Salaudeen, A., Pitkänen, T., Kempe, R., Koivula, M., Pösö, J., yskov, K., Uimari, P., Lidauer, M. (2026) Using deregressed proofs to estimate multivariate variance components for traits from heterogeneous data sources, *Proceedings 13th World Congress on Genetics Applied to Livestock Production* 23700: https://doi.org/10.31274/wcgalp.23700

Using deregressed proofs to estimate multivariate variance components for traits from heterogeneous data sources

A.T. Salaudeen[1,4*], T. Pitkänen[1], R. Kempe[1], M. Koivula[1], J. Pösö[2], K. Byskov[3], P. Uimari[4], and M.H. Lidauer[1]

[1]Natural Resources Institute Finland (Luke), Jokioinen, Finland; [2]Faba Co-op, Vantaa, Finland; [3]SEGES Innovation, Aarhus, Denmark; [4]University of Helsinki, Helsinki, Finland; [*]Corresponding author: ayo.salaudeen@luke.fi

Introduction

Understanding the relationship between economically important traits is essential for effective breeding programs. However, estimating genetic parameters for many traits simultaneously poses substantial computational challenges. Although several approaches such as Monte Carlo Expectation Maximization Restricted Maximum Likelihood (MC EM-REML) methods have been proposed to address high-dimensional variance component (VC) estimation (Matilainen *et al.,* 2012), computational demands increase with increasing number of traits and highly parameterized models.

Goddard *et al.* (1985) introduced an approach for deregressing estimated breeding values into weighted phenotypic values commonly referred to as deregressed proofs (DRPs), to simplify genetic evaluations. Since then, several studies have evaluated the performance of different deregression methods and weighting factors (Canavesi and Miglior, 1999; Calus *et al.,* 2016). Beyond their original application, DRPs can offer potential computational advantages for multivariate VC estimation for complex traits that have phenotypes from heterogeneous data sources.

In this study we evaluated 1) whether multivariate VCs estimated using DRPs are consistent with those estimated using original phenotypes, and 2) whether DRPs developed for separate trait groups can be combined into one single VC estimation analysis to estimate all genetic correlation simultaneously. Specifically, 13 traits were divided into five trait groups, and heritabilities and genetic correlations were estimated via both approaches to assess the feasibility of using DRPs for multi-trait VC estimation. Finally, a VC analysis including all traits was performed using the developed DRPs.

Materials & Methods

Data. Thirteen traits from three different datasets of Nordic Red dairy cattle were included in this study. DAT1 consisted of field data from 18,757 cows for eight traits subdivided into three groups. Group 1 (milk production) included first lactation milk yield (MY), protein yield (PY), and fat yield (FY); Group 2 (cow carcass traits) included cow carcass weight (CCW), cow carcass conformation (CCC), and cow carcass fatness (CCF); Group 3 (cow size traits) included stature (ST) and repeated metabolic body weight (MBW) measurements. DAT2 formed Group 4 (bull carcass trait) and included bull daily carcass gain (DCG), bull carcass conformation (BCC) and bull carcass fatness (BCF) from 45,514 bull calves. DAT3 comprised of research farm data from 898 cows with repeated observations of first lactation MY, PY, FY, MBW, dry matter intake (DMI) and body weight gain (BWG) analyzed as a single trait group (Group 5).

