

Random Regression Models

1 Introduction

All biological creatures grow and perform over their lifetime. Traits that are measured at various times during that life are known as *longitudinal* data. Examples are body weights, body lengths, milk production, feed intake, fat deposition, and egg production. On a biological basis there could be different genes that turn on or turn off as an animal ages causing changes in physiology and performance. Also, an animal's age can be recorded in years, months, weeks, days, hours, minutes, or seconds, so that, in effect, there could be a continuum or continuous range of points in time when an animal could be observed for a trait. These traits have also been called *infinitely dimensional* traits.

Take body weight on gilts from a 60-day growth test as an example.

| Animal | Days on Test | | | | | |
|--------|--------------|-----|-----|-----|-----|-----|
| | 10 | 20 | 30 | 40 | 50 | 60 |
| 1 | 42 | 53 | 60 | 72 | 83 | 94 |
| 2 | 30 | 50 | 58 | 68 | 76 | 85 |
| 3 | 38 | 44 | 51 | 60 | 70 | 77 |
| SD | 1.6 | 3.7 | 3.9 | 5.0 | 5.3 | 5.6 |

The differences among the three animals increase with days on test as the gilts become heavier. As the mean weight increases, so also the standard deviation of weights increases. The weights over time could be modeled as a mean plus covariates of days on test and days on test squared. Depending on the species and trait, perhaps a cubic or spline function would fit the data better. The point is that the means can be fit by a linear model with a certain number of parameters.

2 Multiple Trait Approach

The data presented in the previous table have typically been analyzed such that the weights at each day on test are different traits. If t is the day on test, i.e. 10, 20, 30, 40, 50, or 60, then a model for any one of the weights could be

$$\mathbf{y}_t = \mathbf{X}\mathbf{b}_t + \mathbf{a}_t + \mathbf{e}_t,$$

which is just a simple, single record, animal model. Analyses are usually done so that the genetic and residual variances and covariances are estimated among the six weights. Suppose that an

estimate of the genetic variances and covariances was

$$\mathbf{G} = \begin{pmatrix} 2.5 & 4.9 & 4.6 & 4.6 & 4.3 & 4.0 \\ 4.9 & 13.5 & 12.1 & 12.3 & 11.9 & 10.7 \\ 4.6 & 12.1 & 15.2 & 14.5 & 14.6 & 12.5 \\ 4.6 & 12.3 & 14.5 & 20.0 & 19.0 & 16.9 \\ 4.3 & 11.9 & 14.6 & 19.0 & 25.0 & 20.3 \\ 4.0 & 10.7 & 12.5 & 16.9 & 20.3 & 30.0 \end{pmatrix}.$$

Let the residual covariance matrix be

$$\mathbf{R} = \begin{pmatrix} 3.8 & 7.4 & 6.9 & 6.8 & 6.4 & 6.0 \\ 7.4 & 20.3 & 18.2 & 18.4 & 17.9 & 16.1 \\ 6.9 & 18.2 & 22.8 & 21.8 & 21.9 & 18.8 \\ 6.8 & 18.4 & 21.8 & 30.0 & 28.5 & 25.4 \\ 6.4 & 17.9 & 21.9 & 28.5 & 37.5 & 30.5 \\ 6.0 & 16.1 & 18.8 & 25.4 & 30.5 & 45.0 \end{pmatrix}.$$

Assuming a model with only an intercept, and that the three animals are unrelated, then

$$(\mathbf{X} \ \mathbf{Z}) = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \otimes \mathbf{I}_6,$$

where the identity is of order 6 and \otimes is the direct product operator. The observations would be ordered by days on test within animals, i.e.,

$$\mathbf{y}' = \left(42 \ 53 \ 60 \ 72 \ 83 \ 94 \ \dots \ 60 \ 70 \ 77 \right).$$

The resulting MME would be of order 24 by 24, and the solutions would be as follows.

| Days on | | | | |
|---------|-------|----------|----------|----------|
| Test | Mean | Animal 1 | Animal 2 | Animal 3 |
| 10 | 36.67 | 2.10 | -2.61 | 0.51 |
| 20 | 49.00 | 1.57 | 0.45 | -2.02 |
| 30 | 56.33 | 1.48 | 0.64 | -2.12 |
| 40 | 66.67 | 2.21 | 0.39 | -2.60 |
| 50 | 76.33 | 2.72 | -0.24 | -2.48 |
| 60 | 85.33 | 3.48 | -0.16 | -3.32 |

Animal 1 clearly grew faster than the other two animals and its superiority grew larger with time. Animals 2 and 3 switched rankings after the first 10 days, and Animal 3 was the slower growing animal. The estimates for the mean give an average growth curve for the 3 animals.

A multiple trait approach may be appropriate here because every animal was weighed on exactly the same number of days on test throughout the trial. However, suppose the animals

were of different ages at the start of test, and suppose that instead of days on test, the ages for each weight were given. Assume at start of test that Animal 1 was 18 days old, Animal 2 was 22, and Animal 3 was 25. The multiple trait model could include a factor (classification or covariable) to account for different starting ages. The differences observed at any point in time could be due to the ages of the animals rather than just on the number of days on test. The analysis shown above would have an implied assumption that all animals began the test at the same age.

3 Covariance Functions

Let the example data be as shown below, allowing for the different ages at each test. Note that the ages range from 28 days to 85 days, and that none of the animals were ever weighed at exactly the same age.

| Animal 1 | | Animal 2 | | Animal 3 | |
|----------|----|----------|----|----------|----|
| Age | Wt | Age | Wt | Age | Wt |
| 28 | 42 | 32 | 30 | 35 | 38 |
| 38 | 53 | 42 | 50 | 45 | 44 |
| 48 | 60 | 52 | 58 | 55 | 51 |
| 58 | 72 | 62 | 68 | 65 | 60 |
| 68 | 83 | 72 | 76 | 75 | 70 |
| 78 | 94 | 82 | 85 | 85 | 77 |

Kirkpatrick et al.(1991) proposed the use of covariance functions for longitudinal data of this kind. A covariance function (CF) is a way to model the variances and covariances of a longitudinal trait. Orthogonal polynomials are used in this model and the user must decide the order of fit that is best. Legendre polynomials (1797) are the easiest to apply.

To calculate Legendre polynomials, first define

$$\begin{aligned} P_0(x) &= 1, \text{ and} \\ P_1(x) &= x, \end{aligned}$$

then, in general, the $n + 1$ polynomial is described by the following recursive equation:

$$P_{n+1}(x) = \frac{1}{n+1} ((2n+1)xP_n(x) - nP_{n-1}(x)).$$

These quantities are "normalized" using

$$\phi_n(x) = \left(\frac{2n+1}{2}\right)^{.5} P_n(x).$$

This gives the following series,

$$\begin{aligned}
\phi_0(x) &= \left(\frac{1}{2}\right)^{.5} P_0(x) = .7071 \\
\phi_1(x) &= \left(\frac{3}{2}\right)^{.5} P_1(x) \\
&= 1.2247x \\
P_2(x) &= \frac{1}{2}(3xP_1(x) - 1P_0(x)) \\
\phi_2(x) &= \left(\frac{5}{2}\right)^{.5} \left(\frac{3}{2}x^2 - \frac{1}{2}\right) \\
&= -.7906 + 2.3717x^2,
\end{aligned}$$

and so on. The first six can be put into a matrix, Λ , as

$$\Lambda' = \begin{pmatrix} .7071 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.2247 & 0 & 0 & 0 & 0 \\ -.7906 & 0 & 2.3717 & 0 & 0 & 0 \\ 0 & -2.8062 & 0 & 4.6771 & 0 & 0 \\ .7955 & 0 & -7.9550 & 0 & 9.2808 & 0 \\ 0 & 4.3973 & 0 & -20.5206 & 0 & 18.4685 \end{pmatrix}.$$

Now define another matrix, \mathbf{M} , as a matrix containing the polynomials of standardized time values. Legendre polynomials are defined within the range of values from -1 to +1. Thus, ages or time periods have to be standardized (converted) to the interval between -1 to +1. The formula is

$$q_\ell = -1 + 2 \left(\frac{t_\ell - t_{min}}{t_{max} - t_{min}} \right).$$

Let the minimum starting age for pigs on test be 15 days and the maximum starting age be 28 days, then the maximum age at end of test was 88 days. Thus, $t_{min} = 25 = (15 + 10)$ and $t_{max} = 88 = (28 + 60)$.

