Animal Model - Intraherd Evaluation

Abstract: The discussion starts out with three examples of relationship matrices and their corresponding inverses. This illustrates two points: 1) any complete relationship inverse can be formed from the inverse of the relationships among parents; and 2) the only non-zero elements in the relationship inverse are between parents, parents and progeny, and between mates. Three equivalent models are developed to arrive at breeding value predictions for all animals using all known sources of relative information. A genotypic model, gametic model and finally a combination of the two leads to the reduced animal model which affords the opportunity to solve a set of equations as large as the number of parents in the herd. The equations to be solved are more amenable to iterative solution procedures because most of the progeny information is added to the RHS of the equations instead of through the relationship inverse.

Outline
A. Relationship Matrix and Inverse
   1. Structure - non-zero elements of inverse
      a. parents
      b. parents and progeny
      c. mates
   2. Examples to illustrate formation of a complete relationship inverse from an inverse of relationships among the parents.
B. Three Equivalent Models
   1. Equivalent model theory
   2. Genotypic model
   3. Introduction to gametic model and development of equivalent models for records of a single animal
   4. Reduced animal model
      a. genotypic for parents
      b. gametic for non-parents
   5. Review of equivalent models with example to show equivalence
C. Reduced Animal Model
   1. Properties of solutions
   2. Back solutions for non-parents
      a. average breeding value for future sibs
      b. a prediction of the difference between future sibs and non-parents currently in the herd
   3. How relatives influence animal prediction
   4. Representation of phenotypic, genetic, and environmental trends
What is an animal model?

A statistical model which can be applied to data obtained from dairy cattle, or any other species for that matter, which describes biological processes and effects quantitatively.

Concept of a breeding value

You will need to understand the concept of a breeding value as an expression of relative genetic merit for an individual animal. The sire and dam each contribute ½ of their genes to a young animal. The breeding value of this young animal is the genetic effect of this sample of genes from the sire and dam expressed in the units of measure for the trait (i.e. kg of milk, kg of weight, etc.).

There are many animal models. The appropriate model is determined by the goals for economic improvement, species, trait, and the kind and amount of genetic variation. Some examples are:

1. First lactation milk yield, additive genetic effects.
2. All lactation records for milk yield, additive genetic effects, simple repeatability model.
3. Multiple traits, as in first lactation milk and fat yield.
4. Direct and maternal additive genetic effects, single trait, single records.

What must be known before conducting a genetic evaluation?

The management conditions which affect our records are unknown and these must be estimated so we can make adjustments for the conditions which change from herd to herd. We don't have perfect data to predict the breeding value of all animals. We use all the data available to find the best linear unbiased predictor of the breeding value for all animals. The following information must be known before conducting a genetic evaluation with an animal model.

1. Genetic theory and statistical methods.
   a. Methods enable us to follow what someone else has done.
   b. Theory enables us to manipulate mathematical expressions and chart a new direction.
2. Definition of a correct model.
   This will depend on the breed, traits, and management conditions under which records are made.
3. Complete identification of all ancestors in pedigree.
4. The kind and amount of genetic variation for each trait and breed.

**How do you conduct a genetic evaluation?**

1. Collect data and pedigree information.
2. Standardize records for age at calving, days in milk, and month of calving.
3. Conduct pilot studies.
   a. Define model appropriate for conditions affecting records.
   b. Demonstrate feasibility.
4. Develop computational procedures.
   c. Develop new software.
   d. Import technology and software from other countries.
   e. Modify computing strategies.
      1) Set up equations and solve iteratively.
      2) Adopt indirect procedures which iterate on data.

**Properties of solutions:** The animal model has become the foremost approach to genetic evaluation of animals in the world. The solutions have properties which are defensible by the scientific community (Kennedy et al. 1988) and breeders have found it provides them with a way to improve the genetic merit of their cattle. We don't know how the different management conditions from herd to herd affect each record, therefore these are estimated. The estimates are Best Linear Unbiased Estimates. The true breeding value of every animal is unknown, in fact we will never know the true breeding value. We can, however, obtain a prediction of the breeding value based on the information available. The predictor is the Best Linear Unbiased Predictor.

In practical terms, the animal model allows us to use all available information to estimate management conditions and predict the breeding value. Sources of information which can be included are:

1) The animals own record (i.e. milk and fat yield, weight at fixed age).
2) All progeny information on individual, sire, dam and other relatives.

In addition, breeding values are:

1) Adjusted for heritability and repeatability.
2) Directly comparable across herds and over time.
3) Less subject to change from first to later evaluations as more information on the individual animal becomes available.
4) Adjusted for genetic competition due to genetic merit of mates.
5) The best procedure for ranking animals.
A. Relationship Matrix and Inverse

All animal models use the additive genetic relationship matrix among animals (A) in one form or another. Originally defined by Wright (1922) as the coefficient of parentage among animals, elements of the relationship matrix are twice the probability of identical genes by descent occurring in two individuals. When multiplied by the additive genetic variance ($\sigma_a^2$), $A\sigma_a^2$ is the covariance among breeding values.

The inverse of A, called $A^{-1}$, has very special properties. Knowledge of these properties is helpful when defining alternative animal models and will be used later to explain the composition of an animal’s evaluation. We will begin by assuming we have ordered all animals with progeny in a manner such that they will precede those with no progeny. Second, we will take note of the fact that the structure of the relationship inverse is such that the only non-zero elements are between parents, parents and progeny, and mates. Therefore, $A^{-1}$ is a sparse matrix, although A may not be a sparse matrix.

