Single-step national evaluation using phenotypic, full pedigree and genomic information

Ignacy Misztal, Ignacio Aguilar, Shogo Tsuruta

University of Georgia

Dave Johnson, **LIC, New Zealand**

Andres Legarra, **INRA Toulouse**

Tom Lawlor, **Holstein Association, USA**

Steps in genomic selection

Regular genetic evaluation → Extraction of pseudo observations → Genomic evaluation

→ Creation of index

- Many parameters
- Complicated
- Biases and inflation (Patry and Ducrocq, 2009; VanRaden, 2009)

Equivalent "genomic" equations

Pseudo-obs SNP effects

$$\tilde{y} = \mu + Za + e, \quad \mathrm{var}(a) = D\sigma_a^2$$

breeding values

$$\tilde{y} = \mu + u + e, \quad \mathrm{var}(u) = G\sigma_u^2, \quad G = ZDZ'/k$$

Z – centered design matrix
G – genomic relationship matrix

Genomic information ≈ genomic relationships

Possibility of one-step evaluation (Misztal et al., 2009)

Replace relationship matrix **A** by **H:**

$\mathbf{H}=\mathbf{A}+\mathbf{A}_{\Delta}$

$\mathbf{A}_{\Delta}$ - modifications due to genomic information

Issues

- What is $\mathbf{A}_\Delta$ in $\mathbf{H}=\mathbf{A}+\mathbf{A}_\Delta$?
- How to solve equations with $\mathbf{H}$?

Starting points

- Assume that $\mathbf{A}+\mathbf{A}_{\Delta}$ cannot be inverted

- Product **A** x *vector* easily computed
  - (Colleau, 2002) (*via Bruce Tier*)

Nonsymmetric equations

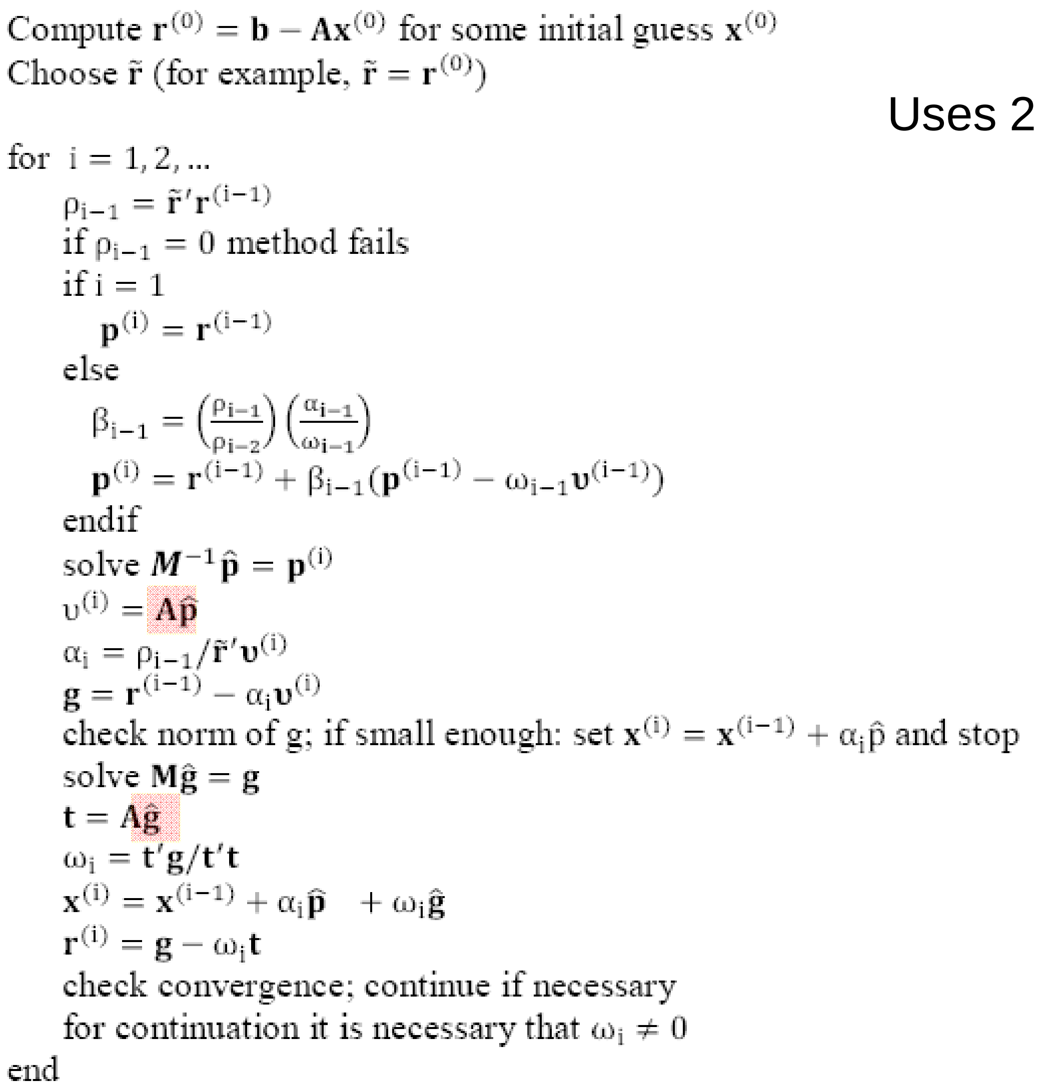
$$\begin{bmatrix} \boldsymbol{X'X} & \boldsymbol{X'Z} \\ \boldsymbol{HZ'X} & \boldsymbol{HZ'Z} + \alpha \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{u}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X'y} \\ \boldsymbol{HZ'y} \end{bmatrix}$$

