

# Pedigree options in ASReml

Arthur Gilmour<sup>a\*</sup> Greg Dutkowski<sup>b</sup>

<sup>a</sup>*Orange Agricultural Institute, Orange NSW 2800 Australia*

<sup>b</sup>*University of Tasmania, Hobart, Tasmania, Australia*

---

## Abstract

ASReml has several options when forming the pedigree. This paper discusses the new developments for Release 2. Briefly, these are constraints on the Genetic groups, changes to the way the Maternal Grandsire model is fitted and provision for a degree of selfing and inbreeding.

*Key words:* BLUP, linear mixed models, REML, variance components, genetic groups, average information, ASReml, Maternal Grandsire Model, Selfing

---

## 1 Introduction

An important development in the use of mixed models was the incorporation of the Numerator Relationship Matrix to account for genetic correlations among animals. The wonder of this approach is that the inverse relationship matrix is in general sparser than the relationship matrix itself and not difficult to form in a recursive manner when parents are listed before their offspring.

However, a few special situations arise in practise which we consider here. First, in some breeding situations, sires and sire groups are tracked but individual dam pedigrees are not tracked. This leads to the Maternal Grandsire model. ASReml Release 1 accomodated this model by simply inserting a dummy dam into the pedigree. We now form the inverse omitting this dummy dam.

In open pollinated forest trees, a proportion of the flowers are self pollinated. We derive the relationship matrix for a given proportion of selfing.

---

\* Corresponding author.

*Email address:* `arthur.gilmour@agric.nsw.gov.au` (Arthur Gilmour).

In other plant species, for example cereals, the plants are mainly self pollinated and lines are selfed for several generations to produce varieties for release. The pedigree relationship needs to take this into account.

ASReml now has 3  $A$ inverse routines accessed by the !METHOD qualifier. Results are presented for !METHOD 0 based on Meuwissen and Lou (1992) which should be slightly faster than the other methods for large pedigrees.

!METHOD 1 should give the same results and is used for testing and comparison. !METHOD 2 has not been adapted for Genetic groups, selfing or the MGS model. Both of these are variations on the method of Quaas (1976).

## 2 Standard relationship matrix

The standard procedures can be summarised as follows. Let  $\mathbf{A} = \{a_{ij}\}$  be the relationship matrix. Let  $\mathbf{a}_{i,-j}$  be the  $i$ th row of  $\mathbf{A}$  except for the  $j$ th element.

- (1) Assume the relationship matrix for the base animals is known for example, are unrelated and not inbred so that their relationship matrix is an Identity matrix.
- (2) The row of the relationship matrix for the progeny of two parents is generated as the average of the relationship matrix rows for the parents.

$$\mathbf{a}_{i,-i} = \frac{1}{2}(\mathbf{a}_{s,-i} + \mathbf{a}_{d,-i})$$

- (3) The diagonal element of this new animal is  $a_{i,i} = 1 + \frac{1}{2}a_{s,d} = 1 + f_i$  where  $f_i$  is the inbreeding coefficient.

Applying this rule for two fullsibs and a half-sib generates the following matrix.

$$\begin{bmatrix} \text{Sire} & 1. & 0. & 0. & 0.5 & 0.5 & 0.5 \\ \text{Dam1} & 0. & 1. & 0. & 0.5 & 0.5 & 0. \\ \text{Dam2} & 0. & 0. & 1. & 0. & 0. & 0.5 \\ \text{SibS1} & 0.5 & 0.5 & 0. & 1. & 0.5 & 0.25 \\ \text{SibS1} & 0.5 & 0.5 & 0. & 0.5 & 1.0 & 0.25 \\ \text{SibS2} & 0.5 & 0. & 0.5 & 0.25 & 0.25 & 1.0 \end{bmatrix}$$

