JOSÉ TEODORO DE PAIVA

GENETIC AND GENOMIC EVALUATION FOR MILK PRODUCTION AND FATTY ACID TRAITS PREDICTED BY MID-INFRARED (MIR) SPECTROSCOPY IN HOLSTEIN DAIRY CATTLE

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

Adviser: Paulo Sávio Lopes

Co-advisers: Nicolas Gengler
Fabyano Fonseca e Silva

VIÇOSA – MINAS GERAIS

2021

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

T

P149g
2021

Paiva, José Teodoro de, 1988-

Genetic and genomic evaluation for milk production and fatty acid traits predicted by mid-infrared (MIR) spectroscopy in Holstein dairy cattle / José Teodoro de Paiva. – Viçosa, MG, 2021.

107 f. : il. ; 29 cm.

Orientador: Paulo Sávio Lopes.

Tese (doutorado) - Universidade Federal de Viçosa.

Inclui bibliografia.

1. Bovinos de leite. 2. Melhoramento genético. 3. Parâmetros genéticos. 4. Leite - Qualidade. 5. Análise de regressão. 6. Confiabilidade. I. Universidade Federal de Viçosa. Departamento de Zootecnia. Programa de Pós-Graduação em Zootecnia. II. Título.

CDD 22. ed. 636.20821

JOSÉ TEODORO DE PAIVA

GENETIC AND GENOMIC EVALUATION FOR MILK PRODUCTION AND FATTY ACID TRAITS PREDICTED BY MID-INFRARED (MIR) SPECTROSCOPY IN HOLSTEIN DAIRY CATTLE

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

APPROVED: February 23th, 2021.

Assent:

José Teodoro de Paiva

José Teodoro de Paiva
Author

Paulo Sávio Lopes

Paulo Sávio Lopes
Adviser

To my dear uncle Ilidio (in memoriam),
for all that I am and have achieved.

I dedicate.

ACKNOWLEDGMENTS

I would like to thank God and the universe for my life and all the connections that allowed me to evolve. I am so thankful to my family, especially my mother (Maria), my father (Teodoro), my sisters (Denise, Joysse, and Andreza), and all other members. To my uncle Ilidio (in memoriam) and my aunt Naná (in memoriam), for believing in me and always help me achieve my dreams.

I am grateful to the Universidade Federal de Viçosa (UFV), especially the Animal Science Graduate Program and the Department of Animal Science (DZO). Also, I am very grateful to the Gembloux Agro-Bio Tech campus from the University of Liège. Both universities gave me the opportunity of carrying out the course and develop my research.

I would like to say a big thank you to my adviser Professor Paulo Sávio Lopes for his enthusiastic guidance, support, and friendship, which have been invaluable for the work in this thesis and beyond. I am extremely grateful for all patience, advice, and opportunities. I am very proud of being your student. You are a great example of professional and person.

My sincere thanks to my co-adviser Professor Nicolas Gengler. Thank you very much for the great opportunity of learning from you and your research group, your valuable support was essential to conduct this research. Thank you for your excellent suggestions, support, and teachings. Thank you for the opportunity to live a new experience in my life. It was very important for my personal and academic growth.

I would like to express my deep gratitude to my co-adviser Professor Fabyano Fonseca e Silva, for all support he has given me since my Master's degree. I am grateful for all attention, suggestions, teachings, and friendship. I have learned a lot from you. You are an amazing person and professional. Thank you very much for everything that you have done.

I also would like to say thank you very much to Dr. Rodrigo Reis Mota for all encouragement and guidance during this great journey, and also to Prof. José Marcelo Soriano Viana, for accepting being part of this defense committee, and for all contributions to this thesis.

My sincere thanks also to Professor Renata Veroneze and Dra. Hinayah Rojas Oliveira for their contributions, attention, support, discussions, and friendship. You helped me a lot to build my academic life and make that dream come true.

I am very grateful to all Professors and staff (especially Aldinha) of the Departament of Animal Science and Departament of Statistics, and from other university, especially to Profa. Simone Elisa Facioni Guimarães, Prof. Mario Luiz Chizotti, Prof. Moysés Nascimento, Profa. Ana Carolina Campana Nascimento, Profa. Camila Ferreira Azevedo, and Prof. Luiz Fernando

Brito, for teaching along my doctorate period. Also, I would like to thank Dra. Maria Gabriela Campolina Diniz Peixoto, Dr. Frank Angelo Tomita Bruneli, and Dr. Cláudio Napolis Costa, for all the support, contributing in some way to my academic life.

I am grateful to the GDMA members for all the insightful discussions and time together and to all friends from my Ph.D program. A special thanks goes to my good friends that I had the opportunity to work more closely, sharing good conversations, barbecues, and volleyball games, who were from “salinha”, especially Hinayah, Hugo, Amanda, Alessandra, Delvan, Sirlene, Giovani, Layla, Ivan, Natália, Pedro Freitas, Matheus, Pedro Vital, Talita, Ândrea, Arícia, Renata Henriques, Renata Rocha, Alex, and Letícia, and those from “Labtec” Karine, Ingrid, Darlene, Daniele, Susana, Haniel, Margareth, and Pâmela.

A big thank you to my roommates Marcelo, Thiago, Eula, and Allan, for friendship and support. A special thanks to the Matheus. To my friends from Viçosa city, Júlia, Isabela, Thaís, Gisele, Lycia, Denise, Larissa, Aline, Hérica, Josélia, Alex, Paula, Priscila, and Gustavo, and from São João del-Rei city, Profa. Leila, Profa. Renata, and Graziela. In special, to my old friends, Thiago, Leonardo, Gisele, Ana, Janaína, and Daniela, for being always the same even far away.

I would like to thank all people from the Gembloux Agro-Bio Tech - University of Liège, especially Sylvie, Hedi, Hélène, Sébastien, Yansen, and Axelle, for receiving me so well in the laboratory, for all help and the opportunity to learn from you. Thanks to good friends Rodrigo, Ana Carolina, Edvaldo, and Darlene, who gave me support and shared good times of happy hour every week in Belgium. Special thanks to Rodrigo and his wife Ana Carolina for giving me generous support since I left Brazil until the last day living in Belgium. Thanks for the excellent reception in their house, and for all the help in everything I needed during the period of my life there. Thank you so much for your friendship.

I am thankful for the scholarship from the National Council of Technological and Scientific Development (CNPq, Brazil) and I appreciate the financial and technical support provided by the National Fund for Scientific Research (Brussels, Belgium), Walloon Breeding Association (AWE, Ciney, Belgium), and Coordination for the Improvement of Higher Education Personnel (CAPES, Brazil).

This journey would not have been possible without the support of all of you!

BIOGRAPHY

José Teodoro de Paiva was born in São Gonçalo, Rio de Janeiro, and he grew up in Capela Nova, Minas Gerais, Brazil. In August 2009, he started his undergraduate studies in Interdisciplinary in Biosystems, and Animal Science at Universidade Federal de São João del-Rei (UFSJ), located in São João del-Rei, Minas Gerais, Brazil. Since the beginning of his Bachelor's degree, he was involved in research, academic, and extension activities, including classes mentoring, internship in cell biology field under the supervision of Prof. Patrícia Maria d'Almeida Lima, and in the animal breeding field under the supervision of the Prof. Leila de Genova Gaya. He helped in the foundation of the Animal Breeding Group (GMA) at the UFSJ. He has completed his internship at the Luiz de Queiroz College of Agriculture (ESALQ-USP) under the supervision of Prof. Gerson Barreto Mourão.

In March 2015, he started his studies in Genetics and Breeding Graduate Program at Universidade Federal de Viçosa (UFV), under the supervision of Prof. Marcos Deon Vilela de Resende. During his Master's degree, he participated as disclosure coordinator in the Genetics and Breeding Study Group (GenMelhor) at the UFV.

In March 2017, he started his Ph.D at Animal Science Graduate Program at UFV, under the supervision of Prof. Paulo Sávio Lopes. During his Ph.D, he participated as vice-coordinator in the Genetic and Animal Breeding Discussion Group (GDMA) at the UFV. He had the opportunity to work on two research projects in partnership with 1) EMBRAPA Dairy Cattle, under the supervision of Dra. Maria Gabriela Campolina Diniz Peixoto, and 2) University of Liège, under the supervision of Prof. Nicolas Gengler. The results from the second project are summarized in this thesis. On February 23, 2021, he will present his Ph.D thesis to the defense committee.

ABSTRACT

PAIVA, José Teodoro de, D.Sc., Universidade Federal de Viçosa, February, 2021. **Genetic and genomic evaluation for milk production and fatty acid traits predicted by mid-infrared (MIR) spectroscopy in Holstein dairy cattle.** Adviser: Paulo Sávio Lopes. Co-advisers: Nicolas Gengler and Fabyano Fonseca e Silva.

Mid-infrared (MIR) spectroscopy is the current main tool that has been used to get access of rapid, cost-effective, and high-throughput data of milk composition. Over the last decade, milk fatty acids (FA) have been predicted by MIR, allowing to record milk quality data at the population level. Interest in milk FA profile has increasing given its nutritional value, technological properties, and its use as biomarker of the cow’s status. The availability of these phenotypes makes possible their inclusion in genomic evaluations, which brings unprecedent and substantial impacts to improve milk quality. Therefore, the general objective of this thesis was to perform genetic and genomic evaluations for milk production and FA traits predicted by MIR using random regression models (RRM) in dairy cattle. Firstly, it was investigated different Legendre polynomials orders to better modeling of random effects in first lactation cows. Third-order Legendre polynomials seem to be most parsimonious and sufficient to describe milk production and FA traits over days in milk. Lower Spearman correlations at the beginning of lactation suggest the optimal model appeared to be even more important in the case of selection in this period. In addition, optimal polynomial orders tend to show lower residual variation, which can help to avoid overestimation at beginning of lactation. The effects of permanent environment and herd-year of calving had a high impact on early lactation. Heritability curves indicated that as long as lactation progressed the proportion of genetic variance increased for all traits. In a second step, it was investigated the potential implications of selection for milk production traits about FA across the first lactation through bi-trait RRMs using pedigree and genomic information. Selection for higher milk yield would decrease FA. Improving the milk FA profile would seem to be an effective way by indirect selection of fat yield, fat and protein content. Subsequently, genomic predictions using the single-step genomic best linear unbiased prediction (ssGBLUP) approach were performed based on RRM. It was investigated the performance of genomic predictions (in terms of reliability and bias) using ssGBLUP approach and it was compared with the parent average (PA) method. Moreover, different scaling and weighting factors to be used in the construction of the **H** matrix were tested. The test-day ssGBLUP approach yielded higher validation reliability compared to PA method for young bulls, even when no scaling and weighting factors were used in the **H** matrix.

In addition, choosing optimal parameters led to less biased prediction (regression coefficient close to 1) for genomic evaluation of milk production traits. Nonetheless, inflated GEBVs were still observed for milk FA. The findings reported in this thesis will contribute to advance on the modeling of milk production and milk FA traits in Walloon Holstein dairy cattle by inclusion of genomic information. Results from this research suggest that changes in milk FA profile can be achieved by the direct selection or indirect by selecting for fat yield and fat content. Moreover, this thesis provides the first results about the impact of different ssGBLUP methods (i.e., different scaling and weighting factors) based on RRM for genomic prediction of milk FA. In summary, our results demonstrated the superiority of ssGBLUP approach based on RRM in prediction performance of milk FA and supports further studies in order to improve reliabilities and reduce bias for milk FA.

Keywords: Genetic parameters. Genomic prediction. Milk quality. Random regression model. Reliability. Single-step GBLUP.

RESUMO

PAIVA, José Teodoro de, D.Sc., Universidade Federal de Viçosa, fevereiro de 2021. **Avaliação genética e genômica para características de produção de leite e ácidos graxos preditos via espectroscopia de infravermelho médio (MIR) em bovinos leiteiros Holstein.** Orientador: Paulo Sávio Lopes. Coorientadores: Nicolas Gengler e Fabyano Fonseca e Silva.

A espectroscopia de infravermelho médio (MIR) é a principal ferramenta atualmente usada para acesso à dados rápidos, econômicos e de alto rendimento da composição do leite. Na última década, o perfil de ácidos graxos (FA) no leite foi predito pela MIR, permitindo a coleta de dados de qualidade do leite em nível populacional. O interesse no perfil de FA tem aumentado devido ao seu valor nutricional, propriedades tecnológicas e seu uso como biomarcador do estado da vaca. A disponibilidade desses fenótipos possibilita sua inclusão em avaliações genômicas, o que traz impactos inéditos e substanciais para a melhoria da qualidade do leite. Portanto, o objetivo geral desta tese foi realizar avaliações genéticas e genômicas para as características de produção do leite e de FA preditas via MIR usando modelo de regressão aleatória (RRM) em bovinos leiteiros. Primeiramente, foram investigadas diferentes ordens de polinômios de Legendre para modelagem de efeitos aleatórios em vacas de primeira lactação. Os polinômios de ordem três parecem ser mais parcimoniosos e suficientes para descrever as características de produção de leite e de FA ao longo dos dias em lactação. Correlações de Spearman mais baixas no início da lactação sugerem que o uso do melhor modelo parece ainda ser mais importante no caso de seleção neste período. Além disso, as melhores ordens de ajuste tendem a apresentar menor variação residual, o que pode ajudar a evitar superestimação no início da lactação. Os efeitos de ambiente permanente e de rebanho-ano de parto tiveram um alto impacto no início da lactação. As curvas de herdabilidade indicaram que à medida que a lactação progredia a proporção da variância genética aumentava para todas as características. Em uma segunda etapa, foram investigadas as potenciais implicações da seleção para as características de produção de leite sobre FA ao longo da primeira lactação por meio de RRMs bicaracterístico usando pedigree e informações genômicas. A hipótese em questão é de que a seleção para maior produção de leite diminuiria os ácidos graxos. O perfil de FA pode ser melhorado por meio de seleção indireta para produção de gordura e teores de gordura e proteína. Posteriormente, predições genômicas usando melhor predição linear não-viesada genômica em passo único (ssGBLUP) foram feitas por meio de RRM. Foi investigado o desempenho da predição genômica (em termos de confiabilidade e viés) usando abordagem ssGBLUP o qual foi comparado com a média dos pais tradicional (PA). Foram testados diferentes fatores de escala e ponderação na construção da matriz **H**. O ssGBLUP baseado em RRMs aumentou a confiabilidade de validação comparado ao método PA para touros jovens, mesmo quando não foi usado nenhum fator de escala e ponderação na matriz **H**. Além disso, a escolha de parâmetros ótimos resultou em menor viés de predição (coeficiente de regressão próximo de 1) na avaliação genômica de características de produção de leite. Todavia, GEBVs inflados ainda foram observados para os FA do leite. Os achados descritos nesta tese contribuirão para avanços na modelagem das características de produção de leite e FA em vacas Holstein da região da Valônia, sul da Bélgica, por meio da inclusão de informações genômicas. Os resultados desta pesquisa sugerem que mudanças no perfil de FA do leite podem ser alcançadas pela seleção direta ou indireta para produção e teor de gordura. Além disso, esta tese fornece os primeiros resultados sobre o impacto de diferentes métodos ssGBLUP (ou seja, diferentes fatores de escala e ponderação) baseados em RRM na predição genômica de FA do leite. Em resumo, nossos resultados demonstraram a superioridade do método ssGBLUP usando RRM na predição genômica de FA do leite e dão suporte a futuros estudos para melhorar a confiabilidade e reduzir o viés para os FA do leite.

Palavras-chave: Confiabilidade. Modelo de regressão aleatória. Parâmetros genéticos. Predição genômica. Qualidade do leite. Single-step GBLUP.

SUMMARY

CHAPTER 1

1.1. General introduction

Milk production is one of the main activities of world agribusiness and global production reached nearly 860 million tonnes in 2020, an increase of 1.4 percent from 2019 (FAO, 2020). Dairy industry has great socioeconomic importance among the food industry sectors and world dairy exports reached 78 million tonnes (milk equivalent) in 2020 (FAO, 2020). The increase is attributed primarily to the persistently high global import demand reflecting increased economic activities. Bovine milk and dairy products have a long tradition and significant importance in human nutrition, being a complex food whose components may have several benefits in health. Milk components take part in metabolism in several ways, by providing essential amino acids, vitamins, minerals, and fatty acids, or by affecting absorption of other nutrients (Haug et al., 2007). Currently, there is an increase in consumers' concern about health-properties of food, and the market of food with special health benefits is growing. Recent trends show that consumers and the dairy industry have shifted the concept of quality and traits like milk coagulation to improve cheese production and fatty acids to enrich the nutritional value for human health became more important (De Marchi et al., 2014).

Interest in the milk fatty acids profile is increasing given its importance for milk quality and nutrition, but also because of the technological properties of milk and dairy products (Soyeurt et al., 2011). Based on the number of double bonds present in their chemical structure, fatty acids can be classified as saturated (no double bond), monounsaturated (one double bond) and polyunsaturated (more than one double bond). Typically, cow's milk contains 70% saturated fatty acids, 25% monounsaturated and 5% polyunsaturated fatty acids (Soyeurt et al., 2006). Studies pointed to an increase in the risk of cardiovascular diseases due to the consumption of some saturated fatty acids, however monounsaturated and polyunsaturated fatty acids have potentially beneficial effects on human health (Hanuš et al., 2018). The oleic and linoleic unsaturated fatty acids and the conjugated linoleic acid isomers are associated with cholesterol reduction and anticarcinogenic effects (Haug et al., 2007). In addition to these effects on human health, the increase in the proportion of unsaturated fatty acids in bovine milk also has a positive impact on the technological properties of dairy products, such as butter (Hanuš et al., 2018). The fatty acids composition in cow's milk does not correspond to the profile of fatty acids provided in the diet and it can be changed through genetic improvement. Several studies reported the existence of additive genetic variation for milk components in dairy cattle and the promisor use of these traits as selection criteria in order to increase the quality of dairy products for human consumption (Soyeurt et al., 2011; Bastin et al., 2013; Narayana et al., 2017).

Changes in milk composition can be used for precise monitoring of the status and health of the cows, and may therefore, it plays the role of indicator traits or milk biomarkers. Major conventional milk components that have been available for a long time, such as fat, protein and lactose, were considered as potential tools to asses the changes in the status of the cows (Gengler et al., 2016). Milk fatty acids are related to stage of lactation and most transition dairy cows are confronted with a negative energy balance (NEB) in early lactation, which may lead to metabolic disorders and impaired fertility (Van Haelst et al., 2008; Stádník et al., 2015). Jorjong et al. (2014) have proposed the use of milk fatty acids (e.g., C18:1 cis-9) as an early warning biomarker for compromised metabolic status in dairy cows. In addition, Hammami et al. (2015) reported that differentiated response of C18:1 cis-9 to heat stress could be an indicator of the equilibrium between feed intake and reserve mobilization under warm conditions. Therefore, these and other findings (e.g., Bastin et al., 2012; Jorjong et al., 2015; Lainé et al., 2017) reinforce that some milk fatty acids reflect body reserve mobilization and fertility.

Until recently, the most widely used method to quantify the qualitative components of milk was based on gas chromatography. Although the method is accurate and generates detailed fatty acid profiles, it is expensive and time-consuming. Thus, gas chromatography is not economically feasible for routine analysis of a large number of samples, which is necessary for the genetic evaluation (Soyeurt et al., 2011). Methods that routinely and accurately measure the animal traits are essential to address emerging challenges in dairy breeding. Mid-infrared spectroscopy (MIR) is fast, cheap and high-throughput method for chemical analysis, which has led to its application on milk quality phenotyping (De Marchi et al., 2014; Gengler et al., 2016). This method is based on the interaction between matter and electromagnetic radiation, with subsequent measurement of energy absorption. MIR has been evaluated as a potential tool to record data at the population level and may become one of the main tools for evaluating milk quality.

Developed countries have been using MIR in quality analysis systems of milk stored on cooling tanks and cows. These systems accurately report the content of protein, fat, and lactose, which are the main components of milk. In addition, MIR spectroscopy has also been used to predict a range of high impact traits for industry, environment, and animal health, such as milk fatty acids (Gengler et al., 2016). MIR has been characterized as a high-performance method to collect phenotypic data, and it is becoming one of the most used tools in dairy science. Several milk analysis laboratories involved in dairy cattle control systems have been storing spectral data to produce "a posteriori" several phenotypes of interest for herds management, reducing production costs, improving milk quality, increase animal welfare, and mitigating environmental impacts (De Marchi et al., 2014).

The availability of these specialized phenotypes makes possible their inclusion in the genetic evaluations, which brings unprecedented and substantial impacts to improve milk quality. Genomic selection has become the main strategy in livestock selective breeding by allowing the selection of candidate bulls at younger ages (Hayes et al., 2009). Single-step genomic best linear unbiased prediction (ssGBLUP; Aguilar et al., 2010; Christensen and Lund, 2010; Misztal et al., 2009) is among the most efficient methods and has been routinely used in animal breeding. This method incorporates all available phenotypic, genotypic and pedigree information in the same step. The use of extra information on non-genotyped animals and the independency of pseudo-phenotypes are responsible for gains in accuracy over other genomic methods (Legarra et al., 2014). Furthermore, higher accuracies were reported by Lourenco et al. (2014) for ssGBLUP compared to multiple-step approaches.

Therefore, associated with the genotyping these new phenotypes can be used for genomic selection which proves even more expressive genetic gains. Moreover, for traits limited by sex (e.g., milk production and fatty acid traits) and expensive-to-measure (e.g., direct individual FA), the prediction of more accurate GEBV for young bulls using phenotypes predicted by MIR from relatives is an interesting promising approach (Freitas et al., 2020). The identification of genetically superior animals for milk quality traits is paramount in view of the current payment policy based on milk quality. An increase in reliability of genomic breeding value (GEBV) over the parent average breeding values (EBV) has been observed for young animals in dairy cattle (Oliveira et al., 2019). However, few studies (e.g., Cesarani et al., 2019; Freitas et al., 2020; Gebreyesus et al., 2019) of genomic selection have been carried out on milk FA in dairy cattle.

1.2. Thesis objectives

The general objective with this thesis was to perform genetic and genomic evaluations for milk production and fatty acid traits predicted by MIR spectroscopy using test-day random regression models in first lactation Walloon Holstein dairy cattle. The specific objectives were to: i) evaluate random regression models using Legendre orthogonal polynomial of different orders for modeling random effects curves (herd-year of calving, additive genetic, and permanent environment effects) in order to identify the most suitable and parsimonious models for routine genetic evaluation, and also to estimate variance components and genetic parameters for milk production and fatty acid traits; ii) estimate genetic correlations among milk production and fatty acid traits over days in milk, and also to investigate the reliability and bias of daily GEBVs estimated using single-step GBLUP method compared to traditional BLUP method.

1.3. References

Aguilar, I., I. Misztal, D.L. Johnson, A. Legarra, S. Tsuruta, and T.J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci. 93:743–752. doi:10.3168/jds.2009-2730.

Bastin, C., D.P. Berry, H. Soyeurt, and N. Gengler. 2012. Genetic correlations of days open with production traits and contents in milk of major fatty acids predicted by mid-infrared spectrometry. J. Dairy Sci. 95:6113–6121. doi:10.3168/jds.2012-5361.

Bastin, C., H. Soyeurt, and N. Gengler. 2013. Genetic parameters of milk production traits and fatty acid contents in milk for Holstein cows in parity 1 - 3. J. Anim. Breed. Genet. 130:118–127. doi:10.1111/jbg.12010.

Cesarani, A., G. Gaspa, F. Correddu, M. Cellesi, C. Dimauro, and N.P.P. Macciotta. 2019. Genomic selection of milk fatty acid composition in Sarda dairy sheep: Effect of different phenotypes and relationship matrices on heritability and breeding value accuracy. J. Dairy Sci. 102:3189–3203. doi:10.3168/jds.2018-15333.

Christensen, O.F., and M.S. Lund. 2010. Genomic relationship matrix when some animals are not genotyped Genomic prediction models. Genet. Sel. Evol. 42:1–8.

FAO. 2020. Food Outlook - Biannual Report on Global Food Markets - November 2020. Rome. https://doi.org/10.4060/cb1993en.

Freitas, P.H.F., H.R. Oliveira, F.F. Silva, A. Fleming, F. Miglior, F.S. Schenkel, and L.F. Brito. 2020. Genomic analyses for predicted milk fatty acid composition throughout lactation in North American Holstein cattle. J. Dairy Sci. 103:6318–6331. doi:10.3168/jds.2019-17628.

Gebreyesus, G., H. Bovenhuis, M.S. Lund, N.A. Poulsen, D. Sun, and B. Buitenhuis. 2019. Reliability of genomic prediction for milk fatty acid composition by using a multi-population reference and incorporating GWAS results. Genet. Sel. Evol. 51:16. doi:10.1186/s12711-019-0460-z.

Gengler, N., H. Soyeurt, F. Dehareng, C. Bastin, F. Colinet, H. Hammami, M.-L. Vanrobays, A. Lainé, S. Vanderick, C. Grelet, A. Vanlierde, E. Froidmont, and P. Dardenne. 2016. Capitalizing on fine milk composition for breeding and management of dairy cows. J.

Dairy Sci. 99:4071–4079. doi:10.3168/jds.2015-10140.

Van Haelst, Y.N.T., A. Beeckman, A.T.M. Van Knegsel, and V. Fievez. 2008. Short Communication: Elevated Concentrations of Oleic Acid and Long-Chain Fatty Acids in Milk Fat of Multiparous Subclinical Ketotic Cows. J. Dairy Sci. 91:4683–4686. doi:10.3168/jds.2008-1375.

Hammami, H., J. Vandenplas, M.-L. Vanrobays, B. Rekik, C. Bastin, and N. Gengler. 2015. Genetic analysis of heat stress effects on yield traits, udder health, and fatty acids of Walloon Holstein cows. J. Dairy Sci. 98:4956–4968. doi:10.3168/jds.2014-9148.

Hanuš, O., E. Samková, L. Křížová, L. Hasoňová, and R. Kala. 2018. Role of Fatty Acids in Milk Fat and the Influence of Selected Factors on Their Variability—A Review. Molecules 23:1636. doi:10.3390/molecules23071636.

Haug, A., A.T. Høstmark, and O.M. Harstad. 2007. Bovine milk in human nutrition – a review. Lipids Health Dis. 6:25. doi:10.1186/1476-511X-6-25.

Hayes, B.J., P.J. Bowman, A.J. Chamberlain, and M.E. Goddard. 2009. Invited review : Genomic selection in dairy cattle : Progress and challenges. J. Dairy Sci. 92:433–443. doi:10.3168/jds.2008-1646.

Jorjong, S., A.T.M. van Knegsel, J. Verwaeren, R.M. Bruckmaier, B. De Baets, B. Kemp, and V. Fievez. 2015. Milk fatty acids as possible biomarkers to diagnose hyperketonemia in early lactation. J. Dairy Sci. 98:5211–5221. doi:10.3168/jds.2014-8728.

Jorjong, S., A.T.M. van Knegsel, J. Verwaeren, M.V. Lahoz, R.M. Bruckmaier, B. De Baets, B. Kemp, and V. Fievez. 2014. Milk fatty acids as possible biomarkers to early diagnose elevated concentrations of blood plasma nonesterified fatty acids in dairy cows. J. Dairy Sci. 97:7054–7064. doi:10.3168/jds.2014-8039.

Lainé, A., C. Bastin, C. Grelet, H. Hammami, F.G. Colinet, L.M. Dale, A. Gillon, J. Vandenplas, F. Dehareng, and N. Gengler. 2017. Assessing the effect of pregnancy stage on milk composition of dairy cows using mid-infrared spectra. J. Dairy Sci. 100:2863–2876. doi:10.3168/jds.2016-11736.

Legarra, A., O.F. Christensen, I. Aguilar, and I. Misztal. 2014. Single Step, a general approach for genomic selection. Livest. Sci. 166:54–65. doi:10.1016/j.livsci.2014.04.029.

Lourenco, D.A.L., I. Misztal, S. Tsuruta, I. Aguilar, E. Ezra, M. Ron, A. Shirak, and J.I. Weller. 2014. Methods for genomic evaluation of a relatively small genotyped dairy population and effect of genotyped cow information in multiparity analyses. J. Dairy Sci. 97:1742–1752. doi:10.3168/jds.2013-6916.

De Marchi, M., V. Toffanin, M. Cassandro, and M. Penasa. 2014. Invited review: Mid-infrared spectroscopy as phenotyping tool for milk traits. J. Dairy Sci. 97:1171–1186. doi:10.3168/jds.2013-6799.

Misztal, I., A. Legarra, and I. Aguilar. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J. Dairy Sci. 92:4648–4655. doi:10.3168/jds.2009-2064.

Narayana, S.G., F.S. Schenkel, A. Fleming, A. Koeck, F. Malchiodi, J. Jamrozik, J. Johnston, M. Sargolzaei, and F. Miglior. 2017. Genetic analysis of groups of mid-infrared predicted fatty acids in milk. J. Dairy Sci. 100:4731–4744. doi:10.3168/jds.2016-12244.

Oliveira, H.R., D.A.L. Lourenco, Y. Masuda, I. Misztal, S. Tsuruta, J. Jamrozik, L.F. Brito, F.F. Silva, and F.S. Schenkel. 2019. Application of single-step genomic evaluation using multiple-trait random regression test-day models in dairy cattle. J. Dairy Sci. 102:2365–2377. doi:10.3168/jds.2018-15466.

