J. Anim. Breed. Genet. ISSN 0931-2668

ORIGINAL ARTICLE

A robust method for simultaneous estimation of single gene and polygenic effects in dairy cows using externally estimated breeding values as prior information

B. Buske[1], M. Szydlowski[1,2] & N. Gengler[1]

[1] Université de Liège, Gembloux Agro-Bio Tech (GxABT), Gembloux, Belgium
[2] Poznan University of Life Sciences, Poznan, Poland

Keywords

Bayesian approach; dairy cows; data-splitting technique; externally estimated breeding values; myostatin gene; test-day model.

Correspondence

B. Buske, University of Liège, Gembloux
Agro-Bio Tech (GxABT), Passage des Déportés 2,
5030 Gembloux Belgium.
Tel: +32(0) 816 22358; Fax: +32(0) 816 22358;
E-mail: bernd.buske@ulg.ac.be

Received: 3 July 2009;
accepted: 20 December 2009

Summary

The aim of this study was to develop a robust method to estimate single gene and random polygenic animal effects simultaneously in a small field dataset with limited pedigree information. The new method was based on a Bayesian approach using additional prior information on the distribution of externally estimated breeding values. The field dataset consisted of 40 269 test-day records for milk performance traits for 1455 genotyped dairy cows for the 11 bp-deletion in the coding sequence of the myostatin gene. For all traits, estimated additive effects of the favoured wild-type allele ('+' allele) were smaller when applying the new method in comparison with the application of a conventional mixed inheritance test-day model. Dominance effects of the myostatin gene showed the same behaviour but were generally lower than additive effects. Robustness of methods was tested using a data-splitting technique, based on the correlation of estimated breeding values from two samples, with one-half of the data eliminated randomly from the first sample and the remaining data eliminated from the second sample. Results for 100 replicates showed that the correlation between split datasets when prior information included was higher than the conventional method. The new method led to more robust estimations for genetic effects and therefore has potential for use when only a small number of genotyped animals with field data and limited pedigree information are available.

Introduction

There is considerable interest in the use of molecular information when estimating breeding values for livestock. This is because knowledge of major single gene effects for quantitative traits (e.g. milk performance traits) and the subsequent selection of animals with desirable genotypes can accelerate breeding progress and can therefore lead to large gain in profits. In principle, the total breeding value for many quantitative traits of any animal can be divided into one or more major single gene effects and a random polygenic component, the latter resulting from a finite number of remaining loci (Fernando *et al.* 1994). Heretofore, candidate gene effects for numerous traits in many kinds of agricultural animals have been estimated. However, evaluation of breeding strategies showed that the use of such genes for marker-assisted selection (MAS) remains difficult and is only performed for a few genes (Hu *et al.* 2009). One reason may be the absence of an appropriate statistical evaluation method for the simultaneous estimation of single gene and random polygenic animal effects, particularly, when field data are used, which include only a

doi:10.1111/j.1439-0388.2010.00854.x

small number of genotyped animals. Moreover, with few genotyped animals, data are limited and often unconnected resulting in less precisely estimated random polygenic effects, which would diminish the ability to reliably estimate single gene effects. Also, when genotyped animals are not closely related or pedigree information is limited, the polygenic effect may be poorly estimated or not included at all, resulting in an overestimation of single gene effects. To overcome these problems, externally estimated breeding values for genotyped animals could be used as prior information because their estimation is based on a large number of non-genotyped relatives and therefore more reliable. Recently, Legarra *et al.* (2007) presented a formalization of the Bayesian method that weights prior estimates, based on external breeding values, relative to information supplied by the internal dataset to evaluate genetic merit of growth traits in beef cattle. They concluded that this method is suitable for populations with limited and unconnected data. Estimating the myostatin gene effect of the Dual Purpose Belgian Blue breed (DP-BBB) is similar. Relatively few cows within the total population are genotyped, and pedigree information is often incomplete. For this study, the myostatin gene was chosen as an example because of availability of data and the use of the knowledge of the 11 bp-deleted allele ('mh' allele) and the wild-type allele ('+' allele) in the DP-BBB cows in the Walloon Region of Belgium.