The matrix \mathbf{G} was based on weights taken on pigs that were all 21 days of age at start of test. The table below shows the ages and standardized time values for the six weigh dates.

| Days on Test | Age | Standardized Value |
|--------------|-----|--------------------|
| 10 | 31 | -1.000 |
| 20 | 41 | -.600 |
| 30 | 51 | -.200 |
| 40 | 61 | .200 |
| 50 | 71 | .600 |
| 60 | 81 | 1.000 |

Therefore,

$$\mathbf{M} = \begin{pmatrix} 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -.600 & .360 & -.216 & .130 & -.078 \\ 1 & -.200 & .040 & -.008 & .002 & -.000 \\ 1 & .200 & .040 & .008 & .002 & .000 \\ 1 & .600 & .360 & .216 & .130 & .078 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

This gives

$$\begin{aligned} \Phi &= \mathbf{M}\Lambda, \\ &= \begin{pmatrix} .7071 & -1.2247 & 1.5811 & -1.8708 & 2.1213 & -2.3452 \\ .7071 & -.7348 & .0632 & .6735 & -.8655 & .3580 \\ .7071 & -.2449 & -.6957 & .5238 & .4921 & -.7212 \\ .7071 & .2449 & -.6957 & -.5238 & .4921 & .7212 \\ .7071 & .7348 & .0632 & -.6735 & -.8655 & -.3580 \\ .7071 & 1.2247 & 1.5811 & 1.8708 & 2.1213 & 2.3452 \end{pmatrix}. \end{aligned}$$

which can be used to specify the elements of \mathbf{G} as

$$\begin{aligned} \mathbf{G} &= \Phi\mathbf{H}\Phi' \\ &= \mathbf{M}(\Lambda\mathbf{H}\Lambda')\mathbf{M}' \\ &= \mathbf{M}\mathbf{T}\mathbf{M}'. \end{aligned}$$

Note that Φ , \mathbf{M} , and Λ are matrices defined by the Legendre polynomial functions and by the standardized time values and do not depend on the data or values in the matrix \mathbf{G} . Therefore, it is possible to estimate either \mathbf{H} or \mathbf{T} ,

$$\begin{aligned} \mathbf{H} &= \Phi^{-1}\mathbf{G}\Phi^{-T}, \\ &= \begin{pmatrix} 27.69 & 5.29 & -1.95 & 0.05 & -1.17 & 0.52 \\ 5.29 & 4.99 & 0.42 & -0.25 & -0.30 & -0.75 \\ -1.95 & 0.42 & 1.51 & 0.20 & -0.33 & -0.07 \\ 0.05 & -0.25 & 0.20 & 1.19 & 0.06 & -0.71 \\ -1.17 & -0.30 & -0.33 & 0.06 & 0.58 & 0.15 \\ 0.52 & -0.75 & -0.07 & -0.71 & 0.15 & 1.12 \end{pmatrix}, \end{aligned}$$

and

$$\begin{aligned} \mathbf{T} &= \mathbf{M}^{-1}\mathbf{G}\mathbf{M}^{-T} \\ &= \begin{pmatrix} 16.44 & 6.48 & -5.93 & -11.49 & -0.93 & 10.02 \\ 6.48 & 49.87 & -2.05 & -155.34 & 1.44 & 111.23 \\ -5.93 & -2.05 & 57.71 & 28.62 & -50.06 & -25.73 \\ -11.49 & -155.34 & 28.62 & 635.49 & -26.91 & -486.90 \\ -0.93 & 1.44 & -50.06 & -26.91 & 49.80 & 26.49 \\ 10.02 & 111.23 & -25.73 & -486.90 & 26.49 & 382.79 \end{pmatrix}. \end{aligned}$$

Why orthogonal polynomials? Convert \mathbf{T} and \mathbf{H} to correlation matrices.

$$\mathbf{T}_{cor} = \begin{pmatrix} 1.00 & .23 & -.19 & -.11 & -.03 & .13 \\ .23 & 1.00 & -.04 & -.87 & .03 & .81 \\ -.19 & -.04 & 1.00 & .15 & -.93 & -.17 \\ -.11 & -.87 & .15 & 1.00 & -.15 & -.99 \\ -.03 & .03 & -.93 & -.15 & 1.00 & .19 \\ .13 & .81 & -.17 & -.99 & .19 & 1.00 \end{pmatrix},$$

and

$$\mathbf{H}_{cor} = \begin{pmatrix} 1.00 & .45 & -.30 & .01 & -.29 & .09 \\ .45 & 1.00 & .15 & -.10 & -.17 & -.32 \\ -.30 & .15 & 1.00 & .15 & -.36 & -.05 \\ .01 & -.10 & .15 & 1.00 & .07 & -.62 \\ -.29 & -.17 & -.36 & .07 & 1.00 & .19 \\ .09 & -.32 & -.05 & -.62 & .19 & 1.00 \end{pmatrix}.$$

The largest absolute correlation in \mathbf{T} was .99, while the largest absolute correlation in \mathbf{H} was only .62. Orthogonal polynomials tend to reduce the correlations between estimated regression coefficients. This is advantageous when trying to estimate \mathbf{H} by REML or Bayesian methods, because the estimates would converge faster to the maximum or appropriate posterior distribution than trying to estimate \mathbf{T} . The matrix \mathbf{T} actually had four correlations greater than 0.80 in absolute value, while \mathbf{H} had none. There are other kinds of orthogonal polynomials, but Legendre polynomials are probably the easiest to calculate and utilize.

\mathbf{H} can be used to calculate the covariance between any two days on test between 10 and 60 days. To compute the covariance between days 25 and 55, calculate the Legendre polynomial covariates as in calculating a row of Φ . The standardized time values for days 25 and 55 are -0.4 and 0.8, respectively. The Legendre polynomials (stored in \mathbf{L} are

$$\mathbf{L} = \begin{pmatrix} .7071 & -.4899 & -.4111 & .8232 & -.2397 & -.6347 \\ .7071 & .9798 & .7273 & .1497 & -.4943 & -.9370 \end{pmatrix}.$$

Then the variances and covariance for those two ages are

$$\mathbf{LHL}' = \begin{pmatrix} 14.4226 & 13.7370 \\ 13.7370 & 28.9395 \end{pmatrix}.$$

Thus, the genetic correlation between days 25 and 55 is 0.67. The same calculations could be repeated for the residual variance-covariance matrix. Let

$$\begin{aligned} \mathbf{S} &= \Phi^{-1}\mathbf{R}\Phi^{-T}, \\ &= \begin{pmatrix} 41.57 & 7.94 & -2.91 & 0.11 & -1.76 & 0.76 \\ 7.94 & 7.45 & 0.62 & -0.41 & -0.44 & -1.07 \\ -2.91 & 0.62 & 2.29 & 0.31 & -0.52 & -0.12 \\ 0.11 & -0.41 & 0.31 & 1.76 & 0.08 & -1.04 \\ -1.76 & -0.44 & -0.52 & 0.08 & 0.88 & 0.24 \\ 0.76 & -1.07 & -0.12 & -1.04 & 0.24 & 1.64 \end{pmatrix}, \end{aligned}$$

then the residual variances and covariances for days 25 and 55 would be

$$\mathbf{LSL}' = \begin{pmatrix} 21.6645 & 20.6166 \\ 20.6166 & 43.3442 \end{pmatrix}.$$

3.1 Reduced Orders of Fit

Although the order of \mathbf{G} in the previous example was six and polynomials of standardized ages to the fifth power were used to derive the covariance functions, perhaps only squared or cubed powers are needed to adequately describe the elements of \mathbf{G} . That is, find Φ^* such that it is rectangular and \mathbf{H}^* has a smaller order, $m < k$, but still

$$\mathbf{G} = \Phi^* \mathbf{H}^* \Phi^{*\prime}.$$

To determine \mathbf{H}^* , first pre-multiply \mathbf{G} by $\Phi^{*\prime}$ and post-multiply that by Φ^* as

$$\begin{aligned} \Phi^{*\prime} \mathbf{G} \Phi^* &= \Phi^{*\prime} (\Phi^* \mathbf{H}^* \Phi^{*\prime}) \Phi^* \\ &= (\Phi^{*\prime} \Phi^*) \mathbf{H}^* (\Phi^{*\prime} \Phi^*). \end{aligned}$$

Now pre- and post- multiply by the inverse of $(\Phi^{*\prime} \Phi^*) = \mathbf{P}$ to determine \mathbf{H}^* ,

$$\mathbf{H}^* = \mathbf{P}^{-1} \Phi^{*\prime} \mathbf{G} \Phi^* \mathbf{P}^{-1}.$$

To illustrate, let $m = 3$, then

$$\Phi^* = \begin{pmatrix} .7071 & -1.2247 & 1.5811 \\ .7071 & -.7348 & .0632 \\ .7071 & -.2449 & -.6957 \\ .7071 & .2449 & -.6957 \\ .7071 & .7348 & .0632 \\ .7071 & 1.2247 & 1.5811 \end{pmatrix},$$

and

$$\begin{aligned} \Phi^{*\prime} \Phi^* &= \begin{pmatrix} 3.0000 & 0.0000 & 1.3415 \\ 0.0000 & 4.1997 & 0.0000 \\ 1.3415 & 0.0000 & 5.9758 \end{pmatrix}, \\ (\Phi^{*\prime} \Phi^*)^{-1} &= \begin{pmatrix} .3705 & .0000 & -.0832 \\ .0000 & .2381 & .0000 \\ -.0832 & .0000 & .1860 \end{pmatrix}. \end{aligned}$$

Also,

$$\Phi^{*\prime} \mathbf{G} \Phi^* = \begin{pmatrix} 220.2958 & 78.0080 & 61.4449 \\ 78.0080 & 67.5670 & 44.9707 \\ 61.4449 & 44.9707 & 50.5819 \end{pmatrix}.$$

The matrix \mathbf{H}^* is then

$$\begin{pmatrix} 26.8082 & 5.9919 & -2.9122 \\ 5.9919 & 3.8309 & .4468 \\ -2.9122 & .4468 & 1.3730 \end{pmatrix}.$$

