Approximating reliabilities of IP based on SNP effects from large ssGBLUP evaluations

Daniela Lourenco, M. Bermann, J.M. Tabet, I. Aguilar,

A. Legarra, S. Tsuruta, I. Misztal

June 19, 2024

ADSA® 2024 Annual Meeting
June 16–19, 2024
West Palm Beach, Florida
#ADSA2024

CDCB database statistics

3.2mil
lactation records integrated in National Cooperator Database in 2023

4.1mil
4.1 million cows on official milk recording (DHI, 2021)

72
countries with animal genotypes included in CDCB database

9mil
dairy animals genotyped (May 3, 2024)

1.4mil
genotypes added in 2023

93%
of genotyped animals are female

12
breeds represented in the genotypic database

89%
of genotypes are Holstein

10%
of genotypes are Jersey

https://uscdcb.com/database-stats/

Material and Methods

	Number of records
Pedigree	93.4M
Genotypes	2M
Cow Conception Rate (CCR)	35.2M
Heifer Conception Rate (HCR)	11.5M
Daughter Pregnancy Rate (DPR)	89.6M
Early First Calving (EFC)	35.4M

- Fertility traits are hard to evaluate due to low heritability, genetic correlations with milk yield, and changing management trends

https://www.hoards.com/

- Equivalence between ssGBLUP and ssSNPBLUP

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'W} \\ \mathbf{W'X} & \mathbf{W'W+H^{-1}}\frac{\sigma_e^2}{\sigma_u^2} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{W'y} \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X}_g'\mathbf{Z}_g\mathbf{M}_g & \mathbf{X}_n'\mathbf{Z}_n \\ \mathbf{M}_g'\mathbf{Z}_g'\mathbf{X}_g & \mathbf{Q} & \mathbf{M}_g'\mathbf{A}^{gn}\frac{\sigma_e^2}{\sigma_g^2} \\ \mathbf{Z}_n'\mathbf{X}_n & \mathbf{A}^{ng}\mathbf{M}_g\frac{\sigma_e^2}{\sigma_g^2} & \mathbf{Z}_n'\mathbf{Z}_n + \mathbf{A}^{nn}\frac{\sigma_e^2}{\sigma_g^2} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\alpha}} \\ \hat{\mathbf{u}}_n \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{M}_g'\mathbf{Z}_g'\mathbf{y}_g \\ \mathbf{Z}_n'\mathbf{y}_n \end{bmatrix}$$

Aguilar et al. (2010)
Christensen & Lund (2010)

Liu et al. (2014)
Fernando et al. (2016)

GEBV

SNP effects

$$\hat{a} = k\mathbf{Z'G^{-1}}\hat{\mathbf{u}}$$

$$\hat{\mathbf{u}} = \mathbf{Z}\hat{a}$$

Indirect Predictions:

$$\hat{\mathbf{u}}_m^* = \mathbf{Z}\hat{a}$$

$$\hat{\mathbf{u}}_m = \hat{\boldsymbol{\mu}} + \hat{\mathbf{u}}_m^*$$

$$\hat{\boldsymbol{\mu}} = \mathbf{1'G^{-1}}\hat{\mathbf{u}}$$

- Reliability of Indirect Predictions

$$\text{REL}_{\text{IP}_j} = \frac{\mathbf{z}_j var(\widehat{\boldsymbol{a}})\mathbf{z}_j{}'}{\sigma_u^2}$$

Liu et al. (2017)

Gualdrón Duarte et al. BMC Bioinformatics 2014, 15:246
http://www.biomedcentral.com/1471-2105/15/246

BMC Bioinformatics

METHODOLOGY ARTICLE   Open Access

Rapid screening for phenotype-genotype
associations by linear transformations of genomic
evaluations

Jose L Gualdrón Duarte[1], Rodolfo JC Cantet[1], Ronald O Bates[2], Catherine W Ernst[2], Nancy E Raney[2]
and Juan P Steibel[2,3*]

- Obtaining $var(\widehat{\boldsymbol{a}})$ from ssGBLUP

$$Var(\widehat{\boldsymbol{a}}) = k\mathbf{Z}'\mathbf{G}^{-1}(\mathbf{G} - \mathbf{C}^{\mathbf{u_2 u_2}})\mathbf{G}^{-1}\mathbf{Z}k$$

