Semiparametric animal models via penalized splines as alternatives to models with contemporary groups[1]

R. J. C. Cantet*†[2], A. N. Birchmeier*, A. W. Canaza Cayo‡, and C. Fioretti§

*Departamento de Producción Animal, Universidad de Buenos Aires, C1417DSE Buenos Aires, Argentina;
†Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina;
‡Universidad del Altiplano, Puno, Perú; and §Estancias y cabaña, Las Lilas, S.A., Argentina

ABSTRACT: Contemporary groups (CG) are used in genetic evaluation to account for systematic environmental effects of management, nutritional level, or any other differentially expressed group effect; however, because the functional form of the distribution of those effects is unknown, CG serve as an approximation to a time-varying mean. Conversely, in semiparametric models, there is no need to assume any functional form for the time-varying effects. In this research, we present a semiparametric animal model (AMS) using the covariate day of birth (DOB) by means of penalized splines (P-splines), as an alternative to fitting CG. In the AMS, the functionality of the data on DOB is expressed by means of a Basic segmented polynomial line (*B*-spline) basis, and proper covariance matrices are used to reflect the serial correlation among the points of support (or knots) at different times. Three different covariance matrices that reflect either short- or long-range dependences among knots are discussed. Different models were fitted to birth weight data from Polled Hereford calves. Models compared were an animal model with CG, an animal model with CG and the covariate DOB nested within CG (CG + DOB), and P-splines with the first difference penalty matrix and three different AMS with 20, 40, 60, 80, or 120 knots. Models were compared using a modified Akaike information criterion ($\boldsymbol{AIC_C}$), which was calculated as a byproduct of the estimation of variance components by REML using the expectation maximization algorithm. All three AMS had smaller (better) values of $\boldsymbol{AIC_C}$ than the regular model with CG, while producing almost the same ranking of predicted breeding values and similar average predicted error variance. In all AMS, the inference and all measures of comparison were similar when the number of knots was equal ≥40. The model CG + DOB had analogous performance to the AMS, but at the expense of using more parameters. It is concluded that the use of penalized regression splines using a *B*-spline basis with proper covariance matrices is a competitive method to the fitting of CG into animal models for genetic evaluation, without having to assume any functional form for the covariate DOB.

Key Words: Contemporary Groups, Penalized Splines, Semiparametric Models

J. Anim. Sci. 2005. 83:2482–2494

Introduction

In genetic evaluation, contemporary group (**CG**) is a discrete explanatory variable that accounts for a high percentage of the total phenotypic variation in most traits. An important issue is the definition of CG by setting cut-off dates to form new CG (Crump et al., 1997; Carabaño et al., 2004); bias may be introduced for records close to a cut-off date between two adjacent CG (Sivajarasingam, 1993). The reason is that the functional form of the relationship between the evaluated trait and the effects of herd environment in time is unknown. Thus, CG is an approximation to a time-varying mean in an attempt to describe the unknown distribution of those effects.

Semiparametric models are employed when the functional form of a covariate is unknown. In this situation, the underlying smooth function is usually a nuisance parameter, and the interest lies in accounting for the effects of the regressor variable (Altman, 2000). Thus, an alternative to CG is to fit a time covariate (e.g., day of birth [**DOB**]) without imposing any particular functional form on its effect on the trait, as suggested by Cantet (2002). Under this approach, one or more

[1] Funding for this research was provided by grants from Secretaría de Ciencia y Técnica, UBA (UBACyT G018—2004–2007) and from Agencia Nacional de Ciencia y Tecnología (FONCyT PICT 08-9502) of Argentina.

[2] Correspondence: Agronomía UBA, Av. San Martin 4453 (phone: 54-011-4514-8000, ext. 8784; fax: 54-011-4514-8000, ext. 8785; e-mail: rcantet@agro.uba.ar).

Received April 29, 2005.
Accepted July 12, 2005.

functions are fitted to the data that account for the irregular behavior of DOB. To fit such a function, Eilers and Marx (1996) proposed penalized splines (**P-splines**), a methodology that is closely connected to mixed models (Ruppert et al., 2003; Wand, 2003). Cantet (2002) suggested that regression splines may be used to substitute CG in genetic evaluation models. His presentation was expository and attempted to use truncated rather than Basic segmented polynomial line (***B*-spline**) basis functions without details on how to perform the numerical computations. Thus, the goals of the present research were 1) to describe an animal model with DOB using P-splines and proper covariance matrices (**AMS**), and 2) to fit the AMS to birth weight records of beef cattle and to compare them with the fit obtained by the regular model with CG for genetic evaluation.

Materials and Methods

P-Splines, B-Splines, and Semiparametric Animal Models

The process we attempt to model is a particular case of functional data in which the time (t) varying covariate DOB follows an underlying continuous time process $\boldsymbol{f}(t)$. This process is described by some random trajectory or trend within a herd, which results from environmental effects: management, grass availability, or weather characteristics such as temperature or rainfall, or a combination of all effects. The idea is to model this trajectory with a smooth function that is fitted using P-splines (Eilers and Marx, 1996). In contrast to the prediction of random regression models for functional breeding values (White et al., 1999; Huisman et al., 2002; Druet et al., 2003; Iwaisaki et al., 2005), in which the unit with repeated observations in time is the individual animal, in the approach used here, the herd is the unit with repeated observations over time. Other time-varying covariates such as age of dam may be modeled using the approach presented here, although a low-order polynomial can be used to describe such data in a more parsimonious fashion. With suitable modifications, the fit of trajectories in two or more dimensions (for example, in space), can be dealt with by using a two-dimensional B-splines basis, but we do not pursue the topic any further.

The P-splines are a combination of regression on a spline basis with a discrete roughness penalty to smooth data series or scatterplots (Eilers and Marx, 1996). The penalty controls the degree of smoothness while fitting the function. For P-splines, Eilers and Marx (1996) used equally spaced B-splines basis functions (De Boor, 1993). When fitted to time-dependent data, a covariate expressed on a B-spline basis results in the union of k-degree polynomial segments that have k − 1 continuous derivatives at the joining points, or knots. The resulting fit is smooth, with better numerical properties than a polynomial fit of high degree (e.g., k > 5; Green and Silverman, 1994). The B-splines of degree k have local support, which means that they are defined only in a small part of the real line between k + 1 knots, being equal to 0 outside this small segment of the real line. Compared with least squares, this prevents any observation that may affect the entire shape of the function. Eilers and Marx (1996) observed that, in P-splines, the number of knots is not a critical parameter as long as "there are enough of them." This lack of sensitivity to the number of knots is due to the control exercised by the penalization. Thus, problems of multicollinearity can be fixed by changing the penalty parameter (P. Eilers, Leiden Univ., The Netherlands and B. D. Marx, Louisiana State Univ.; personal communication). Moreover, Ruppert and Carroll (2000) found that increasing the number of knots to >40 resulted in slight differences in fit after extensive simulation studies. Using a number of knots in the order of 40 has the effect of dramatically decreasing the dimensionality of the parameter space compared with other smoothing spline methods. Notwithstanding this, increases in the number of records may require enlarging the parameter space. We dealt with this issue by comparing models with different numbers of knots.

Covariate DOB Expressed Using a B-Spline Basis

Consider a random vector $\boldsymbol{y}$ ($n \times 1$) of records, which is functionally dependent on time (t) through a general vector valued function $\boldsymbol{f}(t)$. We will consider t to be DOB in Julian days. It is desired that $\boldsymbol{f}(t)$ be a smooth function of t that can be fitted using mixed model theory. To do so, let

$$\boldsymbol{f}(t) = \sum_{i=1}^{nx} B_i^{(k)}\, b_i. \qquad [1]$$

In [1], $\boldsymbol{f}(t)$ is a linear combination of parameters (b_i) that are to be estimated, and elements of B-spline basis functions $B_i^{(k)}$ $i = 1, 2, ..., nx$ (De Boor, 1993). The number of basis functions $B_i^{(k)}$ needed to express DOB is $nx + 6$ (Eilers and Marx, 1996); three additional knots are added to each extreme. Usually the degree of a B-spline is ≤3 ($k \le 3$). In the present research, cubic splines (k = 3) were used because they are sufficiently flexible to retain most features of the data.

Calculations of the $B_i^{(3)}$ coefficients were performed with the recursive algorithm of De Boor (1993). A brief explanation follows. Let $t_1, t_2, ..., t_{nx+6}$ be the set of knots that expand the range of DOB. Then, $B_i^{(0)} = 1$ if DOB is in the interval between t_i and t_{i+1}; otherwise, $B_i^{(0)}$ = 0. For example, suppose t_{10} = 3303.5 and t_{11}= 3390.8, then for DOB = 3353, $B_{10}^{(0)} = 1$, whereas $B_1^{(0)}$, ..., $B_9^{(0)}$, $B_{11}^{(0)}$, ..., and $B_{nx+6}^{(0)}$ are all 0. The next step in the algorithm is to calculate the $B_i^{(1)}$s, then the $B_i^{(2)}$s and finally the $B_i^{(3)}$s, using the scheme displayed in Figure 1, which is called the blossom. Four $B_i^{(3)}$ elements ($B_{i-3}^{(3)}$, $B_{i-2}^{(3)}$,

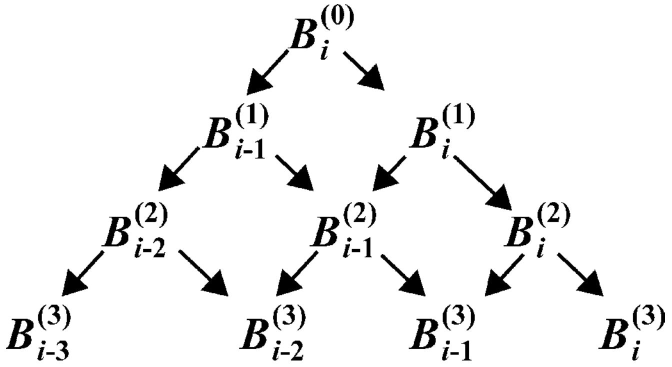

Figure 1. Order of calculation of the Basic segmented polynomial line coefficients.

