

# **Evaluation of two methods for computing approximate accuracies in random regression models for beef growth traits**

J. P. Sánchez, J. K. Bertrand, I. Misztal

## **ABSTRACT**

The objective of this study was to determine the suitability of two methods for computing approximate accuracies when modeling growth traits in beef cattle using random regression models. The first method (SMB) was originally intended to be used in multitrait models, thus to use it with random regression (RR) models requires the clustering of measurements into different traits. The second method (TM) was more general as it accounted for random coefficients other than zeros and ones, thus it could be used directly when fitting RR models. In order to investigate the performance of both approximate methods, their results were compared to the true accuracies using a balanced simulated data set. In general, TM was slightly closer than SMB to the true accuracies for both direct and maternal effects. Also, except for direct and maternal birth weight effect accuracies from TM for males, all approximate methods overestimated the actual accuracies. For direct effects, the average accuracies using SMB were somewhat larger (2.3 to 4.9%) than TM; however for maternal effects, larger variations in the differences between methods were observed. When both approximate methods were compared using a real beef cattle data set, the differences between direct effect accuracies between SMB and TM were on average, across traits and sexes, similar (3.8%) to the differences obtained in the simulation data set; however, the range in the differences across traits and sexes was greater. For maternal effects, SMB consistently produced accuracy values that were higher (average = 15.9%) across traits and sexes in the field data set. TM provided slightly better results compared to SMB; however, it does have higher memory and computation requirements, which could prevent its use in large data sets.

## **INTRODUCTION**

For the purpose of genetic evaluation of growth in beef cattle, there has been an interest to move from the traditional multitrait (MT) approach to random regression (RR) models that can better model the effect of age on the averages and/or variances of weight. Accuracies for the breeding values were computed in genetic evaluation programs to provide an assessment of risk. However, the computation of true accuracies was usually too expensive. Several approximations have been proposed. Tier and Meyer (TM; 2004) provided a method useful either for MT or random regression (RR) models; their approach considered the correlation between traits in the MT model or coefficients in the RR model during the computation of the equivalent number of records, which allowed for the approximation of prediction error covariances as well as variances. An alternative way to calculate accuracies for RR models was to use approaches that had been proposed for MT models, such as the method by Strabel et al. (SMB; 2001). This was possible if the observations were clustered in traits. The objective of this study was to compare the approximate accuracies from TM and SMB.

## MATERIAL AND METHODS

Animal Care and Use Committee approval was not obtained because the data used in this study were simulated or were obtained from an existing data base. Two data sets comprised of beef cattle growth traits were used. The first data set (D1) was a simulated population of 20,000 animals with 3 measurements distributed in four generations. To generate the first generation of measured animals, 5,000 dams were randomly mated to 500 sires. In subsequent generations, 10% of the males and 40% of the females were randomly selected and mated to produce the next generation of 5,000 animals. The average number of offspring and SD per sire and dam was  $16 \pm 6.25$  and  $3.28 \pm 2.4$ , respectively. Measures were assumed to occur at day 1 without variation, and at 205 and 365 days with SD of 20 d for both traits. The simulation model included a fixed contemporary group effect, which were the combination between generation and measure type, a linear spline with 3 knots to account for the effects of age was also considered, the model also include direct genetic effects, maternal genetic effects, permanent environmental effects, and maternal permanent environmental effects. These last four were random effects which were assumed to be three knot, linear spline functions of age. The residual variance was assumed to be heterogeneous, and it was assumed to be described by a linear spline function of age, with three knots, again at 1, 205 and 365 days. The second data set (D2) was comprised of field data provided by the American Gelbvieh Association. Table 1 shows basic statistics describing records in both D1 and D2.

The algorithm by Strabel et al. (SMB; 2001) is an extension of a single-trait algorithm (Misztal and Wiggans, 1988) to multitrait and maternal models. In this procedure during the computation of the direct effect, the maternal variances are added to the residual. Similarly, during the computation of the maternal effect, the direct variances are added to the residual. The direct and maternal accuracies are computed separately, indirectly assuming that these effects are uncorrelated. This method was applied to D1 and D2, considering them always via a MT model. The algorithm of Tier and Meyer (TM; 2004) was designed to compute accuracies either under RR or MT models. This method considers the (co)variance structure among traits (or coefficients in its RR version) during the computation of the equivalent number of records. Once these are known, they are modified for offspring contributions and parent contributions by looping twice through the pedigree. In this method, the maternal and direct genetic effects also are assumed to be uncorrelated. In the implementation of TM for this study, some aspects of the original algorithm were modified. Reduction of records due to repeated measurement for one animal in the same contemporary group in the RR model and discounting of records due to large sizes of sire half-sib families within contemporary group were not considered due to high memory requirements. Also in maternal models, better results were observed when the direct (maternal) genetic variance was moved to the residual variance instead of the permanent environmental (maternal environmental) variance as was suggested in the original study. For random effects in the RR model, two basis functions were used in the analysis, linear splines (RR-ls), as used during the simulation, and cubic Legendre polynomials (RR-lp). In both cases, the residual variance was assumed to be heterogeneous and described by a linear spline function with knots at 1, 205 and 365 days.