Soyeurt, H., P. Dardenne, A. Gillon, C. Croquet, S. Vanderick, P. Mayeres, C. Bertozzi, and N. Gengler. 2006. Variation in Fatty Acid Contents of Milk and Milk Fat Within and Across Breeds. J. Dairy Sci. 89:4858–4865. doi:10.3168/jds.S0022-0302(06)72534-6.

Soyeurt, H., F. Dehareng, N. Gengler, S. McParland, E. Wall, D.P. Berry, M. Coffey, and P. Dardenne. 2011. Mid-infrared prediction of bovine milk fatty acids across multiple breeds, production systems, and countries. J. Dairy Sci. 94:1657–1667. doi:10.3168/jds.2010-3408.

Stádník, L., J. Ducháček, J. Beran, R. Toušová, and M. Ptáček. 2015. Relationships between milk fatty acids composition in early lactation and subsequent reproductive performance in Czech Fleckvieh cows. Anim. Reprod. Sci. 155:75–79. doi:10.1016/j.anireprosci.2015.02.002.

CHAPTER 2

Random regression test-day models to describe milk production and fatty acid traits in first lactation Walloon Holstein cows

Paiva, J.T.[1]; Mota, R.R.[2a]; Lopes, P.S.[1]; Hammami, H.[2]; Vanderick, S.[2]; Oliveira, H.R.[3]; Veroneze, R.[1]; Silva, F.F.[1]; Gengler, N.[2]

[1] Department of Animal Sciences, Universidade Federal de Viçosa, Viçosa, MG, Brazil

[2] Gembloux Agro-Bio Tech, University of Liège, TERRA Teaching and Research Centre, B-5030 Gembloux, Belgium

[3] Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, Ontario, Canada

[a] Present address: Council on Dairy Cattle Breeding – CDCB, Bowie, Maryland, USA

2.1. Abstract

We aimed to investigate the use of different Legendre polynomial orders to model random effects as well as to estimate genetic parameters for milk production and fatty acid (FA) traits. First lactations of Walloon Holstein cows were used in this study (302,684 test-day records). The studied traits were milk yield, fat and protein contents, and fatty acids generated by MIR spectroscopy: C16:0, C18:1 *cis*-9, LCFA, SFA, and UFA. The models included random regression coefficients for herd-year of calving, additive genetic, and permanent environment effects. The selection of the best random regression model (RRM) was based on the Deviance Information Criterion (DIC). Additional residual and Spearman correlations analysis were performed. Variance components and genetic parameters were obtained by using the best fit RRM via Bayesian approach. For all analyzed random effects, DIC values decreased as the order of the Legendre polynomials increased. Best fit models had fifth-order (degree 4) for permanent environment effect, and ranged from second to fifth-order (degree 1 to 4) for additive genetic and herd-year of calving (LEG555 for milk yield and protein content; LEG335 for fat content and SFA; LEG545 for C16:0 and UFA; and LEG535 for C18:1 cis-9 and LCFA). Based on the best fit models, an effect of overcorrection was observed in early lactation (5 to 35 days in milk (DIM)). On the other hand, third-order (LEG333; degree 2) models showed flat residuals trajectory over the lactation. Additive genetic variance used to estimate the heritability (h^2) curves indicated that the genetic variance increased over the lactation. Fat and protein contents (h^2 from 0.11 to 0.58) were more heritable than milk yield (h^2 from 0.12 to 0.29). Milk fatty acids heritabilities ranged from low to high magnitude throughout lactation (h^2 from 0.03 to 0.56). In addition, there were high (values > 0.90 for all bulls and > 0.97 for top 100) Spearman correlations among breeding values for 155 and 305 DIM, between the best RRM of each trait and the third-order (LEG333) model. It was also observed a high percentage of coincidence for top 10% bulls. Moreover, the first three eigenvalues explained over 97.8% of the additive genetic effect, 96.4% of the herd-year of calving effect, and 91.2% of the permanent environment effect for milk production and fatty acid traits. In summary, the use of optimal polynomial orders tends to show lower residual variation, which can help to avoid overestimation at the beginning of lactation. Therefore, third-order Legendre orthogonal polynomials seems to be most parsimonious and sufficient to describe milk production and fatty acid traits over DIM in first lactation Walloon Holstein cows.

Keywords: (co)variance function; dairy cattle; genetic parameters; mid-infrared spectroscopy

2.2. Introduction

Mid-infrared (MIR) spectroscopy is the current main tool that has been used worldwide to get access of rapid, cost-effective, and high-throughput data of milk composition (De Marchi et al., 2014; Gengler et al., 2016). Bovine milk is a complex food whose components may have several effects on health, which might reflect in its nutritional content and market value. Over the years, most dairy cattle genetic selection programs have been prioritizing milk fat and protein contents to determine the value of raw milk in pricing schemes worldwide (Gengler et al., 2016). Moreover, global consumption has move towards the increase of fat and protein yields within lower milk volumes. Most of market signals a projection that will continue to place greater emphasis on the future in selecting milk solids rather than yield (Britt et al., 2018).

Besides milk composition, fatty acids (FA) contents have been investigated as important indicators that allow precise monitoring of fertility and health status of cows, as well as milk technological properties (Bastin et al., 2016). Over the last decades, Soyeurt et al. (2006; 2011) and Grelet et al. (2017), respectively, have developed and improved MIR spectral data calibration equations to predict milk FA with high accuracy. Predicted milk FA can be used in several country's regular milk payments systems (R^2v; validation coefficient of determination >95%) and animal breeding programs (R^2cv; cross-validation coefficient of determination

>75%) (Soyeurt et al., 2011). The interest in milk FA have increased due to its importance for milk quality, as fatty acids influence the milk nutritional value. Studies have pointed out an increase of cardiovascular diseases risk due to the consumption of some saturated fatty acids (SFA), whereas unsaturated fatty acids (UFA) have being potentially beneficial effects on human health (Hanuš et al., 2018). In addition, higher levels of UFA have a positive impact on technological properties of dairy products (De Marchi et al., 2014).

The evolution of milk FA contents over the lactation can be related to the cows' physiology and energy balance status (Dórea, French, & Armentano, 2017). This information can be useful for breeding and management purposes of dairy cows (Gengler et al., 2016). Dairy cattle breeding is permanently adapting to challenges on milk production. Negative energy balance (NEB) remains as one of the most challenges to the ever-increasing lactation productivity (Dórea et al., 2017; Loften et al., 2014). A severe NEB is related to a greater risk of metabolic disorders compromising the cow's health and fertility, and reduced productive lifespan of dairy cows after calving (Xu et al., 2018). Previous studies reported that milk fat contents and fatty acids profiles in early lactation are affected by the postpartum energy balance status. Identified as valuable biomarkers of cow's biological or physiological process, milk fatty acids have been well described in literature as an indicator of metabolic disorders and changes in the cow's status. According to Jorjong et al. (2014), most cows experience a state of NEB during the periparturient period. This may lead to the mobilization of body fat reserves in early lactation, and consequently increase fat percentage in milk and long-chain fatty acid (LCFA) supply in particular.

A preliminary study conducted by Van Haelst et al. (2008) showed that high milk fat C18:1 *cis*-9 concentration as well as higher levels of LCFA indicated subclinical ketosis in cows. Jorjong et al. (2014) supported these findings and additionally reported that milk fat C18:1 *cis*-9 seems to be a potential early warning biomarker for compromised metabolic status in dairy cows during the first two months of lactation. During body fat mobilization, an excessive non-esterified fatty acids (NEFA) amount, particularly rich in LCFA, is transferred to the milk. This leads to higher proportions of the major C16:0 and C18:1 *cis*-9. Jorjong et al. (2015) have reported that the milk fat C18:1 *cis*-9-to-C15:0 ratio revealed the most discriminating factor for the diagnosis of hyperketonemia. Moreover, C16:0 and C18:1 *cis*-9 are among the most abundant FA in bovine milk fat (32.6% weight and 18.0% weight, on average, respectively) (Tzompa-Sosa et al., 2014). Bastin et al. (2012) reported unstable genetic associations between fertility and milk fat content and FA over the lactation curve. These authors suggested that C18:1 *cis*-9 could be a useful indicator of reproductive performance, once NEB is associated with reduced fertility. Hammami et al. (2015) also supported the use of C18:1 *cis*-9 as a good milk biomarker for heat stress in dairy cattle, which express the equilibrium between intake and body reserve mobilization under warm conditions.

Thereby, desirable changes to milk production and FA traits require knowledge on various factors that influence them. Some studies have provided evidences that the additive genetic effect is responsible for a large proportion of the FA phenotypic variation over the lactation curve in dairy cows (Bastin et al., 2013; Freitas et al., 2020; Hammami et al., 2015). The genetic evaluation of longitudinal traits has been mainly focused on using random regression models (RRM). According to Schaeffer (2004), the use of RRM enables to fit random genetic and environmental effects over time. This allows to achieve higher accuracy of estimated breeding values than in multi-trait and repeatability models (Oliveira et al., 2019). Moreover, RRM provides information about temporal variation of biological processes underlying the studied traits to exploit in management and breeding purposes (Oliveira et al., 2019). Although changes in genetic parameters for FA over lactation have been previously reported, studies comparing (co)variance functions to better modeling the random part of milk FA traits in dairy cattle are still scarce (Arnould et al., 2010; Narayana et al., 2017).

To our best knowledge there is no study investigating different orders of Legendre polynomials to describe each random effect in the genetic evaluation of milk FA traits in dairy cattle. Therefore, the aims of this study were to: 1) select the best Legendre polynomial order to describe the (co)variance structure of each random effect, and 2) estimate genetic parameters by using the best fit model for milk production and FA traits in the first lactation of Walloon Holstein cows.

2.3. Material and methods

2.3.1. Phenotypes and pedigree

Phenotypic data were extracted from the Walloon genetic evaluation performed in Belgium, in which milk samples are routinely collected by the Walloon Breeding Association (AWE, Ciney, Belgium). These milk samples are analyzed by using a mid-infrared MilkoScan FT6000 spectrometer (Foss, Hillerød, Denmark), and the generated spectral dataset are used to prediction of FA contents in milk (g/dL of milk) by applying the calibration equations developed by Soyeurt et al. (2011). Fat content (%), protein content (%), C16:0 fatty acid (palmitic acid), C18:1 *cis*-9 fatty acid (oleic acid), long-chain fatty acids (LCFA), saturated fatty acids (SFA), and group unsaturated fatty acids (UFA) were the traits extracted from milk samples and evaluated in this study besides daily milk yield (kg).

Data was recorded between January 2012 and February 2019. The dataset was edited in order to accommodate records between 5 and 305 days in milk (DIM). Cows were required to have records for all traits during at least 3 test-days. Herds with less than 15 cows were removed from the dataset. To eliminate outliers, only predicted FA values above the first and below the 99th percentile were kept (Bastin et al., 2013). The absolute number of cows (%) per number of test-day records were: 3 records = 13,956 (21.80%), 4 records = 16,554 (25.90%), 5 records = 15,812 (24.80%), 6 records = 10,279 (16.10%), 7 records = 4,979 (7.80%), 8 records = 1,702

(2.70%), 9 records = 533 (0.80%), and 10 records = 60 (0.10%). The final dataset comprised 302,684 test-days records from 63,875 first-parity Walloon Holstein cows, from 856 herds. Descriptive statistics of milk production and FA traits are presented in Table 1.

The pedigree was extracted from the official Walloon genetic evaluation data as well. The used pedigree-based relationship matrix contained 228,114 animals, which are 6,509 sires and 114,169 dams. The average inbreeding was 0.01% for all animals (n = 228,114), and 24.47% for inbred animals (n = 96). Although inbreeding ranged from 12.5 to 25% for inbred animals, 99.96% of the animals are not inbred.

Table 1. Descriptive statistics of test-day records for milk production and fatty acids (g/dL of milk) traits in first lactation Walloon Holstein cows

Traits[1]	**Mean**	**SD[1]**	**CV[1] (%)**	**Minimum**	**Maximum**
Milk (kg)	23.32	5.63	24.15	3.40	78.00
Fat (%)	3.95	0.54	13.59	2.28	8.87
Protein (%)	3.39	0.29	8.71	1.91	4.95
C16:0 (x100)	1.263	0.238	18.84	0.678	1.915
C18:1 *cis*-9 (x100)	0.784	0.141	17.98	0.484	1.453
LCFA (x100)	1.592	0.259	16.27	1.004	2.705
SFA (x100)	2.801	0.457	16.31	1.277	4.484
UFA (x100)	1.234	0.201	16.29	0.665	3.197

[1] Milk = milk yield; Fat = fat content; Protein = protein content; 16:0 = palmitic acid; C18:1 *cis*-9 = oleic acid; LCFA = long-chain fatty acids; SFA = saturated fatty acids; UFA = unsaturated fatty acids; SD = standard deviation; CV = coefficient of variation.

2.3.2. Statistical models

The following single-trait random regression test-day model was used based upon the official Walloon genetic evaluations for production traits as described by Auvray and Gengler (2002) and Croquet et al. (2006):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Q}\,(\mathbf{Wh} + \mathbf{Za} + \mathbf{Zp}) + \mathbf{e}, \qquad (1)$$

where $\mathbf{y}$ is the vector of phenotypic records for each trait; $\boldsymbol{\beta}$ is the vector of systematic effects: herd x test-day (28,112 classes), gestation stage (12 classes), minor lactation stage (61 classes of 5 DIM), and major lactation stage (classes of 73 DIM) x age at calving (21–40 months) x season of calving (four seasons: winter from Jan-Mar, spring from Apr-Jun, summer from Jul-Sep and autumn from Oct-Dec) (8,820 classes); $\mathbf{h}$ is the vector of herd-year of calving (4,361 classes) random regression coefficients; $\mathbf{a}$ is the vector of additive genetic (228,114 animals) random regression coefficients; $\mathbf{p}$ is the vector of permanent environment (63,875 animals) random regression coefficients; $\mathbf{X}$, $\mathbf{W}$, and $\mathbf{Z}$ are the incidence matrices assigning observations to effects; $\mathbf{Q}$ is the covariate matrix for second to fifth-order Legendre polynomials; and $\mathbf{e}$ is the vector of random residuals.

The adopted model had the following distributions: $\mathbf{y} \mid \boldsymbol{\beta}, \mathbf{h}, \mathbf{a}, \mathbf{p}, \mathbf{e} \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Q}\,(\mathbf{Wh} + \mathbf{Za} + \mathbf{Zp}), \mathbf{I}\sigma_e^2)$; $\boldsymbol{\beta} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}})$, in which $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ is a diagonal matrix with values 1e+10 (large variances) to represent vague prior knowledge; $\mathbf{h} \mid \mathbf{H_0} \sim N(\mathbf{0}, \mathbf{H_0} \otimes \mathbf{I})$; $\mathbf{a} \mid \mathbf{G_0} \sim N(\mathbf{0}, \mathbf{G_0} \otimes \mathbf{A})$; $\mathbf{p} \mid \mathbf{P_0} \sim N(\mathbf{0}, \mathbf{P_0} \otimes \mathbf{I})$; and $\mathbf{e} \mid \sigma_e^2 \sim N(\mathbf{0}, \mathbf{I}\sigma_e^2)$, where $\mathbf{H_0}$, $\mathbf{G_0}$, and $\mathbf{P_0}$ are the random regression coefficients (co)variances matrices for herd-year of calving, additive genetic, and permanent environment effects of the animal, respectively; $\mathbf{A}$ is the pedigree-based relationship matrix; $\mathbf{I}$ is the identity matrix, σ_e^2 is the residual variance; and $\otimes$ is the Kronecker product. It was assumed that the regression coefficients (co)variance matrices follow an inverted Wishart distribution: $\mathbf{H_0} \mid v_h, \mathbf{V_h} \sim IW(v_h, \mathbf{V_h})$; $\mathbf{G_0} \mid v_a, \mathbf{V_a} \sim IW(v_a, \mathbf{V_a})$; $\mathbf{P_0} \mid v_p, \mathbf{V_p} \sim IW(v_p, \mathbf{V_p})$; with hyperparameters $\mathbf{V_h} = \hat{\mathbf{H}}_0$, $\mathbf{V_a} = \hat{\mathbf{A}}_0$, and $\mathbf{V_p} = \hat{\mathbf{P}}_0$, and the degrees of freedom $v_h = v_a = v_p = 2$, chosen to provide flat priors. The residual variance was assumed to be constant over lactation, which follows an inverted chi-square distribution $\sigma_e^2|S_e, n_e \sim \chi^{-2}(S_e, n_e)$.

Covariance functions for the random effects of herd-year of calving, additive genetic, and permanent environment were fit with Legendre orthogonal polynomials (Kirkpatrick et al., 1990). The models are presented according to the code LEG*hap*, where LEG is the basis function used (Legendre polynomial); *h*, *a*, and *p* are the random regression coefficients matrices orders for herd-year of calving, additive genetic, and permanent environment effects, respectively. Regression curves were fit by using combinations of distinct adjustment orders for Legendre orthogonal polynomials as follows: second-order or degree 1 (2 coefficients), third-order or degree 2 (3 coefficients), fourth-order or degree 3 (4 coefficients), and fifth-order or degree 4 (5 coefficients). A total of 28 RRM were tested by assuming homogeneity of residual variance for each studied trait. The assumption of homogeneous residual variance was evaluated based on the analysis of residual means described in the topic "*Analysis of residuals*", as suggested by Jamrozik et al. (1997) and reported by Soyeurt et al. (2008) using a subset of the data from the same population. The eigenvectors of the genetic, permanent environment and herd-year of calving covariance matrices were estimated using the highest order model and each eigenvalue was expressed as a percentage of the sum of all to determine its importance.

These models were implemented in a Bayesian approach via Gibbs sampling by using the gibbs2f90 software (Misztal et al., 2002). A chain length of 250,000 iterations, considering a burn-in period of 50,000, and a thin interval of 10 iterations were established for all analyses. Thus, 20,000 samples were used to estimate the posterior marginal distribution samples of genetic parameters. Convergence criteria of Gibbs chains were monitored by graphical inspection, and through Geweke (1991) test, both using the "*Bayesian Output Analysis"* package (Smith, 2007) available on R software (R Development Core Team, 2019, https://www.R-project.org/).

2.3.3. Bayesian model comparison

In order to identify the best single-trait RRM based on goodness-of-fit and degrees of parameterization, the Deviance Information Criterion (DIC; Spiegelhalter et al. (2002)) was used. The DIC is defined as follows:

$$\mathrm{DIC} = \bar{\mathrm{D}}(\theta) + \mathrm{p_D} = 2\bar{\mathrm{D}}(\theta) - \mathrm{D}(\theta), \tag{2}$$

where $\bar{\mathrm{D}}(\theta) = \mathrm{E}_{\theta|y}[\mathrm{D}(\theta)]$ (posterior expectation of the Bayesian deviance) and $\mathrm{D}(\theta) = -2\log \mathrm{p}(y \mid \theta)$ are indicative of the goodness of fit of the model, and $\mathrm{p_D}$ is the effective number of parameter representing penalty for increasing model complexity defined as $\mathrm{p_D} = \bar{\mathrm{D}}(\theta) - \mathrm{D}(\bar{\theta})$, which θ is the vector of model parameters and $\mathrm{D}(\bar{\theta})$ is the Bayesian deviance evaluated at the posterior mean of the parameters. Models with lower DIC value are preferred once it indicates a better goodness-of-fit at lower degree of complexity.

2.3.4. Analysis of residual

Additional evaluation of the best fit models was performed using the analysis of residuals. Residuals were calculated for each test-day from all studied traits, and from all cows as the difference between observed and estimated values in each DIM. The means and standard deviations over the lactation were also calculated. Average residuals can be used to determine the accuracy of the model, while variances of residuals can evaluate its precision (Jamrozik et al., 1997). The estimated residuals values were obtained by using the blupf90 software (Misztal et al., 2002).

2.3.5. Variance components and genetic parameters

Models selected as optimal for each trait (based on DIC criterion and analysis of residuals) were used to estimate variance components and genetic parameters over the lactation curve. The genetic (co)variance matrix for all DIM was obtained according to Druet et al. (2003), which is described as:

$$\mathbf{G^*} = \mathbf{Q}\mathbf{G_0}\mathbf{Q}', \qquad (3)$$

where **G*** is a 301 by 301 genetic (co)variance matrix for all DIM ranging from 5 to 305 day, and **Q** is a matrix with the values of the *n* coefficients of the used order Legendre polynomial for each DIM from 5 to 305 day for every trait. Likewise, the (co)variance matrices of herd-year of calving and permanent environment were obtained from the $\mathbf{H_0}$ and $\mathbf{P_0}$ matrices, respectively.

The estimates of heritability (h_j^2), permanent environment (pe_j) and herd-year of calving (hyc_j) effects at test-day *j* were calculated as ratios of variance estimates to the total phenotypic variance, as follow:

$$h_j^2 = \frac{\sigma_{a_j}^2}{\sigma_{hyc_j}^2 + \sigma_{a_j}^2 + \sigma_{p_j}^2 + \sigma_e^2}, \qquad (4)$$

$$pe_j = \frac{\sigma_{p_j}^2}{\sigma_{hyc_j}^2 + \sigma_{a_j}^2 + \sigma_{p_j}^2 + \sigma_e^2}, \qquad (5)$$

$$hyc_j = \frac{\sigma_{hyc_j}^2}{\sigma_{hyc_j}^2 + \sigma_{a_j}^2 + \sigma_{p_j}^2 + \sigma_e^2}; \qquad (6)$$

where $\sigma_{a_j}^2$, $\sigma_{p_j}^2$ and $\sigma_{hyc_j}^2$ are, respectively, the additive genetic, permanent environment and herd-year of calving variances at test-day *j*, and σ_e^2 is the residual variance.

2.3.6. Spearman correlations

Spearman correlations were calculated between the estimated breeding values (EBV) from the best fit Legendre polynomial model and EBVs from the lowest (degree 1) and third (degree 2) order models, in order to verify possible re-raking. Spearman correlations were calculated for the top 100 and all bulls for three different lactation stages: early (5 DIM), middle (155 DIM) and end (305 DIM). For each trait, EBVs of animal *i* at test-day *j* were obtained from a posteriori distribution of random regression coefficient estimates as follow:

$$\mathrm{EBV}_{ij} = \mathbf{C}_j\hat{\mathbf{a}}_i, \tag{7}$$

where $\mathbf{C}_j$ is a matrix of independent covariates associated with the Legendre polynomials for test-day *j*; and $\hat{\mathbf{a}}_i$ is a vector of BLUP solution of the additive genetic effect for Legendre polynomial coefficient for the animal *i*. The accumulated 305 DIM for animal *i* was obtained by summing the EBV for each test-day *j* as:

$$\mathrm{EBV}_{305} = \sum_{j=5}^{305} \mathrm{EBV}_{ij} \tag{8}$$

Complementarily, the lowest (degree 1) and third (degree 2) order models were also contrasted by percentage coincidence of the top 1%, 5%, 10%, and 50% animals.

2.4. Results

2.4.1. Bayesian model comparison

The results from all evaluated random regression models and their DIC values for milk production traits (milk yield, fat and protein contents), individual FA (C16:0 and C18:1 *cis*-9), and FA groups (LCFA, SFA, and UFA) are presented in Table 2.

Table 2. Parameters number (Np) and Deviance Information Criteria (DIC) for random regression models with different Legendre polynomials orders for milk production traits (milk yield, fat and protein contents), individual fatty acids (C16:0 and C18:1 *cis-9*) and fatty acid groups (LCFA, SFA, and UFA)

Traits		**Milk (kg)**	**Fat (%)**	**Protein (%)**	**C16:0**	**C18:1 *cis*-9**	**LCFA**	**SFA**	**UFA**
Model[12]	**Np**	**DIC**	**DIC**	**DIC**	**DIC**	**DIC**	**DIC**	**DIC**	**DIC**
LEG222	10	1398310.926	141755.446	-366806.310	-417279.728	-590462.670	-250227.008	-5943.687	-405844.644
LEG333	19	1382694.490	130092.959	-389272.798	-428804.359	-609067.341	-266599.718	-16957.853	-422942.439
LEG334	21	1375257.596	125512.846	-399806.104	-433417.971	-617643.156	-273596.944	-21307.858	-430461.021
LEG335	28	1368585.113	**121087.320**	-408237.441	-437606.350	-622874.982	-278250.132	**-25740.636**	-435485.139
LEG343	21	1376904.209	126517.713	-398236.434	-432330.262	-616492.963	-272559.664	-20258.854	-429402.795
LEG344	27	1375273.138	125495.260	-399848.955	-433414.413	-617587.149	-273567.297	-21332.651	-430433.553
LEG345	32	1368614.750	121167.133	-408224.800	-437644.529	-622896.526	-278284.366	-25684.615	-435359.268
LEG353	28	1372551.349	123995.113	-404119.049	-434492.806	-621087.252	-276361.555	-22576.337	-433357.418
LEG354	32	1370893.502	122946.009	-406063.944	-435832.567	-621897.059	-277198.325	-23736.020	-434082.826
LEG355	37	1368678.314	121182.685	-408214.670	-437492.038	-622853.598	-278245.683	-25650.942	-435333.623
LEG433	21	1381516.595	129587.747	-390367.100	-429388.937	-610554.071	-267707.604	-17437.848	-424166.476
LEG434	27	1375184.143	125517.909	-399899.836	-433750.752	-617836.662	-273706.786	-21318.184	-430574.375
LEG435	32	1368604.821	121177.157	-408265.940	-437533.730	-623087.475	-273096.516	-25345.890	-435483.299
LEG443	27	1376854.529	126590.816	-398235.210	-432635.693	-616676.950	-272502.461	-20183.379	-429447.556
LEG444	31	1375121.434	125485.646	-399909.833	-433786.164	-617770.264	-273684.760	-21360.744	-430554.015
LEG445	36	1368575.703	121138.204	-408293.225	-437930.159	-623092.655	-278420.737	-24906.193	-435509.472
LEG453	32	1372629.236	124156.834	-404178.940	-434786.034	-621147.799	-276344.841	-22484.490	-433225.686
LEG454	36	1370914.069	123004.746	-406138.799	-435887.959	-621956.138	-277189.308	-23695.308	-434120.034
LEG455	41	1368568.723	121175.190	-408243.615	-438161.863	-622991.585	-278323.648	-25690.524	-435450.311
LEG533	28	1380782.331	128891.524	-391963.679	-429877.121	-612068.182	-268981.997	-18008.934	-425478.802
LEG534	32	1374755.662	125073.884	-400804.145	-434372.581	-618565.103	-274340.616	-21714.937	-431241.692
LEG535	37	1368633.494	121193.562	-408258.076	-438358.151	**-623152.330**	**-278470.668**	-25675.362	-435503.726
LEG543	32	1376442.894	126144.110	-399157.159	-433234.854	-617479.460	-273298.454	-20600.168	-430133.028
LEG544	36	1374700.971	125028.525	-400857.150	-434309.872	-618481.015	-274279.187	-21777.678	-431213.185

LEG545	41	1368570.763	121134.824	-408306.286	**-438479.746**	-623104.967	-278448.096	-25712.383	**-435534.541**
LEG553	37	1372398.779	124003.437	-404173.882	-435030.641	-621294.361	-276486.602	-22585.176	-433443.019
LEG554	41	1370847.554	122837.138	-406170.806	-436375.454	-622051.471	-277265.407	-23799.754	-434351.825
LEG555	46	**1368545.969**	121100.309	**-408327.659**	-438215.857	-623086.553	-278424.890	-25731.975	-435524.536

[1] Models are presented according to the code LEG*hap*, where LEG is the basis function used (Legendre polynomial); *h*, *a*, and *p* are the random regression coefficients matrices orders for herd-year of calving, additive genetic and permanent environment effects, respectively.

[2] Best fit model is shown in bold for each studied trait.

The DIC values decreased as the order of the Legendre polynomials increased for all analyzed random effects, but notably for the permanent environment effect. Adjusted fifth-order (degree 4) Legendre polynomials (LEG555) were the most suitable for random effects modeling for milk yield and protein contents (Table 2). On the other hand, third-order (degree 2) random regressions were enough to model additive genetic and herd-year of calving effects for fat content. Although fifth-order (degree 4) was needed to adjust the permanent environment effect (LEG335).

There was, in general, an improvement in DIC values by increasing the Legendre polynomial order to five (degree 4) for the permanent environment effect. For the other random effects, there was a wide variation of the best fit order for each studied fatty acid trait. Although the herd-year of calving effect seemed to require the same adjust order (fifth-order or degree 4), differences were observed for the additive genetic effect. Thereby, the lowest DIC values indicated fourth-order (degree 3) to be the best model (LEG545) to fit C16:0 and UFA test-day records, and third-order (degree 2) (LEG535) for C18:1 cis-9 and LCFA test-day records (Table 2). For SFA, third-order (degree 2) random regression (LEG335) was supported to best goodness-of-fit of additive genetic and herd-year of calving effects.

Eigenvectors and percentage of variance explained by eigenvalues associated with the random regression coefficients matrix for random effects using the highest order model (LEG555) for each trait are presented in Supplementary Tables S1-S6. The variability of the data for the additive genetic effect was strongly explained by the first three eigenvalues, which were responsible for over 99% (except for SFA, 97.8%) of the variance. In addition, the first three eigenvalues explained over 96.4% and 99.2% of the herd-year of calving variance, respectively, for milk production and fatty acid traits, and over 92% and 91.2% of the permanent environment variance.