Estimating variance components and deregressed proofs. In the first step, VCs were estimated from phenotypic data by applying for each group of traits a linear mixed model of the form:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_1\mathbf{h} + \mathbf{Z}_2\mathbf{p} + \mathbf{Z}_3\mathbf{a} + \mathbf{e} \quad (1)$$

Where $\mathbf{y} = \left[\mathbf{y}_1^{\mathrm{T}}, \ldots, \mathbf{y}_k^{\mathrm{T}}\right]^{\mathrm{T}}$ is a vector of phenotypic observations for all k traits of a specific trait group; $\mathbf{X}$ is the design matrix relating phenotypes to the effects in vector $\mathbf{b} = \left[\mathbf{b}_1^{\mathrm{T}}, \ldots, \mathbf{b}_k^{\mathrm{T}}\right]^{\mathrm{T}}$ that includes trait-specific fixed effects; $\mathbf{Z}_1$, $\mathbf{Z}_2$, and $\mathbf{Z}_3$ are incidence matrices that relate the phenotypes to the effects in vectors **h**, **p**, and **a** that contain the random effects for random herd-year, permanent environmental, and additive genetic effects, respectively; and **e** is the vector of random error terms. All random effects were assumed to be multivariate normally distributed: $\mathrm{var}(\mathbf{h}) \sim \mathrm{MVN}(\mathbf{0}, \mathbf{I} \otimes \mathbf{H})$, $\mathrm{var}(\mathbf{p}) \sim \mathrm{MVN}(\mathbf{0}, \mathbf{I} \otimes \mathbf{P})$, $\mathrm{var}(\mathbf{a}) \sim \mathrm{MVN}(\mathbf{0}, \mathbf{A} \otimes \mathbf{G}_0)$, and $\mathrm{var}(\mathbf{e}) \sim \mathrm{MVN}(\mathbf{0}, \mathbf{I} \otimes \mathbf{R})$, where $\otimes$ denotes the Kronecker product, **I** is an identity matrix of appropriate size, **A** the numerator relationship matrix, and **H**, **P**, $\mathbf{G}_0$ and **R** are the estimated VC matrices of size $k \times k$. For trait groups 1, 2 and 4, animals had at most one observation, and therefore $\mathbf{Z}_2\mathbf{p}$ was excluded from the model. All VC were estimated using the MC EM REML algorithm implemented in the MiX99 software suite (Pitkänen *et al.,* 2022).

In the second step, the estimated multivariate VC were used to obtain estimated breeding values (EBV) and their reliabilities by applying the trait group specific models. Reliabilities were approximated by the method of Tier and Meyer (2004) using MiX99. Multi-trait effective record contributions (ERC) were derived from the reliabilities using the reversed reliability estimation approach described by Zaabza *et al.* (2022). The ERCs were used as weights in the multivariate deregression of the EBVs. The deregression was carried out separately for each trait group using the Broyden method (Broyden, 1965) as implemented in MiX99.

Validation of the approach. For each trait group, VCs were re-estimated using the derived DRPs and the associated ERCs by applying a simple multi-trait model:

$$\mathbf{d} = \mathbf{X\mu} + \mathbf{Za} + \boldsymbol{\varepsilon}, \quad (2)$$

where $\mathbf{d} = \left[\mathbf{d}_1^{\mathrm{T}}, \ldots, \mathbf{d}_k^{\mathrm{T}}\right]^{\mathrm{T}}$ is a vector of DRPs for all k traits of a specific trait group; **X** denotes the design matrix that relates the DRPs to the overall mean effects in vector $\boldsymbol{\mu}$; **Z** and **a** remain as described in Eq.1; and $\boldsymbol{\varepsilon}$ is a vector of random residuals with $\mathrm{var}(\boldsymbol{\varepsilon}) \sim \mathrm{MVN}(\mathbf{0}, \mathbf{E})$, where **E** describes the residual variance covariance structure that accounts for the ERC weights. Finally, a comprehensive analysis including all 13 traits was conducted to estimate the genetic correlations across datasets. The traits shared between datasets (MY, PY, FY, and MBW from Group 5) were merged with their corresponding traits in Groups 1 and 3 and using the DRP of Group 5 to avoid double counting.

Results & Discussion

Heritability estimates based on phenotypic data ranged from 0.19 to 0.84, with generally higher standard errors for Group 5 traits (±0.14 to ±0.21) than for Groups 1 to 4 (±0.02 to ±0.03). Heritability estimates from DRP-based analysis ranged from 0.18 to 0.83 (Tables 1 and 2). Differences between the two methods ranged from 0.01 to 0.07 across all traits, with the highest difference observed for DMI in Group 5.

