

Nordic Holstein single-step test day model using left truncated genomic data

Minna Koivula, Ulrik S. Nielsen, Ismo Strandén, Gert P. Aamand & Esa A. Mäntysaari

Introduction

- Genomic models seem to overvalue the genomic relationship (or SNP) information
 - Causes inflation in the genomic evaluations of the candidate animals
- The problem is generally smaller in the single-step evaluations than in the multi-step evaluations, but in multi-step evaluations more tools are available to handle the problem
- We studied whether the old genotypes cause bias in the single-step evaluation

Materials and methods

- February 2022 HOLSTEIN Test Day data & genotypes
- Genomic data:
 - 384 029 genotyped animals;
 - 292 969 cows
 - 83 594 bulls
 - including ~30 000 Eurogenomic genotypes
- FULL TD data
 - 8.8 million animals with records,
11.4 million animals in pedigree
- REDUCED TD data for the validation (four years reduction)
 - 7.9 million animals with records



Alex Arkink

Models

Test Day model:

Multi-Trait Multi-Lactation

- Milk, protein, fat x 3 lactations = 9 traits, modelled by 15 BV random regression coefficients per animal
- PE effect by 27 random regression coefficients
- HV correction using Meuwissen approach

Models

- Single-step models
 - ssGTBLUP with AF=0.5
 - 176 genetic groups and (partial) QP transformation
 - Residual polygenic proportion $w=0.30$
 - Pedigree inbreeding accounted in \mathbf{A}^{-1} and \mathbf{A}_{22}
 - Matrix \mathbf{G} was scaled so $\text{trace}(\mathbf{G})=\text{trace}(\mathbf{A}_{22})$

Solved using new ssGTaBLUP in MiX99

- Calculations for the T matrix moved from preprocessing program to the solver
i.e. $(\mathbf{Z}'\mathbf{A}_{22}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{A}_{22}^{-1}$
 - So called Partial QP - where QP as $\mathbf{Q}'_2\mathbf{A}_{22}^{-1}$ but not $\mathbf{Q}'_2\mathbf{G}^{-1}$
 - New MiX99 solves simultaneously the SNP effects using $(\mathbf{Z}'\mathbf{A}_{22}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{A}_{22}^{-1}\hat{\mathbf{u}}_2$ (Liu et al. J. Dairy Sci 2014)
 - Simpler prediction program “*predict_GEBV*” reads SNP and animal solutions and calculates estimates of the new candidate GEBVs
(and DGVs for all the genotyped animals)
-
- EBV
 - 176 genetic groups
 - Pedigree inbreeding accounted