In both methods, the only fixed effect considered during the computation of the equivalent number of records was the contemporary group effect. D1 was also used to compute the exact accuracies after a generalized sparse inversion of the MME (GI) from MT, linear spline RR and Legendre Polynomial RR models. TM was applied either assuming a linear spline (TM-ls) or a Legendre polynomial (TM-lp) model, SMB was applied assuming a MT model, and GI was applied assuming the corresponding model used either in TM or SMB (MT, RR-ls or RR-lp) depending on which was applied in the approximated method.

## RESULTS AND DISCUSSION

The averages for the actual and computed accuracies calculated using the different approaches applied to D1 are shown in Table 2. In general, TM was slightly closer than SMB to the true accuracies for both direct and maternal effects. Also, except for direct and maternal birth weight effect accuracies from TM for males, all approximate methods overestimated the actual accuracies. Although not shown, the standard deviations of the computed and exact accuracies were very similar in all the methods and models. For direct effects, the average accuracies in SMB had values between 2.3 and 4.3% greater than the average accuracies for both TM models. For maternal effects, larger variations in the differences between methods were observed, ranging from TM being 4.7% greater than SMB at 365 days to SMB being 13.6% greater than TM at birth.

Table 3 shows the averages for the computed accuracies for direct and maternal effects by sex for animals with offspring and for animals without offspring from D2. In this case the average differences between direct effect accuracies between SMB and TM were, on average, across traits and sexes, similar (3.8%) to the differences obtained with D1; however, the range in the differences for D2, across traits and sexes, was greater than that of D1. For maternal effects, SMB consistently produced accuracy values that were higher (average = 15.9%) across traits and sexes. As expected, traits showing the largest differences between methods were those with variation at the measurement ages, i.e. weaning weight (205 days) and yearling weight (365 days). For animals without offspring, the differences between SMB and TM methods in direct effect accuracies were greater than for males with progeny, but generally less than for females with progeny. However for maternal effect accuracies, animals with no offspring showed the largest differences between the two approximate methods.

Table 4 shows times and memory requirements for the different approximation methods and models for the two data sets. As was expected given its higher dimension, TM-lp was the approach requiring the greatest resources. SMB clearly exhibited better performance in terms of CPU usage and memory requirement than TM-ls when both were used in the field data set.

Overall, TM was slightly closer than SMB to the true accuracies for both direct and maternal effects. Also, computed accuracies for direct effects from an algorithm assuming a MT model, defining traits at birth, weaning and yearling, were on average slightly greater than those from an algorithm specifically designed for RR models. This was observed both under the balanced conditions of a simulated example and with a real data set, where

differences between the methods were greater in the case of real data. It was also apparent that differences in performance between the approximate methods and between the approximate methods and the true accuracy values were greater for maternal effects. However, both approximate methods investigated in this study provided satisfactory accuracies for the most relevant animals, i.e. sires with progeny. Even though there were some differences between the methods, the high memory and computation requirements for the method assuming a RR model could prevent its usage in very large populations.

## **CONCLUSION**

In general for the most relevant animals (sires), both approximate accuracy methods performed satisfactorily. When the results from both approximate methods were compared to each other, the computed accuracies were close for direct effect accuracies; however, the differences for maternal effects accuracies were slightly greater. In spite of the differences between methods, a practical implication of the study was that using an algorithm assuming a MT model to compute accuracies after a RR evaluation could provide quite satisfactory results with less computational usage of resources.

## **LITERATURE CITED**

Misztal, I., and G. R. Wiggans. 1988. Approximation of prediction error variance in large-scale animal models. *J. Dairy Sci.* 71 (Suppl. 2): 27.

Tier, B., and K. Meyer. 2004. Approximating prediction error covariances among additive genetic effects within animals in multiple-trait and random regression models. *J. Anim. Breed. Genet.* 121:77-89.

Strabel, T., I. Misztal, and J. K. Bertrand. 2001. Approximation of reliabilities for multiple-trait model with maternal effects. *J. Anim. Sci.* 79:833-839.