A general representation of the relationship inverse is

$$A^{-1} = \begin{bmatrix}
A^{-1}_{sd} + .25M^\prime NM & -5M^\prime N \\
-5M^\prime N & N
\end{bmatrix}$$

where $A^{-1}_{sd}$ is the inverse of the relationship among parents,

$M$ is an incidence matrix relating parents to progeny, and

$N$ is a diagonal matrix which accounts for the number of known parents (e.g., values of $0 < N \leq 2$ are typically found) plus any inbreeding in the non-parent animals.

Later, these properties will be used to illustrate how solutions for a non-parent depend only on its own record, the solution(s) for fixed effects and the solution for its parents. For example, consider the sire, dam and offspring relationship

$$A = \begin{bmatrix}
1.0 & 0 & .5 \\
0 & 1.0 & .5 \\
.5 & .5 & 1.0
\end{bmatrix}$$

It has in inverse of

$$A^{-1} = \begin{bmatrix}
1.5 & 0.5 & -1.0 \\
0.5 & 1.5 & -1.0 \\
-1.0 & -1.0 & 2.0
\end{bmatrix}$$

which illustrates the generality of the formulas. Two more examples illustrate the non-zero elements in more complex pedigrees.
Example 2. Consider the case where some parents are unknown,

![Diagram](image)

The relationship matrix \((A)\) is

\[
\begin{bmatrix}
S_1 & S_2 & D_0 & D_1 & D_2 & C_1 & C_2 \\
S_1 & 1 & 0 & 0 & 0 & 0 & .5 & 0 \\
S_2 & 0 & 1 & 0 & 0 & 0 & 0 & .5 \\
D_0 & 0 & 0 & 1 & .5 & .5 & .25 & .25 \\
D_1 & 0 & 0 & .5 & 1 & .25 & .5 & .125 \\
D_2 & 0 & 0 & .5 & .25 & 1 & .125 & .5 \\
C_1 & .5 & 0 & .25 & .5 & .125 & 1 & .0625 \\
C_2 & 0 & .5 & .25 & .125 & .5 & .0625 & 1 \\
\end{bmatrix}
\]

And the inverse of the relationship matrix among parents \((A_p^{-1})\) is

\[
\begin{bmatrix}
S_1 & S_2 & D_0 & D_1 & D_2 \\
S_1 & 1 & 0 & 0 & 0 \\
S_2 & 0 & 1 & 0 & 0 \\
D_0 & 0 & 0 & 1.667 & -.667 & -.667 \\
D_1 & 0 & 0 & -.667 & 1.333 & 0 \\
D_2 & 0 & 0 & -.667 & 0 & 1.333 \\
\end{bmatrix}
\]

While the complete inverse \((A^{-1})\) is

\[
\begin{bmatrix}
S_1 & S_2 & D_0 & D_1 & D_2 & C_1 & C_2 \\
S_1 & 1.5 & 0 & 0 & .5 & 0 & -1 & 0 \\
S_2 & 0 & 1.5 & 0 & 0 & .5 & 0 & -1 \\
D_0 & 0 & 0 & 1.667 & -.667 & -.667 & 0 & 0 \\
D_1 & .5 & 0 & -.667 & 1.833 & 0 & -1 & 0 \\
D_2 & 0 & .5 & -.667 & 0 & 1.833 & 0 & -1 \\
C_1 & -1 & 0 & 0 & -1 & 0 & 2 & 0 \\
C_2 & 0 & -1 & 0 & 0 & -1 & 0 & 2 \\
\end{bmatrix}
\]
Example 3. A complete two generation pedigree.

The relationship matrix \((A)\) is

\[
\begin{bmatrix}
1 & 2 & 3 & 4 & 5 & 6 & 7 \\
1 & 1 & 0 & 0 & 0 & .5 & 0 & .25 \\
2 & 0 & 1 & 0 & 0 & .5 & 0 & .25 \\
3 & 0 & 0 & 1 & 0 & 0 & .5 & .25 \\
4 & 0 & 0 & 0 & 1 & 0 & .5 & .25 \\
5 & .5 & .5 & 0 & 0 & 1 & 0 & .5 \\
6 & 0 & 0 & .5 & .5 & 0 & 1 & .5 \\
7 & .25 & .25 & .25 & .25 & .5 & .5 & 1 \\
\end{bmatrix}
\]

And the complete inverse \((A^{-1})\) is

\[
\begin{bmatrix}
1 & 2 & 3 & 4 & 5 & 6 & 7 \\
1 & 1.5 & .5 & 0 & 0 & -1 & 0 & 0 \\
2 & .5 & 1.5 & 0 & 0 & -1 & 0 & 0 \\
3 & 0 & 0 & 1.5 & .5 & 0 & -1 & 0 \\
4 & 0 & 0 & .5 & 1.5 & 0 & -1 & 0 \\
5 & -1 & -1 & 0 & 0 & 2.5 & .5 & -1 \\
6 & 0 & 0 & -1 & -1 & .5 & 2.5 & -1 \\
7 & 0 & 0 & 0 & 0 & -1 & -1 & 2 \\
\end{bmatrix}
\]

Note rows 1 to 4 have non-zero elements corresponding to their mates and progeny rows 5 and 6 have non-zero elements corresponding to their parents (-1), their mate (.5) and progeny (-1), while row seven only has elements relating to parents.