Progressively inverting this matrix gives the sequence:

$$\begin{bmatrix} \text{Sire} & 1 & 0 & 0 \\ \text{Dam1} & 0 & 1 & 0 \\ \text{Dam2} & 0 & 0 & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} \text{Sire} & 1.5 & 0.5 & 0 & -1 \\ \text{Dam1} & 0.5 & 1.5 & 0 & -1 \\ \text{Dam2} & 0 & 0 & 1 & 0 \\ \text{SibS1} & -1 & -1 & 0 & 2 \end{bmatrix} \Rightarrow \begin{bmatrix} \text{Sire} & 2 & 1 & 0 & -1 & -1 \\ \text{Dam1} & 1 & 2 & 0 & -1 & -1 \\ \text{Dam2} & 0 & 0 & 1 & 0 & 0 \\ \text{SibS1} & -1 & -1 & 0 & 2 & 0 \\ \text{SibS1} & -1 & -1 & 0 & 0 & 2 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} \text{Sire} & 2.5 & 1 & 0.5 & -1 & -1 & -1 \\ \text{Dam1} & 1 & 2 & 0 & -1 & -1 & 0 \\ \text{Dam2} & 0.5 & 0 & 1.5 & 0 & 0 & -1 \\ \text{SibS1} & -1 & -1 & 0 & 2 & 0 & 0 \\ \text{SibS1} & -1 & -1 & 0 & 0 & 2 & 0 \\ \text{SibS2} & -1 & 0 & -1 & 0 & 0 & 2 \end{bmatrix}$$

The rule for progressively generating these inverses is derived as follows.

Assume we have  $\mathbf{A}_1$  and  $\mathbf{A}_1^{-1}$  for a set of animals including the parents of a new animal. Let  $\mathbf{p}$  be a vector which performs the operation of averaging the parental rows of  $\mathbf{A}_1$ . It will be all zero except that the positions corresponding to the parents will be 0.5.

Then  $\mathbf{A}_2 = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_1\mathbf{p} \\ \mathbf{p}'\mathbf{A}_1 & 1 + \frac{1}{2}a_{s,d} \end{pmatrix}$  is the expanded relationship matrix.

The standard expression for the inverse of a partitioned matrix is

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{C} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\mathbf{Q}\mathbf{B}'\mathbf{A}^{-1} & \mathbf{A}^{-1}\mathbf{B}\mathbf{Q} \\ \mathbf{Q}\mathbf{B}'\mathbf{A}^{-1} & \mathbf{Q} \end{pmatrix} \text{ where}$$

$\mathbf{Q} = (\mathbf{C} - \mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1}$ . In this case,  $\mathbf{B} = \mathbf{A}_1\mathbf{p}$  so that  $\mathbf{A}^{-1}\mathbf{B} = \mathbf{p}$ .

Thus  $\mathbf{A}_2^{-1} = \begin{pmatrix} \mathbf{A}_1 + q\mathbf{p}\mathbf{p}' & -q\mathbf{p} \\ -q\mathbf{p}' & q \end{pmatrix}$  where  $q = (1 + \frac{1}{2}a_{s,d} - \mathbf{p}'\mathbf{A}_1\mathbf{p})^{-1}$ .

Since  $\mathbf{p}$  has such a simple structure,  $q = 1/(1 - \frac{1}{4}(a_{s,s} + a_{d,d})) = 4/(2 - (f_s + f_d))$  which simply requires the inbreeding coefficients for each parent.

To test **ASReml** we obtain  $\mathbf{A}^{-1}$ , invert it, and check the result. We will use the following pedigree which includes inbreeding and selfing.

```
1 0 0
2 0 0
```

```

3 0 0
4 1 1
5 1 1
6 2 2
7 4 6
8 5 6
9 7 8
10 9 9

```

The following was produced with **ASReml** qualifiers **!METHOD 0 !GIV !DIAG**. First is the relationship matrix obtained by inverting the Inverse relationship matrix. Rounding errors occur because of the moderate precision of the Ainverse .giv file. Then is the Ainverse as returned in the .giv file and finally the inbreeding coefficients as returned in the ainverse.dia file.

```

1 1.0
2 0.0 1.0
3 0.0 0.0 1.0
4 1.0 0.0 0.0 1.500
5 1.0 0.0 0.0 1.000 1.500
6 0.0 1.0 0.0 0.000 0.000 1.500
7 0.5 0.5 0.0 0.750 0.500 0.750 1.0000
8 0.5 0.5 0.0 0.500 0.750 0.750 0.6250 1.0000
9 0.5 0.5 0.0 0.625 0.625 0.750 0.8125 0.8125 1.3125
10 0.5 0.5 0.0 0.625 0.625 0.750 0.8125 0.8125 1.3125 1.65625

