Two blending approaches

MiX99 course on genomic prediction

COURSE DAY 1, March 9th, 2026

Contents

- Introduction
- Theory in brief
- Going through details using small example
- How to get variance components for indices
- Conclusions

What is blending?

- In this presentation, blending means integration of external information, usually from Interbull MACE, to domestic evaluation
- Several different methods exists
    - Bayesian methods, where external information is used as prior for breeding values
    - Pseudo-observation-based methods, where DRPs weighted with ERCs are included to domestic model
- The pseudo-observation-based method will be explained, with two different implementations, for models with and without covariance functions

Steps in blending

Blending approach has three steps

1. Calculation of amount of external information for selected animals
2. Calculation of pseudo-observations for animals
3. Running evaluation model with pseudo-observations

Selecting which animals to blend

- Aim is to choose animals that have additional information in external evaluation compared to domestic evaluation
- Can be chosen by comparing R^2_{dom} to R^2_{ext} , for example by requiring that

$$R^2_{ext} > 0.50 \text{ and } R^2_{ext} - R^2_{dom} > 0.01 \text{ for all traits}$$

- Thresholds for choosing animals can be modified depending on the traits

1. Calculation of amount of external information for selected animals

ERC (effective record contribution)

- **ERC** for external animals is calculated based on R^2 for domestic and external (MACE) evaluations
- Apax99 provides single and multitrait ERC calculation
- The amount of additional information in external evaluation calculated for each trait separately:

$$\mathbf{ERC_{blend} = ERC_{external} - ERC_{domestic}}$$

2. Calculation of pseudo-observations for animals

DRP (deregressed proofs)

- $\mathbf{DRP_{ext}}$ and $\mathbf{DRP_{dom}}$ for selected external animals is calculated based on corresponding **ERC** and **EBV**
- Pseudo-observation for external information

$$DRP_{blend} = \frac{ERC_{ext} * DRP_{ext} - ERC_{dom} * DRP_{dom}}{ERC_{ext} - ERC_{dom}}$$

- Weight for observation

$$Weight = ERC_{ext} - ERC_{dom}$$

- DRP_{blend} contains additional information in external evaluation

3. Running evaluation model with pseudo-observations

This step depends on the model and traits in question

- Data, model definition, and variance components needs to be modified
- Assumes genetic correlation of 1 between external and domestic evaluation
- Next example will illustrate necessary modifications

Example

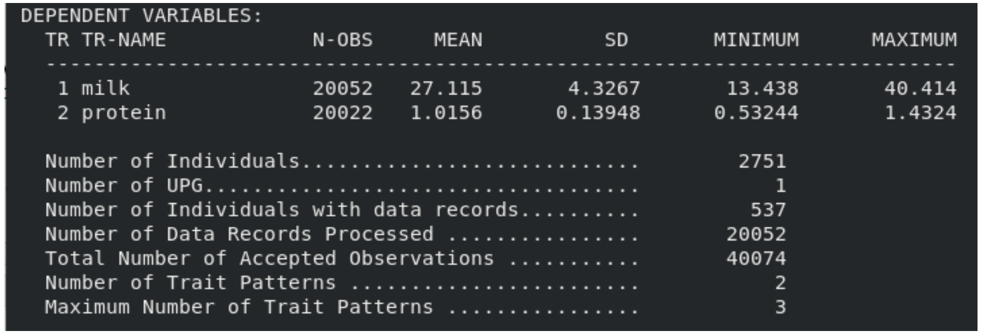
Simple test-day model for milk and protein 1 lactation

```
MODEL
milk     =  CalvAge(age age2)  Lac(t1 t2 t3 t4 t5)  HYS  pe(t1 t2 t5|ID)  g(t1 t2 t5|ID)
protein  =  CalvAge(age age2)  Lac(t1 t2 t3 t4 t5)  HYS  pe(t1 t2 t5|ID)  g(t1 t2 t5|ID)

RANDOM     pe g
PEDIGREE   g  am+p 0.3333
```

- The model is fitted for two data sets: Luke only, and Luke + 3 other herds data (FULL).
- Blending approach is demostrated by blending information from FULL data set to Luke evaluation.
- Information is blended as 305d DRPs.
- Assume genetic correlation 1 between evaluations!

