An animal breeding approach to the estimation of genetic and environmental trends from field populations

D. J. Garrick

The online version of this article, along with updated information and services, is located on the World Wide Web at:

http://www.journalofanimalscience.org/content/88/13_electronic_suppl/E3

American Society of Animal Science

www.asas.org

An animal breeding approach to the estimation of genetic and environmental trends from field populations[1]

D. J. Garrick[2]

Department of Animal Science, Iowa State University, Ames 50011; and Institute of Veterinary, Animal & Biomedical Sciences, Massey University, Palmerston North, New Zealand

ABSTRACT: Observed or phenotypic trends in animal performance can be readily quantified from information collected from research or field populations. Phenotypic performance is determined by the collective impact of systematic effects that vary by trait, but may include herd, year, sex, and age; additive genetic effects; and a remainder that is referred to as the lack-of-fit or unexplained residual. It is of interest to partition observed performance into these respective components to determine the extent to which genetic or environmental trends or both are responsible for any observed phenotypic trends. An animal breeding approach to separate these components from field data involves the use of a linear model that includes fixed effects for systematic terms and random effects for genetic and residual contributions. The fitted random effects are predicted using a shrinkage estimator known as BLUP that relies only on a translation invariant subset of the field data that does not involve the unknown fixed effects. Fixed effects can then be estimated by adjusting observations for estimates of the random effects. Reliable estimation of trends using this approach requires that relevant fixed effects are recorded, cohorts representing different fixed effects classes are genetically related or connected, and that any records used as the basis for selection in the population are included in the data set.

Key words: best linear unbiased prediction, genetic trend, mixed model evaluation, translation invariance

INTRODUCTION

Changes in management, nutrition, disease exposure, and other environmental factors all contribute to variation in animal performance over time, such as for yearling weights in registered American Angus (Figure 1). Food animal populations are also subject to ongoing effects of selection, such as for growth or milk production, increasing these performance attributes and possibly eroding performance in other unfavorably correlated traits, such as reproduction or longevity. It is useful to characterize phenotypic trends in performance for the entire range of economically relevant traits and to decompose any trend into its genetic and environmental causes (Figure 1). In an animal breeding context, this begins by defining a model equation that specifies the various factors that influence the trait. The simplest model equation for field data observed over time may include fixed effects of herd-year and a random genetic effect for each animal. The difficulty in partitioning the genetic and environmental effects has long been recognized, for example in distinguishing the effects of age in first- compared with second-lactation cows that are confounded with the effects of selection whereby poor performing first lactation cows do not get the opportunity to produce a second lactation. That problem stimulated C. R. Henderson to invent the mixed model equations as part of his PhD research, the solutions to which were subsequently shown by Henderson et al. (1959) to have desirable statistical properties and have become routine in the analysis of field data since being popularized by Henderson (1973). The objective of this paper is to present the well-known mixed model equations and demonstrate by example the manner in which the method enables genetic trends to be separated from environmental trends.