What order of reduced fit is sufficient to explain the variances and covariances in \mathbf{G} ? Kirkpatrick et al.(1990) suggested looking at the eigenvalues of the matrix \mathbf{H} from a full rank fit. Below are the values. The sum of all the eigenvalues was , and also shown is the percentage of that total.

| \mathbf{H} | |
|--------------|------------|
| Eigenvalue | Percentage |
| 29.0357 | .7831 |
| 4.2922 | .1158 |
| 1.8161 | .0490 |
| 1.3558 | .0366 |
| .5445 | .0147 |
| .0355 | .0010 |

The majority of change in elements in \mathbf{G} is explained by a constant, and by a linear increment. Both suggest that a quadratic function of the polynomials is probably sufficient. Is there a way to statistically test the reduced orders of fit to determine which is sufficient? A goodness of fit statistic is $\hat{\mathbf{e}}'\hat{\mathbf{e}}$ where

$$\hat{\mathbf{e}} = \mathbf{g} - \hat{\mathbf{g}}$$

and \mathbf{g} is a vector of the half-stored elements of the matrix \mathbf{G} , i.e.,

$$\mathbf{g}' = \left(g_{11} \quad g_{12} \quad \cdots \quad g_{16} \quad g_{22} \quad \cdots \quad g_{66} \right).$$

A half-stored matrix of order k has $k(k+1)/2$ elements. For $k = 6$ there are 21 values. Likewise, $\hat{\mathbf{g}}$ is a vector of half stored elements of the matrix $\Phi^*\mathbf{H}^*\Phi^{*'}.$ Although this matrix also has 21 values, because \mathbf{M} has only $m < k$ columns, the number of independent values is $m(m+1)/2$. For $m = 3$ this number is 6.

The test statistic, $\hat{\mathbf{e}}'\hat{\mathbf{e}}$, has a Chi-square distribution with $k(k+1)/2 - m(m+1)/2$ degrees of freedom. In the example with $m = 3$,

$$\Phi^*\mathbf{H}^*\Phi^{*'} = \begin{pmatrix} 3.9622 & 4.7467 & 5.2006 & 5.3239 & 5.1165 & 4.5786 \\ 4.7467 & 8.9493 & 11.4058 & 12.1162 & 11.0804 & 8.2986 \\ 5.2006 & 11.4058 & 15.2402 & 16.7038 & 15.7966 & 12.5186 \\ 5.3239 & 12.1162 & 16.7038 & 19.0868 & 19.2650 & 17.2386 \\ 5.1165 & 11.0804 & 15.7966 & 19.2650 & 21.4857 & 22.4586 \\ 4.5786 & 8.2986 & 12.5186 & 17.2386 & 22.4586 & 28.1786 \end{pmatrix},$$

and the residuals (differences from the original \mathbf{G}) are

$$\begin{pmatrix} -1.4622 & .1533 & -.6006 & -.7239 & -.8165 & -.5786 \\ .1533 & 4.5507 & .6942 & .1838 & .8196 & 2.4014 \\ -.6006 & .6942 & -.0402 & -2.2038 & -1.1966 & -.0186 \\ -.7239 & .1838 & -2.2038 & .9132 & -.2650 & -.3386 \\ -.8165 & .8196 & -1.1966 & -.2650 & 3.5143 & -2.1586 \\ -.5786 & 2.4014 & -.0186 & -.3386 & -2.1586 & 1.8214 \end{pmatrix},$$

so that the goodness of fit statistic is

$$\hat{\mathbf{e}}'\hat{\mathbf{e}} = 59.3476,$$

with $21-6=15$ degrees of freedom.

Is a fit of order 3 poorer than a fit of order 5? An F-statistic is possible by taking the difference in the goodness of fit statistics, divided by an estimate of the residual variance. The residual variance is estimated from a fit of order $k-1$ or in this case of order 5. The goodness of fit statistic for order 5 was 7.2139 with $21-15=6$ degrees of freedom. Hence the residual variance is

$$\sigma^2 = 7.2139/6 = 1.2023.$$

The F-statistic to test if a fit of order 3 is different from a fit of order 5 is

$$\begin{aligned} F &= \frac{(\hat{\mathbf{e}}'\hat{\mathbf{e}}_{m=3} - \hat{\mathbf{e}}'\hat{\mathbf{e}}_{m=5})/(15-6)}{\sigma^2} \\ &= \frac{(59.3476 - 7.2139)/9}{1.2023} \\ &= 5.7926/1.2023 = 4.8180, \end{aligned}$$

with (9,6) degrees of freedom. The table F-value at the ($P = .05$) level is 4.10. Thus, the difference is significant, and a fit of order 5 is better than a fit of order 3.

4 Basic Structure of RRM

Random regression models have a basic structure that is similar in most applications. A simplified RRM for a single trait can be written as

$$y_{ijkn:t} = F_i + g(t)_j + r(a, x, m1)_k + r(pe, x, m2)_k + e_{ijkn:t},$$

where

$y_{ijkn:t}$ is the n^{th} observation on the k^{th} animal at time t belonging to the i^{th} fixed factor and the j^{th} group;

F_i is a fixed effect that is independent of the time scale for the observations, such as a cage effect, a location effect or a herd-test date effect;

$g(t)_j$ is a function or functions that account for the phenotypic trajectory of the average observations across all animals belonging to the j^{th} group;

$r(a, x, m1)_k = \sum_{\ell=0}^{m1} a_{k\ell} x_{ijk:\ell}$ is the notation adopted for a random regression function. In this case, a denotes the additive genetic effects of the k^{th} animal, x is the vector of time covariates, and $m1$ is the order of the regression function. So that $x_{ijk:\ell}$ are the covariables related to time t , and $a_{k\ell}$ are the animal additive genetic regression coefficients to be estimated;

$r(pe, x, m2)_k = \sum_{\ell=0}^{m2} p_{k\ell} x_{ijk:\ell}$ is a similar random regression function for the permanent environmental (pe) effects of the k^{th} animal; and

$e_{ijkn:t}$ is a random residual effect with mean null and with possibly different variances for each t or functions of t .

The function, $g(t)_j$, can be either linear or nonlinear in t . Such a function is necessary in a RRM to account for the phenotypic relationship between y and the time covariables (or other types of covariables that could be used in a RRM). In a test day model, $g(t)_j$ accounts for different lactation curve shapes for groups of animals defined by years of birth, parity number, and age and season of calving within parities, for example. With growth data, $g(t)_j$ accounts for the growth curve of males or females of breed X or breed Y from young or old dams.

If the shape of the phenotypic relationship is not known or is nonlinear, then $g(t)_j$ could be a set of classification variables. Classification variables take up more degrees of freedom and require a large number of observations per level, but they do not force the user to explicitly define the shape of the trajectory. A mathematical function, on the other hand, does not use many degrees of freedom and gives a smooth trajectory over time regardless of the number of observations. The choice of classification variables or mathematical function is up to the researcher. If data are very numerous, and the mathematical function fits the data well, then

either approach will generally lead to the same results. The phenotypic relationships, $g(t)_j$, are important to a RRM analysis and deserve care and effort in their correct specification.

The random regressions are intended to model the deviations around the phenotypic trajectories. The pattern of variation may be very different in shape or appearance from the phenotypic relationships, and may be more simple than $g(t)_j$. Orthogonal polynomials of standardized units of time have been recommended as covariables (Kirkpatrick et al., 1990). Orthogonal polynomials have computational advantages. The primary general advantage is the reduced correlations among the estimated coefficients. A standardized unit of time, w , ranges from -1 to +1, and is derived as

$$w = \frac{2 * (t - t_{min})}{(t_{max} - t_{min})} - 1,$$

where t_{min} is the earliest date (or the youngest age) and t_{max} is the latest date (or oldest age) represented in the data. The order of the orthogonal polynomials would be m_1 and m_2 , i.e. the highest power of polynomial. Note that m_1 and m_2 do not need to be equal, but often (for simplicity of computing) they are chosen to be the same. Meyer(2000) and Pool et al. (2000), for example, compared many RRM models with different orders of orthogonal polynomials for the genetic and *pe* effects. Several types of orthogonal polynomials are available, but Legendre polynomials have been utilized (Kirkpatrick et al., 1990). The first 6 Legendre polynomial functions of standardized units of time are given in Table 1. Thus, if $w = -0.2$, then the covariables that would go into the model (for order equal to 5) are shown in the last column of Table 1. Covariables based upon orthogonal polynomials are small numbers that reduce problems with rounding errors, and they provide relatively small correlations between the estimated regression coefficients.

The residual variance should not be assumed to be constant from t_{min} to t_{max} . The residual effect is also known as a temporary environmental effect. Changes in residual variance might be predictable depending on the trajectory of the phenotypic data. For example, if RRM were being applied to growth data, weights may increase linearly with age, and the variance of weights may increase quadratically with age. Thus, the residual variance would be expected to increase in a similar manner as the phenotypic variance. Residual variances can be fit with a function of t , or assumed to have an autoregressive structure, or can be grouped into intervals having equal variance within the intervals. Research in this area is needed.