```

```

1 5.0
2 .0 3.0
3 .0 .0 1.0
4 -2.0 .0 .0 3.0
5 -2.0 .0 .0 .0 3.0
6 .0 -2.0 .0 1.0 1.0 4.0
7 .0 .0 .0 -2.0 .0 -2.0 4.5
8 .0 .0 .0 .0 -2.0 -2.0 0.5 4.5
9 .0 .0 .0 .0 .0 .0 -1.0 -1.0 4.90909
10 .0 .0 .0 .0 .0 .0 .0 .0 -2.90909 2.90909

```

Identity	Inbreeding	DiagofAinverse
1	0.0000	5.0000
2	0.0000	3.0000
3	0.0000	1.0000
4	0.50000	3.0000
5	0.50000	3.0000
6	0.50000	4.0000
7	0.0000	4.5000
8	0.0000	4.5000
9	0.31250	4.9091

### 3 Maternal Grand sire Model

In this situation, no dams are identified, or at least the dams are all treated as base animals. Consider the relationship matrix for a Maternal Grand sire, Sire, Dam and Progeny given by

$$\begin{bmatrix} \text{Sire of Dam} & 1 + f_g & r & 0 & (1 + f_g)/2 & (1 + f_g + 2r)/4 \\ \text{Sire} & r & 1 + f_s & 0 & r/2 & (2 + 2f_s + r)/2 \\ \text{Dam of Dam} & 0 & 0 & 1 & 0.5 & 0.25 \\ \text{Dam} & (1 + f_g)/2 & r/2 & 0.5 & 1 & (1 + r/2)/2 \\ \text{Progeny} & \frac{(1+f_g+2r)}{4} & \frac{(2+2f_s+r)}{4} & \frac{1}{4} & \frac{(1+r/2)}{2} & 1 + \frac{r}{4} \end{bmatrix}$$

We see that the Progeny line is half the Sire line plus a quarter of the Maternal Grand sire line except that the diagonal element is 1 plus a quarter of the relationship between the sire and the maternal grandsire.

The algebra is as before except for a different definition of  $\mathbf{p}$  and  $q$ .

For the non-zero elements of  $\mathbf{p}_s$  being  $\frac{1}{2}$  and  $\frac{1}{4}$  for sire and maternal grandsire

respectively,  $\mathbf{A}_m = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_1 \mathbf{p}_s \\ \mathbf{p}'_s \mathbf{A}_1 & 1 + \frac{r}{4} \end{pmatrix}$  and  $\mathbf{A}_m^{-1} = \begin{pmatrix} \mathbf{A}_1^{-1} + \mathbf{p}_s q_s \mathbf{p}'_s & -\mathbf{p}_s q_s \\ -\mathbf{p}'_s q_s & q_s \end{pmatrix}$

where  $q_s = (1 + r/4 - \mathbf{p}'_s \mathbf{A}_s \mathbf{p}_s)^{-1} = (1 - (1 + f_g + 4(1 + f_s))/16)^{-1} = \frac{16}{11 - (f_g + 4f_s)}$ .