MIXED LINEAR MODEL

The first step in the analysis of field data is to define the model equation that specifies the factors that
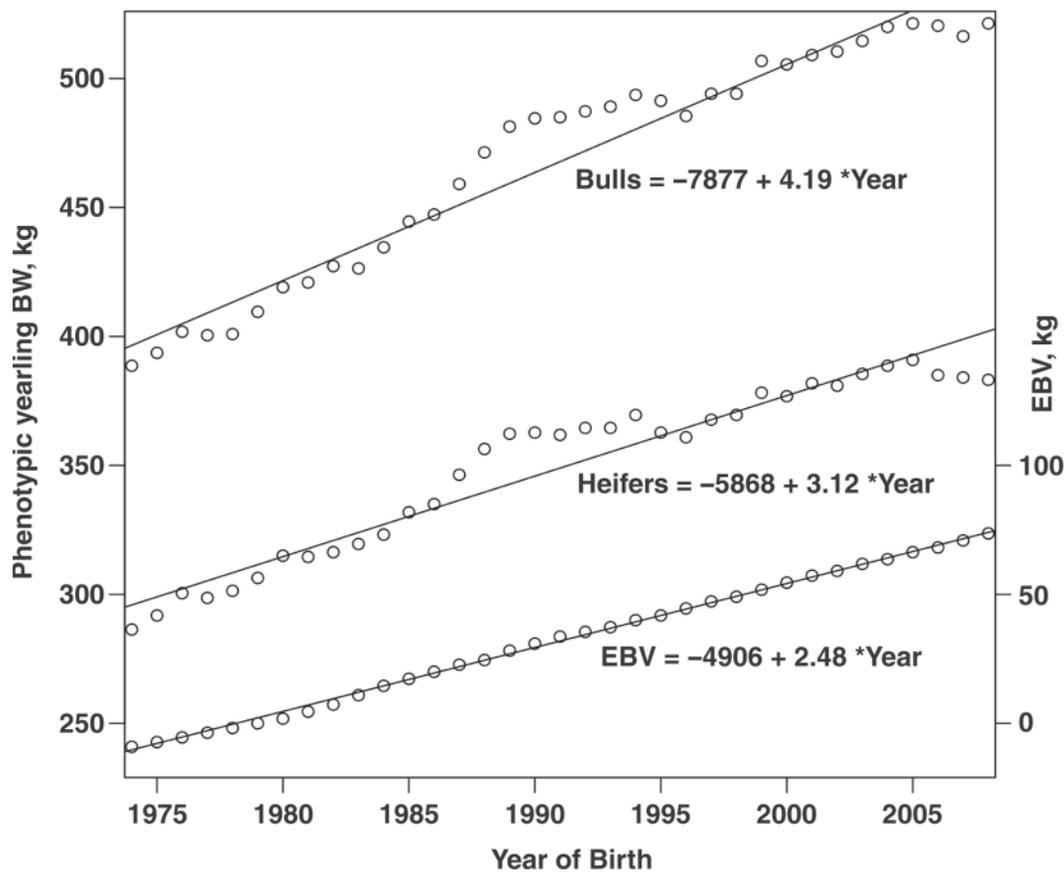
Figure 1. Annual genetic and phenotypic means (kg) and linear trends (kg/yr) in yearling weights of registered American Angus bulls and heifers from 1972 to 2008. Data were derived from American Angus Association (2009) Sire Evaluation Report. Genetic data (i.e., right-hand scale) were converted from EPD to EBV by multiplying by 2. Trends for published adjusted BW obtained from simple linear regression on year of birth after metric conversion of means are depicted in the figure.

influence observed performance. In scalar notation, a possible model equation may be

$$y_{ijk} = \mu_i + s_j + e_{ijk},$$

where y_{ijk} represents an observation on the kth animal, an offspring of the jth sire, observed in the ith cohort, such as herd-year; μ_i represents the unknown fixed effect of the ith cohort; s_j is the random additive genetic effect of the jth sire, and e_{ijk} is the random residual. Although model equations, such as the one above, were motivated by the need to rank candidates for the purposes of selection, they can also be used to quantify trends in genetic merit over time (e.g., Figure 1), for comparison with observed trends that include the collective effects of genetic, environment, and management changes. To fit this model that partitions performance into genetic and environmental contributions, it is not necessary to have every individual observation; we can get the same answers using only the progeny means from each cohort and corresponding numbers of observations by fitting

$$\overline{y}_{ij.} = \mu_i + s_j + \overline{e}_{ij.}.$$

where $\overline{y}_{ij.}$ is the mean of the offspring of sire j in cohort i, $\overline{e}_{ij.}$ is the corresponding mean residual, and n_{ij} is the number of observations contributing to the respective phenotypic and residual means. Assuming that residual effects are independently and identically distributed with $\mathrm{var}(e_{ijk}) = \sigma_e^2$, requiring that dams are non-inbred and unrelated to each other and to the sires, $\mathrm{var}(\overline{e}_{ij.}) = \sigma_e^2/n_{ij}$.

In more general matrix notation, the model equation can be represented as

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e},$$

where $\mathbf{y}$ is a vector of observations; $\mathbf{b}$ is a vector of fixed effects, such as the cohort means; $\mathbf{u}$ is a vector of genetic effects, such as the additive merits of the sires; $\mathbf{X}$ and $\mathbf{Z}$ are incidence matrices relating the effects in $\mathbf{b}$ and $\mathbf{u}$ to the observations in $\mathbf{y}$; and $\mathbf{e}$ is a vector of the random residual effects. The model is not complete un-

til further definition of first and second moments, namely, $E[\mathbf{u}] = \mathbf{0}$, $E[\mathbf{e}] = \mathbf{0}$, $E[\mathbf{y}] = \mathbf{Xb}$; $\text{var}[\mathbf{u}] = \mathbf{G}$, $\text{var}[\mathbf{e}] = \mathbf{R}$, $\text{cov}[\mathbf{u}, \mathbf{e'}] = \mathbf{0}$, and $\text{var}[\mathbf{y}] = \mathbf{ZGZ'} + \mathbf{R}$. In many circumstances, it may also be appropriate to define the distributional properties of the random effects, for example, for hypothesis testing or for likelihood-based analyses. A common assumption is that the genetic and residual effects are normally distributed.

