BEST PREDICTION OF LACTATION YIELD AND PERSISTENCY

P.M. VanRaden
Animal Improvement Programs Laboratory, Agricultural Research Service,
US Department of Agriculture, Beltsville, MD 20705-2350, USA

SUMMARY
Selection index procedures can condense observations of daily yield into estimates of lactation yield and persistency. Condensed data instead of test day data in an animal model might provide many of the same benefits with less computation. The best prediction method uses the phenotypic covariance matrix among daily yields to compute reliabilities and covariances of estimated yield and persistency for any data pattern. The measure of persistency is uncorrelated with yield if persistency coefficients are deviated from a central point of the lactation. Recent records containing both yield and persistency data and earlier records containing just yield data could be combined easily in genetic evaluations.
Keywords: persistency, test day data, lactation yield

INTRODUCTION
Genetic evaluations based on lactation records have ignored differences in the shape of the lactation curve. Test day models account for lactation shape and offer more precision but use many more equations. A two-trait model with just yield and persistency could account for lactation shape with far fewer equations. Best prediction is a method to condense information from many test days into optimal lactation measures of yield and persistency. Computational shortcuts for reporting and analyzing test day data are provided.

DERIVATION
Daily yield can be modeled as management group mean or expected value plus deviation from that mean. The vector $\boldsymbol{\mu}$ contains the expected values at each day of lactation for one trait. Vector $\mathbf{t}$ represents all 305 test day deviations for the trait, and vector $\mathbf{t}_m$ represents only measured deviations. Vector $\mathbf{t}_m$ may be a subset of vector $\mathbf{t}$ (for example, if milk was measured several times before day 305) but would not be a subset if milk was measured after day 305 or if protein and fat were included in a multitrait analysis. Means and variances are both assumed known with $\text{Var}(\mathbf{t}) = \mathbf{V}$ and $\text{Var}(\mathbf{t}_m) = \mathbf{V}_m$. Covariance of $\mathbf{t}$ and $\mathbf{t}_m$ also is assumed known and equal to $\mathbf{C}$.

Lactation yield. A cow's true 305-day yield (y) is the sum of the expected values for each day ($\mathbf{1}'\boldsymbol{\mu}$) plus the sum of her 305 deviations from expectation ($\mathbf{1}'\mathbf{t}$), where $\mathbf{1}'$ is a vector of length 305 and elements of 1.0. The cow's true lactation yield and the best prediction of her lactation yield ($\hat{y}$) are

$$y = E(y) + \mathbf{1}'\mathbf{t} \text{ and}$$
$$\hat{y} = E(y) + \mathbf{1}'\mathbf{C}\mathbf{V}_m^{-1}\mathbf{t}_m.$$

Reliability of $\hat{y}$ is obtained from variances of y and $\hat{y}$, which can be computed from the simple quadratic forms

$$\mathrm{Var}(y) = \mathbf{1'V1},$$
$$\mathrm{Var}(\hat{y}) = \mathbf{1'CV}_m^{-1}\mathbf{C'1}, \text{ and}$$
$$\mathrm{Rel}(\hat{y}) = \mathbf{1'CV}_m^{-1}\mathbf{C'1/1'V1}.$$

Procedures to adjust $\mathbf{V}_m$ if data are from partial days or are computer averages from several days or are recorded by an owner instead of a supervisor were outlined by VanRaden (1997).

Persistency. Cows with high persistency tend to milk more than expected at lactation end and less than expected at the beginning. To measure persistency, test day deviations may be multiplied by the linear function of days in milk (DIM) proposed by Wiggans and Goddard (1996) or perhaps by a nonlinear function. Let **d** represent a vector with elements d_i that represent DIM (or a function of DIM) for that yield. A measure of persistency uncorrelated with yield can be obtained by defining coefficients $q_i = d_i - d_0$, where d_0 is a constant that acts as a balance point between yields in early and later lactation; in vector form, $\mathbf{q'} = \mathbf{d'} - \mathbf{1'}d_0$. With any choice of d_0, covariance between measures of persistency and yield is

$$\mathrm{Cov}(\mathbf{q't},\mathbf{1't}) = \mathbf{q'V1} = (\mathbf{d'} - \mathbf{1'}d_0)\mathbf{V1} = \mathbf{d'V1} - \mathbf{1'V1}d_0.$$

