Use of Bivariate EBV-DGV Model to Combine Genomic and Conventional Breeding Value Evaluations

Esa Mäntysaari

Ismo Strandén

Genetic Research

MTT Agrifood Research

Finland

esa.mantysaari@mtt.fi

Contents of the presentation

Introduction

- Why to combine genomic evaluations with conventional evaluations

- Methods used for combining

Bivariate combining and blending method

An example of use of bivariate combining and blending

Information available:

EBVs

Estimated Breeding Values

- Available for all animals in population
- Contain information from records and pedigree
- Same EBV for all full sibs until own information
- Will not be affected by DGVs of relatives

DGVs

Direct genomic values

Available only for animals genotyped

Will not be (directly) affected by own information (production or daughters, or EBV)

Definition of terms:

GEBV

Genomic Enhanced Breeding value

Estimate of BV that uses both the EBV and DGV information

If calculated for animals genotyped: **Combining**

If calculated for all animals: **Blending**

WHY GEBV?

Additive BV will not be 100% explained by SNP panel

Animals can have both types of information

National genetic evaluations are unbiased only if all information is included !

Methods for combining

Selection index developments

(Van Raden, 2008, Berry, 2009, etc.)

$$\mathbf{GEBV = b_1{*}DGV + b_2{*}EBV_{subset} + b_3{*}EBV_{national}}$$

- EBV_{subset}
  is an EBV calculated using animals in reference data set only

- The **EBV_{subset}** is used to avoid the double counting of information in DGV that is already in **$EBV_{national}$**

Methods for combining and back blending

Genomic *Equivalent Daughter Performances*[§] (EDP^G)

Combine the DGV into national BLUP using **EDP^G** pseudo records

1. Form genomic **EDC**s (genomic equivalent daughter contributions) that reflect amount of information in DGVs
   - The genomic information has to be adjusted to account the information from relationships
2. Form pseudo records from DGVs, i.e. **EDP^G**
3. Plug the G**EDC**s and **EDP^G**s into national evaluations, or pseudo evaluations

Method will automatically combine and blend the genomic information into EBVs of genotyped and non-genotyped animals

[§] Ducrocq and Liu (2009)**,** *Interbull Bulletin 40:172-177*

Methods for combining and back blending

Simultaneous fitting of genomic information and phenotypic records

Christensen et al. 2009, Misztal and Aquilar et al. (2009)

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \alpha\mathbf{H}^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a}_i^G \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix}$$

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

Methods for combining and back blending

Bivariate model

Equivalent daughter performance method has inherent problem on balancing the information from genomics and performance records from relatives

- the DGV are considered being records from daughters while actually being records of individual with a h^2=100%

In the *bivariate model approach* each animal is evaluated for the trait Y and for the DGV of the trait

- Accuracy and information is transferred from DGV to EBV_Y via genetic correlation
- DGV can have a heritability of 100%

Introduction

- Why to combine genomic evaluations with conventional evaluations
- Methods used for combining

Bivariate combining and blending method

An example of use of bivariate combining and blending

Bivariate blending model

Bivariate (sire) model:

$$\begin{bmatrix} DYD \\ DGV \end{bmatrix}_i = \begin{bmatrix} Z_{1\mathrm{i}} & 0 \\ 0 & Z_{2\mathrm{i}} \end{bmatrix} \begin{bmatrix} a_{DYD} \\ a_{DGV} \end{bmatrix}_i + \begin{bmatrix} e_{DYDiIi} \\ e_{DGVi} \end{bmatrix}$$

Assume data scaled so that genetic variances are unity:

$$Var\begin{bmatrix} a_{DYD} \\ a_{DGV} \end{bmatrix} = \begin{bmatrix} \sigma^2{}_{a_{DYD}} & \sigma_{a_{DGV,DYD}} \\ \sigma_{a_{DGV,DYD}} & \sigma^2{}_{a_{DGV}} \end{bmatrix} = \begin{bmatrix} 1 & r_a \\ r_a & 1 \end{bmatrix} = G$$

- Note that:

$$Var[a_{DYD} \mid a_{DGV}] = (1 - r_a^2)$$

- Which suggests the proper r_a to be $\sqrt{R^2_{DGV}}$ i.e. the accuracy of DGV
  - Then the reliability of evaluation (of the trait Y) is R^2_{DGV} for an animal with DGV but without DYD observation