Animal seven is the only non-parent. Therefore, the relationship matrix among parents \((A_p)\) is
And the inverse among parents \((A_p^{-1})\) is

\[
\begin{bmatrix}
1 & 2 & 3 & 4 & 5 & 6 \\
1 & 1.5 & .5 & 0 & 0 & -1 & 0 \\
2 & .5 & 1.5 & 0 & 0 & -1 & 0 \\
3 & 0 & 0 & 1.5 & .5 & 0 & -1 \\
4 & 0 & 0 & .5 & 1.5 & 0 & -1 \\
5 & -1 & -1 & 0 & 0 & 2 & 0 \\
6 & 0 & 0 & -1 & -1 & 0 & 2
\end{bmatrix}
\]

Because the diagonal elements for rows 1 and 2 are smaller in \(A_p^{-1}\) than in \(A^{-1}\) this shows how parents get credit in the full \(A^{-1}\) for every progeny they produce.

B. **Three Animal Models**

All animal models can logically be defined as being either a genotypic, gametic, or reduced animal model. Each model varies by the way random elements are defined. Therefore, equations used for estimating fixed and random effects are different. The primary purpose of having different definitions is to simplify computing. The models are equivalent, however, and this will be discussed later.

**Genotypic model.** This is the simplest form of the animal model. There is one equation for every level of every fixed effect plus an equation for every animal. Some animals may not have values in the vector of observations, but must be included because animals without records are part of the complete inverse of the relationship matrix.

Suppose we order \(y\) and \(u\) as

\[
y = \begin{bmatrix} y_p \\ y_n \end{bmatrix}, \quad u = \begin{bmatrix} u_p \\ u_n \end{bmatrix}
\]

then the genotypic animal model can be developed for all data available. A genotypic model describing these records would be

\[
y = \begin{bmatrix} y_p \\ y_n \end{bmatrix} \quad \begin{bmatrix} \text{parents} \\ \text{non-parents} \end{bmatrix} = u
\]
\[
\begin{bmatrix}
y_p \\
y_n
\end{bmatrix} = \begin{bmatrix} X_p \\ X_n \end{bmatrix} b + \begin{bmatrix} Z_p & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} u_p \\ u_n \end{bmatrix} + \begin{bmatrix} e_p \\ e_n \end{bmatrix}
\]

[1]

where \( \text{Var} \begin{bmatrix} u_p \\ u_n \end{bmatrix} = \begin{bmatrix} A_{pp} & A_{pn} \\ A_{np} & A_{nn} \end{bmatrix} \sigma^2 / \alpha, \text{Var} \begin{bmatrix} e_p \\ e_n \end{bmatrix} = I \sigma_e^2, \) and \( \alpha = \sigma_a^2 / \sigma_e^2 = (1-h^2)/h^2, \)

\( u \) represents breeding values of animals making records plus "base" animals necessary to facilitate computation of \( A^{-1}, \)
\( Z \) is a matrix indicating which parents in \( u \) have records but may have null rows for those parents without records.

The corresponding mixed model equations for this model are

\[
\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + \alpha A^{-1} \end{bmatrix} \begin{bmatrix} b \\ u \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}
\]

[2]

The advantage of this model over those previously presented (e.g., sire models) is the predictions will now incorporate the animals' own performance. Some problems with this model are the large order of the equations (equal to the number of fixed effects plus the number of animals). Thus the number of equations exceeds the number of records. Another problem is associated with the behavior of such equations when iterative solutions are necessary. Experience has shown them to be slow to converge to a final solution.

**Equivalent models for records of a single animal.** Two models are equivalent if they account for the same total variation in the observations. The theory of equivalent models can be used to write an equivalent model for records of a single animal. As a general rule in writing models, any random effects can be put into the error term, as long as they are not correlated with the random effects remaining in the model. The model developed here follows the guidelines intended by Henderson (1974) when he said,

"Try to write the model with as many as possible of the sets of variables mutually uncorrelated as well as uncorrelated with other sets of variables. Examples given previously were use of interaction in place of \( C^2, \) and writing real producing abilities as sums of uncorrelated sire and cow components rather than real producing abilities with non-zero correlations between half-sibs. Remember that a model which will 'generate' the actual variance-covariance matrix of records is equivalent to any other model that generates the same matrix. The choice among equivalent models is dictated by computational efficiency."

The genotypic model for a record on a single calf is

\( y^c = \mu^c + u^c + e^c. \)

To arrive at an equivalent model that does not involve \( u^c \) write the model as

\[
y^c = \mu^c + (\frac{1}{2}u^s + \frac{1}{2}u^d + \phi^c) + e^c
= \mu^c + \frac{1}{2}u^s + \frac{1}{2}u^d + (\phi^c + e^c)
\]
where $\phi^c$ is the Mendelian sample effect which account for the genetic difference between full sibs. As written, this is a gametic model where the breeding value of the individual animals has been written in terms of one-half the breeding value of each of the parents and a Mendelian sampling effect. The two models for records on a single animal are equivalent. This is so by the rule in writing models which allows one to write the model in terms of uncorrelated random variables.

