

## Chapter 18

### Multiple Trait Genetic Evaluation

*Julius van der Werf*

#### Multi trait mixed model

This is an extension of the single trait case. Data on a number of traits are available in  $Y$ , and EBV's are calculated for each trait. The results are generally different from what would be got from a number of separate single-trait BLUPs, because each trait is used to help give information about all other traits, much as with a selection index. In a later chapter, the multiple trait BLUP procedure will be worked out in more detail. The benefit from multiple trait models comes from

- more accuracy as information from correlated traits is used
- less bias as the analysis will take into account that for traits that are measured after sequential rounds of selection, only the better ones are evaluated.

An example of potential selection bias. Compare a good bull and a bad bull, each having 40 progeny at weaning. From the good bull, no progeny are culled, whereas from the bad bull 50% is culled. Comparing the progeny of these bulls at post-weaning will give a huge advantage to the bad bull, as his bad progeny have been removed. Multi-trait BLUP would correct for this bias.

For the genetic evaluation of the animals, we can use information which is available on all traits. Originally the main reason for using information on all traits was to obtain more accurate evaluations. With using information on correlated traits the accuracy of the estimated breeding value increases. A second advantage arose later, namely a multiple trait analysis is the only way to obtain unbiased estimates for a trait, which is observed only on animals selected based on values of a correlated trait. A model including information of the correlated trait, on which selection was based, is able to correct for this type of selection. An example of this is the evaluation of the second lactation productions of dairy cows where selection has been practised based on the first lactation. Only animals that survived the first lactation have a second lactation record, and those are usually only the better animals. Other examples are the analysis of piglets born in second litter, or the analyses of yearling weight after animals have been selected for weaning weight.

Taking again as a starting point the mixed model in its general form:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

With more traits we can now partition the observation vector  $\mathbf{y}$  in a part for each trait. The same can be done with the associated environmental effects. The vector of breeding values is also partitioned for the different traits, so that each animal has a breeding value for each trait in the analysis.

For a 2-trait example, the vector  $\mathbf{y}_1$  represents the  $n_1$  observations for trait 1 and  $\mathbf{y}_2$  represents  $n_2$  observations for trait two. For each trait we can write a mixed model:

$$\mathbf{y}_i = \mathbf{X}_i\mathbf{b}_i + \mathbf{Z}_i\mathbf{u}_i + \mathbf{e}_i,$$

where there are  $p_i$  fixed effects associated with trait  $i$  so that  $\mathbf{X}_i$  is an  $n_i \times p_i$  matrix and  $\mathbf{b}_i$  is a  $p_i \times 1$  dimensional column vector.  $\mathbf{X}_i$  and  $\mathbf{Z}_i$  are incidence matrices for fixed effects and random effects for trait  $i$ , respectively.

The multiple trait model can be represented as follows:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & 0 \\ 0 & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_1 & 0 \\ 0 & \mathbf{Z}_2 \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}$$

Notice that not all animals necessarily have an observation for both traits. Some animals may be represented in  $\mathbf{y}_1$  but not in  $\mathbf{y}_2$ , or vice versa. All animals, however, are represented with a breeding value for each trait in the analysis, irrespective whether they had an observation for that trait. The vectors  $\mathbf{y}_1$  and  $\mathbf{y}_2$  (and  $\mathbf{e}_1$  and  $\mathbf{e}_2$ ) are therefore not necessarily of the same length, but  $\mathbf{u}_1$  and  $\mathbf{u}_2$  are always equally long (with the number of elements equal to the number of animals in the analysis).

To obtain the mixed model equations for estimating fixed effects  $\mathbf{b}$  and breeding values  $\mathbf{u}$ , we need to specify the covariance matrices  $\mathbf{R}$  and  $\mathbf{G}$  associated with the vector  $\mathbf{e} = (\mathbf{e}_1, \mathbf{e}_2)'$  of residual errors and the vector  $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2)'$  of random effects.