- **Define and set residual variance after scaling**

$$Var\begin{bmatrix} e_{DYD} \\ e_{DGV} \end{bmatrix}_i = \begin{bmatrix} \sigma^2_{e,Y} / EDC_i & 0 \\ 0 & 1/100 \end{bmatrix}$$

- **Assumption** $Var[e_{DGV}] = 1/100$ **makes** $h^2_{DGV} = 0.99$

  - **The** $\sigma^2_{e,Y}$ **is scaled to lead to a proper heritability for the trait Y**

Using individual $R^2_{DGV_i}$ each animal

- In above all animals had the same R^2_{DGV}
- This can be relaxed by re-defining the random effects with **random regression coefficients** and independent **random regression breeding values**

$$Var\left\{\begin{bmatrix} Z_{1i} & 0 \\ 0 & Z_{2i} \end{bmatrix}\begin{bmatrix} a_{DYD} \\ a_{DGV} \end{bmatrix}_i\right\} = Z_i\, G\, Z^t{}_i$$

$$\begin{bmatrix} Z_{1i} & 0 \\ 0 & Z_{2i} \end{bmatrix}\begin{bmatrix} L_{11} & 0 \\ L_{12} & L_{22} \end{bmatrix}_i\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} L^t{}_{11} & L^t{}_{12} \\ 0 & L^t{}_{22} \end{bmatrix}_i\begin{bmatrix} Z^t{}_{1i} & 0 \\ 0 & Z^t{}_{2i} \end{bmatrix}$$

- Now choose different decomposition matrix **L** for each animal:

$$Z^*_i = \begin{bmatrix} Z_{1i} & 0 \\ 0 & Z_{2i} \end{bmatrix} \begin{bmatrix} L_{11} & 0 \\ L_{12} & L_{22} \end{bmatrix}_i = \begin{bmatrix} Z_{1i} & 0 \\ 0 & Z_{2i} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \sqrt{R^2_{DGV,i}} & \sqrt{1 - R^2_{DGV,i}} \end{bmatrix}$$

- In principle this is simple RR model where the covariables for DYD are always 1.0 and the covariables for DGV are $\sqrt{R^2_{DGV,i}}$ and $\sqrt{1 - R^2_{DGV,i}}$

Use of DGV of bulls in reference population

- If the bulls’ DYD has been used to derive the prediction equations, the DYD information becomes double counted
- To avoid this we can decrease the information from DYD in the model as

$$EDC_i = EDC_i - df_i$$

- The reduction of information *df* can be derived by absorbing the a_{DGV} equation into a_{DYD} equations in conceptual mixed model equation

$$df_i = R^2_{DGV,i \in ref} / (1 - R^2_{DGV,i \in ref})$$

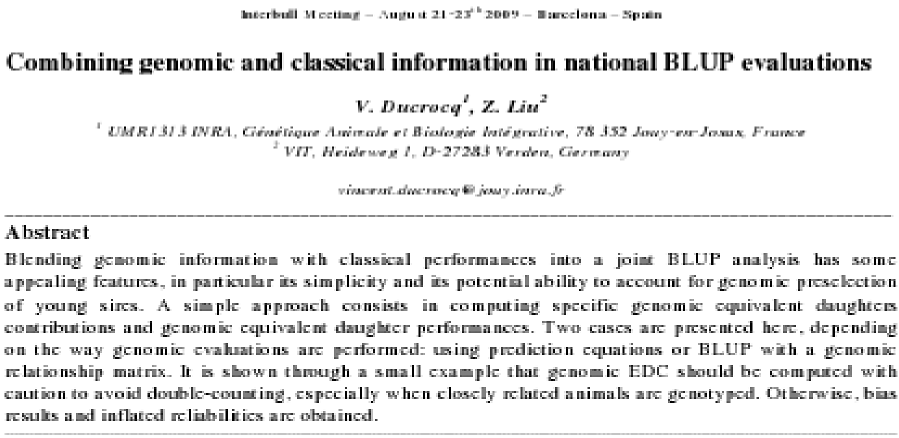
Introduction

- Why genomic evaluations are combined with conventional evaluations
- Previous suggestions for combining

