



Removing data and using metafounders alleviates biases for all traits in Lacaune dairy sheep predictions

F. L. Macedo,^{1,2,3} J. M. Astruc,⁴ T. H. E. Meuwissen,⁵ and A. Legarra^{1*}

¹INRA, GenPhySE, Castanet-Tolosan 31320, France

²Facultad de Veterinaria, Universidad de la República, 11600 Montevideo, Uruguay

³Department of Animal Breeding and Genetics SLU, S-75007 Uppsala, Sweden

⁴Institut de l'Elevage, Castanet-Tolosan 31321, France

⁵Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, 1432 Ås, Norway

ABSTRACT

Bias in dairy genetic evaluations, when it exists, has to be understood and properly addressed. The origin of biases is not always clear. We analyzed 40 yr of records from the Lacaune dairy sheep breeding program to evaluate the extent of bias, assess possible corrections, and emit hypotheses on its origin. The data set included 7 traits (milk yield, fat and protein contents, somatic cell score, teat angle, udder cleft, and udder depth) with records from 600,000 to 5 million depending on the trait, ~1,900,000 animals, and ~5,900 genotyped elite artificial insemination rams. For the ~8% animals with missing sire, we fit 25 unknown parent groups. We used the linear regression method to compare “partial” and “whole” predictions of young rams before and after progeny testing, with 7 cut-off points, and we obtained estimates of their bias, (over)dispersion, and accuracy in early proofs. We tried (1) several scenarios as follows: multiple or single trait, the “official” (routine) evaluation, which is a mixture of both single and multiple trait, and “deletion” of data before 1990; and (2) several models as follows: BLUP and single-step genomic (SSG)BLUP with fixed unknown parent groups or metafounders, where, for metafounders, their relationship matrix gamma was estimated using either a model for inbreeding trend, or base allele frequencies estimated by peeling. The estimate of gamma obtained by modeling the inbreeding trend resulted in an estimated increase of inbreeding, based on markers, faster than the pedigree-based one. The estimated genetic trends were similar for most models and scenarios across all traits, but were shrunken when gamma was estimated by peeling. This was due to shrinking of the estimates of metafounders in the latter case. Across scenarios, all traits showed bias, generally as an overestimate of

genetic trend for milk yield and an underestimate for the other traits. As for the slope, it showed overdispersion of estimated breeding values for all traits. Using multiple-trait models slightly reduced the overestimate of genetic trend and the overdispersion, as did including genomic information (i.e., SSGBLUP) when the gamma matrix was estimated by the model for inbreeding trend. However, only deletion of historical data before 1990 resulted in elimination of both kind of biases. The SSGBLUP resulted in more accurate early proofs than BLUP for all traits. We considered that a snowball effect of small errors in each genetic evaluation, combined with selection, may have resulted in biased evaluations. Improving statistical methods reduced some bias but not all, and a simple solution for this data set was to remove historical records.

Key words: genomic prediction, bias, accuracy, historical data, multiple trait

INTRODUCTION

Among producers of sheep milk, France has better-established breeding programs. Lacaune is the most important breed in the French dairy sheep industry and has had a breeding program operating since the 1960s. Flocks involved in the Lacaune breeding program comprise ~170,000 females, and the selection scheme tested ~450 rams per year until 2014 (Barillet, 2007). Since 2015 and the implementation of genomic selection, rams are genomically preselected at birth, resulting in the use of 250 new genomically selected rams per year (Jean-Michel Astruc, Institut de l'Elevage, Toulouse; personal communication). Today, the genetic evaluation includes the following 10 traits: milk yield (MY), fat and protein contents (FC and PC), fat and protein yields, SCS, and udder morphological traits including teat angle (TA), udder cleft (UC), udder depth (UD), and teat position.

Due to the relevance of the breed for the dairy sheep industry and its pioneering use of genomic selection,

Received June 11, 2021.

Accepted November 23, 2021.

*Corresponding author: andres.legarra@inrae.fr.

the genetic and genomic evaluations of Lacaune have been studied repeatedly. As it happened in dairy cattle (Spelman et al., 2010; Sargolzaei et al., 2012; Tyrisevä et al., 2018), there are concerns of biases in the genetic evaluations (Duchemin et al., 2012; Astruc et al., 2014; Baloche et al., 2014), with overdispersion of young rams' EBV. This would imply selecting too many young animals and hamper the genetic trend.

However, these studies have all been done on a single truncation point, whose single result (e.g., the existence of bias) may be the outcome of chance (Macedo et al., 2020) and should not be used alone for decision making. Also, the set of scenarios and models used in previous works was limited. Particular scenarios that should be addressed include consideration of missing pedigrees (Quaas, 1988; Misztal et al., 2013; Legarra et al., 2015), use of single-trait or multiple-trait evaluations (Ducrocq, 1994), and deletion of old data (Lourenco et al., 2014).

Concerning missing pedigrees, the use of metafounders implies the use of the so-called $\mathbf{\Gamma}$ matrix (Legarra et al., 2015; Garcia-Baccino et al., 2017), typically estimated using estimates of allele frequencies by generalized least squares of base populations (Garcia-Baccino et al., 2017; Macedo et al., 2020). However, for data sets such as the current one, where founders in genetic groups may be several generations removed from genotypes, estimation of allele frequencies by linear models is difficult; in our case, it resulted in estimates out of the parametric space. The difficulty in estimating the $\mathbf{\Gamma}$ matrix is a common problem for introduction of metafounders in genetic evaluation (Kudinov et al., 2020). In this work, we proposed 2 original approaches to estimate $\mathbf{\Gamma}$.

The objective of this work was to measure the extent of bias in BLUP and single-step genomic BLUP (SSGBLUP) evaluations for Lacaune dairy sheep across many traits, scenarios, models, and truncation points. In addition, we tested 2 methods to estimate relationships across metafounders.

MATERIALS AND METHODS

Data

Records and Pedigree. We used all data available until 2019 for the following 7 traits: MY, FC, PC, SCS, TA, UC, and UD. Although official evaluations also include fat and protein yields, we did not include them in our analyses as they are highly correlated with MY, FC, and PC (Barillet, 2007). Also, as the recording for teat position started in 2013, we did not consider this trait because there were only a few years of records.

In Table 1, we summarize the number of records per trait and the first record's year.

There are 1,868,975 animals in the pedigree. There is ~8% missing pedigree, mostly sire of females, because natural mating in groups is used after AI. Hence, we used genetic groups to consider the missing parents. The first genetic group was assigned to the individuals born before 1966; then, the groups were assigned every 5 yr until 1976, and then every 2 yr until 2020, totaling 25 genetic groups. The first few groups are confounded because the earliest data starts in 1978, but their joint effect (which is all that counts for the first years of records) is well estimated. Treatment of genetic groups will be detailed later.

Genomic Information. We included genomic information of 5,851 AI males, all with both parents known and genotyped with the 50K Illumina chip OvineSNP50. The early reference population (2000–2011) was based on stored DNA samples (Baloche et al., 2014), and all new AI rams have been routinely genotyped since then. Young rams that were discarded (did not enter AI) after genomic preselection were not present in the 5,851 males. Only autosomal SNPs were considered. Quality control of genotypes included individual and marker call rate, minor allele frequency higher than 0.05, removal of Mendelian conflicts, and removal of loci deviating from Hardy-Weinberg equilibrium (number of heterozygotes deviating more than 15% from the expectation based on allele frequencies). After quality control, 37,389 SNPs were retained.

Scenarios

We considered different scenarios, and different models for each scenario, to consider missing pedigrees. The scenarios were as follows: (1) official genetic evaluation (**Off**); (2) deletion of historical data (**Del**); (3) single-trait evaluations (**ST**); and (4) multiple-trait evaluations (**MT**).

Herein, we detail the scenarios and then the models. To check bias, we used linear regression (**LR**; Legarra and Reverter, 2018) that we describe in detail later.

Table 1. Number of records per trait¹

Trait	Number of records	Number of animals	Year of first records
MY	5,696,348	1,693,772	1978
FC, PC	2,024,233	1,249,897	1987
SCS	1,640,409	908,073	1999
TA, UC, UD	638,342	638,342	2000

¹MY = milk yield; FC = fat content; PC = protein content; TA = teat angle; UC = udder cleft; UD = udder depth.

The Official Genetic Evaluation. The Off scenario mimics official (routine) genetic evaluations. Traits MY, FC, and PC are evaluated in single-trait evaluations considering heterogeneity of variances across flocks (Meuwissen et al., 1996), whereas SCS is evaluated in a regular single-trait model. Morphology traits (TA, UC, UD) are evaluated together in a multiple-trait evaluation. In this study, the models for the MY, FC, and PC did not account for the heterogeneity of variance, to speed up computations. In preliminary tests, including or removing heterogeneity of variances led to negligible differences between the solutions for rams (either young or progeny-tested). We used “routine” variance components; most often, these have been estimated some years ago using subsets of data (e.g., Barillet, 2007). These are detailed in Supplemental Tables S1 and S2 (<https://doi.org/10.6084/m9.figshare.17702828>). These genetic parameters were used in all scenarios unless otherwise stated.

