

UNIVERSIDADE FEDERAL DE VIÇOSA

HUGO TEIXEIRA SILVA

**AUTOREGRESSIVE MODEL FOR GENETIC EVALUATION OF REPRODUCTIVE
TRAITS IN DAIRY CATTLE**

VIÇOSA – MINAS GERAIS

2020

HUGO TEIXEIRA SILVA

**AUTOREGRESSIVE MODEL FOR GENETIC EVALUATION OF REPRODUCTIVE
TRAITS IN DAIRY 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

2020

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

T

S586a
2020 Silva, Hugo Teixeira, 1987-
Autoregressive model for genetic evaluation of reproductive
traits in dairy cattle / Hugo Teixeira Silva. – Viçosa, MG, 2020.
88 f. : il. ; 29 cm.

Orientador: Paulo Sávio Lopes.
Tese (doutorado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. Genômica. 2. Holstein (Bovino). 3. Melhoramento
genético. 4. Reprodução. I. Universidade Federal de Viçosa.
Departamento de Zootecnia. Programa de Pós-Graduação em
Zootecnia. II. Título.

CDD 22 ed. 636.082


HUGO TEIXEIRA SILVA

**AUTOREGRESSIVE MODEL FOR GENETIC EVALUATION OF REPRODUCTIVE
TRAITS IN DAIRY 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: February 20th, 2020.

Assent:



Hugo Teixeira Silva
Author



Paulo Sávio Lopes
Adviser

I dedicate this work to my parents José Raimundo and Terezinha. You are an example of strength and perseverance

ACKNOWLEDGMENTS

I believe that all the knowledge we acquire and building up during our academic life is not only a consequence of our individual efforts, but also of the people who were at our side and who somehow contributed to overcoming each stage of our trajectory. To each of you who participated in my journey during my PhD, I express my most sincere thanks.

My advisor Paulo Sávio Lopes, thank you for all support that you give me since my undergrad. All my achievements obtained during my academic journey were consequences from opportunities that you gave me. I am very proud of being your student. Thank you very much for everything that you have done for me.

My co-advisor Júlio Carvalheira, thank you for having received me in Portugal and providing all necessary tools for research executing. Your guidance and teaching were essential for my PhD and I will take it with me for my entire life. Thank you for your patience, friendship, advice, and the opportunity to work with you.

My co-advisor Cláudio Napolis Costa, thank you for the discussions, teachings, friendship, suggestions, and opportunities given to me. Your support was essential for the execution and completion of my research.

My co-advisor Fabyano Fonseca e Silva, Thank you for all help, availability, teachings, friendship, suggestions, and opportunities given to me. Our discussions have always inspired me to look for something more and not be stuck in the sameness.

My girlfriend Aline, thanks for being my companion for so long time. You were at my side, at undergrad, masters and now at the PhD, sharing all moments of joy, but also supporting me and giving me strength to face the difficult ones. Without you I would hardly have managed to get where I am. I love you!

My parents José Raimundo and Terezinha, you have always shown me that with strength, persistence and humility we can always achieve our goals. Thank you for the prayers

and for so many times giving up your dreams to help with mine. I am very lucky to have you as parents.

My friends from animal breeding group: Teo, Hinayah, Karine, Delvan, Alessandra, Ivan, Margareth, Darlene, Edinho, Giovani, Eula, Amanda, Matheus, Talita, Pedro, Tulio, Leticia, Suzana, Layla, Arícia, Ingrid, Daniele and Pâmela. Thanks for conversations, discussions and barbecues.

I am grateful to Universidade Federal de Viçosa, 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 and the technical support. The Embrapa Gado de Leite for the opportunity, theoretical and technical support. support. The Portuguese Dairy Cattle Breeders Association and Brazilian Holstein Cattle Breeders Association for providing the data and to have become this work possible. The Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil (CNPq). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoas de Nível Superior – Brasil (CAPES) – Finance Code 88887.163433/2018-00, and Fundação para a Ciência e a Tecnologia – Portugal (FCT).

“Success is born of willingness, determination and persistence in reaching a goal. Even if they do not hit the target, those who seek and overcome obstacles will at least do admirable things”.

José de Alencar

ABSTRACT

SILVA, Hugo Teixeira, D.Sc., Universidade Federal de Viçosa, February, 2020. **Autoregressive model for genetic evaluation of reproductive traits in dairy cattle.** Adviser: Paulo Sávio Lopes. Co-advisers: Júlio Gil Vale Carvalheira, Fabyano Fonseca e Silva and Cláudio Napolis Costa.

In the last decades, there has been a decrease in the reproductive performance of dairy herds; an unfavorable correlated responses in consequence of the intense selection to increase milk production. To overcome this problem, breeding programs worldwide have included reproductive performance in their selection objectives. Dairy cattle breeding programs in Brazil and Portugal have also shown interest in including reproductive traits in their genetic evaluations; however, further studies are required to better understand about traits and statistical methodologies to be used in the genetic evaluation processes. Genetic analyses of longitudinal reproductive traits have been performed by using the traditional repeatability models (REP). Under these models, the environmental components between longitudinal measurements are assumed equally correlated, which is questionable since events closer in time should be more associated than distant ones. A suitable alternative would be fitting an autoregressive covariance structure to describe the different levels of associations that could exist between consecutive events over time. Therefore, the general objective in this thesis was to evaluate the autoregressive model (AR) for genetic analyses of longitudinal reproductive traits in Brazilian and Portuguese Holstein cattle. Firstly it was evaluated the performance of AR model in Portuguese Holstein herd. The reproductive traits considered were interval between calving to first service (ICF), days open (DO), calving interval (CI), and daughter pregnancy rate (DPR) measured during the first three calving orders. In general, the AR model overcome the REP model in all reproductive traits studied, corresponding with better fitting to reproductive data and higher EBV reliabilities. In the second chapter of this thesis, a similar analysis was carried out with the data of Brazilian Holstein herd. The reproductive traits evaluated were DO, CI and DPR. Once more, the AR model was superior to REP model, showing higher reliabilities and a direct response on the ranking of the best bulls of population. In addition, greater genetic variances and smaller residual variances were estimated with the AR model. The results shown in first two studies suggest that AR model is a suitable alternative for genetic evaluations of longitudinal reproductive traits in Dairy cattle. In the last chapter of this thesis, we investigated the applicability of ssGBLUP methodology under the autoregressive model (H-AR) for

genomic evaluation of longitudinal reproductive traits in Portuguese Holstein cattle. The use of H-AR model increased the GEBV reliabilities and reduced the prediction bias when compared to traditional analysis considering pedigree-based relationship matrix. Furthermore, when evaluating the performance of the H-AR model considering different sources of genomic information, it was observed that the inclusion of genotypic cows in the analysis of reproductive traits provided favorable responses for genomic prediction of bulls. However, improved results were obtained when bulls with no daughter and relationship information (non-contributing bulls) were disregarded in the analyses. Overall, the results suggested that the proposed AR model is a suitable alternative for genetic analyses of longitudinal reproductive traits in the Brazilian and Portuguese dairy cattle herds. In addition, they also demonstrated that ssGBLUP methodology under autoregressive model is a feasible and applicable approach to be used in the genomic analysis of reproductive traits.

Keywords: Autocorrelation. Genomic evaluation. Holstein. Repeatability. Longitudinal traits. ssGBLUP

RESUMO

SILVA, Hugo Teixeira, D.Sc., Universidade Federal de Viçosa, fevereiro de 2020. **Modelo autorregressivo para avaliação genética de características reprodutivas em bovinos de leite.** Orientador: Paulo Sávio Lopes. Coorientadores: Júlio Gil Vale Carvalheira, Fabyano Fonseca e Silva e Cláudio Napolis Costa.

Nas últimas décadas têm-se observado a queda na performance reprodutiva dos rebanhos leiteiros; uma consequência das respostas correlacionadas desfavoráveis em decorrência da intensa seleção para aumentar a produção leite. Para contornar este problema, programas de melhoramento genético de bovinos leiteiro no mundo todo tem incluído características reprodutivas em seus objetivos de seleção. Os programas de melhoramento de bovinos de leite do Brasil e de Portugal também têm apresentado interesse em incluir características reprodutivas nas avaliações genéticas. No entanto, ainda são necessários mais estudos em relação as características e metodologias estatísticas a serem implementadas nos processos de avaliação. As análises genéticas envolvendo características reprodutivas longitudinais têm sido realizadas a partir do modelo de repetibilidade tradicional (REP), que assume que os componentes ambientais associados às medidas longitudinais de um animal são igualmente correlacionadas, o que é questionável, visto que espera-se que medidas consecutivas sejam mais associadas do que medidas mais distantes. Uma alternativa seria ajustar uma estrutura de covariância autorregressiva para descrever os diferentes níveis de associações que podem existir entre eventos consecutivos ao longo do tempo. Neste sentido, o objetivo geral nesta tese foi avaliar o modelo autorregressivo (AR) para a análise genética de características reprodutivas longitudinais nos rebanhos de bovinos da raça Holstein do Brasil e de Portugal. Primeiramente foi avaliada a performance do modelo AR na população de bovinos da raça Holandesa de Portugal. As características reprodutivas consideradas foram intervalo entre o parto e o primeiro serviço (IPS), dias em aberto (DA), intervalo de parto (IP) e taxa de prenhes das filhas (TPF), medidas durante as três primeiras ordens de parto. De uma forma geral, o modelo AR foi superior ao modelo REP nas análises de todas características avaliadas, apresentando melhor ajuste aos dados reprodutivos além de maiores confiabilidades para os valores genéticos dos touros. No segundo capítulo desta tese foi realizado uma avaliação similar na população de bovinos da raça Holandesa do Brasil. As características avaliadas foram apenas DA, IP e TPF. Mais uma vez o modelo AR mostrou ser superior ao modelo REP, com maiores acurácias e maior impacto na classificação dos melhores touros da população. Os resultados mostrados nos

dois primeiros estudos desta tese sugerem que o modelo AR proposto é uma alternativa viável para a avaliação genética de características reprodutivas longitudinais em bovinos de leite. No último capítulo desta tese foi realizado um estudo para avaliar a aplicabilidade do método ssGLUBUP associado ao modelo autorregressivo (H-AR) para a avaliação genômica de características reprodutivas em bovinos da raça Holandesa de Portugal. O modelo H-AR apresentou maiores estimativas de confiabilidade para os valores genéticos dos touros quando comparado com as estimativas obtidas a partir do modelo autorregressivo tradicional (A-AR). As análises de validação também foram favoráveis para o modelo H-AR, correspondendo a menores níveis de viés e maiores estimativas de confiabilidade. Além disso, ao avaliar o desempenho do modelo H-AR considerando diferentes fontes de informação genômica, observou-se que a inclusão de vacas genotípicas nas análises das características reprodutivas proporcionou repostas favoráveis para a predição genômica dos touros. No entanto, melhores resultados foram obtidos quando os touros sem informação de filhas ou parentesco foram desconsiderados das análises. Em geral, os resultados apresentados nos dois primeiros estudos desta tese sugerem que o modelo AR é uma alternativa adequada para as análises genéticas das características reprodutivas longitudinais de bovinos da raça Holandesa do Brasil e de Portugal. Além disso, os resultados apresentados no terceiro trabalho demonstraram que a metodologia ssGBLUP associada ao modelo autorregressivo é uma abordagem adequada e viável para ser utilizada nas análises genômicas das características.

Palavras-chave: Avaliação genômica. Autocorrelação. Características longitudinais. Holstein. Repetibilidade. ssGBLUP.

SUMMARY

GENERAL INTRODUCTION	13
References.....	16
CHAPTER 1 - Autoregressive repeatability model for genetic evaluation of longitudinal reproductive traits in dairy cattle	19
1.1. Abstract.....	20
1.2. Introduction.....	21
1.3. Material and methods.....	22
1.4. Results and discussion.....	28
1.5. Conclusion.....	38
1.6. Acknowledgments.....	38
1.7. References.....	39
1.8. Supplementary tables.....	42
CHAPTER 2 - Autoregressive single-step model for genomic evaluation of longitudinal reproductive traits in Portuguese Holstein cattle	43
2.1. Abstract.....	43
2.2. Introduction.....	44
2.3. Material and methods.....	46
2.4. Results.....	53
2.5. Discussion.....	58
2.6. Conclusion.....	63
2.7. Acknowledgments.....	63
2.8. References.....	64
CHAPTER 3 - Autoregressive model for genetic evaluation of longitudinal reproductive traits in Brazilian Holstein cattle	69

3.1. Abstract.....	69
3.2. Introduction.....	70
3.3. Material and methods.....	72
3.4. Results and discussion.....	76
3.5. Conclusion.....	83
3.6. Acknowledgments.....	84
3.7. References.....	84
GENERAL CONCLUSION.....	88

GENERAL INTRODUCTION

In the last hundred years, dairy cattle have undergone a major transformation in order to increase milk production. During this period, the annual milk production per cow has practically doubled, while the size of the herds decreased (Miglior et al., 2017). According with NDFS (2015), the average number of cows managed or volume of milk sold per dairy worker has increased over time and that is particularly true in larger than in smaller dairy herds. The progress in milk production achieved over the years was especially due to genetic selection (Miglior et al., 2017). In this sense, dairy cattle breeding programs worldwide have placed emphasis on production traits. However, due the absence of any direct selection on fertility, linked to intense selection to increase milk yield, there has been a downward genetic trend in fertility associated with selection for yield (Lucy, 2001; Royal et al., 2000). For instance, successful conception in Holstein cows requires approximately 50 days longer than it used to be 50 years ago (VanRaden et al., 2004).

The decrease on reproductive performance of dairy herds has become one of the most important issues faced by modern dairy farming due to extra costs with inseminations, reproductive biotechnologies, and losses attributed to reproductive problems related to involuntary culling (Chiumia et al. 2013; Tiezzi et al. 2012). To mitigate this problems, the dairy breeding programs have been incorporating reproductive traits in different selection indexes. Overall, reproductive traits tend to exhibit low heritability estimates. However, studies have shown the existence of genetic variability within and across breeds, indicating the possibility of improving without substantial genetic gain changes for milk production (Lucy, 2001; Cochran et al., 2013). Currently, the decreasing trend in reproductive traits has been stabilized and reversed in some dairy herds (VanRaden et al., 2004).

In dairy cattle, the genetic evaluations of longitudinal reproductive traits have been widely based on traditional repeatability models (González-Recio *et al.* 2004; VanRaden *et al.* 2004), which assume that the same groups of genes are influencing the expression and/or regulation of these traits during the animals' productive life. Moreover, for an animal, the traditional repeatability model assumes equal environmental correlation between all pairs of records, which is questionable since closer events over time might be more associated than distant ones. An alternative to traditional repeatability model would be to fit an autoregressive covariance structure on the modeling of animal environmental effects. This covariance structure would allow describing the time decaying correlation between consecutive events over time, representing a more realistic approach to deal with the environmental effects associated with the longitudinal records (Quaas, 1984).

Autoregressive test-day models have been used with success for genetic analyses of production traits in dairy cattle (Carvalho *et al.*, 2002). The animals' permanent environmental effects are assumed to follow a first-order autoregressive process between calving (long-term effect) and between test-days within lactations (short-term effect), taking into account the non-genetic correlations due to the repeated performance of cows. According to Carvalho *et al.* (1998) and Carvalho *et al.* (2002), the autocorrelation structure is effective in removing influences that, if ignored, remained partly confounded with the genetic component. Currently, the autoregressive model is used for genetic analyses of production traits in the Portuguese dairy breeding program. In a recent study, Silva *et al.* (2019) also reported the applicability of autoregressive models for genetic analyses in Brazilian dairy herds. However, studies involving this modeling to genetic evaluation of longitudinal reproductive traits are still scarce.

The use of genomic information has become a routine approach in the processes involved with the genetic evaluation of domestic animals. Currently, genomic selection has

become a standard procedure in dairy cattle breeding given its potential to increase selection accuracy and to reduce generation interval (VanRaden, 2008). Most studies addressing genomic selection for reproductive traits have been based on multi-step approach, which in summary, consists in calculating pseudo-phenotypes from the EBV obtained by the traditional genetic evaluation based on pedigree and phenotypic information; to estimate SNP effects to obtain direct genomic values (**DGV**) for genotyped animals. Finalizing, the blending of DGV and parent averages are obtained to predict the genomic breeding values (**GEBV**) (Hayes et al., 2009; VanRaden et al., 2009). However, this approach has shown disadvantages due the loss of information during the process of pseudo-phenotypes estimation. The contribution of genomic information is here reduced due to bias and errors that may be introduced into the evaluations (Legarra et al., 2014). In addition, when the genotyped population small, bias can also be introduced by the fact that not only all information used for selection decisions is accounted for (Vitezica et al., 2011; Guarini et al., 2018). To overcome these problems, the single-step genomic BLUP methodology (ssGBLUP) was proposed, in which pedigree, genotypic and phenotypic information are combined into a single evaluation. Studies with different purpose have shown the applicability of ssGBLUP methodology to genomic evaluation of production and reproductive traits in dairy cattle (Gao et al., 2012; Koivula et al., 2015; Masuda et al., 2018; Silva et al., 2019).

The general objective with this thesis was to evaluate the use autoregressive model (AR) for genetic analyses of longitudinal reproductive traits in Holstein cattle. In addition, the specific objectives were to: i) evaluate the repeatability autoregressive model for genetic evaluation of longitudinal reproductive traits in Portuguese Holstein cattle and compare with the conventional repeatability models; ii) evaluate the repeatability autoregressive model for genetic evaluation of longitudinal reproductive traits in Brazilian Holstein cattle; iii) evaluate

the applicability of ssGBLUP methodology under the autoregressive model for genomic evaluation of longitudinal reproductive traits in Portuguese Holstein cattle.

References

- Carvalho, J. G. V., 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. *Journal of dairy science*, *81*(10), 2738-2751.
- Carvalho, J., Pollak, E. J., Quaas, R. L., & Blake, R. W. (2002). An autoregressive repeatability animal model for test-day records in multiple lactations. *Journal of dairy science*, *85*(8), 2040-2045.
- Chiumia, D., Chagunda, M. G., Macrae, A. I., & Roberts, D. J. (2013). Predisposing factors for involuntary culling in Holstein–Friesian dairy cows. *Journal of dairy research*, *80*(1), 45-50.
- Cochran, S. D., Cole, J. B., Null, D. J., & Hansen, P. J. (2013). Discovery of single nucleotide polymorphisms in candidate genes associated with fertility and production traits in Holstein cattle. *BMC genetics*, *14*(1), 49.
- Gao, H., Christensen, O. F., Madsen, P., Nielsen, U. S., Zhang, Y., Lund, M. S., & Su, G. (2012). Comparison on genomic predictions using three GBLUP methods and two single-step blending methods in the Nordic Holstein population. *Genetics Selection Evolution*, *44*(1), 8.
- González-Recio, O., Pérez-Cabal, M. A., & Alenda, R. (2004). Economic value of female fertility and its relationship with profit in Spanish dairy cattle. *Journal of dairy science*, *87*(9), 3053-3061

- Hayes, B. J., Bowman, P. J., Chamberlain, A. J., & Goddard, M. E. (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of dairy science*, *92*(2), 433-443.
- Koivula, M., Strandén, I., Pösö, J., Aamand, G. P., & Mäntysaari, E. A. (2015). Single-step genomic evaluation using multitrait random regression model and test-day data. *Journal of dairy science*, *98*(4), 2775-2784.
- Legarra, A., Christensen, O. F., Aguilar, I., & Misztal, I. (2014). Single Step, a general approach for genomic selection. *Livestock Science*, *166*, 54-65.
- Lucy, M. C. (2001). Reproductive loss in high-producing dairy cattle: where will it end?. *Journal of dairy science*, *84*(6), 1277-1293.
- Masuda, Y., VanRaden, P. M., Misztal, I., & Lawlor, T. J. (2018). Differing genetic trend estimates from traditional and genomic evaluations of genotyped animals as evidence of preselection bias in US Holsteins. *Journal of dairy science*, *101*(6), 5194-5206.
- Miglior, F., Fleming, A., Malchiodi, F., Brito, L. F., Martin, P., & Baes, C. F. (2017). A 100-Year Review: Identification and genetic selection of economically important traits in dairy cattle. *Journal of dairy science*, *100*(12), 10251-10271.
- NDFS (Northeast Dairy Farm Study). (2015). Farm Credit East. C. Laughton, ed. Accessed March 10, 2017. [http://www.dairychallenge.org/pdfs/2016 Northeast/ 2015 DFS Final .pdf](http://www.dairychallenge.org/pdfs/2016%20Northeast/2015%20DFS%20Final.pdf).
- Quaas, R. L., Anderson, R. D., & Gilmour, A. R. (1984). *BLUP school handbook. In: Use of mixed models for prediction and for estimation of (co) variance components*. Animal Genetics and Breeding Unit, University of New England, New South Wales, Australia.
- Royal, M. D., Darwash, A. O., Flint, A. P. F., Webb, R., Woolliams, J. A., & Lamming, G. E. (2000). Declining fertility in dairy cattle: changes in traditional and endocrine parameters of fertility. *Animal science*, *70*(3), 487-501.