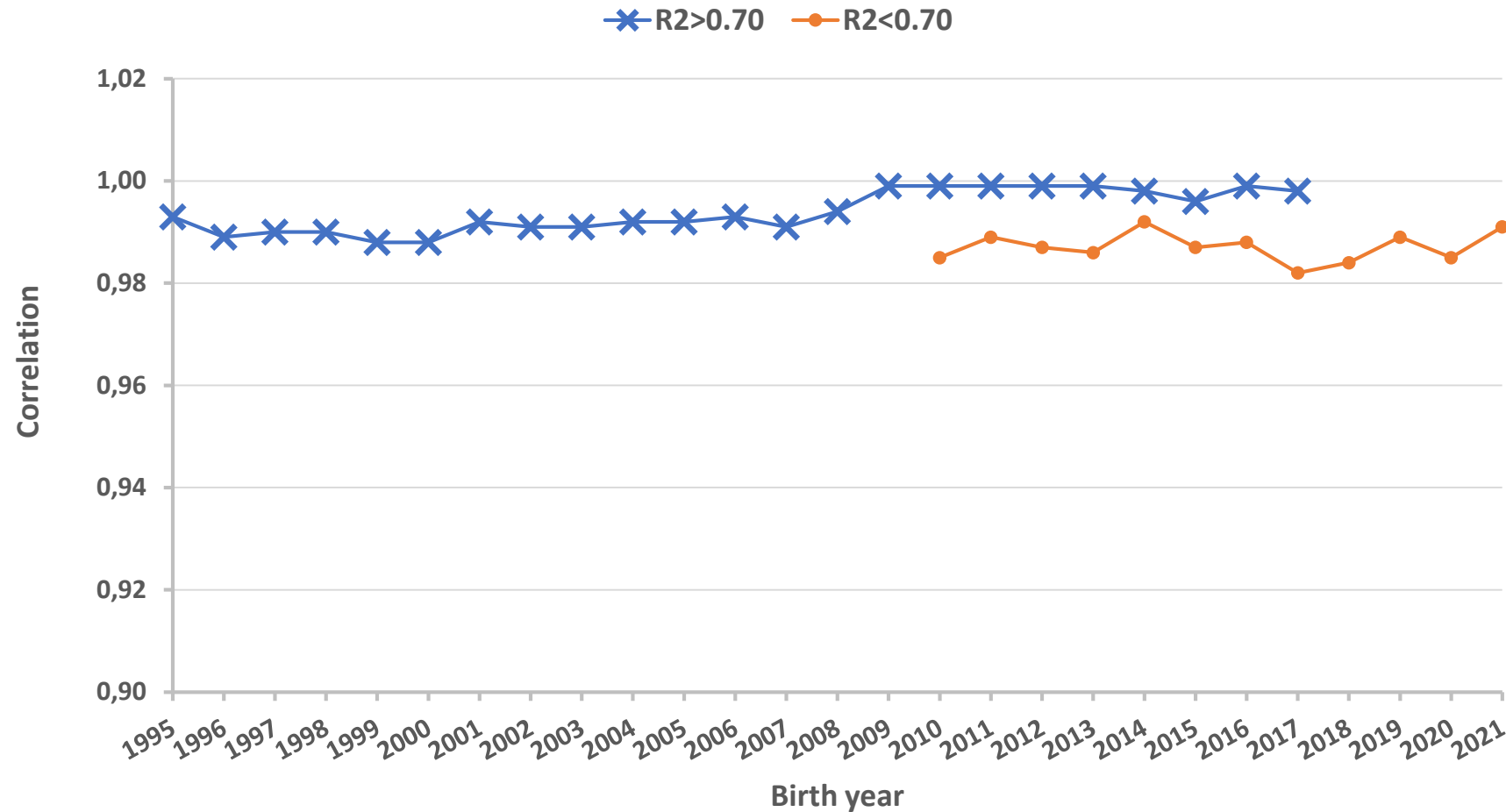
Byear	Bulls	Cows
0 *	2677	565
<1980	110	
1980	24	
1981	30	
1982	27	
1983	23	
1984	23	
1985	34	
1986	231	
1987	312	
1988	324	
1989	365	
1990	451	
1991	518	
1992	554	
1993	602	
1994	728	
1995	729	
1996	812	
1997	854	1
1998	1221	
1999	1625	6
2000	1693	3
2001	1757	5
2002	1824	7
2003	2177	27
2004	2130	123
2005	2260	218
2006	2320	370
2007	2248	462
2008	2545	806

Left truncated genomic data

- Truncated genomic data (**birth years ≤ 2008**)
 - 33 821 genotypes removed
 - 350 208 remained in the analyses
 - Most removed genotypes were HOL breed and RED HOL
 - Most removed bull genotypes were from close by birth years, removed cow genotypes with unknown birth year.

* genotyped with missing birth year have also unknown parents

Correlation between full TD data GEBVs from full and truncated genomic data runs (selected DFS AI bulls only)



Correlation for bulls with R²prot > 0.7 and for bulls with R²prot < 0.7

Bull **first** lactation GEBV validation results using Legarra-Reverter Regression [§]

[§] Regression of (G)EBV_full on PA or GEBV_red

		Bias	b ₁	R ²
Milk1	PA	-212.23	0.86	0.30
	GEBV	-411.96	0.93	0.65
	GEBV_Trunc	-167.65	0.98	0.68
Protein1	PA	-2.85	0.79	0.23
	GEBV	-13.02	0.86	0.60
	GEBV_Trunc	-4.68	0.96	0.64
Fat1	PA	-5.43	0.85	0.34
	GEBV	-18.31	0.85	0.64
	GEBV_Trunc	-9.45	0.94	0.68

