Combination of genotype, pedigree, local and foreign information

J. Vandenplas[1,2], F.G. Colinet[1], N. Gengler[1,*]

[1] University of Liège, Gembloux Agro-Bio Tech, Belgium
[2] National Fund for Scientific Research, Brussels, Belgium

Introduction

- ✓ For genomic prediction
  - ☐ To get reliable GEBV use of multiple sources needed
- ✓ Traditional traits (*i.e.*, in dairy cattle)
  - ☐ Local genomic evaluations ← MACE
- ✓ Novel traits (e.g., milk quality, feed efficiency, methane)
  - ☐ Combining different sources of often locally sparse phenotypic data even more needed

→ Adapting strategies to use multiple sources

Aims

- ✓ To develop strategies to use simultaneously:
    - ☐ Genotypes,
    - ☐ Pedigree,
    - ☐ Local phenotypes and,
    - ☐ Foreign phenotypes
- ✓ By combining and adapting:
    - ☐ Single-step genomic evaluations (**ssGBLUP**) and,
    - ☐ Bayesian procedure to integrate external information*

***Vandenplas, J., & Gengler, N. (2012). *J. Dairy Sci. 95*: 1513-1526**

Aims

✓ To develop strategies to use simultaneously:

  - ☐ Genotypes,
  - ☐ Pedigree,
  - ☐ Phenotypes from multiple sources

→ ssGBAYES

ssGBLUP

- ✓ Single-step genomic evaluation (ssGBLUP)
  - ❑ Allows direct combination of genomic, pedigree and phenotypic information
- ✓ However: current limitation
  - ❑ Only available local phenotypic records can be used

ssGBLUP

- ✓ Single-step genomic evaluation (ssGBLUP)
  - ❑ Allows direct combination of genomic, pedigree and phenotypic information
- ✓ However: current limitation
  - ❑ Only available local phenotypic records can be used

→ In opposition to multi-step methods

(e.g., use of local-EBV and MACE-EBV in prediction equation step)

ssGBLUP

- ✓ Single-step genomic evaluation (ssGBLUP)
  - ☐ Allows direct combination of genomic, pedigree and phenotypic information
- ✓ However: current limitation
  - ☐ Only available local phenotypic records can be used
- ✓ Reason for this limitation
  - ☐ Local ssGBLUP based on modified system of mixed model equations (**MME**)

Modified MME

- ✓ ssGBLUP

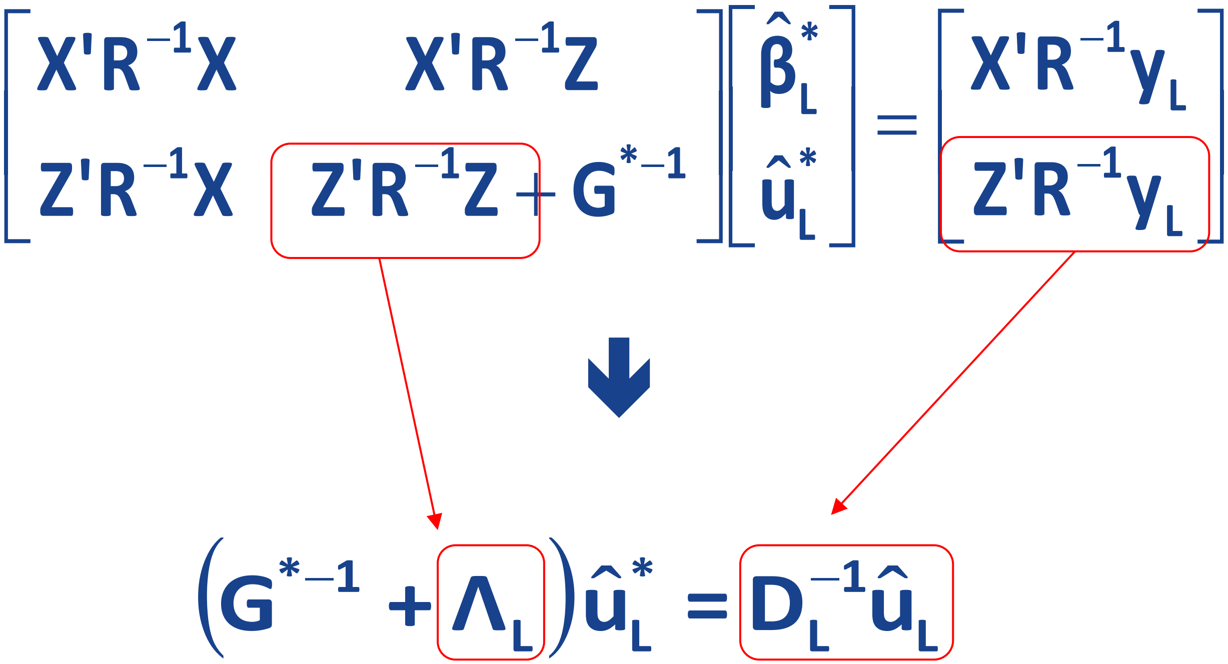
$$\begin{bmatrix} \mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}Z} \\ \mathbf{Z'R^{-1}X} & \mathbf{Z'R^{-1}Z} + \mathbf{G}^{*-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}_L^* \\ \hat{\mathbf{u}}_L^* \end{bmatrix} = \begin{bmatrix} \mathbf{X'R^{-1}y_L} \\ \mathbf{Z'R^{-1}y_L} \end{bmatrix}$$