- Silva, A. A., Silva, D. A., Silva, F. F., Costa, C. N., Lopes, P. S., Caetano, A. R., ... & Carvalheira, J. (2019). Autoregressive single-step test-day model for genomic evaluations of Portuguese Holstein cattle. *Journal of dairy science*.
- Silva, D. A., Costa, C. N., da Silva, A. A., e Silva, F. F., Lopes, P. S., dos Santos, G. G., ... & Carvalheira, J. (2019). Unknown parent and contemporary groups for genetic evaluation of Brazilian Holstein using autoregressive test-day models. *Livestock science*, 220, 1-7.
- Tiezzi, F., Maltecca, C., Cecchinato, A., Penasa, M., & Bittante, G. (2012). Genetic parameters for fertility of dairy heifers and cows at different parities and relationships with production traits in first lactation. *Journal of dairy science*, 95(12), 7355-7362.
- VanRaden, P. M., Sanders, A. H., Tooker, M. E., Miller, R. H., Norman, H. D., Kuhn, M. T., & Wiggans, G. R. (2004). Development of a national genetic evaluation for cow fertility. *Journal of dairy science*, 87(7), 2285-2292.
- VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of dairy science*, 91(11), 4414-4423.
- VanRaden, P. M., Van Tassell, C. P., Wiggans, G. R., Sonstegard, T. S., Schnabel, R. D., Taylor, J. F., & Schenkel, F. S. (2009). Invited review: Reliability of genomic predictions for North American Holstein bulls. *Journal of dairy science*, 92(1), 16-24.

CHAPTER 1

Autoregressive repeatability model for genetic evaluation of longitudinal reproductive traits in dairy cattle

Hugo T. Silva¹, Paulo S. Lopes¹, Claudio N. Costa², Fabyano F. Silva¹, Delvan A. Silva¹,
Alessandra A. Silva¹, Gertrude Thompson^{3,4}, Júlio Carvalheira^{3,4*}

¹Department of Animal Science, Federal University of Viçosa, Viçosa 36570-000, Brazil

²Embrapa Dairy Cattle, Juiz de Fora 36.038-330, Brazil

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

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

Published in Journal of Dairy research, doi: 10.1017/S0022029919000931

1.1. Abstract

We aimed to evaluate the efficiency of the autoregressive repeatability model (AR) for genetic evaluation of longitudinal reproductive traits in Portuguese Holstein cattle and compare the results with those from conventional repeatability model (REP). The data set comprised records taken during the first four calving orders, corresponding to a total of 416, 766, 872 and 766 thousand records for interval between calving to first service, days open, calving interval and daughter pregnancy rate, respectively. Both models included fixed (month and age classes associated to each calving order) and random (herd-year-season, animal and permanent environmental) effects. For AR model, a first-order autoregressive (co)variance structure was fitted for the herd-year-season and permanent environmental effects. The AR outperformed the REP model, with lower Akaike Information Criteria, lower Mean Square Error and Akaike Weights close to unity. Rank correlations between estimated breeding values (EBV) with AR and REP models ranged from 0.95 to 0.97 for all studied reproductive traits, when the total bulls were considered. When considering only the top-100 selected bulls, the rank correlation ranged from 0.72 to 0.88. These results indicate that the re-ranking observed at the top level will provide more opportunities for selecting the best bulls. The EBV reliabilities provided by AR model was larger for all traits, but the magnitudes of the annual genetic progress were similar between two models. Overall, the proposed AR model was suitable for genetic evaluations of longitudinal reproductive traits in dairy cattle, outperforming the REP model.

Keywords: Autocorrelation; Holstein cattle; model evaluation; reproduction performance.

1.2. Introduction

Until recently, the effectiveness of dairy cattle breeding programs were inferred through genetic gains observed in the production traits (Stevenson & Britt, 2017). However, due to unfavorable genetic correlations, continuous selection for increased milk production led to a significant decrease in herd fertility, resulting in negative economic impacts attributed to the increase in reproductive failures, involuntary culling of cows and increased extra costs with reproductive biotechnologies (Sun *et al.* 2010; Tiezzi *et al.* 2012; Chiumia *et al.* 2013). In this context, the number of studies evaluating different classes of reproductive traits (Berry *et al.* 2014; Averill *et al.* 2006) has increased with special emphasis to interval traits, since they address the efficiency of a cow in becoming pregnant or the start of a new reproduction cycle after calving, besides providing early information for genetic evaluations (Jorjani, 2007; Gernand & König, 2017).

Genetic evaluations of reproduction interval traits have been widely based on repeatability models (González-Recio *et al.* 2004; VanRaden *et al.* 2004), which assume that the same groups of genes are influencing the expression and/or regulation of these traits during the animals' productive live. Under these models, the environmental components between longitudinal measurements are assumed equally correlated, which is questionable since events closer in time should be more associated than distant ones. A suitable approach to overcome this problem may be fitting an autoregressive covariance structure to describe the time decaying correlation between consecutive events over time. This methodology has been successfully applied on genetic evaluations of milk yield and its components, where its efficiency in disentangling the genetic components resulted in higher heritabilities, higher reliabilities for estimated breeding values (EBV) and lower residual variances (Carvalheira *et al.*, 1998; Carvalheira *et al.*, 2002; Silva *et al.*, 2019b). Although the positive results obtained in

longitudinal production traits, to the best of our knowledge there are no reports on its application on genetic modeling of longitudinal reproductive traits.

Including reproductive traits in designing selection indexes is essential for the advancement of dairy breeding programs in order to mitigate losses due to diminishing fertility. In this study, we hypothesized that the autoregressive repeatability model (AR) will provide more accurate genetic evaluations to achieve that goal. Therefore, we aimed to evaluate the efficiency of AR models for genetic evaluation of longitudinal reproductive traits in Portuguese Holstein cattle and compare those results with conventional repeatability models (REP).

1.3. Material and methods

Data

The data was provided by the Portuguese Dairy Cattle Breeders Association and comprised pedigree, management and reproduction records from the first three calvings of cows that occurred between 1995 and 2018 in 3,457 herds. Preliminary analysis indicated that the voluntary post-partum waiting period (VWP) corresponded to approximately two estrous cycles (42 days) and for the purpose of this study, we assumed a standard gestation length of 280 ± 15 days.

The interval reproductive traits were calculated as the difference in days between the date of calving and the subsequent event under consideration: interval from calving to first service (ICF), expressed as number of days between the calving and the following first insemination; days open (DO), defined as number of days between calving and conception; calving interval (CI), computed as number of days between two consecutive calvings. Daughter pregnancy rate (DPR) was also evaluated and computed as the linear transformation of the percentage of estrous cycles (of 21 days) that the cow needed to get pregnant. DPR was calculated as $DPR = 0.25(226-DO)$. This formula is an adaptation from the method presented

by VanRaden et al. (2004), taking into account the specific VWP estimated for the Portuguese dairy population, while maintaining the relationship in which each increase of 1% in EBV for DPR equals a decrease of 4 days in EBV for DO.

The total of 495,516 records for ICF, 906,038 for DO and 907,014 for CI, were edited according to trait specifications ensuring that the records of each trait stayed within physiological limits and reliable management criteria. The following minimum and maximum limits were established: 42 to 168 days for ICF, 42 to 252 days for DO and 307 to 730 days for CI. Additionally, ICF and DO values between 21 and 41 days were set to 42 days in order to avoid penalize early-breeder cows, which corresponded to 3.8% and 2.4% of total data for ICF and DO, respectively. All fertile insemination dates were validated by requiring that the number of days to the following calving should be in the range of the assumed gestation length. When the insemination date between two consecutive calving was not available, the DO was assumed equal to CI minus 280 days, the mean gestation period of the population. Finally, it was also required that the cows' data had no missing information between consecutive calving orders. The descriptive statistics for all reproductive traits are presented in Table 1.

Table 1. Phenotypic descriptive statistics for interval between calving to first service (ICF), days open (DO), calving interval (CI) and daughter pregnancy rate (DPR)

Traits	Calving order	N	Mean	SD	Number of cows
ICF (days)	1	165,209	87.42	31.40	
	2	156,265	87.71	31.50	
	3	95,252	88.84	31.30	
	Total	416,726	87.35	31.42	228,937
DO (days)	1	350,206	119.57	53.76	
	2	262,433	121.85	53.64	
	3	154,285	121.99	52.87	
	Total	766,897	120.84	53.55	404,925
CI (days)	1	396,376	426.47	86.66	
	2	293,751	426.21	82.97	
	3	182,476	426.68	81.85	
	Total	872,603	426.43	84.44	431,859
DPR (%)	1	350,206	26.01	13.44	
	2	262,433	26.04	13.41	
	3	154,285	26.00	13.22	
	Total	766,897	26.29	13.39	404,925

Statistical analysis

In a preliminary analysis, we investigated if the observations obtained from each calving order for each trait (ICF, DO, CI and DPR) could be considered as repeated records of the same trait. To perform this study, we applied a multiple trait model considering each calving order as a different trait within ICF, DO, CI and DPR, respectively. As expected, the genetic correlations between calving orders were high for all reproductive traits (Supplementary Table S1), justifying the use of repeatability models for genetic evaluations.

Analysis based on AR and REP models were carried out using the following general model:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Kc} + \mathbf{Za} + \mathbf{Wpe} + \mathbf{e}, \quad (1)$$

where \mathbf{y} is the vector of observations for the first, second and third calving orders; \mathbf{b} is the vector of fixed effects considering twelve month levels and sixteen age-month levels associated to each calving order, which ranged from 18 to 44, 28 to 60 and 40 to 72 months for the first, second and third calving orders, respectively; \mathbf{c} is a vector of random herd-year-season (HYS) of calving effects, corresponding to 73,472 levels for ICF, 137,433 levels for DO and DPR and 141,131 levels for CI; \mathbf{a} is the random additive genetic effect; \mathbf{pe} is a vector of random permanent environmental (PE) effects; \mathbf{e} is a vector of residuals. \mathbf{X} , \mathbf{K} , \mathbf{Z} and \mathbf{W} are the incidence matrices relating records to fixed effects and to random effects of HYS, additive genetic and PE, respectively. The expectations and (co)variance for these models were:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{c} \\ \mathbf{a} \\ \mathbf{pe} \\ \mathbf{e} \end{bmatrix} \sim N \left(\begin{bmatrix} \mathbf{Xb} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}; \begin{bmatrix} \mathbf{V} & \mathbf{JK} & \mathbf{GZ} & \mathbf{PW} & \mathbf{R} \\ & \mathbf{J} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ & & \mathbf{G} & \mathbf{0} & \mathbf{0} \\ & & & \mathbf{P} & \mathbf{0} \\ & & & & \mathbf{R} \end{bmatrix} \right). \quad (2)$$

The differences between AR and REP models are on the modeling of the HYS and PE (\mathbf{J} and \mathbf{P} matrices, respectively) where, as the name implies, first-order autocorrelation structures were fitted on the AR model.

$$\text{For AR model, } \mathbf{J} = \sigma_c^2 \begin{bmatrix} 1 & \rho_c & \rho_c^2 \\ & 1 & \rho_c \\ \text{Sym} & & 1 \end{bmatrix} \otimes \mathbf{I}, \text{ is a block diagonal matrix where } \sigma_c^2 \text{ and } \rho_c$$

are the variance and the autocorrelation coefficient for HYS, respectively (the size of the autocorrelation matrix depends on the number of year-season levels within each herd - 3 in this example). \mathbf{I} is an identity matrix with size equal to the number of herds. $\mathbf{P} =$

$$\sigma_{pe}^2 \begin{bmatrix} 1 & \rho_{pe} & \rho_{pe}^2 \\ & 1 & \rho_{pe} \\ \text{Sym} & & 1 \end{bmatrix} \otimes \mathbf{I}, \text{ is also a block diagonal matrix where } \sigma_{pe}^2 \text{ and } \rho_{pe} \text{ are the variance}$$

and the autocorrelation coefficient for PE, respectively (in this case, the size of the

autocorrelation matrix depends on the number of repeated observations for each cow - 3 in this example). \mathbf{I} is an identity matrix with size equal to the number of animals with records.

On the other hand, for REP model, $\mathbf{J} = \mathbf{I}\sigma_c^2$, where \mathbf{I} is an identity matrix with size equal to number of HYS levels; and $\mathbf{P} = \mathbf{I}\sigma_{pe}^2$, where \mathbf{I} is an identity matrix with size equal to number of animals with records.

For both models, $\mathbf{G} = \mathbf{A}\sigma_a^2$, where σ_a^2 is the additive genetic variance component and \mathbf{A} is the numerator relationship matrix; and $\mathbf{R} = \mathbf{I}\sigma_e^2$, where \mathbf{I} is an identity matrix with size equal to the total number of observations and σ_e^2 is the residual variance component. Finally, $\mathbf{V} = \mathbf{KJK}' + \mathbf{ZGZ}' + \mathbf{WPW}' + \mathbf{R}$, represents the phenotypic (co)variance.

To estimate the variance components and autocorrelations, we used six sub-data sets extracted from the complete data for each trait (ICF, DO, CI and DPR). Each sub-data set consisted of 30 randomly sampled herds with records spanning across the all period of study and representing all regions of Portugal (Supplementary Table S2). All parameters were estimated applying DFREML methodology (Smith & Graser, 1986). Likelihood functions were maximized by the multivariate simplex algorithm (Nelder and Mead, 1965) programmed in MATLAB (Math Works Inc., Natick MA). To achieve convergence, the variance among the simplex $-2\log$ likelihoods had to be less than 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). This process was repeated for each sub-data set and the respective means were used as the variance components and autocorrelations estimates used in all subsequent analyses of each reproductive trait.

Model comparisons

Comparisons between AR and REP models were carried out on each of the six sub data for each trait. The best fit was evaluated by Akaike Information Criterion (AIC), computed as

AIC = $-2 \log(L) + 2p$, where L is the maximum of likelihood function, and p is the number of effective parameters. Since, AIC provides only a qualitative comparison between models (smallest AIC is better), we also calculated the Akaike Weights (AW), presented by Burnham & Anderson (2004), using the formula $AW_i = \exp\left(-\frac{\Delta_i}{2}\right) / \sum_{i=1}^2 \exp\left(-\frac{\Delta_i}{2}\right)$, for $i = 1$ (AR) and 2 (REP). In this formula, Δ_i is the AIC difference between model i and the model that presented the smallest AIC value (the best model). The Δ_i for the best model equals zero (since i and the best model will coincide). With this approach, AW_i represents the probability of model i to be better than the other model evaluated. Additionally, we also evaluated the Mean Square Error (MSE) for each model, calculated as $MSE = n^{-1} \sum (y - \hat{y})^2$, where y and \hat{y} are the observed and predicted vectors for each trait respectively, and n is the number of observations.

The EBVs obtained from AR and REP models were used to evaluate the changes in the ranking of sires using the Spearman's rank correlation among the top 100 and the total sire population. Complementarily, the coincidence percentage between the best 1%, 10% and 50% of bulls was also used for comparisons. The significance of the difference between EBV reliabilities of these bulls, obtained using the Student's t-test at the 5% level of significance, was another criterion of comparison.

Additionally, plots of genetic trends based on the EBV means at year of birth and the respective genetic gain, derived by regressing yearly means for EBV, were used for comparisons of the potential genetic progress provided by the two models. The average EBV of cows born in 2010 was set as the genetic base-year for each trait. Only bulls with at least 10 daughters recorded in at least five distinct herds were considered in these analyses.

1.4. Results and discussion

Model-fitting criteria

The results for model-fitting criteria in the analysis of ICF, DO, CI and DPR are presented in Tables 2 and 3. The AIC estimates for AR model were lower for all reproductive traits, indicating a better fitting when compared to REP model. The AW estimates for AR model were all close to one while for the REP models these values were closed to zero. The MSE estimates were lower for AR, corresponding to the least amount of bias, and therefore, higher predictive ability.

Lower AIC values, AW probability close to unity and the lower MSE estimates are strong indications of the superiority of AR model in fitting (time) reproductive traits. Modeling the HYS and the PE effects as autoregressive variables did contribute to a better explanation of the variation observed in ICF, DO, CI and DPR. These findings are in line with the reports of Quaas *et al.* (1984) and Carvalheira *et al.* (2002), where they suggested relaxing the usual assumption of unitary correlation between repeated events by implementing first-order autoregressive structures, which may result in larger additive genetic variances and heritabilities and estimates of individual genetic merit more accurate for longitudinal traits.

