NATURAL RESOURCES
INSTITUTE FINLAND

Easy implementation of QP transformation in ssGTBLUP

M. Koivula[1], I. Strandén[1], G. P. Aamand[2], E. A. Mäntysaari[1]

1 Natural Resources Institute Finland (Luke), FI-31600 Jokioinen, Finland
2 NAV Nordic Cattle Genetic Evaluation, 8200 Aarhus N, Denmark

Introduction

- We have learned that the genetic groups have significant effect on genetic trends, and, in the single-step genomic BLUP model, to convergence of iterative solving

- Genetic groups can be included in the evaluation model as birth year effects, or regression coefficients as unknown parent contributions

- Computationally more efficient approach is to re-express the parental genetic groups as unknown parent groups (UPG) resulting from QP transformation

Aim: to show options for easy implementation of QP transformation in single-step evaluations

Data

- Official Holstein Nordic TD evaluation data for milk, protein and fat
- Genomic data:

  - 178 177 genotyped animals

- Genotyped animals were associated qith 428 out of 446 genetic groups in the pedigree

FULL TD data

- 8.4 million animals with records, 10.4 million animals in pediree

REDUCED TD data for validation (four years of data reduction)

- 7.3 million animals with records

Methods

Genetic groups as QP transformation can be done alternative ways:

1. QP transformation for $\mathbf{A}^{-1}$ only
2. QP transformation for $\mathbf{A}^{-1}$ and $-\mathbf{A}_{22}^{-1}$
3. QP transformation for $\mathbf{A}^{-1}$ and $\mathbf{G}^{-1}$ and $-\mathbf{A}_{22}^{-1}$

Thus, we tested options;

ssGT_A =ssG**Ta**BLUP with QP in $\mathbf{A}^{-1}$ only
ssGT_AA$_{22}$ =ssG**Ta**BLUP with QP in $\mathbf{A}^{-1}$ and $-\mathbf{A}_{22}^{-1}$
ssGT_H =ssG**Ta**BLUP with QP in $\mathbf{A}^{-1}$ and $\mathbf{G}^{-1}$ and $-\mathbf{A}_{22}^{-1}$
ssG_H =ssGBLUP with QP in $\mathbf{A}^{-1}$ and $\mathbf{G}^{-1}$ and $-\mathbf{A}_{22}^{-1}$

☐ All options included 30% of polygenic effect
☐ ssGBLUP so called “original single-step model where $\mathbf{A}_{22}^{-1}$ and $\mathbf{G}^{-1}$ formed
☐ ssG**Ta**BLUP neither $\mathbf{A}_{22}^{-1}$ and $\mathbf{G}^{-1}$ are formed, and $\mathbf{G}^{-1}$ replaced by $\mathbf{C}^{-1}$-$\mathbf{T}'\mathbf{T}$

QP transformation

In original ssGBLUP

$$\boldsymbol{H}_*^{-1} = \begin{bmatrix} \boldsymbol{A}^{11} & \boldsymbol{A}^{12} & -\boldsymbol{A}^{11}\,\boldsymbol{Q}_1 + \boldsymbol{A}^{12}\,\boldsymbol{Q}_2 \\ \boldsymbol{A}^{21} & \boldsymbol{A}^{22} & -\boldsymbol{A}^{21}\,\boldsymbol{Q}_1 + \boldsymbol{A}^{22}\,\boldsymbol{Q}_2 \\ -(\boldsymbol{Q}'_1\,\boldsymbol{A}^{11} + \boldsymbol{Q}'_2\,\boldsymbol{A}^{21}) & -(\boldsymbol{Q}'_1\,\boldsymbol{A}^{12} + \boldsymbol{Q}'_2\,\boldsymbol{A}^{22}) & \boldsymbol{Q}' \quad \boldsymbol{A}^{-1}\boldsymbol{Q} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (\boldsymbol{G}_w^{-1} - \boldsymbol{A}_{22}^{-1}) & (-\boldsymbol{G}_w^{-1} - \boldsymbol{A}_{22}^{-1})\boldsymbol{Q}_2 \\ \mathbf{0} & -\boldsymbol{Q}'_2\boldsymbol{G}_w^{-1} - \boldsymbol{A}_{22}^{-1} & \boldsymbol{Q}'_2\boldsymbol{G}_w^{-1} - \boldsymbol{A}_{22}^{-1}\boldsymbol{Q}_2 \end{bmatrix}$$

The matrix parts involving the pedigree relationship matrices (**A** and $\mathbf{A}_{22}$) can be easily computed using pedigree information

However, this is impractical if we do ssGTBLUP

QP transformation

In practice, in ssGBLUP an augmented inverse genomic relationship matrix is computed

$$\mathbf{G}_*^{-1} = \begin{bmatrix} \mathbf{G}_w^{-1} & -\mathbf{G}_w^{-1}\mathbf{Q}_2 \\ -\mathbf{Q}_2'\mathbf{G}_w^{-1} & \mathbf{Q}_2'\mathbf{G}_w^{-1}\mathbf{Q}_2 \end{bmatrix}$$

which has elements for the genotyped animals and the genetic groups.