2.4.2. Residual analysis

The average residual means for each day (range 5-305 DIM) and for all studied traits were equal to zero. The evolution of the mean residuals over DIM for lowest (degree 1), third (degree 2) and best (according to DIC) RRM for milk production and FA traits are shown in Supplementary Figure S1. For all analyzed traits, the test-day variances of residuals from the best fit model (LEG555 for milk yield and protein content; LEG335 for fat content and SFA; LEG545 for C16:0 and UFA; and LEG535 for C18:1 cis-9 and LCFA) tended to be largest at the early lactation (5 to 35 DIM) by using second-order (LEG222) models, and with stable values up to the end. By fitting the highest Legendre polynomial order to model random effects, the mean residuals decreased with their standard deviations along the curve, especially at early lactation. On the other hand, third-order (LEG333) models showed residuals trajectory to be flatter from the beginning.

2.4.3. Variance components and genetic parameters

Estimated trajectories of the additive genetic, permanent environment, and herd-year of calving variances over the lactation for milk production and FA traits from the best RRM (LEG555 for milk yield and protein content; LEG335 for fat content and SFA; LEG545 for C16:0 and UFA; and LEG535 for C18:1 cis-9 and LCFA) are shown in Figure 1.

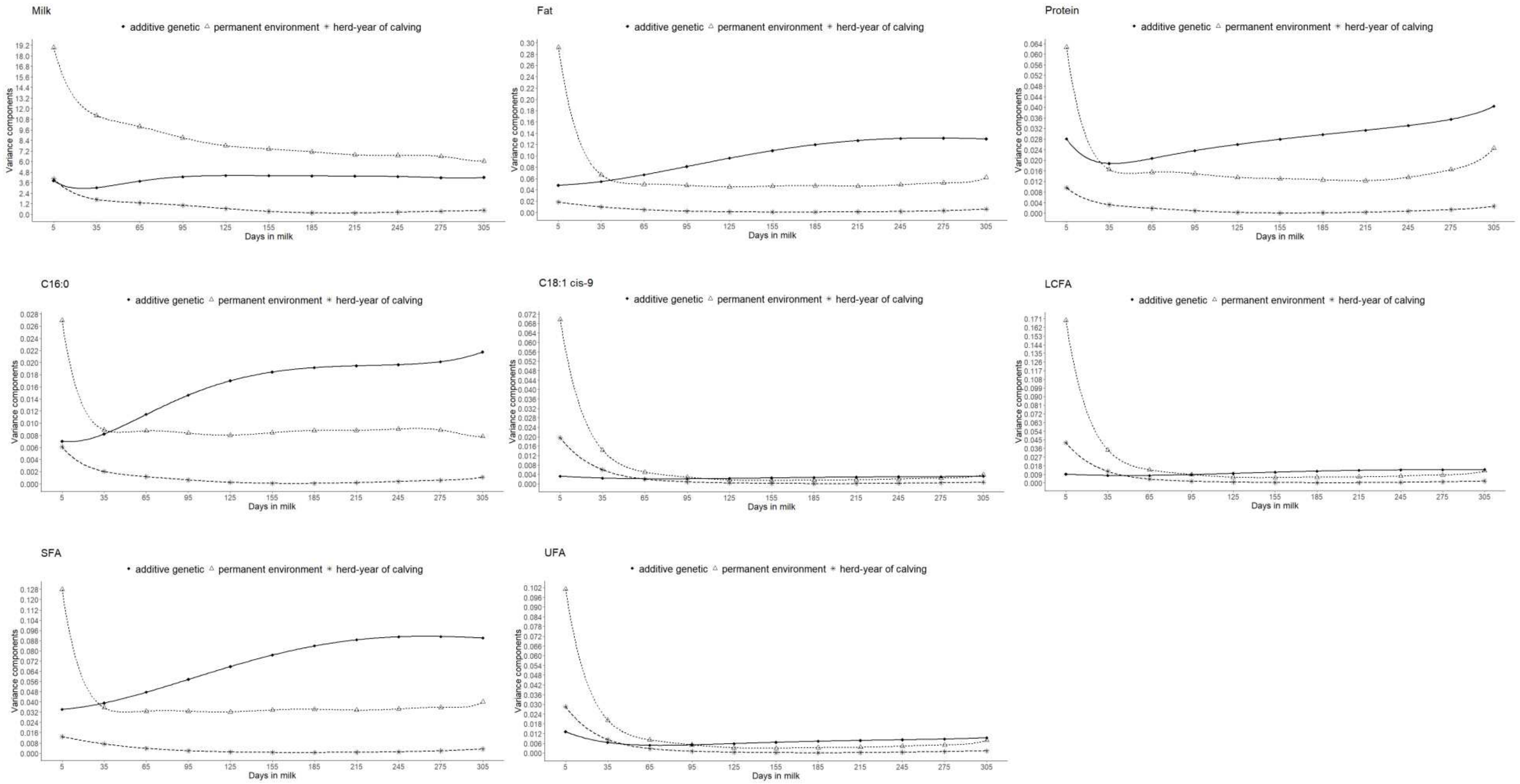


Figure 1. Additive genetic, permanent environment and herd-year of calving variances posterior means over days in milk for milk production (milk yield, fat, and protein contents), individual fatty acids (C16:0 and C18:1 *cis*-9) and fatty acids groups (LCFA, SFA, and UFA).

Overall, variance components posterior means of herd-year of calving and permanent environment decreased quickly starting from the very beginning of lactation. A plateau was also observed from 65 DIM to the end of lactation for all studied traits. The additive genetic variance had higher values in the aforementioned period, except for milk yield (Figure 1), where the permanent environment variance was larger throughout lactation. The general pattern was similar for all traits despite magnitude range. The herd-year of calving variances were higher than additive genetic at early lactation (two months) only for C18:1 cis-9, LCFA, and UFA (Figure 1).

The heritability posterior means for milk production traits (milk yield, fat and protein contents) from the best RRM (LEG555, LEG335 and LEG555, respectively) are shown in Figure 2. The daily heritabilities (Figure 2.a) ranged from 0.12 to 0.29, 0.11 to 0.54, and 0.25 to 0.58, for milk yield, fat and protein contents, respectively. Fat and protein content heritabilities had a similar shape over days in milk, in which presented smaller value in the early and higher in the middle-end lactation. Likewise, milk yield heritability estimates increased from early to the end of lactation, with smaller fluctuations in the middle. The daily permanent environment effect ranged from 0.42 to 0.62, 0.19 to 0.69, and 0.22 to 0.57, respectively, for milk yield, fat and protein contents (Figure 2.b). These values were higher in early lactation, and depicted slightly stable patterns from 125 DIM up to the end of lactation. Fat and protein contents have shown a little rise up in the last 30 days. In addition, a strong decrease in magnitude was observed for fat and protein contents from 5 to 35 DIM. The herd-year of calving effect (Figure 2.c) presented low estimates over the whole lactation (ranged from 0.01 to 0.13). These values decreased up to 215 DIM, when a slight increase was observed.

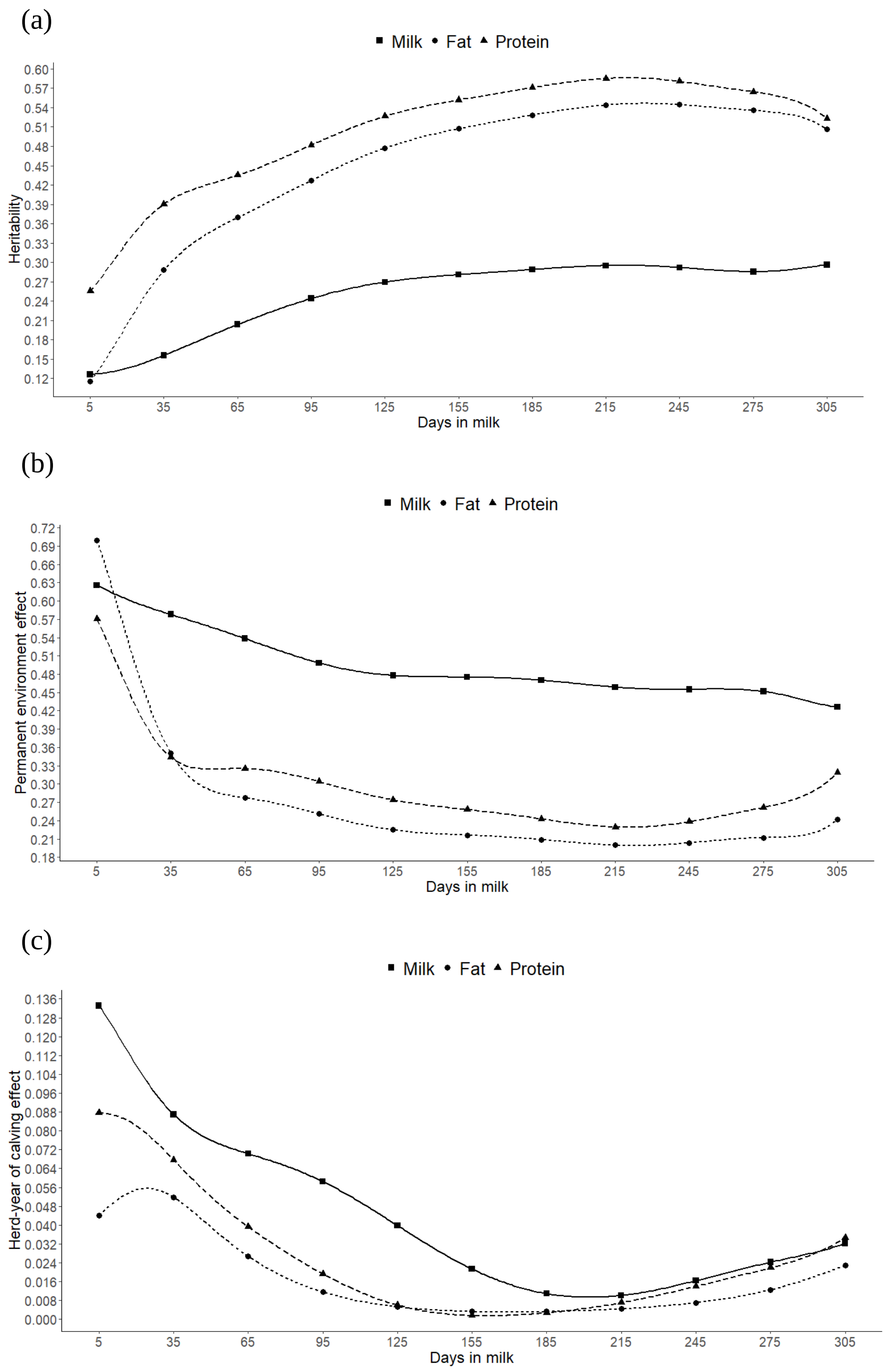


Figure 2. Posterior means of heritability (a), permanent environment (b) and herd-year of calving effects (c) over days in milk for milk production traits (milk yield, fat and protein contents).

The heritability posterior means for individual FA (C16:0 and C18:1 *cis*-9) and FA groups (LCFA, SFA, and UFA) obtained from the best RRM (LEG545, LEG535, LEG535, LEG335, LEG545, respectively) according to DIC criterion are shown in Figure 3. Overall, milk FA followed a similar shape to milk production traits with low estimates at early lactation, and posterior increase throughout almost the entire curve, with a slight decrease at the very end. C16:0 and SFA presented similar and higher estimates over the lactation compared to all studied fatty acids (h^2 range 0.14 – 0.56; Figure 3.a). In general, LCFA and UFA showed similar estimates (h^2 range 0.03 – 0.36), with slight difference in the first two months. The heritability estimates for C18:1 *cis*-9 were the lowest values (h^2 range 0.03 – 0.29), but increased over DIM (Figure 3.a).

The permanent environment effect (Figure 3.b) had high impact on early lactation. This effect showed similar trends along the lactation curve for all studied milk FA. In general, C18:1 *cis*-9, LCFA, and UFA, presented similar estimates. These values were higher at 5 DIM, with abruptly decreased up to 125 DIM, and generally constant up to the end of lactation (daily estimates range 0.15 – 0.71). C16:0 and SFA estimates decreased faster than other milk FA, and have shown the highest and stable pattern from 95 to 275 DIM (daily estimates range 0.19 – 0.60). Although the herd-year of calving effect has shown a slight increase for C18:1 *cis*-9, LCFA, and UFA at the beginning of lactation, it decreased drastically over DIM (range 0.01 – 0.21; Figure 3.c). Finally, for C16:0 and SFA, the herd-year of calving effect had lower estimates throughout lactation period, ranging from 0.03 to 0.12.

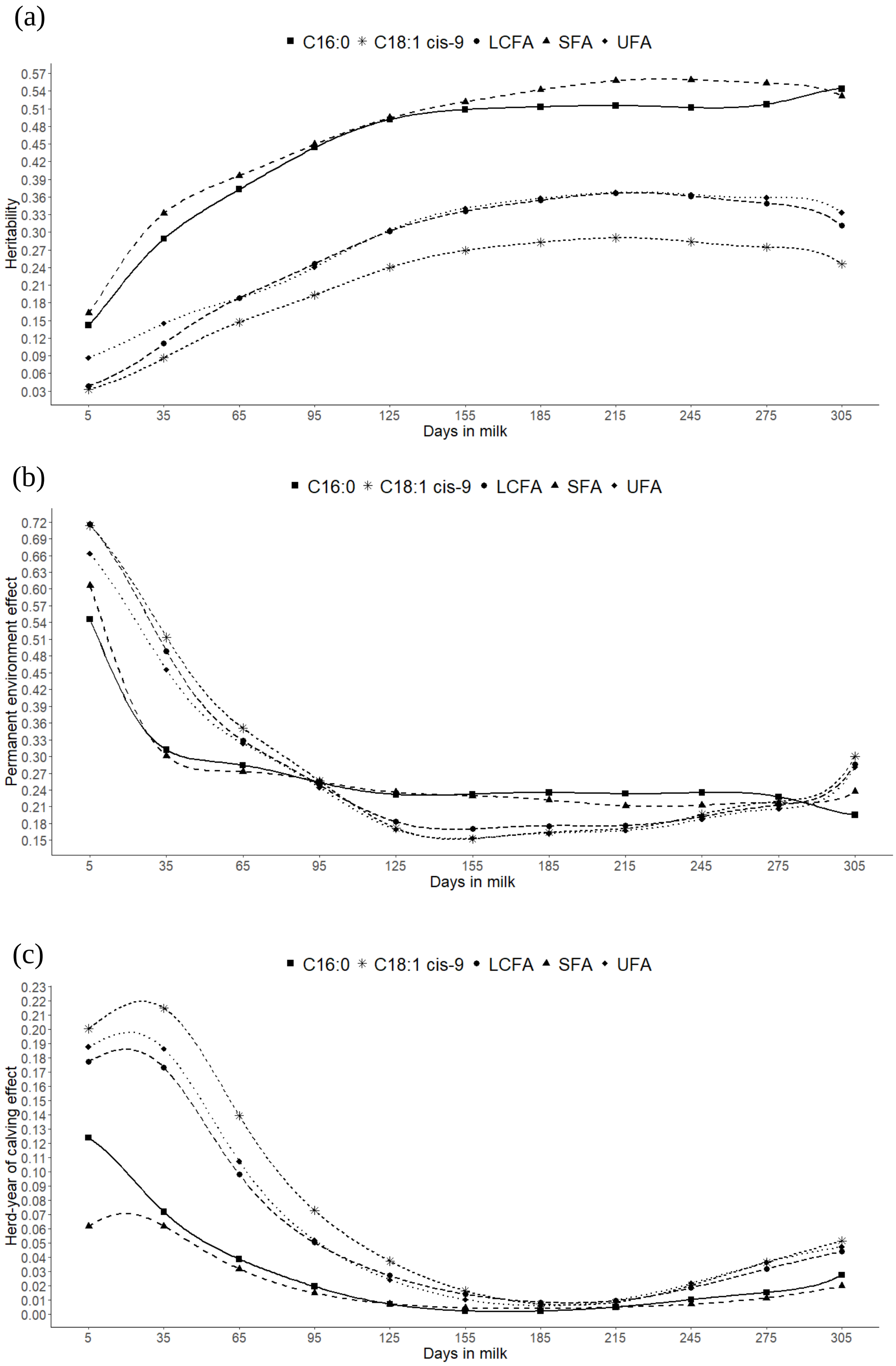


Figure 3. Posterior means of heritability (a), permanent environment (b) and herd-year of calving effects (c) over days in milk individual fatty acids (C16:0 and C18:1 *cis*-9) and fatty acid groups (LCFA, SFA, and UFA).

2.4.4. Rank correlations analysis

Figure 4 presents Spearman correlations among EBV estimated using the best fit models and second-order and third-order for the top 100 and all bulls in the studied population.

In general, Spearman correlations were high among EBVs obtained by the best fit Legendre polynomial model of each trait and third-order (LEG333) model for 155 and 305 DIM (values > 0.90 for all bulls and > 0.97 for top 100 bulls). By using second-order (LEG222) these estimates were, in general, from medium to high magnitude, especially for C16:0 (0.55) for top 100 bulls at 155 DIM. The values estimated for 5 DIM were of low magnitude for the majority of traits. This implies substantial re-ranking for both second (range 0.17 to 0.68) and third-order (range 0.24 to 0.97) models tested, as shown in Figure 4.

The percentages of coincidence between the best fit models and second-order and third-order models in the top 1, 5, 10, and 50% of the bull population for each trait are shown in Figure 5.

For most traits was observed high percentage of coincidence among models for 155 and 305 DIM. The number of common bulls increase as the top percentage increases (except 50%). By using third-order (LEG333) models, the values were high (from 87.96 to 98.46% top 1%, 90.46 to 99.69% top 5% and 94.15 to 99.08 top 10%). For 5 DIM the percentages were low, markedly by using second-order (LEG222) models (range 40.01 to 95.38%).

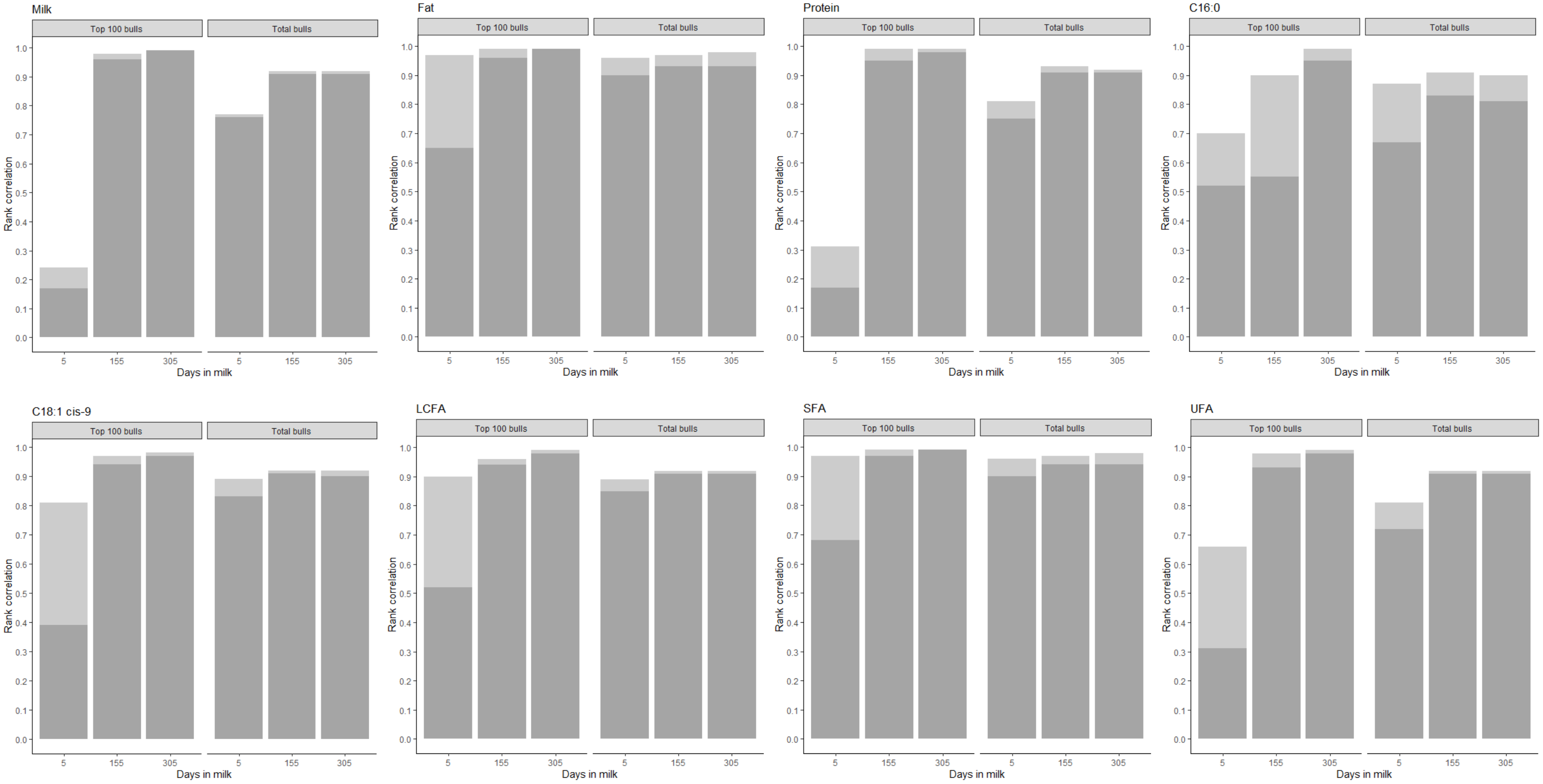


Figure 4. Spearman correlations between the best fit models and second-order (LEG222; dark grey) and third-order (LEG333; light grey) for top 100 and all bulls.

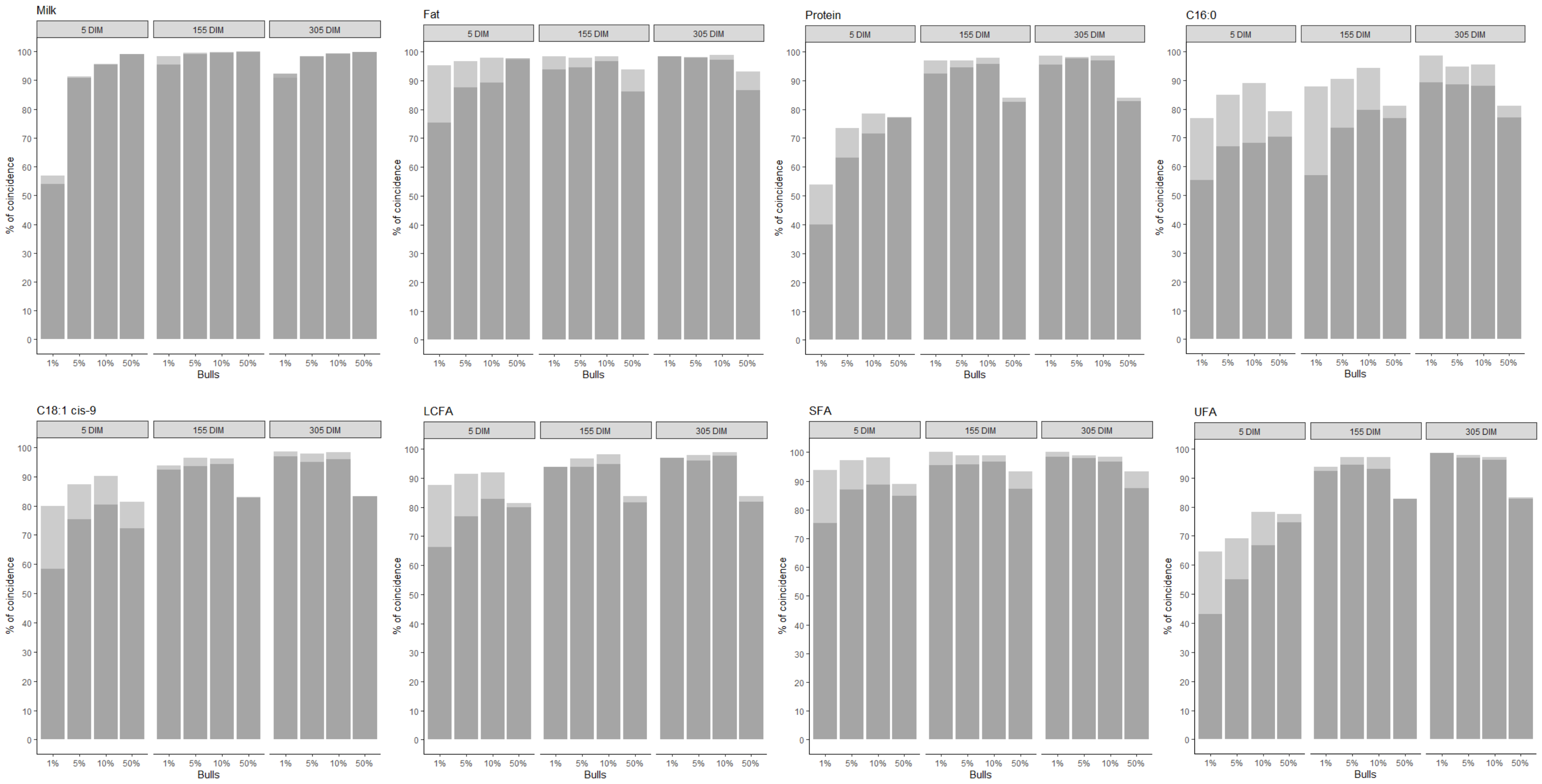


Figure 5. Percentage of bulls selected in common for top 1%, 5%, 10%, and 50% bulls, between the best fit models and second-order (LEG222; dark grey) and third-order (LEG333; light grey) models.

2.5. Discussion

According to Pool et al. (2000) and López-Romero and Carabaño (2003), the genetic (co)variance structure of first parity lactation curves can be modeled using lower Legendre polynomial order for additive genetic compared to permanent environment component. Similarly, our findings showed that differences in DIC were larger among models by using fifth-order (degree 4) Legendre polynomial for the permanent environment effect. This indicates that adjusting the same order or lowest for the additive genetic and herd-year of calving effects resulted in small differences in the goodness-of-fit.

Strabel et al. (2005) also suggested that third-order (degree 2) might be sufficient to capture most genetic variability of daily milk production, while fifth-order (degree 4) was required for age-season of calving and permanent environment curves. Thus, the use of higher order for permanent environment effect could result in a more flexible shape and should be able to model the individual deviations from the average lactation curve more precisely. Similar to our study, for milk yield, Prakash et al. (2017) and Peixoto et al. (2014) reported highest order (degree 4) for permanent environment and third-order (degree 2) for additive genetic effect, respectively, in Sahiwal and Guzera cattle.

Evaluating for fat and protein contents in Brazilian Holstein cattle, Padilha et al. (2016) found a lower order (fourth-order) for the additive genetic and permanent environment effects. For the milk FA traits, Narayana et al. (2017) reported DIC values decreasing with the increase in the degree of Legendre polynomials. According to these authors the fifth-order (degree 4) random regressions present the best fit to model LCFA, SFA and UFA. Similarly, our findings pointed to the same fit order to describe herd-year of calving and permanent environment (co)variance structures for C16:0, C18:1 *cis*-9, LCFA and UFA. Although our results suggested different Legendre polynomials orders for different random effects, this is not a consensus in literature, especially for milk FA (Bastin et al., 2011; Freitas et al., 2020).

Analysis of residuals showed that higher-order models selected by DIC criterion presented changes over DIM. Although the mean residuals trend over DIM were close to zero, it was observed a slight fluctuation in the first month of lactation for all studied traits. Similar trends of residuals for milk yield were observed by Santos et al. (2014) by using a Legendre polynomial model in Guzerá dairy cattle. The increase of the number of parameters for the random effects, especially for permanent environment, decreased the variance of residuals in this period. However, there was a slight overestimation of milk production traits in early lactation, as well as for all the evaluated FA traits. Unlikely, lower-order models presented underestimation and resulted in higher mean residuals and large standard deviations. According to Druet et al. (2003), Legendre polynomial curves present typical border effects at early and end lactation, with waves in the middle. Residuals obtained from models with third-order (degree 2) followed a flatter pattern over the lactation. Thus, the use of the homogenous residual variance across lactation appeared to be justified; these results do support the fit of the proposed model for these traits. In addition, according to first three eigenvalues associated with the highest order model (LEG555) the use of third-order models is sufficient to capture most of the additive genetic herd-year of calving and permanent environment variation of the milk production and fatty acid traits. Moreover, the dimension of the random regression coefficients matrix may be reduced without a substantial loss of information and avoiding some undesirable variations due to over-parameterization, which ensures that the model is relatively robust.

Variance components and genetic parameters for herd-year of calving effect are not constant and can vary with the lactation stage. Neglecting the herd-year of calving effect may cause a large increase in the residual variance or this effect can be confounded with the genetic component, as reported by Bormann et al. (2003). Likewise our findings, Druet et al. (2003) found lower variance and a similar curve for herd-year of calving effect for milk yield in first lactation. Similar curve patterns were also observed across lactation for milk yield, fat, and protein yields by Hammami et al. (2008) and for LCFA, SFA, and UFA by Narayana et al. (2017). According to Hammami et al. (2008), the effect of herd-year of calving is associated with specific environmental influences (management and feeding system) in different herds.

