GENOMIC EVALUATION IN DIFFERENT ENVIRONMENTS IN THE SOUTH AFRICAN HOLSTEIN BREED

By

Michiel van Niekerk

Submitted in fulfilment of the requirements for the degree

DOCTOR OF PHILOSOPHY

In the

Department of Animal Science

Faculty of Natural and Agricultural Sciences

University of Free State

Bloemfontein

Republic of South Africa

Supervisor: Prof. F.W.C. Neser

Co-Supervisor: Prof. V. Ducrocq

November 2024

TABLE OF CONTENTS

DECLARATION

I, Michiel van Niekerk, declare that this thesis submitted for the fulfilment for the degree: Doctor of Philosophy, Animal Science at the University of the Free State titled “GENOMIC EVALUATION OF PRODCUTION TRAITS IN SOUTH AFRICAN HOLSTEIN CATTLE UNDER TWO PRODUCTION SYSTEMS” is my own work. I furthermore declare that this thesis has not been submitted before for any other degree or examination at any other university. I also declare that all sources I have used or quoted have been indicated and acknowledged by complete references.

Signed ________________ 1 November 2024
Michiel van Niekerk Date
Student number: 1997256522

ACKNOWLEDGEMENTS

I would hereby like to acknowledge to following persons and/or institutions with regards to this study, in no particular order.

Firstly, I would like to thank my mother for her love and support, as well as her motivation, to put it into her words, "see me in that red robe at the graduation ceremony". Thank you for all the prayers and well wishes.

My friends, Elmien and Dawid Willemse for their constant motivation throughout all of these years. All the times we were with the cattle, or not with them, or just talking animal breeding over the phone. Thank you for your belief in my abilities as well as many wise words spoken and advice given. Thank you for your kindness and respect towards me.

My friend Henri Maritz for his love, dedication, motivation and making me laugh when I get too deep into my own head. Thank you for the good faith that exists between us, as well as the constant tea brought to my study, the kind-heartedness shown and value you see in me.

To all the other lecturers and admin personnel at the Department of Animal Science during my time as lecturer. Thank you for the help, kind advice and meaningful conversations, and the chance to learn from all of you.

To our farm manager, or rather friend of so many years. Thank you for all the help and keeping everything going during my many absent times. I know I, and so many others can always count on your steadfastness, kindness, trustworthiness and motivation to do the right thing.

My friends Christelle and Martin Barnard, as well as their two wonderful daughters, Chrismarie and Clarissa. Thank you for your support, advice, compassion and willingness to help without question in so many ways.

The University of the Free State and Department of Animal Science for helping to fund the research. A special thank you to my supervisor, Prof. Frikkie Neser for his wisdom, advice, incentive and believing in the study. Also, Prof. Japie van Wyk and Prof. Errol Cason for the valuable advice, support and wise inputs.

The Université Paris-Saclay, AgroParisTech and INRAE Génétique Animale et Biologie Intégrative Unit, Jouy-en-Josas, France. Thank you for helping to fund the study. To all the people I met in this team (unit) for valuable advice and conversations, support and kindness. Some people I worked a little closer with, or met there; Clementine, Roxanne, Alexis Michenet, Marie-Pierre Sanchez, Mecki Bosaha, Thierry Coudert, Manuel Revilla, Didier Boichard, Sophie Mattalia, Judikael Saout, Lidia Chavinskaia and many more. A special thank you to my supervisor Dr Vincent Ducrocq for always being available for invaluable advice, knowledge, helpfulness and support. It was a privilege to learn from him. Thank you for the vital and irreplaceable knowledge of animal breeding that you shared so willingly with me. Also, my gratitude for always trying to make me feel as welcome as possible in France.

My friend, Dr Alfredo Lepori. Thank you for all the advice and limitless conversations, especially during my time as lecturer. Your down to earth, humanity and kind spirit were not lost on me.

ABSTRACT

Fixed regression model (FRM) analyses that consider only fixed, non-genetic effects to vary over the lactation are currently used for genetic evaluation of production traits in South African Holstein. With random regression models (RRM), the random animal and permanent environmental effects are allowed to also vary over the lactation. Hence, RRM can account for an individual component representing changes during the lactation i.e., its persistency (PERS), enabling selection for more persistent cows. Also, test-day (TD) records used for genetic evaluations come from cows in contrasted production systems. The main ones rely on full pasture (PAST) or a total mixed ration (TMR), a choice often depending on local average rainfall where herds are situated. TD records from herds were divided into two datasets based on the production system (PAST or TMR). REML was used to analyse production for each of the first three lactations under different multiple-lactation models for milk, fat and protein production, as well as fat and protein percentage. Various FRM were compared to the current FRM officially used for genetic evaluation in South Africa (saFRM). A FRM that cumulates different curves over the lactation for different fixed effects was retained based on results in the PAST dataset and was also applied to the TMR dataset. This model was then broadened to an alternative RRM (aRRM) combining for each lactation an average production and a PERS effect, after which it was compared to the current saFRM under both production systems. The aRRM for both PAST and TMR had a better goodness of fit than the current saFRM for all traits except protein percentage. The mean squared error of aRRM was lower for all traits. Generally, aRRM heritability estimates were higher than with the saFRM at the beginning and end of lactation for most traits in PAST while being mostly higher during late lactation in TMR. Overall, the h^2 in PAST were mostly higher than in TMR for all traits. Estimates of between-lactations genetic correlations for average production from the aRRM were generally higher. Within-lactations genetic correlations between average production and PERS for TMR from the aRRM were negative and stronger than for PAST. The extra source of information from the aRRM enables a genetic prediction of PERS and is expected to increase accuracy of genetic predictions. Different genetic parameters between the two production systems may denote a genotype x environment interaction.

Dairy cattle in South Africa (SA) must produce in varying environments which can roughly be separated as total mixed ration (TMR) and pasture (PAST) production systems according to the level of rainfall in the herd area. Previous studies aimed to determine a possible genotype x environment interaction (GxE) using 305-day lactation records, finding no interaction, but different heritabilities. An alternative random regression model (aRRM) was developed using SA Holstein test-day records of herds with known production system used (TMR or PAST, analysed independently). The aRRM separates the animal genetic (additive) effect into an average-production (level) and -persistency (slope) effect and shows clear differences in the genetic variances and heritability estimates over the lactation between TMR and PAST. The aim of this study was to further investigate

the existence of a possible GxE interaction, using the same dataset, methods and aRRM to differentiate each trait (milk, fat and protein production for the first three3 lactations) separately for the two production systems (TMR or PAST). The possible reranking of sires on resulting estimated breeding values from the analysis between TMR and PAST were also investigated. Genetic correlations (0.81 to 0.94) of the genetic-average-production effect between TMR and PAST were strong, while corresponding correlations of the genetic-average-persistency effect were weak (Lactation 1, 0.22 to 0.27) to moderate (Lactations 2 and 3, 0.55 to 0.68) for all traits. Minor reranking of sires between TMR and PAST are predicted for genetic-average-production, but with moderate to major reranking for genetic-average-persistency. Analyzing TMR and PAST separately using a reduced rank model fits the data significantly better, adds additional information and allows for more accurate predictions to improve genetic persistency, more so than using within-PAST estimates. The recording of TD records as TMR or PAST should be seriously contemplated as the genetic component of persistency has a significant impact on total production.

Genetic improvement for persistency of milk production in South African Holstein cattle in a total mixed ration or pasture production system will be impeded because of the existence of genotype by environment interaction between the production systems. Previous studies where herds' production system was known (unlike reality) showed that rainfall level indicates the production system preferred by producers. The study aimed to determine the effect of including average rainfall in a random regression model as a possible proxy for production system used. The animal-additive-genetic effect due to annual rainfall (class) was added to the existing additive-genetic-average-production and -persistency effect. Heritability estimates of rainfall class 1 ('low' rainfall) coincided with total mixed ration estimates while rainfall class 5 ('high' rainfall) coincided with pasture estimates in later lactations. Results suggest that genetic background responsible for average production level and persistency over the lactations are not consistent over different rainfall areas.

Genomic evaluations for dairy cattle have become standard practice in 23 developed countries according to Interbull at the end of 2024. The development and use of many (tens to hundreds of thousands) single nucleotide markers (SNPs) spread over the whole genome opened new doors for improved genetic predictions. Together with the relatively newly developed single-step (ss) method that blends phenotypic, pedigree and genotypes (indicated by the SNPs of key animals) a new evaluation encompassing all this information in a single genetic analysis. Such a ss, genomic evaluation has been shown to enhance accuracy of prediction for various traits. SA Holsteins do not employ such genomic evaluations due to the relatively high cost of genotyping animals playing a significant role. Some 1221 SA Holstein animals (1143 cows) were, however, genotyped. The production system under which 535 of these cows were recorded for milk production was known. Because of the significant genotype x environment interaction between the two production systems, cumulative 295-days milk production for lactation 1 and 2 in TMR and PAST were analysed as different traits (i.e., four-trait analyses). Two-trait analyses (lactation 1 and 2, irrespective of production system used) were also carried out. Using REML,

two multi-trait (MT) i.e., four-trait analyses (MT4) were done. In the first analysis (MT4) genomic information was not implemented and yielded estimated breeding values (EBVs) and their accuracies (ACC). The second analysis was the same as the first (MT4) one, except that the single-step method (ssMT4) was used, which incorporated the genomic information and yielded genomically enhanced EBVs (ssGEBVs) and their ACCs. The two-trait analyses were carried out in the same way; a MT2 and ssMT2 analysis. Heritabilities between the ssMT4 vs the MT4 analyses were similar (TMR lactation 1 estimates of 0.23 vs 0.24, respectively; TMR lactation 2 estimates of 0.16, respectively; PAST lactation 1 estimates of 0.31, respectively; PAST lactation 2 estimates of 0.27, respectively). Heritabilities for the ssMT2 vs the MT2 analyses were also similar (0.25, respectively for lactation 1, and 0.17 vs 0.18, respectively for lactation 2). Genetic correlations from the ssMT4 between TMR and PAST for lactation 1 and 2 were 0.81 and 0.70, respectively. Hence, minor to moderate reranking of animals on ssGEBVs between the two production systems might take place for lactation 1 and 2, respectively. On average, the largest and significant increase in ssGEBV vs EBV ACCs were for cows with records (0.17; no progeny in the pedigree) and especially for the corresponding cows with the lowest EBV ACCs (increases of ~0.11 to ~0.40). Dams with records (and progeny in the pedigree) had on average the second largest increase in ssGEBVs (0.06), also with the largest increases in dams with the lowest EBV ACCs (increases of ~0.12 to ~0.33). Results for dams without records followed the same tendency. In this regard, results of the ssGEBV vs EBV ACCs for the ss- and MT2 analyses followed a similar trend. These preliminary results are promising for increasing ACC of genetic prediction using ss genomic evaluations. However, more animals need to be genotyped, especially highly influential sires with genetic merits that represent the whole population. Validation of genomic predictions need to be part of future research in the SA Holstein population.

CHAPTER 1

Introduction and objectives of the study

Increasing the accuracy of genetic predictions for traits of economic importance, included as selection criteria (directly or indirectly) to reach well set up breeding objectives is the mainstay of animal breeding. Accuracy of genetic predictions to identify animals that fit these goals as parents of the next generation are cardinal for favourable genetic change over time.

Since 1987 a sire model was used for the genetic evaluation of South (SA) African Holstein dairy cattle (Mostert et al., 2004). The major drawback of a sire model is that only sires are evaluated, which is based on the performance of their daughters. Secondly, the genetic merit of the daughters' dams are not taken into account and all dams are presumed equal in their genetic merit. This assumption is not accurate and the use of a sire model can cause bias in estimated breeding values (EBVs) if matings of cows and bulls are non-random. (Mrode, 2014). Therefore, where non-random mating is deemed necessary i.e., mating "best to best", or where it is the aim to mate animals that are more complimentary towards each other for an important trait(s), favourable genetic change will be inconsistent using a sire model. Such models were however computationally undemanding. As for the drawbacks of this model, and probably because of the advancements in computing resources, there was a move to animal models.

In an animal model EBVs for all animals in the pedigree, using the numerator relationship matrix, are predicted (Lynch and Walsh, 1998). The use of such a model allows for more accurate selection and mating strategies according to predetermined selection objectives as all animals have genetic predictions. Single-trait (first lactation, cumulative 305-days production) animal models were used for the genetic evaluations of SA Holsteins since 1992 (Mostert et al., 2004). Although single-trait animal models were a step forward from sire models, such analyses still have its shortcomings.

Since 1999, multi-trait (MT) animal models were used for SA Holstein genetic evaluations (Mostert et al., 2004). With such a model it is possible to take advantage of the fact that numerous traits are genetically correlated (Guo et al., 2014), because of some genes that are closely linked (close to each other) on the same chromosome (high linkage disequilibrium i.e., high LD). The farther genes are situated from each other on the same chromosome, the lower LD they have. More important is the effect of pleiotropy where the same genes affect two or more traits. Furthermore, most traits of economic importance in livestock are polygenic, and thus a great number of genes can affect the same trait(s). Analysing traits together (MT) take advantage of these genetic structures and mechanisms and are beneficial for predictions of all animals where only a relatively few are measured i.e., costly, or difficult to measure, or measured (expressed) in only one sex (sex influenced). Low heritable traits usually benefit the most from MT analyses (improved accuracy of predictions) because of genetic correlations (Guo et al., 2014; Bourdon, 2014; Mrode, 2014).

The use of cumulative milk (or fat or protein) yield, usually calculated for a standard 305-day lactation, for the first three parities were the traits analysed in past SA dairy cattle genetic evaluations before the year 2000 (Mostert et al., 2004). The use of test-day (TD) records, taken roughly 30 days apart were used to calculate cumulative 305-day yield. TD records stretch back to 1987 for SA Holsteins. Using this longitudinal (TD) data with ~ 8 to 10 TD records as different, correlated traits in a so-called TD model, for the first three lactations (MT) using all pedigrees (animal model) describes the model that was/is currently used for SA Holstien genetic evaluations. In this model fixed effects are considered either as having the same effect through the lactation, or described using some kind of mathematical curve function to regress their effects over DIM i.e., an animal TD, MT, and fixed regression model (FRM).

The body of this study is made up of four parts, each with the potential to increase rate of genetic change through more accurate and less biased genetic predictions. Put together the four parts build on one another to arrive at improved genetic (genomic) evaluations for the SA Holstein dairy cattle breed. It is written in the form of four articles (Chapters 3 to 6), some of which are already published or in the process of being so. For this reason, the general conclusions and recommendations in Chapter 7 can be seen as a summary of the four main parts. These four parts (objectives) of the study are as follows:

1. The first study (Chapter 3) aimed to improve on the current animal, TD, multi-lactation and fixed regression model that was used for genetic evaluations in the recorded SA Holstein population by developing a model that fit the data better and lead to an increase in genetic prediction accuracy. This was done separately for the two most common production systems used by dairy producers i.e., total mixed ration (TMR) and pasture (PAST) production systems.

2. The second study (Chapter 4) sought to build on the evidence of genotype by environment interaction (GxE) between the TMR and pasture PAST production systems shown in Chapter 3. Such an interaction may have a major impact on the rate of genetic gain through selection.

3. A reaction norm model (Chapter 5) where milk production was grouped in average rainfall classes and used as a proxy for PAST and TMR was considered, to try and accurately predict an animal-additive-genetic effect due to annual rainfall.

4. Lastly, available genomic information (Chapter 6) was introduced in a preliminary investigation into possible changes in rankings of animals, but especially possible changes in accuracy of genetic prediction.

CHAPTER 2

Literature review

2.1 Introduction

Diverse cattle from Holland with regard to type and colour has been introduced as early as during the 17th and 18th century by early settlers to the Cape Colony of South Africa. These imports of cattle were made relatively easy as these settlers still had close racial, political and family relations with Holland. Crossbreeding was still the norm in Europe during this period which ensured these imported cattle to vary in many ways, phenotypically and genetically. It is thought that probably during the early 19th century Holland cattle were started to being separated in different type and breeds in Holland itself. It was during this period that black and white cattle emerged in a way closer to what was (is) known as Frieslands. Reports indicate that Governor van Plettenberg imported what was probably the first cattle of this “more uniform” type in the form of a bull and cow sometime amid 1779 and 1784. These cattle were for the first time called by the term Fries, probably because of the province in Holland from where they came. Seemingly towards the end of the 18th and beginning of the 19th century, a reasonable number of Dutch and Dutch crosses could be found in the Cape and its adjacent districts. These Dutch cattle have earned a good reputation by the mid-19th century to a sufficient degree and imports rose quickly from then onwards. Most of these cattle were seen to be of the black and white Friesian type and their importers became the founders of the first Friesland studs. The Friesland Cattle Breeders’ Association of South Africa (SA) was established during an assembly of involved breeders in Bloemfontein on 21 October 1912. However, the first Frieslands were registered in 1906 in the South African Studbook (Gertenbach, n.d.). Early on, Frieslands were imported to South Africa from Germany, Great Britain, United States of America and Canada. Most of them were however brought in from Holland, specifically from the Friesland and North Holland provinces. (Swart and O’Hagan, n.d.). The Friesland Cattle Breeders’ Association of South Africa changed its name to the SA Holstein Friesland Society during 1990, and to the SA Holstein Cattle Breeders’ Society in 1998. The name change was to reflect on the advancement of the breed (Duvenhage, 2017).

2.2 Fixed regression

At the time of beginning this study the model used for SA Holstein genetic evaluations were a fixed regression model (FRM) analysing test-day (TD) records (Interbull, 2020). This model used the Wilmink (1987) function to describe one fixed (interaction) effect over the lactation, with other fixed effects incorporated as independent of the stage of lactation or time (constant).

In fixed regression models there can be fixed effects (non-genetic) that are regarded to have the same effects over the lactation, but importantly also one or more fixed effects that are non-linear (change) and have a certain curve(s) over the lactation. The lactation is described by the fixed effects over days in milk (DIM) after calving. These curvilinear shape(s) include regressions over DIM from a variety of possible mathematical functions. (Mrode, 2014). Cows that were treated the same over the lactation will have their TD records adjusted in the same manner. The correct mathematical function that gives the best fit of the fixed effect curves influence the accuracy of fixed estimates and consequently also genetic predictions. Cobuci and Costa (2012) reported reranking of sires for milk yield and persistency (of yield) based on the measures used to decide the fixed regression used for the lactation curve.

Silvestre et al. (2006) studied the accuracy of seven functions. These included earlier functions of Wood (1967; WOOD), Wilmink (1987; WIL), and Ali and Schaeffer (1987; ALI) that were originally fitted to describe (phenotypic) lactation curves. According to Silvestre et al. (2006), in later years researchers employed other means that included Legendre polynomials (Kirkpatrick et al., 1994, dairy cattle; Pool et al., 2000), and splines (White et al., 1999) for the mathematical expression of fixed curves. Silvestre et al. (2006) tested Legendre polynomials to the degree of 2, 3 and 4, as well as (cubic) splines. Meyer (2005) stated that after the use of Legendre polynomials by Kirkpatrick et al. (1990) in describing the mean growth trajectory in mice, the use of this function became widespread. Polynomials of a higher order can model fluctuations of means and variances over the length of a continuous variable satisfactory. Yet, these polynomials emphasize observations at the beginning and end of the trajectory, which can be problematic (Meyer, 2005). The is probably because of few records at the ends of the lactation as this is a cause of such Legendre polynomial artefacts (Misztal et al., 2000). Misztal (2006) stated many problems with Legendre polynomials, including sudden changes of the terms with a high order at the extremes, poor mathematical fitting of irregular functions, lack of data to estimate a very large number of parameters, and higher susceptibility of (co)variance curves to each of the numerous parameters.

In its simplest form, a spline function is made up of linear fragments. The points where these fragments join are called knots. A few of these fragments looks like and is called a “broken-stick” curve, and is a basic expansion of parametric, linear regression. These fragments can also be polynomial and are often cubic, like in the case of Silvestre et al. (2006). Their results showed that the accuracy of WOOD, WIL and ALI was influenced markedly by longer intervals (beyond 4 weeks) between TD intervals and when there was an increase in the first TD, beyond 2 months. When this was the case the use of splines was best. Splines also described the lactation curve effectively with fewer TD records compared to ALI and Legendre polynomials. Misztal (2006) cited many studies that was done at the University of Georgia on the use implementation of linear splines (Bohmanova et al., 2005; Iwaisaki et al., 2005; Robbins et al., 2005) for growth in beef. Here, linear splines based on three knots was less complex to configurate and had better numerical characteristics. Misztal (2006) specified that the use of linear splines can be particularly applicable for national genetic evaluations if the covariate functions (of the lactation curve in our case) are known. Importantly, then the correct number of knots as well as the correct position of knots can be established.

Regarding the recording of TD yields during the early lactation stage, Silvestre et al. (2006) showed that all seven mathematical functions tested achieved better results when the first TD record was before 8 or 30 DIM, compared to 60 and 90 days. Macciotta et al. (2005) stressed the importance of TD records before peak yield for accurate estimates of the lactation shape.

The number of fixed regressions will be influenced by the number of fixed effects, to what degree they change over the lactation as well as the trait itself. Cobuci and Costa (2012) showed that multiple regression lactation curves fitted the TD records for milk yield in first lactation best, but that only a single regression curve was the best fit for later lactations.

2.3 Random regression

A serious shortcoming of a FRM is that the random effects, like the animal-genetic (AG) and permanent environmental (PE) effects are assumed to be constant over the lactation (Mrode, 2014). Thus, genetically FRM assume milk production to exactly be the same trait over the lactation. Random regression models (RRM), apart from accommodating fixed regressions, allow for the AG and PE (random) effects to vary over the lactation (Mrode, 2014). For instance, Lund et al. (1999) and Heringstad et al. (2003) showed that gene expression for clinical mastitis differs substantially over the lactation. Veerkamp et al. (2001) proved the AG effect for body condition score (BCS) differed significantly over the lactation while Berry et al. (2003) reported the same for BCS, body weight and milk yield over DIM. This is important as it indicates that the lactation curve can be altered through selection as they are heritable (Oliveira et al. 2024). In addition, Zavadilova et al. (2005) and Li et al. (2020) showed that the PE effect was curvilinear over the lactation. Random effect curves are often described by the same mathematical expressions that is used for fixed effect curves (regressions) like polynomials and Legendre polynomials to the n^{th} degree as well as various types of splines. The popularity of Legendre polynomials were mentioned in the previous section (2.2). Important to note when they are used for the random effects (AG and PE) in random regression models are the possible imperfect emphasis of observations at the beginning and end of the trajectory (Meyer, 2005). Schaeffer and Jamrozik (2008) stated that the problem with

Legendre polynomials are instances of artefacts of unusually high genetic variances at the beginning and end of the lactation with subsequent high heritabilities at these points.

The RRM model was presented by Henderson (1982) and Laird and Ware (1982), but it was Schaeffer and Dekkers (1994) who introduced it to be used for TD records, as it is suited for use in longitudinal data analyses in many cases (Lewis and Brotherstone, 2002, growth in sheep; Akbaş et al., 2004, body weights in quail; Banos et al., 2005a, daily energy balance in dairy cows; Wolc et al., 2007, monthly egg production in laying hens; Wang et al., 2021, residual feed intake in pigs). Mark (2004) reported multi-trait RRM TD models to be the "best potential method" for genetic evaluations of dairy production traits. Liu et al. (1998) showed that FRM are inferior as their residual variance were higher than RRM models. Buttchereit et al. (2010) conferred that as expected, RRM models achieved better results than FRM.

2.4 Persistency of production

With the advent of TD random regression models important functional traits could be distinguished from the lactation curve i.e., peak production and persistency of production. These traits are heritable, meaning that the lactation curve can be altered through selection in a population. (Oliveira et al., 2024). Various definitions and modelling of persistency exist (Grossman et al., 1999). Early on Sanders et al. (1930) defined persistency of production as the rate of decline from peak yield. Gengler (1995) defined persistency of milk production as the "flatness of the lactation curve". Jamrozik et al. (1998) proposed using the average slope of an animal's TD yields between DIM 60 (around peak production) and 280. Other assessments of persistency have been proposed by Sölkner and Fuchs (1987), Jakobsen et al. (2002) and Cobuci et al. (2004). Cobuci et al. (2007) stated that for the most part persistency is evaluated as a "by-product" of random regression TD models. Using EBVs (EBV_t = EBV of DIM_t) from a RRM, these authors tested six measures of persistency (PERS):

$$PERS_1 = (EBV_{280} - EBV_{60}),$$

$$PERS_2 = \left(\sum_{t=106}^{205} EBV_t - \sum_{t=6}^{105} EBV_t \right),$$

$$PERS_3 = \left(\sum_{t=206}^{305} EBV_t - \sum_{t=6}^{105} EBV_t \right),$$

$$PERS_4 = \sum_{t=61}^{280} (EBV_t - EBV_{60}),$$

$$PERS_5 = \sum_{t=60}^{279} (EBV_t - EBV_{280}),$$

$$PERS_6 = (EBV_{290} - EBV_{90}),$$

where low values of $PERS_{1,2,3,4,6}$ indicate high persistency, whereas a high a value for $PERS_5$ indicate the same ($PERS_{1,4}$ from Jamrozik et al., 1997; $PERS_{2,3,5}$ from Jakobsen et al., 2002; $PERS_6$ from Cobuci et al., 2004). Cobuci et al. (2004) recommended $PERS_5$ and $PERS_6$ as their generic correlations with cumulative 305-day milk production were weak (-0.31 and 0.31). Gengler (1996) emphasized that the evaluation of persistency must be unrelated (unbiased) from yield or corrected for the influence of yield. The author concluded that persistency is reliant on yield, especially cumulative yield.

Persistency has a great potential economic impact on the efficiency of milk production, directly and indirectly. Macciotta et al. (2011) stated that because of (total) production increase in milk yield, fertility and health problems in dairy cattle also increased. The overall level of the production curve probably plays an important role. The yield at peak stage of lactation increases health difficulties (Bohmanova et al., 2009) and cows being in a negative energy balance during that stage reduce fertility traits (Friggens et al., 2007). This is supported by Caixeta and Omontese (2021) who specified that roughly a third of dairy cows having at minimum one clinical disease (metabolic and / or infectious), with more than half having no less than one subclinical disease case during the first 90 days of lactation. In a review, Kgari et al. (2020) indicated a decline in fertility traits of dairy cattle for multiple breeds over the world, because of strict selection for milk yield and content traits. Grossman et al. (1999) proposed selection for increased persistency without increasing peak yield and subsequent undesirable stress on the cow.

Extended lactations where there is an intended pause of insemination, may increase the profitability of cows in both conventional and pasture production systems as shown by Rotz et al. (2005) and Butler et al. (2010). If such cows are superior for production persistency this will be of economic advantage to the dairy enterprise. Sehested et al. (2019) argued that an extended lactation strategy will lead to a decrease in number of calvings per year, as well as a decrease in the number of health risk periods related to calving as shown by Ingvartsen et al. (2003). There may also be additional benefits because of reduced insemination, veterinary, and rearing of replacement heifer costs (Bach et al., 2008).

2.5 Genotype by environment interaction

One seemingly same trait that is measured in two different environments or conditions have the potential to be two different traits (Falconer and Latyszewski, 1952). When these environments or conditions are classified as categorical a multi-trait genetic analyses seems sufficient (Calus et al., 2002; Hayes et al., 2016). Such conditions include differential diets, production systems, milking type, production levels, and geographical region (Silva Neto et al., 2024). When these conditions vary enough so that certain genotypes that are significantly more adapted to better survive and (re)produce in a certain environment over other genotypes, the phenomenon of genotype by environment interaction (GxE) might be at play. If the same trait but defined as two in a multi-trait genetic analyses have a genetic correlation below 0.80, GxE between these conditions exist (Robertson, 1959) and will cause reranking of animals within the given population. South Africa is a relatively dry country, with an average annual rainfall of about 464 mm. While the Western Cape gets most of its rainfall in winter, the rest of the country is generally a summer-rainfall region (South African Government, 2024; https://www.gov.za/about-sa/geography-and-climate). According to the Climate Change Knowledge Portal (2024; https://climateknowledgeportal.worldbank.org/country/south-africa/climate-data-historical), the topography varies from desert to semi-desert in the drier northwestern region to sub-humid and wet along the country's eastern coast; approximately half of the country is classified as arid or semi-arid. With regards to the different rainfall regions, from low in the west to significantly higher in the south, southwest and east, production systems used by dairy producers follow the same pattern. A TMR production system is utilised in the drier areas where in the more wet areas a PAST production system is used. These very differing conditions have the potential to favour different genotypes, especially with regards to an exogeneous breed like the Holstein. Saltz et al. (2018) stated that the presence and magnitude of GxE vary dramatically across populations and traits. Bourdon (2014) stated that the extent of possible reranking of animals, especially sires depend on the species and trait. Also, reranking tends to be more in low heritable traits (fertility, survivability, and other functional traits) than in higher heritable traits. This is clear in a previous SA Holstein study where Neser et al. (2014) found a weak genetic correlation of 0.28 for age at first calving, while the corresponding correlation for cumulative (305-day) milk production was 0.90 between TMR and PAST. In later studies with the introduction of (TD) RRM, milk production traits could be disseminated in mainly two genetic parts i.e., production level of the lactation curve and persistency (Togashi and Lin, 2003). Coffey et al. (2001) discussed a close association between persistency and other energy balance traits. To argue that persistency is a fitness trait due to its association with energy balance (mobilization), health and fertility traits, might lead one to believe that a strong GxE magnitude for the persistency part of cumulative (production level plus persistency) yield is more probable. Also, GxE with regards to persistency between SA's two main productions systems have never been investigated.

2.6 Single-step genomic evaluation

As suggested by the acronym, an SNP (single nucleotide polymorphism) marker is just a single base change in a DNA sequence (Vignal et al., 2002). When an animal is genotyped, the specific base pairs that are present at specific points over the whole genome are read. The number of reading positions differ according to the chip used, but generally the most popular chip contains approximately 50 000 SNPs. Wiggans et al. (2009) stated that genotypes from the Illumina BovineSNP50 BeadChip form the foundation for genomic evaluations in the United States and Canada. VanRaden (2008) stated that genomic relationships occur because of alleles that are shared between individuals. These shared alleles as indicated by SNP markers are essential to form the important genomic relationship matrix (G) needed in single-step methodology and is important for the estimation of genomic selection accuracies, depending on how many animals are genotyped (Goddard, 2009). The shared alleles are spread throughout (over) independent chromosome (IC) segments. The role effective population size is of cardinal importance regarding IC segments. Effective population size (N_e) is the size of a hypothetical ideal population with the same level of inbreeding or gene frequency divergence under random genetic drift as the actual population under consideration (Wright, 1938). The smaller the N_e, the fewer the IC segments that must be estimated to get a higher accuracy of genomic selection for the same population size (Goddard, 2009). Misztal et al. (2020) stated that small N_e in farm animals enable genomic prediction. Chunks of DNA from excessively presented ancestors form relatively few IC segments are called linkage disequilibrium (LD) blocks. The authors explain that these segments are difficult to identify because they are broken at slightly different places across siblings (recombination), but they do appear indirectly (e.g., as singular G that needs to be blended to become full rank). (Misztal et al., 2020). Misztal et al. (2009) proposed to join the genomic relationships captured by SNP markers in the G matrix with the pedigree relationship matrix (A) into a new blended (H) matrix that will replace the A matrix.

If the genotyped animals are not genotyped at random, their base population is not the same as the base in the pedigree (Vitezica et al., 2011). This was the case when single-step (ss) methodology was implemented using best linear unbiased prediction (BLUP) using the H matrix, which caused inflation of (or biased in) the subsequent genomic breeding value (GEBV) predictions of genotyped animals (Forni et al., 2011). This causes the G and A (specifically the A_{22} matrix, which are the genotyped animals in the A matrix, needed to form the new blended matrix) matrices to be ill aligned. Aguilar et al. (2010) suggested a scaling factor to reduce the bias of the GEBV.

Christensen (2012) introduced two methods to solve the misaligned G and A_{22} matrix, the first being to adjust the G matrix by using a regression. The author's second method entailed adjusting the whole A matrix to the base of G. Contrariwise, Vitezica et al. (2011) proposed to alter the base of G to that of A_{22}, using the mean difference between A_{22} and G.

An important consideration in genetic evaluations is the modelling of missing pedigrees, which is just as important with single-step genomic BLUP (ssGBLUP). Missing parents of animals are replaced with groups that are determined by time (of birth of "phantom" parent using generation interval) together with genetic path of selection (e.g., sire of sire, sire of dam, dam of sire, dam of dam; Westell et al., 1988). These unknown parent groups (UPGs, also called genetic groups i.e., GGroups) have been fitted in classical animal models as covariates, or some form of classified effects (Westell et al., 1988). The necessity of UPGs is to account for selection that known relations cannot account for (Westell et al., 1988). The approach by Westell et al. (1988) was expanded by Quaas (1988) and alters the inverse of the covariance matrix for the random effects. Bermann et al. (2022) stated that this approach is computationally efficient and the method of choice in ssGBLUP. Misztal et al. (2013) did however point out that UPG solutions are biased when used as per Quaas (1988) in ssGBLUP. These problems are reported to be solved when UPGs are fitted as explicit (separated) covariates (Misztal et al., 2013; Bermann et al., 2022).

What has become to be known as metafounders can also be used to model missing pedigrees in ssGBLUP. Fitting metafounders have the extra advantage of accounting for the misalignment between genomic and pedigree information. (Bermann et al., 2022). The construct of metafounders were presented by Legarra et al. (2015) and were explained as being pseudo-individuals that relate (connect) founders in the pedigree and different populations.

Accuracies achieved and calculated from genetic predictions based on partial data relative to predictions (phenotypes) based on comprehensive data are used to validate genomic evaluations (Daetwyler et al., 2013; Legarra and Reverter, 2018; Misztal et al., 2020). According to Misztal et al. (2020) there are various methods for the validation of genomic predictions that are currently being employed, but that each one is appropriate for a distinct data structure. When population records are few and only one generation is genotyped, separating the population into ***n*** samples and predicting phenotypes of a sample from the ones left over is the k-fold cross validation (Saatchi et al., 2011). In dairy cattle where sires have relatively large numbers of progeny with large numbers of phenotypic records, VanRaden et al. (2009) suggested comparing 'pseudo-observations' of sires (daughter yield deviations or deregressed proofs) with their GEBVs or EBVs acquired without their daughter's data. Masuda et al. (2018) showed that this may be biased by preselection if 'pseudo-observations' are estimated by BLUP in the presence of genomic selection. When validation animals have own records and no progeny, validation is based on correlations between GEBVs obtained without a phenotypic record vs the same record that is adjusted for the fixed effects (Legarra et al., 2008). This method is only suitable for simple models and depends on how accurate the adjustments are calculated (Legarra and Reverter, 2018). Important and indicated by Bijma (2012) and Lourenco et al. (2015) is that accuracies built on validation are reduced by selection, thus are lower than individual theoretical accuracies based on prediction error variance (PEV). Misztal et al. (2020) propose the LR method of Legarra and Reverter (2018) as a viable alternative since it supports any model and data structure. This method uses statistics that compare genetic evaluations including partial and complete datasets. These statistics yield estimators of bias, slope or dispersion and population accuracy for EBVs. Legarra and Reverter (2018) conclude that their LR method is automatic, easy to use and yields the quantities of interest. The LR method and its estimators can also be used when moving from a pedigree based BLUP to a genomic single-step BLUP evaluation as was shown by Bonifazi et al. (2022).

A comprehensive genotyping strategy will have a major impact on reducing bias level (mostly called bias), bias with regards to dispersion (mostly called dispersion/inflation), as well as increasing the population accuracy of the GEBVs using the estimators or the LR method discussed previously. Bias (level) is the disparity between the (G)EBV and the true breeding value (TBV), with a value other than 0 showing over or under estimation, while dispersion (bias) is the slope of the regression of TBV on (G)EBV, with a value deviating from 1 showing over or under dispersion of (G)EBVs (Macedo et al., 2020). The population accuracy which is the correlation between the (G)EBV and the TBV is essential because it can be used to validate the accuracy (reliability) of selection response (Macedo et al., 2020). Population accuracy should be as close to a correlation with a value of 1. This is typically between datasets where one dataset is standard (i.e., without genomic information) and an alternative dataset where certain information (i.e., genomic information in the form of SNP markers) are added. This can also be the validation of datasets where both have genomic information included, but phenotypic information of the youngest generation is taken away in one dataset. Uemoto et al. (2016) reported an increase in the accuracy of GEBVs with ssGBLUP when genotyped cows were included with bulls in their reference population (n > 4000). Their study also showed that it is important to consider the genetic relationships between bulls in the reference (genotyped) population. Thomasen et al. (2020; dairy cattle) showed that the addition of a greater number of genotyped cows (n = 2000) to the sire reference population (n = 1000) can potentially increase the rate of genetic gain and slow down the rate of inbreeding, regardless of the size of the reference population. Lourenco et al. (2015) also stated that adding cows to a current sire reference population were favourable, however small. Important to note is that in most of these studies, the size of the sire reference populations for dairy cattle to which genotyped cows were added were already relatively big. It shows that changing the structure from sires only to including cows is however beneficial. Takeda et al. (2020) stated that the size of the reference population is imperative to the accuracy of GEBV predictions, but also the number of progeny of sires which will increase the quality of information as an additional input to ssGBLUP evaluations. This is because sires that are, and should be genotyped have high accuracy genetic predictions because of large numbers of measured daughters. Lourenco et al. (2015) affirmed that the greatest benefits for genomic evaluations were because of the inclusion of genotypes of sires with high (EBV) accuracy. Whether the trait for which genetic improvement is sought is only expressed in one sex plays an important part. In dairy cattle, sires can achieve very high accuracies for genetic merit for sex limited traits (milk production) because of the potentially large number of measured daughters. Lourenco et al. (2015) showed that in broiler chickens where most traits of importance are measured in both sexes, genotyping males and females increases accuracy of genetic predictions as both sexes add phenotypic information more equally to advance genetic progress. The cost of genotyping large enough numbers of the appropriate animals can be a limiting factor as genotyping animals are relatively expensive, especially in certain countries. Genotyping less animals, but using a chip with a much larger number of markers were investigated by VanRaden et al. (2011) to possible lessen costs to form a reference population. They concluded that increasing the number of animals to be genotyped will lead to better gains than increasing the number of markers. Another possible pitfall that can cause bias in genomic evaluations is called preselection. Jibrila et al. (2021) explained that animals that are selected be to parents of the next generation are mostly selected in several periods of time, with the earlier periods known as preselection. This may be problematic because these animals are selected to be superior. It was concluded that to avoid bias due to preselection in following single-step genomic evaluations, information that characterise the reference data used in the evaluation as well as genotypes and phenotypes of the preselected animals must be included in the following evaluation (Jibrila et al., 2021). Regarding GEBV accuracies, Zhou et al. (2018) presented results of decreasing accuracies as the generation of validation population increased with ssGBLUP. With the aim of improving accuracy of genetic prediction of younger, genotyped animals implies that genotyping strategies should be kept up to date with every generation (Sonessen and Meuwissen, 2009).

2.7 Conclusions

To enable animal breeders to select the correct animals that fit their selection goals, keeping in mind the markets they are producing for as well as the environments their animals must produce in now, and in future, utmost accuracy of genetic prediction is more important than ever. In developing the most suitable model, the use of the most recent and proven technology in the correct order has merit in increasing the accuracy of prediction. These technologies, from the availability of new methodologies build on each other. South African dairy genetic evaluations have been slow to adopt newer models for genetic evaluations and can reap the benefits of using random regression models alone earlier, rather than later. Realising a GxE interaction between the two most important production systems can easily increase genetic gain as much as the inclusion and adoption of genomic predictions. Lastly, finding ways to incorporate “new” models like random regression with methods to better select within the reality of a possible GxE, with genomics is a powerful tool to propel the breeding of genotypes that are more profitable and efficient producers.

CHAPTER 3

Comparison of fixed and random regression models for the analysis of milk production traits in South African Holstein dairy cattle under two production systems[1]

3.1 Introduction

Models for genetic evaluations for dairy cattle in South Africa (SA) moved from a sire model in 1987 for the estimation of BLUP breeding values to a single trait animal model in 1992 and to a multitrait (MT) one in 1999 (Mostert et al., 2004). The advantages of MT BLUP are well established (Schaeffer, 1984; Thompson and Meyer, 1986; Mrode, 2014). The current MT model used for genetic evaluation of production traits of the South African Holstein breed is a test-day (TD), fixed regression model (FRM) (Interbull, 2020). Jamrozik and Schaeffer (1997) listed factors that influence TD yields: breed, region, herd management, management group, weather conditions on the test-day of the year, lactation number, age at calving, month of calving, days in milk (DIM), pregnancy status, medical treatments and milking times per day. Most of them change at each TD. Hence test-day models can consider more non-genetic factors that influence cows in different ways during the lactation. This superiority of test-day models are supported by Swalve (2000) and Jensen (2001). Schaeffer et al. (2000) indicated that the benefits of test-day models also included better accounting of cow-transfers between herds and more accurate genetic selection of animals. This is supported by Sawalha et al. (2005) who reported greater genetic improvement by using a test-day model. Mayeres et al. (2004) stated that test-day models make it possible to evaluate responses to changes in herd management by correcting records with herd-specific correction factors. FRM use TD milk production records directly, rather than calculating 305-day milk production using the available TD production records deliberately (Ptak and Schaeffer, 1993). FRM-TD models make it possible to account for non-genetic effects on persistency over the lactation curve (Jensen, 2001). Persistency of lactation is defined as "the ability of the cow to continue to produce at a high level throughout her lactation" (Cupps, 1966), or "the ability to maintain a more or less constant yield during the lactation" (Gengler, 1996). Dekkers et al. (1998) discussed the importance of considering persistency in relation to feed costs and also outlined the role of persistency in breeding goals. They reported an overall economic value of persistency in relation to production of 10.8% and to 16.4% for an average calving interval of 13 months for a specific system using basic parameters. Results showed that persistency can have significant implications in management decisions like insemination and culling decisions (Dekkers et al., 1998). In FRM the shape of the lactation curve is determined by fixed (non-genetic) effects only, while random effects are assumed to be constant within lactation (Swalve, 2000; Mrode,

---

[1] Published. Van Niekerk et al. Livest. Sci. 267 (2023) p1-11. *https://doi.org/10.1016/j.livsci.2022.105125*.

2014). In contrast, random regression models (RRM) allow this shape to vary with the additive genetic and permanent environmental effects of the cow (Mrode, 2014).

Average milk, fat and protein production were markedly different when South African Holstein herds were grouped into four different regions, according to the global climatic characteristics of the area in which herds are situated. Differences between average milk production were also significant when herds were grouped according to level of rainfall, which was shown to be attributable to the fact that herds in higher rainfall rely more on pasture (PAST) production systems and herds in lower rainfall areas utilize more total mixed ration (TMR) production systems. (Ducrocq et al., 2022).

Important areas for research and reasons for the implementation of an RRM for genetic evaluation of dairy cattle in SA were previously presented by Dzomba et al. (2010). The objectives of our study were to compare a FRM, resembling the current model used for genetic evaluation in South African Holstein according to Interbull (2020), with various FRM and then RRM, as to ascertain the possible use of a RRM instead. These comparisons of models were done separately within each production system (PAST vs TMR) in order to investigate the likelihood of a genotype x environmental interactions due to possible differences in genetic parameters.

3.2 Material and Methods

3.2.1 Data Preparation and Variables Created

Data were obtained from the Logix animal registration and product recording system (Logix, 2019) of SA Studbook as well as the Intergis database of the Agricultural Research Council of South Africa (Intergis, 2000). These two TD databases, together with pedigree information, consist of South African Holstein herds taking part in milk recording. The two data sets were merged to make sure that all possible animals and records were included in a single data set. The raw, unprepared TD dataset consisted of 9,235,326 records. From this data set, records from 140 herds whose type of production system (PAST or TMR) is known were selected. After selection of milk recording participant herds with known production system utilized, the dataset decreased to 2,076,889 TD records. Herds taking part in milk recording for at least 3 consecutive years for the period of 1988 to 2016 were kept. Records where herd was unknown were discarded together with records of animals without birth dates, sex status, cows with unknown sires and calving date(s). Records of traits considered were milk production (Milk), fat production (Fat) and protein production (Protein), measured in kilogram (kg) as well as fat percentage (Fat%) and protein percentage (Protein%). These records were measured within a 24-hour time period and consisted of the sum of 2 or 3 milkings per day (number of milkings, Mi). TD lactation (L) records following the first 3 parities were used (lactations 1, 2 and 3, respectively). Cows that did not have first lactation records were not considered. Records were grouped into 8 calving age classes (AC) as follow: Class 1: 20 to 23 months; Class 2: 24 to 27 months; Class 3: 28 to 31 months; Class 4: 32 to 35 months; Class 5: 36 to 41 months; Class 6: 42 to 47 months; Class 7: 48 to 55 months; Class 8: 56 to 69 months. Previous calving interval for each lactation was calculated and 8 calving interval classes (CIC) were formed with a range from 312 to 730 days. Classes 1 to 6 each had the same range of 42 days (21 days anoestrus plus 1 cycle of 21 days, Opsomer et al. 1996) while class 7 ranged from 564 to 730 days. Class 8 was a separate class for heifers calving for the first time (no previous calving interval). TD records from 6 to 300 days in milk (DIM) were considered for each lactation and cows without a first TD record within day 6 and 65 were discarded for that particular lactation. TD records where the length between two sequential test days were less than 20 days and more than 90 days as well as lactations with less than 7 test days, were not used.

Table 3.1 Descriptive statistics of test-day records for the first 3 lactations of milk production (Milk), fat production (Fat), protein production (Protein), fat content (Fat%) and protein content (Protein%) in the pasture (PAST) and total mixed ration (TMR) production systems, with standard deviations (where applicable) in parentheses

	Lactation 1		Lactation 2		Lactation 3	
Item	PAST	TMR	PAST	TMR	PAST	TMR
Avg. Milk (kg)	21.5 (5.4)	30.4 (8.4)	25.2 (7.1)	34.2 (10.5)	26.8 (7.8)	34.8 (10.9)
Avg. Fat (kg)	0.8 (0.2)	1.1 (0.4)	0.9 (0.3)	1.2 (0.5)	1.0 (0.3)	1.3 (0.5)
Avg. Protein (kg)	0.7 (0.2)	1.0 (0.3)	0.8 (0.2)	1.1 (0.3)	0.8 (0.2)	1.1 (0.3)
Avg. Fat%	3.6 (0.7)	3.6 (0.7)	3.6 (0.7)	3.6 (0.8)	3.6 (0.7)	3.6 (0.8)
Avg. Protein%	3.2 (0.3)	3.2 (0.3)	3.2 (0.3)	3.2 (0.3)	3.2 (0.3)	3.2 (0.3)
n records	89,031	297,871	56,112	150,807	35,223	69,403
n cows with records	11,363	38,340	7,176	19,385	4,537	8,978
Avg. n records per cow	7.9	7.8	7.9	7.8	7.8	7.8
Avg. n records per HTMi	49	99	35	54	24	29
Avg. n daughters per bull	18	36	13	19	9	10

Avg. = average; n = number of; HTMi = Herd x Test-day x Number of milkings interaction effect (contemporary group)

Contemporary group was defined as the interaction between Herd x Test-day x Number of Milkings (HTMi) and groups with less than 5 cows were discarded. Herd x Calving Year x Calving Season (S) x Lactation (HYSL) were also considered. Here (HYSL), S were comprised of four calving seasons as follow: Calving season 1: December, January and February; Calving season 2: March, April and May; Calving season 3: June, July and August; Calving season 4: September, October and November. Calving Month was also considered on its own in chronological order (Calving Month 1: January; Calving Month 2: February; etc.). Calving year x Lactation interaction (CYL) was also formed. The interactions of Herd x Calving Year (Y) x Calving Month (Mo) (HYMo) as well as a calving season x calving age class interaction (SAC) were formed as these are used in the current South African Holstein model (Interbull, 2020). In the latter, they define two calving seasons: April to September and October to March (as in Mostert et al., 2006a).

The prepared data set was divided to form two data sets consisting of TD records for herds using either a PAST or a TMR production system, respectively. For the TMR dataset 40% of records were from 2 milkings and 60% from 3 milkings per day. For the PAST dataset, 91% of records were from 2 and 9% from 3 milkings per day. Most of the cows in the datasets had a sire from the United States (44% and 45% of the PAST and TMR datasets, respectively), followed by South Africa (36% and 34%), Canada (8% and 8%), Netherlands (7% and 6%), Italy (1% and 2%) and France (1% and 1%) in that order. The final data sets consisted of 518,081 and 180,366 records for TMR and PAST (698,447 records in total), respectively (Table 3.1).

3.2.2 Statistical Analyses

All records from herds were analysed using restricted maximum likelihood utilizing the WOMBAT software (Meyer, 2007) under different multiple-lactation FRM and RRM. Five traits, Milk, Fat, Protein, Fat% and Protein% were analysed separately in a multiple-lactation setting, where the traits corresponded to the first 3 lactations. Heterogeneous residual variance due to DIM were defined by 30-day residual classes (RC) over the 3 lactations to form 10 classes per lactation, except for the last class of each lactation which covered 25 days only (295-day lactation period in total i.e., because first TD considered were DIM 6). Partitioning the residual variance in 10 classes over the lactation are in accordance with De Roos et al. (2004; 10 classes, but with different intervals in DIM) as well as with Mostert et al. (2006a) (9 x 30-day classes with the tenth class consisting of 35 days). Olori et al. (1999a) showed that the best fit for their model with regards to the log-likelihood function were obtained when the residual error was assigned in 10 classes over the lactation. Olori et al. (1999b) concluded that assuming the residual variance to be constant over the lactation will bias the residual variance, and that the residual variance will be more accurate by estimating unrelated (individual) estimates at different stages of the lactation as to allow for heterogeneous variances (correlations) between "temporary" environmental effects between test-day records (Meyer, 1998).

Goodness of fit between models was assessed based on the Bayesian Information Criterion (BIC = -2logL + log(N)*k, where logL denote the value of the maximum log likelihood of the model, N is the number of examples in the training dataset, and k is the number of parameters in the model; Hastie et al., 2017) and mean squared error.

3.2.3 Statistical Analyses: Replication of the Current and Given Fixed Regression Model Utilized for South African Holstein

The given and official FRM used for South African Holstein genetic evaluations (saFRM – Interbull, 2020) was replicated and not developed in this study. For each of the 5 traits, the multiple-lactation (lactations 1, 2 and 3), TD and simulated saFRM can be described as:

$$y_{ijkmn} = HTMi_i + HYMo_j + (CIC_k, L) + f_{Wilmink}(SAC_m, DIM, L) + (AG_n, L) + (PE_n, L) + e_{ijkmn} \quad [1]$$

where y is the TD record of 24-h test-day Milk, Fat, Protein, Fat% and Protein% observations for the first 3 lactations, $HTMi_i$ is the ith Herd x Test-day x number of Milkings effect (contemporary group), $HYMo_j$ is the jth Herd x calving Year x calving Month effect, and CIC_k is the kth previous Calving Interval Class effect for Lactation (L) 1, 2 or 3. These 3 variables represent the fixed (class) effects in the model. SAC_m is the mth calving Season (two seasons) x Age at calving Class as a regression over DIM, using the Wilmink (1987) function, for Lactation 1, 2 or 3 and represent the fixed regression in the model. The random effects are the additive genetic animal (AG) and permanent environmental (PE) effects of the nth animal for Lactation 1, 2 or 3. The term e_{ijkmn} represent the random residual effect belonging to observation y_{ijkmn}, utilizing heterogenous residual variances for the stage of lactation. (Interbull, 2020).

The distribution of random effects was assumed to be:

$$\begin{bmatrix} AG \\ PE \\ e \end{bmatrix} \sim N(0, V), \qquad [2],$$

$$\text{With } V = \begin{bmatrix} G \otimes A & 0 & 0 \\ 0 & I \otimes P & 0 \\ 0 & 0 & R_{RC,Lactation} \end{bmatrix}, \qquad [3],$$

where G and P (see equation 5) are covariance matrices for the AG and PE effects, respectively; A is the additive relationship matrix; I is an identity matrix; and $R_{RC,Lactation}$ is the matrix of residual covariances among traits with elements depending on the heterogeneous residual variance class (RC) and lactation. The saFRM model clusters most of the fixed effects together and consider them as constant over the length of the lactation, except for the

SAC interaction effect (Interbull, 2020). Also, the random effects (AG and PE) were assumed to be constant over a lactation.

3.2.4 Statistical Analyses: Comparing Other Fixed Regression Models with the South African Holstein Model

Various other FRM analyses were performed and compared to the saFRM. The use of different types of regression on DIM tested included the Wilmink curve (Wilmink, 1987), Legendre polynomials (of order 3 or 4), cubic splines and natural splines (with 4 or 6 knots). The number of FRM tested in order to find the model with the best BIC fit for the fixed effects part of the model was too great to show and beyond the scope of the article. A final FRM that cumulates different curves over the lactation for different fixed effects was retained based on results in the PAST dataset. Natural splines with 6 knots (DIM 6, 20, 50, 135, 245 and 300) were found to be the best regression function to describe these curves of the various fixed effects over DIM. This FRM was also applied in the TMR dataset.