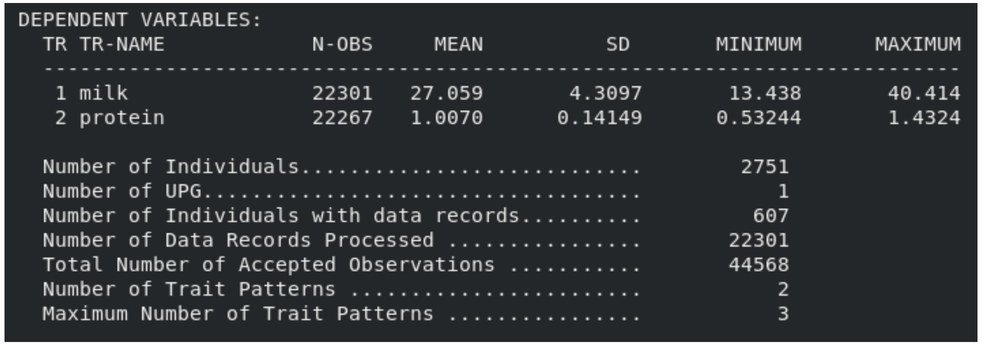
Data description

LUKE data



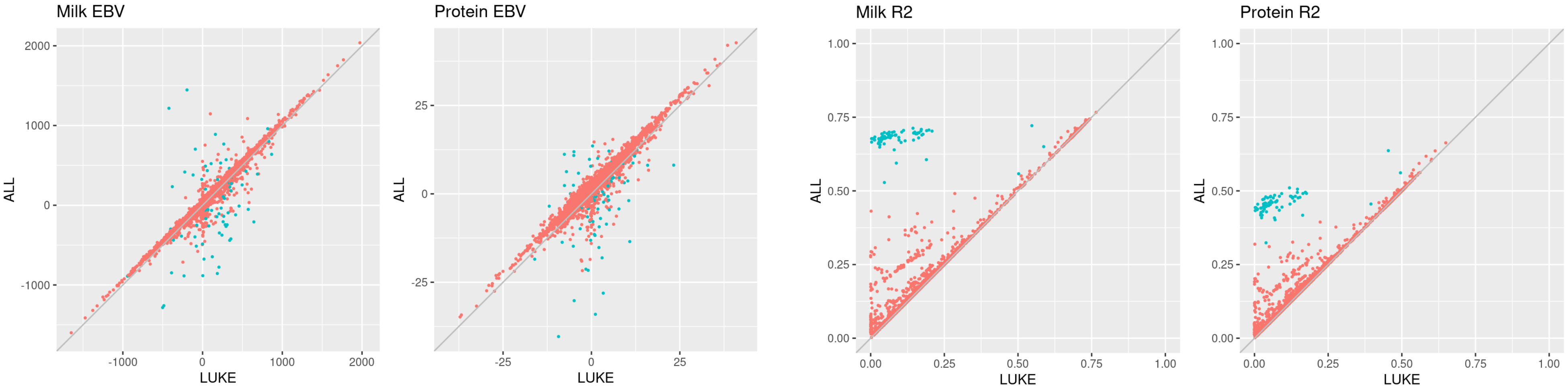
```
DEPENDENT VARIABLES:
  TR TR-NAME             N-OBS     MEAN           SD       MINIMUM       MAXIMUM
  ---------------------------------------------------------------------------
   1 milk                20052   27.115       4.3267        13.438        40.414
   2 protein             20022   1.0156      0.13948       0.53244        1.4324

  Number of Individuals...........................          2751
  Number of UPG...................................             1
  Number of Individuals with data records..........          537
  Number of Data Records Processed ................        20052
  Total Number of Accepted Observations ...........        40074
  Number of Trait Patterns ........................            2
  Maximum Number of Trait Patterns ................            3
```

FULL data



```
DEPENDENT VARIABLES:
  TR TR-NAME             N-OBS     MEAN           SD       MINIMUM       MAXIMUM
  ---------------------------------------------------------------------------
   1 milk                22301   27.059       4.3097        13.438        40.414
   2 protein             22267   1.0070      0.14149       0.53244        1.4324

  Number of Individuals...........................          2751
  Number of UPG...................................             1
  Number of Individuals with data records..........          607
  Number of Data Records Processed ................        22301
  Total Number of Accepted Observations ...........        44568
  Number of Trait Patterns ........................            2
  Maximum Number of Trait Patterns ................            3
```

