>1</sup>	Portugal		Brazil	
	MY $\pm$ SE	SCS $\pm$ SE	MY $\pm$ SE	SCS $\pm$ SE
$\sigma_a^2$	10.14 $\pm$ 1.462	0.55 $\pm$ 0.047	8.89 $\pm$ 0.268	0.63 $\pm$ 0.009
$\sigma_t^2$	10.97 $\pm$ 1.367	1.22 $\pm$ 0.046	18.47 $\pm$ 0.270	1.63 $\pm$ 0.022
$\sigma_c^2$	4.47 $\pm$ 0.494	0.09 $\pm$ 0.012	3.18 $\pm$ 0.101	0.18 $\pm$ 0.005
$\sigma_e^2$	6.94 $\pm$ 0.919	1.10 $\pm$ 0.142	7.88 $\pm$ 0.209	1.22 $\pm$ 0.018
$\rho_t$	0.82 $\pm$ 0.009	0.83 $\pm$ 0.025	0.68 $\pm$ 0.007	0.85 $\pm$ 0.003
$\rho_c$	0.58 $\pm$ 0.025	0.54 $\pm$ 0.070	0.68 $\pm$ 0.011	0.61 $\pm$ 0.005
$h^2$	0.31 $\pm$ 0.037	0.19 $\pm$ 0.013	0.23 $\pm$ 0.004	0.17 $\pm$ 0.003

<sup>1</sup> $\sigma_a^2$ : additive genetic variance, kg<sup>2</sup> for milk yield (MY) and score<sup>2</sup> for somatic cell score (SCS);  $\sigma_t^2$ : short-term environmental variance;  $\sigma_c^2$ : herd-test-day variance;  $\sigma_e^2$ : residual variance;  $\rho_{ti}$  and  $\rho_c$ : autocorrelations for STE and HTD effects;  $h^2$ : heritability.



**Figure 1** Predicted breeding value (EBV) for milk yield (A) and somatic cell score (B) for common bulls between Portugal and Brazil from within-country evaluations using autoregressive model (step1).

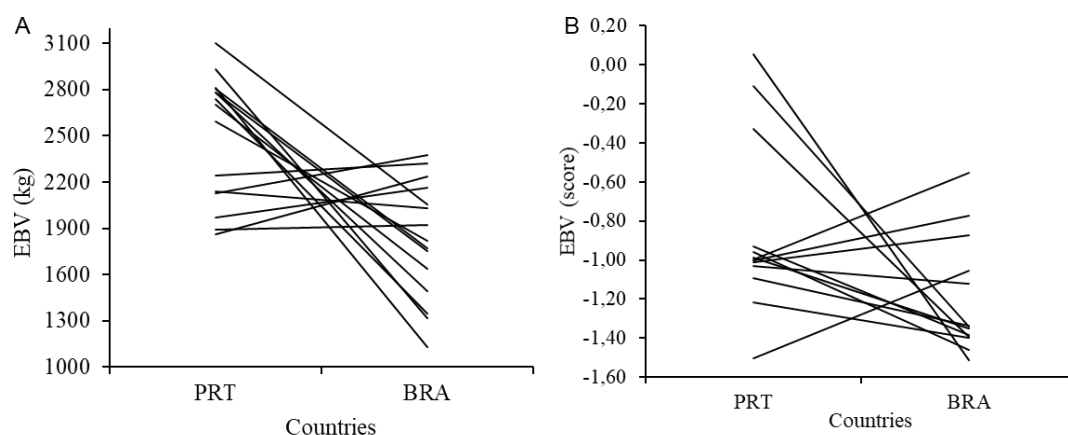
The rank correlation and the percentage of coincidence between EBVs from common bulls obtained from within-country evaluations across different proportions of selection, are given in Table 3. The literature (e.g., Kolmodin et al., 2002; Hayes et al., 2016) indicates that when the differences in phenotypic expression between two genotypes are larger in one of the environments, with a change in ranking, there is a scaling effect of G x E.

By using the rank correlation as consequence of the G x E, our study showed a correlation lower than 0.75, for both traits. In addition, the rank correlation and the percentage of coincidence also decreased as the percentage of bulls selected decreased. A higher re-ranking among the common bulls was observed for SCS compared to MY, between the Portugal and Brazil. The effects of the re-ranking between the two countries are well depicted in Figure 2. Similar results were reported by Hammami et al., (2008), in which the authors found an expressive re-ranking ( $r = 0.41$ ) among common bulls from Luxembourg and the Tunisian population.