Equations for singular **H**: Harville (1978)

Useful for **MACE** where (co)variance matrix singular

Bi-Conjugate Gradient Stabilized algorithm (Bi-CGSTAB)

Uses 2 LHS by vector multiplications



```
Compute $\mathbf{r}^{(0)} = \mathbf{b} - \mathbf{A}\mathbf{x}^{(0)}$ for some initial guess $\mathbf{x}^{(0)}$
Choose $\tilde{\mathbf{r}}$ (for example, $\tilde{\mathbf{r}} = \mathbf{r}^{(0)}$)

for  $i = 1, 2, \ldots$
    $\rho_{i-1} = \tilde{\mathbf{r}}'\mathbf{r}^{(i-1)}$
    if $\rho_{i-1} = 0$ method fails
    if $i = 1$
        $\mathbf{p}^{(i)} = \mathbf{r}^{(i-1)}$
    else
        $\beta_{i-1} = \left(\frac{\rho_{i-1}}{\rho_{i-2}}\right)\left(\frac{\alpha_{i-1}}{\omega_{i-1}}\right)$
        $\mathbf{p}^{(i)} = \mathbf{r}^{(i-1)} + \beta_{i-1}(\mathbf{p}^{(i-1)} - \omega_{i-1}\mathbf{\upsilon}^{(i-1)})$
    endif
    solve $M^{-1}\hat{\mathbf{p}} = \mathbf{p}^{(i)}$
    $\upsilon^{(i)} = \mathbf{A}\hat{\mathbf{p}}$
    $\alpha_i = \rho_{i-1}/\tilde{\mathbf{r}}'\mathbf{\upsilon}^{(i)}$
    $\mathbf{g} = \mathbf{r}^{(i-1)} - \alpha_i\mathbf{\upsilon}^{(i)}$
    check norm of g; if small enough: set $\mathbf{x}^{(i)} = \mathbf{x}^{(i-1)} + \alpha_i\hat{\mathrm{p}}$ and stop
    solve $\mathbf{M}\hat{\mathbf{g}} = \mathbf{g}$
    $\mathbf{t} = \mathbf{A}\hat{\mathbf{g}}$
    $\omega_i = \mathbf{t}'\mathbf{g}/\mathbf{t}'\mathbf{t}$
    $\mathbf{x}^{(i)} = \mathbf{x}^{(i-1)} + \alpha_i\hat{\mathbf{p}} \quad + \omega_i\hat{\mathbf{g}}$
    $\mathbf{r}^{(i)} = \mathbf{g} - \omega_i\mathbf{t}$
    check convergence; continue if necessary
    for continuation it is necessary that $\omega_i \neq 0$
end
```

Computing Left_Hand_Side x vector

$$\begin{bmatrix} \boldsymbol{X'X} & \boldsymbol{X'Z} \\ \boldsymbol{HZ'X} & \boldsymbol{HZ'Z} + \alpha \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{q}_1 \\ \boldsymbol{q}_2 \end{bmatrix} =$$

$$\begin{bmatrix} \boldsymbol{X'Xq}_1 + \boldsymbol{X'Zq}_2 \\ \boldsymbol{HZ'Xq}_1 + \boldsymbol{HZ'Zq}_2 + \alpha \boldsymbol{q}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{c}_1 \\ \boldsymbol{Ac}_2 + \boldsymbol{A}_{\Delta}\boldsymbol{c}_2 + \alpha \boldsymbol{q}_2 \end{bmatrix}$$