- ❑ $\mathbf{G}^{*-1} = \mathbf{H}^{-1} \otimes \mathbf{G}_0^{-1}$: inverse of combined genomic-pedigree based (co)variances matrix
- ❑ $\mathbf{y}_L$: vector of local phenotypes
- ❑ $\hat{\boldsymbol{\beta}}_L^*$: vector of estimated local fixed effects
- ❑ $\hat{\mathbf{u}}_L^*$: vector of estimated local GEBV
- ❑ $\mathbf{p}\left(\mathbf{u}_L^*\right) = MVN\left(\mathbf{0}, \mathbf{G}^{*-1}\right)$

Alternative

- ✓ As alternative to introduce phenotypes
  - ☐ Also allowing for foreign information
- ✓ However avoiding double counting
  - ☐ E.g. MACE-EBV contain our phenotypes
- ✓ Also avoiding deregression
  - ☐ Potential source of trouble
- ✓ Be simple and flexible
  - ☐ Allowing to extend to multiple sources

Concept of "Hybrid" ssGBLUP



$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & \boxed{Z'R^{-1}Z} + G^{*-1} \end{bmatrix} \begin{bmatrix} \hat{\beta}^{*}_{L} \\ \hat{u}^{*}_{L} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y_{L} \\ \boxed{Z'R^{-1}y_{L}} \end{bmatrix}$$

$$\Downarrow$$

$$\left(G^{*-1} + \boxed{\Lambda_{L}}\right)\hat{u}^{*}_{L} = \boxed{D_{L}^{-1}\hat{u}_{L}}$$

➔ More details later

Bayesian priors

- ✓ Assumption: *a priori* known information on $\hat{\mathbf{u}}_L^*$

  ➔ $\mathbf{y}_L$ replaced by $\hat{\mathbf{u}}_L$ and $\mathbf{D}_L$

  - ❑ Source of phenotypic information
    - ➢ Vector of local EBV: $\hat{\mathbf{u}}_L$ 🡐 "external"* information
    - ➢ Prediction error (co)variances matrix: $\mathbf{D}_L$
- ✓ Allows use of simplified models
  - ❑ *E.g.*, Test-day model ➔ lactation EBV

* "external" = outside of original system of MME

"Hybrid" ssGBLUP ⇨ ssGBAYES

✓ Using only local information as source

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{*-1} \end{bmatrix} \begin{bmatrix} \hat{\beta}_L^* \\ \hat{u}_L^* \end{bmatrix} = \begin{bmatrix} X'R^{-1}y_L \\ Z'R^{-1}y_L \end{bmatrix}$$