3.2.5 Statistical Analyses: Extending a Final Fixed Regression Model to a Random Regression Model

Then, the model was extended to a RRM where the random components AG and PE were described separately for each lactation as the sum of a constant and a regression slope function of DIM multiplied by current DIM, therefore combining for each lactation an average production (PROD; first principal component i.e., PC or eigenvector) and an average persistency (PERS; second PC or eigenvector) effect per lactation (6 PCs or eigenvectors in total over the 3 lactations). Eigenvectors were derived from AG and PE and estimated by standardizing DIM between -0.5 to 0.5 as a linear regression (see paragraph after equation [4]). Druet et al. (2003, 2005) also derived the same eigenvectors from AG and PE estimates, however using Legendre polynomials. Various previous studies (Van der Werf et al., 1998; Olori et al., 1999a; Pool et al., 2000; Jamrozik et al., 2002; Druet et al., 2003 and 2005) have shown that by observing eigenfunction tendencies, the inclusion of a second term (regression coefficient) is an intrinsic measure for genetic persistency (PERS). Druet et al. (2003, 2005) argued that the explanation of a third eigenvector (per lactation) was, in biological terms, more intricate and made less sense to include. This model was compared to the current South African Holstein FRM under both production systems and thus had 3 PCs or eigenvectors i.e., one for each of the 3 lactations. The Milk data was used for all analyses to derive the final (RRM) model, which was also used for comparison with the FRM for the other traits. Models were compared for goodness of fit based on their BIC and mean squared error.

For each of the 5 traits, the final multiple-lactation (lactations 1, 2 and 3), TD alternative random regression model (aRRM) can be described as:

$$y_{ijkmnopq} = HTMi_i + HYSL_j + f_{Splines}(Herd_k, DIM, L) + f_{Splines}(CYL_m, DIM, L) + f_{Splines}(Mo_n, DIM, L) + f_{Splines}(CIC_o, DIM, L) + f_{Splines}\left(AC_p, DIM, L\right) + \sum_{r=0}^{1} AG_{qr}\, \varphi_{(r,t,l)} + \sum_{r=0}^{1} PE_{qr}\, \varphi_{(r,t,l)} + e_{ijkmnopq} \qquad [4]$$

where y is the TD record of 24-h test-day Milk, Fat, Protein, Fat% and Protein% observations for the first 3 lactations, $HTMi_i$ is the ith Herd x Test-day x Number of Milkings effect (contemporary group) and $HYSL_j$ is the jth Herd x calving Year x calving Season (4 seasons) x Lactation number effect. These 2 variables represent the fixed (class) effects in the model. Regressions on DIM within various fixed effects, using natural splines with 6 knots as the regression function were used for Lactation (L) 1, 2 or 3. These effects were $Herd_k$ as the kth Herd effect, CYL_m as the mth Calving Year x Lactation effect, Mo_n as the nth calving Month effect, CIC_o as the oth previous Calving Interval Class effect and AC_p as the pth Age of calving Class effect. The random effects are represented by AG_{qr} and PE_{qr}, which are the additive genetic animal and permanent environmental random regression coefficients, respectively of animal q for the rth term of the linear regression. The value of $\varphi_{(r,t,l)}$ is equal to the rth term of the linear regression at DIM t (standardized from -0.5 to 0.5) for Lactation l. The term $e_{ijkmnopq}$ represent the random residual effect belong to observation $y_{ijkmnopq}$. The G and P (co)variance structure can be described as:

$$Var\begin{bmatrix} G_1 \\ G_2 \\ G_3 \\ G_4 \\ G_5 \\ G_6 \\ P_1 \\ P_2 \\ P_3 \\ P_4 \\ P_5 \\ P_6 \end{bmatrix} = \begin{bmatrix} G_1 & G_{12} & G_{13} & G_{14} & G_{15} & G_{16} & 0 & 0 & 0 & 0 & 0 & 0 \\ G_{21} & G_2 & G_{23} & G_{24} & G_{25} & G_{26} & 0 & 0 & 0 & 0 & 0 & 0 \\ G_{31} & G_{32} & G_3 & G_{34} & G_{35} & G_{36} & 0 & 0 & 0 & 0 & 0 & 0 \\ G_{41} & G_{42} & G_{43} & G_4 & G_{45} & G_{46} & 0 & 0 & 0 & 0 & 0 & 0 \\ G_{51} & G_{52} & G_{53} & G_{54} & G_5 & G_{56} & 0 & 0 & 0 & 0 & 0 & 0 \\ G_{61} & G_{62} & G_{63} & G_{64} & G_{65} & G_6 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & P_1 & P_{12} & P_{13} & P_{14} & P_{15} & P_{16} \\ 0 & 0 & 0 & 0 & 0 & 0 & P_{21} & P_2 & P_{23} & P_{24} & P_{25} & P_{26} \\ 0 & 0 & 0 & 0 & 0 & 0 & P_{31} & P_{32} & P_3 & P_{34} & P_{35} & P_{36} \\ 0 & 0 & 0 & 0 & 0 & 0 & P_{41} & P_{42} & P_{43} & P_4 & P_{45} & P_{46} \\ 0 & 0 & 0 & 0 & 0 & 0 & P_{51} & P_{52} & P_{53} & P_{54} & P_5 & P_{56} \\ 0 & 0 & 0 & 0 & 0 & 0 & P_{61} & P_{62} & P_{63} & P_{64} & P_{65} & P_6 \end{bmatrix} \quad [5]$$

where subscripts 1, 3 and 5 are the average production (PROD) variances for Lac1, 2 and 3, respectively, and subscripts 2, 4 and 6 are the average persistency (PERS) variances for Lac1, 2 and 3, respectively.

The distribution of random effects was assumed as in equation [2] with variances as in equation [3] except for G and P. Here (equation 4), these matrices allow the AG and PE random effects, respectively to vary over the lactation, which are not the case in a FRM (and saFRM; equation 1).

3.3 Results and discussion

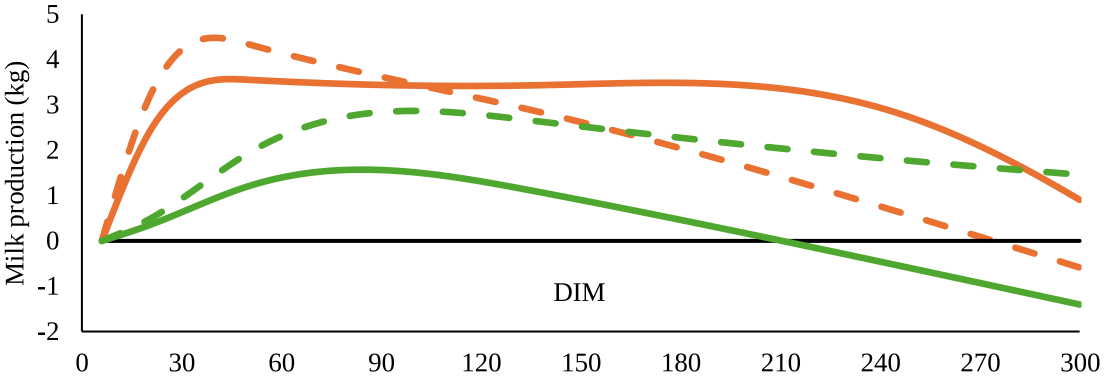
3.3.1 Goodness of Fit

Using the aRRM instead of the saFRM led for both production systems to an increase in the number of parameters estimated. However, taking the increased number of parameters into account, the BIC showed a significant improvement in goodness of fit of the aRRM compared to the saFRM for all production traits (Milk, Fat and Protein) within both PAST and TMR (Table S3.10 in the supplementary material for Milk). This was also the case for Fat% but not for Protein%, where the saFRM had a significantly better fit within both production systems. More research is needed to find the cause of the less suitable fit for Protein%. Specifically, comparing the FRM to the aRRM with the exact same fixed effects structure and thus, where the only difference was the inclusion of a second term (regression coefficient over DIM for each lactation) for the random effects (AG and PE), the BIC of the aRRM was significantly better (improvement of 9,611.77 for Milk in PAST). The mean squared error of the aRRM were lower than the corresponding values of the saFRM for all production and milk composition traits. Swalve (2000) found studies that compared the FRM vs RRM scarce and this still holds true. This author concluded that RRM are superior to FRM as the genetic merit of a cow over her lactation can be estimated and was shown to vary for production. In one of a few studies comparing RRM and FRM, Buttchereit et al. (2010) found random regression models superior to fixed regression models for fat:protein ratio, using the corrected Akaike information criterion (AICc) as evaluation criteria, AICc being AIC with an extra term to penalize for smaller numbers of data points (Burnham and Anderson, 2002). Other evaluation criteria used by these authors were the correlation between the real observation and the estimated value, and residuals outlined alongside DIM. In 2019, RRM were the models of choice for genetic evaluation in 19 out of 32 countries participating in Interbull (Interbull, 2020).

3.3.2 Estimates of Separated Fixed Effects

The inclusion of distinct, cumulative fixed effect regressions on DIM rather than combinations of effects as in the saFRM had a significant and favourable influence on the BIC when included in the model, showing that these effects are not constant over the lactation (Figure 3.1 for different herd effect curves, and Figures S3.1 to S3.4 in the supplementary material for AC, CIC, Mo and CYL effects, respectively). This result coincides with a study by Leclerc et al. (2008) where fixed effects such as days carried calf (effect of gestation), month of calving, length of dry period and calving age were shown to vary over the lactation for multiparous cows. In our study, it was especially true for the individual curves fitted for each herd: they corresponded to the fixed effect that caused the BIC to change the most (increase of 1,785.51 and 5,341.70 when herd curves are removed from the aRRM for Milk in PAST and TMR, respectively). This result is supported by the study of Muller et al. (2014) which reported herd as the fixed effect with the most significant effect when they studied fertility traits in South African Holstein and presumed this effect as an indicator of management level.

Figure 3.1 Estimates of four different Herd effects over days in milk of which two utilize a pasture production (green lines) system and two a total mixed ration (orange lines) production system from the alternative random regression model, over days in milk (DIM) for first lactation.



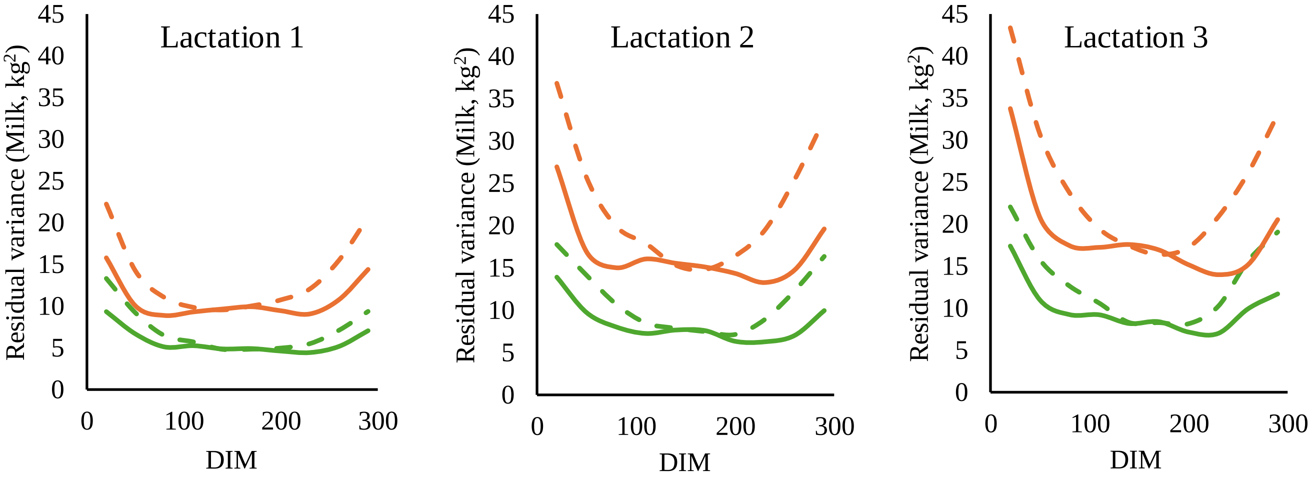
In other studies, Druet et al. (2003) and De Roos et al. (2004) fitted herd x year curves as an additional random effect to the additive genetic and permanent environmental effects. De Roos et al. (2004) recommended the inclusion of herd curves in RRM as there was an unjustly increase of the estimated genetic variance at the borders of the lactation if not included. The concatenation of the herd and additive genetic effects since dams and daughters produce in the same herd may make it problematic for animal models to accurately distinguish between these effects (De Roos et al., 2004). Jamrozik et al. (2001) included a Herd x Year x Season curve effect over the lactation to mitigate the overestimation of genetic variance (early lactation), caused by excluding differing herd lactation curve shapes. Gengler and Van (2001) reported lower heritabilities when incorporating a herd-stage effect over the lactation, indicating that part of the genetic effect was actually a consequence of this herd effect.

3.3.3 Residual Variance Estimates

In accordance with previous studies (Olori et al., 1999a,b; Jamrozik et al., 2001; De Roos et al., 2004; Druet et al., 2005; Mostert et al., 2006b; Hammami et al., 2008a; Miglior et al., 2009) the residual variance is found to be heterogeneous over the lactations using either the saFRM or the aRRM for both production systems (PAST and TMR) and for all traits analysed (Figure 3.2 for Milk, and Figures S3.5 to S3.8 in the supplementary material for Fat, Protein, Fat% and Protein%).

Residual variance estimates with the aRRM were lower than with the saFRM for all traits, especially at the beginning and the end of the lactation for both production systems which is in accordance with Liu et al. (1998) who reported lower residuals for the RRM than for the FRM. The pattern of higher residual variance estimates at the beginning and end of the lactation is in accordance with other studies (Jamrozik et al. 2001; Druet et al. 2005, Mostert et al. 2006b). Residual variance estimates were on average the lowest for lactation 1, increasing each time when moving to the later lactation, irrespective of the model used (saFRM or aRRM). This were also reported by De Roos et al. (2004), Druet et al. (2005) and Mostert et al. (2006b). This was true for all traits except for Protein% where it appeared to be more or less the same for all lactations (Figure S3.8 in the supplementary material). The residual variance over DIM increasing from lactation 1 compared to later lactations is in line with the increase in average production from first lactation to later lactations (Robert-Granié et al., 1999), being an important source of heterogeneous variance over time.

Figure 3.2 Residual variance estimates for milk production (Milk) of the separate pasture production (green lines) and the total mixed ration production (orange lines) system datasets from the South African Holstein fixed regression (dash lines) and the alternative random regression (solid lines) model analyses, over days in milk (DIM) for the first 3 lactations.



3.3.4 *Production Traits: Heritability Estimates (PAST Production System)*

The heritability estimates over DIM for the 3 lactations for Milk using the saFRM were highest during the middle of each lactation and lowest during early and late lactation (Figure 3.3) with a consistent pattern of the estimates between the 3 lactations.

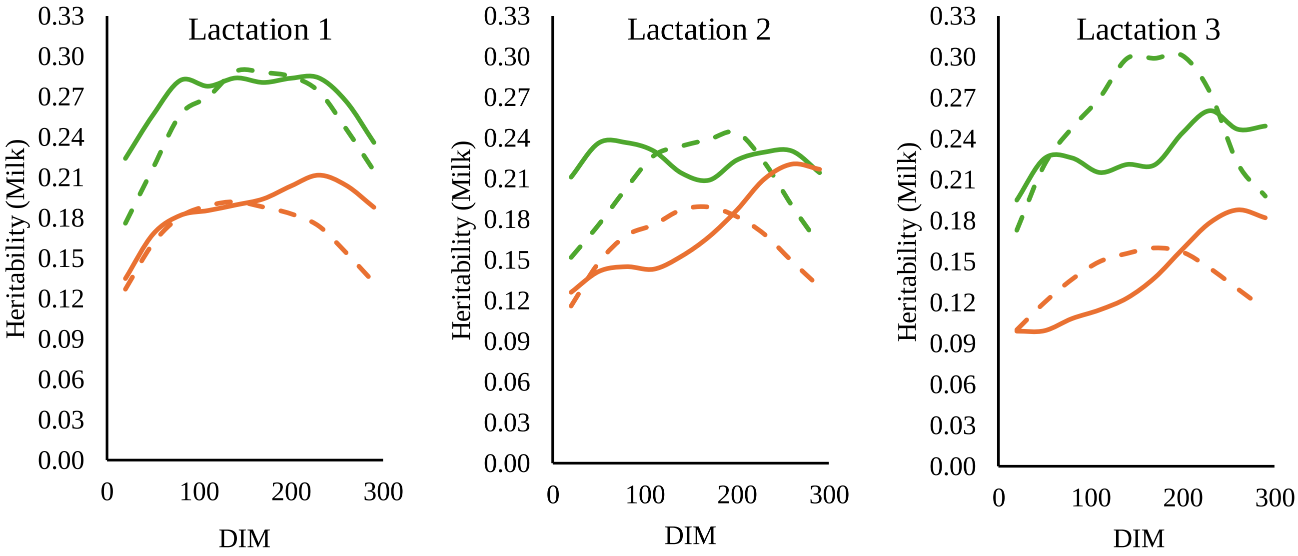
Studies by Mostert et al. (2006a,b) on South African Holstein did not report the pattern of heritability estimates over DIM, but appear to be in accordance with our results, since their trend of the residual variance estimates over DIM (higher at the beginning and end of the lactations is similar to ours (Figure 3.2 for Milk).

Also, since the genetic variance estimates are assumed to be constant over the lactation in FRM, this pattern is to be expected (Figures 3.4a, 3.4c and 3.4e for Milk). The corresponding pattern of heritability estimates over the 3 lactations for all other traits were mostly similar to Milk (Figures S3.10, S3.12, S3.14 and S3.16 in the supplementary material for Fat, Protein, Fat% and Protein%, respectively).

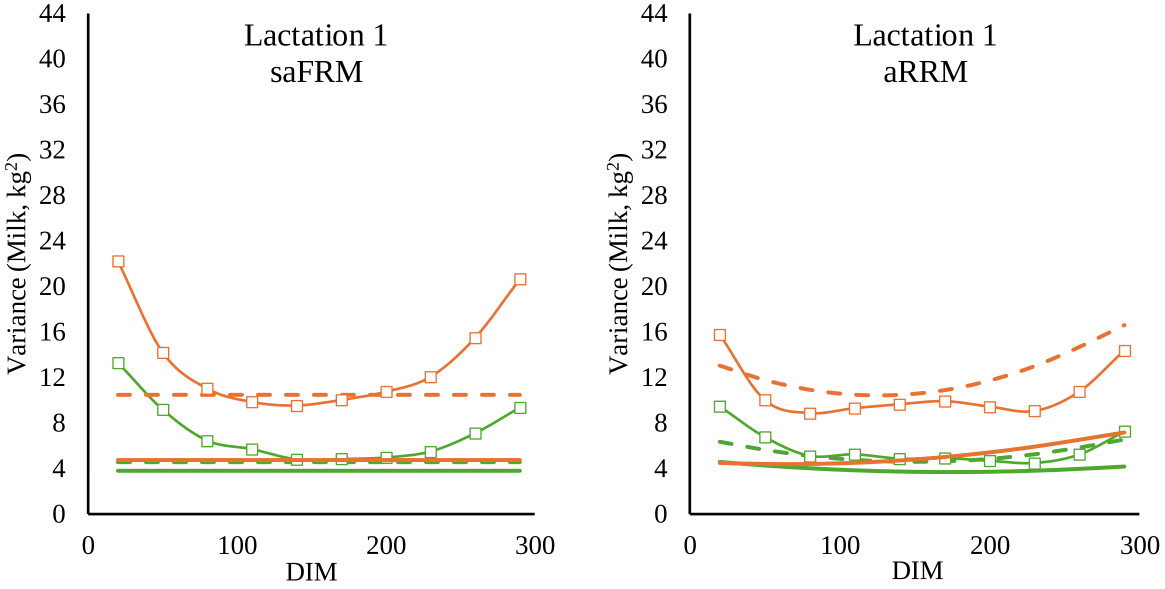
Heritability estimates for Milk using the aRRM were higher than with the saFRM in most cases (lactations 1 and 2), especially at the beginning and the end of the lactation, with estimates of the two models being close to equal during the mid-lactation periods (Figure 3.3). The lower residual variance of the aRRM contributes to these higher heritabilities (Figure 3.2), as well as the higher AG and PE variances at the beginning and end of these lactations (Figures 3.4a to 3.4f). In contrast, heritability estimates for lactation 3 using the aRRM were mostly lower than with the saFRM and can be attributed to the relatively larger PE variance using the aRRM (Figures 3.4a to 3.4f).

Early and late lactation TD records carry a heavier weight and make a bigger contribution towards total lactation production (i.e., 295-day) estimated breeding values (EBVs) utilizing the aRRM (lactations 1 and 2), whereas the same TD records with the saFRM are weighted more conservatively. Lactation 3 TD records (aRRM) are mostly weighted to the same degree as lactation 2 (saFRM TD records bearing generally higher weights), with TD records during late lactation being weighted the heaviest. The aRRM heritability estimates were more consistent (with a narrower range) over the lactation for Milk.

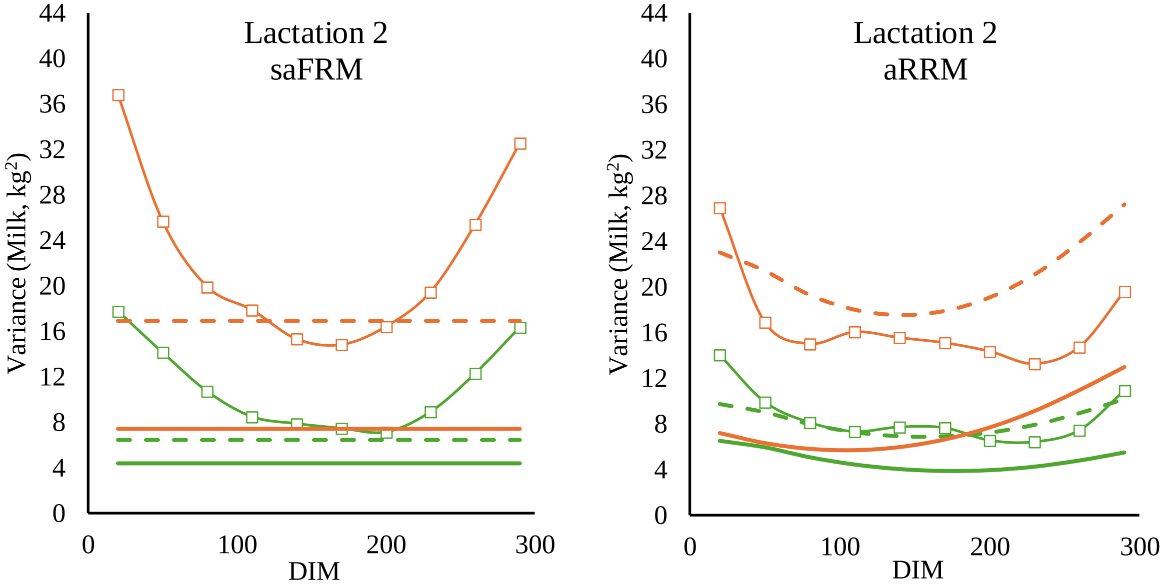
Figure 3.3 Heritability estimates for milk production (Milk) of the separate pasture production (green lines) and the total mixed ration production (orange lines) system datasets from the South African Holstein fixed regression (dash lines) and the alternative random regression (solid lines) model analyses, over days in milk (DIM) for the first 3 lactations. Standard errors ranged from 0.005 to 0.027.



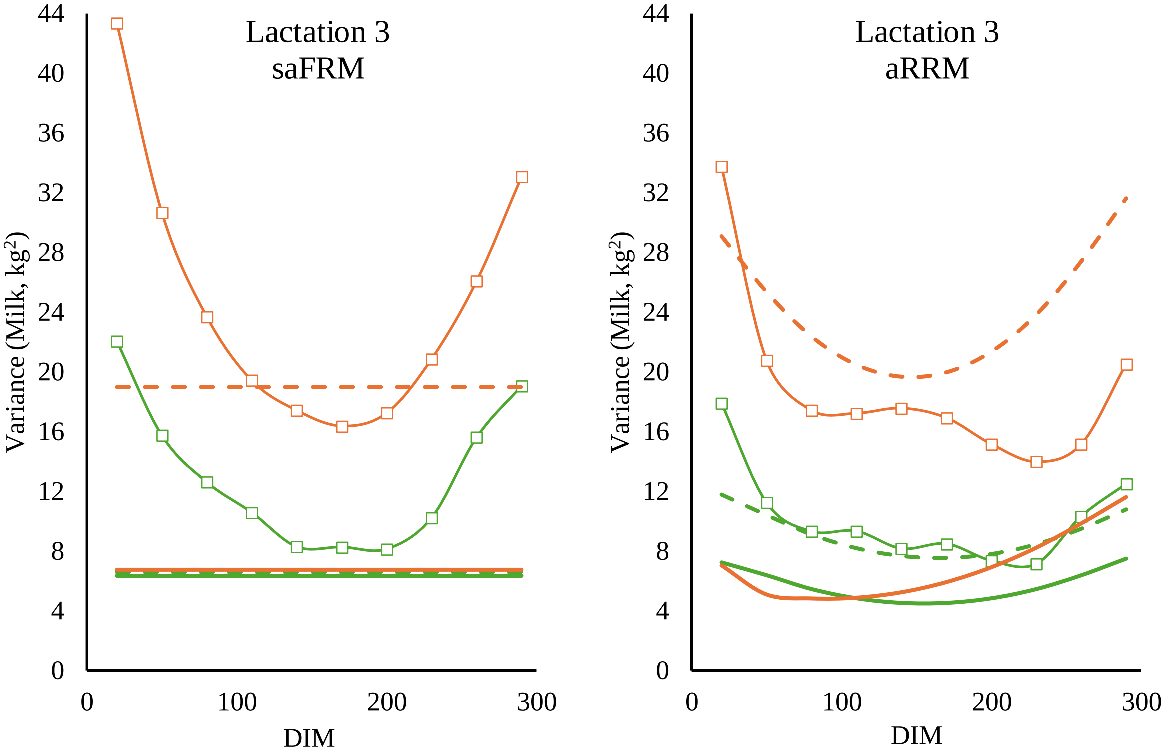
Figures 3.4a and -b Variance components for milk production (Milk) of the genetic (solid lines), permanent environmental (dash lines), residual (solid lines with squared markers) effects for the pasture (green lines) and total mixed ration (orange lines) using the South African fixed regression (saFRM, left) and the alternative random regression (aRRM, right) model analyses, over days in milk (DIM) for lactation 1.



Figures 3.4c and -d Variance components for milk production (Milk) of the genetic (solid lines), permanent environmental (dash lines), residual (solid lines with squared markers) effects for the pasture (green lines) and total mixed ration (orange lines) using the South African fixed regression (saFRM, left) and the alternative random regression (aRRM, right) model analyses, over days in milk (DIM) for lactation 2.



Figures 3.4e and -f Variance components for milk production (Milk) of the genetic (solid lines), permanent environmental (dash lines), residual (solid lines with squared markers) effects for the pasture (green lines) and total mixed ration (orange lines) using the South African fixed regression (saFRM, left) and the alternative random regression (aRRM, right) model analyses, over days in milk (DIM) for lactation 3.



The lactation 1 heritability pattern for Milk in PAST using the aRRM (i.e., heritability estimates lower at the beginning and end of the lactation, and higher during mid-lactation) is in accordance with Jakobsen et al. (2002). Heritability estimates for first lactation is the most published shape and the pattern of heritability estimates over the lactation show considerable variance between studies in literature with contributing factors outlined by Misztal et al. (2000). These include fixing the PE variance and/or modelling the residual variance as homogeneous over the lactation, as well as using a single-trait approach, causing the heritability to typically increase rapidly at both ends of the lactation curve (Misztal et al., 2000). These factors possibly played a role in the differing heritability patterns from our results found by other authors (Jamrozik and Schaeffer, 1997; Van der Werf et al., 1998). Bignardi et al. (2009) reported an opposite heritability pattern (higher at the beginning and end of the lactation). Possible (additional) reasons for differences from our study include the use of polynomial regressions by these authors (opposed to splines), including cows with less records per lactation (limited information of the lactation) (Misztal et al., 2000) and inherent differences between datasets as found by Gengler et al. (1999). Varying genetic parameters can also be due to differences in fixed effects included and described. An applicable effect is gestation which can cause important differences in milk traits (Loker et al., 2009; Penasa et al., 2016; Lainé et al., 2017) and subsequent heritability estimates (pattern and overall level) but was not considered (per se) in this study. Druet et al. (2003) stated that this effect might influence the end of lactation and create large differences between pregnant and non-pregnant cows, causing the residual variance to be greater at the end of lactation. Leclerc et al. (2008) showed a negative effect of pregnancy that began at 112, 131 and 132 days in calf for first, second and third parity cows, respectively. Our aRRM model does however include a previous calving interval effect (CIC) which partially consider the effect (of stage) of gestation. Data on insemination dates would be ideal but was not available (recorded) in the South African Holstein dataset.

On average, the heritability estimates for Milk using the aRRM show a higher level for first lactation than for later lactations (lactations 2 and 3). This is to be expected as the non-additive genetic and environmental variances (PE and residual) and the AG variance are more similar between later lactations than in lactation 1. Additionally, the AG variance explains more of the phenotypic variance at the beginning and end of later lactations than in lactation 1 (Figures 3.4b, d and f). This is in contradiction with the saFRM where on average heritability estimates for lactations 1 and 3 are similar (both with an average of 0.25), with estimates for lactation 2 (average 0.20) being lower. The lower lactation 2 heritability is caused by the increase in the PE and residual variances from lactation 1 to lactation 2, with the AG variance staying approximately the same. Mostert et al. (2006b) reported a heritability estimate of 0.24 for Milk when using data adjusted for heterogeneous variances in a fixed regression TD model in the South African Holstein population. The average aRRM heritability estimates for Milk in PAST for lactations 1 and 2 are higher than what is currently used in the genetic evaluation (0.27 vs 0.22 for lactation 1 and 0.22 vs 0.19 for lactation 2; Interbull, 2020). The current genetic evaluation heritability estimates for lactation 3 is higher (0.27 vs 0.23). The range of the aRRM heritability estimates (Milk, PAST) from this study were mostly in accordance with Holsteins maintained in a tropical environment (Meseret and Negussie, 2017) and relatively close to Italian Holsteins (Muir et al., 2007). Hammami et al. (2008a) and Strabel and Jamrozik (2006) reported lower estimates for Milk for the first 3 lactations. According to Hammami et al. (2008a) high estimates of AG variances and heritabilities are associated with high milk production levels.

The pattern in heritability estimates for Protein (Figure S3.12 in the supplementary material) were similar to Milk (aRRM and saFRM). Heritability patterns for Fat (Figure S3.10 in the supplementary material) were slightly different during late lactation, where heritability estimates were higher than in Milk. In general, heritability estimates for Protein were slightly lower than for Milk, with Fat having the lowest heritability estimates.

The permanent environmental effect was in most cases the random effect with the largest influence on production traits (Milk, Fat and Protein) over all lactations for both PAST and TMR production systems (Figures S3.9, S3.11 and S3.13 in the supplementary material). Various authors (Mostert et al., 2006b, Strabel and Jamrozik, 2006;

Savegnago et al., 2013) also reported higher PE variances compared to the AG variances as proportions of the total phenotypic variances.

3.3.5 Production Traits: Eigenvectors (PAST Production System)

The first two eigenvectors (using aRRM), representing respectively the average production (PROD) and persistency (PERS) genetic effects for lactation 1, explained 90.8% to 94.0% of the total AG variance effect (Table S1 in the supplementary material) in PAST. This fraction represented close to 100% after lactation 2 (99.5% to 99.7% for production traits) i.e., considering the third and fourth eigenvectors. These results are in accordance with Jamrozik et al. (2002) who reported a value of ~90% for the first two eigenvectors in Milk over several countries (Australia, Canada, Italy and New Zealand). These values were higher in the aRRM where a persistency component is included than when using the saFRM (87.7% to 88.4% explained by the first eigenvector, representing lactation 1, and 96.7% to 97.8% after the second eigenvector representing lactation 2, without a persistency component). Druet et al. (2005) reported values of 91.6% and 6.6% for the first and second eigenvectors for first lactation milk yield, explaining 98.2% of the AG variation, compared to 91.3% (lactation 1 Milk in PAST) in this study (63.5% and 27.8% for the first and second eigenvector, respectively). Leclerc et al. (2008) reported a value of 98.4% for four eigenvectors to explain the total genetic variation in the first 3 lactations when reducing the rank of the AG variance matrix from six as previously estimated by Druet et al. (2005). Compared to the results of the study by Druet et al. (2005), the persistency component of the current study is of a higher magnitude and seem to play a more prominent role in the South African Holstein breed (PAST).

3.3.6 Production Traits: Average Production Genetic Correlations (PAST Production System)

Genetic correlations between average production (Milk, Fat and Protein) of all lactations using the aRRM were stronger than the corresponding correlations using the saFRM (Table 3.2 for Milk, and Tables S3.2 and S3.3 in the supplementary material for Fat and Protein, respectively). This was the case for both PAST and TMR production systems. SaFRM genetic correlations ranged from 0.66 to 0.96 while aRRM genetic correlations were 0.83 to 0.99. Apart from a different fixed effects structure, different curve functions and obvious differences between FRM and RRM with regards to random effects, another reason for these differences in genetic correlations is the absence of a persistency component in the saFRM.

Average production aRRM genetic correlations for all production traits in PAST were considerably higher than the parameters used in the current genetic evaluation (Interbull, 2020; MT, TD and FRM without a genetic persistency component) ranging between 0.93 to 0.96 vs 0.56 to 0.67 between lactations 1 and 2, 0.89 to 0.93 vs 0.53 to 0.59 between lactations 1 and 3, and 0.97 to 0.99 vs 0.67 to 0.74 between lactations 2 and 3, respectively.

Average Milk genetic correlations from the aRRM for lactations contiguous to each other (lactations 1,2 and lactations 2,3) were higher than lactations more distant from each other (lactations 1,3) as were also found by Strabel and Jamrozik (2006), as well as in other studies, which reported genetic correlations on a 305-day production basis (De Roos et al., 2004; Hammami et al., 2008a; Meseret and Negussie, 2017). Our estimates (Milk) were in accordance with Druet et al. (2005), Strabel and Jamrozik (2006) and Togashi et al. (2008).

Table 3.2 Genetic correlations between the first 3 lactations for average milk production for the South African Holstein fixed regression model (saFRM) vs the alternative random regression model (aRRM) for the pasture (PAST) and total mixed ration (TMR) datasets, with standard errors in parentheses

	PAST		TMR	
Lactations	saFRM	aRRM	saFRM	aRRM
1,2	0.77 (0.03)	0.95 (0.02)	0.77 (0.03)	0.87 (0.02)
1,3	0.69 (0.04)	0.89 (0.03)	0.77 (0.04)	0.88 (0.03)
2,3	0.93 (0.02)	0.97 (0.02)	0.92 (0.03)	0.99 (0.02)

ij = between lactations i and j

3.3.7 Production Traits: Average Production and Persistency Genetic Correlations (PAST Production System)

Genetic correlations between average production (PROD) and their corresponding average persistency (PERS) estimates within each of the first 3 lactations for all production traits were weak, ranging from -0.16 to 0.09 (Table 3.3 for Milk, and Tables S3.6 and S3.7 in the supplementary material for Fat and Protein, respectively). Different studies define persistency of milk production traits in various ways and in our study, persistency is estimated as the average regression slope of DIM over the lactation.

Genetic correlations between average production and their corresponding persistencies for production traits in South African Holstein could not be found in literature. Sölkner and Fuchs (1987) and Gengler (1996) reasoned that persistency should be estimated independently from milk yield, as is the case in this study. Mostert et al. (2008; SA Jersey data) found a negative and weak genetic correlation (-0.14) between 60- and 280-day yield EBVs as an indicator of persistency (Schaeffer et al., 2000) only after 305-day milk yield EBVs were included as covariates in their genetic model, with an estimate of -0.81 before this addition. Our results were in line with Wasike et al. (2014) who found a weak genetic correlation between persistency and peak milk yield as well as with total milk yield in Kenyan dairy cattle, irrespective of differing persistency descriptors. Yamazaki et al. (2013) reported weak genetic correlations between 305-day milk yield and persistency and their results, together with those from Strabel and Jamrozik (2006) support findings from this study in that selection for persistency (PAST) of Milk can be successful without changing the average level of production.

Genetic correlations between persistency estimates in PAST of the 3 lactations for Milk (Table 3.3) and Protein (Table S3.7) followed the same tendencies, being positive throughout and of low magnitude between lactation 1 and later lactations (range 0.29 to 0.36) but of high magnitude between later lactations (range 0.96 to 0.97). Lactation 1 EBVs for persistency of production is genetically not a good indicator of persistency in later lactations and selection on lactation 1 persistency EBVs will be inconsistent and unreliable to improve persistency in later lactations. However, selection on persistency for lactation 2 will be successful in improving persistency in lactation 3. Fat (Table S3.6) also followed the same tendencies as Milk, but some genetic correlations between lactation 1 and later lactations for PAST tended to be more moderate (0.47 and 0.46, respectively). Yamazaki et al. (2013), analysing only the first 2 lactations found a slightly stronger correlation (0.63) in Milk.

Table 3.3 Genetic correlations between the two components of the random regression model for milk production [(average milk production (PROD) and persistency (PERS)] within and between the first 3 lactations for the pasture (PAST) and the total mixed ration (TMR) production systems, with standard errors in parentheses

		TMR					
PAST		Lactation 1		Lactation 2		Lactation 3	
		PROD	PERS	PROD	PERS	PROD	PERS
Lactation 1	PROD		-0.32 (0.05)	0.87 (0.02)	-0.14 (0.05)	0.88 (0.03)	-0.20 (0.07)
	PERS	0.04 (0.07)		-0.46 (0.05)	0.27 (0.06)	-0.41 (0.07)	0.13 (0.09)
Lactation 2	PROD	0.95 (0.02)	-0.06 (0.08)		-0.31 (0.05)	0.99 (0.02)	-0.32 (0.07)
	PERS	0.12 (0.06)	0.33 (0.08)	0.09 (0.07)		-0.32 (0.07)	0.87 (0.05)
Lactation 3	PROD	0.89 (0.03)	-0.12 (0.08)	0.97 (0.02)	-0.06 (0.08)		-0.38 (0.08)
	PERS	0.09 (0.07)	0.32 (0.09)	0.12 (0.07)	0.97 (0.03)	-0.03 (0.08)	

Genetic correlations between lactation 1 and later lactations reported by De Roos et al. (2004; first 3 lactations) for milk production traits (Milk, Fat and Protein) were also positive and of moderate, ranging from 0.46 to 0.52. These last two studies found much lower genetic correlations between later lactations than we did (0.98), but their measure of persistency was very different (decrease between 60 and 240/245 DIM).

Genetic correlations of the aRRM between average milk production (Milk, Table 3.3) of a lactation and the persistency of later lactations (i.e., lactations 1 vs 2, 1 vs 3 and 2 vs 3) were weak and often close to zero (-0.32 to 0.12) for both PAST and TMR production systems. This was mostly true for the other production traits as well (Tables S3.6 and S3.7 for Fat and Protein, respectively).

Genetic correlations of the aRRM between the persistency of a lactation and the average production (Milk, Fat and Protein) of later lactations were negative, weak and often close to zero in PAST (-0.29 to -0.01). Corresponding genetic correlations in TMR were also negative but ranged from weak to medium strength (-0.48 to -0.19). (Table 3.3 for Milk and Tables S3.6 and S3.7 for Fat and Protein, respectively).

3.3.8 Production Traits: Differences in Genetic Parameters in the TMR Production System

Differences in the pattern of heritability estimates in the TMR vs PAST production systems are presented in Figures 3.3 (for Milk), and S3.12 in the supplementary material (for Protein), respectively. TMR estimates from the aRRM were close to equal than the saFRM during the second half of the lactation for lactation 1 (average of 0.20 vs 0.16 during the last 145 DIM for Milk), and also during the late lactation period in later lactations (average of 0.21 vs 0.15 for lactation 2, and 0.18 vs 0.13 for lactation 3, during the last 95 DIM for Milk, respectively). In estimating 295-day EBVs, the emphasis will thus be slightly more on the second half of the lactation (lactation 1) and late lactation period (later lactations) with higher accuracy of prediction during these periods than with the saFRM. Heritability estimates (aRRM) were always highest during late lactation in TMR vs PAST. The main cause of the different heritability estimates between TMR vs PAST is a much greater increase in AG variance after early lactation in TMR. This was especially true for the later lactations (Figures 3.4b, d and f for Milk, and Figures S3.18a to S3.18c and S3.19a to S3.19c for Fat and Protein, respectively, in the supplementary material). The aRRM heritability patterns for Fat in TMR (Figure S3.10 in the supplementary material) were close to the same than for Milk and Protein. The pattern of the aRRM TMR heritability estimates for production traits coincides with several other studies which also reported an increase in the AG variance and heritability estimates over the lactation (Zavadilova et al., 2005; Muir et al., 2007; Aliloo et al., 2014).

When comparing the overall level of the heritability estimates for all production traits, those of TMR were lower than PAST estimates throughout all 3 lactations when using both models (saFRM or aRRM). This is caused by the higher PE (Figures S3.9, S3.11 and S3.13 in the supplementary material for Milk, Fat and Protein, respectively) and residual variances in TMR relative to the AG variance (Figures 3.4a to 3.4f for Milk, and Figures S3.18a to S3.18c and S3.19a to S3.19c in the supplementary material for Fat and Protein, respectively). Our results show that in TMR, milk and protein production is mainly influenced by the PE and residual effects (temporary environmental component), especially at the beginning and end of the lactations, and overall, much more so than in PAST. Large herds that utilize a TMR production system mainly feed their cows according to their individual phenotypic milk yield (level) quite possibly more so in the early lactation period, including peak production and during at least mid-lactation (considering the results of this study). This explanation makes sense as it is recommended in SA that higher producing cows are fed at a higher level than low producing cows in order to increase profitability, and that cows should be fed according to the lactation curve (The Dairy Farming Handbook, 2017). These recommendations also show the importance of high-quality roughage together with high energy concentrates, as well as the correct (increasing) feeding level during the first 56 to 70 DIM, specific for each cow according to peak milk yield (The Dairy Farming Handbook, 2017). Individual cow feeding management information (within herd) is not sent to the genetic evaluation service provider(s) and thus cannot be considered in the genetic evaluation of the breed. Consequently, the model used cannot accurately distinguish if production level is the result of the AG effect or the individual (phenotypic group) nutrition level. Other than with TMR, cows in the PAST production system have an equal chance to be exposed to the same nutritional level.

The amount of the total AG variance explained by the first two eigenvectors (lactation 1) for production traits (Table S1) using the aRRM in TMR were lower than for PAST (87.6% to 88.9% vs 90.8% to 94%), being fairly equal for lactation 2 (third and fourth eigenvector), but still high (98.8% to 99.2% vs 99.5% to 99.7%).

3.3.9 Milk Composition Traits

Heritability estimates for Fat% and Protein% in both production systems (aRRM) were higher compared to the saFRM from approximately the middle to late lactation period and onwards (Figures S3.14 and S3.16 in the supplementary material). These results imply that with the calculation of aRRM total lactation i.e., 295-day EBVs, TD records during the first half of the lactation will be weighted more conservatively than those during the second half in comparison to those from the saFRM. Thus, selection progress for increased total milk production per lactation will be faster because of later production than early production (in the lactation). The aRRM heritability patterns over DIM (lactation 1) are in accordance with Bastin et al. (2011) and to a great degree with Mohammadi et al. (2014).

Regarding the amount of total AG variance explained by the first two eigenvectors (aRRM, lactation 1) for the two milk composition traits, values were mostly higher (range 91.5% to 93.5%) than for the milk production traits in both production systems, especially TMR (Table S1).

In contrast to Milk, genetic correlations between average milk composition traits and their corresponding persistency estimates within each of the first 3 lactations were consistently negative and of medium strength over both production systems (Table S3.8 and S3.9 in the supplementary material).

In accordance with milk production traits, genetic correlations between persistencies in the 3 lactations for Protein% (Table S3.9) were also positive throughout, but generally stronger in both PAST and TMR, ranging from 0.74 to 0.85 between lactation 1 and later lactations. Genetic correlations between later lactations are higher (0.88 and 0.95 for PAST and TMR, respectively). Druet et al. (2005) reported more moderate genetic correlations between lactations for protein percentage (0.56 and 0.66 between lactations 1 and 2, and lactations 1 and 3, respectively) with a genetic correlation of 0.75 between later lactations. Persistency genetic correlations for Fat% between lactation 1 and later lactations were slightly lower than for Protein%, ranging from 0.64 to 0.75 while being 0.65 and 0.83 between later lactations for PAST and TMR, respectively. Druet et al. (2005) reported higher genetic correlations of 0.90 (between lactations 1 and 2) and 0.83 (between lactations 1 and 3), and 0.97 between later lactations for fat percentage.

3.4 Conclusions

The partitioning of fixed effects into a sum of separate curves over the lactation (i.e., curves as used in an aRRM, based on 6-knot natural splines instead of the Wilmink function) explains the fixed effects significantly better and led to a better model fit. Including the AG and PE effects as random over the lactations (aRRM) and not fixed (FRM) led to a further better fit of the model as well as lower residual variances than the current saFRM. The aRRM showed that the AG and PE random effects should not be considered constant over the lactation with both effects explaining more of the phenotypic variance at the beginning and end of the lactation. The extension of the aRRM with the addition of a measure of persistency as a genetic component allowed for a more accurate estimation of the AG effect. It is an important, extra source of information which increases the accuracy of genetic prediction. Additionally, the genetic prediction for persistency allows for selection over the lactation as a trait of economic importance. Increasing accuracy of genetic prediction using the aRRM instead of the saFRM is also possible due to higher genetic correlations between average productions, especially between lactation 1 and later lactations. The implementation of the alternative random regression model for genetic evaluations of SA Holsteins is recommended.

The overall lower heritability estimates in this study compared to some other studies might be explained by a less rigorous on-farm recording system without an external supervisor who visit dairy farms on a monthly basis, as it is for instance the case in most European countries. A possible larger amount of incorrect pedigree information in South African Holstein, possibly because of much larger herd sizes compared to European countries, might also play an important role.

The quite distinct genetic parameters of the TMR compared to the PAST production system denote a genotype by environment interaction between the two production systems. Disregarding such an interaction decreases genetic gain and biases EBVs to a certain degree, the magnitude, possible origin, and ways to mitigate the interaction must be studied further using a test-day, multi-lactation, and random regression model.

CHAPTER 4

Genotype by environment interaction under two production systems utilizing a random regression model for milk production traits in multiparous South African Holstein dairy cattle[2]

4.1 Introduction

Selection response under different “environments” in the presence of genotype by environment interaction (GxE) can vary substantially. This can be attributed to different genes that may be necessary to adapt and thrive in different environments (Falconer, 1960). South Africa (SA) is a country with diverse environments, classified into 13 distinct climate zones according to the Köppen-Geiger classification method (CSIR, 2023). Despite the varying climatic conditions and regions in SA, Van Niekerk et al. (2006) found no GxE interaction in SA Jersey cattle for milk production (Milk) in various groups based on different production systems and climatic areas, using a bivariate (lactation-based) animal model. However, Theron and Mostert (2009) compared SA dairy cows on a concentrate, mixed or pasture (PAST) ration and showed significant phenotypic differences in production between these feeding systems and speculated about a possible GxE interaction. Various other studies reported mostly minor to no GxE interactions for production traits between some type of control (conventional or confined environment) vs grazing herds (Weigel et al., 1999; Boettcher et al., 2003; Kearney et al., 2004; Fahey et al., 2007). It is quite possible that the lack of GxE interactions could have been that the differing environments in each study did not vary enough, as the magnitude of GxE between environments is influenced by, amongst other factors the amount of variation between the climatic areas or management systems (Bourdon, 2014).

All these studies used whole lactation (mostly 305-day) records, which prevents the differentiation between production-level and -persistency as shown by Van der Werf et al. (1998), Olori et al., (1999a) and Jamrozik et al. (2002). Previously, Sölkner and Fuchs (1987) evaluated various persistency measures using test-day (TD) milk yield records. They showed that more persistent cows could be more efficient (and/or profitable) as they require less concentrate to produce the same amount of milk than less persistent cows. According to Gengler (1995) the concentrate reduction amounts to 3%, with a further 7% reduction in health and reproduction cost. Dekkers et al. (1998) indicated an economic value for persistency of 3.4% relative to the corresponding value of 305-day production at a calving interval of 12.4 months, with a threefold increase in the economic value

---

[2] Submitted. Van Niekerk et al. Genet. Sel. Evol. (2025)

of persistency when the average calving interval was 13 months. They stated that taking health and reproductive costs into account, the economic importance of persistency will increase further.

Banga et al. (2009) compiled economic values for SA Holstein and Jersey cattle of different payment systems that included milk volume, fat and protein yield as production traits, as well as live weight, longevity, calving interval and somatic cell score (SCS). However, the effect of lactation persistency per se was not part of the study. Banga et al. (2014) emphasized the importance of protein yield and (milk) volume, longevity and SCS as traits to be included into breeding objectives for SA Holstein cattle. There exists a genetic relationship between all these traits and persistency. A persistency index for SA dairy breeds based on the Canadian Persistency Index was implemented by Mostert et al. (2008).

A preliminary study of GxE interaction in SA Holstein was conducted for 305-day production and age at first calving between total mixed ration (TMR) and PAST feeding system herds with no such interaction for milk production (Neser et al., 2014). Neser et al. (2014) did, however, highlight the possible existence of a GxE interaction for production traits due to the large difference between the heritabilities for the same trait in the two feeding systems. Van Niekerk et al. (2023) confirmed this result using a TD, multi-lactation, and random regression (RR) model with differing heritability levels and patterns over the lactation for production traits between TMR and PAST. Ducrocq et al. (2022) also concluded that an important GxE interaction exists in SA Holstein cattle which could result in reranking of sires between different feeding systems (TMR and PAST). Their study showed that the average annual rainfall of the region in which the herds are kept are a determinant of production (feeding) system used. Ducrocq et al. (2022) emphasized that GxE interaction research in SA dairy cattle should be expanded by using more complex genetic evaluation models.

The objective of this study was to investigate the previous indications of a possible GxE in milk production traits under a TMR versus a PAST production system, specifically using a TD, multi-lactation model based on the alternative random regression (aRR) developed in Van Niekerk et al. (2023). This model can give new insights into the existence and magnitude of a possible GxE as it includes a genetic average-production (level; PROD) and -persistency (slope; PERS) effect. TMR and PAST productions were never analyzed together as different traits to investigate GxE interaction in SA before. The second objective was to consider a potential reranking of sires in the different production systems.

4.2 Material and methods

4.2.1 Dataset

TD (n = 2 167 022) and pedigree records from all herds (140) participating in milk recording were made available by the South African Holstein Cattle Breeders' Society for the study. Only herds that participated in milk recording for 3 consecutive years were included in the final dataset. TD records for milk, fat and protein production (Milk, Fat and Protein; kg) of cows with sufficient first lactation records, milked 2 or 3 times a day (number of milkings) for the first 3 lactations (Lac1, Lac2 and Lac3, with Lac2 and Lac3 demarcated as "later lactations") were used. Cows without calving dates and less than 7 TD records were excluded from the dataset. Each lactation consisted of TD records that ranged from 6 to 300 days in milk (DIM). Lactations where the first TD record were not within 65 DIM, and where consecutive TD records were shorter than 21 days, or longer than 90 days, were discarded.

Table 4.1 Descriptive statistics

	Lac1		**Lac2**		**Lac3**	
Item	**TMR**	**PAST**	**TMR**	**PAST**	**TMR**	**PAST**
Avg. Milk (kg)	31.4 (8.3)	21.1 (5.0)	35.0 (10.6)	24.8 (6.8)	35.6 (10.9)	26.5 (7.6)
Avg. Fat (kg)	1.1 (0.4)	0.7 (0.2)	1.3 (0.5)	0.9 (0.3)	1.3 (0.5)	0.9 (0.3)
Avg. Protein (kg)	1.0 (0.3)	0.7 (0.2)	1.1 (0.3)	0.8 (0.2)	1.1 (0.3)	0.8 (0.2)
Records^1	272 318	82 591	137 944	52 446	63 712	33 365
Cows^1	35 011	10 548	17 716	6 703	8 236	4 295
Records^2 / Cow	7.8	7.9	7.8	7.9	7.8	7.8
Records^2 / Group	105	49	58	35	31	24
Sires^1 Total	1022		998		900	
Sires^1 / System	972	570	929	513	814	455
Sires 5^1 / System	933	424	778	331	475	249
Sires 10^1 / System	853	303	507	196	266	124
Shared sires 5^1	335		226		151	
Daughters^2 / Sire	36	19	19	13	10	9

Standard deviations in parentheses; Laci = Lactation i; Avg. = Average; Milk = Milk production; Fat = Fat production; Protein = Protein production; Superscript 1 = Number of; Superscript 2 = Average number of; Group = Contemporary group defined by the Herd x Test-day x Number of milkings interaction effect; Sires Total = Total number of sires irrespective of production system; Sires / System = Total number of sires in the corresponding production system; Sires 5 / System = Sires with a minimum of 5 daughters in the corresponding production system; Sires 10 / System = Sires with a minimum of 10 daughters in the corresponding production system; Shared sires 5 = Sires with a minimum of 5 daughters in the pasture and total mixed ration production systems each.