Deletion of Historical Data. There is some evidence that by deleting historical data, the bias or overdispersion of young predictions decreases (Lourenco et al., 2014). Previous works in Lacaune showed overdispersion of EBV, so we considered that elimination of historical data could reduce this bias. The reasons for this are unclear, but a possible reason is some discordance between the actual reality and the model being fitted, such that bias for selected traits may accumulate with time. Removing old data would alleviate this snowball effect.

The Del scenario was similar to Off, but we deleted all records and pedigree before 1991, pretending that pedigree and performance recording had started from scratch that year. However, in official evaluations, the earliest performances (MY) started in 1978, and all pedigrees since the 1960s were taken into account. This resulted in a lack of relationships in the first 4 to 5 yr until a minimal pedigree structure was built up (generation interval is ~4 yr). As individuals born before 1991 were eliminated, the number of genetic groups reduced from 25 to 13. We expected a large effect on MY under this scenario, as the deleted data volume was very large (~30%). The other traits started recording more recently and should, in principle, be less affected. Because we focused on rams with well-known pedigrees, we expected the effect to be smaller than in females, but we also expected an effect on genetic trend due to confounding of unknown parent groups and contemporary groups, as every flock had ~8% missing pedigrees.

Single-Trait Evaluations. Another approach was to evaluate all traits as single traits using “routine” variance components. Under the ST scenario, the changes only affect udder morphology traits, as the other traits are already evaluated in single-trait models.

Multiple-Trait Evaluations. A multiple-trait genetic evaluation considers relationship between all traits. This should increase all accuracies and bias due to selection on correlated traits (Ducrocq, 1994). In Lacaune, there are estimations of (co)variances in the first lactation (Barillet, 2007), but not at the level of permanent animal effect. Therefore, we estimated the (co)variances for all traits by Gibbs sampling (12,000 iterations discarding initial 1,200 as burn-in) using all available data and pedigree relationships (without marker information).

Models

All genetic evaluations were performed using the basic model as follows:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{W}_u\mathbf{u} + \mathbf{W}_p\mathbf{p} + \mathbf{e},$$

where \mathbf{y} is the vector of records for the trait, \mathbf{b} is a vector of the fixed effect, \mathbf{u} is a vector of breeding values, \mathbf{p} is a vector of permanent animal effects, \mathbf{e} is a vector of residuals, and \mathbf{X} , \mathbf{W}_u , and \mathbf{W}_p are the incidence matrices for fixed effects, breeding values, and permanent animal effects, respectively. The main fixed effect is contemporary group based on flock, year, and parity number. In particular, interaction effects of age \times year, and parity \times year, are included. This should remove biases due to incorrect modelization of age-parity, and possible confounding with fixed effects (Powell and Wiggans, 1994). The effects are detailed in several papers (Barillet et al., 1992, 2001; Marie-Etancelin et al., 2005). For the udder morphology traits, each animal is scored only once, and thus no permanent animal effect was fitted.

Based on this basic animal model, different genetic evaluations were performed according to scenarios above. We also considered BLUP and SSGBLUP evaluations. The BLUP evaluations used “fixed effect” genetic groups in a traditional manner by expanding \mathbf{A}^{-1} through the QP transformation (Quaas, 1988).

For SSGBLUP, we used (fixed) unknown parent groups using the QP transformation (Misztal et al., 2013). Note that another, more recent, option uses the QP transformation only for the \mathbf{A}^{-1} relationship matrix but not for the \mathbf{G}^{-1} , which seems to avoid double counting and results in better predictions (Cesarani et al., 2021); we have not considered it. Alternatively, to using unknown parent groups, we used the theory of metafounders, which explicitly includes relationships across genetic groups (Legarra et al., 2015). The relatedness of metafounders is described in the so-called $\mathbf{\Gamma}$ matrix (Legarra et al., 2015; Garcia-Baccino et al., 2017), typically estimated using estimates of allele

frequencies by generalized least squares of base population (Garcia-Baccino et al., 2017; Macedo et al., 2020). However, for data sets such as the current one, where founders in genetic groups may be several generations removed from genotypes, estimation of allele frequencies by linear models is difficult and resulted (in our case) in estimates out of the parametric space. The difficulty in estimating the $\mathbf{\Gamma}$ matrix is a common problem for the introduction of metafounders in genetic evaluation (Kudinov et al., 2020). Thus, we considered here 2 (“trend” and “peeling”) manners of estimating $\mathbf{\Gamma}$ without using linear models for allele frequencies of base populations. We present them below.

Thus, we have 4 models for each of the scenarios, as follows:

- BLUP-UPGA with fixed unknown parent groups.
- SSGBLUP-MF-trend, which uses a smooth trend to estimate $\mathbf{\Gamma}$.
- SSGBLUP-MF-peeling, which uses peeling to estimate $\mathbf{\Gamma}$.
- SSGBLUP-UPGH with fixed unknown parent groups.

The 2 MF options are detailed below.

Gamma Estimated by “Trend.” Lacaune is a closed breed; therefore, we expect overall relationship and inbreeding to increase in a steady manner. Sorensen and Kennedy (1983) showed under certain assumptions that the structure of average relationships \bar{A} across generations is as follows:

$$\begin{pmatrix} \bar{A}_{t-1} & \bar{A}_{t-1} & \bar{A}_{t-1} \\ \bar{A}_{t-1} & \bar{A}_t & \bar{A}_t \\ \bar{A}_{t-1} & \bar{A}_t & \bar{A}_{t+1} \end{pmatrix},$$

Because inbreeding is a half relationship, $\bar{A}_t - \bar{A}_{t-1} \approx 2\Delta F$. This leads to the following model for $\mathbf{\Gamma}$, starting from time 0, and which includes a steady trend of relationship as follows:

$$\mathbf{\Gamma} = \begin{pmatrix} \Gamma_0 & \Gamma_0 & \Gamma_0 & \Gamma_0 & \dots \\ \Gamma_0 & \Gamma_0 + 2\Delta F_{(\gamma)} & \Gamma_0 + 2\Delta F_{(\gamma)} & \Gamma_0 + 2\Delta F_{(\gamma)} & \dots \\ \Gamma_0 & \Gamma_0 + 2\Delta F_{(\gamma)} & \Gamma_0 + 4\Delta F_{(\gamma)} & \Gamma_0 + 4\Delta F_{(\gamma)} & \dots \\ \Gamma_0 & \Gamma_0 + 2\Delta F_{(\gamma)} & \Gamma_0 + 4\Delta F_{(\gamma)} & \Gamma_0 + 6\Delta F_{(\gamma)} & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}.$$

Thus, only 2 values are needed to build the whole $\mathbf{\Gamma}$: $\mathbf{\Gamma}_0$ and $\Delta F_{(\gamma)}$. Note that $\Delta F_{(\gamma)}$ is not the same as ΔF [e.g., with a single metafounder, $\Delta F_{(\gamma)} = \Delta F(1 - \gamma/2)$]. We

found, by grid search, the values of $\mathbf{\Gamma}_0$ and $\Delta F_{(\gamma)}$ that maximized (assuming, as an approximation, normality of gene content) the likelihood of observed marker genotypes given $\mathbf{\Gamma}$ (Legarra et al., 2015).

Gamma Estimated by “Peeling.” Matrix $\mathbf{\Gamma}$ contains covariances of allele frequencies at base populations. Use of linear models to estimate allele frequencies yields only approximates results because genotypes follow Mendelian transmission. Thus, the optimal way to estimate allele frequencies is to maximize a Mendelian likelihood including all pedigree and marker information. To estimate base population allele frequencies, we used single-locus iterative peeling (Fernando et al., 1993; Kerr and Kinghorn, 1996) as programmed in linkage disequilibrium multilocus iterative peeling (LDMIP; Meuwissen and Goddard, 2010). Due to time constraints, we considered a random subset of 4,000 markers.

Model Checking

We used method LR (Legarra and Reverter, 2018) to compare the models. The method LR is based on the comparison between genomic (\mathbf{G})EBV with partial data (earlier) breeding values (\hat{u}_p), estimated at the time of selection, versus whole data (later) breeding values (\hat{u}_w) of a group of individuals, called focal individuals. The focal individuals in this work were the cohorts of AI rams that, every year, were selected (based on parent average up to 2015) and put to progeny testing. In the following years, these rams have progeny (females with records), although the number of progeny varies across traits. Therefore, we worked with 9 focal groups composed of AI rams. The number of rams and number of daughters per ram per year of birth of the 9 focal groups is detailed in Table 2.