PRACTICAL EXAMPLE

Suppose that there are records on pulmonary arterial pressure (Shirley et al., 2008) of yearling cattle, measured in high altitude conditions to predict risk of brisket disease, from offspring of 3 sires with only 1 of the 3 represented in both of 2 cohorts. For the purpose of this example cohorts will be considered to be years, but they could equally represent herds or herd-year classes. In keeping with field data, the assumption is that each sire has a different number of offspring in each cohort. The distribution of sires across years, their progeny average, and numbers of offspring are shown along with the mean phenotypes in Table 1.

In matrix notation, the model equation for these progeny means can be represented as follows:

$$
\begin{bmatrix} 40 \\ 50 \\ 25 \\ 20 \end{bmatrix} = \begin{bmatrix} \bar{y}_{11.} \\ \bar{y}_{12.} \\ \bar{y}_{22.} \\ \bar{y}_{23.} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} + \begin{bmatrix} \bar{e}_{11.} \\ \bar{e}_{12.} \\ \bar{e}_{22.} \\ \bar{e}_{23.} \end{bmatrix},
$$

with $\text{var}[\mathbf{u}] = \mathbf{I}\sigma_s^2$ if the 3 sires are non-inbred and unrelated and

$$
\text{var} \begin{bmatrix} \bar{e}_{11.} \\ \bar{e}_{12.} \\ \bar{e}_{22.} \\ \bar{e}_{23.} \end{bmatrix}^{-1} = \mathbf{R}^{-1} =
$$

$$
\begin{bmatrix} n_{11} & 0 & 0 & 0 \\ 0 & n_{12} & 0 & 0 \\ 0 & 0 & n_{22} & 0 \\ 0 & 0 & 0 & n_{23} \end{bmatrix} \sigma_e^{-2} = \begin{bmatrix} 10 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & 15 & 0 \\ 0 & 0 & 0 & 20 \end{bmatrix} \sigma_e^{-2}
$$

if dams are non-inbred and unrelated.

MIXED MODEL EQUATIONS

The statistical basis for estimating best linear unbiased estimates (**BLUE**) of the fixed effects was well established using generalized least squares (**GLS**). However, that approach required the construction of

Table 1. Example data set for pulmonary arterial pressure (mm Hg) representing average offspring performance (number of offspring) from 3 sires whose progeny are observed in 1 or 2 yr[1]

Year	Sire 1	Sire 2	Sire 3
Yr 1	40 (10)	50 (5)	
Yr 2		25 (15)	20 (20)

[1]Data are presented as $\bar{y}_{ij.}(n_{ij})$, where $\bar{y}_{ij.}$ is the mean of the offspring of sire j in cohort i and n_{ij} is the number of observations contributing to the respective phenotypic and residual means.

the phenotypic variance-covariance matrix, $\mathbf{V}$, and then used its inverse to form the left- and right-hand side of the GLS equations, namely, $\mathbf{X'V}^{-1}\mathbf{Xb} = \mathbf{X'V}^{-1}\mathbf{y}$. This is not practical in most circumstances with field data. Henderson's mixed model equations (**MME**) do not involve $\mathbf{V}^{-1}$ and were shown by Henderson et al. (1959) to produce BLUE of estimable functions of the fixed effects while simultaneously providing BLUP of the random effects. The MME are routinely used in animal breeding and other contexts that involve the fitting of linear mixed models that include fixed effects other than the mean along with random effects other than the residual. The general form of the MME is

$$
\begin{bmatrix} \mathbf{X'R}^{-1}\mathbf{X} & \mathbf{X'R}^{-1}\mathbf{Z} \\ \mathbf{Z'R}^{-1}\mathbf{X} & \mathbf{Z'R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'R}^{-1}\mathbf{y} \\ \mathbf{Z'R}^{-1}\mathbf{y} \end{bmatrix},
$$

which are like weighted least squares equations except that the inverse of the variance-covariance matrix of the random effects, $\mathbf{G}^{-1}$, is added to a submatrix of the left-hand side. An appealing feature of the MME is that they involve the more tractable matrices $\mathbf{G}^{-1}$ and $\mathbf{R}^{-1}$ rather than the formidable $\mathbf{V}^{-1}$, and the former are typically sparse and easy to form even from large sets of field data. The equations are particularly simple in single trait settings, such as for the data structured as in Table 1 that has MME given by