Contemporary groups (Herd x TD x Number of milkings) had to consist of at least 5 cows. Calving ages of cows (in months) were grouped into 8 calving age classes as follow: Class 1, 20 to 23; Class 2, 24 to 27; Class 3, 28 to 31; Class 4, 32 to 35; Class 5, 36 to 41; Class 6, 42 to 47; Class 7, 48 to 55; Class 8, 56 to 69. Calving month variable were formed and followed the natural 12 months of the year, while calving season was also used and followed the natural seasons in South Africa (Summer, Season 1: December, January, and February; Autumn, Season 2: March, April and May; Winter, Season 3: June, July and August; Spring, Season 4: September, October and November).

The calving year ranged from 1988 to 2016. Previous calving interval classes were categorized and ranged from 312 to 730 days (Class 1: 312 to 353; Class 2: 354 to 395; Class 3: 396 to 437; Class 4: 438 to 479; Class 5: 480 to 521; Class 6: 522 to 563; Class 7: 564 to 730; Class 8: no previous calving interval – first calf cows). Sires had to have at least 10 daughters in the dataset, irrespective of whether daughters produced in TMR or PAST. Cows with unknown sires were discarded.

The final dataset consisted of 26 herds with a total of 473 974 TD records utilizing a TMR production system and 11 herds with a total of 168 402 TD records utilizing a PAST production system (Table 4.1). The total number of sires in the TD dataset was 1022 (≥ 10 daughters in total). By far the largest proportion (53%) of these sires were from North America (United States and Canada). This proportion varied little between TMR and PAST.

Table 4.1 shows the total number of sires irrespective of production system, total number of sires per production system, the number of sires with at least 5 daughters, as well as at least 10 daughters in TMR and PAST. There were 520 shared sires between TMR and PAST with an average of 44 and 18 daughters in the two production systems, respectively. Out of these, 335 (64%) had at least 5 daughters in both TMR (average of 53) and PAST (average of 26) for Lac1. These sires are termed Shared sires 5[1] (Table 4.1).

4.2.2 Model for analysis

Milk, Fat and Protein were analyzed separately, using WOMBAT (Meyer, 2007). A TD, multi-lactation (first 3 lactations) RR model was used to analyse each of the traits. The residual variance was considered as heterogeneous and TD records were assigned to 9 residual effect classes (CLASS) of 30 days each, and 1 class of 25 days at the end of the lactation (10 classes in total; 295-day lactation). The model is described in Table 4.2.

Table 4.2 Description of the random regression (RR) model used to analyse milk, fat and protein production

Fixed effects (class)	**Fixed regressions on DIM**	**Random regressions on DIM**
Herd x TD x Number of milkings per day	Herd	Animal Genetic (additive; AG)
	Calving year x Lactation	Permanent Environmental (PE)
Herd x Calving year x Calving season x Lactation	Calving month	Heterogeneous residual variance (10 classes)
	Previous calving interval	
	Calving age	

Class = classified; TD = Test-day; DIM = Days in milk

Including the Herd x Calving year x Calving season x Lactation fixed effect in the model will adjust records for stage of lactation where the cows in a group calved within the same ~90-day window. Regressions on DIM within various fixed effects, using natural splines with 6 knots (DIM 6, 20, 50, 135, 245, 300) as the regression functions were used for lactation (Lac) 1, 2 or 3. AG and PE effects were assumed normally distributed and were fitted as random regression effects. The AG and PE effects described each lactation separately, and each effect consisted of a constant as well as a regression slope multiplied by current DIM. The constant term describes the (additive) genetic-average-production (PROD) effect, while the regression slope describes the genetic-average-persistency (PERS) effect over the whole lactation (DIM 6 to 300). Thus, these AG effects result in estimated breeding values [EBV(s)] for both components (i.e., EBV-PROD and EBV-PERS) for each animal. The sum of these two components are thus a total production EBV i.e., EBV-TOTAL. These two elements are the first and second principal components for the first lactation, respectively, resulting in 6 principal components over the 3 lactations for each of the traits.

Using the model in Table 4.2, TD records obtained either in a TMR or a PAST production system, were treated as two separate traits, and was analysed together in a multiple trait analysis resulting in 12 principal components (TMR-PAST12 model). These principal components were derived in the same manner from the G matrix as in Van Niekerk et al. (2023), where we fitted standardized DIM directly as linear regressions. This analysis was followed by a reduced rank model resulting in 8 principal components (TMR-PAST model) and represent the two genetic components (PROD and PERS) of each of the first 3 lactations in TMR, plus the two genetic components (PROD and PERS) of Lac1 in PAST (all traits). The estimates from the TMR-PAST model are considered in the results. The 8 principal components were based on the largest eigenvalues after decomposition of the 12 by 12 (co)variance matrix of the AG effect (full rank).

The Bayesian Information Criterion (BIC = -2logL + log(N)*k, where **logL** denote the value of the maximum log likelihood of the model, **N** is the sample size of the training dataset, and **k** is the number of parameters in the model) were used to assess the differences in goodness of fit between the different analyses.

4.3 Results and discussion

4.3.1 Goodness of fit: Production systems analysed as different traits

The TMR-PAST model fitted the data best for all traits. This reduced rank model showed a significant improvement of 266, 359 and 266 in the BIC of Milk, Fat and Protein, respectively when compared to the TMR-PAST12 full rank model. This was because of the reduced number of effects that had to be estimated. Consequently, using Milk as example, the Central Processing Unit (CPU) time of the reduced rank model decreased by ~100 hours, with faster and easier convergence (within ~55 hours), compared to the full rank model. Eigenvalues showed that 99.8% of the total genetic variance for Milk, Fat, and 100% for Protein is explained by the first 8 principal components used in the TMR-PAST model.

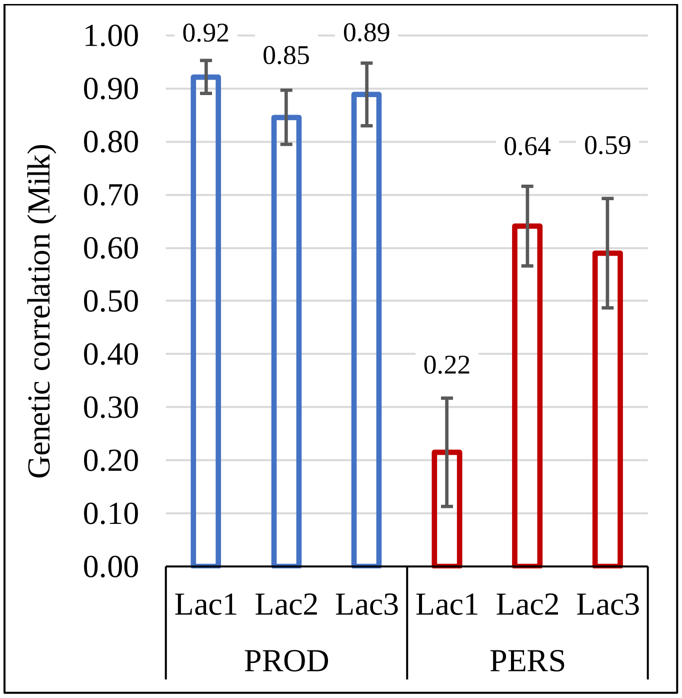
4.3.2 Genetic correlations between TMR and PAST for Milk production

The low to moderate genetic correlations (Figure 4.1; 0.22, 0.64 and 0.59) obtained for PERS between the production systems indicate that PERS in a TMR system is a different trait as PERS in a PAST system. The genetic correlations are significantly different (weaker) from 0.80 (Robertson, 1959), showing substantial GxE interaction between TMR and PAST for persistency. These results indicate that different genes are probably at play in the different production systems and that selection in one of the production systems could thus lead to a sub-optimal genetic trend for the trait in the other production system.

Contrary to PERS, the genetic correlations (Milk) between the PROD effects (Figure 4.1) were strong (0.92, 0.85 and 0.89), indicating that the same trait is measured in the two productions systems. Both the correlations in Lac1 and Lac2 differ significantly from 1 ($P<0.05$), but not for Lac3. Corresponding results for PROD and PERS between TMR and PAST in Fat and Protein were mostly similar (Figures S4.1 and S4.2, supplementary material).

Similar studies using TD, multi-lactation RR models, where records were obtained from different production systems or feeding regimes, and where both persistency (PERS) and production (PROD) were included in the study, are relatively scarce within countries. Zwald et al. (2001) concluded that herds and countries can be grouped into different production systems according to various parameters (peak milk yield, days to peak milk yield, persistency, percentage of North American Holstein genes, standard deviation of milk yield, sire predicted transmitting ability and fat:protein (%) ratio).

Figure 4.1 Genetic correlations (standard errors) between the total mixed ration and pasture production systems



Milk = Milk production; PROD = Average-production (blue); PERS = Average-persistency (red); Laci = Lactation i

Compared to Zwald et al. (2001), milk production in SA using Holsteins resemble, to a greater extent, countries that use more concentrates (United States, Israel and Canada) in contrast to the rotational grazing systems used in New Zealand and Ireland, but probably also Estonia and Czech Republic. Zwald et al. (2003) argued that different countries can be treated as different production environments. Jamrozik et al. (2002) indicated a significant GxE interaction between New Zealand and countries with more intensive management systems (Italy and Canada) for total (milk) yield (r_g 0.66 and 0.72, respectively) and persistency (defined similarly in our study; r_g 0.16 and 0.28, respectively), as well as between all 4 countries (Australia as the 4th) in the study for persistency (r_g 0.16 to 0.55). Hammami et al. (2008b) also showed significant GxE between Luxembourg (high-input system) and Tunisian Holsteins (low- to medium-input systems) for 305-day milk yield (r_g 0.60) and persistency (r_g 0.36; defined as the difference between DIM 80 and 280). Using a TD, single-lactation RR model over different production systems (i.e., low, medium, and high) based on the herds' average 305-day milk yield, Wahinya et al.

(27; multi-breed cows in Kenya) supported the results in the current study, with milk yield genetic correlations varying from 0.04 to 0.54 over different DIM. In a bivariate analysis (confined vs grazing cows) using 305-day milk yield records, Ramírez-Valverde et al. (2010) obtained a genetic correlation of 0.76 for Jersey cattle in Mexico, indicating a possible GxE interaction. Studying factors that influence lactation curves in different herds, Caccamo et al. (2010) concluded that different crude protein and dry matter content in diets, and more specifically their interaction influenced persistency of milk production between herds.

Using experimental data, other studies (Kolver et al., 2002; Horan et al., 2005; Fulkerson et al., 2008) also found significant genotype by diet, feeding systems and feeding level interactions, respectively, for milk yield in dairy cattle.

In contrast to the current study, Weigel et al. (1999) and Boettcher et al. (2003) found negligible differences between their applicable systems (r_g close to 0.90 or greater). However, these studies used 305-day lactation records and not TD, RR models, resulting in an inability to independently distinguish between the genetic persistency vs the genetic (level of) production effects. Most likely due to this reason, Neser et al. (2014) also did not find a significant GxE interaction in a previous preliminary SA Holstein study of milk production in TMR vs PAST ($r_g = 0.90 \pm 0.03$). The same reasoning could likely explain the high genetic correlations ($r_g \geq 0.88$) between TMR and PAST, as well as between other groups clustered according to genetic levels, management practices, herd size, geographical location, and environmental influences in SA Jersey dairy cattle (Van Niekerk et al., 2006).

According to Slagboom et al. (2021), these weak to moderate genetic correlations, specifically for persistency from the above-mentioned studies (Jamrozik et al., 2002, and Hammami et al., 2008b), as well as in the current study may very well require different breeding goals. These authors showed that if genetic correlations between production systems, a mainstream and an organic niche production environment in their case, are below 0.6, it is better to consider GxE interaction in genetic evaluations.

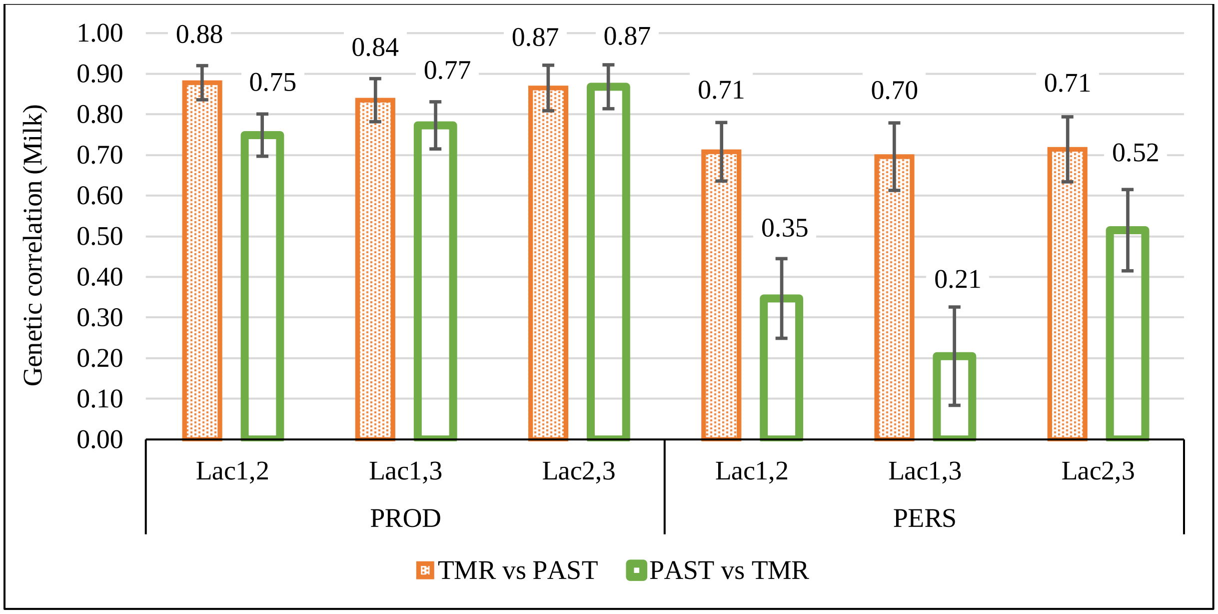
Genetic correlations between TMR and PAST (and vice versa), and at the same time between PROD and PERS were weak and varied from -0.29 to 0.27, often times not being significantly different from zero (Tables S4.1 to S4.3 for Milk, Fat and Protein, respectively, supplementary material). Correspondingly, within-TMR and -PAST genetic correlations between PROD and PERS were also weak and ranged from -0.41 to 0.14, also regularly not being significantly different from zero. Thus, independent genetic improvement of PROD from PERS and vice versa should be easily attainable.

4.3.3 Genetic correlations between lactations for Milk production

Between lactations PERS genetic correlations (Milk) for PAST Lac1 vs TMR later lactations (0.35 and 0.21, respectively; Figure 4.2; Figures S4.3 and S4.4 for Fat and Protein, respectively, supplementary material) were weak and significantly more so compared to TMR Lac1 vs PAST later lactations (0.71 and 0.70, respectively). These TMR Lac1 vs PAST later lactations genetic correlations show that TMR Lac1 is a better indicator of PAST later lactations than is the case within PAST (0.34 and 0.37, respectively; Table S3, supplementary material).

The higher genetic correlation between TMR Lac1 and PAST later lactations can be an indication of TMR first parity cows that are anatomically and physiologically more developed and thus more similar to later parity cows in PAST, rather than to their first parity PAST counterparts. The latter might be due to possible differing rearing environments between TMR and PAST systems.

Figure 4.2 Genetic correlations (standard errors) between lactations for the total mixed ration vs pasture production systems



Milk = Milk production; PROD = Average-production; PERS = Average-persistency; Laci,j = Lactation i,j; TMR = Total Mixed Ration production system (orange); PAST = Pasture production system (green)

MacDonald et al. (2007; Holsteins) concluded that North American heifers (closer to TMR) have a greater mature weight and age at puberty than those from New Zealand (closer to PAST), which possibly suggests that pasture conditions limit their growth rate as they are unable to meet their metabolizable energy requirements.

The complete genetic correlations matrix between- and within-TMR and -PAST for Fat, Protein and Milk production are shown in Table S4.1 to S4.3, respectively in the supplementary material.

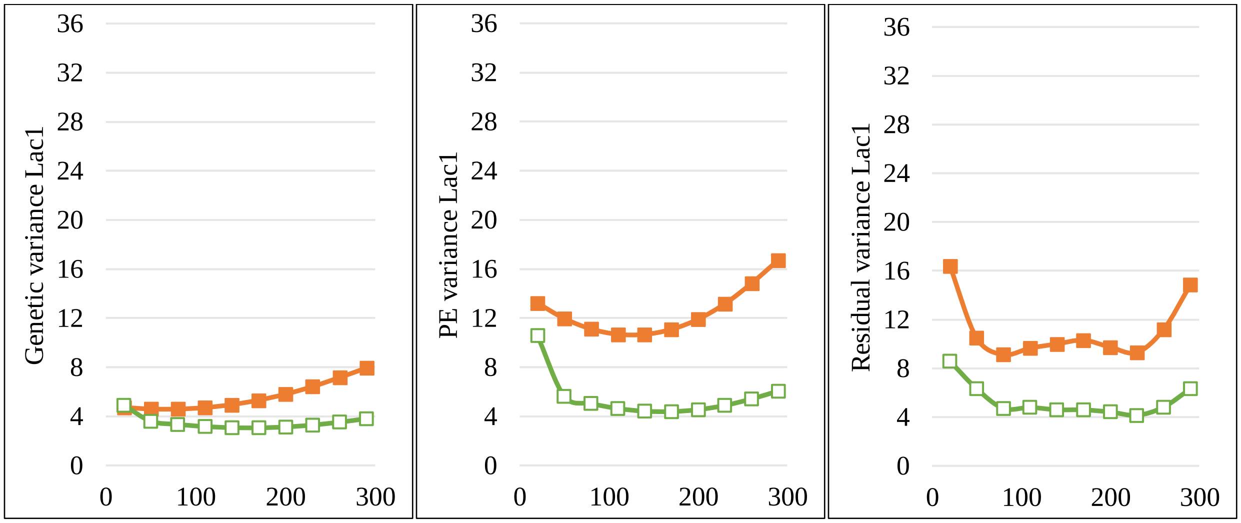
Important to note, the majority (91%) of PAST sires also have daughters (with greater number on average) in TMR, than in PAST. On the other hand, half (53%) of TMR sires also have daughters (with on average a smaller number) in PAST than in TMR.

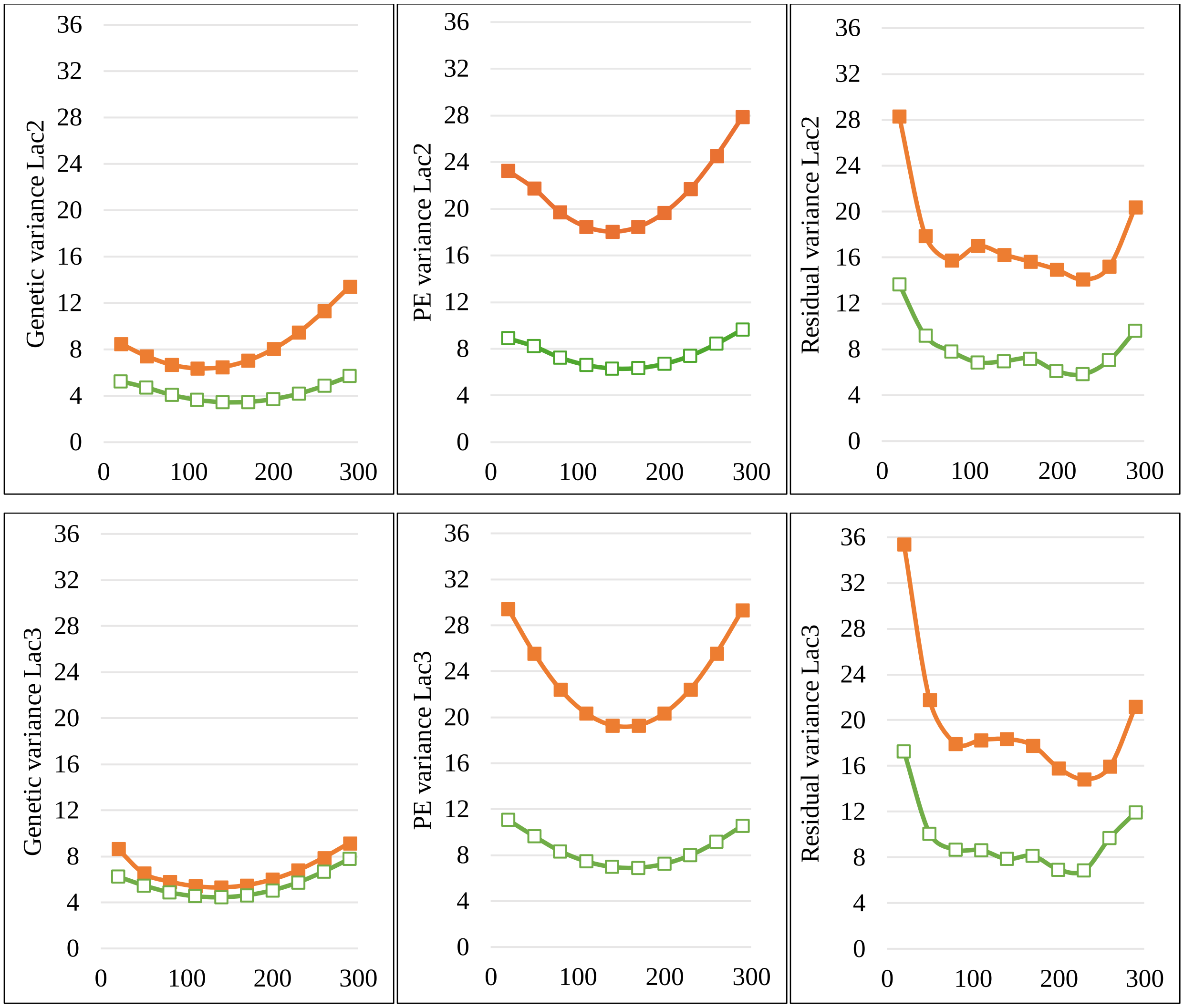
4.3.4 Genetic and other variances for Milk production

The genetic, permanent environmental and residual variance components of TMR and PAST over DIM 6 to 300 in each of the first 3 lactations for Milk are presented in Figures 4.3, 4.4 and 4.5, respectively. Corresponding variance components for Fat and Protein are presented in Figures S4.5 to S4.10 (supplementary material). TMR genetic variances (Milk) were higher than corresponding PAST genetic variances for all lactations, with the largest differences being in Lac1 and Lac2.

Jamrozik et al. (2002) reported differing random regression coefficients for Milk that described the curves of the genetic effects over the lactation between intensive management (Canada and Italy) and rotational grazing systems (Australia and New Zealand). The (additive) genetic variances for Milk over lactations also differed among low-, medium- and high-production systems with regards to their shape as shown by Wahinya et al. (2020). Moreover, our results are in accordance with Jamrozik et al. (2002) and Wahinya et al. (2020) who also showed the tendency of the genetic variance to be higher in the more optimal environment (TMR in our study). Similarly, Santos et al. (2019) showed increasing genetic variance as the environment improves. Dos Santos Daltro et al. (2022) argued that genetic differences are more detectable in more favourable environments as animals are more able to express their full genetic merit. In this regard, the environmental variances (PE and residual) indicate that TMR can be considered a more optimal production system compared to PAST for production.

Figures 4.3, 4.4 and 4.5 Genetic, permanent environmental and residual variances for Milk production over days in milk





y-axis values = kg^2; x-axis values = days in milk; PE = Permanent environmental; Orange blocks (solid) = Total Mixed Ration; Green blocks (open) = Pasture; Laci = Lactation i.

The differing genetic variances over the lactation, within and between TMR and PAST can be explained by cows' dissimilar genetic energy-balances and nutrient partition profiles, postpartum. Body condition score (BCS) can be used to predict total body energy content (lipid and protein) change over the lactation (Banos et al., 2005b). Loker et al. (2011) showed increasing genetic variance for BCS as lactation progressed, illustrating genetic differences in cows' capability to regulate the utilisation of body reserves, especially at the end of the lactation.

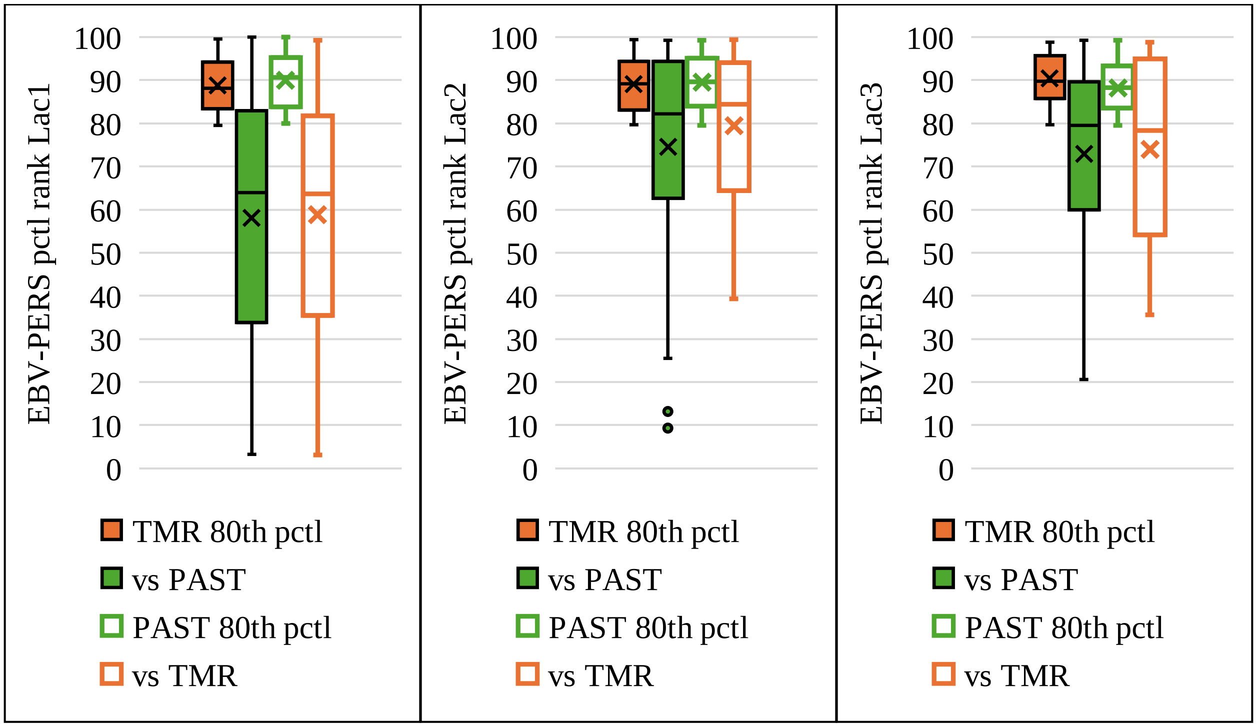
4.3.5 EBV rankings of sires between TMR and PAST for Milk production

Using Milk as example, the distribution of the percentile rank of sires (within lactation) with at least 5 daughters each in TMR and PAST (i.e., the better-connected sires), based on their EBV-PERS are presented in

Figure 4.6. The most significant within lactation reranking of sires in the 80th percentile in TMR vs their percentile rank in PAST take place in Lac1. The same is true contrary-wise, where major reranking is also the case for sires in the 80th percentile for PAST vs their rank in TMR. In these two instances roughly half of 80th percentile sires in TMR will be below the 60th percentile in PAST, with approximately a quarter of the sires below the ~37th percentile. Reranking in Lac2 and Lac3 will be less, but still significant. These results are in line with the EBV-PERS percentile rank correlations between TMR and PAST for all sires with at least 5 daughters each in TMR and PAST (shared sires 5; Table 4.1), which were 0.16, 0.65 and 0.54 for Milk Lac1 to Lac3, respectively (not shown). The equivalent correlations for Fat Lac1 to Lac3 were 0.34, 0.65 and 0.59, respectively and 0.29, 0.23 and 0.54, respectively for Protein.

Unlike EBV-PERS, reranking of "shared sires 5" (Table 4.1) for EBV-PROD should be minor with strong percentile rank correlations ranging from 0.94 to 0.99 for all lactations in Milk, Fat and Protein (not shown).

Figure 4.6 Reranking of sires between the total mixed ration and pasture production systems for Milk production



Laci = Lactation i; EBV-PERS = Estimated breeding value for average-persistency; TMR = Total Mixed Ration production system; PAST = Pasture production system; Solid blocks = Sires in the 80th percentile (pctl) rank for EBV-PERS in TMR (orange) vs their percentile rank distribution in PAST (green); Open blocks = Sires in the 80th percentile (pctl) rank for EBV-PERS in PAST (green) vs their percentile rank distribution in TMR (orange).

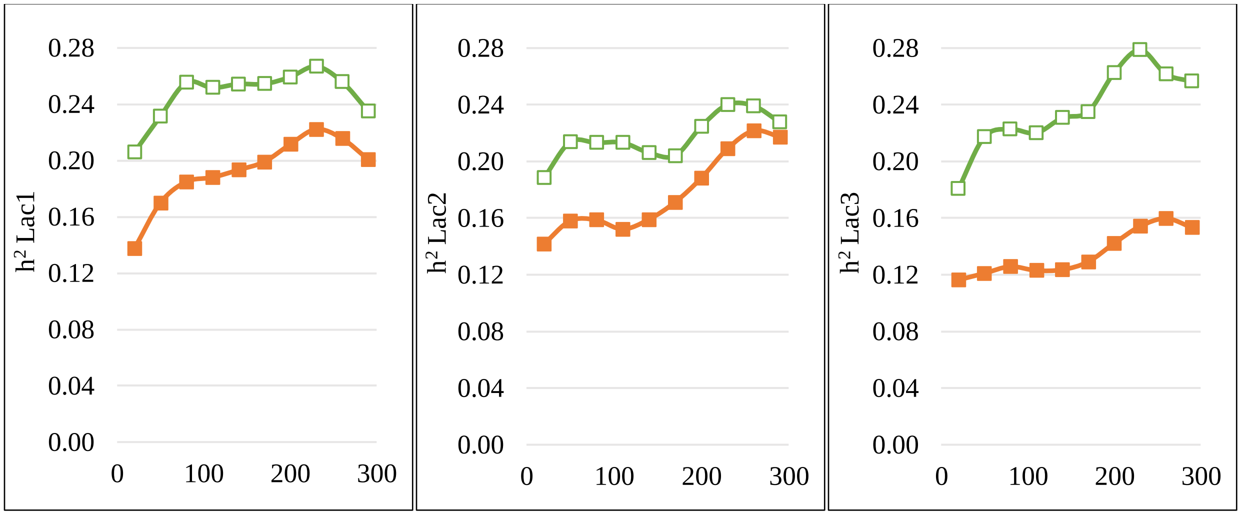
The EBV-TOTAL percentile rank correlations between TMR and PAST for all "shared sires 5" were 0.67, 0.75 and 0.75 for Milk Lac1 to Lac3, respectively (not shown). The corresponding correlations for Fat were 0.72, 0.69 and 0.76, respectively and 0.76, 0.41 and 0.74, respectively for Protein. Moderate reranking of sires between TMR and PAST will probably still take place based on EBV-TOTAL. The corresponding EBV-TOTAL correlations between TMR and PAST for all sires (n = 1022) were similar.

In their GxE study of production traits in confined vs grazing herds of primiparous US Holsteins, Kearney et al. (2004) obtained EBV rank correlations of 0.59, 0.63 and 0.66, also for sires with at least 5 daughters producing in both environments. Across herd management levels for Luxembourg and Tunisian Holsteins (primiparous cows), Hammami et al. (2009) showed EBV rank correlations for common sires to be weak, ranging from 0.19 to 0.41 for 305-day milk yield. In that study, the Luxembourgish dairy farms can be defined as mostly high-input production systems with maize silage, brewers grains feeds and concentrates, which is likely similar to TMR systems in the current study.

4.3.6 Heritability estimates, permanent environmental and residual variance ratios for Milk production

PAST heritability estimates were higher than TMR estimates throughout all lactations (Milk; Figure 4.7), more so during the first 100 days in Lac1 (average of 0.24 vs 0.17, respectively) and Lac2 (average of 0.21 vs 0.15, respectively), and during the last 95 days in Lac3 (average of 0.27 vs 0.15, respectively). The biggest difference between PAST and TMR heritability estimates were during Lac3 (average of 0.24 vs 0.14, respectively). Corresponding estimates for Fat and Protein are shown in Figures S4.11 and S4.12 in the supplementary material.

Figure 4.7 Heritability estimates over days in milk for Milk production

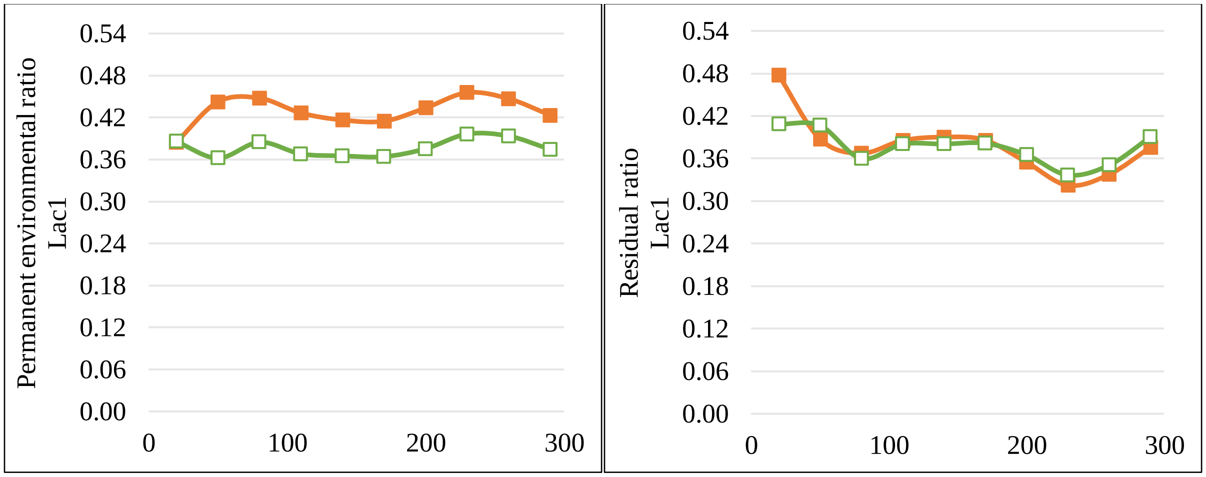


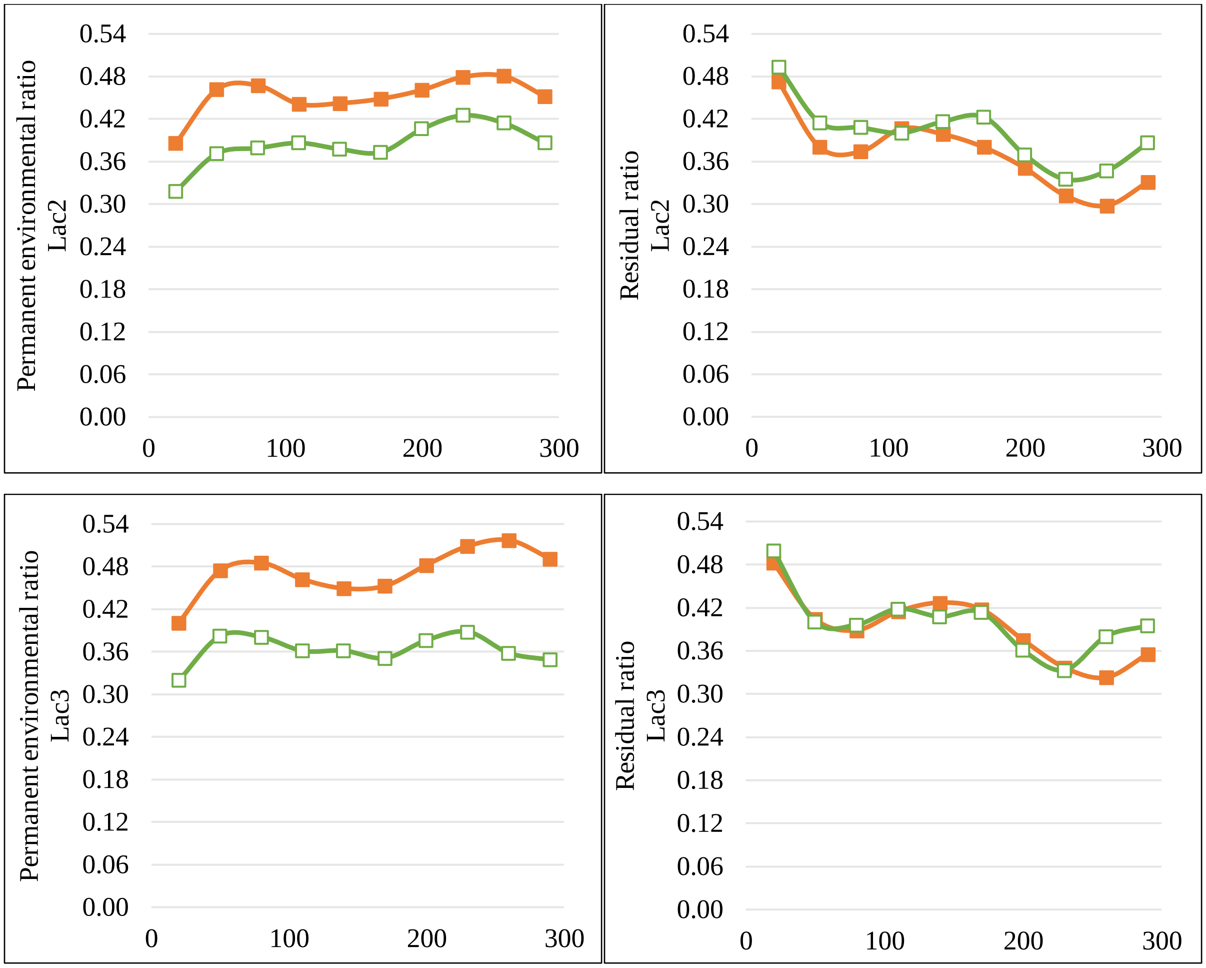
h^2 = heritability; x-axis values = days in milk; Laci = Lactation i; Orange blocks (solid) = Total Mixed Ration; Green blocks (open) = Pasture; Standard errors for heritability estimates ranged from 0.01 to 0.02.

Average heritability estimates for Milk, Protein and Fat in PAST and TMR were essentially the same as those reported by Van Niekerk et al. (2023) in separate analyses of PAST and TMR using the same model. PAST heritability estimates (0.25 for Lac1) for Milk were closest to estimates from various authors in primiparous dairy cows on a 305-day milk yield basis, with an estimate of 0.19 from Hammami et al. (2008b) and 0.15 from Ramírez-Valverde et al. (2010). These two authors reported heritability estimates of 0.42 for the high-input system (Hammami et al., 2008b; Luxembourg) and 0.27 for the "fed in confinement" system (Ramírez-Valverde et al., 2010). Overall, heritability estimates in this study are similar to Holsteins kept in a tropical environment (Meseret and Negussie, 2017) and somewhat higher than Hammami et al. (2008a) and Strabel and Jamrozik (2006).

The higher Milk (and Protein) heritability estimates in PAST compared to TMR were mostly because of the higher proportion of the total (phenotypic) variance that was explained by the PE variance (ratios; Figure 4.8) in TMR (on average) with a resulting drop in the genetic variance and are supported by the results in Chapter 3, section 3.3.8. This higher PE ratio tendency in TMR estimates is most likely because TMR cows are fed according to individual phenotypic milk yield, and quite understandably more so during early lactation, including peak lactation up to mid-lactation. Individual cow feeding information is not sent to the genetic evaluation service provider(s) and thus cannot be adjusted for through the fixed effects equations in BLUP. Accordingly, BLUP cannot accurately distinguish superior or inferior production as the result of the AG or the individual nutrition level effect. Other than with TMR, cows in PAST have an equal chance to be exposed the same nutritional level.

Figures 4.8 and 4.9 Permanent environmental and residual variance ratios over days in milk for Milk production





x-axis values = days in milk; Laci = Lactation i; Orange blocks (solid) = Total Mixed Ration; Green blocks (open) = Pasture; Standard errors for heritability estimates ranged from 0.01 to 0.02.

This can also be seen in Van Niekerk et al. (2023) with the same corresponding results with regards to the PE and residual variance (ratios; Figure 4.9 for the residual ratios in this study) as reasons for differing TMR and PAST heritability estimates. Ramírez-Valverde et al. (2010) also implied a higher PE ratio for confined vs grazing herds while Jamrozik et al. (2002) reported a higher PE variance for more intensive production systems (Canada and Italy) vs rational grazing systems (New Zealand and Australia).

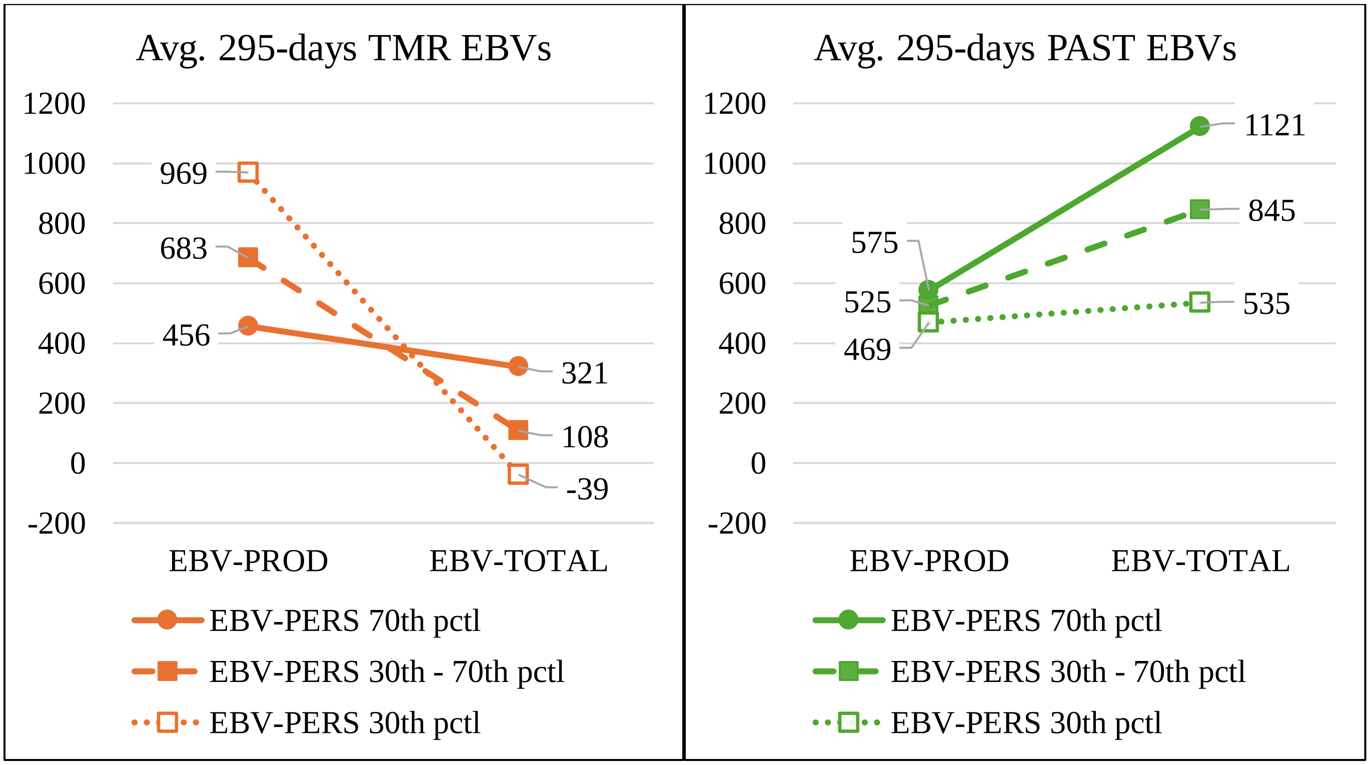
Different from Milk, the most important cause of the higher Fat PAST heritability estimates compared to TMR were on average the smaller residual variance ratios in PAST with a resulting increase in the heritability. In the same way, the smaller PE variance ratios in PAST also contributed to the genetic variance explaining more of the total variance.

The corresponding PE and residual variance ratios for Fat and Protein are shown in Figures S4.13 to S4.16, respectively in the supplementary material.

4.3.7 The genetic effect (EBVs) of persistency on total production (Milk)

The PERS effect on genetic total production for Milk in Lac1 are shown in Figure 4.10, using the average EBVs of three different sire groups. The groups only included sires with at least 5 daughters each in TMR and PAST (total of 335 bulls). The groups were differentiated as superior sires (70th percentile; pctl), average sires (30th to 70th pctl) and inferior sires (30th pctl) for TMR (left figure) and PAST (right figure). In the three graphs (Figure 4.10) each sire was allocated to the appropriate sire group based on its percentile rank for EBV-PERS. Each graph in Figure 4.10 shows the sire group average EBV-PROD vs the corresponding sire group EBV-PROD + EBV-PERS i.e., total production or EBV-TOTAL on a 295-days Milk basis. Thus, the slope of each graph is the average EBV-PERS of the specific sire group.

Figure 4.10 Effect of 295-days EBV-PERS (Milk) for superior, middle and inferior sire groups



Avg. = Average; y-axis values are in kilograms (kg Milk) on a 295-days EBV basis in lactation 1; EBV-PERS = Estimated breeding value for average-persistency; EBV-PROD = Estimated breeding value for average-production; EVB-TOTAL = Estimated breeding value for average-production plus -persistency i.e., EBV-TOTAL; Milk = Milk production; pctl = percentile; Sire groups (average EBVs): superior (solid lines, round markers) = sires in the 70th pctl for EBV-PERS; average (dashed lines, solid square markers) = sires in the 30th to 70th pctl for EBV-PERS; inferior (dotted lines, open squared markers) = sires in the 30th pctl for EBV-PERS; TMR = Total mixed ration production system; PAST = Pasture production system.

In developing breeding objectives for SA Holstein cattle, Banga et al. (2014) stated that protein yield was generally the most economically important trait, but also acknowledged that milk yield (volume), amongst other traits, was more important in some (payment system) instances. Persistency of production (genetic) is an integral part of milk volume produced, especially in South African Holsteins. Van Niekerk et al. (2023; supplementary material) reported that the persistency component (PERS) explained a more significant proportion of the total genetic variance in TMR, and even more so in PAST than some other Holstein populations, with eigenvalues (%) of 25%, 22% and 23% for Milk, Fat and Protein, respectively in TMR Lac1, and 28%, 25%, 30% for the same traits and lactation, respectively in PAST.

Figure 4.10 highlight the effect of genetic-average-persistency (EBV-PERS) on genetic-total-production (EBV-TOTAL) as a deviation from genetic-average-production (level) i.e., EBV-PROD, on a 295-days Milk basis, as a possible indication on the importance of persistency on production. Regarding the average EBV-PERS sire groups, their EBV-PERS resulted in EBV-TOTAL changes (decrease or increase) of -575 kg (-84%) and +320 kg (+38%) in relation to their EBV-PROD in TMR and PAST, respectively. The corresponding results for the inferior EBV-PERS sire group were -1008 kg (-104%) and +66 kg (+14%). Pertaining to the superior EBV-PERS sire groups, their EBV-PERS brought about EBV-TOTAL changes of -135 kg (-30%) and +546 kg (+95%) relative to their EBV-PROD in TMR and PAST, respectively. In TMR, although the average EBV-PROD of the inferior EBV-PERS sire group was the highest, because of their inferior genetic merit (EBV-PERS) for persistency, genetically production could not be maintained and their EBV-TOTAL over the lactation was the lowest. On the contrary, EBV-TOTAL was highest for the superior EBV-PERS sire group, notwithstanding that this sire group had the lowest average EBV-PROD. In PAST, the three EBV-PERS sire groups did not differ much in their average EBV-PROD and consequently the superior EBV-PERS sire group had the highest EBV-TOTAL.

4.4 Conclusions

Including test-day records of cows explicitly considering production system of their herd (total mixed ration or pasture management), as well as implementing a reduced rank model for the genetic evaluation of milk production traits fits the dataset best by a large margin.

Results showed that a significant genotype by environment interaction for production traits between total mixed ration and pasture production systems exist in the South African Holstein population for the genetic-average-persistency (of production) effect for all lactations (especially first lactation) and traits: all genetic correlations were significantly different from 1 and most from 0.80. These results show that there is major (first lactation) to intermediate (second and third lactation) reranking of sires between the two production systems for EBV-PERS. Using EBV-TOTAL moderate reranking of sires are likely to take place as all EBV-TOTAL percentile rank correlations are below 0.80 for all lactations in all traits. The effect of EBV-PROD (minor GxE) to compensate for the major GxE in EBV-PERS do not seem large enough to significantly lessen the reranking of sires on EBV-TOTAL.

Deriving from the weak to zero genetic correlations between PROD and PERS, publication of EBV-PROD separately from -PERS enable producers to be better positioned to select animals that produce milk in a “certain way” over the lactation. This will depend on their selection objectives. For instance, selecting cows that have a lower genetic production level, and not have the genetic merit to produce as much milk at peak yield, but to have a “flatter” lactation curve (superior genetic persistency) will lead to a cow herd that become pregnant earlier again, having more favourable negative energy balance (better body condition score), less health problems and utilizing less feed during the first 100 days in milk. In our study such a cow (lower average-production, higher average-persistency) are more likely to produce more milk in both TMR and PAST. After determining their economic importance and derived weights, both EBV-PROD and -PERS can be included in a selection index. This will simplify selection for cumulative production (over the lactation) for producers. EBV-PROD and -PERS can be published alongside the index and EBV-TOTAL.

Regarding genetic-average-persistency, a significant advantage of analysing the total mixed ration and pasture systems as different traits is the considerable increase in reliability to better predict and select for pasture production in later lactations because of the significantly higher genetic correlations with first lactation in total mixed ration, compared to weak within-pasture genetic evaluation estimates. This is because the pasture system is genetically very well represented in the total mixed ration system through shared sires, while the opposite is less true. In this way, because of the notably higher number of daughters of shared sires that are performance tested in total mixed ration, the genetic evaluation in this population benefits from increased information. Ideally more herds, especially PAST herds should be included in the study, together with bulls that have sufficient progeny in both production systems. Also, genetically average persistency (slope) has a significant influence on the total volume of milk produced over the lactation.

CHAPTER 5

Inclusion of average rainfall in genetic evaluation of SA Holsteins to mitigate genotype by environment interaction[3]

5.1 Introduction

Genotype by environment interaction has been shown to exist in South African (SA) Holstein dairy cattle in previous studies by Ducrocq et al. (unpublished data) and Van Niekerk et al. (2019, 2021). In the latter two studies, clear differences in genetic parameters between herds utilising a total mixed ration (TMR) or pasture (PAST) production system were shown. Specifically, genotype by environment interaction for persistency of production between the two production systems were indicated with genetic correlations for the first three lactations being below 0.64 throughout. This mean that animals whose progeny perform well in TMR will not necessary perform well under PAST conditions with regards to milk production traits. Selecting sires whose progeny will perform satisfactory or superior in both or one of the production systems will thus not be accurate, as the production system used by herds are not known and thus excluded in genetic evaluations of SA Holsteins (Interbull, 2020, 2022). Ducrocq et al. (2015) showed that the production system used by herds are dependent on the long-term-annual, average rainfall (AvgRain) of the herd’s location (nearest town) in SA. This study also showed that AvgRain had the largest effect on the genetic variability of milk production using a reaction norm model. The aim of this study was to investigate the effects of including AvgRain of the towns nearest to the applicable herds (which is known), as a proxy for the production system used, utilising a previously developed alternative random regression model (aRRM; Van Niekerk et al., 2019).

5.2 Materials & Methods

Full access to the complete pedigree and milk recording data that were used in this study were given by the South African (SA) Holstein Cattle Breeders’ Society.

5.2.1 Data

---

[3] Published. Van Niekerk et al., Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP) 2022. https://doi.org/10.3920/978-90-8686-940-4_807.

