IT Solutions for
Animal Production

Computing strategies for a single step SNP model with an across country reference population

Z. Liu, M. E. Goddard
F. Reinhardt and R. Reents

vit w.V., Heideweg 1, 27283 Verden, Germany
University of Melbourne, Australia

OUTLINE

- Introduction
  - Current multiple step genomic models
  - Single step GBLUP model
- A single step SNP model
- Computational issues of the SSS model
  - SNP effect estimation
  - Interim genomic evaluation without new phenotypes
- Technical issues for implementation in Holsteins
  - Test-day data of production traits
  - Integrating MACE phenotypes
- Discussion and Conclusions

Introduction: current genomic evaluation models

- Genomic model (Meuwissen et al. 2001) revolutionises animal breeding, particularly for Holsteins
- A multiple step genomic model for German Holsteins
  - Conventional bull proofs deregressed
    - National and international MACE evaluations
  - Deregressed bull proofs for SNP effect estimation with an across-country genomic reference population (EuroGenomics)
  - Male pedigree index excluded overestimated EBV of bull dams
  - DGV of candidates combined with conventional male pedigree index

- Advantages and drawbacks of the current multi-step genomic model
  - Simple for implementation
  - Genomic reference population customised as wished
    - Only progeny-tested bulls with a minimum EDC
    - No cows included due to possible overestimated EBV
  - Conventional EBV will be biased by genomic pre-selection

Introduction: single step genomic model

- Invention of $\mathbf{H}^{-1}$ matrix (Misztal et al., Christensen & Lund)
  - Accurate integration of genotyped animals into conventional evaluation

$$\mathbf{H} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{G}_{22} \end{bmatrix} = \mathbf{A} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{22} - \mathbf{A}_{22} \end{bmatrix}$$

$$\mathbf{H}^{-1} = \begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} \\ \mathbf{A}^{21} & \mathbf{G}_{22}^{-1} + \mathbf{A}^{22} - \mathbf{A}_{22}^{-1} \end{bmatrix} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{22}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}$$

- Computing strategies for large populations (Legarra & Ducrocq)
- Single step GBLUP model works perfectly for closed populations with all genotypes and phenotypes in one hand (Misztal et al.)
- Further developments for ‘open’ Holstein genomic evaluations
  - Using deregressed MACE EBV of foreign reference bulls
    - as substitutes of original phenotype data of foreign cows
  - Efficient interim genomic evaluations without new phenotypes
    - Instead of running the whole genomic evaluation

Reducing the impact of inflated EBV of genotyped cows

A single step SNP model (Goddard & Liu, 2012)

- A mixed linear model in a general form

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_p\mathbf{p} + \mathbf{Wu} + \mathbf{e} \qquad \text{var}(\mathbf{p}) = \mathbf{I}\sigma_p^2 \quad \text{var}(\mathbf{e}) = \mathbf{I}\sigma_e^2$$

- For genotyped animals (group 2)

$$\mathbf{u}_2 = \mathbf{Zg} + \mathbf{a}_2$$

- Distribution of SNP marker effects (BLUP or Bayesian models)
  - $\text{var}(\mathbf{g}) = \mathbf{B}\sigma_g^2$ (e.g. BLUP SNP model: $\mathbf{B} = b\mathbf{I} = \frac{1-k}{m}\mathbf{I}$)
- Residual polygenic effects of genotyped animals

$$\text{var}(\mathbf{a}_2) = \mathbf{A}_{22}k\sigma_g^2 \qquad \text{var}(\mathbf{u}_2) = \mathbf{G}_{22}\sigma_g^2 = (\mathbf{ZBZ}' + k\mathbf{A}_{22})\sigma_g^2$$

- Conditional distribution for non-genotyped animals (group 1)

$\mathbf{u}_1 = \mathbf{T}\mathbf{u}_2 + \mathbf{d}$ with transmission matrix $\mathbf{T} = \mathbf{A}_{12}\mathbf{A}_{22}^{-1}$

a deviation effect $\text{Var}(\mathbf{d}) = \mathbf{D}\sigma_g^2$

- Joint distribution for genotyped and non-genotyped animals

$$\mathbf{G} = \text{Var}(\mathbf{u}) = \text{Var}\begin{pmatrix}\mathbf{u}_1\\ \mathbf{u}_2\end{pmatrix} = \begin{bmatrix}\mathbf{T}\mathbf{G}_{22}\mathbf{T}' + \mathbf{D} & \mathbf{T}\mathbf{G}_{22}\\ \mathbf{G}_{22}\mathbf{T}' & \mathbf{G}_{22}\end{bmatrix}\sigma_g^2$$

SSS model (Goddard & Liu, 2012)

- Inverse of (co)variance matrix for **u**

$$\mathbf{G}^{-1} = \begin{bmatrix} \mathbf{D}^{-1} & -\mathbf{D}^{-1}\mathbf{T} \\ -\mathbf{T}'\mathbf{D}^{-1} & \mathbf{G}_{22}^{-1} + \mathbf{T}'\mathbf{D}\mathbf{T} \end{bmatrix} \sigma_g^{-2} \qquad \mathbf{G}^{-1} = \begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} \\ \mathbf{A}^{21} & \mathbf{G}_{22}^{-1} + \mathbf{A}^{22} - \mathbf{A}_{22}^{-1} \end{bmatrix} \sigma_g^{-2}$$