EBVs and R2 and choosing which animals are blended



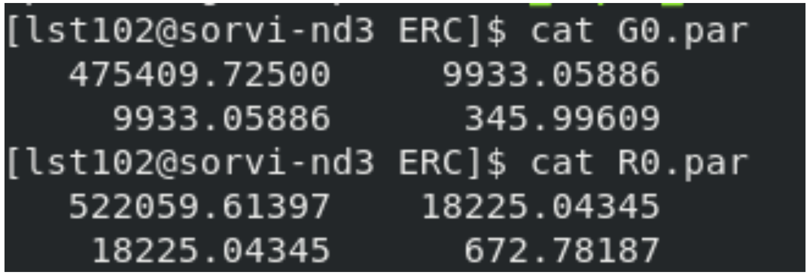
- Animals having **FULL MILK R2 – LUKE MILK R2 > 0.05** and **FULL MILK R2 > 0.50** were chosen to be blended. (Marked with different color)
- In total 70 cows and 3 bulls were chosen

Calculating ERC

```
#-----------------------------------------------------------------------------
# AccurType:     Reliability method
               40  p   # multitrait ERC calculations
# Reliabilities: Filename of the reliability information (IDs and r2s)
               R2forBLEND.dat
# Pedigree:
                resfarm.pedi
# Number of ERCs:
               2
# G-matrix:
               G0.par
# R-matrix:
               R0.par
# tol_ERC:       tolERC MaxIter_ERC smallestERC tolNewton MaxIter_Newton
               1e-8   100         1e-10        1e-4      20
# tol_r2:       tolr2  MinIter Proportion_of_iter Num_LastIter_meet_tolr2
               1e-7   10      0.5                 10
#-----------------------------------------------------------------------------
```

Separately for FULL and Luke evaluation:

- Full pedigree
- Only R2s for 73 animals!
- G0 and R0 for 305d variances

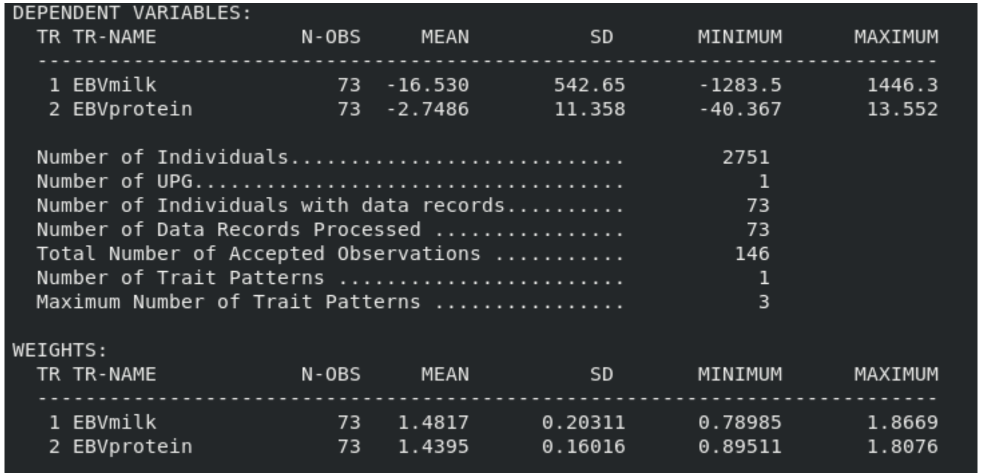


```
[lst102@sorvi-nd3 ERC]$ cat G0.par
   475409.72500      9933.05886
     9933.05886       345.99609
[lst102@sorvi-nd3 ERC]$ cat R0.par
   522059.61397     18225.04345
    18225.04345       672.78187
```

