Variance Component Re-Estimation and Model Simplification in a Single-Step Random Regression Test-Day Model for Nordic Red Dairy Cattle

T. Pitkänen[1], M.H. Lidauer[1], M. Taskinen[1], J. Pösö[2], A. Kudinov[3], K. Byskov[4], H. Splittorff[5], M. Koivula[1]

[1]Natural Resources Institute Finland (Luke), FI-31600 Jokioinen, Finland

[2] Faba co-op, P.O. BOX 40, FIN-01301 Vantaa, Finland

[3]Växa, Box 288, 75105 Uppsala, Sweden

[4] SEGES Innovation P/S, Agro Food Park 15, 8200 Aarhus N, Denmark

[5] NAV Nordic Cattle Genetic Evaluation, 8200 Aarhus N, Denmark

Corresponding author: e-mail address: timo.j.pitkanen@luke.fi

Abstract

The Nordic Red dairy cattle random regression test-day model (RRM) has been updated to reflect increased genetic similarity among the original country subpopulations (Finland, Sweden and Denmark) and changes in data structure over the past two decades. Data prior to 2005 were excluded, and previously country-specific models were harmonized into a simplified nine-trait multi-trait single-step RRM using a metafounder approach. The objective of this study was to describe the implementation of the updated model. This included re-estimation of variance components for parameter-reduced covariance functions for random effects identified as optimal in a previous study, accounting for heterogeneous measurement error (co)variances across eight milk recording schemes, and evaluation of genomic estimated breeding value reliability using forward cross-validation. Variance components were estimated from a harmonized subset of data sampled across Finland, Sweden and Denmark, applying the updated model. Compared with previously used variance components derived from Finnish-only data, heritability estimates for 305-day yields increased by 8 to 46% across all traits. Substantial differences in measurement error variances and residual correlations among daily milk, protein, and fat yields were observed across the eight milk recording schemes, reflecting heterogeneous milking systems and historical protocol changes in automated milking systems in Finland. For Finland, four residual covariance matrices were required to model measurements from conventional milking systems and three automated milking system protocols. In contrast, two residual classes (automated and conventional milking systems) were sufficient for Denmark and Sweden. The selected covariance functions enabled a substantial reduction in variance component parameters while maintaining model stability and predictive ability. The model was solved using a fully component-wise ssGTBLUP approach with metafounders. The updated model provides a computationally efficient basis for routine genetic evaluation of Nordic Red dairy cattle.

Key words: Nordic Red dairy cattle, test-day model, yield evaluation, variance components

Introduction

In 2006, the Nordic countries, Finland (FIN), Denmark (DNK), and Sweden (SWE), introduced across-country genetic yield evaluations for their dairy cattle populations (Lidauer et al. 2015).

Revising the random regression test-day model (RRM) for Red dairy cattle (RDC) evaluation was considered necessary as the genetic similarity among the original country subpopulations has increased and changes in data structure over the past two decades have made the original model outdated.

The objective of this study was to describe the implementation of the updated Nordic RDC yield evaluation model. This included re-estimation of variance components, deriving parameter-reduced covariance functions for random effects (Pitkänen et al. 2026), accounting for heterogeneous measurement error (co)variances across eight milk recording schemes, updating pedigree and including metafounders. Finally, the impact of the new variance components on genetic parameters are validated.

Materials and Methods

Data

The model update started with the data and pedigree modifications (Table 1). We decided to truncate the test-day (TD) data from year 2005 as older data had much more variation in the countries. From Finnish data, also Holstein records were removed, as well as later lactation (3+) records. This simplifies the model. Including Holstein and later lactation data were important in the past to increase contemporary group sizes for Finnish herds. As herd sizes have increased, this is not needed anymore.

