

Reliability of breeding values for single- and multi-trait models of dam pig breed obtained using direct inversion and approximation methods

Emil Krupa*, Jiří Bauer, Eliška Žáková, Zuzana Krupová
Institute of Animal Science, Prague – Uhřetěves, Czech Republic

Article Details: Received: 2016-11-09 | Accepted: 2016-12-03 | Available online: 2017-12-31

<http://dx.doi.org/10.15414/afz.2017.20.04.90–94>

The objective of this study was to predict breeding values for single- and multi-trait animal models and compute their reliabilities using a direct inversion method (DIM), and single (ST-APM) and multi-trait approximate methods (MT-APM). Variance and covariance components of lean meat (LM) content, average daily gain (ADG) from birth until the end of the field test, and number of piglets born alive at first (NBA1) and second and subsequent parities (NBA2), were estimated for the analyses of Czech Large White pigs (390,734 records), using single- and four-trait animal models. The average reliabilities estimated by DIM for all considered animals were 0.514 ± 0.069 , 0.406 ± 0.070 , 0.050 ± 0.044 , and 0.321 ± 0.090 for LM, ADG, NBA1, and NBA2, respectively. Values of 0.576 ± 0.087 , 0.150 ± 0.078 , 0.228 ± 0.078 , and 0.323 ± 0.099 , were obtained for the ST-APM for LM, ADG, NBA1, and NBA2, respectively. The use of MT-APM slightly increases the reliability of breeding values by 4%, 6%, 14% and 8% for LM, ADG, NBA1, and NBA2, respectively. In addition, the dependence of the reliability values on the number of offspring of breeding boars is obtained; the reliability increases from 0.215 for less than 5 offspring to 0.989 for more than 400 offspring for the LM trait. Calculated Pearson's and Spearman's correlation coefficients between the employed methods were, in general, high, positive, and highly statistically significant. The multi-trait approximation method can be used for the calculation of reliabilities of breeding values as an alternative for direct inversion method that has computational limitations.

Keywords: pig, breeding value, reliability, direct inversion, approximation

1 Introduction

Numerical reliabilities of predicted individual breeding values are required for the optimisation of the selection of next generation parents. The reliabilities of predicted breeding values are important and they can be calculated by inverting the left-hand side of the BLUP system of equations (VanRaden (2008); Misztal and Perez-Enciso (1993)). However, this is not always feasible due to the large population size, which leads to enormous computational requirements, hence this approach is often suitable only for a restricted number of equations. Therefore, alternative methods should be used to obtain reasonable values of predicted breeding value reliabilities. Several approximation methods have been developed for non-genomic data. Misztal and Wiggans (1988) developed a method that is computationally simple and iterative, based on the effective number of records and sum of contributions to an animal from its parents. This iterative method is computationally easy. A non-iterative modification of this method has been proposed by VanRaden and Wiggans (1991). Wiggans et al. (1988) and Misztal et al. (1993) extended the approximation method of Misztal and Wiggans (1988) to repeatability

models. Strabel et al. (2001) extended the method to the multiple-trait models that include maternal effects; it has been applied to random regression models, as reported by Sánchez et al. (2008). In addition, an approximation method for a single-step genomic evaluation has been reported by Misztal et al. (2013). Tier and Meyer (2004) described a method for approximating prediction error variances and covariances among estimates of individual animals' genetic effects for multiple-trait and random regression models. The advantage of the approximation method is its computational simplicity. The computations of breeding value reliabilities of pigs in the Czech Republic are based on a direct inversion method (DIM). This approach reaches the computational limits for a large number of dam breed Czech Large White (CLW) pigs. Therefore, as soon as possible, novel methods should be developed, tested, and applied in routine genetic evaluations.

This study aims to predict breeding values for single- and multi-trait animal models, and compute and compare the values of their reliabilities using direct and approximate methods.