Number of validation bulls, 366

Note:
Results are without Inclusion of MACE data on EG bulls

- bias - mean(Full_(G)EBV – reduced_(G)EBV)
- GEBV - ssGTBLUP with QP and w30 and AF 0.5
- GEBV_Trunc - ssGTBLUP , w30 and AF 0.5, genomic data truncation
- PA - EBV w. no daugh



Bull **combined** lactation GEBV validation results using Legarra-Reverter Regression [§]

[§] Regression of (G)EBV_full on PA or GEBV_red

		Bias	b ₁	R ²
Milk	PA	-148.09	0.81	0.30
	GEBV	-457.26	0.86	0.64
	GEBV_Trunc	-148.04	0.93	0.68
Protein	PA	0.26	0.72	0.23
	GEBV	-14.60	0.77	0.59
	GEBV_Trunc	-3.96	0.88	0.62
Fat	PA	-2.83	0.82	0.36
	GEBV	-21.01	0.79	0.66
	GEBV_Trunc	-9.53	0.88	0.69

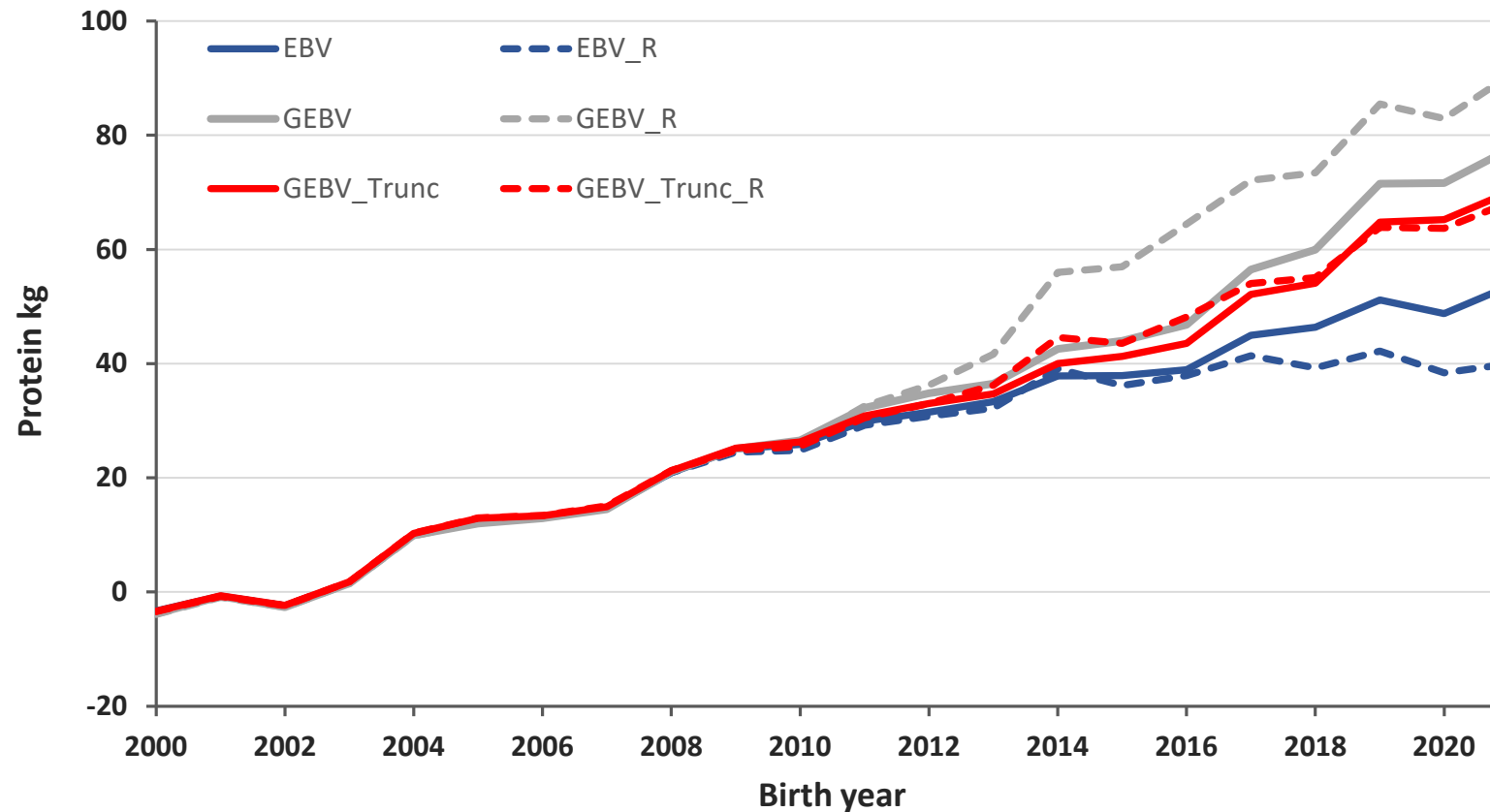
Number of validation bulls, 366

Note:
Results are without Inclusion of MACE data on EG bulls

- bias - mean(Full_(G)EBV – reduced_(G)EBV)
- GEBV - ssGTBLUP with QP and w30 and AF 0.5
- GEBV_Trunc - ssGTBLUP , w30 and AF 0.5, genomic data truncation
- PA - EBV w. no daugh