Genetic correlations among Groups 1-4 ranged from -0.07 to 0.81 when using phenotypes and -0.06 to 0.81 when using DRPs (Table 4). Differences between methods were consistently small (0.00 to 0.02) across these groups. For Group 5 traits, genetic correlations ranged from -0.25 to 0.87 from analyses with phenotypic data, and from -0.24 to 0.87 when DRPs were used (Table 3). Differences between methods ranged from -0.09 to 0.09, with the largest discrepancies observed

for correlations involving DMI, MBW, and BWG, which were also the traits with the fewest records. In contrast, well-recorded traits (MY, PY, FY) showed minimal differences (0.00 to 0.02). The similarity between phenotypic and DRP-based estimates improved with increasing sample size indicating that DRP-based variance component estimation is most reliable for well-recorded traits.

Table 1. Heritability estimates for Group 1 (milk production), 2 (cow carcass), 3 (cow size) and 4 (bull carcass) traits in Nordic Red dairy cattle by variance component estimation (VCE) approach.

Trait[1]	MY	PY	FY	CCW	CCC	CCF	STA	MBW	DCG	BCC	BCF
P[2]	0.29	0.20	0.21	0.46	0.24	0.27	0.55	0.65	0.34	0.19	0.20
DRP[2]	0.33	0.23	0.23	0.47	0.25	0.28	0.59	0.59	0.35	0.21	0.21

[1] MY=milk yield, PY=protein yield, FY=fat yield, CCW=cow carcass weight, CCC=cow carcass conformation, CCF=cow carcass fatness, ST=stature, MBW=metabolic body weight, DCG=bull daily carcass gain, BCC=bull carcass conformation, BCF=bull carcass fatness.
[2] P=VCE based on phenotypes; DRP=VCE based on deregressed proofs.

Table 2. Heritability estimates for Group 5 (feed efficiency) traits in Nordic Red dairy cattle by variance component estimation (VCE) approach.

Trait[1]	MY	PY	FY	MBW	DMI	BWG
VCE based on phenotypes	0.46	0.28	0.25	0.84	0.27	0.22
VCE based on deregressed proofs	0.42	0.24	0.19	0.83	0.34	0.18

[1] MY=milk yield, PY=protein yield, FY=fat yield, MBW=metabolic body weight, DMI=dry matter intake, BWG=body weight gain.

Table 3. Genetic correlations between Group 5 (feed efficiency) traits in Nordic Red dairy cattle based on variance component analyses using phenotypes (in lower triangle) or deregressed proofs.

Traits[1]	MY	PY	FY	MBW	DMI	BWG
MY		0.87	0.67	0.26	0.65	-0.24
PY	0.87		0.66	0.35	0.78	-0.14
FY	0.68	0.64		0.20	0.55	-0.03
MBW	0.21	0.29	0.11		0.72	0.42
DMI	0.66	0.79	0.50	0.63		0.05
BWG	-0.25	-0.12	-0.12	0.42	0.04	

[1] Definition of traits is given in Table 2.

Table 4. Genetic correlations between Group 1 (milk production), 2 (cow carcass and stature), and 4 (bull carcass) traits in Nordic Red dairy cattle based on variance component analyses using phenotypes (in lower triangle) or deregressed proofs.

	Group1				Group2					Group 4		
	MY	PY	FY		CCW	CCC	CCF	STA		DCG	BCC	BCF
MY		0.81	0.63	CCW		0.66	0.47	0.67	DCG		0.42	0.02
PY	0.81		0.76	CCC	0.66		0.53	0.06	BCC	0.42		0.14
FY	0.64	0.77		CCF	0.47	0.53		-0.06	BCF	0.02	0.13	
				STA	0.66	0.04	-0.07					

[1] Definition of traits is given in Table 1.