In matrix notation the RRM is

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}_1\mathbf{a} + \mathbf{Z}_2\mathbf{p} + \mathbf{e},$$

where \mathbf{b} contains F_i and $g(t)_j$ effects, \mathbf{a} contains $m_1 + 1$ additive genetic regression coefficients for each animal, \mathbf{p} contains $m_2 + 1$ permanent environmental regression coefficients for each animal with data, and \mathbf{e} contains the temporary environmental effects. Also,

$$Var \begin{pmatrix} \mathbf{a} \\ \mathbf{p} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{A} \otimes \mathbf{G} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \otimes \mathbf{P} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R} \end{pmatrix},$$

where \mathbf{G} is the variance-covariance matrix of the additive genetic random regression coefficients of order $m_1 + 1$; \mathbf{P} is the variance-covariance matrix of the permanent environmental random regression coefficients of order $m_2 + 1$; and \mathbf{R} is a diagonal matrix of temporary environmental variances which could vary depending on t , or \mathbf{R} could be block diagonal with an autocorrelation structure for each animal's records. The mixed model equations (MME) are represented as

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}_1 & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}_2 \\ \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{Z}_1 + \mathbf{A}^{-1} \otimes \mathbf{G}^{-1} & \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{Z}_2 \\ \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{Z}_1 & \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{Z}_2 + \mathbf{I} \otimes \mathbf{P}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{p}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix}.$$

Assumptions about the distributions of \mathbf{y} and other random variables are not necessary to derive best linear unbiased predictors (BLUP)(Goldberger, 1962; Henderson, 1984) or the MME, but when \mathbf{y} is normally distributed then BLUP is also BLP if the model is correct and variances and covariances are known. In order to estimate the elements of \mathbf{G} , \mathbf{P} , and \mathbf{R} via Bayesian methods or restricted maximum likelihood, then normality of the random variables must be assumed (See for example Jamrozik and Schaeffer, 1997). This paper will concentrate on the applications of RRM and not on the estimation of (co)variance parameters, nor on the computational details of estimating (co)variances or solving mixed model equations. Some of the applications in this paper have applied RRM to discrete data, and therefore, a BLUP analysis would not be optimum. However, the presentation of these ideas may stimulate others to find better solutions.

5 Example Data Analysis By RRM

Below are the data structure and pedigrees of four dairy cows. Given is the age at which they were observed for a trait during four visits to one herd.

| Cow | Sire | Dam | Age, Obs. at Visit | | | |
|-----|------|-----|--------------------|---------|---------|---------|
| | | | Visit 1 | Visit 2 | Visit 3 | Visit 4 |
| 1 | 7 | 5 | 22,224 | 34,236 | 47,239 | |
| 2 | 7 | 6 | 30,244 | 42,247 | 55,241 | 66,244 |
| 3 | 8 | 5 | 28,224 | 40,242 | | |
| 4 | 8 | 1 | | 20,220 | 33,234 | 44,228 |

The model equation might be

$$\begin{aligned} y_{jik:t} &= V_j + b_0 + b_1(A) + b_2(A)^2 \\ &\quad + (a_{i0}z_0 + a_{i1}z_1 + a_{i2}z_2) \\ &\quad + (p_{i0}z_0 + p_{i1}z_1 + p_{i2}z_2) + e_{jik:t} \end{aligned}$$

where

V_j is a random contemporary group effect which is assumed to follow a normal distribution with mean 0 and variance, $\sigma_c^2 = 4$.

b_0 , b_1 , and b_2 are fixed regression coefficients on $(A) = \text{age}$ and age squared which describes the general relationship between age and the observations,

a_{i0} , a_{i1} , and a_{i2} are random regression coefficients for animal i additive genetic effects, assumed to follow a multivariate normal distribution with mean vector null and variance-covariance matrix, \mathbf{G} ,

p_{i0} , p_{i1} , and p_{i2} are random regression coefficients for animal i permanent environmental effects, assumed to follow a multivariate normal distribution with mean vector null and variance-covariance matrix, \mathbf{P} ,

z_0 , z_1 , and z_2 are the Legendre polynomials based on standardized ages and derived as indicated earlier. The minimum age was set at 18 and the maximum age was set at 68 for calculating the Legendre polynomials.

and e_{jlk} is a temporary residual error term assumed to follow a normal distribution with mean 0 and variance, $\sigma_e^2 = 9$. In this example, the residual variance is assumed to be constant across ages.

The model in matrix notation is

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{W}\mathbf{v} + \mathbf{Z}\mathbf{a} + \mathbf{Z}\mathbf{p} + \mathbf{e},$$

where

$$\mathbf{X} = \begin{pmatrix} 1 & 22 & 484 \\ 1 & 30 & 900 \\ 1 & 28 & 784 \\ 1 & 34 & 1156 \\ 1 & 42 & 1764 \\ 1 & 40 & 1600 \\ 1 & 20 & 400 \\ 1 & 47 & 2209 \\ 1 & 55 & 3025 \\ 1 & 33 & 1089 \\ 1 & 66 & 4356 \\ 1 & 44 & 1936 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} 224 \\ 244 \\ 224 \\ 236 \\ 247 \\ 242 \\ 220 \\ 239 \\ 241 \\ 234 \\ 244 \\ 228 \end{pmatrix}, \quad \mathbf{W} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

and

$$\mathbf{Z} = \begin{pmatrix} .7071 & -1.0288 & .8829 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & .7071 & -.6369 & -.1493 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .7071 & -.7348 & .0632 & 0 & 0 & 0 \\ .7071 & -.4409 & -.4832 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & .7071 & -.0490 & -.7868 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .7071 & -.1470 & -.7564 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .7071 & -1.1268 & 1.2168 \\ .7071 & .1960 & -.7299 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & .7071 & .5879 & -.2441 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .7071 & -.4899 & -.4111 \\ 0 & 0 & 0 & .7071 & 1.1268 & 1.2168 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .7071 & .0490 & -.7868 \end{pmatrix}.$$

In order to reduce rounding errors the covariates of age for the fixed regressions can be forced to have a mean of approximately zero by subtracting 38 from all ages and 1642 from all ages squared. Then

$$\mathbf{X} = \begin{pmatrix} 1 & -16 & -1158 \\ 1 & -8 & -742 \\ 1 & -10 & -858 \\ 1 & -4 & -486 \\ 1 & 4 & 122 \\ 1 & 2 & -42 \\ 1 & -18 & -1242 \\ 1 & 9 & 567 \\ 1 & 17 & 1383 \\ 1 & -5 & -553 \\ 1 & 28 & 2714 \\ 1 & 6 & 294 \end{pmatrix}.$$

The mixed model equations that need to be constructed to provide estimated breeding values are as follows;

$$\begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{W} & \mathbf{X}'\mathbf{Z} & \mathbf{0} & \mathbf{X}'\mathbf{Z} \\ \mathbf{W}'\mathbf{X} & \mathbf{W}'\mathbf{W} + \mathbf{I}_4^9 & \mathbf{W}'\mathbf{Z} & \mathbf{0} & \mathbf{W}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{W} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{nn} \otimes \mathbf{G}^{-1} & \mathbf{A}^{nb} \otimes \mathbf{G}^{-1} & \mathbf{Z}'\mathbf{Z} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}^{bn} \otimes \mathbf{G}^{-1} & \mathbf{A}^{bb} \otimes \mathbf{G}^{-1} & \mathbf{0} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{W} & \mathbf{Z}'\mathbf{Z} & \mathbf{0} & \mathbf{Z}'\mathbf{Z} + \mathbf{I} \otimes \mathbf{P}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{c}} \\ \hat{\mathbf{a}}_n \\ \hat{\mathbf{a}}_b \\ \hat{\mathbf{p}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{W}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \\ \mathbf{0} \\ \mathbf{Z}'\mathbf{y} \end{pmatrix}.$$

The entire MME can not be presented, but parts of the MME are given below.

$$\mathbf{W}'\mathbf{W} = \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix},$$

$$\mathbf{W}'\mathbf{X} = \begin{pmatrix} 3 & -34 & -2758 \\ 4 & -16 & -1648 \\ 3 & 21 & 1397 \\ 2 & 34 & 3008 \end{pmatrix},$$

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} 12 & 5 & -1 \\ 5 & 1995 & 166,883 \\ -1 & 166,883 & 14,415,319 \end{pmatrix},$$

$\mathbf{Z}'\mathbf{Z}$ is composed of the following four blocks of order 3, for the four animals with records;

$$\text{Animal 1} \quad \begin{pmatrix} 1.5 & -.9006 & -.2335 \\ -.9006 & 1.2912 & -.8383 \\ -.2335 & -.8383 & 1.5457 \end{pmatrix},$$

$$\text{Animal 2} \quad \begin{pmatrix} 2 & .7275 & .0259 \\ .7275 & 2.0233 & 1.3612 \\ .0259 & 1.3612 & 2.1815 \end{pmatrix},$$

$$\text{Animal 3} \quad \begin{pmatrix} 1 & -.6235 & -.4902 \\ -.6235 & .5615 & .0648 \\ -.4902 & .0648 & .5761 \end{pmatrix},$$

$$\text{Animal 4} \quad \begin{pmatrix} 1.5 & -1.1085 & .0134 \\ -1.1085 & 1.5121 & -1.2082 \\ .0134 & -1.2082 & 2.2687 \end{pmatrix}.$$

and $\mathbf{Z}'\mathbf{X}$ is

$$\mathbf{Z}'\mathbf{X} = \begin{pmatrix} 2.1213 & -7.7781 & -761.5467 \\ -1.2737 & 19.9884 & 1516.7598 \\ -.3302 & -18.7627 & -1201.416 \\ 2.8284 & 28.9911 & 2458.5867 \\ 1.0288 & 46.4439 & 4337.8027 \\ .0366 & 27.9679 & 2979.5959 \\ 1.4142 & -5.6568 & -636.3900 \\ -.8818 & 7.0540 & 636.6324 \\ -.6932 & -2.1448 & -22.4568 \\ 2.1213 & -12.0207 & -1061.3570 \\ -1.5677 & 23.0259 & 1684.8063 \\ .0189 & -24.5677 & -1515.2470 \end{pmatrix}.$$