The next question is whether they are equivalent for records on different animals. Suppose we have records of both a dam and her calf. The gametic model for these two records is

$$
\begin{align*}
y^d &= \mu^d + \frac{1}{2}u^mgs + \frac{1}{2}u^mgd + (\phi^d + e^d) \\
y^c &= \mu^c + \frac{1}{2}u^s + \frac{1}{2}u^d + (\phi^c + e^c)
\end{align*}
$$

In this case the error of the dam's record ($\phi^d + e^d$) is correlated with $\frac{1}{2}u^d$ of the calf's record. That is, when both parents and non-parents have records a rule for writing models is violated. Recall the rule which states "any random effects can be put into the error term, as long as they are not correlated to the random effects remaining in the model". From this one concludes the gametic model may only be used for non-parent records when simultaneous solutions for both parents and non-parents are desires. While only non-parents may be expressed with the gametic model, any record may be expressed with the genotypic model without difficulty. Thus we now use a combination of the two in the model for all records. That is, parents are expressed in terms of the genotypic model and nonparents in terms of the gametic model. The model is

$$
\begin{bmatrix}
y_p \\
y_n
\end{bmatrix} =
\begin{bmatrix}
X_p & Z_p \\
X_n & I
\end{bmatrix}
\begin{bmatrix}
b \\
0
\end{bmatrix} +
\begin{bmatrix}
\phi_n \\
e_p
\end{bmatrix}
$$

where $P_n$ is a matrix of zeroes and ones relating non-parents to parents. Rows correspond to non-parent and columns to parents. The breeding value of parents are $u_p$ and $\phi_n$ is the prediction of the Mendelian sampling effect of the non-parents. The predicted breeding value of non-parents is formed from $u_n = \frac{1}{2}P_n u_p + \phi_n$. To complete the specification of the model the variances and covariances are

$$
\text{Var} \begin{bmatrix} u_p \\ \phi_n \end{bmatrix} = \begin{bmatrix} A_{pp} & 0 \\ 0 & D \end{bmatrix} \sigma^2 / \alpha, \quad \text{Var} [\epsilon] = I \sigma^2 \text{ and } \alpha = \sigma_c^2 / \sigma_a^2.
$$

D = 1.0, 0.75, or 0.5 for no parents in $u_p$, one, or both, respectively.

The mixed model equations for this model are

$$
\begin{bmatrix}
X_p'X_p + X_n'X_n \\
Z_p'Z_p + \frac{1}{2}X_p'P_n \\
X_n
\end{bmatrix} + \frac{1}{2}P_n'P_n + \alpha A^{-1} \begin{bmatrix}
X_p'Y_p + X_n'Y_n \\
Z_p'Y_p + \frac{1}{2}P_n'Y_n \\
Y_n
\end{bmatrix} = \begin{bmatrix}
X_p'Y_p + X_n'Y_n \\
Z_p'Y_p + \frac{1}{2}P_n'Y_n \\
Y_n
\end{bmatrix}
$$
What is Mendelian sampling?

Mendelian sampling is the effect due to random segregation and recombination of genes from the sire and dam.

“Mendelian sampling is the random sampling of parental genes caused by segregation and independent assortment of genes during germ cell formation and by random selection of gametes in the formation of the embryo. Bourdon (1997)”

Randomness of inheritance is critically important from an evolutionary standpoint and is also vitally important to the success of artificial selection. Nevertheless it creates a problem for animal breeders – it reduces our ability to control the outcomes of matings.

What about linkage? How does linkage affect Mendelian sampling?

The breeding values of either the sire or dam estimates the effect of linked genes. Linkage determines which genes are transmitted from parent to offspring.

Reduced animal model

Primary interest is in prediction of \( u_p \) and only some \( u_n \) are of interest, equivalent model theory is us used further to arrive at a model which involved only the fixed effects and the parents. Recall that any random effects can be put into the error term, as long as they are not correlated to the random effects remaining in the model. This was accomplished by expressing the parents with the genotypic model and non-parents with the gametic model. Since we now have two sets of uncorrelated random variables in the model we can absorb one set into the remaining equations to arrive at a reduced model and the remaining solutions will be exactly equivalent to those obtained from solving the full set of equations. The difference between models is in the composition of the error term.

The model is now written as

\[
\begin{bmatrix}
    y_p \\
    y_n
\end{bmatrix} =
\begin{bmatrix}
    X_p \\
    X_n
\end{bmatrix} b +
\begin{bmatrix}
    Z_p \\
    \frac{1}{2} P_n
\end{bmatrix} u_p +
\begin{bmatrix}
    e_p \\
    \phi_n + e_n
\end{bmatrix}
\]

where \( \text{Var}(u_p) = A_{pp} \sigma_a^2 / \alpha, \) and \( \alpha = \sigma_e^2 / \sigma_n^2. \)

\[
\text{Var}
\begin{bmatrix}
    e_p \\
    \phi_n + e_n
\end{bmatrix} =
\begin{bmatrix}
    I \\
    0
\end{bmatrix} \alpha^{-1} D_n + I
\begin{bmatrix}
    \sigma_e^2 \\
    0
\end{bmatrix} =
\begin{bmatrix}
    R_p \\
    0
\end{bmatrix}
\]

The mixed model equations for the reduced animal model are now

\[
\begin{bmatrix}
    X_p'X_p + X_n' R_n^{-1} X_n \\
    X_p'Z_p + \frac{1}{2} X_n' R_n^{-1} P_n \\
    Z_p'Z_p + \frac{1}{4} P_n' P_n^{-1} P_n + \alpha A_{pp}^{-1}
\end{bmatrix}
\begin{bmatrix}
    b \\
    \hat{u}_p
\end{bmatrix} =
\begin{bmatrix}
    X_p'Y_p + X_n' R_n^{-1} Y_n \\
    Z_p'Y_p + \frac{1}{2} P_n' R_n^{-1} Y_n
\end{bmatrix}
\]
The advantages are:

1. Order is reduced by the number of non-parents.
2. The equations are more suitable for iteration since the progeny information (most) has a direct effect on the right-hand-sides rather than through $A^{-1}$.