$B_{i-1}^{(3)}$, and $B_i^{(3)}$) are needed to express one value of DOB in terms of a cubic *B*-splines basis. These four values add up to 1, and the recursion formula (De Boor, 1993) to calculate them is equal to

$$B_i^{(1)} = \begin{cases} 1, \text{ if DOB} \in [t_i, t_{i+1}) \\ 0, \text{ otherwise} \end{cases} \quad B_i^{(m)} = \frac{\text{DOB} - t_i}{t_{i+m-1} - t_i} B_i^{(m-1)} + \frac{t_{i+m} - \text{DOB}}{t_{i+m} - t_{i+1}} B_{i+1}^{(m-1)} \; m = 2, 3. \quad [2]$$

For the sake of completeness, truncated bases are an alternative to *B*-splines bases. Using truncated bases, the function in [1] is expressed as

$$f(t_{ij}) = b_0 + \sum_{l=1}^{m} b_l^l t_{ijl}^l + \sum_{l=1}^{m} \sum_{k=1}^{nx} u_{kl} \, (t_{ij} - \kappa_k)_+^l,$$

where t_{ij} is the DOB measure for herd *i* at time *j*, *m* is the order of fit (usually $m = 1$, 2, or 3), and b_0, b_j, and the u_{kl} are the parameters to estimate. The values of the knots are κ_k ($k = 1, 2 \ldots, nx$), and the quantity $(t_{ij} - \kappa_k)_+$ is taken to be equal to $t_{ij} - \kappa_k$ whenever $t_{ij} > \kappa_k$, or 0 otherwise. If $l = 1$, the fit is linear; it is quadratic for $l = 2$ and cubic for $l = 3$.

Expression [1] can be written in matrix form as $\boldsymbol{Bb}$, where $\boldsymbol{B}$ is the $n \times nx$ matrix that contains the $B_i^{(3)}$, and $\boldsymbol{b}$ is the parametric vector ($nx \times 1$) containing the b_i for $\boldsymbol{f}(t)$. Each row of $\boldsymbol{B}$ has all elements equal to zero except for basis coefficients $B_{i-3}^{(3)}$, $B_{i-2}^{(3)}$, $B_{i-1}^{(3)}$, and $B_i^{(3)}$ in columns $i - 3$, $i - 2$, $i - 1$, and *i*, respectively. Thus, each value of the covariable DOB is transformed into four *B*-spline coefficients in the interval (0, 1) for each animal with a record in $\boldsymbol{y}$.

P-Splines

To obtain the estimators of $\boldsymbol{b}$ in the model

$$\boldsymbol{y} = \boldsymbol{B}\,\boldsymbol{b} + \boldsymbol{e}, \quad [3]$$

Eilers and Marx (1996) proposed maximizing the likelihood while penalizing the sum of squares of the differences between adjacent b_i, either for first $\sum_{i=1}^{nx} (b_i - b_{i+1})^2$ or for second squared differences $\sum_{i=1}^{nx} (b_i - 2b_{i+1} + b_{i+2})^2$. The resulting function is then proportional to

$$(\boldsymbol{y} - \boldsymbol{B}\,\boldsymbol{b})'(\boldsymbol{y} - \boldsymbol{B}\,\boldsymbol{b}) + \lambda\, \boldsymbol{b}'\, \boldsymbol{D}'\, \boldsymbol{D}\, \boldsymbol{b}. \quad [4]$$

The scalar λ controls the amount of smoothing. For first differences, matrices $\boldsymbol{D}$ ($nx - 1 \times nx$) and $\boldsymbol{D'D}$ are respectively equal to

$$\boldsymbol{D} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ . & . & . & . & . \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix} \quad [5]$$

$$\boldsymbol{D'D} = \begin{bmatrix} 1 & -1 & 0 & . & 0 & 0 \\ -1 & 2 & -1 & . & 0 & 0 \\ 0 & -1 & 2 & . & 0 & 0 \\ 0 & 0 & -1 & . & -1 & 0 \\ 0 & 0 & 0 & . & 2 & -1 \\ 0 & 0 & 0 & . & -1 & 1 \end{bmatrix}.$$

Notice that $\boldsymbol{D}'\,\boldsymbol{D}$ ($nx \times nx$) is singular. If second differences are considered, then matrices $\boldsymbol{D}$ and $\boldsymbol{D'D}$ are

$$\boldsymbol{D} = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ . & . & . & . & . \\ 0 & 0 & 1 & -2 & 1 \end{bmatrix} \quad [6]$$

$$\boldsymbol{D'D} = \begin{bmatrix} 1 & -2 & 1 & . & 0 & 0 \\ -2 & 5 & -4 & . & 0 & 0 \\ 1 & -4 & 6 & . & 1 & 0 \\ 0 & 1 & -4 & . & -4 & 1 \\ 0 & 0 & 1 & . & 5 & -2 \\ 0 & 0 & 0 & . & -2 & 1 \end{bmatrix}.$$

In [6], $\boldsymbol{D}$ is of order ($nx - 2 \times nx$), and $\boldsymbol{D'D}$ is again singular. Eilers and Marx (1996) obtained the penalized estimator of $\boldsymbol{b}$ as the solution of the following system of equations:

$$(\boldsymbol{B'B} + \lambda\, \boldsymbol{D'D})\hat{\boldsymbol{b}} = \boldsymbol{B'y}. \quad [7]$$

The effect of the penalty is to shrink $\boldsymbol{b}$ in an amount proportional to λ. The connection between P-splines and mixed models (e.g., Ruppert et al., 2003; Wand, 2003) is now apparent. In a mixed-model setting, $\boldsymbol{b}$ can be viewed as random effects, λ to the ratio of the error variance to the variance of the $\boldsymbol{b}$s (Ruppert et al., 2003), whereas $\boldsymbol{D'D}$ may be interpreted as some g-inverse of the variance-covariance matrix of the linear spline parameters. From a Bayesian viewpoint, a singular $\boldsymbol{D'D}$ induces the prior distribution of the linear spline parameters to be improper, which, in turn, causes the posterior distribution to be improper (Hobert and Casella, 1996; Lang and Brezger, 2004). A way around this problem is to formulate an equivalent mixed model (Henderson, 1984) such as the one discussed by Currie and Durban (2002); however, the fitting of such a model is computationally involved, as it makes the mixed model equations extremely dense and does not behave in a numerically stable manner. Alternatively, proper covariance matrices of the bs at the knots provide a better fit to the model than singular penalty matrices as shown subsequently.

Animal Models with a Functional Covariate Using a B-Spline Fit

To specify an animal model suitable for genetic evaluation, we now incorporate fixed effects (e.g., age of dam and sex) and breeding values to the specification of DOB in [1] and [3]. The resulting animal model is equal to

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Bb} + \boldsymbol{Za} + \boldsymbol{e}. \qquad [8]$$

The incidence matrix $\boldsymbol{X}$ ($n \times p$) associates data to the $p \times 1$ parametric vector β of fixed effects; $\boldsymbol{Z}$ ($n \times q$) relates elements of $\boldsymbol{y}$ to the random vector $\boldsymbol{a}$ ($q \times 1$) of breeding values, whereas $\boldsymbol{e}$ is the $n \times 1$ vector of random errors. We assume a full-rank parametrization in β so that rank $[\boldsymbol{X}] = p$. The Gaussian vectors $\boldsymbol{a}$ and $\boldsymbol{e}$ are stochastically independent; both have zero expectations and covariance matrices equal to $\boldsymbol{A}\sigma_A^2$ and $\boldsymbol{I}\sigma_e^2$, respectively. Matrix $\boldsymbol{A}$ contains the additive relationships. The random variables that relate to the knots are assumed to be distributed as $\boldsymbol{b} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{G}_S\sigma_b^2)$, where σ_b^2 is the variance component between knots and $\boldsymbol{G}_S$ is a positive definite matrix that portrays the covariance structure in time for the knots. Three alternative specifications for $\boldsymbol{G}_S$ are described in the next section. The expectations and the variances and covariances of all random vectors in [8] are equal to

$$\mathrm{E}\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{b}\\ \boldsymbol{a}\\ \boldsymbol{e}\end{bmatrix} = \begin{bmatrix}\boldsymbol{X}\beta\\ \boldsymbol{0}\\ \boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix} \quad \mathrm{Var}\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{b}\\ \boldsymbol{a}\\ \boldsymbol{e}\end{bmatrix} = \qquad [9]$$