⬇

$$\left(G^{*-1} + \boxed{\Lambda_L}\right)\hat{u}_L^* = \boxed{D_L^{-1}\hat{u}_L}$$

Least square part of LHS of hypothetical BLUP generating local information

RHS of hypothetical BLUP generating local information

Bayesian priors

- ✓ Assumption: *a priori* known information on $\hat{\mathbf{u}}_L^*$
  - ❑ Extended to 2 sources of phenotypic information
    - ➢ Vector of local EBV: $\hat{\mathbf{u}}_L$ 🡐 1st "external"* information
      Vector of foreign EBV: $\hat{\mathbf{u}}_F$ 🡐 2nd "external"* information
    - ➢ Prediction error (co)variances matrices: $\mathbf{D}_L$, $\mathbf{D}_F$
  - ❑ Issue: only available for some animals

➔ $\hat{\mathbf{u}}_L$, $\hat{\mathbf{u}}_F$, $\mathbf{D}_L$ and $\mathbf{D}_F$: (partially) unknown

* "external" = outside of original system of MME

Methods

- ✓ For both sources: estimation of $\hat{\mathbf{u}}_{\mathbf{i}}$ (i=L,F)
  - ❑ Available
    - ➢ EBV of some animals (so-called “external” $\hat{\mathbf{u}}_{\mathbf{i}_E}$)
  - ❑ “Internal” animals: prediction of EBV ($\hat{\mathbf{u}}_{\mathbf{i}_I}$)

$$p\left(\hat{\mathbf{u}}_{\mathbf{i}_I} \middle| \hat{\mathbf{u}}_{\mathbf{i}_E}\right) = MVN\left(\mathbf{G}_{\mathbf{i}_{IE}} \mathbf{G}_{\mathbf{i}_{EE}}^{-1} \hat{\mathbf{u}}_{\mathbf{i}_E}, \left(\mathbf{G}^{\mathbf{i}_{II}}\right)^{-1}\right)$$

$\mathbf{G} = \mathbf{A} \otimes \mathbf{G}_0$: genetic (co)variances matrix

- ❑ $\hat{\mathbf{u}}_{\mathbf{i}} = \begin{bmatrix} \hat{\mathbf{u}}_{\mathbf{i}_E}' & \hat{\mathbf{u}}_{\mathbf{i}_I}' \end{bmatrix}'$

➔ Correct propagation of information

Methods

- ✓ For both sources: estimation of $\mathbf{D}_\mathrm{i}$ (i=L,F)

$$\mathbf{D}_\mathrm{i} = \mathbf{G}^{-1} + \mathbf{\Lambda}_\mathrm{i}$$

$$\mathbf{\Lambda}_\mathrm{i} = block\,diag\left(\mathbf{\Delta}_\mathrm{j}\mathbf{R}_\mathbf{0}^{-1}\mathbf{\Delta}_\mathrm{j}\right); j = 1,\ldots,n \text{ animals}$$

$$\begin{cases} \text{For external animals}: \mathbf{\Delta}_\mathrm{j} = diag\left(\sqrt{RE_k}\right); k = 1,\ldots,t \text{ traits} \\ \text{For internal animals}: \mathbf{\Delta}_\mathrm{j} = \mathbf{0} \end{cases}$$

➔ All matrices $\mathbf{\Lambda}_\mathrm{i}$ depend only on contributions due to own records

Issue

- ✓ Non-Independence of information sources
  - ❑ E.g., Local information included in MACE-EBV

$$\mathbf{D}_{Ff}^{-1}\boxed{\hat{\mathbf{u}}_{Ff}} = \mathbf{D}_{F}^{-1}\boxed{\hat{\mathbf{u}}_{F}} - \mathbf{D}_{Lc}^{-1}\boxed{\hat{\mathbf{u}}_{Lc}}$$

