IT-Solutions for
Animal Production

A deregression method for single-step model using all genotype data

Zengting Liu and Yutaka Masuda
IT Solutions for Animal Production (**vit**),Germany
University of Georgia, USA

Generating pseudo-phenotype data via proof deregression

- EBV deregression is to undo the regression step in genetic evaluation (Jairath et al. 1998)
  - Non-iterative deregression on animal-by-animal basis: sub-optimal (Calus et al. 2016)
  - Deregression using pedigree (Jairath et al. 1998) for MACE evaluation since 1995
    - A reversibility study for validation (Madsen & Mark, 2002)
- Bull MACE EBV deregression on a country scale
  - EDC calculation using all national EDC, h^2 and country correlations (Liu, Interbull Workshop, Guelph, 2011)
  - Routinely used for German Holstein genomic evaluation since 2010
- Cow national EBV deregression for adding cows into genomic reference population
  - Routinely used in German Holstein since 2019
- Systematic validation of the deregressed proofs for all traits in German Holstein
  - For milk production and some conformation traits on cow EBV deregression (Liu, EuroGenetics, Sep 2016)
    - Cows with data, bulls with daughters, female ancestors
    - Reliability comparison to conventional evaluation
  - For all 60+ traits for national cows or 38 MACE traits for all bulls in MACE (**vit**, October 2020)

Genomic-free EBV from single-step evaluation for bull MACE evaluation

- Interbull working group meetings on various approaches
  - Mäntysaari, Liu both proposed using matrix $\mathbf{H}^{-1}$ for GEBV deregression (2019)
  - A concern about ignoring genotype data of 100,000s reference cows

- Interbull webinar in 11 Feb 2021 on Genomic-free EBV
  - Masuda independently proposed using $\mathbf{H}^{-1}$ for deregressing single-step GEBV
  - A presentation at this meeting (Masuda and Liu, 2021)

- Generating pseudo-phenotype data for conventional bull MACE evaluation
  - Important, because bull GEBV are not pure ‘phenotypic’

- Other use of deregressed GEBV as genomic-free pseudo-phenotype data
  - Deregressed cow proofs
  - Integration of foreign bull MACE data in national evaluation: adjusting deregressed proofs and EDC
  - Model checking and validation, etc.

Options for generating genomic-free EBV for bull MACE evaluation (I)

- Options proposed at Interbull webinar on 11 Feb 2021 (Sullivan 2021)

- 1) Keep running conventional besides single-step evaluation
- 2) Adjusting observations for fixed effects from single-step model and y-Xb_{ss}
  - Genetic trends between conventional and single-step models
  - An applicable, short-term solution with two evaluations

- 3) Calculating yield deviations of cows or daughter yield deviations of bulls
  - Perfect solution if the same trait definition used in both national and MACE evaluations
  - MACE trait definition for milk production traits is an index of lactation EBV on 305-day basis
  - Complicated by national evaluation models, e.g. random regression model
    - DYD or YD of test-day records to be converted to 305-day lactation
    - Problem of short lactations or missing later lactations
  - The direct-maternal effect model for calving evaluation
    - A single YD / DYD for the two direct and maternal effects
    - Separate MACE evaluation for direct or maternal effect (no multi-effect MACE evaluation)

Options for generating genomic-free EBV for bull MACE evaluation (II)

- Options proposed at Interbull webinar on 11 Feb 2021 (Sullivan 2021)
- 4) Deregressing genomic EBV from single-step model
  - Straightforward for the same trait definition in MACE and national evaluation: e.g. conformation
  - In case of different trait definitions in MACE and national evaluations
    - Deregression using the trait definition for MACE, e.g. an index of lactation EBVs for milk yield
    - Bypassing the complex national model, such as a multi-variate random regression model
  - For calving traits with direct and maternal effects
    - Deregressing direct or maternal effect separately