For the breeding values we can write

$$\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \text{ and } \text{var}(\mathbf{u}) = \mathbf{G} = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix}$$

If  $\sigma_{g_{ii}}^2$  is the genetic variance of trait  $i$ , and  $\sigma_{g_{ij}}$  is the genetic covariance between the two traits (within one animal), we can define a 2 by 2 genetic covariance matrix

$$\mathbf{G}_0 = \begin{pmatrix} \sigma_{g_{11}}^2 & \sigma_{g_{12}} \\ \sigma_{g_{21}} & \sigma_{g_{22}}^2 \end{pmatrix}$$

Each part of  $\mathbf{G}$  is obtained by multiplying the relationships matrix with either the variance of a trait (diagonal blocks  $\mathbf{g}_{i,A}$ ) or the covariance between the traits (off diagonal blocks  $\mathbf{g}_{ij,A}$ ) where  $\mathbf{g}_{ij}$  is an element of  $\mathbf{G}_0$ . The covariance between the breeding value of trait  $i$  on individual  $k$  and the breeding value of trait  $j$  in individual  $l$  is the additive genetic covariance between traits  $i$  and  $j$  multiplied by the additive genetic relationship between individuals  $k$  and  $l$ .

### Multiple Trait Mixed Model Equations

The mixed model equations for a multiple trait model can be written according to the general principle of setting up mixed model equations. However, they are extended for the  $\mathbf{G}$ - and the  $\mathbf{R}$ -matrices.

For the mixed model equations we will need the inverse of  $\mathbf{G}$ . In the multiple trait mixed model this becomes:

$$\mathbf{G}^{-1} = \begin{pmatrix} G^{11} & G^{12} \\ G^{21} & G^{22} \end{pmatrix} \text{ where } G^{ij} = \mathbf{g}^{ij}\mathbf{A}^{-1},$$

and where  $g^{ij}$  is (i,j)-element of the inverse of the 2 by 2 genetic covariances matrix  $G_0$  and  $A^{-1}$  is the inverse of the relationships matrix as it can be setup directly.

The residual covariance matrix  $R$  has the same form, but  $A$  is replaced by an identity matrix  $I$  assuming there are no correlations between the residuals of different animals. While residual deviations for a given trait measured on different individuals are often assumed to be uncorrelated, this is not necessarily the case for different traits measured on the same individual. The phenotypic correlation between traits is often the result of correlation between genetic as well as environmental effects. When all traits are measured on all individuals ( $n_1=n_2=n$ ), the covariance matrix between  $e_i$  and  $e_j$  can be written as  $\sigma(e_i, e_j)=r_{ij}I$ , where  $r_{ij} = \sigma_e(i,j)$  is the environmental covariance between traits  $i$  and  $j$  as expressed in the same individual. The resulting  $n.2 \times n.2$  variance-covariance matrix for the total error vector  $e = (e_1, e_2)'$  becomes:

$$\text{and the inverse is } R^{-1} = \begin{bmatrix} I r^{11} & I r^{12} \\ I r^{21} & I r^{22} \end{bmatrix}$$

where  $r^{ij}$  is i-j element of the inverse of the 2 by 2 environmental covariances matrix between the two traits:  $R_0$ .

$$R = \begin{bmatrix} \sigma(e_1, e_1) & \sigma(e_1, e_2) \\ \sigma(e_2, e_1) & \sigma(e_2, e_2) \end{bmatrix} = \begin{bmatrix} I r_{11} & I r_{12} \\ I r_{21} & I r_{22} \end{bmatrix}$$