We performed genetic evaluations with truncation dates given by years of birth in Table 2 (e.g., for the focal group born in 2000, the “partial” evaluation is obtained deleting all records after 2000). The last evaluation had all available data until 2019. For the Del scenario, the first evaluation was in 1995; otherwise, everything was identical. We obtained EBV with BLUP-UPGA for every evaluation, whereas SSGBLUP models were used only in genetic evaluations after 2000. For each focal group, there were several comparisons. For instance, consider the focal individuals for MY born in 1985. For them, we obtained 9 sets of bias, slope, and accuracy estimates (\hat{u}_p from 1985 vs. \hat{u}_w from 1990, \hat{u}_p from 1985 vs. \hat{u}_w from 1995 ... \hat{u}_p from 1985 vs. \hat{u}_w from 2019).

We used the estimators described by Legarra and Reverter (2018) and refined by Macedo et al. (2020).

We described them briefly. An estimation of the bias was obtained, for each cohort of focal animals, by the simple difference between the averages of old and recent EBV as $\hat{\Delta}_p = (\hat{\mathbf{u}}_p) - (\hat{\mathbf{u}}_w)$ with expected value of 0. In other words, for a group of contemporary animals selected equally, their average expected and observed genetic value (i.e., estimated genetic) gain should be identical (and identical to the true value); this is a property of BLUP. Dispersion (regression of true on estimated breeding value) is estimated from the slope of the regression of \hat{u}_w on \hat{u}_p : $\hat{b}_p = \frac{Cov(\hat{\mathbf{u}}_p, \hat{\mathbf{u}}_w)}{Var(\hat{\mathbf{u}}_p)}$, with an expectation of 1.

Finally, an estimator of the ratio of accuracies (acc) is the correlation between EBV, where

$$\hat{\rho}_{p,w} = \frac{Cov(\hat{\mathbf{u}}_p, \hat{\mathbf{u}}_w)}{\sqrt{Var(\hat{\mathbf{u}}_p)Var(\hat{\mathbf{u}}_w)}}, \text{ which (on expectation) equals}$$

$\frac{acc_p}{acc_w}$. The relative gain in accuracy from a partial data set to a whole one can be obtained by

$$\frac{1}{\hat{\rho}_{p,w}} - 1 = \frac{acc_w - acc_p}{acc_p} \text{ (Bermann et al., 2021). Also, assuming that } acc_w(BLUP) \approx acc_w(SSGBLUP), \text{ which is}$$

a reasonable assumption for progeny-tested rams as is the case in this work (but not generally), then $\hat{\rho}_{p,w}(SSGBLUP) - \hat{\rho}_{p,w}(BLUP)$ indicates the increase in relative accuracies from BLUP to SSGBLUP.

For each year of partial data sets, we obtained a set of estimators. However, this set varies for each focal group. For instance, for the focal group born in 1985, for the MY and BLUP-UPGA model, we made 9 comparisons, whereas for the focal group born in 2016, we made a single one. Therefore, to summarize all comparisons, the use of raw averages is not correct. The final results of the estimators were obtained following the linear model $s_{ij} = p_i + w_j + e_{ij}$, where s_{ij} is the observed statistic for year of “partial” i and year of

“whole” j , and p_i and w_j are the respective effects. The final estimate was obtained as the estimable function “average s_{ij} if the design was balanced” (Macedo et al., 2020).

RESULTS

Genetic Parameter Estimates

In Table 3 and Supplemental Tables S3 and S4 (<https://doi.org/10.6084/m9.figshare.17702828>), we present the estimated genetic, residual, and permanent ratios of variances including correlations, as well as total phenotypic variances. These results are very similar to reported estimates (Barillet, 2007); the most relevant results, but already known, are the opposition of MY with contents, SCC, and UD, and the moderate genetic correlations across udder traits and SCC. These correlations may influence the bias for some trait when selection on other correlated traits is ignored in single-trait evaluations.

Estimates of Metafounders Relationships

Gamma Estimated by “Trend.” The values of Γ_0 and $\Delta F_{(\gamma)}$ that maximize the likelihood of marker genotypes, given Γ (assuming normality), were 0.3 and 0.025, respectively. Note that these values yield an increase in “regular” inbreeding of $\Delta_F \approx \Delta F_{(\gamma)} / (1 - \Gamma_0 / 2) = 0.03$, from one genetic group to the next, every 2 yr, resulting in an increase of inbreeding of 0.015 per year. Rodríguez-Ramilo et al. (2019) reported that Δ_F per generation (generation interval ~ 3 yr) is 0.022, resulting in an increase of inbreeding of 0.007 per year. Thus, our method overestimated the increase of inbreeding and should be refined. Our estimates of Γ_0 and $\Delta F_{(\gamma)}$ lead to the Γ matrix presented below that we used for further analyses:

Table 2. Number of rams per focal group per year of birth¹

Year	Milk yield		SCS		Udder depth	
	Rams	Med; IQ	Rams	Med; IQ	Rams	Med; IQ
1995	392	45; 31				
2000	383	41; 27	377	44; 30		
2005	407	47; 31	398	38; 26	332	35; 31
2010	417	44; 27	406	45; 26	327	38; 23
2012	368	47; 28	366	43; 25	316	39; 17
2014	338	47; 33	338	45; 24	318	40; 28
2016	241	39; 39	240	46; 31	236	38; 20

¹Med = median number of daughters with records in the following years; IQ = interquartile range.

Table 3. Genetic correlations (off-diagonal) and heritabilities (diagonal) estimated by Gibbs sampling¹

Trait	MY	FC	PC	SCS	TA	UC	UD
MY	0.34	-0.35	-0.46	0.05	0.03	0.21	-0.47
FC	-0.35	0.48	0.59	0.04	0.13	-0.14	-0.01
PC	-0.46	0.59	0.59	0.06	0.04	-0.10	0.07
SCS	0.05	0.04	0.06	0.18	0.18	-0.25	-0.32
TA	0.03	0.13	0.04	0.18	0.42	-0.33	-0.22
UC	0.21	-0.14	-0.10	-0.25	-0.33	0.40	0.04
UD	-0.47	-0.01	0.07	-0.32	-0.22	0.04	0.37

¹MY = milk yield; FC = fat content; PC = protein content; TA = teat angle; UC = udder cleft; UD = udder depth. Posterior SD < 0.01.

$$\begin{pmatrix} 0.3 & 0.3 & 0.3 & 0.3 & \dots & 0.3 \\ 0.3 & 0.3425 & 0.3425 & 0.3425 & \dots & \vdots \\ 0.3 & 0.3425 & 0.3850 & 0.3850 & \dots & \vdots \\ 0.3 & 0.3425 & 0.3850 & 0.4275 & \dots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & 1.2775 \\ 0.3 & \dots & \dots & \dots & 1.2775 & 1.3200 \end{pmatrix}.$$

This matrix resulted in correlations across metafounders of ~ 0.95 .

Gamma Estimated by “Peeling.” Estimates of allelic frequencies were very highly correlated across base populations, resulting in $\mathbf{\Gamma}$ with diagonal values of ~ 0.89 and off-diagonal values of ~ 0.88 . This resulted in very high a priori covariances across metafounders (the correlation was ~ 0.99), which in turn was expected to shrink all their estimates toward 0. The reason for these highly correlated estimates of allele frequencies was probably lack of information; it would seem that peeling split the observed genotypes equally across different origins.

Although the method correctly considered the Mendelian distribution of the genotypes, it was still not optimal for several reasons as follows: (1) iterative peeling is approximate, (2) using single-marker likelihoods ignores information from neighboring markers, and especially (3) an unbiased estimate of marker allele frequencies does not imply an unbiased estimate of $\mathbf{\Gamma}$. However, use of peeling results can clearly be improved. For instance, instead of using the square of the estimate of allele frequencies to estimate the diagonal of $\mathbf{\Gamma}$, one could use average individual estimated homozygosities. We consider that exploring more options is very time consuming and out of the scope of this paper.

Genetic Trends and Estimates of Unknown Parent Groups and Metafounders

In Figure 1, we present the genetic trend (average across females EBV per year of birth) for all analyzed

traits, obtained from BLUP-UPGA in the Off scenario. Note that all traits have desirable trends in the recent years (for TA, lower values are better). We expected BLUP-UPGA to give unbiased estimates of genetic trends, given that genomic selection started only in 2015. There was a genetic gain for all traits, stronger for MY as expected, given its heritability and weight in the selection index. The trends showed the changes in the selection objectives toward increasing milk contents, less mastitis, and a better conformed udder (Barillet, 2007).

