Pitkänen, T., Taskinen, M., Koivula, M., Pösö, J., Kudinov, A., Nielsen, U., Byskov, K., Lidauer, M. (2026) Deriving covariance functions (CF) for Nordic Red dairy cattle yield evaluation model, *Proceedings 13th World Congress on Genetics Applied to Livestock Production* 23946: https://doi.org/10.31274/wcgalp.23946

Deriving covariance functions (CF) for Nordic Red dairy cattle yield evaluation model

T. Pitkänen[1*], M. Taskinen[1], M. Koivula[1], J. Pösö[2], A. Kudinov[3], U.S. Nielsen[4], K. Byskov[4], and M.H. Lidauer[1]

[1]*Natural Resources Institute Finland (Luke), Jokioinen, Finland;* [2]*Faba co-op, Vantaa, Finland;* [3]*Växa, Uppsala, Sweden;* [4]*SEGES Innovation, Aarhus N, Denmark;*
**Corresponding author: timo.j.pitkanen@luke.fi*

Introduction

Random regression test-day models are widely used in dairy cattle genetic evaluations because they allow genetic and environmental effects to vary continuously over days in milk (DIM) and provide a biologically meaningful description of lactation curves. Key components of these models are the applied random regression functions, which define the (co)variance structure of random effects across lactation and directly affects estimates of genetic parameters and breeding values (Meyer, 1998).

To model time-dependent (co)variances, polynomial-based functions such as Legendre polynomials (Kirkpatrick *et al.*, 1994) and parametric functions such as the Wilmink curve (Wilmink, 1987) have been widely applied. In large-scale routine evaluations, however, highly parameterized models can lead to substantial computational demands and numerical challenges. Consequently, parameter-reduced covariance functions (CF) derived through eigenvalue decomposition of estimated variance–covariance matrices have been introduced (Kirkpatrick and Meyer, 2004).

In the Nordic Red dairy cattle (RDC) routine production traits evaluation, parameter-reduced CF are applied in the test-day (TD) model for modelling additive and non-additive effects across traits and lactations, allowing smooth genetic variance curves and efficient computation in a multi-trait setting (Lidauer *et al.*, 2015). Previous studies have shown that moderate parameter reduction has limited impact on estimated breeding values and genetic trends (Meyer, 2001). However, with the increasing integration of genomic information into national evaluations, the consequences of CF choice and parameter reduction for genomic estimated breeding values (GEBV) and their reliability are less studied.

The aim of our study was to evaluate alternative CFs with respect to their fit to observed variance components and their effects on GEBV and validation outcomes for milk, protein, and fat yields. For RDC, the currently applied CF that comprises of a 2nd order Legendre and the exponential term exp(−0.04·DIM) (LW) was compared with a CF based on a Wilmink function (W), and a CF based on a repeatability model (R), as well as with stronger parameter-reduced LW CFs. Forward validation was carried out by applying a linear regression method and calculating Pearson's correlations between 305d yield deviations (YD) and earlier GEBVs.

Materials & Methods

Test-day data and variance components. The official Nordic milk production evaluation data were obtained from Nordic Cattle Genetic Evaluation (NAV) and used in all analyses. TD milk yield, protein yield, and fat yield observations from the first three lactations of all Danish, Finnish and Swedish RDC cows recorded since 2005 were modelled by a 9-traits multi-trait random regression TD model, and including in total 32.7 million TD records and 2.7 million animals in the pedigree.

The necessary variance components for the model were estimated by applying a random regression model that fitted LW functions on herd, additive genetic and non-additive animal