Table 2. Akaike Information Criterion (AIC) and Akaike Weights (AW) for the Autoregressive repeatability model (AR) and the traditional repeatability model (REP) from the analyses of six sub data of interval between calving to first service (ICF), days open (DO), calving interval (CI) and daughter pregnancy rate (DPR)

Traits	Sub data	AIC_{REP}	AIC_{AR}	AW_{REP}	AW_{AR}
ICF	1	152,194.79	151,909.44	≈ 0	≈ 1
	2	151,301.85	151,168.82	≈ 0	≈ 1
	3	146,529.06	146,356.67	≈ 0	≈ 1
	4	158,615.14	158,317.99	≈ 0	≈ 1
	5	131,498.71	131,252.86	≈ 0	≈ 1
	6	165,648.23	165,404.17	≈ 0	≈ 1
DO	1	276,276.45	275,958.92	≈ 0	≈ 1
	2	214,765.49	214,632.17	≈ 0	≈ 1
	3	290,234.06	290,042.47	≈ 0	≈ 1
	4	283,179.4	282,805.49	≈ 0	≈ 1
	5	247,111.05	247,090.51	≈ 0	≈ 1
	6	199,326.89	199,143.32	≈ 0	≈ 1
CI	1	316,149.88	315,666.29	≈ 0	≈ 1
	2	285,674.55	285,156.53	≈ 0	≈ 1
	3	279,907.48	279,345.11	≈ 0	≈ 1
	4	280,756.82	280,321.47	≈ 0	≈ 1
	5	311,376.86	310,920.39	≈ 0	≈ 1
	6	359,830.31	359,274.86	≈ 0	≈ 1
DPR	1	151,463.69	151,190.39	≈ 0	≈ 1
	2	180,739.68	180,578.75	≈ 0	≈ 1
	3	171,980.99	171,679.61	≈ 0	≈ 1
	4	161,608.04	161,430.39	≈ 0	≈ 1
	5	161,632.33	161,485.40	≈ 0	≈ 1
	6	196,817.01	196,632.79	≈ 0	≈ 1

Table 3. Mean Square Error (MSE) for the autoregressive repeatability model (AR) and the traditional repeatability model (REP) from the analyses of interval between calving to first service (ICF), days open (DO), calving interval (CI) and daughter pregnancy rate (DPR)

Model	ICF	DO	CI	DPR
REP	5024.27	2164.45	5024.27	136.09
AR	4588.23	2029.87	4588.23	126.55

Rank correlation, percentage of coincidence and reliabilities of EBVs for bulls

Re-ranking between the breeding values predicted by AR and REP models were relatively small when considering the total population of bulls in all traits (rank correlation varied between 0.95 and 0.97 as shown in Figure 1). On the other hand, for the top-100 bulls the rank correlation were significantly lower (no overlapping of 95% confidence intervals was observed) and ranged from 0.71 to 0.88 (Figure 1). These results are consistent with those presented by Sawalha *et al.* (2005), which reported rank correlation estimates ranging from 0.71 to 0.87 for top animals of population when evaluating the inclusion of autoregressive covariance structure in the test-day models for analyses of productive traits in dairy cattle. The moderate rank correlation estimates for the top 100 bulls may indicate a different ranking when considering one or another model, which may lead to changes in the choice of bulls by the farmers depending on the model used (Sawalha *et al.*, 2005; Sun *et al.*, 2010). In this sense, the evaluation of reproductive efficiency using AR model could create more opportunities for selecting the best bulls and, therefore, promoting greater genetic progress.

To quantify the EBV re-ranking between the two models, we also computed the percentage of coincidence in the top 1%, 10% and 50% of the bull population for each trait (Figure 1). The percentage of coincidence increased with the increase in the number of bulls compared (from 80% to 84.6% in the top 1% and from 90% to 93% in the top 50%). The high percentage of coincidence observed indicates that both models captured almost the same top

animals, which emphasizes the importance of the reliabilities of EBV predictions as an auxiliary tool to increase the selection pressure.

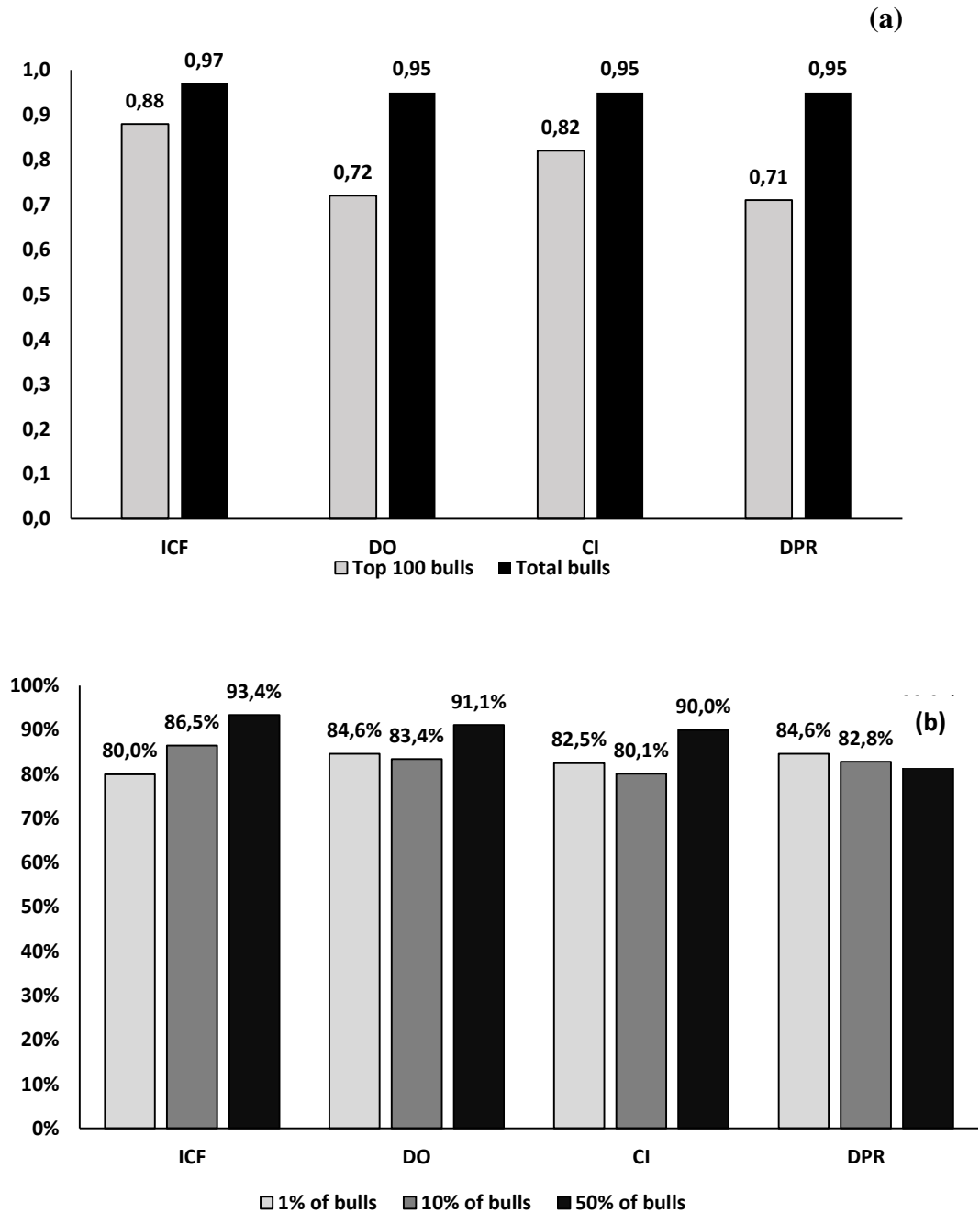


Figure 1. Rank correlation for Top 100 and Total population of bulls (a) and percentage of coincidence between the best 1%, 10% and 50% of bulls (b) ranked according to solutions predicted by the Autoregressive repeatability model (AR) and the traditional repeatability

model (REP) models for interval between calving to first service (ICF), days open (DO), calving interval (CI) and daughter pregnancy rate (DPR).

The bulls' EBV reliabilities predicted by REP and AR models were significantly different ($P < 0.05$) in all traits evaluated, favoring the AR (Table 4). According to Chegini *et al.* (2019), when genetic values are predicted with higher accuracy and precision, the greater will be the selection differential and, consequently, a higher response to selection are expected. Thus, the results presented in our study indicate that greater genetic progress may occur with the use of EBVs predicted by the AR model in the genetic evaluations of the reproductive traits evaluated.

Table 4. Means of EBV reliabilities (Rel) for the best 1%, 10% and 50% of bulls obtained from the autoregressive repeatability model (AR) and the traditional repeatability model (REP) for interval between calving to first service (ICF), days open (DO), calving interval (CI) and daughter pregnancy rate (DPR)

Proportions	ICF		DO		CI		DPR	
	Rel _{AR}	Rel _{REP}	Rel _{AR}	Rel _{REP}	Rel _{AR}	Rel _{REP}	Rel _{AR}	Rel _{REP}
1%	0.95*	0.87	0.95*	0.88	0.95*	0.87	0.96*	0.87
10%	0.93*	0.84	0.94*	0.85	0.94*	0.86	0.93*	0.82
50%	0.93*	0.81	0.93*	0.82	0.93*	0.84	0.93*	0.81

*Significant difference ($P < 0.05$) between Rel_{AR} and Rel_{REP} within reproductive trait.

Genetic parameters

The variance components and autocorrelations estimated with AR and REP models for ICF, DO, CI and DPR, are presented in Table 5. In general, the estimates of the variance components were similar between models, with heritabilities ranging from 0.08 to 0.11 for AR

and from 0.06 to 0.09 for REP. In the present study, the HYS was the most auto-correlated effect (ranging from 0.7 to 0.9) probably because of the shorter time lag between events (between seasons) compared with the PE effect where the events are much more distant (between calvings). A similar pattern was reported in autoregressive analysis of production traits (e.g., milk yield) where the autocorrelations between lactation orders (long term effect) were negligible compared with the short term autocorrelations between test-days within lactations (Carvalho *et al.*, 2002; Costa *et al.*, 2009; Silva *et al.*, 2019b).

The heritabilities estimated in our study are in agreement with previous results reported in the literature, which presented values ranging from 0.02 to 0.05 for ICF, from 0.02 to 0.07 for DO and CI and from 0.03 to 0.07 for DPR (Kadarmideen *et al.* 2003; Ghiasi *et al.* 2011; Haile-Mariam *et al.* 2013; Frioni *et al.* 2017; Gernand & König, 2017).

Genetic trends for bulls and cows

The genetic trends of reproductive traits resulting from the REP and AR analysis are depicted in Figure 2, for bulls and cows, respectively. It is possible to identify two distinct periods for each class of animal. The intervals between 1980 and 2000 (in bulls) and from 1980 to 2008 (in cows) are characterized by a drop in reproductive efficiency of the population. However, in the years following those periods, there was a recovery in these traits with a decrease in the EBV means of ICF, DO and CI and an increase for DPR in both, bulls and cows. As expected, the shape and magnitude of the curves are very similar between models, reflecting the high correlation between EBVs. The genetic gains expressed by regressing yearly means for EBV on year of birth were also close between the two models evaluated, as presented in the Table 6.

Table 5. Variance components, autocorrelation coefficients and genetic parameters (\pm SE) estimated by the Autoregressive repeatability model (AR) and the traditional repeatability model (REP) for interval between calving to first service (ICF), days open (DO), calving interval (CI) and daughter pregnancy rate (DPR)

Parameters ^a	AR				REP			
	ICF	DO	CI	DPR	ICF	DO	CI	DPR
σ_a^2	64.2 \pm 8.4	182.4 \pm 7.8	652.3 \pm 24.6	11.8 \pm 0.6	64.5 \pm 4.9	204.7 \pm 12.7	687.0 \pm 38.1	13.1 \pm 0.6
σ_c^2	193.0 \pm 4.5	179.5 \pm 11.1	539.3 \pm 37.4	9.3 \pm 1.0	201.1 \pm 7.4	158.0 \pm 6.6	475.6 \pm 37.0	8.9 \pm 0.4
σ_{pe}^2	31.5 \pm 6.5	64.9 \pm 9.2	213.9 \pm 15.7	4.4 \pm 0.7	26.6 \pm 6.5	57.7 \pm 7.7	216.2 \pm 20.5	4.1 \pm 0.8
σ_e^2	674.1 \pm 14.7	2,479.9 \pm 18.7	6,023.1 \pm 144.8	155.6 \pm 1.2	670.2 \pm 11.9	2,446.8 \pm 16.8	5,961.0 \pm 133.9	153.2 \pm 0.9
ρ_c	0.7 \pm 0.02	0.9 \pm 0.004	0.8 \pm 0.02	0.9 \pm 0.005	-	-	-	-
ρ_{pe}	\approx 0.0 \pm 0.0	\approx 0.0 \pm 0.0	\approx 0.0 \pm 0.0	\approx 0.0 \pm 0.0	-	-	-	-
h^2	0.07 \pm 0.008	0.06 \pm 0.002	0.09 \pm 0.003	0.06 \pm 0.003	0.07 \pm 0.004	0.07 \pm 0.004	0.09 \pm 0.003	0.07 \pm 0.003
r	0.10 \pm 0.001	0.09 \pm 0.003	0.11 \pm 0.005	0.09 \pm 0.006	0.09 \pm 0.007	0.09 \pm 0.002	0.12 \pm 0.004	0.09 \pm 0.003

^a σ_a^2 = additive genetic variance; σ_c^2 = herd-year-season variance; σ_{pe}^2 = permanent environmental variance; σ_e^2 = residual variance;

ρ_c = autocorrelation coefficient for herd-year-season effect; ρ_{pe} = autocorrelation coefficient for permanent environmental effect;

h^2 = heritability; r = repeatability.

Table 6. Genetic gains expressed by regression coefficients (b), with the respective standard deviations and coefficients of determination (R^2) estimated from analyses with the Autoregressive repeatability model (AR) and the traditional repeatability model (REP) for interval between calving to first service (ICF), days open (DO), calving interval (CI) and daughter pregnancy rate (DPR) considering two different year intervals for bulls and cows

		1980 – 2000				2000 - 2012				
		Trait	b_{AR}	R^2_{AR}	b_{REP}	R^2_{REP}	b_{AR}	R^2_{AR}	b_{REP}	R^2_{REP}
Bulls	ICF		$-0.06 \pm 0.058^*$	0.12	$-0.05 \pm 0.063^*$	0.10	$-0.46 \pm 0.076^*$	0.76	$-0.39 \pm 0.076^*$	0.70
	DO		$0.35 \pm 0.076^*$	0.54	$0.20 \pm 0.079^*$	0.26	$-1.18 \pm 0.095^*$	0.93	$-1.29 \pm 0.109^*$	0.92
	CI		$0.40 \pm 0.148^*$	0.47	$0.35 \pm 0.151^*$	0.23	$-2.68 \pm 0.193^*$	0.94	$-2.94 \pm 0.198^*$	0.95
	DPR		$-0.08 \pm 0.019^*$	0.53	$-0.05 \pm 0.020^*$	0.25	$0.30 \pm 0.024^*$	0.93	$0.33 \pm 0.027^*$	0.93
		1980 – 2008				2008 – 2014				
		Trait	b_{AR}	R^2_{AR}	b_{REP}	R^2_{REP}	b_{AR}	R^2_{AR}	b_{REP}	R^2_{REP}
Cows	ICF		$0.09 \pm 0.006^*$	0.87	$0.07 \pm 0.005^*$	0.86	$-0.51 \pm 0.057^*$	0.94	$-0.45 \pm 0.047^*$	0.94
	DO		$0.24 \pm 0.016^*$	0.87	$0.21 \pm 0.013^*$	0.90	$-0.98 \pm 0.137^*$	0.91	$-1.10 \pm 0.149^*$	0.92
	CI		$0.35 \pm 0.031^*$	0.83	$0.30 \pm 0.023^*$	0.85	$-2.09 \pm 0.288^*$	0.91	$-2.57 \pm 0.289^*$	0.94
	DPR		$-0.05 \pm 0.004^*$	0.87	$-0.05 \pm 0.003^*$	0.90	$0.25 \pm 0.035^*$	0.90	$0.29 \pm 0.038^*$	0.92

* $p < 0.05$.

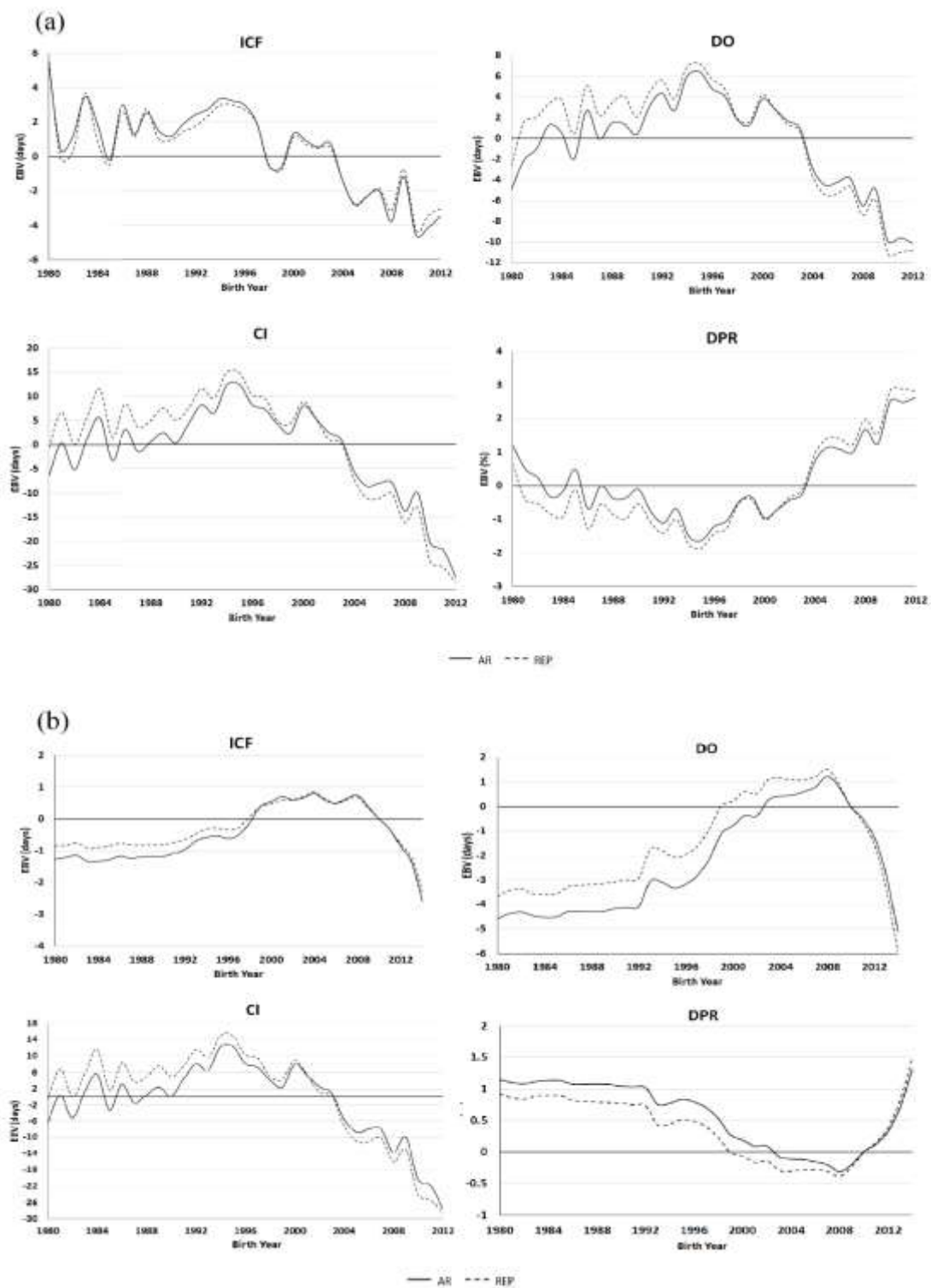


Figure 2. Genetic trends in the bull (a) and cows (b) populations estimated from EBVs obtained by the Autoregressive repeatability model (AR) and the traditional repeatability model (REP) for interval between calving to first service (ICF), days open (DO), calving interval (CI) and daughter pregnancy rate (DPR).