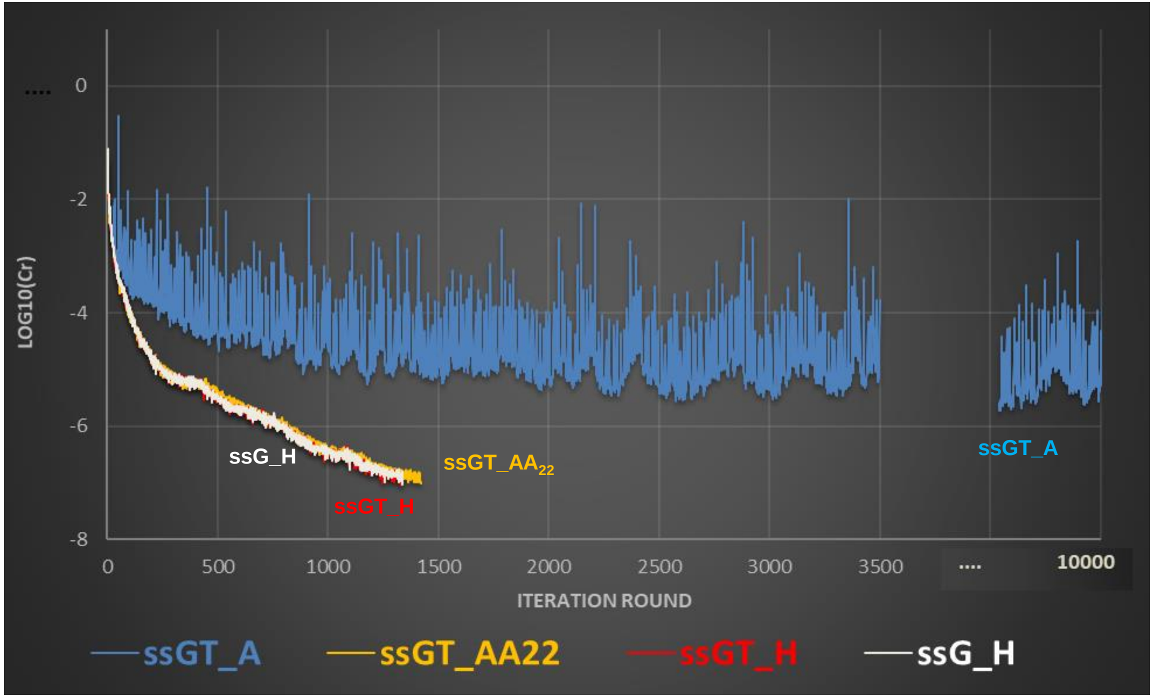
In ssGTBLUP, this can be done accordingly to relationship **T**-matrix by augmententing the new columns to original $\mathbf{T}_w$

Part for $\mathbf{A}_{22}$ can be done with an equivalent sparse formulation by reading the pedigree and including UPGs into the $\mathbf{A}_{22}$

Convergence statistic in PCG according to log10($Cr < 10^{-7}$)

Convergence criteria was $C_r<10^{-7}$ where Cr is a Euclidean norm of difference between the true right-hand side (RHS) of the MME and the one predicted by the current solutions Relative to the norm of the true RHS.

QP transformation is needed in $\mathbf{G}^{-1}$ and/or $-\mathbf{A}_{22}^{1}$ for good convergence of the single-step model



ssGT_A =ssGT**a**BLUP with QP in $\mathbf{A}^{-1}$ only
ssGT_AA$_{22}$ =ssGT**a**BLUP with QP in $\mathbf{A}^{-1}$ and $-\mathbf{A}_{22}^{-1}$
ssGT_H =ssGT**a**BLUP with QP in $\mathbf{A}^{-1}$ and $\mathbf{G}^{-1}$ and $-\mathbf{A}_{22}^{-1}$
ssG_H =ssGBLUP with QP in $\mathbf{A}^{-1}$ and $\mathbf{G}^{-1}$ and $-\mathbf{A}_{22}^{-1}$