- Joint distribution of **u** and SNP effects

$$\mathbf{H} = \mathrm{var}\begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \mathbf{g} \end{bmatrix} = \begin{bmatrix} \mathbf{T}\mathbf{G}_{22}\mathbf{T}'+\mathbf{D} & \mathbf{T}\mathbf{G}_{22} & \mathbf{T}\mathbf{Z}\mathbf{B} \\ \mathbf{G}_{22}\mathbf{T}' & \mathbf{G}_{22} & \mathbf{Z}\mathbf{B} \\ \mathbf{B}\mathbf{Z}'\mathbf{T}' & \mathbf{B}\mathbf{Z}' & \mathbf{B} \end{bmatrix} \sigma_g^2$$

$$\mathbf{H}^{-1} = \begin{bmatrix} \mathbf{D}^{-1} & -\mathbf{D}^{-1}\mathbf{T} & \mathbf{0} \\ -\mathbf{T}'\mathbf{D}^{-1} & \frac{1}{k}\mathbf{A}_{22}^{-1} + \mathbf{T}'\mathbf{D}^{-1}\mathbf{T} & -\frac{1}{k}\mathbf{A}_{22}^{-1}\mathbf{Z} \\ \mathbf{0} & -\frac{1}{k}\mathbf{Z}'\mathbf{A}_{22}^{-1} & \mathbf{B}^{-1} + \frac{1}{k}\mathbf{Z}'\mathbf{A}_{22}^{-1}\mathbf{Z} \end{bmatrix} \sigma_g^{-2}$$

$$= \begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} & \mathbf{0} \\ \mathbf{A}^{21} & \mathbf{A}^{22} + (\frac{1}{k} - 1)\mathbf{A}_{22}^{-1} & -\frac{1}{k}\mathbf{A}_{22}^{-1}\mathbf{Z} \\ \mathbf{0} & -\frac{1}{k}\mathbf{Z}'\mathbf{A}_{22}^{-1} & \mathbf{B}^{-1} + \frac{1}{k}\mathbf{Z}'\mathbf{A}_{22}^{-1}\mathbf{Z} \end{bmatrix} \sigma_g^{-2}$$

SSS model (Goddard & Liu, 2012)

- Mixed model equations for all the effects

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z}_p & \mathbf{X}'\mathbf{W} \\ \mathbf{Z}_p'\mathbf{X} & \mathbf{Z}_p'\mathbf{Z}_p + \mathbf{I}\delta & \mathbf{Z}_p'\mathbf{W} \\ \mathbf{W}'\mathbf{X} & \mathbf{W}'\mathbf{Z}_p & \mathbf{W}'\mathbf{W} + \mathbf{H}^{-1}\sigma_e^2 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{p}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}_p'\mathbf{y} \\ \mathbf{W}'\mathbf{y} \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z}_p & \mathbf{X}'\mathbf{W}_1 & \mathbf{X}'\mathbf{W}_2 & \mathbf{0} \\ \mathbf{Z}_p'\mathbf{X} & \mathbf{Z}_p'\mathbf{Z}_p + \mathbf{I}\delta & \mathbf{Z}_p'\mathbf{W}_1 & \mathbf{Z}_p'\mathbf{W}_2 & \mathbf{0} \\ \mathbf{W}_1'\mathbf{X} & \mathbf{W}_1'\mathbf{Z}_p & \mathbf{W}_1'\mathbf{W}_1 + \lambda\mathbf{A}^{11} & \lambda\mathbf{A}^{12} & \mathbf{0} \\ \mathbf{W}_2'\mathbf{X} & \mathbf{W}_2'\mathbf{Z}_p & \lambda\mathbf{A}^{21} & \mathbf{W}_2'\mathbf{W}_2 + \lambda(\mathbf{A}^{22} + (\tfrac{1}{k} - 1)\mathbf{A}_{22}^{-1}) & -\tfrac{1}{k}\lambda\mathbf{A}_{22}^{-1}\mathbf{Z} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\tfrac{1}{k}\lambda\mathbf{Z}'\mathbf{A}_{22}^{-1} & \lambda(\mathbf{B}^{-1} + \tfrac{1}{k}\mathbf{Z}'\mathbf{A}_{22}^{-1}\mathbf{Z}) \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{p}} \\ \hat{\mathbf{u}}_1 \\ \hat{\mathbf{u}}_2 \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}_p'\mathbf{y} \\ \mathbf{W}_1'\mathbf{y} \\ \mathbf{W}_2'\mathbf{y} \\ \mathbf{0} \end{bmatrix}$$

- Solve two sets of equations iteratively:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z}_p & \mathbf{X}'\mathbf{W}_1 & \mathbf{X}'\mathbf{W}_2 \\ \mathbf{Z}_p'\mathbf{X} & \mathbf{Z}_p'\mathbf{Z}_p + \mathbf{I}\delta & \mathbf{Z}_p'\mathbf{W}_1 & \mathbf{Z}_p'\mathbf{W}_2 \\ \mathbf{W}_1'\mathbf{X} & \mathbf{W}_1'\mathbf{Z}_p & \mathbf{W}_1'\mathbf{W}_1 + \lambda\mathbf{A}^{11} & \lambda\mathbf{A}^{12} \\ \mathbf{W}_2'\mathbf{X} & \mathbf{W}_2'\mathbf{Z}_p & \lambda\mathbf{A}^{21} & \mathbf{W}_2'\mathbf{W}_2 + \lambda(\mathbf{A}^{22} + (\tfrac{1}{k} - 1)\mathbf{A}_{22}^{-1}) \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{p}} \\ \hat{\mathbf{u}}_1 \\ \hat{\mathbf{u}}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}_p'\mathbf{y} \\ \mathbf{W}_1'\mathbf{y} \\ \mathbf{W}_2'\mathbf{y} + \tfrac{1}{k}\lambda\mathbf{A}_{22}^{-1}\mathbf{Z}\hat{\mathbf{g}} \end{bmatrix}$$

$$\left(\mathbf{B}^{-1} + \tfrac{1}{k}\mathbf{Z}'\mathbf{A}_{22}^{-1}\mathbf{Z}\right)\hat{\mathbf{g}} = \tfrac{1}{k}\mathbf{Z}'\mathbf{A}_{22}^{-1}\hat{\mathbf{u}}_2$$