***Corresponding Author:** Emil Krupa, Institute of Animal Science, Přátelství 748, Prague-Uhřetěves, 10400, Czech Republic, e-mail: krupa.emil@vuzv.cz

2 Material and methods

2.1 Data

The analysis employed performance test data (production and reproduction traits) for breed CLWs from 1995 to 2015. The total number of records was 390,734. The considered traits were the lean meat (LM) content (in %) at the end of the performance, average daily gain (ADG) (in g/d) from birth until the end of the field test, number of piglets born alive at the first parity (NBA1), and number of piglets born alive at the second and subsequent parities (NBA2). The litter size data is retained for the analysis when the litter are purebred CLWs. The gestation length was ranged from 105 d to 125 d. The minimum sow age at first farrowing was 300 d. Parities greater than 10 were not considered. The number of born piglets was at least four. The farrowing interval was between 130 d and 300 d. More detailed description of CLW population structure is reported in Krupa et al. (2015).

2.2 Breeding value estimation

The statistical model in a matrix notation is:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Wz} + \mathbf{e}$$

where:

- y** – the vector of observations of the investigated traits
- X, Z** and **W** – are known incidence matrices for the fixed effects, i.e., the random additive genetic animal and remaining random effects
- b** – denotes the vector of the fixed effects
- a** – the vector of additive genetic animal effects
- z** – the vector of further random effects
- e** – the vector of residuals

The vector **b** contains fixed factors and covariates, while the vector **z** contains the random factors. Detailed description of the employed effects in the models is reported in Krupa and Wolf (2013). It is worth noting that the pedigree was traced back to 1985. Variance and covariance components were estimated by a restricted maximum likelihood (REML) and optimisation using a quasi-Newton algorithm with analytical gradients (Neumaier and Groeneveld (1998)), as implemented in the VCE 6.0 program (Groeneveld et al. (2008)). The breeding values were estimated using PEST 4.2 (Groeneveld et al. (1990)). The calculations were performed for four single-trait animal models and one four-trait animal model.

2.3 Reliability estimation

Two different methods for the estimation of breeding value reliability were used. DIM is based on computations of standard errors of breeding value predictions; it is described in detail in Mrode (2014). The general formula for the estimation of reliabilities (r^2) using DIM is:

$$r^2 = 1 - \frac{SEP^2}{\sigma_a^2}$$

where:

SEP – the standard error of the breeding value prediction:

$$SEP = \sqrt{d_i \sigma_e^2}$$

where:

d_i – the i^{th} diagonal element of the matrix \mathbf{C}^{22} , which is described in Mrode (2005) PEST 4.2 (Groeneveld et al., 1990) was used for computing the *SEP* for each pedigree animal

The employed approximation method was based on the procedures reported in Misztal et al. (1993), implemented in the SAS IML program

The method is based on approximating reliabilities for the entire population using the iterative approach. The four single-trait models and one multi-trait model were tested. The effective number for the animal i that accounts for the production (reproduction) data and its relationship to animal j were used for the single-trait models. The approximation of reliabilities was based on estimating diagonal elements of matrix **C** for the multi-trait model, computed as:

$$r^2 = 1 - \alpha C_{ii}$$

where:

α – the ratio between error variance and additive genetic variance

The same variance and covariance component values of the random effects were assumed for both direct inversion and approximation methods. The procedure CORR of statistical package SAS 9.3 was used for the calculation of Pearson's and Spearman's correlation coefficients.

3 Results and discussion

The employed variance and covariance components for the additive random animal effect are summarised in Table 1. We assumed the same component values for each run of the reliability computations. Strabel et al. (2001) compared the direct inversion, single-, and multi-trait approximation methods, and reported that the results for the comparison of the methods were strongly affected by the variance components used in the models.

Notably, not all the calculated reliabilities can be used. Some of them, obtained by DIM, have negative values. This could be caused by inbreeding that was ignored in the calculations of the inverse relationship matrix; negative reliabilities were excluded from the analysis.

Table 2 provides the basic statistical values for the calculated reliabilities of all analysed traits and for all methods used averaged for all animals studied. The single-trait approximation method slightly overestimated the reliabilities, compared to those obtained by direct inversion method for lean meat content by 0.062, and for number of piglets born alive after first parity by 0.092. The reliability for NBA1 was highly overestimated, by 0.178, when ST-APM was used. The average ADG reliabilities were higher when direct inversion method was used (by 0.256 and 0.250, compared with the ST-APM and MT-APM, respectively). Further, the breeding value reliabilities computed using the MT-APM were higher than those computed by ST-APM; they were increased by 4%, 6%, 14% and 8%, for LM, ADG, NBA1, and NBA2, respectively. This result is due to the extra information from other correlated traits in MT-APM.