The lactation stage along with the energy balance of dairy cows has also great influence on the milk FA profile. The increased energy requirement is evident in the postpartum period and dairy cows often experience a period of NEB at the beginning of lactation. This results in higher mobilization of adipose FA and fat percentage in milk (Gross et al., 2011). These high variations for herd-year of calving effect can be associated with differences among cows in body fat mobilization, especially in early lactation. The high uptake of long-chain fatty acids by mammary gland inhibits de novo synthesis of fatty acids (e.g C16:0 and SFA), and therefore, some increases in the proportion of preformed milk FA, such as C18:1 *cis*-9, LCFA, and UFA (Hanuš et al., 2018) are observed. From middle towards late lactation, the FA profile became more stable, may be due to improvements in energy balance or higher feed intake. Garnsworthy et al. (2006) reported cows in middle lactation producing the same FA profile as in late lactation. Overall, the herd-year of calving random effect allows analyzing changes in environmental conditions that influenced the shape of the lactation. According to Strabel et al. (2005) the solution for this effect may help to make management decisions, to further indicate in which part of the lactation high levels of production can be achieved.

In general, estimated permanent environment variances were higher at the beginning, have decreased up to the middle and kept stable to the end of lactation. Similar trends were observed for milk, fat and protein yield in several studies (Cobuci et al., 2011; Costa et al., 2008; Druet, Jaffrézic, & Ducrocq, 2005; Kheirabadi, 2019) for permanent environment effect. This indicates that environmental factors have a high influence on test-day records at the beginning of lactation in first lactation Holstein cows. Likewise, Narayana et al. (2017) and Soyeurt et al. (2008) reported similar pattern for the permanent environment variance curves for LCFA, SFA and UFA. According to Ødegard et al. (2003), the permanent environment effect absorbed most of the heterogeneity of residual variance, particularly on high-order Legendre polynomial models.

Overall, additive genetic variances estimates were consistently larger than permanent environment variances from the middle to the end of lactation, except for milk yield. Similar results for milk yield were reported for first lactation Holstein cows in Tunisia (Hammami et al., 2008), Brazil (Cobuci et al., 2011) and United Kingdom (Olori et al., 1999). However, opposite trends were reported by Druet et al. (2003), in which they obtained lower additive genetic variance estimates at the beginning and the end of the lactation. Strabel et al. (2005) and Narayana et al. (2017) showed extreme values at the peripheries of lactation, similar to found for LCFA, SFA and UFA, by using higher Legendre polynomial order (degree 4). Thus, additive genetic variance can follow oscillatory patterns, as seen in results from different studies, depending upon the RRM fit and data size. Particularly, the use of Legendre polynomials has been described to result in overestimation of the additive genetic at the extremes of lactation not only in dairy cattle (Pool et al., 2000; López-Romero and Carabaño, 2003; Strabel, 2004) but also in beef cattle (Mota et al., 2013), and small ruminants (Oliveira et al., 2016).

Regards on genetic parameters, there was an increase in heritability from the beginning to end of the lactation period, with a slight decrease in the last 30 days for most traits (fat and protein contents, C18:1 *cis*-9, LCFA, SFA, and UFA). The increase in heritability estimates is associated not only with increases on the additive genetic variances but also with the reductions of permanent environment and herd-year of calving components. Similarly, Bormann et al. (2003) found herd-stage effect variance higher in early lactation, which contributed to lower estimates of heritability in this period.

In addition, heritabilities and their tendency towards increasing over days noticed for milk production traits are in agreement with Druet et al. (2005). Likewise, similar curves were found by Soyeurt et al. (2008) and Bastin et al. (2011). Heritabilities similar curve patterns and magnitudes for milk yield were also reported by Bormann et al. (2003) and Kheirabadi (2019) in first lactation dairy cattle, with values ranging from 0.14 to 0.27 and 0.09 to 0.25, respectively. Fat and protein contents were more heritable (h^2 maximum value of 0.54 and 0.58, respectively) around 215 DIM with similar trajectories to the findings of Padilha et al. (2016). Nevertheless, these authors reported an increase from the beginning up to 215-245 DIM, and a decreased thereafter to the end of lactation. According to Druet et al. (2005) this behavior is expected, once heritability for yields increases while for contents decreases in the final period. Currently, milk fat and protein still have been one of most important traits in the dairy industry because they determine the value of raw milk in pricing schemes all over the world (Gengler et al., 2016). Thus, the high heritabilities indicates that genetic gains will be effective when these traits were used as selection criteria in breeding programs of Walloon Holstein cows.

As seen for milk production traits, daily heritabilities for FA were lower in early lactation. Similar to Soyeurt et al. (2008) and Bastin et al. (2011) results, all studied FA were more heritable from the middle (around 155 DIM) to late lactation stage. In general, our estimates were close to the results reported by Hammami et al. (2015), however higher than those reported by Petrini et al. (2016). The later authors reported mean estimates ranging from 0.07 to 0.26. Narayana et al. (2017) showed differences in heritability curves for UFA (h^2 mean 0.21) and LCFA (h^2 mean 0.23), and these estimates decreased at early lactation up to 55 DIM, where an increase started to be observed. The UFA and LCFA estimates reported by these authors were very close to ours all over the lactation. Moreover, all FA were higher from middle to end lactation stage, and SFA (h^2 mean 0.33) presented the higher estimates over DIM. In agreement with our results, Freitas et al. (2020) reported similar heritabilities curves for LCFA, UFA and SFA, as well their estimates were higher for SFA over all DIM. The observed differences in estimates in comparison to previous studies might be explained by the different analytical methods and calibration equations to predict the milk FA, modeling of lactation curves, sample sizes, and dairy cattle populations.

Our results reinforced that de novo synthesized FA (SFA, especially C16:0) are under more genetic control than preformed fatty acids. This can be explained by the different origins of milk FA. While half of C16:0 is synthesized de novo in the mammary gland by the enzymes acetyl-coenzyme A carboxylase and FA synthetase, the LCFA, and C18:1 *cis*-9 are originated mainly from the dietary lipids and body fat mobilization (Hanuš et al., 2018). This indicates that early lactation lipolysis is genetically controlled. Furthermore, the effects of genes that influence milk FA can change along lactation stages, and hence, larger improving of these traits could be achieved by genetic selection in specific times over lactation.

Lower Spearman correlations between EBV estimated using different models for early lactation suggest that larger impacts can occur at the beginning of lactation, indicating greater re-ranking of the animals under selection. Re-ranking was observed, especially for top 1–10% common bulls, for all traits evaluated at 5 DIM. As the selection intensity increases, the animals selected in common became more dissimilar. It indicates that how higher the selection intensity applied to the herd at the beginning of lactation greater is the need using an appropriate model. Therefore, the selection of high EBV for milk production and FA traits at the beginning of lactation by using Legendre polynomials lower order could compromise future genetic gains because it would not represent the best bulls in this population. In addition, it seems to be worse if second-order models are used, which would decrease more the percentage of bull selected in common for top 1–10%. Thus, re-ranking of common bulls was more important at 5 DIM lactation stage compared to total yield.

On the other hand, the high Spearman correlations for 105 or 305 DIM indicate that the ranking of bulls by using third-order Legendre polynomial showed the same ability to rank bulls in comparison to the best fit model. Third-order Legendre polynomials have been used in Walloon official genetic evaluation for milk production traits. Likewise, it was observed a high coincidence percentage in the classification of EBVs for them. Therefore, this high similarity indicates that there is no re-ranking among bulls when using different models for 105 or 305 DIM. Pereira et al. (2013) also reported high Spearman correlation (0.99) between accumulated 305 DIM bull EBV from a lower (degree 4) and higher (degree 5) order model in Gyr dairy cattle. Usually, the selection is made at 305 days in milk, thus the genetic progress in the population would have similar trends even using different polynomial Legendre models. This aspect should be taken into account during selection of bulls using EBV estimated from RRM since the genetic merit of these animals is predicted based on daily milk yield records.

Legendre polynomials RRM have become a standard model used in most national genetic evaluations and the adjustment order implemented varies among countries. In general, as seen previously, higher orders models usually present best fit according to DIC criterion. Besides this aspect, it should also be considered flexibility and robustness of random regression models to making decisions regarding their application in genetic evaluation (Druet et al., 2003; Cobuci et al., 2011). Therefore, third-order RRM (LEG333) showed lower residual mean over DIM, and the highest Spearman correlation between EBV and lower re-ranking among bulls in comparison to best fit models, which imply that milk production and FA traits could be sufficiently modeled by using only three coefficients of Legendre polynomials. In addition, further implementation of the multiple lactation model and especially multi-trait would result in a substantial reduction in the number of parameters to be estimated, which will be less demanding and computationally feasible. Hence, as have been used routinely in Walloon official genetic evaluation for milk yield, third-order seems most appropriate to model milk production and FA traits with homogenous residual variance.

Although the economic value of milk FA has not been yet used as a bonus for milk producers in most countries, its importance to the dairy industry has been increasing over the years due to nutritional, physical, and organoleptic properties (Hanuš et al., 2018). Several studies have reported the role of milk FA as a valuable early biomarker on the health status of cows (Dórea et al., 2017; Jorjong et al., 2015; Stoop et al., 2008; Van Haelst et al., 2008). According to Bastin et al. (2013) and Hammami et al. (2015), the content of C18:1 *cis*-9 can be used as a biomarker of energy balance status in early lactation and is associated with fertility traits and high sensitivity to heat stress. In this context, the application of the most appropriate random regression test-day models in genetic evaluation for those FA results in higher EBV accuracy. Moreover, the direction of selection for milk production traits and FA in breeding programs depends upon the purpose of the milk final production as well as changes in management of the herds.

2.6. Conclusion

Third-order Legendre orthogonal polynomials seem to be most parsimonious and feasible to describe milk production and fatty acid traits over DIM in first lactation Walloon Holstein cows. The first three eigenvectors of the genetic, permanent environment and herd-year of calving matrices explained most of the variance for all traits. The use of an optimal model appeared to be even more important in case of selection in early lactation. In addition, optimal polynomial orders tend to show lower residual variation compared to other models, which can help to avoid overestimation at the beginning of lactation. Heritability curves indicated that as long as lactation progressed the proportion of genetic variance increased for all traits.

2.7. Acknowledgments

The authors acknowledge the support of the National Fund for Scientific Research (Brussels, Belgium) under the grants (Projects: GplusE and Indiku). The authors also gratefully acknowledge the financial and technical support provided by the Walloon Breeding Association (AWE, Ciney, Belgium). Computational resources have been provided by the Plateau de Calcul et Modélisation informatique (CAMI) of the University of Liège, Gembloux Agro-Bio Tech and the Consortium des Equipements de Calcul Intensif (CECI) of the Federation Wallonia-Brussels (Brussels, Belgium), funded by the National Fund for Scientific Research (Brussels, Belgium) funded under grant 2.5020.11. The authors also acknowledge the CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior).

2.8. References

Arnould, V.M.-R., H. Hammami, H. Soyeurt, and N. Gengler. 2010. Short communication: Genetic variation of saturated fatty acids in Holsteins in the Walloon region of Belgium. J. Dairy Sci. 93:4391–4397. doi:10.3168/jds.2010-3049.

Auvray, B., and N. Gengler. 2002. Feasibility of a Wallon test-day model and study of its potential as tool for selection and management. Interbull Bull. 29:123–127.

Bastin, C., D.P. Berry, H. Soyeurt, and N. Gengler. 2012. Genetic correlations of days open with production traits and contents in milk of major fatty acids predicted by mid-infrared spectrometry. J. Dairy Sci. 95:6113–6121. doi:10.3168/jds.2012-5361.

Bastin, C., N. Gengler, and H. Soyeurt. 2011. Phenotypic and genetic variability of production traits and milk fatty acid contents across days in milk for Walloon Holstein first-parity cows. J. Dairy Sci. 94:4152–4163. doi:10.3168/jds.2010-4108.

Bastin, C., H. Soyeurt, and N. Gengler. 2013. Genetic parameters of milk production traits and fatty acid contents in milk for Holstein cows in parity 1 - 3. J. Anim. Breed. Genet. 130:118–127. doi:10.1111/jbg.12010.

Bastin, C., L. Théron, A. Lainé, and N. Gengler. 2016. On the role of mid-infrared predicted phenotypes in fertility and health dairy breeding programs. J. Dairy Sci. 99:4080–4094. doi:10.3168/jds.2015-10087.

Bormann, J., G.R. Wiggans, T. Druet, and N. Gengler. 2003. Within-Herd Effects of Age at Test Day and Lactation Stage on Test-Day Yields. J. Dairy Sci. 86:3765–3774. doi:10.3168/jds.S0022-0302(03)73983-6.

Britt, J.H., R.A. Cushman, C.D. Dechow, H. Dobson, P. Humblot, M.F. Hutjens, G.A. Jones, P.S. Ruegg, I.M. Sheldon, and J.S. Stevenson. 2018. Invited review: Learning from the future—A vision for dairy farms and cows in 2067. J. Dairy Sci. 101:3722–3741. doi:10.3168/jds.2017-14025.

Cobuci, J.A., C.N. Costa, J. Braccini Neto, and A.F. de Freitas. 2011. Genetic parameters for milk production by using random regression models with different alternatives of fixed regression modeling. Rev. Bras. Zootec. 40:557–567. doi:10.1590/S1516-35982011000300013.

Costa, C.N., C.M.R. de Melo, I.U. Packer, A.F. de Freitas, N.M. Teixeira, and J.A. Cobuci. 2008. Genetic parameters for test day milk yield of first lactation Holstein cows estimated by random regression using Legendre polynomials. Rev. Bras. Zootec. 37:602–608. doi:10.1590/S1516-35982008000400003.

Croquet, C., P. Mayeres, A. Gillon, S. Vanderick, and N. Gengler. 2006. Inbreeding Depression for Global and Partial Economic Indexes, Production, Type, and Functional Traits. J. Dairy Sci. 89:2257–2267. doi:10.3168/jds.S0022-0302(06)72297-4.

Dórea, J.R.R., E.A. French, and L.E. Armentano. 2017. Use of milk fatty acids to estimate plasma nonesterified fatty acid concentrations as an indicator of animal energy balance. J. Dairy Sci. 100:6164–6176. doi:10.3168/jds.2016-12466.

Druet, T., F. Jaffrézic, D. Boichard, and V. Ducrocq. 2003. Modeling Lactation Curves and Estimation of Genetic Parameters for First Lactation Test-Day Records of French Holstein Cows. J. Dairy Sci. 86:2480–2490. doi:10.3168/jds.S0022-0302(03)73842-9.

Druet, T., F. Jaffrézic, and V. Ducrocq. 2005. Estimation of genetic parameters for test day records of dairy traits in the first three lactations. Genet. Sel. Evol. 37:257–271. doi:10.1051/gse:2005001.

Freitas, P.H.F., H.R. Oliveira, F.F. Silva, A. Fleming, F. Miglior, F.S. Schenkel, and L.F. Brito. 2020. Genomic analyses for predicted milk fatty acid composition throughout lactation in North American Holstein cattle. J. Dairy Sci. 103:6318–6331. doi:10.3168/jds.2019-17628.

Garnsworthy, P.C., L.L. Masson, A.L. Lock, and T.T. Mottram. 2006. Variation of Milk Citrate with Stage of Lactation and De Novo Fatty Acid Synthesis in Dairy Cows. J. Dairy Sci. 89:1604–1612. doi:10.3168/jds.S0022-0302(06)72227-5.

Gengler, N., H. Soyeurt, F. Dehareng, C. Bastin, F. Colinet, H. Hammami, M.-L. Vanrobays, A. Lainé, S. Vanderick, C. Grelet, A. Vanlierde, E. Froidmont, and P. Dardenne. 2016. Capitalizing on fine milk composition for breeding and management of dairy cows. J. Dairy Sci. 99:4071–4079. doi:10.3168/jds.2015-10140.

Geweke, J. 1991. Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. In Bayesian Statistics 4 (J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, ed.) Clarendon Press, Oxford, UK.

Grelet, C., J.A.F. Pierna, P. Dardenne, H. Soyeurt, A. Vanlierde, F. Colinet, C. Bastin, N. Gengler, V. Baeten, and F. Dehareng. 2017. Standardization of milk mid-infrared spectrometers for the transfer and use of multiple models. J. Dairy Sci. 100:7910–7921. doi:10.3168/jds.2017-12720.

Gross, J., H.A. van Dorland, R.M. Bruckmaier, and F.J. Schwarz. 2011. Milk fatty acid profile related to energy balance in dairy cows. J. Dairy Res. 78:479–488. doi:10.1017/S0022029911000550.

Hammami, H., B. Rekik, H. Soyeurt, A. Ben Gara, and N. Gengler. 2008. Genetic Parameters for Tunisian Holsteins Using a Test-Day Random Regression Model. J. Dairy Sci. 91:2118–2126. doi:10.3168/jds.2007-0382.

Hammami, H., J. Vandenplas, M.-L. Vanrobays, B. Rekik, C. Bastin, and N. Gengler. 2015. Genetic analysis of heat stress effects on yield traits, udder health, and fatty acids of Walloon Holstein cows. J. Dairy Sci. 98:4956–4968. doi:10.3168/jds.2014-9148.

Hanuš, O., E. Samková, L. Křížová, L. Hasoňová, and R. Kala. 2018. Role of Fatty Acids in Milk Fat and the Influence of Selected Factors on Their Variability—A Review. Molecules 23:1636. doi:10.3390/molecules23071636.

Jamrozik, J., G.J. Kistemaker, J.C.M. Dekkers, and L.R. Schaeffer. 1997. Comparison of Possible Covariates for Use in a Random Regression Model for Analyses of Test Day Yields. J. Dairy Sci. 80:2550–2556. doi:10.3168/jds.S0022-0302(97)76210-6.

Jorjong, S., A.T.M. van Knegsel, J. Verwaeren, R.M. Bruckmaier, B. De Baets, B. Kemp, and V. Fievez. 2015. Milk fatty acids as possible biomarkers to diagnose hyperketonemia in early lactation. J. Dairy Sci. 98:5211–5221. doi:10.3168/jds.2014-8728.

Jorjong, S., A.T.M. van Knegsel, J. Verwaeren, M.V. Lahoz, R.M. Bruckmaier, B. De Baets, B. Kemp, and V. Fievez. 2014. Milk fatty acids as possible biomarkers to early diagnose elevated concentrations of blood plasma nonesterified fatty acids in dairy cows. J. Dairy Sci. 97:7054–7064. doi:10.3168/jds.2014-8039.

Kheirabadi, K. 2019. Estimates of genetic trends for daily milk yield and somatic cell score of primiparous Holstein cattle in Iran. J. Appl. Anim. Res. 47:467–473. doi:10.1080/09712119.2019.1663741.

Kirkpatrick, M., D. Lofsvold, and M. Bulmer. 1990. Analysis of the inheritance, selection and evolution of growth trajectories. Genetics 124:979–993.

Loften, J.R., J.G. Linn, J.K. Drackley, T.C. Jenkins, C.G. Soderholm, and A.F. Kertz. 2014. Invited review: Palmitic and stearic acid metabolism in lactating dairy cows. J. Dairy Sci. 97:4661–4674. doi:10.3168/jds.2014-7919.

López-Romero, P., and M.J. Carabaño. 2003. Comparing alternative random regression models to analyse first lactation daily milk yield data in Holstein–Friesian cattle. Livest. Prod. Sci. 82:81–96. doi:10.1016/S0301-6226(03)00003-4.

De Marchi, M., V. Toffanin, M. Cassandro, and M. Penasa. 2014. Invited review: Mid-infrared spectroscopy as phenotyping tool for milk traits. J. Dairy Sci. 97:1171–1186. doi:10.3168/jds.2013-6799.

Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, and D.H. Lee. 2002. Blupf90 and Related Programs (Bgf90). Pages 21-22 in Proc. 7th World Congr. Genet. Appl. Livest. Prod. Mantpeliier, France. Editions Quae, Montpellier, France.

Mota, R.R., L.F.A. Marques, P.S. Lopes, L.P. da Silva, F.R.A. Neto, M.D.V. de Resende, and R.A. Torres. 2013. Genetic evaluation using multi-trait and random regression models in Simmental beef cattle. Genet. Mol. Res. 12:2465–2480. doi:10.4238/2013.July.24.2.

Narayana, S.G., F.S. Schenkel, A. Fleming, A. Koeck, F. Malchiodi, J. Jamrozik, J. Johnston,

M. Sargolzaei, and F. Miglior. 2017. Genetic analysis of groups of mid-infrared predicted fatty acids in milk. J. Dairy Sci. 100:4731–4744. doi:10.3168/jds.2016-12244.

Ødegard, J., J. Jensen, G. Klemetsdal, P. Madsen, and B. Heringstad. 2003. Genetic Analysis of Somatic Cell Score in Norwegian Cattle Using Random Regression Test-Day Models. J. Dairy Sci. 86:4103–4114. doi:10.3168/jds.S0022-0302(03)74024-7.

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

Oliveira, H.R., F.F. Silva, O.H.G.B.D. Siqueira, N.O. Souza, V.S. Junqueira, M.D. V. Resende, R.R.A. Borquis, and M.T. Rodrigues. 2016. Combining different functions to describe milk, fat, and protein yield in goats using Bayesian multiple-trait random regression models1. J. Anim. Sci. 94:1865–1874. doi:10.2527/jas.2015-0150.

Olori, V.E., W.G. Hill, B.J. McGuirk, and S. Brotherstone. 1999. Estimating variance components for test day milk records by restricted maximum likelihood with a random regression animal model. Livest. Prod. Sci. 61:53–63. doi:10.1016/S0301-6226(99)00052-4.

Padilha, A.H., J.A. Cobuci, D. dos S. Daltro, and J. Braccini Neto. 2016. Reliability of breeding values between random regression and 305-day lactation models. Pesqui. Agropecuária Bras. 51:1848–1856. doi:10.1590/s0100-204x2016001100007.

Peixoto, M.G.C.D., D.J. de A. Santos, R.R.A. Borquis, F.Â.T. Bruneli, J.C. do C. Panetto, and H. Tonhati. 2014. Random regression models to estimate genetic parameters for milk production of Guzerat cows using orthogonal Legendre polynomials. Pesqui. Agropecuária Bras. 49:372–383. doi:10.1590/S0100-204X2014000500007.

Pereira, R.J., A.B. Bignardi, L. El Faro, R.S. Verneque, A.E. Vercesi Filho, and L.G. Albuquerque. 2013. Random regression models using Legendre polynomials or linear splines for test-day milk yield of dairy Gyr (Bos indicus) cattle. J. Dairy Sci. 96:565–574. doi:10.3168/jds.2011-5051.

Petrini, J., L.H.S. Iung, M.A.P. Rodriguez, M. Salvian, F. Pértille, G.A. Rovadoscki, L.D. Cassoli, L.L. Coutinho, P.F. Machado, G.R. Wiggans, and G.B. Mourão. 2016. Genetic parameters for milk fatty acids, milk yield and quality traits of a Holstein cattle population reared under tropical conditions. J. Anim. Breed. Genet. 133:384–395. doi:10.1111/jbg.12205.

Pool, M.H., L.L.G. Janss, and T.H.E. Meuwissen. 2000. Genetic Parameters of Legendre Polynomials for First Parity Lactation Curves. J. Dairy Sci. 83:2640–2649. doi:10.3168/jds.S0022-0302(00)75157-5.

Prakash, V., A.K. Gupta, A. Gupta, R.S. Gandhi, A. Singh, and A.K. Chakravarty. 2017. Random regression model with heterogeneous residual variance reduces polynomial order fitted for permanent environmental effect but does not affect genetic parameters for milk production in Sahiwal cattle. Anim. Prod. Sci. 57:1022. doi:10.1071/AN15347.

R Development Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.r-project.org.

Santos, D.J.A., M.G.C.D. Peixoto, R.R. Aspilcueta Borquis, J.C.C. Panetto, L. El Faro, and H. Tonhati. 2014. Predicting breeding values for milk yield of Guzerá (Bos indicus) cows using random regression models. Livest. Sci. 167:41–50. doi:10.1016/j.livsci.2014.05.023.

Schaeffer, L.R. 2004. Application of random regression models in animal breeding. Livest. Prod. Sci. 86:35–45. doi:10.1016/S0301-6226(03)00151-9.

Smith, B.J. 2007. boa: An R Package for MCMC Output Convergence Assessment and Posterior Inference. J. Stat. Softw. 21:1–37.

Soyeurt, H., P. Dardenne, F. Dehareng, C. Bastin, and N. Gengler. 2008. Genetic Parameters of Saturated and Monounsaturated Fatty Acid Content and the Ratio of Saturated to Unsaturated Fatty Acids in Bovine Milk. J. Dairy Sci. 91:3611–3626. doi:10.3168/jds.2007-0971.

Soyeurt, H., P. Dardenne, F. Dehareng, G. Lognay, D. Veselko, M. Marlier, C. Bertozzi, P. Mayeres, and N. Gengler. 2006. Estimating Fatty Acid Content in Cow Milk Using Mid-Infrared Spectrometry. J. Dairy Sci. 89:3690–3695. doi:10.3168/jds.S0022-0302(06)72409-2.

Soyeurt, H., F. Dehareng, N. Gengler, S. McParland, E. Wall, D.P. Berry, M. Coffey, and P. Dardenne. 2011. Mid-infrared prediction of bovine milk fatty acids across multiple breeds, production systems, and countries. J. Dairy Sci. 94:1657–1667. doi:10.3168/jds.2010-3408.

Spiegelhalter, D.J., N.G. Best, B.P. Carlin, and A. van der Linde. 2002. Bayesian measures of model complexity and fit. J. R. Stat. Soc. Ser. B (Statistical Methodol. 64:583–639. doi:10.1111/1467-9868.00353.

Stoop, W.M., J.A.M. van Arendonk, J.M.L. Heck, H.J.F. van Valenberg, and H. Bovenhuis. 2008. Genetic Parameters for Major Milk Fatty Acids and Milk Production Traits of Dutch Holstein-Friesians. J. Dairy Sci. 91:385–394. doi:10.3168/jds.2007-0181.

Strabel, T., J. Szyda, E. Ptak, and J. Jamrozik. 2005. Comparison of Random Regression Test-Day Models for Polish Black and White Cattle. J. Dairy Sci. 88:3688–3699. doi:10.3168/jds.S0022-0302(05)73055-1.

Tzompa-Sosa, D.A., G.A. van Aken, A.C.M. van Hooijdonk, and H.J.F. van Valenberg. 2014. Influence of C16:0 and long-chain saturated fatty acids on normal variation of bovine milk fat triacylglycerol structure. J. Dairy Sci. 97:4542–4551. doi:10.3168/jds.2014-7937.

Van Haelst, Y.N.T., A. Beeckman, A.T.M. Van Knegsel, and V. Fievez. 2008. Short Communication: Elevated Concentrations of Oleic Acid and Long-Chain Fatty Acids in Milk Fat of Multiparous Subclinical Ketotic Cows. J. Dairy Sci. 91:4683–4686. doi:10.3168/jds.2008-1375.

Xu, W., J. Vervoort, E. Saccenti, R. van Hoeij, B. Kemp, and A. van Knegsel. 2018. Milk Metabolomics Data Reveal the Energy Balance of Individual Dairy Cows in Early Lactation. Sci. Rep. 8:15828. doi:10.1038/s41598-018-34190-4.