Deregressing single-step GEBV using a special single-step SNP BLUP model

- A national single-step evaluation without bull MACE data
  - Based on a single-step SNP BLUP model: SNP effects available
  - Based on a single-step GBLUP model → back-solve SNP effects
- Selection of animals in the GEBV deregression process
  - All genotyped animals including **young candidates** AND
  - All animals with own phenotype data: cows with records / bulls with daughters
  - GEBV of genotyped or phenotyped animals as input data and conventional EDC/ERC as weights
- Using **the same** genotype and pedigree data from the single-step evaluation
- A single-step SNP BLUP model (Liu-Goddard model) for GEBV deregression
  - No matrix like $\mathbf{H}^{-1}$ to be set up, no approximation for large genotyped population needed
  - Feasible for millions of genotyped animals with 2-bit genotype representation (Vandenplas et al. 2019)
  - 'Application of a single-step SNP BLUP model to conformation traits of German Holstein' (Alkhoder & Liu, at this meeting)

A single-step SNP BLUP model (Liu-Goddard) for GEBV deregression

- The GEBV dergression model

$$\boldsymbol{y} = \mu\mathbf{1} + \boldsymbol{u} + \boldsymbol{e}$$

with residuals:

$$[var(\boldsymbol{e})]^{-1} = \mathbf{D}\sigma_e^{-2} = diag\{n_i\}\,\sigma_e^{-2}$$

- GEBV of genotyped animals

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

- Residual polygenic effect

$$\mathbf{a}_2 \sim N(\mathbf{0}, k\sigma_u^2\mathbf{A}_{22})$$

- SNP effects

$$\mathbf{g} \sim N(\mathbf{0}, (1-k)\sigma_u^2\mathbf{B})$$

with matrix **B**

$$\mathbf{B} = \left(\sum_{j=1}^{m} 2p_j(1-p_j)\right)^{-1}\mathbf{I}$$

- Variance ratio

$$\lambda = \sigma_e^2/\sigma_u^2$$

Deregressing single-step GEBV by solving the mixed model equations

$$
\begin{bmatrix}
\sum_{i=1}^{n} n_i & \mathbf{0} & \mathbf{0} & n_{11} \quad n_{21} \quad \cdots \quad n_{(n-n_g)1} & n_{21} \quad n_{22} \quad \cdots \quad n_{n_g 2} & \mathbf{0} \\
 & \lambda\mathbf{A}^{tt} + (\frac{1}{k} - 1)\mathbf{Q}_2{}'\mathbf{A}_{22}^{-1}\mathbf{Q}_2 & \lambda\mathbf{A}^{t0} & \lambda\mathbf{A}^{t1} & \lambda\mathbf{A}^{t2} - (\frac{1}{k} - 1)\mathbf{Q}_2{}'\mathbf{A}_{22}^{-1} & \frac{1}{k}\mathbf{Q}_2{}'\mathbf{A}_{22}^{-1}\mathbf{Z} \\
 & & \lambda\mathbf{A}^{00} & \lambda\mathbf{A}^{01} & \lambda\mathbf{A}^{02} & \mathbf{0} \\
 & & & diag\{n_{i1}\} + \lambda\mathbf{A}^{11} & \lambda\mathbf{A}^{12} & \mathbf{0} \\
 & symm. & & & diag\{n_{i2}\} + \lambda(\mathbf{A}^{22} + (\frac{1}{k} - 1)\mathbf{A}_{22}^{-1}) & -\lambda\frac{1}{k}\mathbf{A}_{22}^{-1}\mathbf{Z} \\
 & & & & & \lambda(\frac{1}{1-k}\mathbf{B}^{-1} + \frac{1}{k}\mathbf{Z}'\mathbf{A}_{22}^{-1}\mathbf{Z})
\end{bmatrix}
$$

gen. groups

ancestors

SNP effects

general mean

$$
\mathrm{X}
\begin{bmatrix}
\hat{\mu} \\ \hat{\mathbf{t}} \\ \hat{\mathbf{u}}_0 \\ \hat{\mathbf{u}}_1 \\ \hat{\mathbf{u}}_2 \\ \hat{\mathbf{g}}
\end{bmatrix}
=
\begin{bmatrix}
\sum_{i=1}^{n} n_i y_i \\ \mathbf{0} \\ \mathbf{0} \\ \{n_{i1} y_i\} \\ \{n_{i2} y_i\} \\ \mathbf{0}
\end{bmatrix}
$$