effects, and 12 lactation-stage specific residual effects. The model was applied on a subset of the data including 306,296 records from 14,141 animals of 74 randomly sampled herds. The variance components were estimated by a Monte Carlo expectation maximization REML algorithm, implemented into MiX99 (Pitkänen *et al.*, 2022). The applied model can be presented as: $\mathbf{y}=\mathbf{Xb}+\mathbf{Wh}+\mathbf{Wa}+\mathbf{Wp}+\mathbf{e}$, where **y** is a vector of phenotypes, **X** denotes the design matrix for the fixed effects in vector **b**, **W** is a design covariable matrix that relates TD observations with DIM-specific LW coefficients to the random effects in vectors **h**, **p**, and **a** that contain the random herd, additive genetic, and non-additive genetic animal effects, respectively, and **e** is the vector of random residuals. The random effects were assumed to be multivariate normal distributed: $\mathbf{h}\sim\text{MVN}(\mathbf{0},\mathbf{I}\otimes\mathbf{P}_h)$, $\mathbf{a}\sim\text{MVN}(\mathbf{0},\mathbf{A}\otimes\mathbf{P}_a)$, $\mathbf{p}\sim\text{MVN}(\mathbf{0},\mathbf{I}\otimes\mathbf{P}_p)$, $\mathbf{e}\sim\text{MVN}(\mathbf{0},\mathbf{I}\otimes\mathbf{R})$, where $\otimes$ denotes the Kronecker product, **A** the numerator relationship matrix, $\mathbf{P}_i$ is the estimated variance component matrix of size 36×36 with $i = h,a,p$, and **R** has 3×3 block diagonal form with lactation stage specific residual (co)variance estimates.

Parameter-reduced covariance functions. In a first step, the coefficient matrix $\mathbf{K}_i$ of the CF_i was obtained by fitting CF_i is $\mathbf{\Phi}_0\mathbf{K}_i\mathbf{\Phi}_0^T=\mathbf{\Phi}_1\mathbf{P}_i\mathbf{\Phi}_1^T$, where $\mathbf{\Phi}_0$ and $\mathbf{\Phi}_1$ are matrices containing covariates for the new and original CF_i for selected DIM. For obtaining $\mathbf{K}_i$, where $i=p$, an expectation maximization algorithm was applied that combined random herd, non-genetic additive animal and residual effects, resulting in a $\mathbf{K}_p$ and a measurement error (co)variance matrix **M** (Lidauer *et al*., 2015). The covariates in $\mathbf{\Phi}_0$ were based on the functions LW, W, or R, resulting in a coefficient matrix $\mathbf{K}_i$ of rank 36, 27, or 9, respectively. In the following, parameter-reduced CFs were developed by assessing the eigenvalues of the correlation matrix ($\mathbf{C}_i$) of $\mathbf{K}_i$, i.e., $\mathbf{C}_i=\mathbf{S}_i^{-1}\mathbf{K}_i\mathbf{S}_i^{-1}$ with $\mathbf{S}_i=\text{diag}(\mathbf{K}_i^{0.5})$. Based on the eigenvalue decomposition $\mathbf{C}_i=\mathbf{V}_i\mathbf{D}_i\mathbf{V}_i^T$, a parameter-reduced CF was built that retained at least 99.5% of the variance of $\mathbf{D}_i$ into $\mathbf{D}_i^*$ that ignored the remaining small eigenvalues. This resulted in:

$$\mathbf{\Phi}_0\mathbf{K}_i\mathbf{\Phi}_0^T\approx\mathbf{\Phi}_0\mathbf{S}_i\,\mathbf{V}_i^*\,\mathbf{D}_i^{*0.5}\mathbf{I}_i\mathbf{D}_i^{*0.5}\mathbf{V}_i^{*T}\mathbf{S}_i\mathbf{\Phi}_0^T=\mathbf{U}_i\mathbf{I}_i\mathbf{U}_i^T, \quad (1)$$

where $\mathbf{U}_i$ contains the covariates of the parameter-reduced CF_i that are applied in the single-step TD model for random effect i and $\mathbf{I}_i$ denotes the applied variances for random effect i with vector size equal to the rank of CF_i. For the LW CF, additional parameter reductions, retaining 99.0, 98.0, and 97.0% of the variance, were also considered.

Validation of covariance functions. All studied CFs were incorporated into single-step TD models that were solved by a fully component-wise ssGTABLUP approach (Mäntysaari *et al*., 2017; Vanderplas *et al* 2023) implemented into MiX99 (Strandén and Lidauer, 1999). The genomic relationship matrix was based on VanRaden (2008) method 1 with a 10% residual polygenic proportion. The impact of the chosen CF on GEBVs was assessed using the linear regression validation method (Legarra and Reverter, 2018) that is regressing GEBVs obtained from the full dataset on GEBVs estimated from a reduced dataset excluding the recent 4 years of data. The validation cohort consisted of 56,272 genotyped cows that had no own observations in the reduced data. A combined GEBV was derived from the estimated breeding value coefficients, by forming 305d GEBVs and applying lactation weights 0.30, 0.25 and 0.45. Validation metrics included the regression coefficient (b_1) and the coefficient of determination (R^2). Furthermore, Pearson's correlations between combined 305d YD from the full dataset and GEBVs from the reduced dataset were used as an additional measure of prediction accuracy.