The comprehensive analysis including all 13 traits produced within-group correlations that are comparable with results from group-wise analyses (Table 5). However, when inspecting the convergence of estimated genetic covariances between DAT1 and DAT3 traits, and between DAT2

and DAT3 traits, we found that those estimates had not converged after 1,000 EM steps and would need additional iterations to reach convergence. Thus, the corresponding genetic correlations in Table 5 are expected to change with further iterations. This highlights that comprehensive multi-trait analysis using DRPs is feasible but computationally intensive for between-group correlations.

Table 5. Genetic correlations between Group 1 (milk production), 2 (cow carcass), 3 (cow size), 4 (bull carcass) and 5 (feed efficiency) traits in Nordic Red dairy cattle based on DRP analysis.

T[1]	PY	FY	CCW	CCC	CCF	ST	MBW	DCG	BCC	BCF	DMI	BWG
1	0.81	0.62	0.00	-0.27	-0.36	0.25	0.18	0.23	-0.05	-0.11	0.42	-0.08
2		0.70	0.08	-0.24	-0.30	0.28	0.23	0.25	-0.08	-0.10	0.39	-0.07
3			-0.04	-0.30	-0.34	0.19	0.12	0.18	-0.06	0.01	0.30	-0.11
4				0.67	0.47	0.68	0.80	0.64	0.16	-0.11	0.25	0.14
5					0.56	0.05	0.37	0.33	0.47	-0.02	0.02	0.14
6						-0.05	0.25	0.13	0.11	0.34	-0.09	0.10
7							0.68	0.51	-0.17	-0.23	0.32	0.08
8								0.59	0.07	-0.09	0.32	0.11
9									0.42	0.02	0.33	0.04
10										0.13	0.04	0.05
11											-0.07	-0.08
12												0.05

[1] T=Trait: definition of traits is given in Tables 1 and 2.

Conclusions

This study shows that DRP can be successfully used for VC estimation. Within-group genetic correlations showed minimal differences between DRP and phenotype-based analyses for well-recorded traits, whereas larger differences were observed for traits with limited data. DRP based multi-trait VC analysis is practically advantageous when traits originate from different data sources or when modelling original phenotypes would require highly parameterized models.

Acknowledgements

The work was cofounded by Naudanjalostussäätiö sr (Finland), VikingGenetics Ltd. (Denmark), Nordic Cattle Genetic Evaluation NAV (Denmark), and Faba Co-op (Finland).

References

Broyden, C. G. (1965). Math. Comp. 19:577–593. DOI: https://doi.org/10.1090/S0025-5718-1965-0198670-6

Calus M.P.L., Vandenplas J., ten Napel J., *et al*. (2016) J. Dairy Sci. 99:6403-6419. DOI:doi.org/10.3168/jds.2016-11028

Canavesi F., and Miglior, F. (2019) Interbull Bulletin, 49. journal.interbull.org/index.php/ib/article/view/607

Goddard M. (1985) Livest. Prod. Sci. 13:321-331. DOI: doi.org/10.1016/0301-6226(85)90024-7

Matilainen K., Mäntysaari E.A., Lidauer M.H., *et al*. (2012) J. Anim. Breed. Genet. 129:457-468. DOI: doi.org/10.1111/j.1439-0388.2012.01000.x

Pitkänen T. J., Gao H., Kudinov A., *et al*. (2022) Proc. 12th WCGALP, Rotterdam, the Netherlands.

Tier B., and Meyer K. (2004) J. Anim. Breed. Genet. 121:77–89. DOI: doi.org/10.1111/j.1439-0388.2003.00444.x

Zaabza H.B., Taskinen M., Mäntysaari E.A., *et al*. (2022) J. Dairy Sci. 105:5221-5237. DOI: doi.org/10.3168/jds.2021-21016