Computing strategies for the SSS model

- Re-arranging equation for $\mathbf{u}_2$

$$\mathbf{W}_2'\mathbf{X}\hat{\mathbf{b}} + \mathbf{W}_2'\mathbf{Z}_p\hat{\mathbf{p}} + \lambda\mathbf{A}^{21}\hat{\mathbf{u}}_1 + (\mathbf{W}_2'\mathbf{W}_2 + \lambda\mathbf{A}^{22})\hat{\mathbf{u}}_2 = \mathbf{W}_2'\mathbf{y} + \lambda\mathbf{A}_{22}^{-1}(\hat{\mathbf{u}}_2 - \tfrac{1}{k}\hat{\mathbf{a}}_2)$$

  - New part on top of conventional MME: pure genomic contribution

$$\mathbf{A}_{22}^{-1}(\hat{\mathbf{u}}_2 - \tfrac{1}{k}\hat{\mathbf{a}}_2) = \mathbf{A}_{22}^{-1}\hat{\mathbf{u}}_2^* = \hat{\varphi}$$

- Re-arranging SNP equations

$$\mathbf{B}^{-1}\hat{\mathbf{g}} = \tfrac{1}{k}\mathbf{Z}'\mathbf{A}_{22}^{-1}(\hat{\mathbf{u}}_2 - \mathbf{Z}\hat{\mathbf{g}}) = \tfrac{1}{k}\mathbf{Z}'\mathbf{A}_{22}^{-1}\hat{\mathbf{a}}_2 = \tfrac{1}{k}\mathbf{Z}'\hat{\gamma}$$

$$\hat{\mathbf{g}} = \tfrac{1}{k}\mathbf{B}\mathbf{Z}'\hat{\gamma}$$

- Two core calculations $\hat{\varphi} = \mathbf{A}_{22}^{-1}\hat{\mathbf{u}}_2^*$ and $\hat{\gamma} = \mathbf{A}_{22}^{-1}\hat{\mathbf{a}}_2$ can be done by solving equations (Legarra & Ducrocq, 2012) with Gauss-Jacobi (VanRaden)

$$\mathbf{A}_{22}\hat{\varphi} = \hat{\mathbf{u}}_2^* \qquad \mathbf{A}_{22}\hat{\gamma} = \hat{\mathbf{a}}_2$$

- A direct algorithm for computing φ and γ (Liu & Goddard)
  - Calculating ‘special’ EBV of non-genotyped relatives
  - Additional decomposing $\mathbf{A}^{-1*}$ (besides $\mathbf{A}^*$) using Colleau’s method
  - No setup of $\mathbf{A}_{22}$ or A_{22}^{-1} needed

Features of the SSS model

- A simple and closed form of $\mathbf{H}^{-1}$ (including SNP effects)
- No large matrix or product of large matrices in MME
- No need for genomic relationship matrix **G** or $\mathbf{G}^{-1}$ or $\mathbf{A}_{22}$ or A_{22}^{-1}
  - No limits on genotyped animals
- Suited for *Iteration on Data* technique for populations of any size
- Flexible SNP effect modelling: Bayesian or BLUP SNP models
  - One step Bayesian model (Goddard & Liu 2012)
- A residual polygenic effect in the SSS model
  - Analogue to SSGBlup using weighted **G** matrix: $\text{var}(\mathbf{u}_2) = \mathbf{G}_{22}\,\sigma_g^2 = (\mathbf{ZBZ}' + k\mathbf{A}_{22})\sigma_g^2$
  - Removed overestimation bias of genomic prediction (Liu et al. 2011)
  - Numerical equivalence: k=0.0001 as no residual polygenic effect
  - k=0.9999 as no SNP/DGV effects
  - RPG connects genotyped animals to phenotyped population
  - Removed large matrix multiplications (e.g. $\mathbf{Z'A}^{22}\mathbf{Z}$, $\mathbf{Z'A}^{21}$)
  - Residual polygenic variance estimated or determined via genomic validation
- Similar to Gengler's model (EAAP 2012) but with a different derivation

SNP effect estimation: reference population

- Single step GBLUP has no SNP effect estimation step and thus no direct control of information flow from reference pop. to candidates
- But genomic prediction can be improved, if RP is controlled by:
  - Removing bulls with limited data & less reliable EBV (with biases)
  - Deleting bull dams or cows with preferential treatments
  - Genotyped candidates without phenotypes or with imputed genotypes

  Animals of other breeds in a multi-breed evaluation
- Introduce a filter: **F** = *diag*{1, 0, 0, 1, …, 1, 0} to SNP equation:

$$\hat{\mathbf{g}} = \frac{1}{k}\mathbf{BZ}'\mathbf{A}_{22}^{-1}\hat{\mathbf{a}}_2 \longrightarrow \hat{\mathbf{g}} = \frac{1}{k}\mathbf{BZ}'\mathbf{FA}_{22}^{-1}\hat{\mathbf{a}}_2$$

  - 1 / 0: genotyped animal is included / excluded in reference population
  - 0.9 for imputed genotypes from a low density chip
- Meanwhile keep ALL animals (ref. or not) in **u**, $\mathbf{a}_2$ and $\mathbf{A}_{22}$

Impact of genomic pre-selection on **u** is not influenced by the selection of reference animals

Estimate SNP effects with a special algorithm

A 'large p and small n' computational problem

An efficient Gauss-Seidel algorithm with a special residual update (GSRU, Legarra & Misztal 2008)

- For a given set of $\mathbf{u}_2$ estimates:

$$\mathbf{Z}(\hat{\mathbf{g}}^{[j+1]} - \hat{\mathbf{g}}^{[j]}) = (\hat{\mathbf{u}}_2 - \hat{\mathbf{a}}_2^{[j+1]}) - (\hat{\mathbf{u}}_2 - \hat{\mathbf{a}}_2^{[j]}) = \hat{\mathbf{a}}_2^{[j]} - \hat{\mathbf{a}}_2^{[j+1]}$$

$$\hat{\mathbf{a}}_2^{[j+1]} = \hat{\mathbf{a}}_2^{[j]} - \mathbf{Z}(\hat{\mathbf{g}}^{[j+1]} - \hat{\mathbf{g}}^{[j]})$$