Ac_2 efficiently computed by Colleau (2002) algorithm

Preconditioner for symmetric equations

$$\boldsymbol{M}^{-1} = diag(\boldsymbol{LHS})^{-1} = \begin{bmatrix} \boldsymbol{D}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{D}_2 \end{bmatrix}$$

Preconditioner for unsymmetric equations

$$\mathbf{M}_{\mathrm{M}}{}^{-1} = \begin{bmatrix} \boldsymbol{D}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{D}_2\boldsymbol{A}^{-1} \end{bmatrix}$$

Progression of H

Initial (Misztal et al., 2009)

$$\mathrm{H}=\mathrm{A}+\begin{bmatrix} 0 & 0 \\ 0 & \mathrm{G}\text{-}\mathrm{A}_{22} \end{bmatrix}$$

1-ungenotyped animals
2-genotyped animals

Comprehensive (Legarra ,2009)

$$\mathbf{H} = \mathbf{A} + \begin{bmatrix} \mathbf{A}_{12}\mathbf{A}_{22}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ \mathbf{I} \end{bmatrix} \left[\mathbf{G} - \mathbf{A}_{22}\right] \left[\mathbf{I} \quad \mathbf{I}\right] \begin{bmatrix} \mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

Inverse of Comprehensive (Johnson, 2009)

$$\mathrm{H}^{-1}=\mathrm{A}^{-1}+\begin{bmatrix} 0 & 0 \\ 0 & \mathrm{G}^{-1}\text{-}\mathrm{A}_{22}^{-1} \end{bmatrix}$$

Data

- 10.5 million final score of US Holsteins
- 6,2 million cows
- 9.1 million pedigrees

- Genotypes for 6,508 bulls (Illumina BovineSNP50 BeadChip)

Repeatability models

- **P2009** – 2009 regular evaluations
- **P2004** – 2004 regular evaluations

- **Multi** – 2004 genomic prediction using multiple steps (VanRaden et al., 2009)

- **Single** – 2004 single-step genomic prediction

Genomic relationship matrix

$$\mathbf{G}_b = \frac{\mathbf{ZZ'}}{k}$$

$$z_{ij} = \begin{cases} 0 - 2p_j & \text{- homozygous 11} \\ 1 - 2p_j & \text{- heterozygous 12 - 21} \\ 2 - 2p_j & \text{- homozygous 22} \end{cases}$$

- Options for p
  - p=0.5 (G05)
  - base generation allele frequencies (GB) (Gengler et al., 2007)
  - current generation allele frequencies (GP)

Computing

- $\mathbf{A}_{22}^{-1}$ by Colleau algorithm
- $\mathbf{G} = 0.95\mathbf{G}_b + 0.05\mathbf{A}_{22}$

- Solutions by modified BLUP90IOD (Tsuruta et al.. 2009)
  - Iteration on data
  - Preconditioned Conjugate Gradient
  - Used for national evaluations in dairy, beef, swine, chicken,…

Regression for comparisons

$X2009 = \mu + \delta\ X2004 + e$

0.0 1.0 small (R^2 high) Desirables

X2009 – 2009 proofs or DYD

X2004 - proofs from various 2004 analyses

Inflation =$1/\delta$

Goodness of fit for several predictions

Prediction	EBV2009	
	R^2	δ (Regr)
PA	36	0.82
Multistep	50	0.83
Single-step		
G05	49	0.71
GB	45	0.64
GP	45	0.66