A comparison of the employed methods for breeding boars, the most common breeding animals, is shown in Table 3. As expected, the mean reliabilities were higher

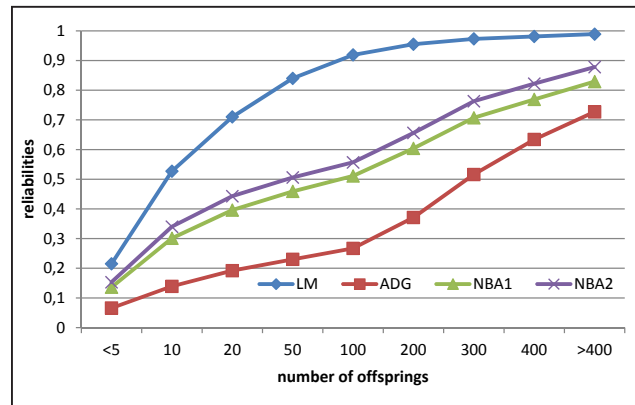


Figure 1 Dependence of the approximated reliabilities (by MT-APM) of breeding boars on the number of their progenies

than those for other animals. The differences of the results between the employed methods have the same tendencies as those for other animals; however, their values are lower (0.004, 0.091, and 0.020, between ST-APM and DIM for

Table 1 Variance (on diagonal) and covariance (off diagonal) components for random animals effect for all traits

Trait	LM	ADG	NBA1	NBA2
LM ¹	1.187	5.777	0.025	0.016
ADG ²	5.777	615.372	2.107	3.845
NBA1 ³	0.025	2.107	0.681	0.639
NBA2 ⁴	0.016	3.845	0.639	0.795

¹ – lean meat content; ² – average daily gain in field test; ³ – number of piglets born alive at first parity; ⁴ – number of piglets born alive at second and subsequent parity

Table 2 Basic statistics for the reliabilities of traits obtained by direct inversion method and approximation methods in single-trait and multi-trait models for all considered animals

Trait	DIM ¹				ST-APM ²				MT-APM ³
	mean	std.	min.	max.	mean	std.	min.	max.	%
LM	0.514	0.069	0.000	0.945	0.576	0.087	0.000	0.995	4
ADG	0.406	0.070	0.000	0.951	0.150	0.078	0.000	0.889	6
NBA1	0.050	0.044	0.000	0.641	0.228	0.078	0.000	0.911	14
NBA2	0.231	0.090	0.000	0.877	0.323	0.099	0.000	0.960	8

¹ – direct inversion method for the multi-trait model; ² – approximation method for the single-trait model; ³ – average increase in reliability for traits in the multi-trait model (%)

Table 3 Basic statistics for the reliabilities of traits obtained by direct inversion method and approximation methods in single-trait and multi-trait models for breeding boars

Trait	DIM ¹				ST-APM ²				MT-APM ³
	mean	std.	min.	max.	mean	std.	min.	max.	%
LM	0.603	0.288	0.000	0.945	0.607	0.352	0.000	0.995	1.9
ADG	0.496	0.291	0.000	0.951	0.168	0.185	0.000	0.889	5.7
NBA1	0.171	0.137	0.000	0.641	0.262	0.220	0.000	0.911	12.5
NBA2	0.351	0.205	0.000	0.877	0.371	0.258	0.000	0.960	5.1

¹ – direct iteration method for the multi-trait model; ² – approximation method for the single-trait models; ³ – average increase in reliability for the traits in the multi-trait model (%)

LM, NBA1, and NBA2, respectively). In contrast, for ADG the difference between reliabilities obtained by ST-APM and DIM is -0.311. For a detailed analysis of breeding boar reliabilities, we divided them into groups depending on the number of their tested progenies (<5, 6–10, 11–20, 21–50, 51–100, 101–200, 201–300, 301–400, and >400). One of

the sources of difference between analysed methods may be caused by fact, that approximate method not generate negative values. Figure 1 shows the relationship between the breeding boar reliabilities (computed using ST-APM) and number of their progenies. As expected, the reliabilities increase with the number of tested progenies.