Garcia et al. Genetics Selection Evolution   (2022) 54:66
https://doi.org/10.1186/s12711-022-00752-4

GSE Genetics
Selection
Evolution

RESEARCH ARTICLE   Open Access

Theoretical accuracy for indirect predictions
based on SNP effects from single-step GBLUP

Andre Garcia[1*], Ignacio Aguilar[2], Andres Legarra[3], Shogo Tsuruta[1], Ignacy Misztal[1] and Daniela Lourenco[1]

High cost – inverse of LHS of MME

Feasible for small datasets – up to 35k genotyped animals

How to overcome this limitation?
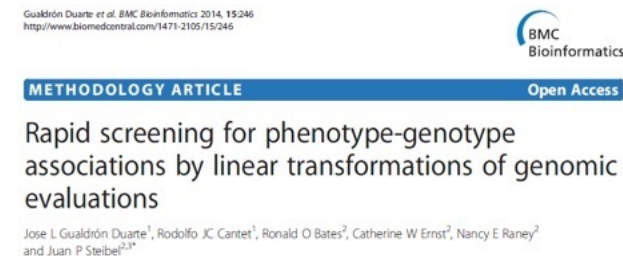
Using the APY framework

Theoretical accuracy for indirect predictions based on SNP effects from single-step GBLUP

Andre Garcia[1]*, Ignacio Aguilar[2], Andres Legarra[3], Shogo Tsuruta[1], Ignacy Misztal[1] and Daniela Lourenco[1]

On the equivalence between marker effect models and breeding value models and direct genomic values with the Algorithm for Proven and Young

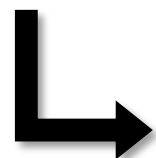Matias Bermann[1]*, Daniela Lourenco[1], Natalia S. Forneris[2,3], Andres Legarra[4] and Ignacy Misztal[1]

- If using APY in ssGBLUP
  - Numerical equivalence
  - $\hat{\mathbf{u}} = \mathbf{Z}\hat{\boldsymbol{a}}$
  - $\hat{\boldsymbol{a}} = k\mathbf{Z}'\mathbf{G}_{\text{APY}}^{-1}\hat{\mathbf{u}}$
  - $Var(\hat{\boldsymbol{a}}) = k\mathbf{Z}'\mathbf{G}_{\text{APY}}^{-1}(\mathbf{G} - \mathbf{C}^{\mathbf{u_2 u_2}})\mathbf{G}_{\text{APY}}^{-1}\mathbf{Z}k$

  → Function of all genotyped animals

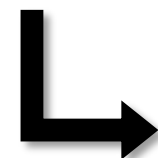- If using APY in ssGBLUP
  - Equivalent APY ssSNPBLUP model
  - $\hat{\mathbf{u}} = \mathbf{Z}^{\dagger}\hat{\boldsymbol{a}}$
  - $\hat{\boldsymbol{a}} = k\mathbf{Z}^{\dagger'}\mathbf{G}_{\text{APY}}^{-1}\hat{\mathbf{u}} = k\mathbf{Z}_{\text{c}}'\mathbf{G}_{\text{cc}}^{-1}\hat{\mathbf{u}}_{\text{c}}$
  - $Var(\hat{\boldsymbol{a}}) = k\mathbf{Z}_{\text{c}}'\mathbf{G}_{\text{cc}}^{-1}(\mathbf{G}_{\text{cc}} - \mathbf{C}^{\mathbf{u_{2c} u_{2c}}})\mathbf{G}_{\text{cc}}^{-1}\mathbf{Z}_{\text{c}}k$

  → Function of CORE animals

$$Var(\widehat{\mathbf{a}}) = k\mathbf{Z}'\mathbf{G}_{\text{APY}}^{-1}(\mathbf{G} - \mathbf{C}^{\mathbf{u_2 u_2}})\mathbf{G}_{\text{APY}}^{-1}\mathbf{Z}k$$

$$Var(\widehat{\mathbf{a}}) = k\mathbf{Z}_c'\mathbf{G}_{cc}^{-1}(\mathbf{G}_{cc} - \boxed{\mathbf{C}^{\mathbf{u_{2c} u_{2c}}}})\mathbf{G}_{cc}^{-1}\mathbf{Z}_c k$$

?