The following results are from ASReml using the same pedigree file as before but with qualifiers !METHOD 0 !MGS !GIV !DIAG where the !MGS qualifier causes the third field to be interpreted as the maternal grandsire.

```

1 1.0000
2 0.0000 1.0000
3 0.0000 0.0000 1.
4 0.7500 0.0000 0. 1.2500
5 0.7500 0.0000 0. 0.5625 1.2500
6 0.0000 0.7500 0. 0.0000 0.0000 1.2500
7 0.3750 0.1875 0. 0.6250 0.2812 0.3125 1.0000
8 0.3750 0.1875 0. 0.2813 0.6250 0.3125 0.2188 1.0000
9 0.2812 0.1406 0. 0.3828 0.2969 0.2344 0.5547 0.3594 1.0547
10 0.2109 0.1055 0. 0.2871 0.2227 0.1758 0.4160 0.2695 0.7910 1.2637
```

1	1	2.63636						
2	2	1.81818						
3	3	1.00000						
4	1	-1.09091	4	1.86480				
5	1	-1.09091	5	1.86480				
6	2	-1.09091	4	0.205128	5	0.205128	6	1.65967
7	4	-0.820513	6	-0.410256	7	2.00466		
8	5	-0.820513	6	-0.410256	7	0.181818	8	1.73193
9	7	-0.727273	8	-0.363636	9	2.29358		
10	9	-1.11872	10	1.49162				
Identity	Inbreeding	DiagofAinverse						
	1	0.0000		2.6364				
	2	0.0000		1.8182				
	3	0.0000		1.0000				
	4	0.25000		1.8648				
	5	0.25000		1.8648				
	6	0.25000		1.6597				
	7	0.0000		2.0047				
	8	0.0000		1.7319				
	9	0.54688E-01		2.2936				
	10	0.26367		1.4916				

#### 4 Genetic groups

The **ASReml** qualifier `!GROUPS g` indicates that the first  $g$  identifiers in the pedigree file relate to genetic groups rather than to individuals in the population. When genetic groups are present, the SIRE and DAM fields should both be zero for the group lines. All other lines must specify one of the genetic groups as SIRE or DAM if the actual parent is unknown.

In release 1.62, the option was added to apply 'sum to zero' constraints on group effects. If the pedigree file includes genetic groups without constraint, the 'constant term' will be included in the genetic group effects. The constant term is unlikely to be the population average and may generate singularities in under predictable locations because of the ordering of the 'sparse' equations. This may be undesirable.

'Sum to Zero' constraints are applied to sets of genetic group effects by following the set with a dummy genetic group (one with no animals in it). ASReml then modifies the A inverse to apply a constraint in place of this dummy genetic group. That is, when there are two or more groups containing animals followed by a group with no animals, the A inverse line for the empty group is replaced with a line containing 1's for the preceding set of groups, and zero otherwise. This Lagrangian constrains the genetic group effects to sum to zero.

The following results from ASReml !METHOD 0 !GRP 3 !GIV !DIAG. We use the same pedigree file as before but now the first three lines are genetic groups. Notice that there are no individuals assigned to the third group. When ASReml has formed the A-inverse, it notes that this 'group' is empty and inserts Lagrangian off diagonal elements for this equation.

For the purpose of checking the inverse, the A matrix was calculated after zeroing the intersection between groups and individuals.

```
%ex/ex11/amg00.txt
\input /data/ex/ex11/amg00.txt
 1 0.33333
 2-0.33333 0.33333
 3 0.33333 0.66667-0.66667
 4 0.00000 0.00000 0.00000 1.00
 5 0.00000 0.00000 0.00000 0.00 1.00
 6 0.00000 0.00000 0.00000 0.00 0.00 1.0
 7 0.00000 0.00000 0.00000 0.50 0.00 0.5 1.000
 8 0.00000 0.00000 0.00000 0.00 0.50 0.5 0.250 1.000
 9 0.00000 0.00000 0.00000 0.25 0.25 0.5 0.625 0.625 1.125
10 0.00000 0.00000 0.00000 0.25 0.25 0.5 0.625 0.625 1.125 1.5625

 1 1 2.00000
 2 2 1.00000
 3 1 1.00000 2 1.00000 3 0.00000
 4 1 -1.00000 4 1.50000
 5 1 -1.00000 5 1.50000
 6 2 -1.00000 4 0.50000 5 0.50000 6 2.00000
 7 4 -1.00000 6 -1.00000 7 2.50000
 8 5 -1.00000 6 -1.00000 7 0.50000 8 2.50000
 9 7 -1.00000 8 -1.00000 9 4.28571
10 9 -2.28571 10 2.28571
Identity Inbreeding DiagofAinverse
 1 -1.0000 2.0000
 2 -1.0000 1.0000
 3 -1.0000 0.0000
 4 0.0000 1.5000
 5 0.0000 1.5000
 6 0.0000 2.0000
 7 0.0000 2.5000
 8 0.0000 2.5000
 9 0.12500 4.2857
10 0.56250 2.2857
Wed 08/09/2004
```