Calculating DRP

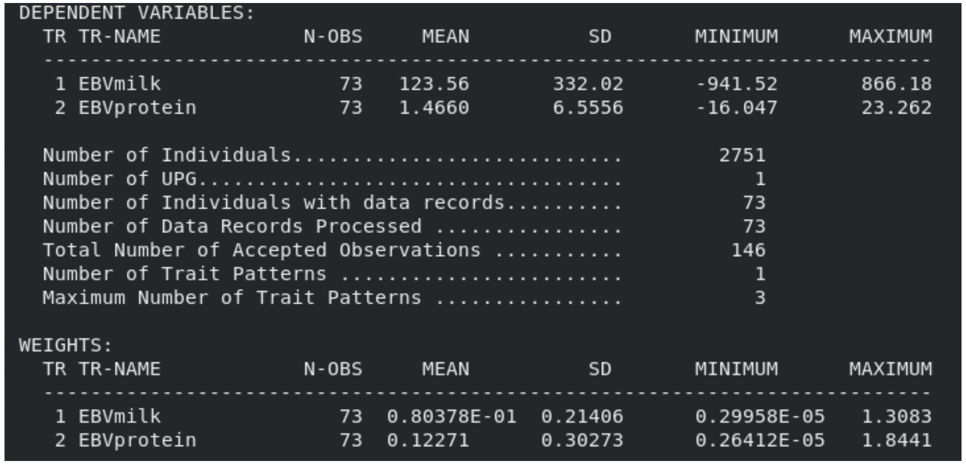
```
TITLE       MT-DRP calculation

DATAFILE    DRPINPUT.dat
PEDFILE     resfarm.pedi
PARFILE     DRPcalculation.para

INTEGER     ID MEAN
REAL        EBVmilk EBVprotein ERCmilk ERCprotein
MISSING     -99999.0

MODEL
EBVmilk    = MEAN  g(ID) ! WEIGHT=ERCmilk
EBVprotein = MEAN  g(ID) ! WEIGHT=ERCprotein

RANDOM      g
PEDIGREE    g  am+p 0.3333
TMPDIR      ./tmpMiX
```

FULL



```
DEPENDENT VARIABLES:
  TR TR-NAME             N-OBS      MEAN           SD        MINIMUM       MAXIMUM
  ------------------------------------------------------------------------------
   1 EBVmilk                73   -16.530       542.65        -1283.5        1446.3
   2 EBVprotein             73   -2.7486       11.358        -40.367        13.552

  Number of Individuals...........................        2751
  Number of UPG...................................           1
  Number of Individuals with data records.........          73
  Number of Data Records Processed ...............          73
  Total Number of Accepted Observations ..........         146
  Number of Trait Patterns .......................           1
  Maximum Number of Trait Patterns ...............           3

WEIGHTS:
  TR TR-NAME             N-OBS      MEAN           SD        MINIMUM       MAXIMUM
  ------------------------------------------------------------------------------
   1 EBVmilk                73    1.4817      0.20311        0.78985        1.8669
   2 EBVprotein             73    1.4395      0.16016        0.89511        1.8076
```

LUKE



```
DEPENDENT VARIABLES:
  TR TR-NAME             N-OBS      MEAN           SD        MINIMUM       MAXIMUM
  ------------------------------------------------------------------------------
   1 EBVmilk                73    123.56       332.02        -941.52        866.18
   2 EBVprotein             73    1.4660       6.5556        -16.047        23.262

  Number of Individuals...........................        2751
  Number of UPG...................................           1
  Number of Individuals with data records.........          73
  Number of Data Records Processed ...............          73
  Total Number of Accepted Observations ..........         146
  Number of Trait Patterns .......................           1
  Maximum Number of Trait Patterns ...............           3

WEIGHTS:
  TR TR-NAME             N-OBS      MEAN           SD        MINIMUM       MAXIMUM
  ------------------------------------------------------------------------------
   1 EBVmilk                73  0.80378E-01  0.21406       0.29958E-05   1.3083
   2 EBVprotein             73  0.12271      0.30273       0.26412E-05   1.8441
```

Calculating pseudo-observation and weight

Separately for milk and protein:

- $DRP_{blend} = \frac{ERC_{ext} * DRP_{ext} - ERC_{dom} * DRP_{dom}}{ERC_{ext} - ERC_{dom}}$
- $Weight = ERC_{ext} - ERC_{dom}$

- For 3 animal weight for milk was negative
- observations for these animals were omitted in blending

⇒In total 70 animals will be blended

Model allowing DRPs to be included

- Observations are test-day records, modelled by random regression curve

$$\boldsymbol{Y_{i,dim}} = \boldsymbol{X\beta} + \boldsymbol{\Phi_{dim} p_i} + \boldsymbol{\Phi_{dim} u_i} + \boldsymbol{\epsilon_{i,dim}}$$

- Model DRPs with Φ_{305} :

$$\boldsymbol{DRP_i} = \boldsymbol{\mu} + \boldsymbol{\Phi_{305} u_i} + \boldsymbol{\epsilon_{i,drp}},$$

- Assume the same u_i for test-day observations and DRPs!, ie genetic correlation = 1
- Variation due to PE and residual is included to $\epsilon_{i,drp}$.