Table 1.- Descriptive statistic for the studied traits in the three data sets considered

<b>D1<sup>a</sup></b>					
	<b>N</b>	<b>Mean</b>	<b>SD</b>	<b>Average Age, days</b>	<b>Age SD</b>
<b>BWT<sup>b</sup></b>	20,000	40	5	1	0
<b>WWT</b>	20,000	266	31	201	20
<b>YWT</b>	20,000	430	44	364	20
<b>D2<sup>a</sup></b>					
	<b>N</b>	<b>Mean</b>	<b>SD</b>	<b>Average Age, days</b>	<b>Age SD</b>
<b>BWT</b>	841,848	40	5	1	0
<b>WWT</b>	749,697	268	46	201	30
<b>YWT</b>	221,326	428	78	360	33

<sup>a</sup> D1= simulated dataset, D2 = Gelbvieh dataset

<sup>b</sup> BWT = Birth Weight, WWT = Weaning Weight, WWT = Weaning Weight

Table 2.- Means of accuracies obtained from the different methods and models using simulated data set (D1).

	Direct effects					
	MALES			FEMALES		
	1 day	205 days	365 days	1 day	205 days	365 days
<b>TM-ls<sup>a</sup></b>	0.534	0.461	0.440	0.398	0.343	0.326
<b>GI-ls</b>	0.535	0.454	0.438	0.387	0.326	0.317
<b>TM-lp</b>	0.536	0.467	0.439	0.400	0.347	0.326
<b>GI-lp</b>	0.537	0.460	0.439	0.388	0.330	0.318
<b>SMB</b>	0.554	0.478	0.454	0.415	0.356	0.338
<b>GI-mt</b>	0.534	0.457	0.440	0.386	0.327	0.319
	Maternal effects					
	MALES			FEMALES		
	1 day	205 days	365 days	1 day	205 days	365 days
<b>TM-ls</b>	0.178	0.140	0.132	0.212	0.166	0.155
<b>GI-ls</b>	0.179	0.121	0.106	0.198	0.134	0.117
<b>TM-lp</b>	0.177	0.140	0.129	0.210	0.166	0.152
<b>GI-lp</b>	0.178	0.123	0.103	0.196	0.136	0.114
<b>SMB</b>	0.201	0.155	0.141	0.225	0.164	0.148
<b>GI-mt</b>	0.179	0.123	0.107	0.197	0.136	0.118

<sup>a</sup> SMB = method of Strabel et al. (2001), TM = method of Tier and Meyer (2004), GI = generalized inverse, ls= linear spline model, lp = Legendre polynomial model, mt = multi-trait model.

Table 3.- Means of accuracies obtained from the different methods and models using Gelbvieh data set (D2).

<b>Direct effects</b>									
<b>ANIMALS WITH OFFSPRING</b>						<b>ANIMALS WITHOUT OFFSPRING</b>			
	<b>MALES</b>			<b>FEMALES</b>			<b>OFFSPRING</b>		
	<b>1<sup>a</sup></b>	<b>205</b>	<b>365</b>	<b>1</b>	<b>205</b>	<b>365</b>	<b>1</b>	<b>205</b>	<b>365</b>
<b>TM-ls<sup>b</sup></b>	0.483	0.429	0.374	0.34	0.284	0.250	0.512	0.432	0.366
<b>TM-lp</b>	0.483	0.434	0.365	0.341	0.298	0.243	0.514	0.439	0.355
<b>SMB</b>	0.488	0.441	0.377	0.353	0.313	0.260	0.527	0.460	0.382
<b>Maternal effects</b>									
<b>ANIMALS WITH OFFSPRING</b>						<b>ANIMALS WITHOUT OFFSPRING</b>			
	<b>MALES</b>			<b>FEMALES</b>			<b>OFFSPRING</b>		
	<b>1</b>	<b>205</b>	<b>365</b>	<b>1</b>	<b>205</b>	<b>365</b>	<b>1</b>	<b>205</b>	<b>365</b>
<b>TM-ls</b>	0.273	0.236	0.218	0.269	0.212	0.193	0.233	0.202	0.187
<b>TM-lp</b>	0.272	0.238	0.215	0.268	0.216	0.191	0.232	0.204	0.185
<b>SMB</b>	0.307	0.282	0.255	0.284	0.250	0.222	0.288	0.269	0.245

<sup>a</sup> 1, 205, and 365 days

<sup>b</sup> SMB = method of Strabel et al. (2001), TM = method of Tier and Meyer (2004), ls= linear spline model, lp = Legendre polynomial model.

Table 4. RAM and CPU time for different approximate methods and data sets.

	<i>Memory (Mb.)</i>		<b>CPU Time</b>	
	D1	D2	D1 (Sec.)	D2 (Min.)
<b>SMB<sup>a</sup></b>	12	326	6.9	6.8
<b>TM-ls</b>	24	887	3.8	8.6
<b>TM-lp</b>	36	1451	5.2	9.9

<sup>b</sup>D1 = simulated data set; D2 = Gelbvieh field data set.

<sup>a</sup>SMB = method of Strabel et al. (2001), TM = method of Tier and Meyer (2004), ls= linear spline model, lp = Legendre polynomial model.