$$
\begin{bmatrix} n_{1.} & 0 & n_{11} & n_{12} & 0 \\ 0 & n_{2.} & 0 & n_{22} & n_{23} \\ n_{11} & 0 & n_{.1}+\lambda & 0 & 0 \\ n_{12} & n_{22} & 0 & n_{.2}+\lambda & 0 \\ 0 & n_{23} & 0 & 0 & n_{.3}+\lambda \end{bmatrix}^{\sigma_e^{-2}} \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \end{bmatrix} =
$$

$$
\begin{bmatrix} y_{1..} \\ y_{2..} \\ y_{.1.} \\ y_{.2.} \\ y_{.3.} \end{bmatrix}^{\sigma_e^{-2}}
$$

with $\lambda = \dfrac{\sigma_e^2}{\sigma_s^2} = \dfrac{4-h^2}{h^2} = 15$ for $h^2 = 0.25$. Factoring out the scalar inverse of the residual variance, the left-hand side of the MME simply includes elements based on counts of numbers of observations and the right-hand side involves sums of animal performance. These equations are identical to those that would be formed if every individual observation was used rather than just the progeny means and counts. Inspection of a typical equation for fixed effects, such as that from row 1 of the MME, shows that

$$\begin{bmatrix} n_{1.} & 0 & n_{11} & n_{12} & 0 \end{bmatrix} \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \end{bmatrix} = \begin{bmatrix} y_{1..} \end{bmatrix},$$

with fixed effect solution

$$\hat{\mu}_1 = \frac{y_{1..} - \left(n_{11}\hat{s}_1 + n_{12}\hat{s}_2\right)}{n_{1.}}.$$

This solution makes sense in that the estimate of the year effect is simply the mean performance in that year, adjusted for the merit of the sires whose progeny were represented in that year. This adjustment for sire merit ensures that year effects are not biased upwards or downwards due to having above- or below-average sires represented in a particular year. The estimate of the year effects represents the performance that would have been observed if the same sires were equally represented in every year.

A similar inspection of an equation for a sire effect, such as that for sire 2 shows

$$\begin{bmatrix} n_{12} & n_{22} & 0 & \left(n_{.2} + \lambda\right) & 0 \end{bmatrix} \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \end{bmatrix} = \begin{bmatrix} y_{.2.} \end{bmatrix},$$

with random effect solution

$$\hat{s}_2 = \frac{y_{.2.} - \left(n_{12}\hat{\mu}_1 + n_{22}\hat{\mu}_2\right)}{n_{.2} + \lambda}.$$

This solution is more complex than the equation for solving a fixed effect because the denominator includes the addition of a variance ratio. If it were not for the

variance ratio, the solution would be the deviation defined by the average performance of the offspring of the sire, adjusted for the effect of the year in which the progeny were observed. The addition of the variance ratio shrinks the estimate, the amount of shrinkage depending upon the number of progeny of the sire in relation to the size of the variance ratio (i.e., $\lambda = \dfrac{4-h^2}{h^2} = 15$ in this example). Given a trait with $h^2 = 0.25$, such as in this example, the deviation is shrunk by a little more than the fraction $n_{.j}/\left(n_{.j}+15\right)$, the actual value depending upon contemporary group size and distribution of paternity in the contemporary groups as these influence loss of information from unknown fixed effects. If the sire has say 100 progeny in large contemporary groups, the shrunk deviation will be some 4/5 the adjusted deviation; that is, little shrinkage will occur. If the sire has 15 or less progeny (i.e., $n \leq \lambda$), the deviation in progeny performance will be shrunk by at least one-half.

Populating these equations scaled by the residual variance leads to the following MME

$$\begin{bmatrix} 15 & 0 & 10 & 5 & 0 \\ 0 & 35 & 0 & 15 & 20 \\ 10 & 0 & 10+15 & 0 & 0 \\ 5 & 15 & 0 & 20+15 & 0 \\ 0 & 20 & 0 & 0 & 20+15 \end{bmatrix} \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \end{bmatrix} = \begin{bmatrix} 650 \\ 775 \\ 400 \\ 625 \\ 400 \end{bmatrix}.$$