CLIM requires some changes

Original model

```
MODEL
milk     =  CalvAge(age age2)  Lac(t1 t2 t3 t4 t5)  HYS  pe(t1 t2 t5|ID)  g(t1 t2 t5|ID)
protein  =  CalvAge(age age2)  Lac(t1 t2 t3 t4 t5)  HYS  pe(t1 t2 t5|ID)  g(t1 t2 t5|ID)
```

Blending model

```
MODEL
milk(1)     =  CalvAge(age age2)  Lac(t1:5)  HYS  pe(t1 t2 t5|ID)  g(t1 t6 t2 t6 t5 t6 |ID)@1
protein(1)  =  CalvAge(age age2)  Lac(t1:5)  HYS  pe(t1 t2 t5|ID)  g(t6 t1 t6 t2 t6 t5 |ID)@1
#DRPS
milk(2)     =  CalvAge( -     - )  Lac( D5 )  HYS  pe(-  -  - )     g(t1 t6 t2 t6 t5 t6 |ID)@1 ! WEIGHT = mwght
protein(2)  =  CalvAge( -     - )  Lac( D5 )  HYS  pe(-  -  - )     g(t6 t1 t6 t2 t6 t5 |ID)@1 ! WEIGHT = pwght
```

- () after trait name refers to trait group. 1 = test-day observation, 2 = DRP
- The ”@1” combines unknowns accros traits, here for genetic effect
- Covariable t6 is always zero.
- HYS for DRPs is always 1
- Covariable values for DRPs are 305d sum of DIM covariables!

Genetic effect in original model

```
MODEL
milk      = … g(t1 t2 t5|ID)
protein   = … g(t1 t2 t5|ID)
```

Breeding value for given DIM:

$$\mathrm{S}_{\mathrm{dim}}^{\mathrm{orig}} = \begin{bmatrix} \mathrm{t1}_{\mathrm{dim}} & 0 & \mathrm{t2}_{\mathrm{dim}} & 0 & \mathrm{t5}_{dim} & 0 \\ 0 & \mathrm{t1}_{\mathrm{dim}} & 0 & \mathrm{t2}_{\mathrm{dim}} & 0 & \mathrm{t5}_{dim} \end{bmatrix}$$

$$\mathrm{EBV}_{\mathrm{dim},i}^{\mathrm{orig}} = \mathrm{S}_{\mathrm{dim}}^{\mathrm{orig}}\, u_i$$

```
"Solani"-File: Solutions for the Individual Genetic Effects
    Column  |  Description
   ----------------------
       1       Individual ID
       2       Number of Descendants
       3       Number of Observations
       4       Solution for Trait  1  milk  and Factor  T1(ID)
       5       Solution for Trait  2  protein  and Factor  T1(ID)
       6       Solution for Trait  1  milk  and Factor  T2(ID)
       7       Solution for Trait  2  protein  and Factor  T2(ID)
       8       Solution for Trait  1  milk  and Factor  T5(ID)
       9       Solution for Trait  2  protein  and Factor  T5(ID)
```

Genetic effect in BLEND model

```
MODEL
milk(1)     = g(t1 t6 t2 t6 t5 t6 |ID)@1
protein(1) = g(t6 t1 t6 t2 t6 t5 |ID)@1
#DRPS
milk(2)      = g(t1 t6 t2 t6 t5 t6 |ID)@1 ! WEIGHT = mwght
protein(2)  = g(t6 t1 t6 t2 t6 t5 |ID)@1 ! WEIGHT = pwght
```

6 unknowns for genetic effect

Breeding value for given DIM:

$$\mathrm{S}_{\mathrm{dim}}^{\mathrm{orig}} = \begin{bmatrix} \mathrm{t1}_{\mathrm{dim}} & 0 & \mathrm{t2}_{\mathrm{dim}} & 0 & \mathrm{t5}_{dim} & 0 \\ 0 & \mathrm{t1}_{\mathrm{dim}} & 0 & \mathrm{t2}_{\mathrm{dim}} & 0 & \mathrm{t5}_{dim} \end{bmatrix}$$

$$\mathrm{EBV}_{\mathrm{dim},i}^{\mathrm{blend}} = \mathrm{S}_{\mathrm{dim}}^{\mathrm{orig}} u_i$$

```
 "Solani"-File: Solutions for the Individual Genetic Effects

    Column  |  Description
    ----------------------
       1       Individual ID
       2       Number of Descendants
       3       Number of Observations
       4       Solution for Trait  1  milk  and Factor  T1(ID)
       5       Solution for Trait  1  milk  and Factor  T6(ID)
       6       Solution for Trait  1  milk  and Factor  T2(ID)
       7       Solution for Trait  1  milk  and Factor  T6(ID)
       8       Solution for Trait  1  milk  and Factor  T5(ID)
       9       Solution for Trait  1  milk  and Factor  T6(ID)
```