Automated milking system recording protocols have changed historically in Finland but remained stable in Denmark and Sweden (Pitkänen et al. 2024). In Finland, three AMS recording protocols were used: 1) the daily average of a month's milkings, 2) a 24-h method based on two successive milkings (Peeters and Galesloot, 2002), and 3) a 96-h method (Lazenby et al., 2002). The first protocol was used until 2015, and the second from January 2015 to June 2022. However, from January 2017 onwards, AMS herds gradually moved to the 96-h method which is the current method for AMS records. Information on the milking system used for test-day recording was corrected for a small proportion of records because of database issues in Finland and Sweden.

Finally, unknown parent groups were redefined to support the metafounder approach. The revised grouping structure was designed to better reflect the historical relationships among the Nordic RDC subpopulations and provide a more realistic representation of the genetic base population.

Model

The new updated multiple-trait RRM describes TD milk, protein, and fat yields for the first three lactations using nine model traits instead of 27 in the previous model. Each trait has a random regression function for additive genetic and permanent environmental effects. The model includes fixed effects nested within countries, heterosis and recombination loss adjustments, and correction for heterogeneous variance (HV). Due to rank reduction, 16 cow-specific coefficients define all nine breeding value curves (Pitkänen et al. 2026).

First steps in the model revision were to recalculate heterosis and recombination effects (Pitkänen et al. 2025) and correct the calving age effects and new herd-year season effects. Because of moving from 27 country specific traits to nine traits, country interactions were included in some fixed effects.

Table 1: Main changes implemented in the updated Nordic RDC yield evaluation model.

Area	Main update
Data	TD data truncated from 2005, Finnish HOL and lactations 3+ removed
Pedigree	Metafounder approach implemented
Traits	27 traits reduced to 9 traits
Fixed effects	Updated heterosis, recombination, calving age and herd-year-season effects
Recording schemes	AMS/CMS classes revised
Variance components	Nordic RDC specific variance components and parameter-reduced covariance functions
Evaluation	ssGTBLUP with metafounders

Revising the variance components

Variance component estimation was based on a representative sample of Nordic Red dairy cattle from Finland, Denmark, and Sweden,

including TD records from both conventional milking systems (CMS) and automated milking systems (AMS).

Animals were selected according to strict criteria. All records had to originate from a single herd, and complete lactation sequences were required. Cows could contribute data from only the first lactation, first and second lactations, or the first three lactations, with no missing lactations before the final observed lactation. Herds were required to have at least eight but fewer than 100 suitable first-lactation cows per year on average and at least 12 years of data beginning in 2010.

The final sample size was approximately 100,000 animals. Herds were sampled independently within each country × milking-system class, with allocations proportional to the number of first-lactation cows in each class. Variance components were estimated using a pedigree-based random regression TD model implemented in MiX99 (Gao et al., 2026). The statistical model for VCE was defined as:

$$y = \mathbf{X}b + \mathbf{Z_h}h + \mathbf{Zc}c + \mathbf{Z_p}p + \mathbf{Z_a}a + e,$$

The model included fixed effects of heterosis, recombination, herd-year-season, year-month × milking system × country, days carried calf, and a linear herd effect. The random effects included herd-test-day (HTD), herd-lactation curve (HC), permanent environmental (PE), additive genetic (A), and residual effects. The HTD effect was fitted as an intercept effect. The HC effect was modeled using a second-order Legendre polynomial without intercept but combined with a Wilmink function exponential term, with common regression coefficients for the second and third lactations. Both additive genetic and permanent environmental effects were modeled using third-order Legendre polynomials combined with a Wilmink function exponential term. Residual variances were assumed to be heterogeneous across 12 lactation intervals.

To account for differences in recording protocols among countries and milking systems, residual covariance matrices were estimated separately for each country × milking-system class. In total, eight residual classes were defined, including one CMS and three AMS classes in Finland corresponding to different recording protocols, and one AMS and one CMS class in both Denmark and Sweden. This structure allowed heterogeneous measurement error variances and residual correlations to be modelled consistently across recording environments.