Correlation between GEBVs in full data analyses

Regression of (G)EBV on PA or GEBV_red

Mean correlations by birthyear between GEBVs from different single-step methods

- Above diagonal genotyped reference bulls
- below diagonal genotyped validation bulls

	Trait	ssGT_A	ssGT_AA$_{22}$	ssGT_H	ssG_H
ssGT_A	Milk		0.98	0.98	0.98
	Protein		0.98	0.98	0.98
	Fat		0.99	0.99	0.99
ssGT_AA$_{22}$	Milk	0.95		1.00	1.00
	Protein	0.95		1.00	1.00
	Fat	0.95		1.00	1.00
ssGT_H	Milk	0.95	1.00		1.00
	Protein	0.94	1.00		1.00
	Fat	0.94	1.00		1.00
ssG_H	Milk	0.95	1.00	1.00	
	Protein	0.94	1.00	1.00	
	Fat	0.94	1.00	1.00	

ssGT_A =ssG**Ta**BLUP with QP in $\mathbf{A}^{-1}$ only
ssGT_AA$_{22}$ =ssG**Ta**BLUP with QP in $\mathbf{A}^{-1}$ and $-\mathbf{A}_{22}^{-1}$
ssGT_H =ssG**Ta**BLUP with QP in $\mathbf{A}^{-1}$ and $\mathbf{G}^{-1}$ and $-\mathbf{A}_{22}^{-1}$
ssG_H =ssGBLUP with QP in $\mathbf{A}^{-1}$ and $\mathbf{G}^{-1}$ and $-\mathbf{A}_{22}^{-1}$

Validation with GEBV Legarra-Reverter regression, 643 DFS bulls

Regression of (G)EBV on PA or GEBV_red

EBVs and GEBVs were standardized with mean of cows born 2007

Validation bulls have at least 20 daughters in the full data and zero daughters in the reduced data

DFS= Denmark, Finland, Sweden

	model	mean(Full-Red)	b_1	R^2
Milk	PA	-129.81	0.77	0.22
	ssGT_A	-459.69	0.61	0.48
	ssGT_AA_{22}	-364.58	0.86	0.64
	ssGT_H	-352.13	0.86	0.64
	ssG_H	--378.49	0.85	0.64
Protein	PA	-0.71	0.68	0.18
	ssGT_A	-17.60	0.43	0.37
	ssGT_AA_{22}	-13.38	0.79	0.58
	ssGT_H	-13.10	0.79	0.58
	ssG_H	-13.75	0.79	0.58
Fat	PA	-3.86	0.71	0.22
	ssGT_A	-21.83	0.54	0.47
	ssGT_AA_{22}	-16.63	0.79	0.61
	ssGT_H	-16.59	0.79	0.60
	ssG_H	-17.20	0.78	0.60

ssGT_A =ssG**Ta**BLUP with QP in $\mathbf{A}^{-1}$ only
ssGT_AA_{22} =ssG**Ta**BLUP with QP in $\mathbf{A}^{-1}$ and -$\mathbf{A}_{22}^{-1}$
ssGT_H =ssG**Ta**BLUP with QP in $\mathbf{A}^{-1}$ and $\mathbf{G}^{-1}$ and –$\mathbf{A}_{22}^{-1}$
ssG_H =ssGBLUP with QP in $\mathbf{A}^{-1}$ and $\mathbf{G}^{-1}$ and –$\mathbf{A}_{22}^{-1}$

