Incorporating external information as observations in Dutch/Flemish genetic evaluations

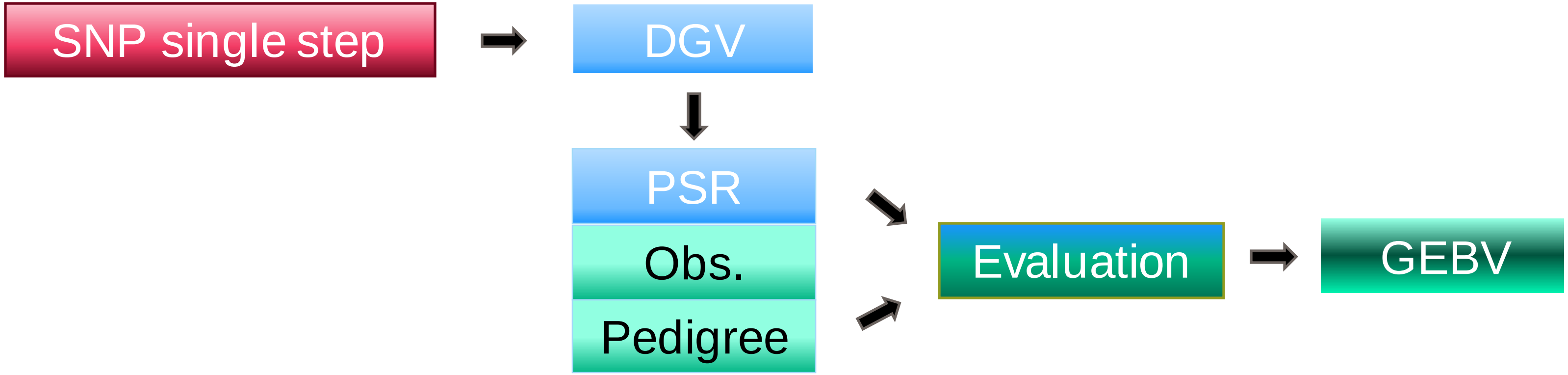
Herwin Eding | Interbull, May 19, 2024

CRV Coöperatie
BETTER COWS > BETTER LIFE

Introduction I

Introduction single step genomic evaluations
December 7, 2023

Pubished GEBV from pseudo-record system
DGV fitted as correlated traits



Coöperatie
BETTER COWS > BETTER LIFE

Introduction II

Trigger

- Incorporate DGV information of underlying traits
  - – Without (increasing) number of correlated traits (runtime!)
  - – Improved transfer of DGV to official evaluations.

- Deregressed proofs as observations on actual traits
  - No extra correlated pseudo-traits necessary
  - Limiting run time evaluation (avoid fitting extra traits)

Challenges

1. Observations ➔ Breeding values
2. Number of repeat records ➔ Reliabilities

Coöperatie
BETTER COWS > BETTER LIFE

Observations

Deriving observation records from GEBV

Two step process

1. Deregression
   – Linear deregression to remove national information from MACE proof

2. Transformation
   – Based on approach already in use.
   – Somewhat more formalized and simplified:

$$\mathbf{o} = \mathbf{Tb}$$

Coöperatie
BETTER COWS > BETTER LIFE

Observation: Deregression

For a list of eurogenomics bulls

External BV	XBV
National BV	EBV

Observation: Deregression

For a list of eurogenomics bulls

External BV XBV => $DRP_x = PA + (XBV-PA)/r_x$; $EOCx = \alpha * r_x/(1-r_x)$

National BV EBV => $DRP_e = PA + (EBV-PA)/r_e$; $EOCe = \alpha * r_e/(1-r_e)$

EOC = Expected Own Contribution; expected number of observation records

Coöperatie

BETTER COWS > BETTER LIFE

Observation: Deregression

For a list of eurogenomics bulls

External BV XBV => $DRP_x = PA + (XBV-PA)/r_x$; $EOCx = \alpha * r_x/(1-r_x)$

National BV EBV => $DRP_e = PA + (EBV-PA)/r_e$; $EOCe = \alpha * r_e/(1-r_e)$

EOC = Expected Own Contribution; expected number of observation records

Final result

EOC = (EOCx – EOCe)

DRP = [DRPx * EOCx – DRPe*EOCe] / EOC

(Pitkänen et al, 2019)

Coöperatie

BETTER COWS > BETTER LIFE

Observations: Transformation

The function

- **b** is a vector with *n* input DRP
- **o** is a vector with *m* output observations
- **T** is a *m* x *n* transformation matrix

$$\mathbf{o} = \mathbf{T}\mathbf{b}$$

Coöperatie
BETTER COWS > BETTER LIFE

Observations: Transformation

The function

$$\mathbf{o} = \mathbf{T}\mathbf{b}$$

– **b** is a vector with *n* input DRP
– **o** is a vector with *m* output observations
– **T** is a *m* x *n* transformation matrix

The transformation matrix

Important to distinguish between:

- Input trait: Which traits have DRP records?
- Analyzed traits: Which traits are in the evaluation?
- Observed traits: Which traits have observations?