It is worth noting that until the present, the reproductive performance was never included in any selection index for the improvement of dairy cattle in Portugal. This suggests that the genetic trends observed along those years were a result of indirect selection on production traits. In fact, the trends here observed are almost the opposite of those of milk production for the same period (Silva *et al.*, 2019a). As in other countries, the Portuguese dairy breeding program has as its main goal, the increase of milk production and its components. Thus, the positive genetic gains between 1980 and 2000 for bulls, and between 1980 and 2008 for cows, are probably consequences of the genetic progress achieved with selection for improvement of productive traits, emphasizing the negative genetic correlation between these two classes of traits. Similar results were found in the dairy herds of Canada, United States, Denmark and Australia, where improved productive performance also led to a fall in the reproductive potential of herds (VaRaden *et al.*, 2004; Sewalem *et al.*, 2010; Sun *et al.*, 2010; Haile-Mariam *et al.*, 2013).

The negative genetic gains observed from the year of birth of 2000 for bulls (with practical effects 6 to 8 years after) and from the year 2008 for cows (with their first calving about 2010), may be indirectly associated with the economic scenario during that period. Starting in the year 2008-2010, Portugal underwent a strong economic recession, affecting all sectors of the economy, including the importation of genetics (frozen semen and heifers). Silva *et al.* (2019a) reported a drop in the genetic tendencies for productive traits for this period. Therefore, the response observed in the genetic trends for reproductive traits could be actually a correlated response from the selection on the productive traits.

Although the genetic improvement observed in the reproductive traits during the last years, the expectation is that this scenario will not remain given the tendency to prioritize the improvement of productive traits that will end up causing deterioration on the animals' reproductive performance. In general, reproductive traits tend to have low heritability estimates,

indicating a strong influence of management and environmental effects on the respective phenotypic expressions. However, due to their impact on the profitability of dairy farmers, different breeding programs have considered the reproductive traits in the selection process, and genetic gains have been observed with the inclusion of these traits within selection indexes (Gernand e König, 2017). In this sense, the planning of strategies for inclusion of reproductive traits in the genetic evaluation process becomes of great importance for the Portuguese dairy farming, by making possible genetic gains for the productive traits while at the same time softening the negative effects on the reproductive performance.

1.5. Conclusion

In conclusion, results from the application of AR models outperformed the REP models in evaluating time reproductive traits for all statistical fitting criteria providing lower AIC estimates, AW close to unit and lower MSE values. AR model also provided higher EBV reliabilities, which for selection purposes, will permit a better discernment among the best bulls. Thus, we conclude that AR model represents a viable approach for genetic evaluations of longitudinal reproductive traits in dairy cattle, which correspond to more reliable results and consequently, faster genetic progress.

1.6. Acknowledgments

Portuguese Dairy Cattle Breeders Association for providing data for this study. was funded by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES/ FCT, no. 99999.008462/2014-03 and 88887.163433/2018-00), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq 465377/ 2014-9 – PROGRAMA INCT)

1.7. References

- Averill, T., Rekaya, R., & Weigel, K. (2006). Random regression models for male and female fertility evaluation using longitudinal binary data. *Journal of dairy science*, *89*(9), 3681-3689.
- Berry, D. P., Wall, E., & Pryce, J. E. (2014). Genetics and genomics of reproductive performance in dairy and beef cattle. *animal*, *8*(s1), 105-121.
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological methods & research*, *33*(2), 261-304.
- Carvalho, J. G. V., 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. *Journal of dairy science*, *81*(10), 2738-2751.
- Carvalho, J., Pollak, E. J., Quaas, R. L., & Blake, R. W. (2002). An autoregressive repeatability animal model for test-day records in multiple lactations. *Journal of dairy science*, *85*(8), 2040-2045.
- Costa, C. N., Carvalho, J., Cobuci, J. A., Freitas, A. F., & Thompson, G. (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 Journal of Animal Science*, *39*(1), 165-168.
- Chiumia, D., Chagunda, M. G., Macrae, A. I., & Roberts, D. J. (2013). Predisposing factors for involuntary culling in Holstein–Friesian dairy cows. *Journal of dairy research*, *80*(1), 45-50.
- Frioni, N., Rovere, G., Aguilar, I., & Urioste, J. I. (2017). Genetic parameters and correlations between days open and production traits across lactations in pasture based dairy production systems. *Livestock science*, *204*, 104-109.

- Gernand, E., & König, S. (2017). Genetic relationships among female fertility disorders, female fertility traits and productivity of Holstein dairy cows in the early lactation period. *Journal of Animal Breeding and Genetics*, 134(5), 353-363.
- González-Recio, O., Pérez-Cabal, M. A., & Alenda, R. (2004). Economic value of female fertility and its relationship with profit in Spanish dairy cattle. *Journal of dairy science*, 87(9), 3053-3061
- Ghiasi, H., Pakdel, A., Nejati-Javaremi, A., Mehrabani-Yeganeh, H., Honarvar, M., González-Recio, O., ... & Alenda, R. (2011). Genetic variance components for female fertility in Iranian Holstein cows. *Livestock science*, 139(3), 277-280.
- Haile-Mariam, M., Bowman, P. J., & Pryce, J. E. (2013). Genetic analyses of fertility and predictor traits in Holstein herds with low and high mean calving intervals and in Jersey herds. *Journal of dairy science*, 96(1), 655-667.
- Jorjani, H. (2007). International genetic evaluation of female fertility traits in five major breeds. *Interbull Bulletin*, (37), 144-144.
- Kadarmideen, H. N., Thompson, R., Coffey, M. P., & Kossaibati, M. A. (2003). Genetic parameters and evaluations from single-and multiple-trait analysis of dairy cow fertility and milk production. *Livestock Production Science*, 81(2-3), 183-195.
- Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The computer journal*, 7(4), 308-313.
- Quaas, R. L., Anderson, R. D., & Gilmour, A. R. (1984). *BLUP school handbook. In: Use of mixed models for prediction and for estimation of (co) variance components*. Animal Genetics and Breeding Unit, University of New England, New South Wales, Australia.
- Sawalha RM, Keown JF, Kachman SD & Van Vleck LD 2005 Genetic evaluation of dairy cattle with test-day models with autoregressive covariance structures and with a 305-d model. *Journal of Dairy Science* **88** 3346-53.

- Sawalha, R. M., Keown, J. F., Kachman, S. D., & Van Vleck, L. D. (2005). Genetic evaluation of dairy cattle with test-day models with autoregressive covariance structures and with a 305-d model. *Journal of dairy science*, 88(9), 3346-3353.
- Silva, A. A., Silva, D. A., Silva, F. F., Costa, C. N., Lopes, P. S., Caetano, A. R., ... & Carvalheira, J. (2019). Autoregressive single-step test-day model for genomic evaluations of Portuguese Holstein cattle. *Journal of dairy science*.
- Silva, D. A., Costa, C. N., da Silva, A. A., e Silva, F. F., Lopes, P. S., dos Santos, G. G., ... & Carvalheira, J. (2019). Unknown parent and contemporary groups for genetic evaluation of Brazilian Holstein using autoregressive test-day models. *Livestock science*, 220, 1-7.
- Sun, C., Madsen, P., Lund, M. S., Zhang, Y., Nielsen, U. S., & Su, G. (2010). Improvement in genetic evaluation of female fertility in dairy cattle using multiple-trait models including milk production traits. *Journal of animal science*, 88(3), 871-878.
- Smith, S. P., & Graser, H. U. (1986). Estimating variance components in a class of mixed models by restricted maximum likelihood. *Journal of Dairy Science*, 69(4), 1156-1165.
- Stevenson, J. S., & Britt, J. H. (2017). A 100-year review: Practical female reproductive management. *Journal of dairy science*, 100(12), 10292-10313.
- Tiezzi, F., Maltecca, C., Cecchinato, A., Penasa, M., & Bittante, G. (2012). Genetic parameters for fertility of dairy heifers and cows at different parities and relationships with production traits in first lactation. *Journal of dairy science*, 95(12), 7355-7362.
- VanRaden, P. M., Sanders, A. H., Tooker, M. E., Miller, R. H., Norman, H. D., Kuhn, M. T., & Wiggans, G. R. (2004). Development of a national genetic evaluation for cow fertility. *Journal of dairy science*, 87(7), 2285-2292.

1.8. Supplementary tables

Table S1. Means (\pm SE) of genetic correlations between calving orders for interval between calving to first service (ICF), days open (DO), calving interval (CI) and daughter pregnancy rate (DPR) estimated using multiple-trait analyses

Calving orders	Reproductive traits			
	ICF	DO	CI	DPR
1,2	0.81 \pm 0.04	0.88 \pm 0.02	0.90 \pm 0.01	0.90 \pm 0.02
1,3	0.81 \pm 0.05	0.80 \pm 0.03	0.81 \pm 0.07	0.80 \pm 0.03
2,3	0.88 \pm 0.03	0.87 \pm 0.04	0.91 \pm 0.03	0.82 \pm 0.07

Table S2. Number of records and cows in the six sub-data sets used to estimate variance components and autocorrelations of interval between calving to first service (ICF), days open (DO), calving interval (CI) and daughter pregnancy rate (DPR) considering two different year intervals for bulls and cows

Samples	ICF		DO		CI		DPR	
	Records	Cows	Records	Cows	Records	Cows	Records	Cows
1	19,691	11,236	30,000	17,252	32,001	16,217	24,537	13,593
2	19,669	10,865	24,056	13,416	28,720	14,891	29,252	16,202
3	18,862	10,756	32,647	17,916	28,459	14,258	27,983	15,270
4	20,409	11,618	31,678	16,824	28,563	14,487	26,195	14,095
5	17,009	9,878	27,711	14,828	31,753	16,161	26,237	14,338
6	21,536	12,407	22,292	12,192	36,575	18,601	31,948	16,914

CHAPTER 2

Autoregressive single-step model for genomic evaluation of longitudinal reproductive traits in Portuguese Holstein cattle

2.1. Abstract

We investigated the applicability of ssGBLUP methodology under the autoregressive model (H-AR) for genomic evaluation of longitudinal reproductive traits in Portuguese Holstein cattle. The genotype data of 1,230 bulls and 1,645 cows were considered in our study. The reproductive traits evaluated were interval from calving to first service (ICF), calving interval (CI), and daughter pregnancy rate (DPR) measured during the first four calving orders. Reliability and rank correlation were used to compare the H-AR with the traditional pedigree-based autoregressive models (A-AR). In addition, a validation study was performed considering different scenarios. Higher genomic estimated breeding values (GEBV) reliabilities were obtained for genotyped bulls when evaluated under the H-AR model, with emphasis on bulls with less than 9 daughters. For this group, the averages of GEBV reliabilities corresponded to 0.62, 0.69, and 0.62 for ICF, CI, and DPR, respectively, while the averages obtained by the A-AR model were 0.27, 0.15, and 0.16. The validation study was favorable to H-AR. The best results were observed in the scenario where genotyped cows were combined with contributing bulls (genotyped bulls with daughter or relationship information in the population). Overall, the results suggest that ssGBLUP methodology under the autoregressive model is a feasible and applicable approach to be used in genomic analyses of longitudinal reproductive traits in Portuguese Holstein cattle.

Keywords: autocorrelation, dairy cattle; genomic prediction; ssGBLUP

2.2. Introduction

The concept of genomic selection introduced by Meuwissen et al. (2001) arose as a revolutionary proposition for the process involving genetic evaluation in domestic animals. Due to the possibility of greater genetic gains by improving the accuracy of breeding values and reducing the generation interval, genomic selection has been widely implemented in dairy cattle breeding programs (Hayes et al., 2009; Wiggans et al., 2017).

Most methods used in genomic evaluation of dairy cattle are based on the multistep procedure (VanRaden 2008; Lund et al., 2011), which makes use of pseudo-phenotypes (de-regressed estimated breeding values or daughter yield deviation) as variable responses in a prediction model, and direct genomic breeding values (DGV) are estimated for genotyped animals. After this process, DVG are combined with the parent averages (PA) to obtain the genomic estimated breeding values (GEBV). Despite the widespread use of multistep approach, it has some disadvantages attributed to the loss of information during the process of pseudo-phenotypes estimation and the introduction of bias and errors, which reduce the contribution of the genomic information into evaluations (Legarra et al., 2014). Moreover, if selection decisions are not fully accounted as usually occurs in small genotyped populations, bias can also be introduced (Vitezica et al. 2011; Guarini et al., 2018).

An alternative to the multistep procedure is simultaneously taken into account phenotypic, pedigree and genomic data to directly estimate GEBV by using the single-step genomic methodology (ssGBLUP) (Misztal et al., 2009; Aguilar et al., 2010; Christensen and Lun, 2010). The ssGBLUP makes use of the same equation structure applied in analyses using the pedigree-based relationship matrix, where the latter is replaced by a new one, created by integrating the pedigree- and the genomic-based relationship matrices. Currently, different

studies have reported the success of ssGBLUP for genomic evaluations of production and reproductive traits in dairy cattle, due to reduced bias and increased reliability of genomic predictions, when compared to methodologies based on the multistep approach (Fragomeni et al., 2019; Guarini et al., 2018; Ma et al., 2015).

In Portugal, the genetic evaluations of production (reproduction) traits are performed using the multiple lactations (multiple calving) autoregressive model (AR-Model). For production traits, for example, a first-order autoregressive structure is implemented within (short-term) and between (long-term) lactations to fit environmental covariances among repeated test-day records (Carvalho et al., 2002). Under this model, the longitudinal records are assumed to be repeated expression of the same trait. Differently from the traditional repeatability model, the autoregressive structure associated with environmental effects permits describing the time decaying correlations between events further apart in time, representing a more realistic alternative to deal with the environmental effects associated with longitudinal records. Analysis of reproductive traits follows the same modeling approach. In a recent study, Silva et al. (2020) demonstrated the applicability of the autoregressive model for genetic evaluation of longitudinal reproductive traits in dairy cattle. They reported better results of the AR-Model when compared with the traditional repeatability model, providing better fitting criteria and higher EBV reliabilities.

The next step for the Portuguese dairy cattle breeding program will be the inclusion of genomic data in their genetic evaluations. Recently, Silva et al. (2019) demonstrated the feasibility of using the ssGBLUP methodology under the AR-Model for genomic evaluation of milk production traits in the Portuguese Holstein cattle. They reported higher EBV reliabilities for young bulls when compared with the traditional autoregressive model (using pedigree-based relationship matrix). In addition, the results of the validation studies indicated that the GEBV predicted from the autoregressive single-step model were less biased. Thus, the purpose of this

study was to evaluate the applicability of ssGBLUP methodology under the autoregressive model for genomic evaluation of longitudinal reproductive traits in Holstein cattle.

2.3. Material and methods

Description of data

The phenotypic and pedigree data used in this study were provided by the Portuguese Dairy Cattle Breeders Association. The data comprised management and reproductive records that occurred during the first four calvings between 1995 and 2018. The reproductive traits evaluated were interval from calving to first service (ICF), defined as the number of days between calving and the following first insemination and calving interval (CI), defined as the number of days between two consecutive calvings. Daughter pregnancy rate (DPR) was computed as a linear transformation of days open (number of days between calving and conception), expressed by the equation adapted from VanRaden et al. (2004), where [DPR = $0.25 \times (266 - \text{days open})$]. In addition, preliminary analyses indicated a voluntary waiting post-partum period of approximately 42 days. For purpose of this study, it was assumed a standard gestation length of 280 ± 15 days.

The raw data were edited according to particulates of each reproductive trait, ensuring that the records are within physiological limits and reliable management procedures. DPR was calculated only after editing days open (DO) records. Therefore, the following minimum and maximum limits were established: 42 to 168 days for ICF, 42 to 252 days for DO and 307 to 730 days for CI. Additionally, ICF and DO values between 21 and 41 days were set to 42 days to avoid penalize early-breeder cows, which corresponded to 3.8% and 2.4% of total data for ICF and DO, respectively.

All fertile insemination dates were validated by requiring that the number of days to the following calving should be in the range of the assumed gestation length. When the insemination date between two consecutive calving was not available, the DO was assumed

equal to CI minus 280 days (gestation period). Finally, it was also required that the cows' data had no missing information between consecutive calving orders. After editing, 461,726, 872,603, and 766,897 records of ICF, CI and DPR, respectively, were used to calculate GEBVs and EBVs. Furthermore, to evaluate GEBV predictions, a reduced data set was prepared, in which cows' records measured in the last 5 years were removed. The reduced data set comprised a total of 304,052, 758,719, and 658,794 records for ICF, CI, and DPR, respectively.

A total of 2,875 animals were genotyped using SNP chip panels of different densities: LDv1, LDv2, LDv3, LDv4, HDv1, HDv2, and HDv3 (GeneSeek Genomic Profiler, Neogen Corp., Lincoln, NE, USA, BeadChips); 50Kv1 and 50Kv2 (Illumina, San Diego, CA, USA); 57K (USDA Illumina, San Diego, CA, USA); and Ax58K (Affymetrix, Santa Clara, CA, USA). All cows were genotyped with the Ax58K. The SNP chips of low-, medium- and high-density were imputed to 50Kv2 panel using the FImpute 2.2 software (Sargolzaei et al., 2014). The imputed genotypes from different panels were validated with a minimum value for concordance rate of 96% and correlation between imputed and true SNP of 0.95. Analyses of quality control were performed using PLINK software (Purcell et al., 2007). SNPs with minor allele frequency < 0.01, SNPs and animals with call rates < 0.9 and Hardy-Weinberg disequilibrium test ($P < 1 \times 10^{-6}$) were excluded from the analyses. Moreover, Mendelian conflicts and errors of sex disagreement were checked. SNPs with unknown position or located on sex chromosomes were disregarded. After quality control, 39,372 SNP were used in genomic analyses. The number of animals in the pedigree, animals with records and genotyped animals are shown in Table 1.

Table 1. Number of animals in pedigree, phenotypic and genotypic data used in the analyses for interval from calving to first service (ICF), calving interval (CI) and daughter pregnancy rate (DPR)

	ICF		CI		DPR	
	Bulls	Cows	Bulls	Cows	Bulls	Cows
Animals in pedigree	25,434	461,819	26,731	621,240	26,794	627,541
Animals with records						
Full data set	-	228,937	-	404,925	-	431,859
Reduced data set	-	168,362	-	339,260	-	363,622
Genotyped animals						
Bulls with 0 to 9 daughter	609	-	599	-	599	-
Bulls with more than 9 daughter	621	-	631	-	631	-
Total	1,230	1,645	1,230	1,645	1,230	1,645
Non-genotyped animals						
Non-genotyped bulls with 0 to 9 daughter	20,560	-	21,350	-	21,366	-
Non-genotyped with more than 9 daughter	3,644	-	4,151	-	4,200	-
Total	24,204	460,174	25,501	619,595	25,566	625,896

Statistical models

The analyses of ICF, CI and DPR were performed using the pedigree-based (A-AR) and the genomic (ssGBLUP methodology) relationship matrix (H-AR). The models can be described as follows:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Kc} + \mathbf{Za} + \mathbf{Wpe} + \mathbf{e}, \quad (1)$$

where \mathbf{y} is the vector of observations for the first, second and third calving orders; \mathbf{b} is the vector of fixed effect (calving month and age-month classes at calving); \mathbf{c} is the vector of random herd-year-season (HYS) of calving effects; \mathbf{a} is the random additive genetic effect; \mathbf{pe} is the vector of random permanent environmental effects; \mathbf{e} is the vector of residuals. \mathbf{X} , \mathbf{K} , \mathbf{Z} ,

and \mathbf{W} are the incidence matrices relating records to the fixed and random effects, respectively.