Test-day (TD) records of milk production (kg; Milk), from 1988 to 2016 were used for the analysis. The dataset consisted of the first three lactations of cows from 37 herds, of which 26 utilize a total mixed ration (TMR) and 11 a pasture (PAST) production system. Cows with sufficient (seven or more) TD records per lactation and completed their first lactation were used in the study. Sires with less than 10 daughters and contemporary groups with less than five cows were excluded from the dataset.

5.2.2 Statistical analyses

Restricted maximum likelihood was used to analyse the records utilising the WOMBAT software (Meyer, 2007). The analysis for Milk were multi-lactation where an alternative random regression model (aRRM), developed by Van Niekerk et al. (2019) were expanded and applied. This model was created as a better substitute to the fixed regression model used for genetic evaluations of SA Holsteins (Interbull, 2022). In the aRRM the animal-additive-genetic (AG) component (random effect) were defined as separate for each lactation and expressed as a function of days in milk (DIM), containing a constant plus a regression (standardized from -0.5 to 0.5) multiplied by current DIM. Consequently, each lactation is a combination of an (additive) genetic-average-production (GPr) and -persistency (GPe) (first and second eigenvectors, respectively) effect (6 eigenvectors in total over the first three lactations). In the analysis for this study a third term (regression) over each DIM was added for the AG effect. This term represents the genetic effect due to long-term-average, annual rainfall (GAR). Average, annual rainfall (AvgRain) was categorised in five classes as follow: Class 1: 413 to 492 mm; Class 2: 493 to 597 mm; Class 3: 598 to 672 mm; Class 4: 673 to 787 mm; Class 5: 788 to 1008 mm. Classes were defined according to the distribution of herds over AvgRain in the prepared dataset. The permanent environmental (PE) effect was included in the same way (GAR effect included) as the AG effect. Residual variance due to DIM were considered as heterogeneous over and between each lactation and partitioned in 10 classes per lactation. This aRRM, TD and multi-lactation model that included the GAR effect (GAR model) for Milk can be described as follows:

$$\mathbf{y} = \mathbf{Tt} + \mathbf{Ss} + \mathbf{Hh} + \mathbf{Ll} + \mathbf{Mm} + \mathbf{Aa} + \mathbf{Cc} + \mathbf{Qu} + \mathbf{Zp} + \mathbf{e} \qquad [1],$$

where **y** is a vector of 24-h test-day Milk observations for the lactations; **t** is a vector of fixed Herd x Test-day x Number of milkings interaction (contemporary group) effects; **s** is a vector of fixed Herd x Calving year x Calving season x Lactation number interaction effects. The following vectors of fixed regression coefficients are included in equation (1): **h** for Herd effects; **l** for Calving year x Lactation number interaction effects; **m** for Calving month effects; **a** for Calving age class effects (8 classes); **c** for previous calving interval class effects (8 classes). The random regression coefficients for the animal additive genetic and permanent environmental effects were **u** and **p** (two terms for each effect i.e., of order 1 for each effect), respectively and **e** is a vector of residual effects. The matrices **T**, **S**, **H**, **L**, **M**, **A**, **C**, **Q** and **Z** are incidence matrices that relate observations to their respective effects.

The Bayesian Information Criterion (BIC) was used to compare the goodness of fit of the GAR model (this analysis), to results from previous analyses by Van Niekerk et al. (2021), where the same dataset was used. In the previous study by Van Niekerk et al. (2021) a very similar aRRM were used in two analyses. These two analyses differed in the following way: (1) TD records from herds using a TMR and PAST production system were defined as separate traits and analysed together, without the GAR effect (TMR-PAST model), so to get accurate TMR and PAST production system genetic parameters; (2) TD records were undefined from a production system and without the GAR effect, so to get genetic parameters as would be the case in a 'national SA Holstein genetic evaluation' (nGE model).

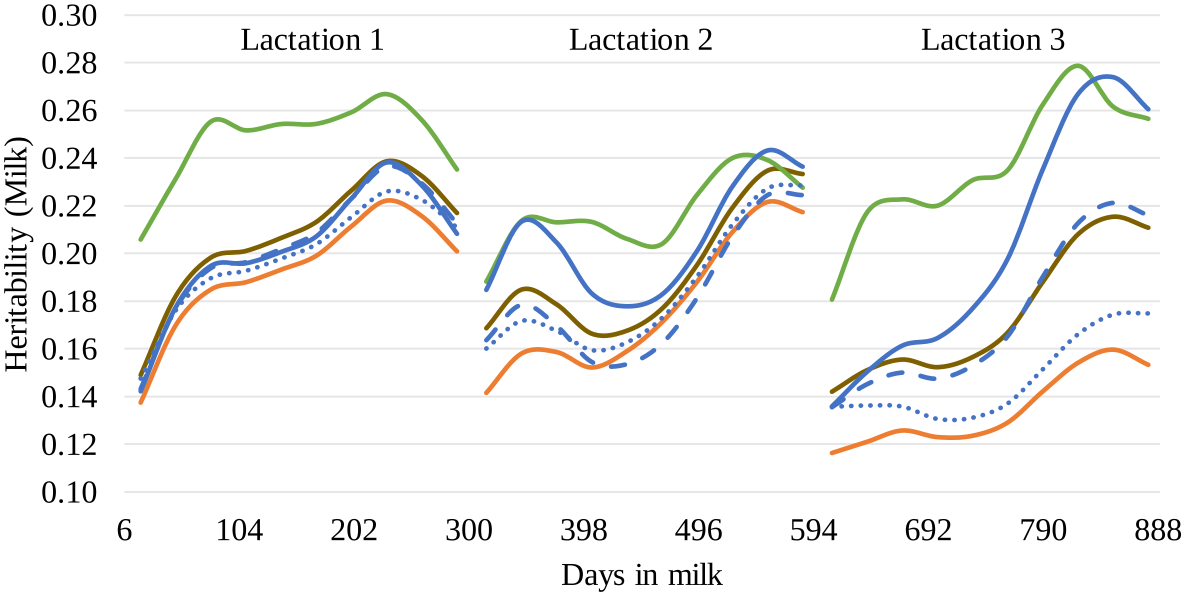
5.3 Results and Discussion

The current model that included the GAR effect in this study, did not have a better BIC (increase of 27,316.38) comparing to results from the previous TMR-PAST model, where the TD records from TMR and PAST were treated as different traits. However, there was an improvement of the BIC (decrease of 2,879.55) of the current model comparing to results from the previous nGE model without the GAR effect. (Van Niekerk et al., 2021). The GAR effect explained 7.80% of the animal-additive-genetic effect for Milk in lactation 1 and decreased to 0.89% and 0.00% for lactation 2 and 3, respectively.

In lactation 1 (Figure 5.1), heritability estimates from the current GAR model were close to similar between AvgRain class 1, 3 and 5, and in close agreement to previous TMR and nGE estimates. In lactation 2 the AvgRain class 1 and 3 estimates were still close to each other and to previous TMR and nGE estimates. However, AvgRain class 5 estimates were higher during early- to mid-lactation and closer to previous heritability estimates from the PAST production system. In lactation 3 the differences between the heritability estimates of AvgRain class 1, 3 and 5 are more pronounced during mid- to late-lactation with estimates from AvgRain class 5 being closest to previous PAST estimates, and AvgRain class 1 estimates still being closest to previous TMR estimates (as in lactation 1 and 2). (Van Niekerk et al., 2021).

Previous nGE heritability estimates are near the TMR estimates as most records analysed with the nGE model came from herds using a TMR production system (Van Niekerk et al., 2021). AvgRain class 1 heritability estimates being close to previous TMR heritability estimates make sense as 100% of herds in this class use a TMR production system. This also make sense to a lesser degree, for AvgRain class 3 estimates as 78% of herds utilise a TMR production system (86% of herds in AvgRain class 2 herds use TMR). 75% of herds employing a PAST production system in AvgRain class 5 explains the closeness of these estimates with previous PAST estimates (67% of herds in AvgRain Class 4 herds use PAST). AvgRain classes 3 and 5 heritability estimates seem to follow previous nGE and PAST estimates more precisely in later lactations, specifically lactation 3. (Van Niekerk et al., 2021).

Figure 5.1 Heritability estimates for milk production over the first three lactations using the current model including the **genetic effect due to average rainfall (class 1: blue dotted; class 3: blue dashed; class 5: blue solid; GAR model)**, compared to the previous model where the two production systems were defined as different traits **(total mixed ration: orange; pasture: green;** TMR-PAST model), as well as the **'national SA Holstein genetic evaluation'** model excluding the genetic effect due to average rainfall and TD records undefined for production system **(brown; nGE model).**



Genetic correlations (Table 5.1) between the GAR and GPr effects were positive but weak, ranging from 0.43 to 0.32 for lactation 1 and 2, and 0.19 for lactation 3. The genetic correlation between the GAR and GPe effects were negative and weak for lactation 1 and 2 (-0.28 and -0.30, respectively), but positive and close to zero for lactation 3 (0.10).

The GPr and GPe components are made up of the average-additive-genetic level and persistency respectively, of performance of animals and their progeny, irrespective of the rainfall area(s) in which they produce and are measured in, over each of the lactations. When the additive-genetic performance of animals and their progeny are further regressed for AvgRain, the genetic correlations mentioned suggest that the genes involved in Milk are not consistent and reliable over different AvgRain areas.

Table 5.1 Genetic correlations between the different animal-additive-genetic effects for milk production (Milk) over the first three lactations (standard errors in parentheses)

Milk		Lac1[1]			Lac2[1]			Lac3[1]		
		GPr[2]	GPe[3]	GAR[4]	GPr[2]	GPe[3]	GAR[4]	GPr[2]	GPe[3]	GAR[4]
Lac1[1]	GPr[2]	1.00								
		NA								
	GPe[3]	-0.29	1.00							
		(0.04)	NA							
	GAR[4]	0.43	-0.28	1.00						
		(0.09)	(0.12)	NA						
Lac2[1]	GPr[2]	0.90	-0.41	0.43	1.00					
		(0.02)	(0.05)	(0.11)	NA					
	GPe[3]	-0.07	0.12	-0.03	-0.15	1.00				
		(0.04)	(0.05)	(0.12)	(0.05)	NA				
	GAR[4]	0.20	-0.41	0.88	0.32	-0.30	1.00			
		(0.10)	(0.13)	(0.11)	(0.32)	(0.12)	NA			
Lac3[1]	GPr[2]	0.88	-0.35	0.32	0.98	-0.24	0.25	1.00		
		(0.03)	(0.06)	(0.12)	(0.02)	(0.05)	(0.12)	NA		
	GPe[3]	-0.08	0.03	-0.07	-0.14	0.95	-0.26	-0.22	1.00	
		(0.05)	(0.06)	(0.13)	(0.06)	(0.02)	(0.14)	(0.06)	NA	
	GAR[4]	0.26	-0.15	0.52	0.27	0.03	0.38	0.19	0.10	1.00
		(0.20)	(0.22)	(0.30)	(0.20)	(0.21)	(0.30)	(0.19)	(0.23)	NA

[1] Lac$_i$ – Lactation 1, 2 or 3; [2] GPr – (additive) genetic-average-production effect; [3] GPe – (additive) genetic-average-persistency effect; [4] GAR – (additive) genetic effect due to long-term-average, annual rainfall; NA – Not Applicable.

5.4 Conclusions

Considering the heritability estimates in this study and supported by previous research, including AvgRain as an important climatic variable in the proposed aRRM for Milk seem to be a good proxy for production system used by Holstein herds in SA, especially for later lactations. The way in which the AvgRain classes are constructed may have an important effect on the heritability estimates. Another method for construction of classes is to fit annual rainfall classes as an interaction with production systems defined as TMR and PAST to possibly counteract the confounding of AvgRain classes with TMR or PAST better. All AvgRain classes, except class 1 did however contain TMR and PAST herds. More herds in high rainfall areas that utilise a PAST production system should be identified and incorporated in the study to increase accuracy of genetic parameters.

CHAPTER 6

The effect of including genomic information from primarily cows on the reranking and accuracies of genetic predictions for SA Holsteins in two production systems

6.1 Introduction

According to Interbull (2024), 23 countries use genomic evaluations for genetic predictions of production traits and can be seen as almost standard practice in the developed world. South African (SA) Holsteins do not utilize such genomic evaluations with the relatively high cost of animal genotyping playing a substantial role. The Dairy Genomics Programme (DGP) was initiated in 2016 where animal genotyping of the three main dairy breeds (Holstein, Jersey and Ayrshire) would be financed by government (Mafolo et al., 2021). One of the key objectives of the DGP initiative was genomic research with possible future genomic evaluations as an outcome (Mafolo et al., 2021). The initiative realised 1221 genotyped Holstein animals.

Genomic evaluations, and more specific the single-step (ss) genomic evaluations were proposed by Misztal et al. (2009), blending phenotypic, pedigree and genotypes to form a single evaluation. Specifically, the pedigree relationship matrix in traditional best linear unbiased predictions (BLUP) are replaced by a more complete matrix (H), which merges the pedigree and genomic relationships (Lourenco et al., 2020). The genomic relationship matrix (G) measures the proportion of alleles shared between animals as indicated by the observed single nucleotide polymorphism markers (SNPs). The new H matrix was, however, rather complex and after Aguilar et al. (2010), and Christensen and Lund (2010) established that the inverse of the H matrix is reasonably simple, single-step genomic BLUP (ssGBLUP) became one of the most widely used methods for genomic evaluations. Legarra et al. (2009) showed that in the H matrix genomic information is passed through to all ungenotyped individuals through the covariances among them. By its very nature (structure), the G matrix accounts for the probability that two random individuals share the same alleles (identical by state) as shown by the SNP markers, irrespective of origin. The pedigree-based relationship matrix can only consider the degree of relatedness through recorded pedigrees (identical by descent; Lourenco et al., 2020). This, together with the G matrix that calculates more complete (exact) relationships, not relying on average probabilities between ancestors, descendants and siblings, increases the accuracy (ACC) of genetic predictions.

Significant genotype by environment interaction between production traits in total mixed ration (TMR) and pasture (PAST) production systems have previously been shown for lactation persistency in SA Holsteins (Van Niekerk et al., 2021). For this reason, studying the inclusion of genomic information, we continued to distinguish TMR and PAST milk production (Milk) as two distinct traits. Although in this study we could not distinguish between production-level and -persistency (cumulative Milk production were used and not the test-day records per se), the influence of persistency on 'total' production over the lactation is still present. Van Niekerk et al. (2023) reported that production-persistency explained 27.8% of the genetic variance for first lactation (Milk) in the pasture production system.

At the start of dairy genomic evaluations, reference populations were mainly made up of bulls with phenotyped progeny (Koivula et al., 2016). However, in small dairy populations such sires can be relatively scarce, and in the case of SA Holsteins where a lot of sires come from foreign countries, local proven bulls for inclusion in genomic evaluations may be in short stock. The SA DGP project led to 78 bulls and 1143 cows that were genotyped, and Mafolo et al. (2020) also reported limited availability of local proven sires to include. In this instance, the use of genotyped cows in genomic evaluations are necessary. Pryce and Daetwyler (2013) and Ding et al. (2013) stated that this is a desirable solution to extend a reference population. Also, to avoid bias animals that were identified under the DGP for genotyping were not just elite according to their genetic predictions but represented the whole variability with regards to their genetic merit.

Regarding the incorporation of genomic information, this study can be seen as a preliminary investigation in SA for Holsteins. The objective was to investigate the difference in genetic predictions (EBVs compared to ssGEBVs) after the inclusion of SNPs from available genotyped animals using the single-step method where Milk in TMR and PAST are defined as separate traits. Specifically, the change in rankings of cows (dams) and sires as well as the shift in corresponding genetic prediction ACCs for TMR and PAST i.e., a bivariate analysis, but also when analysing Milk in the two production systems as the same trait in a more "standard" genetic evaluation i.e., univariate analysis.

6.2 Material and methods

6.2.1 Data

Test-day records (n = 2 167 022) and pedigree information were made available by the South African (SA) Holstein Association, together with the production system used by 140 herds. Production systems utilized by herds included a total mixed ration (TMR) and a pasture (PAST) production system. Records used came from herds that participated in milk recording for at least 3 consecutive years. Milk production (Milk; kg over 24 hours) from cows that were milked 2 or 3 times a day (number of milkings) were included and cows retained in the dataset had to have sufficient (see below) first lactation (Lac1) records. Second lactation (Lac2), if present for these cows were also used in the analyses. Cows without calving dates and those with less than 7 test-day records were discarded. Test-day records from 6 to 300 days in milk were used and retained cows' first test-day record had to be within the first 65 days in milk. Additionally, cows whose consecutive test-day records were shorter than 21 days, or longer than 90 days between each other were discarded. Retained cows' test-day records were used to calculate cumulative 295-days Milk using the test interval method from Sargent (1968) as described by Interbull (2024).

Contemporary groups were formed by concatenating Herd x calving Year x calving Season (HYS) where the following natural four seasons in South Africa were used: Summer, Season 1: December, January and February; Autumn, Season 2: March, April and May; Winter, Season 3: June, July and August; Spring, Season 4: September, October and November. Only cows that were part of a contemporary group with at least 5 cows were kept. The calving year of cows ranged from 1999 to 2019. Calving ages in months, of cows were grouped into 8 groups as follows: Class 1, 20 to 23; Class 2, 24 to 27; Class 3, 28 to 31; Class 4, 32 to 35; Class 5, 36 to 41; Class 6, 42 to 47; Class 7, 48 to 55; Class 8, 56 to 69. Previous calving interval in days, of cows before each lactation were also classified in 8 groups as follows: Class 1, 312 to 353; Class 2, 354 to 395; Class 3, 396 to 437; Class 4, 438 to 479; Class 5, 480 to 521, Class 6, 522 to 563; Class 7, 564 to 730; Class 8, no previous calving interval – i.e., first calf cows. Cows that were not from a sire with at least 10 daughters as well as cows without a sire were removed.

Table 6.1 Descriptive statistics for Milk production data

	Lac1		**Lac2**	
Item	**TMR**	**PAST**	**TMR**	**PAST**
Avg. Milk ± SD (kg)	8 652.8 ± 1 923.8	6 833.8 ± 1 334.1	9 840.5 ± 2 274.4	8 263.0 ± 1 569.8
Records^1	26 943	8 089	10 588	4 000
Records^2 / Group	20	17	13	13
Sires^1	802	460	758	368
Sires 5^1	754	329	561	218
Shared sires 5^1	239		138	
Daughters^2 / Sire	33	18	14	11

SD = Standard deviation; Laci = Lactation i; Avg. = Average; Milk = Milk production; Superscript 1 = Number of; Superscript 2 = Average number of; Group = Contemporary group defined by the Herd x calving Year x calving Season interaction effect; Sires 5 = Sires with a minimum of 5 daughters; Shared sires 5 = Sires with a minimum of 5 daughters in the pasture and total mixed ration production systems each.

Pedigree information from animals born in 1998 to 2018, together with their parents and grandparents (3 generations) were used. This was done to reduce the decline in genetic prediction ACC, which happens with every generation because of recombination, especially for long-range linkage disequilibrium between markers and quantitative trait loci (QTL; Boichard et al., 2022). Hollifield et al. (2021) showed a deterioration in ACC of -0.19 to -0.13 from generation 1 (after the 'training' population) to 9. Nine unknown parent groups (GGroups) were formed according to country groups and birth year of animals with one or more unknown parents. These were included explicitly as additional random effects. The final dataset consisted of 26 943 cows from 44 TMR, and 8 089 cows from 20 PAST production system herds, respectively. The total number of sires with recorded daughters was 844 of which 56% came from North America (Canada and the United States). The proportion of sires from North America of all the sires in the final 3-generation pedigree dataset (2 452 sires; 65 862 animals) were 48%, whilst the foreign country that contributed the third most sires are the Netherlands (5%). South African bulls comprised 40% of all sires.

In order to understand the data structure, cows were grouped on whether they had Milk records or not (Table 6.2). These cows were subdivided on whether they had progeny (DAMS) or no progeny (COWS) in the pedigree. COWS are not heifers. Not all calves of cows, especially bull calves are registered by producers.

Table 6.2 Genotyping patterns for number of animals over TMR and PAST

Animals	**Genotyped**		**Non-genotyped**	
Cows with Milk records	**TMR**	**PAST**	**TMR**	**PAST**
… COWS	244	1	18 636	5 881
… DAMS	289	1	7 774	2 206
… their sires[1]	25	0	359	42
… shared sires[2]	22		396	
Cows without Milk records	**TMR and PAST**		**TMR and PAST**	
… DAMS	141		28 208	
… their sires	20		1 617	

COWS = cows with no progeny in the pedigree; DAMS = cows with progeny in the pedigree; TMR = Total Mixed Ration; PAST = Pasture; Superscript 1 = Sires with recorded daughters only in TMR or PAST; Superscript 2 = Sires with recorded daughters in TMR and PAST.

Genomic data (markers) incorporated in the analyses were 53 218 SNPs (Single Nucleotide Polymorphisms) from each of the 1233 genotyped animals before quality control. The Illumina 50K chip v3

(Illumina Inc., San Diego, CA, USA) were used to genotype animals. Sex chromosomes and variants with a minor allele frequency of less than 2% were removed. The Hardy-Weinberg equilibrium test was applied at P < 10E-6 and a call rate of at least 95% was used. After quality control 50 661 SNPs for 1173 animals were available. Only the genomic information from animals that are in the 3-generation pedigree dataset were kept which led to a total of 743 genotyped animals (676 cows and 67 sires) whose SNPs were incorporated in the applicable analyses. Doing quality control on genomic data, needed to construct the genomic relationship matrix (G; see below) avoids biases and losses of ACC in genomic predictions (Lourenco et al., 2020).

6.2.2 Model and analyses

Cumulative 295-days Milk from the first two lactations were analysed utilizing multi-trait (MT) models with the WOMBAT (Meyer, 2007) software. Multiple analyses with either two or four traits (MT2 or MT4) were performed. The model used is described in Table 6.3.

Table 6.3 Description of the multi-trait (MT2 and MT4) model used to analyse Milk production

Fixed effects (classified)	**Random effects**
Herd x calving Year x calving Season (HYS)	Animal Genetic (additive; AG)
calving Age class (AC)	Permanent Environmental (PE)
previous Calving Interval class (CIC)	Genetic Group (GGroups i.e., Unknown parent groups)
number of Milkings (24 hours; Mi)	Residual

The inclusion of the HYS fixed effect in the model adjusts cow records to a same contemporary group and stage of lactation effect where groups of cows calved in the same 3-month window (season). Shorter calving window periods (one or two months) for the seasonal effect within calving year and herd were tested but did not fit the model significantly ($P \geq 0.05$) better as indicated by the corrected Akaike Information Criterion (AICc). The AICc were mainly used to assess the goodness of fit for the production data when including different fixed effects. All fixed effects in the model were found to have a significant ($P < 0.05$) effect on Milk.

In the analyses where SNPs of genotyped animals were included, the single-step (ss) method to form the so-called H^{-1} matrix, was used. The H^{-1} matrix replaces the A^{-1} (inverse of the pedigree relationship matrix) in the mixed model equations in the animal model. Initially, the G_M matrix is calculated from the SNP marker counts as:

$$G_M = \frac{(M-2P)(M-2P)'}{s} \qquad [1]$$

following Method 1 of VanRaden (2008), where M = the matrix of marker counts (number of genotyped animals x number of markers i.e., 743 animals x 50 661 SNPs in our study), P = corresponding matrix of assumed frequencies p_i, and $s = 2\sum_{i=1}^{m} p_i\,(1- p_i)$. In our case G_M should be positive semidefinite using this method as the number of markers are not smaller than the number of genotyped animals (VanRaden, 2008; Lourenco et al., 2020). However, G_M is commonly adjusted as to ensure its inverse exists, to assure that the G (genomic relationship; see below) matrix line up better with the A (pedigree relationship) matrix, and to account for the residual polygenic variation (Meyer, 2022). Meyer (2022) summarise the general types of adjustments as:

$$G = \lambda\,[\beta(G_M + \epsilon I) + \alpha J] + (1-\lambda)\,A_{22} \qquad [2]$$

Where $0 \le \boldsymbol{\lambda} \ge 1$ represents the fraction of total genetic variance due to the SNP effects, $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are alignment factors (blending parameters) proposed by Vitezica et al. (2011) and Christensen (2012), and reviewed by Lourenco et al. (2020), $\boldsymbol{J}$ is the matrix with all elements equal to 1, $\boldsymbol{I}$ is the identity matrix, $\boldsymbol{\epsilon}$ is a small constant, and $\boldsymbol{A_{22}}$ is the pedigree relationship matrix among genotyped animals. In this study a value of 1 for both $\boldsymbol{\lambda}$ and $\boldsymbol{\beta}$ was assumed. Method 1 of VanRaden (2008) was implemented, which is the default method using WOMBAT (Meyer, 2007). To ensure a positive definite matrix using this method the allele counts are centred by observed frequencies and adding $\boldsymbol{\epsilon}$ (0.1) to the diagonal elements of $\mathbf{G}$. The default WOMBAT (Meyer, 2007) method of Colleau (2002) was used to calculate the submatrix $\boldsymbol{A_{22}}$. To modify $\mathbf{G}$ to better align with $\boldsymbol{A}$, the option to add the term $\boldsymbol{\alpha J}$ to G_M was used, where $\boldsymbol{\alpha}$ was the average difference between the elements of $\boldsymbol{A_{22}}$ and $\boldsymbol{G_M}$ (Vitezica et al., 2011). Finally, the $\mathbf{H}^{-1}$ used in the analyses was:

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \tau G^{-1} - \omega A_{22}^{-1} \end{bmatrix} \qquad [3]$$

where $\boldsymbol{A^{-1}}$ is the inverse of the pedigree relationship matrix, $\boldsymbol{G^{-1}}$ is the inverse of the genomic relationship matrix, and $\boldsymbol{A_{22}^{-1}}$ is the inverse of the pedigree relationship matrix among genotyped animals. Heuristic scale factors, $\boldsymbol{\tau}$ and $\boldsymbol{\omega}$ in [3] were proposed to reduce bias in EBVs (Tsuruta et al., 2011), but they were not used. Thus, the $\boldsymbol{H^{-1}}$ matrix unifies the pedigree and genomic relationship matrices. Using the $\boldsymbol{H^{-1}}$ matrix, SNP information is also propagated to individuals that are not genotyped (Lourenco et al., 2020).

6.2.3 Four-trait analyses

Milk records from cows were separated according to the production system utilized in their corresponding herds (TMR or PAST). Lac1 and Lac2 from cows that produced under a TMR production system (traits 1 and 2) were defined as two different traits compared to Lac1 and Lac2 from cows that produced under a PAST production system (traits 3 and 4). Using REML, two MT4 analyses were done. The first analysis (MT4) was implemented without genomic information, with resulting EBVs and their corresponding ACCs. The second analysis (single-step; ssMT4) incorporated the available genomic information (Table 6.2), with subsequent ssGEBVs and their ACCs. Using WOMBAT, standard error predictions of the AG effect solutions together with diagonal elements of the general relationship matrix were used to calculate ACCs (Meyer, 2007).

6.2.4 Two-trait analyses

The production system (TMR or PAST) under which cows produced were ignored and REML was used to perform two, MT2 (Lac1 and 2) analyses. As in the four-trait analyses, the first analysis (MT2) was done without genomic information and the second analysis with genomic information using single-step (ssMT2). The same pedigree information and production dataset were used for all analyses. ACCs were calculated the same way as in the four-trait analyses (section 6.2.3).

6.3 Results and discussion

6.3.1 (Co)variance ratios

Estimates of all (co)variance components between the ssMT4 and MT4 analyses were identical or very similar, as well as between the ssMT2 and MT2 analyses (Table 6.4). All estimates from the two trait analyses were also in line with the estimates for TMR, from the four traits analyses.

Table 6.4 Variance ratios from the single-step multi-trait (ssMT4 and ssMT2) and multi-trait (MT4 and MT2; in parentheses if different) models for Milk

	ssMT4 and MT4 analyses i.e., with TMR and PAST							
	Heritability		**PE**		**GGroups**		**Residual**	
	TMR	**PAST**	**TMR**	**PAST**	**TMR**	**PAST**	**TMR**	**PAST**
Lac1	0.23 (0.24)	0.31	0.25	0.67	0.02	0.01	0.49	0.01
Lac2	0.16	0.27	0.26	0.11	0.03	0.00	0.55	0.62

	ssMT2 and MT2 analyses i.e., without TMR and PAST			
	Heritability	**PE**	**GGroups**	**Residual**
Lac1	0.25	0.24	0.02	0.50 (0.49)
Lac2	0.17 (0.18)	0.20	0.02	0.61

ss = single-step; Laci = Lactation i; PE = Permanent Environmental; GGroups = Genetic Groups; TMR = Total Mixed Ration production system; PAST = Pasture production system

Senarath et al. (2022) also reported similar heritabilities for Milk between BLUP and ssGBLUP within five different regions in Thailand with estimates ranging from 0.23 to 0.40 for Holsteins. The authors reasoned that the genetic parameters are not altered when genomic and pedigree matrices for the genotyped animals are alike. Boonkum et al. (2023) showed very comparable estimates of 0.04 (BLUP) vs 0.05 (ssGBLUP), but with regards to number of services per conception. The heritability estimates for Milk in PAST was higher than in TMR as well as in Lac1 compared to Lac2 (all analyses). Results from Van Niekerk et al. (2023) confirm this tendency, but using a test-day (TD), multi-lactation, and random regression model where TMR and PAST were analysed separately. The heritabilities (per se), especially for TMR corresponded well with Önder et al. (2023) with an estimate of 0.24 for 305-day milk yield in Polish Holsteins. Buch et al. (2012) demonstrated that with a higher heritability (0.15 vs 0.05 in their case) a reference population of cows will be more advantageous than a reference population of bulls. When a trait has a relatively higher heritability (own record is a more accurate indicator of phenotypic performance), together with the fact that it is measured in the animals (cows) that directly provide the genomic information, the stronger the affiliation between the phenotype, genetic merit and marker(s) should be. This is due to no recombination (effect) that take place between the markers and phenotype. This help explain results on ACCs of genetic predictions in the current study where MILK, because of its higher heritability estimates (h^2 range 0.23 to 0.31; Table 6.4) compared to Buch et al. (2012) are a more accurate indicator of cows' phenotypic performance (Bourdon, 2014). Bulls on the other hand, and more specifically their genomic information (genetic markers) are removed from the phenotypic performance, measured in their daughters because of recombination loss. If the trait in question has a relatively low heritability, the weaker of an indicator his daughters' phenotypic performance are of his genetic merit, and the larger number of measured daughters for the trait the sire needs for a more accurate genetic prediction. For bulls this should translate in a weaker affiliation between phenotypes (measured in his daughters), his genetic merit and markers. A relatively low average number of daughters per sire (Table 6.1) in our study is thus not ideal.

Compared to our previous results where TMR and PAST were analysed as separate traits (Van Niekerk et al., submitted), and considering that a TD, multi-lactation, and random regression model was used, results for TMR and to a lesser degree PAST, were broadly in line. However, the PAST PE ratio had a much higher estimate for Lac1 and a relatively low estimate for Lac2 (0.67 and 0.11, respectively; ss- and MT4). Conversely, the corresponding residual ratios for PAST (0.01 and 0.62 for Lac1 and Lac2, respectively) were opposite than the PE ratios (0.67 and 0.11 for Lac 1 and Lac2, respectively). It seems that in this case, using cumulative production, and a very different model and methodology, the ss- and MT4 analyses could not accurately differentiate the two environmental effects (permanent and "temporary"), specifically in the PAST environment.

6.3.2 Genetic correlations

Genetic correlations between estimates from the ss- and MT4 analyses (Table 6.5) were similar. Corresponding correlations between Lac1 and 2 were 0.90 (not shown) for both the ss- and MT2 analyses. Within TMR and PAST genetic correlations between Lac1 and 2 (0.85 and 0.97, respectively; ssMT4) were strong. These genetic correlations (Lac1,2) are similar to the estimate of 0.91 for 305-day Milk by Lee et al. (2020; ssGBLUP). The between TMR and PAST genetic correlation for Lac1 were also strong (0.81; ssMT4) and above the threshold of 0.80 as per Robertson (1959), indicating no significant reranking of animals between the production systems for cumulative Milk in Lac1. However, the corresponding genetic correlation for Lac2 (0.70; ssMT4) indicates that there will be moderate reranking of animals between the production systems for Lac2. This result is in line with Van Niekerk et al. (submitted; comparable dataset) where the authors also found evidence of genotype by environment interaction between TMR and PAST with a genetic correlation of 0.64 for Lac2 persistency of Milk.

Table 6.5 Genetic correlations (below diagonal) and phenotypic correlations (above diagonal) from the single-step multi-trait (ssMT4) and multi-trait (MT4; in parentheses if different) models for Milk

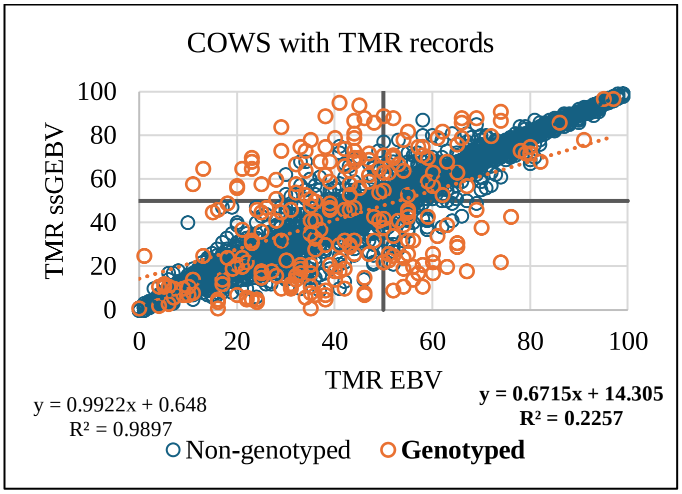
		TMR		**PAST**	
		Lac1	**Lac2**	**Lac1**	**Lac2**
TMR	**Lac1**	-	0.48	0.23 (0.24)	0.23
	Lac2	0.85	-	0.13 (0.14)	0.15
PAST	**Lac1**	0.81 (0.84)	0.53 (0.58)	-	0.53
	Lac2	0.89 (0.90)	0.70 (0.71)	0.97 (0.96)	-

TMR = Total Mixed Ration production system; PAST = Pasture production system; Laci = Lactation i

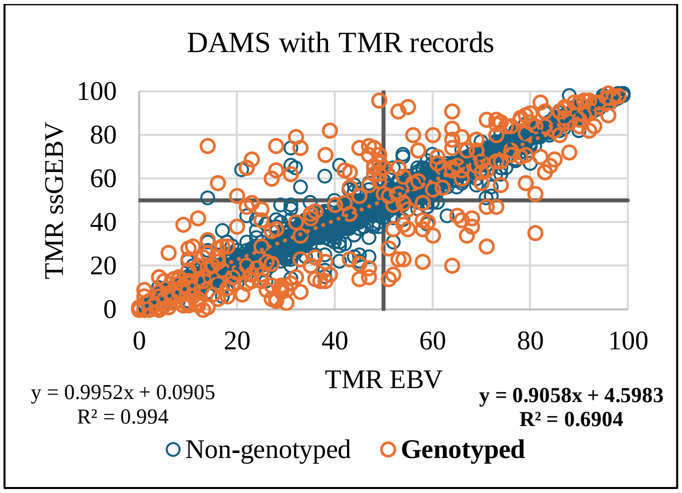
6.3.3 Change in genetic prediction rank (percentile) for COWS and DAMS: ss- and MT4 analyses

The percentile (pctl) rank of EBV vs ssGEBVs for Milk Lac1 from the two MT4 analyses for cows are depicted in Figures 6.1a to d (left). Cows were also grouped on whether they had TMR or PAST records, where applicable. Non-genotyped animals (left labels) are distinguished from genotyped animals (right labels) in each graph, each with their own regression line and coefficient of determination (R^2). The mean difference together with the mean difference standard deviation (SD) for genotyped vs non-genotyped animals in each group were also calculated and shown under each graph. Correspondingly, Figures 6.2a to d (right) display the same information as Figures 6.1a to d, but for the EBV and ssGEBV ACCs for the equivalent (left) group of animals.

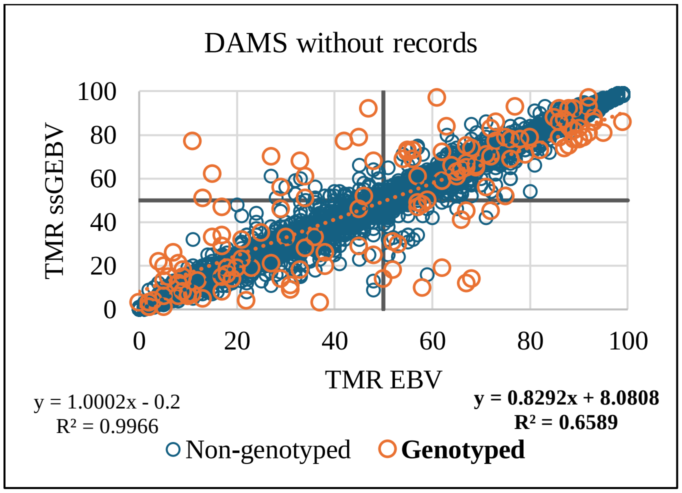
Figures 6.1a, -b, -c and -d (left) Lactation 1 Milk EBV vs ssGEBV percentile rank: COWS and DAMS



Mean difference ± SD;
genotyped vs non-genotyped = **0.73 ± 24.08** vs 0.31 ± 2.48

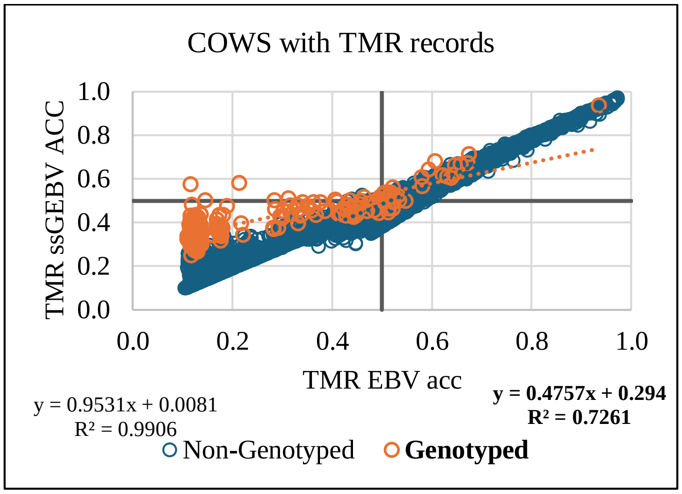


Mean difference ± SD;
genotyped vs non-genotyped = **0.50 ± 17.53** vs -0.15 ± 2.37

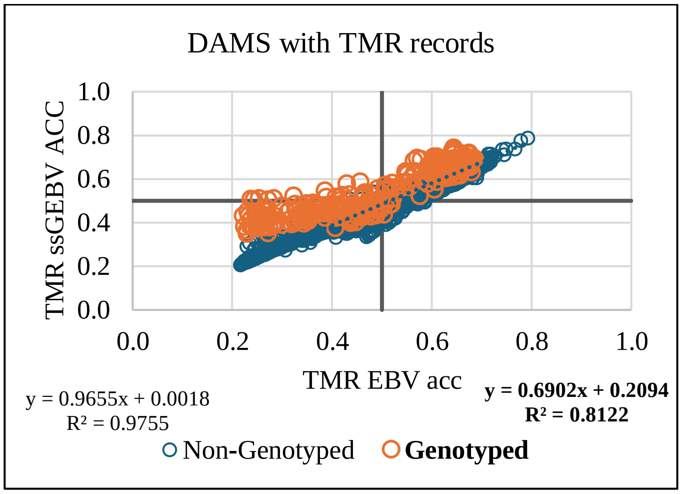


Mean difference ± SD;
genotyped vs non-genotyped = **0.02 ± 18.47** vs -0.19 ± 1.79

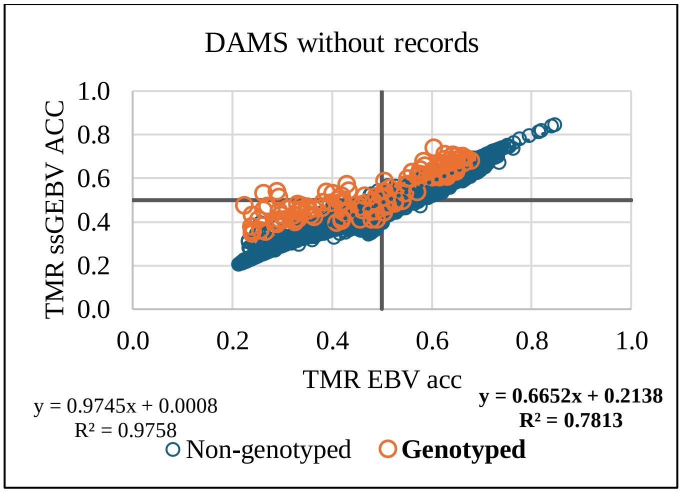
Figures 6.2a, -b, -c and -d (right) Lactation 1 Milk EBV vs ssGEBV accuracy (ACC): COWS and DAMS



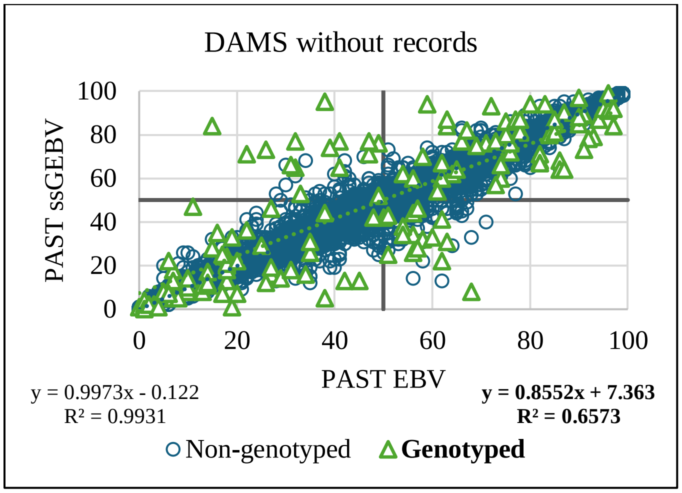
Mean difference ± SD;
genotyped vs non-genotyped = **0.17 ± 0.10** vs -0.01 ± 0.02



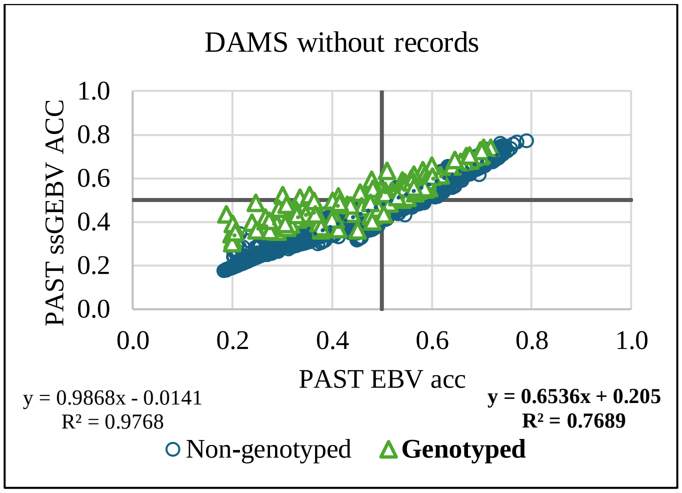
Mean difference ± SD;
genotyped vs non-genotyped = **0.06 ± 0.06** vs -0.02 ± 0.02



Mean difference ± SD;
genotyped vs non-genotyped = **0.05 ± 0.07** vs -0.01 ± 0.02



Mean difference ± SD;
genotyped vs non-genotyped = **0.56 ± 18.97** vs -0.27 ± 2.56



Mean difference ± SD;
genotyped vs non-genotyped = **0.04 ± 0.07** vs -0.02 ± 0.02

The relationship between the EBV and ssGEBV pctl ranks and ACCs for COWS and DAMS with PAST records are not shown as there was only 1 animal in each of these two groups that were genotyped (see Table 6.2).

When adding SNP information and switching from a MT4 to a ssMT4 analysis, the largest changes in predicted genetic merit were found amongst the genotyped COWS with Milk records in TMR (Figure 1a). The correlation between the TMR EBV and ssGEBV pctl ranks for COWS were 0.48 compared to nearly 1 (0.996) for the non-genotyped COWS. Significant reranking will take place amid the genotyped COWS because of the inconsistent correlation between their EBVs and ssGEBVs.

The correlation between the TMR EBV and ssGEBV pctl ranks for the genotyped DAMS (with or without TMR records; 0.83 and 0.81, respectively; Figures 6.1b and c) were stronger and more consistent than the COWS with records. Non-genotyped DAMS in these groups' corresponding correlations were also close to 1 (0.994 and 0.997, respectively). Minor reranking midst these genotyped DAMS will take place. The PAST pctl ranks for the DAMS without records (Figure 1d) followed the same tendency.

The mean difference between the EBV and ssGEBV pctl ranks for all cows (Figure 1a to d) were small (below 1 pctl), with a negligible increase for genotyped cows compared to non-genotyped cows. The variation of the mean difference of the genotyped cows were however much larger (7 to 10 times) than the corresponding variation in the non-genotyped cows. The maximum increases in TMR EBV vs ssGEBV pctl rank for genotyped cows from Figure 1a to d ranged from 55 to 69, while the maximum decreases ranged from -60 to -46. Maximum differences for the non-genotyped animals were smaller, but still substantial with corresponding increases ranging from 34 to 43, and maximum decreases from -49 to -23.

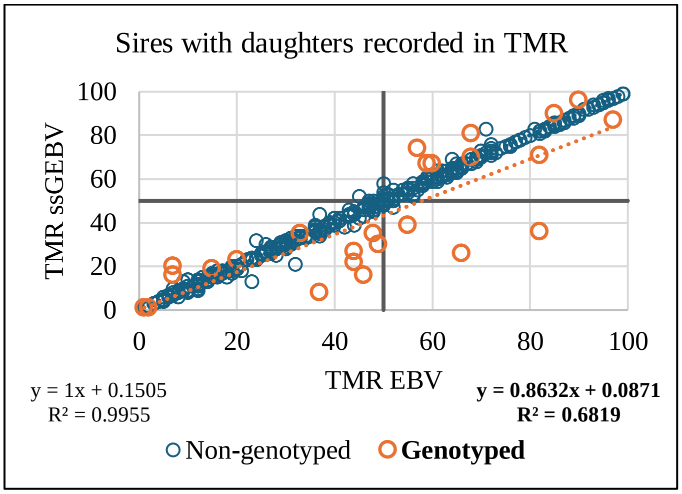
Meyer et al. (2018) stated that often there is overdispersion reported for GEBVs (genomic estimated breeding values). Regarding dispersion, the SD of the EBV vs the GEBV pctl rankings when moving from the MT4 to the ssMT4 analyses were similar for all animals in the specified groups (Table 6.2) when not distinguishing between genotyped and non-genotyped animals. This was also the case for the MT2 vs the ssMT2 analyses. For example, the difference between the SDs of the EBV vs GEBV pctl ranks (MT4 vs ssMT4) for all COWS with TMR records (Figure 1a), irrespective of being genotyped (i.e., 244 + 18 636 = 18 880 animals; Table 6.2) was negligible (24.36 vs 24.39, respectively; not shown). Furthermore, the SD of the GEBV pctl rank of specifically the genotyped COWS (in the same group) are also very similar (26.45; not shown) to the EBV and GEBV values of all COWS. This shows no significant overdispersion of the GEBV pctl ranks, even for the genotyped animals compared to the dispersion for the entire group (regardless of genotyped or not) they are in. This was the tendency for all groups (Table 6.2) as well as for the MT2 vs the ssMT2 analyses.

6.3.4 Change in genetic prediction accuracies for COWS and DAMS

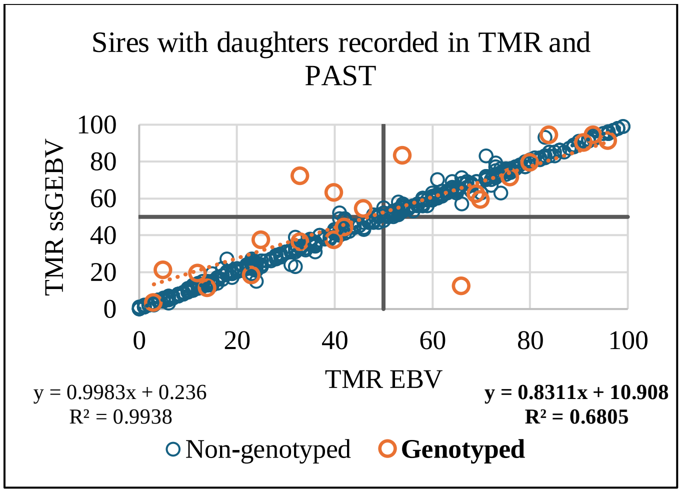
The correlation between the TMR EBV and ssGEBV ACCs for the genotyped COWS with records (Figure 2a) were 0.85 while the same correlation for the non-genotyped COWS were close to 1 (0.995). The largest and main increase in ssGEBV ACCs were amongst the genotyped COWS with the lowest EBV ACCs (~0.11 to ~0.40). The EBV ACCs in this range all increased to ssGEBV ACCs of ~0.51 (i.e., an increase of ~0.11 to ~0.40). This group of cows had the highest mean difference (increase) in genetic prediction ACCs (0.17) as well as the highest SD (0.10) and the largest maximum increase (0.46). The corresponding values for the non-genotyped COWS were -0.01 ± 0.02. Lee et al. (2020) reported an increase of 0.05 in Milk EBV vs GEBV ACC for cows with records in Korean Holsteins where more (n = 2007) animals, mostly cows, were genotyped. The number of genotyped animals in a population, together with the quality of phenotypes does affect the GEBV ACCs (Lund et al., 2011). Some non-genotyped COWS in this group also had ACC increases, the largest being 0.20. There were however also ACC decreases with the largest being -0.06 amid the genotyped and -0.14 amid the non-genotyped COWS.

Results for the genotyped DAMS with TMR records (Figure 2b), or DAMS without records for their TMR (Figure 2c) and PAST (Figure 2d) estimates, mostly followed the same tendency regarding ACCs compared to the COWS with records (Figure 2a). The TMR EBV vs ssGEBV ACCs correlation were slightly stronger in the two DAM groups with values of 0.90, 0.88 and 0.88 for DAMS with TMR records and, the TMR and PAST estimations for DAMS without records, respectively. Apart from increases in ssGEBV ACCs midst the genotyped DAMS with the lowest EBV ACCs (~0.22 to ~0.43, all increasing to ssGEBV ACCs of ~0.55 i.e., an increase of ~0.12 to ~0.33), results for the two dam groups also showed additional gains in the ssGEBV ACCs of DAMS with EBV ACCs from ~0.43 onwards, especially when compared to corresponding non-genotyped DAMS. This can be seen especially for genotyped DAMS with records in Figure 2b. The mean difference (increase) and SD between the EBV and ssGEBV ACCs were 0.06 ± 0.06, 0.05 ± 0.07 and 0.04 ± 0.07 for DAMS with TMR records' TMR estimates, and DAMS without records' TMR and PAST estimates, respectively. Correspondingly, the non-genotyped mean estimates were lower and with less variation i.e., -0.02 ± 0.02, -0.01 ± 0.02 and -0.02 ± 0.02, respectively. Lee et al. (2020) showed an increase of 0.12 in Milk EBV vs GEBV ACC for heifers without records. Luštrek et al. (2021; Slovenian Brown Swiss cattle) reported an average MILK ACC increase from EBVs to GEBVs of 0.22 for female animals. These animals did, however, not have phenotypic records of their own and the heritability in this population was higher than in this study (0.33 vs 0.23). They incorporated 2 166 genotyped animals.