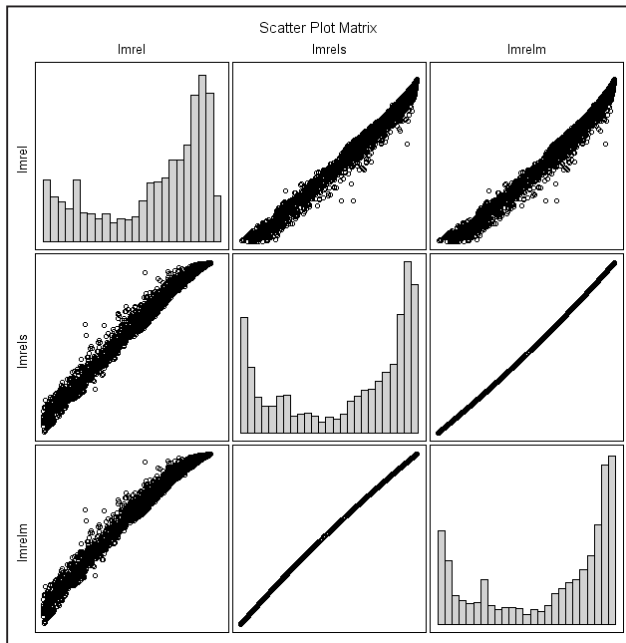


Figure 2 Trait distributions (histograms, on diagonal) and individual reliability pair-wise comparisons (scatter plots, off-diagonal) for the lean meat content

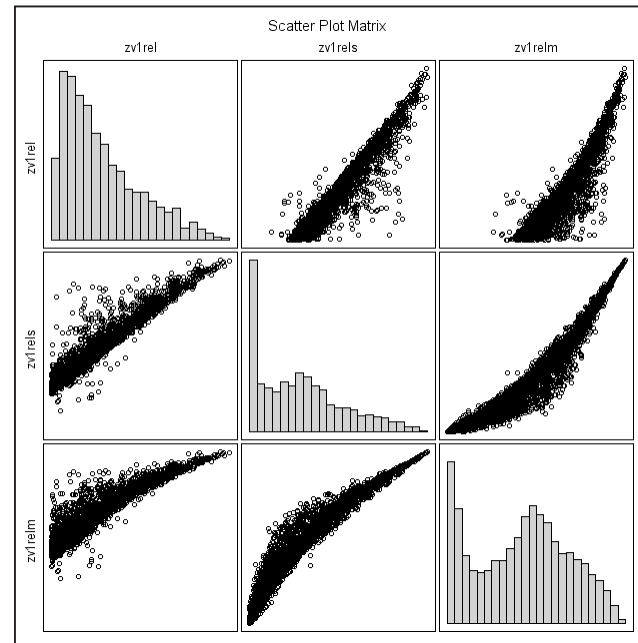


Figure 4 Trait distributions (histograms, on diagonal) and individual reliability pair-wise comparisons (scatter plots, off-diagonal) for number of piglets born alive at first parity