**Table 3** EBV rank correlations and percentage of coincidence for common bulls in Portugal and Brazil from within-country evaluations using autoregressive model (step 1) for milk yield (MY) and somatic cell score (SCS)

Common bulls selected (%)	MY		SCS	
	Rank <sup>1</sup>	Coincidence (%)	Rank <sup>1</sup>	Coincidence (%)
1	0.09	12.5	0.33	37.5
10	0.35	45.0	0.46	40.5
50	0.41	78.7	0.43	71.0
100	0.75	100	0.62	100

<sup>1</sup>Rank correlations for common bulls ranked against predictions from Portugal.



**Figure 2** Predicted breeding value (EBV) for milk yield (A) and somatic cell score (B) of top 1% bulls (n=8) in Portugal (PRT) and Brazil (BRA) from within-country evaluations using autoregressive model (step 1).

#### Step 2 –reaction norms evaluation

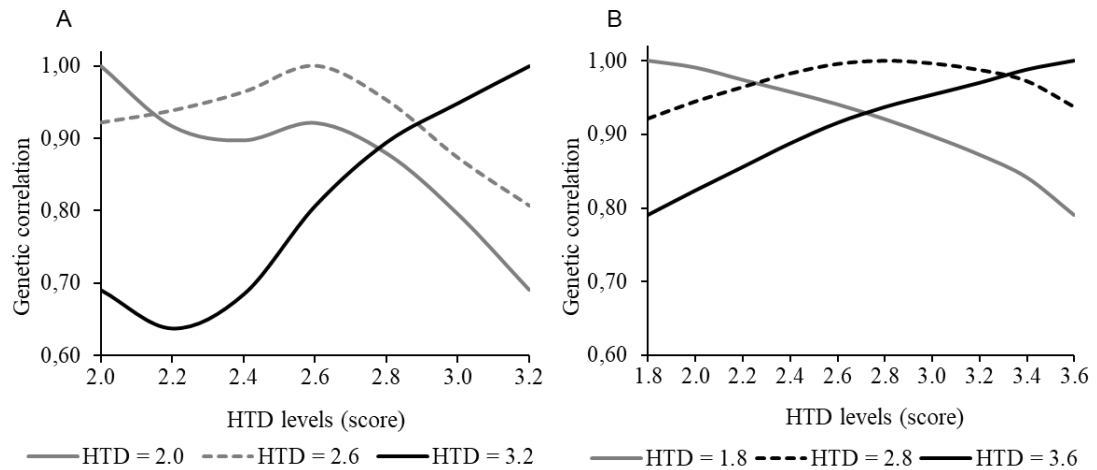
The herd test-day solutions (HTD levels) from the within-country evaluation was used as environment gradient in our study. The HTD levels were very similar between Portugal and Brazil and ranged from 21 to 30 kg for MY and from 1.8 to 3.6 scores for SCS. For milk yield, high HTD levels indicated more favorable environments (higher level of production), while the low HTD levels indicated more

unfavorable environments (lower level of production). By using the HTD levels as an environmental gradient for SCS, it was assumed that they reflect the level of hygiene in the farms, especially the infection pressure on the udder (Streit et al., 2012).

A genetic correlation of 0.80 was chosen as a biological threshold of G x E (Hayes et al., 2016), in which values lower than 0.80 suggests evidence of G x E. For MY, the genetic correlation between HTD levels within countries were higher than 0.92 for Portugal and 0.98 for Brazil, indicating the absence of G x E between HTD levels within country. For SCS, the genetic correlation between HTD levels ranged from 0.64 to 0.99 for Portugal and from 0.79 to 0.99 for Brazil. Genetic correlations were plotted for all levels of HTD with the HTD levels of 2.0 (highest hygiene), 2.6 (intermediate hygiene) and 3.2 (poorest hygiene) for Portugal (Figure 3A) and values of 1.8 (highest hygiene), 2.8 (intermediate hygiene) and 3.6 (poorest hygiene) for Brazil (Figure 3B). For Portugal, the genetic correlation between extreme HTD levels (2.0 and 3.2) was remarkably low, which indicated substantial G x E.