Figures 6.3a, -b, -c, -d and -e (left) Lactation 1 Milk EBV vs ssGEBV percentile rank: sires

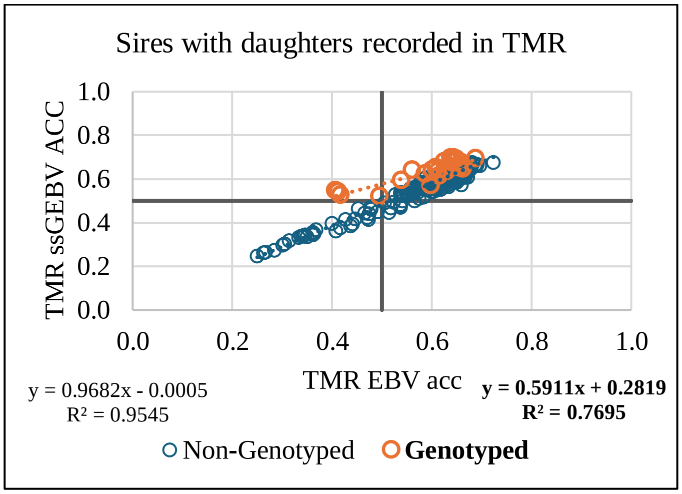


Mean difference ± SD;
genotyped vs non-genotyped = **-6.60 ± 17.16** vs 0.15 ± 1.84

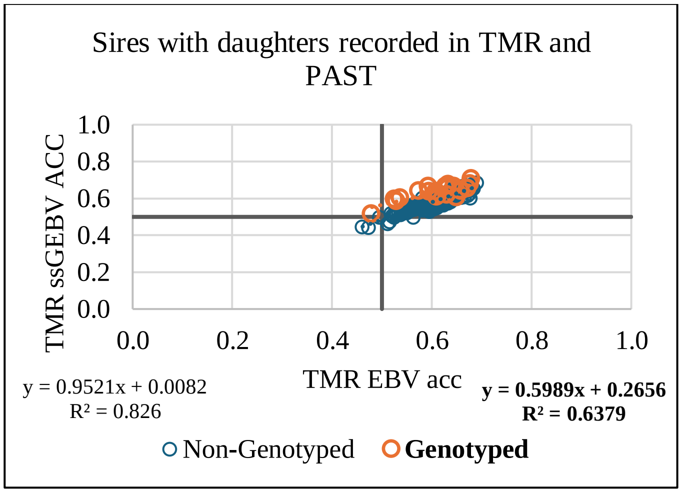


Mean difference ± SD;
genotyped vs non-genotyped = **2.50 ± 17.68** vs 0.16 ± 2.23

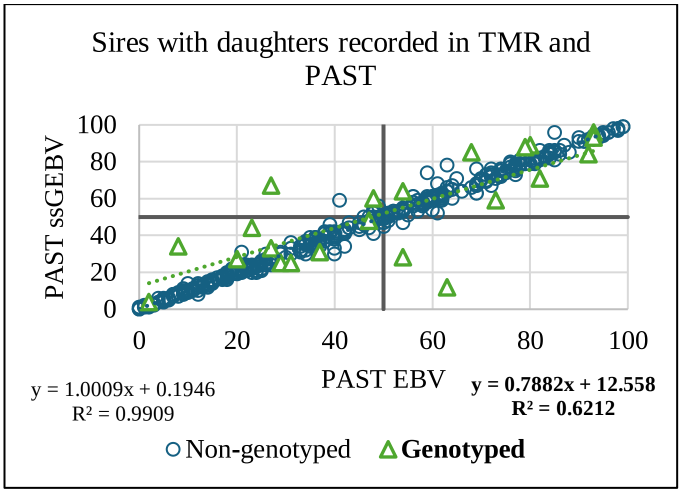
Figures 6.4a, -b, -c, -d and -e (right) Lactation 1 Milk EBV vs ssGEBV accuracy (ACC): sires



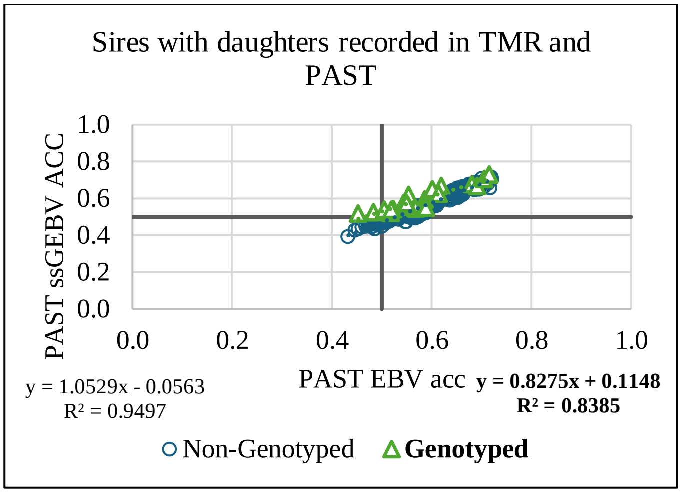
Mean difference ± SD;
Genotyped vs non-genotyped = **0.04 ± 0.04** vs -0.02 ± 0.02



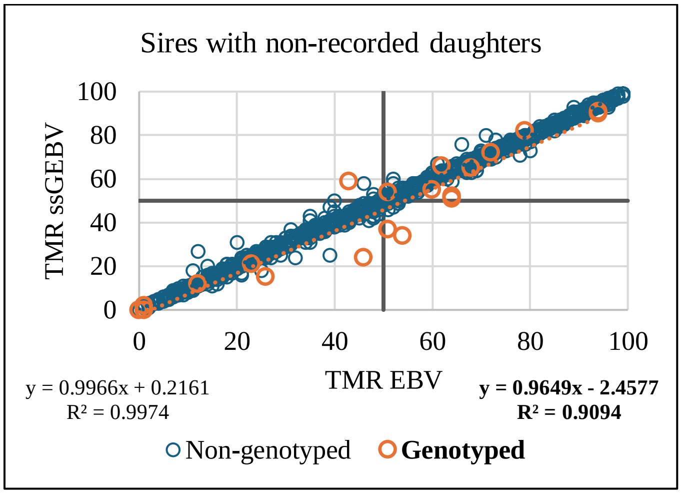
Mean difference ± SD;
genotyped vs non-genotyped = **0.02 ± 0.03** vs -0.02 ± 0.02


Sires with daughters recorded in TMR and PAST

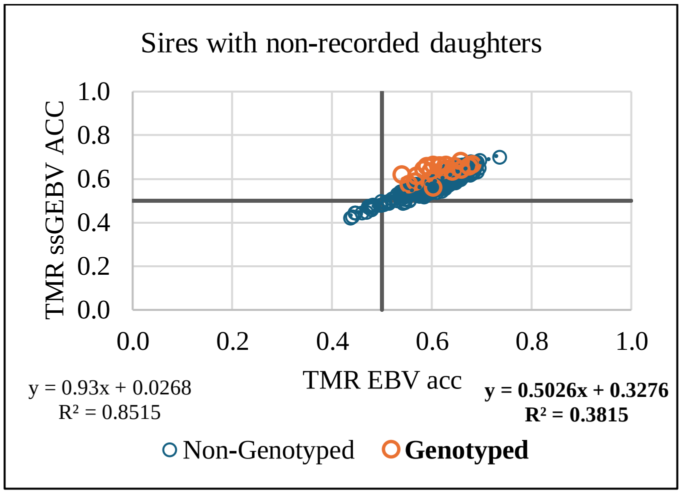
Mean difference ± SD;
genotyped vs non-genotyped = **1.68 ± 18.51** vs 0.24 ± 2.67


Sires with daughters recorded in TMR and PAST

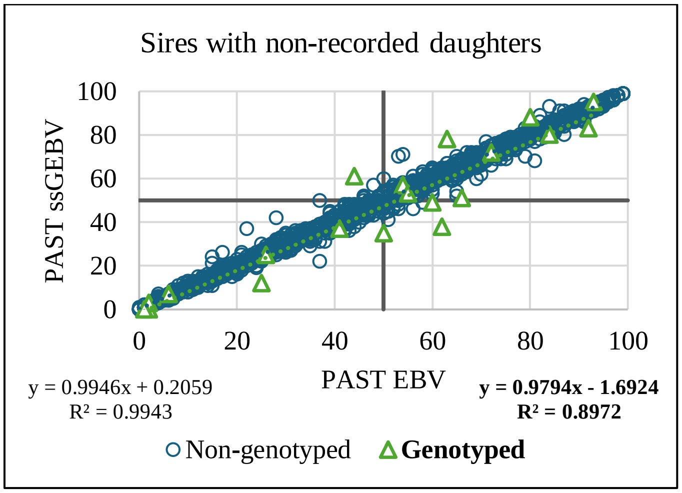
Mean difference ± SD;
genotyped vs non-genotyped = **0.02 ± 0.03** vs -0.03 ± 0.02


Sires with non-recorded daughters

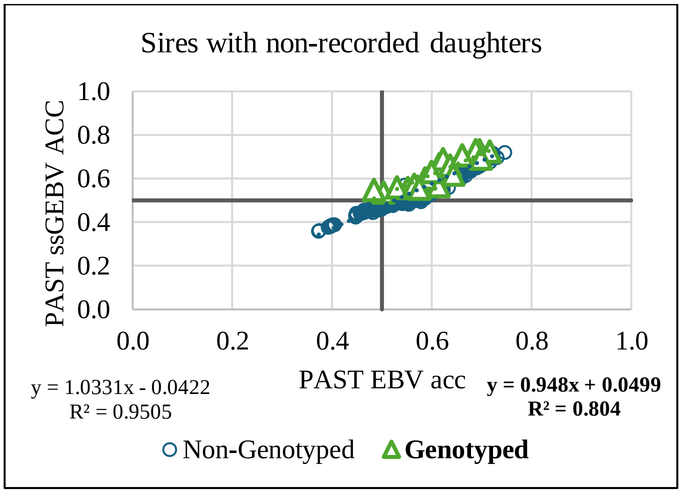
Mean difference ± SD;
genotyped vs non-genotyped = **-4.15 ± 8.96** vs 0.04 ± 1.50


Sires with non-recorded daughters

Mean difference ± SD;
genotyped vs non-genotyped = **0.02 ± 0.03** vs -0.02 ± 0.01


Sires with non-recorded daughters

Mean difference ± SD;
genotyped vs non-genotyped = **-2.70 ± 9.96** vs -0.07 ± 2.21


Sires with non-recorded daughters

Mean difference ± SD;
genotyped vs non-genotyped = **0.02 ± 0.03** vs -0.02 ± 0.01

6.3.5 Change in genetic prediction rank (percentile) and accuracies for sires: ss- and MT4 analyses

Figures 6.3a to d (left) show the Milk Lac1 EBV vs ssGEBV percentile (pctl) ranks from the four trait analyses for sires. Non-genotyped vs genotyped animals are distinguished from each other as with the COW and DAM groups, with the same basic statistics under each graph. Correspondingly, Figures 6.4a to d (right) show the same information as Figures 6.3a to d, but for the EBV and ssGEBV ACCs for the equivalent (left) group of animals. The relationship between the EBV and ssGEBV pctl ranks and ACCs for sires with daughters recorded in PAST only are not shown as none of them were genotyped.

The number of sires in the dataset that were genotyped were much lower than the COWS and DAMS, which is not ideal. Sires (recorded daughters in TMR) and shared sires (recorded daughters in TMR and PAST) followed the same results than the two DAM groups with regards to the TMR EBV vs ssGEBV pctl rank correlations for genotyped sires, being fairly strong for sires (Figure 3a; 0.83) and shared sires regarding their TMR (Figure 3b; 0.82) and PAST (Figure 3c; 0.79) pctl rankings. Non-genotyped counterparts' correlations were also close to 1. Some re-ordering did however take place between these genotyped sires' EBV and ssGEBV pctl rankings. The sires with recorded TMR daughters had the largest mean difference between the EBV and ssGEBV pctl rank (-6.60) as there were some major reranking of ~9 sires, downwards from their EBV. Sires with recorded daughters also had the largest mean difference SD amid sires with the variation in their pctl rankings being 7 to 9 times larger than their non-genotyped counterparts. Sires and shared sires (recorded daughters) had maximum pctl rank increases ranging from 17 to 40 for genotyped sires vs 12 to 18 for non-genotyped sires. Corresponding pctl rank decreases that took place ranged from -46 to -51 for genotyped sires vs -10 to -11 for non-genotyped sires.

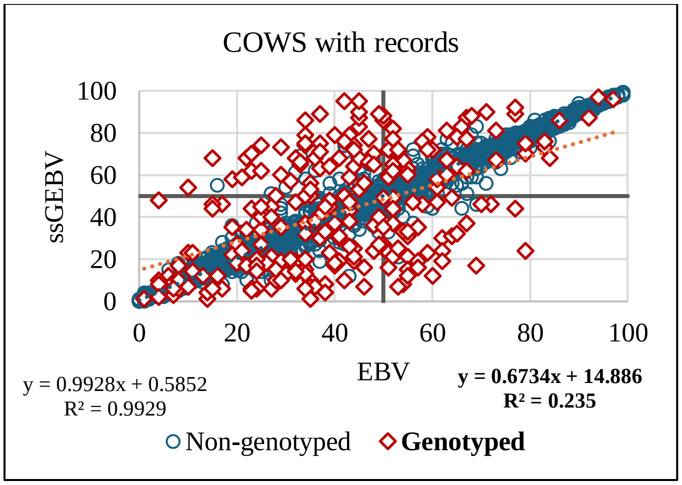
The TMR (Figure 3d) and PAST (Figure 3e) EBV vs ssGEBV pctl rank correlations for genotyped sires with non-recorded daughters were nearly as close to 1 (0.95 and 0.95, respectively) than their non-genotyped counterparts (0.998 and 0.994, respectively).

The correlation between the EBV and ssGEBV ACCs were all fairly strong (0.80) to strong (0.90) for most genotyped sires (Figures 6.4a, b, c and e), except between the TMR estimates of sires with non-recorded daughters (Figure 4d). The greatest increase in ACC for the ssGEBVs followed the same tendency as for the cows, but to a lesser degree. The greatest increase in ACCs took place amongst the genotyped sires that had the lower EBV ACCs, although the increases were not large. The mean differences between the EBV and ssGEBV ACCs were lower than for the cows and with less variation (SD), with mean values of 0.04 to 0.02 and SDs of 0.03 to 0.04. Corresponding values for non-genotyped sires were slightly lower and ranged from -0.03 to -0.02 and SDs of 0.01 to 0.02. Using substantially more and only cows (n = 3 087) in their reference population, Ding et al. (2013) reported an increase of 0.10 to 0.17 for GEBVs compared to conventional EBVs for young bulls. In a study by Buch et al. (2012) the difference between using progeny tested bulls or their genotyped daughters with their phenotypes in a reference population, were investigated. ACC was higher when using the daughters as there is a closer and more specific association between SNP markers and trait performance in these cows. The importance of genotyping cows with phenotypes to predict SNP effects led to increased correlations between phenotypes and GEBVs compared with having reference animals only consisting of just genotyped bulls (Jenko et al., 2017). Lee et al. (2020) reported increases in (Milk) ACCs of GEBVs vs EBVs of 0.03 for sires with progeny and 0.15 for bulls without progeny. This result (Lee et al., 2020) aligns with the primary goal of genomic selection, which is to provide as accurately as possible genetic predictions for young animals, which do not yet have phenotypic measurements (Meuwissen et al., 2001; Schaeffer, 2006).

6.3.6 Change in genetic prediction rank (percentile) and accuracies without production systems (TMR and PAST): ss- and MT2 analyses

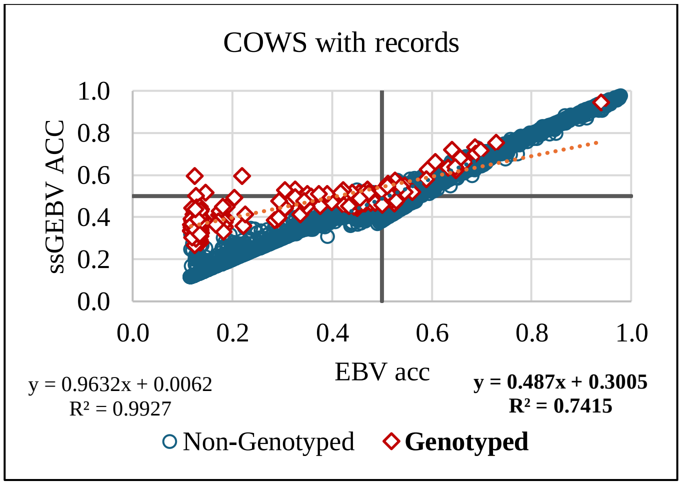
Figures 6.5a to c (left) show the Milk Lac1 EBV percentile (pctl) rank (MT2) vs ssGEBVs pctl rank (ssMT2) analyses for COWS and DAMS, irrespective of production system used. Non-genotyped vs genotyped animals are distinguished from each other as with the previous cow (dam) and sire groups, with the same basic statistics under each graph. Correspondingly, Figures 6.6a to c (right) show the same information as Figures 6.5a to c, but for the EBV and ssGEBV ACCs for the equivalent (left) group of animals. Correspondingly, Figures 6.7a and b as well as Figures 6.8a and b are the same as for above mentioned COWS and DAMS, but only for sires with or without recorded daughters.

Figures 6.5a, -b and -c (left) Lactation 1 Milk EBV vs ssGEBV percentile rank: COWS and DAMS (irrespective of production system)

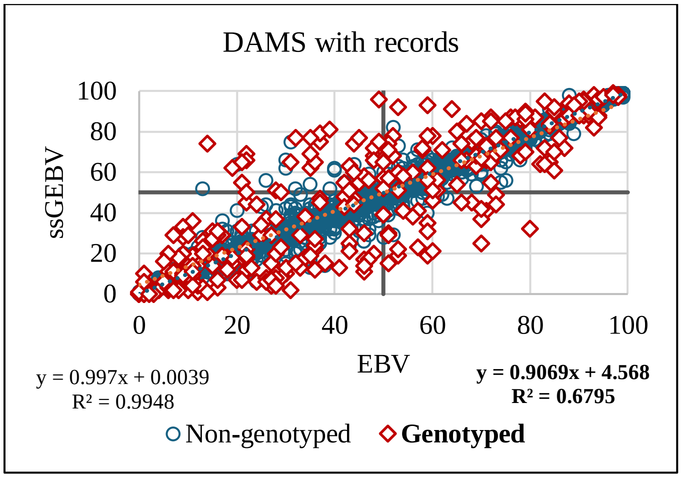


Mean difference ± SD;
genotyped vs non-genotyped = **1.31 ± 24.25** vs 0.27 ± 2.40

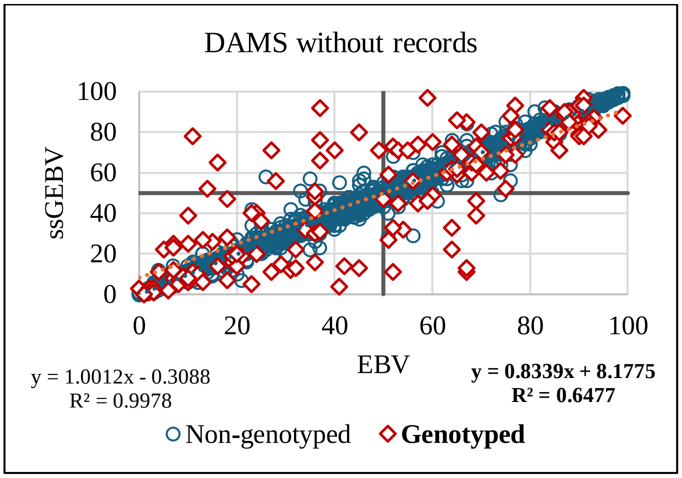
Figures 6.6a, -b and -c (right) Lactation 1 Milk EBV vs ssGEBV accuracy (ACC): COWS and DAMS (irrespective of production system)



Mean difference ± SD;
genotyped vs non-genotyped = **0.17 ± 0.10** vs -0.01 ± 0.02

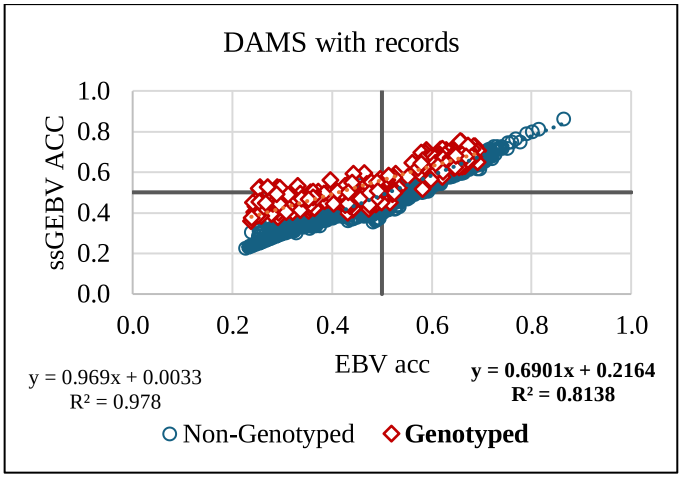


Mean difference ± SD;
genotyped vs non-genotyped = **0.52 ± 17.79** vs -0.15 ± 2.21

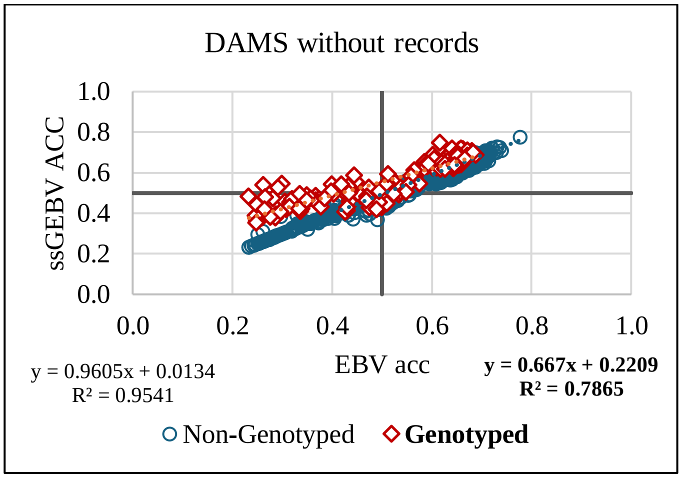


Mean difference ± SD;
genotyped vs non-genotyped = **0.31 ± 18.87** vs -0.19 ± 1.69

Figures 6.7a and -b (left) Lactation 1 Milk EBV vs ssGEBV percentile rank: sires (irrespective of production system)

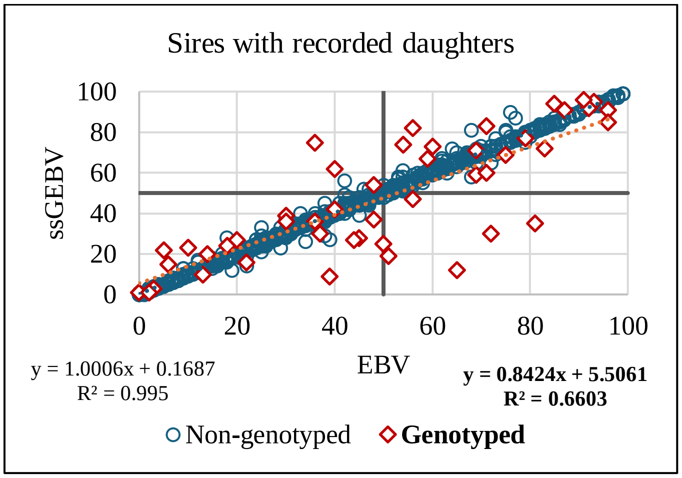


Mean difference ± SD;
genotyped vs non-genotyped = **0.06 ± 0.06** vs -0.01 ± 0.02

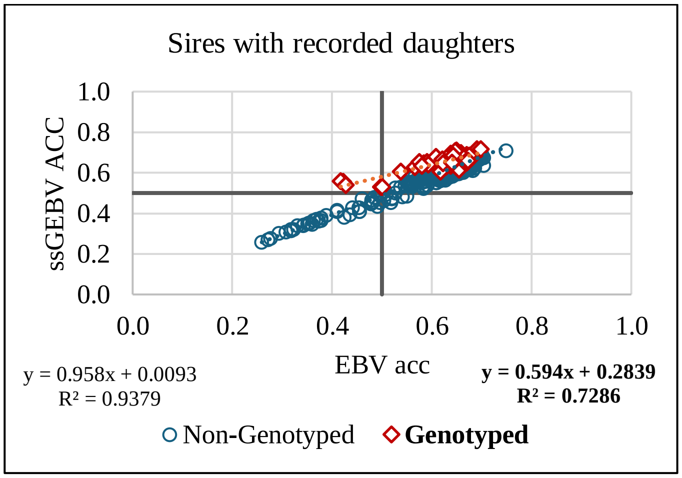


Mean difference ± SD;
genotyped vs non-genotyped = **0.05 ± 0.07** vs -0.01 ± 0.01

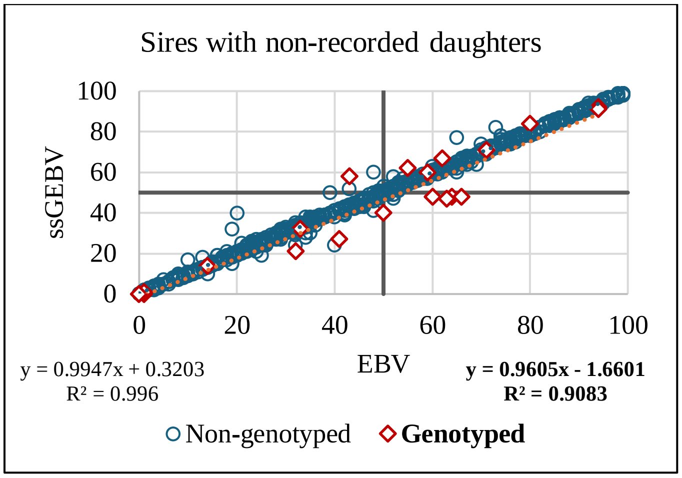
Figures 6.8a and -b (right) Lactation 1 Milk EBV vs ssGEBV accuracy (ACC): sires (irrespective of production system)



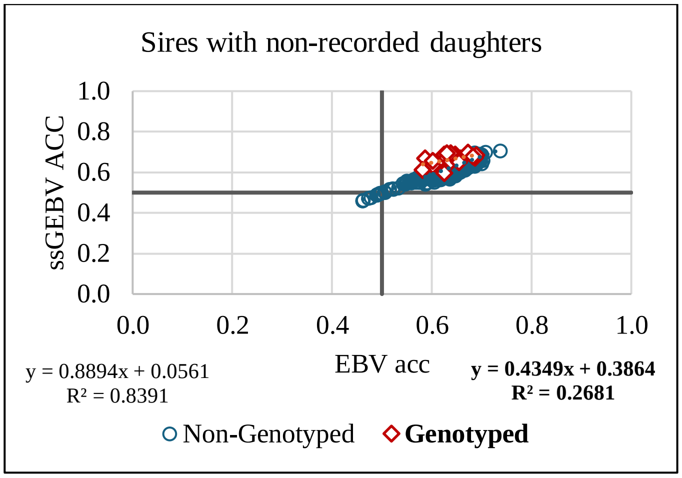
Mean difference ± SD;
genotyped vs non-genotyped = **-2.36 ± 17.93** vs 0.19 ± 1.95



Mean difference ± SD;
genotyped vs non-genotyped = **0.03 ± 0.03** vs -0.01 ± 0.01



Mean difference ± SD;
genotyped vs non-genotyped = **-3.60 ± 8.79** vs 0.02 ± 1.47



Mean difference ± SD;
genotyped vs non-genotyped = **0.03 ± 0.04** vs -0.02 ± 0.02

Results from the ss- and MT2 analyses followed the same pattern for all COWS, DAMS and sires. This can be expected as results from the analyses with production system could mainly be based on TMR results, which forms the majority of the data analysed. Lac2 results from the ss- and MT4 analyses (with production systems; shown as supplementary material), as well as Lac2 results from the ss- and MT2 analyses (not shown) also followed the same corresponding tendencies. Sungkhapreecha et al. (2021) reported an increase in Milk ACC of 0.18 from BLUP vs ssGBLUP genetic predictions incorporating 882 genotyped animals in a relatively small population.

Results from this study show negative differences between the EBV vs the GEBV ACCs. This probably points to sub-optimal blending (scaling) between the G and A_{22} matrices. In the method used the G matrix is adjusted to align with A_{22} by moving genomic relationships with a constant (Vitezica eta al., 2011). However, other methods also exist (Meyer et al., 2018) and include 'shrinking' the G matrix towards the corresponding part of the A matrix as described by McMillan et al. (2017) and Zhang et al. (2017). The A matrix can also be adjusted to match the G matrix (see Christensen et al., 2012 and Legarra et al., 2015). 'Correction factors' to build H^{-1} (Meyer et al., 2018) can also be used where the G^{-1} matrix is weighed by value τ and the A_{22}^{-1} matrix by value ω. These values are determined heuristically. With the results of Martini et al. (2018) they aimed to improve the comprehension of the use of τ and ω to reduce inflation of ssGBLUP predictions.

Also, the use of GGroups (unknown parent groups) in our study are validated by Tsuruta et al. (2014) and Matilainen et al. (2018) in that such groups do reduce inflation. However, in a review by Masuda et al. (2022) they stated that GGroups may not fully remove bias and inflation because of inadequate relationships and inbreeding in matrix A when missing pedigrees are present. This might not be the case in our study as we did use the method by Vitezica et al. (2011) as suggested by Masuda et al. (2022). Buaban et al. (2021) also concluded that inflation tended to be reduced by using cow genotypes as in the current study.

Finally, results from Buch et al. (2012) and Jenko et al. (2016) support that although the inclusion of recorded and genotyped cows are important, the best results with regards to reduced inflation and increased GEBV ACCs are realized when the reference population is made up of sufficient number of genotyped bulls as well as cows with records.

6.4 Conclusions

In our subset of the SA Holstein population, the largest mean increases in genetic predictions (EBV vs ssGEBV) ACC were for genotyped COWS (with records, but no progeny in the pedigree). Among these genotyped COWS the largest increases occurred mainly for COWS with a lower EBV ACC. The same tendency could be observed for genotyped DAMS (with own records, and progeny in the pedigree) with lower EBV ACC. The mean increase in genetic predictions for all DAMS (also those without own records) were lower than for COWS. Mean genetic prediction ACC increases for COWS and DAMS varied from being in line with other comparable studies to being somewhat lower. Results for female animals were significant when incorporating SNPs in ssGBLUP analyses, even with the relative limited number of genotyped animals in the study. Comparable studies did have more genotyped females, and in some instances, males incorporated in their genomic analyses. To determine and utilize the full benefits of incorporating genomic data it is recommended that a far greater number of recorded cows and sires must be genotyped. This should especially include more cows producing in the PAST production system because of the significant GxE for persistency of production between TMR and PAST shown in previous chapters.

CHAPTER 7

General conclusion and recommendations

New model development for genetic evaluations is an ongoing process in animal breeding as research progresses and new methodologies are developed, tested and recognised. Previous research has shown the importance of distinguishing milk records originating from either of the two main production systems (TMR or PAST) used for more accurate genetic predictions. In this study it seemed logical to study the development of a possibly improved model for genetic evaluations that would be beneficial for both systems.

Non-genetic (fixed) effects are documented together with milk records. These effects do not have the same effect on milk production over the lactation and it was shown that six linear splines describe these fixed effects better than four splines, third and fourth order polynomials as well as the original Wilmink curve functions previously used. Apart from contemporary group, dissociating fixed effects and explaining them with six natural splines at DIM 6, 20, 50, 135, 245 and 300 fitted the data better, rather than concatenating such effects and attempting to describe them with one curve function. Records after DIM 300 were too scarce and created curves for fixed effect that made little biological sense. Too much erosion of number of records took place the closer to DIM 300. Milk records closer to the end of the lactation is of cardinal importance for the accurate prediction of production persistency. Where cows are milked for longer than ~300 days (e.g., cows that did not become pregnant), milk recording should continue and can be valuable for research on genetic predictions and the viability of extended lactations.

Apparent from this SA Holstein study compared to similar studies from different countries, is the inability to include variables that may help to better explain the environmental variances (residual and PE) in the SA Holstein population. These include number of days pregnant, gestation length, dry period, rearing status, and management groups with regards to feeding, production and health level. Considering these effects may have a significant impact on the SA Holstein genetic evaluation and may help to improve the lower heritability estimates, especially in TMR overall, but also during the early lactation period up to ~100 DIM. To include these effects requires the recording of additional information such as insemination dates and number, dry-off dates, and the definition of different management groups according to rearing status, feeding regime and health treatment. These extra information recordings must be accompanied by whether the recording was made under TMR or PAST, as a seemingly same non-genetic variable might not have the same effect under the two different production systems. Regarding the associated costs of recording and inclusion of these additional information (traits), if there is not realised improvement in accuracy and genetic progress, these additional recordings can at least add valuable management information to producers.

The "new" model where most (more) fixed effects were fitted as regressions over the lactation, compared to the previous SA Holstein model for genetic evaluations were extended with regards to the random effects (AG and PE). Instead of assuming that these effects, especially the genetic (AG) effect do not change over the lactation (a feature of the previous saFRM), they were allowed to vary over the lactation and described by regressions over DIM. The most significant feature of this aRRM was the fact that it allowed each TD record to be a function of a constant (average-production level) and a slope (average-persistency). The genetic merit for milk production traits are expressed in these two terms and enables the prediction of an average-production and average-persistency EBV over the lactation. These two EBVs facilitate the selection of the manner in which the quantity of milk at the end of the lactation can be produced i.e., genetically higher or lower average level of production and/or genetically higher or lower average persistency of production over the lactation. In other words, it allows for more selection strategies on how to produce milk, possibly in a more efficient way. The new aRRM fit the PAST as well as the TMR data better and will improve genetic prediction accuracy.

The aRRM showed that in addition to different levels, the shape of the AG variance, as well as the heritabilities over DIM between TMR and PAST differ significantly. This is an intrinsic property of random regression models and a shortcoming of fixed regression models. This tendency indicates GxE between the two production systems, but with the magnitude of such an interaction still unknown. Further investigation was performed, using the aRRM and analysing TMR and PAST together, but as different traits. The genetic correlation for PROD between TMR and PAST was strong (0.81 to 0.94), but weak to moderate (0.22 to 0.68) for PERS for all production traits. PERS in the two production systems are not the same trait, with different genes and/or genetic mechanisms at play over the lactation. Significant reranking of animals based on their EBV-PERS between TMR and PAST will take place.

The low to moderate genetic correlations for PERS between the TMR and PAST systems are at a level that justify the publication of two separate EBVs for persistency for each production system. This is supported by research from other authors as well. This will aid producers to select for sires that are more resilient and breed daughters that are genetically superior for PERS over TMR and PAST. Contrariwise, some producers might aim to select for sires to breed superior daughters in one or the other production system i.e., more specialised genetic lines. Whatever the case may be, enabling producers for more effective selection according to their breeding objectives implies that each participating herd must record their production system. Ideally in the future, it will be more accurate to record each test-day as coming from a TMR or PAST system, as some herds may change the feeding regime of cows according to lactation stage.

The economic importance of PERS and its effect on total (cumulative) 295-days production in both production systems are significant. Distinguishing between PROD and PERS, producers have a wider choice of selection strategies that can be used in order to produce milk in different ways within their production system. In this study it was shown that superior sires for EBV-PERS are also superior for EBV-TOTAL, irrespective of having on average the lowest EBV-PROD. Such cows with better PERS have been associated with decreased peak yield that may improve fertility and health problems. Studies have also shown that cows' different PROD and PERS profiles influence their ability to recovery from the negative energy balance after parity. More research should be done on the importance of the role of PERS in efficient milk production, as well as its genetic association with other traits of economic importance. Selection for both traits is also possible by including them in an economic selection index after economic weights have been established.

Considering the significant GxE between TMR and PAST, together with previous studies showing the average rainfall of the region of the herds' location to be a good indicator of production system used, a reaction norm model was applied. When the genetic performance of animals are regressed not only for PROD and PERS, but also genetic-producing-ability for average rainfall (GAR), the weak genetic correlations with the other two genetic components suggest that the genes involved in milk production are not consistent over different rainfall areas. Including the GAR effect in the aRRM did result in heritability estimates for low average rainfall to be closer to TMR, and high average rainfall heritabilities to be closer to PAST heritabilities. More herds in high rainfall areas that utilise a PAST system should be identified and incorporated in the study to increase accuracy of genetic parameters. Other variables that can be used as possible proxies for the two production systems should be investigated for use in a reaction norm model. However, including TMR and PAST as different traits in the aRRM has yielded by the most significant margin the best goodness of fit for the data. Publishing separate TMR and PAST genetic predictions, only for traits where significant GxE are present still seem to be the best option.

In a preliminary investigation the effect of including genomic information of the seemingly small number of genotyped animals (mostly cows with few sires) on certain groups showed surprisingly positive results for female animals. COWS (with records, but no progeny in the pedigree) showed a significant mean increase in ssGEBV compared to EBV ACC, most of all for animals that had the lowest EBVs. This is promising in the sense that the result is in line with the objective of ssGBLUP, which is to more accurately predict the genetic merit of animals that still have low ACC, albeit more for younger animals without production records. All DAMS (also those without own records) saw a modest mean increase in ssGEBV ACCs, which were in line to somewhat lower than comparable studies. These studies did however have more genotyped cows. Also, the smaller ACC increase is expected as DAMS have measured progeny, thereby leading to an already higher EBV accuracy than COWS.

Our study focused on building a dataset from cows and dams with records whose production system (TMR or PAST) had to be known and that were genotyped. The feasibility of applying a ssGBLUP to SA data were demonstrated. However, it can be argued that a comprehensive data build around all the genotyped animals may prove more beneficial to demonstrate the benefits of genomics in SA Holsteins. Although the addition of and availability of animals will only be a few hundred.

The addition of more genotyped cows, as well as related sires, whether local or foreign, is essential for improving genomic prediction ACC gains. These sires must be representative of the whole population with regards to genetic merit. The sample of the population for this study are the result of many North American bulls and with the addition of their SNP genotypes and deregressed proofs may show significant increases in ssGEBV ACCs. Possible collaborations and strategies to include such bulls must be pursued.

In cases where collaborations are challenging, implementing female genomic selection to identify superior cows. These cows can be used to breed bulls that are superior to foreign bulls with regards to production in more challenging SA climatic conditions. Such bulls can also be selected to better fit the market requirements of SA. However, it must be iterated that for such strategies to succeed, a very comprehensive and official milk recording system, together with more genotyping of cows (and bulls) be in place.

Validation of genetic predictions need to be part of the study to gauge the performance of the model used. The LR method of Legarra and Reverter (2018) is chosen in many studies because of ability to support any model and data structure. A partial dataset (p), where phenotypes of a group of genotyped animals (youngest) are removed from a complete dataset (c), where these animals do have their phenotypes, are used. Also let $\hat{u}$ be the mean GEBV. If $u_{cp} = \hat{u}_p - \hat{u}_c = 0$, the evaluation is unbiased. The slope (dispersion) of the regression of $\hat{u}_c$ on $\hat{u}_c$ will be 1 if there is not over or under dispersion (inflation or deflation). It can be calculated as $b_{c,p} = cov(\hat{u}_c, \hat{u}_p)/var(\hat{u}_p)$. To calculate the consistency between subsequent evaluations the correlation between $\hat{u}_p$ and $\hat{u}_c$ can be calculated as $r_{p,c} = cov(\hat{u}_p, \hat{u}_c)/\sqrt{var(\hat{u}_{c)}var(\hat{u}_p)}$. Lastly, the accuracy of the GEBVs can be calculated as $\widehat{ACC} = \sqrt{cov(\hat{u}_c, \hat{u}_p)/(1 - \bar{F})\sigma_a^2}$, where $\bar{F}$ is the average inbreeding coefficient of the validation animals, and σ_a^2 is the additive genetic variance. Examples where the LR calculations were used was by Kluska et al. (2021) and Sungkhapreecha et al. (2021). This might also assist in indicating the correct scaling of the G and A_{22} matrices, as well as the effectiveness of the unknown parent groups used. Some validation parameters might also point to whether the use of metafounders should be considered but should be investigated nevertheless, moving forwards. Validation is also important to make sure the reference population is kept up to date, through strategic genotyping of influential parent animals, to ensure optimal future predictions of the youngest selection candidates. Further studies should focus on the magnitude of ACC change of genetic prediction over the parent average for young animals without records.

List of tables and figures

ADDENDUM

CHAPTER 3 Supplementary material

Comparison of fixed and random regression models for the analysis of milk production traits in South African Holstein dairy cattle under two production systems

Van Niekerk, M., Neser, F.W.C., Van Wyk, J.B., Ducrocq, V., 2023. *Livest. Sci.267:1-11.*

Figure S3.1 The effect of Calving Age Class 5 (between 35.5 and 41.5 months, dotted lines) and Class 7 (between 47.5 and 55.5 months, solid lines) in relation to Class 2 (between 23.5 and 27.5, the reference, set to 0) for milk production of the separate pasture production (green lines) and the total mixed ration production (orange lines) system datasets from the alternative random regression model analyses, over days in milk (DIM) for first lactation.

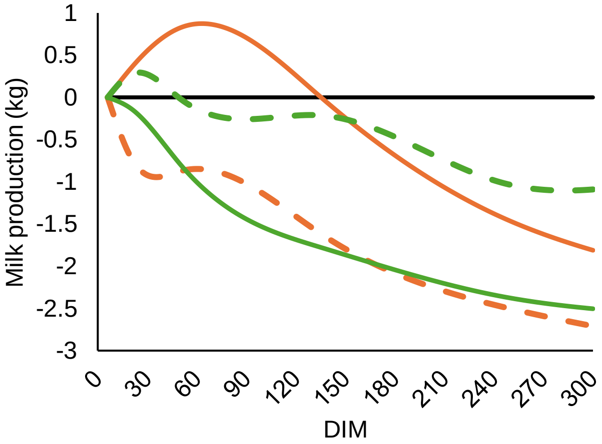


Figure S3.2 The effect of Previous Calving Interval Class 1 (between 312 and 353 days, dotted lines) and Class 3 (396 and 437 days, solid lines) in relation to Class 2 (between 354 and 395 days, the reference, set to 0) for milk production of the separate pasture production (green lines) and the total mixed ration production (orange lines) system datasets from the alternative random regression model analyses, over days in milk (DIM) for first lactation.

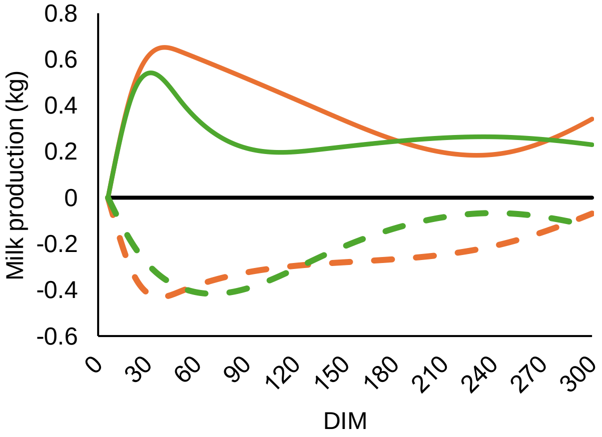


Figure S3.3 The effect of Calving Month 5 (May, dotted orange line) and Month 1 (January, solid orange line) in relation to Month 2 (February, the reference for the total mixed ration production system dataset and set to 0) and Calving Month 4 (April, short dash green line) and Month 11 (November, long dash green line) in relation to Month 8 (August, the reference for the pasture production system dataset and set to 0) for milk production from the alternative random regression model analyses, over days in milk (DIM) for first lactation.

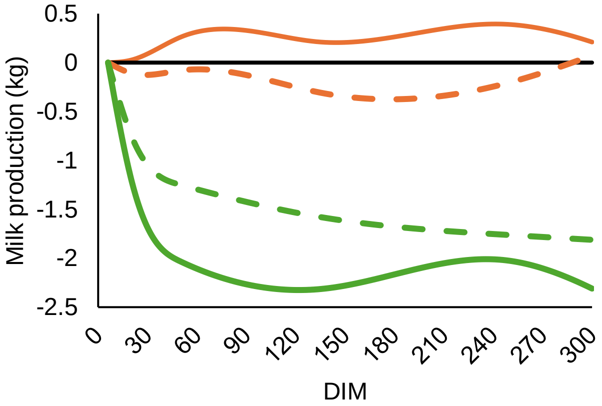


Figure S3.4 The effect of Calving Year x Lactation interaction (CYL) 65 (Calving Year 2010 x Lactation 2, solid orange line) and CYL 25 (Calving Year 1997 x Lactation 1, long dash orange line) in relation to CYL 64 (Calving year 2010 x Lactation 1, set to 0) and CYL 16 (Calving Year 1994 x Lactation 3, short dash green line) and CYL 63 (Calving Year 2010 x Lactation 2, dotted green line) in relation to CYL 62 (Calving Year 2010 x Lactation 1, set to 0) for milk production from the alternative random regression model analyses, over days in milk (DIM) for first lactation.

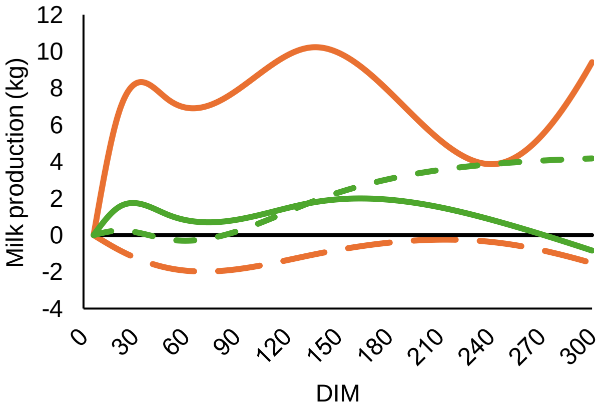


Figure S3.5 Residual variance estimates for fat production (Fat) of the separate pasture production (green lines) and the total mixed ration production (orange lines) system datasets from the South African Holstein fixed regression (dash lines) and the alternative random regression (solid lines) model analyses, over days in milk (DIM) for the first 3 lactations.

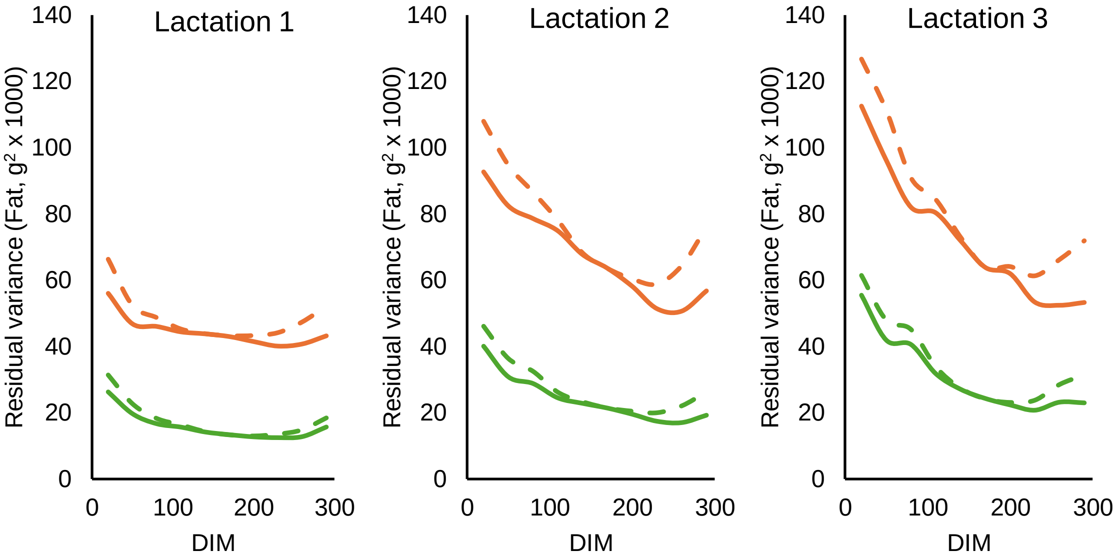


Figure S3.6 Residual variance estimates for protein production (Protein) of the separate pasture production (green lines) and the total mixed ration production (orange lines) system datasets from the South African Holstein fixed regression (dash lines) and the alternative random regression (solid lines) model analyses, over days in milk (DIM) for the first 3 lactations.

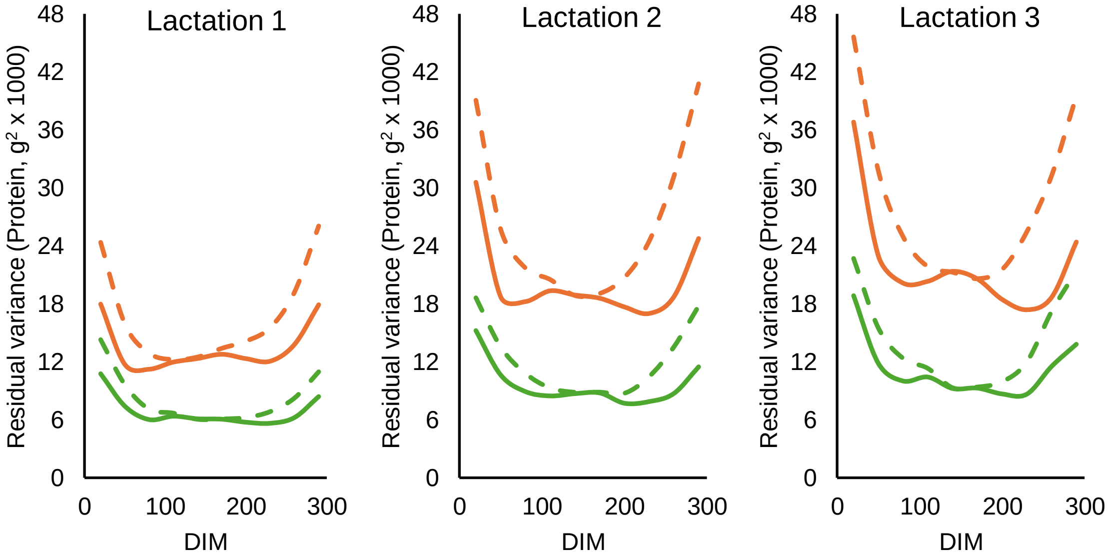


Figure S3.7 Residual variance estimates for fat content (Fat%) of the separate pasture production (green lines) and the total mixed ration production (orange lines) system datasets from the South African Holstein fixed regression (dash lines) and the alternative random regression (solid lines) model analyses, over days in milk (DIM) for the first 3 lactations.

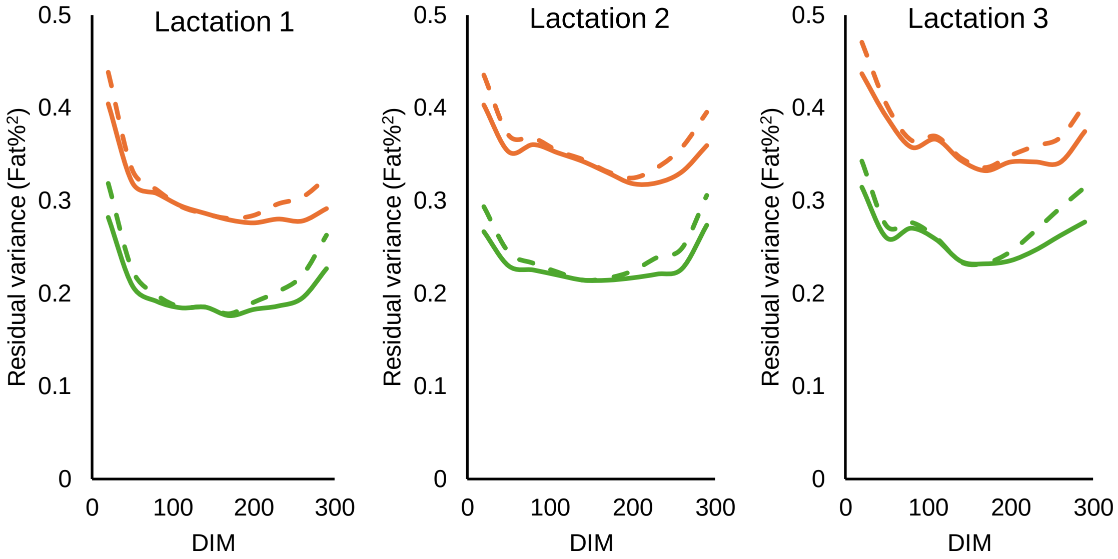


Figure S3.8 Residual variance estimates for protein content (Protein%) of the separate pasture production (green lines) and the total mixed ration production (orange lines) system datasets from the South African Holstein fixed regression (dash lines) and the alternative random regression (solid lines) model analyses, over days in milk (DIM) for the first 3 lactations.

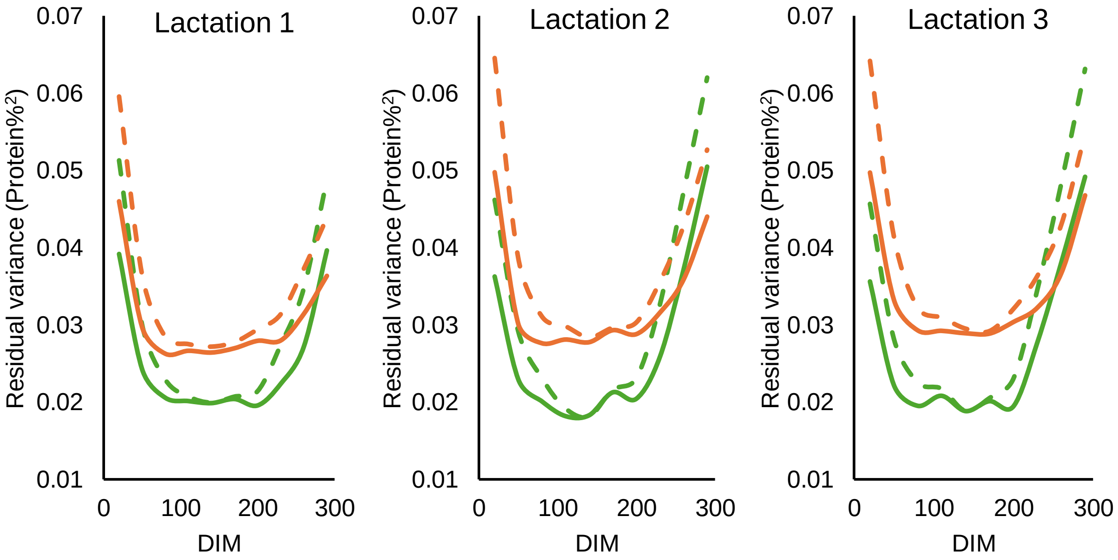


Figure S3.9 Permanent environmental variance estimates as a proportion (ratio) of the phenotypic variance for milk production (Milk) of the separate pasture production (green lines) and the total mixed ration production (orange lines) system datasets from the South African Holstein fixed regression (dash lines) and the alternative random regression (solid lines) model analyses, over days in milk (DIM) for the first 3 lactations. Standard errors ranged from 0.006 to 0.018.

