Canadian Journal of Forest Research

Note

Recognizing van Deusen's mixed estimator for annual forest inventory as a linear mixed model

David L.R. Affleck[a] **and George C. Gaines, III**[b]

[a]Department of Forest Management, University of Montana, Missoula, MT 59812, USA; [b]USDA Forest Service, Rocky Mountain Research Station Forest Inventory & Analysis, Missoula, MT 59808, USA

Corresponding author: **David L.R. Affleck** (email: david.affleck@umontana.edu)

Abstract

The mixed estimator (ME) for annual forest inventory proposed by van Deusen (1999; *Can. J. For. Res.* **29**: 1824–1828) is reformulated as a linear mixed model. This equivalent structure admits an interpretation of the ME as a polynomial regression on year with correlated year-specific random effects. It also uncovers the necessary criterion for maximum likelihood (ML) estimation. The improved performance of the ME under ML estimation is illustrated through simulations and application to inventory data from Montana, USA. Limitations of the ME relating to model-misspecification are also discussed.

Key words: forest inventory, mixed models, penalized estimation, small area estimaton, Fay–Herriot models

Introduction

Continuous forest inventory (CFI) programs are motivated by and must contend with variation in forest structure across space and time. Numerous sampling designs have been developed to spread and balance CFI sample points spatially across a region so as to improve estimation accuracy (e.g., Grafström and Matei 2018), but the distribution of measurements over time generally follows one of two strategies. In the case of a periodic CFI, the complete set of sample points is measured in a single year (or short span of consecutive years), with recurrent full remeasurement planned on some longer-term basis. Alternatively, the sample may be divided into spatially-interpenetrating panels with each panel (re)measured in a given year and the full set of panels completed only after a multi-year inventory cycle has elapsed. These two strategies impose different workforce and budgetary demands. The first necessitates periodically re-assembling measurement crews and securing a significant increase or re-allocation of budget for field sampling. By contrast, the panel strategy requires an agency to continuously maintain personnel and field sampling efforts, along with attendant budgets. The other major difference between the two strategies, and the focus of this research, is that the panel approach allows for estimation of the status of forest resources in any given year, as well as of interannual trends.

Estimates of inventory attributes made from a single CFI panel for a specific year are generally insufficiently precise. This is because the sample size corresponding to the desired accuracy for the resources of interest is typically split across panels and achieved only when the full set of panels is combined. For example, the USDA Forest Service's Forest Inventory and Analysis (FIA) program derived an areal sampling intensity to meet accuracy standards on forest volume and area, then divided the plot network into panels such that approximately 1/5th of the plots are measured every year in eastern states and 1/10th in the west (Bechtold and Patterson 2005). Substituting temporal for spatial domains, the challenge of annual estimation for a panel CFI is analogous to that of small area estimation (SAE; Rao and Molina 2015), wherein a sample size is determined for regional estimation but estimates are then sought for subregions. Moreover, as with SAE, an important consideration is whether annual estimates can be improved by borrowing strength from data and/or estimates pertaining to other, proximal years. Several techniques have been advanced to do so.

Imputation methods have been evaluated for supplementing data from an individual panel with those from other panels in order to improve annual estimates (McRoberts 2001; Eskelson et al. 2009). Such methods draw on measurements from past years or other regions using, for example, similarities in remotely-sensed attributes to locations not measured in the panel for the year of interest. Alternatively, model-based techniques can be used to update or predict values for plots not measured in the year of interest using regression techniques (Johnson et al. 2003). Of course, data from previously measured panels can also be bundled, without adjustment, with those of the panel pertaining to the year of interest. Yet doing so necessarily biases annual estimates toward a periodic mean (van Deusen 2002).

In contrast to the imputation methods surveyed above, van Deusen (1999) proposed a mixed estimator (ME) to draw on temporal associations among annual inventory estimates. Like a Fay–Herriot model for SAE (Rao and Molina 2015, section 4.2), van Deusen's approach effectively models variation

over time and then applies empirical best linear unbiased predictors (EBLUPs) to combine modeled trend and direct panel-based estimates. However, the only auxiliary variable used by the ME is time (i.e., year of measurement), and constraints are specified to restrict the complexity of the temporal trend. Hou et al. (2021) applied the method to estimate annual levels and trends in forest area over parts of the USA. Yet the ME otherwise appears to have been infrequently used, despite longstanding calls for its evaluation for use with the FIA program (Bechtold and Patterson 2005; Westfall et al. 2022). As suggested by Edgar and Westfall (2022), we believe this is in part owing to a perceived complexity as well as to the uniqueness of van Deusen's derivation of the ME from a model with stochastic constraints.