Results & Discussion

The number of unknowns per animal in the original model was 36 for both non-genetic and genetic animal effects. Running evaluation model with original VCE would not be feasible due to computing resource requirements. Reducing the dimensionality of the LW CF from 99.5 to

97.0% of the explained variance substantially decreased the number of parameters, with only minor changes in model fit. The Log Likelihood values differed little among the LW CFs, whereas the W CF and R CF showed more pronounced reductions in Log Likelihood values (Table 1).

Table 1. Log Likelihood and number of parameters (in parentheses) of fitted covariance function (CF) for additive genetic ($\mathbf{U}_a\mathbf{U}_a^T$) and non-additive genetic ($\mathbf{U}_p\mathbf{U}_p^T$) random effects.

CF[1]	LW99.5	LW99	LW98	LW97	W	R
$\mathbf{U}_a\mathbf{U}_a^T$	1072.7 (20)	1068.0 (18)	1039.1 (15)	1034.9 (14)	899.6 (15)	-1628.6 (5)
$\mathbf{U}_p\mathbf{U}_p^T$	173.9 (25)	152.6 (21)	136.7 (18)	97.1 (15)	137.4 (19)	75.5 (6)

[1] The CF based on the Legendre Wilmink (LW) function was reduced by eigenvalue decomposition retaining 99.5% (LW99.5), 99% (LW99), 98% (LW98), or 97% (LW97) of the variance. For the CF based on the Wilmink function (W) or a repeatability model (R) 99.5% of the variance was retained.

Daily heritability estimates derived from LW CF closely followed those obtained from the original variance components, regardless of the degree of parameter reduction. The W CF produced slightly altered heritability trajectories across DIM, while the R CF resulted in a constant heritability across lactation (Figure 1).

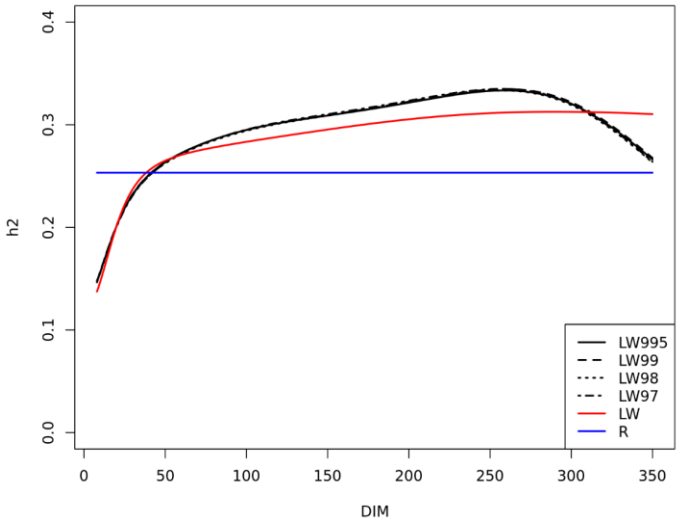

Figure 1. Daily heritability for first lactation milk yield by covariance function (CF). The CF based on the Legendre Wilmink (LW) function was reduced by eigenvalue decomposition retaining 99.5% (LW99.5), 99% (LW99), 98% (LW98), or 97% (LW97) of the variance. For the CF based on the Wilmink function (W) or a repeatability model (R) 99.5% of the variance was retained.

Validation results revealed clear differences among applied CFs. For milk yield, the R CF performed weakest ($b_1 = 0.95$, $R^2 = 0.73$), whereas the LW CF retaining 99.5% of the variance achieved the best results ($b_1 = 0.98$, $R^2 = 0.78$). The W CF showed intermediate performance (Table 2). Similar patterns were observed for protein and fat yields. Reducing the LW CF to preserve at least 97.0% of the variance had only marginal effects on validation results. When correlations between combined 305d YD and combined 305d GEBVs from the reduced data sets were examined, the pattern was less clear. We obtained correlations of above 0.36 for all LW CFs and for the W CF, but a slightly higher correlation of 0.37 for the R CF. Correlations also tended to slightly increase as the degree of rank reduction became stronger.