The particular choice of d_0 that makes yield and persistency phenotypically uncorrelated is obtained by setting $\mathrm{Cov}(\mathbf{q't},\mathbf{1't})$ to 0.0 and then solving for d_0:

$$d_0 = \mathbf{d'V1/1'V1}.$$

Solutions for d_0 were 128, 146, and 135 DIM for milk, fat, and protein, respectively, based on 650 records of daily yield. True persistency (p), predicted persistency ($\hat{p}$), and expected value of persistency are given by

$$p = E(p) + \mathbf{q't},$$
$$\hat{p} = E(p) + \mathbf{q'CV}_m^{-1}\mathbf{t}_m, \text{ and}$$
$$E(p) = \mathbf{q'}\boldsymbol{\mu} = (\mathbf{d'} - \mathbf{1'}d_0)\boldsymbol{\mu} = \mathbf{d'}\boldsymbol{\mu} - d_0E(y).$$

Variances of p and $\hat{p}$ have forms similar to those for lactation yield, and reliability again is the ratio of predicted to true variance:

$$\mathrm{Var}(\hat{p}) = \mathbf{q'CV}_m^{-1}\mathbf{C'q},$$
$$\mathrm{Var}(p) = \mathrm{Var}(\mathbf{q't}) = \mathrm{Var}[(\mathbf{d'} - \mathbf{1'}d_0)\mathbf{t}] = \mathbf{d'Vd} - 2\mathbf{d'V1}d_0 + \mathbf{1'V1}d_0^2 = \mathbf{q'Vq}, \text{ and}$$
$$\mathrm{Rel}(\hat{p}) = \mathrm{Var}(\hat{p})/\mathrm{Var}(p) = \mathbf{q'CV}_m^{-1}\mathbf{C'q/q'Vq}.$$

A standardized estimate of persistency ($\hat{s}$) may be more appealing and can be obtained by subtracting population mean for persistency (μ_p) and dividing by within-herd phenotypic standard deviation:

$$\hat{s} = (\hat{p} - \mu_p)/\sqrt{\mathrm{Var}(p)}.$$

Variance of persistency then is 1.0. Alternatively, genetic standard deviations could be used to standardize persistency.

Expected values. Lactation curves may differ according to age, parity, breed, time, herd, and their interactions. In the past, adjustment factors were used to standardize lactation records. Let vector $\boldsymbol{\mu}$ contain the mature-equivalent or standard lactation curve, and let vector $\boldsymbol{\beta}$ contain the 305 expected values for the age, parity, season, year, and herd of interest. If $\boldsymbol{\mu}$ and $\boldsymbol{\beta}$ differ, adjustment factors can be used to standardize 305-day yield and persistency. Multiplicative factors are favored for yield because variances become more nearly equal. Multiplicative factors should not be used for persistency because division by zero could occur and because differences in variance can be removed by creating a unitless trait. For yield, the additive adjustment is $\mathbf{1}'(\boldsymbol{\mu} - \boldsymbol{\beta})$, and the multiplicative adjustment is $\mathbf{1}'\boldsymbol{\mu}/\mathbf{1}'\boldsymbol{\beta}$. For persistency, the additive adjustment is $\mathbf{q}'(\boldsymbol{\mu} - \boldsymbol{\beta})$ or equivalently $(\mathbf{d}' - \mathbf{1}'d_0)(\boldsymbol{\mu} - \boldsymbol{\beta})$. This approach is simple, but curve shape is not fully preserved. The lactaion curve is scaled vertically by the yield adjustment and rotated by the persistency factor. For any group of interest, the assumed curve is $[\boldsymbol{\mu} - \mathbf{q}'(\boldsymbol{\mu} - \boldsymbol{\beta})](\mathbf{1}'\boldsymbol{\beta}/\mathbf{1}'\boldsymbol{\mu})$, which is the standard curve minus the persistency factor and then divided by the yield factor. New factors for persistency are needed, but current factors for yield can be kept.

Expansion. Predicted yields and persistencies are easier to model if they are first expanded to conform to assumptions of the statistical model. While keeping means constant, division of each predicted value by its reliability provides an expanded yield ($\bar{y}$) and expanded persistency ($\bar{p}$) that contain the corresponding true value plus an independent error. Formulas of VanRaden *et al.* (1991) were adapted to provide

$$\bar{y} = E(y) + [\hat{y} - E(y)]/\mathrm{Rel}(\hat{y}) \text{ and}$$
$$\bar{p} = E(p) + [\hat{p} - E(p)]/\mathrm{Rel}(\hat{p}).$$