This research was motivated by an interest in connecting van Deusen's (1999) ME to more commonly used statistical modelling techniques. Our primary objective is to demonstrate that van Deusen's model can be specified as a linear mixed model (LMM). In particular, a LMM with fixed effects tied to a temporal polynomial regression with dimension set by the selected constraint, plus correlated year-specific random effects. The best linear unbiased predictor (BLUP) of annual means under this LMM structure then corresponds directly to van Deusen's ME. Secondly, we show that the maximum likelihood (ML) estimator of the variance parameter is distinct from the estimator proposed by van Deusen. This distinction is important because the latter skews toward smaller values. Third, we evaluate the performance of the methodology through simulation and with empirical data, and shed light on some of its limitations. The following section bridges the models and estimators, which are then compared via a small simulation study and an application to forest carbon estimation for Montana, USA.

Models and estimators

Consider a CFI program based on a sequence of K interpenetrating panels of plots, the kth of which is measured in years k, $k+K$, $k+2K$, etc. Let y_t be a direct estimate of a forest attribute for year t that draws exclusively on data from a single panel. The $n \times 1$ vector $\boldsymbol{y}$ arranges the direct estimates chronologically by year, and the estimated covariance matrix of $\boldsymbol{y}$ is denoted by $\widehat{\boldsymbol{\Sigma}}$. The latter may be diagonal for short spans of time, but should have nonzero off-diagonal elements when $n > K$ because direct estimates separated by K years are obtained by remeasurement of the same panel of plots. Within this framework, van Deusen (1999) suggested the following model to link the sequence of direct estimates $\boldsymbol{y}$ to an underlying set of annual means $\boldsymbol{b}$:

$$\text{(1a)} \quad \boldsymbol{y} = \boldsymbol{b} + \boldsymbol{e}$$

$$\text{(1b)} \quad \mathbf{R}_p \boldsymbol{b} = \boldsymbol{v}$$

where $\boldsymbol{e}$ is a vector of sampling errors with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$; $\mathbf{R}_p$ is a $(n-p) \times n$ constraint matrix; and $\boldsymbol{v}$ is a $(n-p) \times 1$ vector of transition errors (independent of $\boldsymbol{e}$) with mean $\mathbf{0}$ and diagonal covariance matrix $q\,\boldsymbol{\Omega}$, where $q \geqslant 0$. van Deusen described a family of constraint matrices that could be applied, with $\mathbf{R}_p$ corresponding to the pth-order numerical derivative operator (Appendix A). He further suggested equating $\boldsymbol{\Sigma}$ to its empirical estimate $\widehat{\boldsymbol{\Sigma}}$ and the diagonal of the unscaled covariance matrix $\boldsymbol{\Omega}$ to the last $(n-p)$ elements of the diagonal of $\widehat{\boldsymbol{\Sigma}}$. This left q as the only variance parameter to be estimated.

Within this framework, van Deusen (1999) established that the optimal estimator of $\boldsymbol{b}$ is given by his ME:

$$\text{(2a)} \quad \widehat{\boldsymbol{b}} = \left[\boldsymbol{\Sigma}^{-1} + \frac{1}{q}\mathbf{R}_p^T \boldsymbol{\Omega}^{-1} \mathbf{R}_p\right]^{-1} \boldsymbol{\Sigma}^{-1} \boldsymbol{y}$$

for $q > 0$ and

$$\text{(2b)} \quad \widehat{\boldsymbol{b}} = \left[\mathbf{I} - \boldsymbol{\Sigma}\mathbf{R}_p^T \left(\mathbf{R}_p \boldsymbol{\Sigma} \mathbf{R}_p^T\right)^{-1} \mathbf{R}_p\right] \boldsymbol{y}$$