For udder traits, genetic trends were almost identical across models and scenarios. However, this was not true for other traits. We show results for Off and Del for the traits MY (Figure 2) and SCS (Figure 3). Milk yield is a trait that has been selected since the beginning of the scheme in the early 1970s, whereas SCS was effectively selected since ~ 2005 . The slope of genetic trend for MY was steeper for Off than for Del (Figure 2). This was a bit surprising because one would think that deleting data before 1991 should not affect genetic trend estimated in 2010. In our opinion, this could mean that there was a small systematic bias that cumulated in time in genetic evaluations for MY. In addition, the genetic trend for MY was lower for MF-peeling; this probably reflected the fact that, for this data set, the estimated $\mathbf{\Gamma}$ using peeling forced the genetic groups to be a priori very similar to each other, which resulted in strongly shrunken estimates of the metafounders (Figure 4). This, in turn, penalized the genetic trend, even if the number of animals with missing pedigree was rather small ($\sim 8\%$). The strategy MF-trend did not suffer from this drawback.

As for SCS, the genetic trend shown in Figure 3 depended not on the scenario, but on the model. Thus, the use of SSGBLUP resulted in faster genetic trend, regardless of the scenario. This perhaps indicated that the extra accuracy by SSGBLUP, especially given that *SOCS2*, a major gene for SCS segregating in this population, explained 12% of the genetic variance and that its effect is well captured with genomic predictions

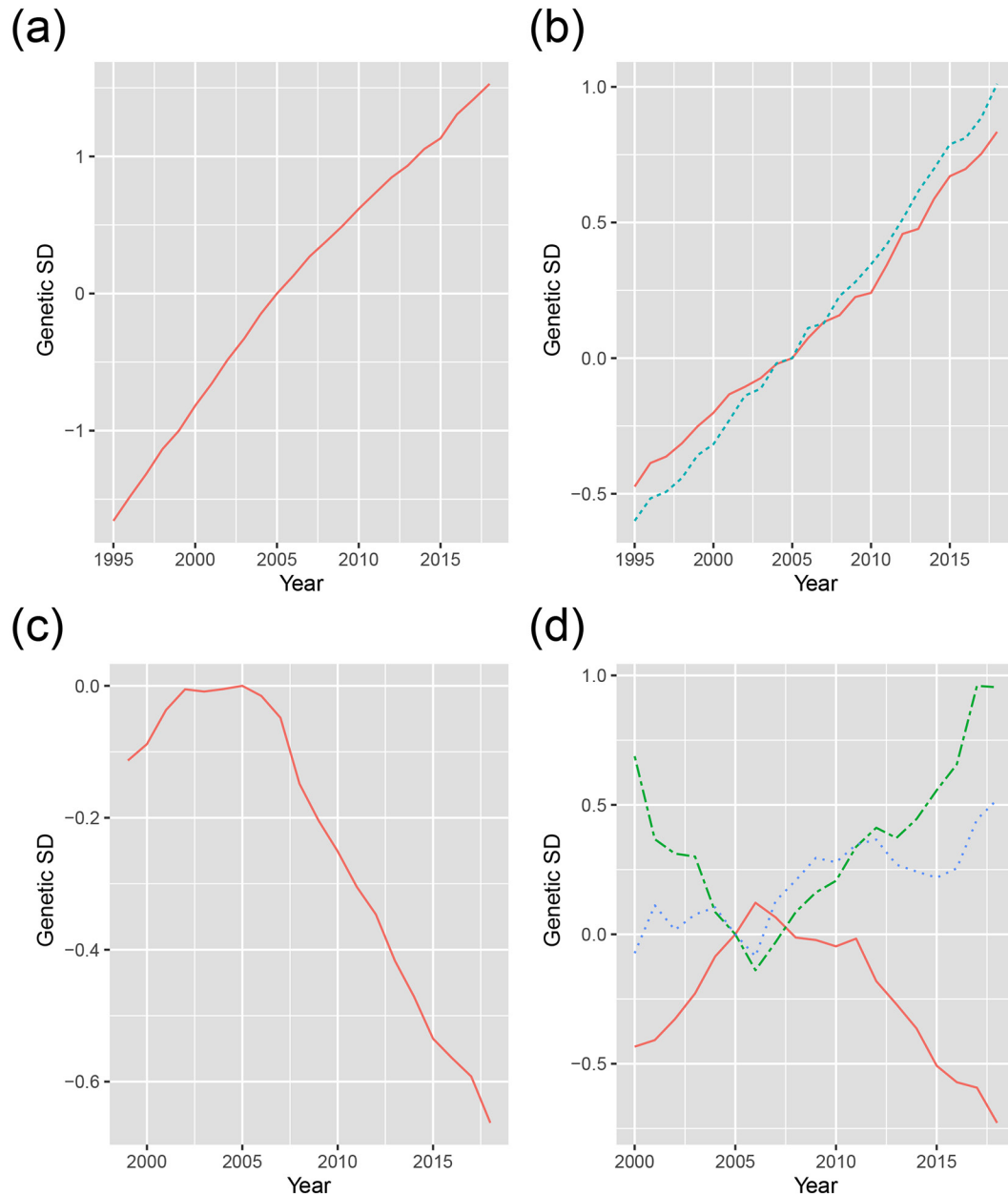


Figure 1. Genetic trend for all traits expressed in genetic standard deviations. (a) Milk yield; (b) fat content (red solid line) and protein content (blue dotted line); (c) SCS; (d) teat angle (red solid line), udder cleft (green dot-dash line), and udder depth (blue dotted line).

(Oget et al., 2019). For this trait, there was not much difference between MF-peeling and MF-trend.

Figure 5 shows, for MY, the variation of the estimators of slope \hat{b}_p depending on the pairs of points partial-whole that we compared for BLUP-UPGA under the Del scenario. The slope fluctuated between 0.80 and 1 and had no clear pattern. These observations of the high variability of the estimators across years agreed with previous studies done in Manech Tête Rousse

(Macedo et al., 2020), and they highlighted the need of several cut-off tests to assess unbiasedness.

Table 4 presents final estimates of bias across all traits, scenarios, and models. The only trait that is consistently overestimated is MY, which was the one with faster genetic progress, whereas the other traits tended to be underestimated. In the Del scenario, biases were mildly reduced for most traits and strongly reduced for MY, as the bias dropped from 0.10 to 0.20 genetic stan-