Validation with GEBV Legarra-Reverter regression, 643 DFS bulls

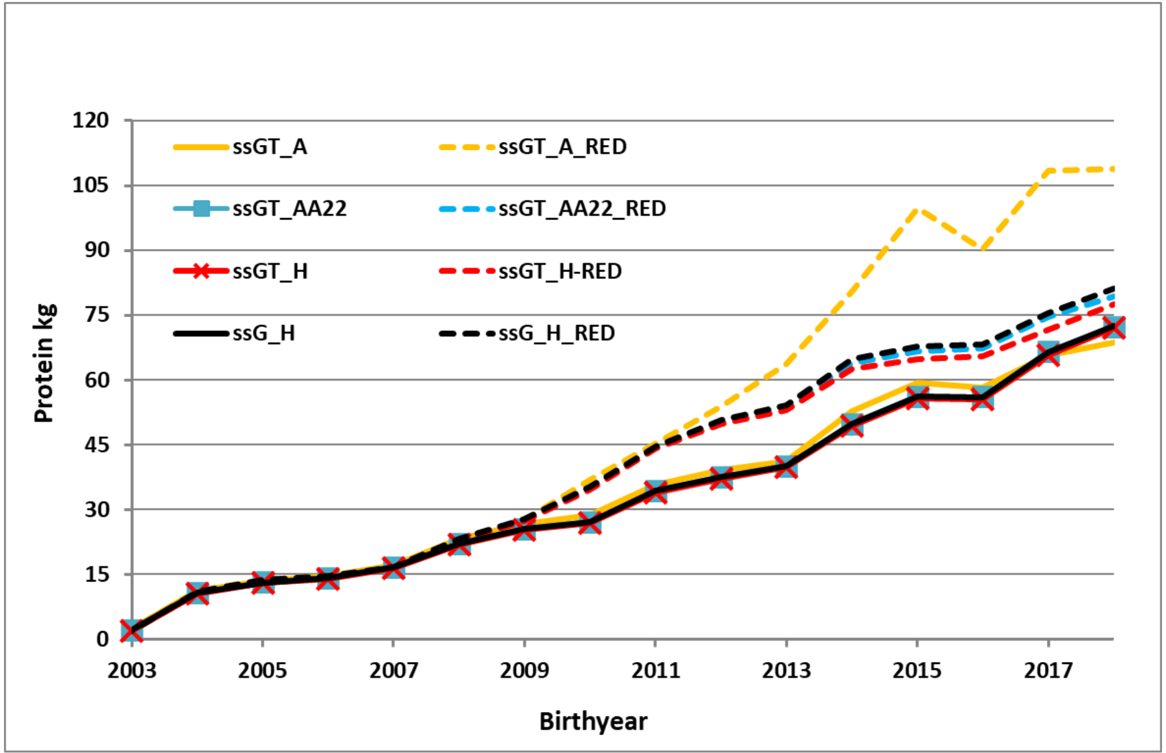
Regression of (G)EBV on PA or GEBV_red

	model	mean(Full-Red)	b_1	R^2
Milk	PA	-129.81	0.77	0.22
	ssGT_A	-459.69	0.61	0.48
	ssGT_AA_{22}	**-364.58**	**0.86**	**0.64**
	ssGT_H	**-352.13**	**0.86**	**0.64**
	ssG_H	**--378.49**	**0.85**	**0.64**
Protein	PA	-0.71	0.68	0.18
	ssGT_A	-17.60	0.43	0.37
	ssGT_AA_{22}	**-13.38**	**0.79**	**0.58**
	ssGT_H	**-13.10**	**0.79**	**0.58**
	ssG_H	**-13.75**	**0.79**	**0.58**
Fat	PA	-3.86	0.71	0.22
	ssGT_A	-21.83	0.54	0.47
	ssGT_AA_{22}	**-16.63**	**0.79**	**0.61**
	ssGT_H	**-16.59**	**0.79**	**0.60**
	ssG_H	**-17.20**	**0.78**	**0.60**

ssGT_AA22, ssGT_H and ssG_H give similar regression results,
but results from ssGT_A deviate from other because of poor convergence

ssGT_A	=ssG**Ta**BLUP with QP in $\mathbf{A}^{-1}$ only
ssGT_AA_{22}	=ssG**Ta**BLUP with QP in $\mathbf{A}^{-1}$ and -$\mathbf{A}_{22}^{-1}$
ssGT_H	=ssG**Ta**BLUP with QP in $\mathbf{A}^{-1}$ and $\mathbf{G}^{-1}$ and –$\mathbf{A}_{22}^{-1}$
ssG_H	=ssGBLUP with QP in $\mathbf{A}^{-1}$ and $\mathbf{G}^{-1}$ and –$\mathbf{A}_{22}^{-1}$

Genetic trend of protein (kg)