Vector of foreign EBV free of local information

Vector of foreign EBV

Vector of local EBV contributing to foreign information

➔ Estimation of external information free from local information

Methods

✓ Combination of genotype, pedigree, local and foreign information (ssGBayes)

$$\left(\mathbf{G}^{*-1} + \boxed{\mathbf{\Lambda}_{L}} + \boxed{\mathbf{\Lambda}_{F} - \mathbf{\Lambda}_{Lc}}\right)\hat{\mathbf{u}}_{L}^{*} = \boxed{\mathbf{D}_{L}^{-1}\hat{\mathbf{u}}_{L}} + \boxed{\mathbf{D}_{F}^{-1}\hat{\mathbf{u}}_{F} - \mathbf{D}_{Lc}^{-1}\hat{\mathbf{u}}_{Lc}}$$

Least square part of LHS of hypothetical BLUP of local information

Least square part of LHS of hypothetical BLUP of foreign information free of local information

RHS of hypothetical BLUP of local information

RHS of hypothetical BLUP of foreign information free of local information

→ No double counting of contributions

Example : Walloon genomic evaluation

- ✓ ssGBAYES results for milk yield
  - ❑ But already used for all INTERBULL traits
- ✓ 1,909 genotyped Holstein bulls and cows
- ✓ 16,234 animals (genotyped and ancestors)
  - ❑ 12,046 animals with Walloon EBV (EBV_W)
  - ❑ 1,981 bulls with MACE EBV (EBV_M)
  - ❑ 601 bulls with Walloon EBV contributing to MACE (EBV_{Wc})
- ✓ Reliabilities (REL) for GEBV obtained through inversion of left-hand side

Results

- ✓ Averaged REL (SD) associated to EBV_W, $GEBV_W$ and $GEBV_{W+M-Wc}$ for genotyped bulls for milk yield

Class of REL_W	N	EBV_W [1]	No. of publishable bulls	$GEBV_W$ [2]	No. of publishable bulls	$GEBV_{W+M-Wc}$ [3]	No. of publishable bulls
< 0.50	647	0.25 (0.12)	0	0.44 (0.09)	194	0.80 (0.09)	628
0.50 – 0.74	173	0.63 (0.07)	173	0.69 (0.06)	173	0.87 (0.05)	173
≥ 0.75	390	0.90 (0.07)	390	0.91 (0.06)	390	0.94 (0.04)	390