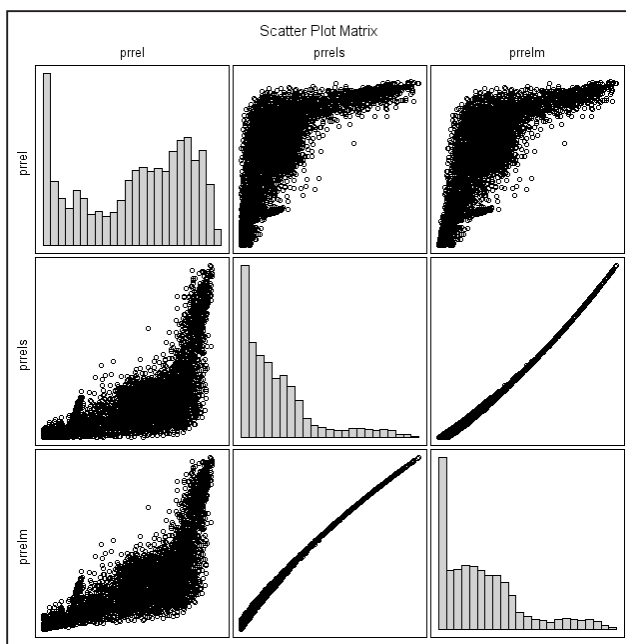


Figure 3 Trait distributions (histograms, on diagonal) and individual reliability pair-wise comparisons (scatter plots, off-diagonal) for average daily gain in the field test

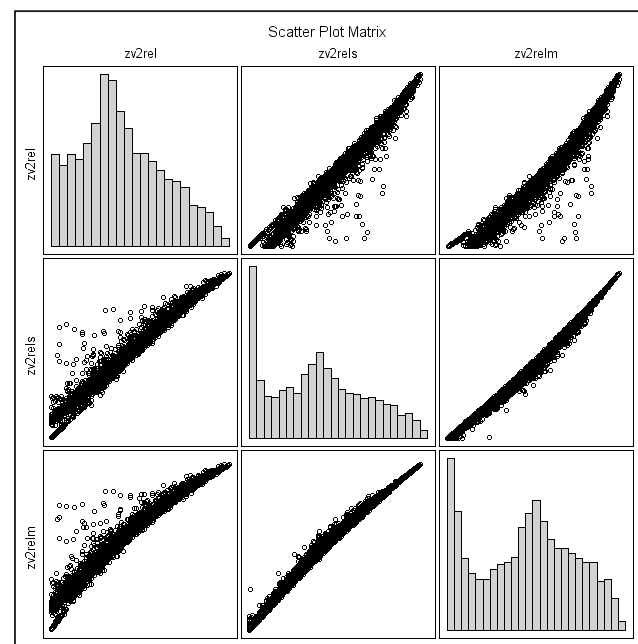


Figure 5 Trait distributions (histograms, on diagonal) and individual reliability pair-wise comparisons (scatter plots, off-diagonal) for number of piglets born alive at second and subsequent parity

Table 4 Spearman's correlation coefficients of breeding boars

		ST-APM ⁶	MT-APM ⁷
LM ¹	DIM ⁵	0.992	0.992
	ST-APM	–	1.000
ADG ²	DIM	0.830	0.854
	ST-APM	–	0.997
NBA1 ³	DIM	0.943	0.906
	ST-APM	–	0.959
NBA2 ⁴	DIM	0.981	0.976
	ST-APM	–	0.994

1 – lean meat content; 2 – average daily gain in field test; 3 – number of piglets born alive at first parity; 4 – number of piglets born alive at second and subsequent parity; 5 – direct iteration method for multi-trait model; 6 – approximation method for single trait models; 7 – approximation method for multi-trait model

Figures 2–4 show trait reliability distributions and comparisons of individual reliability (in scatter plots). In general, the plots show a low variability for traits LM, NBA1, and NBA2. Similar results were obtained by Strabel et al. (2001) for beef cattle reliabilities.

Table 4 summarises calculated Spearman's correlation coefficients between the analysed methods for appropriate traits. All the correlation coefficients are positive, large, and highly statistically significant. The highest correlations were obtained for LM content (1.000 between ST-APM and MT-APM). Nevertheless, ADG has the highest differences in mean reliabilities; the correlation coefficients maintained high values (0.830 and 0.854, between DIM and ST-APM, and DIM and MT-APM, respectively).

4 Conclusions

The reliabilities obtained by multi-trait approximation method are the most valuable for all the analysed traits, except average daily gain, for which a detailed analysis is required in order to understand the obtained results. In contrast to DIM, the approximation methods have no computational limits. In addition, as the correlation coefficients between DIM and MT-APM results were positive and large, the MT-APM can be employed for a routine genetic evaluation of CLW pig breeds in the Czech Republic.

Acknowledgments

We would like to thank the Pig Breeders Association of the Czech Republic for providing us with the required data. This study is part of a research supported by the projects MZERO0717 and QJ1310109 of the Ministry of Agriculture of the Czech Republic.