$$\begin{bmatrix}\boldsymbol{V} & \boldsymbol{BG}_S\sigma_b^2 & \boldsymbol{ZA}\sigma_A^2 & \boldsymbol{I}\sigma_e^2\\ \boldsymbol{G}_S\boldsymbol{B}'\sigma_b^2 & \boldsymbol{G}_S\sigma_b^2 & \boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{AZ}'\sigma_A^2 & \boldsymbol{0} & \boldsymbol{A}\sigma_A^2 & \boldsymbol{0}\\ \boldsymbol{I}\sigma_e^2 & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}\sigma_e^2\end{bmatrix},$$

where $\boldsymbol{V} = \boldsymbol{ZAZ}'\sigma_A^2 + \boldsymbol{BG}_S\boldsymbol{B}'\sigma_b^2 + \boldsymbol{I}\sigma_e^2$. The variance components are the additive genetic variance σ_A^2, the variance between knots σ_b^2, and the error variance σ_e^2. Mixed-model equations for [8] are:

$$\begin{bmatrix}\boldsymbol{X'X} & \boldsymbol{X'B} & \boldsymbol{X'Z}\\ \boldsymbol{B'X} & \boldsymbol{B'B} + \boldsymbol{G}_S^{-1}\lambda & \boldsymbol{B'Z}\\ \boldsymbol{Z'X} & \boldsymbol{Z'B} & \boldsymbol{Z'Z} + \boldsymbol{A}^{-1}\alpha\end{bmatrix}\begin{bmatrix}\hat{\beta}\\ \hat{\boldsymbol{b}}\\ \hat{\boldsymbol{a}}\end{bmatrix} = \begin{bmatrix}\boldsymbol{X'y}\\ \boldsymbol{B'y}\\ \boldsymbol{Z'y}\end{bmatrix}, \qquad [10]$$

where $\lambda = \sigma_e^2/\sigma_b^2$ and $\alpha = \sigma_e^2/\sigma_A^2$.

Covariance Matrices of the Elements in $\boldsymbol{b}$

There are several considerations in the search for a suitable covariance matrix $\boldsymbol{G}_S$. First, it should be able to account for "serial correlation" (Diggle et al., 1994) in such a way that pairs of knots that are closer in time are likely to be more strongly correlated than pairs farther distant in time. A second consideration is that correlation structures that, after inversion, are similar to the banded matrices of differences [5] and [6] would behave similarly to the original formulation of P-splines. It also was desirable to choose a linear covariance structure (i.e., $\boldsymbol{G}_S = \boldsymbol{P}\sigma_b^2$) from a regular multivariate normal density rather than from a "curved" normal density (Lehmann, 1983). In this latter case, the covariance structure depends on dispersion parameters in a nonlinear fashion, such as in an autoregressive process. Linear dispersion structures allow the use of REML-expectation maximization or Gibbs sampling, avoiding the need for more involved algorithms such as Metropolis-Hastings. Finally, it was considered to be important that inversion of $\boldsymbol{G}_S$ not be computationally expensive. As a result, we used the covariance structure ($\boldsymbol{P}$) that accounts for a linear decay of the correlation with time originally proposed by Cantet (2002). Off-diagonal elements of $\boldsymbol{P}$ ($\boldsymbol{P}_{ij}$) are functions of time lag $t_j - t_i$, between the knots i and j. Then, for any pair of knots at times t and $t + w$, $\mathrm{cov}(b_t, b_{t+w}) = \left(1 - \dfrac{w}{nx}\right)\sigma_b^2$. This covariance structure is stationary, as it depends on time lag w but not on the time moments t or $t + w$. If the time measures (expressed in days, weeks, or months) are $t_1 < t_2 < ... < t_{nx-1} < t_{nx}$, diagonal elements of $\boldsymbol{P}$ are equal to 1 and off-diagonals are $\left(1 - \dfrac{t_j - t_i}{nx}\right)$ for $t_j > t_i$. This has the effect of mapping the difference $t_j - t_i$ into the interval [0,1) such that, as $t_j - t_i$ decreases (i.e., knots are positioned closer), the correlation between the spline effects increases linearly. This formulation of $\boldsymbol{P}$ also allows dealing with irregular timings, whereas the inverse of $\boldsymbol{P}$ can be computed proportional to $O(nx)$ calculations such as for $\boldsymbol{A}^{-1}$, a development shown in Appendix A. To model $\boldsymbol{P}$ with equally spaced knots, let $\boldsymbol{j}$ be a $nx \times 1$ vector with all elements equal to 1; then $t_{nx} = (t_j - t_i)$ nx, and $\boldsymbol{P}$ is equal to

$$\boldsymbol{jj}' - \frac{1}{nx}\begin{bmatrix} 0 & 1 & 2 & . & (nx-2) & (nx-1)\\ 1 & 0 & 1 & . & (nx-3) & (nx-2)\\ 2 & 1 & 0 & . & (nx-4) & (nx-3)\\ . & . & . & . & . & .\\ (nx-2) & (nx-3) & (nx-4) & . & 0 & 1\\ (nx-1) & (nx-2) & (nx-3) & . & 1 & 0 \end{bmatrix}. \qquad [11]$$

To exemplify, let $nx = 5$, then $\boldsymbol{P}$ is

$$\begin{bmatrix} 1 & 0.8 & 0.6 & 0.4 & 0.2\\ 0.8 & 1 & 0.8 & 0.6 & 0.4\\ 0.6 & 0.8 & 1 & 0.8 & 0.6\\ 0.4 & 0.6 & 0.8 & 1 & 0.8\\ 0.2 & 0.4 & 0.6 & 0.8 & 1 \end{bmatrix},$$

which is a matrix with Toeplitz structure (Marin and Dhorne, 2002).

The second formulation considered is related to the stochastic interpretation of $\boldsymbol{f}(t)$ given by Wahba (1990), who showed that a polynomial smoothing spline is the solution to the stochastic differential equation of a Wiener process. Wecker and Ansley (1983) and De Jong and Mazzi (2001) pursued this idea and obtained an expression for the variance-covariance matrix of the spline function and its derivative. Guo (2002) used this formulation to write down a mixed model for functional data. Finally, Hyndman et al. (2005) used the expression for the variance of the stochastic process involving splines that was obtained by De Jong and Mazzi (2001) to derive the following expression for the covariance between knots i and j: $\mathrm{cov}(b_i, b_j) = \dfrac{i^2(3j - i)}{6nx^3}\sigma_b^2$. In terms of a covariance matrix for the cubic spline functions, we have

$$\boldsymbol{\Sigma}\sigma_b^2 = \frac{\sigma_b^2}{6(nx)^3}\begin{bmatrix} 2 & 5 & 8 & . & 3nx-1\\ 5 & 16 & 28 & . & .\\ 8 & 28 & 54 & . & .\\ . & . & . & . & .\\ 3nx-1 & . & . & . & 2(nx)^3 \end{bmatrix}. \qquad [12]$$

Matrix $\boldsymbol{\Sigma}$ also shows a decay of correlation in time, but this is not linear as in $\boldsymbol{P}$. We were not able to find an algorithm that allows inverting $\boldsymbol{\Sigma}$ as simply as $\boldsymbol{P}$.

In the correlation structures induced by matrices $\boldsymbol{P}$ and $\boldsymbol{\Sigma}$, all off-diagonal elements are non-zero, indicating dependency among all random variables in $\boldsymbol{b}$, although long-range covariances may be quite small. Alternatively, one may want to model a correlation structure in which there is covariance only with the next neighbor knot. For this formulation, Durban et al. (2001) presented a covariance matrix for the spline function using a decomposition of the penalty matrix discussed by Green and Silverman (1994). The resulting matrix is tridiagonal, with main diagonal elements equal to $\boldsymbol{U}_{i,i} = 2/3$ $i = 1, 2, \ldots, nx$, and off-diagonals $\boldsymbol{U}_{i+1,i} = \boldsymbol{U}_{i,i+1} = 1/6$, for $i = 1, 2, \ldots, nx - 1$, being 0; otherwise,

$$\boldsymbol{U}\sigma_b^2 = \frac{\sigma_b^2}{6}\begin{bmatrix} 4 & 1 & 0 & . & 0 & 0\\ 1 & 4 & 1 & . & 0 & 0\\ 0 & 1 & 4 & . & 0 & 0\\ 0 & 0 & 1 & . & 1 & 0\\ 0 & 0 & 0 & . & 4 & 1\\ 0 & 0 & 0 & . & 1 & 4 \end{bmatrix}. \qquad [13]$$

The matrix $\boldsymbol{U}$ can be viewed as the covariance structure of a first-order moving average process with correlation between adjacent knots equal to ¼ and 0 thereafter. All AMS discussed in this section, plus the regular animal model with CG and the original P-spline formulation with matrix $\boldsymbol{D'D}$ in [5], were fit to a data set containing birth weights of beef cattle.