2.9. Supplementary material

Supplementary Table S1. Eigenvectors and percentage of variance explained by eigenvalues of the random regression coefficients matrix for additive genetic effect using highest order model (LEG555) for milk production traits

Eigenvectors (e_i)	**Eigenvalues (λ_i)**				
	λ_1 (%)	**λ_2 (%)**	**λ_3 (%)**	**λ_4 (%)**	**λ_5 (%)**
Milk yield					
	7.4413 (91.5)	**0.4489 (5.5)**	**0.2113 (2.6)**	**0.0241 (0.3)**	**0.0107 (0.1)**
e_1	0.9913	0.1031	0.0784	-0.0222	0.0017
e_2	0.0742	-0.9435	0.3089	0.0153	-0.0935
e_3	-0.1007	0.2611	0.7894	-0.5119	-0.1910
e_4	0.0338	-0.1763	-0.4408	-0.8543	0.2091
e_5	-0.0220	-0.0018	0.2846	0.0862	0.9545
Fat content					
	0.1914 (93.6)	**0.0095 (4.7)**	**0.0029 (1.4)**	**0.0003 (0.2)**	**0.0001 (0.1)**
e_1	0.9787	0.2016	0.0322	0.0151	-0.0151
e_2	0.1858	-0.8204	-0.5389	0.0317	-0.0325
e_3	-0.0850	0.4718	-0.7052	0.4499	-0.2655
e_4	0.0200	-0.2204	0.3903	0.8906	0.0746
e_5	-0.0034	0.1229	-0.2429	0.0565	0.9606
Protein content					
	0.0501 (90.1)	**0.0035 (6.4)**	**0.0016 (3.0)**	**0.0002 (0.4)**	**0.0001 (0.1)**
e_1	0.9875	0.1563	0.0212	-0.0071	0.0010
e_2	0.1480	-0.9646	0.2148	-0.0184	-0.0351
e_3	-0.0533	0.1974	0.8858	-0.4164	-0.0113
e_4	0.0138	-0.0787	-0.3618	-0.8204	0.4355
e_5	-0.0027	0.0027	0.1946	0.3913	0.8995

Supplementary Table S2. Eigenvectors and percentage of variance explained by eigenvalues of the random regression coefficients matrix for herd-year of calving effect using highest order model (LEG555) for milk production traits

Eigenvectors (e_i)	**Eigenvalues (λ_i)**				
	λ_1 (%)	**λ_2 (%)**	**λ_3 (%)**	**λ_4 (%)**	**λ_5 (%)**
			Milk yield		
	1.0808 (69.9)	**0.3126 (20.0)**	**0.1152 (7.4)**	**0.0383 (2.5)**	**0.0000 (0.1)**
e_1	0.5802	0.1533	0.7654	-0.0148	-0.2322
e_2	-0.8123	0.1618	0.5288	-0.0583	-0.1759
e_3	-0.0150	-0.8251	0.1145	-0.5259	-0.1712
e_4	0.0552	0.4861	-0.2935	-0.6766	-0.4656
e_5	-0.0183	-0.1823	-0.1879	0.5119	-0.8180
			Fat content		
	0.0066 (70.7)	**0.0018 (20.2)**	**0.0005 (5.5)**	**0.0002 (2.5)**	**0.0001 (1.1)**
e_1	0.2384	0.1078	0.1382	0.9531	0.0640
e_2	-0.8348	-0.4798	-0.0260	0.2631	0.0548
e_3	0.4288	-0.6714	-0.5664	0.0369	0.2078
e_4	-0.2219	0.5057	-0.5295	0.0319	0.6431
e_5	0.1149	-0.2271	0.6156	-0.1415	0.7322
			Protein content		
	0.0017 (63.9)	**0.0006 (23.1)**	**0.0003 (10.3)**	**0.0001 (2.3)**	**0.0001 (0.4)**
e_1	0.3001	-0.4781	-0.5579	0.2649	0.5476
e_2	-0.9153	-0.3619	-0.1767	0.0112	0.0002
e_3	0.2683	-0.6848	-0.0017	-0.2162	-0.6421
e_4	-0.0140	0.3730	-0.6682	0.3701	-0.5265
e_5	-0.0070	-0.1798	0.4593	0.8637	-0.1031

Supplementary Table S3. Eigenvectors and percentage of variance explained by eigenvalues of the random regression coefficients matrix for permanent environment effect using highest order model (LEG555) for milk production traits

Eigenvectors (e_i)	**Eigenvalues (λ_i)**				
	λ_1 (%)	**λ_2 (%)**	**λ_3 (%)**	**λ_4 (%)**	**λ_5 (%)**
			Milk content		
	12.3280 (74.1)	**2.6026 (15.7)**	**1.0801 (6.5)**	**0.5720 (3.4)**	**0.0584 (0.3)**
e_1	0.9939	-0.0810	-0.0651	-0.0177	-0.0311
e_2	-0.0997	-0.9281	-0.2786	-0.2174	-0.0622
e_3	-0.0385	0.3542	-0.7294	-0.4550	-0.3660
e_4	0.0125	-0.0151	0.6194	-0.5857	-0.5224
e_5	-0.0224	-0.0802	-0.0486	0.6344	-0.7670
			Fat content		
	0.0656 (59.0)	**0.0267 (24.0)**	**0.0118 (10.7)**	**0.0061 (5.5)**	**0.0009 (0.8)**
e_1	0.9883	0.1418	0.0046	0.0291	-0.0479
e_2	0.1091	-0.8364	0.4677	0.2634	-0.0190
e_3	-0.1037	0.4611	0.4453	0.6941	-0.3107
e_4	0.0004	-0.2576	-0.6662	0.3200	-0.6224
e_5	-0.0244	0.0365	0.3729	-0.5878	-0.7166
			Protein content		
	0.0191 (59.0)	**0.0068 (22.0)**	**0.0035 (11.0)**	**0.0021 (7.0)**	**0.0006 (1.0)**
e_1	0.9913	0.1071	-0.0753	-0.0159	0.0016
e_2	0.0616	-0.8899	-0.4516	-0.0124	0.0108
e_3	-0.0928	0.3190	-0.6182	-0.6917	0.1705
e_4	0.0643	-0.2741	0.5767	-0.5090	0.5737
e_5	-0.0291	0.1402	-0.2751	0.5119	0.8011

Supplementary Table S4. Eigenvectors and percentage of variance explained by eigenvalues of the random regression coefficients matrix for additive genetic effect using highest order model (LEG555) for milk fatty acids

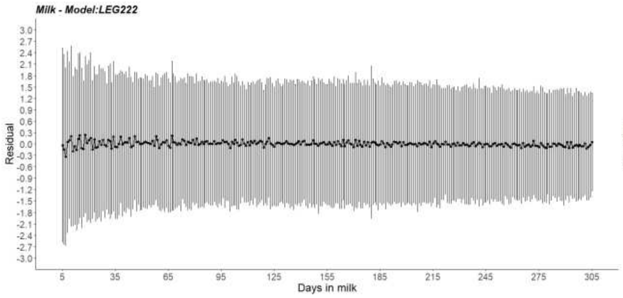
Eigenvectors (e_i)	**Eigenvalues (λ_i)**				
	λ_1 (%)	**λ_2 (%)**	**λ_3 (%)**	**λ_4 (%)**	**λ_5 (%)**
			C16:0		
	0.0274 (93.1)	**0.0013 (4.7)**	**0.0005 (1.5)**	**0.0002 (0.6)**	**0.0001 (0.1)**
e_1	-0.9924	0.1190	0.0201	0.2201	-0.0210
e_2	-0.1007	-0.8934	0.3564	-0.2539	0.0070
e_3	0.0643	0.3952	0.5599	-0.6396	-0.3419
e_4	-0.0215	-0.1676	-0.6743	-0.3651	-0.6190
e_5	0.0163	-0.0567	0.3228	0.6269	-0.7066
			C18:1 *cis-9*		
	0.0031 (85.3)	**0.0004 (9.5)**	**0.0002 (4.7)**	**0.0001 (0.4)**	**0.0001 (0.1)**
e_1	0.4688	0.7594	-0.3281	0.2191	-0.2185
e_2	-0.7332	0.1250	-0.6601	0.0642	-0.0827
e_3	0.4518	-0.3650	-0.6484	-0.2496	0.4239
e_4	-0.1882	0.4503	0.1845	-0.0312	0.8524
e_5	0.0545	-0.2674	-0.0444	0.9404	0.1974
			LCFA		
	0.0064 (81.7)	**0.0009 (11.2)**	**0.0005 (6.5)**	**0.0002 (0.5)**	**0.0001 (0.1)**
e_1	0.3848	0.7904	0.3189	0.2758	-0.2217
e_2	-0.7694	0.0866	0.6280	0.0541	-0.0557
e_3	0.4683	-0.2862	0.6737	-0.2896	0.4010
e_4	-0.1919	0.4450	-0.2146	-0.0806	0.8441
e_5	0.0611	-0.2960	0.0613	0.9113	0.2726
			SFA		
	0.0041 (65.6)	**0.0012 (19.9)**	**0.0007 (12.3)**	**0.0002 (2.1)**	**0.0001 (0.1)**
e_1	0.4028	-0.2044	0.7074	-0.2765	-0.4679
e_2	-0.8905	0.1200	0.2947	-0.1237	-0.3003
e_3	0.2039	0.7867	-0.2657	-0.1056	-0.5075
e_4	-0.0380	-0.5206	-0.4401	0.3172	-0.6580
e_5	0.0390	0.2314	0.3850	0.8924	-0.0127
			UFA		
	0.0045 (84.0)	**0.0005 (9.8)**	**0.0003 (5.5)**	**0.0001 (0.6)**	**0.0001 (0.1)**
e_1	0.4403	0.7367	-0.3790	-0.3459	-0.0046
e_2	-0.7464	0.0557	-0.6525	-0.1162	-0.0198
e_3	0.4525	-0.3767	-0.6384	0.4711	0.1541
e_4	-0.2015	0.4669	0.1506	0.5641	0.6326
e_5	0.0593	-0.3069	-0.0153	-0.5713	0.7586

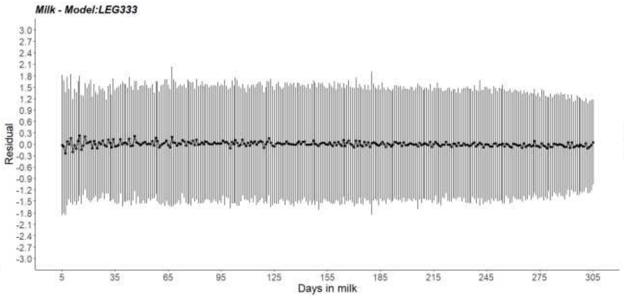
Supplementary Table S5. Eigenvectors and percentage of variance explained by eigenvalues of the random regression coefficients matrix for herd-year of calving effect using highest order model (LEG555) for milk fatty acids

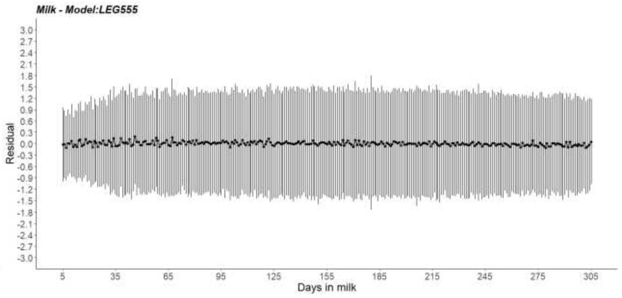
Eigenvectors (e_i)	**Eigenvalues (λ_i)**				
	λ_1 (%)	**λ_2 (%)**	**λ_3 (%)**	**λ_4 (%)**	**λ_5 (%)**
			C16:0		
	0.0348 (94.5)	**0.0014 (3.9)**	**0.0005 (1.4)**	**0.0001 (0.1)**	**0.0001 (0.1)**
e_1	0.9823	0.1768	-0.0535	0.0119	-0.0260
e_2	0.1609	-0.9632	-0.2123	0.0282	-0.0209
e_3	-0.0918	0.1844	-0.8166	0.4264	-0.3298
e_4	0.0258	-0.0684	0.4492	0.8903	-0.0065
e_5	-0.0011	0.0474	-0.2885	0.1562	0.9434
			C18:1 *cis-9*		
	0.0044 (70.8)	**0.0016 (25.6)**	**0.0002 (2.8)**	**0.0002 (0.7)**	**0.0001 (0.1)**
e_1	0.9902	0.1349	0.0261	0.0152	-0.0165
e_2	0.1215	-0.7556	-0.6386	-0.0652	0.0449
e_3	-0.0561	0.5617	-0.6432	-0.4673	-0.2217
e_4	0.0238	-0.2867	0.3324	-0.4312	-0.7878
e_5	-0.0300	0.1138	-0.2591	0.7688	-0.5726
			LCFA		
	0.0196 (79.8)	**0.0042 (16.8)**	**0.0007 (2.7)**	**0.0002 (0.6)**	**0.0001 (0.1)**
e_1	0.9825	0.1789	0.0460	-0.0001	-0.0217
e_2	0.1684	-0.7637	-0.6153	-0.0954	0.0227
e_3	-0.0752	0.5412	-0.6276	-0.4824	-0.2733
e_4	0.0157	-0.2821	0.3801	-0.3574	-0.8049
e_5	-0.0183	0.1101	-0.2842	0.7930	-0.5257
			SFA		
	0.1346 (94.3)	**0.0059 (4.1)**	**0.0018 (1.3)**	**0.0003 (0.2)**	**0.0001 (0.1)**
e_1	0.9805	0.1940	-0.0150	0.0177	-0.0173
e_2	0.1770	-0.8622	0.4731	0.0190	-0.0295
e_3	-0.0820	0.4047	0.7327	0.4700	-0.2675
e_4	0.0202	-0.2054	-0.4108	0.8815	0.1065
e_5	-0.0020	0.1129	0.2649	0.0342	0.9570
			UFA		
	0.016 (77.9)	**0.0028 (18.8)**	**0.0004 (2.8)**	**0.0001 (0.4)**	**0.0001 (0.1)**
e_1	0.9828	0.1704	0.0660	-0.0077	-0.0206
e_2	0.1736	-0.7520	-0.6313	-0.0511	0.0555
e_3	-0.0589	0.5588	-0.6469	-0.5102	-0.7269
e_4	0.0082	-0.2829	0.3377	-0.6510	-0.6180
e_5	-0.0146	0.1139	-0.2538	0.5595	-0.7805

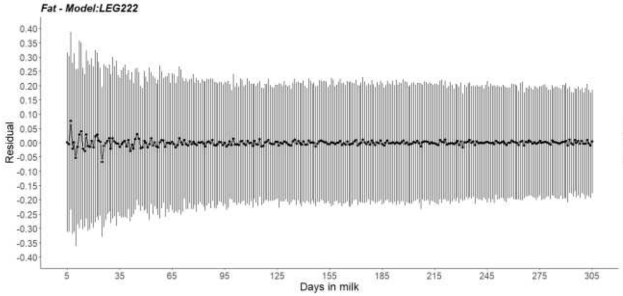
Supplementary Table S6. Eigenvectors and percentage of variance explained by eigenvalues of the random regression coefficients matrix for permanent environment effect using highest order model (LEG555) for milk fatty acids

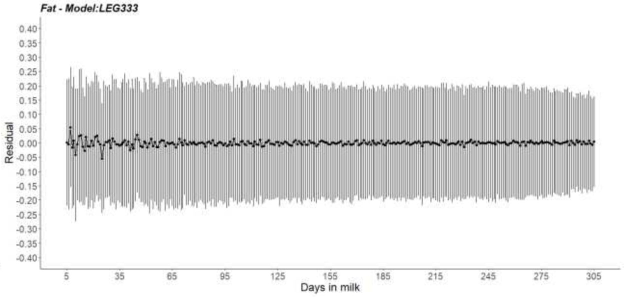
Eigenvectors (e_i)	**Eigenvalues (λ_i)**				
	λ_1 (%)	**λ_2 (%)**	**λ_3 (%)**	**λ_4 (%)**	**λ_5 (%)**
			C16:0		
	0.010 (64.8)	**0.0030 (18.7)**	**0.0015 (9.3)**	**0.0011 (7.1)**	**0.0001 (0.1)**
e_1	0.9694	0.1690	0.1486	-0.0347	-0.0905
e_2	0.1426	-0.9561	0.0677	-0.2404	-0.0540
e_3	-0.1939	0.1791	0.6727	-0.5409	-0.4302
e_4	0.0322	0.1118	-0.7216	-0.5326	-0.4266
e_5	-0.0327	-0.1120	0.0017	0.6083	-0.7885
			C18:1 *cis*-9		
	0.0071 (64.8)	**0.0021 (18.5)**	**0.0013 (12.4)**	**0.0004 (3.5)**	**0.0001 (0.8)**
e_1	0.4278	0.8844	0.1671	0.0570	0.0579
e_2	-0.6641	0.2233	0.6348	-0.3250	-0.0150
e_3	0.5099	-0.2574	0.2424	-0.7548	0.2123
e_4	-0.3219	0.2322	-0.5456	-0.2797	0.6830
e_5	0.1102	-0.2185	0.4610	0.4928	0.6962
			LCFA		
	0.0171 (56.2)	**0.0078 (26.0)**	**0.0039 (12.7)**	**0.0013 (4.3)**	**0.0002 (0.8)**
e_1	0.4815	0.8743	0.0338	0.0221	-0.0448
e_2	-0.6388	0.3377	0.6083	-0.3274	0.0214
e_3	0.4993	-0.2779	0.2872	-0.7406	-0.2057
e_4	-0.3146	0.1679	-0.5640	-0.3051	-0.6793
e_5	0.1077	-0.1265	0.4776	0.5005	-0.7026
			SFA		
	0.0466 (65.3)	**0.0138 (19.5)**	**0.0063 (8.9)**	**0.0040 (5.7)**	**0.0005 (0.6)**
e_1	0.9833	0.1613	0.0463	-0.0395	-0.0574
e_2	0.1276	-0.9254	0.2682	-0.2321	-0.0380
e_3	-0.1276	0.3248	0.5700	-0.6461	-0.3683
e_4	0.0028	-0.0934	-0.7389	-0.3914	-0.5403
e_5	-0.0210	-0.0574	0.2341	0.6114	-0.7533
			UFA		
	0.0107 (58.4)	**0.0043 (23.9)**	**0.0023 (12.5)**	**0.0008 (4.2)**	**0.0001 (1.0)**
e_1	0.4186	0.9027	0.0683	0.0532	0.0478
e_2	-0.6707	0.2856	0.6070	-0.3138	-0.0381
e_3	0.5068	-0.2231	0.2876	-0.7544	0.2034
e_4	-0.3259	0.1738	-0.5514	-0.2952	0.6871
e_5	0.1082	-0.1531	0.4898	0.4921	0.6947

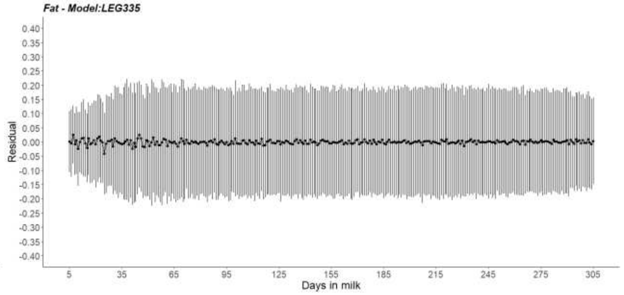

Milk - Model:LEG222
Residual
3.0 2.7 2.4 2.1 1.8 1.5 1.2 0.9 0.6 0.3 0.0 -0.3 -0.6 -0.9 -1.2 -1.5 -1.8 -2.1 -2.4 -2.7 -3.0
5 35 65 95 125 155 185 215 245 275 305
Days in milk

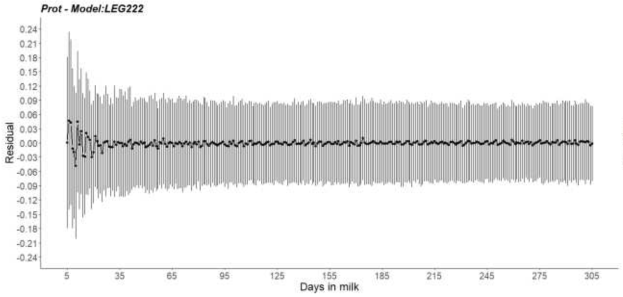

Milk - Model:LEG333
Residual
3.0 2.7 2.4 2.1 1.8 1.5 1.2 0.9 0.6 0.3 0.0 -0.3 -0.6 -0.9 -1.2 -1.5 -1.8 -2.1 -2.4 -2.7 -3.0
5 35 65 95 125 155 185 215 245 275 305
Days in milk

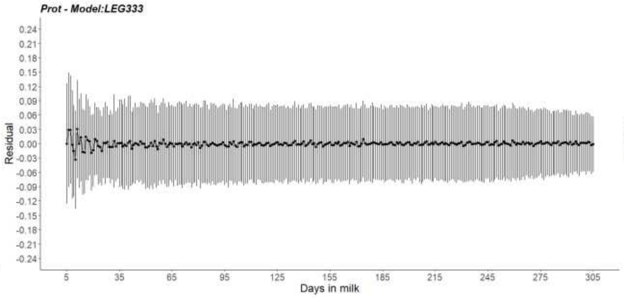

Milk - Model:LEG555
Residual
3.0 2.7 2.4 2.1 1.8 1.5 1.2 0.9 0.6 0.3 0.0 -0.3 -0.6 -0.9 -1.2 -1.5 -1.8 -2.1 -2.4 -2.7 -3.0
5 35 65 95 125 155 185 215 245 275 305
Days in milk

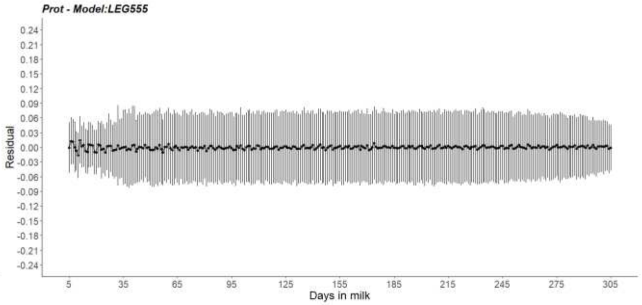

Fat - Model:LEG222
Residual
0.40 0.35 0.30 0.25 0.20 0.15 0.10 0.05 0.00 -0.05 -0.10 -0.15 -0.20 -0.25 -0.30 -0.35 -0.40
5 35 65 95 125 155 185 215 245 275 305
Days in milk

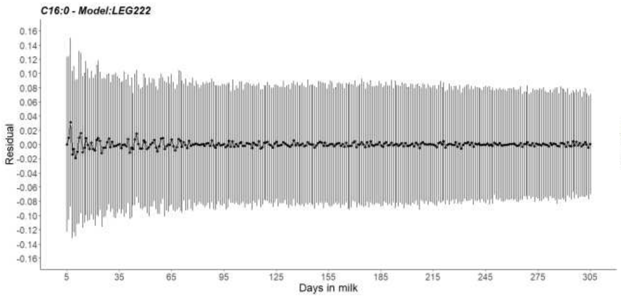

Fat - Model:LEG333
Residual
0.40 0.35 0.30 0.25 0.20 0.15 0.10 0.05 0.00 -0.05 -0.10 -0.15 -0.20 -0.25 -0.30 -0.35 -0.40
5 35 65 95 125 155 185 215 245 275 305
Days in milk


Fat - Model:LEG335
Residual
0.40 0.35 0.30 0.25 0.20 0.15 0.10 0.05 0.00 -0.05 -0.10 -0.15 -0.20 -0.25 -0.30 -0.35 -0.40
5 35 65 95 125 155 185 215 245 275 305
Days in milk

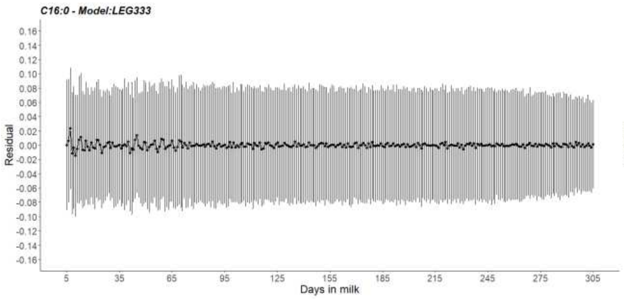

Prot - Model:LEG222
Residual
0.24 0.21 0.18 0.15 0.12 0.09 0.06 0.03 0.00 -0.03 -0.06 -0.09 -0.12 -0.15 -0.18 -0.21 -0.24
5 35 65 95 125 155 185 215 245 275 305
Days in milk


Prot - Model:LEG333
Residual
0.24 0.21 0.18 0.15 0.12 0.09 0.06 0.03 0.00 -0.03 -0.06 -0.09 -0.12 -0.15 -0.18 -0.21 -0.24
5 35 65 95 125 155 185 215 245 275 305
Days in milk

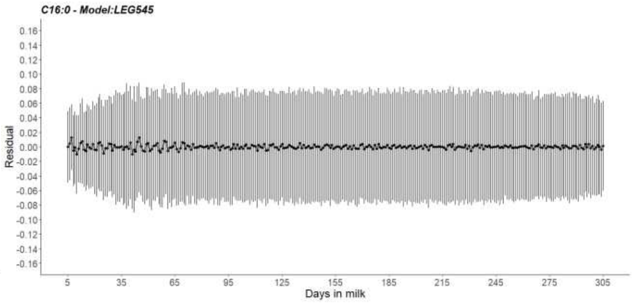

Prot - Model:LEG555
Residual
0.24 0.21 0.18 0.15 0.12 0.09 0.06 0.03 0.00 -0.03 -0.06 -0.09 -0.12 -0.15 -0.18 -0.21 -0.24
5 35 65 95 125 155 185 215 245 275 305
Days in milk



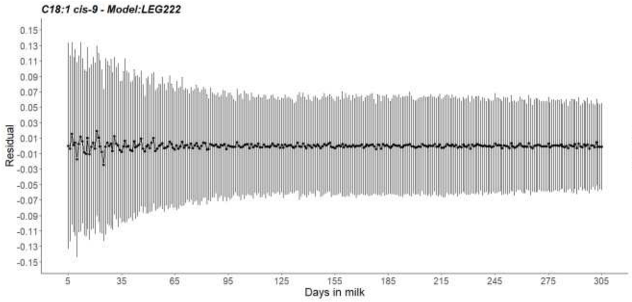
C16:0 - Model:LEG222



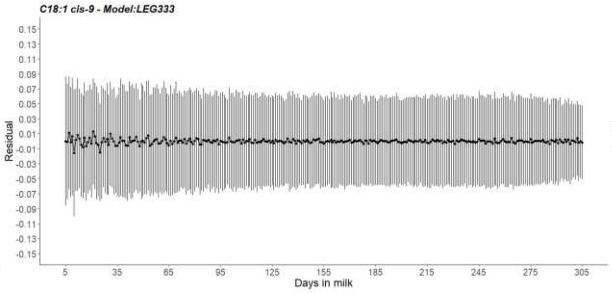
C16:0 - Model:LEG333



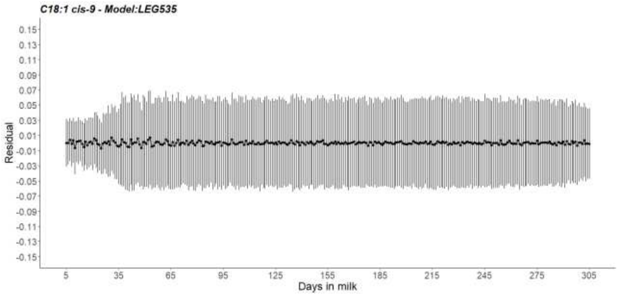
C16:0 - Model:LEG545



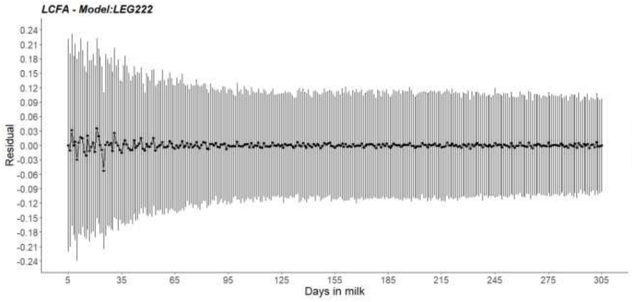
C18:1 cis-9 - Model:LEG222



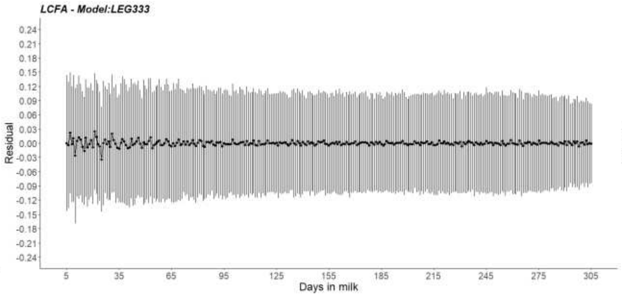
C18:1 cis-9 - Model:LEG333



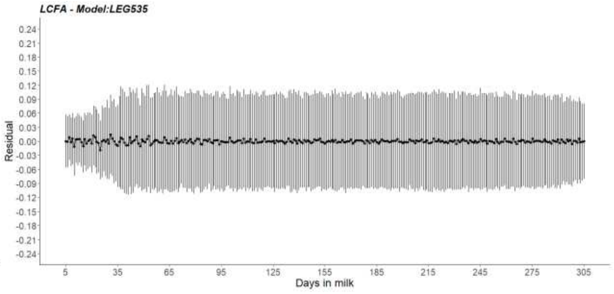
C18:1 cis-9 - Model:LEG535



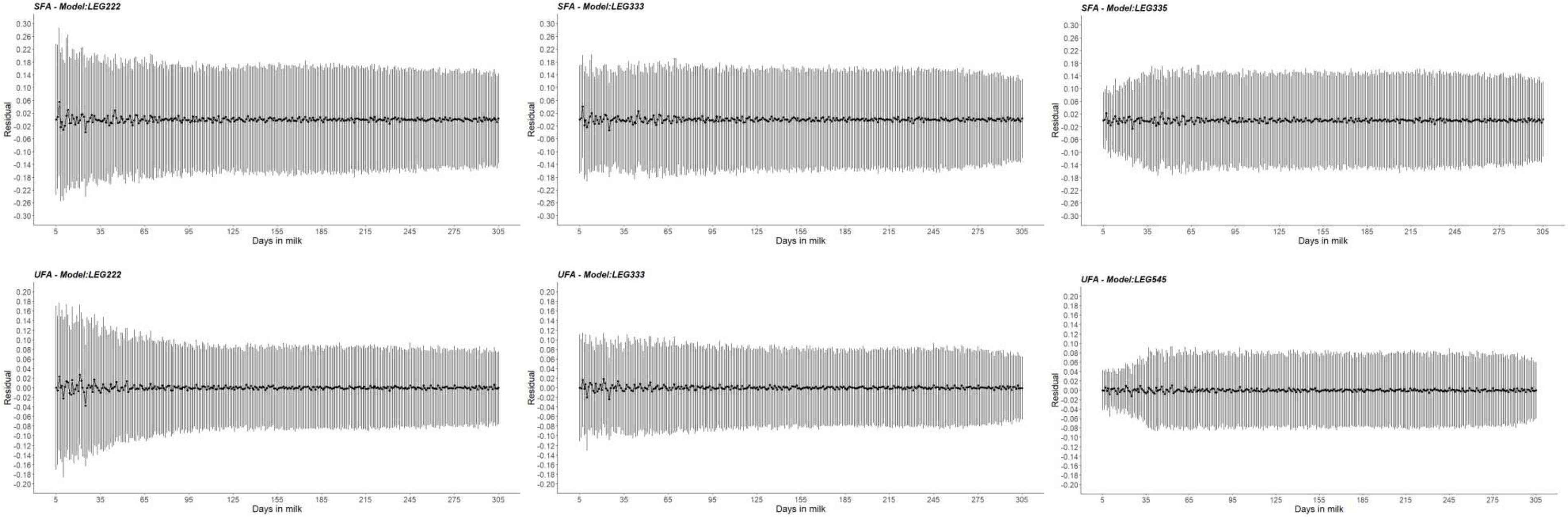
LCFA - Model:LEG222



LCFA - Model:LEG333



LCFA - Model:LEG535



Supplementary Figure S1. Mean residuals and standard deviation over days in milk for random regression models with the second (on the left), third (in the middle) and best fit (on the right) orders.