The set of multiple trait mixed model equations are given in the next figure. It is not the idea to memorise these equations, but to give you a how single trait mixed model

$$\begin{bmatrix} X_1' r^{11} X_1 & X_1' r^{12} X_2 & X_1' r^{11} Z_1 & X_1' r^{12} Z_2 \\ X_2' r^{21} X_1 & X_2' r^{22} X_2 & X_2' r^{21} Z_1 & X_2' r^{22} Z_2 \\ Z_1' r^{11} X_1 & Z_1' r^{12} X_2 & Z_1' r^{11} Z_1 + g^{11} A^{-1} & Z_1' r^{12} Z_2 + g^{12} A^{-1} \\ Z_2' r^{21} X_1 & Z_2' r^{22} X_2 & Z_2' r^{21} Z_1 + g^{21} A^{-1} & Z_2' r^{22} Z_2 + g^{22} A^{-1} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} X_1'(r^{11} y_1 + r^{12} y_2) \\ X_2'(r^{21} y_1 + r^{22} y_2) \\ Z_1'(r^{11} y_1 + r^{12} y_2) \\ Z_2'(r^{21} y_1 + r^{22} y_2) \end{bmatrix}$$

equations are 'blown up' to multiple trait mixed model equations. This has a rather large impact on the number of equations that has to be solved. Roughly, computer time for solving multiple trait mixed models goes up quadratically with the number of traits!

If not all traits are recorded for all animals, the inverse of the residual covariance matrix  $R$  becomes a bit trickier. The reason is that animals with one record only do not have a residual covariance with another trait. The covariance matrix between the residuals of the different traits ( $\sigma(e_1, e_2)$ ) can no longer be written as a diagonal matrix (a multiple of  $I$ ). When some observations are missing, the matrix  $X_1'$  can not be directly multiplied with  $X_2$ , i.e. the number of columns ( $= n_1$ ) does not correspond with the number of rows ( $= n_2$ ). This can be solved by using  $X_1' r^{12} I_{12} X_2$  where  $I_{12}$  identifies when two observations are on the same individual (only in those case we have an environmental covariance). When both

traits are measured on all animals  $\mathbf{I}_{12} = \mathbf{I}$  and  $\mathbf{X}_1 \mathbf{r}^{12} \mathbf{I}_{12} \mathbf{X}_2$  reduces to  $\mathbf{X}_1' \mathbf{r}^{12} \mathbf{X}_2$ .

The rules for building up multiple trait mixed model equations are outlined hereafter, as a reference for the further interested reader.

**Rules for building mixed model equations:**

*(this section is only for reference)*

For small examples the mixed model equations can be build from the corresponding design matrices. For larger data sets, however, this becomes complicated. Rules have been developed to build the mixed model equation without explicitly setting up the design matrices. These rules for building the mixed model equations with a multiple trait model (per animal) are:

- 1) If both  $y_1$  and  $y_2$  are known for an animal;  
The values for  $r^{11}$ ,  $r^{12}$ ,  $r^{21}$ , and  $r^{22}$  are added to the particular sections for each trait in the fixed and random part of the coefficient matrix. For instance for 2 effects (herd and animal), we have to add these four numbers to each of  $\mathbf{X}'\mathbf{X}$ ,  $\mathbf{X}\mathbf{Z}$ ,  $\mathbf{Z}\mathbf{X}$  and  $\mathbf{Z}\mathbf{Z}$ . To the vectors with the totals (right hand sides) we add  $r^{11}y_1 + r^{12}y_2$  and  $r^{12}y_1 + r^{22}y_2$  to each trait partition of the two vectors ( $\mathbf{X}'\mathbf{y}$  and  $\mathbf{Z}'\mathbf{y}$ ).  
In a single trait model, we would have added only one figure to the 4 partial matrices for a trait. For the totals (right hand sights), we would add only  $y$  to the partial vectors for herd and animal.
- 2) When only one observation for one trait on the animal is available;  
The values  $(r_{11})^{-1}$  or  $(r_{22})^{-1}$  are added to each of the relevant partial matrices in the coefficient matrix, while  $y_1(r_{11})^{-1}$  or  $y_2(r_{22})^{-1}$  are added to the relevant parts of the right hand sides.
- 3) Independent of the pattern of traits measured, we add the relationships matrix multiplied by  $g^{ij}$  to the  $i$ - $j$  block of the random effects of the coefficient matrix.