The right hand sides of the MME are

$$\mathbf{X}'\mathbf{y} = \begin{pmatrix} 2823 \\ 2070 \\ 68,064 \end{pmatrix},$$

$$\mathbf{W}'\mathbf{y} = \begin{pmatrix} 692 \\ 945 \\ 714 \\ 472 \end{pmatrix},$$

and

$$\mathbf{Z}'\mathbf{y} = \begin{pmatrix} 494.2629 \\ -287.6596 \\ -90.7117 \\ 690.1296 \\ 249.1165 \\ 7.3023 \\ 329.5086 \\ -200.1692 \\ -168.8920 \\ 482.2422 \\ -351.3606 \\ -7.8918 \end{pmatrix}.$$

The variance-covariance matrices of the additive and permanent environmental effects need to be known for BLUP. Normally, these are not well known and must be estimated simultaneously with the other effects of the model. Let

$$\mathbf{G} = \begin{pmatrix} 94.0000 & -3.8500 & .03098 \\ -3.8500 & 1.5000 & -.0144 \\ .03098 & -.0144 & .0014 \end{pmatrix},$$

and

$$\mathbf{P} = \begin{pmatrix} 63.0000 & -2.1263 & .0447 \\ -2.1263 & .5058 & -.00486 \\ .0447 & -.00486 & .0005 \end{pmatrix}.$$

The solutions to MME are

$$\hat{\mathbf{b}}' = \begin{pmatrix} 234.9797 & 1.4670 & -.01399 \end{pmatrix},$$

$$\hat{\mathbf{c}}' = \begin{pmatrix} -.8630 \\ 1.2885 \\ .1443 \\ -.5698 \end{pmatrix}.$$

Let the solutions for the animal additive genetic random regression coefficients be presented in tabular form as follows.

| Animal | a_0 | a_1 | a_2 |
|--------|-----------|-----------|----------|
| 1 | -2.021529 | .175532 | -.002696 |
| 2 | 5.751601 | -2.139115 | .025848 |
| 3 | -2.474456 | 2.554412 | -.029269 |
| 4 | -5.376687 | -.370873 | .002174 |
| 5 | -1.886714 | 1.464975 | -.016963 |
| 6 | 3.333268 | -1.065525 | .013047 |
| 7 | 1.503398 | -1.081654 | .012555 |
| 8 | -2.948511 | .681643 | -.008633 |

Similarly, the solutions for the animal permanent environmental random regression coefficients can be given in tabular form.

| Animal | p_0 | p_1 | p_2 |
|--------|-----------|----------|----------|
| 1 | -.296786 | .246946 | -.002521 |
| 2 | 3.968256 | -.730659 | .009430 |
| 3 | -.834765 | .925329 | -.008164 |
| 4 | -4.505439 | -.441805 | .001257 |

The problem is to rank the animals for selection purposes. If animals are ranked on the basis of a_0 , then animal 2 would be the highest (if that was desirable). If ranked on the basis of a_1 , then animal 3 would be the highest, and if ranked on the basis of a_2 , then animal 2 would be the highest. To properly rank the animals, an EBV at different ages could be calculated, and then these could be combined with appropriate economic weights. Calculate EBVs for 24, 36, and 48 mo of age, and use economic weights of 2, 1, and .5, respectively, for the three EBVs. A Total Economic Value can be calculated as

$$\text{TEV} = 2 * \text{EBV}(24) + 1 * \text{EBV}(36) + .5 * \text{EBV}(48).$$

The Legendre polynomials for ages 24, 36, and 48 mo are given in the rows of the following matrix \mathbf{L} ,

$$\mathbf{L} = \begin{pmatrix} .7071 & -.8328 & .3061 \\ .7071 & -.3429 & -.6046 \\ .7071 & .2449 & -.6957 \end{pmatrix}.$$

The results are shown in the following table.

| Animal | EBV(24) | EBV(36) | EBV(48) | TEV |
|--------|---------|---------|---------|--------|
| 1 | -1.58 | -1.49 | -1.38 | -5.33 |
| 2 | 5.86 | 4.78 | 3.53 | 18.26 |
| 3 | -3.89 | -2.61 | -1.10 | -10.93 |
| 4 | -3.49 | -3.68 | -3.89 | -12.61 |
| 5 | -2.56 | -1.83 | -.96 | -7.43 |
| 6 | 3.25 | 2.71 | 2.09 | 10.25 |
| 7 | 1.97 | 1.43 | .79 | 5.76 |
| 8 | -2.66 | -2.31 | -1.91 | -8.58 |

The animal with the highest TEV was animal 2. All animals ranked rather similarly at each age on their EBVs. Rankings of animals could change with age. Thus, the pattern of growth could be changed one that is desirable.

Estimation of the residual variance is

$$\hat{\sigma}_e^2 = (\mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} - \hat{\mathbf{c}}'\mathbf{W}'\mathbf{y} - \hat{\mathbf{a}}'_n\mathbf{M}'\mathbf{y} - \hat{\mathbf{p}}'\mathbf{M}'\mathbf{y}) / (N - r(\mathbf{X})),$$

where

$$\begin{aligned} \mathbf{y}'\mathbf{y} &= 665,035, \\ \hat{\beta}'\mathbf{W}'\mathbf{y} &= 664902.89, \\ N - r(\mathbf{X}) &= 12 - 3 = 9, \\ \hat{\sigma}_e^2 &= 14.6788. \end{aligned}$$

6 EXERCISES

1. A biologist studied chipmunks in southern Ontario. He planted small TV cameras inside four nests of chipmunks. The territories of the occupants of the four nests did not overlap. With the cameras he had students monitor the nests for a day at several times during the summer and fall. Students counted the number of nuts that were collected and stored in the nest on a given day during a set 8 hour period. Below are the observations on the four nests for various days of the year.

Number of nuts collected in 8 hours.

| Day of Year | Nest | Number of Nuts | Day of Year | Nest | Nuts |
|-------------|------|----------------|-------------|------|------|
| 123 | 1 | 25 | 161 | 1 | 11 |
| 124 | 2 | 37 | 155 | 2 | 14 |
| 127 | 3 | 16 | 153 | 3 | 9 |
| 129 | 4 | 42 | 151 | 4 | 12 |
| 192 | 1 | 13 | 225 | 1 | 38 |
| 194 | 2 | 15 | 227 | 2 | 29 |
| 198 | 3 | 10 | 233 | 3 | 37 |
| 200 | 4 | 16 | 246 | 4 | 44 |

- (a) Write a complete random regression model for these data, including the assumptions and limitations.
- (b) Compute a simple 4 by 4 phenotypic covariance matrix for the four nest sites. Then use covariance functions and tests to determine the appropriate order for a random regression model, and to obtain initial covariance matrices for the model.
- (c) Apply the random regression model to the data, and plot the differences between the nests.

7 Application to Test Day Records

Several reviews on the use of RRM for the analysis of test day records of dairy cattle have been given by Swalve (1995, 1998, 2000) and Jensen (2001). A detailed summary of the Canadian application of a test day model (TDM) was given by Schaeffer et al. (2000). Research still continues on determining the order of fit for the random regressions for additive genetic and permanent environmental effects that explains the most variation in the observations. Pool et al. (2000) concluded that the additive genetic effects require a lower order of Legendre polynomials than the pe effects. The use of lactational shape functions for the covariables of time (i.e. days in milk) have largely been abandoned because the estimated covariance matrices, \mathbf{G} and \mathbf{P} , usually have very high correlations between the coefficients which can hinder the estimation process. Orthogonal polynomials of time have much lower correlations among the coefficients and provide estimates of the covariance matrices that tend to be more robust over different data sets. There could be differences in computing properties between different types of orthogonal polynomials (e.g. Legendre versus Chebyshev polynomials).

RRM have been used with multiple trait models to analyze test day records. Each lactation is treated as a separate trait, and each of milk, fat, and protein yields, and somatic cell scores are separate traits within each lactation. Feed intake, body condition scores, and heart girth also have been measured on test day and analyzed by multiple trait RRM (Liu, 1998; Jones et al., 1999; Gallo et al., 2001; Veerkamp et al., 2001b). Only the first three lactations have been included in the Canadian test day model due to computational limitations, but research on different models that might be able to handle more lactations is ongoing. Jensen (2001), for example, suggested a multidimensional RRM for test day records which has been studied by Guo et al. (2002). The approach has a set of random regressions for parity number and a set of random regressions for days in milk. Thus, persistency is assumed to be the same across parities. To circumvent this problem the covariables for parity number should be multiplied by those for days in milk. If the order of the polynomials for parity number was 4 (i.e. 5 covariables) and the order of the polynomials for days in milk was 4, then there would be a total of 25 covariables. Multiply this by 4 traits and there would be 100 coefficients per animal to estimate, and \mathbf{G} would be of order 100. Lifetime production ability could be computed from these coefficients. The computing issues might be resolved by considering two traits at a time (e.g. milk and fat, milk and protein, and milk and SCS) rather than all four at once. Reduced rank models may be possible to reduce the number of coefficients to be estimated per animal.