Genetic trend for Protein combined GEBV for genotyped DFS bulls

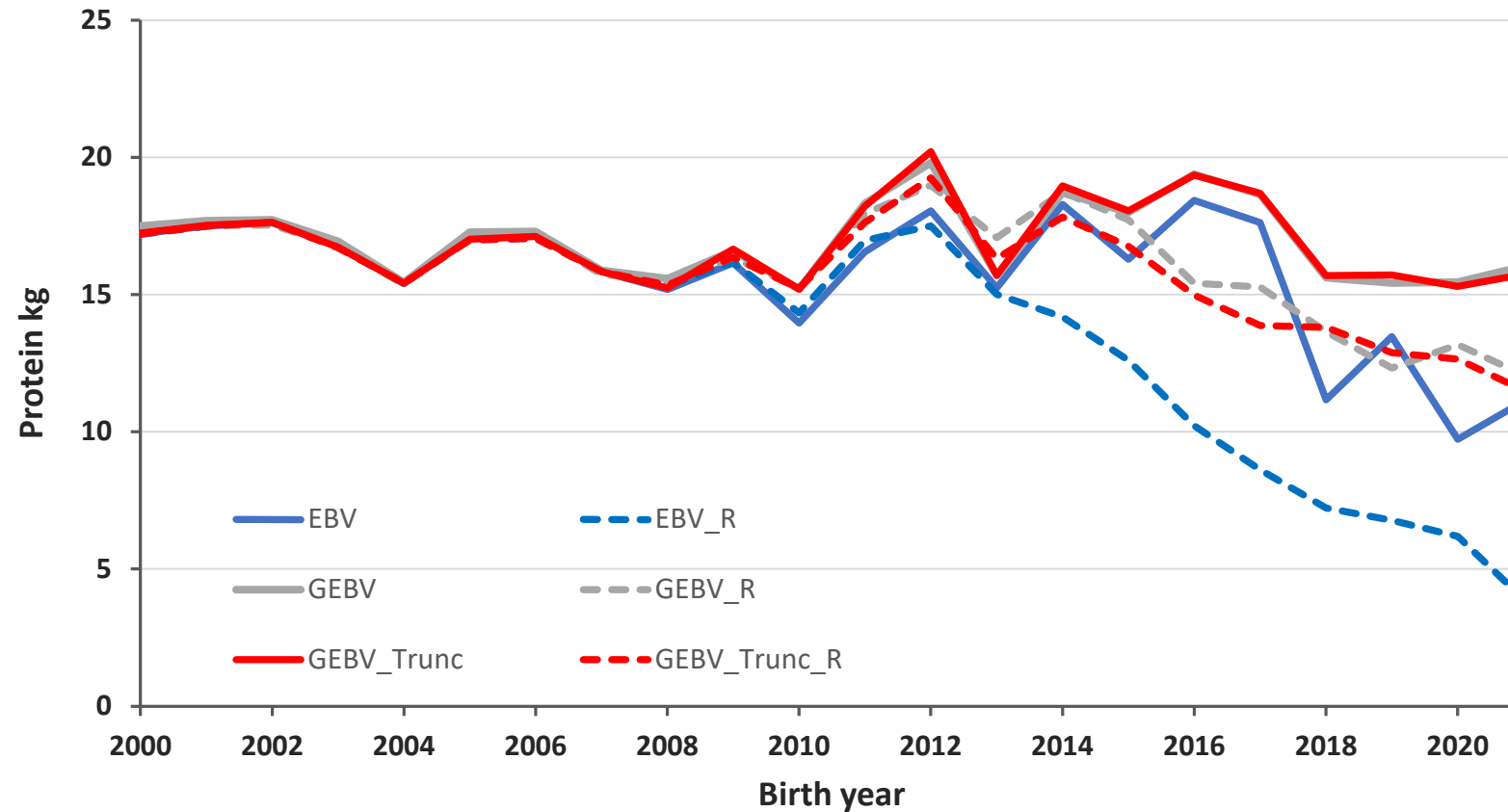
Bulls registered
in data base



GEBV – ssGTBLUP with QP and w30 and AF 0.5
 GEBV_Trunc – ssGTBLUP, w30 and AF 0.5, genomic data truncation
 EBV – animal model