## 5 Partial selfing

$$\begin{bmatrix} \text{Female Parent} & 1 + f_f & r & 1 + f_f & (1 + f_f + r)/2 \\ \text{Male Parent} & r & 1 + f_m & r & (1 + f_m + r)/2 \\ \text{Selfed} & 1 + f_f & r & 1 + \frac{1+f_f}{2} & (1 + f_f + r)/2 \\ \text{Crossed} & \frac{1+f_f+r}{2} & \frac{1+f_m+r}{2} & \frac{1+f_f+r}{2} & 1 + r/2 \end{bmatrix}$$

Let  $s$  be the proportion selfed. The contribution of the female parent is then  $(s + \frac{1}{2}(1 - s))\mathbf{a}_f = \frac{1+s}{2}\mathbf{a}_f$ . The contribution from the male parent is  $\frac{1-s}{2}\mathbf{a}_m$ . The diagonal element will be  $s(1 + \frac{1+f_f}{2}) + (1 - s)(1 + r/2) = 1 + \frac{s(1+f_f)+(1-s)r}{2}$ .

Thus, the non zero elements of  $\mathbf{p}_w$  are  $\left(\frac{1+s}{2} \frac{1-s}{2}\right)$  and

$$\begin{aligned} q_w^{-1} &= 1 + \frac{s(1+f_f)+(1-s)r}{2} - \left(\frac{1+s}{2} \frac{1-s}{2}\right) \begin{pmatrix} 1 + f_f & r \\ r & 1 + f_m \end{pmatrix} \begin{pmatrix} \frac{1+s}{2} \\ \frac{1-s}{2} \end{pmatrix} \\ &= 1 - \frac{1}{4}((1 + s^2)(1 + f_f) + 2sr(1 - s) + (1 - s)^2(1 + f_m)) \end{aligned}$$

Now this expression for  $q_w$  involves the covariance between the parents which previously cancelled out and which is not easily available from the existing algorithm. However, we expect that if the plant is outcrossed, the male parent is not known. We could therefore nominate a degree of relationship for seeds collected in the wild, or just assume the male parent is unrelated.

As a starting point, ASReml assumes that if the male parent is unknown, there is proportional selfing but if the male parent is identified, that a controlled mating has occurred. This complicates the algorithm because in the former case, the nonzero element of  $\mathbf{p}$  is  $\frac{1}{2}(1 + s)$  while in the other case, there are two values of  $\frac{1}{2}$

The following results from ASReml !METHOD 0 !SELF 0.3 !GIV !DIAG for the pedigree displayed at the bottom.

```

1 1.0000
2 0.0000 1.0000
3 0.0000 0.0000 1.
4 1.0000 0.0000 0. 1.5000
5 0.6500 0.0000 0. 0.9750 1.2250
6 0.0000 0.6500 0. 0.0000 0.0000 1.1500
7 0.5000 0.3250 0. 0.7500 0.4875 0.5750 1.0000
8 0.3250 0.3250 0. 0.4875 0.6125 0.5750 0.5312 1.0000
9 0.4125 0.3250 0. 0.6188 0.5500 0.5750 0.7656 0.7656 1.2656
10 0.2681 0.2112 0. 0.4022 0.3575 0.3737 0.4977 0.4977 0.8227 1.1898

1      1      3.00000
```