An issue with all RRM analyses is to choose the appropriate range of time or ages for study. In the Canadian TDM, for example, only TD records between 5 to 305 days in milk are included, but longer lactations of 330, 345, or 400 days have been chosen in some countries. Given that cows are producing milk longer, the need to restrict test day records between 5 to 305-d is unnecessary. All countries should allow test day records to go up to 400-d, but EBVs may be expressed for any length of lactation that is suitable for that country. Changing the endpoint of the time interval results in changes to the (co)variance matrices, \mathbf{G} and \mathbf{P} , too. Thus, if every country agreed to a 400-d upper limit then comparisons of (co)variance matrices would become easier. Estimation of (co)variance matrices in RRM is sufficiently time consuming that care should be taken in choosing the time or age range for a study.

Jamrozik et al.(2002) applied a multiple country RRM to first lactation test day records from Canada, Italy, New Zealand and Australia. The management systems for Canada and Italy were similar, but different from the pasture based systems in New Zealand and Australia. Hence the fixed lactation curves differed in the same manner. A RRM was used in order to deal with differences in lactation lengths between countries. New Zealand tended to have shorter lactations than the other countries. The joint country multiple trait model was used as a basis for the international comparison of dairy bulls. Extension of procedures to deal with over 20 countries would not be computationally possible unless those countries could be grouped into five or fewer groups. However, the main time limiting factor would be the huge number of cows and test day records to be processed. The study showed that an international RRM was possible, but may not be practical for Holsteins.

8 Applications Similar to Test Day Models

Besides dairy cattle, test day models have also been applied to dairy goats and dairy sheep. The following applications utilize the same basic structure of RRM as has been applied to test day yields of dairy cattle. Differences would be in the definition of subclasses for the fixed curves, the intervals of time, and the order of the random regressions that might be used. A difficulty with goats and sheep is that the offspring are often allowed to suckle during the first week or more, so that test day records in early lactation are not readily available. This period could be critical for fitting the phenotypic trajectories.

Egg production in chickens also follows a lactation-like curve, and therefore, RRM can be applied to weekly or monthly egg production (Anang et al.,2000). The exact same issues would arise as with test day production. Similarly, annual wool production of sheep, litter size of sows per parity, and annual antler yield of deer or elk could also be analyzed by RRM similar to that of test day models.

Another application could be to semen fertility of dairy bulls where the observations can be the volume of semen, the percentage motility of the sperm, or the percentage of dead sperm, per ejaculate. Semen samples of dairy bulls are collected once or twice per week after the bulls have been returned to service. There would be many observations for bulls that were older than 6 years. Age at collection, handler, season of the year, interval of time between collections, and diet could be factors included in a model. RRM could be applied to male semen fertility in essentially any species of livestock.

In nutritional studies, blood samples of animals are often collected every 5 to 15 minutes to measure various substances in the blood in order to determine absorption of nutrients or the rate of release of enzymes. Hormones may have random spikes depending on when they are released in the blood and RRM may not deal with these appropriately. Due to costs not very many animals are measured in this way. RRM could be used to model these changes, but usually there are not enough animals collected to do any kind of genetic study. The emphasis would be more on the dietary effects and the fixed "curves". The animal random regressions would be

nuisance effects and would be a combination of genetic and *pe* effects.

9 Application to Growth Traits

The pattern of animal growth over time can be modeled by random regressions. In livestock species, growth is generally an economically important trait. Associated with growth are feed intake, feed efficiency, fat deposition, muscle development, bone length, degree of maturity, and body condition. Growth is slightly different from test day milk yields because body weights are cumulative over time. This would be analogous to accumulating daily milk yields of cows through the lactation rather than having individual test day yields on given days in the lactation. Accumulated weights have part-whole correlations from one weighing to the next, but will likely continue to be measured and analyzed as such.

One of the first applications of RRM to growth in pigs was made by Andersen and Pedersen (1996). In their study, pigs were weighed twice weekly from 30 kg live weight to 115 kg live weight. Machines monitored individual feed intake even though animals were in pens of twelve individuals. Thus, pigs started the test at different ages and consequently were weighed at different days on test. Weight and weight gains were modeled as a function of time, but were also modeled as a function of feed intake from which a measure of feed efficiency was derived. That is, the genetic merit for growth was a function of the amount of feed intake. Usually growth and feed intake are highly correlated both phenotypically and genetically, so that the genetic variation in growth remaining after accounting for feed intake would be reduced. The fixed curves of the model, $g(t)_j$, were a fourth order polynomial of days on test (not orthogonal polynomials), while the order of random regressions was 2. Growth rate was fairly linear between 30 to 115 kg, but did decrease between 30 and 50 days on test, and further decreased between 50 and 80 days on test, for both gilts and castrated males. Rather than model weights against feed intake, a multiple trait RRM model having both weight and feed intake traits against time on test would be a better way to examine feed efficiency without reducing the genetic variation in weight. A multiple trait RRM would simultaneously account for the changes in genetic and residual variation in each trait while allowing both traits and the relationship between those traits to vary together with time. The general concept would be not to model one trait against another if they are genetically correlated.

Maternal genetic effects of growth traits are known to be important in beef cattle. Albuquerque and Meyer(2001) studied growth in Nelore cattle from birth to 630 days of age. The general RRM structure was augmented to include random regressions for maternal genetic effects and maternal permanent environmental effects. Let $r(ma, x, m3)$ and $r(mp, x, m4)$ denote the random regressions on maternal genetic of order $m3$ and maternal *pe* effects of order $m4$, respectively. However, Albuquerque and Meyer (2001) assumed zero correlations between direct and maternal genetic effects at all time points in order to simplify computations. Different orders of fit for the random regressions were applied to three different data sets. Using their notation, one of the favoured models was $k = 6\ 6\ 6\ 4$ which refers to the order (plus one) of the Legendre polynomials for direct genetic, maternal genetic, animal PE, and maternal PE effects,

respectively, i.e. $k = (m_1 + 1) (m_3 + 1) (m_2 + 1) (m_4 + 1)$. Such a model has 77 (co)variances in addition to the residual variances to be estimated. Another favoured model based on Bayesian Information Content (BIC) was $k = 4 \ 4 \ 6 \ 3$ with 51 parameters to be estimated. With either model, maternal genetic variance increased from birth to around 115 d of age and decreased thereafter, while direct genetic variance increased throughout from birth to 630 d of age and was generally much larger than the maternal genetic variance. Residual variances were small and increased only slightly with age. The effect of zero correlations between direct and maternal genetic effects was not examined, but perhaps may not be too important in these particular data.

Besides Andersen and Pedersen (1996), RRM have been applied to growth traits by Schnyder et al. (2001), Meyer (1999, 2000), Magnabosco et al. (2000), Schenkel et al. (2002), McKay et al. (2002), Veerkamp and Thompson (1999), and Uribe et al. (2000). The key issues in application of RRM to growth traits are the number of times individuals need to be measured, at what times in their lives, and what will be the upper age range. The costs of collecting these measurements would also play a role in determining how often and when to measure growth. A RRM provides some freedom in this regard, and animals are generally weighed at all ages from birth to maturity. Some animals could have many weights recorded while other animals may have only a few. Fixed growth curves should be estimated for each sex, within years of birth, within breeds or breed crosses, and within different parities of dams. Much work remains in applications to growth.

Besides animals, RRM could be applied to growth of plants, such as crops (which grow quickly) or trees (which grow slowly). RRM could be used to model growth of bacterial populations grown under certain conditions. Similar to growth would be a decay function such as the degradation of nutrients in the gut as they were digested in various parts of the gastrointestinal tract. Application of RRM to growth traits is in itself a growing area of research.

10 Application to Genotype by Environment Interaction Studies

A genotype by environment interaction implies that animals would rank differently in different environments. If those environments can be ordered such as by a temperature-humidity index (THI), or by elevation, or by average herd production, then a RRM could be applied with regressions on these variables rather than on time or age. RRM are useful when the environments cover a range of different conditions rather than just a few categories. Ravagnolo and Misztal (2000) used the THI to obtain different heritabilities and different rankings of animals at different heat stress levels. Test day yields for two particular days in milk (DIM) classes were analyzed by the following model:

$$y_{ijklmn} = HTD_i + A_j + D_k + F_l$$

$$+r(a, f(i), 1)_m + r(pe, f(i), 1)_m + e_{ijklmn},$$

where HTD_i is a herd-test date effect, A_j is an age at calving class, D_k is a DIM class effect, F_l is a milking frequency class, a and pe are the random animal additive genetic and permanent environmental effects, respectively, $f(i)$ is a heat stress value (ranging from 70 to 93 and averaging 78), and $r(a, f(i), 1)_m$ and $r(pe, f(i), 1)_m$ are random regressions for animal additive genetic and permanent environmental linear effects, respectively. This model differs from the general structure of RRM in that there were no fixed curves, $g(t)_j$ in the model for $f(i)$. However, because a HTD is associated with only one THI value the equations for $g(t)_j$ would be completely confounded with the HTD equations. The $f(i)$ values were not standardized and converted to orthogonal polynomials. Because only the linear component was considered, there was no need to use orthogonal polynomials. If higher order polynomials were included, then orthogonal polynomials should be considered. A common residual variance was assumed for all records. Only first parity cows were included. Heritability increased as THI increased, and correlations between predicted breeding values (i.e. $\hat{a}_{m,1}$ with $\hat{a}_{m,1} + \hat{a}_{m,2} \times f(i)$) decreased as THI increased. Lawlor et al. (2002) used a RRM to model the change in genetic variances over time. The model implies that genetic values of animals also change with time and that re-ranking among cows calving in the same year might occur with time (if they could be re-observed), but cows are only observed in one year and do not have repeated records. Years 1981 to 1996 were standardized to values between -1 to 1. Production traits, productive life, and four linear type scores were analyzed jointly in a multiple trait model. Only first order Legendre polynomials were used with random regressions for herds, herd by sire interactions, and animal additive genetic effects. The model did not include any fixed curves for changes in traits over time. The authors related changes in variances and correlations to linkage disequilibrium and selection practices. Because there were no fixed curves considered in their model, changes in variances could be associated with changes in phenotypic means rather than due to linkage disequilibrium.