Bivariate combining and blending method

An example of use of bivariate combining and blending

Example data and pedigree



Interbull Meeting – August 21-23th 2009 – Barcelona – Spain

Combining genomic and classical information in national BLUP evaluations

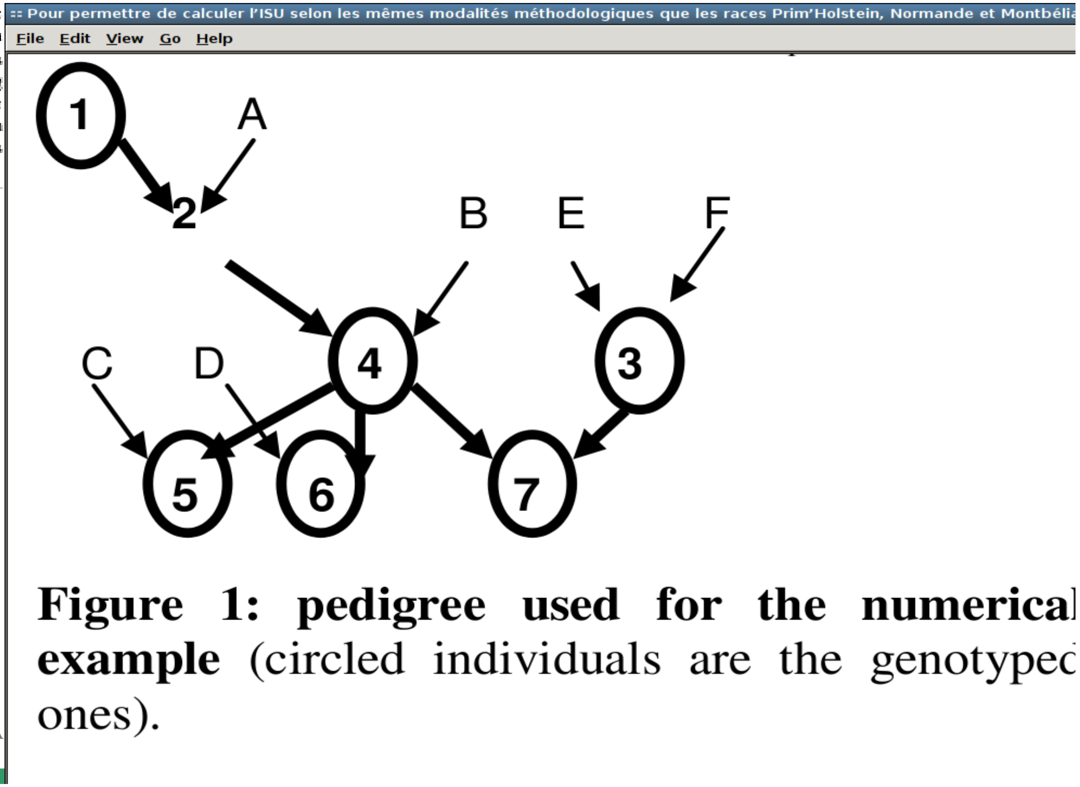
V. Ducrocq[1], Z. Liu[2]

[1] UMR1313 INRA, Génétique Animale et Biologie Intégrative, 78 352 Jouy-en-Josas, France
[2] VIT, Heideweg 1, D-27283 Verden, Germany

vincent.ducrocq@jouy.inra.fr

Abstract

Blending genomic information with classical performances into a joint BLUP analysis has some appealing features, in particular its simplicity and its potential ability to account for genomic preselection of young sires. A simple approach consists in computing specific genomic equivalent daughters contributions and genomic equivalent daughter performances. Two cases are presented here, depending on the way genomic evaluations are performed: using prediction equations or BLUP with a genomic relationship matrix. It is shown through a small example that genomic EDC should be computed with caution to avoid double-counting, especially when closely related animals are genotyped. Otherwise, bias results and inflated reliabilities are obtained.



:: Pour permettre de calculer l'ISU selon les mêmes modalités méthodologiques que les races Prim'Holstein, Normande et Montbélia

Figure 1: pedigree used for the numerical example (circled individuals are the genotyped ones).