for $q = 0$. In eq. 2*a*, q can be viewed as a weighting factor that, as it decreases, shrinks $\widehat{\boldsymbol{b}}$ away from the direct estimates $\boldsymbol{y}$ and toward a temporal trend whose complexity is limited by the selected $\mathbf{R}_p$. In the second case ($q = 0$), $\boldsymbol{b}$ and $\widehat{\boldsymbol{b}}$ are fully constrained to that trend because the transition errors $\boldsymbol{v}$ are necessarily equal to $\mathbf{0}$. Feasible estimators of $\boldsymbol{b}$ are obtained by inserting into eq. 2 an estimate $\widehat{q}$ of q, as well as the empirical estimates of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$. Furthermore, working from Theil (1971, p. 349), van Deusen advanced the following variance expression for $\widehat{\boldsymbol{b}}$ when $q > 0$:

$$\text{(3a)} \quad \mathbb{V}\left(\widehat{\boldsymbol{b}}\right) = \left[\boldsymbol{\Sigma}^{-1} + \frac{1}{q}\mathbf{R}_p^T \boldsymbol{\Omega}^{-1} \mathbf{R}_p\right]^{-1}$$

and it follows from eq. 2*b* that

$$\text{(3b)} \quad \mathbb{V}\left(\widehat{\boldsymbol{b}}\right) = \boldsymbol{\Sigma} - \boldsymbol{\Sigma}\mathbf{R}_p^T \left(\mathbf{R}_p \boldsymbol{\Sigma} \mathbf{R}_p^T\right)^{-1} \mathbf{R}_p \boldsymbol{\Sigma}$$

when $q = 0$. Again, with empirical estimates of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ in hand, these expressions require only an estimate of q to provide feasible estimators of the standard errors for $\widehat{\boldsymbol{b}}$.

To estimate q, van Deusen (1999) suggested regarding both $\boldsymbol{e}$ and $\boldsymbol{v}$ as Gaussian distributed and maximizing the joint density of $\boldsymbol{y}$ and $\boldsymbol{b}$. The logarithm of the joint density of $\boldsymbol{y}$ and $\boldsymbol{v}$ can be written

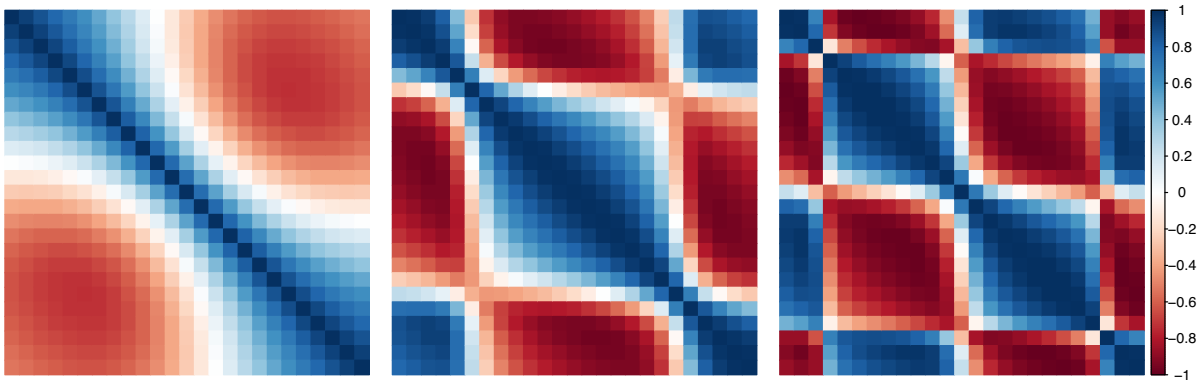
$$\begin{aligned}\mathcal{V}(q; \boldsymbol{y}, \boldsymbol{v}) = &-\frac{2n-p}{2}\ln 2\pi - \frac{1}{2}\ln\left|\begin{matrix}\boldsymbol{\Sigma} & \mathbf{0} \\ \mathbf{0} & q\,\boldsymbol{\Omega}\end{matrix}\right| \\ &- \frac{1}{2}\begin{bmatrix}\boldsymbol{y}-\boldsymbol{b} \\ \boldsymbol{v}\end{bmatrix}^T \begin{bmatrix}\boldsymbol{\Sigma} & \mathbf{0} \\ \mathbf{0} & q\,\boldsymbol{\Omega}\end{bmatrix}^{-1} \begin{bmatrix}\boldsymbol{y}-\boldsymbol{b} \\ \boldsymbol{v}\end{bmatrix} \\ \propto &-(n-p)\ln q - \ln|\boldsymbol{\Omega}| - \ln|\boldsymbol{\Sigma}| \\ &- (\boldsymbol{y}-\boldsymbol{b})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}-\boldsymbol{b}) - \frac{1}{q}\boldsymbol{v}^T \boldsymbol{\Omega}^{-1} \boldsymbol{v}\end{aligned}$$