Blending CLIM file

```
DATAFILE   BLENDINGFULL.dat
PEDFILE    resfarm.pedi   # same as in original
PARFILE    BLENDING.par   # Updated
TABLEFILE  BLEND.cov      # Updated
TABLEINDEX DIM

INTEGER    ID TGRP HERD HYS DIM  # TGRP included
REAL       age age2 milk protein mwght pwght  # Weights included
MISSING    -99999.0

TRAITGROUP TGRP    # TRAITGROUP included

DEFINE  D5  - - - - -

MODEL
milk(1)     =  CalvAge(age age2)  Lac(t1:5)  HYS  pe(t1 t2 t5|ID)  g(t1 t6 t2 t6 t5 t6 |ID)@1
protein(1)  =  CalvAge(age age2)  Lac(t1:5)  HYS  pe(t1 t2 t5|ID)  g(t6 t1 t6 t2 t6 t5 |ID)@1
milk(2)     =  CalvAge( -    - )  Lac( D5 )  HYS  pe(-  -  - )     g(t1 t6 t2 t6 t5 t6 |ID)@1  ! WEIGHT = mwght
protein(2)  =  CalvAge( -    - )  Lac( D5 )  HYS  pe(-  -  - )     g(t6 t1 t6 t2 t6 t5 |ID)@1  ! WEIGHT = pwght

WITHINBLOCKORDER g pe HYS
RANDOM     pe g
PEDIGREE   g  am+p 0.3333
```

Intercept for DRPs modelled in HYS effect

Blending data file

- Trait group column:

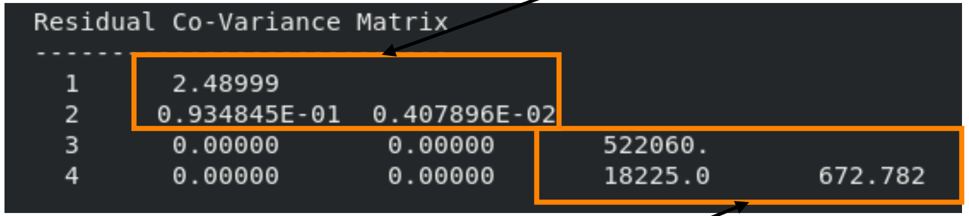
  1 for test-day records,

  2 for drps
- HYS for DRPs always 1.
- DIM for DRPs always 366
- Weights for TD records (tgrp =1) allways 1

```
# ID        TGRP HERD   HYS      DIM    age   age2      milk   protein mwght pwght
1320041726   1    2   20202403 298   0.334 0.11      27.97     1.23  1     1      # TDREC
1320041726   1    2   20202403 305   0.334 0.11      27.17     1.20  1     1      # TDREC
1210425818   2   99          1 366   0     0       2078.24    39.96  1.48  1.39   # DRPS
1210425820   2   99          1 366   0     0      -1824.14   -69.07  1.46  1.37   # DRPS
```

Blending parfile

- Variance components for PE and GE the same as in original
- Residual variance



```
Residual Co-Variance Matrix
------
  1     2.48999
  2   0.934845E-01  0.407896E-02
  3     0.00000        0.00000        522060.
  4     0.00000        0.00000        18225.0        672.782
```

Original residual covariance for test-day observations

Residual covariance matrix is for 305d observation and includes variation due to PE + residual

$$\mathbf{R}_{305} = \mathbf{S}\mathbf{K}_{\mathrm{p}}\mathbf{S}' + 305\times\mathbf{K}_{\mathrm{e}}$$

$$\mathbf{S} = \begin{bmatrix} 215.7 & 0 & -58.1 & 0 & 16.0 & 0 \\ 0 & 215.7 & 0 & -58.1 & 0 & 16.0 \end{bmatrix}$$