2	2	1.58076						
3	3	1.00000	1	-2.00000				
4	4	3.45533						
5	4	-1.09937	5	2.30672				
6	2	-0.893471	4	0.740741	5	0.615385	6	2.73070
7	4	-1.48148	6	-1.48148	7	3.46296		
8	5	-1.23077	6	-1.23077	7	0.500000	8	2.96154
9	7	-1.00000	8	-1.00000	9	2.64492		
10	9	-0.992189	10	1.52644				
Identity	Female	Male	Inbreeding	DiagofAinverse				
1	0	0	0.0000	3.0000				
2	0	0	0.0000	1.5808				
3	0	0	0.0000	1.0000				
4	1	1	0.50000	3.4553				
5	4	0	0.22500	2.3067				
6	2	0	0.15000	2.7307				
7	4	6	0.0000	3.4630				
8	5	6	0.0000	2.9615				
9	7	8	0.26562	2.6449				
10	9	0	0.18984	1.5264				

Fri 17/09/2004

## 6 Inbred lines

In this situation, after a cross is made, it is selfed a large number of times to produce an inbred line. The procedure is basically as before except that the inbreeding is always 1.

For the non-zero elements of  $\mathbf{p}_i$  being  $\frac{1}{2}$  and  $\frac{1}{2}$ ,  $\mathbf{A}_i = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_1\mathbf{p}_i \\ \mathbf{p}'_i\mathbf{A}_1 & 2 \end{pmatrix}$  and

$\mathbf{A}_i^{-1} = \begin{pmatrix} \mathbf{A}_1^{-1} + \mathbf{p}_i q_i \mathbf{p}'_i & -\mathbf{p}_i q_i \\ -\mathbf{p}'_i q_i & q_i \end{pmatrix}$  where  $q_i^{-1} = 2 - \mathbf{p}'_i \mathbf{A}_1 \mathbf{p}_i = 1 - \frac{1}{2} a_{mf}$ . This

last term comes about from  $2 - \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 2 & a_{mf} \\ a_{mf} & 2 \end{pmatrix} \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$

The special cases are that where no parents are known,  $\mathbf{p}_i = (0)$  so that  $q_i^{-1} = 2$  and if only one parent is known,  $\mathbf{p}_i$  contains a single  $\frac{1}{2}$  so that  $q_i^{-1} = 2 - \frac{1}{2} 2 \frac{1}{2} = 1.5$

The following results are from ASReml using the pedigree file displayed below with qualifiers !METHOD 0 !INBRED 1.0 !GIV !DIAG where the !INBRED

qualifier implies inbred lines. Note that 'selfing' is not permitted in the pedigree with inbred lines as it will generate a singularity.

```

1 2.0000
2 0.0000 2.0000
3 0.0000 0.0000 2.00000
4 1.0000 0.0000 1.0000 2.0000
5 0.0000 1.0000 1.0000 0.5000 2.0000
6 0.5000 0.5000 1.0000 1.2500 1.2500 2.0000
7 0.7500 0.2500 1.0000 1.6250 0.8750 1.6250 2.0000
8 1.3750 0.1250 0.5000 1.3125 0.4375 1.0625 1.3750 2.0000
9 1.0625 0.1875 0.7500 1.4687 0.6562 1.3437 1.6875 1.6875 2.0000
10 0.7812 0.3437 0.8750 1.3594 0.9531 1.6719 1.6562 1.3750 1.6719 2.0

```

```

1      1      1.15000
2      2      0.750000
3      1 0.250000      2 0.250000      3      1.00000
4      1 -0.500000      3 -0.500000      4      2.00000
5      2 -0.500000      3 -0.500000      4 0.333333      5      1.33333
6      4      0.00000      5 -0.666667      6      2.76190
7      1 0.400000      4 -1.333333      6 -1.333333      7      3.86667
8      1 -0.800000      7      0.00000      8      2.40000
9      6 0.761905      7 -1.60000      8 -1.60000      9      3.96190
10     6 -1.52381      9 -1.52381      0      3.04762