Parameter reduction of covariance functions

The estimated variance components were subsequently reduced for implementation in routine single-step genetic evaluations. Covariance functions were fitted to the variance components. The genetic covariance function was reduced from a full rank of 36 to rank 16 while retaining approximately 99% of the original genetic variation. Similarly, the permanent environmental covariance function was reduced from rank 36 to rank 21. The permanent environmental covariance function was designed to absorb variation associated with permanent environmental effects, herd-lactation curve effects, and within-lactation residual variation. For more details, see Pitkänen et al. (2026). The parameter-reduced covariance functions were used in the updated Nordic RDC evaluation model.

Validation

The impact of the updated variance components was assessed with TD data from May 2026. The data included approximately 34.7 million records from 1.78 million animals, and a pedigree with 3.1 million animals. Genotype data for RDC included 315 448 animals. The single-step RRM was solved by preconditioned conjugate gradient iteration on data (Strandén and Lidauer, 1999) implemented in MiX99. The model was solved using a fully component-wise formulation of ssGTBLUP (Mäntysaari et al., 2017) combined with metafounders. Metafounders were used to account for genetic

differences among base populations and to model relationships among founder groups.

Validation was based on a four-year data truncation. The linear regression (LR) method (Legarra and Reverter, 2018) was used for validation. Validation statistics included prediction bias, regression coefficients (b_1), and coefficients of determination (R^2) for milk, protein, and fat yield traits.

Validation bulls were required to have at least 20 daughters in the complete dataset and no daughters in the reduced dataset. For cows, the validation set included animals with at least five TD records in the full data and no records in the reduced data.

Results

The updated variance components resulted in clearly higher heritability estimates compared with the previous model (Table 2). For milk yield, heritability increased from 0.39 to 0.44 in first lactation, from 0.28 to 0.35 in second lactation, and from 0.25 to 0.31 in third lactation. Similar increases were observed for protein and fat yield traits.

The largest relative increase was observed for third-lactation fat yield, where heritability increased by approximately 46%. For the other traits, increases ranged from 12% to 26%.

Table 2: Old and new 305d heritability estimates for milk, protein and fat in lactation 1-3, and change in heritability as %.

Trait	New	Old	Change %
Lactation1			
Milk	0.44	0.39	13
Protein	0.37	0.33	12
Fat	0.38	0.34	12
Lactation 2			
Milk	0.35	0.28	26
Protein	0.32	0.27	19
Fat	0.30	0.25	20
Lactation 3			
Milk	0.31	0.25	24
Protein	0.29	0.24	21
Fat	0.29	0.20	46

Substantial differences in residual variances were observed among milking-system classes (Figure 1). In Finland, the residual variance for milk yield differed markedly between CMS and the three AMS recording protocols. The highest residual variances were generally observed for the 24-h AMS protocol, whereas the monthly average showed reduced residual variation. In Denmark and Sweden, residual variance profiles were more stable because recording protocols remained unchanged throughout the study period.

These results indicate that a single residual covariance structure is insufficient for describing the heterogeneous recording environments present in the Nordic RDC population. Accounting for country-specific AMS and CMS recording schemes improved the biological interpretation of residual variation and provided a more realistic description of measurement error in the updated evaluation model.

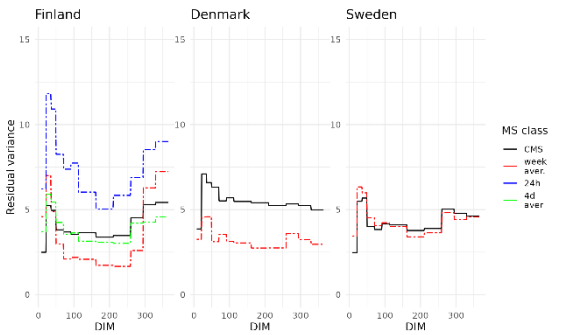


Figure 1. Residual variances for first-lactation milk yield across country × milking-system classes. Finnish AMS observations were divided according to recording protocol (week average, 24-h equivalent and 96-h average).