The aim of this study was to develop a robust method to simultaneously estimate single gene and random polygenic animal effects in a small genotyped population with limited pedigree information and a small field dataset. The new method is based on a conventional mixed inheritance test-day model using externally estimated breeding values and their distribution as prior information. Results were compared to the same model without using prior information. Robustness of the estimation of single gene and random polygenic effects was tested for both methods by applying a data-splitting technique.

Materials and methods

A total of 1455 genotyped DP-BBB cows with 40 269 test-day (TD) records serving as the 'internal dataset' were available. All cows were genotyped for the 11 bp-deletion in the coding sequence of the myostatin gene using a method adapted from Fahrenkrug *et al.* (1999). Genotype and allele frequencies were 0.181 (+/+), 0.371 (mh/+) and 0.447 (mh/mh) as well as 0.37 (+) and 0.63 (mh), respectively. Genotype frequencies departed slightly from the expected frequencies under Hardy–Weinberg equilibrium (0.137, 0.466 and 0.397), probably because matings were not random.

Number of lactations for cows varied between 1 and 13 and number of TD records per lactation varied between 1 and 22 and were sampled between 1991 and 2008. TD records within the first 5 days after calving were excluded from the dataset. Cows came from 72 herds (average of 21 cows per herd). Seventy-seven percentage of the cows were progeny of 132 known sires and the remainder had unknown sires. A moderate deviation of genotype frequencies for cows with unknown sires (0.263 for +/+, 0.457 for mh/+, 0.280 for mh/mh) was observed. These records were retained so that the sample closely reflected the current population.

For simplicity, the following single trait mixed inheritance test-day model was used, which is the basis for both the conventional and the new method:

$$\mathbf{y} = \mu + \mathbf{X}\beta + \mathbf{Hh} + \mathbf{Wi} + \mathbf{Zp} + \mathbf{Zu} + \mathbf{ZQg} + \mathbf{e}$$

where **y** is a vector of TD records representing the phenotype of the animal, $\boldsymbol{\mu}$ is the overall mean, $\boldsymbol{\beta}$ is a vector of fixed effects, **h** is a vector for random herd × test-day effect, **i** for random intralactation effect, **p** for random permanent environment effect and **e** represents the residual. The vector **u** stands for the random polygenic animal effect, and **g** represents the myostatin genotype effect. Genotype effect was considered fixed including an additive effect (**a**) defined as the estimated value for one copy of the '+' allele and a dominance effect (**d**) defined as the estimated value for the deviation of the heterozygous genotype from the mean of both homozygous genotypes. The incidence matrices **X, H**, **W** and **Z** link the records to the fixed effects, herd × test-day, animal × lactation number and animals, respectively, whereas **Q** is a matrix linking animals to their myostatin genotype. The equations for the estimates of $\boldsymbol{\beta}$, **u** and **g**, using the conventional method, are:

$$\begin{bmatrix} \mathbf{X'R^{-1}X} & \vdots & \mathbf{X'R^{-1}Z} & \mathbf{X'R^{-1}ZQ} \\ \cdots & \vdots & \cdots & \cdots \\ \mathbf{Z'R^{-1}X} & \vdots & \mathbf{Z'R^{-1}Z + G^{-1}} & \mathbf{Z'R^{-1}ZQ} \\ \mathbf{Q'Z'R^{-1}X} & \vdots & \mathbf{Q'Z'R^{-1}Z} & \mathbf{Q'Z'R^{-1}ZQ} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \cdot\cdot \\ \hat{\mathbf{u}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'R^{-1}y} \\ \cdots \\ \mathbf{Z'R^{-1}y} \\ \mathbf{Q'Z'R^{-1}y} \end{bmatrix} \quad (1)$$