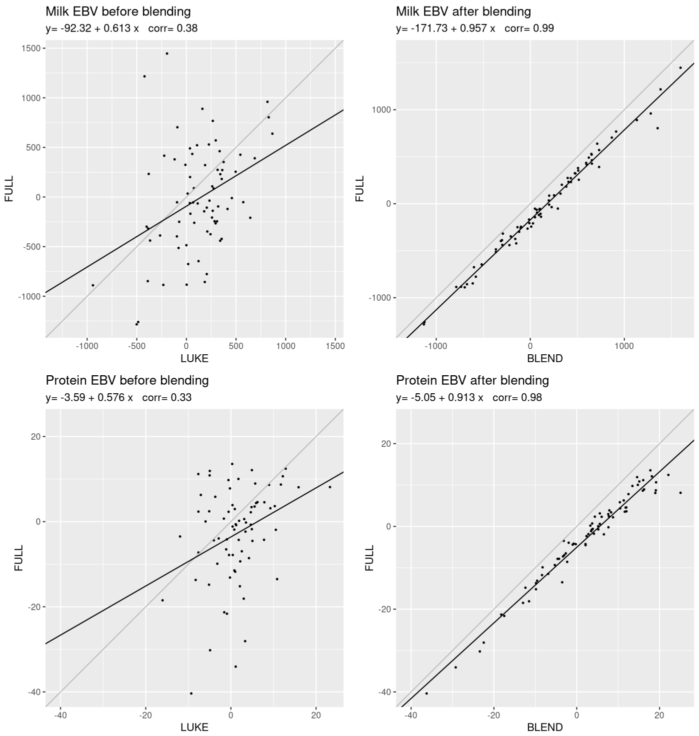
Blending covariable file

```
DIM      t1                 t2              t3                t4               t5              t6
363 0.707106781186548 1.21113659504279 1.52872700590177  1.74783261354842 1.31085650470662e-08 0
364 0.707106781186548 1.21794073321719 1.55485971712122  1.80889999717125 1.2469252785751e-08  0
365 0.707106781186548 1.22474487139159 1.58113883008419  1.87082869338697 1.18611201513438e-08 0
366 215.667568261897 -58.1073400093565 -50.546665969895 -31.5044729042412 15.9686571167008     0
```

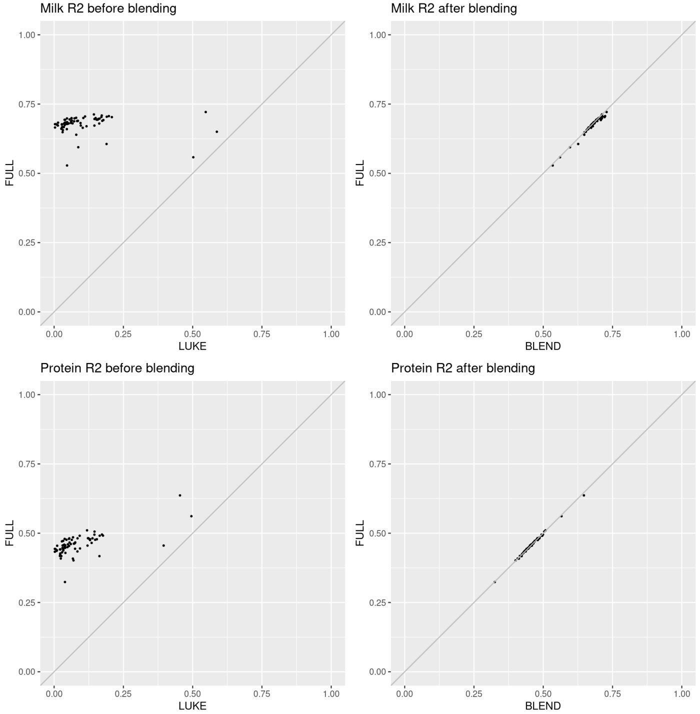
- Covariables 1-5 for DRP are calculated as: $t_{i;366} = \sum_{dim=8}^{312} t_{i;dim}$

Results for blended animals

EBVS



Reliabilities



Mu estimates:
-169.23 -4.8080

Conclusion, example

- The required steps were illustrated
- Blending method works as expected
- When blending MACE information, R2 for MACE evaluation is usually much higher compared to this example
- R2 for domestic evaluation can be low
- 3 animals had negative

$$Weight = ERC_{ext} - ERC_{dom}$$

- Those were omitted from blening

- In the example, 305d DRP for traits were

Alternative 2

Alternative 2: Use covariance functions

```
MODEL
milk(1)      =  CalvAge(age age2)  Lac(t1:5)  HYS  pe(t1 t2 t5|ID)  g( t7:12|ID)@1
protein(1)   =  CalvAge(age age2)  Lac(t1:5)  HYS  pe(t1 t2 t5|ID)  g(t13:18|ID)@1
#DRPS
milk(2)      =  CalvAge( -    - )  Lac( D5 )  HYS  pe(-  -  - )     g(t7:12 |ID)@1 ! WEIGHT = mwght
protein(2)   =  CalvAge( -    - )  Lac( D5 )  HYS  pe(-  -  - )     g(t13:18 |ID)@1 ! WEIGHT = pwght
```