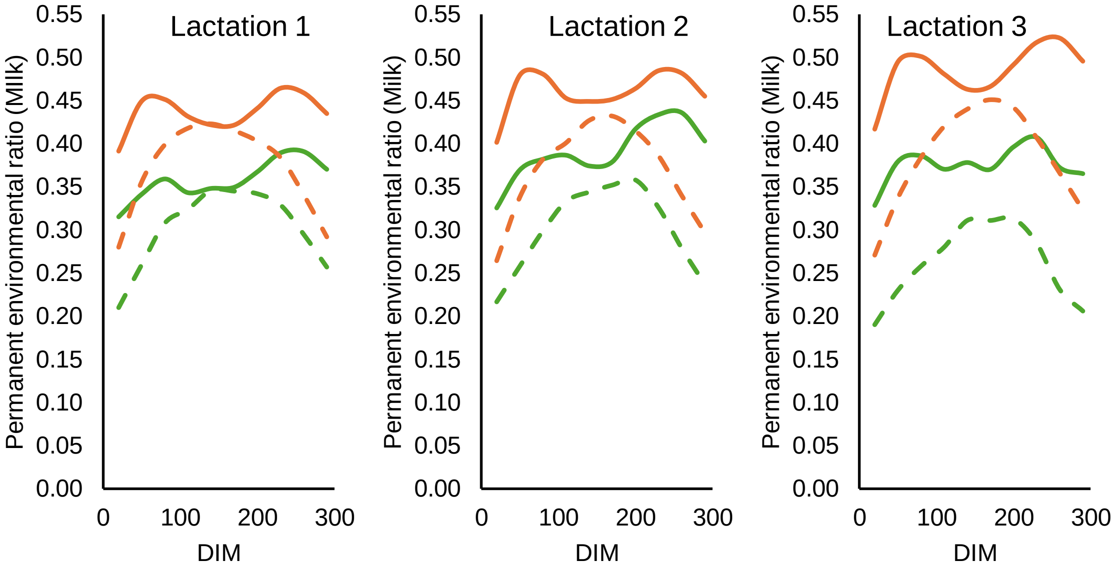


Figure S3.10 Heritability estimates for fat production (Fat) of the separate pasture production (green lines) and the total mixed ration production (orange lines) system datasets from the South African Holstein fixed regression (dash lines) and the alternative random regression (solid lines) model analyses, over days in milk (DIM) for the first 3 lactations. Standard errors ranged from 0.003 to 0.027.

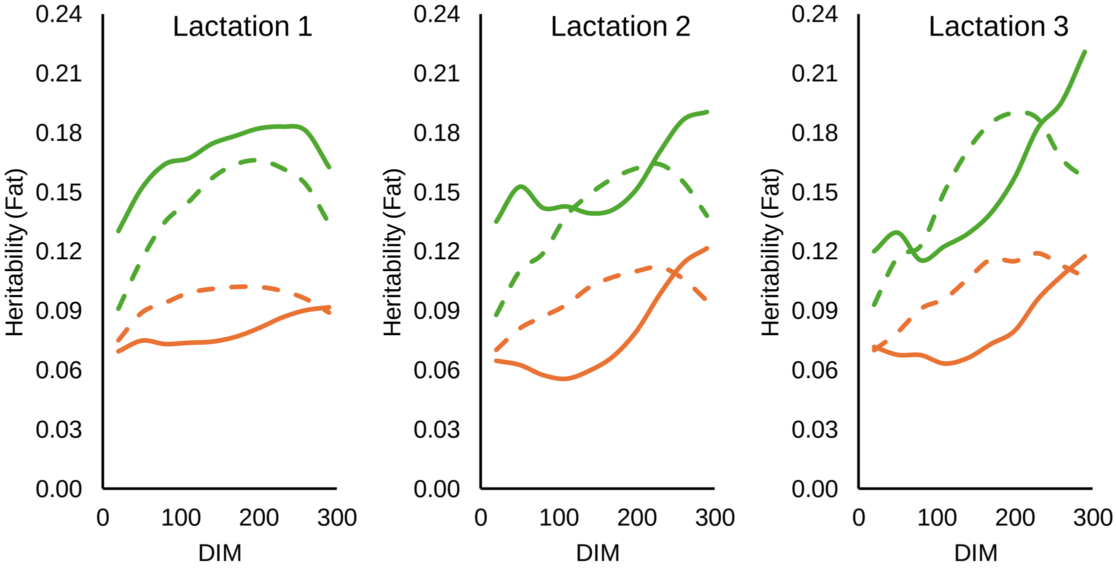


Figure S3.11 Permanent environmental variance estimates as a proportion (ratio) of the phenotypic variance for fat production (Fat) of the separate pasture production (green lines) and the total mixed ration production (orange lines) system datasets from the South African Holstein fixed regression (dash lines) and the alternative random regression (solid lines) model analyses, over days in milk (DIM) for the first 3 lactations. Standard errors ranged from 0.003 to 0.025.

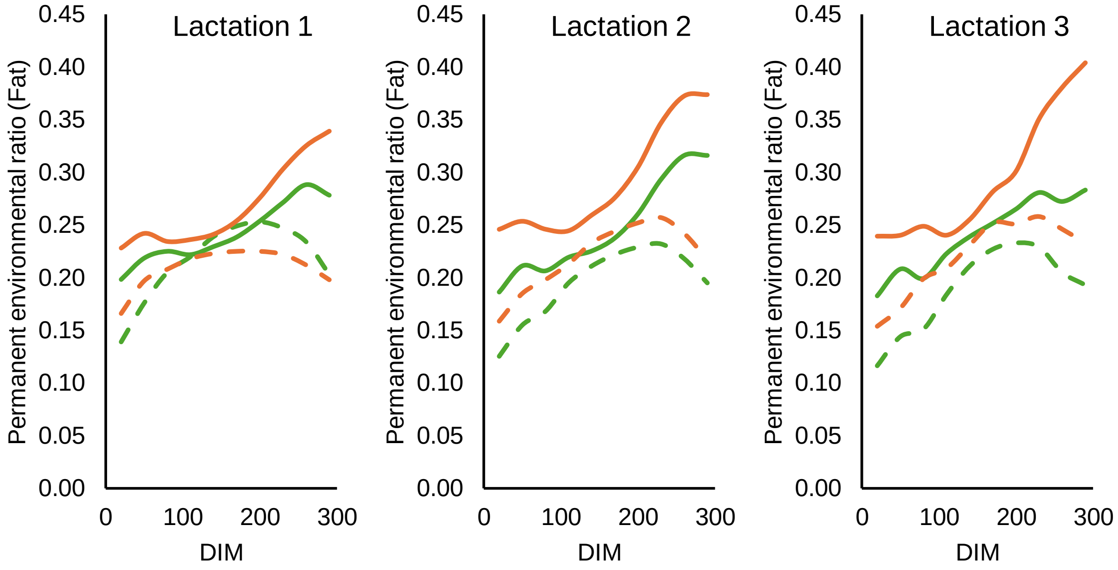


Figure S3.12 Heritability estimates for protein production (Protein) of the separate pasture production (green lines) and the total mixed ration production (orange lines) system datasets from the South African Holstein fixed regression (dash lines) and the alternative random regression (solid lines) model analyses, over days in milk (DIM) for the first 3 lactations. Standard errors ranged from 0.004 to 0.024.

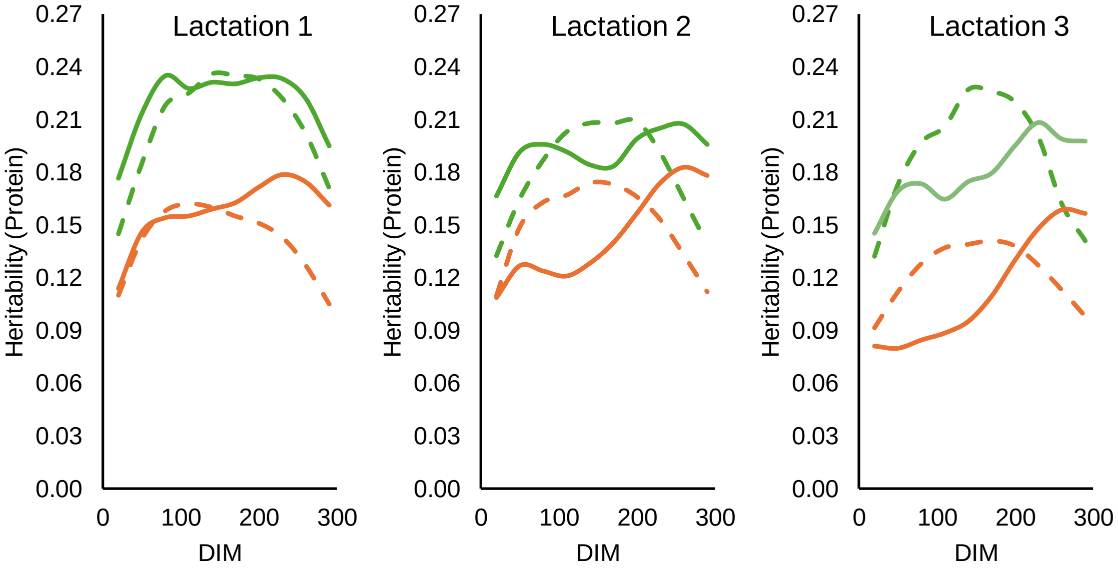


Figure S3.13 Permanent environmental variance estimates as a proportion (ratio) of the phenotypic variance for protein production (Protein) of the separate pasture production (green lines) and the total mixed ration production (orange lines) system datasets from the South African Holstein fixed regression (dash lines) and the alternative random regression (solid lines) model analyses, over days in milk (DIM) for the first 3 lactations. Standard errors ranged from 0.004 to 0.023.

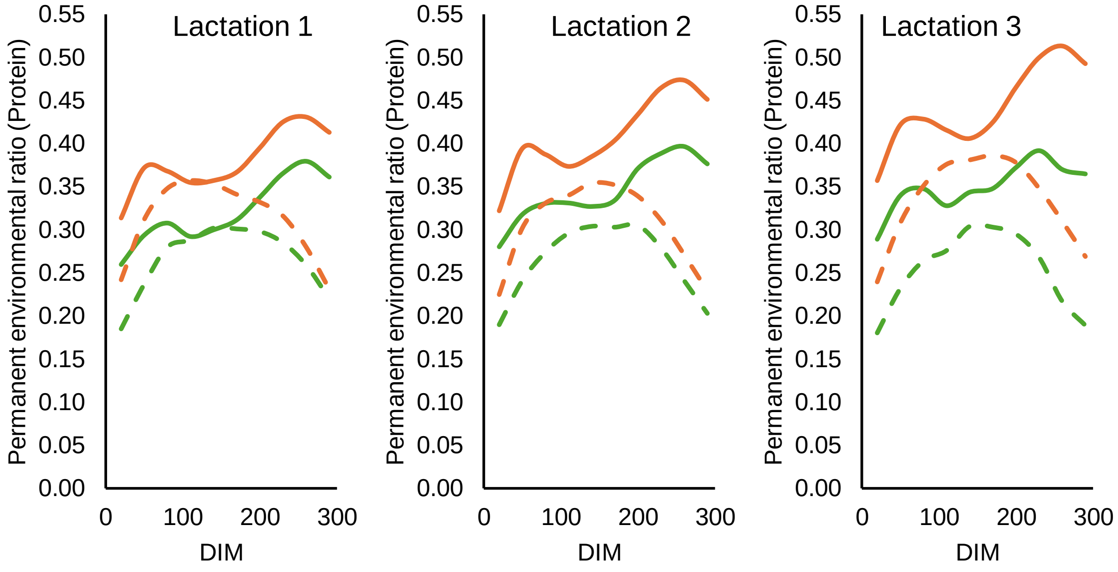


Figure S3.14 Heritability estimates for fat content (Fat%) of the separate pasture production (green lines) and the total mixed ration production (orange lines) system datasets from the South African Holstein fixed regression (dash lines) and the alternative random regression (solid lines) model analyses, over days in milk (DIM) for the first 3 lactations. Standard errors ranged from 0.004 to 0.022.

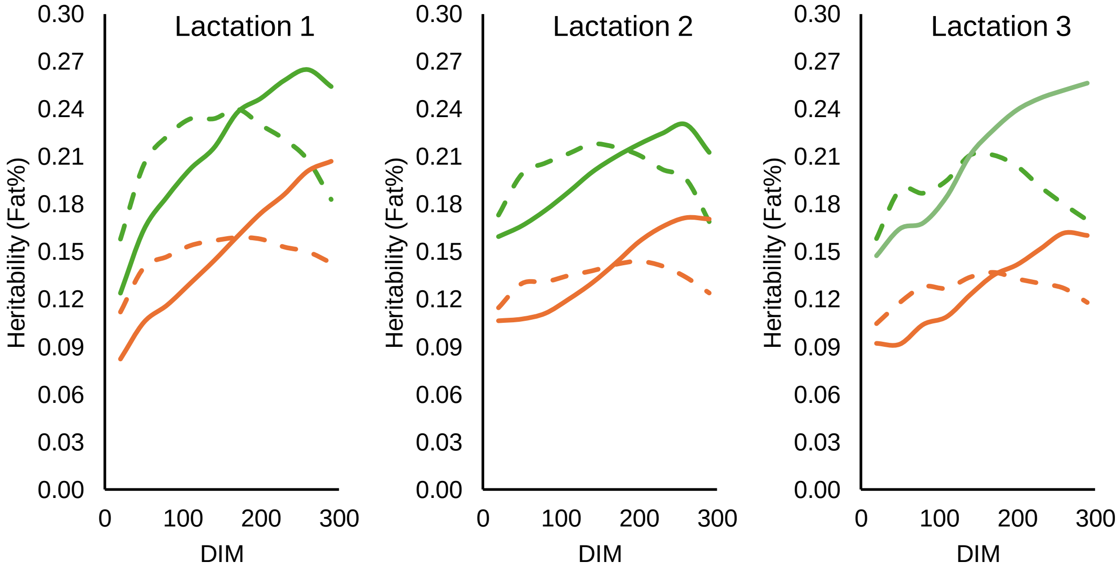


Figure S3.15 Permanent environmental variance estimates as a proportion (ratio) of the phenotypic variance for fat content (Fat%) of the separate pasture production (green lines) and the total mixed ration production (orange lines) system datasets from the South African Holstein fixed regression (dash lines) and the alternative random regression (solid lines) model analyses, over days in milk (DIM) for the first 3 lactations. Standard errors ranged from 0.003 to 0.019.

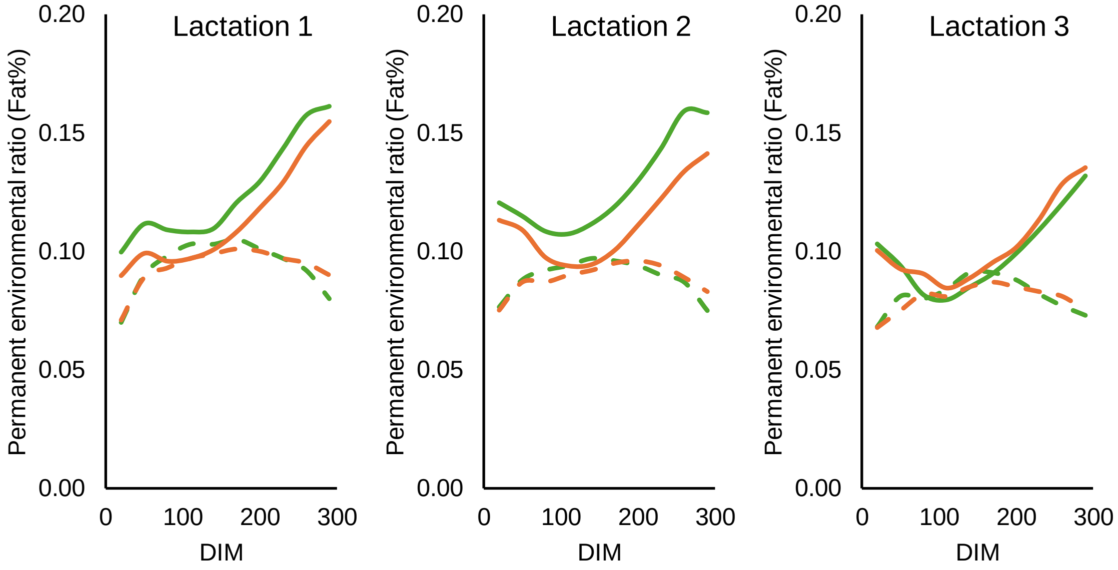


Figure S3.16 Heritability estimates for protein content (Protein%) of the separate pasture production (green lines) and the total mixed ration production (orange lines) system datasets from the South African Holstein fixed regression (dash lines) and the alternative random regression (solid lines) model analyses, over days in milk (DIM) for the first 3 lactations. Standard errors ranged from 0.005 to 0.024.

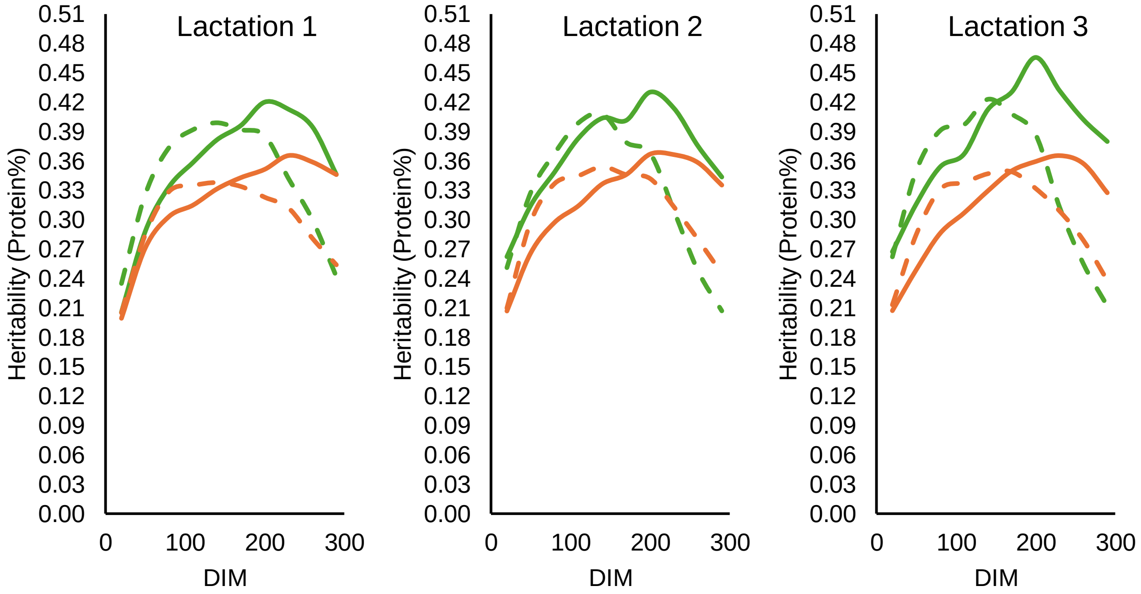


Figure S3.17 Permanent environmental variance estimates as a proportion (ratio) of the phenotypic variance for protein content (Protein%) of the separate pasture production (green lines) and the total mixed ration production (orange lines) system datasets from the South African Holstein fixed regression (dash lines) and the alternative random regression (solid lines) model analyses, over days in milk (DIM) for the first 3 lactations. Standard errors ranged from 0.003 to 0.021.

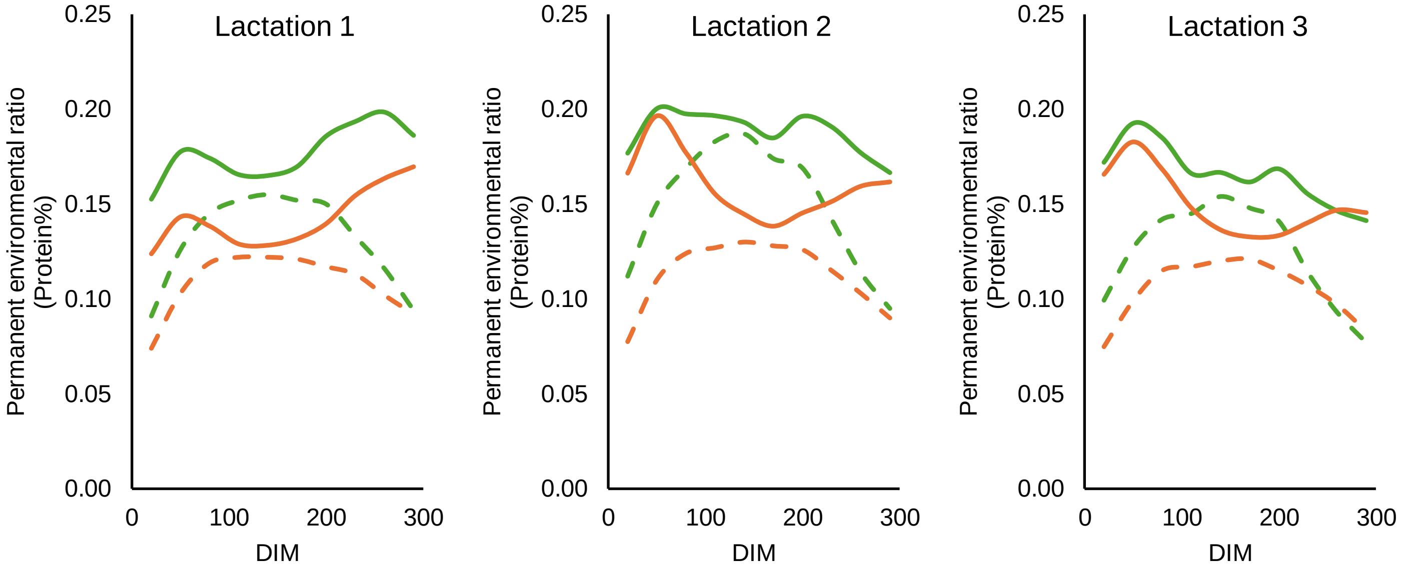


Table S3.1 Eigenvalues of the covariance matrix as a percentage of the total genetic (AG) and permanent environmental (PE) variance for milk (Milk), fat (Fat) and protein (Protein) production as well as fat (Fat%) and protein (Protein%) content using the alternative random regression model (aRRM) for the pasture (PAST) and total mixed ration (TMR) production system

Effect	Trait	Production System	nth Eigenvector using the aRRM								nth Eigenvector using the saFRM			
			1	2		3	4		5	6	1	2		3
			Lactation 1		Total[a]	Lactation 2		Total[b]	Lactation 3		Lactation 1	Lactation 2	Total[c]	Lactation 3
AG	Milk	PAST	63.5	27.8	91.3	6.8	1.4	99.5	0.4	0.0	87.7	10.1	97.8	2.2
		TMR	62.8	25.3	88.1	7.7	3.0	98.8	1.3	0.1	89.0	7.8	96.8	3.2
	Fat	PAST	68.9	25.1	94.0	4.8	0.9	99.7	0.4	0.0	88.4	8.3	96.7	3.3
		TMR	66.3	22.3	88.9	8.9	1.4	99.2	0.8	0.1	87.7	10.8	98.5	1.5
	Protein	PAST	60.6	30.2	90.8	7.1	1.8	99.7	0.4	0.0	87.7	9.7	97.4	2.6
		TMR	65.1	22.5	87.6	7.3	4.0	98.9	1.0	0.2	86.5	10.6	97.1	2.9
	Fat%	PAST	73.2	18.3	91.5	5.1	2.9	99.5	0.3	0.2	97.9	1.8	99.7	0.3
		TMR	73.6	17.9	91.5	5.8	2.5	99.8	0.2	0.0	96.1	3.3	99.4	0.7
	Protein%	PAST	73.5	19.7	93.2	3.5	2.2	98.9	0.7	0.4	93.2	5.7	98.9	1.1
		TMR	72.0	21.5	93.5	4.2	0.9	98.6	0.9	0.5	94.1	4.8	98.9	1.1
PE	Milk	PAST	35.5	22.0	57.5	18.7	12.7	88.9	6.2	4.8	62.4	21.4	83.8	16.2
		TMR	37.2	21.4	58.6	18.0	11.8	88.4	7.0	4.6	60.3	24.1	84.4	15.6
	Fat	PAST	33.1	22.3	55.4	19.4	14.4	89.2	6.2	4.6	67.0	18.5	85.5	14.5
		TMR	38.7	21.4	60.1	17.9	11.6	89.6	6.1	4.3	62.2	23.4	85.6	14.4
	Protein	PAST	37.4	21.3	58.7	18.8	13.3	90.8	5.3	3.9	66.4	19.5	85.9	14.1
		TMR	39.2	23.3	62.5	17	11.3	90.8	5.4	3.8	61.8	23.5	85.3	14.7
	Fat%	PAST	44.9	22.1	67.0	15.6	12.6	95.2	3.2	1.6	78.1	14.1	92.2	7.9
		TMR	39.8	23.6	63.4	15.6	14.2	93.2	4.1	2.7	77.5	13.7	91.2	8.7
	Protein%	PAST	49.5	18.3	67.8	14.1	10.5	92.4	4.2	3.4	71.8	17.2	89.0	11.0
		TMR	44.7	19.5	64.2	17.0	11.6	92.8	3.8	3.5	71.9	17.3	89.2	10.8

[a] = the total of eigenvalues (%) 1 and 2 (lactation 1); [b] = the total of eigenvalues (%) 1, 2, 3 and 4 (lactation 1 and 2); [c] = the total of eigenvalues (%) 1 and 2 (lactation 1 and 2)

Table S3.2 Genetic correlations between the first 3 lactations for average fat production for the South African Holstein fixed regression model (saFRM) versus the alternative random regression model (aRRM) for the pasture (PAST) and total mixed ration (TMR) datasets, with standard errors in parentheses

	PAST		TMR	
Lactations	saFRM	aRRM	saFRM	aRRM
1,2	0.79 (0.04)	0.96 (0.02)	0.67 (0.04)	0.83 (0.03)
1,3	0.71 (0.05)	0.93 (0.03)	0.66 (0.04)	0.87 (0.05)
2,3	0.90 (0.03)	0.99 (0.03)	0.96 (0.02)	0.99 (0.04)

ij = between lactations i and j

Table S3.3 Genetic correlations between the first 3 lactations for average protein production for the South African Holstein fixed regression model (saFRM) versus the alternative random regression model (aRRM) for the pasture (PAST) and total mixed ration (TMR) datasets, with standard errors in parentheses

	PAST		TMR	
Lactations	saFRM	aRRM	saFRM	aRRM
1,2	0.76 (0.03)	0.93 (0.02)	0.70 (0.03)	0.86 (0.02)
1,3	0.71 (0.04)	0.89 (0.03)	0.66 (0.04)	0.86 (0.04)
2,3	0.92 (0.02)	0.98 (0.02)	0.92 (0.03)	0.98 (0.03)

ij = between lactations i and j

Table S3.4 Genetic correlations between the first 3 lactations for average fat content (%) for the South African Holstein fixed regression model (saFRM) versus the alternative random regression model (aRRM) for the pasture (PAST) and total mixed ration (TMR) datasets, with standard errors in parentheses

	PAST		TMR	
Lactations	saFRM	aRRM	saFRM	aRRM
1,2	0.97 (0.01)	0.98 (0.01)	0.94 (0.01)	0.99 (0.01)
1,3	0.95 (0.02)	0.97 (0.01)	0.91 (0.02)	0.99 (0.01)
2,3	0.99 (0.01)	0.99 (0.01)	0.97 (0.01)	0.99 (0.01)

ij = between lactations i and j

Table S3.5 Genetic correlations between the first 3 lactations for average protein content (%) for the South African Holstein fixed regression model (saFRM) versus the alternative random regression model (aRRM) for the pasture (PAST) and total mixed ration (TMR) datasets, with standard errors in parentheses

	PAST		TMR	
Lactations	saFRM	aRRM	saFRM	aRRM
1,2	0.88 (0.02)	0.94 (0.01)	0.90 (0.01)	0.95 (0.01)
1,3	0.84 (0.02)	0.94 (0.01)	0.86 (0.01)	0.95 (0.01)
2,3	0.96 (0.01)	0.96 (0.01)	0.97 (0.01)	0.96 (0.01)

ij = between lactations i and j

Table S3.6 Genetic correlations between the two components of the random regression model for fat production [average fat production (PROD) and persistency (PERS)] within and between the first 3 lactations for the pasture (PAST) and the total mixed ration (TMR) production systems, with standard errors in parentheses

		TMR					
PAST		Lactation 1		Lactation 2		Lactation 3	
		PROD	PERS	PROD	PERS	PROD	PERS
Lactation 1	PROD		-0.15 (0.07)	0.83 (0.03)	-0.15 (0.07)	0.87 (0.05)	-0.07 (0.08)
	PERS	0.07 (0.09)		-0.30 (0.08)	0.29 (0.08)	-0.22 (0.09)	0.08 (0.11)
Lactation 2	PROD	0.96 (0.02)	-0.01 (0.10)		-0.23 (0.07)	0.99 (0.04)	-0.13 (0.09)
	PERS	0.12 (0.07)	0.47 (0.10)	0.07 (0.08)		-0.19 (0.09)	0.95 (0.05)
Lactation 3	PROD	0.93 (0.03)	-0.07 (0.11)	0.99 (0.03)	-0.10 (0.09)		-0.10 (0.10)
	PERS	0.12 (0.07)	0.46 (0.11)	0.05 (0.08)	0.98 (0.04)	-0.04 (0.09)	

Table S3.7 Genetic correlations between the two components of the random regression model for protein production [average protein production (PROD) and persistency (PERS)] within and between the first 3 lactations for the pasture (PAST) and the total mixed ration (TMR) production systems, with standard errors in parentheses

PAST		TMR					
		Lactation 1		Lactation 2		Lactation 3	
		PROD	PERS	PROD	PERS	PROD	PERS
Lactation 1	PROD		-0.41 (0.05)	0.86 (0.02)	-0.08 (0.05)	0.86 (0.04)	-0.21 (0.07)
	PERS	-0.08 (0.08)		-0.48 (0.05)	0.25 (0.07)	-0.43 (0.07)	0.18 (0.09)
Lactation 2	PROD	0.93 (0.02)	-0.22 (0.08)		-0.38 (0.05)	0.98 (0.03)	-0.43 (0.07)
	PERS	0.09 (0.07)	0.36 (0.10)	-0.07 (0.08)		-0.40 (0.07)	0.85 (0.05)
Lactation 3	PROD	0.89 (0.03)	-0.29 (0.09)	0.98 (0.02)	-0.20 (0.08)		-0.47 (0.08)
	PERS	0.05 (0.08)	0.29 (0.11)	-0.01 (0.08)	0.96 (0.04)	-0.16 (0.09)	

Table S3.8 Genetic correlations between the two components of the random regression model for fat content (%) [average fat content (PROD) and persistency (PERS)] within and between the first 3 lactations for the pasture (PAST) and the total mixed ration (TMR) production systems, with standard errors in parentheses

PAST		TMR					
		Lactation 1		Lactation 2		Lactation 3	
		PROD	PERS	PROD	PERS	PROD	PERS
Lactation 1	PROD		-0.58 (0.04)	0.99 (0.01)	-0.40 (0.07)	0.99 (0.01)	-0.49 (0.10)
	PERS	-0.55 (0.06)		-0.60 (0.05)	0.64 (0.09)	-0.58 (0.05)	0.65 (0.13)
Lactation 2	PROD	0.98 (0.01)	-0.55 (0.06)		-0.48 (0.07)	0.99 (0.01)	-0.49 (0.10)
	PERS	-0.52 (0.09)	0.75 (0.11)	-0.47 (0.09)		-0.48 (0.08)	0.83 (0.13)
Lactation 3	PROD	0.97 (0.01)	-0.51 (0.07)	0.99 (0.01)	-0.41 (0.10)		-0.51 (0.11)
	PERS	-0.50 (0.09)	0.75 (0.12)	-0.53 (0.10)	0.65 (0.15)	-0.50 (0.10)	

Table S3.9 Genetic correlations between the two components of the random regression model for protein content (%) [average protein content (PROD) and persistency (PERS)] within and between the first 3 lactations for the pasture (PAST) and the total mixed ration (TMR) production systems, with standard errors in parentheses

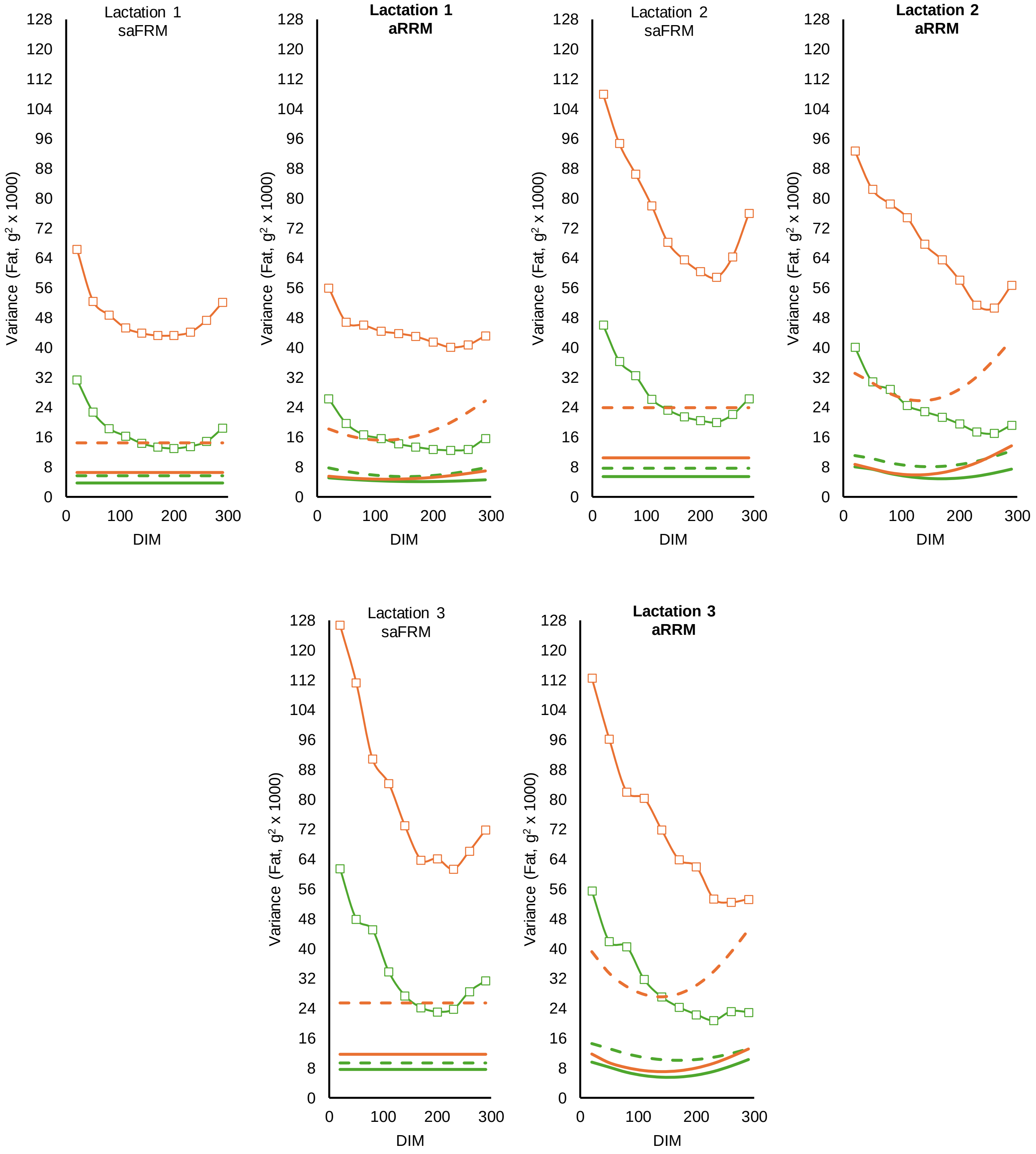
		TMR					
		Lactation 1		Lactation 2		Lactation 3	
PAST		PROD	PERS	PROD	PERS	PROD	PERS
Lactation 1	PROD		-0.45 (0.03)	0.95 (0.01)	-0.44 (0.04)	0.95 (0.01)	-0.48 (0.05)
	PERS	-0.55 (0.04)		-0.55 (0.03)	0.82 (0.04)	-0.50 (0.04)	0.74 (0.06)
Lactation 2	PROD	0.94 (0.01)	-0.59 (0.04)		-0.49 (0.04)	0.96 (0.01)	-0.53 (0.06)
	PERS	-0.52 (0.04)	0.78 (0.05)	-0.49 (0.05)		-0.48 (0.04)	0.95 (0.05)
Lactation 3	PROD	0.94 (0.01)	-0.61 (0.04)	0.96 (0.01)	-0.55 (0.05)		-0.51 (0.06)
	PERS	-0.48 (0.05)	0.85 (0.05)	-0.54 (0.05)	0.88 (0.04)	-0.54 (0.05)	

Table S3.10 The comparison of the Bayesian Information Criterion (BIC) of some of the fixed regression models (FRM) applied to milk production in the pasture production system, beginning with the current official FRM used for South African Holstein genetic evaluations (shown here as saFRMw). The second model shown are a FRM with the alternative fixed effects structure (a), still using the Wilmink (w) regression function (aFRMw). The third model only differs from the second with regards to the fixed regression function (splines; s) used (aFRMs). The last model shown are the final, alternative random regression model (shown here as aRRM6) used in this study, where the only difference to the aFRMs is the regression coefficient that was fitted over the lactation for the random effects.

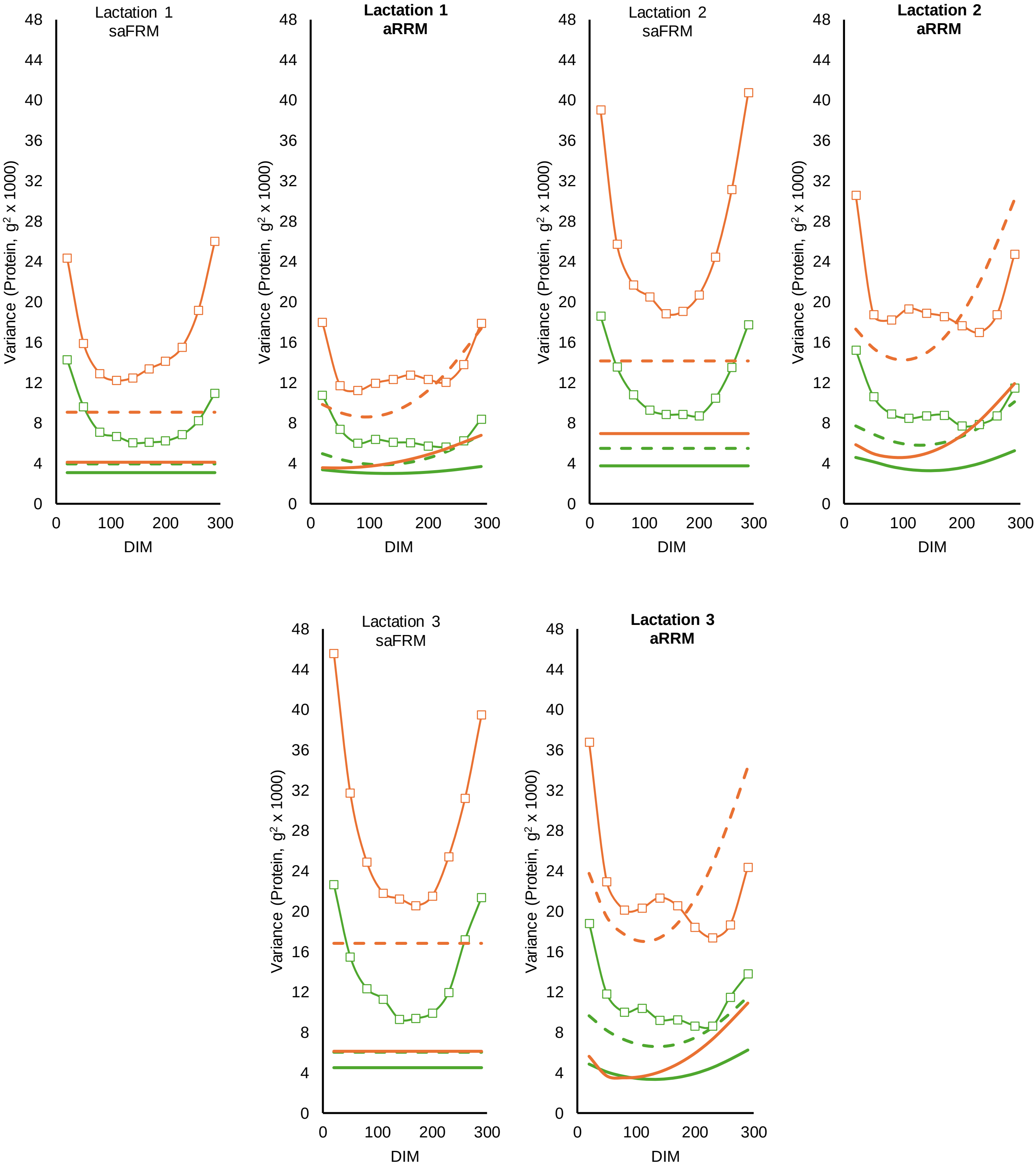
Model	Fixed effects (class)	Fixed regressions on DIM	Fixed regression curve function	Random effects (constant)	Random regressions on DIM	Random regression function	Difference in BIC compared to saFRMw
saFRMw	HTMi HYMo CIC	SAC	Wilmink	AG PE	n/a	n/a	0
aFRMw	HTMi **HYSP**	**Herd** **CYL** **Mo** **AC** **CIC**	Wilmink	AG PE	n/a	n/a	-2,294.46
aFRMs	HTMi HYSP	Herd CYL Mo AC CIC	**Natural splines with 6 knots**	AG PE	n/a	n/a	-2,634.59
aRRMs	HTMi HYSP	Herd CYL Mo AC CIC	Natural splines with 6 knots	n/a	**AG** **PE**	**Constant and a slope**	-12,246.35

HTMi = Herd x Test-day x nr of Milkings; HYMo = Herd x calving Year x calving Month; HYSP = Herd x calving Season x Parity; SAC = calving Season x Age Class; CYL = Calving Year x Lactation nr; Mo = calving Month; AC = calving Age Class; CIC = previous Calving Interval Class; AG = animal Additive Genetic; PE = Permanent Environmental; n/a = not applicable; DIM = Days In Milk; Wilmink = Wilmink function *(Wilmink, J.B.M., 1987. Livest. Prod. Sci. 16:335-348).*

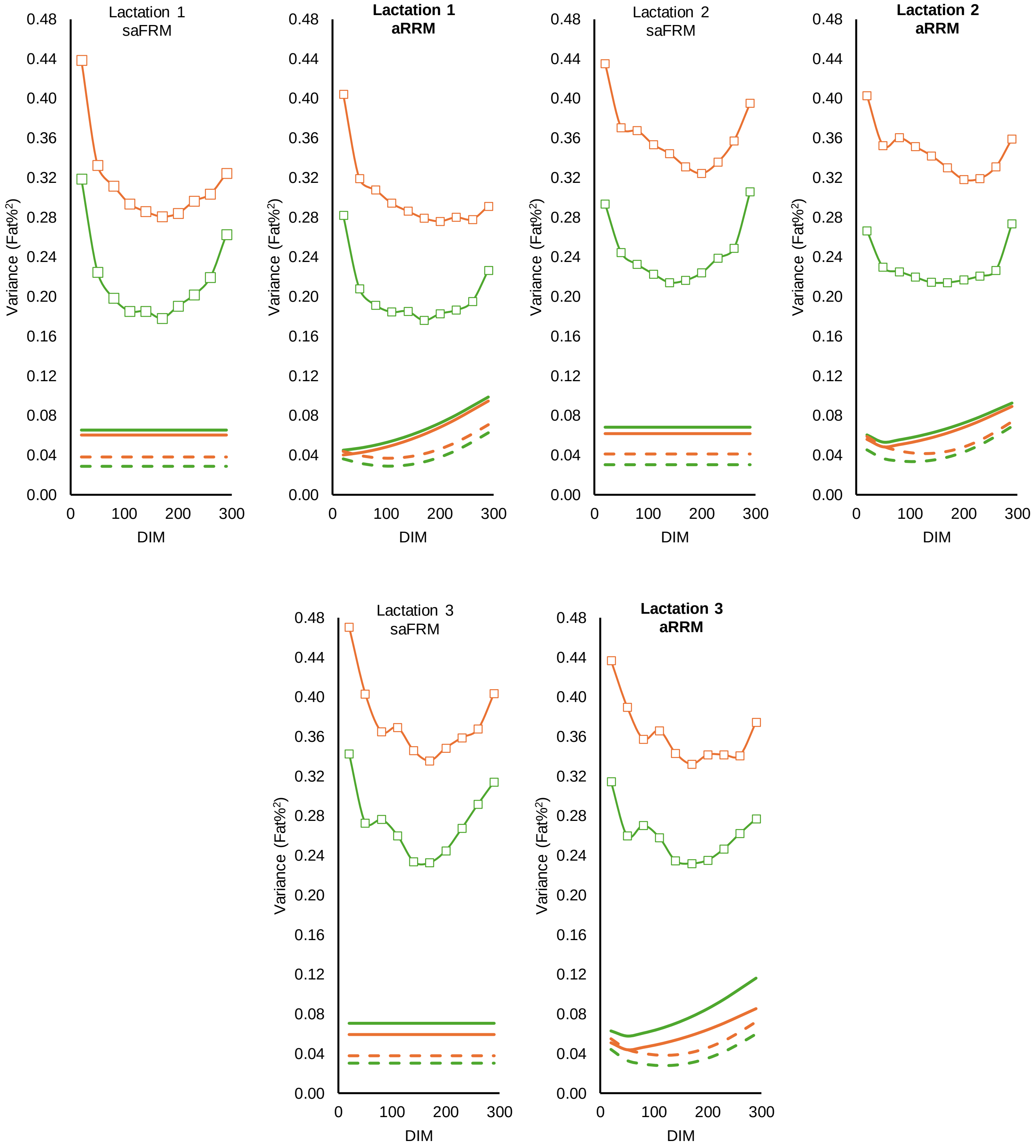
Figures S3.18a, -b (top) and -c (bottom). Variance components for fat production (Fat) of the genetic (solid lines), permanent environmental (dash lines), residual (solid lines with squared markers) effects the pasture (green lines) and total mixed ration (orange lines) using the South African fixed regression (saFRM, left in each lactation) and the alternative random regression (aRRM, right in each lactation) model analyses, over days in milk (DIM) for the first 3 lactations.



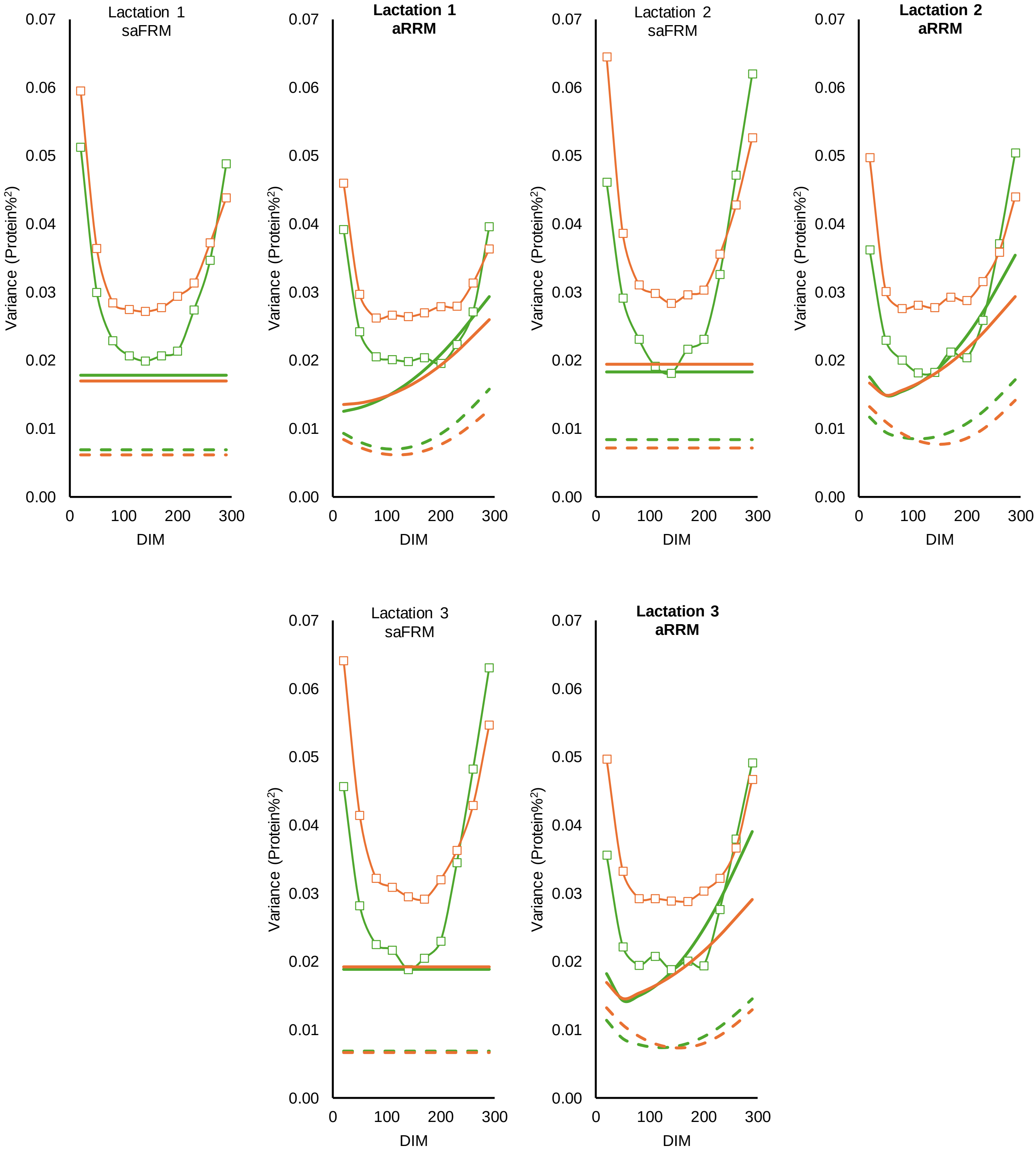
Figures S3.19a, -b (top) and -c (bottom). Variance components for protein production (Protein) of the genetic (solid lines), permanent environmental (dash lines), residual (solid lines with squared markers) effects the pasture (green lines) and total mixed ration (orange lines) using the South African fixed regression (saFRM, left in each lactation) and the alternative random regression (aRRM, right in each lactation) model analyses, over days in milk (DIM) for the first 3 lactations.



Figures S3.20a, -b (top) and -c (bottom). Variance components for fat content (Fat%) of the genetic (solid lines), permanent environmental (dash lines), residual (solid lines with squared markers) effects the pasture (green lines) and total mixed ration (orange lines) using the South African fixed regression (saFRM, left in each lactation) and the alternative random regression (aRRM, right in each lactation) model analyses, over days in milk (DIM) for the first 3 lactations.