G is the additive genetic (co)variance matrix proportional to additive relationship between animals and **R** is the residual variance matrix. The following five traits were considered: milk, fat and protein yield (kg) per lactation period (comprising of 305 days) and fat and protein content (%). Apart from the genotype effect, further fixed effects were herd (72 levels), the combination between lactation class (3 levels) and lactation stage (25 levels), the season in month (11 levels), the year of sampling (10 levels) and the age of cows in years at the test-day (7 levels). Concerning lactation class, third and later lactation numbers were combined. For the season, July and August were also combined because of low number of TD records in August. Age of cow at test-day classes was defined yearly; however, cows older than 6 years were combined to one class, whereas cows up to 2.5 years and cows between 2.5 and 3 years were assigned to an extra class. The pedigree consisted of 3511 animals including genotyped cows and their non-genotyped relatives and was extracted from the complete pedigree comprising over 956 000 animals, which is permanently updated and used for the official Walloon genetic evaluations (Croquet *et al.* 2006). Variance components for all random effects were assumed to be uncorrelated and were estimated with the restricted maximum likelihood (REML) procedure using only the internal dataset. Calculations for fixed and random effects were performed simultaneously using a sparse matrix-solving procedure (Misztal *et al.* 2002) because of the possibility of obtaining prediction error variances (**PEV**) directly from the inverse of the matrix of coefficients.

The new method was the same as described above, except that externally estimated breeding values (so called 'priors') were introduced by modifying the mixed model equations as proposed by Legarra *et al.* (2007). These externally estimated breeding values with their corresponding reliabilities were obtained via the routine evaluation system in the Walloon Region of Belgium and were calculated by a multi-lactation, multi-trait random regression test-day model described in Croquet *et al.* (2006). As the externally estimated breeding values and their reliabilities were not directly comparable to those obtained only from internal data applying Equation (1), the former were precorrected before they were used as prior information for the new method (for calculation details see the Appendix). Besides the inclusion of precorrected priors $\tilde{\boldsymbol{\mu}}_{0E}$ on the right hand side of the mixed model equations, the matrix $\mathbf{G}^{-1}$ from the conventional method was replaced by $\mathbf{G}^{*-1}$. The matrix $\mathbf{G}^{*-1}$ is the full-ranked additive genetic relationship matrix obtained by the internal dataset as before, but modified taking into account the distribution of precorrected priors $\tilde{\mu}_{0E}$. Thus, the diagonal elements $diag\{\mathbf{G}^{*-1}\}$ are the additive genetic variances from internal data $diag\{\mathbf{G}^{-1}\}$ plus **PEV**'s from external data $diag\{\mathbf{D}^{-1}\}$ minus the additive genetic variance obtained only by external proofs $diag\{\mathbf{G}_E^{-1}\}$ and are calculated as follows: $diag\{\mathbf{G}^{*-1}\} = diag\{\mathbf{G}^{-1}\} + diag\{\mathbf{D}^{-1}\} - diag\{\mathbf{G}_E^{-1}\}$ in which the diagonal matrix $\mathbf{D}^{-1}$ represents the **PEV**'s for genotyped cows obtained by the inverse of the diagonal elements of the variance matrix $\mathbf{Z}_E{}'\mathbf{R}^{-1}\mathbf{Z}_E + \mathbf{G}_E^{-1}$ from external evaluations also including non-genotyped animals. The final equations for the estimates of $\boldsymbol{\beta}$, $\mathbf{u}$ and $\mathbf{g}$ using the adapted method are:

$$\begin{bmatrix} \mathbf{X'R^{-1}X} & \vdots & \mathbf{X'R^{-1}Z} & \mathbf{X'R^{-1}Z}Q \\ \cdots & \vdots & \cdots & \cdots \\ \mathbf{Z'R^{-1}X} & \vdots & \mathbf{Z'R^{-1}Z + G^{*-1}} & \mathbf{Z'R^{-1}ZQ} \\ \mathbf{Q'Z'R^{-1}X} & \vdots & \mathbf{Q'Z'R^{-1}Z} & \mathbf{Q'Z'R^{-1}ZQ} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \cdot\cdot \\ \hat{\mathbf{u}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'R^{-1}y} \\ \cdots \\ \mathbf{Z'R^{-1}y + G^{*-1}}\tilde{\boldsymbol{\mu}}_{\mathbf{0E}} \\ \mathbf{Q'Z'R^{-1}y} \end{bmatrix} \quad (2)$$

(Correction added after online publication 26 February 2010: Equation (2) was realigned.)