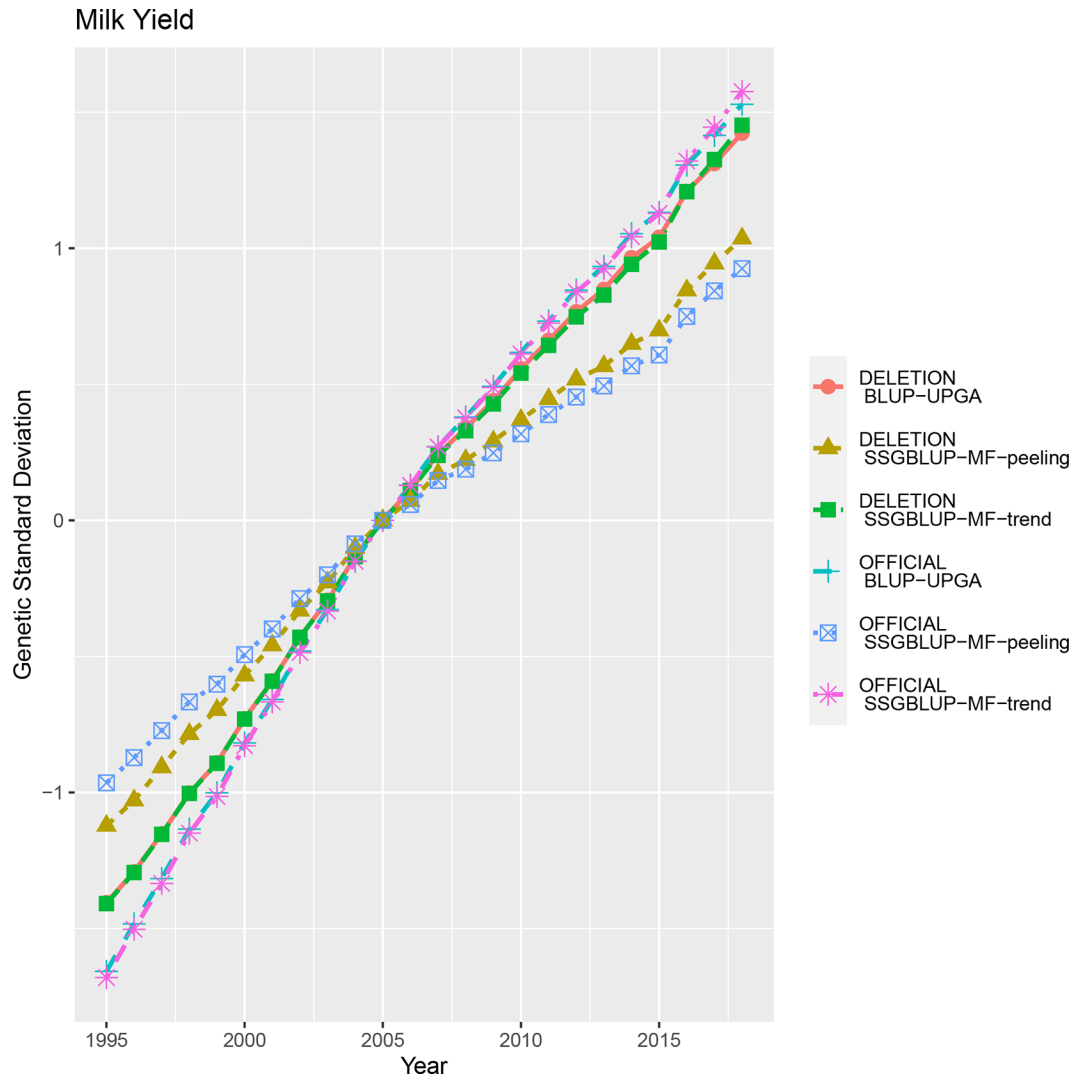


Figure 2. Estimated genetic trends for milk yield for different data sets, truncated at 1991 (Deletion) or not (Official) using BLUP or single-step genomic BLUP (SSGBLUP). Models: BLUP UPGA (with fixed unknown parent groups in A); SSGBLUP-MF-trend, with a smooth trend to estimate Γ across metafounders (MF); and SSGBLUP-MF-peeling, which uses peeling to estimate Γ .

dard deviation to almost 0. On the other hand, fitting ST, MT, or Off evaluations resulted in similar biases.

As for the models, there was little difference except for MY. There was more bias for BLUP-UPGA and SSGBLUP-MF-peeling than for SSGBLUP-MF-trend and SSGBLUP-UPGH. This was surprising because the 2 biased models (BLUP-UPGA and SSGBLUP-MF-peeling) did not share features. Again, in the Del scenario, these biases disappeared; therefore, they had to do with including old data and perhaps the existence of small, cumulating, biases.

The values of \hat{b}_p are summarized in Table 5. Most traits showed overdispersion ($\hat{b}_p < 1$), which can be quite large in some combinations of trait, scenario, and method. The highest overdispersions were, again, ob-

served for MY, with values as low as 0.60. This agreed with previous results in the same breed (Baloche et al., 2014). On the scenario MT, the slope did not perform better; as the opposite, it presented a slight magnification of the overdispersion. However, by deleting historical data in Del, the overdispersion was almost eliminated in production traits and SCS, and greatly diminished for udder morphology traits. As for the latter, these traits were not recorded before 1990, so that the changes were entirely due to removal of pedigree and probably of unknown parent groups or metafounders that were difficult to estimate.

Inside each scenario, there were important differences across models, in particular for MY and PC, which are traits with strong genetic trend. The model that

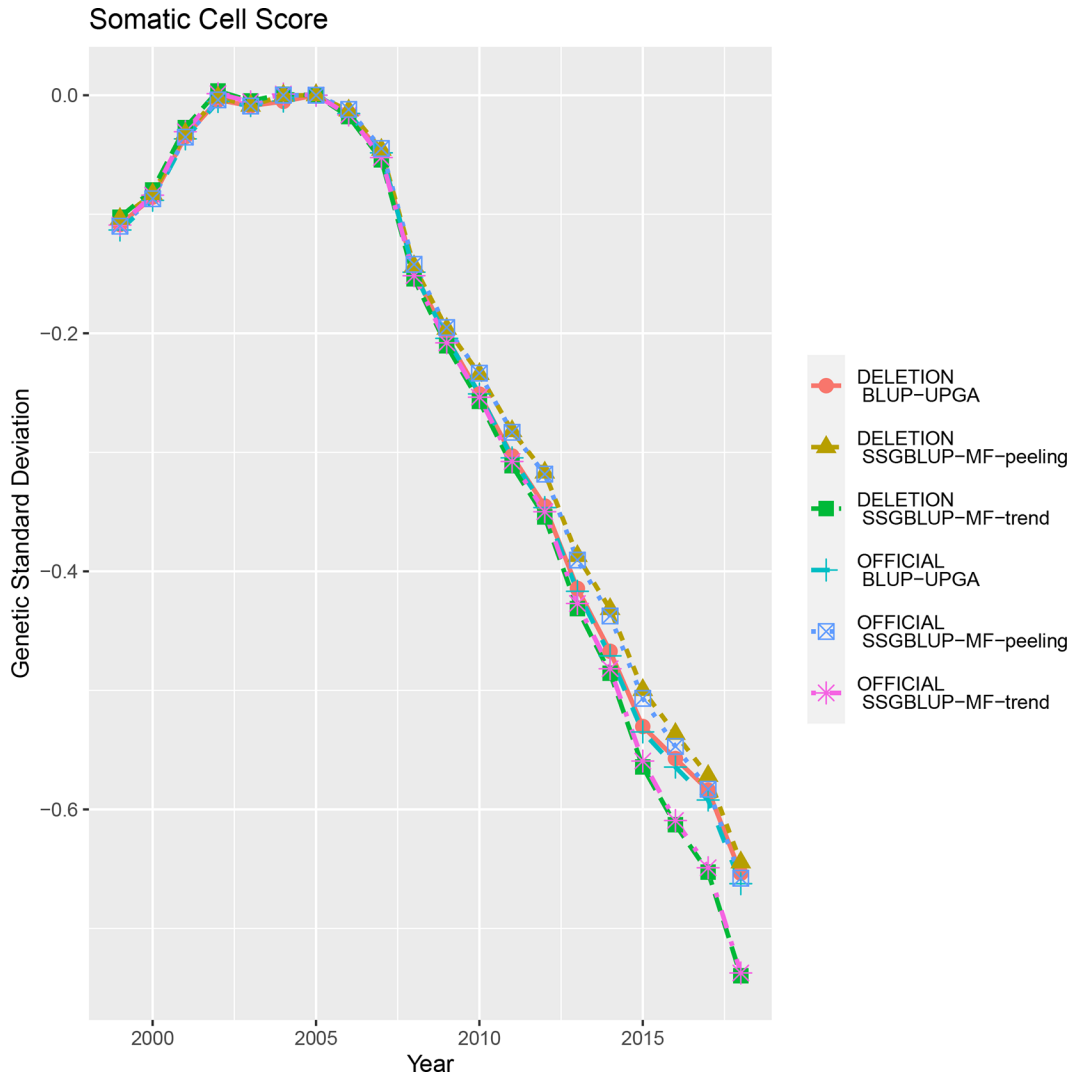


Figure 3. Estimated genetic trends for SCS for different data sets, truncated at 1991 (Deletion) or not (Official), using BLUP or single-step genomic BLUP (SSGBLUP). Models: BLUP UPGA (with fixed unknown parent groups in A); SSGBLUP-MF-trend, with a smooth trend to estimate Γ across metafounders (MF); SSGBLUP-MF-peeling, which uses peeling to estimate Γ ; and SSGBLUP-UPGH with fixed unknown parent groups in H.

presented more overdispersion was SSGBLUP-UPGH, followed by BLUP-UPGA, probably because UPG were fit as fixed effects with no shrinkage in both cases. On the other hand, the model that presented less overdispersion was SSGBLUP-MF-trend. However, all differences in biases across models almost disappeared when historical data were deleted (Del).

In Table 6, we present ratios of accuracies $\hat{\rho}_{p,w}$ for the 2 least biased models, BLUP-UPGA and SSGBLUP-MF-trend, for 2 of the scenarios (ST and MT were very similar to Off). The high values observed for Del were puzzling, and we do not have a clear explanation. Nevertheless, in both cases (Off or Del), the difference $\hat{\rho}_{p,w}(\text{SSGBLUP}) - \hat{\rho}_{p,w}(\text{BLUP})$ results in values around

0.03 to 0.20 of extra relative accuracy for SSGBLUP compared with BLUP, which agreed with previous results in the breed. As for the different modeling of UPG and MF, generally SSGBLUP_MF_trend was as accurate or more accurate than the other options for Off, whereas all 3 models were equally accurate for Del.