Number of Knots

Although knot selection methods exist, P-splines rely on using a large number of knots (Eilers and Marx, 1996), and on limiting the effect of the knots, while constraining the size of the spline coefficients (the b_i). However, for genetic evaluation purposes, the variable DOB expands several years and calving seasons, and its range increases as new data arrive. Therefore, the dimension of the parameter space (the number of knots) may have to increase when the number of records increases. Brumback and Rice (1998) observed that the curve $\sum_{i=1}^{nx} B_i^{(k)} b_i$ is a finite nx-dimensional approximation of the function $\boldsymbol{f}(t)$, which lies in an infinite dimensional parameter space. An appropriate reduction of the parameter space can be produced using the method of sieves (Chen, 2005). By sieves, it is meant a sequence of approximating, and significantly less complex parameter spaces, which optimize some criterion function. The consistency of the method is ensured by increasing the dimension of the parameter space with the increase in sample size. Whereas the asymptotics of the method of sieves are complicated because of the dual approach to infinity of the dimension of the parameter space and the number of data ($nx \to \infty$ and $n \to \infty$), the implementation is easily achieved by fitting models with increased numbers of knots and by comparing them using, for example, the Akaike information criterion ($\boldsymbol{AIC_C}$) as a criterion function. In general, if too many knots are used, the fit tend to be unstable and highly variable (the curve is too "wiggly"). Conversely, fitting too few knots leads to bias. To quantify the effect of the increase in the number of knots, models with 20, 40, 80, or 120 equally spaced knots may be fitted and compared according to convergence, $\boldsymbol{AIC_C}$, and the way they handle periods in which DOB are not observed.

Differential Management Effects

For multiple herd evaluations or to model interruptions of measures, different herds, or types of management, $\boldsymbol{G}_{\mathrm{S}}$ can be taken to be block-diagonal. Whenever some records in a herd cannot be considered as part of the stochastic process related to herd-time effects, the fitting of DOB requires a modification of model [8]–[9]. For example, suppose the trait of interest is weaning weight and, throughout the years, a few animals received a different nutritional management than the majority of them. One possible solution is to modify [8] to incorporate unrelated time-varying parameters to the function in [1]. Let the vector of these CG effects be $\boldsymbol{b}_{\mathrm{G}}$, which is related to the records by the incidence matrix $\boldsymbol{B}_{\mathrm{G}}$. Rows in $\boldsymbol{B}_{\mathrm{G}}$ will be equal to zero, except for those belonging to animals from the differential management, which will have a 1 in the column related to the CG in $\boldsymbol{b}_{\mathrm{G}}$. This vector may be treated as either fixed or random. In the latter case, $\boldsymbol{b}_{\mathrm{G}} \sim N(\boldsymbol{0}, \boldsymbol{I}\sigma_c^2)$, with σ_c^2 either being equal to or different from σ_b^2, in which case this variance has to be estimated too.

Data

Records were 5,175 birth weights from a purebred Polled Hereford herd, belonging to Estancias y Cabaña Las Lilas S.A. and located in Pasteur, western Buenos Aires province, Argentina. Data were collected from 1972 to 2000. A total of 9,742 animals were included in the pedigree file. The records were from 2,739 males and 2,436 females calves, respectively, sired by 177 bulls and from 1,825 dams. The cow herd was kept in cultivated pastures without supplemental feeding throughout the entire year. During most years, there were calving seasons in spring and fall. Therefore, a CG was defined for each calving season within year, creating 53 CG. The average CG day span was 54 d (range = 12 to 107 d), and the average number of calves per CG was 98 (range = 9 to 204).

Models of Analysis. Six animal models were fitted to the records. All models included fixed effects of sex of calf and age of dam and random breeding values and error terms. Four models had DOB as a covariate expressed with a cubic B-spline basis, such as in [1] and [8]. The fifth model included CG as a fixed classification variable, and the sixth also included the covariate DOB nested within CG, as suggested by a reviewer. For $\boldsymbol{X}$ to have full column rank, the sex effect of females was set to 0. In the two models with CG, the last age of dam also was set to 0.

B-Spline Fit of DOB. Covariate DOB was calculated using Julian days. Day 1 was the day the first registered calf in the herd was born (July 2, 1972). The B-spline coefficients for DOB were calculated using the FORTRAN subroutine BSLPVN (De Boor, 1977), with 46 (= nx + 6, or nx = 40) equally spaced knots; however, the order of the vector $\boldsymbol{b}$, as well as of the penalty matrix $\boldsymbol{D'D}$ or the covariance matrices $\boldsymbol{P}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{U}$, was equal to nx = 40.

Estimation of Variance Components. The procedure used to estimate the dispersion parameters was REML (Patterson and Thompson, 1971) by means of the expectation maximization algorithm (Dempster et al., 1977). A program was written in FORTRAN to perform all calculations. The estimator of the additive variance σ_A^2 at the r-iteration was equal to

$$\sigma_A^{2[r]} = \frac{[\hat{\boldsymbol{a}}'\boldsymbol{A}^{-1}\hat{\boldsymbol{a}}]^{[r-1]} + \mathrm{tr}[(\boldsymbol{A}^{-1}\boldsymbol{C}^{\boldsymbol{aa}})\sigma_e^2]^{[r-1]}}{q},$$

where $\hat{\boldsymbol{a}}$ is BLUP($\boldsymbol{a}$), and $\boldsymbol{C}^{\boldsymbol{aa}}$ is the portion of the inverse of the mixed model equations associated with $\boldsymbol{a}$ in the inverse matrix:

$$\begin{bmatrix} \boldsymbol{X'X} & \boldsymbol{X'B} & \boldsymbol{X'Z} \\ \boldsymbol{B'X} & \boldsymbol{B'B} + \boldsymbol{G}_{\mathrm{S}}^{-1}\lambda & \boldsymbol{B'Z} \\ \boldsymbol{Z'X} & \boldsymbol{Z'B} & \boldsymbol{Z'Z} + \boldsymbol{A}^{-1}\alpha \end{bmatrix}^{-1} = \qquad [14]$$

$$\begin{bmatrix} \boldsymbol{C}^{\beta\beta} & \boldsymbol{C}^{\beta\boldsymbol{b}} & \boldsymbol{C}^{\beta\boldsymbol{a}} \\ \boldsymbol{C}^{\boldsymbol{b}\beta} & \boldsymbol{C}^{\boldsymbol{bb}} & \boldsymbol{C}^{\boldsymbol{ba}} \\ \boldsymbol{C}^{\boldsymbol{a}\beta} & \boldsymbol{C}^{\boldsymbol{ab}} & \boldsymbol{C}^{\boldsymbol{aa}} \end{bmatrix}.$$

The REML-EM estimator of σ_b^2 is

$$\sigma_b^{2[r]} = \frac{[\hat{\boldsymbol{b}}'\boldsymbol{G}_{\mathrm{S}}^{-1}\hat{\boldsymbol{b}}]^{[r-1]} + \mathrm{tr}[\boldsymbol{C}^{\boldsymbol{bb}}\boldsymbol{G}_{\mathrm{S}}^{-1}\sigma_e^2]^{[r-1]}}{nx},$$

where $\hat{\boldsymbol{b}}$ = BLUP($\boldsymbol{b}$), and $\boldsymbol{C}^{\boldsymbol{bb}}$ is as in [14]. Finally, the error variances in all AMS were estimated using the formula

$$\sigma_e^{2[r]} = \frac{[\hat{\boldsymbol{e}}'\hat{\boldsymbol{e}}]^{[r-1]} + [\{p + nx + q - \mathrm{tr}(\boldsymbol{C}^{\boldsymbol{bb}})\lambda - \mathrm{tr}(\boldsymbol{A}^{-1}\boldsymbol{C}^{\boldsymbol{aa}})\alpha\}\sigma_e^2]^{[r-1]}}{n};$$

whereas in the two models with GC, the estimated error variance was equal to

$$\sigma_e^{2[r]} = \frac{[\hat{\boldsymbol{e}}'\hat{\boldsymbol{e}}]^{[r-1]} + [\{p + k + q - \mathrm{tr}(\boldsymbol{A}^{-1}\boldsymbol{C}^{\boldsymbol{aa}}\alpha)\}\sigma_e^2]^{[r-1]}}{n}.$$

In all cases n = 5,175, and $\hat{\boldsymbol{e}}$ = BLUP($\boldsymbol{e}$). The algorithm was performed until the difference in the estimates from two successive iterates for any variance component was <10^{-2} kg^2. To make the estimates from all models comparable, regardless of the definition of either fixed CG or random DOB effects, heritability (h^2) was estimated as $\hat{h}^2 = \hat{\sigma}_A^2/(\hat{\sigma}_A^2 + \hat{\sigma}_e^2)$.