Garcia et al. Genetics Selection Evolution   (2022) 54:66
https://doi.org/10.1186/s12711-022-00752-4

GSE Genetics Selection Evolution

RESEARCH ARTICLE   Open Access

Theoretical accuracy for indirect predictions based on SNP effects from single-step GBLUP

Andre Garcia[1], Ignacio Aguilar[2], Andres Legarra[3], Shogo Tsuruta[1], Ignacy Misztal[1] and Daniela Lourenco[1]

Bermann et al. Genetics Selection Evolution   (2022) 54:52
https://doi.org/10.1186/s12711-022-00741-7

GSE Genetics Selection Evolution

RESEARCH ARTICLE   Open Access

On the equivalence between marker effect models and breeding value models and direct genomic values with the Algorithm for Proven and Young

Matias Bermann[1], Daniela Lourenco[1], Natalia S. Forneris[2,3], Andres Legarra[4] and Ignacy Misztal[1]
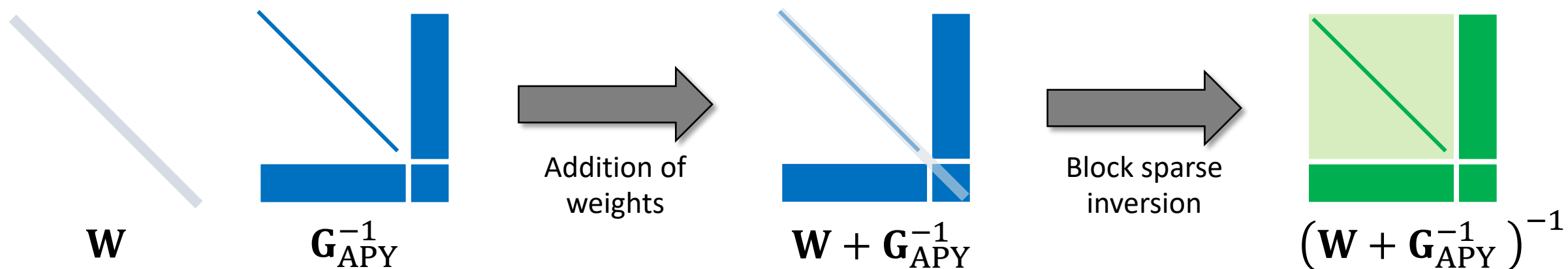
- Approximated reliabilities of GEBV

7

- Reliability based on PEV
  - Approximated for large populations
  - Weights based on approximations
  - Block sparse inversion with APY

$$\mathbf{W} \qquad \mathbf{G}_{\mathrm{APY}}^{-1}$$

Addition of weights

$$\mathbf{W} + \mathbf{G}_{\mathrm{APY}}^{-1}$$

Block sparse inversion

$$\left(\mathbf{W} + \mathbf{G}_{\mathrm{APY}}^{-1}\right)^{-1}$$

- Genomic evaluation process
  - GEBV using APY ssGBLUP + reliability using block sparse inversion

$$\mathbf{C}^{\mathbf{u}_{2c}\mathbf{u}_{2c}} = \left(\mathbf{W} + \mathbf{G}_{\mathbf{APY}}^{-1}\right)^{-1}$$
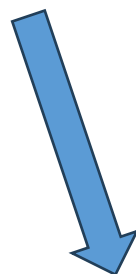
$$Var(\widehat{\mathbf{a}}) = k\mathbf{Z}_c'\mathbf{G}_{cc}^{-1}\left(\mathbf{G}_{cc} - \mathbf{C}^{\mathbf{u}_{2c}\mathbf{u}_{2c}}\right)\mathbf{G}_{cc}^{-1}\mathbf{Z}_c k$$

Pipeline

- `blup90iod3`
  - GEBV

- `accf90GS3`
  - GEBV reliability and $\mathbf{C}^{\mathbf{u}_{2c}\mathbf{u}_{2c}}$

- `postGSf90`
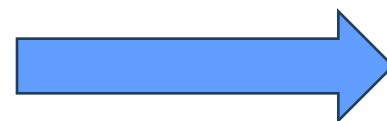 - Backsolve $\mathbf{C}^{\mathbf{u}_{2c}\mathbf{u}_{2c}}$ to $var(\hat{a})$