© Natural Resources Institute Finland

Trend in SD of Protein combined GEBV for genotyped DFS bulls

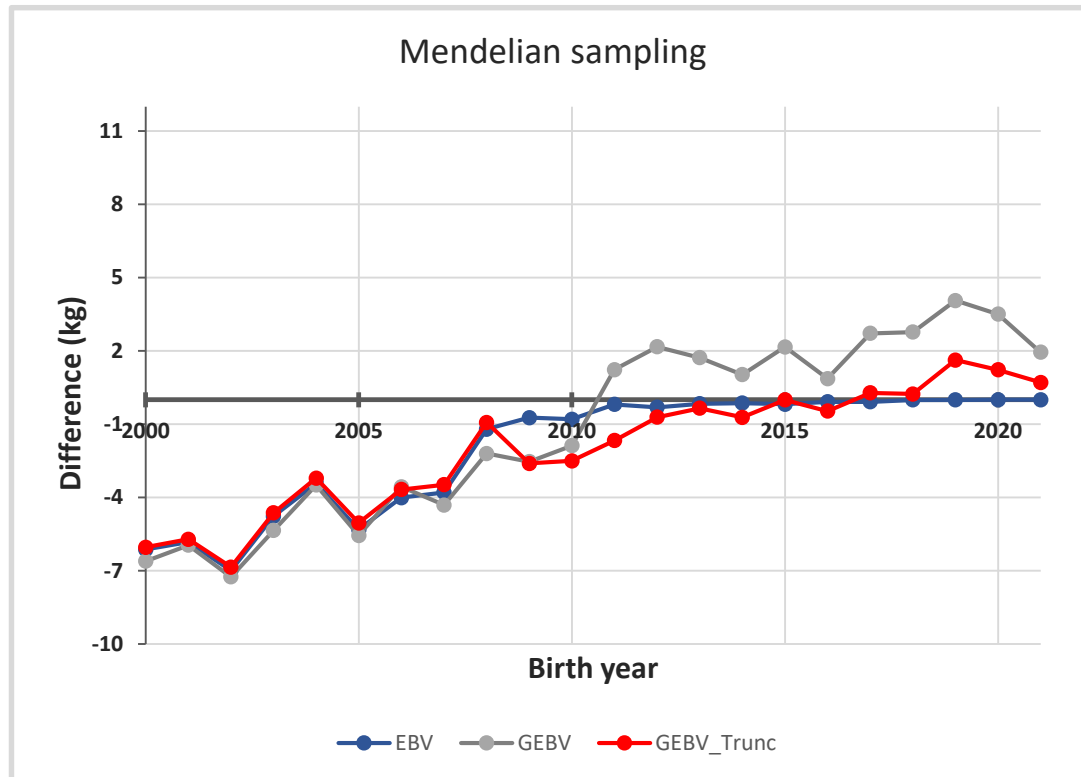


GEBV – ssGTBLUP with QP and w30 and AF 0.5
GEBV_Trunc – ssGTBLUP , w30 and AF 0.5, genomic data truncation
EBV – animal model