Usually 2 or all 3 categories are identical, but not always!

Coöperatie
BETTER COWS > BETTER LIFE

Observations: Transformation matrix

$$\mathbf{o} = \mathbf{Tb}$$

Transformation matrix T can be obtained relatively easily

- Two 'phi'-matrices are needed that describe the relations between traits
  - Matrix **F** describes relation between input and analyzed traits
  - Matrix **D** describes relation between analyzed and output traits

- Additionally a genetic matrix **G** is needed (the one in the evaluation)

$$\mathbf{T} = (\mathbf{DGF'})(\mathbf{FGF'})^{-1}$$

Coöperatie
BETTER COWS > BETTER LIFE

Observaties: Transformation matrix conformation

For evaluations like Conformation
Single observation (lactation), single DRP
Input, analyzed and observed traits are identical

$$\mathbf{o} = \mathbf{Tb}$$

- **F** = **I** (identity matrix)
- **D** = **I**

- Result: $\mathbf{T} = (\mathbf{DGF'})(\mathbf{FGF'})^{-1} = \mathbf{GG}^{-1} = \mathbf{I}$

$$\mathbf{o} = \mathbf{Tb} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{b} = \mathbf{b}$$

CRV Coöperatie
BETTER COWS > BETTER LIFE

Observations: Transformation matrix index trait

For evaluations like Fertility, Udder health
Single GEBV/DRP; multiple underlying observed traits

$$\mathbf{o} = \mathbf{Tb}$$

- $\mathbf{F} = \mathbf{w}' = [\ 0.41\ 0.33\ 0.26\ ]$ ← lactation specific weights
- $\mathbf{D} = \mathbf{I}$

- Result: $\mathbf{T} = (\mathbf{DGF'})(\mathbf{FGF'})^{-1} = (\mathbf{Gw})(\mathbf{w'Gw})^{-1}$

Coöperatie
BETTER COWS > BETTER LIFE

Observations: Transformation matrix index trait

For evaluations like Fertility, Udder health
Single GEBV/DRP; multiple underlying observed traits

o = Tb

- **F** = **w**' = [0.41 0.33 0.26] 🡐 lactation specific weights
- **D** = **I**

- Result: **T** = (**DGF'**)(**FGF'**)$^{-1}$ = **(Gw)(w'Gw)**$^{-1}$

- Example:

$$\mathbf{C} = \begin{bmatrix} 1 & 0.7 & 0.6 \\ 0.7 & 1 & 0.8 \\ 0.6 & 0.8 & 1 \end{bmatrix} \quad \mathbf{v} = \begin{bmatrix} 9 \\ 16 \\ 25 \end{bmatrix} \quad \rightarrow \mathbf{G} = \begin{bmatrix} 9 & 8.4 & 9.0 \\ 8.4 & 16 & 16 \\ 9.0 & 16 & 25 \end{bmatrix}$$

$$\mathbf{T} = \begin{bmatrix} 0.933 \\ 1.048 \\ 1.045 \end{bmatrix} \quad \mathbf{b} = [5.0] \quad \mathbf{o} = \mathbf{Tb} = \begin{bmatrix} 4.7 \\ 5.2 \\ 5.2 \end{bmatrix}$$

Observations: Transformation Matrix Test Day Model

For random regression evaluations like Fertility, Udder health
Production test day model is an example that needs both **F** and **D** matrix

– INPUT : DRP from cumulative 305 day breeding values
– ANALYZED : Legendre regression polynomes
– OUTPUT : Production on day 60 of lactation

Construction of matrices

- For **F** 305 day factors needed: **s** = [212.84 -102.16 -57.96 -1.12 32.48]'

- For **D** factors needed for day 60: **t** = [0.707 -0.900 0.491 0.206 -0.794]'

Coöperatie
BETTER COWS > BETTER LIFE

Observations: Transformation matrix TDM

Example: Three lactations, 5 legendre regressions per lactations

$\mathbf{F} = \mathbf{s} \otimes \mathbf{I}_3 =$

212.8391	0	0	-102.156	0	0	-57.9619	0	0	-1.1099	0	0	32.4796	0	0
0	212.8391	0	0	-102.156	0	0	-57.9619	0	0	-1.1099	0	0	32.4796	0
0	0	212.8391	0	0	-102.156	0	0	-57.9619	0	0	-1.1099	0	0	32.4796

$\mathbf{D} = \mathbf{t} \otimes \mathbf{I}_3 =$

0,70711	0	0	-0,90011	0	0	0,49048	0	0	0,20577	0	0	-0,79362	0	0
0	0,70711	0	0	-0,90011	0	0	0,49048	0	0	0,20577	0	0	-0,79362	0
0	0	0,70711	0	0	-0,90011	0	0	0,49048	0	0	0,20577	0	0	-0,79362

