CDCB™
COUNCIL ON DAIRY CATTLE BREEDING

Lazy computation of Residual Polygenic effects

A Legarra[1,2], S Savoia[1], G Jansen[1], M Bermann[2], D Lourenco [2]
Council on Dairy Cattle Breeding, Bowie, MD[1] University of Georgia, Athens, GA[2]

Interbull meeting, June 2025

Gerald

- This is maybe the last "appearance" of Gerald Jansen in an Interbull meeting
- He is enjoying retirement now…
- He worked at U Guelph, Anafi, and then as consultant for Anafi, Interbull and CDCB (probably other places too)
- LOTS of what we do in this room use Gerald's programs or ideas
- I found great pleasure working with him, and I learnt a lot

Residual polygenic effects

- Single step methods use a blended relationship matrix

$$G^* = (1-\alpha)\frac{ZZ'}{k} + \alpha A_{22}$$

- α (typically 5%-20%) represents "relationships in pedigree but not in markers"
- "blending" protects from overdispersion / bias
- As a result, the final GEBV is composed of
  - a fraction due to "purely genomics" (often called DGV)
  - a fraction due to "pedigree"(RPG, Residual Polygenic Effect)

RPGs in Indirect Predictions

- "Nominators" send "young animals" (no records or progeny with records) DNA readings and pedigree to National Centers like CDCB
- National Centers often send back (initially) a so-called "Indirect Prediction"
  - IP = DGV (sum of SNP effects) + RPG
  - RPG[new_animal]=0.5*(RPG[sire[new_animal]]+RPG[dam[new_animal]])
  - In other words, RPG[new_animal]=PA(RPGs)
  - back in the pedigree, the RPGs eventually come from ssGBLUP equations
- The algebra of DGV and RPG can be found in
  - AL (that's me) et al. notes "Bases of Genomic Prediction"
  - Vandenplas et al. Genetics Selection Evolution(2023) 55:37 https://doi.org/10.1186/s12711-023-00808-z (very complete and detailed)

Algorithms to get RPGs from ssGBLUP

- get RPGs = (GEBVs-DGVs) for genotyped animals in the MME

$$\hat{d}_g = \hat{u}_g - Z\hat{g}$$

- what is the RPG for a non genotyped animal [or a metafounder]?
- solve for non-genotyped animals (+metafounders) using selection index

$$\hat{d}_n = A_{n,g} A_{g,g}^{-1} \hat{d}_g$$

- or similar alternative equations. In all cases this involves big pedigrees, Colleau's algorithm, etc. Doable but tedious and may require more programming.
- Both Vandenplas et al. and Liu et al. (2024) also propose approximate expressions with good results

New idea

- see RPG $\hat{d}_g$ of genotyped animals as “data”
- use a BLUP solver with MME as follows:

$$\begin{pmatrix} A^{n,n}\lambda & A^{n,g}\lambda \\ A^{g,n}\lambda & A^{g,g}\lambda + I \end{pmatrix} \begin{pmatrix} \hat{d}_n \\ \hat{d}_g^* \end{pmatrix} = \begin{pmatrix} 0 \\ \hat{d}_g \end{pmatrix}$$

- same animals as in your ssGBLUP equations
- for $h^2 \approx 1$, $\hat{d}_g^* \approx \hat{d}_g$ and we get $\hat{d}_n$ too!
- $\hat{d}_n$ can be verified to be $\hat{d}_n = A_{n,g} A_{g,g}^{-1} \hat{d}_g$

Advantages

- $\begin{pmatrix} A^{n,n}\lambda & A^{n,g}\lambda \\ A^{g,n}\lambda & A^{g,g}\lambda + I \end{pmatrix} \begin{pmatrix} \hat{d}_n \\ \hat{d}_g^* \end{pmatrix} = \begin{pmatrix} 0 \\ \hat{d}_g \end{pmatrix}$
- we don’t restrict the equation to genotyped + ancestors
- we get RPG for **all** animals in the ssGBLUP equations
- easy to implement from existing ssGBLUP pedigree, GEBVs, genotypes, SNP solutions

does it run?

- Example MY, FY, PY
- ~50M animals in pedigree, ~2.5M animals genotyped
- we have 2.5M RPG "records" ($\hat{d}_g = \hat{u}_g - Z\hat{g}$) and 48M animals with "no records"
- 150M equations
- PCG iteration in blup90iod3, 8 threads

does it run?

- The PCG iteration seems very robust
- pseudo-true solution: PCG convergence $<10^{-14}$
- Time per round: 20 seconds
- reaching 10E-10 in 900 iterations takes 5h

PCG convergence	corr with pseudo-true	iteration
10E-06	0.916	113
10E-07	0.958	244
10E-08	0.978	381
10E-09	0.995	655
10E-10	0.9995	898
10E-11	0.999988	1177
10E-12	0.999999	1385
10E-13	1	1549
10E-14	1	1808

- Across all our trait groups, rounds to reach 10E-12 went from 100 to 1500
- Number of rounds depends on the size of the pedigree and proportion of genotyped animals in it.

Alternative new idea

- Maybe could add a general mean ?

$$\begin{pmatrix} 1'1 & 0 & 1' \\ 0 & A^{n,n}\lambda & A^{n,g}\lambda \\ 1 & A^{g,n}\lambda & A^{g,g}\lambda + I \end{pmatrix} \begin{pmatrix} \hat{\mu} \\ \hat{d}_n \\ \hat{d}_g^* \end{pmatrix} = \begin{pmatrix} 1'\hat{d}_g \\ 0 \\ \hat{d}_g \end{pmatrix}$$

- If we add back $\hat{\mu}$ we obtain the same solutions
- The PCG convergence is faster yet the "real" convergence is slower

Alternative new idea: does it run?

- Compare with previous pseudo-true solution
- In fact, we need a stricter criterion to reach the pseudo-true solution
- Total time is about the same as the first idea

PCG convergence	corr with pseudo-true	iteration
10E-10	0.872	31
10E-11	0.895	46
10E-12	0.9137	79
10E-13	0.950	213
10E-14	0.9947	568
10E-15	0.99918	712
10E-16	0.9996	804

- Improved PCG convergence might be misleading for some models

Thoughts & Conclusion

- My personal experience is that approximate expressions may eventually end up biting you
- …and that having and maintaining ONE Swiss Army knife is better than having many dedicated programs
- We think that using BLUP for RPG effects is a good option to get all RPGs, without much extra cost in time or programming
- relaxed convergence criterion of 10-E10 is probably enough

Acknowledgments

- Participating **dairy producers** for supplying data
- **DHI** organizations and **DRPCs** for processing and relaying the information to the Council on Dairy Cattle Breeding (CDCB)
- **Purebred breed associations** for providing pedigree data

- Mention of trade names or commercial products is solely for the purpose of providing specific information and does not imply recommendation or endorsement by CDCB
- CDCB is an equal opportunity provider and employer

Holstein in Oderitz, Navarre (my picture)

CDCB™

COUNCIL ON DAIRY CATTLE BREEDING

Thoughts & Conclusion

- My personal experience is that approximate expressions may eventually end up biting you
- …and that having and maintaining ONE Swiss Army knife is better than having many dedicated programs
- We think that using BLUP for RPG effects is a good option to get all RPGs, without much extra cost in time or programming
- relaxed convergence criterion of 10-E10 is probably enough