References

- BAUER, J., PŘIBYL, J. and VOŠTRÝ, L. (2015) Contribution of domestic and Interbull records to reliabilities of single-step genomic breeding values in dairy cattle. In *Czech J. Anim. Sci.*, vol. 60, no. 6, pp. 263–267
- GROENEVELD, E., KOVAČ M. and MIELENZ, N. (2008) VCE User's Guide and Reference Manual, Version 6.0. Available from ftp://ftp.tzv.fal.de/pub/latest_vce/doc/ (accessed Aug 1, 2011).
- GROENEVELD E., KOVAC M. and WANG T. (1990) PEST, a general purpose BLUP package for multivariate prediction and estimation. In *Proc. 4th World Congr. Genet. Appl. Livest. Prod., Edinburgh*, 13, pp. 488–491.
- KRUPA, E. and WOLF, J. (2013) Simultaneous estimation of genetic parameters for production and litter size traits in Czech Large White and Czech Landrace pigs. In *Czech J. Anim. Sci.*, vol. 58, pp. 429–436.
- KRUPA, E., ŽÁKOVÁ, E. and KRUPOVÁ, Z. (2015) Evaluation of inbreeding and genetic variability of five pig breeds in Czech Republic. *Asian – Australas. In J. Anim. Sci.*, vol. 28, pp. 25–36.
- MISZTAL, I., LAWLOR, T. J. and SHORT, T. H. (1993) Implementation of single- and multiple-trait animal models for genetic evaluation of Holstein type traits. In *J. Dairy Sci.*, vol. 76, pp. 1421–1432.
- MISZTAL, I. and WIGGANS, G. R. (1988) Approximation of prediction error variance in large-scale animal models. In *J. Dairy Sci.*, vol. 71, no. 2, pp. 27–32.
- MISZTAL, I. and PEREZ-ENCISO, P. (1993) Sparse Matrix Inversion for Restricted Maximum Likelihood Estimation of Variance Components by Expectation-Maximization. In *J. Dairy Sci.*, vol. 76, no. 5, pp. 1479–1483
- MISZTAL, I. et al. (2013) Methods to approximate reliabilities in single-step genomic evaluation. In *J. Dairy Sci.*, vol. 96, pp. 647–654.
- MRODE, R. A. (2005) *Linear Models for the Prediction of Animal Breeding Values*. 2nd ed., Wallingford: CABI.
- MRODE, R. A. (2014) *Linear Models for the Prediction of Animal Breeding Values*. 3rd ed., Wallingford: CABI.
- NEUMAIER, A. and GROENEVELD, E. (1998) Restricted maximum likelihood estimation of covariances in sparse linear models. In *Genet. Sel. Evol.*, vol. 30, pp. 3–26.
- SÁNCHEZ, J. P., MISZTAL, I. and BERTRAND, J. K. (2008) Evaluation of methods for computing approximate accuracies in maternal random regression models for growth trait in beef. In *J. Anim. Sci.*, vol. 86, pp. 1057–1066.
- STRABEL, T., MISZTAL, I. and BERTRAND, J. K. (2001) Approximation of reliabilities for multiple-trait models with maternal effects. In *J. Anim. Sci.*, vol. 79, pp. 833–839.
- TIER, B. and MEYER, K. (2004) Approximating prediction error covariances among additive genetic effects within animals in multiple-trait and random regression models. In *J. Anim. Breed. Genetics*, vol. 121, no. 2, pp. 77–89.
- VANRADEN, P. M. (2008) Efficient methods to compute genomic predictions. In *J. of Dairy Sci.*, vol. 91, pp. 4414–4423.
- VANRADEN, P. M. and WIGGANS, G. R. (1991) Derivation, calculation and use of national animal model information. In *J. Dairy Sci.*, vol. 74, pp. 2737–2746.
- WIGGANS, G. R., MISZTAL, I. and van VLECK, L. D. (1988) Animal model evaluation of Ayrshire milk yield with all lactations, herd-sire interaction, and groups based on unknown parents. In *J. Dairy Sci.*, vol. 71, pp. 1319–1329.