Further, one can make the following generalizations:

1. $R^{-1}$ is diagonal which implies that we can still absorb fixed effects.
2. If there are no records on the parents, the model reduces to the sire and dam model.
3. If there are no records on the parents and dams are related only through sires, the model reduces to the sire and maternal grandsire model.
4. If no parents have records and dams are unrelated (only single progeny), the model reduces to the sire model.

**Back solution for nonparents**

The reduced animal model uses all data on all animals to estimate effects and breeding values. Realized values of predicted breeding values for non-parents are obtained by adjusting the record for the non-parent by using estimates of the fixed effects in the model, estimating the Mendelian sampling effect, and then calculating the breeding value from one-half the sum of the breeding value of the two parents plus the estimate of the Mendelian sampling effect.

Recall the model for a single record is

$$y_{in} = X_{in}b + \frac{1}{2}u_s + \frac{1}{2}u_d + \phi + e$$

where

$y_{in}$ is the record on the $i^{th}$ nonparent,
$X_{in}$ is the incidence matrix for the $i^{th}$ animal with respect to $\beta$,
$u_s$ and $u_d$ are the breeding values of the two parents,
$\phi$ is the Mendelian sampling effect, and
$e$ is random error.

The predicted breeding value for a young animal (any non-parent) is

$$\hat{u}_n = \frac{1}{2} \hat{u}_s + \frac{1}{2} \hat{u}_d + \phi.$$ 

Computation of $u_{in}$, Best Linear Unbiased Predictor of an individual not in the solution for $u_p$ but providing data in $y$, is simple. First,

$$\hat{\varepsilon}_{in} = \hat{\phi}_n + \hat{e}_{in} = y_{in} - X_{in}b - \frac{1}{2} (\hat{u}_s + \hat{u}_d ).$$ \[7\]

then

$$\hat{u}_{in} = \frac{1}{2} (\hat{u}_s + \hat{u}_d ) + k_i \hat{\varepsilon}_{in} .$$ \[8\]

where $k_i = D_n \sigma_a^2 / (D_n \sigma_a^2 + \sigma_e^2) = D_n h^2 / [D_n h^2 + (1-h^2)]$.

$D_n$ needs clarification. It is the fraction of the additive genetic variance in the error associated with a record. Further, the value depends on the presence or absence of parental
information and the amount of this information that is included in the evaluation.

\[
D = \begin{cases} 
0.50 & \text{for progeny with both parents in } u_p, \\
0.75 & \text{for progeny with one parent in } u_p, \text{ and} \\
1.00 & \text{for progeny with no parents in } u_p.
\end{cases}
\]

Values of \(k_i\) for two levels of heritability are illustrated in Table 1.

<table>
<thead>
<tr>
<th>(D)</th>
<th>(h^2 = 0.50)</th>
<th>(h^2 = 0.30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.50</td>
<td>0.3333</td>
<td>0.1765</td>
</tr>
<tr>
<td>0.75</td>
<td>0.4286</td>
<td>0.2432</td>
</tr>
<tr>
<td>1.00</td>
<td>0.5000</td>
<td>0.3000</td>
</tr>
</tbody>
</table>

**Example**

Application and concepts employed in the reduced animal model (RAM) are illustrated using the data in Table 2. The data are typical weaning weights over three contemporary groups. The first five animals are parents of the animals in subsequent contemporary groups.

<table>
<thead>
<tr>
<th>Animal</th>
<th>Sire</th>
<th>Dam</th>
<th>Year</th>
<th>Season</th>
<th>Weaning Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>79</td>
<td>-</td>
<td>523</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
<td>79</td>
<td>-</td>
<td>554</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>79</td>
<td>-</td>
<td>551</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>-</td>
<td>79</td>
<td>-</td>
<td>566</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>-</td>
<td>79</td>
<td>-</td>
<td>537</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>3</td>
<td>80</td>
<td>-</td>
<td>464</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>4</td>
<td>80</td>
<td>-</td>
<td>494</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>3</td>
<td>81</td>
<td>-</td>
<td>476</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>5</td>
<td>81</td>
<td>-</td>
<td>476</td>
</tr>
</tbody>
</table>

The data are representative of a typical performance test for beef cattle. There is a simple structure to the relationship matrix (Table 3) among the nine animals in the example. This was done on purpose to make it easier to see the contribution of an animal’s information to the RAM equations.

Note that the inverse of relationships among the parents is an identity matrix. Therefore, the inverse of the relationships among the parents that is used in the RAM model is also an identity matrix. This will not necessarily be true for more complex data structures. This segment of the inverse is no longer an identity matrix in the complete inverse of the relationships as given in Table 4. In the complete inverse of relationships, the parents get credit for progeny; there are additional elements between mates; and finally the only remaining elements are between parents and progeny.