Nevertheless, for Brazil, the genetic correlations were very close to 0.80, which indicated the absence of G x E between the HTD levels. In this context, Zwald et al. (2003) highlighted the role of environment as the possible predominant factor in the effect on production, provided that the manifestations of the genetic components are influenced by the variables (e.g. temperature differences between the regions). The Brazilian territory is vast, presenting regions with different climatic conditions and varied production systems. Such scenarios could lead to different performances of genotypes, which characterizes G x E. Nevertheless, no G x E effect between HTD levels was observed in our study for Brazil. A possible reason could be that 98% of herds used in this study were from the South (69%) and Southeast (29%) regions of Brazil, which could reduce the differences between environmental gradients. This is

supported by results of Moreira et al. (2019), in which the authors evaluated the G x E on milk traits of Holstein cows in southern Brazil and also did not find significant interactions.



**Figure 3** Genetic correlation estimates for somatic cell score across different herd-test-day (HTD) levels in Portugal (A) and Brazil (B) from reaction norm evaluations (step 2).

The genetic correlation estimates between HTD levels from Portugal with HTD levels from Brazil are shown in Tables 4 and 5, for MY and SCS, respectively. In support to the rank correlations from step 1 (0.75 for MY and 0.62 for SCS), the average of genetic correlations was 0.73 for MY and 0.57 for SCS. This indicates expressive G x E between Portugal and Brazil. These results confirm that superior animals for MY and SCS in Portugal are not necessarily the best in Brazil and vice versa. The presence of G x E is expected between both countries due mainly to differences in production systems and climate conditions. These results emphasize the need and importance of including Brazil in international genetic evaluations that accounts for the interaction of the genotypes with the Brazilian environments, complementing the national evaluation (Costa et al., 2000).

**Table 4** Genetic correlation for milk yield between HTD levels from Brazil with HTD levels from Portugal

HTD levels in Brazil	HTD levels in Portugal									
	21	22	23	24	25	26	27	28	29	30
22	0.69	0.72	0.73	0.74	0.74	0.74	0.74	0.74	0.73	0.72
23	0.71	0.73	0.74	0.75	0.75	0.75	0.74	0.74	0.73	0.72
24	0.71	0.73	0.74	0.75	0.75	0.75	0.74	0.74	0.73	0.71
25	0.71	0.73	0.74	0.75	0.75	0.75	0.74	0.74	0.73	0.72
26	0.71	0.73	0.74	0.75	0.75	0.75	0.74	0.74	0.73	0.72
27	0.70	0.72	0.74	0.74	0.75	0.75	0.74	0.74	0.73	0.72
28	0.70	0.72	0.73	0.74	0.75	0.75	0.74	0.74	0.73	0.72
29	0.70	0.72	0.73	0.74	0.75	0.75	0.75	0.74	0.73	0.72
30	0.71	0.73	0.74	0.74	0.75	0.75	0.75	0.74	0.74	0.72

HTD = herd-test-day

**Table 5** Genetic correlation for somatic cell score between HTD levels from Brazil with HTD levels from Portugal

HTD levels in Brazil	HTD levels in Portugal						
	2.0	2.2	2.4	2.6	2.8	3.0	3.2
1.8	0.49	0.50	0.54	0.59	0.60	0.58	0.55
2.0	0.52	0.53	0.56	0.61	0.60	0.58	0.55
2.2	0.53	0.54	0.57	0.61	0.61	0.58	0.55
2.4	0.55	0.54	0.57	0.62	0.62	0.59	0.56
2.6	0.55	0.54	0.57	0.62	0.63	0.60	0.57
2.8	0.55	0.53	0.56	0.62	0.63	0.61	0.58
3.0	0.54	0.52	0.55	0.62	0.63	0.62	0.59
3.2	0.53	0.51	0.54	0.61	0.63	0.62	0.59
3.4	0.52	0.50	0.53	0.60	0.62	0.61	0.59
3.6	0.51	0.49	0.52	0.57	0.59	0.58	0.57

HTD = herd-test-day

There are several studies reporting the effect of G x E between pairs of countries with different production systems and climatic conditions for Holstein cattle. Genetic correlations for MY of 0.63 was reported between USA and Mexico (Cienfuegos-Rivas et al., 1999), 0.67 between Luxembourg and Tunisia (Hammami et

al., 2009a) and 0.63 between Brazil and European countries (Nordic and France) (Li et al., 2016). On the other hand, Costa et al. (2000) did not observe G x E between USA and Brazil (genetic correlation of 0.85).

### **3.5. Conclusion**

Our results indicated the presence of genotype by environment interaction in Holstein cattle considering different production systems and climatic conditions from the temperate environment (Portugal) and the tropical environment (Brazil). The low genetic correlation in the Portuguese population indicated the presence of G x E between extreme environmental gradient (HTD levels) for somatic cell score. This study emphasizes the importance of testing the genotypes under local environmental circumstances for imported semen investments.