for $q > 0$. Making the substitution $\boldsymbol{v} = \mathbf{R}_p \boldsymbol{b}$, van Deusen re-expressed this as

$$\begin{aligned}\text{(4)} \quad \mathcal{V}(q; \boldsymbol{y}, \boldsymbol{b}) \propto &-(n-p)\ln q - \ln|\boldsymbol{\Omega}| - \ln|\boldsymbol{\Sigma}| \\ &-(\boldsymbol{y}-\boldsymbol{b})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}-\boldsymbol{b}) - \frac{1}{q}\boldsymbol{b}^T \mathbf{R}_p^T \boldsymbol{\Omega}^{-1} \mathbf{R}_p \boldsymbol{b}\end{aligned}$$

Note, however, that $\boldsymbol{b}$ is not a one-to-one function of $\boldsymbol{v}$ and therefore a valid probability distribution $F_{\boldsymbol{b}}(\boldsymbol{b})$ for $\boldsymbol{b}$ can not be obtained from the aforementioned substitution into the Gaussian distribution $F_{\boldsymbol{v}}(\boldsymbol{v})$ for $\boldsymbol{v}$. That is, $F_{\boldsymbol{b}}(\boldsymbol{b}) \neq F_{\boldsymbol{v}}(\mathbf{R}_p \boldsymbol{b})$ and consequently eq. 4 is not proportional to the log-likelihood of $(\boldsymbol{y}, \boldsymbol{b})$. Yet the maximum of eq. 4 in $\boldsymbol{b}$ given $q > 0$, which can

Fig. 1. Correlation of random year effects in eq. 6 illustrated for a $n = 25$ year sequence with constant transition variance ($\boldsymbol{\Omega} = \mathbf{I}$) and constraint matrix $\mathbf{R}_1$ (left), $\mathbf{R}_2$ (center), and $\mathbf{R}_3$ (right).



be obtained analytically, still yields eq. 2*a*. In contrast, the maximum of eq. 4 in q must be obtained numerically and, importantly, does not equate to the ML estimator (as shown below). We highlight in particular that eq. 4 contains a term proportional to $-\ln q$. This term necessarily rises rapidly as q approaches 0 and we show through simulations that this skews estimates to small values. Indeed, its tendency to yield estimates of q within machine precision of 0 motivated our interest in alternative representations of model 1.

To help elucidate the structure imposed on $\boldsymbol{b}$ by a given constraint matrix, note that $\mathbf{R}_p$ restricts any trend in $\boldsymbol{b}$ such that numerical derivatives of order p (and higher) equal 0, at least up to the random error $\boldsymbol{v}$ (i.e., $\mathbb{E}\left[\mathbf{R}_p\boldsymbol{b}\right] = 0$). To cast this differently, define $\boldsymbol{t}$ as the $n \times 1$ ordered sequence of years for which direct estimates are available and let $\mathbf{X}$ be an $n \times p$ matrix of full column rank with columns corresponding to functions of $\boldsymbol{t}$. A linear model $\mathbf{X}\,\boldsymbol{\beta}$ for the temporal trend with the same constraints can then be specified if $\mathbf{X}$ is chosen such that $\mathbf{R}_p\,\mathbf{X} = 0$. One way to ensure the latter equality is to construct $\mathbf{X}$ from polynomial functions of $\boldsymbol{t}$ up to order $p-1$. Thus, define $\mathbf{X}_p$ from raw powers of $\boldsymbol{t}$