CHAPTER 3

Genome-enabled prediction for milk production and fatty acid traits in Walloon Holstein cattle via random regression models

Paiva, J.T.[1]; Mota, R.R.[2a]; Lopes, P.S.[1]; Hammami, H.[2]; Vanderick, S.[2]; Oliveira, H.R.[3]; Veroneze, R.[1]; Silva, F.F.[1]; Gengler, N.[2]

[1] Department of Animal Sciences, Universidade Federal de Viçosa, Viçosa, MG, Brazil

[2] Gembloux Agro-Bio Tech, University of Liège, TERRA Teaching and Research Centre, B-5030 Gembloux, Belgium

[3] Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, Ontario, Canada

[a] Present address: Council on Dairy Cattle Breeding – CDCB, Bowie, Maryland, USA

3.1. Abstract

Combining phenotypic records, pedigree, and genomic information simultaneously in the genetic evaluation, the single-step genomic best linear unbiased prediction (ssGBLUP) approach can lead to more accurate and less biased GEBV. In addition, the use of ssGBLUP based on RRM enable the estimation of GEBV for all days in milk, which may increase the reliability of genomic predictions for test-day traits. Therefore, the aims of this study were to: 1) estimate genetic parameters for milk production traits (milk, fat and protein yield, and fat and protein content) and FA (C16:0, C18:1 *cis*-9, LCFA, SFA, and UFA) over days in milk, 2) investigate the performance of genomic predictions (in terms of reliability and bias) using the single-step GBLUP (ssGBLUP) and to compare with parent average (PA) approach in Walloon Holstein dairy cattle; and 3) identify the optimal scaling and weighting factors to be used in the construction of the **H** matrix in this population. A total of 302,684 test-day records of 63.875 first lactation Walloon Holstein cows were used in the ssGBLUP evaluations based on random regression models (RRM). Daily heritability range from low to high magnitude (from 0.11 to 0.59 for milk production traits, and from 0.09 to 0.57 for milk FA) and tended to increase over lactation. Average genetic correlations between milk yield, and fat and protein yield, ranged from 0.46 to 0.85. On the other hand, negative genetic correlations ranging from -0.22 to -0.59 were estimated between fat and protein contents, and from -0.22 to -0.62 between milk yield and milk FA. Positive genetic correlations (ranging from 0.17 to 0.47) were estimated between fat yield and milk FA, while negative estimates (average from -0.11 to -0.19) were found between protein yield and milk FA. The selection for high fat content increases milk FA throughout lactation (average from 0.61 to 0.98). Genetic correlations among all milk FA traits were positive over days in milk. The test-day ssGBLUP approach showed considerably higher prediction reliability compared to PA method for all milk production and FA traits, even when no scaling and weighting factors were used in the **H** matrix. The highest validation reliabilities

(r^2 ranged from 0.09 to 0.38) and less biased predictions (b_1 ranged from 0.76 to 0.92) were obtained using the optimal parameters (i.e., $\omega = 0.7$ and $\alpha = 0.6$) for the genomic evaluation of milk production traits. For milk FA, the optimal parameters were $\omega = 0.6$ and $\alpha = 0.6$, however, biased predictions were still observed (b_1 from 0.32 to 0.81) using them for most FA. The findings reported in this study suggest that using ssGBLUP based on RRM is feasible for genomic prediction of daily milk production and FA traits in Walloon Holstein dairy cattle. However, further studies are required to improve reliabilities and reduce bias for milk FA.

Keywords: genetic correlation; genomic prediction; MIR; random regression model; test-day; single-step GBLUP

3.2. Introduction

Since the 1950s, changes to the milk pricing structure were proposed and most milk markets have used milk solids to define price, with an economic value projected especially for fat and protein production (Miglior et al., 2017; Britt et al., 2018). The triacylglycerols (TAG) account for about 95% of the lipid fraction in milk solids and are composed of fatty acids (FA) of different lengths and saturation (Haug et al., 2007). Currently, the prediction of milk FA using mid-infrared (MIR) spectroscopy has showed high prediction accuracy in the Walloon region of Belgium (Soyeurt et al., 2011) and other countries (Rodriguez et al., 2014; Fleming et al., 2017). Interest in the milk FA profile is increasing given its important nutritional value for human health, and due to its influence on the technological properties of milk and dairy products (Gengler et al., 2016; Miglior et al., 2017). In addition, increased levels of unsaturated FA (UFA) content in milk can contribute to lower stability and oxidation, which directly impact dairy production (Hanuš et al., 2018). On the other hand, C18:1 *cis*-9 (oleic acid) is the unsaturated FA with the highest concentration in milk, which was considered to be favorable for human health (Haug et al., 2007). Van Haelst et al. (2008) and Jorjong et al. (2014) reported that C18:1 *cis*-9 shows potential as an early warning biomarker for metabolic status and subclinical ketosis in dairy cows. More recently, Mäntysaari et al. (2019) reported that the use of milk FA alone predicted postpartum energy status of cows better than milk yield, milk fat:protein ratio, and body traits combined. Thus, despite the milk components traditionally used for breeding and management in the farm (e.g., fat and protein contents) monitoring milk FA profiles has become paramount (Gengler et al., 2016).

Studies have been shown that milk FA profile is under genetic control in dairy cattle (Paiva et al., 2021; Bastin et al., 2013; Petrini et al., 2016; Narayana et al., 2017) and thus, changes in FA composition can be made through genetic selection. Moreover, the availability of these specialized phenotypes makes possible their inclusion in the genetic and genomic evaluations, which brings unprecedented and substantial impacts to improve milk quality. However, potential implications of selection for milk production and FA traits requires the estimation of genetic parameters in the population studied. Evaluating the impact of selecting one trait instead of another one is essential to predict indirect genetic gains in order to determine the best selection strategies in a breeding program. Heritability and genetic correlations are expected to change when animals are under selection, with the amount of change dependent on the intensity of selection and initial genetic variance. Thus, the genetic gains as predicted using old parameters may not be realized in the population (Hidalgo et al., 2020).

The identification of genetically superior animals for milk quality traits is paramount in view of the current trend of payment policy based on milk quality in many countries. Nowadays, with the wide availability of genomic information, the genetic similarity between relatives can be determined more accurately (VanRaden et al., 2009). In addition, the variance components estimated without genomic information could be biased for populations undergone genomic selection (Hidalgo et al., 2020). According to Weller, Ezra, & Ron (2017), the implementation of genomic evaluation has also lead to inclusion of more traits in the selection indices, and reduction of the emphasis on fat and protein production. In this context, to our best knowledge, there are few studies evaluating the impact of the selection of milk production traits on the FA profile over days in milk (e.g., Bastin et al., 2011; Fleming et al., 2018). It is important to mention that the previous authors did not perform genomic analysis, only pedigree-based.

The inclusion of genomic information has allowed an improvement of genomic estimated breeding value (GEBV) accuracies over the traditional pedigree-based estimated breeding values (EBV; Legarra et al., 2014). Single-step genomic best linear unbiased prediction (ssGBLUP) is a unified approach that optimally combines phenotypic records, pedigree information, and genomic information in the calculation of GEBV (Misztal et al., 2009; Aguilar et al., 2010; Christensen and Lund, 2010). Currently, given the limitation of genotyped animals in some real populations, ssGBLUP is among the most efficient methods and has been routinely used in animal breeding (Lourenco et al., 2020). The theoretical advantage of ssGBLUP is that it uses all phenotypic information for which either the pedigree or marker effect are known (Wang et al., 2014). The use of extra information on non-genotyped animals and the independency f pseudo-phenotypes are responsible for gains in accuracy over multiple-step approaches (Lourenco et al., 2014). However, this approach usually requires the use of optimal scaling factors (Vitezica et al., 2011; Misztal et al., 2013, 2017; Lourenco et al., 2020) to assure the ideal compatibility between the marker-based relationship and pedigree-based relationship matrices. Adjusting the genomic relationship matrix toward their expected values in the pedigree-based matrix can improve the accuracy and reduce bias of GEBV. In addition, genomic predictions for various time points can be simultaneously done using random regression model (RRM), which enables selection of breeding animals based on the complete pattern of the production curve. In general, studies using ssGBLUP based on RRM provided higher accuracy and less biased GEBV compared with other methods (Koivula et al., 2015;

Kang et al., 2017, 2018; Oliveira et al., 2019a). To our best knowledge, there is no study investigating the optimal scaling factors to perform the ssGBLUP evaluation for milk FA using RRM in dairy cattle.

Therefore, the aims of this study were to: 1) estimate heritability and genetic correlation for milk production traits (milk, fat and protein yield, and fat and protein content) and FA (C16:0, C18:1 *cis*-9, LCFA, SFA, and UFA) over days in milk using pedigree and genomic information; 2) investigate the performance of genomic prediction (in terms of reliability and bias) using the ssGBLUP approach and to compare with the pedigree-based method in Walloon Holstein cattle; and 3) identify the optimal scaling factors to be used in the construction of the **H** matrix in this population.

3.3. Material and methods

3.3.1. Datasets

Phenotypic and genotypic data were extracted from the Walloon genetic evaluation performed in Belgium. Milk samples are routinely collected by the Walloon Breeding Association (AWE; Ciney, Belgium) and are analyzed by a mid-infrared MilkoScan FT6000 spectrometer (Foss, Hillerød, Denmark). The generated spectral dataset is routinely used to predict FA contents in milk (g/dL of milk), based on the calibration equations developed by Soyeurt et al. (2011). Fat yield (kg), protein yield (kg), fat content (%), protein content (%), C16:0 fatty acid (palmitic acid), C18:1 *cis*-9 fatty acid (oleic acid), long-chain fatty acid (LCFA), saturated fatty acids (SFA), and group unsaturated fatty acids (UFA) were the traits predicted from the test-day milk samples evaluated in this study besides daily milk yield (kg). More details about the dataset edition used in this study can be found in the previous chapter.

The final dataset comprised 302,684 test-days records from 63,875 first-parity Walloon Holstein cows, from 856 herds. The pedigree data used to create the pedigree-based relationship matrix contained 228,820 animals, which are 6,562 sires and 114,830 dams. Descriptive statistics of milk production and FA traits are presented in Table 1.

Table 1. Descriptive statistics of test-day records for milk production and fatty acids (g/dL of milk) traits in first lactation Walloon Holstein cows

Traits[1]	**Mean**	**SD[1]**	**CV[1] (%)**	**Minimum**	**Maximum**
Milk yield (kg)	23.32	5.63	24.15	3.40	78.00
Fat yield (kg)	0.91	0.21	23.07	0.10	2.83
Protein yield (kg)	0.78	0.19	24.35	0.11	2.51
Fat content (%)	3.95	0.54	13.59	2.28	8.87
Protein content (%)	3.39	0.29	8.71	1.91	4.95
C16:0 (x100)	1.263	0.238	18.84	0.678	1.915
C18:1 *cis*-9 (x100)	0.784	0.141	17.98	0.484	1.453
LCFA (x100)	1.592	0.259	16.27	1.004	2.705
SFA (x100)	2.801	0.457	16.31	1.277	4.484
UFA (x100)	1.234	0.201	16.29	0.665	3.197

[1] C16:0 = palmitic acid; C18:1 *cis*-9 = oleic acid; LCFA = long-chain fatty acids; SFA = saturated fatty acids; UFA = unsaturated fatty acids; SD = standard deviation; CV = coefficient of variation.

Routine evaluations genotypes for Walloon Holsteins come from 11 different SNP chips (Illu50K, Illu50K2, Illu50K3, IlluHD, GGP150K, IDB3, EuroG10K, IlluHD3, EuroG_MD) of different densities. Therefore, imputation and edits are done in order to come up with a common list of single nucleotide polymorphism (SNP) markers. The current list contains 19,468 SNPs in common with all 11 chips in the data base. In this study, a total of 5,057 animals were genotyped (620 sires and 1,931 dams in the pedigree). Genotypic quality control was performed using the preGSf90 software (Aguilar et al., 2014; Misztal et al., 2018). SNP markers with minor allele frequency lower than 0.05 (n = 1,167), call rate lower than 0.90 (n = 494), monomorphic (n = 0), and SNP with Mendelian conflicts (n = 3) were removed. Likewise, individual samples with call rate lower than 0.90 (n = 0) and with parent-progeny Mendelian conflicts (n = 20) were also excluded. After the quality control, the genotypic dataset included 5,037 genotyped animals (2,865 cows and 525 sires in the datafile) and 17,866 SNPs, which were used for further analysis.

3.3.2. Statistical model

The random regression test-day bi-trait model (RRM) was based on the statistical model used in the official Walloon genetic evaluations for milk production traits (Auvray and Gengler, 2002; Croquet et al., 2006), and as described in the previous chapter for milk FA:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Q}\,(\mathbf{Why} + \mathbf{Za} + \mathbf{Zp}) + \mathbf{e}, \qquad (1)$$

where **y** is the vector of phenotypic records for each trait; $\boldsymbol{\beta}$ is the vector of systematic effects, which included herd x test-day (28,112 classes), gestation stage (12 classes), minor lactation stage (61 classes of 5 DIM), and major lactation stage (classes of 73 DIM) x age at calving (21–40 months) x season of calving (four seasons: winter from Jan-Mar, spring from Apr-Jun, summer from Jul-Sep and autumn from Oct-Dec) (8,820 classes); **hy** is the vector of herd-year of calving (4,361 classes) random regression coefficients; **a** is the vector of additive genetic (228,820 animals) random regression coefficients; **p** is the vector of permanent environment (63,875 animals) random regression coefficients; **X**, **W**, and **Z** are the incidence matrices assigning the observations to effects; **Q** is the covariate matrix for the third-order Legendre polynomials; and **e** is the vector of random residuals.

The expectations and covariance structure for the random effects were given by:

$$\mathrm{E}[\mathbf{y}] = \begin{bmatrix} \mathbf{X\beta} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \text{and Var} \begin{bmatrix} \mathbf{hy} \\ \mathbf{a} \\ \mathbf{p} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{HY}_0 \otimes \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_0 \otimes \mathbf{H} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{P}_0 \otimes \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R}_0 \otimes \mathbf{I} \end{bmatrix}, \tag{2}$$

where $\mathbf{HY_0}$, $\mathbf{G_0}$, and $\mathbf{P_0}$ are the random regression coefficients (co)variances matrices for herd-year of calving, additive genetic, and permanent environment effects, respectively; **I** is the identity matrix; **H** is a matrix that combines pedigree and genomic information; $\mathbf{R}_0$ is the (co)variance matrix of residual effects among traits, assumed to be homogeneous over the lactation; and $\otimes$ is the Kronecker product. More details about the distribution of regression coefficients (co)variance matrices can be found in Paiva et al. (2021).

The inverse of the **H** matrix was created as Aguilar et al. (2010), and Christensen and Lund (2010):

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \tau(\alpha\mathbf{G} + \beta\mathbf{A}_{22})^{-1} - \omega\mathbf{A}_{22}^{-1} \end{bmatrix}, \tag{3}$$

where $\mathbf{A}^{-1}$ is the inverse of the pedigree-based numerator relationship matrix, **G** is the genomic relationship matrix; and $\mathbf{A}_{22}$ is the proportion of $\mathbf{A}^{-1}$ for the genotyped animals. The genomic relationship matrix was created as proposed by VanRaden (2008):

$$\mathbf{G} = \frac{\mathbf{ZZ'}}{2\sum \mathrm{p}_i(1-\mathrm{p}_i)}, \tag{4}$$

where **Z** = **M** – **P**; **M** is a matrix of genotypes centered for allele frequencies (i.e., 0, 1 and 2 to represent *aa*, *Aa* and *AA*, respectively), with dimensions equal to a number of animals by number of SNP; and **P** contains the allele frequency for SNP *i* (p_i) in its column *i*th, expressed as $2(\mathrm{p}_i - 0.5)$, and $2\sum \mathrm{p}_i(1-\mathrm{p}_i)$ is a scaling parameter. The default scaling (i.e., $\tau = 1$ and $\omega = 1$) and weighting (i.e., $\alpha = 0.95$ and $\beta = 0.05$) factors were used to estimate the genetic parameters.

3.3.3. Genetic parameter estimation

The (co)variance components and genetic parameters (i.e., heritabilities and genetic correlations) were estimated through bi-trait analysis via Gibbs sampling algorithm using the GIBBS2F90 software (Misztal et al., 2002). A chain length of 250,000 iterations was established, with a burn-in period of 50,000 cycles, and a sampling interval of 10 cycles, corresponding to 20,000 samples for subsequent analysis. Convergence criteria were monitored by Geweke (1991) criterion using the "*Bayesian Output Analysis*" package (Smith, 2007) available on R software (R Development Core Team, 2019, https://www.R-project.org/).

The genetic (co)variance matrix for all DIM was obtained according to Druet et al. (2003), which is described as:

$$\mathbf{G^* = QG_0Q'} \,, \tag{5}$$

where **G*** is a genetic (co)variance matrix among the traits for all DIM ranging from 5 to 305 day; $\mathbf{G_0}$ is a (co)variance matrix of genetic regression coefficients; and **Q** is a matrix with the values of the three coefficients of the third-order Legendre polynomial for each DIM. Likewise, the (co)variance matrices of herd-year of calving and permanent environment were obtained from the $\mathbf{HY_0}$ and $\mathbf{P_0}$ matrices, respectively.

The posterior marginal distribution samples for heritability (h_j^2) and genetic correlation (r_{g_j}) at test-day *j* were calculated:

$$h_j^2 = \frac{\sigma_{a_j}^2}{\sigma_{hyc_j}^2 + \sigma_{a_j}^2 + \sigma_{p_j}^2 + \sigma_e^2} \text{ and } r_{g_j} = \frac{\sigma_{a12j}}{\sqrt{\sigma_{a1_j}^2 \times \sigma_{a2_j}^2}} \,, \tag{6}$$

where $\sigma_{a_j}^2$, $\sigma_{p_j}^2$ and $\sigma_{hyc_j}^2$ are the additive genetic, permanent environment and herd-year of calving variances at test-day *j*, respectively; σ_e^2 is the residual variance, σ_{a12j} is the additive genetic covariance between trait 1 and 2 at test-day *j*, and $\sigma_{a1_j}^2$ and $\sigma_{a2_j}^2$ are the additive genetic variances of trait 1 and 2, respectively, at test-day *j*. Daily heritabilities for each trait studied were presented as the average of all estimates obtained in each bi-trait analysis.

3.3.4. Breeding values prediction

Breeding values for random regression coefficients of each animal were predicted using both the traditional BLUP (**A** matrix) and the ssGBLUP (**H** matrix) approaches on single-trait analysis. The optimal values for the scaling factors ($\tau = 1$ and $\omega = 0.70$) reported by Tsuruta et al. (2011) and used by (Colinet et al., 2018), in a subset of the same Walloon Holstein dairy population, were used in this study for the genomic evaluation of milk production traits. In summary, the ω parameter helps to reduce the GEBV inflation and compensates for pedigree incompleteness (Masuda et al., 2016; Lourenco et al., 2020). Moreover, τ and ω are used to account for the reduced genetic variance and for different depths of pedigree in order to make $\mathbf{G}^{-1}$ compatible with $\mathbf{A}_{22}^{-1}$ and also $\mathbf{A}^{-1}$. The weighting factors ($\alpha = 0.60$ and $\beta = 0.40$) were used for milk production traits as suggested by Colinet et al. (2018), because these values reflect the partition of the genetic variance between the SNP markers and residual polygenic parts. The scaling and weighting factors described previously have been used in the official Walloon evaluation for milk production traits.

As changes in scaling and weighting factors were not investigated for milk FA in the population analyzed in this study, scenarios (see Table 2) with different values of scaling factors were tested for ω (i.e., 0.6, 0.7, 0.8, 0.9, and 1.0). In this analysis, the proportion of markers variance was fixed to $\alpha = 0.95$ and the polygenic variance was fixed to $\beta = 0.05$ (default values). Also, various scenarios with fixed $\omega = 1$ (default values) and different values of α (i.e., 0.6, 0.7, 0.8, 0.95) were tested. It is worth noting that for all scenarios $\tau = 1$ (default values) was used, since changes in this parameter have been reported to have small effects on the bias of GEBV (Oliveira et al., 2019b). The best combination of these parameters for each trait analyzed was chosen according to the validation reliabilities and regression coefficients (bias; see details in the "Prediction Reliability and Bias" topic).

For each trait, EBV and GEBV of animal *i* at test-day *j* were obtained from a posteriori distribution of additive genetic (estimated by BLUP) and genomic (estimated by ssGBLUP) random regression coefficients as follows:

$$\mathrm{EBV}_{ij} = \mathbf{C}_j \hat{\mathbf{a}}_i \quad \text{and} \quad \mathrm{GEBV}_{ij} = \mathbf{C}_j \hat{\mathbf{g}}_i \ , \tag{7}$$

where $\mathbf{C}_j$ is a matrix of independent covariates associated with the Legendre polynomials for test-day *j*; and $\hat{\mathbf{a}}_i$ and $\hat{\mathbf{g}}_i$ are the vector of predicted additive genetic and genomic coefficients for each animal *i*, respectively. The blupf90 software (Misztal et al., 2002) was used to obtain the solutions of the additive genetic and genomic for the random regression coefficients.

3.3.5. Prediction reliability and bias

A reduced data set was created in which the last four years of the full data were cut off (i.e., excluding records of daughters of young bulls). Therefore, the reduced data set included all phenotypic information available until 2015 and it was used to predict GEBV using ssGBLUP, and parent average (PA) for validation animals using the traditional BLUP. This scenario allows us to mimic the Interbull GEBV test (Mäntysaari et al., 2010). The full data set was used to predict current EBV, which was used as a benchmark to validate GEBV and PA obtained from the reduced data set for validation animals by assessing the reliability and bias of genomic predictions. Genotyped bulls that only had daughters born after 2015 were defined as validation animals (151 bulls). These 151 validation bulls had 3,583 daughters with a total of 17,177 test-day records. All animals born before 2015 (excluding the validation bulls) were included in the training population. In summary, 3,556 bulls (374 genotyped bulls) and 285,507 test-day records from 60,292 cows (2,695 genotyped cows) were included in the training population. The total number of phenotyped and genotyped animals and test-day records in full, training, and validation data sets are given in Table 2.

Table 2. Structure of full, training and validation data sets with the total number of test-day records and number of genotyped animals for prediction analysis

Dataset	**Phenotypes**	**Genotypes**
Full	302,684 test-days 63,875 cows 3,707 sires 50,855 dams	14,915 test-days 2,865 cows 525 sires 1,153 dams
Training	285,507 test-days 60,292 cows 3,556 sires 48,350 dams	14,043 test-days 2,695 cows 374 sires 890 dams
Validation	17,177 test-days 3,583 cows 151 sires 3,478 dams	872 test-days 170 cows 151 sires 319 dams

The validation reliability for each trait was calculated as the squared Pearson correlation coefficient (r^2) between daily GEBV estimated based on the reduced data set (i.e., excluding phenotypes for the validation animals or their descendants) and daily EBV estimated based on the full data set, considering only animals in the validation population. Full-data set EBV have been also used to validate the performance of genomic predictions using RRM in other studies (Oliveira et al., 2019b; Freitas et al., 2020). Likewise, only the bulls from the validation population were considered to assess the genomic prediction bias, which was calculated by obtaining the regression coefficient (b_1) estimated using a linear regression of the daily EBV from the full data set on GEBV from the reduced data set (i.e., $EBV = b_0 + b_1 \times GEBV$) (Mäntysaari et al., 2010). To evaluate the impact of including genomic information and to compare prediction reliability and bias of GEBV to those of EBV from traditional genetic evaluation, the parent average (PA) was predicted for validation animals. The PA was used to calculate r^2 (i.e., squared Pearson correlation coefficient between EBV and PA) and b_1 (i.e., obtained from EBV = $b_0 + b_1 \times$ PA) using daily PA and daily EBV predicted based on the reduced and full data set, respectively.

3.4. Results

3.4.1. Genetic parameter estimates

Daily posterior mean heritability ranged from 0.18 to 0.59 for fat and protein content and from 0.11 to 0.30 for milk, fat and protein yield, throughout the first lactation of Walloon Holstein cows. Figure 1.a shows the heritability estimates over days in milk for milk production traits. Average daily heritabilities are shown in Table 3. These values were 0.46 and 0.50 for fat and protein content, respectively, and 0.25, 0.21 and 0.19 for milk, fat and protein yield. Heritabilities showed a similar trend increasing along the curve of lactation and a slight decline at the end of lactation, with estimates of 0.51 for fat and protein content at 305DIM, and 0.25, 0.23 and 0.20, respectively, for milk, fat and protein yield (Supplementary Table S1).

Likewise, heritability for milk FA traits increased from the beginning of lactation to the end. The heritabilities over days in milk for milk fatty acid traits are shown in Figure 1.b. Heritabilities estimated for C16:0 and SFA test-day records were the highest among all FA traits (they ranged from 0.20 to 0.57). Although the curves patterns were similar, estimates for C18:1 *cis*-9 tended to be lower than for the other FA, ranging from 0.09 to 0.25. LCFA and UFA showed values closest over lactation and tended to be somewhat higher (range 0.10-0 .37). Average daily heritabilities were 0.46 and 0.48 for C16:0 and SFA, respectively, 0.23 for C18:1 *cis*-9, and 0.29 for LCFA and UFA (Table 3). Heritabilities at 305DIM for all traits ranged from low to moderate (0.25 to 0.53; Supplementary Table S1).

Table 3. Average daily heritability (on the diagonal) and genetic correlations estimates (above the diagonal) of milk production and FA traits[1] in first lactation Walloon Holstein cows

	Milk yield	Fat yield	Protein yield	Fat content	Protein content	C16:0	C18:1 *cis*-9	LCFA	SFA	UFA
Milk yield	0.25	0.52	0.81	-0.52	-0.55	-0.47	-0.38	-0.49	-0.48	-0.49
Fat yield		0.21	0.70	0.46	0.09	0.47	0.17	0.31	0.47	0.26
Protein yield			0.19	-0.13	0.03	-0.11	-0.18	-0.19	-0.12	-0.15
Fat content				0.46	0.67	0.96	0.61	0.83	0.98	0.79
Protein content					0.50	0.62	0.41	0.57	0.64	0.63
C16:0						0.46	0.44	0.69	0.98	0.62
C18:1 *cis*-9							0.23	0.91	0.45	0.93
LCFA								0.29	0.72	0.97
SFA									0.48	0.65
UFA										0.29

[1] C16:0 = palmitic acid; C18:1 *cis*-9 = oleic acid; LCFA = long-chain fatty acids; SFA = saturated fatty acids; UFA = unsaturated fatty acids.

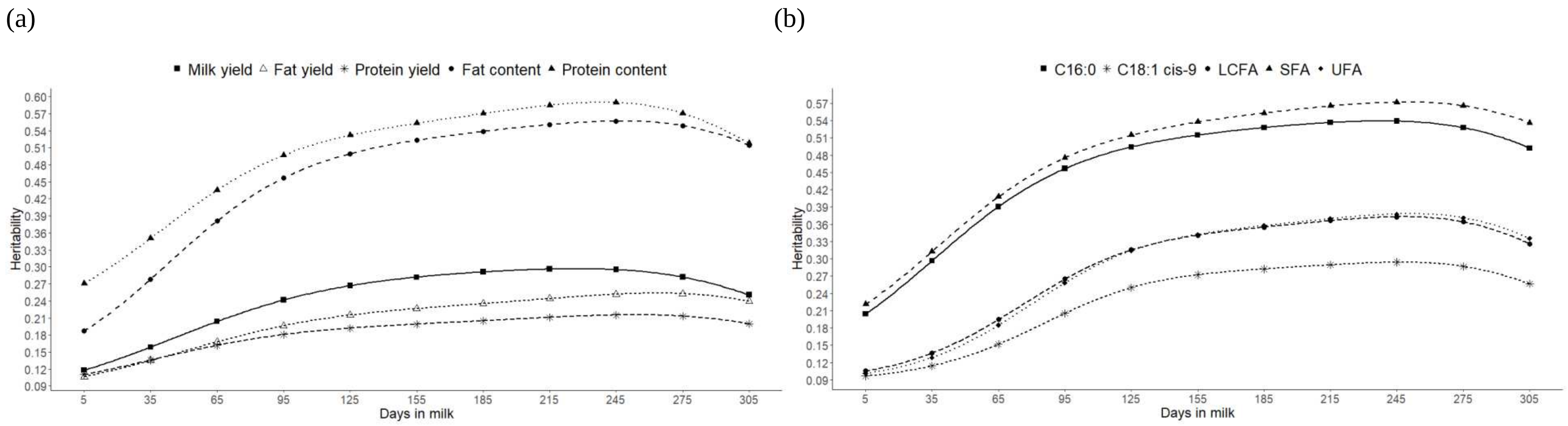


Figure 1. Posterior means of heritability over days in milk for (a) milk production traits (milk, fat, and protein yield; and fat and protein content) and (b) milk individual FA (C16:0 and C18:1 *cis*-9) and FA groups (LCFA, SFA, and UFA).