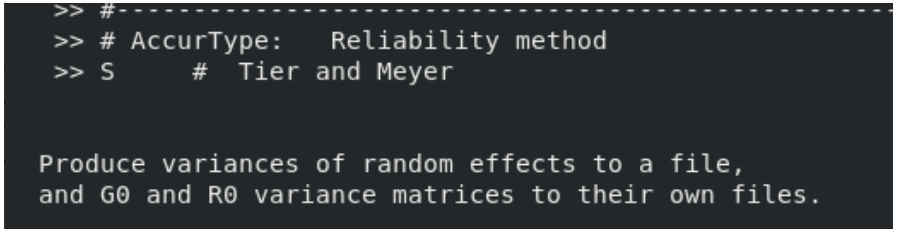
- Here var(a) = I
- The covariance structure between random effects is explained by covariables
- Covariance functions are used in Nordic test-day model
- Covariables for DRPs: $t_{i;366} = \sum_{dim=8}^{312} t_{i;dim}, i = 7, \ldots, 18$
- First steps are the same as for alternative 1

How to get 305d variance components

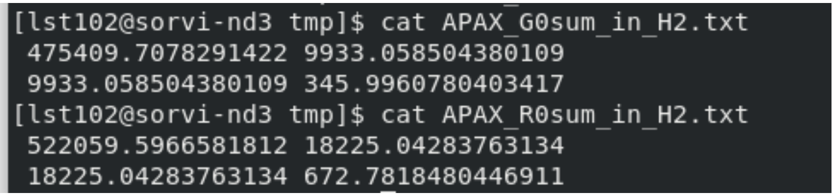
How to get variance components for EBVs or indices from apax99

```
•  # AccurType:   Reliability method
•                 S     #  Will calculate VC for indices defined
•  # MaxNonZ:     Number of non-zeros in sparse matrix
•                 100000
•  # firstDIM:    DIM of first row in covariable table
•                 5
•  # PHImatrix:   For each taits: first DIM, number of DIMs, step
   size
•                 5   305    1
•                 5   305    1
•  # OriginalDir: MiX99 directives file
•                 -
•  # NumBVs:      Number of breeding values for which r2 should be
   approximated
•                 2
•  # Weights:     Weights applied for the breeding values
•                 1 0
•                 0 1
•  # H2calc:      Random effects used for heritability calculation
•                 1     # also 1st random effect (pe) is included
•  # JFilter:     Within-block effect number from which approximate
   absorption is used
•                       #   1  2   3
•                       #   g  pe  HYS
•                 4
```

- Specify method ”S”
- It will calculate genetic and residual variance based on PHImatrix and Weights.
- In beta testing at the moment!



```
  >> #------------------------------------------------------------
  >> # AccurType:   Reliability method
  >> S     #  Tier and Meyer


 Produce variances of random effects to a file,
 and G0 and R0 variance matrices to their own files.
```



```
[lst102@sorvi-nd3 tmp]$ cat APAX_G0sum_in_H2.txt
 475409.7078291422 9933.058504380109
 9933.058504380109 345.9960780403417
[lst102@sorvi-nd3 tmp]$ cat APAX_R0sum_in_H2.txt
 522059.5966581812 18225.04283763134
 18225.04283763134 672.7818480446911
```

Conclusions

Conclusions

- Alternative 1 and 2 have been successfully implemented for multitrait test-day evaluations
- DRPs can be blended also to single step models
- In that case, rember that ERC and DRP calculation assumes simple animal model!

- If domestic ERC or DRP go to zero for an animal, it is best to set:

$$ERC_{dom} = 0,$$

$$DRP_{blend} = \frac{ERC_{ext} * DRP_{ext} - 0 * DRP_{dom}}{ERC_{ext} - 0} = DRP_{ext},$$

$$Weight = ERC_{ext}$$

Discussion