Simultaneously solving these equations provides estimates of the cohort means $\begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{bmatrix} = \begin{bmatrix} 43.5 \\ 21.7 \end{bmatrix}$, year trend $\hat{\mu}_1 - \hat{\mu}_2 = 21.8$, and sire effects $\begin{bmatrix} \hat{s}_1 & \hat{s}_2 & \hat{s}_3 \end{bmatrix} = \begin{bmatrix} -1.39 & 2.35 & -0.96 \end{bmatrix}$. Based on these evaluations, the ranking of the 3 sires, recognizing that smaller arterial pressures are better because they are associated with reduced incidence of brisket disease, would be $1 > 3 > 2$. A producer that simply considered raw performance (Table 1) would have ranked the 3 sires differently as $3 > 2 > 1$. A producer that compared the sires on a within-cohort basis would have concluded that $1 > 2$ (yr 1) and $3 > 2$ (yr 2) but would have had more difficulty relating the merit of sires 1 and 3 that were not directly compared. Note that the differences between the sires have been shrunk in relation to the differences between the sire averages, reflecting the shrinkage property of BLUP. Less shrinkage would occur if the progeny means were based on larger numbers of offspring.

The dependence of the solutions to the fixed effects on solutions to the random effects and vice versa is apparent in the above representations, showing that the equations need to be solved simultaneously. In the con-

text of genetic evaluation, fixed effects can be viewed as nuisance parameters, of no particular interest. It is possible to absorb the equations for the fixed effects, leading to a reduced set of equations for random effects that will produce identical solutions to the full MME. These reduced equations whereby solutions to the fixed effects have been eliminated are $\left[\mathbf{Z'SZ} + \mathbf{G}^{-1}\right]\left[\hat{\mathbf{u}}\right] = \left[\mathbf{Z'Sy}\right]$

with $\quad \mathbf{S} = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{X}\left(\mathbf{X'R}^{-1}\mathbf{X}\right)^{-}\mathbf{X'R}^{-1} \quad$ representing the so-called absorption matrix. In the scenario of this particular example, these equations are

$$\begin{bmatrix} r^{11} + \lambda & -r^{11} & 0 \\ -r^{11} & r^{11} + r^{22} + \lambda & -r^{22} \\ 0 & -r^{22} & r^{22} + \lambda \end{bmatrix}\begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \end{bmatrix} = \\ \begin{bmatrix} \left(\overline{y}_{11.} - \overline{y}_{12.}\right)r^{11} \\ \left(\overline{y}_{12.} - \overline{y}_{11.}\right)r^{11} + \left(\overline{y}_{22.} - \overline{y}_{23.}\right)r^{22} \\ \left(\overline{y}_{23.} - \overline{y}_{22.}\right)r^{22} \end{bmatrix}$$

with $r^{11} = \dfrac{n_{11}n_{12}}{\left(n_{11} + n_{12}\right)}$ and $r^{22} = \dfrac{n_{22}n_{23}}{\left(n_{22} + n_{23}\right)}$.

These equations with cohort effects absorbed were widely used in fitting sire and sire-maternal grandsire models, but are seldom used in modern evaluations based on animal models because they increase the number of nonzero off-diagonal elements in the MME.

TRANSLATION INVARIANCE

Random effects predicted using BLUP are translation invariant, meaning that the predictions do not change if for example any constant is added to all the observations in a particular cohort. The manner in which this is achieved is by ignoring some of the data in predicting the random effects. It is not typically efficient to ignore data in an analysis, but the data ignored in this context contain no valuable information about the predictions of random effects when the true or parametric values of the fixed effects are unknown. This is best demonstrated by example.

Recall the 4 progeny-year means for pulmonary arterial pressure represented in Table 1 from the viewpoint of predicting sire effects. Consider the contrast between 2 sires with progeny in different years, such as the performance of sire 1 in yr 1 compared with the performance of sire 3 in yr 2. These 2 means represent the effect of sire 1 in yr 1 and the effect of sire 3 in year 2, respectively. Accordingly, their difference or contrast involves the difference in merit between sire 1 and 3 plus the difference in the effects of yr 1 and yr 2. In this contrast, sire and year effects are said to be confounded. Therefore, there is no useful information about the merit of these 2 unrelated sires when the year effects