Three sets of solutions are provided to illustrate the impact of including various sources of information. The first ignores the relationships among animals and the second set of solutions from
a genotypic model or a full animal model. The latter are included to establish the equivalence of solutions with those from the reduced animal model.

Table 3. Relationship matrix.

<table>
<thead>
<tr>
<th>Animal</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>6</td>
<td>0.5</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.25</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
<td>0.25</td>
<td>0.25</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.25</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4. Inverse of relationship matrix.

<table>
<thead>
<tr>
<th>Animal</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>0.5</td>
<td>-1</td>
<td>0</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0.5</td>
<td>0.5</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
<td>1.5</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>6</td>
<td>-1</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

Solutions for the fixed effects are of interest, since they are used to calculate environmental trends and form the predicted breeding values for nonparent animals in the reduced animal model. These solutions are given in Table 5. Note the change in solutions as more additive genetic information is included in the model.

Table 5. Contemporary group solutions.

<table>
<thead>
<tr>
<th>Contemporary</th>
<th>Without Relationship</th>
<th>With Relationship</th>
<th>RAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>79</td>
<td>546.2</td>
<td>546.200</td>
<td>546.200</td>
</tr>
<tr>
<td>80</td>
<td>479.0</td>
<td>478.063</td>
<td>478.063</td>
</tr>
<tr>
<td>81</td>
<td>476.0</td>
<td>478.813</td>
<td>478.813</td>
</tr>
</tbody>
</table>

The three contemporary groups are different, a difference caused by two things. The first concerns differences in weather, feed, etc., or strictly environmental differences. The second differences are in the breeding value of animals in the three contemporary groups. The solutions for the contemporary group means are 546.2, 478.063, and 478.813. These solutions differences
represent the environmental differences between contemporary groups because genetic differences among animals are reflected in the breeding value solution.

Breeding value solutions for the animals in the example are given in Table 6. There is no difference between breeding value solutions for animals 8 and 9 when relationships are ignored. Solutions for other animals change as well when a full animal model is used. This reflects changes due to the adjustment for mates and other relative information. Solutions for parents from the RAM are identical to those from the full animal model. This is as it should be.

Table 6. Breeding value solutions with and without using relationships.

<table>
<thead>
<tr>
<th>Animal</th>
<th>without relationships</th>
<th>with relationships</th>
<th>RAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-11.6</td>
<td>-12.1</td>
<td>-12.1</td>
</tr>
<tr>
<td>2</td>
<td>3.9</td>
<td>4.4</td>
<td>4.4</td>
</tr>
<tr>
<td>3</td>
<td>2.4</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>4</td>
<td>9.9</td>
<td>11.2</td>
<td>11.2</td>
</tr>
<tr>
<td>5</td>
<td>-4.6</td>
<td>-3.7</td>
<td>-3.7</td>
</tr>
<tr>
<td>6</td>
<td>-7.5</td>
<td>-8.6</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>7.5</td>
<td>10.5</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>0.0</td>
<td>0.6</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>0.0</td>
<td>-6.2</td>
<td>-</td>
</tr>
</tbody>
</table>

In many cases it is not necessary to have solutions for all non-parents. Therefore a back solution procedure is used to predict breeding values for the non-parents of interest. The Mendelian Sampling and error part of each record is in \( \varepsilon_{in} \). This term is regressed to find the Mendelian Sampling part and this is added to half the sum of the predicted breeding value of the two parents. These computations are given in Table 7.

Table 7. Computation of non-parent breeding values.

<table>
<thead>
<tr>
<th>Animal</th>
<th>( y_{in} )</th>
<th>( \hat{X}_{in} \hat{b} )</th>
<th>( \frac{1}{2}(\hat{\mu}<em>{s} + \hat{\mu}</em>{a}) )</th>
<th>( \hat{\varepsilon}_{in} )</th>
<th>( \hat{\kappa}<em>{e</em>{in}} )</th>
<th>( \hat{\mu}_{in} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>464</td>
<td>478.06</td>
<td>-5.95</td>
<td>-8.1</td>
<td>-2.7</td>
<td>-8.6</td>
</tr>
<tr>
<td>7</td>
<td>494</td>
<td>478.06</td>
<td>7.80</td>
<td>8.1</td>
<td>2.7</td>
<td>10.5</td>
</tr>
<tr>
<td>8</td>
<td>476</td>
<td>478.81</td>
<td>2.30</td>
<td>-5.1</td>
<td>-1.7</td>
<td>0.6</td>
</tr>
<tr>
<td>9</td>
<td>476</td>
<td>478.81</td>
<td>-7.90</td>
<td>5.1</td>
<td>1.7</td>
<td>-6.2</td>
</tr>
</tbody>
</table>

Composition of an animal’s evaluation

The example is completed by looking at the composition of an animal’s evaluation. To do this we will use the full animal model. In the case where all animals have a single record, \( Z \) is the identity matrix and the mixed model equations are:

\[
\begin{bmatrix}
X'X & X'
\end{bmatrix}
\begin{bmatrix}
\hat{b}
\end{bmatrix} =
\begin{bmatrix}
X'y
\end{bmatrix}
\]

\[
\begin{bmatrix}
X'X & X'
\end{bmatrix}
\begin{bmatrix}
\hat{b}
\end{bmatrix} =
\begin{bmatrix}
X'y
\end{bmatrix}
\]
The equation for the \(i^{th}\) animal is:

\[
\left(1 + a_{ii}^\alpha\right)\hat{u}_i + \sum_{j \neq i} a_{ij}^\alpha \hat{u}_j = y - X\hat{b}
\]

where \(a_{ij}^\alpha\) represents the \(ij^{th}\) element from \(A^{-1}\). Thus, the animal’s evaluation is:

\[
\hat{u}_i = \left[1 / \left(1 + a_{ii}^\alpha\right)\right] (y - X\hat{b}) + \left[1 / \left(1 + a_{ii}^\alpha\right)\right] \sum_{j \neq i} a_{ij}^\alpha \hat{u}_j
\]