```

Identity	Female	Male	Inbreeding	DiagofAinverse
1	0	0	1.0000	1.1500
2	0	0	1.0000	0.75000
3	0	0	1.0000	1.0000
4	1	3	1.0000	2.0000
5	2	3	1.0000	1.3333
6	4	5	1.0000	2.7619
7	6	4	1.0000	3.8667
8	7	1	1.0000	2.4000
9	8	7	1.0000	3.9619
10	9	6	1.0000	3.0476

Thu 16/09/2004

For example, the coefficient  $3.0476 = 1/(1 - 43/64)$  where  $43/64 = \frac{1}{2}a_{96}$ . The calculation of this relationship required development of a recursive routine which works as follows. Define three vectors  $\mathbf{H}$  to hold the index of the higher numbered parent,  $\mathbf{L}$  to hold the index of the other parent, and  $\mathbf{c}$  to hold the contributions to relationship. Initialise  $a_{69} = 0$ ,  $k = 1$ ,  $H_1 = 9$ , and  $L_1 = 6$ .

While  $k > 0$ ,

if  $H_k$  equals  $L_k$ , add  $c_k$  to  $a$  and decrement  $k$ .

else identify sire and dam of  $H_k$  ( $s$  and  $d$ ) and replace  $(H_k, L_k, c_k)$  with  $(\max(s, L_k), \min(s, L_k, \frac{1}{2}c_k))$  and  $(H_{k+1}, L_{k+1}, c_{k+1})$  with  $(\max(d, L_k), \min(d, L_k, \frac{1}{2}c_k))$  (but omit operations if  $s$  and/or  $d$  are 0, adjusting  $k$  accordingly).

H,L,c	k=1	k=2	k=3	k=4	k=5	k=6	a
	9,6,1						
	8,6,1/2	7,6,1/2					
	8,6,1/2	6,4,1/4	6,6,1/4				
	8,6,1/2	5,4,1/8	4,4,1/8				1/4
	8,6,1/2	4,3,1/16	4,2,1/16				3/8
	8,6,1/2	4,3,1/16	3,2,1/32	2,1,1/32			3/8
	8,6,1/2	3,3,1/32	3,1,1/32				3/8
	7,6,1/4	6,1,1/4					13/32
	7,6,1/4	4,1,1/8	5,1,1/8				13/32
	7,6,1/4	4,1,1/8	2,1,1/16	3,1,1/16			13/32
	7,6,1/4	1,1,1/16	3,1,1/16				13/32
	6,6,1/8	6,4,1/8					15/32
	6,6,1/8	4,4,1/16	5,4,1/16				15/32
	6,6,1/8	4,4,1/16	4,2,1/32	4,3,1/32			15/32
	6,6,1/8	4,4,1/16	4,2,1/32	3,3,1/64	3,1,1/64		15/32
	6,6,1/8	4,4,1/16	1,2,1/64	3,2,1/64			31/64
	6,6,1/8						35/64
							43/64

While the ASReml syntax allows for an inbreeding coefficient less than one to be specified, the algebra has not been worked through for that case.

## 7 Mixed Model equations

If  $\mathbf{y}$  denotes the  $n \times 1$  vector of observations, the linear mixed model can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad (1)$$

where  $\boldsymbol{\tau}$  is the  $p \times 1$  vector of fixed effects,  $\mathbf{X}$  is an  $n \times p$  design matrix of full column rank which associates observations with the appropriate combination of fixed effects,  $\mathbf{u}$  is the  $q \times 1$  vector of random effects,  $\mathbf{Z}$  is the  $n \times q$  design matrix which associates observations with the appropriate combination of random effects, and  $\mathbf{e}$  is the  $n \times 1$  vector of residual errors.

The model (1) is called a linear mixed model or linear mixed effects model. It is assumed

$$\begin{bmatrix} \mathbf{u} \\ \mathbf{e} \end{bmatrix} \sim N \left( \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{G}(\boldsymbol{\gamma}) & \mathbf{0} \\ \mathbf{0} & \mathbf{R}(\boldsymbol{\phi}) \end{bmatrix} \right) \quad (2)$$

where the matrices  $\mathbf{G}$  and  $\mathbf{R}$  are functions of parameters  $\boldsymbol{\gamma}$  and  $\boldsymbol{\phi}$ , respectively.