$$\mathbf{X}_p = \left[\boldsymbol{t}^0\ \boldsymbol{t}^1\ \cdots\ \boldsymbol{t}^{p-1}\right] \tag{5}$$

or, for improved numeric stability, as a corresponding set of orthogonal polynomials. Then the temporal trend model $\mathbf{X}_p\,\boldsymbol{\beta}$ carries the same constraint as that imposed on $\boldsymbol{b}$ by $\mathbf{R}_p$. For example, with $p = 1$ all numerical derivatives of the temporal trend are 0 in expectation, which is equivalent to simply $\mathbb{E}[\boldsymbol{b}] = \mathbf{1}_n\,\beta$. On the other hand, if $p = 2$ then second- and higher-order numerical derivatives are 0 in expectation, and a linear regression structure must be augmented to $\mathbb{E}[\boldsymbol{b}] = \left[\mathbf{1}_n\ \boldsymbol{t}\right]\boldsymbol{\beta}$. Using this connection, we can recast eq. 1 as an LMM.

An LMM equivalent to eq. 1 can be written as

$$\boldsymbol{y} = \mathbf{X}_p\boldsymbol{\beta} + \boldsymbol{u} + \boldsymbol{e} \tag{6a}$$

$$\mathbf{R}_p\mathbf{X}_p = \mathbf{0} \tag{6b}$$

$$\mathbf{R}_p\boldsymbol{u} = \boldsymbol{v} \tag{6c}$$

where $\mathbf{X}_p$ is a $n \times p$ polynomial design matrix (per eq. 5), $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed effects, $\boldsymbol{u}$ is a $n \times 1$ vector of random effects, and the other terms are as defined previously. In Appendix B we show that for a given q, LMM 6 and the standard BLUP of $\mathbf{X}_p\boldsymbol{\beta} + \boldsymbol{u}$ are equivalent to van Deusen's (1999) model 1 and ME 2. That is, van Deusen's approach with constraint matrix $\mathbf{R}_p$ can be realized by fitting a $(p-1)$-order polynomial regression $\mathbf{X}_p\boldsymbol{\beta}$ with random year effects $\boldsymbol{u}$, and for any $\widehat{q} \geq 0$ the corresponding EBLUP $\mathbf{X}_p\widehat{\boldsymbol{\beta}} + \widehat{u}$ is equal to $\widehat{\boldsymbol{b}}$ from eq. 2.

The LMM structure is more easily recognized when eq. 6*c* is inverted to write the random effects $\boldsymbol{u}$ as a function of the transition errors $\boldsymbol{v}$

$$\boldsymbol{u} = \left(\mathbf{R}_p^T\boldsymbol{\Omega}^{-1}\mathbf{R}_p\right)^+\mathbf{R}_p^T\boldsymbol{\Omega}^{-1}\boldsymbol{v} = \mathbf{G}\mathbf{R}_p^T\boldsymbol{\Omega}^{-1}\boldsymbol{v}$$

where $\mathbf{G} = \left(\mathbf{R}_p^T\boldsymbol{\Omega}^{-1}\mathbf{R}_p\right)^+$ is the Moore–Penrose inverse of the matrix product inside parentheses. Note that $\mathbf{G}$ is positive semi-definite owing to the underlying quadratic form, but not of full rank n because $\mathbf{R}_p$ is only of rank $n-p$. The covariance matrix of $\boldsymbol{u}$ can then be derived as

$$\mathbb{V}(\boldsymbol{u}) = \mathbf{G}\mathbf{R}_p^T\boldsymbol{\Omega}^{-1}\mathbb{V}(\boldsymbol{v})\,\boldsymbol{\Omega}^{-1}\mathbf{R}_p\mathbf{G} = q\,\mathbf{G}$$

This structure implies a banded, nondiagonal correlation matrix for $\boldsymbol{u}$, with correlations not dependent on q. Correlation matrices are illustrated in Fig. 1 for $\mathbf{R}_1$ through $\mathbf{R}_3$, fixing $n = 25$ years and $\boldsymbol{\Omega} = \mathbf{I}$ to facilitate visualization. The number of bands increases with the order of the constraint imposed by eq. 6*c*, but the undulatory band structure is such that correlations are not simply a function of the temporal separation between elements of $\boldsymbol{u}$.