The general expectations and (co)variance were:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{c} \\ \mathbf{a} \\ \mathbf{pe} \\ \mathbf{e} \end{bmatrix} \sim N \left(\begin{bmatrix} \mathbf{Xb} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{V} & \mathbf{JK} & \mathbf{GZ} & \mathbf{PW} & \mathbf{R} \\ & \mathbf{J} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ & & \mathbf{G} & \mathbf{0} & \mathbf{0} \\ & \text{Sym} & & \mathbf{P} & \mathbf{0} \\ & & & & \mathbf{R} \end{bmatrix} \right),$$

where, $\mathbf{J} = \sigma_c^2 \begin{bmatrix} 1 & \rho_c & \rho_c^2 \\ & 1 & \rho_c \\ \text{Sym} & & 1 \end{bmatrix} \otimes \mathbf{I}$, is a block diagonal matrix where σ_c^2 and ρ_c are the variance

and the autocorrelation coefficient for HYS, respectively (the size of the autocorrelation matrix depends on the number of year-season levels within each herd - 3 in this example). \mathbf{I} is an identity matrix with order equal to the number of herds.

$\mathbf{P} = \sigma_{pe}^2 \begin{bmatrix} 1 & \rho_{pe} & \rho_{pe}^2 \\ & 1 & \rho_{pe} \\ \text{Sym} & & 1 \end{bmatrix} \otimes \mathbf{I}$ is also a block diagonal matrix where σ_{pe}^2 and ρ_{pe} are the

variance and the autocorrelation coefficient for \mathbf{pe} , respectively (in this case, the size of the autocorrelation matrix depends on the number of repeated observations for each cow - 3 in this example). \mathbf{I} is an identity matrix with order equal to the number of animals with records.

For A-AR model $\mathbf{G} = \mathbf{A}\sigma_a^2$, where σ_a^2 is the additive genetic variance and \mathbf{A} is the pedigree-based relationship matrix. For H-AR model $\mathbf{G} = \mathbf{H}\sigma_a^2$, where \mathbf{H} is the modified relationship matrix that combines pedigree- and genomic-based relationships (Aguilar et al., 2010); and $\mathbf{R} = \mathbf{I}\sigma_e^2$, where \mathbf{I} is an identity matrix with size equal to the total number of records and σ_e^2 is the residual variance component. Finally, $\mathbf{V} = \mathbf{KJK}' + \mathbf{ZGZ}' + \mathbf{WPW}' + \mathbf{R}$ represents the phenotypic (co)variance.

The inverse of the \mathbf{H} matrix was calculated using the preGSF90 software (Misztal et al., 2014), as follows:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}_w^{-1} - \omega \mathbf{A}_{22}^{-1} \end{bmatrix},$$

where \mathbf{A}^{-1} and \mathbf{A}_{22}^{-1} are the inverse of pedigree-based relationship matrix among all animals and genotyped animals, respectively. $\mathbf{G}_w = w\mathbf{G} + (1 - w)\mathbf{A}_{22}$, where w represents the weight whose purpose is to avoid singularity problems related to polygenic effect, assuming the value of 0.95 (VanRaden, 2008). Finally, ω is the scaling factor used to avoid converge problems and bias. Preliminary analyses indicated 0.6, 0.3, and 1.0 as optimal ω values for ICF, CI, and DPR, respectively.

Variance components were obtained by using a program implemented in MATLAB software (MathWorks Inc., Natick MA) considering phenotypes and pedigree data. Preliminary analyses showed to be computationally impracticable the use of complete data set to estimate the variance components. A large demand for time and memory was required to store the Cholesky factor of the coefficient matrix from the mixed model equations. For this reason, six sub-data sets were extracted from the complete data set to estimate the variance components and the autocorrelation coefficients for ICF, CI, and DPR. Each sub-data set consisted of 30 randomly sampled herds with records spanning across all periods of study and representing all regions of Portugal. All parameters were estimated applying DFREML methodology (Smith & Graser, 1986). Likelihood functions were maximized by the multivariate simplex algorithm (Nelder and Mead, 1965) programmed in MATLAB (Math Works Inc., Natick MA). To achieve convergence, the variance among the simplex $-2\log$ likelihoods had to be less than 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). This process was repeated for each sub-data set and the respective averages of variance components and autocorrelations estimates were used in the prediction analyses by the H-AR and A-AR models. The variance components, autocorrelation coefficients, and heritability estimates are shown in Table 2.

Table 2. Variance components, autocorrelation coefficients and genetic parameters with the respective standard errors (SE) estimated by the traditional autoregressive model (A-AR) for interval from calving to first service (ICF), calving interval (CI) and daughter pregnancy rate (DPR)

Parameters [†]	ICF ± SE	CI ± SE	DPR ± SE
σ_a^2	64.2 ± 8.4	652.3 ± 24.6	11.8 ± 0.6
σ_c^2	193.0 ± 4.5	539.3 ± 37.4	9.3 ± 1.0
σ_{pe}^2	31.5 ± 6.5	213.9 ± 15.7	4.4 ± 0.7
σ_e^2	674.1 ± 14.7	6,023.1 ± 144.8	155.6 ± 1.2
ρ_c	0.7 ± 0.02	0.8 ± 0.02	0.9 ± 0.005
ρ_{pe}	≈ 0.0 ± 0.0	≈ 0.0 ± 0.0	≈ 0.0 ± 0.0
h^2	0.07 ± 0.008	0.09 ± 0.003	0.06 ± 0.003

[†] σ_a^2 = additive genetic variance; σ_c^2 = herd-year-season variance; σ_{pe}^2 = permanent environmental variance; σ_e^2 = residual variance; ρ_c = autocorrelation coefficient for herd-year-season effect; ρ_{pe} = autocorrelation coefficient for permanent environmental effect; h^2 = heritability.

Evaluation and validation of the genomic prediction

To evaluate the performance of the H-AR and A-AR models the results of analyses for ICF, CI, and DPR were compared considering different groups of bulls: non-genotyped bulls with 0 to 9 daughters (NG_0_9), non-genotyped bulls with more than 9 daughters (NG_M9), genotyped bulls with 0 to 9 daughters (G_0_9) and genotyped bulls with more than 9 daughters (G_M9). More than nine daughters as the threshold size for comparisons was considered because it represents the minimum information necessary for the Portuguese dairy cattle breeding program to make the results of the genetic evaluation available.

The GEBV and EBV reliabilities for the different groups of bulls, assessed as expected reliability, were obtained by taking the diagonal of the inverse of the coefficient matrix from the mixed model equations and calculated as follows:

$$\mathbf{R}_e^2 = 1 - \frac{\mathbf{PEV}}{\sigma_a^2},$$

where **PEV** is the vector of prediction error variance and σ_a^2 is the additive genetic variance. Furthermore, the rank correlations between GEBV and EBV for the groups of Top 10 and Top 100 bulls were calculated using the estimates from the completed data set.

The data included genotype information from national bulls and cows, including foreign bulls, with or without daughters in Portuguese herds. To evaluate different combinations of the available genomic information and to compare the GEBV (H-AR) and EBV (A-AR) predictions, a validation study was performed considering five different scenarios (Table 3).

Table 3. Scenarios for validation study considering the single-step autoregressive model (H-AR) and the traditional autoregressive model (A-AR)

Models	Scenarios	N [†]	Animal category [‡]	
			Bulls	Cows
A-AR	S0	-	-	-
H-AR	S1	759	B1	-
	S2	1230	B1+B2	-
	S3	2404	B1	C
	S4	2875	B1+B2	C

[†]N = number of genotyped animals.

[‡]B1 = 759 genotyped bulls with daughter or relationship information the population (contributing bulls); B2 = 471 genotyped bulls with no daughter or relationship information in the complete data set (non-contributing bulls); C = 1,645 genotyped cows.

The validation study was carried out using the complete and reduced data sets. From that, genotyped bulls with no daughters in the reduced data set, but with at least 20 daughters in the completed data set (validation bulls), were used in the analyses for ICF, CI, and DPR, which corresponded to 112, 117 and 114 bulls, respectively. Although there were no genotyped animals in scenario S0, the same bulls used in the other scenarios were considered for

validation. The validations were tested as proposed by Mäntysaari et al. (2010), where daughter yield deviation (DYD) of bulls calculated in the completed data set are regressed on their GEBV or PA obtained in the reduced data set. The regression model is described as follows:

$$\mathbf{y}_c = \mathbf{b}_0 + \mathbf{b}_1 \mathbf{X}_p + \boldsymbol{\varepsilon},$$

where y_c is the DYD of validation bulls calculated in the completed data set; b_0 is the intercept; b_1 is the linear regression coefficient indicating the bias of prediction; \mathbf{X}_p is the GEBV (S1, S2, S3, and S4) or PA (S0) of bulls obtained in the reduced data set; and $\boldsymbol{\varepsilon}$ is the residual. Validation reliabilities (\mathbf{R}_{val}) were calculated according to Meuwissen et al. (2015), as follows:

$$\mathbf{R}_{val} = \text{Cor}(\text{GEBV or PA, DYD}) / \sqrt{\overline{\mathbf{R}_{DYD}^2}},$$

where $\overline{\mathbf{R}_{DYD}^2}$ is the mean of DYD reliabilities referring to the validation bulls, in which $\mathbf{R}_{DYD}^2 = \text{EDC}/(\text{EDC} + K)$. EDC is the effective daughter contribution, which was in turn calculated as $\text{EDC} = K \times \mathbf{R}_e^2 / (1 - \mathbf{R}_e^2)$ and $K = (4 - h^2)/h^2$ (Mäntysaari et al., 2010).

2.4. Results

Reliabilities and Rank correlations

The expected reliabilities (\mathbf{R}_e^2) of GEBV and EBV are shown in Figure 1. Complementarily, \mathbf{R}_e^2 averages are shown in Table 4. In general, the \mathbf{R}_e^2 estimates obtained for non-genotyped bulls with 0 to 9 daughters (NG_0_9) and non-genotyped bulls with more than 9 daughters (NG_M9) were similar between H-AR and A-AR models, which were reflected on the curves with linear coefficients closed to the unit (Figure 1). However, the \mathbf{R}_e^2 estimates obtained for genotyped bulls with 0 to 9 daughters (G_0_9) and genotyped bulls with more than 9 daughters (G_M9) were visually higher when evaluated using the H-AR model, causing the curves for these groups to deviate from those observed for non-genotyped ones.

The results showed in Table 4 confirm those displayed in Figure 1. The R_e^2 averages for NG_0_9 and NG_M9 were similar between the two evaluated models ($P > 0.001$). However, when considering the groups of genotyped bulls (G_0_9 and G_M9) the R_e^2 averages were higher when estimated from the H-AR models ($P < 0.001$).

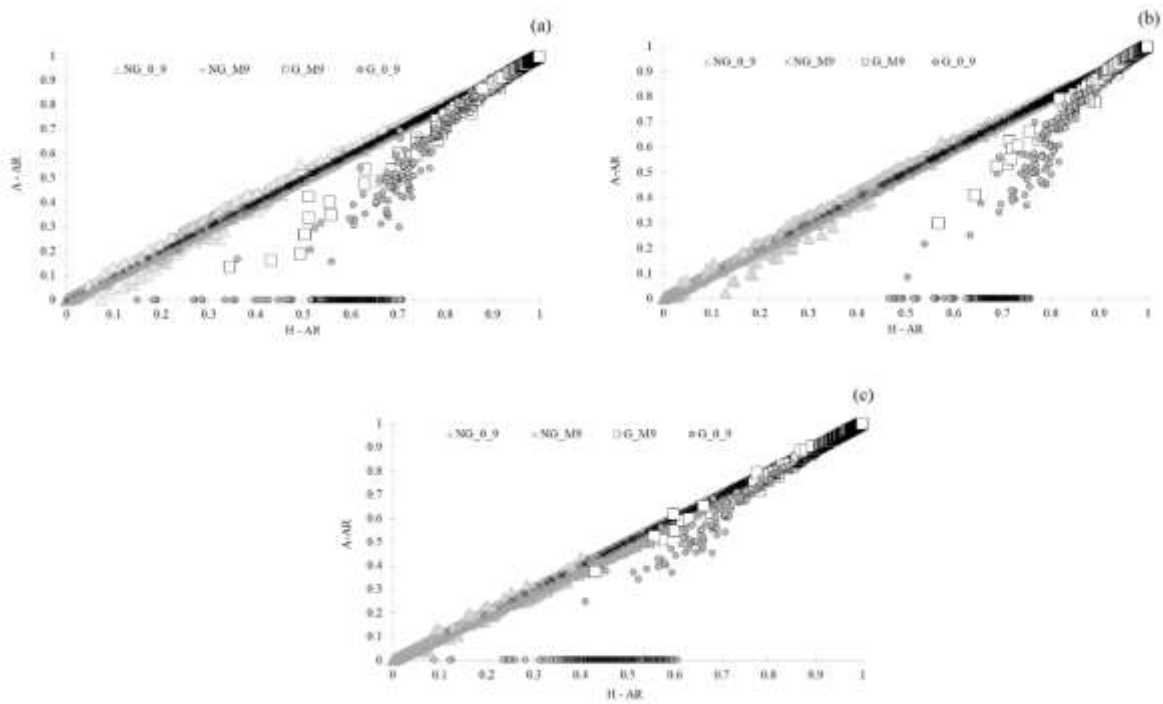


Figure 1. GEBV and EBV reliabilities obtained from the single-step autoregressive model (H-AR) and the traditional autoregressive model (A-AR) for interval from calving to first service (a), calving interval (b) and daughter pregnancy rate (c), considering genotyped bulls with 0 to 9 daughters (NG_0_9), genotyped bulls with more than 9 daughters (G_M9), non-genotyped bulls with 0 to 9 daughters (NG_0_9) and non-genotyped bulls with more than 9 daughters (NG_M9).

Table 4. Averages of GEBV and EBV reliabilities with respective standard deviations (in parentheses) obtained from the single-step autoregressive model (H-AR) and the traditional autoregressive model (A-AR) for interval from calving to first service (ICF), calving interval (CI) and daughter pregnancy rate (DPR) considering different group of bulls

Group of bulls [†]	ICF			CI			DPR		
	H-AR	A-AR	P-value	H-AR	A-AR	P-value	H-AR	A-AR	P-value
NG_0_9	0.34 (0.18)	0.34 (0.19)	> 0.001	0.36 (0.15)	0.36 (0.15)	> 0.001	0.34 (0.14)	0.34 (0.12)	> 0.001
NG_M9	0.84 (0.17)	0.83 (0.19)	> 0.001	0.87 (0.11)	0.86 (0.10)	> 0.001	0.83 (0.10)	0.83 (0.11)	> 0.001
G_0_9	0.62 (0.10)	0.27 (0.26)	< 0.001	0.69 (0.10)	0.15 (0.27)	< 0.001	0.62 (0.13)	0.16 (0.25)	< 0.001
G_M9	0.95 (0.07)	0.90 (0.11)	< 0.001	0.96 (0.04)	0.90 (0.07)	< 0.001	0.96 (0.06)	0.90 (0.07)	< 0.001

[†]NG_0_9 = non-genotyped bulls with 0 to 9 daughters; NG_M9 = non-genotyped bulls with more than 9 daughters; G_0_9 = genotyped bulls with

0 to 9 daughters; G_M9 = genotyped bulls with more than 9 daughter.

As the differences between R_e^2 estimated from the H-AR and A-AR models were observed only for genotyped bulls, we investigated the gain in reliabilities provided by the use of genomic information in these groups of bulls. The difference between reliabilities estimated was plotted (Figure 2) with the H-AR and A-AR models ($R_{e_{H-AR}}^2 - R_{e_{A-AR}}^2$) versus the respective number of daughters. The maximum gains in reliability were obtained for bulls without daughters, corresponding to values up to 0.72, 0.76, and 0.60 for ICF, CI, and DPR, respectively. However, from progeny size of 18, 25, and 13 daughters for ICF, CI, and DPR, respectively, the reliabilities estimated from the H-AR and A-AR models were practically the same.

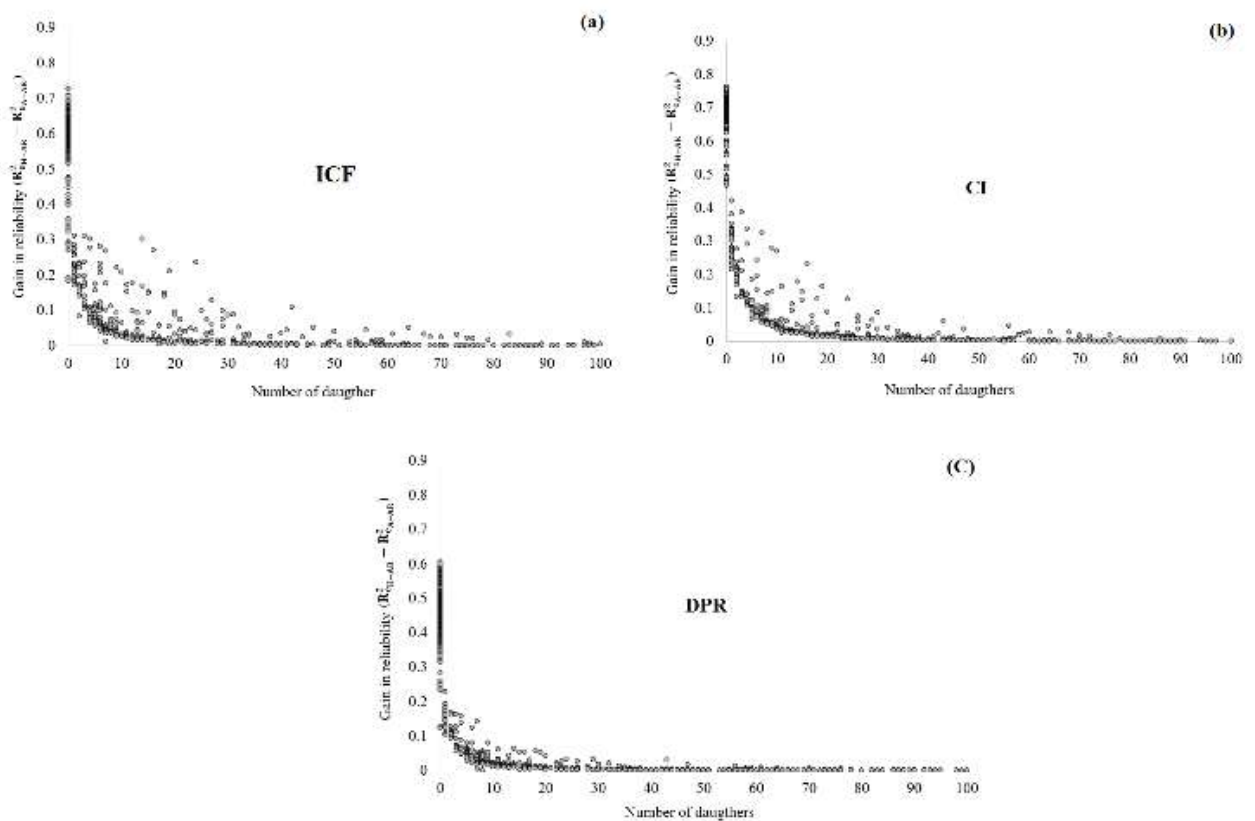


Figure 2. Gain in reliability for genotyped bulls versus number of daughters when comparing the single-step autoregressive model (H-AR) with the traditional autoregressive model (A-AR) for interval from calving to first service (a), calving interval (b) and daughter pregnancy rate (c).