The authors are aware that solutions for the second method give only an approximation for polygenic effects and therefore they could be slightly biased for two reasons. First, the term $\mathbf{Z}_E{}'\mathbf{R}^{-1}\mathbf{Z}_E + \mathbf{G}_E^{-1}$ is generally not invertible and thus, the matrix $\mathbf{D}^{-1}$ has only the form of a diagonal structure, disregarding covariances between animals in the external evaluation. Second, the model assumption for external evaluations was based on an infinitesimal model, whereas our model assumes random polygenic effects as well as a major single gene effect. As the prior information was based on an infinitesimal model, the externally estimated breeding values still included the myostatin genotype effect. Therefore, Equation (2) needed to be solved iteratively so that the adjustment to the part $\mathbf{G}^{*-1}\,\tilde{\boldsymbol{\mu}}_{0E}$ of the right hand side was modified to be free of the myostatin genotype effect. After each round *n*, precorrected priors $\hat{\boldsymbol{\mu}}_{0E}$were corrected for the new internally estimated additive part $\hat{\mathbf{a}}_I$ of the myostatin genotype effect depending on the genotype of each cow as follows:

$$\tilde{\boldsymbol{\mu}}_{\mathbf{0}\boldsymbol{E}} = \hat{\boldsymbol{\mu}}_{\mathbf{0}\boldsymbol{E}} - \mathbf{Q}\hat{\mathbf{a}}_{\boldsymbol{I}} \quad \textbf{with} \quad \hat{\mathbf{a}}_{\boldsymbol{I}} = \frac{\hat{\mathbf{a}}_{\boldsymbol{I}_{n-1}} + \hat{\mathbf{a}}_{\boldsymbol{I}_n}}{\mathbf{2}}$$

The corrected priors were then used for the next round until convergence (i.e. priors $\tilde{\boldsymbol{\mu}}_{\mathbf{0E}}$ and additive effect $\hat{\mathbf{a}}_I$ remained stable) was reached. Pre-investigations showed that convergence for all traits was reached using the method of successive under-relaxation. In this study, the additive myostatin genotype effect $\hat{\mathbf{a}}_I$ from the current round *n* and from the previous round *n-1* was averaged to correct the priors used for the next round.

Mean bias (MB), mean square prediction error (MSPE), correlation between estimated and observed yields (r) and the coefficient of model determination (CD) using the full dataset were calculated to compare accuracy and precision of the conventional and the new method (Tedeschi 2006). Also, new estimated total breeding values were compared between methods by means of correlation between new estimated and externally estimated breeding values, whereas the total breeding value of each cow was defined as the sum of the random polygenic breeding value and the additive part of the myostatin genotype effect. This was performed because it was assumed that externally estimated breeding values are more reliable because of the inclusion of valuable information of many related non-genotyped animals. Thus, a high correlation between new estimated total breeding values and externally estimated breeding values should indicate a reliable estimate for the genetic part of milk performance traits.

Model stability for the prediction of polygenic and total breeding values was tested using a data-splitting technique as in Ramirez-Valverde *et al.* (2001). Generally, all genotyped cows with all their records were randomly assigned to two complementary subsets A and B so that cows in subset B were those not chosen for subset A. Thus, all records for a given cow were only present in one of the two subsets. This procedure was repeated 100 times. Cows were distributed between the subsets rather than TD records for two reasons. First, in real-life situations, it is more likely that one has few genotyped animals with many observations, compared to many genotyped animals with only one observation or with few observations but with 'gaps' between them. Second, the applied procedure leads to subsets, representing subpopulations with different genotype frequencies but with complete data within these subpopulations. Breeding values for each genotyped cow, even if not in a subset, were predicted from the remaining cows according to own performance and pedigree information. Pedigree was the same for each subset and contained the relationship of all 3511 animals as explained earlier. Estimated genotype effects for both subsets were calculated and compared with the results obtained by the full dataset. Also, correlations for both polygenic and total breeding values between both subsets were calculated for the conventional and the newly adapted method and reported correlations were the average of the 100 replicates.

Results and discussion

Pre-examination of the complete field dataset showed that it was impossible to generate a standardized subset in which non-genetic effects could be excluded reliably. For example, most herds were not fully informative (e.g. only some small herds contained cows with all three genotypes), or number of genotyped cows per sire was different. Time-frame of TD records was long, and cows differed in lactation numbers and TD records per lactation, or even changed herds in their productive life. In this case, a reduction in the number of cows or TD records at the expense of information loss was not reasonable, because such a procedure did not lead to an improvement of standardization for non-genetic parameters. Thus, all genotyped cows with their TD records were retained.