- An efficient estimation procedure:
  - At an outer iteration round calculate for ALL genotyped animals

$$\hat{\mathbf{a}}_2 = \hat{\mathbf{u}}_2 - \mathbf{Z}\hat{\mathbf{g}}$$

  - An inner loop (j-th round) for separating SNP from RPG effects
    - Step 1. estimate SNP effects

$$\hat{\mathbf{g}}^{[j+1]} = \tfrac{1}{k}\mathbf{B}\mathbf{Z}'\mathbf{F}\mathbf{A}_{22}^{-1}\hat{\mathbf{a}}_2^{[j]}$$

    - Step 2. update residual polygenic effects

$$\hat{\mathbf{a}}_2^{[j+1]} = \hat{\mathbf{a}}_2^{[j]} - \mathbf{Z}(\hat{\mathbf{g}}^{[j+1]} - \hat{\mathbf{g}}^{[j]})$$

Interim genomic evaluation w/o new phenotypes

- In contrast to conventional evaluation, genomic prediction is a more continuous process, monthly, weekly, or on-demand (just-in-time)
- Using SNP effect estimates can easily provide genomic evaluations as genotypes available any time between two major evaluations
- **Simple formulae for GEBV** instead of running the complete system
- Equation $\mathbf{u}_2$ is simplified ($\mathbf{y} = \mathbf{0}$ for candidates)

$$\mathbf{W}_2'\mathbf{X}\hat{\mathbf{b}} + \mathbf{W}_2'\mathbf{Z}_p\hat{\mathbf{p}} + \lambda\mathbf{A}^{21}\hat{\mathbf{u}}_1 + (\mathbf{W}_2'\mathbf{W}_2 + \lambda(\mathbf{A}^{22} + (\tfrac{1}{k} - 1)\mathbf{A}_{22}^{-1}))\hat{\mathbf{u}}_2 = \mathbf{W}_2'\mathbf{y} + \tfrac{1}{k}\lambda\mathbf{A}_{22}^{-1}\mathbf{Z}\hat{\mathbf{g}}$$

$$\mathbf{A}^{21}\hat{\mathbf{u}}_1 + \mathbf{A}^{22}\hat{\mathbf{u}}_2 - \mathbf{A}_{22}^{-1}(\hat{\mathbf{u}}_2 - \tfrac{1}{k}\hat{\mathbf{a}}_2) = \mathbf{0}$$

- A candidate (l) has sire (s) and dam (d)

$$u_l = \frac{A}{A+B}\mathbf{z}'\mathbf{g} + \frac{d_l}{A+B}\tfrac{1}{2}(u_s + u_d) + \frac{1}{A+B}\sum_{m=1}^{n} a_{22}^{lm}(u_m - \tfrac{1}{k}a_m)$$

SNP/DGV effects **Parental average** **Correction for genotyped relatives**

Interim genomic evaluation w/o new phenotypes

- A general formula for GEBV of candidate *l*

$$u_l = \frac{A}{A+B}\mathbf{z}'\mathbf{g} + \frac{d_l}{A+B}\tfrac{1}{2}(u_s + u_d) + \frac{1}{A+B}\sum_{m=1}^{n} a_{22}^{lm}(u_m - \tfrac{1}{k}a_m)$$

- If all ancestors (both parents) of the candidate are genotyped (appr.)

$$u_l = \mathbf{z}'\mathbf{g} + \frac{d_l^*}{1+d_l^*}\tfrac{1}{2}(a_s + a_d)$$

- Genotyped relatives influence the correction term:

$$\frac{1}{A+B}\sum_{m=1}^{n} a_{22}^{lm}(u_m - \tfrac{1}{k}a_m)$$

  - A good approximation using nucleus family: genotyped sire, dam/MGS, (direct progeny and mate) of the genotyped animal
- Alternative: select index method may be used to combine DGV and parental average as in case of multiple step genomic model

Assumption: contribution of new candidate genotypes to GEBV of genotyped population is negligible

Genomic and phenotypic data for DEU Holsteins

- Pedigree data for national and international evaluations
  - 76 million animals in **vit** database
  - 571,000 animals in Interbull Holstein bull pedigree
- Phenotypic test-day data (milk yield, August 2013 evaluation)
  - 19 mln DEU cows with test-day records (25 mln animals in pedigree)
  - 340 mln test-day records
- MACE phenotype (August 2013 MACE evaluation)
  - 133,028 Holstein bulls (representing 70 mln cows worldwide)
- Genotype data (45,613 SNPs selected from Illumina 50K v2)
  - 93,233 genotyped animals (278,000 animals in pedigree)
    - 6978 cows with test-day data
    - 26,361 Holstein bulls in EuroGenomics RP (c.a. 34 mln cows)
      - 18,497 bulls with only or more MACE info
- Combined (inter)national genotype, phenotype and pedigree data sets for genomic evaluation using the SSS model

Integration of MACE with test-day data

- A three-lactation random regression test-day model for each of three production traits for German Holsteins
- DEU random regression model uses Legendre polynomials with three terms: $u_i = t_1 * c_{1i} + t_2 * c_{2i} + t_3 * c_{3i}$
- Every animal has 3 lactations x 3 coefficients = 9 EBVs
- But: MACE phenotype is a single trait deregressed proof (DRP) on a combined lactation basis ($EBV_{comb}=w_1*u_1 + w_2*u_2 + w_3*u_3$)
- Majority of reference bulls (> 67% in case of DEU Holstein) have only one single DRP available for genomic evaluation
  - Little info for SNP effect in form of random regression coefficients
  - **SNP effect on a combined lactation basis**
- However, all animals, including candidates and foreign bulls without domestic daughters, have GEBV expressed in random regression coefficient form