The estimates of rank correlation between GEBV and EBV (H-AR and A-AR) in the Top 10 and Top 100 bulls are shown in Table 5. For all reproductive traits, the rank correlation considering the Top 10 bulls were of medium to low magnitude, ranging from 0.41 to 0.72 and 0.58 to 0.85 within the G_0_9 and G_M9 groups, respectively. The rank correlation estimates increased by increasing the number of bulls (Top 100). However, few ranking changes occurred in the non-genotyped groups (NG_0_9 and NG_M9), where the rank correlation values were greater than 0.87 in the Top 10 and Top 100 bulls.

Table 5. Rank correlation between GEBV and EBV obtained from the single-step autoregressive model (H-AR) and the traditional autoregressive model (A-AR) for interval from calving to first service (ICF), calving interval (CI) and daughter pregnancy rate (DPR) considering the Top 10 and Top 100 bulls

Group of bulls [†]	ICF		CI		DPR	
	Top 10	Top 100	Top 10	Top 100	Top 10	Top 100
NG_0_9	0.91	0.98	0.87	0.96	0.92	0.98
NG_M9	0.96	0.99	0.90	0.92	0.97	0.98
G_0_9	0.72	0.91	0.51	0.78	0.41	0.87
G_M9	0.85	0.90	0.58	0.90	0.83	0.90

[†]NG_0_9 = non-genotyped bulls with 0 to 9 daughters; NG_M9 = non-genotyped bulls with more than 9 daughters; G_0_9 = genotyped bulls with 0 to 9 daughters; G_M9 = genotyped bulls with more than 9 daughter.

Validation

The regression coefficients (b_1) and validation reliabilities (\mathbf{R}_{val}^2) obtained from the analyses in the different validation scenarios are shown in Table 6. When comparing the scenarios that included genotyped animals (S1, S2, S3, and S4) with that comprising only non-genotyped ones (S0), the b_1 estimates were closer to the unit, indicating a reduction in prediction bias by the H-AR model. The GEBVs predicted from the H-AR model were at least

0.17, 0.24 and 0.09 less biased than EBVs predicted from the A-AR model for ICF, CI, and DPR, respectively. In addition, the validation reliabilities obtained from analyses using the H-AR model in S1, S2, S3 and S4 outperformed the estimates obtained from the A-AR model in S0.

The validation considering only contributing bulls (S1) was less biased than in S2, with non-contributing bulls also added. The differences between b_1 estimates were 0.03, 0.03 and 0.11 in favor of S1 for ICF, CI and DPR, respectively. The inclusion of genotyped cows reduced the bias in both scenarios, S3 and S4 (cows plus S1 and S2). Estimates of b_1 were closer to unit when in S3 (genotyped cows were combined with only contributing bulls), than with S4 (when they were combined with contributing and non-contributing bulls simultaneously). The R_{val}^2 estimates followed the same trend of b_1 , with higher values when genotyped cows were included.

Table 6. Regression coefficient (b_1) and empirical reliability (R_{val}) for different validation scenarios, considering the single-step autoregressive model (H-AR) and the traditional autoregressive model (A-AR) for interval from calving to first service (ICF), calving interval (CI) and daughter pregnancy rate (DPR)

Models	Scenarios	ICF		CI		DPR	
		b_1	R_{val}	b_1	R_{val}	b_1	R_{val}
A-AR	S0	0.72	0.23	0.64	0.20	1.41	0.32
	S1	0.92	0.26	0.93	0.24	1.21	0.38
H-AR	S2	0.89	0.25	0.90	0.26	1.32	0.39
	S3	0.94	0.29	1.02	0.36	1.12	0.43
	S4	0.93	0.28	1.12	0.36	1.17	0.42

2.5. Discussion

Different studies (Legarra et al. 2014; Ma et al., 2015; VanRaden and Wright, 2013) discuss the advantages of the single-step methodology in genetic evaluations of dairy cattle.

However, this possible superiority is more evident in small genotyped populations. In these circumstances, the accuracies of GEBV obtained from multistep methodologies could be as low as those from traditional BLUP evaluations (Van Grevenhof et al., 2012).

Lourenco et al. (2014) evaluated different methods for genomic evaluation in a small genotyped dairy population and concluded that the accuracy of DGV obtained from the multistep method may be less accurate than PA. In addition, the results from the ssGBLUP have shown consistency in providing reliable accuracies. Currently, the genotyped population in Portuguese Holstein cattle is relatively small and, although our study did not investigate the relationship between the size of genotyped population and methodologies for genomic prediction, the higher R_e^2 estimates in the G_0_9 and G_M9 groups from the H-AR model are in agreement with those authors. Moreover, from the results shown in Figure 1 and Table 4, it is possible to observe that the R_e^2 estimates were especially higher in the G_0_9 groups, in line with previous studies (De Ross et al., 2011; Hayes et al., 2009).

One of the most important contribution of genomic selection for the genetic progress in dairy cattle is the possibility to improve the EBV reliabilities of young bulls. These animals may not have sufficient time to accrue the number of daughters necessary to achieve similar levels of reliabilities of older bulls. The information from genomic data makes it possible to improve the GEBV reliabilities for young bulls. Although the H-AR model has provided higher reliability estimates for genotyped bulls (G_0_9 and G_M9), the estimates obtained for non-genotyped bulls were practically the same for the two models evaluated. This result is in agreement with other studies evaluating the ssGBLUP and conventional BLUP methodologies in small genotyped populations (Oh et al., 2017, Silva et al., 2019).

The gain in reliability showed in Figure 2 indicates how much the genomic information may contribute to R_e^2 estimates when compared with the traditional approach. These results suggest that the use of genomic information by using the H-AR model may be equivalent to

records of up to 18, 25 and 13 daughters in terms of EBV reliabilities for bulls in the analyses of ICF, CI, and DPR, respectively. Low heritability traits, such as reproductive ones, benefit most from the inclusion of genomic information, especially due to the higher relative increase in reliabilities of predictions achieved for these traits (Wiggans et al., 2017). Besides the benefits of genomic information in the evaluation of low heritability traits are potentially greater in small breeding populations. In these situations, the opportunities to obtain a sufficiently large number of daughters to generate highly reliable EBVs are limited. Therefore, the information from genomic data may recover genetic relationship that otherwise would not be considered with the pedigree-based analysis (Gaddis et al., 2014; Mészáros et al., 2015).

The rank correlation between GEBV and EBV obtained respectively with H-AR and A-AR model followed the same trend of R_e^2 results where, greater changes in ranking occurred in groups of genotyped bulls, especially for the Top 10 (Table 5). Similar results were reported by Silva et al. (2019) when evaluating genomic and traditional models for analyses of milk production in dairy cattle.

The quality of results obtained through genomic selection is directly related to the number of genotyped animals considered in the analyses. In addition to the combination of bulls and cows data, sharing genotype data among countries is a very important strategy to increase the genomic information available to the analyses. Despite the small size of the genotyped population, this situation is also observed in the genotyped data available in Portugal. Therefore, the validation study was carried out to investigate the performance of H-AR model considering different sources of genomic information (scenarios).

The scenario S0 (without genomic information) was considered as a reference to compare the H-AR and A-AR models. The other scenarios (S1 – S4) were designed to evaluate the impact of different genomic information on the prediction accuracy of bulls. In this context, scenario S1 included genotypes from contributing bulls (bulls with daughters or relationship

information in the Portuguese herd), scenario S2 included genotypes from all bulls (S1 plus from non-contributing bulls - without daughter or relationship information in the Portuguese herd), scenario S3 included genotypes from cows plus S1 and finally, scenario S4 included genotypes from cows plus S2. The inclusion of any genomic information had a favorable impact on the prediction bias and reliabilities.

All scenarios analyzed under the H-AR model (S1 - S4) outperformed the A-AR model (S0), highlighting S3, where lower bias and higher reliabilities were obtained (Table 6). These results are similar to those observed in other studies. Evaluating the use of genomic information for genetic analyses of interval from calving to first insemination, interval from first to last insemination and non-return rate, Matilainen et al. (2018) reported that ssGBLUP improved the predictions compared to the traditional BLUP methodology. This is especially true for lowly heritable traits in Holstein cattle. Guarini et al. (2018) showed that the inclusion of genomic information in the analyses, improved the average predictions reliabilities by 7.0% and 13.0% for reproduction and workability traits, respectively. Complementarily, the regression coefficients for the evaluated traits were, on average, 0.05 closer to the unit when using ssGBLUP. The results from the ssGBLUP under the autoregressive model, in the present study, promoted an average improvement of up to 30% and 26% in the reliabilities and predictions, respectively (scenario S3).

The use of non-contributing bulls did not provide any improvement in predictive ability. On the other hand, lower bias levels and higher validation reliabilities were observed when these animals were not included in the analyses (S1 and S3). These results are in agreement with those presented by Silva et al. (2019) when evaluating the ssGBLUP methodology for milk production. According to those authors, the scenarios where only contributing bulls were considered, the validation reliabilities and regression coefficient were on average 29.0% and 16.0% better than the scenario where non-contributing bulls were added. In theory, the inclusion

of all genotyped information in ssGBLUP seems to be a reasonable strategy, since all information applied to selection decisions should be accounted for in genetic evaluation. According to Koivula et al. (2018) using all information in the genomic evaluations makes it possible to take into account genomic preselection. However, in genomic preselection, young bulls are culled based on GEBV and therefore do not provide any phenotypic information about selection. In the Portuguese dairy herds, there was no preselection based on genomic information, the non-contributing bulls in scenarios S2 and S4 are foreign animals without progeny records. Therefore, they did not provide any phenotypic information about selection, similar to the situation where genomic information is used for preselection of young bulls. In this way, the non-contributing bulls do not need to be included in the genomic analyses of ICF, CI, and DPR using the H-AR model.

The scenarios S3 and S4 permitted to investigate the possible contributions of including genotyped cows in the genomic prediction of bulls. Combining genomic information of cows and bulls (S3 and S4) outperformed the analyses where only genotyped bulls were included (S1 and S2), especially for scenario S3. In this scenario, b_1 and R_{val} estimates were on average 6.6% and 9.1%, and 11.7% and 9.0% better than those obtained in S1 and S2, respectively. The use of genotyped cows for bulls GEBV prediction has been an approach considered in small dairy cattle populations in which a small number of genotyped progeny-tested bulls are available. In these populations, the use of genotyped cows may contribute to increasing the size of the reference population (Uemoto et al., 2017; Calus et al., 2013). The results from the scenarios S3 and S4 indicate that the inclusion of genotyped cows had a positive effect in genomic evaluation of bulls. Guarini et al. (2018) reported similar results in the analyses of various workability and reproductive traits under the ssGBLUP methodology. However, there are also studies reporting a decrease of GEBV accuracies and increased bias with the inclusion of genotyped cows in the analyses (Dassonneville et al., 2012; Uemoto et al., 2017), which may

be partially attributed to the methodology used. The genomic analyses in these studies were performed using multistep approaches. Under these methodologies, the data are pre-adjusted to avoid problems related to preferential treatment of cows (Wiggans et al., 2012). However, pre-adjustments are approximations and errors may be introduced, decreasing the GEBV accuracy and increasing bias (Uemoto et al., 2017). Working directly with records, the ssGBLUP does not require any pre-adjustment (Tsuruta et al. 2013) and the use of genotyped cows may represent an important advantage for the evaluations as shown in the present study.

2.6. Conclusion

The ssGBLUP autoregressive model (H-AR) resulted in improved reliabilities and reduced bias of breeding value predictions compared with the traditional analyses based on the pedigree relationship matrix. In general, the inclusion of genotyped cows provided new genetic connections and contributed to an increased response in the accuracy of the bulls genomic breeding values. This increase in accuracy was especially evident when using only the genotyped bulls with daughters in the database. Therefore, bulls with no relationships in the data do not need to be included in the genomic analyses performed by using the H-AR model. The results from the present study suggest that the ssGBLUP methodology under the autoregressive model (H-AR) is a feasible and applicable approach to be used in the genomic analysis of reproductive traits in Portuguese Holstein cattle.

2.7 Acknowledgments

The authors acknowledge the Portuguese Dairy Cattle Breeders Association for providing the data for this study. This study was funded by Coordination for Improvement of Higher Education Personnel and Portuguese National Funding Agency for Science, Research and Technology (CAPES/ FCT, no. 99999.008462/2014-03 and 88887.163433/2018-00), and

National Council of Technology and Scientific Development (CNPq 465377/2014-9 – PROGRAMA INCT and CNPq 140491/2016-3).

2.8. References

- Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., & Lawlor, T. J. (2010). Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science*, *93*(2), 743-752.
- Calus, M. P. L., De Haas, Y., & Veerkamp, R. F. (2013). Combining cow and bull reference populations to increase accuracy of genomic prediction and genome-wide association studies. *Journal of Dairy Science*, *96*(10), 6703-6715.
- Carvalho, J., Pollak, E. J., Quaas, R. L., & Blake, R. W. (2002). An autoregressive repeatability animal model for test-day records in multiple lactations. *Journal of Dairy Science*, *85*(8), 2040-2045.
- Christensen, O. F., & Lund, M. S. (2010). Genomic prediction when some animals are not genotyped. *Genetics Selection Evolution*, *42*(1), 2.
- Dassonneville, R., Baur, A., Fritz, S., Boichard, D., & Ducrocq, V. (2012). Inclusion of cow records in genomic evaluations and impact on bias due to preferential treatment. *Genetics Selection Evolution*, *44*(1), 40.
- De Roos, A. P. W., Schrooten, C., Veerkamp, R. F., & Van Arendonk, J. A. M. (2011). Effects of genomic selection on genetic improvement, inbreeding, and merit of young versus proven bulls. *Journal of Dairy Science*, *94*(3), 1559-1567.
- Fragomeni, B. O., Lourenco, D. A. L., Legarra, A., VanRaden, P. M., & Misztal, I. (2019). Alternative SNP weighting for single-step genomic best linear unbiased predictor evaluation of stature in US Holsteins in the presence of selected sequence variants. *Journal of Dairy Science*, *102*(11), 10012-10019.

- Gaddis, K. P., Cole, J. B., Clay, J. S., & Maltecca, C. (2014). Genomic selection for producer-recorded health event data in US dairy cattle. *Journal of Dairy Science*, *97*(5), 3190-3199.
- Guarini, A. R., Lourenco, D. A. L., Brito, L. F., Sargolzaei, M., Baes, C. F., Miglior, F., ... & Schenkel, F. S. (2018). Comparison of genomic predictions for lowly heritable traits using multi-step and single-step genomic best linear unbiased predictor in Holstein cattle. *Journal of Dairy Science*, *101*(9), 8076-8086.
- Hayes, B. J., Bowman, P. J., Chamberlain, A. J., & Goddard, M. E. (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science*, *92*(2), 433-443.
- Koivula, M., Strandén, I., Aamand, G. P., & Mäntysaari, E. A. (2018). Reducing bias in the dairy cattle single-step genomic evaluation by ignoring bulls without progeny. *Journal of Animal Breeding and Genetics*, *135*(2), 107-115.
- Legarra, A., Christensen, O. F., Aguilar, I., & Misztal, I. (2014). Single Step, a general approach for genomic selection. *Livestock Science*, *166*, 54-65.
- Lourenco, D. A. L., Misztal, I., Tsuruta, S., Aguilar, I., Lawlor, T. J., Forni, S., & Weller, J. I. (2014). Are evaluations on young genotyped animals benefiting from the past generations?. *Journal of Dairy Science*, *97*(6), 3930-3942.
- Lund, M. S., de Roos, A. P., de Vries, A. G., Druet, T., Ducrocq, V., Fritz, S., Guillaume, F., Guldbandsen, B., Liu, Z., Reents, R., Schrooten, C., Seefired, F., & Su, G. (2011). A common reference population from four European Holstein populations increases reliability of genomic predictions. *Genetics Selection Evolution*, *43*(1), 43.
- Ma, P., Lund, M. S., Nielsen, U. S., Aamand, G. P., & Su, G. (2015). Single-step genomic model improved reliability and reduced the bias of genomic predictions in Danish Jersey. *Journal of Dairy Science*, *98*(12), 9026-9034.

- Mäntysaari, E. A., Liu, Z., & VanRaden, P. (2010). Interbull validation test for genomic evaluations. *Interbull Bulletin*, (41), 17-17.
- Matilainen, K., Strandén, I., Aamand, G. P., & Mäntysaari, E. A. (2018). Single-step genomic evaluation for female fertility in Nordic Red dairy cattle. *Journal of Animal Breeding and Genetics*, 135(5), 337-348.
- Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157(4), 1819-1829.
- Meuwissen, T. H. E., Svendsen, M., Solberg, T., & Ødegård, J. (2015). Genomic predictions based on animal models using genotype imputation on a national scale in Norwegian Red cattle. *Genetics Selection Evolution*, 47(1), 79.
- Mészáros, G., Boison, S. A., Pérez O'Brien, A. M., Ferenčaković, M., Curik, I., Da Silva, M. V. B., ... & Sölkner, J. (2015). Genomic analysis for managing small and endangered populations: a case study in Tyrol Grey cattle. *Frontiers in Genetics*, 6, 173.
- Misztal, I., Legarra, A., & Aguilar, I. (2009). Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *Journal of Dairy Science*, 92(9), 4648-4655.
- Misztal, I., Tsuruta, S., Lourenco, D., Aguilar, I., Legarra, A., & Vitezica, Z. (2014). Manual for BLUPF90 family of programs. *Athens: University of Georgia*.
- Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, 7(4), 308-313.
- Oh, J. D., Na, C. S., & Park, K. D. (2017). Validation of selection accuracy for the total number of piglets born in Landrace pigs using genomic selection. *Asian-Australasian Journal of Animal Sciences*, 30(2), 149.

- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., ... Sham, P. C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, *81*(3), 559–575.
- Sargolzaei, M., Chesnais, J. P., & Schenkel, F. S. (2014). A new approach for efficient genotype imputation using information from relatives. *BMC Genomics*, *15*(1), 478.
- Silva, A. A., Silva, D. A., Silva, F. F., Costa, C. N., Lopes, P. S., Caetano, A. R., ... & Carvalheira, J. (2019). Autoregressive single-step test-day model for genomic evaluations of Portuguese Holstein cattle. *Journal of Dairy Science*, *102*(7), 6330-6339.
- Silva, H. T., Lopes, P. S., Costa, C. N., Silva, F. F., Silva, D. A., Silva, A. A., ... & Carvalheira, J. (2020). Autoregressive repeatability model for genetic evaluation of longitudinal reproductive traits in dairy cattle. *Journal of Dairy Research*, *87*(1), 37-44.
- Smith, S. P., & Graser, H. U. (1986). Estimating variance components in a class of mixed models by restricted maximum likelihood. *Journal of Dairy Science*, *69*(4), 1156-1165.
- Tsuruta, S., Misztal, I., & Lawlor, T. J. (2013). Genomic evaluations of final score for US Holsteins benefit from the inclusion of genotypes on cows. *Journal of Dairy Science*, *96*(5), 3332-3335.
- Uemoto, Y., Osawa, T., & Saburi, J. (2017). Effect of genotyped cows in the reference population on the genomic evaluation of Holstein cattle. *Animal*, *11*(3), 382-393.
- Van Grevenhof, E. M., Van Arendonk, J. A., & Bijma, P. (2012). Response to genomic selection: The Bulmer effect and the potential of genomic selection when the number of phenotypic records is limiting. *Genetics Selection Evolution*, *44*(1), 26.
- VanRaden, P. M., Sanders, A. H., Tooker, M. E., Miller, R. H., Norman, H. D., Kuhn, M. T., & Wiggans, G. R. (2004). Development of a national genetic evaluation for cow fertility. *Journal of Dairy Science*, *87*(7), 2285-2292.

- VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, *91*(11), 4414-4423.
- VanRaden, P. M., & Wright, J. R. (2013). Measuring genomic pre-selection in theory and in practice. *Interbull Bulletin*, (47).
- Vitezica, Z. G., Aguilar, I., Misztal, I., & Legarra, A. (2011). Bias in genomic predictions for populations under selection. *Genetics Research*, *93*(5), 357-366.
- Wiggans, G. R., VanRaden, P. M., & Cooper, T. A. (2012). Adjustment of all cow evaluations for yield traits to be comparable with bull evaluations. *Journal of Dairy Science*, *95*(6), 3444-3447.
- Wiggans, G. R., Cole, J. B., Hubbard, S. M., & Sonstegard, T. S. (2017). Genomic selection in dairy cattle: the USDA experience. *Annual Review of Animal Biosciences*, *5*, 309-327.

CHAPTER 3

Autoregressive model for genetic evaluation of longitudinal reproductive traits in Brazilian Holstein cattle

3.1. Abstract

We evaluated the use of autoregressive covariance structures associated with modeling of environmental effects for genetic analyses of longitudinal reproductive traits in Brazilian Holstein cattle. The analyses were performed comparing the autoregressive (AR) and classical repeatability (REP) models. The reproductive traits evaluated were days open (DO), calving interval (CI), and daughter pregnancy rate (DPR) during the first three calving orders. The results of model-fitting criteria were favorable to the AR model. The Akaike Information Criterion estimated in the analyses of DO, CI and DPR were, in average, 35.30 (\pm 4.90), 44.82 (\pm 6.43), and 20.10 (\pm 3.34) lower than those obtained from the REP model. Lower Mean Square Error and higher Akaike Weight probabilities were also obtained under the AR model. Additionally, the AR model provided higher EBV reliabilities for the top bulls. The rank correlation between EBVs obtained from the two evaluated models ranged from 0.72 to 0.76 when considering the Top 10 bulls; and increased when the Top 100 bulls were considered, ranging from 0.98 to 0.99. Lower residual variance and higher additive genetic variances were estimated under the AR model. However, the heritability and repeatability estimates were similar between the two models, ranging from 0.03 to 0.07 among the evaluated traits. The results in our study suggest that the inclusion of autoregressive covariance structures associated with modeling of herd-year-season and permanent environmental effects is a feasible and

applicable approach to be used for genetic analyses of longitudinal reproductive traits in Brazilian Holstein cattle.

Keywords: autocorrelation, interval reproductive traits, model evaluation, reproductive performance.

3.2. Introduction

Over the years, milk production have composed the main selection objectives of dairy cattle breeding programs. However, the intensive genetic selection for production traits has negatively affected the reproductive performance of Holstein herds, an indirect response due to unfavorable genetic correlations between these two classes of traits (Martin et al., 2019; Rocha et al., 2018; Rodríguez-Bermúdez et al., 2019). In this context, the decline of cows fertility has become one of the main challenges in the modern dairy farming due to extra costs with inseminations, reproductive biotechnologies and losses attributed to reproductive disorders and involuntary culling of cows (Chiumia et al., 2013; Mee, 2012; Tiezzi et al., 2012; De Vries and Marcondes, 2020).

In order to mitigate the negative impacts attributed to reproductive problems, different reproductive traits have been evaluated in the last decades to be applied in dairy cattle breeding programs (Jorjani, 2007; Gernand and König, 2017). Currently, they are routinely included in national breeding goals to improve reproductive performance while still achieving genetic gain for milk production (Miglior et al., 2005). However, the genetic evaluation system for Holstein cattle in Brazil have only included milk, fat, and protein yields and conformation traits (Costa et al., 2013). There are no recent information about genetic parameters and breeding value prediction for reproductive traits in national Holstein cattle herd.

Most of studies evaluating reproductive traits in dairy cattle are based on repeatability models, in which repeated measurements over time are considered as expressions of the same trait. In addition, under the classical repeatability model is also assumed a correlation equal to a unit among the environmental components associated with longitudinal records, which may not be a realistic approach. In this situation, it is expected that events closer in time are more correlated than distant ones (Carvalho et al., 2002; Sawalha et al., 2005). An alternative to the classical approach is to fit a first-order autoregressive covariance structure associated with the modeling of environmental effects, reflecting a decaying correlation between events over the time. Autoregressive models have been successfully used for genetic analyses of milk traits in Holstein cattle, corresponding to higher genetic variance, higher reliabilities, and lower residual variance estimates (Carvalho et al., 2002), and recently they were also validated in dairy cattle herd from Brazil. Silva et al. (2019a) demonstrated the applicability of the autoregressive test-day model under different approach for genetic evaluation of milk, fat, and protein yield and somatic cell scores in Brazilian Holstein cattle. In addition, Silva et al. (2019b) reported that in terms of predictive ability, autoregressive and random regression models performed similarly in the analyses of milk production and somatic cell scores. However, the autoregressive model represented a more parsimonious alternative for analyses of test-day records. Although studies in Brazil comprising autoregressive models have only considered production traits and somatic cell score, recently Silva et al. (2020) reported the feasibility of this methodology in genetic evaluation of reproductive traits in Portuguese Holstein cattle. Therefore, the use of the autoregressive model for genetic evaluation in Brazilian dairy cattle could also represent a promising alternative for analyses of reproductive traits.

The challenge to overcome the decline of reproductive performance in dairy herds is determinant for dairy farming profitability. Therefore, it is crucial to look for methodologies that provide more accurate genetic evaluations. In this context, we aimed to evaluate the use of

autoregressive covariance structures associated with modeling of environmental effects for genetic analyses of reproductive traits in Brazilian Holstein cattle comparing with the classical repeatability model.

3.3. Material and methods

Data

The data were provided by the Brazilian Association of Holstein Cattle comprising reproductive records from the first four calving of cows born between 1985 and 2014, that occurred between 1987 to 2016, in 1469 herds.

This study included reproductive traits based on intervals, as follows: days open (DO), computed as the interval (days) between calving and the successful (fertile) insemination; calving interval (CI) expressed as the interval (days) between two consecutive calving; and daughter pregnancy rate (DPR), proposed by VanRaden et al. (2004) and computed as a linear transformation of DO: $[DPR = 0.25*(233 - DO)]$.

The data were edited according to the specifications of each reproductive trait ensuring that all stayed within physiological limits. For management reasons, minimum and maximum limits were defined as 60 and 252 days for DO, and 307 and 730 days for CI. In order to avoid penalizing early-breeder cows, DO values between 21 and 59 days were set to 60 days. All fertile insemination dates were validated by requiring that the number of days to the following calving should be in the range of the assumed gestation length (280 days). In case of missing information for the fertile insemination date between two consecutive calving, DO was estimated as CI minus 280 days. Finally, it was required that cows had no missing information between consecutive calving orders. The descriptive statistics for the reproductive traits are presented in Table 1.

Table 1. Descriptive statistics for days open (DO), calving interval (CI), and daughter pregnancy rate (DPR)

Traits	Calving order	Records	Means	SD ^a	Cows
DO (days)	1	52,195	122.53	51.47	
	2	51,098	128.26	53.12	
	3	23,987	128.73	53.24	
	Total	127,280	125.05	53.29	68,286
CI (days)	1	69,675	427.28	88.98	
	2	68,221	438.22	91.49	
	3	35,196	439.28	90.55	
	Total	173,092	434.03	90.47	77,638
DPR (%)	1	52,155	27.86	13.11	
	2	51,098	26.18	13.50	
	3	23,987	26.06	13.53	
	Total	127,280	26.85	13.37	68,286

^aSD = standard deviation.

Models

The edited data were analyzed from the autoregressive (AR) and classical repeatability (REP) models, as described by the following general model:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Kc} + \mathbf{Za} + \mathbf{Wpe} + \mathbf{e},$$

where \mathbf{y} is the vector of observations for the first, second and third calving orders; \mathbf{b} is the vector of fixed effects for calving month and sixteen age at calving (class 1, < 20 months; classes 2–15 were formed for every 4-month interval from 20 to 75 months; and class 16 for cows over 75 months old at calving); \mathbf{c} is the vector of random herd-year-season (HYS) of calving effects; \mathbf{a} is the vector of random additive genetic effect; \mathbf{pe} is the vector of random permanent environmental effects (PE); \mathbf{e} is the vector of residuals; \mathbf{X} , \mathbf{K} , \mathbf{Z} and \mathbf{W} are the incidence matrices relating \mathbf{y} to the fixed and random effects. It was assumed that:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{c} \\ \mathbf{a} \\ \mathbf{pe} \\ \mathbf{e} \end{bmatrix} \sim N \left(\begin{bmatrix} \mathbf{Xb} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}; \begin{bmatrix} \mathbf{V} & & & & \\ & \mathbf{JK} & & & \\ & & \mathbf{GZ} & & \\ & & & \mathbf{PW} & \\ & & & & \mathbf{R} \end{bmatrix} \right).$$

The differences between AR and REP models are on the modeling of HYS and PE (\mathbf{J} and \mathbf{P} matrices, respectively), where a first-order autocorrelation structure was fitted in the AR models.

For the AR model, \mathbf{J} (\mathbf{P}) is a block diagonal matrix where the number of blocks equal the number of herds (cows with records), each fitted with a first-order autocorrelation matrix with size equal to the respective number of year-season levels (repeated records). Exemplifying

for a herd (cow) with 3 year-season levels (3 repeated records), \mathbf{J} (\mathbf{P}) = $\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 \\ & 1 & \rho \\ \text{Sym} & & 1 \end{bmatrix}$,

where σ^2 is σ_c^2 (σ_{pe}^2) and ρ is ρ_c (ρ_{pe}), the variance components and autocorrelation coefficients for \mathbf{c} (\mathbf{pe}).

For the REP model, $\mathbf{J} = \mathbf{I}\sigma_c^2$ and $\mathbf{P} = \mathbf{I}\sigma_{pe}^2$, where \mathbf{I} in both cases, is an identity matrix of order equal to the level numbers of HYS and PE. For both models $\mathbf{G} = \mathbf{A}\sigma_a^2$, where σ_a^2 is the additive genetic variance component and \mathbf{A} the numerator relationship matrix; $\mathbf{R} = \mathbf{I}\sigma_e^2$, where σ_e^2 is the residual variance component and \mathbf{I} an identity matrix of order equal the total number of records; and the phenotypic variance $\mathbf{V} = \mathbf{KJK}' + \mathbf{ZGZ}' + \mathbf{WPW}' + \mathbf{R}$.

Variance components and autocorrelation coefficients were obtained applying the DFREML methodology (Smith and Graser, 1986) implemented in the MATLAB software (MathWorks Inc., Natick MA). Preliminary analyses showed that a large demand for time and memory was required to store the Cholesky factor of the coefficient matrix from the mixed model equations when considering complete data set. Therefore, the analyses for variance components and autocorrelations coefficient were performed by suing six sub-data sets extracted from the complete data set of DO, CI, and DPR. Each sub-data set included data from 30 herds randomly sampled from Brazilian milking recording regions. Likelihood functions were maximized by the multivariate simplex algorithm (Nelder and Mead, 1965). For convergence criterion, it was considered that the variance among the simplex $-2\log$ likelihoods

had to be lower than 10^{-8} . The occurrence of the local maximum was verified by five consecutive cold starts without changes in the log-likelihood (up to four decimal places). The means of the variance components and autocorrelations estimated from the six samples of DO, CI, and DPR were posteriorly used for the prediction analyses.

Model evaluations

The AR and REP models were evaluated using the Akaike Information Criteria (AIC), represented as $AIC = -2 \log(L) + 2p$, in which L is the likelihood function and p is the number of parameters. However, AIC provides only a qualitative comparison between models (smallest AIC is better). In order to complement this information, we also used the Akaike Weights (AW), presented by Burnham & Anderson (2004), and computed as $AW_t = \frac{\exp\left(-\frac{\Delta_t}{2}\right)}{\sum_{t=1}^2 \exp\left(-\frac{\Delta_t}{2}\right)}$, where t are the AR ($t = 1$) and REP ($t = 2$) models; Δ_t is the difference between AIC of model t and the best model (with smallest AIC values). The Δ_t value for the best model is equal to zero. The level of bias of each model was estimated by the Mean Square Error (MSE), calculated as $MSE = n^{-1} \sum (y - \hat{y})^2$, where y and \hat{y} are the observed and predicted values for each trait, and n is the respective number of records.

Another factor of great importance for studies of methodologies applied to genetic evaluation is the respective impact on the EBV ranking of animals. In this sense, Spearman's rank correlation among EBVs predicted from AR and REP models were used to evaluate the changes in the ranking in the Top 10, Top 100 and total bulls. The percentage of coincidence between bulls in the Top 10 and Top 100 were also analyzed. In addition, the individual EBV reliabilities obtained from the AR and REP models were compared considering the Top 100 bulls ranked according to the solutions of the best fitted model for each trait. The individual

EBV reliabilities were computed as $\mathbf{R}_e^2 = 1 - \frac{\text{PEV}}{\sigma_a^2}$, where PEV is the prediction error variance and σ_a^2 is the additive genetic variance.

3.4. Results and discussion

Model evaluations

The results of the model-fitting criterion are shown in Table 2 and Table 3. The AIC obtained from the AR model were in average 35.30 (± 4.90), 44.82 (± 6.43), and 20.10 (± 3.34) lower than those obtained from the REP model in the analyses of DO, CI, and DPR, respectively. Furthermore, AW estimated from the AR model were closer to one for all reproductive traits, and in all sub-data sets. The information provided by the AW criterion allows the comparison of models in terms of probability. The closer to one, the greater the chance of the model be superior to the other tested models. Moreover, the AW complemented the AIC results, indicating a high probability of the AR model to be superior to the REP model. The results of MSE were also favorable to AR model, corresponding to lower estimates when compared with the REP model, which suggest lower bias in the genetic evaluations and, consequently, higher predictive ability.

These results also suggest that the inclusion of autoregressive covariance structures for environmental effects (HYS and permanent environmental) allowed a better explanation of the effects inherent to data fitting by the AR model. According to Carvalheira et al. (2002), factors that may affect the level of correlation between the longitudinal measurements of each cow have a temporal influence along with calving orders, but with different degrees of intensities. In this context, the autoregressive structure made it possible to consider the effects of autocorrelations that decay proportionally to the temporal distance between non-subsequent records (Quaas et al., 1984), which may provide more accurate genetic evaluations.

Table 2. Akaike Information Criteria (AIC) and Akaike Weigth (AW) from the analyses of the six sub-data of days open (DO), calving interval (CI) and daughter pregnancy rate (DPR) with the autoregressive model (AR) and classical repeatability (REP) models

Traits	Sub-data	AIC _{AR}	AIC _{REP}	AW _{AR}	AW _{REP}
DO	1	103,494.27	103,578.90	1.00	0.00
	2	112,399.35	112,409.44	1.00	0.00
	3	120,190.48	120,235.56	1.00	0.00
	4	112,052.22	112,060.50	0.99	0.01
	5	138,927.05	138,933.96	0.99	0.01
	6	87,471.61	87,528.32	1.00	0.00
CI	1	197,767.50	197,851.30	1.00	0.00
	2	159,494.60	159,617.30	0.99	0.01
	3	208,216.30	208,251.50	1.00	0.00
	4	179,075.20	179,207.80	1.00	0.00
	5	127,001.70	127,041.10	0.99	0.01
	6	186,113.40	186,166.60	1.00	0.00
DPR	1	71,312.38	71,314.59	0.90	0.10
	2	77,521.96	77,530.27	0.99	0.01
	3	83,025.47	83,067.56	1.00	0.00
	4	77,266.50	77,274.37	0.99	0.01
	5	95,881.38	95,892.29	1.00	0.00
	6	60,322.10	60,371.01	1.00	0.00

Table 3. Mean Square Error (MSE) obtained from the autoregressive model (AR) and the classical repeatability model (REP) for days open (DO), calving interval (CI) and daughter pregnancy rate (DPR)

Models	DO	CI	DPR
AR	2,172.01	5,784.17	137.48
REP	2,241.53	5,995.19	140.97

Spearman's rank correlation, percentage of coincidence and EBV reliabilities

In Table 4 and Figure 1 are shown the rank correlations and the percentages of coincidence for different groups of bulls, classified according to EBVs obtained from the AR and REP models, for each reproductive trait. The rank correlation estimates were moderate when considering the Top 10 bulls, ranging from 0.72 to 0.76 among the evaluated traits.

However, the estimates increased as more bulls were considered, reaching the values of 0.98 for CI, and 0.99 for DO and DPR when considering the total bulls. The percentages of coincidence followed the same trend in the rank correlation results. For the group of Top 10 bulls the estimates ranged from 70% to 80% among the evaluated traits. When more bulls were considered the percentage of coincidence also increased, ranging from 77% to 89% for the Top 100 bulls.

Table 4. Spearman’s rank correlation between the EBV of Top 10, Top 100 and Total bulls obtained from the autoregressive model (AR) and the classical repeatability model (REP) for days open (DO), calving interval (CI), and daughter pregnancy rate (DPR)

Group of bulls	DO	CI	DPR
Top 10	0.76	0.72	0.75
Top 100	0.88	0.80	0.89
Total	0.99	0.98	0.99

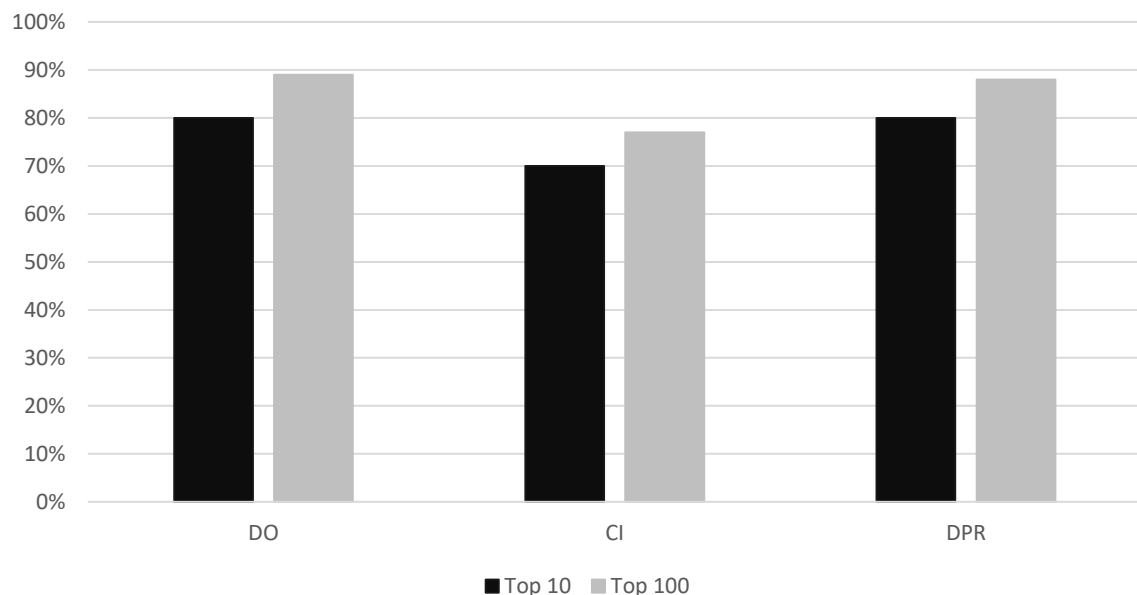


Figure 1. Percentage of coincidence within the Top 10 and Top 100 bulls ranked according to the EBV obtained from the autoregressive model (AR) and the classical repeatability model (REP) for days open (DO), calving interval (CI) and daughter pregnancy rate (DPR).