are unknown. Suppose we now consider the contrast between progeny means of different sires in a particular year, such as sire 1 in yr 1 and sire 2 in yr 1. Because both of these progeny groups were measured in yr 1, the difference in their average performance does not involve the unknown year effect. Similarly, the difference in average performance of sire 2 in yr 2 and sire 3 in yr 2 provides useful information about sire differences, unencumbered by the unknown yr 2 effect. It turns out that these 2 contrasts between the 4 progeny-year means contains all the available information about sire differences when year effects are treated as fixed and unknown. This can be shown by defining 2 linear contrasts to represent the informative differences in off-spring performance within cohorts. Define $\mathbf{y'} = \begin{bmatrix} \overline{y}_{11.} & \overline{y}_{12.} & \overline{y}_{22.} & \overline{y}_{23.} \end{bmatrix}$ and the contrast between sires in yr 1 as $\mathbf{k_1'} = \begin{bmatrix} 1 & -1 & 0 & 0 \end{bmatrix}$ and sires in yr 2 as $\mathbf{k_2'} = \begin{bmatrix} 0 & 0 & 1 & -1 \end{bmatrix}$ with these 2 contrasts accumulated in columns of the matrix $\mathbf{K}$ such that

$$\mathbf{K'y} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}\begin{bmatrix} \overline{y}_{11.} \\ \overline{y}_{12.} \\ \overline{y}_{22.} \\ \overline{y}_{23.} \end{bmatrix} = \begin{bmatrix} \overline{y}_{11.} & -\overline{y}_{12.} \\ \overline{y}_{22.} & -\overline{y}_{23.} \end{bmatrix}.$$

Rather than fitting a model for $\mathbf{y}$ we will now fit the model $\mathbf{y}^* = \mathbf{K'y} = \mathbf{K'Xb} + \mathbf{K'Zu} + \mathbf{K'e}$, recognizing the modifications that are required to use the MME. First, note that

$$\mathbf{K'X} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

so the incidence matrix for fixed effects for fitting our contrasts is null. Accordingly, there are no fixed effects to fit for $\mathbf{K'y}$. Second, $\mathbf{Z}^* = \mathbf{K'Z}$. Third, the residuals are $\mathbf{K'e}$ with $\mathrm{var}[\mathbf{K'e}] = \mathbf{K'}\,\mathrm{var}[\mathbf{e}]\mathbf{K} = \mathbf{K'RK} = \mathbf{R}^*$. The MME are therefore $\left[\mathbf{Z}^{*\prime}\mathbf{R}^{*-1}\mathbf{Z}^* + \mathbf{G}^{-1}\right]\left[\hat{\mathbf{u}}\right] = \left[\mathbf{Z}^{*\prime}\mathbf{R}^{*-1}\mathbf{y}^*\right]$, which in the context of our example are

$$\begin{bmatrix} r^{11} + \lambda & -r^{11} & 0 \\ -r^{11} & r^{11} + r^{22} + \lambda & -r^{22} \\ 0 & -r^{22} & r^{22} + \lambda \end{bmatrix}\begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \end{bmatrix} = \\ \begin{bmatrix} \left(\overline{y}_{11.} - \overline{y}_{12.}\right)r^{11} \\ \left(\overline{y}_{12.} - \overline{y}_{11.}\right)r^{11} + \left(\overline{y}_{22.} - \overline{y}_{23.}\right)r^{22} \\ \left(\overline{y}_{23.} - \overline{y}_{22.}\right)r^{22} \end{bmatrix},$$

exactly the same equations when all the data were used and the fixed effects were absorbed. This demonstrates that the only information used to predict the random effects are linear contrasts that do not involve the unknown fixed effects. Given estimates of the random effects, it is possible to recover all the information from every year to estimate the fixed effects by adjusting that data for the now quantified estimates of the random effects.

Modern evaluations typically comprise overlapping generations with some individuals being mildly inbred and many mates being related. In those circumstances, it is more convenient to fit a so-called animal model that includes an equation for every animal represented in the data file or the pedigree file, rather than just fitting an equation for each sire in the data file as in the example.

PRACTICAL CONSIDERATIONS

It was demonstrated in the example discussed previously that mixed model analyses take advantage of a subset of possible linear combinations of the data to estimate the random effects. The linear contrasts chosen are those that do not involve functions of the unknown fixed effects. Genetic trends are obtained from estimates of differences between sires, and these differences will be reliably estimated when sires have many offspring that compete directly with offspring of other sires in the same cohort groups. Indirect connections can be used to estimate the difference between sires whose offspring were never directly compared in the same cohort. These linkages rely on common sires to connect cohorts. Such sires might be reference sires used specifically for the purpose, or any widely used sire, such as legacy AI bulls that are well-represented in many pedigrees. In some cases, the cross-classification of sires among contemporary groups may be incomplete such that certain groups of animals have never been compared with other groups of animals. In that case, the data are said to be disconnected. Figure 2 depicts well-connected, weakly connected, and disconnected data sets. Strictly speaking, analysts should check for connectedness (Fernando et al., 1983) before publishing evaluation results. Even when data are connected, so that any animal can legitimately be compared with every other animal, the strength of the comparison will be limited by the quality of the connection. The quality of connections can be represented by the prediction error covariances between animals (Kennedy and Trus, 1993), but there are so many of these that they are not routinely calculated or published.