DISCUSSION

Most evaluations tended to be biased either for bias $\hat{\Delta}_p \neq 0$ (sometimes over- and sometimes underestimation) or for slope \hat{b}_p (generally $\hat{b}_p < 1$, implying overestimation of selected candidates). This was a problem, as expected genetic gains based on recent trends were not

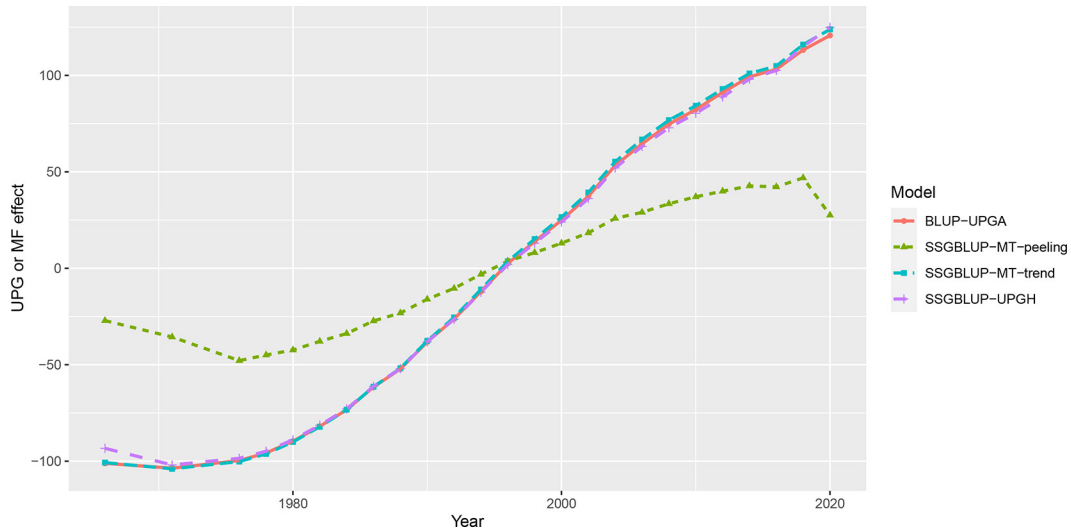


Figure 4. Estimates of unknown parent groups (UPG) and metafounders effect for milk yield for the official genetic evaluation scenario. Models: BLUP-UPGA, with fixed unknown parent groups; SSGBLUP-MF-trend, with a smooth trend to estimate Γ across metafounders (MF); SSGBLUP-MF-peeling, which uses peeling to estimate Γ ; and SSGBLUP-UPGH with fixed unknown parent groups.

accomplished, although an economic quantification of the losses was missing.

Some, but not all, bias was reduced using MT models. The source of these remaining biases was largely unknown, as the Lacaune breed has a well-organized performance recording and low level of missing pedigree. We hypothesized that bias came from inherently imperfect statistical models, and that accumulation of small biases during many generations (more than 10 since the 1970s) led to a snowball effect and observed

existing bias. For instance, small noises in estimation of breeding values may lead the genetic evaluation system to infer that the average breeding value of the next generation is, say, 11% better instead of the true 10%. The 1-percentage-point difference enters into parent averages that, in turn, enter into prediction of genetic and environmental effects. The same effect would be achieved if, for the same trait, the genetic correlation across distant generations was not 1 (Tsuruta et al., 2004), for instance, due to genotype by environment

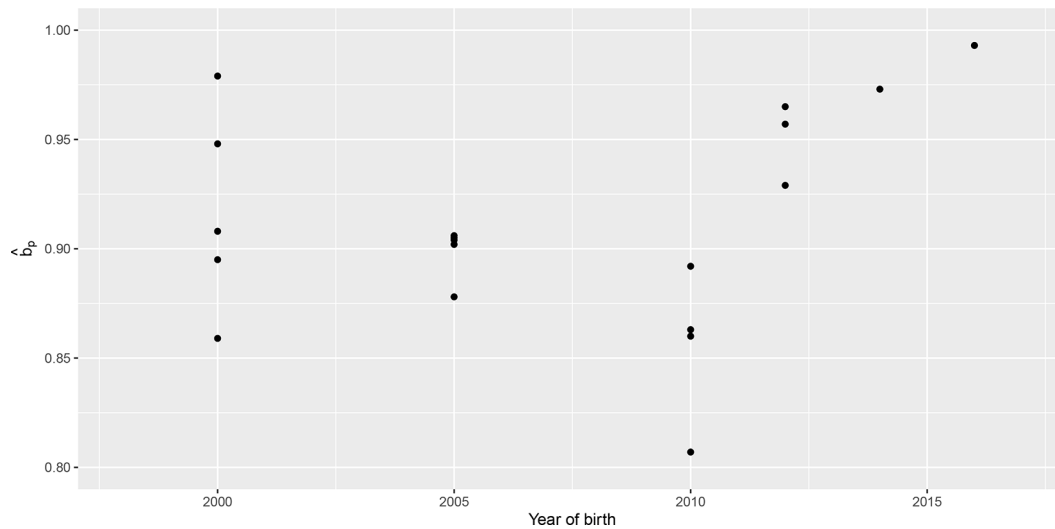


Figure 5. Estimated slope \hat{b}_p (regression of EBV_w on EBV_p) as function of year of birth of focal individuals. EBV_w = EBV based on whole data; EBV_p = EBV based on partial data. Genetic evaluation: BLUP-UPGA (with fixed unknown parent groups) in scenario with deletion of historical data.

Table 4. Bias ($\hat{\Delta}_p$) expressed in genetic standard deviations

Evaluation ¹	Model ²	Trait ³						
		MY	FC	PC	SCS	TA	UC	UD
Official	BLUP-UPGA	0.17	-0.13	-0.25	-0.10	-0.04	-0.04	-0.07
Official	SSGBLUP-MF-trend	0.11	-0.09	-0.15	-0.08	0.06	-0.10	-0.06
Official	SSGBLUP-MF-peel	0.36	0.00	-0.05	-0.13	0.03	-0.05	-0.06
Official	SSGBLUP-UPGH	0.10	-0.10	-0.16	-0.08	0.06	-0.09	-0.06
Deletion	BLUP-UPGA	-0.01	-0.14	-0.21	-0.05	-0.04	-0.03	-0.04
Deletion	SSGBLUP-MF-trend	-0.02	-0.09	-0.11	-0.04	-0.05	-0.03	-0.03
Deletion	SSGBLUP-MF-peeling	0.05	-0.06	-0.09	-0.05	-0.05	-0.03	-0.03
Deletion	SSGBLUP-UPGH	-0.02	-0.09	-0.11	-0.03	-0.04	-0.03	-0.03
ST	BLUP-UPGA					-0.06	-0.04	-0.07
ST	SSGBLUP-MF-trend					-0.06	-0.04	-0.08
ST	SSGBLUP-MF-peeling					-0.08	-0.05	-0.10
ST	SSGBLUP-UPGH					-0.06	-0.04	-0.08
MT	BLUP-UPGA	0.18	-0.05	-0.14	-0.06	0.07	-0.05	-0.14
MT	SSGBLUP-MF-trend	0.09	-0.03	-0.08	-0.04	0.05	-0.04	-0.12
MT	SSGBLUP-MF-peel	0.29	0.02	-0.02	-0.10	0.03	0.01	-0.13
MT	SSGBLUP-UPGH	0.07	-0.03	-0.09	-0.03	0.06	-0.04	-0.13

¹Scenarios: Official = official evaluation; Deletion = deletion of historical data; ST = single-trait evaluation; MT = full multiple-trait evaluation.

²Models: BLUP-UPGA, with fixed unknown parent groups; single-step genomic (SSG)BLUP-MF-trend = uses a smooth trend to estimate Γ ; SSGBLUP-MF-peeling, which uses peeling to estimate Γ ; and SSGBLUP-UPGH with fixed unknown parent groups.

³MY = milk yield; FC = fat content; PC = protein content; TA = teat angle; UC = udder cleft; UD = udder depth. Standard errors between 0.01 and 0.03.

interactions. Thus, selection differentials at one generation are not fully reflected in the following generations. If we're correct, this effect would explain the different genetic trend in Off and in Del. However, this hypothesis has not been verified through this study and should be properly addressed elsewhere.

As for the means of correcting the bias, the most obvious (improving the model) is not always easy as

linear models have their limitations, for instance, if important factors are unregistered or unaccounted for. Indeed, an outdated correcting factor (age – parity) introduced serious bias in US dairy evaluation (Powell and Wiggans, 1994). This particular factor (age – parity) is taken into account in the Lacaune evaluation, but it still may remain inaccurate due to confounding (e.g., with herd-year-parity) as well as other factors.