Single gene additive and dominance effects

Results for estimated additive gene effects for the favoured '+' allele were 425.41 kg per lactation for milk yield, 0.059% for fat content, and −0.007% for protein content for the conventional method versus 120.26 kg per lactation for milk yield, 0.020% for fat content, and −0.001% for protein content for the new method (Table 1). The low values for protein content were expected because the phenotypic correlation between milk and protein yield was very high (>0.96). Our results showed that estimated additive effects differed strongly between applied methods and were very high for milk, protein and fat yield when the conventional method was used. To our knowledge, there is no study that investigated the influence of the myostatin gene on milk performance traits. Liefers *et al.* (2002) reported comparable results for an intronic polymorphism in the leptin gene for milk yield in dairy cows, using a similar conventional model. However, when dry matter intake was considered as a covariate in their statistical model, a significant reduction in milk yield was observed implying that feeding effects were not negligible for milk production. Because the current study used field data, it is possible that there are feeding effects confounding the additive effects. Although a herd × test-day effect was included in our statistical model, which in a broader sense represents nutrition and management effects, it might be possible that +/+ cows showed a different dry matter intake behaviour in comparison with mh/+ or mh/mh cows, which was not measured.

Dominance effects were generally lower than additive effects, but considering milk, protein and fat yield, dominance effects were not negligible in either method, whereas fat and protein content were rather uninfluenced by genetic dominance (Table 1). How-

Table 1 Additive and dominance effects[a] of the myostatin gene for the conventional mixed inheritance test-day model (BLUP) and the new method using externally estimated breeding values as prior information (Bayesian) for milk production traits applying the full dataset

Trait	Milk yield kg		Fat yield kg		Fat content %		Protein yield kg		Protein content %	
Method	BLUP	Bayesian	BLUP	Bayesian	BLUP	Bayesian	BLUP	Bayesian	BLUP	Bayesian
Additive effect[b]	425.41	120.26	18.953	5.521	0.059	0.020	13.532	3.960	−0.007	−0.001
Dominance effect	147.96	89.15	5.524	2.715	0.010	−0.002	4.948	3.172	0.007	0.007

[a] Per lactation period comprising 305 days.
[b] Additive effect for one copy of the '+' allele.

ever, there were remarkable differences between the two applied methods. For example, although results for fat content were generally small for both methods, the application of the conventional method led to an extensively higher estimation compared to the new method, making it difficult to decide whether dominance effects play a role for this trait.

A comparison between the two applied methods resulted in basic agreement concerning the magnitude of myostatin genotype effects among all traits, although the results of genotype effects were generally lower for the new method. This was not surprising because the inclusion of corrected externally estimated breeding values as priors influenced the distribution between the random polygenic and the additive single gene effect by a considerable amount. Therefore, it could be assumed that the correlation between new and externally estimated breeding values should be higher for the new method than for the conventional method. Our results clearly confirmed this expectation and showed further that using the new method, new estimated polygenic breeding values and iteratively corrected priors were even more highly correlated than newly estimated total breeding values and non-corrected priors (Table 2).

Model adequacy applying the full dataset

The comparison of mean bias (MB), mean square prediction error (MSPE), correlation of estimated and observed yields (r) and coefficient of model determination (CD) as indicators for model adequacy showed no remarkable differences between both methods (Table 3). This was also observed when the myostatin effect was removed from the model (data not shown). Both methods estimated the error solutions in the same range although there was a slight tendency in favour of the new method for MB. Thus, our test of model adequacy indicated that only the intragenetic distribution between single gene and random polygenic effects was influenced by the different methods, but not the estimation of residuals or expected values.