That is, the animal’s evaluation is the regressed deviation of the animal’s record \((y - X\hat{b})\) minus the regressed weighted function of all relatives’ evaluation. The nature of \(A^{-1}\) allows us to further define the latter component. In \(A^{-1}\) (Table 4) there are nonzero elements other than the diagonals only between parents, offspring and mates. Thus, the evaluation of animal \(i\) contains information on its sire and dam (if known) and progeny (if available), including adjustments for the merit of mates. Values for each of the terms is given in Table 8.

Table 8. Composition of animal’s evaluations.

<table>
<thead>
<tr>
<th>Animal</th>
<th>Weaning</th>
<th>Contemporary</th>
<th><a href="y-X%5Chat%7Bb%7D">1(1+a(^{\alpha}))</a></th>
<th>(-\sum_{j \neq i} a_{ij}^\alpha \hat{u}_j)</th>
<th>Breeding Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>523</td>
<td>546.200</td>
<td>-7.7</td>
<td>-0.5, -4.9</td>
<td>-12.1</td>
</tr>
<tr>
<td>2</td>
<td>554</td>
<td>546.200</td>
<td>2.6</td>
<td>-1.9, 3.7</td>
<td>4.4</td>
</tr>
<tr>
<td>3</td>
<td>551</td>
<td>546.200</td>
<td>1.6</td>
<td>1.3, -2.7</td>
<td>0.2</td>
</tr>
<tr>
<td>4</td>
<td>566</td>
<td>546.200</td>
<td>7.8</td>
<td>-0.8, 4.2</td>
<td>11.2</td>
</tr>
<tr>
<td>5</td>
<td>537</td>
<td>546.200</td>
<td>-3.7</td>
<td>2.4, -2.4</td>
<td>-3.7</td>
</tr>
<tr>
<td>6</td>
<td>464</td>
<td>478.063</td>
<td>-4.6</td>
<td>-4.0</td>
<td>-8.6</td>
</tr>
<tr>
<td>7</td>
<td>494</td>
<td>478.063</td>
<td>5.3</td>
<td>5.2</td>
<td>10.5</td>
</tr>
<tr>
<td>8</td>
<td>476</td>
<td>478.813</td>
<td>-0.9</td>
<td>1.5</td>
<td>0.6</td>
</tr>
<tr>
<td>9</td>
<td>476</td>
<td>478.813</td>
<td>-0.9</td>
<td>-5.3</td>
<td>-6.2</td>
</tr>
</tbody>
</table>

FIGURE 1. Pedigree and phenotypic records of animals from Kennedy and Sorensen (1988) for illustration of an animal model; records are in parentheses beside animal number.

Properties of solutions. (This section was completely adapted from Kennedy et al. (1988).

A common feature of all animal models is that they use the additive genetic relationship matrix \((A)\) either directly or indirectly. Diagonal elements of \(A\) equal \(1+F_i\), where \(F_i\) is the coefficient of inbreeding for animal \(i\); off diagonal elements equal the numerator of Wright’s (1922) coefficient of relationship between animals \(i\) and \(j\). When multiplied by the additive genetic variance \((\sigma_a^2)\), \(A\sigma_a^2\) describes the variance-covariance structure among additive genetic
(breeding) values of animals (a). Also, with most animal models, average breeding value, $E(a) = 0$.

In animal populations, however, genetic mean and variance seldom are constant but change as a consequence of finite population size or selection (Falconer, 1981). In closed populations of finite size, inbreeding accumulates each generation (as does genetic drift) and affects both genetic means and variances. Genetic means and variances also are changed by selection. Genetic variances change because of changes in gene frequency and establishment of covariances between genotypic values of different loci (linkage disequilibrium) (Bulmer 1971).

Kennedy et al. (1988) have shown that animal models by use of $A$ or $A^{-1}$ can accommodate some of the changes in genetic mean and variance. The breeding value of animal i ($a_i$) can be expressed as:

$$a_i = .5a_s + .5a_d + w_i \quad [9]$$

where $a_s$ and $a_d$ are breeding values of the sire and dam of the animal, .5 is the regression of $a_i$ on $a_s$ or $a_d$, and $w_i$ represents the Mendelian sampling contribution to $a_i$. The Mendelian sampling terms have an expected value of 0 and variance of $0.5(1 - F)\sigma_a^2$, where $F$ is the average inbreeding coefficient of the animal's parents. Selection affects the distribution of $a_s$ and $a_d$ but not $w_i$, which is normal and independent of parental breeding values if the number of loci is large enough, strictly infinite (Bulmer 1971, 1980). If a parent of an animal is not known or not included in $A$, then half its breeding value is included in $w_i$, and the distribution of $w_i$ reflects Mendelian sampling terms plus the contribution of the unknown parent. If the unknown parents are unselected, then the expected value of $w_i$ still is 0. However, variance of $w_i$ is $\sigma_a^2$ if both parents are not known or $(.75-.25F_p)\sigma_a^2$ if only one parent is not known, assuming the unknown parents are not inbred.