Example of a Multiple Trait Model

Consider a situation where we have the following measurements on 6 unrelated and non-inbred individuals, performing in two different herds. Both traits on an animal are measured in the same herd.

Individual	Herd	Trait 1	Trait 2
1	1	160	-
2	1	180	320
3	1	210	330
4	2	190	-
5	2	228	360
6	2	210	350

The phenotypic standard deviations for weaning weight and yearling are 20 and 40 kg, respectively, the heritabilities are 0.42 and 0.39 and the genetic correlation is 0.769. The 2

$$G_0 = \begin{bmatrix} 169 & 250 \\ 250 & 625 \end{bmatrix} \text{ which corresponds with } G_0^{-1} = \begin{bmatrix} .0145 & -.0058 \\ -.0058 & .0039 \end{bmatrix}$$

by 2 matrix with additive genetic covariances (within an individual) are:

The within-individual environmental covariances are:

$$R_0 = \begin{bmatrix} 231 & 285 \\ 285 & 975 \end{bmatrix} \text{ which corresponds with } R_0^{-1} = \begin{bmatrix} .0068 & -.0020 \\ -.0020 & .0016 \end{bmatrix}$$

The design matrices for the first trait are straightforward.  $Z_1$  is an identity matrix since all animals have a record for the first trait. For the second trait, however, more attention is needed. The matrix  $Z_2$  has one column for each breeding value (i.e. 6 columns) and one row for each observation (i.e. 4 rows) which results in:

$$Z_2 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The right hand side (RHS) for the example are (transposed):

[2.05 2.38 0.271 0.272 0.69 0.58 0.77 0.82 0.83 0.73 0 0.16 0.11 0 0.13 0.15]

The first 4 elements are for the fixed effects (2 herds for 2 traits).

The values are all scaled by multiplying with residual (co)variances. For example: the

RHS-value for the second animal for the first trait (6<sup>th</sup> element) is obtained as:

$r^{11}y_{12} + r^{12}y_{22} = 0.0068*180 + (-0.002)*320 = 0.584$ , where  $y_{12}$  (=180) and  $y_{22}$  (=320) are the record for the first and second trait for the second animal. The first animal has only the first trait measured, and its RHS value (5<sup>th</sup> element) becomes

$(r_{11})^{-1} y_{11} = 0.0043*160 = 0.69$ . Notice that when an animal has only one trait recorded, we multiply it by  $(r_{11})^{-1}$  (the inverse of the 1-1 element of the residual covariance matrix) and not by  $r^{11}$  (the (1,1)-element of the inverse of the residual covariance matrix). Notice also that animals with no record for a given trait have a zero in the RHS.

The solutions for the fixed effects and the multiple trait BLUP EBV's are:

$$b_1 = [183 \quad 209] \quad b_2 = [309 \quad 342]$$

$$u_1 = \begin{bmatrix} -9.86 \\ -1.00 \\ 10.86 \\ -8.17 \\ 7.70 \\ 0.47 \end{bmatrix} \quad u_2 = \begin{bmatrix} -14.58 \\ 2.87 \\ 11.72 \\ -12.08 \\ 9.35 \\ 2.73 \end{bmatrix}$$

where  $b_1$  are the solutions for the herd effects for weaning weight, and  $u_1$  are EBV's for weaning weight, and  $b_2$  and  $u_2$  refer to yearling weight.

When performing two single trait evaluations for the two traits the following solutions were found for Single Trait BLUP:

$$\text{Single trait : } b_1 = [183 \quad 209] \quad b_2 = [325 \quad 355]$$

$$u_1 = \begin{bmatrix} -9.86 \\ -1.41 \\ 11.27 \\ -8.17 \\ 7.89 \\ 0.28 \end{bmatrix} \quad u_2 = \begin{bmatrix} 0 \\ -1.95 \\ 1.95 \\ 0 \\ 1.95 \\ -1.95 \end{bmatrix}$$

Notes to the solutions:

- 1) The average breeding value for both traits is equal to zero within herd. This is to be expected because animals are assumed to be unrelated. This illustrates that it is impossible to make a fair comparison of the average breeding value of animals in herds when there are no genetic ties (e.g. offspring from a common sire).
- 2) Animal 1 has no observation for trait 2. Consequently its breeding value is entirely based on the information from the correlated trait 1. Animal 1 has a value for trait 1 which lies below the herd average and as a result of the positive genetic correlation between traits its breeding value is also below average, i.e. negative. The same is true for animal 4 in herd 2.
- 3) The single trait breeding values (and fixed effect solutions) deviate from the multiple trait solutions. As to be expected the single trait breeding values of animal 1 and 6 for yearling weight are equal to zero. There is no information to estimate the breeding value and consequently the animals get the average breeding value. The difference in breeding value for yearling weight between animal 2 and 3 (and between 5 and 6) is larger in the multiple trait case. The reason is that the information from weaning weight (the correlated trait) gives additional evidence that these animals are different in breeding value.
- 4) The difference between the average herd effect for weaning weight and yearling weight is larger in the single trait analysis. In fact, this difference is overestimated, since it is biased by the fact that for yearling weight we only recorded the best animals (=selection). The multiple trait evaluation takes this into account. From using the information on the first trait, the model knows that only the better animals had a yearling weight measured.
- 5) In the multiple trait EBV's we see that the animals that were not culled have an average EBV's above zero. This makes sense, because from the information on trait 1 we know that these are actually the better animals. Single trait evaluation would not use information on weaning weight, and consider the yearlings that were weighted as average animals. This shows that multiple trait evaluation is able to correct for sequential selection.

### **Advantages of Multiple Trait BLUP evaluation**

In general, using the multiple trait model gives an increase in accuracy of estimated breeding values. Furthermore, in many cases it is the only way to correct for selection on correlated trait.

The importance of increase of accuracy by using extra information, i.e. the importance of using a multiple trait (MT) model, depends on several aspects:

- *the information available on each animal*

If few or no observations are available for a particular trait, using observations on another trait when both traits are genetically correlated can increase the accuracy.

- *parameter structure*

If genetic and environmental correlations are small, the multiple trait model has few advantages. Furthermore, in a situation with a high  $h^2$ , only a few observations are needed for an accurate estimate of the breeding value. In other words, information of other traits is less important in that case. Besides, the difference between  $r_g$  and  $r_e$  is important; the larger the difference, the larger is the contribution of a correlated trait to the reduction of the Prediction Error Variance. The contribution of correlated traits to the accuracy of estimating breeding values can be examined with the selection index method.

- *correctness of parameters;*

In multiple trait model we make use of estimated values of the genetic parameters (heritabilities, correlations). This variance-covariance (VCV) matrix has to be checked on incorrectness (or consistency).

Schaeffer (1984) discussed the effects of incorrect estimated parameters. He distinguished two kinds of mistakes. First, the VCV matrices may not be valid, i.e. within the parameter space. A valid VCV matrix, by definition, is a positive definite matrix. This can be checked by looking at all the eigenvalues of the matrix. Eigenvalues of covariance matrices all have to be positive, making the matrix "positive definite". The second and most common mistake, mentioned by Schaeffer, is that estimates used in the model, could be greatly different from the underlying true values. Assume that the true parameters give the maximum response of selection. The realised response then depends on the difference with the parameters used, namely  $(r_g - \hat{r}_g)$  and  $(r_e - \hat{r}_e)$ .

In this respect, it is good to realise that single trait models are MT models with the assumption that  $\hat{r}_e = \hat{r}_g = 0$ . Therefore, inaccurate correlations are often still closer to the true values than zero correlations!

- *Correction for selection*

The example illustrated selection on sequentially recorded traits leads to culling and missing records for traits that are recorded in a later stage. Multiple trait evaluation was able to avoid selection bias.