Details of the AI algorithm for REML estimation can be found in several places including Gilmour et al. (1995). It revolves around the mixed model equations derived from the objective function

$$\log f_{\mathbf{Y}}(\mathbf{y} | \mathbf{u}; \boldsymbol{\tau}, \mathbf{R}) + \log f_{\mathbf{U}}(\mathbf{u}; \mathbf{G}) \quad .$$

which is the log-joint distribution of  $(\mathbf{Y}, \mathbf{u})$ . Differentiating with respect to  $\boldsymbol{\tau}$  and  $\mathbf{u}$  leads to the mixed model equations (Robinson, 1991) which are given by

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\tau}} \\ \tilde{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}. \quad (3)$$

These can be written as

$$\mathbf{C}\tilde{\boldsymbol{\beta}} = \mathbf{W}\mathbf{R}^{-1}\mathbf{y}$$

where  $\mathbf{C} = \mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \mathbf{G}^*$ ,  $\tilde{\boldsymbol{\beta}} = [\boldsymbol{\tau}' \ \mathbf{u}']'$  and  $\mathbf{G}^* = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} \end{bmatrix}$ .

The solution of (3) requires values for the variance parameters  $\boldsymbol{\gamma}$  and  $\boldsymbol{\phi}$ . In practice we replace  $\boldsymbol{\gamma}$  and  $\boldsymbol{\phi}$  by their REML estimates  $\hat{\boldsymbol{\gamma}}$  and  $\hat{\boldsymbol{\phi}}$ .

The AI algorithm uses the trace( $(\partial\mathbf{C})\mathbf{C}^{-1}$ ) when calculating the score and uses  $\mathbf{Y}'\mathbf{P}\mathbf{Y}$  as the the Average Information matrix where  $\mathbf{y}'\mathbf{P}\mathbf{y}$  is the usual residual sum of squares and  $\mathbf{Y}$  is a matrix of working variables, one for each variance parameter, given by either  $(\partial\mathbf{R})\mathbf{R}^{-1}\mathbf{e}$  for parameters of  $\mathbf{R}$  or  $\mathbf{Z}(\partial\mathbf{G})\mathbf{G}^{-1}\mathbf{u}$  for parameters of  $\mathbf{G}$ .

$\mathbf{Y}'\mathbf{P}\mathbf{Y}$  is formed by an absorption process (Gilmour et al 1995) which is quite efficient when  $\mathbf{C}$  is sparse provided a judicious ordering is used. The other part of the process requires calculation of trace( $(\partial\mathbf{C})\mathbf{C}^{-1}$ ). We only need the elements of  $\mathbf{C}^{-1}$  which correspond to non-zero elements in  $\mathbf{C}$  when calculating this trace since the derivative  $\partial\mathbf{C}$  is zero for all zero elements of  $\mathbf{C}$ . There is a therefore a huge advantage in forming as few extra elements as possible when

forming the required elements of  $\mathbf{C}^{-1}$ . This is controlled by the order in which the equations are processed.

The AI algorithm requires space for  $\mathbf{C}^{-1}$  and for working variables and their cross-products. ASReml uses the same memory to form the mixed model equations, absorb them and then hold the (partial) inverse.

For convenience, we have described the situation where both  $\mathbf{R}$  and  $\mathbf{G}$  are positive definite but we in fact also allow special cases where  $\mathbf{G}$  is singular or negative definite (for example Thompson et al 2003).

## 8 Discussion

## 9 Acknowledgements

## 10 References

Gilmour, A.R., Thompson, R. and Cullis, B.R. (1995). Average Information, an efficient algorithm for REML estimation in linear mixed models. *Biometrics* **51**, 1440-50.

Gilmour, A.R., Cullis, B.R., Gogel, B.J., Welham, S.J. and Thompson, R. (2002). ASREML User Guide. VSN International, Hemel Hempstead, United Kingdom. <http://www.asreml.co.uk>