Robustness of methods by data splitting

Results of average correlation for the prediction of polygenic and total breeding values including 100 replicates showed large differences between the two methods (Table 4). For the conventional method, correlations ranged from 0.133 (protein yield) to 0.367 (protein content) when only polygenic effects were considered. Such low correlations were unexpected but might be explained by the use of field data with limited pedigree information. Because pedigree information was poor because of the lack of parent and particularly of valuable sire information for several genotyped cows, the estimation of breeding values in one subset depended strongly on own performance and for cows being removed in this subset, on the performance of their remaining female relatives with sometimes limited records. For example, a removed cow, which has itself many records of poor performance for a given trait, could be assigned a high breeding value because of one remaining related female (e.g. a half sister with a common dam) with good, but few records for this

Table 2 Correlation between new and externally estimated breeding values (priors) for the conventional mixed inheritance test-day model (BLUP) and the new method using externally estimated breeding values as prior information (Bayesian) applying the full dataset

Breeding value	Method	Milk yield	Fat yield	Fat content	Protein yield	Protein content
Total	BLUP	0.762	0.766	0.747	0.745	0.737
Total	Bayesian	0.914	0.902	0.785	0.926	0.806
Polygenic[a]	Bayesian	0.966	0.959	0.812	0.986	0.805

[a] correlation between polygenic breeding values obtained by the new method (Bayesian) and iteratively corrected externally estimated polygenic breeding values.

Table 3 Mean bias (MB), mean square prediction error (MSPE), correlation between observed and estimated yields ($r_{y:\hat{y}}$) and coefficient of model determination (CD) for the conventional mixed inheritance test-day model (BLUP) and the new method using externally estimated breeding values as prior information (Bayesian) applying the full dataset

Trait	Milk yield		Fat yield		Fat content		Protein yield		Protein content	
Method	BLUP	Bayesian	BLUP	Bayesian	BLUP	Bayesian	BLUP	Bayesian	BLUP	Bayesian
MB	$1.4e^{-5}$	$-1.9e^{-8}$	$-6.2e^{-7}$	$6.2e^{-9}$	$-1.0e^{-6}$	$8.7e^{-9}$	$2.6e^{-7}$	$5.5e^{-9}$	$-9.9e^{-6}$	$-1.2e^{-8}$
MSPE	3.983	3.982	0.009	0.009	0.172	0.172	0.004	0.004	0.037	0.037
$r_{y:\hat{y}}$	0.944	0.944	0.897	0.897	0.734	0.734	0.930	0.930	0.871	0.871
CD	1.176	1.175	1.350	1.348	2.346	2.322	1.230	1.229	1.448	1.443

Table 4 Correlations[a] between split datasets for milk performance polygenic and total breeding value solutions of genotyped cows by the conventional mixed inheritance test-day model (BLUP) and the new method using externally estimated breeding values as prior information (Bayesian). Standard deviation (SD) of 100 replicates is given in parenthesis

Breeding	Milk yield		Fat yield		Fat content		Protein yield		Protein content	
value	BLUP	Bayesian	BLUP	Bayesian	BLUP	Bayesian	BLUP	Bayesian	BLUP	Bayesian
Polygenic	0.194	0.967	0.141	0.957	0.272	0.677	0.133	0.986	0.367	0.713
	(0.042)	(0.001)	(0.032)	(0.001)	(0.032)	(0.014)	(0.044)	(0.010)	(0.033)	(0.012)
Total	0.751	0.973	0.752	0.966	0.355	0.688	0.791	0.989	0.363	0.709
	(0.019)	(0.002)	(0.014)	(0.002)	(0.029)	(0.019)	(0.017)	(0.002)	(0.034)	(0.017)

[a]Correlation estimates are means from 100 replicates.