To compare the fit obtained with the different models, $\boldsymbol{AIC_C}$, as adapted by Hurvich et al. (1998) to local smoothing spline estimators, was calculated as

Table 1. Estimates of the variance components, heritability (h^2), and a modified Akaike information criterion (AIC_C)

		Parameter[b]				
Item[a]	No.	$\hat{\sigma}_A^2$, kg^2	$\hat{\sigma}_b^2$, kg^2	$\hat{\sigma}_e^2$, kg^2	$\hat{h}^2$	AIC_C
Fixed CG	53[c]	11.15	—	14.34	0.43	4.77
Fixed CG + DOB(CG)	106[c]	10.42	—	14.16	0.42	4.62
D′D	20[d]	12.99	21.24	13.99	0.48	4.74
	40[d]	13.71	25.43	13.09	0.51	4.80
	60[d]	14.96	26.78	11.93	0.55	4.92
	80[d]	14.21	25.82	11.91	0.54	4.85
	120[d]	13.78	26.30	11.82	0.53	4.81
P	20[d]	11.68	12.67	14.81	0.44	4.67
	40[d]	11.27	5.82	14.71	0.43	4.65
	60[d]	11.13	7.86	14.32	0.43	4.64
	80[d]	11.04	8.41	14.11	0.43	4.64
	120[d]	11.02	4.49	14.00	0.44	4.64
Σ	20[d]	11.12	0.06	15.10	0.42	4.63
	40[d]	10.87	5.45	14.85	0.42	4.62
	60[d]	10.74	59.15	14.44	0.42	4.62
	80[de]	—	—	—	—	—
	120[de]	—	—	—	—	—
U	20[d]	11.50	57.60	14.90	0.44	4.66
	40[d]	11.10	96.07	14.69	0.43	4.64
	60[d]	11.09	71.93	14.22	0.44	4.64
	80[d]	10.99	94.14	13.92	0.44	4.64
	120[d]	11.00	101.28	13.69	0.45	4.65

[a] CG = contemporary group, DOB = day of birth, *D′D* = model with first-order penalty matrix, *P* = model with covariance matrix proposed by Cantet (2002), *Σ* = model with covariance matrix proposed by Hyndman et al. (2005), and *U* = model with covariance matrix proposed by Durban et al. (2001).

[b] $\hat{\sigma}_A^2$ = estimated additive genetic variance, $\hat{\sigma}_b^2$ = estimated variance of *B*-spline coefficients, $\hat{\sigma}_e^2$ = estimated error variance, $\hat{h}^2$ = estimated heritability, and AIC_C = Akaike information criterion modified by Hurvich et al. (1998).

[c] Number of fixed CG and fixed regression coefficients.

[d] Number of *B*-spline random coefficients.

[e] Did not meet the convergence criterion (10^{-2} kg^2 for all three variance components) after 1,000 iterations, and parameters could not be estimated.

$$\boldsymbol{AIC_C} = \log \hat{\sigma}_e^2 + 1 + \frac{2(\mathrm{tr}(\boldsymbol{H}) + 1)}{n - \mathrm{tr}(\boldsymbol{H}) - 2}; \qquad [15]$$

the estimated error variance and $\boldsymbol{H}$ is a symmetric matrix such that $\hat{\boldsymbol{y}} = \boldsymbol{Hy}$. The statistic $\boldsymbol{AIC_C}$ was calculated as a byproduct of the EM algorithm when estimating the variance components using the expression (see Appendix B for its derivation):

$$\mathrm{tr}(\boldsymbol{H}) = p + nx + q - [\mathrm{tr}(\boldsymbol{U}^{-1}\boldsymbol{C}^{\boldsymbol{bb}})\lambda + \mathrm{tr}(\boldsymbol{A}^{-1}\boldsymbol{C}^{\boldsymbol{aa}})\alpha].$$

The model with the smallest value of $\boldsymbol{AIC_C}$ is to be selected (Hurvich et al., 1998).

Results

To improve convergence for the model with $\boldsymbol{G}_\mathrm{S} = \boldsymbol{P}$, $\boldsymbol{P}^{-1}$ was divided by its (1, 1) or (nx, nx) element, which is equal to $2/nx$ (see A1 in Appendix A). This is equivalent to Var($\boldsymbol{b}$) = $\boldsymbol{P}\sigma_{b'}^2$, where $\sigma_{b'}^2 = (nx/2)\sigma_b^2$. Notice that the resulting matrix is similar to the first difference penalty matrix $\boldsymbol{D'D}$ in [5], except for the non-zero elements in (1, nx) or (nx, 1), which makes $\boldsymbol{P}$ be nonsingular. In addition, the matrix $\boldsymbol{\Sigma}$ was multiplied by the cube of the constant interval between equally spaced knots $(t_{i+1} - t_i)^3$ to improve convergence. Having done that, the four models with DOB and 20, 40, and 60 knots and the two models with CG met the criterion of convergence in <200 rounds of iteration. The convergence criterion was not met after 1,000 rounds of iteration for the models that included matrix $\boldsymbol{\Sigma}$ with 80 or 120 knots. The REML-EM estimates of σ_A^2, σ_b^2, σ_e^2, h^2, and the value of $\boldsymbol{AIC_C}$ from all models are displayed in Table 1. Within any AMS, estimates of σ_A^2 and σ_e^2 were similar across models with different numbers of knots, and the model with matrix $\boldsymbol{D'D}$ had the largest range of estimates. As a result, $\hat{h}^2$ values in all AMS were similar to the values observed for both models with CG, whereas the model with $\boldsymbol{D'D}$ showed the largest value of $\hat{h}^2$. The estimates of σ_b^2 were different in the four AMS and tended to increase as the number of knots increased. This was not the case for the model with $\boldsymbol{P}$, where $\hat{\sigma}_b^2$ showed a somewhat erratic downward trend.

Comparisons using $\boldsymbol{AIC_C}$ (last column of Table 1) favored the model with CG and the covariate DOB nested within CG ($\boldsymbol{AIC_C}$ = 4.62) and all AMS, with respect to the models with CG ($\boldsymbol{AIC_C}$ = 4.77) or the $\boldsymbol{D'D}$ penalty ($\boldsymbol{AIC_C}$ = 4.74 to 4.92). Among the three AMS,

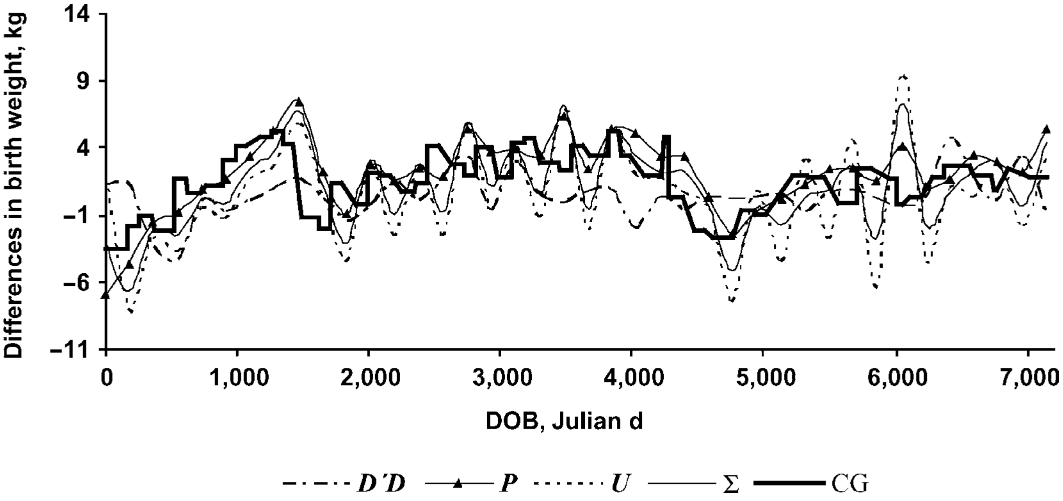

Figure 2. Differences in birth weight plotted against Julian day of birth (DOB) for different models that differ in the covariance matrices for *B*-spline coefficients. Covariance matrices are ***D′D***, first-order penalty matrix; ***P***, as proposed by Cantet (2002); ***Σ***, as proposed by Hyndman et al. (2005); ***U***, as proposed by Durban et al. (2001), and the usual model with fixed contemporary group (CG).

the model with ***Σ*** showed the smallest set of values of $\boldsymbol{AIC_C}$ whenever convergence was met ($\boldsymbol{AIC_C}$ = 4.62 to 4.63), being slightly higher than the range for the models with ***P*** ($\boldsymbol{AIC_C}$ = 4.64 to 4.67) and ***U*** ($\boldsymbol{AIC_C}$ = 4.64 to 4.65). The Pearson and Spearman correlations among the predicted breeding values (BLUP($\boldsymbol{a}$)) for the 20 models in Table 1 were all ≥0.96, so that animals ranked similarly using the predictions across all models. The average PEV($\boldsymbol{a}$) and accuracy were respectively equal to 0.39 and 0.67 for all models, except for those with the ***D′D*** penalty, which ranged from 0.53 to 0.55 for PEV($\boldsymbol{a}$) and from 0.70 to 0.71 for accuracy, reflecting the higher $\hat{h}^2$ in the model with penalty matrix [5].

Figure 2 displays the solutions of CG effects (the thickest line with irregularities), and the BLUP($\boldsymbol{b}$) of all AMS from the models with 40 knots. The shapes of the fitted curves were affected by the model. The curve from the AMS with matrix ***U*** was more wiggly, whereas the one from the model with penalty matrix ***D′D*** was smoother than the curves from the other models. The diagonal lines in the curve for the model with CG are caused by a lack of observations at that time. Of particular interest is the period expanding from 5,012 to 5,219 d of DOB, in which no birth weights were recorded. Notice that the solution of the CG on the left of 5,012 d is lower than the one on the right of 5,219 d. The curve from the model with penalty matrix ***D′D*** passes above the solution of the CG on the left and below the CG on the right. The path from the AMS with ***P*** went up in a more or less straight fashion; however, the curves from the models with ***Σ*** and ***U*** went down after 5,012 d to a local minimum at 5,127 d and then increased to 5,219 d. Thus, this effect seems to be an artifact of both models, which is not supported by the data. During most time intervals with observed birth weights, the curves from the AMS with matrices ***P*** and ***Σ*** tended to have a similar shape.