Strandberg et al. (2000) used a RRM to study genotype by environment interactions between populations of Nordic Red and White cattle. The independent variable was the herd-year average as a deviation from the overall year averages across the three countries (i.e. Finland, Norway, and Sweden). The model used was

$$y_{ij} = g(X_{ij}, 1) + r(s, X_{ij}, 1)_i + e_{ij},$$

where y_{ij} was either 305-d protein yield, days open, or interval from calving to last insemination; X_{ij} was a herd-year average deviation; $r(s, X_{ij}, 1)$ are sire random regressions of order 1; $g(X_{ij}, 1)$ is the fixed regression function on X_{ij} ; and e_{ij} was a random residual effect assumed to have a common variance across countries. Nearly 1.5 million records were analyzed. Heritabilities of all traits increased as herd-year deviations increased. Considerable re-ranking of sires occurred in herd-year averages that were 1.5 SD or more units below the average, while re-rankings were not as severely affected by herd-year averages that were more than 1.5 SD units above the mean environment. By having a regression on a measure of the herd environment rather than a fixed herd effect, then the problem of lack of genetic connections between countries is circumvented, but perhaps other problems arise. The random regressions should also have considered higher order polynomials even if the phenotypic relationship was expected to be linear.

While there have been attempts to study genotype by environment interactions with RRM, improvements should be made to the models that have been used. Studies should include fixed curves to account for the phenotypic relationships, and random regressions on higher order orthogonal polynomials need to be explored. Because animals in these studies have only one observation (in one particular time interval or one environment), the hope is that a RRM can pull out a correlation or trend through the connections created by sires. A simulation study should be conducted to determine if a RRM can do this adequately.

Table 1
The first 6 Legendre polynomial functions of
standardized units of time, w .

| Order | | for $w = -0.2$ |
|-------|--|----------------|
| 0 | $0.7071 w^0$ | 0.7071 |
| 1 | $1.2247 w^1$ | -0.2449 |
| 2 | $-0.7906 w^0 + 2.3717 w^2$ | -0.6957 |
| 3 | $-2.8062 w^1 + 4.6771 w^3$ | 0.5238 |
| 4 | $0.7955 w^0 - 7.9550 w^2 + 9.2808 w^4$ | 0.4921 |
| 5 | $4.3973 w^1 - 20.5206 w^3 + 18.4685 w^5$ | -0.7212 |

11 References

- Albuquerque, L. G., Meyer, K. 2001. Estimates of covariance functions for growth from birth to 630 days of age in Nelore cattle. *J. Anim. Sci.* 79:2776-2789.
- Anang, A., Mielenz, N., Schuler, L. 2000. Genetic and phenotypic parameters for monthly egg production in White Leghorn hens. *J. Anim. Breed. Genet.* 117:407-415.
- Andersen, S., Pedersen, B. 1996. Growth and food intake curves for group-housed gilts and castrated male pigs. *Animal Science* 63:457-464.
- Ducrocq, V. 1987. An analysis of productive life in dairy cattle. Ph.D. Diss., Cornell University, Ithaca, NY.
- Ducrocq, V. 1994. Statistical analysis of length of productive life for dairy cows in the Normande breed. *J. Dairy Sci.* 77:855-865.
- Ducrocq, V., Solkner, J. 1998. "The Survival Kit V3.0", a package for large analyses of survival data. Pages 447-450 in Proc. 6th World Congress on Genetics Applied to Livestock Production, Armidale, Vol. 27.
- Gallo, L., Carnier, P., Cassandro, M., Dal Zotto, R., Bittante, G. 2001. Test day genetic analysis of condition score and heart girth in Holstein Friesian cows. *J. Dairy Sci.* 84:2321-2326.
- Guo, Z., Lund, M. S., Madsen, P., Korsgaard, I., Jensen, J. 2002. Genetic parameter estimation for milk yield over multiple parities and various lengths of lactation in Danish Jerseys by random regression models. *J. Dairy Sci.* 85:1596-1606.

- Goldberger, A. S. 1962. Best linear unbiased prediction in the generalized linear regression model. *J. Am. Stat. Assoc.* 57:369-377.
- Henderson, Jr., C. R. 1982. Analysis of covariance in the mixed model: Higher level, nonhomogeneous, and random regressions. *Biometrics* 38:623-640.
- Henderson, C. R. 1984. *Applications of Linear Models in Animal Breeding*. University of Guelph.
- Hill, W. G., Brotherstone, S. 1999. Advances in methodology for utilizing sequential records. *British Soc. of Animal Science*. Publication No. 24: 55-61.
- Jamrozik, J., Schaeffer, L. R. 1997. Estimates of genetic parameters for a test day model with random regressions for production of first lactation Holsteins. *J. Dairy Sci.* 80:762-770.
- Jamrozik, J., Schaeffer, L. R., Weigel, K. A. 2002. Genetic evaluation of bulls and cows with single- and multiple-country test-day models. *J. Dairy Sci.* 85:1617-1622.
- Jensen, J. 2001. Genetic evaluation of dairy cattle using test day models. *J. Dairy Sci.* 84:2803-2812.
- Jones, H. E., White, I. M. S., Brotherstone, S. 1999. Genetic evaluation of Holstein Friesian sires for daughter condition score changes using a random regression model. *Animal Science* 68:467-475.
- Kirkpatrick, M., Lofsvold, D., Bulmer, M. 1990. Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics* 124:979-993.
- Laird, N. M., Ware, J. H. 1982. Random effects models for longitudinal data. *Biometrics* 38:963-974.
- Lawlor, T. J., Tsuruta, S., Klei, L., Misztal, I. 2002. Use of a random regression model to investigate changes in genetic parameters over time. 7th World Congress of Genetics Applied to Livestock Production, paper 252.
- Liu, X. 1998. The estimation of genetic parameters of test day dry matter intake, energy intake and milk yield of Holstein cows. M.Sc. Thesis, University of Guelph, Guelph, Ontario, Canada.
- Magnabosco, C. de U., Lobo, R. B., Famula, T. R. 2000. Bayesian inference for genetic parameter estimation on growth traits for Nelore cattle in Brazil, using the Gibbs sampler. *J. Anim. Breed. Genet.* 117:169-188.
- McKay, L. R., Schaeffer, L. R., McMillan, I. 2002. Analysis of growth curves in rainbow trout using random regression. 7th World Congress of Genetics Applied to Livestock Production, paper 241.
- Meyer, K. 1999. Estimates of genetic and phenotypic covariance functions for postweaning growth and mature weight of beef cows. *J. Anim. Breed. Genet.* 116:181-205.
- Meyer, K. 2000. Random regressions to model phenotypic variation in monthly weights of Australian beef cows. *Livest. Prod. Sci.* 65:19-38.