Data with increasing complexity

BASE

Set up as in discussion of Ducrocq and Liu (2009)

①

- R^2_{DGV}=0.4 for all animals
- NOTE: Animal **4** had 10 sons altogether

Variable $R^2_{DGV,i}$

②

Pedigree and data as in BASE

but three bulls (4,5,6) had R^2_{DGV}=0.6

Data with increasing complexity 2.

BASE

Variable $R^2_{DGV,i}$

EBVs combined

Pedigree and data as in “variable $R^2_{DGV,i}$“,
but four bulls (1,4,5,6) had also 50 daughters

③

EDCs of reference bulls penalized

Pedigree and data as in “EBVs combined”,
but three of the bulls (1,4,5) had also been in the
reference population

④

Results: Reliabilities of the EBVs of animals after including more complexity to data

Animal	EDP^G (by D&L)	BASE 1	R^2_i 2	EDC_i 3	EDC_i-df_i 4
1	0.400	0.400	0.404	0.852	0.835
2	0.197	0.160	0.213	0.355	0.348
3	0.400	0.400	0.400	0.400	0.400
4	0.490	0.400	0.640	0.888	0.863
5	0.400	0.400	0.610	0.875	0.845
6	0.400	0.400	0.610	0.875	0.874
7	0.400	0.400	0.441	0.509	0.501
8-14	0.400	0.400	0.441	0.509	0.501

Results

Animal	EDP^G (by D&L)	BASE 1
1	0.400	0.400
2	0.197	0.160
3	0.400	0.400
4	0.490	0.400
5	0.400	0.400
6	0.400	0.400
7	0.400	0.400
8-14	0.400	0.400

Base scenario

- Animal 4 has 10 genotyped sons:
  - Equivalent to 100 grand daughters
  - Discounting of information by D&L can not keep the R^2 in 0.4 anymore
- Animal 2 has sire and son genotyped.
  - This gives R^2 =0.4 for it's DGV and $R^2_Y = 0.16$

Results

Animal	BASE 1	R^2_i 2
1	0.400	0.404
2	0.160	0.213
3	0.400	0.400
4	0.400	0.640
5	0.400	0.610
6	0.400	0.610
7	0.400	0.441
8-14	0.400	0.441

Individual $R^2_{DGV,i}$ for bulls

- Animals 4, 5 and 6 have $R^2_{DGV} = 0.6$
  - Surprisingly the individual R^2_{DGV} will converge to higher value than the original (0.60 → 0.61)

Results

Animal	R^2_i 2	EDC_i 3
1	0.404	0.852
2	0.213	0.355
3	0.400	0.400
4	0.640	0.888
5	0.610	0.875
6	0.610	0.875
7	0.441	0.509
8-14	0.441	0.509

Individual R^2_{DGV}

+ Animals 1, 4, 5 and 6 have EDC_i=50 daughters

- Accuracy will be about double
- Increase in also in R^2_{DGV} of sons of bull 4

Results

Animal	**EDC$_i$** **3**	**EDC$_i$-df$_i$** **4**
1	0.852	0.835
2	0.355	0.348
3	0.400	0.400
4	0.888	0.863
5	0.875	0.845
6	0.875	0.874
7	0.509	0.501
8-14	0.509	0.501

- Individual R^2_{DGV}

- Animals 1, 4, 5 and 6 EDC$_i$= 50

+ Animals 1, 4 and 5 been in reference group

- Small decrease in R^2_{DGV} but not very large
- Note that accuracy from 50 daughters would yield R^2_{DGV} of 0.83

Conclusions

- Bivariate blending gives logical accuracies for all combinations of records
- Transfers genomic information to non-genotyped relatives

- Bivariate blending is easy to do:
- Needs
  - DGV and their corresponding R^2_{DGV}
  - DYD or de-regressed proofs and their EDCs

- A BLUP program that allows RR and weights
  - Requires extra BLUP run, but with a simple model

$$Z_i = \begin{bmatrix} Z_{1i} & 0 \\ 0 & Z_{2i} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \sqrt{R^2_{DGV,i}} & \sqrt{1 - R^2_{DGV,i}} \end{bmatrix}$$

Acknowledgement

Nordic Genomic Evaluation Implementation group

Guosheng Su

Ole Fredslund Christensen

Rasmus Froberg Brøndum

Ulrik Sander Nielsen

Gert Pedersen Aamand

NAV

• Nordisk Avlsværdi Vurdering