The effect of the number of knots on the fit is shown in Figure 3, in which the curves for the models with matrix ***P*** and 20, 40, 80, or 120 knots are shown. The effects of the number of knots for the other AMS were similar to the one in Figure 3, so they are not shown here. Increasing the number of knots from 20 to 120 resulted in more wiggly curves. In the interval between 2,000 and 3,000 d of DOB, the AMS with 20 knots went upward softly, missing the peaks observed in Figure 2 for the model with CG. Notice that those peaks were not missed by the models with ≥40 knots. The curves from the AMS with 80 and 120 knots were similar. At the interval in DOB between 5,012 and 5,219 d, a minimum of 5,062 d is observed for the AMS with 80 or 120 knots, which also is not supported by the data.

Discussion

Whether herd environmental conditions manifest themselves in a continuous or in a discrete fashion is debatable, but certainly time is the central criterion to form CG within a herd or management unit. Although time is inherently a continuous variable, it is discretized for the purposes of genetic evaluation. Nonetheless, the categorization of a continuous variable into a discrete variable in linear models results in a loss of statistical information and creates bias (Taylor and Yu, 2002). This is especially true for animals with records near the cut-off date after which a new CG is created. This bias arises by the introduction of a "sudden jump" in expected value at the cut-off point (Taylor and Yu, 2002): all records in the CG to the left of the cut-off have an equal mean that is different from the CG mean to the right of the cut-off. Sivarajasingam (1993) recog-

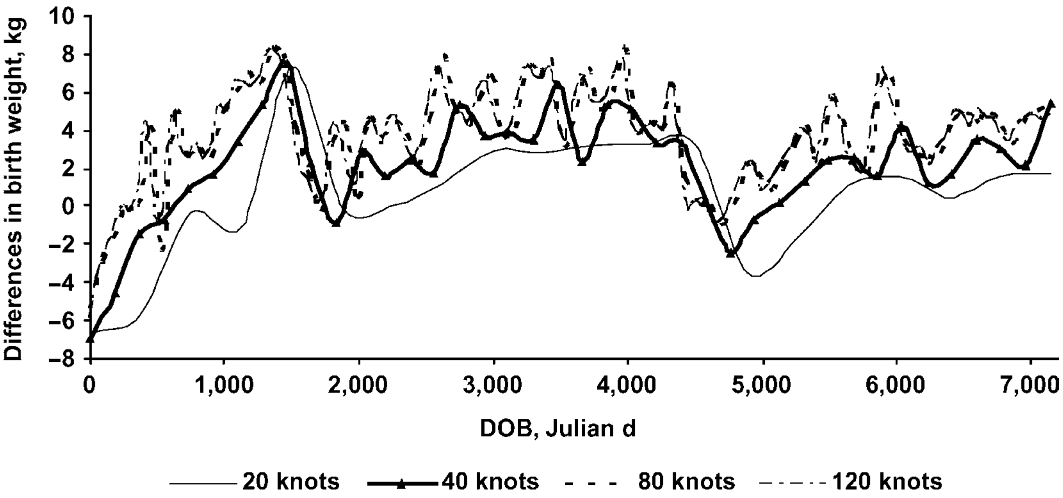

Figure 3. Differences in birth weight plotted against Julian day of birth (DOB) for the models with covariance matrix for *B*-spline coefficients $\boldsymbol{P}\ \sigma_b^2$, as proposed by Cantet (2002), for different numbers of knots.

nized this effect as problematic for CG formation and suggested the use of a similarity matrix to link various observations across CG. Although this formulation mitigates the effects of bias on borderline records from contiguous CG, it still assumes a parametric form for these effects. Conversely, employing a semiparametric fit of DOB does not require assuming any functional form of the covariate, and the need for discretization does not arise. In this regard, we used P-splines with proper covariance matrices to avoid fitting CG (or herd-year-seasons) into an animal model for genetic evaluation. All AMS had smaller (better) values of $\boldsymbol{AIC_C}$, produced almost the same ranking of predicted breeding values, and had similar average PEV($\boldsymbol{a}$) compared with the regular model with CG. Including the linear effects of the covariate DOB nested within CG in the model produced a similar value of $\boldsymbol{AIC_C}$ as those obtained with the AMS, but at the expense of doubling the number of parameters for CG. Carabaño et al. (2004) argued that other procedures to mitigate the cut-off date effect described previously, such as the one proposed by Sivarajansigam (1993), or the modeling of a covariance structure for random CG effects, while still defining CG-classes (Chauhan and Thompson, 1986), found no clear advantage with respect to the classic formulation of fixed CG. From an animal breeding view of model comparison, the criteria that we used ($\hat{h}^2$ correlation between BLUP($\boldsymbol{a}$) or their rankings, average PEV($\boldsymbol{a}$) or accuracy) permit a similar conclusion. However, in the data structure we used, 1) there were a large number of animals in any CG; 2) indicators of connectedness suggested records were extremely well distributed—most CG had calves from several sires, and at least one-half of the sires were repeated every year; and 3) each CG represented a calving season within a year so that 95% were ≤89 d, 75% were ≤65 d, and 5% (one CG) had a spread of 12 d. Therefore, it would be difficult for any other well-posed model to produce differences in predicted breeding values or PEV($\boldsymbol{a}$); however, all AMS performed similarly to models with CG using fewer "parameters" (40 knots compared with 53 or 106). Moreover, the AMS require neither definition of cut-off date as models with CG nor the assumption of any parametric form of the effect being modeled. Finally, in the current application, the unit in which repeated measures are made is the herd and not the animal. As CG are defined on a within-herd basis, herds had more records than CG classes, allowing more direct comparisons of animals, which in theory should decrease problems of connectedness across herds. Overall, the AMS are more parsimonious; fitting DOB with P-splines does not imply any assumption on the functional form of the curve, and its performance in terms of genetic evaluation would be similar to the regular animal model with CG.

Most other applications of splines in animal breeding are based on modeling functional breeding values (White et al., 1999; Huisman et al., 2002; Druet et al., 2003; Iwaisaki et al., 2005). The literature on smoothing methods with splines has become abundant, and the fit of semiparametric methods using mixed linear model software has become popular (Ruppert et al., 2003; Wand, 2003). This is especially true for P-splines (Eilers and Marx, 1996), either viewed from the frequentist (Wand, 2003) or the Bayesian (Lang and Brezger, 2004) camps. There are many reasons for the appeal of this methodology that generalizes ordinary smoothing splines to knot sequences, which are usually much smaller than the response variable. The P-splines have the stable numerical properties of *B*-splines compared with other basis function approaches, such as truncated basis (Eilers and Marx, 2005). We have shown the flexibility of P-splines to accommodate different specifications of covariance matrices as penalties. Moreover, the fit of the function does not require any assumption on the parametrization of the curve.

In this research, we compared curves fitted with P-splines in models with different numbers of knots but at equal intervals or spacings. Within any type of penalty matrix or covariance structure, the values of all statistics used for comparison ($\hat{\sigma}_A^2$, $\hat{\sigma}_e^2$, $\hat{h}^2$, $\boldsymbol{AIC_C}$, average PEV($\boldsymbol{a}$), average accuracy, and ranking of predicted breeding values) were similar for models with different numbers of knots. The only parameter that showed sizeable changes with the increase in the number of knots was $\hat{\sigma}_b^2$, which is the denominator of the smoothing parameter λ. These results are consistent with the observation of Eilers and Marx (1996) that the number of knots is not a critical parameter in P-splines but the smoothing parameter λ is. After extensive simulation, Ruppert and Carroll (2000) found that increasing the number of knots to >40 resulted in marginal decreases in mean square error for all models and examples they worked out. Furthermore, setting the knots at equal spacings performed slightly better than knot sequential procedures. In the words of Altman (2000), “penalized regression splines appear to be robust to the choice of knots (as long as there are sufficiently many).” The curves fitted with matrix $\boldsymbol{P}$ (Figure 3) suggest that 20 knots are not adequate to take care of all features of the trend, whereas 80 or 120 are less parsimonious than 40 knots, but give some inconsistencies in intervals of DOB with no birth weight recorded. In cases where more than one curve has to be fit (for example, a curve per herd or per system of management within a herd), the *B*-spline coefficients and the knots have to be calculated on a herd-by-herd basis, and Var($\boldsymbol{b}$) will be a block diagonal matrix.

The choice of covariance matrix in the AMS attempted to take into account the situations of either complete dependence among the random variables related to the knots (covariance structures from matrices $\boldsymbol{P}$ and $\boldsymbol{\Sigma}$) or to reflect covariance only with the next neighbor knot (matrix $\boldsymbol{U}$) in the other extreme. The performance of all three AMS, in terms of the estimates of the variance components and h^2, PEV($\boldsymbol{a}$), and $\boldsymbol{AIC_C}$, were similar. However, there were differences in the shape of the fitted curves (model with matrix $\boldsymbol{U}$ produced more wiggly curves) and in the approach to convergence when the number of knots increased (model with matrix $\boldsymbol{\Sigma}$ not converging when 80 or 120 knots were fitted). All things considered, the AMS with matrix $\boldsymbol{P}$ and 40 knots seems to be the model of choice for this data.