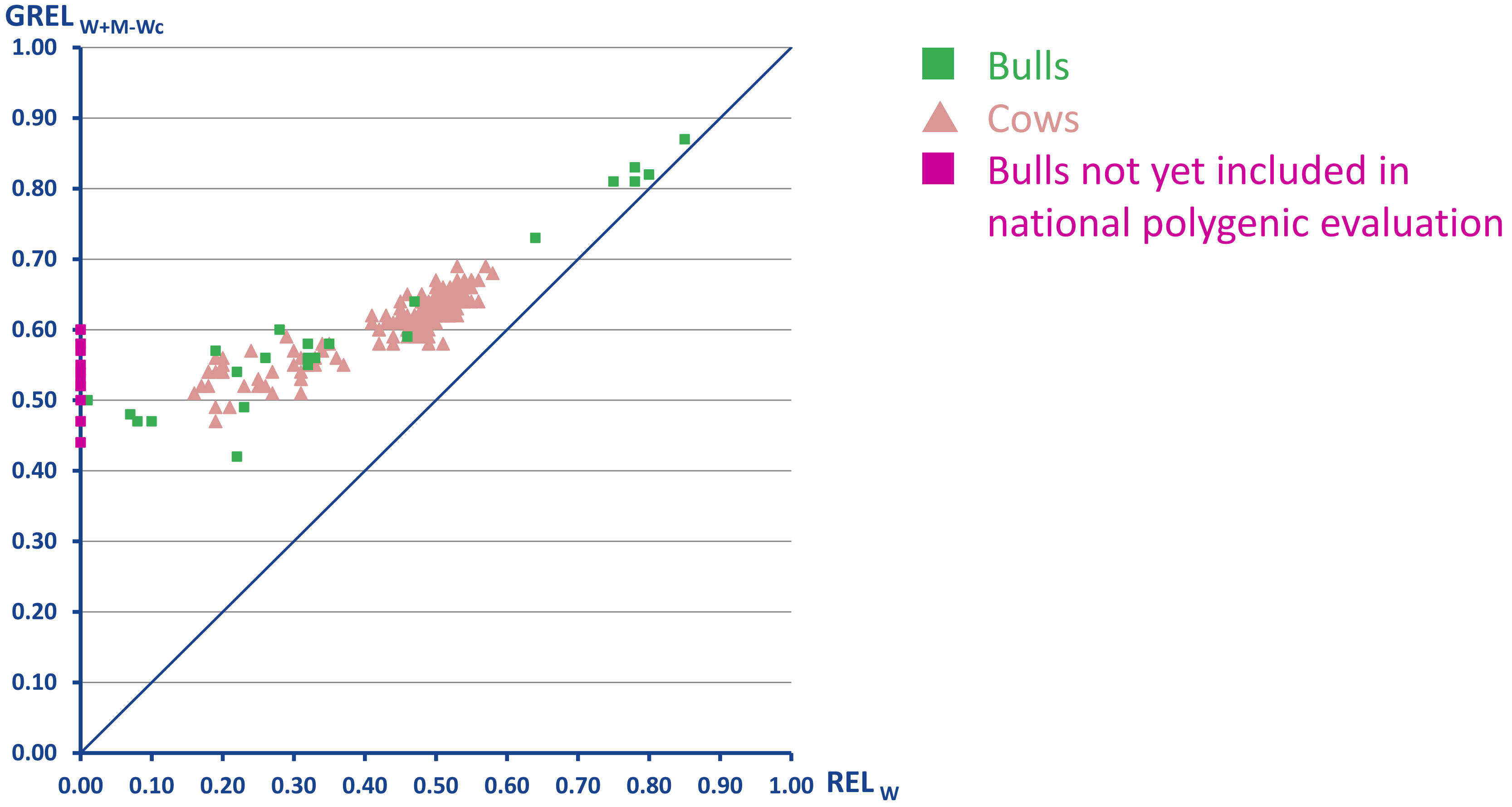
[1] REL obtained from Walloon polygenic evaluation

[2] REL obtained from Walloon genomic evaluation using only EBV_W

[3] REL obtained from Walloon genomic evaluation using EBV_W, EBV_M and EBV_{Wc}

Results

- ✓ Increase of REL for genotyped animals without MACE results and sired by genotyped bulls with MACE results



Conclusions

- ✓ Applied to Walloon dairy genomic evaluations
  - ☐ Bayesian approach integrates well MACE results into ssGBLUP

→ Recovers large amount of phenotypic information

  - ☐ More accurate predictions for genotyped animals and their progeny

→ Correct propagation of all available information

General notation

✓ Combining “s” sources of information

$$\left(G^{*-1} + \sum_{i=1,s} \boxed{\Lambda_i} \right) \hat{u}_c^{*} = \sum_{i=1,s} \boxed{D_i^{-1} \hat{u}_i}$$

Least square part of LHS of hypothetical BLUP
for information source i

RHS of hypothetical BLUP
for information source i

→ Potential to improve current genomic prediction strategies

Further implications ssGBayes

- ✓ No deregression
  - ☐ Direct use of EBV from multiple sources
- ✓ Applicable to multi-trait models
  - ☐ *E.g.*, external information for correlated and/or predictor traits
- ✓ Open general framework, can be modified to accomodate latest genomic models, *e.g.:*
  - ☐ GWAS models based on ssGBLUP
  - ☐ SNP based single-step models

Thank you for your attention

Wallonie

DGO 3

WBI
Wallonie-Bruxelles International

fnrs
LA LIBERTÉ DE CHERCHER

✓Acknowledgments for financial support

- Service Public of Wallonia SPW – DGO3 through project D31-1274 DairySNP
- National Fund for Scientific Research and Wallonie Brussels International for scientific stays

✓Acknowledgments

- Animal Breeding Association for providing data
- CECI for computational resources
- Animal and Dairy Science Department, University of Georgia, Athens, USA
- Animal Science Department, University of Ljubljana, Slovenia