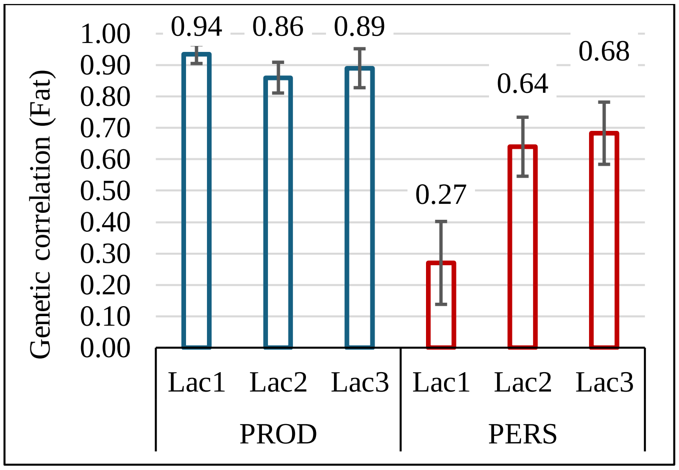


Figures S3.21a, -b (top) and -c (bottom). Variance components for fat content (Protein%) of the genetic (solid lines), permanent environmental (dash lines), residual (solid lines with squared markers) effects the pasture (green lines) and total mixed ration (orange lines) using the South African fixed regression (saFRM, left in each lactation) and the alternative random regression (aRRM, right in each lactation) model analyses, over days in milk (DIM) for the first 3 lactations.



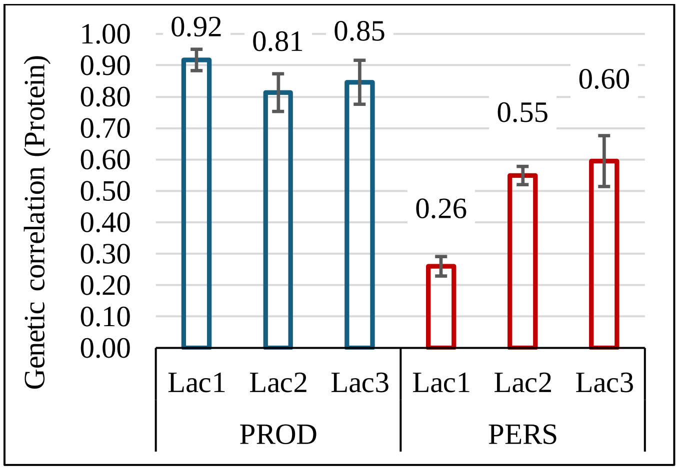
CHAPTER 4 Supplementary material

Figure S4.1 Genetic correlations (standard errors) between the total mixed ration and pasture production systems for Fat production



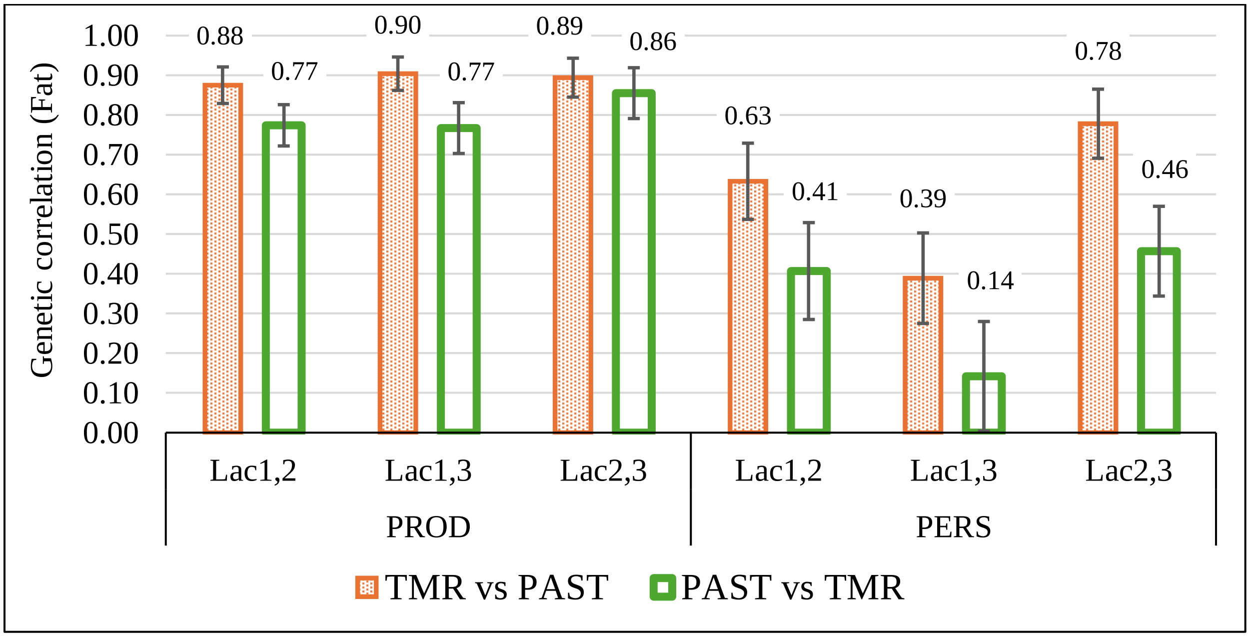
Fat = Fat production; PROD = Average-production (blue); PERS = Average-persistency (red); Laci = Lactation i

Figure S4.2 Genetic correlations (standard errors) between the total mixed ration and pasture production systems for Protein production



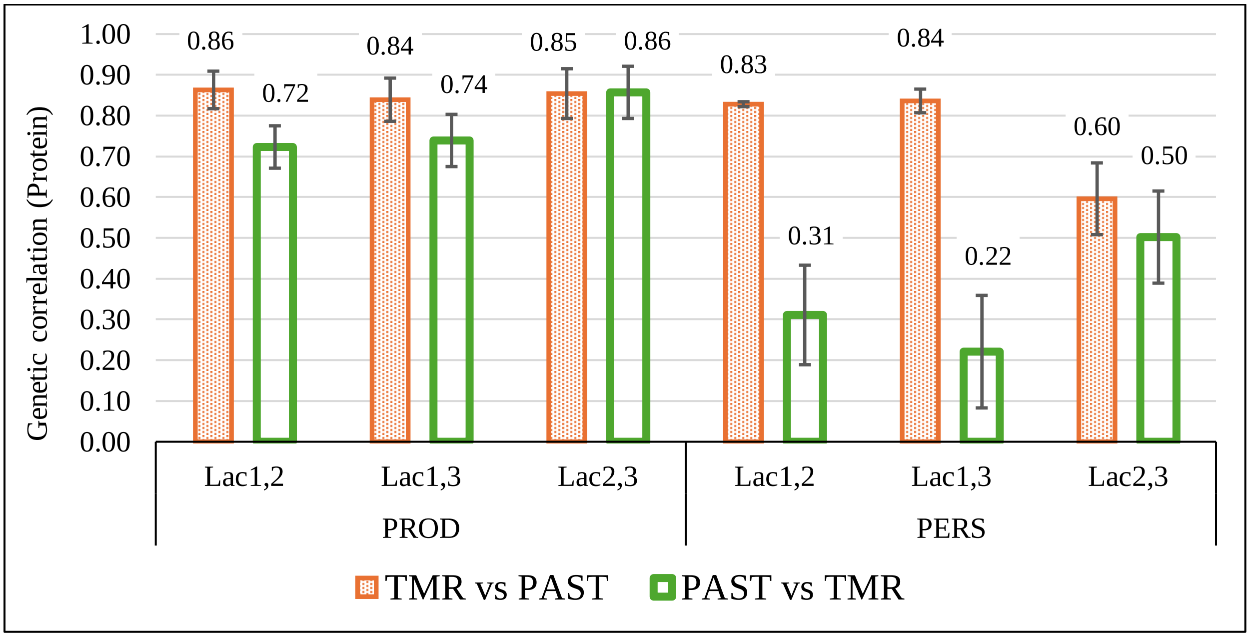
Protein = Protein production; PROD = Average-production (blue); PERS = Average-persistency (red); Laci = Lactation i

Figure S4.3 Genetic correlations (standard errors) between lactations for the total mixed ration vs pasture production systems for Fat production



Fat = Fat production; PROD = Average-production; PERS = Average-persistency; Laci,j = Lactation i,j; TMR = Total Mixed Ration production system (orange); PAST = Pasture production system (green)

Figure S4.4 Genetic correlations (standard errors) between lactations for the total mixed ration vs pasture production systems for Protein production



Protein = Protein production; PROD = Average-production; PERS = Average-persistency; Laci,j = Lactation i,j; TMR = Total Mixed Ration production system (orange); PAST = Pasture production system (green)

Table S4.1 Genetic correlations (Milk) within and between the TMR and PAST production systems

			TMR						PAST					
			Lac1		**Lac2**		**Lac3**		**Lac1**		**Lac2**		**Lac3**	
			PROD	**PERS**	**PROD**	**PERS**	**PROD**	**PERS**	**PROD**	**PERS**	**PROD**	**PERS**	**PROD**	**PERS**
TMR	**Lac1**	**PROD**	1.00											
		PERS	-0.33_5	1.00										
	Lac2	**PROD**	0.89_2	-0.47_5	1.00									
		PERS	-0.12_5	0.27_6	-0.24_5	1.00								
	Lac3	**PROD**	0.86_4	-0.42_7	0.95_3	-0.19_7	1.00							
		PERS	-0.09_7	0.09_8	-0.19_7	0.90_4	-0.12_9	1.00						
PAST	**Lac1**	**PROD**	**0.92_3**	-0.24_7^e	0.75_5^b	0.01_7^e	0.77_6^b	-0.07_9^e	1.00					
		PERS	-0.03_8^e	**0.22_{10}**	-0.07_9^e	0.35_{10}^d	0.11_{11}^e	0.21_{12}^d	0.01_7	1.00				
	Lac2	**PROD**	0.88_4^a	-0.27_8^e	**0.85_5**	-0.01_8^e	0.87_5^b	-0.08_{10}^e	0.94_2	-0.08_7	1.00			
		PERS	-0.03_7^e	0.71_7^c	-0.17_8^e	**0.64_8^b**	-0.16_9^e	0.52_{10}^d	0.02_6	0.34_8	-0.03_7	1.00		
	Lac3	**PROD**	0.84_5^a	-0.49_8^e	0.87_6^a	-0.04_9^e	**0.89_6**	-0.08_{11}^e	0.89_3	-0.05_8	0.96_2	-0.15_8	1.00	
		PERS	-0.15_8^e	0.70_8^c	-0.23_9^e	0.71_8^c	-0.16_{10}^e	**0.59_{10}**	-0.04_7	0.37_8	-0.03_8	0.95_3	-0.13_8	1.00

Milk = Milk production; TMR = Total mixed production system; PAST = Pasture production system; Laci = Lactation i; PROD = Genetic-average-production effect; PERS = Genetic-average-persistency effect; Standard errors: Subscript i * 10^{-2} ; **Bold** = Within lactation genetic correlations, PROD and PERS effects between TMR and PAST**;** Superscripts: a = Between lactations genetic correlations of PROD effects, TMR vs PAST; b = Between lactations genetic correlations of PROD effects, PAST vs TMR; c = Between lactations genetic correlations of PERS effects, TMR vs PAST; d = Between lactations genetic correlations of PERS effects, PAST vs TMR; e = Within and between lactation genetic correlations, PROD vs PERS effects (and vice versa) between TMR and PAST.

Table S4.2 Genetic correlations (Fat) within and between the TMR and PAST production systems

			TMR						PAST					
			Lac1		**Lac2**		**Lac3**		**Lac1**		**Lac2**		**Lac3**	
			PROD	**PERS**	**PROD**	**PERS**	**PROD**	**PERS**	**PROD**	**PERS**	**PROD**	**PERS**	**PROD**	**PERS**
TMR	**Lac1**	**PROD**	1.00											
		PERS	-0.23_{6}	1.00										
	Lac2	**PROD**	0.86_{3}	-0.36_{7}	1.00									
		PERS	-0.05_{6}	0.24_{8}	-0.13_{7}	1.00								
	Lac3	**PROD**	0.87_{4}	-0.27_{8}	0.97_{2}	-0.06_{8}	1.00							
		PERS	0.06_{7}	-0.03_{9}	0.06_{8}	0.87_{5}	0.14_{9}	1.00						
PAST	**Lac1**	**PROD**	**0.94_{3}**	-0.09_{9}^{e}	0.77_{5}^{b}	0.11_{8}^{e}	0.77_{6}^{b}	0.11_{10}^{e}	1.00					
		PERS	0.08_{10}^{e}	**0.27_{13}**	0.03_{11}^{e}	0.41_{12}^{d}	0.03_{13}^{e}	0.14_{14}^{d}	0.03_{9}	1.00				
	Lac2	**PROD**	0.87_{5}^{a}	-0.08_{10}^{e}	**0.86_{5}**	0.12_{9}^{e}	0.86_{6}^{b}	0.15_{10}^{e}	0.95_{2}	-0.02_{9}	1.00			
		PERS	-0.16_{8}^{e}	0.63_{9}^{c}	-0.29_{9}^{e}	**0.64_{9}**	-0.32_{10}^{e}	0.46_{11}^{d}	-0.01_{7}	0.44_{10}	-0.06_{8}	1.00		
	Lac3	**PROD**	0.90_{4}^{a}	-0.24_{10}^{e}	0.89_{5}^{a}	0.15_{9}^{e}	**0.89_{6}**	0.27_{10}^{e}	0.91_{3}	-0.02_{10}	0.96_{3}	-0.09_{8}	1.00	
		PERS	-0.19_{9}^{e}	0.39_{11}^{c}	-0.26_{10}^{e}	0.78_{9}^{c}	-0.28_{11}^{e}	**0.68_{10}**	-0.06_{8}	0.44_{10}	-0.10_{8}	0.93_{5}	-0.09_{9}	1.00

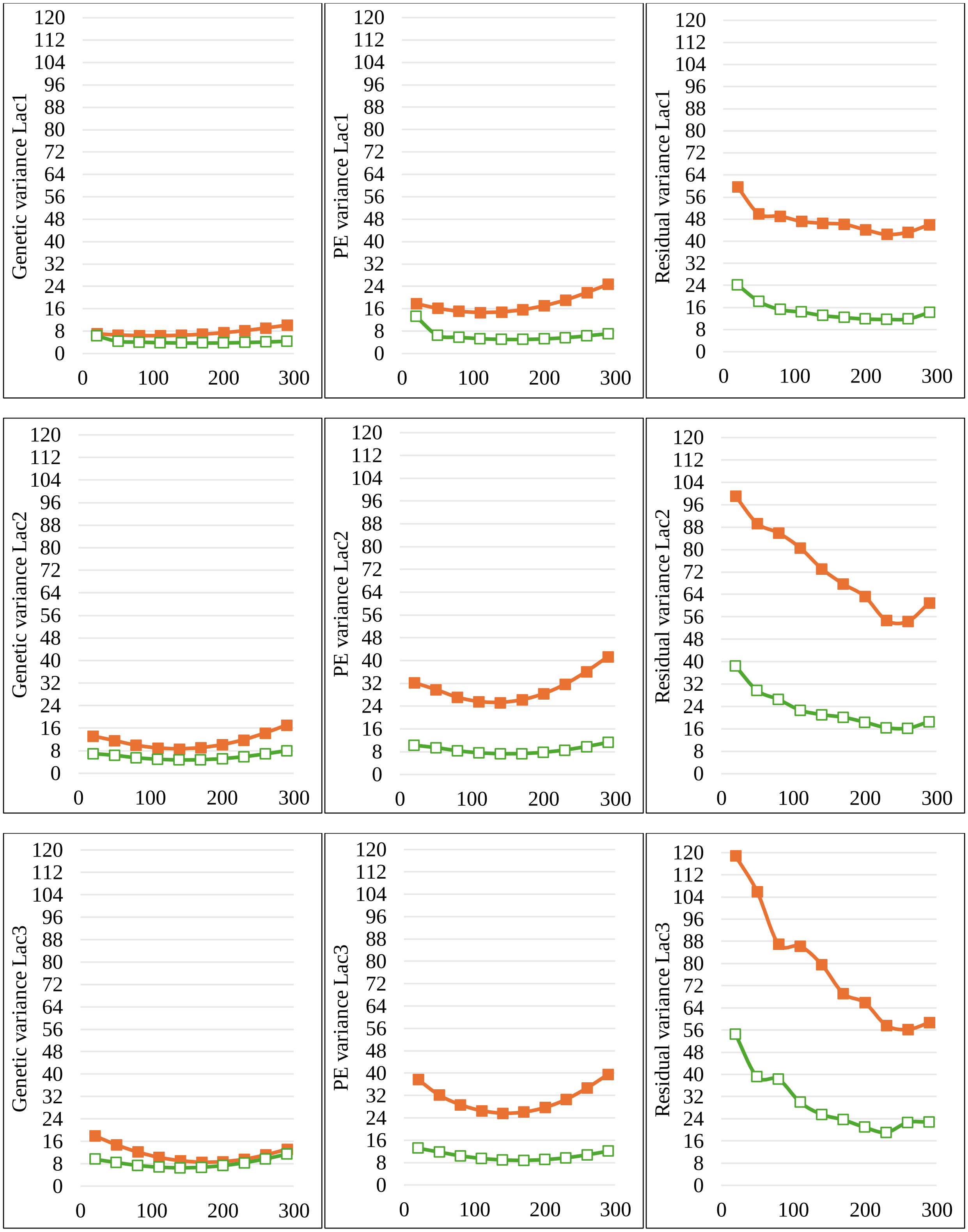
Fat = Fat production; TMR = Total mixed production system; PAST = Pasture production system; Laci = Lactation i; PROD = Genetic-average-production effect; PERS = Genetic-average-persistency effect; Standard errors: Subscript i * 10^{-2}; **Bold** = Within lactation genetic correlations, PROD and PERS effects between TMR and PAST**;** Superscripts: a = Between lactations genetic correlations of PROD effects, TMR vs PAST; b = Between lactations genetic correlations of PROD effects, PAST vs TMR; c = Between lactations genetic correlations of PERS effects, TMR vs PAST; d = Between lactations genetic correlations of PERS effects, PAST vs TMR; e = Within and between lactation genetic correlations, PROD vs PERS effects (and vice versa) between TMR and PAST.

Table S4.3 Genetic correlations (Protein) within and between the TMR and PAST production systems

			TMR						**PAST**					
			Lac1		**Lac2**		**Lac3**		**Lac1**		**Lac2**		**Lac3**	
			PROD	**PERS**	**PROD**	**PERS**	**PROD**	**PERS**	**PROD**	**PERS**	**PROD**	**PERS**	**PROD**	**PERS**
TMR	**Lac1**	**PROD**	1.00											
		PERS	-0.41_{1}	1.00										
	Lac2	**PROD**	0.87_{2}	-0.50_{1}	1.00									
		PERS	-0.09_{5}	0.28_{2}	-0.35_{5}	1.00								
	Lac3	**PROD**	0.83_{4}	-0.44_{2}	0.94_{3}	-0.29_{7}	1.00							
		PERS	-0.10_{7}	0.18_{2}	-0.34_{7}	0.88_{5}	-0.26_{9}	1.00						
PAST	**Lac1**	**PROD**	**0.92_{3}**	-0.27_{1}^{e}	0.72_{6}^{b}	0.07_{8}^{e}	0.74_{6}^{b}	-0.04_{9}^{e}	1.00					
		PERS	-0.04_{9}^{e}	**0.26_{3}**	-0.11_{10}^{e}	0.31_{11}^{d}	0.08_{12}^{e}	0.22_{13}^{d}	-0.03_{8}	1.00				
	Lac2	**PROD**	0.86_{5}^{a}	-0.34_{2}^{e}	**0.81_{6}**	-0.02_{9}^{e}	0.86_{6}^{b}	-0.08_{10}^{e}	0.93_{2}	-0.15_{8}	1.00			
		PERS	-0.07_{2}^{e}	0.83_{1}^{c}	-0.27_{3}^{e}	**0.55_{3}**	-0.26_{3}^{e}	0.50_{4}^{d}	0.02_{2}	0.35_{4}	-0.14_{3}	1.00		
	Lac3	**PROD**	0.84_{5}^{a}	-0.52_{2}^{e}	0.85_{6}^{a}	-0.04_{9}^{e}	**0.85_{7}**	-0.13_{10}^{e}	0.88_{3}	-0.20_{9}	0.96_{2}	-0.28_{3}	1.00	
		PERS	-0.22_{6}^{e}	0.84_{3}^{c}	-0.33_{7}^{e}	0.60_{9}^{c}	-0.27_{9}^{e}	**0.60_{8}**	-0.10_{7}	0.29_{9}	-0.14_{8}	0.94_{4}	-0.29_{7}	1.00

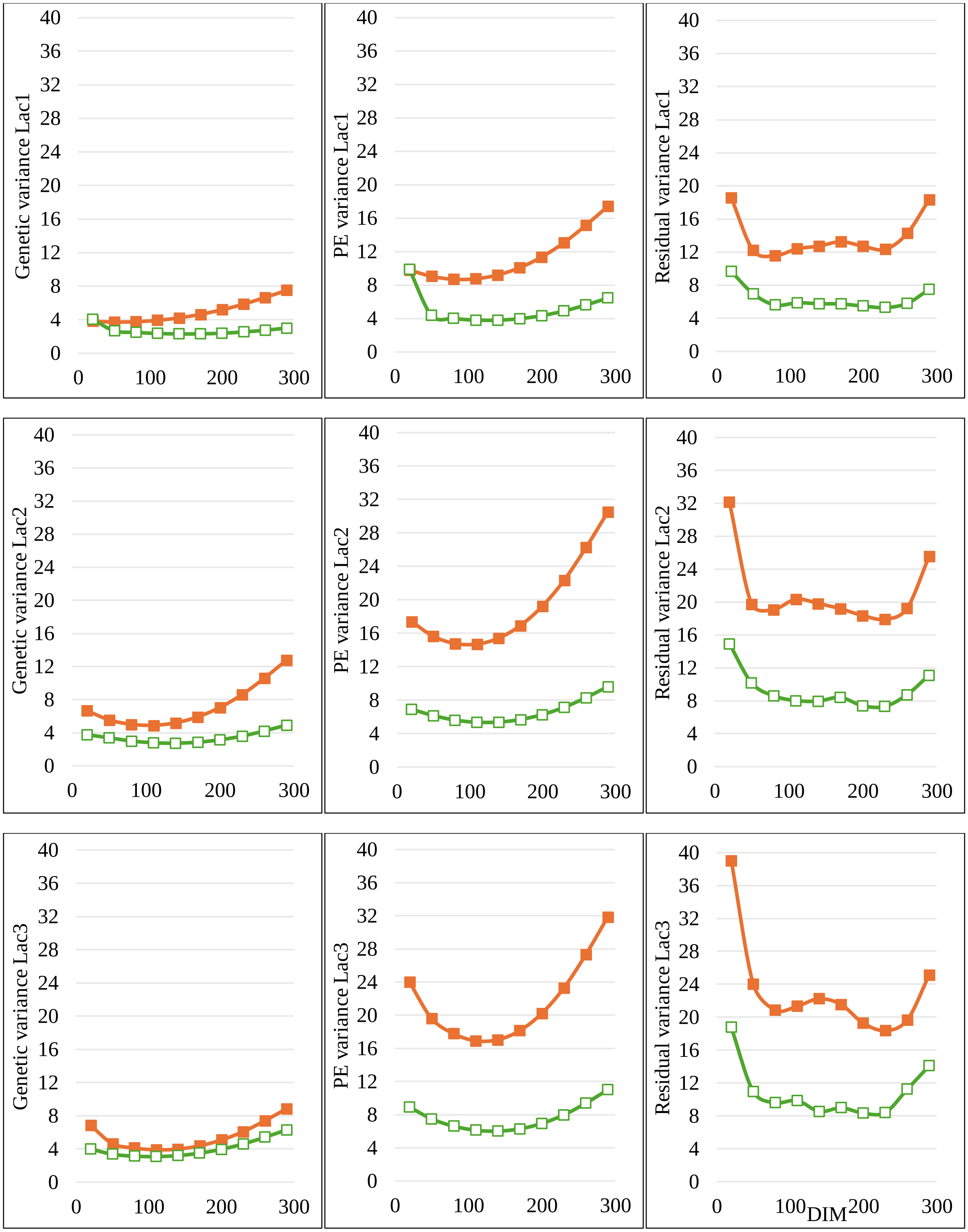
Protein = Protein production; TMR = Total mixed production system; PAST = Pasture production system; Laci = Lactation i; PROD = Genetic-average-production effect; PERS = Genetic-average-persistency effect; Standard errors: Subscript i * 10^{-2}; **Bold** = Within lactation genetic correlations, PROD and PERS effects between TMR and PAST**;** Superscripts: a = Between lactations genetic correlations of PROD effects, TMR vs PAST; b = Between lactations genetic correlations of PROD effects, PAST vs TMR; c = Between lactations genetic correlations of PERS effects, TMR vs PAST; d = Between lactations genetic correlations of PERS effects, PAST vs TMR; e = Within and between lactation genetic correlations, PROD vs PERS effects (and vice versa) between TMR and PAST.

Figures S4.5, S4.6 and S4.7 Genetic, permanent environmental and residual variances for Fat production over days in milk



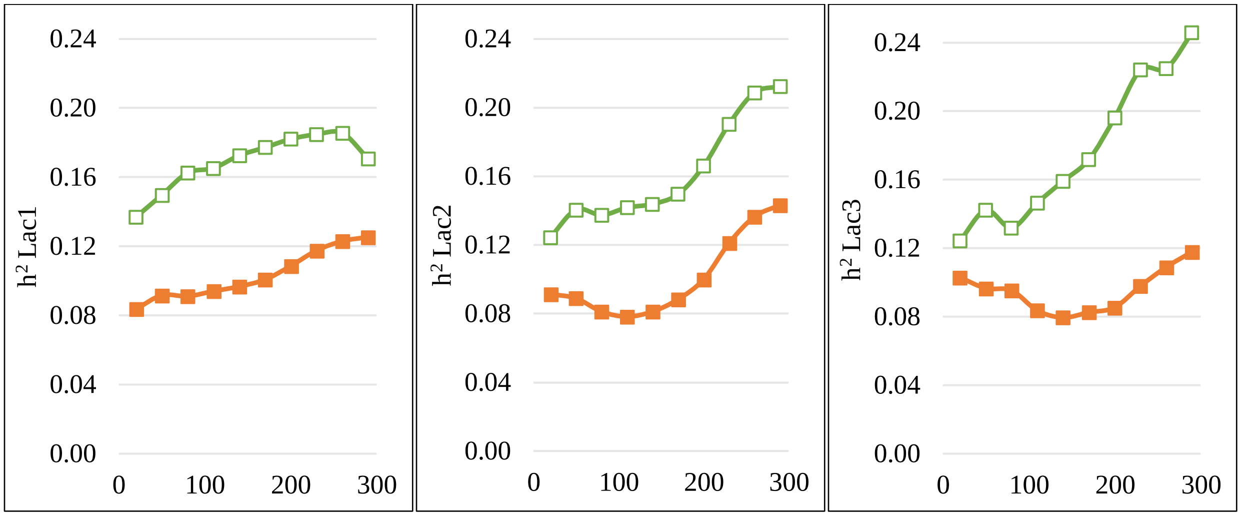
y-axis values = kg^2; x-axis values = days in milk; PE = Permanent environmental; Orange blocks (solid) = Total Mixed Ration; Green blocks (open) = Pasture; Laci = Lactation i

Figures S4.8, S4.9 and S4.10 Genetic, permanent environmental and residual variances for Protein production over days in milk



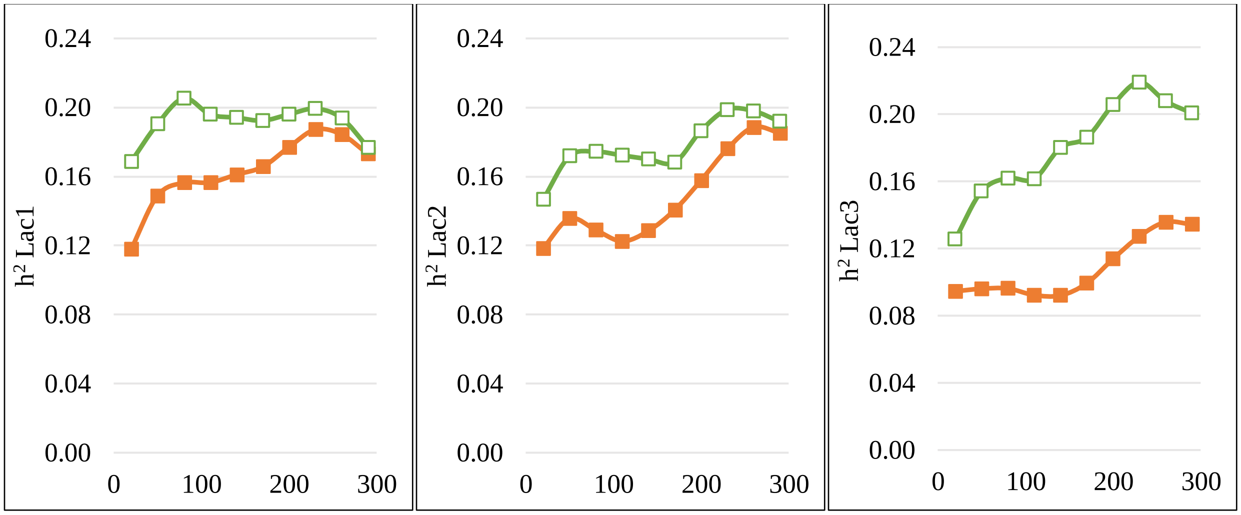
y-axis values = kg^2; x-axis values = days in milk; PE = Permanent environmental; Orange blocks (solid) = Total Mixed Ration; Green blocks (open) = Pasture; Laci = Lactation i

Figure S4.11 Heritability estimates over days in milk for Fat production



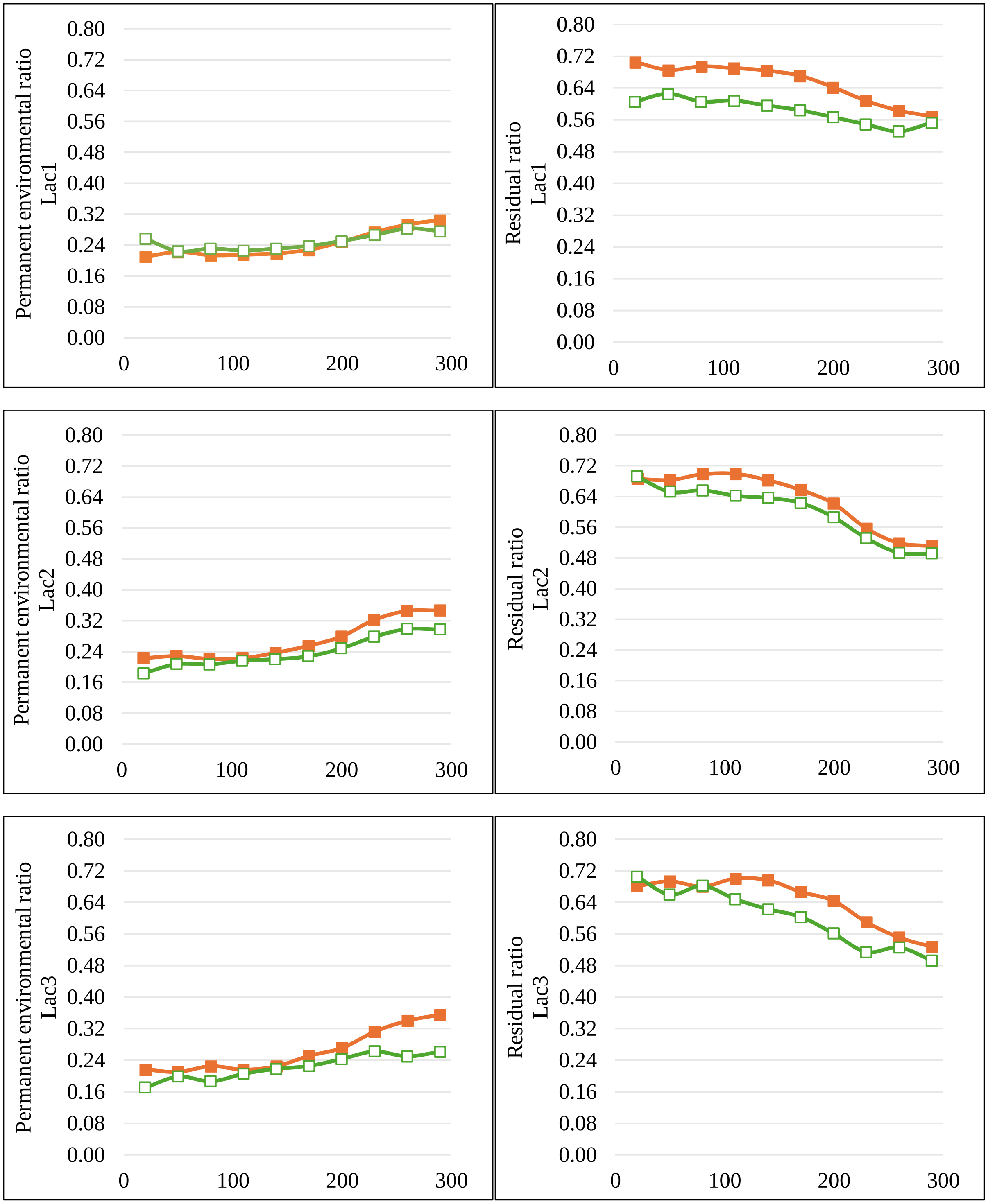
h^2 = heritability; x-axis values = days in milk; Laci = Lactation i; Orange blocks (solid) = Total Mixed Ration; Green blocks (open) = Pasture; Standard errors for heritability estimates ranged from 0.01 (TMR), and 0.01 to 0.03 (PAST), respectively.

Figure S4.12 Heritability estimates over days in milk for Protein production



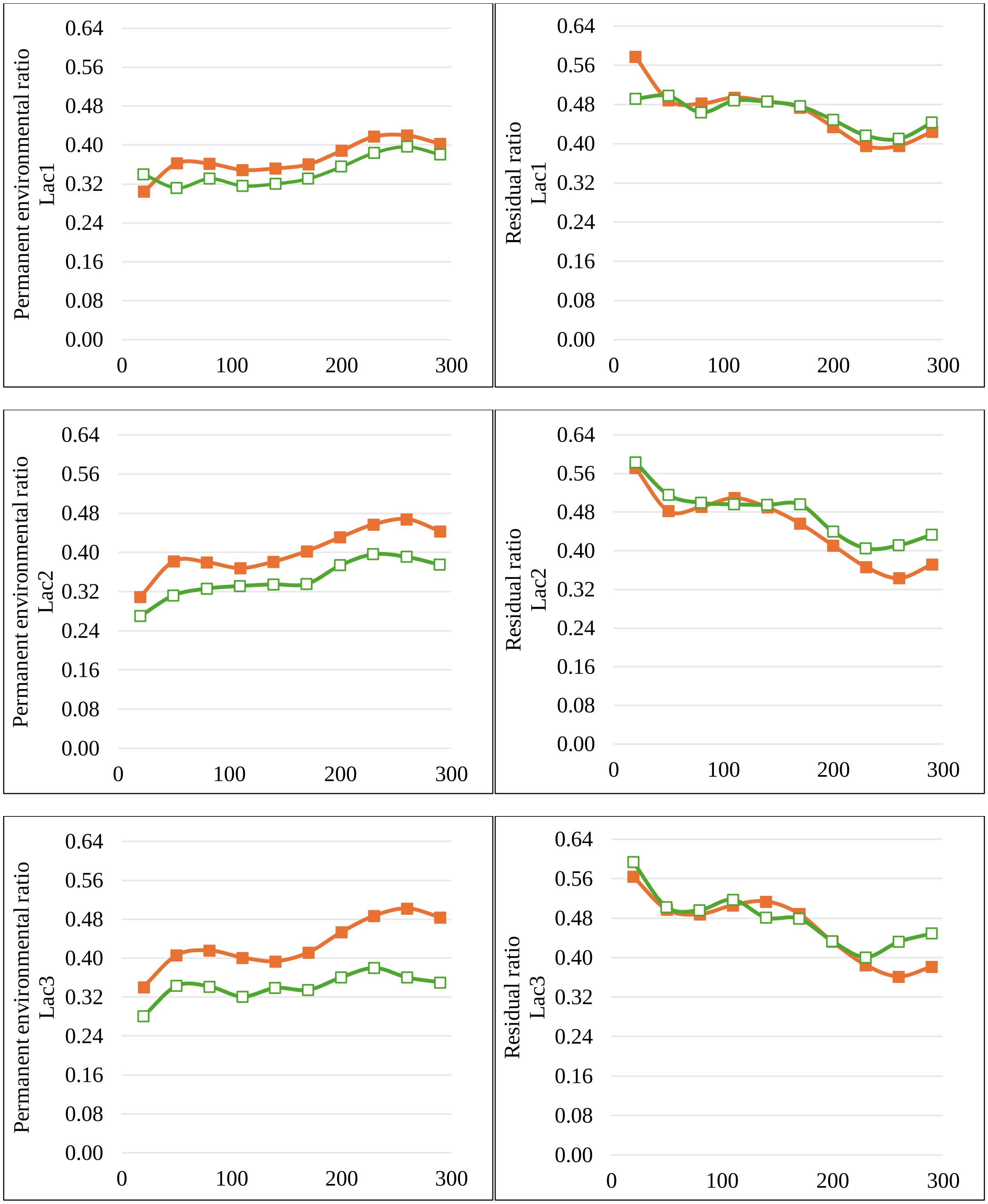
h^2 = heritability; x-axis values = days in milk; Laci = Lactation i; Orange blocks (solid) = Total Mixed Ration; Green blocks (open) = Pasture; Standard errors for heritability estimates ranged from 0.01 to 0.02.

Figures S4.13 and S4.14 Permanent environmental and residual variance ratios over days in milk for Fat production



x-axis values = days in milk; Laci = Lactation i; Orange blocks (solid) = Total Mixed Ration; Green blocks (open) = Pasture; Standard errors for permanent environmental ratio estimates ranged from 0.01 to 0.02.

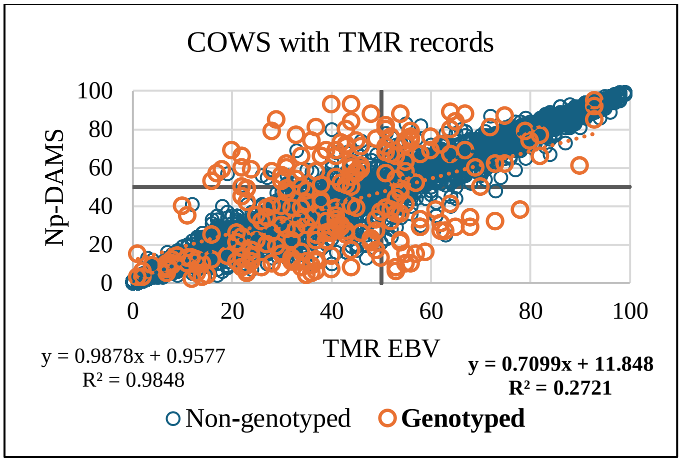
Figures S4.15 and S4.16 Permanent environmental and residual variance ratios over days in milk for Protein production



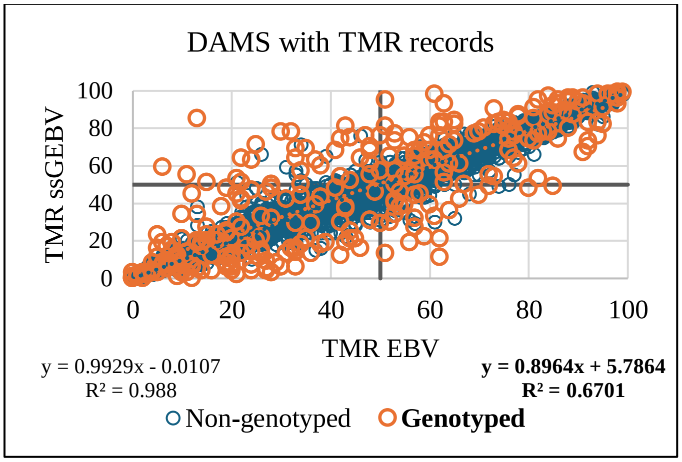
x-axis values = days in milk; Laci = Lactation i; Orange blocks (solid) = Total Mixed Ration; Green blocks (open) = Pasture; Standard errors for permanent environmental ratio estimates ranged from 0.01 to 0.02.

CHAPTER 6 Supplementary material

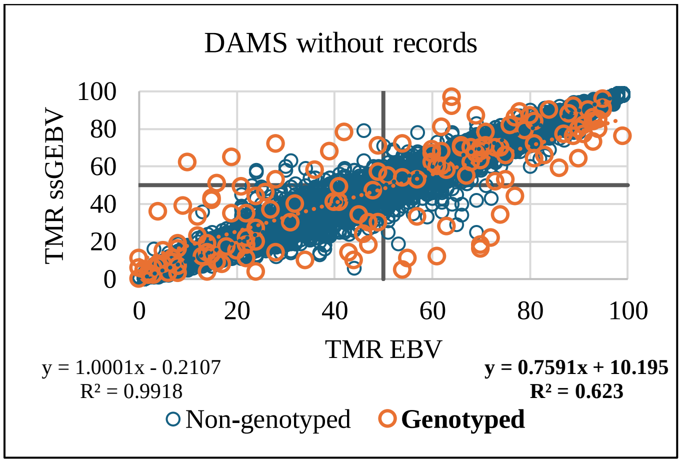
Figures S6.1a, -b, -c and -d (left) Lactation 2 Milk EBV vs ssGEBV percentile rank: COWS and DAMS



Mean difference ± SD;
genotyped vs non-genotyped = **0.19 ± 22.46** vs 0.42 ± 3.04

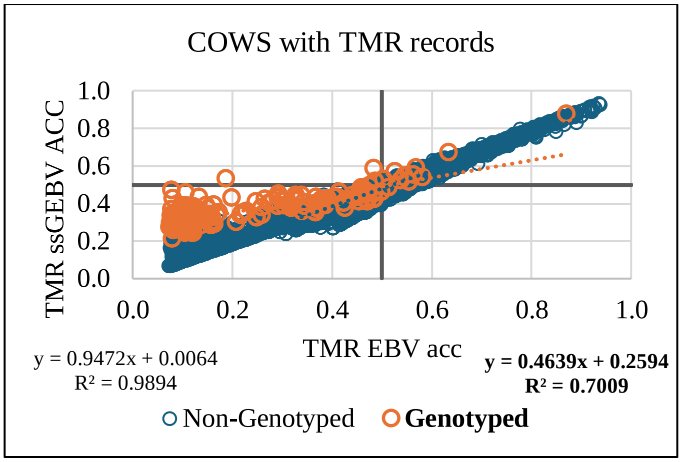


Mean difference ± SD;
genotyped vs non-genotyped = **0.17 ± 18.15** vs -0.37 ± 3.36

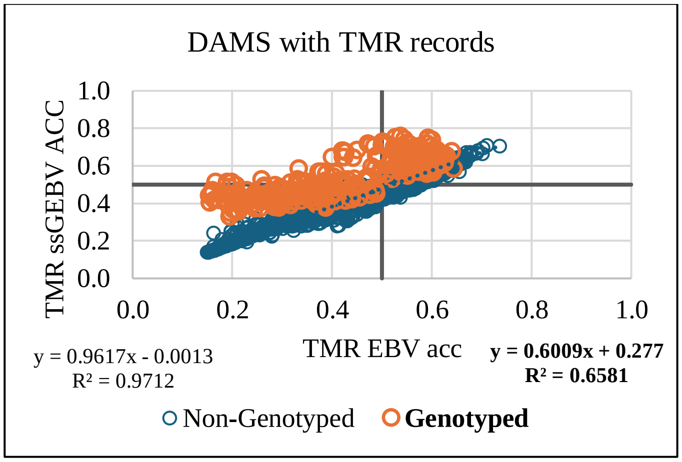


Mean difference ± SD;
genotyped vs non = **1.30 ± 19.33** vs -0.21 ± 2.7

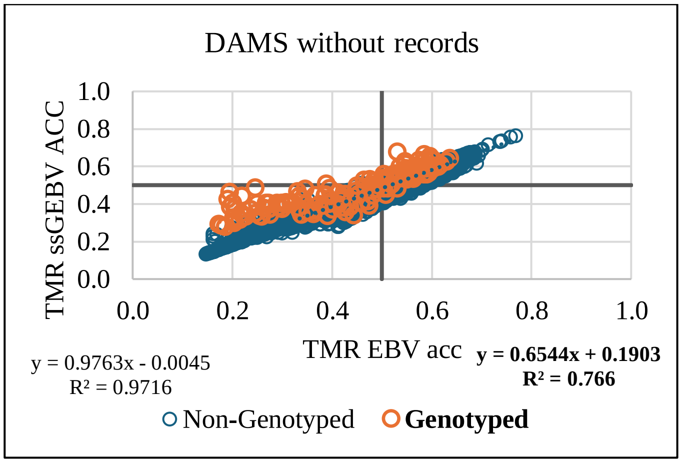
Figures S6.2a, -b, -c and -d (right) Lactation 2 Milk EBV vs ssGEBV accuracy (ACC): COWS and DAMS



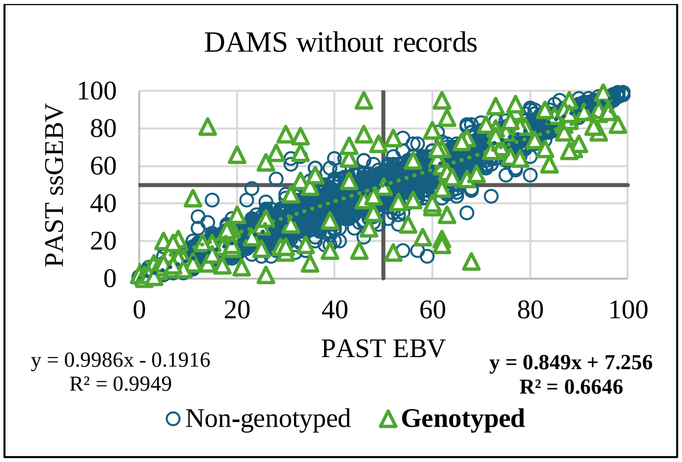
Mean difference ± SD;
genotyped vs non-genotyped = **0.15 ± 0.09** vs -0.01 ± 0.02



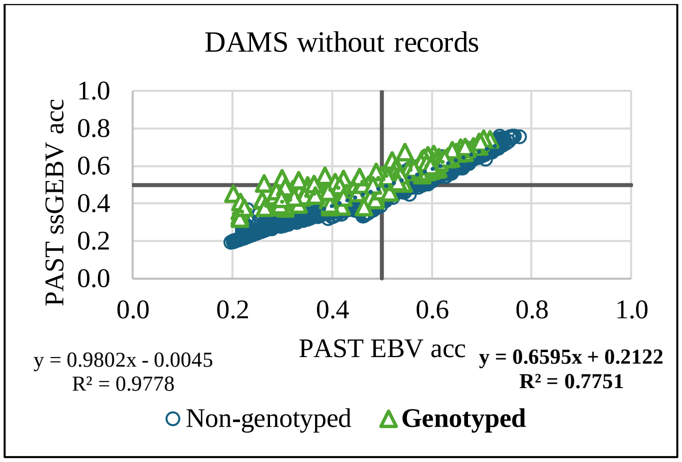
Mean difference ± SD;
genotyped vs non-genotyped = **0.05 ± 0.06** vs -0.02 ± 0.02



Mean difference ± SD;
genotyped vs non-genotyped = **0.04 ± 0.07** vs -0.02 ± 0.02



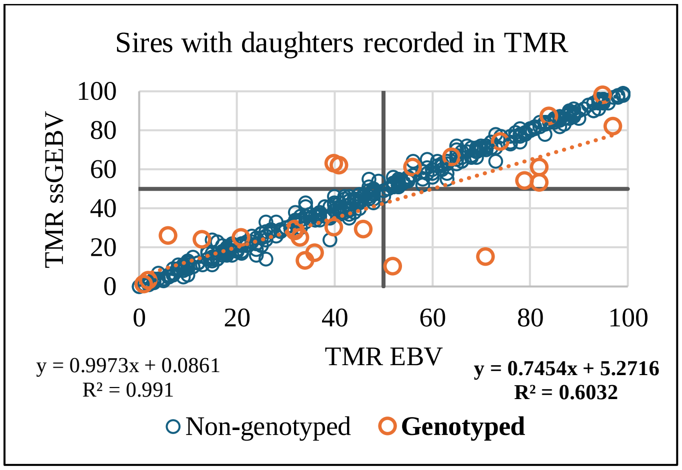
Mean difference ± SD;
genotyped vs non-genotyped = **0.16 ± 18.57** vs -0.27 ± 2.21



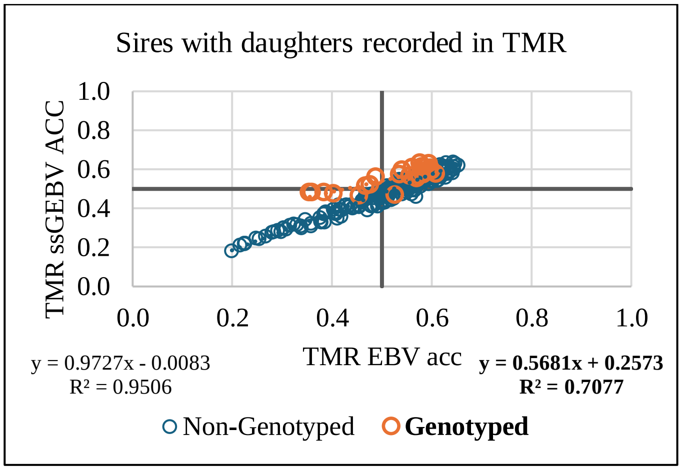
Mean difference ± SD;
genotyped vs non-genotyped = **0.05 ± 0.07** vs -0.02 ± 0.01

Figures S6.3a, -b, -c, -d and -e (left) Lactation 2 Milk EBV vs ssGEBV percentile rank: Sires

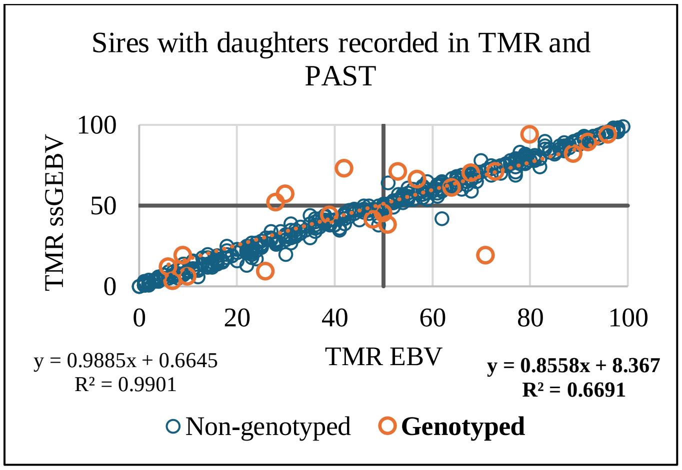
Figures S6.4a, -b,- c, -d and -e (right) Lactation 2 Milk EBV vs ssGEBV accuracy (ACC): Sires



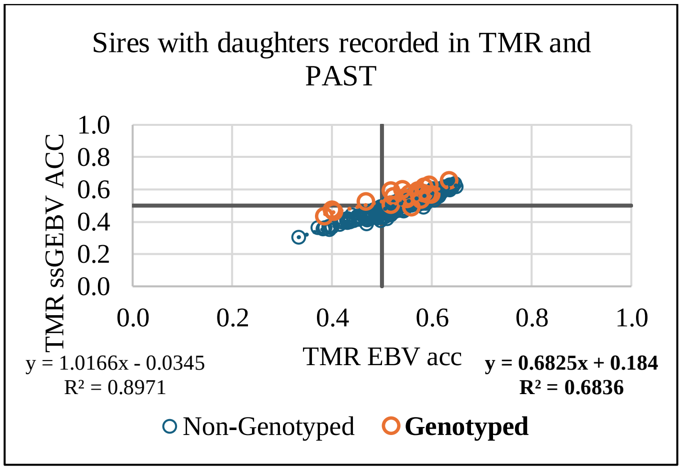
Mean difference ± SD;
genotyped vs non-genotyped = **-7.08 ± 18.94** vs -0.04 ± 2.69



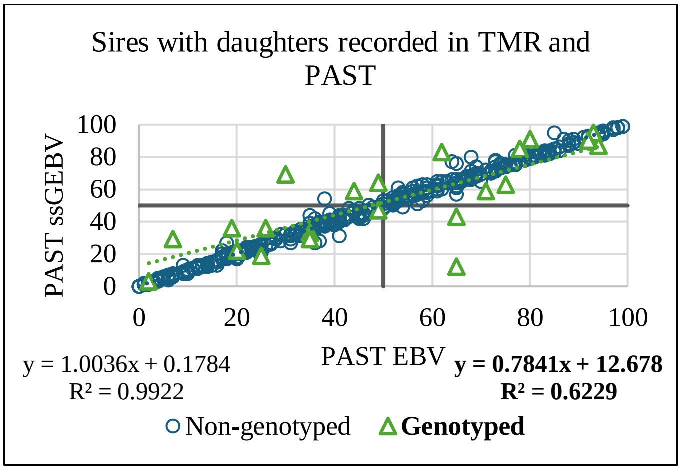
Mean difference ± SD;
genotyped vs non-genotyped = **0.03 ± 0.05** vs -0.02 ± 0.02



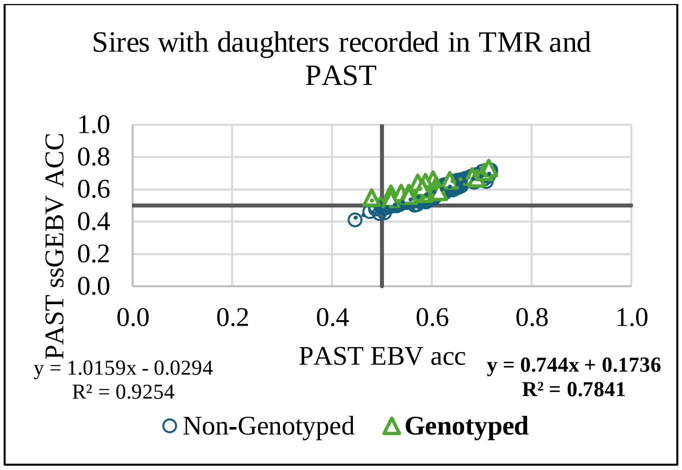
Mean difference ± SD;
genotyped vs non-genotyped = **1.23 ± 17.38** vs 0.14 ± 2.92



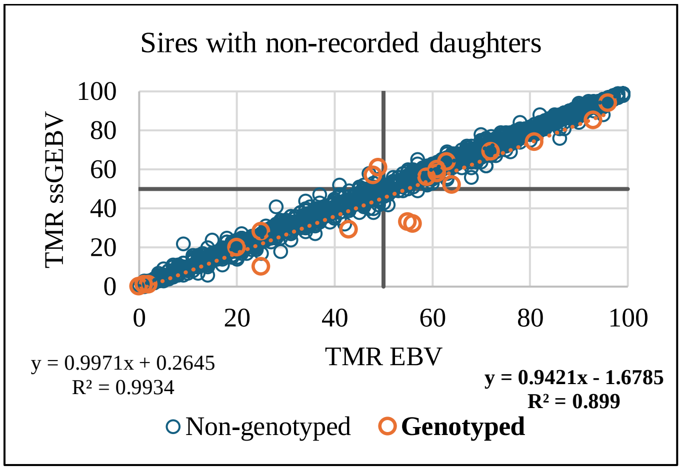
Mean difference ± SD;
genotyped vs non-genotyped = **0.01 ± 0.04** vs -0.02 ± 0.02



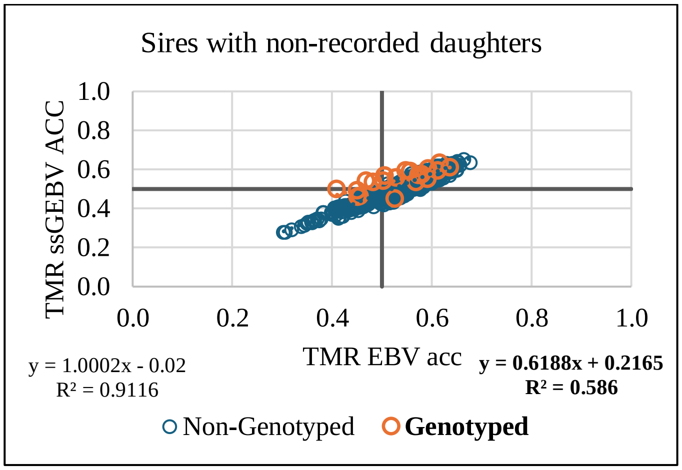
Mean difference ± SD;
genotyped vs non-genotyped = **1.72 ± 18.40** vs 0.27 ± 2.21



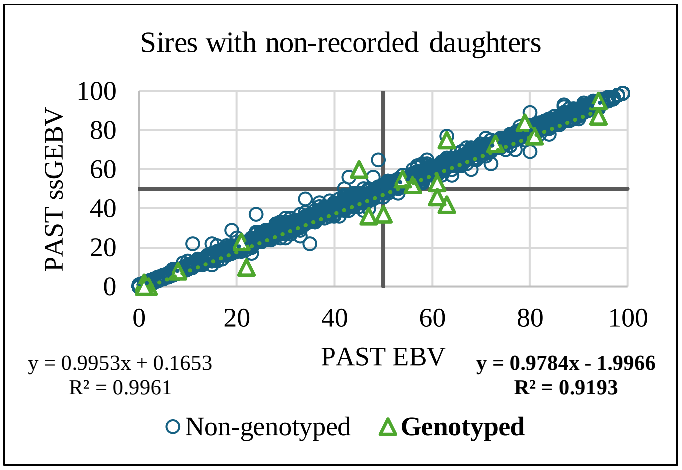
Mean difference ± SD;
genotyped vs non-genotyped = **0.02 ± 0.03** vs -0.02 ±0.01



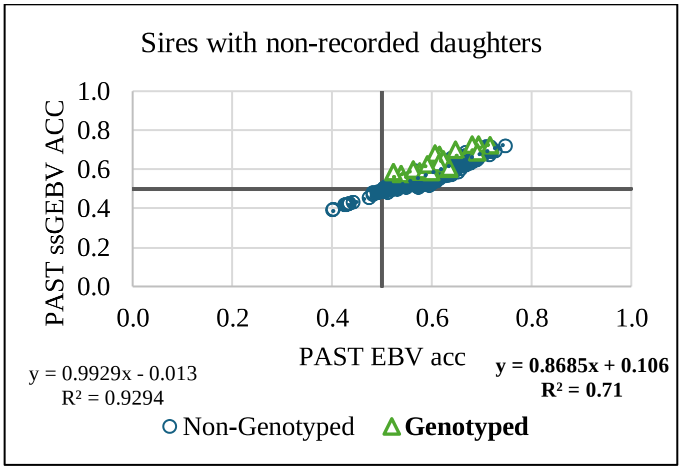
Mean difference ± SD;
genotyped vs non-genotyped = **-4.15 ± 8.96** vs 0.04 ± 1.50



Mean difference ± SD;
genotyped vs non-genotyped = **0.02 ± 0.03** vs -0.02 ±0.01



Mean difference ± SD;
genotyped vs non-genotyped = **-2.70 ± 9.96** vs -0.07 ± 2.21



Mean difference ± SD;
genotyped vs non-genotyped = **0.02 ± 0.03** vs -0.02 ± 0.01

REFERENCES

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

Akbaş, Y., Takma, Ç., Yaylak, E., 2004. Genetic parameters for quail body weights using a random regression model. S. Afr. J. Anim. Sci. 34 (2), 104-109.