While earlier studies have reported limited effects of covariance function rank reduction on estimated breeding values (Kirkpatrick and Meyer, 2004), the present study demonstrates that this conclusion also holds for genomic evaluations. To our knowledge, this is the first study to

systematically quantify the impact of covariance function structure and rank reduction on genomic predictions.

Table 2. Linear regression validation results for combined 305d genomic estimated breeding values (GEBV) of cows for milk, protein and fat yield by model. Regression coefficients (b_1) and coefficient of correlation (R^2), as well as mean bias in GEBVs (b_0) with standard error (±).

	Milk yield			Protein yield			Fat yield		
Model[1]	b_0	b_1	R^2	b_0	b_1	R^2	b_0	b_1	R^2
LW99.5	-185±310	0.98	0.78	-6.35±10.1	0.94	0.76	-7.31±11.9	0.93	0.79
LW99	-196±312	0.98	0.78	-6.64±10.2	0.93	0.76	-7.96±11.9	0.92	0.79
LW98	-167±308	0.97	0.78	-6.04±10.3	0.94	0.75	-6.97±11.9	0.93	0.79
LW97	-159±311	0.97	0.78	-5.87±10.4	0.94	0.75	-7.13±12.3	0.92	0.78
W	-208±320	0.97	0.78	-7.24±10.4	0.93	0.75	-8.52±12.1	0.92	0.78
R	-212±335	0.95	0.73	-7.31±10.9	0.91	0.71	-8.36±13.0	0.90	0.75

[1] Applied CF based on the Legendre Wilmink (LW) function was reduced by eigenvalue decomposition retaining 99.5% (LW99.5), 99% (LW99), 98% (LW98), or 97% (LW97) of the variance. For the CF based on the Wilmink function (W) or a repeatability model (R) 99.5% of the variance was retained.

[2] $b0 = \text{mean}(GEBV_{FULL} - GEBV_{REDUCED})$.

Conclusions

The results demonstrate that CF structure has impact on genomic prediction accuracy in TD models. Parameter reduction within the LW CF had minor effects on model fit and GEBV reliability, supporting its use in routine genomic evaluations. In contrast, replacing the LW CF with simpler structures, particularly the repeatability model, resulted in reduced validation results. The differences in Log Likelihood among models were small relative to differences observed in validation results, indicating that model fit alone may not fully reflect genomic prediction performance. These findings highlight the importance of maintaining sufficient flexibility in the covariance structure to capture time-dependent genetic variation. Overall, the study confirms that conservative parameter reduction within a CF is preferable to adopting overly simplified covariance structures in single-step TD model evaluations.

References

Kirkpatrick M., Hill W.G., and Thompson R. (1994) Genetic Research 64:57–69. doi.org/10.1017/S0016672300032559

Kirkpatrick, M., and Meyer K. (2004). Genetics 168:2295–2306. doi.org/10.1534/genetics.104.029181

Legarra A., and Reverter A. (2018). Genetics Selection Evolution 50:53. doi.org/10.1186/s12711-018-0426-6

Lidauer M.H., Pösö J., Pedersen J., *et al*. (2015). Journal of Dairy Science 98:1296–1309. doi.org/10.3168/jds.2014-8307

Meyer, K. (1998). Genetics Selection Evolution 30:221-240. doi.org/10.1186/1297-9686-30-3-221

Meyer, K. (2001). Genetics Selection Evolution 30:557. doi.org/10.1186/1297-9686-33-6-557

Mäntysaari E.A., Evans R.D., and Strandén I. (2017) Journal of Animal Science 95:4728-4737. doi.org/10.2527/jas2017.1912

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

Strandén I., and Lidauer M. (1999) Journal of Dairy Science 82:2779-2787. doi.org/10.3168/jds.S0022-0302(99)75535-9

Vandenplas J., ten Napel J., Darbaghshahi S.N., *et al*. (2023). Genetics Selection Evolution 55:37. doi.org/10.1186/s12711-023-00808-z

VanRaden, P. M. (2008). Journal of Dairy Science 91(11):4414-4423. doi.org/10.3168/jds.2007-0980

Wilmink J.B.M. (1987). Livestock Production Science 16:335–348. doi.org/10.1016/0301-6226(87)90003-0