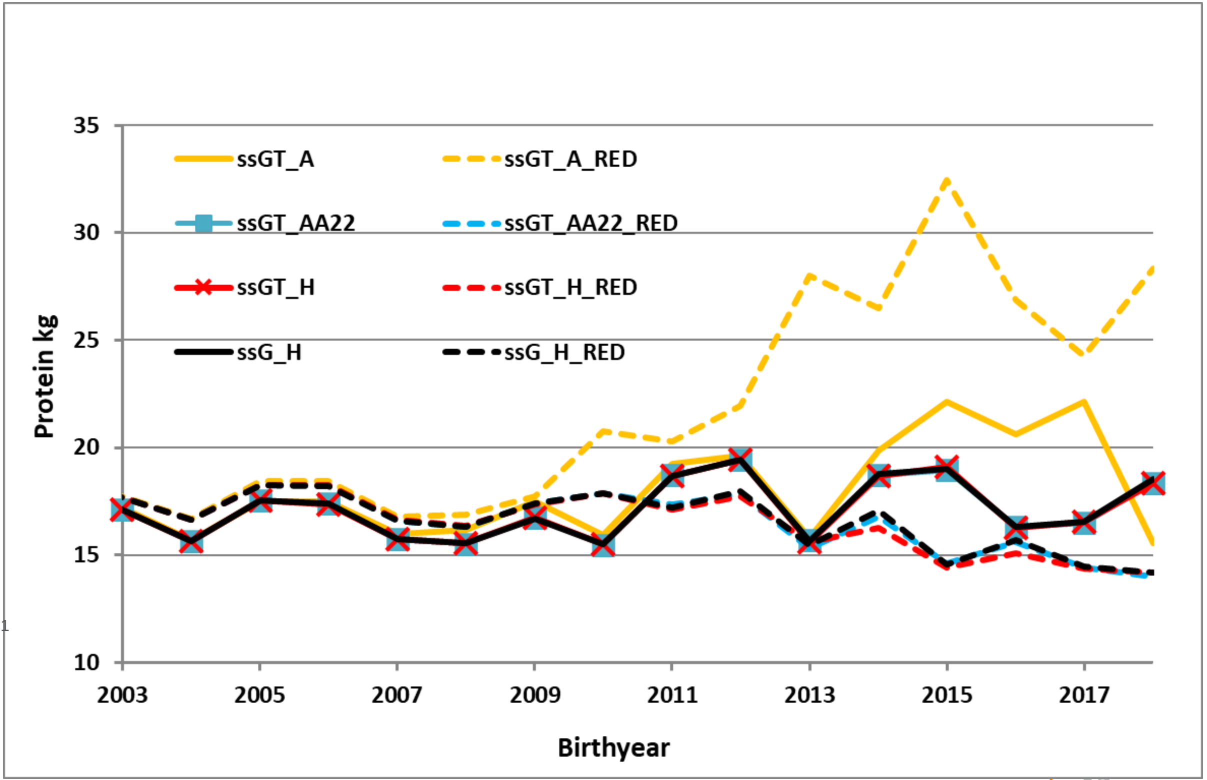


ssGT_A =ssGT**Ta**BLUP with QP in $\mathbf{A}^{-1}$ only
ssGT_AA$_{22}$ =ssGT**Ta**BLUP with QP in $\mathbf{A}^{-1}$ and -$\mathbf{A}_{22}^{-1}$
ssGT_H =ssGT**Ta**BLUP with QP in $\mathbf{A}^{-1}$ and $\mathbf{G}^{-1}$ and –$\mathbf{A}_{22}^{-1}$
ssG_H =ssGBLUP with QP in $\mathbf{A}^{-1}$ and $\mathbf{G}^{-1}$ and –$\mathbf{A}_{22}^{-1}$

*_RED reduced data runs

Genetic SD trend of protein (kg)

- Genetic trends, as well as standard deviations of the GEBVs by birthyear, were the same for methods ssGT_AA$_{22}$, ssGT_H and ssG_H
  - Trends for ssGT_A deviated from others both in full and reduced runs



ssGT_A	=ssGT**Ta**BLUP with QP in **A**$^{-1}$ only
ssGT_AA$_{22}$	=ssGT**Ta**BLUP with QP in **A**$^{-1}$ and -**A**$_{22}^{-1}$
ssGT_H	=ssGT**Ta**BLUP with QP in **A**$^{-1}$ and **G**$^{-1}$ and –**A**$_{22}^{-1}$
ssG_H	=ssGBLUP with QP in **A**$^{-1}$ and **G**$^{-1}$ and –**A**$_{22}^{-1}$

*_RED reduced data runs

Conclusions

1. QP transformation is needed in $\mathbf{A}^{-1}$ and $-\mathbf{A}_{22}^{-1}$ for a good convergence of the single-step model

2. Different methods to make QP transformation to genotyped animals give similar results
   - method with QP only to $\mathbf{A}^{-1}$ gives unreliable result (because model did not converge)

3. QP transformation is easy to implement also in the ssGTBLUP which with large genomic data is computationally efficient

Luke
NATURAL RESOURCES
INSTITUTE FINLAND

Luke

NATURAL RESOURCES
INSTITUTE FINLAND