- `Predf90`
  - IP and IP reliability

Official Evaluations

Added portion

- **Pedigree**: 2,240,568 animals

- **Milk Yield**: 1,422,330 Records
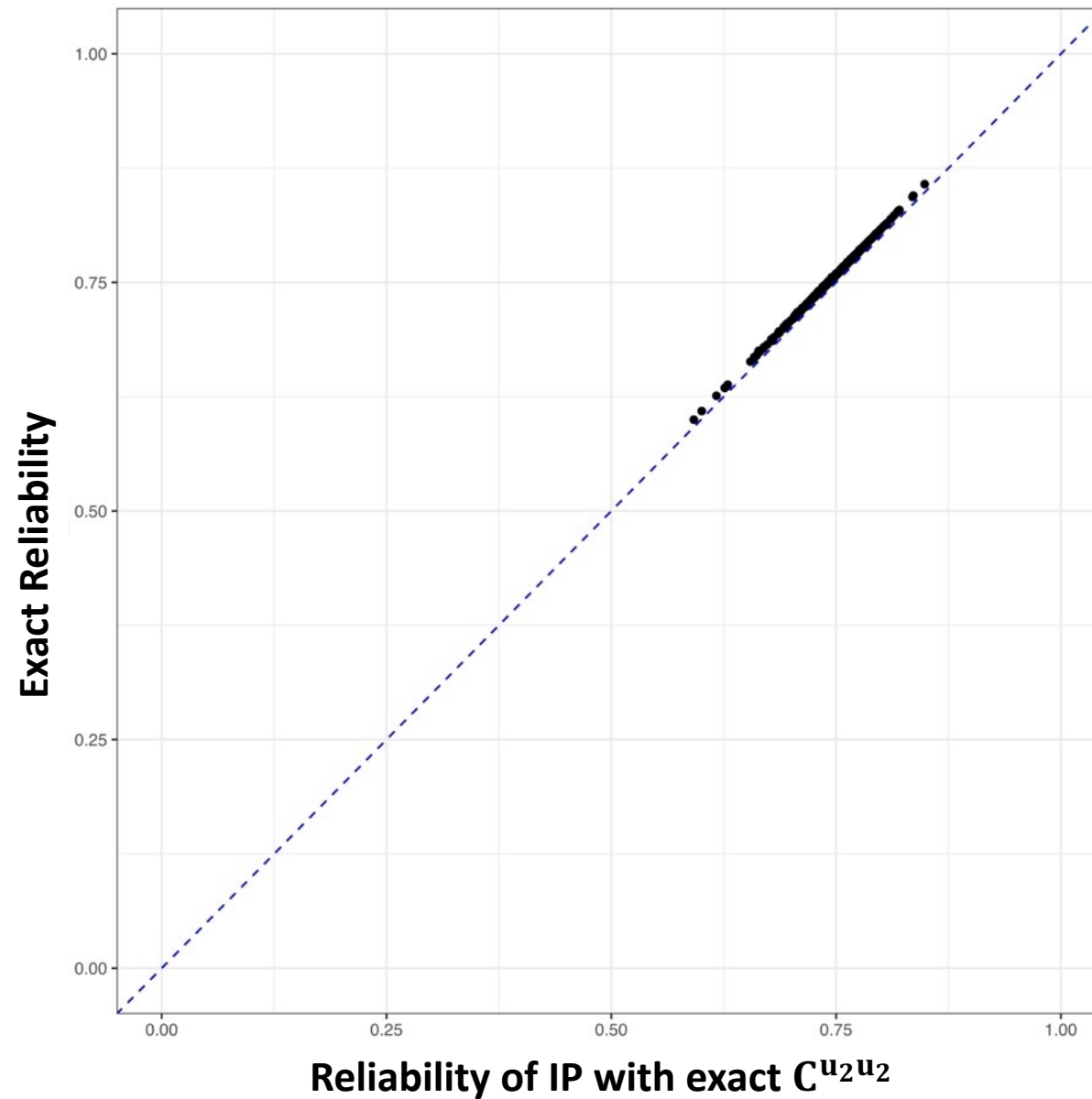
- **Genotypes**: Total: 33,338

    Training: 32,570 bulls

    Validation (2017): 768 bulls

- **Exact reliabilities based on the inverse** (training + validation)

- **Reliabilities of IP** (for validation) **with exact $C^{u_2 u_2}$** (from training)