Ali, T.E., Schaeffer, L.R., 1987. Accounting for covariances among test day milk yields in dairy cows. Can. J. Anim. Sci. 67, 637-644.

Aliloo, H., Miraie-Ashtiani, S.R., Moradi Shahrbabak, M., Urioste, J.I., Sadeghi, M., 2014. Accounting for heterogeneity of phenotypic variance in Iranian Holstein test-day milk yield records. Livest. Sci. 167, 25-32.

Bach, A., Valls, N., Solans, A., Torrent, T., 2008. Associations between non-dietary factors and dairy herd performance. J. Dairy Sci. 91, 3259-3267.

Banga, C.B., Neser, F.W.C., Van der Westhuizen, J., Garrick, D.J., 2009. Economic values for dairy production traits under different milk payment systems in South Africa. S. Afr. J. Anim. Sci. 39 (supplement 1), 112-115.

Banga, C.B., Neser, F.W.C., Garrick, D.J., 2014. Breeding objectives for Holstein cattle in South Africa. S. Afr. J. Anim. Sci. 44 (3), 199-214.

Banos, G., Coffey, G., Brotherstone, S., 2005a. Modeling daily energy balance of dairy cows in the first three lactations. J. Dairy Sci. 88, 2226-2237.

Banos, G., Brotherstone, S., Coffey, M.P., 2005b. Genetic profile of total body energy content of Holstein cows in the first three lactations. J. Dairy Sci. 88, 2616-2623.

Bastin, C., Gengler, N., Soyeurt, H., 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.

Bermann, M., Cesarani, A., Misztal, I., Lourenco, D., 2022. Mis. Ital. J. Anim. Sci. 21 (1), 673-685.

Berry, D.P., Buckley, F., Dillon, P., Evans, R.D., Rath, M., Veerkamp, R.F., 2003. Genetic parameters for body condition score, body weight, milk yield and fertility estimated using random regression models. J. Dairy Sci. 86, 3704-3717.

Bignardi, A.B., El Faro, L., Cardoso, V.L., Machado, P.F., de Albuquergue, L.G., 2009. Random regression models to estimate test-day milk yield genetic parameters Holstein cows in Southeastern Brazil. Livest. Sci. 123, 1-7.

Bijma, P., 2012. Accuracies of estimated breeding values from ordinary genetic evaluations do not reflect the correlation between true and estimated breeding values in selected populations. J. Anim. Breed. Genet. 129, 345-358.

Boettcher, P.J., Fatehi, J., Schutz, M.M., 2003. Genotype x environment interactions in conventional versus pasture-based dairies in Canada. J. Dairy Sci. 86, 383-389.

Bohmanova, J., Misztal, I., Bertrand, J.K., 2005. Studies on multiple trait and random regression models for genetic evaluation of beef cattle for growth. J. Anim. Sci. 83, 62-67.

Bohmanova, J., Miglior, F., Jamrozik, J., 2009. Use of test-day records beyond three hundred five days for estimation of three hundred five-day breeding values for production traits and somatic cell score of Canadian Holsteins. J. Dairy Sci. 92, 5314-5325.

Boichard, D., Fritz, S., Croiseau, P., Ducrocq, V., Cuyabano, B., Tribout, T., 2022. Long-distance associations generate erosion of genomic breeding values of candidates for selection. Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP), p1213-1216. https://doi.org/10.3920/978-90-8686-940-4.

Bonifazi, R., Calus, M.P.L., ten Napel, J., Veerkamp, R.F., Michenet, A., Savoia, S., Cromie, A., Vandenplas, J., 2022. International single-step SNPBLUP beef cattle evaluations for Limousin weaning weight. Genet. Sel. Evol. 54, 57.

Boonkum, W., Chankitisakul, V., Duangjinda, M., Buaban, S., Sumreddee, P., Sungkhapreecha, P., 2023. Genomic selection using single-step genomic BLUP on the number of services per conception trait in Thai-Holstein Crossbreeds. Animals 13, 3609.

Bourdon, R.M., 2014. Large-scale genetic evaluation. In: Understanding animal breeding. Essex: Pearson Education Limited. p. 262-78.

Buaban, S., Prempree, S., Sumreddee, P, Duangjinda, M., Masuda, Y., 2021. Genomic prediction of milk-production traits and somatic cell score using single-step genomic best linear unbiased predictor with random regression test-day model in Thai dairy cattle. J. Dairy Sci. 104, 12713-12723.

Buch, L.H., Kargo, M., Berg, P., Lassen, J., Sørensen, A.C., 2012. The value of cows in reference populations for genomic selection of new functional traits. Animal 6 (6), 880-886.

Burnham, K.P., Anderson, D.R., 2002. Model selection and multimodel inference: A practical information-theoretic approach (2nd ed.), Springer-Verlag.

Butler, S.T., Shallo, L., Murphy, J.J., 2010. Extended lactations in a seasonal -calving pastoral system of production to modulate the effects of reproductive failure. J. Dairy Sci. 93, 1283-1295.

Buttchereit, E., Stamer, E., Junge, W., Thaller, G., 2010. Evaluation of five lactation curve models fitted for fat:protein ratio of milk and daily energy balance. J. Dairy Sci. 93, 1702-1712.

Caccamo, M., Veerkamp, R.F., Ferguson, J.D., Petriglieri, R., La Terra, F., Licitra, G., 2010. Associations of breed and feeding management with milk production curves at herd level using a random regression test-day model. J. Dairy Sci. 93, 4986-4995.

Caixeta, L.S., Omontese, B.O., 2021. Monitoring and improving the metabolic health of dairy cows during the transition period. Animals 11 (352), 1-17.

Calus, M.P.L., Groen, A.F., de Jong, G., 2002. Genotype x Environment interaction for protein yield in Dutch dairy cattle as quantified by different models. J. Dairy Sci. 85, 3115-3123.

Christensen, O.F., Lund, M.S., 2010. Genomic prediction when some animals are not genotyped. Genet. Sel. Evol. 42, 2.

Christensen, O.F., 2012. Compatibility of pedigree-based and marker-based relationship matrices for single-step genetic evaluation. Genet. Sel. Evol. 44, 37.

Christensen, O.F., Madsen, P., Nielsen, B., Ostersen, T., Su, G., 2012. Single-step methods for genomic evaluation in pigs. Animal 6 (10), 1565-1571.

Cobuci, J.A., Euclydes, R.F., Costa, C.N., Lopes, P.S., Torres, R.A., Perreira, C.S., 2004. Análises da persistência na lactação de vacas da raça Holandesa, usandon produção no dia do controle e modelo de regressão aleatória. Brazilian Journal of Animal Science 33, 546-554.

Cobuci, J.A., Euclydes, R.F., Costa, C.N., de Almeida Torres, R., Lopes, P.S., Pereira, C.S., 2007. Genetic evaluation for persistency of lactation in Holstein cows using a random regression model. Genet. Mol. Biol. 30 (2), 349-355.

Cobuci, J.A., Costa, C.N., 2012. Persistency of lactation using random regression models and different fixed regression modelling approaches. R. Bras. Zootec. 41 (9), 1996-2004.

Coffey, M.P., Emmans, G.C., Brotherstone, S., 2001. Genetic evaluation of dairy bulls for energy balance traits using random regression. Anim. Sci. 73, 29-40.

Colleau, J.J., 2002. An indirect approach to the extensive calculation of relationship coefficients. Genet. Sel. Evol. 34, 409–421.

CSIR, 2023. StepSA. Climate indicators. Köppen-Geiger climate classification. http://stepsa.csir.co.za/climate_koppen_geiger.html. Accessed Aug 2023.

Cupps, P.T., 1966. Breeds of dairy cattle. Page 149 in Introduction to Livestock Production. 2nd ed. H. H. Cole, ed. W. H. Freeman and Co., San Francisco, CA.

Dekkers, J.C.M., Ten Hag, J.H., Weersink, A., 1998. Economic aspects of persistency of lactation in dairy cattle. Livest. Prod. Sci. 53, 237-252.

De Roos, A.P.W., Harbers, A.G.F., De Jong, G., 2004. Random herd curves in a test-day for milk, fat, and protein production of dairy cattle in the Netherlands. J. Dairy Sci. 87, 2693-2701.

Ding, X., Zhang, Z., Li, X., Wang, S., Wu, X., Sun, D., Yu, Y., Liu, J., Wang, Y., Zhang, Y., Zhang, S., Zhang, Y., Zhang, Q., 2013. Accuracy of genomic prediction for milk production traits in the Chinese Holstein population using a reference population consisting of cows. J. Dairy Sci. 96, 5315-5323.

Dos Santos Daltro, D., Ambrosini, D.P., Negri, R., Vinícius, M., Da Silva, G.B., Cobuci, J.A., 2022. Reaction norm model to describe environmental sensitivity in Girolando cattle. Livest. Sci. 263, 1-13.

Druet, T., Jaffrézic, F., Boichard, D., Ducrocq, V., 2003. Modelling lactation curves and estimation of genetic parameters for first lactation test-day records of French Holstein cows. J. Dairy Sci. 86, 2480-2490.

Druet, T., Jaffrézic, F., Ducrocq, V., 2005. Estimation of genetic parameters for test day records of dairy traits in the first three lactations. Genet. Sel. Evol. 37, 257-271.

Ducrocq, V., Cadet, A., Patry, C., Pienaar, L., Van Wyk, J.B., 2015. Interbull Meeting, Orlando, Florida, United States of America. https://interbull.org/static/web/Ducrocq_GxE2015.pdf.

Ducrocq, V., Cadet, A., Patry, C., Van der Westhuizen, L., Van Wyk, J.B., Neser, F.W.C., 2022. Two approaches to account for genotype-by-environment interactions for production traits and age at first calving in South African Holstein cattle. Genet. Sel. Evol. 54(43), 1-8.

Duvenhage, H., 2017. How well do you know your Holstein history? The Dairy Mail 24 (2). https://hdl.handle.net/10520/EJC-59ed8f8ff.

Dzomba, E.F., Nephawe, K.A., Maiwashe, A.N., Cloete, S.W.P., Chimonyo, M., Banga, C.B., Muller, C.J.C., Dzama., K., 2010. Random regression test-day model for the analysis of dairy cattle production data in South Africa: Creating the framework. S. Afr. J. Anim. Sci. 40(4), 273-284.

Fahey, A.G., Schutz, M.M., Lofgren, D.L., Schinckel, A.P., Stewart, T.S., 2007. Genotype by environment interaction for production traits while accounting for heteroscedasticity. J. Dairy Sci. 90, 3889-3899.

Falconer, D.S., Latyszewski, M., 1952. The environment in relation to selection for size in mice. J. Genet. 51, 67–80.

Falconer, D.S., 1960. Correlated characters. In: Introduction to quantitative genetics. London: Longman group limited. p. 322-324.

Forni, S., Aguilar, I., Misztal, I., 2011. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. Genet. Sel. Evol. 43 (1), 1-7.

Friggens, N.C., Berg, P., Theilgaard, P., Korsgaard, I.R., Ingvartsen, K.L., Løvendahl, P., Jensen, J., 2007. Breed and parity effects on energy balance profiles through lactation: Evidence of genetically driven body energy change. J. Dairy Sci. 90, 5291-5305.

Fulkerson, W.J., Davison, T.M., Garcia, S.C., Hough, G., Goddard, M.E., Dobos, R., Blockey, M., 2008. Holstein-Friesian dairy cows under a predominantly grazing system: interaction between genotype and environment. J. Dairy Sci. 91, 826-839.

Gengler, N., 1995. Use of mixed models to appreciate the persistency of yields during the lactation of milk cows. Ph.D. Diss. Faculté Univ. Sci. Agron., Gembloux, Belgium.

Gengler, N., 1996. Persistency of lactation yields: a review. Interbull Bull. 12, 87-96.

Gengler, N., Tijani, A., Wiggans, G.R., Misztal, I., 1999. Estimation of (co)variance function coefficients for test day yields with Expectation-Maximization Restricted Maximum Likelihood Algorithm. J. Dairy Sci. 82 (Aug.) Online. http://www.journalofdairyscience.org.

Gengler, N., Wiggans, G.R., 2001. Variance of effects of lactation stage within herd by herd yield. J. Dairy Sci. 84 (Suppl. 1), 216. (Abstr.)

Gertenbach, W.D., no date. Dairying in Kwazulu-Natal. Breeds of dairy cattle. Department of Agriculture and Rural Development, South Africa. https://www.kzndard.gov.za/images/Documents/RESOURCE_CENTRE/GUIDELINE_DOCUMENTS/PRODUCTION_GUIDELINES/Dairying_in_KwaZulu-Natal/Breeds%20of%20Dairy%20Cattle.pdf.

Goddard, M.E., 2009. Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 136, 245-257.

Grossman, M., Hartz, S.M., Koops, W.J., 1999. Persistency of lactation yield: A novel approach. J. Dairy Sci. 82, 2192-2197.

Guo, G., Zhao, F., Wang, Y., Zhang, Y., Du, L., Su, G., 2014. Comparison of single-trait and multiple-trait genomic prediction models. Genetics 15 (30), 1-7.

Hammami, H., Rekik, B., Soyeurt, H., Ben Gara, A., Gengler, N., 2008a. Genetic parameters for Tunisian Holsteins using a test-day random regression model. J. Dairy Sci. 9, 2118-2126.

Hammami, H., Rekik, B., Soyeurt, H., Bastin, C., Stoll, J., Gengler, N., 2008b. Genotype x environment interaction for milk yield in Holsteins using Luxembourg and Tunisian populations. J. Dairy Sci. 91, 3661-3671.

Hammami, H., Rekik, B., Gengler, N., 2009. Genotype by environment interaction in dairy cattle. Biotechnol. Agron. Soc. Environ. 13 (1), 155-164.

Hastie, T., Tibshirani, R., Friedman, J., 2017. The Elements of Statistical Learning, second ed. Springer, New York.

Hayes, B.J., Daetwyler, H.D., Goddard, M.E., 2016. Models for genome x environment interaction: Examples in livestock. Crop Sci. 56, 1-9.

Henderson, C.R. Jr., 1982. Analysis of covariance in the mixed model: higher-level, nonhomogeneous, and random regressions. Biometrics 38 (3), 623-640.

Heringstad, B., Chang, Y.M., Gianola, D., Klemetsdal, G., 2003. Genetic analysis of longitudinal trajectory of clinical mastitis in first -lactation Norwegian cattle. J. Dairy Sci. 86, 2676-2683.

Hollifield, M.K., Lourenco, D., Bermann, M., Howard, J.T., Misztal, I., 2021. Determining the stability of accuracy of genomic estimated breeding values in future generations in commercial pig populations. J. Anim. Sci. 99 (4), 1-8.

Horan, B., Dillon, P., Faverdin, P., Delaby, L., Buckley, F., Rath, M., 2005. The interaction of strain of Holstein-Friesian cows and pasture-based feed systems on milk yield, body weight, and body condition score. J. Dairy Sci. 88, 1231-1243.

Ingvartsen, K.L., Dewhurst, R.J., Friggens, N., 2003. On the relationship between lactational performance and health: is it yield or metabolic imbalance that cause production disease in dairy cattle? A position paper. Livest. Prod. Sci. 83, 277-308.

Interbull, 2020. National genetic evaluation. Rep. of South Africa. https://interbull.org/ib/geforms

Interbull, 2022. National genetic evaluation. Rep. of South Africa. https://interbull.org/ib/geforms

Interbull, 2024. National genomic evaluation forms provided by countries. https://interbull.org/ib/nationalgenoforms.

Intergis, 2000. Integrated registration and genetic information system. Animal production. Agricultural Research Council. South Africa. http://www.arc.agric.za/arc-api/Pages/ARC-API-Homepage.aspx.

Iwaisaki, H., Tsuruta, S., Misztal, I., Bertrand, J.K., 2005. Genetic parameters estimated with multitrait and linear spline-random regression models using Gelbvieh early growth data. J. Anim. Sci. 83, 757–763.

Jakobsen, J. H., Madsen, P., Jensen, J., Pedersen, J., Christensen, L.G., Sorensen, D.A., 2002. Genetic parameters for milk production and persistency for Danish Holsteins estimated in random regression models using REML. J. Dairy Sci. 85, 1607-1616.

Jamrozik, J., Schaeffer, L.R., Dekkers, J.C.M., 1997. Genetic evaluation of dairy cattle using test day yields and random regression model. J. Dairy Sci. 80, 1217-1226.

Jamrozik, J., Schaeffer, L.R., 1997. Estimates of genetic parameters for a test day model with random regression for yield traits of first lactation Holsteins. J. Dairy Sci. 80, 762-770.

Jamrozik, J., Janse, G., Schaeffer, L.R., Liu, Z., 1998. Analysis of persistency of lactation calculated from a random regression test day model. Interbull Bull. No. 17, 64-69.

Jamrozik, J., Gianola, D., Schaeffer, L.R., 2001. Bayesian estimation of genetic parameters for test day records in dairy cattle using linear hierarchical models. Livest. Prod. Sci. 71, 223-240.

Jamrozik, J., Schaeffer, L.R., Weigel, K.A., 2002. Estimates of genetic parameters for single- and multiple-country test-day models. J. Dairy Sci. 85, 3131-3141.

Jenko, J., Wiggans, G.R., Cooper, T.A., Eaglen, S.A.E., Luff, W.G.D., Bichard, M., Pong-Wong, R., Woolliams, J.A., 2016. Cow genotyping strategies for genomic selection in a small dairy cattle population. J. Dairy Sci. 100, 439-452.

Jensen, J., 2001. Genetic evaluation of dairy cattle using test-day models. J. Dairy Sci. 84, 2803-2812.

Jibrila, I., Vandenplas, J., ten Napel, J., Veerkamp, R.F., 2021. Avoiding preselection bias in subsequent single-step genomic BLUP evaluations of genomically preselected animals. J. Anim. Breed. Genet. 138, 432-441.

Kearney, J.F., Schutz, M.M., Boettcher, P.J., Weigel, K.A., 2004. Genotype x environment interaction for grazing versus confinement. I. Production traits. J. Dairy Sci. 87, 501-509.

Kgari, R.D., Muller, C.J.C., Dzama, K., Makgahlela, M.L., 2020. Evaluation of female fertility in dairy cattle enterprises – A review. S. Afr. J. Anim. Sci. 50 (6), 819-829.

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

Kirkpatrick, M., Hill, W.G., Thompson, R., 1994. Estimating the covariance structure of traits during growth and ageing, illustrated with lactation in dairy cattle. Genetical Research 64 (1), 57-69.

Kluska, S., Masuda, Y., Ferraz, J.B.S., Tsuruta, S., Eler, J.P., Baldi, F., Lourenco, D., 2021. Metafounders may reduce bias in composite cattle genomic predictions. Front. Genet. 12, 678587.

Koivula, M., Strandén, I., Aamand, G.P., Mäntysaari, E.A., 2016. Effect of cow reference group on validation reliability of genomic evaluation. Animal 10 (6), 1061-1066.

Kolver, E.S., Roche, J.R., De Veth, M.J., Thorne, P.L., Napper, A.R., 2002. Total mixed rations versus pasture diets: Evidence for a genotype x diet interaction in dairy cow performance. Proc. N.Z. Soc. Anim. Prod. 62, 246-251.

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

Laird, N.M., Ware, J.H., 1982. Random-Effects Models for Longitudinal Data. Biometrics 38, 963-974.

Leclerc, H., Duclos, D., Barbat, A., Druet, T., Ducrocq, V., 2008. Environmental effects on lactation curves included in a test-day model genetic evaluation. Animal 2(3), 344-353.

Lee, Y., Dang, C., Alam, M.Z., Kim, Y., Cho, K., Park, K., Kim, J., 2020. The effectiveness of genomic selection for milk production traits of Holstein dairy cattle. Asian-Australas J. Anim. Sci. 33 (3), 382-389.

Legarra, A., Robert-Granié, C., Manfredi, E., Elsen, J.M., 2008. Performance of genomic selection in mice. Genetics 180, 611-618.

Legarra, A., Aguilar, I., Misztal, I., 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92, 4656-4663.

Legarra, A., Christensen, O.F., Vitezica, Z.G., Aguilar, I., Misztal, I., 2015. Ancestral relationships using metafounders: finite ancestral populations and across population relationships. Genetics 200, 455-468.

Legarra, A., Reverter, A., 2018. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. Genet. Sel. Evol. 50 (53), 1-18.

Lewis, R.M., Brotherstone, S., 2002. A genetic evaluation of growth in sheep using random regression techniques. Faculty Papers and Publications in Animal Science. 828. http://digitalcommons.unl.edu/animalscifacpub/828.

Li, J., Gao, H., Madsen, P., Li, R., Liu, W., Bao, P., Xue, G., Gao, Y., Di, X., Su, G., 2020. Impact of the order of Legendre polynomials in random regression model on genetic evaluation for milk yield in dairy cattle population. Front. Genet. 11 (586155), 1-8.

Liu, Z., Jamrozik, J., Jansen G., 1998. A comparison of fixed and random regression models applied to dairy test day production data. Interbull Bull. No. 17, 60-63.

Logix, 2019. Livestock Operation and Genetic Information Exchange. South African Studbook Association. http://www.sastudbook.co.za/p112/services/logix-animal-recording-services.html

Loker, S., Miglior, F., Bohmanova, J., Jamrozik, J., Schaeffer, L.R., 2009. Phenotypic analysis of pregnancy effect on milk, fat, and protein yields of Canadian Ayrshire, Jersey, Brown Swiss, and Guernsey breeds. J. Dairy Sci. 92, 1300-1312.

Loker, S., Bastin, C., Miglior, F., Sewalem, A., Schaeffer, L.R., Jamrozik, J., Osborne, V., 2011. Short communication: Estimates of genetic parameters of body condition score in the first 3 lactations using a random regression animal model. J. Dairy Sci. 94, 3693-3699.

Lourenco, D.A., Fragomeni, B.O., Tsuruta, S., Aquilar, I., Zumbach, B., Hawken, R.J., Legarra, A., Misztal, I., 2015. Accuracy of estimated breeding values with genomic information on males, females, or both: an example on broiler chicken. Genet. Sel. Evol. 47 (56), 1-9.

Lourenco, D., Legarra, A., Tsuruta, S., Masuda, Y., Aguilar, I., Misztal, I., 2020. (Review) Single-step genomic evaluations from theory to practice: using SNP Chips and sequence data in BLUPF90. Genes 11, 790, 1-32.

Lund, M.S., Jensen, J., Petersen, P.H., 1999. Estimation of genetic and phenotypic parameters for clinical mastitis, somatic cell production deviance and protein yield in dairy cattle using Gibbs sampling. J. Dairy Sci. 82, 1045-1051.

Lund, M.S., de Roos, A.P.W., de Vries, A.G., Druet, T., Ducrocq, V., Fritz, S., Guillaume, F., Guldbrandtsen, B., Liu, Z., Reents, R., Schrooten, C., Seefried, F., Su, G., 2011. A common reference population from four European Holstein populations increases reliability of genomic predictions. Genet. Sel. Evol. 43, 43.

Luštrek, B., Vandenplas, J., Gorjanc, G., Potočnik, K., 2020. Genomic evaluation of Brown Swiss dairy cattle with limited national genotype data and integrated external information. J. Dairy Sci. 104, 5738-5754.

Lynch, M., Walsh, B., 1998. Genetics and analysis of quantitative traits. Sinauer Associates, Inc., 23 Plumtree Road, Sunderland, Massachusetts 01375, United States of America.

Macciotta, N.P.P., Vicario, D., Cappio-Borlino, A., 2005. Detection of different shapes of lactation curve for milk yield in dairy cattle by empirical mathematical models. J. Dairy Sci. 88, 1178-1191.

Macciotta, N.P.P., Dimauro, C., Rassu, S.P.G., Steri, R., Pulina, G., 2011. The mathematical description of lactation curves in dairy cattle. Ital. J. Anim. Sci. 10 (51), 213-223.

MacDonald, K.A., McNaughton, L.R., Verkerk, G.A., Penno, J.W., Burton, L.J., Berry, D.P., Gore, P.J.S., Lancaster, J.A.S., Holmes, C.W., 2007. A comparison of three strains of Holstein-Friesian cows grazed on pasture: growth, development, and puberty. J. Dairy Sci. 90, 3993-4003.

Mafolo, K.S., Neser, F.W.C., Makgahlela, M.L., 2021. Virtual ICAR-Interbull (meeting). 26 – 30 April 2021. https://interbull.org/static/web/2.5_Kgaogelo_Mafolo.pdf.

Mark, T., 2004. Applied genetic evaluations for production and functional traits in dairy cattle. J. Dairy Sci. 87, 2641-2652.

Martini, J.W.R., Schrauf, M.F., Garcia-Baccino, C.A., Pimentel, E.C.G., Munilla, S., Rogberg-Muñoz A., Cantet, R.J.C., Reimer, C., Gao, N., Wimmer, V., Simianer, H., 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.

Masuda, Y., Misztal, I., VanRaden, P.M., Lawlor, T.J., 2018. Differing genetic trend estimates from traditional and genomic evaluations of genotyped animals as evidence of preselection bias in US Holsteins. ADSA Annual Meeting, Knoxville, Tennessee, p. 5194-5206.

Masuda, Y., VanRaden, P.M., Tsuruta, S., Lourenco, D.A.L., Misztal, I., 2021. Invited review: Unknown-parent groups and metafounders in single-step genomic BLUP. J. Dairy Sci. 105, 923-939.

Matilainen K., Strandén I., Aamand G.P., Mäntysaari E.A., 2018. Single step genomic evaluation for female fertility in Nordic Red dairy cattle. J. Anim. Breed. Genet. 135, 5, 337–348.

Mayeres, P., Stoll, J., Bormann, J., Reents, R., Gengler, N., 2004. Prediction of daily milk, fat, and protein production by a random regression test-day model. J. Dairy Sci. 87, 1925-1933.

McMillan, A.J., Swan, A.A., 2017. Weighting of genomic and pedigree relationships in single step evaluation of carcass traits in Australian sheep. Proc. Assoc. Advmt. Anim. Breed. Genet. 22, 557-560.

Meseret, S., Negussie, E., 2017. Genetic parameters for test-day milk yield in tropical Holstein Friesian cattle fitting a multiple-lactation random regression animal model. S. Afr. J. Anim. Sci. 47(3), 352-361.

Meyer, K., 1998. Estimating covariance functions for longitudinal data using a random regression model. Genet. Sel. Evol. 30, 221-240.

Meyer, K., 2005. Random regression analyses using B-splines to model growth of Australian Angus cattle. Genet. Sel. Evol. 37, 473-500.

Meyer, K., 2007. WOMBAT – A tool for mixed model analyses in quantitative genetics by REML. J. Zhejiang Uni. SCIENCE B 8:815-821. [doi:10.1631/jzus.2007.B0815]

Meyer, K., Tier, B., Swan, A., 2018. Estimates of genetic trend for single-step genomic evaluations. Genet. Sel. Evol. 50, 39.

Meyer, K., 2022. Download WOMBAT examples. http://didgeridoo.une.edu.au/km/wmbdownload2.php. Example 20. RNote_WOMBATHinverse.pdf. Calculations involving genomic relationship matrices and friends available in WOMBAT.

Meuwissen, T.H.E., Hayes, B.J., Goddard, M.E., 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157, 1819-1829.

Miglior, F., Gong, W., Wang, Y., Kistemaker, G.J., Sewalem, A., Jamrozik, J., 2009. Short communication: Genetic parameters of production traits in Chinese Holsteins using a random regression test-day model. J. Dairy Sci. 92, 4697-4706.

Misztal, I., Strabel, T., Jamrozik, J., Mäntysaari, E.A., Meuwissen, T.H.E., 2000. Strategies for estimating the parameters needed for different test-day models. J. Dairy Sci. 83, 1125-1134.

Misztal, I., 2006. Properties of random regression models using linear splines. J. Anim. Breed. Genet. 123, 74-80.

Misztal, I., Legarra, A., Aguilar, I., 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J. Dairy Sci. 92, 4648-4655.

Misztal, I., Vitezica, Z.G., Legarra, A., Aquilar, I., Swan, A.A., 2013. Unknown-parent groups in single-step genomic evaluation. J. Anim. Breed. Genet. 130 (4), 252-258.

Misztal, I., Lourenco, D., Legarra, A., 2020. Board invited review. Current status of genomic evaluation. J. Anim. Sci. 98 (4), 1-14.

Mohammadi, A., Alijani, S., Daghighkia, H., 2014. Comparison of different polynomial functions in random regression model for milk production traits of Iranian Holstein dairy cattle. Ann. Anim. Sci. 14(1), 55-68.

Mostert, B.E., Groeneveld, E., Kanfer, F.H.J., 2004. Testday models for production traits in dairy cattle. J. Dairy Sci. 34(Suppl. 2), 35-37.

Mostert, B.E., Theron, H.E., Kanfer, F.H.J., Van Marle-Köster, E., 2006a. Test-day models for South African dairy cattle for participation in international evaluations. S. Afr. J. Anim. Sci. 36(1), 58-70.

Mostert, B.E., Theron, H.E., Kanfer, F.H.J., Van Marle-Köster, E., 2006b. Adjustment of heterogeneous variances and a calving year effect in test-day models for national genetic evaluations of dairy cattle in South Africa. S. Afr. J. Anim. Sci. 36(3), 165-174.

Mostert, B.E., Van der Westhuizen, R.R., Theron, H.E., 2008. Procedures for estimation of genetic persistency indices for milk production for the South African dairy industry. S. Afr. J. Anim. Sci. 38(3), 224-230.

Mrode, R.A., 2014. Linear models for the prediction of animal breeding values. 3rd ed. CABI, Nosworthy Way, Wallingford, Oxfordshire, UK.

Muir, B.L., Kistemaker, G., Jamrozik, J., Canavesi, F., 2007. Genetic parameters for a multiple-trait multiple-lactation random regression test-day model in Italian Holsteins. J. Dairy Sci. 90, 1564-1574.

Muller, C.J.C., Potgieter, J.P., Cloete, S.W.P., Dzama, K., 2014. Non-genetic factors affecting fertility traits in South African Holstein cows. S. Afr. J. Anim. Sci. 44(1), 54-63.

Neser, F.W.C., Van Wyk, J.B., Ducrocq, V., 2014. A preliminary investigation into genotype x environment interaction in South African Holstein cattle for reproduction and production traits. S. Afr. J. Anim. Sci. 44 (Issue 5, supplement 1), 75-79.

Oliviera, H.R., Campos, G.S., Lazaro, S.F., Jamrozik, J., Schinckel, A., Brito, L.F., 2024. Phenotypic and genomic modeling of lactation curves: A longitudinal perspective. Joint Breeding and Genetics and Lactation Biology Symposium. J. Dairy Sci. Communications 5, 241-246.

Olori, V.E., Hill, W.G., McGuirk, B.J., Brotherstone, S., 1999a. Estimating variance components for test day milk records by restricted maximum likelihood with a random regression animal model. Livest. Prod. Sci. 61, 53-63.

Olori, V.E., Hill, W.G., Brotherstone, S., 1999b. The structure of the residual error variance of test day milk yield in random regression model. In: Proc. Computational Cattle Breeding Workshop 1999. Tuusula, Finland.

Önder, H., Sitskowska, B., Kurnaz, B., Piwczyński, D., Kolenda, M., Şen, U., Tırınk, C., Boğa D.C., 2023. Multi-trait single-step genomic prediction for milk yield and milk components for Polish Holstein population. Animals 13, 3070.

Opsomer, G., Mijten, P., Coryn, M., De Kruif, A., 1996. Post-partum anoestrus in dairy cows: A review. Vet. Q. 18(2), 68-75.

Penasa, M., De Marchi, M., Cassandro, M., 2016. Short communication: Effects of pregnancy on milk yield, composition traits, and coagulation properties of Holstein cows. J. Dairy Sci. 99, 4868-4869.

Pool, M.H., Janss, L.L.G., Meuwissen, T.H.E., 2000. Genetic parameters of Legendre Polynomials for first parity lactation curves. J. Dairy Sci. 83, 2640-2649.

Ptak, E., Schaeffer, L.R., 1993. Use of test day yields for genetic evaluation of dairy sires and cows. Livest. Prod. Sci. 34, 23-34.

Pryce, J.E., Daetwyler, H.D., 2012. Designing dairy cattle breeding schemes under genomic selection: a review of international research. Anim. Prod. Sci. 52, 107–114.

Quaas, R.L., 1988. Additive genetic model with groups and relationships. J. Dairy Sci. 71, 1338-1345.

Ramírez-Valverde, R., Peralta-Aban, J.A., Núñez-Domínguez, R., Ruiz-Flores, A., García-Muñiz, J.G., García-Peniche, T.B., 2010. Genotype by feeding system interaction in the genetic evaluation of Jersey cattle for milk yield. Animal 4 (12), 1971-1975.

Robbins, K.R., Misztal, I., Bertrand, J.K., 2005. A practical longitudinal model for evaluating growth in Gelbvieh cattle. J. Anim. Sci. 83, 29–33.

Robert-Granié, C., Bonaïti, B., Boichard, D., Barbat, A., 1999. Accounting for variance heterogeneity in French dairy cattle genetic evaluation. Livest. Prod. Sci. 60, 343-357.

Robertson, A., 1959. The sample variance of the genetic correlation coefficient. Biometrics 15, 469-475.

Rotz, C.A., Zartman, D.I., Crandall, K.I., 2005. Economic and environmental feasibility of a perennial cow dairy farm. J. Dairy Sci. 88, 3009-3019.

Saatchi, M., McClure, M.C., McKay, S.D., Rolf, M.M., Kim, J., Decker, J.E., Taxis, T.M., Chapple, R.H., Ramey, H.R., Northcutt, S.L., 2011. Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. Genet. Sel. Evol. 43 (40), 1-16.

Saltz, J.B., Bell, A.M., Flint, J., Gomulkiewicz, R., Hughes, K.A., Keagy, J., 2018. Why does the magnitude of genotype-by-environment interaction vary? Ecol. Evol. 8, 6342-6353.

Sanders, H.G., 1930. The analysis of the lactation curve into maximum yield and persistency. J. Agric. Sci. 20, 145-185

Santos, J., Malhado, C.H., Ambrosini, D., Filho, R., Araujo, A., Carneiro, P.L., 2019. Normas de reação para peso aos 365 e 550 dias de idade em bovinos Guzerá no Nordeste do Brasil. Arch. Zootec. 68, 562-570.

Sargent, F.D., Lyton, V.H., Wall, O.G., 1968. Test interval method of calculating Dairy Herd Improvement Association records. J. Dairy Sci. 51, 170.

Savegnago, R.P., Rosa, G.J.M., Valente, B.D., Herrera, L.G.G., Carneiro, R.L.R., Sesana, R.C., El Faro, C., Munari, D.P., 2013. Estimates of genetic parameters and eigenvector indices for milk production of Holstein cows. J. Dairy Sci. 96, 7284-7293.

Sawalha, R.M., Keown, J.F., Kachman, S.D., Van Vleck, L.D., 2005. Genetic evaluation of dairy cattle with test-day models with autoregressive covariance structures and with a 305-d model. J. Dairy Sci. 88, 3346-3353.

Schaeffer, L.R., 1984. Sire and cow evaluation under multiple trait models. J. Dairy Sci. 67, 1567-1580.

Schaeffer, L.R., Dekkers, J.C., 1994. Random regressions in animal models for test-day production in dairy cattle. Proc. 5th World Congr. Genet. Appl. Livest. Prod., Guelph, ON, Canada, p 443–446.

Schaeffer, L.R., Jamrozik, J., Kistemaker, G.J., Van Doormaal, B.J., 2000. Experience with a test-day model. J. Dairy Sci. 83, 1135-1144.

Schaeffer, L.R., 2006. Strategy for applying genome-wide selection in dairy cattle. J. Anim. Breed. Genet. 123, 218‐223.

Schaeffer, L.R., Jamrozik, J., 2008. Editorial. Random regression models: a longitudinal perspective. J. Anim. Breed. Genet. 125, 145-146.

Sehested, J., Gaillard, C., Lehmann, J.O., Maciel, G.M., Vestergaard, M., Weisbjerg, M.R., Mogensen, L., Larsen, L.B., Poulsen, N.A., Kristensen, T., 2019. Review: extended lactation in dairy cattle. Animal 13 (S1), 65-74.

Senarath, N., Sungkhapreecha, P., Buaban, S., Chankitisakul, V., Boonkum, W., 2021. Integrating genomic selection for rapid improvement of milk yield in small-scale dairy farms. Applied Animal Science 38, 246-251.

Silva Neto, J.B., Mota, L.F.M., Londoño-Gil, M., Schmidt, P.I., Rodrigues, G.R.D., Ligori, V.A., Arikawa, L.M., Magnabosco, C.U., Brito, L.F., Baldi, F., 2024. Genotype-by-environment interactions in beef and dairy cattle populations: A review of methodologies and perspectives on research and applications. Anim. Genet. Epub ahead of print. PMID: 39377556. doi:10.1111/age.13483.

Silvestre, A.M., Petim-Batista, F., Colaço, J., 2006. The accuracy of seven mathematical functions in modeling dairy cattle lactation curves based on test-day records from varying sample schemes. J. Dairy Sci. 89, 1813-1821.

Slagboom, M., Sørensen, A.C., Thomasen, J.R., Liu, H., Kargo, M., Hjortø, L., 2021. Ignoring genotype by environment interaction in the genetic evaluation of dairy cattle reduces accuracy but may increase selection intensity. J. Dairy Sci. 104, 12756-12764.

Sölkner, J., Fuchs, W., 1987. A comparison of different measures of persistency with special respect to variation of test-day milk yields. Livest. Prod. Sci. 16, 335-348.

Sonesson, A.K., Meuwissen, T.H.E., 2009. Testing strategies for genomic selection in aquaculture breeding programs. Genet. Sel. Evol. 41, 37.

Strabel, T., Jamrozik, J., 2006. Genetic analysis of milk production traits of Polish Black and White cattle using large-scale random regression test-day models. J. Dairy Sci. 89, 3152-3163.

Sungkhapreecha, P., Misztal, I., Hidalgo, J., Lourenco, D., Buaban, S., Chankitisakul, V., Boonkum, W., 2021. Validation of single-step genomic predictions using the linear regression method for milk yield and heat tolerance in a Thai-Holstein population. Veterinary World 14, 3119-3125.

Swart, J.C., O'Hagan, L.O., no date. The South African Friesland, its background and history. South African Holstein Cattle Breeders' Society. 141 President Reitz Avenue, Westdene, Bloemfontein, 9301.

Swalve, H.H., 2000. Symposium: Test-day models. Theoretical basis and computational methods for different test-day genetic evaluation methods. J. Dairy Sci. 83, 1115-1124.

The Dairy Farming Handbook. 2017. Compiled by Dr C.J.C. Muller. Directorate: Animal Sciences. Research and Technology Development Services (RTDS). Western Cape Department of Agriculture. https://www.elsenburg.com/resource-library/e-book-publications. ISBN: 978-0-9947024-4-9. South Africa.

Theron, H.E., Mostert, B.E., 2009. Production and breeding performance of South African dairy herds. S. Afr. J. Anim. Sci. 39 (Supplement 1), 206-210.

Thomasen, J.R., Liu, H., Sørensen, A.C., 2020. Genotyping more cows increases genetic gain and reduces rate of true inbreeding in a dairy cattle breeding scheme using female reproductive technologies. J. Dairy Sci. 103, 597-606.

Thompson, R., Meyer, K., 1986. A review of theoretical aspects in the estimation of breeding values for multi-trait selection. Livest. Prod. Sci. 15, 299-313.

Togashi, K., Lin, C.Y., 2003. Efficiency of different selection criteria for persistency and lactation milk yield. J. Dairy Sci. 87, 1528-1535.

Togashi, K., Lin, C.Y., Atagi, Y., Hagiya, K., Sato, J., Nakanishi, T., 2008. Genetic characteristics of Japanese Holstein cows based on multiple-lactation random regression test-day animal models. Livest. Sci. 114, 194-201.

Tsuruta, S., Misztal, I., Aguilar, I., Lawlor, T.J., 2011. Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holstein. J. Dairy Sci. 94, 4198-4204.

Tsuruta, S., Misztal, I., Lourenco, D. A. L., and Lawlor, T. J., 2014. Assigning unknown parent groups to reduce bias in genomic evaluations of final score in US Holsteins. J. Dairy Sci. 97, 5814–5821.

Uemoto, Y., Osawa, T., Saburi, J., 2016. Effect of genotyped cows in the reference population on the genomic evaluation of Holstein. Animal 11 (3), 382-393.

Van Der Werf, J.H.J., Goddard, M.E., Meyer, K., 1998. The use of covariance functions and random regressions for genetic evaluation of milk production based on test day records. J. Dairy Sci. 81, 3300-3308.

Van Niekerk, D.J., Neser, F.W.C., Van der Westhuizen, J., 2006. An investigation into the possibility of a genotype by environment interaction for milk production in the South African Jersey population. S. Afr. J. Anim. Sci. 36 (Issue 5, supplement 1), 107-110.

Van Niekerk, M., Neser, F.W.C., Ducrocq, V., 2019. 51st South African Society for Animal Science (SASAS) congress, University of the Free State, Bloemfontein, South Africa. http://dx.doi.org/10.13140/RG.2.2.17663.15520.

Van Niekerk, M., Neser, F.W.C., Van Wyk, J. B., Ducrocq, V., 2021. Genotype by environment interaction for persistency of production in multiparous SA Holsteins. 72nd Annual Meeting of the European Federation of Animal Science, Davos, Switzerland. http://dx.doi.org/10.13140/RG.2.2.17981.38883.

Van Niekerk, M., Neser, F.W.C., Van Wyk, J.B., Ducrocq, V., 2023. Comparison of fixed and random regression models for the analysis of milk production traits in South African Holstein dairy cattle under two production systems. Livest. Sci. 267, 1-11.

Van Niekerk, M., Neser, F.W.C., Van Wyk, J.B., Ducrocq, V., submitted. Genotype by environment interaction under two production systems utilizing a random regression model for milk production traits in multiparous South African Holstein dairy cattle. Genet. Sel. Evol.

VanRaden, P.M., 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91, 4414-4423.

VanRaden, P.M., Van Tassell, C.P., Wiggans, G.R., Sonstegard, T.S., Schnabel, R.D., Taylor, J.F., Schenkel, F.S., 2009. Invited review: reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92, 16-24.

VanRaden, P.M., O’Connell J.R., Wiggans, G.R., Weigel, K.A., 2011. Genomic evaluations with many more genotypes. Gen. Sel. Evol. 43, 10.

VanRaden, P.M., Wright, J.R., 2013. Measuring genomic pre-selection in theory and in practice. Interbull Bull. No. 47, 147-150.

Veerkamp, R.F., Koenen, E.P.C., De Jong, G., 2001. Genetic correlations among body condition score, yield, and fertility in first-parity cows estimated by random regression models. J. Dairy Sci. 84, 2327-2335.

Vignal, A., Milan, D., SanCristobal, M., Eggen, A., 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Genet. Sel. Evol. 34, 275-305.

Vitezica, Z.G., Aguilar, I., Misztal, I., Legarra, A., 2011. Bias in genomic predictions for populations under selection. Genet. Res. Camb. 93, 357-366.

Wahinya, P.K., Jeyaruban, M.G., Swan, A.A., Gilmour, A.R., Magothe, T.M., 2020. Genetic parameters for test-day milk yield, lactation persistency, and fertility in low-, medium-, and high-production systems in Kenya. J. Dairy Sci. 103, 10399-10413.

Wang, Y., Diao, C., Kang, H., Hao, W., Mrode, R., Chen, J., Liu, J., Zhou, L., 2021. A random regression model based on a single-step method for improving the genomic prediction accuracy of residual feed intake in pigs. Front. Genet. 12, 769849.

Wasike, C.B., Kahi, A.K., Peters, K.J., 2014. Genetic relationship between lactation curve traits in the first three parities of dairy cattle. S. Afr. J. Anim. Sci. 44(3), 245-253.

Weigel, K.A., Kriegl, T., Pohlman, A.L., 1999. Genetic analysis of dairy cattle production traits in a management intensive rotational grazing environment. J. Dairy Sci. 82, 191-195.

Westell, R.A., Quaas, R.L., Van Vleck, L.D., 1988. Genetic groups in an animal model. J. Dairy Sci. 71 (5), 1310-1318.

White, I.M.S., Thompson, R., Brotherstone, S., 1999. Genetic and environmental smoothing of lactation curves with cubic splines. J. Dairy Sci. 82, 632–638.

Wiggans, G.R., Sonstegard, T.S., VanRaden, P.M., Matukumalli, L.K., Schnabel, R.D., Taylor, J.F., Schenkel, F.S., Van Tassell, C.P., 2009. Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. J. Dairy Sci. 92, 3431-3436.

Wilmink, J.B.M., 1987. Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. Livest. Prod. Sci. 16, 335-348.

Wolc, A., Lisowski, M., Szwaczkowski, T., 2007. Genetic evaluation of laying hens under fixed regression animal models. Arch. Tierz., Dummerstorf 50 (3), 279-287.

Wood, P., 1967. Algebraic Model of the Lactation Curve in Cattle. Nature **216**, 164–165.

Wright, S., 1938. Proceedings of the National Academy of Sciences 24 (7), 253-259.

Yamazaki, T., Hagiya, K., Takeda, H., Sasaki, O., Yamaguchi, S., Sogabe, M., Saito, Y., Nakagawa, S., Togashi, K., Suzuki, K., Nagamine, Y., 2013. Genetic correlations between milk production traits and somatic cell scores on test day within and across first and second lactations in Holstein cows. Livest. Sci. 152, 120-126.

Zavadilova, L., Jamrozik, J., Schaeffer, L.R., 2005. Genetic parameters for test-day model with random regressions for production traits of Czech Holstein cattle. Czech J. Anim. Sci. 50, 142-154.

Zhang, Y.D., Swan, A.A., Johnston, D.J., Girard, C.J., 2017. Weighting factors for genomic information used in single-step genomic selection in Australian beef. Proc. Assoc. Advmt. Anim. Breed. Genet. 22, 309-312.

Zhou, L., Mrode, R., Zhang, S., Zhang, Q., Li, B., Liu, J., 2018. Factors affecting GEBV accuracy with single-step Bayesian models. Heredity 120, 100-109.

Zwald, N.R., Weigel, K.A., Fikse, W.F., Rekaya, R., 2001. Characterization of dairy production systems in countries that participate in the international bull evaluation service. J. Dairy Sci. 84, 2530-2534.

Zwald, N.R., Weigel, K.A., Fikse, W.F., Rekaya, R., 2003. Identification of factors that cause genotype by environment interaction between herds of Holstein cattle in seventeen countries. J. Dairy Sci. 86, 1009-1018.