Validation results showed comparable performance between the old and new variance components (Table 3). The old variance components resulted in slightly better regression coefficients for all traits, differences were small. Also, coefficients of determination were nearly same for the old and the new variance components, and overall, predictive ability remained at a similar level for both models.

The updated variance components resulted in slightly higher genetic trends and larger genetic standard deviations for both bulls and cows (Figure 2). Consequently, greater differentiation among animals was observed, leading to some re-ranking of breeding candidates compared with evaluations based on the previous variance components.

Table 3: Linear regression (LR) validation in bulls and cows with new (NewVCE) or old (OldVCE) variance components. The values in the table are: bias(kg) = mean ($GEBV_{Full} - GEBV_{Reduced}$) ±SD, b_1 regression coefficient and R^2 coefficient of determination for milk, protein and fat, respectively.

	Model	bias (kg)	b_1	R^2
Bulls	OldVCE	-267.21 ±377.8	1.06	0.75
	NewVCE	-296.67 ±397.7	1.02	0.73
	OldVCE	-9.48 ±13.4	0.99	0.64
	NewVCE	-11.00 ±13.9	0.96	0.63
	OldVCE	-11.12 ±15.8	0.97	0.67
	NewVCE	-13.60 ±16.8	0.97	0.68
	Model	**bias(kg)**	**b_1**	**R^2**
Cows	OldVCE	-120.82 ±279.1	1.03	0.83
	NewVCE	-147.94 ±303.6	1.01	0.81
	OldVCE	-3.75 ±9.8	0.98	0.78
	NewVCE	-4.92 ±10.1	0.96	0.77
	OldVCE	-3.84 ±11.8	0.97	0.79
	NewVCE	-5.69 ±12.5	0.97	0.80

Discussion and conclusions

The updated Nordic RDC evaluation model includes several important improvements in data structure, pedigree handling, trait definitions, and variance component estimation. The use of newly estimated Nordic-specific variance components provides a more realistic representation of the current Nordic RDC population and production environment. In addition to updated variance components, the implementation of metafounders provides a more appropriate framework for modelling the increasingly connected Nordic RDC population.

Higher heritability estimates suggest that the previous model underestimated the proportion of genetic variation, particularly in later lactations. The estimation of separate residual covariance matrices for different country × milking-system classes revealed substantial heterogeneity in measurement error structures. This was particularly evident in Finland, where changes in AMS recording protocols over time resulted in distinct residual variance patterns. These findings support the use of multiple residual covariance structures in Nordic evaluations and suggest that part of the residual variation previously attributed to unexplained environmental effects was related to differences in recording procedures.

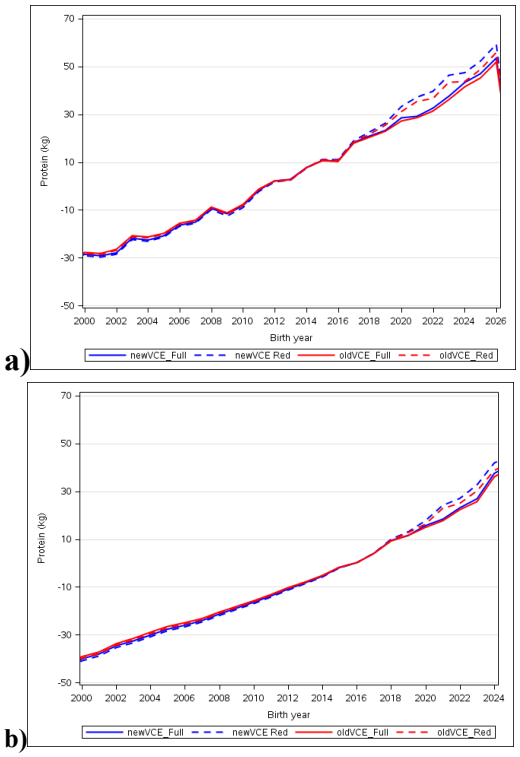


Figure 2. Genetic trends for protein a) bulls and b) cows by birth year. The model ssGTBLUP with MF using old (OldVCE) or new variance components (NewVCE).