This reflects a more general rule, also applicable in single trait genetic evaluation, that to avoid selection bias, all information that was used to base selection decisions on, should be included in the analysis. This is not only the case with missing records in sequentially recorded traits. Assume the situation when two traits are recorded simultaneously, and all animals have records for all trait, but selection is only for one of the traits. Single trait evaluation of one trait only would lead to biased EBV's and generally to an underestimate of the genetic trend for the correlated trait (although this depends on the genetic and environmental correlation between the traits). Since selection is usually on an index (a linear combination on all traits), single trait evaluation leads to incorrect estimates of the genetic trend in most of the cases!

### Computational considerations

Computer requirements quickly increase with the application of multiple trait BLUP genetic evaluation procedures. Suppose we want to carry out a 5-trait BLUP analysis. The multiple trait mixed model equations require nearly 25 times more coefficient to be handled compared to single trait BLUP. Solving the mixed model equations when multiple traits are present can be greatly simplified by constructing a transformation for the traits being considered (this is called 'canonical transformation'). This transformation constructs a new set of uncorrelated variables, which can be analysed in independent single trait evaluations. Such a transformation is possible when all animals had observations for all traits. Recently, algorithms have been developed to handle transformations also for the case of missing observations on some traits. Multiple trait models can still be quite cumbersome if more random effects are included (e.g. maternal effects for some traits). However, The combination of more efficient computing algorithms with the rapid increase of computing power has lead to a situation that multiple trait BLUP is the method of choice for more and more genetic evaluation systems.

### Software

There are software packages available that can be used to implement multiple trait genetic evaluations. A commonly used package for breeding value estimation is PEST (Prediction and ESTimation) written by Groeneveld et al. (1994). A more versatile and increasingly used package is ASREML (Gilmour et al., 1996: This package is most suitable for estimation of genetic parameters in animal breeding data for a wide variety of models. There are also genetic evaluation services around that provide the whole package of delivering multiple trait EBV's.

An ASREML example:

```
bwt wwt~ Trait at(Trait,1).bt at(Trait,2).rt Trait.dage !r Trait.anim!f Trait.grp
1 2 1 #R struct: 1 site, dimension Ro, 1 G structure
0 #order R (?), ASREML figures out if put to zero
2 0 US 12 0 14 !GP # order Ro, 0, model, starting_values
Trait.anim 2 #G structure: model term, dimension
2 0 US 4.9 0 4.5 !GP #order Go, 0, model starting_values
anim
```

In the model statement, some effects are fitted for both traits: Trait.dage  
 Other effects are fitted for one trait only at(Trait,2).rt  
 !GP means that the matrix (R of G) has to be positive definite

A multi-trait model can also have more random effects, e.g. a maternal effect:

```
bwt wwt~ Trait at(Trait,1).bt at(Trait,2).rt Trait.dage !r Trait.anim Trait.dam !f
Trait.grp
1 2 1 #R struct: 1 site, dimension Ro, 1 G structure
0 0 ID #nrec (= outer dim. Of R), ASREML figures out if put to
zero
Trait 0 US 12 0 14 !GPUP # order Ro (equal to nr. of traits), 0, model, start_values
Trait.anim 2 #G structure: model term, dimension
4 0 US !GP #order Go, 0, model starting_values
4.9
2 9.5
0 0 4.5
0 0 2 4.2
```

18: Multiple Trait Models

anim 0 AINV

The  $G_0$  has now dimension 4. The definition of the  $G_0$  can be spelled out in some more detail:

```
4 0 US 4.9 2 9.5 0 0 4.5 0 0 2 4.2 !GPUPFFPFFUP
```

```
4      order of Go
0      always a zero here
US     unstructured Go
4.9    following is lower Go starting values
2 9.5
0 0 4.5
0 0 2 4.2
!GPUPFFPFFUP      indicating whether the components should be Positive,
                   Unstructured, or Fixed at the starting value
```

the same line could be replaced by:

```
4 0 US !+10 !GPUPFFPFFUP
4.9    following is lower Go starting values
2 9.5
0 0 4.5
0 0 2 4.2
```

!GPUPFFPFFUP could be replaced by !GP if we simply want Go to be positive definite