Table 5. Slope (\hat{b}_p) of the regression of EBV_w (whole data set) on EBV_p (partial data set)

Evaluation ¹	Model ²	Trait ³						
		MY	FC	PC	SCS	TA	UC	UD
Official	BLUP-UPGA	0.71	0.89	0.85	0.88	0.76	0.92	0.72
Official	SSGBLUP-MF-trend	0.86	0.93	0.91	0.87	0.83	0.79	0.72
Official	SSGBLUP-MF-peeling	0.75	0.89	0.88	0.82	0.78	0.74	0.68
Official	SSGBLUP-UPGH	0.59	0.87	0.78	0.87	0.83	0.79	0.73
Deletion	BLUP-UPGA	0.97	0.99	0.94	0.96	0.90	0.90	0.92
Deletion	SSGBLUP-MF-trend	0.99	0.99	0.98	0.98	0.89	0.89	0.91
Deletion	SSGBLUP-MF-peeling	0.94	0.97	0.95	0.96	0.87	0.87	0.90
Deletion	SSGBLUP-UPGH	0.98	0.99	0.97	0.98	0.89	0.88	0.91
ST	BLUP-UPGA	0.71	0.89	0.85	0.88	0.85	0.94	0.74
ST	SSGBLUP-MF-trend	0.86	0.93	0.92	0.87	1.30	1.44	1.29
ST	SSGBLUP-MF-peeling	0.75	0.89	0.88	0.82	1.17	1.24	1.09
ST	SSGBLUP-UPGH	0.59	0.87	0.78	0.87	1.06	1.12	1.01
MT	BLUP-UPGA	0.69	0.82	0.81	0.80	0.81	0.76	0.59
MT	SSGBLUP-MF-trend	0.83	0.88	0.87	0.81	0.81	0.76	0.65
MT	SSGBLUP-MF-peeling	0.73	0.84	0.83	0.75	0.76	0.70	0.60
MT	SSGBLUP-UPGH	0.60	0.84	0.77	0.81	0.80	0.76	0.65

¹Scenarios: Official = official evaluation; Deletion = deletion of historical data; ST = single-trait evaluation; MT = full multiple-trait evaluation.

²Models: BLUP-UPGA, with fixed unknown parent groups; single-step genomic (SSG)BLUP-MF-trend = uses a smooth trend to estimate Γ ; SSGBLUP-MF-peeling, which uses peeling to estimate Γ ; and SSGBLUP-UPGH with fixed unknown parent groups.

³MY = milk yield; FC = fat content; PC = protein content; TA = teat angle; UC = udder cleft; UD = udder depth. Standard errors between 0.01 and 0.02.

Thus, it seems hard to fully debug a linear model. However, from our results, a simple solution was deleting old data (scenario Del). This simple procedure reduced the bias to almost nothing for all traits and both statistics ($\hat{\Delta}_p$ and \hat{b}_p) and is simple to implement. This was already suggested by Lourenco et al. (2014). Cesarani et al. (2021) showed approximately unbiased evaluations, regardless of data truncation, when the most correct model (an SSGBLUP method) was used. However, for seemingly incorrect models (another SSGBLUP method or BLUP), Cesarani et al. (2021) observed a positive effect (reduction of overdispersion) when truncating old data; the effect was marked in cows and rather small in bulls. The reason why we observed a more marked effect is probably because AI rams did not have as many daughters as AI bulls (hundreds for AI rams compared with thousands or more for AI bulls). It can be difficult to delete data because only really old data can be removed (i.e. no breeding animals should be removed), and it removes the possibility of observing long-time genetic trends.

Another possible explanation for bias in genomic predictions is selective or scarce genotyping (although this does not explain why in some cases BLUP evaluations are biased). This was not the case in this work because there was no selective genotyping (all AI rams, pre- or postgenomic selection, were genotyped) and these rams covered all the genetic diversity of the population, as next generations were either offspring (through AI) or grand-offspring (through natural mating rams' sons of AI rams) of genotyped AI rams. We do not believe that compatibility of pedigree and genomic information was an important source of bias in our case (at least for the models that we tested), given that we found that BLUP was also biased.

Except in the case of Del, udder traits always showed some bias as shown by values of \hat{b}_p . Strangely, moving from ST to MT changed \hat{b}_p from $\hat{b}_p > 1$ (underdispersion) to $\hat{b}_p < 1$ (overdispersion). In a way, if for most traits there is overdispersion ($\hat{b}_p < 1$) due to selection, using a MT model changes udder traits from an anomalous situation of underdispersion ($\hat{b}_p > 1$) to a “normal” situation of overdispersion ($\hat{b}_p < 1$). Perhaps the reason for overdispersion for ST was because udder traits, and in particular UD, were genetically correlated with MY and SCS, and this results in underdispersion for ST models where the effect of such selection was ignored.

The use of metafounders seemed to slightly reduce biases, in particular for the slope \hat{b}_p , compared with “fixed” UPG as shown in SSGBLUP-MF-trend versus either BLUP-UPGA or SSGBLUP-UPGH. We attributed this to the shrinkage of MF estimates, which resulted in shrinkage of genetic trends. Indeed, in Del, all methods were unbiased. Thus, the simplest alternative to remove bias in this data set was to remove data earlier than 1990. This has the desirable byproduct that all models for MF and UPG were, roughly, equally accurate. In other words, modeling genetic trend when parentships are missing becomes much easier. As for the method to estimate Γ , the method using trend of inbreeding gave a better answer than using peeling (or our particular use of peeling results). This method using trend of inbreeding was straightforward to use in single-breed populations, although it should be modified for use in multiple populations, possible crosses, and more complicated UPG structures than those solely based on year of birth.

When we compare bias $\hat{\Delta}_p$ (Table 4) with the genetic trends (Figure 1), it seemed that the stronger the trend,

Table 6. Correlation ($\hat{\rho}_{p,w}$) of the regression of EBV_w (whole data set) on EBV_p (partial data set)

Evaluation ¹	Model ²	Trait ³						
		MY	FC	PC	SCS	TA	UC	UD
Official	BLUP-UPGA	0.45	0.57	0.59	0.52	0.69	0.75	0.61
Official	SSGBLUP-MF-trend	0.65	0.72	0.73	0.71	0.68	0.66	0.62
Official	SSGBLUP-MF-peeling	0.61	0.71	0.72	0.68	0.65	0.63	0.59
Official	SSGBLUP-UPGH	0.54	0.70	0.68	0.71	0.68	0.67	0.63
Deletion	BLUP-UPGA	0.85	0.84	0.84	0.84	0.79	0.83	0.85
Deletion	SSGBLUP-MF-trend	0.91	0.92	0.92	0.91	0.82	0.85	0.87
Deletion	SSGBLUP-MF-peeling	0.89	0.91	0.91	0.90	0.81	0.84	0.87
Deletion	SSGBLUP-UPGH	0.90	0.92	0.91	0.91	0.81	0.85	0.88

¹Scenarios: Official = official evaluation; Deletion = deletion of historical data.

²Models: BLUP-UPGA, with fixed unknown parent groups; single-step genomic (SSG)BLUP-MF-trend = uses a smooth trend to estimate Γ ; SSGBLUP-MF-peeling, which uses peeling to estimate Γ ; and SSGBLUP-UPGH with fixed unknown parent groups.

³MY = milk yield; FC = fat content; PC = protein content; TA = teat angle; UC = udder cleft; UD = udder depth. Standard errors between 0.01 and 0.02.

the higher $\hat{\Delta}_p$. This was as expected as an unselected trait would not change the mean, and thus $\hat{\Delta}_p$ was expected to be 0. For a selected trait, the realized selection response (seen in the whole data set) was generally lower than the expected response (inferred from partial data), resulting in nonzero $\hat{\Delta}_p$ values.

CONCLUSIONS

Our results have some consequences for the dairy sheep industry, and perhaps genetic evaluation methods in all species. If, as our data seems to suggest, bias accumulates due to a snowball effect of small biases (due perhaps to incomplete modeling), then deleting old data is a simple and efficient option. This could eliminate problems of bias in the dairy industry. Use of metafounders through the new method “MF-trend” to estimate Γ is an interesting refinement but does not eliminate all bias. Multiple-trait models can be run in large data sets (comparable to most breeds that are not Holstein and most countries outside of the United States) but it does not eliminate bias.

ACKNOWLEDGMENTS

F. L. Macedo acknowledges financing from Région Occitanie (France) and INRAE SelGen metaprogram. This project has received funding from the European Unions’ Horizon 2020 Research and Innovation program under grant agreement N°772787 -SMARTER. We are grateful to the genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, <https://doi.org/10.15454/1.5572369328961167E12>) for providing help, computing, and storage resources. We are also grateful to the organizations responsible for the Lacaune breeding program (Upa Lacaune, Ovitest, Confederation Générale de Roquefort) for allowing us to use the database (pedigree, performance, and genotypes). T. M. acknowledges financing from the Norwegian Research Council projects 225297 and 309611. The authors have not stated any conflicts of interest.

