IT Solutions for
Animal Production

Experiences with the Interbull genomic reliability method using a mixed reference population

Zengting Liu, H. Alkhoder and F. Reinhardt
IT Solutions for Animal Production (**vit**), Germany

Introduction

- Interbull working group Genomic Reliability Calculation (GREL)
  - A standardized statistical method developed (Liu et al. 2017)
  - Issues on applicability for single-step genomic models and double counting cows phenotypes in case of a mixed reference population

- Fourth video conference of GREL WG on 03 May 2018 after WCGALP
  - A supplementary document (distributed on 05 June 2018 by Interbull)
    - Definition of genomic reference population for single-step models
    - EDC adjustment for reference bulls with reference daughters
      - Bulls representing non-genotyped daughters
      - Avoid double counting of data contribution by reference daughters

- New options in ***snp_blup_rel*** (Strandén, LUKE, Finland)
  - More compact formats, no space & in Fortran style, for genotype data
  - Compact binary and faster format for the inverse matrix of LHS of MME
  - Process genotyped animals in groups to allow any # genotyped animals

Data materials of DEU HOL with a mixed reference pop.

- Phenotype data from April 2018 national and MACE evaluations
  - National cows with deregressed EBV and ERC
    - a test-day model for production trait (***mil***) with 21,073,615 Holstein cows
  - Bulls with MACE or national deregressed EBV and EDC
    - 152,560 Holstein bulls worldwide
  - 43 regular traits + 13 direct health traits

- Genotype and pedigree data from July 2018 monthly genomic evaluation
  - 489,255 genotyped Holstein animals
  - 45,613 SNP markers
  - 36, 892 reference bulls + 90,737 reference cows = 127,629 for trait ***mil***

- Genotype & phenotype data as if stemming from a single-step evaluation
  - 21,453,082 genotyped animals or phenotyped cows (Holsteins)
  - 27,230,923 animals in pedigree

Data materials for a genomic validation study

- The majority of reference cows born in the recent 3-4 years
- Data truncation for a genomic validation
  - Reference bulls from April 2016 evaluation
  - Reference cows born in the recent 2 birth years due to the short history of cow genotyping
- Genomic reference population (trait ***mil***) for the truncated validation
  - Reference cows: 44,891 (full evaluation: 90,737)
  - Reference bulls: 33,683 (full evaluation: 36,892)
- Interbull GEBV Test performed and passed with the mixed reference pop.
  - Based on validation bulls only

EDC adjustment for bulls for a mixed reference pop.

- Cow's own data contribution to herself
  - Reliability $R_i(o)$ $= h^2$ with complete data information
  - $\varphi_i = \lambda R_i(o)/(1 - R_i(o))$
- Cow's data contribution to her sire
  - $R_i(o)/(4 - R_i(o)\ R_{dam}(o))$ $= R_i(o)\ /\ 4$ if dam has no data or is missing
  - ${\varphi_i}^{d} = \lambda\ R_i(o)/(4 - R_i(o)\ [1 + R_{dam}(o)])$
- Bull's EDC from two daughter types: reference and non-reference cows
  - $\varphi_s = \sum_{i=1}^{n_s} {\varphi_i}^{d} = \sum_{i=1}^{n_g} {\varphi_i}^{d} + \sum_{i=1}^{n_n} {\varphi_i}^{d}$
- Adjust bull's EDC for his daughters in the same reference population
  - ${\varphi_s}^{mod} = \varphi_s - \sum_{i=1}^{n_g} {\varphi_i}^{d}$ for national data only, or
  - ${\varphi_s}^{mod} = {\varphi_s}^{MACE} - \sum_{i=1}^{n_g} {\varphi_i}^{d}$ for MACE data

Results and Discussion

- Computing resources by Step 1 of the GREL method (via ***snp_blup_rel***)
  - Using option –memlow 100000
  - Size of the inverse matrix reduced 7.8 Gb from 25 Gb for 45,613 SNP markers
  - Milk yield (***mil***) with 127,629 reference animals:
    - → Sub-step 1: Inverting LHS of MME for reference population

    Peak RAM 87 Gb, 73 minutes total time with 27 cores (Intel(R) Xeon(R) CPU E5-2690 v4 @ 2.60GHz)
    - → Sub-Step 2: Calculating SNP reliabilities for new animals using

    19 cores of 1 server Intel(R) Xeon(R) CPU E5-2690 v2 @ 3.00GHz

# newly genotyped animals		Peak RAM (Gb)	Total clock time (min.)
ALL	489,225	85	55
May-Jun 2018	41,189	45	10
Weekly	10,000	24	3.5
Weekly	5,000	20	1.5
Weekly	3,000	19	1.2

Results and Discussion

- Computing resources by Step 2 to 6
  - Run on a single core, many traits run on different cores in parallel
  - Step 5 of propagation of genomic information to non-genotyped relatives: most time consuming:
    - Tracing ancestors and coding all 27.2 mln animals (***mil***): 10 minutes
    - Calculating propagated GREL: 4 minutes
  - The propagation step is NOT needed for tightly timed weekly evaluation

- Total time required by all the steps is acceptable for weekly genomic evaluation

Results and Discussion

- Correlation between the new and current GREL methods (trait: mil)
  - Reference animals: 0.702
  - Genotyped candidates: 0.630 (males: 0.715, females: 0.548)
- The new GREL method gave 3.5% higher for validation bulls, lower heritability traits increased more
- Large GREL differences between reference and candidate animals for traits with a high residual polygenic variance
- Reference cows (mostly genotyped with LD chips) have high GREL values, as imputation accuracy is > 0.99
  - Special data structure caused by the short history of cow genotyping
- Not all results are validated for all the traits yet, specifically related to the GREL after the adjustment procedure

Further developments

- Software ***snp_blup_rel***: feasible for weekly or daily genomic evaluation
  - For processing all traits, read (very large) genotype data only once
- All the later steps may be made faster by using multiple cores like Step 1
- Adjusted genomic reliability versus validation R^2 values
- GREL of candidates decrease with increasing distance to reference pop.
  - More discounted for second or third generation candidates?

Conclusions

- Software ***snp_blup_rel***: feasible for weekly or daily genomic evaluation
  - No approximation is needed, given the **vit** computing resources
  - Propagation to non-genotyped relatives may not be needed for tightly timed frequent genomic evaluation
- Double counting of the data contribution of reference cows in a mixed reference population is avoided
- The standardized GREL method clarified the definition of reference animals for single-step genomic models
- Overall, the standardized GREL method seems to work efficiently and results make sense
- Further investigation and testing are needed

Country feedbacks

- DEA Bavaria (see presentation by Malena Erbe)
  - Application to a single-step model evaluation
- NLD genetic evaluation unit (Herwin Eding)
  - Application to the NLD genomic evaluation
- Spain genetic evaluation centre
  - Noureddine Charfeddine Saeid Naderi
- Italian ANAPRI breeders association
  - Alberto Cesarani on use of ***snp_blup_rel***
- **vit**, German genetic evaluation centre

Acknowledgements

- LUKE of Finland (Ismo Strandén, Esa Mantysaari and Martin Lidauer)
  - New features or options
- National genetic evaluation centres for testing the software snp_blup_rel and Interbull standardized GREL method
  - Feedbacks from more countries are needed
- Interbull genomic reliability calculation working group
  - P. M. VanRaden, M. H. Lidauer, M. P. Calus, H. Benhajali, H. Jorjani & Z. Liu

IT Solutions for Animal Production