In this research, the estimation of dispersion parameters and the model comparison procedure ($\boldsymbol{AIC_C}$) followed a likelihood approach; however, to estimate the variance components, a Bayesian method and Gibbs sampling (Sorensen and Gianola, 2002) also are employed in any of the models considered. In fact, models [8]–[9] with covariance matrices [11], [12], or [13] are Bayesian P-splines models (Lang and Brezger, 2004), which results from an *nx*-variate normal prior for $\boldsymbol{b}$ such that a priori $\boldsymbol{b} \sim N(\boldsymbol{0}, \boldsymbol{G}_S\sigma_b^2)$ with $\boldsymbol{G}_S = \boldsymbol{P}$, $\boldsymbol{\Sigma}$ or $\boldsymbol{U}$.

Implications

Semiparametric animal models with a penalized-spline fit of the covariate day of birth of the animal and a proper covariance matrix of the spline coefficients had a better fit than the regular model with contemporary groups. Moreover, the penalized-splines models produced similar ranking of predicted breeding values and average accuracy for birth weights of beef cattle. Thus, the use of penalized regression splines using a *B*-spline basis with proper covariance matrices is a competitive method to avoid fitting contemporary groups (or herd-year-seasons) into animal models for genetic evaluation without having to assume any parametric form for the environmental effects of herd in time.

Literature Cited

Altman, N. 2000. Krige, smooth, both or neither? Aust. N. Z. J. Stat. 42:441–461.

Brumback, B. A., and J. A. Rice. 1998. Smoothing spline models for the analysis of nested and crossed samples of curves. J. Am. Stat. Assoc. 93:961–976.

Cantet, R. J. C. 2002. B-spline fitting of time as an alternative to contemporary group effect. Communication No. 20-11 in Proc. 7th World Cong. Genet. Appl. Livest. Prod., Montpellier, France.

Carabaño, M. J., A. Moreno, P. López-Romero, and C. Díaz. 2004. Comparing alternative definitions of the contemporary group effect in Avileña Negra Ibérica beef cattle using classical and Bayesian criteria. J. Anim. Sci. 82:3447–3457.

Chauhan, V. P. S., and R. Thompson. 1986. Dairy sire evaluation using a ‘rolling months’ model. J. Anim. Breed. Genet. 103:321–333.

Chen, X. 2005. Large sample sieve estimation of semi-nonparametric models. Handbook of Econometrics volume 6. Available: http://athens.src.uchicago.edu/jenni/handbookv6/Chen/. Accessed June 28, 2005.

Crump, R. E., N. R. Wray, R. Thompson, and G. Simm. 1997. Assigning pedigree beef performance records to contemporary groups taking account of within-herd calving patterns. Anim. Sci. 65:193–198.

Currie, I., and M. Durban. 2002. Flexible smoothing with P-splines: A unified approach. Stat. Model. 4:333–349.

De Boor, C. 1977. Package for calculating with B-splines. SIAM J. Numerical Anal. 14:441–472.

De Boor, C. 1993. B(asic)-spline basics. Pages 27–49 in Fundamental Developments of Computer-Aided Geometric Modeling. L. Piegl, ed. Academic Press, San Diego, CA.

De Jong, P., and S. Mazzi. 2001. Modeling and smoothing unequally spaced sequence data. Stat. Inference Stochastic Processes 4:53–71.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. J. Royal Stat. Soc., Series B 39:1–38.

Diggle, P. J., K. Y. Liang, and S. L. Zeger. 1994. Analysis of Longitudinal Data. Clarendon Press, Oxford, UK.

Druet, T., F. Jaffrézic, D. Boichard, and V. Ducrocq. 2003. Modeling lactation curves and estimation of genetic parameters for first lactation test-day records of French Holstein cows. J. Dairy Sci. 86:2480–2490.

Durban, M., I. Currie, and R. Kempton. 2001. Adjusting for fertility and competition in variety trials. J. Agric. Sci. (Camb.) 136:129–140.

Eilers, P. H. C., and B. D. Marx. 1996. Flexible smoothing with *B*-splines and penalties (with comments and rejoinder). Stat. Sci. 11:89–121.

Eilers, P. H. C., and B. D. Marx. 2005. Splines, knots and penalties. Technical report. Available: http://www.stat.lsu.edu/faculty/marx/. Accessed June 23, 2005.

Green, P. J., and B. W. Silverman. 1994. Nonparametric Regression and Generalized Linear Model. Chapman & Hall, London, UK.

Guo, W. 2002. Functional mixed models. Biometrics 58:121–128.

Harville, D. A. 1997. Matrix Algebra from a Statistician's Perspective. Springer-Verlag, New York, NY.

Henderson, C. R. 1984. Applications of Linear Models in Animal Breeding. Univ. of Guelph, Ontario, Canada.

Hobert, J. P., and G. Casella. 1996. The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. J. Am. Stat. Assoc. 91:1461–1473.

Huisman, A. E., R. F. Veerkamp, and J. A. M. van Arendonk. 2002. Genetic parameters for various random regression models to describe the weight data of pigs. J. Anim. Sci. 80:575–582.

Hurvich, C. M., J. S. Simonoff, and C. Tsai. 1998. Smoothing parameter selection in nonparametric regression using an improved Akaike Information Criterion. J. Royal Stat. Soc., Series B 60:271–293.

Hyndman, R. J., M. L. King, I. Pitrun, and B. Billah. 2005. Local lineal forecasts using cubic smoothing splines. Aust. N. Z. J. Stat. 47:87–99.

Iwaisaki, H., S. Tsuruta, I. Misztal, and J. K. Bertrand. 2005. Genetic parameters estimated with multitrait and linear spline-random regression models using Gelbvieh early growth data. J. Anim. Sci. 83:757–763.

Lang, S., and A. Brezger. 2004. Bayesian P-splines. J. Comput. Graphic. Stat. 13:183–212.

Lehmann, E. L. 1983. Theory of Point Estimation. J. Wiley & Sons, New York, NY.

Marin, J. M., and T. Dhorne. 2002. Linear Toeplitz covariance structure models with optimal estimators of variance components. Linear Algebra Appl. 354:195–212.

Patterson, H. D., and R. T. Thompson. 1971. Recovery of inter-block information when block sizes are unequal. Biometrika 58:545–554.

Ruppert, D., and R. J. Carroll. 2000. Spatially-adaptive penalties for spline fitting. Aust. N. Z. J. Stat. 42:205–253.

Ruppert, D., M. P. Wand, and R. J. Carroll. 2003. Semiparametric Regression. Cambridge Univ. Press, Cambridge, UK.

Rybicki, G. B., and W. H. Press. 1992. Interpolation, realization, and reconstruction of noisy, irregularly sampled data. Astrophys. J. 398:169–176.

Sivarajansingam, S. 1993. Comparison of alternative methods of handling contemporary group effects in animal model prediction. J. Anim. Breed. Genet. 110:401–411.

Sorensen, D., and D. Gianola. 2002. Likelihood, Bayesian and MCMC Methods in Genetics. Springer Verlag, Berlin, Germany.

Taylor, J. M. G., and M. Yu. 2002. Bias and efficiency loss due to categorizing an explanatory variable. J. Multivariate Anal. 83:248–263.

Wand, M. P. 2003. Smoothing and mixed models. Comput. Stat. 18:223–249.

Wahba, G. 1990. Spline Models for Observational Data. Soc. Ind. Appl. Math., Philadelphia, PA.

Wecker, W. E., and C. F. Ansley. 1983. The signal extraction approach to nonlinear regression and spline smoothing. J. Am. Stat. Assoc. 78:81–89.

White, I. M. S., R. Thompson, and S. Brotherstone. 1999. Genetic and environmental smoothing of lactation curves with cubic splines. J. Dairy Sci. 82:632–638.

Appendix A: Elements of $\boldsymbol{P}^{-1}$

Let the times of measure (expressed in days, weeks, or months) be $t_1 < t_2 < \ldots < t_{nx-1} < t_{nx}$. All diagonal elements of $\boldsymbol{P}$ are 1, whereas off-diagonals are equal to $1 - [(t_j - t_i)/t_{nx}]$ whenever $t_j > t_i$. Let the matrix of knot differences be

$$\psi = \begin{bmatrix} 0 & t_2 - t_1 & t_3 - t_1 & . & t_{nx-1} - t_1 & t_{nx} - t_1 \\ t_2 - t_1 & 0 & t_3 - t_2 & . & t_{nx-1} - t_2 & t_{nx} - t_2 \\ t_3 - t_1 & t_3 - t_2 & 0 & . & t_{nx-1} - t_3 & t_{nx} - t_3 \\ . & . & . & . & . & . \\ t_{nx-1} - t_1 & t_{nx-1} - t_2 & t_{nx-1} - t_3 & . & 0 & t_{nx} - t_{nx-1} \\ t_{nx} - t_1 & t_{nx} - t_2 & t_{nx} - t_3 & . & t_{nx} - t_{nx-1} & 0 \end{bmatrix}.$$