trait. When now the complementary subset is analysed, the breeding value for that cow excluded from the first subset will be low, leading to low correlations between estimated polygenic breeding values for the same animal. Therefore, estimation of breeding values was not stable using the conventional method, as it probably would if all sires were available, assuming a large number of phenotyped daughters per sire. Ramirez-Valverde *et al.* (2001) compared animal versus sire-maternal-grand-sire models for the estimation of breeding values for calving difficulties. They observed that provided that sires have few progeny, the correlations of breeding values between complementary data subsets were higher for the animal model in which the full pedigree information is considered. Their result is important for this study, as it implies a better prediction of breeding values when the pedigree is complete and should be even more important, when a small (genotyped) population is used. However, pedigree information was limited in this study, and it was assumed that the inclusion of externally estimated breeding values into the dataset would lead to more reliable predictions of polygenic and total breeding values. Our assumption was confirmed because the inclusion of precorrected externally estimated breeding values as priors led to moderate (e.g. 0.677 for fat content) to very high correlations (e.g. 0.986 for protein yield) between new estimated polygenic breeding values. Obviously, iteratively corrected externally estimated breeding values stabilized the new estimation of polygenic breeding values, particularly for cows with few own records, which also influenced the breeding value estimation of their relatives. Concerning total breeding values, correlations between both subsets became slightly higher using the new method, but much higher using the conventional method, except for protein content, for which no difference between correlation solutions for polygenic and total breeding values in either of the methods was observed. This was expected as the inclusion of major single gene effects generally stabilizes total breeding values depending on their magnitude. Because protein content was uninfluenced by the myostatin gene, correlations between subsets were similar for total and polygenic breeding values for both methods.

Robustness of the additive myostatin gene effect prediction was tested by comparing the estimates of the 100 pairs of complementary subsets. We assumed that the estimated effect from the subsets should reflect the estimated effect using the complete dataset as precisely as possible. For the conventional method, the additive effect was slightly overestimated for all traits in each subset (Table 5). By contrast, when applying the new method, slight overestimations in one subset were almost compensated by the complementary subset. This observation should be the case assuming that the estimated single gene effect applying the full dataset is the true effect. However, as the true single gene effect is unknown, these results should be interpreted with caution and show only a trend in favour of the new method.

Implications

As the simultaneous estimation of single gene and random polygenic effects is crucial when field data are used, which derive from a small, non-environmental standardized population with limited pedigree, the idea was to include externally estimated breeding values and their distribution as prior information into the statistical model. Results showed that the utilization of the new method led to more robust estimates in comparison with the conventional method. For further research, genotyped cows with a complete pedigree should be used to investigate if the large differences for estimated random polygenic and single gene effects between both methods remain at the same range. Another promising strategy might be to use genotyped bulls with a high number of evaluated daughters instead of genotyped cows. Such a strategy could save genotyping costs and computation time and could also serve as a verification of the current results. The current results show that the use of externally estimated breeding values as priors has the potential to estimate more robust genetic effects in a small genotyped population with limited field data and incomplete pedigree.

Acknowledgements

The authors gratefully acknowledge the financial support (project D31/5593) of the Ministry of Agriculture of the Walloon Region of Belgium. The first author and Nicolas Gengler, research associate at the National Fund for Scientific Research (Brussels, Belgium), acknowledge the financial support provided by the National Fund for Scientific Research through grants F.4552.05 and 2.4507.02 F (2) for processing facilities. The authors thank S. Loker, University of Guelph (Canada), for reviewing the manuscript.

Appendix

Calculation of reliabilities r^2 for the breeding values concerning the internal dataset was performed for each trait separately using the equation $r^2 = 1 - \frac{PEV}{\sigma_g^2}$ where **PEV** are the prediction error variances directly obtained as diagonal elements from the inverse matrix of coefficients (**C**-matrix) of the conventional mixed inheritance test-day model and σ_g^2 is the additive genetic variance obtained by the REML procedure.

Because the basis for the calculation of breeding values differed between internal and external data because of inclusion of additional phenotypic information for non-genotyped animals in the external dataset, externally estimated breeding values were precorrected before they were used as priors. This precorrection was performed by adding the mean of internally estimated breeding values to each externally estimated breeding value followed by subtracting the mean of externally estimated breeding values from each externally estimated breeding value. The corresponding equation for the vector of precorrected externally estimated breeding values $\hat{\mu}_{0E}$ is: $\hat{\mu}_{0E} = \frac{\hat{u}_E}{305} + \frac{1\left(1'\hat{u}_I - 1'\left(\frac{\hat{u}_E}{305}\right)\right)}{n}$ where n is the 1455 estimated breeding values for the genotyped cows and $\hat{\mathbf{u}}_I$ and $\hat{\mathbf{u}}_E$ are internally and externally