To obtain reliable comparison of different environments or locations, genetic links or connections are required across environments or locations. To obtain reliable estimation of genetic trend, genetic links or connections are required across time. It is not necessary for the same individuals to provide all the connections, as would occur for example if a control sire were used continuously across time. From a national population perspective, a team of AI sires provides the within-breed links across locations and time. Such teams typically have some new members added each year, whereas others are removed. Connections across long periods, such as 2 to 3 decades, are achieved by the collective action of many short-term connections of 2 to 5 yr involving different sires.

In principle it is not sires themselves that must be cross-classified with environments or time. Rather, it is necessary that there is a known genetic covariance between animals in different cohorts. Due to the fact that sires tend to be used more widely and are better identified than individual dams, it is sire relationships that contribute much of the genetic covariance. Even if new sires were used every year, for only 1 yr, the fact that dams are used across years would provide linkages to enable the prediction of genetic trend. The availability of genomic markers, such as high-density SNP panels, have now made it possible to compute genetic covariances without any pedigree information. The extent to which alleles are shared on average between animals allows them to be categorized by breed and often by sire. This will open up new opportunities for extracting useful information from field data without requiring pedigree information.

Given that linear contrasts among offspring of different sires in the same cohort are the driving information for genetic evaluation, it stands to reason that no information is obtained on the relative merit of sires whose progeny are represented in a single-sire contemporary group. However, single sire contemporary groups may provide information on the relative merits of dams and

Yr	Sire 1	Sire 2	Sire 3	Sire 4	Sire 5	Sire 6	Sire 7	Sire 1	Sire 2	Sire 3	Sire 4	Sire 5	Sire 6	Sire 7	Sire 1	Sire 2	Sire 3	Sire 4	Sire 5	Sire 6	Sire 7
Yr 1	X	X						X		X	X				X		X	X			
Yr 2	X	X	X					X	X		X				X	X		X			
Yr 3		X		X	X			X	X	X					X	X	X				
Yr 4			X	X		X			X		X	X				X		X			
Yr 5				X	X	X						X		X					X		X
Yr 6					X	X	X					X	X						X	X	
Yr 7						X	X						X	X						X	X

Figure 2. Cross-classification of sires and years to demonstrate well-connected (left), weakly connected (center), and disconnected (right) data sets. In weakly connected data, only sire 5 in yr 4 provides connections between sires 1 to 4 in yr 1 to 4 and sires 6 to 7 in yr 5 to 7. In the disconnected data, sires 1 to 4 in yr 1 to 4 are disconnected from sires 5 to 7 in yr 5 to 7.

therefore contribute indirectly to paternal differences through relationships such as maternal grandsires.

Given that collective differences between sires represented together in individual cohorts contribute to the overall comparison of every sire with each other, it is apparent that one must assume that the merit of particular sires does not change with cohort. There are several circumstances when this might not be the case. First, preselection of the best offspring of 1 or more sires to enter a particular cohort can bias results unless the information used in the preselection is appropriately incorporated in the MME. Second, cohorts might represent different years of birth or different environments, such as herd or region. The analysis, therefore, assumes that the merit of the sire does not change over time. Strictly speaking, the additive genetic merit is the sum of average effects of genes and average effects are influenced by gene frequency in the presence of non-additive gene action, such as dominance or epistasis. However, changes in gene frequency are likely to be gradual and unlikely to modify average effects unless cohorts represent different breeds or the trait is subject to the influence of 1 or more major genes rather than a large or infinitesimal number of small genes. Third, the use of sires in extreme environments, such as those that vary widely in stressors such as climate, nutrition, or disease, might exhibit genotype-environment interaction and, therefore, compromise calculation of trends.

Reliable estimation of genetic trends requires reliable partitioning of phenotypic performance into systematic effects, and its genetic and random components. This cannot be achieved when information to characterize the systematic effects is not available. For example, if feedlot groups represent different birth and backgrounding cohorts, these effects must be taken into account to get reliable partitioning of effects based on subsequent performance. Collecting information, such as birth date, birth cohort, age of dam, and management groups, can be particularly challenging in field circumstances in which animals can change ownership and may also change their system of individual identification over time.