The moderate rank correlation in the Top 10 bulls indicate a re-ranking of the best animals when AR or REP model were considered. In a similar study, Silva et al. (2020) have also reported a re-ranking of bulls when evaluating AR and REP models for genetic analyses of reproductive traits in dairy cattle. According to the authors, the rank correlation between the Top 100 bulls ranged from 0.71 to 0.88 among the evaluated traits. Similar results were shown by Sawalha et al. (2005) when evaluating the inclusion of autoregressive covariance structure for genetic analyses of production traits. Although this study diverges from ours in terms of genetic nature of the evaluated traits, they also reported a re-ranking of elite bulls when autoregressive covariance structures were considered. Moreover, the results for percentage of coincidence in our study suggest that, among the evaluated traits, about 30% to 20% of bulls would not be present into the Top 10 group and 23% to 11% into Top 100 group when comparing the ranking obtained from AR and REP models. These results are evidences that genetic analyses of reproductive traits using the AR model may provide a better discernment among the best bulls, which may directly influence the choice of animals to be used for genetic improvement purpose.

The AR model also provided higher individual EBV reliabilities for the best bulls in the population (top EBVs) for all reproductive traits (Figure 2). The averages of EBV reliabilities obtained from the AR model corresponded to 0.89 (\pm 0.09), 0.81 (\pm 0.15), and 0.89 (\pm 0.09) when considering the Top 100 bulls for DO, CI and DPR, respectively. The estimates obtained from the REP model were markedly lower ($P < 0.001$) for the same group of bulls, corresponding to 0.56 (\pm 0.13), 0.65 (\pm 0.14), and 0.57 (\pm 0.13) for DO, CI, and DPR, respectively. These results are in agreement with that obtained in the Portuguese Holstein herd, where the reliabilities for the best 1%, 10%, 50% bulls were higher when obtained from the AR model (Silva et al., 2020). According to Carvalho et al. (2002), including autoregressive covariance structure associated with environmental effects could provide higher EBV

reliabilities, which is in agreement with our study. Thus, the use of autoregressive models confirm to be a more realistic approach for genetic analyses of reproductive traits, representing a lower risk approach to the accuracy of genetic evaluations. Furthermore, the results of EBV reliabilities associated with rank correlation and percentage of coincidence suggest that the genetic analyses under the AR model may provide a greater opportunity for genetic progress of reproductive traits due to the possibility of obtaining more reliable predictions and to reduce the chances of selecting animals with lower genetic potential.

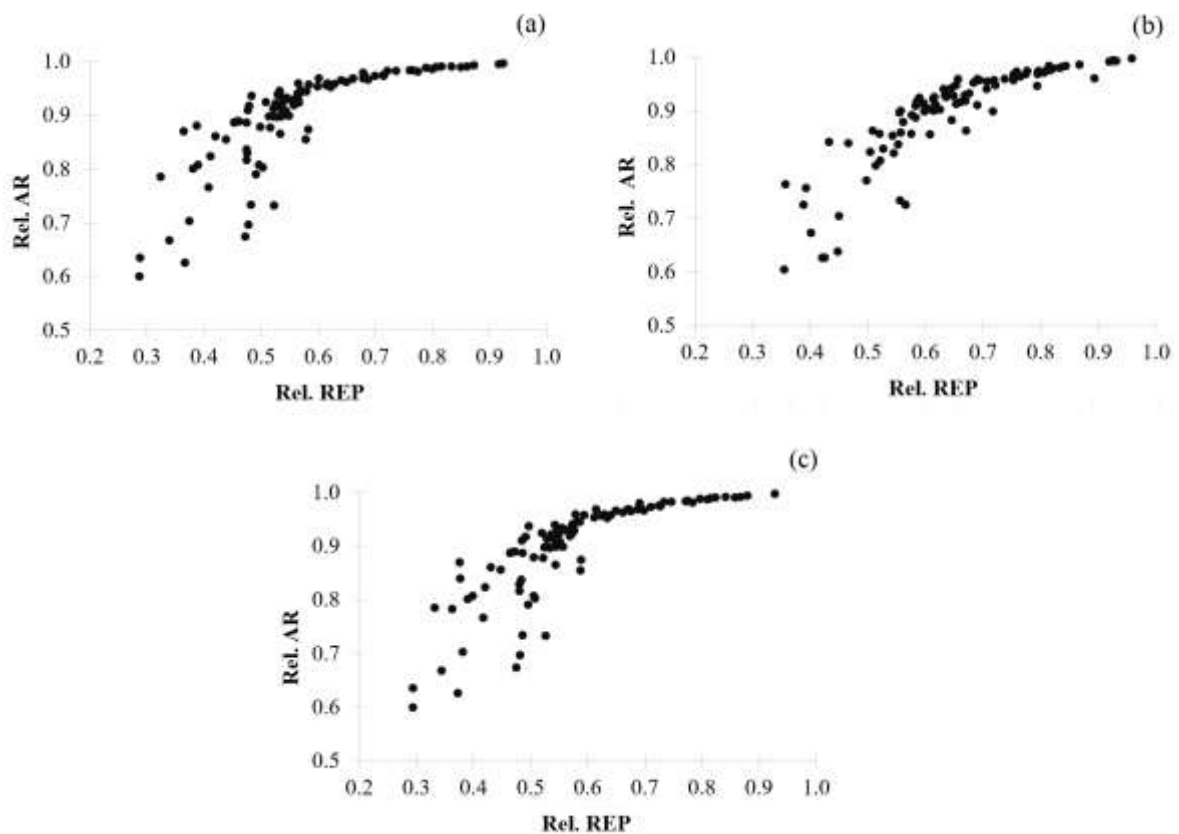


Figure 2. EBV reliabilities (Rel.) for the Top 100 bulls obtained from the autoregressive (AR) and the classical repeatability model (REP) for days open (a), calving interval (b) and daughter pregnancy rate (c).

Variance components and genetic parameters

Variance components and genetic parameters estimated from the AR and REP models are shown in Table 5. Higher additive genetic variances were obtained under the AR model, corresponding to estimates 26.5 %, 24.3%, and 23.4% greater than those obtained from the REP model in the analyses of DO, CI, and DPR, respectively. Besides that, the residual variances were slightly smaller when estimated from the AR model. The variance estimated for HYS were higher for CI, and lower for DO and DPR from the AR model. On the other hand, the variances estimated for PE were higher for DO and lower for CI when analyzed from the REP model. For DPR, the variances estimated for PE were practically the same in both models.

The autocorrelation coefficients associated to HYS effect ranged from 0.58 to 0.92 in the analyses of reproductive traits, which indicate high association among year-season classes. However, the estimates for the coefficients associated to PE effect were practically null. The heritability and repeatability estimates were similar between the two models, for all reproductive traits, with values ranging from 0.03 to 0.07 for heritability and from 0.06 to 0.12 for repeatability.

The differences in additive genetic variances between the two models evaluated suggest a contribution of the autoregressive covariance structure (AR model) in removing confounded variation of genetic component, which were subestimated when obtained from REP model. In addition, the lower residual variances estimated from the AR model, even if to a small percentage, complement the hypothesis about the favorable contribution of autoregressive covariance structure associated with HYS and PE. These results are consistent with those reported in studies involving analyses of production traits in Holstein cattle, where the inclusion of autoregressive covariance structure associated with the environmental effects indicated the possibility of obtaining higher additive genetic and lower residual variances (Carvalho et al., 1998; Carvalho et al., 2002; Silva et al., 2019b).

Table 5. Variance components, autocorrelation coefficients and genetic parameters (\pm SE) estimated from the Autoregressive models (AR) and the classical repeatability model (REP) for days open (DO), calving interval (CI) and daughter pregnancy rate (DPR)

Parameters ^a	AR			REP		
	DO	CI	DPR	DO	CI	DPR
σ_a^2	119.5 \pm 10.7	635.1 \pm 29.3	7.9 \pm 0.7	94.5 \pm 12.2	510.7 \pm 36.2	6.4 \pm 0.8
σ_c^2	119.7 \pm 15.4	387.6 \pm 27.2	7.5 \pm 0.8	140.31 \pm 11.4	353.5 \pm 15.1	8.6 \pm 0.6
σ_{pe}^2	117.0 \pm 26.5	466.5 \pm 42.7	6.0 \pm 0.8	82.7 \pm 22.0	496.4 \pm 16.9	6.2 \pm 0.9
σ_e^2	2,446.6 \pm 26.4	7,171.9 \pm 62.8	157.5 \pm 3.9	2,464.4 \pm 25.5	7,212.3 \pm 31.0	151.5 \pm 2.6
ρ_c	0.6 \pm 0.1	0.92 \pm 0.04	0.58 \pm 0.1	-	-	-
ρ_{pe}	\approx 0.0 \pm 0.0	\approx 0.0 \pm 0.0	\approx 0.0 \pm 0.0	-	-	-
h^2	0.04 \pm 0.003	0.07 \pm 0.003	0.04 \pm 0.004	0.03 \pm 0.004	0.06 \pm 0.003	0.04 \pm 0.005
r	0.08 \pm 0.005	0.12 \pm 0.003	0.07 \pm 0.06	0.06 \pm 0.006	0.11 \pm 0.001	0.07 \pm 0.006

σ_a^2 = additive genetic variance; σ_c^2 = herd-year-season variance; σ_{pe}^2 = permanent environmental variance; σ_e^2 = residual variance; ρ_c = autocorrelation coefficient for herd-year-season effect; ρ_{pe} = autocorrelation coefficient for permanent environmental effect; h^2 = heritability; r = repeatability.

The heritabilities estimated for reproductive traits were similar between the AR and REP models, with minimum and maximum values corresponding to 0.03 and 0.07 (Table 5). The estimates obtained in our study were close those obtained by Silva et al, (2020), which ranged from 0.06 to 0.09 among the reproductive traits evaluated. Other authors, reported heritability estimates ranging from 0.03 to 0.09 for DO, 0.02 to 0.07 for CI, and 0.03 to 0.07 for DPR (Kadarmideen et al., 2003; Ghiasi et al., 2011; Haile-Mariam et al., 2013; Gibson et al., 2018; Ali et al., 2019). Contrasting to the higher additive genetic and smaller residual variances obtained by the AR model, the similarity between heritabilities estimated from the two models suggests that the autoregressive structure provided a redistribution of variance component among the random effects in the model, but retained their respective proportionality, as observed in the REP model. However, the superiority of the AR model in the analyses of model-fitting criterion associated to the highest EBV reliabilities and the direct impact on the bulls ranking suggest the superiority of this methodology.

Autoregressive models have been used with success in genetic evaluations of production and reproductive traits in Portuguese, which is a member of International Bull Evaluation Service (Interbull). Moreover, recently was verified the applicability of the autoregressive single-step test-day model for genomic evaluations production traits in dairy cattle (Silva et al, 2019). Thus, the results in the present study associated with those reported by other authors, support the indication of the AR model as a promising alternative for future strategies to be implemented in the genetic analyses of the Brazilian Holstein cattle breeding program.

3.5. Conclusion

The results in our study suggest that the inclusion of autoregressive covariance structures associated with modeling of herd-year-season and permanent environmental effects is a feasible and applicable approach to be used for genetic analyses of longitudinal reproductive

traits. Better model-fitting and higher EBV accuracies were obtained when compared with the REP model. In addition, the AR model provided a re-ranking of bulls, making possible a better discernment among the best animals in the population. Therefore, The AR model represents a promising alternative to be implemented in the Brazilian Holstein cattle breeding program for genetic evaluations of longitudinal reproductive traits.

3.6. Acknowledgments

The authors acknowledge the Brazilian Holstein Cattle Breeders Association (ABCBRH) for providing data for this study. This study was funded by Coordination for Improvement of Higher Education Personnel and Portuguese National Funding Agency for Science, Research and Technology (CAPES/ FCT, no. 99999.008462/2014-03 and 88887.163433/2018-00), and National Council of Technology and Scientific Development (CNPq 465377/2014-9 – PROGRAMA INCT and CNPq 140491/2016-3).

3.7. References

- Ali, I., Muhammad Suhail, S., & Shafiq, M. (2019). Heritability estimates and genetic correlations of various production and reproductive traits of different grades of dairy cattle reared under subtropical condition. *Reproduction in Domestic Animals*, 55(7).
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological methods & research*, 33(2), 261-304.
- Carvalho, J. G. V., 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. *Journal of dairy science*, 81(10), 2738-2751.

- Carvalho, J., Pollak, E. J., Quaas, R. L., & Blake, R. W. (2002). An autoregressive repeatability animal model for test-day records in multiple lactations. *Journal of dairy science*, 85(8), 2040-2045.
- Chiumia, D., Chagunda, M. G., Macrae, A. I., & Roberts, D. J. (2013). Predisposing factors for involuntary culling in Holstein–Friesian dairy cows. *Journal of dairy research*, 80(1), 45-50.
- Costa, C. N., Teixeira, N. M., Freitas, A. F., Cobuci, J. A., Barra, R. B., Valloto, A. A., ... & Marques, A. (2013). Sumário Nacional de Touros da Raça Holandesa - 2013. *Embrapa Gado de Leite-Documentos*, 56, (INFOTECA-E).
- Gernand, E., & König, S. (2017). Genetic relationships among female fertility disorders, female fertility traits and productivity of Holstein dairy cows in the early lactation period. *Journal of Animal Breeding and Genetics*, 134(5), 353-363.
- Ghiasi, H., Pakdel, A., Nejati-Javaremi, A., Mehrabani-Yeganeh, H., Honarvar, M., González-Recio, O., Carabaño, M. J. & Alenda, R. (2011). Genetic variance components for female fertility in Iranian Holstein cows. *Livestock science*, 139(3), 277-280.
- Gibson, K. D., & Dechow, C. D. (2018). Genetic parameters for yield, fitness, and type traits in US Brown Swiss dairy cattle. *Journal of dairy science*, 101(2), 1251-1257.
- Haile-Mariam, M., Bowman, P. J., & Pryce, J. E. (2013). Genetic analyses of fertility and predictor traits in Holstein herds with low and high mean calving intervals and in Jersey herds. *Journal of dairy science*, 96(1), 655-667.
- Jorjani, H. (2007). International genetic evaluation of female fertility traits in five major breeds. *Interbull Bulletin*, (37), 144-144.
- Kadarmideen, H. N., Thompson, R., Coffey, M. P., & Kossaibati, M. A. (2003). Genetic parameters and evaluations from single-and multiple-trait analysis of dairy cow fertility and milk production. *Livestock Production Science*, 81(2-3), 183-195.

- Martin, P., Baes, C., Houlihan, K., Richardson, C., Jamrozik, J., & Miglior, F. (2019). Genetic Correlations among Selected Traits in Canadian Holsteins. *Canadian Journal of Animal Science*, 1-12.
- Mee, J. F. (2012). Reproductive issues arising from different management systems in the dairy industry. *Reproduction in Domestic Animals*, 47, 42-50.
- Miglior, F., Muir, B. L., & Van Doormaal, B. J. (2005). Selection indices in Holstein cattle of various countries. *Journal of dairy science*, 88(3), 1255-1263.
- Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The computer journal*, 7(4), 308-313.
- Quaas, R. L., Anderson, R. D., & Gilmour, A. R. (1984). *BLUP school handbook. In: Use of mixed models for prediction and for estimation of (co) variance components*. Animal Genetics and Breeding Unit, University of New England, New South Wales, Australia.
- Rocha, G. M., Cobuci, J. A., Costa, C. N., Abreu, L. R., Mota, L. F., Pires, A. V., ... & Martins, P. G. (2018). Genetic association between stayability, and productive and reproductive traits in Holstein cows. *Animal Production Science*, 58(10), 1788-1793.
- Rodríguez-Bermúdez, R., Fouz, R., Miranda, M., Orjales, I., Minervino, A. H. H., & López-Alonso, M. (2019). Organic or conventional dairy farming in northern Spain: Impacts on cow reproductive performance. *Reproduction in Domestic Animals*.
- Sawalha, R. M., Keown, J. F., Kachman, S. D., & Van Vleck, L. D. (2005). Genetic evaluation of dairy cattle with test-day models with autoregressive covariance structures and with a 305-d model. *Journal of dairy science*, 88(9), 3346-3353.
- Smith, S. P., & Graser, H. U. (1986). Estimating variance components in a class of mixed models by restricted maximum likelihood. *Journal of Dairy Science*, 69(4), 1156-1165.

- Silva, A. A., Silva, D. A., Silva, F. F., Costa, C. N., Lopes, P. S., Caetano, A. R., ... & Carvalheira, J. (2019). Autoregressive single-step test-day model for genomic evaluations of Portuguese Holstein cattle. *Journal of dairy science*.
- Silva, D. A., Costa, C. N., da Silva, A. A., e Silva, F. F., Lopes, P. S., dos Santos, G. G., ... & Carvalheira, J. (2019a). Unknown parent and contemporary groups for genetic evaluation of Brazilian Holstein using autoregressive test-day models. *Livestock science*, 220, 1-7.
- Silva, D. A., Costa, C. N., Silva, A. A., Silva, H. T., Lopes, P. S., Silva, F. F., ... & Carvalheira, J. (2019b). Autoregressive and random regression test-day models for multiple lactations in genetic evaluation of Brazilian Holstein cattle. *Journal of Animal Breeding and Genetics*.
- Silva, H. T., Lopes, P. S., Costa, C. N., Silva, F. F., Silva, D. A., Silva, A. A., ... & Carvalheira, J. (2020). Autoregressive repeatability model for genetic evaluation of longitudinal reproductive traits in dairy cattle. *Journal of Dairy Research*, 1-8.
- Tiezzi, F., Maltecca, C., Cecchinato, A., Penasa, M., & Bittante, G. (2012). Genetic parameters for fertility of dairy heifers and cows at different parities and relationships with production traits in first lactation. *Journal of dairy science*, 95(12), 7355-7362.
- VanRaden, P. M., Sanders, A. H., Tooker, M. E., Miller, R. H., Norman, H. D., Kuhn, M. T., & Wiggans, G. R. (2004). Development of a national genetic evaluation for cow fertility. *Journal of dairy science*, 87(7), 2285-2292
- De Vries, A., & Marcondes, M. I. (2020). Overview of factors affecting productive lifespan of dairy cows. *animal*, 14(S1), s155-s164.

GENERAL CONCLUSION

In this thesis, we have investigated the use of the repeatability autoregressive model as an alternative for genetic evaluation of longitudinal reproductive traits in Holstein cattle from Brazil and Portugal. Additionally, we have also evaluated the use of ssGBLUP methodology under the autoregressive model for genomic evaluation of the same longitudinal reproductive traits.

The results in our study suggest that repeatability autoregressive model is a suitable and feasible alternative for genetic evaluation of longitudinal reproductive traits in Holstein from Brazil and Portugal, corresponding to better fitting to reproductive data, higher EBV accuracies, and; consequently, better discernment among the best bulls of the population. In addition, the ssGBLUP under the autoregressive model shows to be a feasible and applicable approach for genomic analyses of the same reproductive traits. In addition, the ssGBLUP under the autoregressive model shows to be a feasible and applicable approach to be used in the genomic evaluation of the same reproductive, improving the EBV reliabilities and reducing the bias of prediction when compared with the traditional analyses based on the pedigree relationship matrix.