SSS & test-day models: trait definition changes

- SNP effect estimation with a single trait model
  - Majority of reference bulls have a single MACE EBV
  - **Condense** 3 x 3 = 9 RRC per animal to 1 combined lactation EBV

$$\hat{\mathbf{g}} = \frac{1}{k}\mathbf{B}\mathbf{Z}'\mathbf{F}\hat{\boldsymbol{\gamma}} = \frac{1}{k}\mathbf{B}\mathbf{Z}'\mathbf{F}\mathbf{A}_{22}^{-1}\hat{\mathbf{a}}_2$$

- Adjusting conventional MME for genomic contribution
  - Conventional MME is expressed on a 3 x 3 = 9 RRC basis $\hat{u}_2 = \sum_{l=1}^{3}\sum_{m=1}^{3} t_m c_{ij}$
  - **Expand** $\mathbf{z}'\mathbf{g}$ (DGV) from one combined lactation to RRC basis

    $\delta = \mathbf{z}'\mathbf{g} / \hat{u}_2$ on a single trait (combined lactation) basis

    RRC for $\hat{\mathbf{u}}_2$: c_{ij}

    RRC for $DGV\,(\mathbf{z}'\mathbf{g}) = \delta\, c_{ij}$

Genotypes do not change lactation curve shapes, only curve areas

Lactation curve shapes determined only by phenotypes

Conclusions and Summary

- Our single step SNP model provides
  - Useful SNP effects estimates
  - Flexible SNP effect modelling (Bayes or BLUP, diagonal matrix **B**)
  - No need for **G** or $\mathbf{G}^{-1}$, no direct setup for $\mathbf{A}_{22}$ or $\mathbf{A}_{22}^{-1}$
  - Suited for iteration on data techniques because of no large matrices
- A residual polygenic effect for reducing prediction bias
- Unlike SSGblup, our SSS model has a SNP effect estimation step
  - Identical modelling GEBV as SSGblup
- Avoids bias in evaluations caused by genomic pre-selection
- Active control of genomic information flow from reference population to candidates by incl./excl. animals from reference population
- Simple formulae for frequent interim genomic evaluations
- Applicable for ‘open’ Holstein system with mixed (inter)national phenotypes, genotypes and pedigree

Acknowledgements

- Mike Goddard for deriving the single step SNP model
- Paul VanRaden and Vincent Ducrocq for discussions on the calculation of $\hat{\varphi} = \mathbf{A}_{22}^{-1}\hat{\mathbf{u}}_2^*$

THANK YOU!

vit

IT-Solutions for Animal Production

SSS model without RPG (Goddard & Liu, 2012)

- For genotyped animals $\hat{\mathbf{u}}_2 = \mathbf{Z}\hat{\mathbf{g}}$
- Mixed model equations for all the effects

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z}_p & \mathbf{X}'\mathbf{W}_1 & \mathbf{0} \\ \mathbf{Z}_p'\mathbf{X} & \mathbf{Z}_p'\mathbf{Z}_p + \mathbf{I}\delta & \mathbf{Z}_p'\mathbf{W}_1 & \mathbf{0} \\ \mathbf{W}_1'\mathbf{X} & \mathbf{W}_1'\mathbf{Z}_p & \mathbf{W}_1'\mathbf{W}_1 + \lambda\mathbf{A}^{11} & \lambda\mathbf{A}^{12}\mathbf{Z} \\ \mathbf{0} & \mathbf{0} & \lambda\mathbf{Z}'\mathbf{A}^{21} & \mathbf{Z}'\mathbf{W}_2'\mathbf{W}_2\mathbf{Z} + \lambda\mathbf{Z}'(\mathbf{A}^{22} - \mathbf{A}_{22}^{-1})\mathbf{Z} + \lambda\mathbf{B}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{p}} \\ \hat{\mathbf{u}}_1 \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}_p'\mathbf{y} \\ \mathbf{W}_1'\mathbf{y} \\ \mathbf{Z}'\mathbf{W}_2'\mathbf{y} \end{bmatrix}$$

- Solve two sets of equations iteratively:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z}_p & \mathbf{X}'\mathbf{W}_1 \\ \mathbf{Z}_p'\mathbf{X} & \mathbf{Z}_p'\mathbf{Z}_p + \mathbf{I}\delta & \mathbf{Z}_p'\mathbf{W}_1 \\ \mathbf{W}_1'\mathbf{X} & \mathbf{W}_1'\mathbf{Z}_p & \mathbf{W}_1'\mathbf{W}_1 + \lambda\mathbf{A}^{11} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{p}} \\ \hat{\mathbf{u}}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}_p'\mathbf{y} \\ \mathbf{W}_1'\mathbf{y} - \lambda\mathbf{A}^{12}\mathbf{Z}\hat{\mathbf{g}} \end{bmatrix}$$

$$(\mathbf{Z}'\mathbf{W}_2'\mathbf{W}_2\mathbf{Z} + \lambda\mathbf{Z}'(\mathbf{A}^{22} - \mathbf{A}_{22}^{-1})\mathbf{Z} + \lambda\mathbf{B}^{-1})\hat{\mathbf{g}} = \mathbf{Z}'\mathbf{W}_2'\mathbf{y} - \lambda\mathbf{Z}'\mathbf{A}^{21}\hat{\mathbf{u}}_1$$

- Similar to equations (Legarra & Ducrocq, 2012)

Comparing single step SNP & SSGblup models

- Legarra and Ducrocq (2012)
  - Genomic contribution to RHS of MME

$$\alpha_u \hat{\varphi} - \alpha_u \hat{\gamma} = \alpha_u \mathbf{A}_{22}^{-1} \hat{\mathbf{u}}_2 - \alpha_u \mathbf{G}^{-1} \hat{\mathbf{u}}_2 \quad [1]$$

  - SNP effect estimation (Strandén & Garrick 2009)

$$\hat{\mathbf{g}} = \mathbf{D}\mathbf{Z}'\mathbf{G}^{-1}\hat{\mathbf{u}}_2 \quad [2]$$