non-genotyped animals

genotyped animals

conventional EDC

Deregressed GEBV for animals with phenotype data

- For non-genotyped animals
  - RHS of its equation

$$\Delta_1 = \{n_1\}\hat{\mu} + \lambda\mathbf{A}^{1\mathrm{t}}\hat{\mathbf{t}} + \lambda\mathbf{A}^{10}\hat{\mathbf{u}}_0 + (diag\{n_1\} + \lambda\mathbf{A}^{11})\hat{\mathbf{u}}_1 + \lambda\mathbf{A}^{12}\hat{\mathbf{u}}_2$$

  - Deregressed proof for animal i

$$\hat{y}_{i1} = \Delta_{i1}/n_{i1}$$

- For genotyped animals
  - RHS of its equation

$$\Delta_2 = \{n_2\}\hat{\mu} + \lambda(\mathbf{A}^{2\mathrm{t}} - \left(\tfrac{1}{k} - 1\right)\mathbf{A}_{22}^{-1}\mathbf{Q}_2)\hat{\mathbf{t}} + \lambda\mathbf{A}^{20}\hat{\mathbf{u}}_0 + \lambda\mathbf{A}^{21}\hat{\mathbf{u}}_1 + \left(diag\{n_2\} + \lambda(\mathbf{A}^{22} + \left(\tfrac{1}{k} - 1\right)\mathbf{A}_{22}^{-1}\right)\hat{\mathbf{u}}_2 - \lambda\frac{1}{k}\mathbf{A}_{22}^{-1}\mathbf{Z}\hat{\mathbf{g}}$$

  - Deregressed proof for animal i

$$\hat{y}_{i2} = \Delta_{i2}/n_{i2}$$

A validation study for GEBV deregression (reversibility test)

- A reversibility test for deregressed GEBV for cows with own records (Case 1)
  - A special single-step evaluation using deregressed GEBV of all the cows as phenotypes
    - With same genotype and pedigree data and SNP effect estimates
  - Resulting GEBV of the cows must equal the original **national single-step** GEBV

- A reversibility test for deregressed GEBV for bulls with daughters (Case 2)
  - A special single-step evaluation using deregressed GEBV of all bulls and reference cows as phenotypes
    - With same genotype and pedigree data and SNP effect estimates
  - Resulting GEBV of the bulls must equal the original **national single-step** GEBV

Discussion issues

- GEBV deregression leading to snowballing of genomic evaluations?
  - Avoidable, if the GEBV deregression uses only national phenotype data (excluding bull MACE data)
  - Using accurate GEBV deregression methods
    - An exact reverse engineering of the single-step evaluation
    - Passing the validation test

- GEBV deregression by national genetic evaluation centres due to genotype data
  - In contrast, conventional bull EBV deregression by Interbull

- A common software is required for all NGECs
  - Coordination of R&D projects
  - Similar to Interbull's SNPMace project

Summary and conclusions

- Generating genomic-free EBV for bull MACE evaluation
  - Short-term solutions: run 2 evaluations or using adjusted records y-Xb for the second evaluation
    - More workload, resulting in lower GEBV in production traits (Mäntysaari)
  - (Daughter) yield deviation calculation too difficult for complex statistical models

- GEBV deregression provides an accurate and efficient alternative for generating pseudo-phenotypes
  - Using information from national single-step evaluation, no extra data needed
  - A post-evaluation step to be done by each NGEC
  - Circumvent complex national models, like random regression or maternal effects models
  - Using deregressed GEBV will lead to equal GEBV as from the single-step evaluation (reversibility test)

- The proposed GEBV deregression method with a single-step SNP BLUP model
  - Feasible for millions of genotyped animals

IT-Solutions for Animal Production