05 – equal gene frequencies B-base population P-genotyped population

Reduction of inflation

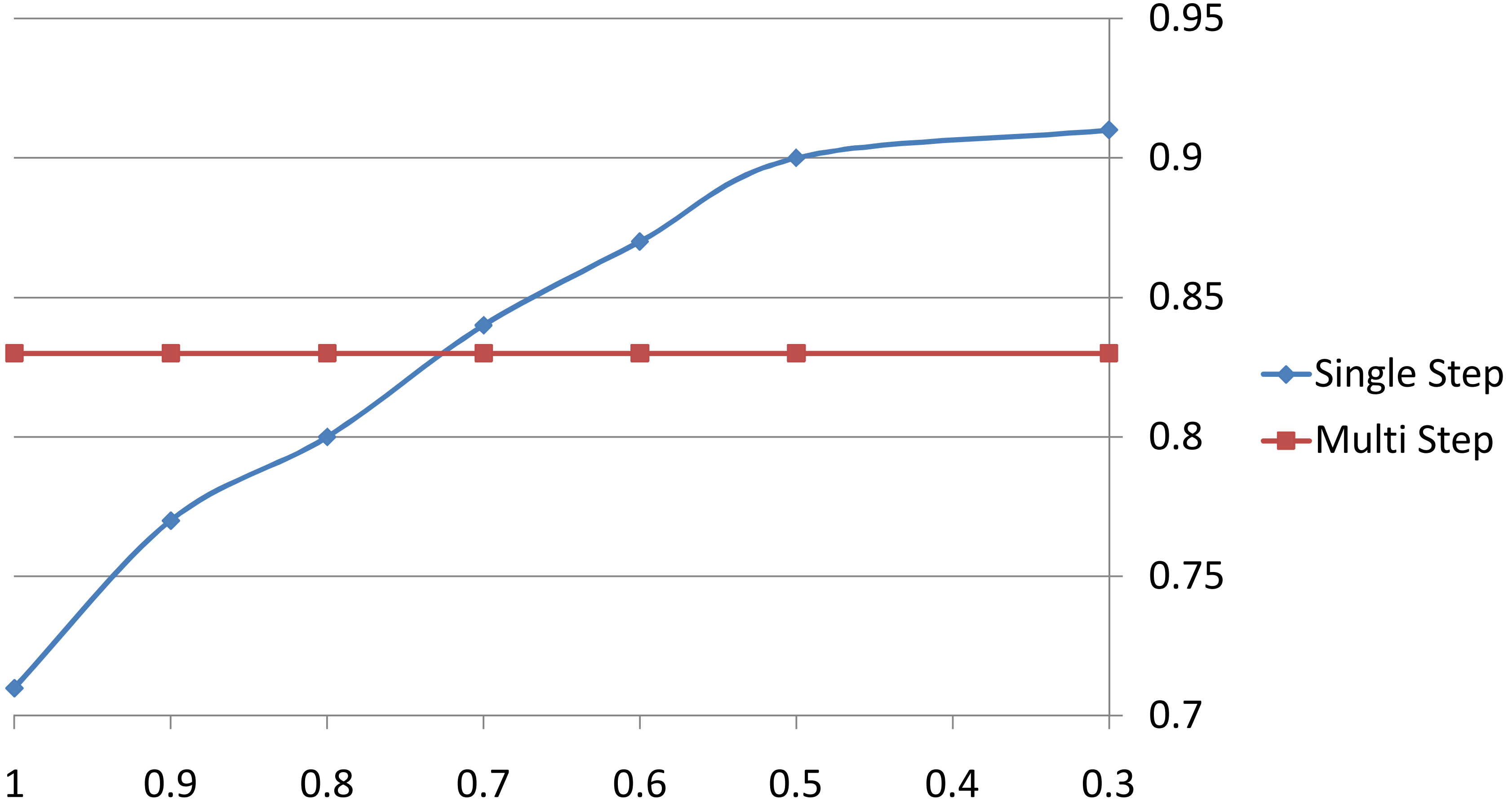
$$u_i = \frac{u_s + u_d + \sum_{j, j \neq i} (a_{22}^{ij} - g^{ij}) u_j}{2 + \lambda (g^{ii} - a_{22}^{ii})} = \text{wPA+GBV-GPI}$$