- **Reliabilities of IP** (for validation) **with approximated $C^{u_{2c} u_{2c}}$** (from training)
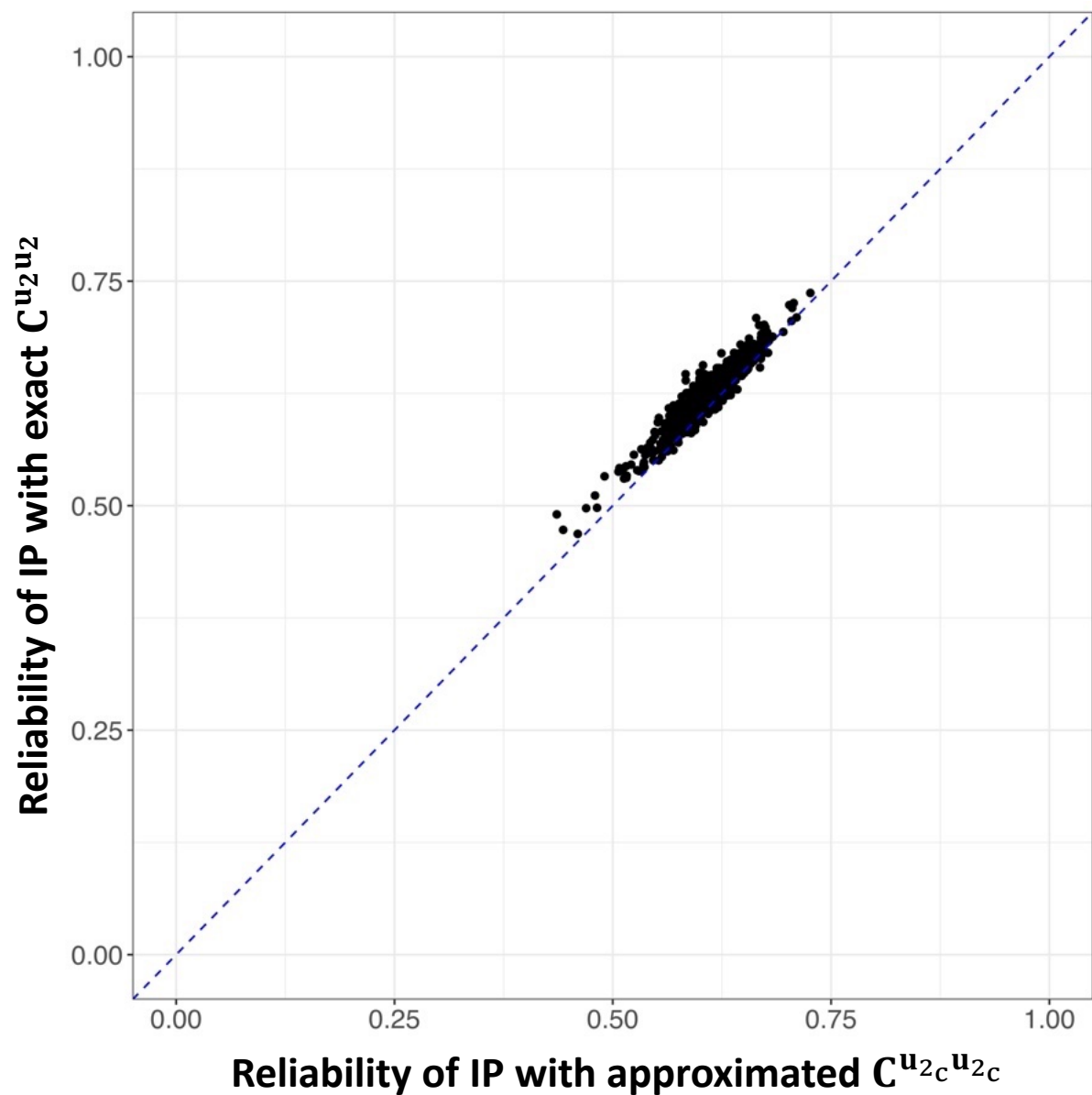
Exact vs. IP reliabilities

$b_0 = 0.01$

$b_1 = 0.99$

$Cor = 0.99$

$b_0 = 0.07$

$b_1 = 0.90$

Cor $= 0.94$

- Reliability of IP can be computed from the official ssGBLUP runs

  - Based on approximated $\mathbf{C^{u_{2c}u_{2c}}}$

  - Using the existent pipeline

  - Already implemented in BLUPF90

- Next steps:

  - Add the residual polygenic effect

  - Genotyped animals with own or progeny records

  - Multibreed