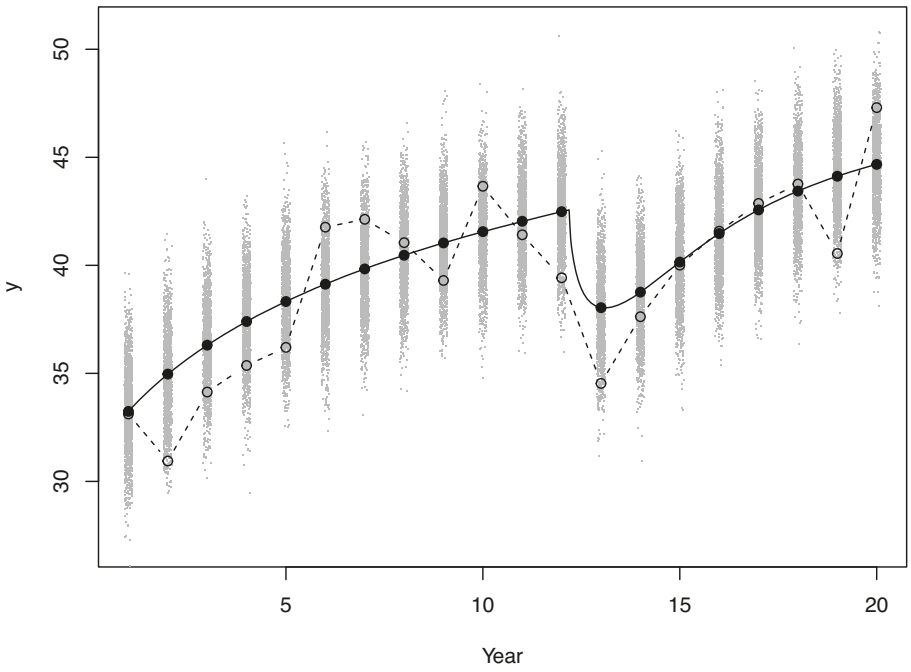
The covariance matrix of $\boldsymbol{y}$ in eq. 6 is

$$\mathbf{V} = \mathbb{V}(\boldsymbol{y}) = \mathbb{E}[\mathbb{V}(\boldsymbol{y} \mid \boldsymbol{u})] + \mathbb{V}[\mathbb{E}(\boldsymbol{y} \mid \boldsymbol{u})] = \boldsymbol{\Sigma} + q\,\mathbf{G} \tag{7}$$

and is necessarily positive definite because $\boldsymbol{\Sigma}$ is of full rank, even though $\mathbf{G}$ is not. Thus, the log-likelihood for model 6 assuming Gaussian errors and random effects can be written

$$\begin{aligned}\mathcal{L}(q, \boldsymbol{\beta}; \boldsymbol{y}) &= -\frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|\mathbf{V}| - \frac{1}{2}(\boldsymbol{y} - \mathbf{X}_p\boldsymbol{\beta})^T\mathbf{V}^{-1}(\boldsymbol{y} - \mathbf{X}_p\boldsymbol{\beta}) \\ &\propto -\ln|\boldsymbol{\Sigma} + q\,\mathbf{G}| - (\boldsymbol{y} - \mathbf{X}_p\boldsymbol{\beta})^T(\boldsymbol{\Sigma} + q\,\mathbf{G})^{-1}(\boldsymbol{y} - \mathbf{X}_p\boldsymbol{\beta})\end{aligned} \tag{8}$$

Fig. 2. Simulated temporal population trend (solid line and points) and distributions of direct estimates for simulated standard error level 5% (grey point clouds). A single simulated sequence of direct estimates is shown by the dashed line and open points.



for $q \geqslant 0$. This can be maximized numerically in q and $\boldsymbol{\beta}$ to obtain ML estimators. Alternatively, analytic solutions for the ML estimators of $\boldsymbol{\beta}$ (given q) can substituted into eq. 8 to obtain a profiled log-likelihood that is a function only of q (see Appendix C). With either approach, we note that the objective function 8 does not contain a term proportional to $\ln q$. That has been replaced with the term $\ln|\boldsymbol{\Sigma} + q\mathbf{G}|$ which collapses to $\ln|\boldsymbol{\Sigma}|$ as q goes to 0. As a result, the ML estimator of q for a given $\mathbf{R}_p$ can be quite different from the estimate obtained by maximizing van Deusen's 1999 objective function 4. This is demonstrated empirically below.