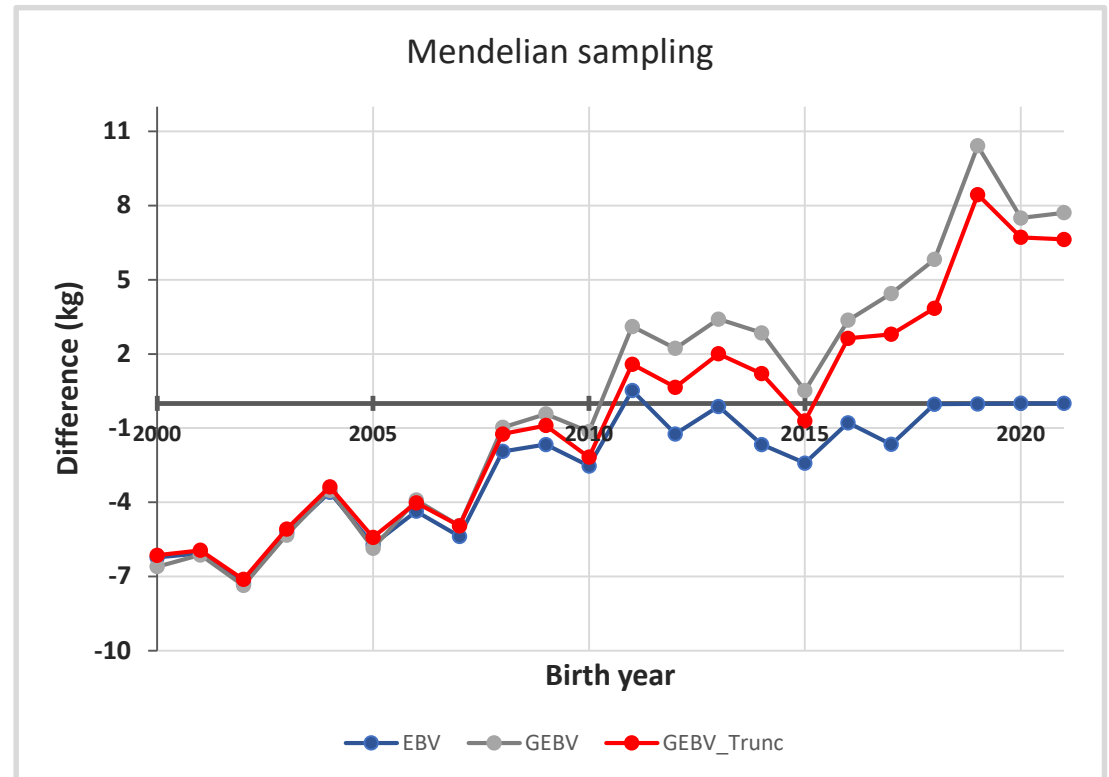
© Natural Resources Institute Finland

Average Mendelian Sampling terms for protein genotyped bulls – full TD data

All genotyped bulls



Selected bulls



GEBV – ssGTBLUP with QP and w30 and AF 0.5

GEBV_Trunc – ssGTBLUP, w30 and AF 0.5, genomic data truncation

EBV

Conclusions

- Removal of the genotypes of animals born before 2009 gave better validation results compared to full genomic data
 - The MS term averages were also closer to mean of zero
- For older and reliably evaluated bulls, correlation between GEBVs with full genomic data and truncated genomic data was high (bulls had own progeny)
- There was no considerable reranking of the candidate bulls
- Truncation of genomic data removed the over prediction of recent year classes of bulls and reduced the amount of overdispersion in candidate evaluations into a level acceptable in practice



Acknowledgements

Phenotypic and Genomic data
NAV Nordic Cattle Genetic Evaluation,



Support:
NAV, Viking Genetics