So we can write $\boldsymbol{P} = \boldsymbol{jj}' - \Psi(t_{nx})^{-1}$, where $\boldsymbol{j}$ is a $nx \times 1$ vector with all elements equal to 1. To invert $\boldsymbol{P}$, we use the formula for the inverse of a sum of matrices (Harville, 1997):

$$\boldsymbol{P}^{-1} = \left(\boldsymbol{jj}' - \frac{\psi}{t_{nx}}\right)^{-1} = (-t_{nx})\left[\boldsymbol{\psi}^{-1} + \boldsymbol{\psi}^{-1}\boldsymbol{j}\left(\frac{1}{1 - \boldsymbol{j}'\boldsymbol{\psi}^{-1}\boldsymbol{j}}\right)\boldsymbol{j}'\boldsymbol{\psi}^{-1}\right].$$

Rybicki and Press (1992) observed that Ψ^{-1} is equal to

$$\boldsymbol{\psi}^{-1} = \frac{1}{2}\begin{bmatrix} \frac{1}{t_{nx} - t_1} - \frac{1}{t_2 - t_1} & \frac{1}{t_2 - t_1} & 0 & . & 0 & \frac{1}{t_{nx} - t_1} \\ \frac{1}{t_2 - t_1} & -\frac{1}{t_3 - t_2} - \frac{1}{t_2 - t_1} & \frac{1}{t_3 - t_2} & . & 0 & 0 \\ 0 & \frac{1}{t_3 - t_2} & -\frac{1}{t_4 - t_3} - \frac{1}{t_3 - t_2} & . & 0 & 0 \\ . & . & . & . & . & . \\ 0 & 0 & 0 & . & -\frac{1}{t_{nx} - t_{nx-1}} - \frac{1}{t_{nx-1} - t_{nx-2}} & \frac{1}{t_{nx} - t_{nx-1}} \\ \frac{1}{t_{nx} - t_1} & 0 & 0 & . & \frac{1}{t_{nx} - t_{nx-1}} & \frac{1}{t_{nx} - t_1} - \frac{1}{t_{nx} - t_{nx-1}} \end{bmatrix}.$$

Although $\boldsymbol{P}$ may have all elements different from 0, its inverse shares the structure of $\boldsymbol{\Psi}^{-1}$. Again, on using the inverse of a sum formula and after algebra, the elements of $\boldsymbol{P}^{-1}$ are

$$\boldsymbol{P}^{-1}_{1,1} = \frac{t_{nx}(t_{nx} + t_2)}{2(t_2 - t_1)(t_{nx} + t_1)} \quad \boldsymbol{P}^{-1}_{1,nx} = \boldsymbol{P}^{-1}_{nx,1} = \frac{t_{nx}}{2(t_{nx} + t_1)},$$

$$\boldsymbol{P}^{-1}_{i,i} = \frac{t_{nx}(t_{i+1} - t_{i-1})}{2(t_{i+1} - t_i)(t_i - t_{i-1})} \quad i = 2, \ldots, nx - 1,$$

$$\boldsymbol{P}^{-1}_{nx,nx} = \frac{t_{nx}(2t_{nx} - t_{nx-1} + t_1)}{2(t_{nx} - t_{nx-1})(t_{nx} + t_1)} \quad \boldsymbol{P}^{-1}_{i,i+1} = \boldsymbol{P}^{-1}_{i+1,i} = \frac{-t_{nx}}{2(\mathrm{t}_{i+1} - t_i)}.$$

With equally spaced knots, all differences t_j– t_i are equal, and $\boldsymbol{P}^{-1}$ is equal to

$$\boldsymbol{P}^{-1} = \begin{bmatrix} 0.5nx & -0.5(nx-1) & 0 & . & 0 & 0.5 \\ -0.5(nx-1) & (nx-1) & -0.5(nx-1) & . & 0 & 0 \\ 0 & -0.5(nx-1) & (nx-1) & . & 0 & 0 \\ . & . & . & . & . & . \\ 0 & 0 & 0 & . & (nx-1) & -0.5(nx-1) \\ 0.5 & 0 & 0 & . & -0.5(nx-1) & 0.5nx \end{bmatrix}. \qquad [A1]$$

Appendix B: Corrected Akaike Information Criterion (AIC_C)

Hurvich et al. (1998) wrote the $\boldsymbol{AIC_C}$ statistics as follows:

$$\boldsymbol{AIC_C} = \log \hat{\sigma}^2_e + 1 + \frac{2[\mathrm{tr}(\boldsymbol{H}) + 1]}{n - \mathrm{tr}(\boldsymbol{H}) - 2}, \qquad [B1]$$

where $\hat{\sigma}^2_e$ is the estimated error variance, n is the number of data in $\boldsymbol{y}$, and $\boldsymbol{H}$ is a symmetric matrix such that the predicted data vector is $\hat{\boldsymbol{y}} = \boldsymbol{Hy}$. Under model [8], $\hat{\boldsymbol{y}}$ is equal to

$$\boldsymbol{Hy} = \boldsymbol{X}\hat{\beta} + \boldsymbol{B}\hat{\boldsymbol{b}} + \boldsymbol{Z}\hat{\boldsymbol{a}}. \qquad [B2]$$

The left-hand side of [B2] can be written as

$$[\boldsymbol{X}|\boldsymbol{B}|\boldsymbol{Z}]\begin{bmatrix} \hat{\beta} \\ \hat{\boldsymbol{b}} \\ \hat{\boldsymbol{a}} \end{bmatrix} = [\boldsymbol{X}|\boldsymbol{B}|\boldsymbol{Z}]\boldsymbol{M}\begin{bmatrix} \boldsymbol{X}' \\ \boldsymbol{B}' \\ \boldsymbol{Z}' \end{bmatrix}\boldsymbol{y},$$

where $\boldsymbol{H} = [\boldsymbol{X}|\boldsymbol{B}|\boldsymbol{Z}]\boldsymbol{M}\begin{bmatrix} \boldsymbol{X}' \\ \boldsymbol{B}' \\ \boldsymbol{Z}' \end{bmatrix}$ and $\boldsymbol{M} = \begin{bmatrix} \boldsymbol{C}^{\beta\beta} & \boldsymbol{C}^{\beta\boldsymbol{b}} & \boldsymbol{C}^{\beta\boldsymbol{a}} \\ \boldsymbol{C}^{\boldsymbol{b}\beta} & \boldsymbol{C}^{\boldsymbol{bb}} & \boldsymbol{C}^{\boldsymbol{ba}} \\ \boldsymbol{C}^{\boldsymbol{a}\beta} & \boldsymbol{C}^{\boldsymbol{ab}} & \boldsymbol{C}^{\boldsymbol{aa}} \end{bmatrix}$ as in [14].

Then, take $\mathrm{tr}(\boldsymbol{H}) = \mathrm{tr}\left[[\boldsymbol{X}|\boldsymbol{B}|\boldsymbol{Z}]\boldsymbol{M}\begin{bmatrix} \boldsymbol{X}' \\ \boldsymbol{B}' \\ \boldsymbol{Z}' \end{bmatrix}\right]$ and rotate it so as to obtain

$$\mathrm{tr}(\boldsymbol{H}) = \mathrm{tr}\left[\begin{bmatrix} \boldsymbol{X}' \\ \boldsymbol{B}' \\ \boldsymbol{Z}' \end{bmatrix}[\boldsymbol{X}|\boldsymbol{B}|\boldsymbol{Z}]\boldsymbol{M}\right] = \mathrm{tr}\left[\begin{bmatrix} \boldsymbol{X'X} & \boldsymbol{X'B} & \boldsymbol{X'Z} \\ \boldsymbol{B'X} & \boldsymbol{B'B} & \boldsymbol{B'Z} \\ \boldsymbol{Z'X} & \boldsymbol{Z'B} & \boldsymbol{Z'Z} \end{bmatrix}\boldsymbol{M}\right]. \qquad [B3]$$

Now, let $\boldsymbol{S}$ be equal to

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{U}^{-1}\lambda & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}^{-1}\alpha \end{bmatrix}$$

and add and subtract $\boldsymbol{S}$ from [B3] in tr($\boldsymbol{H}$). We then have

$$\mathrm{tr}(\boldsymbol{H}) = [\boldsymbol{M}(\boldsymbol{L} + \boldsymbol{S} - \boldsymbol{S})] = \mathrm{tr}[\boldsymbol{M}[(\boldsymbol{L} + \boldsymbol{S}) - \boldsymbol{S}]] = \mathrm{tr}[\boldsymbol{M}(\boldsymbol{L} + \boldsymbol{S}) - \boldsymbol{M}\mathrm{S}],$$

where $\boldsymbol{MS}$ is equal to $\boldsymbol{MS} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{C}^{\boldsymbol{bb}}\boldsymbol{U}^{-1}\lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{C}^{\boldsymbol{aa}}\mathbf{A}^{-1}\alpha \end{bmatrix}$.

Finally,

$$\mathrm{tr}(\boldsymbol{H}) = \mathrm{tr}[\boldsymbol{I} - \boldsymbol{MS}] = \mathrm{tr}(\boldsymbol{I}_{p+nx+q}) - \mathrm{tr}(\boldsymbol{MS})$$

or

$$\mathrm{tr}(\boldsymbol{H}) = p + nx + q - [\mathrm{tr}(\boldsymbol{U}^{-1}\boldsymbol{C}^{\boldsymbol{bb}})\lambda + \mathrm{tr}(\boldsymbol{A}^{-1}\boldsymbol{C}^{\boldsymbol{aa}})\alpha]. \quad \text{[B4]}$$

Expression [B4] is the operational formula to calculate tr($\boldsymbol{H}$).