  - Another formulation using the terms by Goddard & Liu (2012)

$$\lambda\hat{\varphi} - \lambda\hat{\gamma} = \lambda\mathbf{A}_{22}^{-1}\hat{\mathbf{u}}_2 - \lambda\mathbf{G}_{22}^{-1}\hat{\mathbf{u}}_2 \quad [3]$$

$$\hat{\mathbf{g}} = \mathbf{B}\mathbf{Z}'\mathbf{G}_{22}^{-1}\hat{\mathbf{u}}_2 \quad [4]$$

- Corresponding equations by Goddard and Liu (2012)
  - Genomic contribution to RHS of MME

$$\lambda\mathbf{A}_{22}^{-1}\hat{\mathbf{u}}_2 - \tfrac{1}{k}\lambda\mathbf{A}_{22}^{-1}\hat{\mathbf{a}}_2 \quad [5]$$

  - SNP effect estimation

$$\hat{\mathbf{g}} = \tfrac{1}{k}\mathbf{B}\mathbf{Z}'\mathbf{A}_{22}^{-1}\hat{\mathbf{a}}_2 \quad [6]$$

Algorithm for calculating $\varphi_2 = \mathbf{A}_{22}^{-1}\mathbf{u}_2$

- Misztal et al. (2009) and Ducrocq et al. (2012) transformed the matrix inversion into an equation solving issue: $\mathbf{A}_{22}\,\varphi_2 = \mathbf{u}_2$
- Adding relatives of genotyped animals to the equation $\mathbf{u}_2$:

$$\begin{bmatrix} \mathbf{A}_{00} & \mathbf{A}_{02} \\ \mathbf{A}_{20} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \varphi_2 \end{bmatrix} = \begin{bmatrix} \mathbf{u}_0 \\ \mathbf{u}_2 \end{bmatrix} \longrightarrow \mathbf{A}\hat{\varphi} = \mathbf{u}$$

  - where $\begin{bmatrix} \mathbf{A}_{00} & \mathbf{A}_{02} \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \varphi_2 \end{bmatrix} = \mathbf{u}_0 \longrightarrow \mathbf{u}_0 = \mathbf{A}_{02}\varphi_2$

- Inverting the complete relationship matrix **A** (Mrode, 2005):

$$\hat{\varphi} = \mathbf{T}^{-T}(\mathbf{D}^{-1}\mathbf{T}^{-1}\mathbf{u})$$

- Solve the equations by reading pedigree twice (Colleau, 2002)

$$\hat{\boldsymbol{\eta}} = \mathbf{D}^{-1}\mathbf{T}^{-1}\mathbf{u}$$

$$\hat{\eta}_i = d^i\left[u_i - \tfrac{1}{2}(u_s + u_d)\right]$$

from oldest to youngest

$$\hat{\varphi} = \mathbf{T}^{-T}\hat{\boldsymbol{\eta}}$$

$$\hat{\varphi}_i = \hat{\varphi}_i + \hat{\eta}_i$$

$$\hat{\varphi}_s = \hat{\varphi}_s + (-\tfrac{1}{2})\hat{\eta}_i$$

$$\hat{\varphi}_d = \hat{\varphi}_d + (-\tfrac{1}{2})\hat{\eta}_i$$

from youngest to oldest

Algorithm for $\mathbf{u}_0$ for non-genotyped relatives

- Calculating $\mathbf{u}_0$ for non-genotyped relatives

$$\mathbf{u} = \begin{bmatrix} \mathbf{u}_0 \\ \mathbf{u}_2 \end{bmatrix} = \mathbf{A}\hat{\varphi} = \mathbf{TDT}'\hat{\varphi} = \mathbf{TDT}' \begin{bmatrix} 0 \\ \hat{\varphi}_2 \end{bmatrix}$$

- Solve equations by reading pedigree twice (Colleau, 2002)

$$\hat{\zeta} = \mathbf{T}'\hat{\varphi} \qquad \mathbf{u} = \mathbf{TD}\hat{\zeta}$$

$$\hat{\xi}_i = \hat{\xi}_i + \hat{\varphi}_i$$
$$\hat{\xi}_s = \hat{\xi}_s + \tfrac{1}{2}\hat{\xi}_i$$
$$\hat{\xi}_d = \hat{\xi}_d + \tfrac{1}{2}\hat{\xi}_i$$

from youngest to oldest

$$u_i = d_i\hat{\xi}_i + \tfrac{1}{2}(u_s + u_d)$$

from oldest to youngest

Algorithms for calculating $\varphi_2 = \mathbf{A}_{22}^{-1}\mathbf{u}_2$

- An iterative solving procedure (Misztal et al, Legarra & Ducrocq)

  $\mathbf{A}\hat{\varphi}^* = \mathbf{u}$ $\qquad\qquad$ $\mathbf{v} = \mathbf{TDT}'\tilde{\varphi}^*$

  - Gauss-Jacobi solving: $\varphi_j = ((\mathbf{u}_2 - \mathbf{v}_2)_j + diag(\mathbf{A}_{22})_j * \tilde{\varphi}_j) / diag(\mathbf{A}_{22})_j$
    (VanRaden, personal communication, 2012)
  - Straightforward, not using $\mathbf{A}^{-1}$.

- A direct solving procedure using the inverse relationship matrix
  - 1. choose starting values for genotyped animals, eg. $\tilde{\varphi}_2 = \mathbf{u}_2$
  - 2. calculate for non-genotyped relatives: $\mathbf{u}_0 = \mathbf{A}_{02}\varphi_2$
  - 3. estimate using $\mathbf{A}^{-1}$: $\hat{\phi} = \mathbf{T}^{-T}(\mathbf{D}^{-1}\mathbf{T}^{-1}\mathbf{u})$
  - 4. $\varphi_2^{[i]} = w\varphi_2^{[i]} + (1-w)\varphi_2^{[i-1]}$ e.g. w=0.5
  - 5. check convergence of $\hat{\varphi}_2$ only for genotyped animals
  - 6. if not converged, repeat steps 2 to 4 until converged

- The direct procedure may be optimised by estimating w.