### **3.6. Acknowledgements**

The authors acknowledge Portuguese Dairy Cattle Breeders Association and Brazilian Holstein Cattle Breeders Association (ABCBRH) for providing data for this study. This study was partially financed by Coordination for the Improvement of Higher Education Personnel and Portuguese National Funding Agency for Science, Research and Technology (CAPES/FCT, n° 99999.008462/2014-03 and 88887.125450/2016-00), and National Council of Technological and Scientific Development (CNPq 465377/2014-9 - PROGRAMA INCT and CNPq 142467/2015-4).

### **3.7. References**

ASBIA. 2017. Index - Associação Brasileira de Insiminação Artificial 2017. Accessed. <http://www.asbia.org.br/wp-content/uploads/2018/10/INDEX-ASBIA->

2017\_completo.pdf.

- Burnham, K.P., and D.R. Anderson. 2004. Multimodel inference: Understanding AIC and BIC in model selection. *Sociol. Methods Res.* 33:261–304. doi:10.1177/0049124104268644.
- Calus, M.P.L., A.F. Groen, and G. de Jong. 2002. Genotype  $\times$  Environment Interaction for Protein Yield in Dutch Dairy Cattle as Quantified by Different Models. *J. Dairy Sci.* 85:3115–3123. doi:10.3168/jds.s0022-0302(02)74399-3.
- Carvalho, J., E.J. Pollak, R.L. Quaas, and R.W. Blake. 2002. An autoregressive repeatability animal model for test-day records in multiple lactations. *J. Dairy Sci.* 85:2040–2045. doi:10.3168/jds.S0022-0302(02)74281-1.
- Carvalho, J.G., R.W. Blake, E.J. Pollak, R.L. Quaas, and C. V. Duran-Castro. 1998. Application of an autoregressive process to estimate genetic parameters and breeding values for daily milk yield in a tropical herd of Lucerna cattle and in United States Holstein herds. *J. Dairy Sci.* 81:2738–2751. doi:10.3168/jds.S0022-0302(98)75831-X.
- Cienfuegos-Rivas, E.G., P.A. Oltenacu, R.W. Blake, S.J. Schwager, H. Castillo-Juarez, and F.J. Ruiz. 1999. Interaction between milk yield of Holstein cows in Mexico and the United States. *J. Dairy Sci.* 82:2218–2223. doi:10.3168/jds.s0022-0302(99)75468-8.
- Costa, C.N., R.W. Blake, E.J. Pollak, P. a Oltenacu, R.L. Quaas, and S.R. Searle. 2000. Genetic analysis of Holstein cattle populations in Brazil and the United States.. *J. Dairy Sci.* 83:2963–2974. doi:10.3168/jds.S0022-0302(00)75196-4.
- Hammami, H., B. Rekik, C. Bastin, H. Soyeurt, J. Bormann, J. Stoll, and N. Gengler. 2009a. Environmental sensitivity for milk yield in Luxembourg and Tunisian Holsteins by herd management level. *J. Dairy Sci.* 92:4604–4612. doi:10.3168/jds.2008-1513.
- Hammami, H., B. Rekik, H. Soyeurt, C. Bastin, E. Bay, J. Stoll, and N. Gengler. 2009b. Accessing genotype by environment interaction using Within- and Across-country Test-day random regression sire models. *J. Anim. Breed. Genet.* 126:366–377. doi:10.1111/j.1439-0388.2008.00794.x.
- Hammami, H., B. Rekik, H. Soyeurt, C. Bastin, J. Stoll, and N. Gengler. 2008. Genotype  $\times$  environment interaction for milk yield in Holsteins using Luxembourg and Tunisian populations.. *J. Dairy Sci.* 91:3661–3671. doi:10.3168/jds.2008-1147.