Parameter reduction successfully simplified the covariance functions while preserving nearly all biologically relevant variation. This reduction is important for maintaining computational efficiency in routine single-step evaluations involving large datasets and genomic information.

Acknowledgements

This work was a part of the Nordic project "Modelling of phenotypes and ancestral relationships in genomic prediction for yield traits in Nordic Red dairy cattle". The Nordic breeding organizations (Viking Genetics, Randers, Denmark; Nordic Cattle Genetic Evaluation (NAV), Aarhus, Denmark), are acknowledged for providing the genotype data and the TD data, and for cofounding the project.

References

Gao, H., Taskinen, M., Pitkänen, T.J., Ylitalo, A.-K., Vargas Jurado, N., Lidauer, M.H., Strandén, I. 2026. Estimation of variance components in single-step models using MiX99. Proc. 13th WCGALP, Madison, Wisconsin, USA, July 12-17, 2026. doi: 10.31274/wcgalp.23537.

Lazenby, D., Bohlsen, E., Hand, K.J., Kelton, D.F., Miglior, F. and Lissemore, K.D. 2002. Methods to estimate 24-hour yields for milk, fat and protein in robotic milking herds. In: *Performance Recording of Animals: State of the Art 2002*. EAAP/ICAR Technical Series, pp. 65-71. doi: 10.3920/9789086865116_010

Legarra, A., Reverter, A. 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. doi: 10.1186/s12711-018-0426-6.

Lidauer, M., Pösö, J., Pederson, J., Lassen, J., Madsen, P., Mäntysaari, E.A., Nielsen, U., Eriksson, J-Å., Johansson, K., Pitkänen, T., Strandén, I., Aamand, G.P. 2015. Across-country test-day model evaluations for Nordic Holstein, Red Cattle and Jersey. *J. Dairy Sci. 98*, 1296–1309. doi: 10.3168/jds.2014-8307.

Mäntysaari, E.A., Evans, R.D., Strandén, I. 2017. Efficient single-step genomic evaluation for a multibreed beef cattle population having many genotyped animals. *J. Anim. Sci. 95*, 4728-4737. doi: 10.2527/jas2017.1912.

Peeters R, Galesloot P. 2002. Estimating Daily Fat Yield from a Single Milking on Test Day for Herds with a Robotic Milking System. *J. Anim. Sci.,* 85, 682-688. doi: 10.3168/jds.S0022-0302(02)74124-6.

Pitkänen, T, Lidauer, M, Pösö,J, Fikse, W.F., Nielsen, U.S. Aamand, G.P., Koivula, M. 2024. Revision of random regression test-day model has improved genomic prediction for Nordic Red dairy cattle. *Interbull Bulletin,* 60, 20-21. https://journal.interbull.org/index.php/ib/article/view/1932.

Pitkänen, Koivula, M, Taskinen, M, Pösö, J, Nielsen, U.S, Kudinov, A, Lidauer, M.H. 2025. Dynamic breed definition for admixture Nordic Red dairy cattle – An ad hoc approach for modelling heterosis in genomic evaluation for production traits. In: *Book of abstracts of the 67th annual meeting of EAAP*, Innsbruck, Austria. 25- 29 Aug. p. 584.

Pitkänen, T, Taskinen, M, Koivula, M, Pösö, J, Kudinov, A, Nielsen, U.S., Byskov, K, Lidauer, M.H. 2026. Deriving covariance functions (CF) for Nordic Red dairy cattle yield evaluation model. *Proc. 13th WCGALP*, Madison, Wisconsin, USA, July 12-17, 2026. doi: 10.31274/wcgalp.23946.

Strandén, I., Lidauer, M 1999. Solving large mixed models using preconditioned conjugate gradient iteration. *J. Dairy Sci*. 82, 2779-2787. doi: 10.3168/jds.S0022-0302(99)75535-9.