Table 5 Estimated additive myostatin genotype effects[a] for milk performance traits using split datasets by the conventional mixed inheritance test-day model (BLUP) and the new method using externally estimated breeding values as prior information (Bayesian). Standard deviation (SD) of 100 replicates is given in parenthesis

	Milk yield kg		Fat yield kg		Fat content %		Protein yield kg		Protein content %	
Method	BLUP	Bayesian	BLUP	Bayesian	BLUP	Bayesian	BLUP	Bayesian	BLUP	Bayesian
Subset A	432.77	121.27	19.190	5.563	0.062	0.019	13.768	4.023	−0.005	−0.002
	(36.631)	(14.533)	(1.436)	(0.567)	(0.0171)	(0.0070)	(1.098)	(0.437)	(0.0085)	(0.0034)
Difference[b]	7.361	1.013	0.249	0.045	0.003	−0.001	0.226	0.088	0.002	<0.001
Subset B	425.56	117.02	19.078	5.439	0.067	0.019	13.576	3.897	−0.003	−0.001
	(38.695)	(14.949)	(1.579)	(0.599)	(0.0173)	(0.0073)	(1.186)	(0.457)	(0.0090)	(0.0038)
Difference[b]	0.151	−3.238	0.138	−0.081	0.008	<0.001	0.034	−0.037	0.004	<0.001

[a] Results are means from 100 replicates and refer to lactation period comprising 305 days.

[b] Differences are means from 100 replicates between solutions from subsets and the solution from the complete dataset.

estimated breeding values, respectively. For fat and protein content, breeding values were not divided by 305. Externally estimated reliabilities r_E^2 were corrected by multiplying the provided externally estimated reliability for each cow by a factor α. This factor was calculated by the assumption that the coefficient of variances of estimated internal and external breeding values $\frac{\sigma^2_{\hat{u}_I}}{\sigma^2_{\hat{u}_E}}$ is proportional to the coefficient of the corresponding means of reliabilities $\frac{r_I^2}{r_E^2}$. Hence, α can be calculated with $\alpha = \frac{r_I^2}{r_E^2} \times \frac{\sigma^2_{\hat{u}_E}}{\sigma^2_{\hat{u}_I}}$.

References

Croquet C., Mayeres P., Gillon A., Vanderick S., Gengler N. (2006) Inbreeding depression for global and partial economic indexes, production, type, and functional traits. *J. Dairy Sci.*, **89**, 2257–2267.

Fahrenkrug S. C., Casas E., Keele J. W., Smith T. P. (1999) Technical Note: direct genotyping of the double-muscling locus (mh) in Piedmontese and Belgian Blue cattle by fluorescent PCR. *J. Anim. Sci.*, **77**, 2028–2030.

Fernando R.L., Stricker C., Elston R. C. (1994) The finite polygenic mixed model: an alternative formulation for the mixed model of inheritance. *Theor. Appl. Genet.*, **88**, 573–580.

Hu X., Gao Y., Feng C., Liu Q., Wang X., Du Z., Wang Q., Li N. (2009) Advanced technologies for genomic analysis in farm animals and its application for QTL mapping. *Genetica*, **136**, 371–386.

Legarra A., Bertrand J. K., Strabel T., Sapp R. L., Sánchez J. P., Misztal I. (2007) Multi-breed genetic evaluation in a Gelbvieh population. *J. Anim. Breed. Genet.*, **124**, 286–295.

Liefers S.C., te Pas M. F. W., Veerkamp R. F., van der Lende T. (2002) Associations between Leptin gene polymorphisms and production, live weight, energy balance, feed intake, and fertility in Holstein heifers. *J. Dairy Sci.*, **85**, 1633–1638.

Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T., Lee D.H. (2002) BLUPF90 and related programs (BGF90). Proc. 7thWorld Congr. Genet. Appl. Livest. Prod., Montpellier, France. CD-ROM Commun. 28:07.

Ramirez-Valverde R., Misztal I., Bertrand J. K. (2001) Comparison of threshold vs linear and animal vs sire models for predicting direct and maternal genetic effects on calving difficulty in beef cattle. *J. Anim. Sci.*, **79**, 333–338.

Tedeschi L. O. (2006) Assessment of the adequacy of mathematical models. *Agric. Syst.*, **89**, 225–247.