In addition to knowledge of the model equation and relevant information to construct the incidence matrices, a requirement for BLUP is that the variance components be known without error (Henderson, 1973). In practice, genetic evaluations are carried out as if the variance components are known, often using values obtained from REML estimation of variance components from the field data itself, or using assumed values based on published literature. Results may be influenced by the variance used, for example the heritability in a single-trait analysis (Blair and Pollak, 1984; Sorensen and Kennedy, 1984; Thompson, 1986), the influence depending upon the amount of information available on candidates at the time of selection.

Comparison of genetic trends with phenotypic trends allows prediction of trends in management and the environment. These can be useful, provided the model equation and other assumptions hold, to retrospectively determine the underlying causes of change. The example in Figure 1 using registered Angus yearling weight records adjusted for systematic effects, such as age of dam, shows that the genetic trend is 2.48 kg/yr, whereas the adjusted phenotypic trends are greater in both sexes, being 3.12 kg/yr in heifers and 4.19 kg/yr in bulls. These data demonstrate that management or environmental factors fluctuate more widely from year to year than do genetic changes and shows that nongenetic factors, perhaps nutritional regimens, are improving by an average 0.64 kg/yr in heifers (i.e., 3.12 kg/yr phenotypic less 2.48 kg/yr genetic) and by a greater amount of 1.71 kg/yr in bulls (i.e., 4.19 to 2.48 kg/yr). Those sex differences are perhaps not surprising given that registered yearling bulls are principally being prepared for sale as sires, whereas yearling females are principally being managed for their own reproductive performance.

SUMMARY

Phenotypic trends can be partitioned into genetic and environmental components through fitting of mixed linear models that include fixed cohort and random genetic effects. This procedure requires that factors that influence variation in performance within a cohort are known, and recorded, along with relevant information, such as herd, year, and management groups, to define the cohorts themselves. Further it requires nonzero genetic covariance between cohorts, which has historically been created by use of common sires across herds, years, and management groups, but could be derived in the future using genomic similarity.

Estimates of the environmental effects rely on estimates for the genetic effects and vice versa, requiring a system of equations for all factors to be solved simultaneously. It can be shown that estimates of the genetic effects are obtained from only a subset of the observed data, with that subset representing the set of contrasts that do not involve linear functions of unknown fixed effects such as cohorts which identify animals that performed together in the same herd or year. Given estimates of the random effects, estimates of the fixed effects can be recovered from all the data by adjusting observations for these estimates of the random effects.

LITERATURE CITED

American Angus Association. 2009. Fall 2009 Sire Evaluation Report. Am. Angus Assoc., St Joseph, MO.

Blair, H. T., and E. J. Pollak. 1984. Estimation of genetic trend in a selected population with and without the use of a control population. J. Anim. Sci. 58:878–886.

Fernando, R. L., D. Gianola, and M. Grossman. 1983. Identifying all connected subsets in a two-way classification without interaction. J. Dairy Sci. 66:1399–1402.

Henderson, C. R. 1973. Sire evaluation and genetic trends. Pages 10–41 in Proc. Anim. Breeding Genet. Symp. in Honor of Dr. J. L. Lush. ASAS and ADSA, Champaign, IL.

Henderson, C. R., O. Kempthorne, S. R. Searle, and C. N. Von Krosigk. 1959. Estimation of environmental and genetic trends from records subject to culling. Biometrics 13:192–218.

Kennedy, B. W., and D. Trus. 1993. Considerations on genetic connectedness between management units under an animal model. J. Anim. Sci. 71:2341–2352.

Shirley, K. L., D. W. Beckman, and D. J. Garrick. 2008. Inheritance of pulmonary arterial pressure in Angus cattle and its correlation with growth. J. Anim. Sci. 86:815–819.

Sorenson, D. A., and B. W. Kennedy. 1984. Estimation of response to selection using least-squares and mixed model methodology. J. Anim. Sci. 58:1097–1106.

Thompson, R. 1986. Estimation of realized heritability in a selected population using mixed model methods. Genet. Sel. Evol. 18:475–484.

References	This article cites 7 articles, 4 of which you can access for free at:
	http://www.journalofanimalscience.org/content/88/13_electronic_suppl/E3 #BIBL
Citations	This article has been cited by 1 HighWire-hosted articles:
	http://www.journalofanimalscience.org/content/88/13_electronic_suppl/E3 #otherarticles