These expanded variables have greater variance than the true values, whereas predictions have less variance. Variances of $\bar{y}$ and $\bar{p}$ are

$$\mathrm{Var}(\bar{y}) = \mathbf{1}'\mathbf{C}\mathbf{V}_m^{-1}\mathbf{C}'\mathbf{1}/[\mathrm{Rel}(\hat{y})]^2 = \mathbf{1}'\mathbf{V}\mathbf{1}/\mathrm{Rel}(\hat{y}) \text{ and}$$
$$\mathrm{Var}(\bar{p}) = \mathbf{q}'\mathbf{C}\mathbf{V}_m^{-1}\mathbf{C}'\mathbf{q}/[\mathrm{Rel}(\hat{p})]^2 = \mathbf{q}'\mathbf{V}\mathbf{q}/\mathrm{Rel}(\hat{p}).$$

Although Cov(y,p) = 0, best predictions or expanded estimates of y and p may covary if tests do not represent the entire lactation, such as with records in progress. Covariance of $\bar{y}$ and $\bar{p}$ is

$$\mathrm{Cov}(\bar{y},\bar{p}) = \mathbf{1}'\mathbf{C}\mathbf{V}_m^{-1}\mathbf{C}'\mathbf{q}/[\mathrm{Rel}(\hat{y})\mathrm{Rel}(\hat{p})].$$

Expanded yield and persistency records contain the normal environmental variance present in a true record plus an additional measurement error independent of the true record (VanRaden *et al.* 1991). Total error variances for yield $[\mathrm{Var}(\bar{y} - u_y)]$ and persistency $[\mathrm{Var}(\bar{p} - u_p)]$ then are

$$\mathrm{Var}(\bar{y} - u_y) = \mathrm{Var}(\bar{y}) + \mathrm{Var}(u_y) - 2\mathrm{Cov}(\bar{y},u_y) \text{ and}$$
$$\mathrm{Var}(\bar{p} - u_p) = \mathrm{Var}(\bar{p}) + \mathrm{Var}(u_p) - 2\mathrm{Cov}(\bar{p},u_p),$$

where u_y and u_p are the sum of random effects other than error contained in the models for yield and persistency, respectively. For example, u_y and u_p might contain genetic effects plus permanent environmental effects. If measurement errors $\bar{y} - u_y$ and $\bar{p} - u_p$ are uncorrelated with u_y and u_p, respectively, the covariance terms above are variances of random effects u_y and u_p. Variances and covariances then reduce to

$$\mathrm{Var}(\bar{y} - u_y) = \mathrm{Var}(\bar{y}) - \mathrm{Var}(u_y),$$
$$\mathrm{Var}(\bar{p} - u_p) = \mathrm{Var}(\bar{p}) - \mathrm{Var}(u_p), \text{ and}$$
$$\mathrm{Cov}(\bar{y} - u_y, \bar{p} - u_p) = \mathrm{Cov}(\bar{y},\bar{p}) - \mathrm{Cov}(u_y,u_p).$$

Multitrait prediction. Analysis of just two traits should be simpler than a complete test day model. Yields and persistencies of milk, fat, and protein can be estimated separately or jointly (Schaeffer and Jamrozik 1996). Multitrait best prediction was presented for yields (VanRaden 1997) but not for persistencies. Multitrait predictions increase accuracy of phenotypic selection, but expanded single-trait records may provide a better source of data for mixed model equations. Instead of combining correlated information from all traits into each lactation measure, traits are kept separate and their correlations included later in the lactation model. For example, milk yield can be used to predict protein yield even if no samples are taken. This predicted protein yield is not needed in the mixed model equations because the covariance matrix transfers information from milk to protein at that time.

APPLICATIONS

Many farmers will continue to base their breeding, marketing, and culling decisions on lactation yields instead of individual test days. Data from any test plan can be condensed quickly into optimal measures of lactation yield and persistency. Reliabilities of lactation yield and persistency can be used to label records and may be a first step toward computing reliability of breeding value. Expanded yields and persistencies may allow faster estimation of some variance components and adjustment factors without reprocessing of daily data. Finally, extremely large populations might be evaluated more easily if condensed data were substituted for test day data.

REFERENCES

Schaeffer, L.R. and Jamrozik, J. (1996) *J. Dairy Sci.* **79**:2044-2055.

VanRaden, P.M. (1997) *J. Dairy Sci.* **80**:(accepted).

VanRaden, P.M., Wiggans, G.R. and Ernst, C.A. (1991) *J. Dairy Sci.* **74**:4344-4349.

Wiggans, G.R. and Goddard, M.E. (1996) *Proc. New Zealand Soc. Anim. Prod.* **56**:19-21.