Average daily genetic correlations were positive between milk and fat (0.52) and milk and protein (0.81) yield (Table 3). On the other hand, negative and moderate average genetic correlations were estimated for fat (-0.52) and protein (-0.55) contents with milk yield, as well as FA and milk yield (ranged from -0.38 to -0.49; Table 3). Likewise, at 305DIM those estimates ranged from -0.37 to -0.58 (Supplementary Table S1). Figure 2.a shows the genetic correlations of fat and protein yield with milk yield (range from 0.46 to 0.85) over days in milk, as well as of fat and protein contents (range from -0.22 to -0.59).

Posterior mean genetic correlations of FA with milk yield over days in milk are shown in Figure 2.b. In general, most estimates were negative, and higher magnitudes were observed in the middle of lactation. Genetic correlations between milk yield and C18:1 *cis*-9, LCFA, and UFA were low and positive (maximum value of 0.22, 0.04, and 0.13, respectively) at the beginning of lactation (from 5 to 35DIM). The genetic correlations curves of milk yield with fat yield and content, C16:0, and SFA had similar patterns across lactation, with estimates ranging from -0.22 to -0.59.

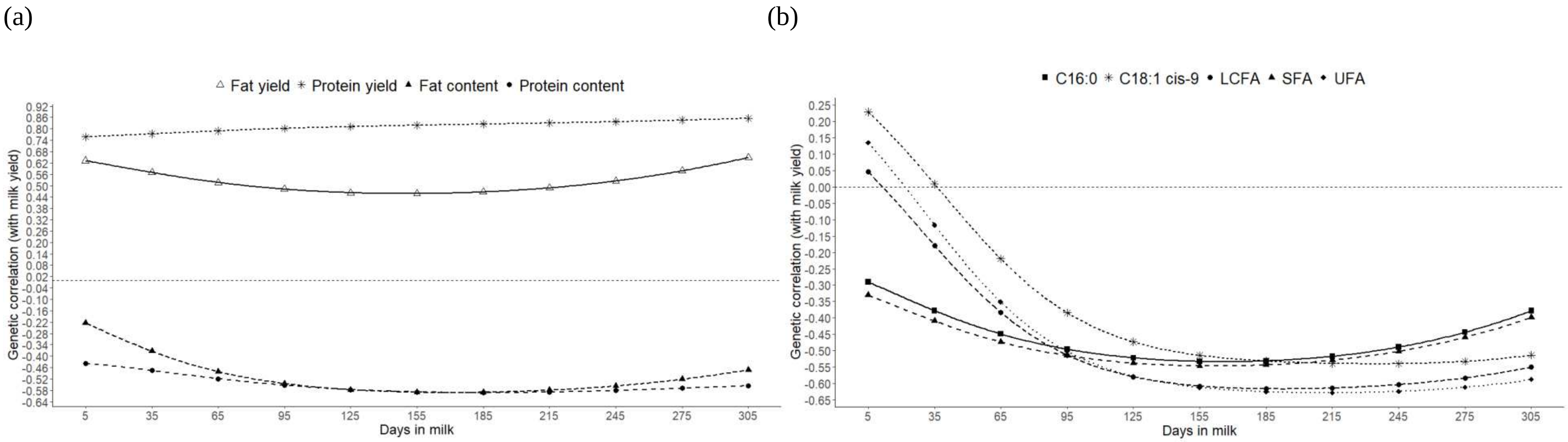


Figure 2. Posterior means of genetic correlations over days in milk estimated between milk yield and (a) milk production and (b) FA traits.

Positive genetic correlations were estimated between fat yield and milk FA, and they tended to decrease across lactation, as shown in Figure 3.a. Higher estimates were found for C16:0 (range from 0.43 to 0.49) and for other FA the estimates curve were similar with lowest values at 305DIM (Supplementary Table S1). Average daily genetic correlations between fat yield and FA were moderate, 0.47 for C16:0 and SFA, 0.31 for LCFA, 0.26 for UFA, and 0.17 for C18:1 *cis*-9 (Table 3). Genetic correlations of milk FA with protein yield presented similar pattern over days in milk, as shown in Figure 3.b. Positive estimates were found at beginning of lactation (35-65DIM) and became negative thereafter, with average daily genetic correlations ranging from -0.11 to -0.19 (Table 3).

Posterior mean genetic correlation estimated between milk FA and fat content were positive across lactation and had a similar pattern, as shown in Figure 3.c. The estimates increased until the middle of lactation reaching its maximum, and they followed slightly stable until the end of lactation. Genetic correlations estimated between fat content and C16:0, and SFA were similar and high, ranging from 0.89 to 0.98 across lactation. On the other hand, lower estimates were found between fat content and protein content (range from 0.55 to 0.72), and C18:1 *cis*-9 (range from 0.44 to 0.66). Moderate genetic correlations were found between fat content and LCFA (range from 0.69 to 0.87), and between fat content and UFA (range from 0.62 to 0.82). Overall, the average daily genetic correlations had moderate to high magnitude (range from 0.61 to 0.98), as seen in Table 2.

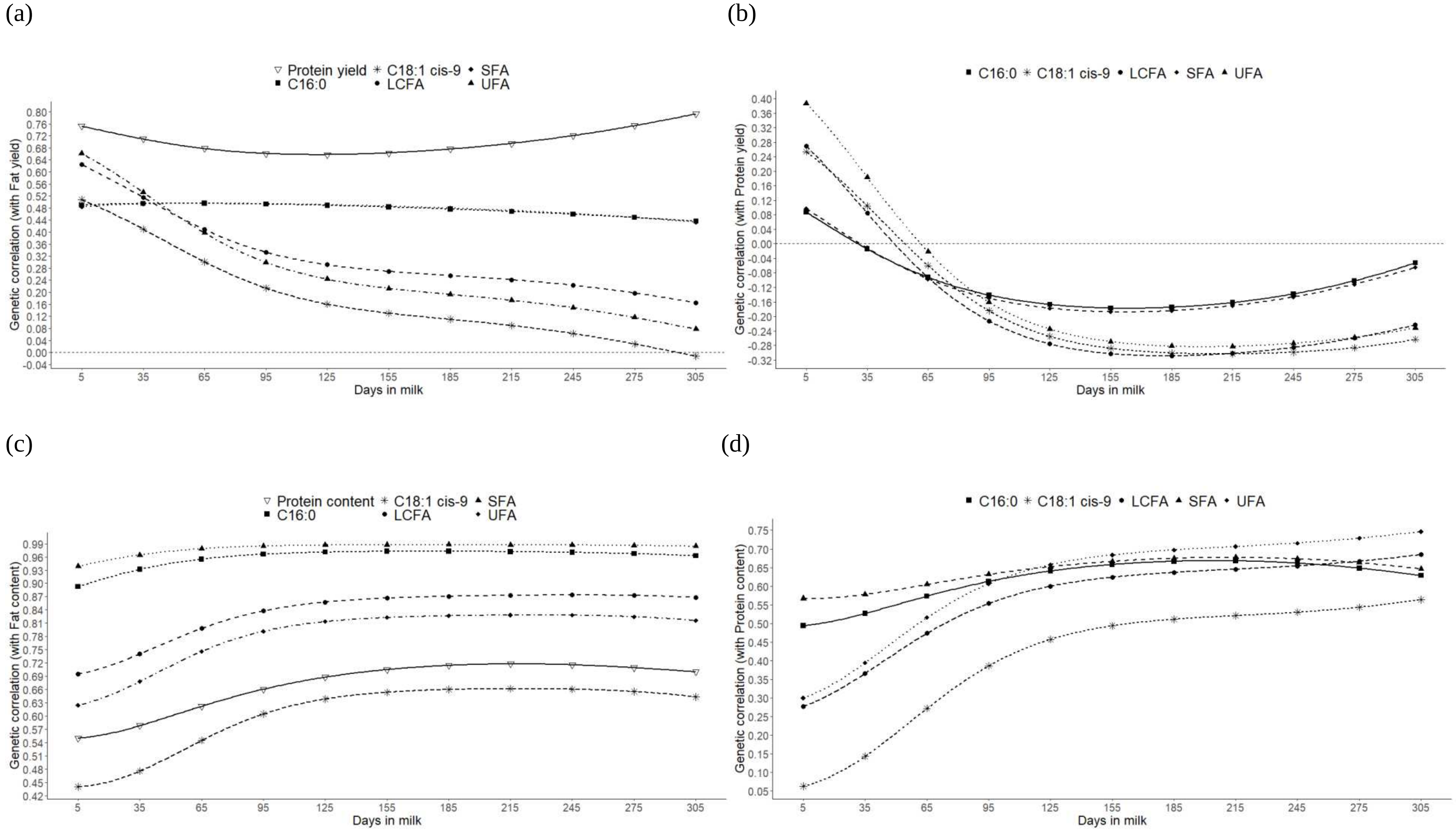


Figure 3. Posterior means of genetic correlations over days in milk estimated between milk production and FA traits and (a) fat yield, (b) protein yield, (c) fat content, and (d) protein content.

Posterior mean genetic correlations estimated between protein content and milk FA are shown in Figure 3.d. Estimates were positive throughout lactation (range from 0.06 to 0.74) and showed the same trend of low values in early lactation, increasing thereafter. From the middle to the end of lactation, estimated genetic correlations between protein content and C16:0, LCFA, SFA, and UFA were moderate to high and slightly close over days in milk, ranging from 0.56 to 0.74 (Figure 3.d). Genetic correlations between protein content and C18:1 *cis*-9 tended to be low since the beginning of lactation, as they ranged from 0.06 to 0.56. Average genetic correlations between protein content and milk FA were moderate, ranging from 0.41 to 0.64, as shown in Table 2.

Posterior mean genetic correlations among all FA traits were positive over days in milk, with higher magnitude estimates observed from the middle (around 125DIM) to the end of lactation, as shown in Figure 4a-e. The highest genetic correlations for FA were found between C16:0 and SFA, C18:1 *cis*-9 and LCFA, C18:1 *cis*-9 and UFA, and between LCFA and UFA, which tended to be slightly stable along the lactation curve (they ranged from 0.91 to 0.98; Table 2). Genetic correlations among the other FA had similar patterns, with low values in early lactation that increased as lactation progressed up to the middle (Figure 4a-e). Genetic correlations estimated between C18:1 *cis*-9 and C16:0, and between C18:1 *cis*-9 and SFA were similar, ranging from 0.07 to 0.53 (Figure 4b). Moreover, similar genetic correlations across lactation were found between C16:0 and LCFA (ranged from 0.35 to 0.76), and between C16:0 and UFA (ranged from 0.25 to 0.70) (Figure 4). Likewise, the genetic correlation estimated between SFA and LCFA, and between SFA and UFA ranged from 0.41 to 0.78 and from 0.31 to 0.72 (Figure 4c-e), respectively.

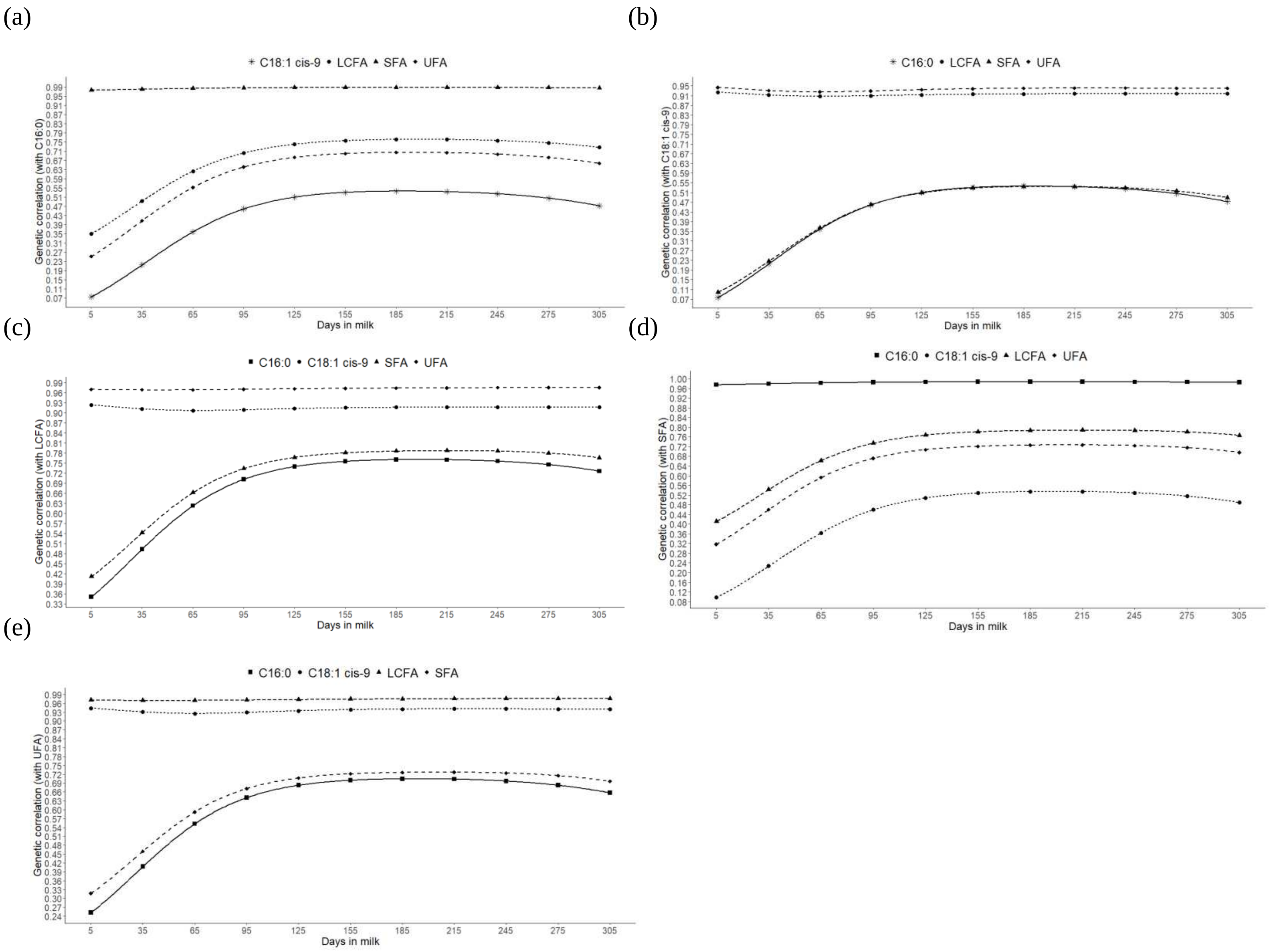


Figure 4. Posterior means of genetic correlations over days in milk among different milk FA traits. Correlations were estimated between (a) C16:0, (b) C18:1 *cis*-9, (c) LCFA, (d) SFA, and (e) UFA; and the other FA.

3.4.2. Prediction reliability and bias

Average validation reliability and bias (with its respective standard deviation) of EBV and GEBV estimated for milk production and FA traits are shown in Tables 3 and 4, respectively. Overall, the use of ssGBLUP increase the reliabilities compared to the traditional BLUP (i.e., parent average) for the validation bulls even when no scaling and weighting factors (default; $\omega = 1$ and $\alpha = 0.95$) were used for milk production traits. However, the highest validation reliabilities were obtained using scaling and weighting factors $\omega = 0.7$ and $\alpha = 0.6$ (Table 3) to combine $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$ in the genomic evaluations for milk production traits. Average reliabilities were especially higher for the milk yield, fat and protein contents (r^2; 0.38, 0.19, and 0.18, respectively), while for fat and protein yield they tended to be lower (r^2; 0.14 and 0.09, respectively). In addition, the use of optimal scaling factors yielded the least biased prediction for milk production traits, with average b_1 coefficients ranged from 0.76 to 0.92 (Table 3).

For milk FA, the ssGBLUP models achieved higher average validation reliability than parent average (Table 4). Similar validation reliabilities were estimated using different scaling and weighting factors for all milk FA (Table 4). The largest gain in genomic reliability was observed for the C16:0 and SFA (r^2; 0.11 and 0.17, respectively), as well as in all scenarios considered (Table 4). Validation average reliability was low for C18:1 cis-9 (0.04), LCFA (0.05) and UFA (0.05). However, a greater effect of scaling and weighting factors was observed in the regression coefficients for milk FA (Table 4). The inclusion of optimal values of $\omega = 0.6$ and $\alpha = 0.6$ for genomic evaluation of milk FA greatly decreased bias (i.e., average b_1 coefficients ranged from 0.32 to 0.81). Even though an improvement was observed, the GEBV predicted for most milk FA traits were still inflated, when using the optimal scaling and weighting factors (Table 4).

Table 3. Mean validation reliabilities ($\mathbf{r}^2$) and regression coefficients ($\hat{\mathbf{b}}_1$) for parentage average (PA) and genomic breeding value estimated assuming different* scaling (ω) and weighting (α) factors for milk production traits for the Walloon Holstein bulls.

Scenario	**Milk yield**		**Fat yield**		**Protein yield**		**Fat content**		**Protein content**	
	$\mathbf{r}^2$	$\hat{\mathbf{b}}_1$	$\mathbf{r}^2$	$\hat{\mathbf{b}}_1$	$\mathbf{r}^2$	$\hat{\mathbf{b}}_1$	$\mathbf{r}^2$	$\hat{\mathbf{b}}_1$	$\mathbf{r}^2$	$\hat{\mathbf{b}}_1$
PA	0.35 (0.04)	0.91 (0.15)	0.12 (0.03)	0.70 (0.13)	0.07 (0.03)	0.71 (0.15)	0.12 (0.02)	0.71 (0.08)	0.10 (0.04)	0.73 (0.19)
ω1α0.95[1]	0.36 (0.06)	0.91 (0.12)	0.08 (0.03)	0.47 (0.10)	0.08 (0.03)	0.59 (0.11)	0.16 (0.09)	0.65 (0.21)	0.16 (0.09)	0.60 (0.18)
ω0.7α0.6[2]	0.38 (0.04)	0.92 (0.15)	0.14 (0.04)	0.78 (0.16)	0.09 (0.04)	0.77 (0.15)	0.19 (0.09)	0.86 (0.38)	0.18 (0.10)	0.76 (0.24)

[*] τ scaling factor used for genomic relationship matrix (1.0$\mathbf{G}^{-1}$); ω = scaling factor used for pedigree relationship matrix (ω $\mathbf{A}_{22}^{-1}$); α and β = weighting factors used for polygenic effect (ssGBLUP - β$\mathbf{A}_{22}$).

[1] ω = 1 and α = 0.95; default values.

[2] ω = 0.70 and α = 0.60; the optimal values assumed for scaling and weighting factors proposed by Colinet et al. (2017).

Table 4. Mean validation reliabilities ($\mathbf{r}^2$) and regression coefficients ($\hat{\mathbf{b}}_1$) for parentage average (PA) and genomic breeding value estimated assuming different* scaling (ω) and weighting (α) factors for milk FA traits for the Walloon Holstein bulls.

Scenario	**C16:0**		**C18:1 cis-9**		**LCFA**		**SFA**		**UFA**	
	$\mathbf{r}^2$	$\hat{\mathbf{b}}_1$	$\mathbf{r}^2$	$\hat{\mathbf{b}}_1$	$\mathbf{r}^2$	$\hat{\mathbf{b}}_1$	$\mathbf{r}^2$	$\hat{\mathbf{b}}_1$	$\mathbf{r}^2$	$\hat{\mathbf{b}}_1$
PA	0.06 (0.02)	0.54 (0.14)	0.03 (0.02)	0.35 (0.23)	0.03 (0.02)	0.43 (0.25)	0.10 (0.02)	0.69 (0.08)	0.04 (0.03)	0.42 (0.25)
ω1α0.95	0.10 (0.07)	0.34 (0.21)	0.04 (0.03)	0.13 (0.09)	0.04 (0.03)	0.19 (0.11)	0.16 (0.09)	0.46 (0.21)	0.05 (0.03)	0.28 (0.11)
ω1α0.8	0.10 (0.08)	0.39 (0.23)	0.03 (0.02)	0.17 (0.10)	0.04 (0.03)	0.17 (0.09)	0.16 (0.09)	0.51 (0.21)	0.04 (0.03)	0.23 (0.12)
ω1α0.7	0.10 (0.07)	0.42 (0.23)	0.03 (0.02)	0.19 (0.11)	0.04 (0.03)	0.19 (0.10)	0.16 (0.09)	0.54 (0.21)	0.04 (0.03)	0.25 (0.13)
ω1α0.6	0.10 (0.07)	0.45 (0.24)	0.03 (0.02)	0.21 (0.12)	0.04 (0.03)	0.22 (0.11)	0.16 (0.08)	0.58 (0.21)	0.05 (0.03)	0.28 (0.14)
ω0.9α0.95	0.10 (0.08)	0.37 (0.22)	0.03 (0.02)	0.14 (0.10)	0.04 (0.03)	0.16 (0.09)	0.16 (0.10)	0.51 (0.23)	0.04 (0.03)	0.20 (0.11)
ω0.8α0.95	0.10 (0.08)	0.40 (0.24)	0.03 (0.02)	0.15 (0.10)	0.04 (0.03)	0.17 (0.09)	0.17 (0.10)	0.56 (0.24)	0.04 (0.03)	0.22 (0.12)
ω0.7α0.95	0.11 (0.09)	0.43 (0.25)	0.03 (0.02)	0.16 (0.11)	0.04 (0.03)	0.19 (0.10)	0.17 (0.10)	0.60 (0.26)	0.04 (0.03)	0.23 (0.12)
ω0.6α0.95	0.11 (0.09)	0.47 (0.26)	0.03 (0.02)	0.17 (0.11)	0.04 (0.03)	0.20 (0.10)	0.17 (0.10)	0.65 (0.27)	0.04 (0.03)	0.25 (0.13)
ω0.6α0.6	0.11 (0.07)	0.63 (0.28)	0.04 (0.03)	0.32 (0.16)	0.05 (0.03)	0.34 (0.15)	0.17 (0.08)	0.81 (0.25)	0.05 (003)	0.41 (0.18)
ω0.7α0.6	0.11 (0.07)	0.60 (0.28)	0.04 (0.03)	0.30 (0.14)	0.05 (0.03)	0.32 (0.14)	0.17 (0.08)	0.77 (0.24)	0.05 (0.03)	0.38 (0.17)

$^*\tau$ = scaling factor used for genomic relationship matrix ($1.0\mathbf{G}^{-1}$); ω = scaling factor used for pedigree relationship matrix ($\omega\,\mathbf{A}_{22}^{-1}$); α and β = weighting factors used for polygenic effect (ssGBLUP – $\beta\mathbf{A}_{22}$)

3.5. Discussion

3.5.1. Genetic parameters

According to the heritabilities estimated over days in milk for milk production and FA traits, genetic effects increased until the middle of lactation and trended to follow slightly stable thereafter (Figure 1). Heritability estimates were slightly higher using genomic information than pedigree-based analyses throughout lactation, using the same database and third-order Legendre polynomials (results not shown) based on the study by Paiva et al. (2021). Thus, our results shown that ssGBLUP method allows the most accurate estimates of genetic variance, as reported by Veerkamp et al. (2011). In addition, for selective genotyping or when the genotyped population is not representative of many previous generations, the ssGBLUP produces the most accurate estimates compared to pedigree-based method (Cesarani et al., 2018). According to Hidalgo et al. (2020), in populations undergone genomic selection, variance components estimated using only pedigree and phenotypes are biased due to preselection.

The heritabilities estimated in this study ranged from low to high. Our results reinforce an opportunity to improve milk production and FA traits over days in milk through genetic selection. These results agreed with previous studies available in the literature for dairy cattle (Bastin et al., 2011; Narayana et al., 2017). In this context, as milk FA are being also under genetic control (h^2 from 0.09 to 0.57 throughout lactation), genetic selection can be used to improve milk quality derived from MIR data. Similar results across lactations were found by Freitas et al. (2020), who reported a similar average heritability for SFA (h^2 0.47), and with a more upward curve and values lower for LCFA (h^2 0.25) and UFA (h^2 0.24). Lower estimates were reported by Petrini et al. (2016), 0.26 for C16:0, 0.07 for C18:1 *cis*-9, 0.25 for SFA, and 0.08 for UFA, using a repeatability model.

Most milk market employs a component pricing system to producer payment based on milk solids, especially fat and protein contents, which are associated with dairy products processing and directly impact on farm income. In addition, these major conventional components have been used for a long time to monitor the status and health of the cows (Gengler et al., 2016; Mäntysaari et al., 2019). One century of selection has mostly focused on increasing milk production (Miglior et al., 2017), however, an antagonist association across lactation (r_g from -0.22 to -0.59) was found between milk yield and, fat and protein content. Thus, herds selected to increase milk volume can disfavor simultaneous milk contents genetic gain although will increase fat and protein yield. Likewise, Hammami et al. (2008) and Petrini et al. (2016) reported similar genetic associations among these traits. Hence, challenges come and are likely to be overcome by cautious evaluation of selection criteria chosen in the dairy breeding program. Decisions should be made in order to achieve the main selection objectives, particularly, based on the payment system for the farmer, and it could change according to each country.

In general, daily genetic correlations between milk yield and FA were negative (they ranged from -0.38 to -0.49) which indicates that selection for improved milk yield would affect to a greater or lesser extent, all FA traits in milk throughout lactation. Although in early lactation the genetic correlations were positive and weak for LCFA, C18:1 *cis*-9, and UFA, the selection for higher milk yield would decrease FA in milk as soon as progressed lactation, showing a strong influence mostly on C18:1 *cis*-9 which the correlation becomes negative after 35 DIM. As already been reported by Bastin et al. (2013), a variation of milk FA profile is prominent in early lactation, and genetic correlations with milk yield are weaker than in late lactation. The general pattern of the estimates from this study was in line with the results of approximate genetic correlations reported by Bastin et al. (2011, 2013) and Fleming et al. (2018).

Regards the results reported in this study, heritabilities estimated for C16:0 and SFA were similar to that observed for fat and protein content over days in milk. This could be explained by the highly positive genetic correlations existing between these FA and fat content (average above 0.96) and also due to positive association with protein content (average above 0.62). Similar estimates were reported by Penasa et al. (2015) through multitrait analyses, as they reported genetic correlations of 0.99 (SFA and fat content) and 0.60 (SFA and protein content) in Holstein cows. The strong genetic correlation between fat content and C16:0 could be explained because of its important role in the synthesis of triacylglycerol in the mammary gland. Bovine milk fat is mainly composed of triacylglycerols (95%) and the most abundant fatty acid is C16:0. The increase in fat content has led to higher amounts of SFA and especially of C16:0, which affects the triacylglycerol structure and consequently, has an effect on the thermal properties of milk fat (Tzompa-Sosa et al., 2014). Tzompa-Sosa et al. (2014) reported that a modification in the triacylglycerol structure suggests that the distribution of FA could be triggered by differences in blood-derived fatty acids or by changes in the activity of enzymes related to the fat synthesis that respond to the availability of FA for triacylglycerol synthesis. In addition, those authors suggested that an increase in the availability of C16:0 for lipid synthesis in mammary epithelial cells will increase the activity of glycerol-3-phospate acyltransferase in the mammary gland.

Overall, the selection for higher fat yield and fat content would increase all underlying fatty acids in milk throughout lactation (r_g from 0.17 to 0.98). Likewise, there were positive correlations between milk protein content and FA, ranging from lower at the beginning (r_g from 0.06 to 0.57) to higher magnitude at the end of lactation (r_g from 0.56 to 0.68). Proteins are synthesized in mammary tissue from amino acids and are associated to TAG synthesis pathway such as DGAT1 which is one of many proteins composing this pathway (Bionaz and Loor, 2008). On the other hand, genetic correlation between protein yield and FA were negative (r_g from -0.11 to -0.19). Similar curves across lactation for milk FA with protein yield, and content also were showed by Fleming et al. (2018) using fifth-order Legendre polynomial. Nevertheless, they found weaker estimates and close to zero for the association between protein content and LCFA, and UFA in early lactation, as well as lower values in late lactation. However, differences among the genetic correlation estimated in this study from literature could be due to data set size (number of cows and phenotypes), different statistical models used, measurement methods of fatty acids, calibration equations, but also estimation methodologies.

There was a strong genetic correlation estimated between milk FA from the middle to the end of lactation, which suggest that selection for one-time point will likely results in genetic gains for all lactation stages. Besides, individual FA were strongly associated with the group to which they belonged, as seen by average genetic correlation between C16:0 and SFA (0.98), and between C18:1 *cis*-9 and LCFA (0.91), and UFA (0.93). Likewise, Bastin et al. (2013) presented similar average approximate daily estimates among the FA content across days in milk and three parities. C18:1 *cis*-9 is the single UFA with highest concentration in milk accounting for about 8g/liter whole milk (Haug et al., 2007), which could explain the high association found. Genetic correlations between C18:1 *cis*-9 and C16:0, and SFA had about the same average estimates (i.e., 0.44 and 0.45, respectively), and therefore, selection in favor of C18:1 *cis*-9 are likely to yield similar results on these FA. It seems that the genetic mechanism that was driving de novo (i.e., half of C16:0) FA synthesis was also driving the FA originating from other syntheses. Previous results were expected as well as estimated high genetic correlation between C16:0 and SFA (0.98) because the palmitic acid is a 16-carbon SFA and the most common found in animals (Loften et al., 2014). Furthermore, similar genetic gains would have expected for LCFA and UFA about selection on C18:1 *cis*-9 (r_g 0.91 and 0.93, respectively).