Observations: Transformation matrix TDM

Application example: Three lactations of milk production in kg

$$\mathbf{T} = (\mathbf{DGF'})(\mathbf{FGF'})^{-1}$$

$$\mathbf{T} = \begin{bmatrix} 3.55\times10^{-3} & -3.59\times10^{-4} & -1.92\times10^{-4} \\ 6.01\times10^{-5} & 3.52\times10^{-3} & -5.07\times10^{-4} \\ 4.50\times10^{-4} & -1.33\times10^{-4} & 2.85\times10^{-3} \end{bmatrix}$$

$$\mathbf{b} = \begin{bmatrix} +1100 \\ +1300 \\ +1400 \end{bmatrix} \quad \Rightarrow \quad \mathbf{o} = \mathbf{Tb} = \begin{bmatrix} +3.2 \\ +3.9 \\ +4.3 \end{bmatrix}$$

CRV Coöperatie
BETTER COWS > BETTER LIFE

Observations

Transformation for every animal with DRP with a simple function

Transformation matrix **T** is a constant, needs to be constructed only once.

Accounts for genetic correlations among traits

Applicable to a variety of models, input DRP

Repeat records

Coöperatie
BETTER COWS > BETTER LIFE

Repeat records

Number of repeat records determines reliability of DRP information in genetic evaluations

At a reliability *r* of DRP we can calculate *expected own contributions* (EOC)

$$\mathrm{EOC} = \frac{1-h^2}{h^2} \times \frac{r}{1-r}$$

Example: If h^2 = 0.20 and r = 0.75 then EOC = 3 ➔ Number of repeat records

Coöperatie
BETTER COWS > BETTER LIFE

Repeat records

Number of repeat records determines reliability of DRP information in genetic evaluations

At a reliability *r* of DRP we can calculate *expected own contributions* (EOC)

$$\mathrm{EOC} = \frac{1-h^2}{h^2} \times \frac{r}{1-r}$$

Example: If h^2 = 0.20 and r = 0.75 then EOC = 3 ➔ Number of repeat records

But…
Calculated this way, the EOC is valid for single trait analysis only.

- Does not account for correlations between traits in MT evaluations

Determining number of repeat records

- Reliability is a function **r** = *rel_liu* (**G**, **F**, **Y**)
  - **G** is the genetic covariance matrix
  - **F** is a ‘phi’ – matrix comparable to before (for TDM: 305 day matrix)
  - **Y** is a MT-EDC matrix following Liu *et al*. (2001)
    - But **Y** is enumerated using **D** (for TDM: day 60 matrix)

The objective is to find a $\mathbf{Y}_{est}$ such that $\mathbf{r}_{est} \approx \mathbf{r}_{DRP}$

Coöperatie
BETTER COWS > BETTER LIFE

Determining number of repeat records

- Reliability is a function **r** = *rel_liu* (**G**, **F**, **Y**)
  - **G** is the genetic covariance matrix
  - **F** is a 'phi' – matrix comparable to before (for TDM: 305 day matrix)
  - **Y** is a MT-EDC matrix following Liu *et al*. (2001)
    - But **Y** is enumerated using **D** (for TDM: day 60 matrix)

The objective is to find a $\mathbf{Y}_{est}$ such that $\mathbf{r}_{est} \approx \mathbf{r}_{DRP}$

$$\mathbf{Y} = 4\mathbf{O}\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}$$

- **Z** is a genetic effect matrix comparable to **D** (day 60 matrix) in the previous
- **R** is the residual covariance matrix (the one used in analysis)
- **O** is a diagonal matrix with number of repeats per trait

Coöperatie
BETTER COWS > BETTER LIFE

Repeat records: Iterative approach

Goal: To optimize the matrix **O** such that $\mathbf{r}_{est} \approx \mathbf{r}_{DRP}$

Start: Let **O** = **E** (single trait EOC of input traits, DRP)

1. Calculate $\mathbf{Y} = 4(\mathbf{OZ})'\mathbf{R}^{-1}\mathbf{Z}$
2. Calculate $\mathbf{r}_{est}$ = *rel_liu*(**G**, **F**, **Y**)
3. Compare $\mathbf{r}_{est}$ with $\mathbf{r}_{DRP}$
   a. If $\mathbf{r}(i)_{est} > \mathbf{r}(i)_{DRP}$ ➔ $\mathbf{O}(i,i) = \mathbf{O}(i,i) - 1$ (minimum value 0)
   b. If $\mathbf{r}(i)_{est} < \mathbf{r}(i)_{DRP}$ ➔ $\mathbf{O}(i,i) = \mathbf{O}(i,i) + 1$
4. Repeat until convergence or until **O** stops changing.

CRV Coöperatie
BETTER COWS > BETTER LIFE

Final remarks

Method provides capability to use DRP as observations on existing traits

Unified approach, valid for all types of models, DRP

- Single trait, indices, random regressions

No additional correlated traits

- No need for additional software
- Same statistical model, parameters and matrices
- Same output format

Approach is being implemented

Application to all test day model traits and claw health traits

Further implementation for all evaluations planned for later this year

Coöperatie
BETTER COWS > BETTER LIFE

Coöperatie
BETTER COWS > BETTER LIFE