- Hayes, B.J., H.D. Daetwyler, and M.E. Goddard. 2016. Models for genome  $\times$  environment interaction: examples in livestock. *Crop Sci.* 56:2251–2259. doi:10.2135/cropsci2015.07.0451.
- Kolmodin, R., E. Strandberg, P. Madsen, J. Jensen, and H. Jorjani. 2002. Genotype by Environment Interaction in Nordic Dairy Cattle Studied Using Reaction Norms. *Acta Agric. Scand.* 52:11–24. doi:10.1080/09064700252806380.
- Li, X., M.S. Lund, Q. Zhang, C.N. Costa, V. Ducrocq, and G. Su. 2016. Short communication: Improving accuracy of predicting breeding values in Brazilian Holstein population by adding data from Nordic and French Holstein populations. *J. Dairy Sci.* 99:4574–4579. doi:10.3168/jds.2015-10609.
- Misztal, I., S. Tsuruta, D. Lourenco, I. Aguilar, A. Legarra, and Z. Vitezica. 2015. Manual for BLUPF90 Family of Programs. Accessed October 10, 2017. [http://nce.ads.uga.edu/wiki/doku.php?id=application\\_programs](http://nce.ads.uga.edu/wiki/doku.php?id=application_programs).
- Moreira, R.P., L.F.B. Pinto, A.A. Valloto, and V.B. Pedrosa. 2019. Evaluation of genotype by environment interactions on milk production traits of Holstein cows in southern Brazil. *Asian-Australasian J. Anim. Sci.* 32:459–466. doi:10.5713/ajas.18.0174.
- Nelder, J.A., and R. Mead. 1965. A Simplex method for function minimization. *Comput. J.* 7:308–313. doi:10.1093/comjnl/7.4.308.
- Silva, D.A., C.N. Costa, A.A. Silva, F.F. Silva, P.S. Lopes, G.G. dos Santos, G. Thompson, and J. Carvalheira. 2019. Unknown parent and contemporary groups for genetic evaluation of Brazilian Holstein using autoregressive test-day models. *Livest. Sci.* 220:1–7. doi:10.1016/j.livsci.2018.11.022.
- Silva, F.F., H.A. Mulder, E.F. Knol, M.S. Lopes, S.E.F. Guimaraes, P.S. Lopes, P.K. Mathur, J.M.S. Viana, and J.W.M. Bastiaansen. 2014. Sire evaluation for total number born in pigs using a genomic reaction norms approach. *J. Anim. Sci.* 92:3825–3834. doi:10.2527/jas.2013-6486.
- Smith, S.P., and H.U. Graser. 1986. Estimating variance components in a class of mixed models by restricted maximum likelihood. *J. Dairy Sci.* 69:1156–1165. doi:10.3168/jds.S0022-0302(86)80516-1.
- Streit, M., F. Reinhardt, G. Thaller, and J. Bennewitz. 2012. Reaction norms and genotype-by-environment interaction in the German Holstein dairy cattle. *J. Anim. Breed. Genet.* 129:380–389. doi:10.1111/j.1439-0388.2012.00999.x.
- Vasconcelos, J., F. Santos, A. Bagnato, and J. Carvalheira. 2008. Effects of clustering

herds with small-sized contemporary groups in dairy cattle genetic evaluations. *J. Dairy Sci.* 91:377–384. doi:10.3168/jds.2007-0202.

Zwald, N.R., K.A. Weigel, W.F. Fikse, and R. Rekaya. 2003. Identification of factors that cause genotype by environment interaction between herds of Holstein cattle in seventeen countries. *J. Dairy Sci.* 86:1009–1018. doi:10.3168/jds.s0022-0302(03)73684-4.

## GENERAL CONCLUSION

In Brazil, the majority of dairy herds supervised by the official Holstein milk recording system are characterized by small sizes. In addition, the genetic composition of the Brazilian Holstein population is characterized by the continuous importation of genotypes as well as animals originated from commercial Holstein herds, previously not registered in the official identification system. In this context, the amount of information recovered by using random contemporary groups and the correction of the base population with the inclusion of unknown phantom groups showed a strong and positive effect on the genetic evaluations of longitudinal traits in Brazilian Holstein cattle.

Autoregressive and random regression test-day models for multiple lactations performed well and may be used for genetic parameter estimation and genetic evaluations of milk yield and somatic cell score (SCS) in multiple-lactation of the Holstein cattle in Brazil. In general, the autoregressive models were more efficient and, given the lower number of parameters to estimate and its suitability to fit data from small herds (especially relevant to avoid loss of information due to data editing), these models are more parsimonious and should be chosen for genetic evaluations of longitudinal traits of the Holstein cattle in Brazil.

We reported the presence of genotype by environment interaction in Holstein cattle considering different production systems and climatic conditions from the temperate (Portugal) and the tropical environment (Brazil). The low genetic correlation in the Portuguese population indicated the presence of G x E between extreme environmental gradient for SCS. This study emphasizes the importance of testing the genotypes under local environmental circumstances before imported semen investments.