The lowest genetic correlations were estimated between C18:1 *cis*-9 and C16:0, and between C18:1 *cis*-9 and SFA at beginning of lactation. These estimates can likely be explained by the different origins of FA and its dynamic pattern influenced by lactation stage, energy balance and dietary composition. Milk fat is the main component determining energy expenditure for milk production in dairy cows and most of ruminant adipose tissue is stored as triglycerides, which comprise mainly C16:0, C18:0, and C18:1 *cis*-9 (Chilliard et al., 2000; Gross et al., 2011). In early lactation, dairy cows mobilize their body reserves to deliver the energy requirement for milk synthesis and secretion in the mammary gland. Thus, the FA in milk originate from both the mammary gland uptake of preformed FA from circulating blood (around 60%) and de novo synthesis within the mammary gland (around 40%) (Chilliard et al., 2000). C18:1 *cis*-9 is the predominant FA in adipocytes and primarily released through lipolysis during negative energy balance (NEB). Plasma nonesterified fatty acids (NEFA) and triglycerides are utilized by the mammary gland for milk FA synthesis. When lipolysis is high, the FA composition of milk has a much higher proportion of C18:0 and C18:1 *cis*-9 (Barber et al., 1997). The high uptake of LCFA by mammary gland tissue inhibits de novo synthesis of FAs (e.g. SFAs, especially C16:0) through the inhibition of acetyl-coenzyme A carboxylase (Hanuš et al., 2018).

In early lactation, C16:0 is originated primarily from mobilized body fat, and thereafter the cow achieves a positive energy balance and most of them should be produced within the mammary cells from acetate (Loften et al., 2014). Furthermore, the lower genetic correlation found between C16:0 and LCFA in early lactation can be associated with this mobilization body fat. Especially LCFA are derived from plasma and incorporated into milk, which inhibits the de novo synthesis (part of C16:0) by the mammary gland. Selecting for C18:1 *cis*-9 would have a great increase of UFA and LCFA (genetic correlations above than 0.90) in milk. As seen, the increased mobilization of lipids is associated with higher NEFA levels which are particularly rich in LCFA, especially C18:1 *cis*-9 (Jorjong et al., 2014). Most high-producing dairy cows are confronted with a NEB in early lactation, resulting in increased mobilization of adipose FA and the incorporation of these FA in the milk (Jorjong et al., 2014; Loften et al., 2014). Jorjong et al. (2014) identified C18:1 *cis*-9 as potential to be useful for diagnosis and early warning of cows suffering from a severe NEB. In addition, Martin et al. (2015) found high amount of C18:1 *cis*-9 in milk in the group of cows experiencing late onset of luteal activity, which provide information about their potential reproductive performance.

Efforts to carry out practical improvements of these milk FA are usually driven by its several roles as possible biomarkers to early lactation metabolic disorders (Jorjong et al., 2014), from specifically nutritional effects to benefit consumers (Haug et al., 2007), and also technological properties influencing the phenomena such as oxidation and possible sensory changes (Hanuš et al., 2018). Overall, the FA profile has been seen as a potential factor to contribute to added value in the milk market. Given the genetic correlation estimates over days in milk found in this study, it is possible to make desirable changes to the FA through selection strategies in this dairy cattle population.

3.5.2. Prediction reliability and bias

Genomic predictions using optimal scaling and weighting factors in ssGBLUP approach led to greater validation reliability and lesser bias compared to the traditional BLUP for most milk production traits. These findings are in agreement with previous studies in dairy cattle (Lourenco et al., 2014; Koivula et al., 2015; Baba et al., 2017; Kang et al., 2018; Oliveira et al., 2019b). The ssGBLUP models using the optimal scaling and weighting factors improved the reliability by 0.03 for milk yield, 0.02 for fat and protein yield, 0.07 and 0.08, respectively, for fat and protein content, which represent an increase ranging from 8.57% to 80%. The use of optimal factor scaling to combine $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$ are required for a better model fit (increase r^2 and decrease b_1) and may better account for differences in genetic architecture of each trait analyzed (Oliveira et al., 2019b). According to Gao et al. (2012), the bias of prediction tended to decrease with increasing polygenic weights in the **G** matrix and could be explained by the fact that genetic markers are not able to explain the total genetic variance. Thus, the polygenic effect would account for the residual genetic variance, which is not accounted for by using only genetic markers (Guarini et al., 2018). As seen in our results, it becomes important the optimization of the weighting factor α used in the combination of the raw genomic, and the pedigree relationship matrix. Similar findings were reported by Colinet et al. (2018), in which those showed that the weight has a large influence on the partition of total genetic variance into variances explained by SNP effects and by the residual polygenic effects.

Overall, differences in validation reliabilities r^2 among different scenarios to build the **H** matrix were smaller than effects on regression coefficients b_1, which is consistent with previous studies (Guarini et al., 2018; Kang et al., 2018). According to Koivula et al. (2015), the degree of inflation in GEBV is affected by the method used in construction of the **H** matrix. Decreasing of ω led to an increase of the regression coefficient, and consequently, decrease the inflation, and according to Martini et al. (2018), it can be explained because decreasing ω tends to reduce the variance of the GEBV. Thus, the scaling factors could be chosen in order to achieve smaller bias (degree of inflation) and it was expected for young animals (Masuda et al., 2016). Furthermore, decreasing ω increases the importance of the pedigree information in the genomic prediction, and it is also dependent on the completeness of the pedigree. As reported by Misztal et al. (2017), the ssGBLUP evaluations are inflated when the pedigree is deep but incomplete. The best ω parameter assumed in this study was 0.70 (lower than 1.00) for milk production traits, which increases the importance of pedigree information on GEBV prediction. Likewise, Tsuruta et al. (2011) reported that smaller values for ω (0.70) could be used in order to reduce the inflation of US Holstein genomic evaluations for young bulls without affecting accuracy. Baba et al. (2017) found ω equal to 0.30 to provide reduced biases prediction and a little effect on reliability in Japanese Holstein cows. Therefore, seems to be a good alternative choosing a small weight to account for contributions from the inverse of the numerator relationship matrix for genotyped bulls, but it should be carefully chosen.

For all milk FA, the inclusion of genomic information based on ssGBLUP approach also improved the reliabilities for young bulls, in all scenarios evaluated. Among all tested scenarios, the combination of ω equal to 0.60 associated with a polygenic effect α equal to 0.60 yielded the least biased GEBVs predicted for milk FA. These ssGBLUP models improved the accuracy by 0.05 for C16:0, 0.01 for C18:1 cis-9, 0.02 for LCFA, 0.07 for SFA, and 0.01 for UFA, representing an increase from 25% to 70%. Similarly, gains in prediction by using ssGBLUP for milk FA profile predicted by MIR were also reported by Freitas et al. (2020) and Cesarani et al. (2019), respectively, in dairy cattle and dairy sheep. Predictions reliabilities very low (ranging from 0.005 to 0.19) for C16:0 and C18:1 cis-9 predicted by gas chromatography were presented by Gebreyesus et al. (2019), using GBLUP model in Chinese, Danish and Dutch Holstein cows. The effect of a small training population size and the lower heritability estimates may reflect lower prediction reliability of GEBV. According to Guarini et al. (2018), predictions for lowly heritable traits benefit greatly from genomic information, especially by using ssGBLUP approach. Improvement in genomic prediction may be achieved by increasing the numbers of both genotyped and phenotyped animals as well as using optimal scaling and weighting factors to maximize the observed accuracy of the GEBVs for milk FA in our population.

3.6. Conclusion

Changes in milk production and FA traits can be achieved using genomic selection over days in milk. Selection for higher milk yield would decrease fat and protein content, as well as all FA (C16:0, C18:1 *cis*-9, LCFA, SFA, and UFA). Improving the milk FA profile (specially based on C16:0 and SFA) seems to be an effective way to indirect selecting for fat yield and fat content. The ssGBLUP approach yielded higher reliabilities than the traditional BLUP for young bulls. Less biased GEBV by choosing optimal scaling factors in construction of the **H** matrix were found. Therefore, ssGBLUP based on RRM is feasible for genomic prediction of milk production and FA traits in Walloon Holstein cattle.

3.7. Acknowledgments

The authors acknowledge the support of the National Fund for Scientific Research (Brussels, Belgium) under the grants (Projects: GplusE and Indiku). The authors also gratefully acknowledge the financial and technical support provided by the Walloon Breeding Association (AWE, Ciney, Belgium). Computational resources have been provided by the Plateau de Calcul et Modélisation Informatique (CAMI) of the University of Liège, Gembloux Agro-Bio Tech and the Consortium des Equipements de Calcul Intensif (CECI) of the Federation Wallonia-Brussels (Brussels, Belgium), funded by the National Fund for Scientific Research (Brussels, Belgium) funded under grant 2.5020.11. The authors also acknowledge the CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior).

3.8. References

Aguilar, I., I. Misztal, D.L. Johnson, A. Legarra, S. Tsuruta, and T.J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci. 93:743–752. doi:10.3168/jds.2009-2730.

Aguilar, I., I. Misztal, S. Tsuruta, and A. Legarra. 2014. PREGSF90 – POSTGSF90 : Computational Tools for the Implementation of Single-step Genomic Selection and Genome-wide .... Proceedings, 10th World Congr. Genet. Appl. to Livest. Prod. PREGSF90 0–3. doi:10.13140/2.1.4801.5045.

Auvray, B., and N. Gengler. 2002. Feasibility of a Wallon test-day model and study of its potential as tool for selection and management. Interbull Bull. 29:123–127.

Baba, T., Y. Gotoh, S. Yamaguchi, S. Nakagawa, H. Abe, Y. Masuda, and T. Kawahara. 2017. Application of single-step genomic best linear unbiased prediction with a multiple-lactation random regression test-day model for Japanese Holsteins. Anim. Sci. J. 88:1226–1231. doi:10.1111/asj.12760.

Barber, M.C., R.A. Clegg, M.T. Travers, and R.G. Vernon. 1997. Lipid metabolism in the lactating mammary gland. Biochim. Biophys. Acta - Lipids Lipid Metab. 1347:101–126. doi:10.1016/S0005-2760(97)00079-9.

Bastin, C., N. Gengler, and H. Soyeurt. 2011. Phenotypic and genetic variability of production traits and milk fatty acid contents across days in milk for Walloon Holstein first-parity cows. J. Dairy Sci. 94:4152–4163. doi:10.3168/jds.2010-4108.

Bastin, C., H. Soyeurt, and N. Gengler. 2013. Genetic parameters of milk production traits and fatty acid contents in milk for Holstein cows in parity 1 - 3. J. Anim. Breed. Genet. 130:118–127. doi:10.1111/jbg.12010.

Bionaz, M., and J.J. Loor. 2008. Gene networks driving bovine milk fat synthesis during the lactation cycle. BMC Genomics 9:366. doi:10.1186/1471-2164-9-366.

Britt, J.H., R.A. Cushman, C.D. Dechow, H. Dobson, P. Humblot, M.F. Hutjens, G.A. Jones, P.S. Ruegg, I.M. Sheldon, and J.S. Stevenson. 2018. Invited review: Learning from the future—A vision for dairy farms and cows in 2067. J. Dairy Sci. 101:3722–3741. doi:10.3168/jds.2017-14025.

Cesarani, A., G. Gaspa, F. Correddu, M. Cellesi, C. Dimauro, and N.P.P. Macciotta. 2019. Genomic selection of milk fatty acid composition in Sarda dairy sheep: Effect of different phenotypes and relationship matrices on heritability and breeding value accuracy. J. Dairy Sci. 102:3189–3203. doi:10.3168/jds.2018-15333.

Cesarani, A., I. Pocrnic, N.P.P. Macciotta, B.O. Fragomeni, I. Misztal, and D.A.L. Lourenco. 2018. Bias in heritability estimates from genomic restricted maximum likelihood methods under different genotyping strategies. J. Anim. Breed. Genet. 1–11. doi:10.1111/jbg.12367.

Chilliard, Y., A. Ferlay, R.M. Mansbridge, and M. Doreau. 2000. Ruminant milk fat plasticity: nutritional control of saturated, polyunsaturated, trans and conjugated fatty acids. Ann. Zootech. 49:181–205. doi:10.1051/animres:2000117.

Christensen, O.F., and M.S. Lund. 2010. Genomic relationship matrix when some animals are not genotyped Genomic prediction models. Genet. Sel. Evol. 42:1–8.

Colinet, F.G., J. Vandenplas, S. Vanderick, H. Hammami, R.R. Mota, A. Gillon, X. Hubin, C. Bertozzi, and N. Gengler. 2018. Bayesian single-step genomic evaluations combining local and foreign information in Walloon Holsteins. Animal 12:898–905. doi:10.1017/S1751731117002324.

Croquet, C., P. Mayeres, A. Gillon, S. Vanderick, and N. Gengler. 2006. Inbreeding Depression for Global and Partial Economic Indexes, Production, Type, and Functional Traits. J. Dairy Sci. 89:2257–2267. doi:10.3168/jds.S0022-0302(06)72297-4.

Druet, T., F. Jaffrézic, D. Boichard, and V. Ducrocq. 2003. Modeling Lactation Curves and Estimation of Genetic Parameters for First Lactation Test-Day Records of French Holstein Cows. J. Dairy Sci. 86:2480–2490. doi:10.3168/jds.S0022-0302(03)73842-9.

Fleming, A., F.S. Schenkel, J. Chen, F. Malchiodi, V. Bonfatti, R.A. Ali, B. Mallard, M. Corredig, and F. Miglior. 2017. Prediction of milk fatty acid content with mid-infrared spectroscopy in Canadian dairy cattle using differently distributed model development sets. J. Dairy Sci. 100:5073–5081. doi:10.3168/jds.2016-12102.

Fleming, A., F.S. Schenkel, F. Malchiodi, R.A. Ali, B. Mallard, M. Sargolzaei, J. Jamrozik, J. Johnston, and F. Miglior. 2018. Genetic correlations of mid-infrared-predicted milk fatty acid groups with milk production traits. J. Dairy Sci. 101:4295–4306. doi:10.3168/jds.2017-14089.

Freitas, P.H.F., H.R. Oliveira, F.F. Silva, A. Fleming, F. Miglior, F.S. Schenkel, and L.F. Brito. 2020. Genomic analyses for predicted milk fatty acid composition throughout lactation in North American Holstein cattle. J. Dairy Sci. 103:6318–6331. doi:10.3168/jds.2019-17628.

Gao, H., O.F. Christensen, P. Madsen, U.S. Nielsen, Y. Zhang, M.S. Lund, and G. Su. 2012. Comparison on genomic predictions using three GBLUP methods and two single-step blending methods in the Nordic Holstein population. Genet. Sel. Evol. 44:8. doi:10.1186/1297-9686-44-8.

Gebreyesus, G., H. Bovenhuis, M.S. Lund, N.A. Poulsen, D. Sun, and B. Buitenhuis. 2019. Reliability of genomic prediction for milk fatty acid composition by using a multi-population reference and incorporating GWAS results. Genet. Sel. Evol. 51:16. doi:10.1186/s12711-019-0460-z.

Gengler, N., H. Soyeurt, F. Dehareng, C. Bastin, F. Colinet, H. Hammami, M.-L. Vanrobays, A. Lainé, S. Vanderick, C. Grelet, A. Vanlierde, E. Froidmont, and P. Dardenne. 2016. Capitalizing on fine milk composition for breeding and management of dairy cows. J. Dairy Sci. 99:4071–4079. doi:10.3168/jds.2015-10140.

Geweke, J. 1991. Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. In Bayesian Statistics (4nd ed.). Clarendon Press, Oxford, UK.

Gross, J., H.A. van Dorland, R.M. Bruckmaier, and F.J. Schwarz. 2011. Milk fatty acid profile related to energy balance in dairy cows. J. Dairy Res. 78:479–488. doi:10.1017/S0022029911000550.

Guarini, A.R., D.A.L. Lourenco, L.F. Brito, M. Sargolzaei, C.F. Baes, F. Miglior, I. Misztal, and F.S. Schenkel. 2018. Comparison of genomic predictions for lowly heritable traits using multi-step and single-step genomic best linear unbiased predictor in Holstein cattle. J. Dairy Sci. 101:8076–8086. doi:10.3168/jds.2017-14193.

Van Haelst, Y.N.T., A. Beeckman, A.T.M. Van Knegsel, and V. Fievez. 2008. Short Communication: Elevated Concentrations of Oleic Acid and Long-Chain Fatty Acids in Milk Fat of Multiparous Subclinical Ketotic Cows. J. Dairy Sci. 91:4683–4686. doi:10.3168/jds.2008-1375.

Hanuš, O., E. Samková, L. Křížová, L. Hasoňová, and R. Kala. 2018. Role of Fatty Acids in Milk Fat and the Influence of Selected Factors on Their Variability—A Review. Molecules 23:1636. doi:10.3390/molecules23071636.

Haug, A., A.T. Høstmark, and O.M. Harstad. 2007. Bovine milk in human nutrition – a review. Lipids Health Dis. 6:25. doi:10.1186/1476-511X-6-25.

Hidalgo, J., S. Tsuruta, D. Lourenco, Y. Masuda, Y. Huang, K.A. Gray, and I. Misztal. 2020. Changes in genetic parameters for fitness and growth traits in pigs under genomic selection. J. Anim. Sci. 98:1–12. doi:10.1093/jas/skaa032.

Jorjong, S., A.T.M. van Knegsel, J. Verwaeren, M.V. Lahoz, R.M. Bruckmaier, B. De Baets, B. Kemp, and V. Fievez. 2014. Milk fatty acids as possible biomarkers to early diagnose elevated concentrations of blood plasma nonesterified fatty acids in dairy cows. J. Dairy Sci. 97:7054–7064. doi:10.3168/jds.2014-8039.

Kang, H., C. Ning, L. Zhou, S. Zhang, Q. Yan, and J.-F. Liu. 2018. Short communication: Single-step genomic evaluation of milk production traits using multiple-trait random regression model in Chinese Holsteins. J. Dairy Sci. 101:11143–11149. doi:10.3168/jds.2018-15090.

Kang, H., L. Zhou, R. Mrode, Q. Zhang, and J.-F. Liu. 2017. Incorporating the single-step strategy into a random regression model to enhance genomic prediction of longitudinal traits. Heredity (Edinb). 119:459–467. doi:10.1038/hdy.2016.91.

Koivula, M., I. Strandén, J. Pösö, G.P. Aamand, and E.A. Mäntysaari. 2015. Single-step genomic evaluation using multitrait random regression model and test-day data. J. Dairy Sci. 98:2775–2784. doi:10.3168/jds.2014-8975.

Legarra, A., O.F. Christensen, I. Aguilar, and I. Misztal. 2014. Single Step, a general approach for genomic selection. Livest. Sci. 166:54–65. doi:10.1016/j.livsci.2014.04.029.

Loften, J.R., J.G. Linn, J.K. Drackley, T.C. Jenkins, C.G. Soderholm, and A.F. Kertz. 2014. Invited review: Palmitic and stearic acid metabolism in lactating dairy cows. J. Dairy Sci. 97:4661–4674. doi:10.3168/jds.2014-7919.

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

Lourenco, D.A.L., B.O. Fragomeni, H.L. Bradford, I.R. Menezes, J.B.S. Ferraz, I. Aguilar, S. Tsuruta, and I. Misztal. 2017. Implications of SNP weighting on single-step genomic predictions for different reference population sizes. J. Anim. Breed. Genet. 134:463–471. doi:10.1111/jbg.12288.

Lourenco, D.A.L., I. Misztal, S. Tsuruta, I. Aguilar, E. Ezra, M. Ron, A. Shirak, and J.I. Weller. 2014. Methods for genomic evaluation of a relatively small genotyped dairy population and effect of genotyped cow information in multiparity analyses. J. Dairy Sci. 97:1742–1752. doi:10.3168/jds.2013-6916.

Mäntysaari, E., Z. Liu, and P.M. VanRaden. 2010. Interbull Validation Test for Genomic Evaluations. Interbull Bull. 17–22.

Mäntysaari, P., E.A. Mäntysaari, T. Kokkonen, T. Mehtiö, S. Kajava, C. Grelet, P. Lidauer, and M.H. Lidauer. 2019. Body and milk traits as indicators of dairy cow energy status in early lactation. J. Dairy Sci. 102:7904–7916. doi:10.3168/jds.2018-15792.

Martin, A.D., N.K. Afseth, A. Kohler, Å. Randby, M. Eknæs, A. Waldmann, G. Dørum, I. Måge, and O. Reksen. 2015. The relationship between fatty acid profiles in milk identified by Fourier transform infrared spectroscopy and onset of luteal activity in Norwegian dairy cattle. J. Dairy Sci. 98:5374–5384. doi:10.3168/jds.2015-9343.

Martini, J.W.R., M.F. Schrauf, C.A. Garcia-Baccino, E.C.G. Pimentel, S. Munilla, A. Rogberg-Muñoz, R.J.C. Cantet, C. Reimer, N. Gao, V. Wimmer, and H. Simianer. 2018. The effect of the H−1 scaling factors τ and ω on the structure of H in the single-step procedure. Genet. Sel. Evol. 50:16. doi:10.1186/s12711-018-0386-x.

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

Miglior, F., A. Fleming, F. Malchiodi, L.F. Brito, P. Martin, and C.F. Baes. 2017. A 100-Year Review: Identification and genetic selection of economically important traits in dairy cattle. J. Dairy Sci. 100:10251–10271. doi:10.3168/jds.2017-12968.

Misztal, I., H.L. Bradford, D.A.L. Lourenco, S. Tsuruta, Y. Masuda, and A. Legarra. 2017. Studies on inflation of GEBV in single-step GBLUP for type. Interbull Bull. 0.

Misztal, I., A. Legarra, and I. Aguilar. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J. Dairy Sci. 92:4648–4655. doi:10.3168/jds.2009-2064.

Misztal, I., S. Tsuruta, I. Aguilar, A. Legarra, P.M. VanRaden, and T.J. Lawlor. 2013. Methods to approximate reliabilities in single-step genomic evaluation. J. Dairy Sci. 96:647–654. doi:10.3168/jds.2012-5656.

Misztal, I., S. Tsuruta, D.A.L. Lourenco, Y. Masuda, I. Aguilar, A. Legarra, and Z. Vitezica. 2018. Manual for BLUPF90 family programs. University of Georgia. Page.

Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, and D.H. Lee. 2002. Blupf90 and Related Programs (Bgf90) 2001–2002.

Narayana, S.G., F.S. Schenkel, A. Fleming, A. Koeck, F. Malchiodi, J. Jamrozik, J. Johnston, M. Sargolzaei, and F. Miglior. 2017. Genetic analysis of groups of mid-infrared predicted fatty acids in milk. J. Dairy Sci. 100:4731–4744. doi:10.3168/jds.2016-12244.

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

Oliveira, H.R., D.A.L. Lourenco, Y. Masuda, I. Misztal, S. Tsuruta, J. Jamrozik, L.F. Brito, F.F. Silva, and F.S. Schenkel. 2019b. Application of single-step genomic evaluation using multiple-trait random regression test-day models in dairy cattle. J. Dairy Sci. 102:2365–2377. doi:10.3168/jds.2018-15466.

Penasa, M., F. Tiezzi, P. Gottardo, M. Cassandro, and M. De Marchi. 2015. Genetics of milk fatty acid groups predicted during routine data recording in Holstein dairy cattle. Livest. Sci. 173:9–13. doi:10.1016/j.livsci.2014.12.014.

Petrini, J., L.H.S. Iung, M.A.P. Rodriguez, M. Salvian, F. Pértille, G.A. Rovadoscki, L.D. Cassoli, L.L. Coutinho, P.F. Machado, G.R. Wiggans, and G.B. Mourão. 2016. Genetic parameters for milk fatty acids, milk yield and quality traits of a Holstein cattle population reared under tropical conditions. J. Anim. Breed. Genet. 133:384–395. doi:10.1111/jbg.12205.

R Development Core Team. (2019). *R:* A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.r-project.org.

Rodriguez, M.A.P., J. Petrini, E.M. Ferreira, L.R.M.B. Mourão, M. Salvian, L.D. Cassoli, A.V. Pires, P.F. Machado, and G.B. Mourão. 2014. Concordance analysis between estimation methods of milk fatty acid content. Food Chem. 156:170–175. doi:10.1016/j.foodchem.2014.01.092.

Smith, B.J. 2007. boa: An R Package for MCMC Output Convergence Assessment and Posterior Inference. J. Stat. Softw. 21:1–37.

Soyeurt, H., F. Dehareng, N. Gengler, S. McParland, E. Wall, D.P. Berry, M. Coffey, and P. Dardenne. 2011. Mid-infrared prediction of bovine milk fatty acids across multiple breeds, production systems, and countries. J. Dairy Sci. 94:1657–1667. doi:10.3168/jds.2010-3408.

Tsuruta, S., I. Misztal, I. Aguilar, and T.J. Lawlor. 2011. Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. J. Dairy Sci. 94:4198–4204. doi:10.3168/jds.2011-4256.

Tzompa-Sosa, D.A., G.A. van Aken, A.C.M. van Hooijdonk, and H.J.F. van Valenberg. 2014. Influence of C16:0 and long-chain saturated fatty acids on normal variation of bovine milk fat triacylglycerol structure. J. Dairy Sci. 97:4542–4551. doi:10.3168/jds.2014-7937.

VanRaden, P.M. 2008. Efficient Methods to Compute Genomic Predictions. J. Dairy Sci. 91:4414–4423. doi:10.3168/jds.2007-0980.

VanRaden, P.M., C.P. Van Tassell, G.R. Wiggans, T.S. Sonstegard, R.D. Schnabel, J.F. Taylor, and F.S. Schenkel. 2009. Invited Review: Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92:16–24. doi:10.3168/jds.2008-1514.

Veerkamp, R.F., H.A. Mulder, R. Thompson, and M.P.L. Calus. 2011. Genomic and pedigree-based genetic parameters for scarcely recorded traits when some animals are genotyped. J. Dairy Sci. 94:4189–4197. doi:10.3168/jds.2011-4223.

Vitezica, Z.G., I. Aguilar, I. Misztal, and A. Legarra. 2011. Bias in genomic predictions for populations under selection. Genet. Res. (Camb). 93:357–366. doi:10.1017/S001667231100022X.

Wang, H., I. Misztal, I. Aguilar, A. Legarra, R.L. Fernando, Z. Vitezica, R. Okimoto, T. Wing, R. Hawken, and W.M. Muir. 2014. Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. Front. Genet. 5:1–10. doi:10.3389/fgene.2014.00134.

Weller, J.I., E. Ezra, and M. Ron. 2017. Invited review: A perspective on the future of genomic selection in dairy cattle. J. Dairy Sci. 100:8633–8644. doi:10.3168/jds.2017-12879.

3.9. Supplementary material

Supplementary Table S1. Posterior mean heritability (on the diagonal) and genetic correlations estimates (below the diagonal) at 305DIM of milk production and FA traits in first lactation Walloon Holstein cows

	Milk yield	Fat yield	Protein yield	Fat content	Protein content	C16:0	C18:1 *cis*-9	LCFA	SFA	UFA
Milk yield	0.25									
Fat yield	0.64	0.23								
Protein yield	0.85	0.79	0.20							
Fat content	-0.47	0.36	-0.11	0.51						
Protein content	-0.55	0.02	-0.04	0.70	0.51					
C16:0	-0.37	0.43	-0.05	0.96	0.63	0.49				
C18:1 *cis*-9	-0.51	-0.02	-0.26	0.64	0.56	0.47	0.25			
LCFA	-0.55	0.16	-0.22	0.86	0.68	0.72	0.92	0.32		
SFA	-0.39	0.43	-0.06	0.98	0.64	0.98	0.49	0.76	0.53	
UFA	-0.58	0.07	-0.23	0.81	0.74	0.65	0.94	0.97	0.69	0.33

CHAPTER 4

4.1. General conclusions

The most important milk production and fatty acid (FA) traits in dairy cattle can be measured over time and random regression models (RRM) have been an optimal method to analyze longitudinal traits. RRM based on third-order Legendre orthogonal polynomials (i.e., 3 covariates) provides the best fit and should be used to describe the following random effects genetic additive, permanent environment, and herd-year of calving for milk production and FA traits of Walloon Holstein dairy cattle. In addition, using the optimal model will affect strongly the selection of bulls in early lactation and avoid generate inflated variances at the beginning. Genetic and genomic RRM enable to account for changes in genetic variation in production over days in milk. According to our results, the genetic selection may be done at any time point, with substantial gains especially from the middle to the end of lactation. Selection for higher milk yield would decrease FA over days in milk. The indirect selection for fat yield, fat, and protein contents would seem to be an effective way to improve milk FA profile. Genomic predictions using single-step genomic best linear unbiased prediction (ssGBLUP) based on RRM increased the validation reliability compared to the parent average method for young bulls. In addition, choosing optimal scaling and weighting factors in the construction of the **H** matrix yielded less biased prediction (regression coefficient close to 1). Thus, the ssGBLUP applied to RRM is feasible, and selecting young animals based on GEBV seems possible for milk production and FA traits in Holstein cattle. The findings from this thesis will contribute to advance on the modeling of milk production and FA traits in Walloon Holstein dairy cattle by the inclusion of genomic information. Our results contributed to a better understanding of genetic and genomic aspects of milk FA traits over the first lactation in dairy cattle. Moreover, this thesis provides the first results about the impact of different ssGBLUP methods based on RRM on genomic prediction for milk FA, which supports further studies.