- Meyer, K., Hill, W. G. 1997. Estimation of genetic and phenotypic covariance functions for longitudinal or 'repeated' records by restricted maximum likelihood. *Livest. Prod. Sci.* 47:185-200.
- Pool, M. H., Janss, L. L. G., Meuwissen, T. H. E. 2000. Genetic parameters of Legendre polynomials for first parity lactation curves. *J. Dairy Sci.* 83:2640-2649.
- Ravagnolo, O., Misztal, I. 2000. Genetic component of heat stress in dairy cattle, parameter estimation. *J. Dairy Sci.* 83:2126-2130.
- Schaeffer, L. R., Dekkers, J. C. M. 1994. Random regressions in animal models for test-day production in dairy cattle. *Proc. 5th World Congress of Genetics Applied to Livestock Production.* Guelph, Ontario, Canada XVIII:443-446.
- Schaeffer, L. R., Jamrozik, J., Kistemaker, G. J., Van Doormaal, B. J. 2000. Experience with a test-day model. *J. Dairy Sci.* 83:1135-1144.
- Schenkel, F. S., Miller, S. P., Jamrozik, J., Wilton, J. W. 2002. Two step and random regression analyses of weight gain of station tested beef bulls. *J. Anim. Sci.* 80:1497-1507.
- Schnyder, U., Hofer, A., Kunzi, N. 2001. Impact of variation in length of individual testing periods on estimation of covariance components of a random regression model for feed intake of growing pigs. *J. Anim. Breed. Genet.* 118:235-246.
- Strandberg, E., Kolmodin, R., Madsen, P., Jensen, J., Jorjani, H. 2000. Genotype by environment interaction in Nordic dairy cattle studied by use of reaction norms. *Proc. Interbull Meeting, Bled, Slovenia.* Bulletin 25:41-45.
- Swalve, H. H. 1995. Test day models in the analysis of dairy production data: A review. *Arch. Tierz.* 38:591-612.
- Swalve, H. H. 1998. Use of test day records for genetic evaluation. *Proc. 6th World Congress of Genetics Applied to Livestock Production.* Armidale, NSW, Australia. XXIII:295-302.
- Swalve, H. H. 2000. Theoretical basis and computational methods for different test-day genetic evaluation methods. *J. Dairy Sci.* 83:1115-1124.
- Uribe, H., Schaeffer, L. R., Jamrozik, J., Lawlor, T. J. 2000. Genetic evaluation of dairy cattle for conformation traits using random regression models. *J. Anim. Breed. Genet.* 117:247-259.
- Veerkamp, R. F., Thompson, R. 1999. A covariance function for feed intake, live weight, and milk yield estimated using a random regression model. *J. Dairy Sci.* 82:1565-1573.
- Veerkamp, R. F., Brotherstone, S., Meuwissen, T. H. E. 1999. Survival analysis using random regression models. *Proc. International Workshop on EU Concerted Action Genetic Improvement of Functional Traits in Cattle; Longevity.* Interbull Bulletin 21:36-40.
- Veerkamp, R. F., Koenen, E. P. C., DeJong, G. 2001. Genetic correlations among body condition score, yield, and fertility in first parity cows estimated by random regression models. *J. Dairy Sci.* 84:2327-2335.

Wilmink, J. B. M., 1987. Adjustment of test day milk, fat, and protein yield for age, season and stage of lactation. *Livest. Prod. Sci.* 16, 335-348.

12 Application to Survival Analyses

Survival is a trait that could be observed every day of an animal's life, but only the point at which an animal dies or is culled is usually recorded in the data. Survival analyses are typically undertaken using a non-linear proportional hazard model (Ducrocq 1987, 1994; Ducrocq and Solkner, 1998). However, Veerkamp et al. (1999) proposed that a RRM could be used for survival analyses. At any given point in time an animal is either alive, (1), or dead, (0). The survival 'curve' for an individual animal is a straight line at the value 1 from birth until the animal is culled (i.e. removed from the population via death, accident or management decision), after which a value 0 is assigned and continues as a straight line to the end of the observation period. An average of the survival 'curves' of many animals gives a population survival curve of the expected survival rate at various ages. At the beginning of the observation period, usually birth or first calving, nearly all animals are still alive, so that the phenotypic variance is very close to zero. The phenotypic variance becomes larger with time until 50% of the population remains, after which the variance decreases towards zero again at the end of the observation period when nearly all animals have been culled.

Suppose the observation period for dairy cattle begins at first calving and ends 75 months later. Thus, every cow that has been culled has 75 monthly observations for survival. For example, if a cow was culled in month 41, then the observation vector for this cow would have the first 40 elements being 1, and the next 35 being 0. If a cow has not had an opportunity to be culled and was currently milking in month 40, then this cow would have an observation vector of length 40, with all elements equal to 1.

A RRM for survival might be as follows:

$$y_{ijklno:t} = (YS : Ht)_{ij} + (YSAP : t)_{ikl} \\ + r(a, x, m1)_n + r(pe, x, m2)_n + e_{ijklno:t}$$

where YS refers to the i^{th} year-season of first calving; H refers to the j^{th} herd of first calving; A refers to the k^{th} age at first calving groups; P refers to the l^{th} production level as a deviation from herd average in first lactation; x are the appropriate orthogonal polynomials of time, t , after first calving; a are the random regression coefficients for additive genetic value of animal n ; pe are the random regression coefficients for permanent environmental effects of animal n ; and $e_{ijklno:t}$ is a residual effect. The appropriate order of the orthogonal polynomials has yet to be determined.

The herd-time effect within year-season, $(YS : Ht)_{ij}$, specifies a group of cows first calving in the same year-season, i , now located within herd j in time period t . This contemporary group would conceptually face the same management decision factors at that point in time. Cows could change herds and would be part of a different group of contemporaries born in the same year-season, but now in a different herd and time period. This effect also implies that herd management decisions are not consistent over years or seasons within years, and the change in

size of the herd would also be considered with this effect. Rather than having t effects within each ($YS : H$), these effects could be modeled as random regressions with $mh + 1$ coefficients per ($YS : H$). These effects would be similar to herd-test date effects in a test day model for production traits.

The ($YSAP : t$) $_{ikl}$ effects are the fixed survival curves for different ages at first calving and production levels. Consequently, the additive genetic random regression coefficients for animals would represent survival ability (involuntary culling) after accounting for production. The fixed curves are nested within year-season of first calving to allow for changes in population decision strategies over time. This would account for the general declining population sizes within a country.

This model allows the inclusion of animals that have not yet been culled (missing observations), and provides genetic evaluations of cows for survival, which were not initially available from the Survival Kit (Ducrocq and Solkner 1998). Genetic evaluations (EBV) could be expressed as the percentage survival rate at a specific number of months after first calving (e.g. the one that gives the highest genetic variance). Alternatively, EBV could be reported as the number of months after first calving that results in 50% culling of a bull's progeny (i.e. half-life EBV). Research on RRM models for survival are just beginning.

The reviewers correctly pointed out that the observations on survival are binary (discrete) rather than continuous, and therefore, a BLUP analysis of these data will not be optimum. Perhaps the Survival Kit could be altered to include random regressions to cover the time dependent variables. Otherwise a generalized linear mixed model with its additional computational complexity could be applied.

Application to Fertility in Dairy Cattle

Reproductive performance in dairy cattle is economically important, but typically has a low heritability. Reproductive performance includes the age at first estrus, the age at first service, the number of services to conception, age at conception, age at calving, the ease of calving, if the calf was stillborn, or if the cow aborted. These traits are observed before every parturition. Thus, there are observations on the same animal over a period of years. Genetic merit for reproductive performance could change over time as well. Let parity number be the unit of time being observed. Most cows have only 3-4 parities, but some cows can have as many as 10 or more. Let the data files be limited to a maximum of ten parities and standardize the parity numbers from -1 (for first parity) to +1 (for tenth parity). A RRM for many of the fertility traits, say age at first estrus/ days to first estrus after calving, for example, might be as follows:

$$\begin{aligned}
 y_{ijkmno:t} &= (YSP : t)_{ij} + (YH : t)_{ik} \\
 &\quad + r(a, parity, m1)_m + r(pe, parity, m2)_m \\
 &\quad + r(ma, parity, m3)_n + r(mp, parity, m4)_n + e_{ijkmno:t}.
 \end{aligned}$$

where $y_{ijkmno:t}$ is an observation on age at first estrus/ days to first estrus after calving; Y refers

to year of birth/calving; S refers to season of birth/calving; H refers to herd of birth/calving; P refers to production level as a deviation from herd average for that lactation; the random regressions are as described for earlier models for additive genetic, maternal genetic, animal pe , and maternal pe effects; and $e_{ijkmno:t}$ is a residual effect.

A multiple trait RRM should be applied to all reproductive traits simultaneously. For some traits the sex of calf and service sire effects might be added to the model. Adjustment for production levels may be useful. Conformation traits like feet and legs could affect the ability of a cow to become pregnant; rump traits could affect calving ease and fertility; and body condition scores have been shown to have a relationship to fertility. Development of an appropriate model for each trait may require substantial work.

Estimation of the (co)variance matrices for the random regressions could be problematic because data would not be available on all seven parities for all animals. Data should be required to have first parity information for all cows. Analysis of cows with data on all seven parities would firstly give a relatively small number of cows, and secondly would be a group of cows that have had exceptional reproductive performance (otherwise they would have been culled). In order to obtain unbiased estimates, data on all cows will have to be used. Covariances between direct and maternal genetic effects may need to be assumed to be zero, as was made by Albuquerque and Meyer (2001) in their growth studies.

The results of such an analysis would be a different genetic value for each animal for each parity. An issue would be how to combine the genetic values of each parity. First lactation information should probably be weighted more heavily than that from later lactations. Breeding values could be expressed relative to an optimal reproductive performance perspective. All of the traits analyzed could be combined into an economic fertility index.

Reproductive traits like number of inseminations, calving ease, and stillbirths are not continuous traits, and so BLUP analyses would not be optimum. Studies comparing these approximations to more suitable types of analyses are needed. If the approximations are computationally more feasible and lead to selection of the same animals as the theoretically correct methodology, then an optimal solution may be unnecessary.

The time variable in the above RRM was parity number which ranged from 1 to 10, so that there are only 10 possible values of the x variable. This is not a problem for a RRM, conceptually or computationally. The time variable in a test day model is days in milk which go from 5 to 305 days, or 301 possible values. Time is usually recorded in discrete units (days, weeks, months, years, etc.) even though time is considered to be continuous. Parity number could be substituted by age at calving (in months), but this variable may be highly related to one of the reproductive traits. Parity number was chosen for this model because reproductive performance of cows can change from one parity to the next, and the performance in previous parities could influence the reproductive performance in the current parity, and a RRM is a good way to study this variability.