Reliability approximation for SSS model

- Single step GBLUP model provides direct info than SSS model
  - Properly scaled genomic relationship matrix $\mathbf{G}_{22}$
- Genomic added value: $\varpi\left[\mathbf{G}_{22}^{-1} - \mathbf{A}_{22}^{-1}\right]$
  - Obtained from genomic validations
- Approximation method avoiding matrix inversion and even direct forming both matrices in core
- One unified reliability procedure for ALL groups of animals
- Theoretical genomic reliability is adjusted to realised genomic reliability
- Currently, most countries assume ONE single constant genomic EDC for pure genomic contribution
- Candidates with sires having no phenotype should have lower reliability than those with sires in reference population
- Reliability approximation also for interim genomic evaluation

Adjusting overestimated cow EBV

- Current multi-step genomic model uses DRP of bulls
- Overestimated bull dams have little impact on SNP effects
- Male pedigree index of candidates do not use bull dam EBV directly
- In single step genomic model, phenotypes of bull dams can no longer be excluded
  - Using filter **F** to exclude genotyped bull dams from SNP effect estimation
  - Impact of overestimated bull dam EBV still exists in candidate GEBV
- Inflated bull dams seem to exist only in production traits
- Strategy for adjusting possible inflated bull dam EBV
  - Identify cows or bull dams as potentially preferentially treated
  - Fit special fixed lactation curves to test-day data

Model differences between national and MACE

- MACE evaluation uses a single trait model on bull level
- DEU national evaluations apply multi-trait models on animal level
  - a three-lactation random regression test-day model (MFP+SCS)
  - a multiple trait fertility model
  - a three-parity animal model with direct and maternal effects (calving)
  - a multiple trait model for conformation and workability traits
  - a non-linear survival model for direct functional longevity
- Integrating (previous) MACE evaluation of ALL (foreign) bulls into single step genomic model is technically challenging
- Foreign bulls enter the SSS system as animals with records
- Modify least squares part of LHS and RHS of mixed model equations to integrate MACE phenotypic information
- Bulls in 3 groups with only national phenotypic information, only foreign data, and both

Three groups of bulls with different data sources

- Conventional phenotypic information per bull
  - EDC: φ^{NAT}, φ^{MACE},
  - Deregressed MACE proof: DRP
- Add one equation to mixed model equations per bull
- ONLY foreign daughter phenotypes are to be added to MME
- Group 1: bulls without daughters/progeny in home country
  - LHS = φ^{MACE}, RHS = φ^{MACE} * DRP (φ^{NAT}=0)
- Group 2: bulls with daughters/progeny ONLY in home country: add nothing (φ^{MACE} - φ^{NAT} = 0)
- Group 3: bulls with daughters/progeny in BOTH home and foreign countries
  - LHS = φ^{MACE} - φ^{NAT}, RHS = (φ^{MACE} - φ^{NAT})* DRP

Discussion: genomic pre-selection

- Evaluation bias due to genomic pre-selection can be avoided only when all culled candidates are also considered (Patry & Ducrocq)
- Equation $\mathbf{u}_2$ is simplified ($\mathbf{y} = \mathbf{0}$) for candidates

$$\mathbf{W}_2'\mathbf{X}\hat{\mathbf{b}} + \mathbf{W}_2'\mathbf{Z}_p\hat{\mathbf{p}} + \lambda\mathbf{A}^{21}\hat{\mathbf{u}}_1 + (\mathbf{W}_2'\mathbf{W}_2 + \lambda(\mathbf{A}^{22} + (\tfrac{1}{k} - 1)\mathbf{A}_{22}^{-1}))\hat{\mathbf{u}}_2 = \mathbf{W}_2'\mathbf{y} + \tfrac{1}{k}\lambda\mathbf{A}_{22}^{-1}\mathbf{Z}\hat{\mathbf{g}}$$

$$\downarrow$$

$$\mathbf{A}^{21}\hat{\mathbf{u}}_1 + \mathbf{A}^{22}\hat{\mathbf{u}}_2 - \mathbf{A}_{22}^{-1}\hat{\mathbf{u}}_2 + \tfrac{1}{k}\mathbf{A}_{22}^{-1}\hat{\mathbf{u}}_2 - \tfrac{1}{k}\mathbf{A}_{22}^{-1}\mathbf{Z}\hat{\mathbf{g}} = \mathbf{0}$$

$$\downarrow$$

$$\mathbf{A}^{21}\hat{\mathbf{u}}_1 + \mathbf{A}^{22}\hat{\mathbf{u}}_2 = \mathbf{A}_{22}^{-1}(\hat{\mathbf{u}}_2 - \tfrac{1}{k}\hat{\mathbf{a}}_2)$$

$$\downarrow$$

$$\mathbf{A}^{21}\hat{\mathbf{u}}_1 + \mathbf{A}^{22}\hat{\mathbf{u}}_2 = \mathbf{A}_{22}^{-1}\hat{\mathbf{u}}_2^* = \hat{\boldsymbol{\varphi}}$$

- Pure genomic contribution $\hat{\varphi}$ of the candidates is propagated to all relatives via pedigree
- In fact, this process for candidates applies also to genotyped sires of domestic cows with raw records
- But the candidates can be excluded from SNP effect estimation