REFERENCES

Astruc, J. M., G. Baloche, F. Barillet, and A. Legarra. 2014. Genomic evaluation validation test proposed by Interbull is necessary but not sufficient because it does not check the correct genetic trend. Page 50 in Proc. 39th ICAR Annual Conference. ICAR, Arthur van Schendelstraat.

Baloche, G., A. Legarra, G. Sallé, H. Larroque, J. M. Astruc, C. Robert-Granié, and F. Barillet. 2014. Assessment of accuracy of genomic prediction for French Lacaune dairy sheep. *J. Dairy Sci.* 97:1107–1116. <https://doi.org/10.3168/jds.2013-7135>.

Barillet, F. 2007. Genetic improvement for dairy production in sheep and goats. *Small Rumin. Res.* 70:60–75. <https://doi.org/10.1016/j.smallrumres.2007.01.004>.

Barillet, F., D. Boichard, A. Barbat, J. M. Astruc, and B. Bonaiti. 1992. Use of an animal model for genetic evaluation of the Lacaune dairy sheep. *Livest. Prod. Sci.* 31:287–299. [https://doi.org/10.1016/0301-6226\(92\)90023-W](https://doi.org/10.1016/0301-6226(92)90023-W).

Barillet, F., R. Rupp, S. Mignon-Graudeau, J.-M. Astruc, and M. Jacquelin. 2001. Genetic analysis for mastitis resistance and milk somatic cell score in French Lacaune dairy sheep. *Genet. Sel. Evol.* 33:397–415. <https://doi.org/10.1186/1297-9686-33-4-397>.

Bermann, M., A. Legarra, M. K. Hollifield, Y. Masuda, D. Lourenco, and I. Misztal. 2021. Validation of single-step GBLUP genomic predictions from threshold models using the linear regression method: An application in chicken mortality. *J. Anim. Breed. Genet.* 138:4–13. <https://doi.org/10.1111/jbg.12507>.

Cesarani, A., Y. Masuda, S. Tsuruta, E. L. Nicolazzi, P. M. VanRaden, D. Lourenco, and I. Misztal. 2021. Genomic predictions for yield traits in US Holsteins with unknown parent groups. *J. Dairy Sci.* 104:5843–5853. <https://doi.org/10.3168/jds.2020-19789>.

Duchemin, S. I., C. Colombani, A. Legarra, G. Baloche, H. Larroque, J.-M. Astruc, F. Barillet, C. Robert-Granié, and E. Manfredi. 2012. Genomic selection in the French Lacaune dairy sheep breed. *J. Dairy Sci.* 95:2723–2733. <https://doi.org/10.3168/jds.2011-4980>.

Ducrocq, V. 1994. Multiple trait prediction: principles and problems. Pages 455–462 in Proceedings of the World Congress on Genetics applied to Livestock Production. University of Guelph.

Fernando, R. L., C. Stricker, and R. Elston. 1993. An efficient algorithm to compute the posterior genotypic distribution for every member of a pedigree without loops. *Theor. Appl. Genet.* 87:89–93. <https://doi.org/10.1007/BF00223750>.

Garcia-Baccino, C. A., A. Legarra, O. F. Christensen, I. Misztal, I. Pocrnic, Z. G. Vitezica, and R. J. Cantet. 2017. Metafounders are related to F_{st} fixation indices and reduce bias in single-step genomic evaluations. *Genet. Sel. Evol.* 49:34. <https://doi.org/10.1186/s12711-017-0309-2>.

Kerr, R., and B. Kinghorn. 1996. An efficient algorithm for segregation analysis in large populations. *J. Anim. Breed. Genet.* 113:457–469. <https://doi.org/10.1111/j.1439-0388.1996.tb00636.x>.

Kudinov, A. A., E. A. Mäntysaari, G. P. Aamand, P. Uimari, and I. Strandén. 2020. Metafounder approach for single-step genomic evaluations of Red Dairy cattle. *J. Dairy Sci.* 103:6299–6310. <https://doi.org/10.3168/jds.2019-17483>.

Legarra, A., O. F. Christensen, Z. G. Vitezica, I. Aguilar, and I. Misztal. 2015. Ancestral relationships using metafounders: Finite ancestral populations and across population relationships. *Genetics* 200:455–468. <https://doi.org/10.1534/genetics.115.177014>.

Legarra, A., and A. Reverter. 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. <https://doi.org/10.1186/s12711-018-0426-6>.

Lourenco, D. A., I. Misztal, S. Tsuruta, I. Aguilar, T. Lawlor, S. Forni, and J. Weller. 2014. Are evaluations on young genotyped animals benefiting from the past generations? *J. Dairy Sci.* 97:3930–3942. <https://doi.org/10.3168/jds.2013-7769>.

Macedo, F. L., O. F. Christensen, J.-M. Astruc, I. Aguilar, Y. Masuda, and A. Legarra. 2020. Bias and accuracy of dairy sheep evaluations using BLUP and SSGBLUP with metafounders and unknown parent groups. *Genet. Sel. Evol.* 52:47. <https://doi.org/10.1186/s12711-020-00567-1>.

Marie-Etancelin, C., J. M. Astruc, D. Porte, H. Larroque, and C. Robert-Granié. 2005. Multiple-trait genetic parameters and genetic evaluation of udder-type traits in Lacaune dairy ewes. *Livest. Prod. Sci.* 97:211–218. <https://doi.org/10.1016/j.livprodsci.2005.04.005>.

Meuwissen, T. H. E., G. De Jong, and B. Engel. 1996. Joint estimation of breeding values and heterogeneous variances of large data files. *J. Dairy Sci.* 79:310–316.

Meuwissen, T., and M. Goddard. 2010. The use of family relationships and linkage disequilibrium to impute phase and missing genotypes in up to whole-genome sequence density genotypic data. *Genetics* 185:1441–1449. <https://doi.org/10.1534/genetics.110.113936>.

Misztal, I., Z.-G. Vitezica, A. Legarra, I. Aguilar, and A. Swan. 2013. Unknown-parent groups in single-step genomic evaluation. *J.*

- Anim. Breed. Genet. 130:252–258. <https://doi.org/10.1111/jbg.12025>.
- Oget, C., M. Teissier, J.-M. Astruc, G. Tosser-Klopp, and R. Rupp. 2019. Alternative methods improve the accuracy of genomic prediction using information from a causal point mutation in a dairy sheep model. *BMC Genomics* 20:719. <https://doi.org/10.1186/s12864-019-6068-4>.
- Powell, R. L., and G. R. Wiggans. 1994. Impact of changes in U.S. evaluations on conversions and comparisons. *Interbull Bull.* 10.
- Quaas, R. L. 1988. Additive genetic model with groups and relationships. *J. Dairy Sci.* 71:1338–1345. [https://doi.org/10.3168/jds.S0022-0302\(88\)79691-5](https://doi.org/10.3168/jds.S0022-0302(88)79691-5).
- Rodríguez-Ramilo, S. T., J. M. Elsen, and A. Legarra. 2019. Inbreeding and effective population size in French dairy sheep: Comparison between genomic and pedigree estimates. *J. Dairy Sci.* 102:4227–4237. <https://doi.org/10.3168/jds.2018-15405>.
- Sargolzaei, M., J. Chesnais, and F.S. Schenkel. 2012. Assessing the bias in top GPA bulls. https://www.edn.ca/Articles/GEBOCT2012/Jacques_Bias_in_Top_Bulls.pdf.
- Sorensen, D. A., and B. W. Kennedy. 1983. The use of the relationship matrix to account for genetic drift variance in the analysis of genetic experiments. *Theor. Appl. Genet.* 66:217–220.
- Spelman, R. J., J. Arias, M. D. Keehan, V. Obolonkin, A. M. Winkelman, D. L. Johnson, and B. L. Harris. 2010. Application of genomic selection in the New Zealand dairy cattle industry. Proceedings of the 9th World Congress on Genetics Applied to Livestock Production.
- Tsuruta, S., I. Misztal, and T. J. Lawlor. 2004. Genetic correlations among production, body size, udder, and productive life traits over time in Holsteins. *J. Dairy Sci.* 87:1457–1468. [https://doi.org/10.3168/jds.S0022-0302\(04\)73297-X](https://doi.org/10.3168/jds.S0022-0302(04)73297-X).
- Tyrisevä, A.-M., E. A. Mäntysaari, J. Jakobsen, G. P. Aamand, J. Dürr, W. F. Fikse, and M. H. Lidauer. 2018. Detection of evaluation bias caused by genomic preselection. *J. Dairy Sci.* 101:3155–3163. <https://doi.org/10.3168/jds.2017-13527>.

ORCID

F. L. Macedo  <https://orcid.org/0000-0002-1949-9214>

A. Legarra  <https://orcid.org/0000-0001-8893-7620>