The model for $[a]$ can be extended back one generation such that $a_i$ can be written as a function of grandparental breeding values and additional Mendelian sampling terms:

$$a_i = .25a_{pgs} + .25a_{pgd} + .25a_{mgs} + .25a_{mgd} + .5w_s + .5w_d + w_i$$

where $p_{gs}$, $p_{gd}$, $m_{gs}$, and $m_{gd}$ identify paternal grandsire, paternal granddam, maternal grandsire, and maternal granddam, and $s$ and $d$ identify sire and dam. Again, Mendelian sampling terms are unaffected by selection in the previous generation and are independent of ancestral breeding values.

The vector of breeding values ($a$), can be expressed as the product of a matrix of appropriate regression coefficients ($T$) and a vector of ancestral breeding values and Mendelian sampling terms from subsequent generations ($w$):

$$a = Tw$$

Accordingly,

$$\text{Var}(a) = A\sigma_a^2 = T[\text{Var}(w)]T'\sigma_a^2 = TW'T\sigma_a^2$$
where $W$ is a diagonal matrix with diagonal elements \( .5(1-F)\sigma^2_a, (.75-.25F_p)\sigma^2_a, \) or \( \sigma^2_a \), depending if both, one, or no parents are identified.

Decomposition of $A$ into $TWT'$ is illustrated by the example pedigree of eight animals in Figure 1 from Kennedy and Sorensen (1988). Animals 1 through 4 have unknown parents and are assumed unrelated. Both parents of animals 5 through 8 are known. For this example:

$$W' = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & m_5 & m_6 & m_7 & m_8 \end{bmatrix}$$

where $a_1$, $a_2$, $a_3$, and $a_4$ are breeding values of ancestors with unknown parents and $m_5$, $m_6$, $m_7$, and $m_8$ are Mendelian sampling contributions to breeding values of their descendants.

Numerically,

$$W = \text{diag} [1 \ 1 \ 1 \ 1 \ .5 \ .5 \ .5 \ .5]$$

Off-diagonal elements of $W$ are 0 because animals 1 to 4 are unrelated and subsequent Mendelian sampling terms are independent of each other and ancestral breeding values even if selection has occurred (Bulmer, 1971).

If animals are ordered genealogically, $T$ is a lower triangular matrix with diagonal values of 1:

$$T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 4 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 2 & 5 & .5 & .5 & 0 & 0 & 1 \\ 1 & 2 & 6 & .5 & .5 & 0 & 0 & 0 \\ 3 & 4 & 7 & 0 & 0 & .5 & .5 & 0 \\ 5 & 6 & 8 & .5 & .5 & 0 & 0 & 5 \\ 5 & 6 & 8 & .5 & .5 & 0 & 0 & 5 \end{bmatrix}$$

Elements of $T$ are:

$$t_{ij} = .5(t_{sj} + t_{dj})$$

for $j=1,...,i-1$ if both parents are identified:

$$t_{ij} = .5(t_{s(d)j})$$

if only one parent is identified, and:

$$t_{ij} = 0$$

if neither parent is identified; where $s$ and $d$ refer to the sire and dam of animal $i$. For example,

$$t_{81} = .5(t_{s1} + t_{d1}) = .5 (.5 + .5) = .5$$

Breeding values of the eight animals in the example can be expressed as:
For example, the breeding value of animal 8 is expressed as regressions on grandparental breeding values ($a_1$ and $a_2$), Mendelian sampling contributions to parental breeding values ($m_5$ and $m_6$), and the animal's own Mendelian sampling component ($m_8$). The variance of $a_8$ is

$$\text{Var}(0.5a_1 + 0.5a_2 + 0.5m_5 + 0.5m_6 + m_8)$$

If animals 1 and 2 are not inbred, not related, and not selected, then

$$\text{Var}(a_8) = [0.25 + 0.25 + 0.25(0.5) + 0.25(0.5) + (0.5)]\sigma_a^2 = 1.25\sigma_a^2$$

The .25 reflects the contribution of inbreeding ($F_8=.25$).

Under a model of infinite loci, changes in genetic variance from both inbreeding and selection subsequent to establishment of the base population (ancestors with unknown parents) clearly are accounted for by $A\sigma_a^2=\text{TWT}'\sigma_a^2$. The constant $\sigma_a^2$ refers to additive genetic variance in the base population prior to subsequent selection and inbreeding. As shown, breeding values of all animals can be expressed as linear functions of breeding values of ancestral or base population animals and subsequent Mendelian sampling terms. If the number of loci are finite, however, changes in genetic variance from changes in gene frequency are not accounted for by $A\sigma_a^2$.

Are control populations necessary to measure genetic response?

Are control populations necessary to measure genetic response?

Are breeding values comparable across generations?

Claim:

1. Predicted breeding values represent all additive genetic difference among animals.
2. Environmental effects are measured by other factors in mixed models (e.g., contemporary groups, generations).

References


Bulmer, M.G. 1980. The mathematical theory of quantitative genetics. Clarendon Press,
Kennedy, B.W., L.R. Schaeffer, and D.A. Sorensen. 1988. Genetic properties of animal models.
Kennedy, B.W. and D.A. Sorensen. 1988. Properties of mixed model methods for prediction of