Interim genomic evaluation w/o new phenotypes

- Monthly, weekly or on-demand (just-in-time) genomic evaluations between two major evaluations (no new phenotypes available)
- Equation $\mathbf{u}_2$ is simplified ($\mathbf{y} = \mathbf{0}$)

$$\mathbf{W}_2'\mathbf{X}\hat{\mathbf{b}} + \mathbf{W}_2'\mathbf{Z}_p\hat{\mathbf{p}} + \lambda\mathbf{A}^{21}\hat{\mathbf{u}}_1 + (\mathbf{W}_2'\mathbf{W}_2 + \lambda(\mathbf{A}^{22} + (\tfrac{1}{k} - 1)\mathbf{A}_{22}^{-1}))\hat{\mathbf{u}}_2 = \mathbf{W}_2'\mathbf{y} + \tfrac{1}{k}\lambda\mathbf{A}_{22}^{-1}\mathbf{Z}\hat{\mathbf{g}}$$

$$\mathbf{A}^{21}\hat{\mathbf{u}}_1 + \mathbf{A}^{22}\hat{\mathbf{u}}_2 - \mathbf{A}_{22}^{-1}\hat{\mathbf{u}}_2 + \tfrac{1}{k}\mathbf{A}_{22}^{-1}\hat{\mathbf{u}}_2 - \tfrac{1}{k}\mathbf{A}_{22}^{-1}\mathbf{Z}\hat{\mathbf{g}} = \mathbf{0}$$

$$\mathbf{A}^{21}\hat{\mathbf{u}}_1 + \mathbf{A}^{22}\hat{\mathbf{u}}_2 - \mathbf{A}_{22}^{-1}(\hat{\mathbf{u}}_2 - \tfrac{1}{k}\hat{\mathbf{a}}_2) = \mathbf{0}$$

- A candidate (l) has sire (s) and dam (d)

$$diag(\mathbf{A}^{22})_l = 1 + d_l \qquad 1/d_l = 0.5 - (f_s + f_d)/4 \qquad diag(\mathbf{A}_{22}^{-1})_l = 1 + d_l^*$$

$$[(d_l - d_l^*) + \tfrac{1}{k}(1 + d_l^*)]u_l - \tfrac{1}{2}d_l(u_s + u_d) - \tfrac{1}{k}(1 + d_l^*)\mathbf{z}'\mathbf{g} + \sum_{m=1}^{n} a_{22}^{lm}(-u_m + \tfrac{1}{k}a_m) = 0$$

$$[(d_l - d_l^*) + \tfrac{1}{k}(1 + d_l^*)]u_l = \tfrac{1}{k}(1 + d_l^*)\mathbf{z}'\mathbf{g} + \tfrac{1}{2}d_l(u_s + u_d) + \sum_{m=1}^{n} a_{22}^{lm}(u_m - \tfrac{1}{k}a_m) = 0$$

$$u_l = \frac{A}{A+B}\mathbf{z}'\mathbf{g} + \frac{d_l}{A+B}\tfrac{1}{2}(u_s + u_d) + \frac{1}{A+B}\sum_{m=1}^{n} a_{22}^{lm}(u_m - \tfrac{1}{k}a_m)$$

SNP/DGV effects **Parental average** **Correction for genotyped relatives**

(small variance)

$$A = \tfrac{1}{k}(1 + d_l^*) \qquad B = d_l - d_l^*$$

SSS & test-day models: solving strategy

- Foreign bulls without domestic daughters
  - Entering the SSS model as animals with own data
  - Only one single MACE EBV/DRP on a combined lactation basis
  - **Starting values** for random regression coefficients
    - $q_{comb} = q_1 = q_2 = q_3$
    - $q_i = t_1 * c_1 + t_2 * c_2 + t_3 * c_3 = t_1 * c_1$
    - $c_{1i} = q_{comb} / t_1$ and $c_{2i} = c_{3i} = 0$
- Procedures for solving MME of the SSS model
  - Iterate conventional MME for some rounds or using solution priors
  - Add MACE phenotypes of foreign bulls
    - Single trait model on combined lactation basis: DRP = μ + EBV + ε
  - Add genomic correction term of RHS

$$\mathbf{W}_2'\mathbf{X}\hat{\mathbf{b}} + \mathbf{W}_2'\mathbf{Z}_p\hat{\mathbf{p}} + \lambda\mathbf{A}^{21}\hat{\mathbf{u}}_1 + (\mathbf{W}_2'\mathbf{W}_2 + \lambda\mathbf{A}^{22})\hat{\mathbf{u}}_2 = \mathbf{W}_2'\mathbf{y} + \lambda\mathbf{A}_{22}^{-1}(\hat{\mathbf{u}}_2 - \tfrac{1}{k}\hat{\mathbf{a}}_2)$$

  - Convert $\mathbf{u}_2$ to SNP effects

Estimate SNP effects with a special algorithm

- A 'large p and small n' computational problem
- An efficient Gauss-Seidel algorithm with a special residual update (GSRU, Legarra & Misztal 2008)
- For a given set of $\mathbf{u}_2$ estimates:

$$\mathbf{Z}(\hat{\mathbf{g}}_i^{[j+1]} - \hat{\mathbf{g}}_i^{[j]}) = (\hat{\mathbf{u}}_2 - \hat{\mathbf{a}}_{2(+i)}^{[j]}) - (\hat{\mathbf{u}}_2 - \hat{\mathbf{a}}_{2(-i)}^{[j]}) = \hat{\mathbf{a}}_{2(-i)}^{[j]} - \hat{\mathbf{a}}_{2(+i)}^{[j]}$$

- Computing procedure:
  - At j-th iteration round calculate for ALL genotyped animals

$$\hat{\mathbf{a}}_2^{[j]} = \hat{\mathbf{u}}_2^{[j]} - \mathbf{Z}\hat{\mathbf{g}}^{[j]}$$

  - An inner loop over SNP (i =1,…,m) sorted by heterozygosity (D)
    - Step 1. $\hat{g}_i^{[j+1]} = \dfrac{1}{(\mathbf{B}^{-1})_{ii}} \frac{1}{k} \mathbf{Z}' \mathbf{A}_{22}^{-1} \hat{\mathbf{a}}_{2(-i)}^{[j]} = \dfrac{1}{(\mathbf{B}^{-1})_{ii}} \frac{1}{k} \mathbf{Z}' \hat{\boldsymbol{\gamma}}^{[j]}$
    - Step 2. update residual polygenic effects

$$\hat{\mathbf{a}}_{2(+i)}^{[j]} = \hat{\mathbf{a}}_{2(-i)}^{[j]} + \mathbf{z}_i(\hat{g}_i^{[j]} - \hat{g}_i^{[j+1]})$$