Modification

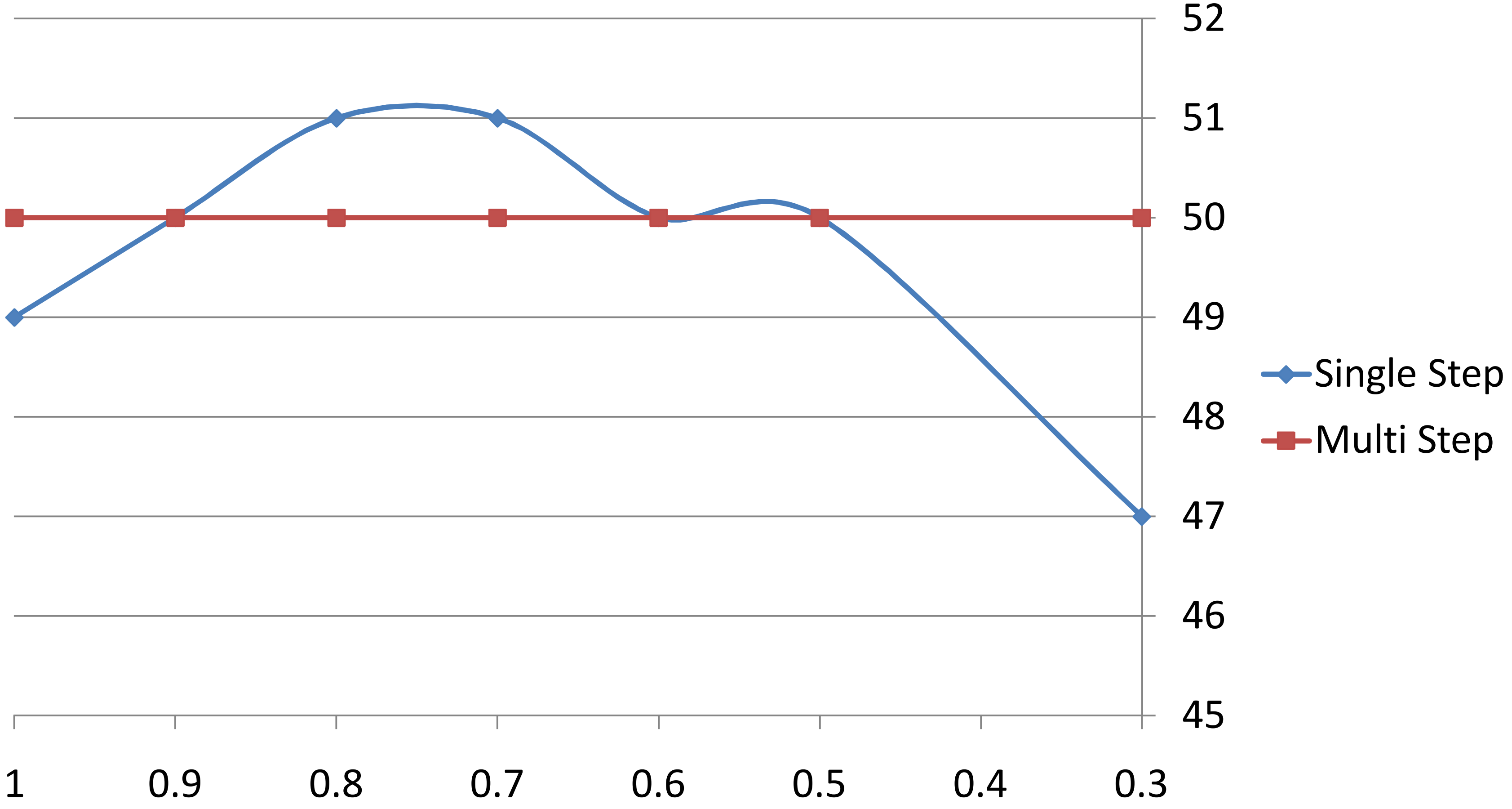
$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \lambda\left(\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}\right) \end{bmatrix}$$

λ determines weight on genomic information
(poster S28-16 by Legarra et al.)

Regression as function of λ (EBV)



R^2 as function of λ (EBV)



Timing

A_{22}	< 1 min
G	5 min
G^{-1}	3 min
A_{22}^{-1}	3 min

Evaluation – 2 hrs (2% more than regular)

Details by Aguilar et al, 2009; ADSA2009

Features of one-step evaluation

- Simplicity
- Improvement of accuracies of nongenotyped animals
- Greater resistance to genomic selection bias

- Ability to work with any model

- Special utility when information on genotyped animals low and varied (e.g., chicken)

BLUP with a new relationship matrix

Issues

- Correspondence to BayesA, BayesB,….
- Optimal selection of weighting factor
- Development of best “G”
- New tools for fine-tuning

Conclusions

- One-step genomic evaluation with $\mathbf{H}^{-1}$ much simpler, more versatile and more accurate than multistep
- Genomic selection becomes mature technology
- Lots of research to do
  - Postdocs? Graduate students?

Acknowledgements

Guilherme Rosa

Paul VanRaden

Curt VanTassel

George Wiggans

Rohan Fernando

Bruce Tier

Holstein Association

Beef Cattle Consortium

Smithfield Premium Genetics

PIC