Simulation study

Simulations were undertaken to evaluate the ME using alternative estimators of q. Both polynomial and nonpolynomial temporal trends were simulated, but only results for the latter are reported here. Specifically, the simulations below were developed around the trend $\mu(t) = 26.65 + 6\ln(t+2)(1 - 1.25\chi^2(t - 12.2, 3))$, where $\chi^2(\cdot, d)$ is the χ^2 density function with d degrees of freedom. This defined an increasing sequence of annual means $\mu(t)$ ($t = 1, 2, \ldots, 20$) over a 20-year period, but with a sharp drop ahead of year 13 (Fig. 2). Around this trend, data were simulated for 10 annual panels of 500 plots each. Each plot was assigned an initial measurement and a 10 year remeasurement. These pairs were obtained by adding correlated ($\rho = 0.8$), normally-distributed errors to the appropriate means $\mu(t)$ and $\mu(t+10)$. Measurements separated by more or fewer than 10 years were independent as they came from distinct plots in different panels. Measurement-level standard deviations of 22.4, 44.7, or 89.4 were used in simulations, with these values selected to obtain average standard errors of 2.5%, 5%, or 10% for direct estimates. For each level of variation, 25 000 simulations (of 10 panels of 500 plots) were generated.

For each simulation s, the sequence of direct estimates $\boldsymbol{y}_s$ was obtained by averaging plot measurements. The matrix $\widehat{\boldsymbol{\Sigma}}_s$ was obtained from empirical variances and 10-year covariances (covariances for other lags were fixed at 0). ML estimates $\widehat{q}_s^{\mathrm{ML}}$ of q were then obtained by profiling log-likelihood 8 and applying the optimize function in R 4.4.0 (R Core Team 2024). Also, estimates $\widehat{q}_s^{\mathrm{V}}$ were obtained using the procedure suggested by van Deusen (1999). The latter consists of a discrete optimization of eq. 4 over values for q ranging from 0.01 to 20 (in steps of 0.01). With these estimates of q, the ME 2 was evaluated to provide updated annual means $\widehat{\boldsymbol{b}}_s^{\mathrm{ML}}$ and $\widehat{\boldsymbol{b}}_s^{\mathrm{V}}$. Estimation was carried out using constraint matrices $\mathbf{R}_p$ for $p = 1$ through $p = 15$.

Percent bias and root mean squared error were calculated by year and estimation method as

$$\mathrm{bias}_t = \frac{100}{\mu(t)} \frac{\sum_{s=1}^{25000} \left(\widehat{b}_{s_t} - \mu(t)\right)}{25000}$$

and

$$\mathrm{rmse}_t = \frac{100}{\mu(t)} \sqrt{\frac{\sum_{s=1}^{25000} \left(\widehat{b}_{s_t} - \mu(t)\right)^2}{25000}}$$

where $\widehat{b}_{s_t}$ is an element of $\widehat{\boldsymbol{b}}_s^{\mathrm{ML}}$ or $\widehat{\boldsymbol{b}}_s^{\mathrm{V}}$. In addition, feasible estimates of standard errors were calculated from eq. 3 for each simulation. This allowed construction of pointwise 95% confidence intervals for the updated

annual means (±1.96 estimated standard errors) for each simulated sequence, and evaluation of coverage rates across simulations.

Table 1 summarizes the distributions of estimates of q for constraint matrices $\mathbf{R}_1$ through $\mathbf{R}_3$. It illustrates how estimates of q decline as the order of the constraint matrix increases, corresponding to increased flexibility in the equivalent regression structure $\mathbf{X}_p\,\boldsymbol{\beta}$. Also evident is that estimates of q fall as the variation in the system increases. When direct estimates have larger standard errors, the ME shrinks closer to the fixed effects structure $\mathbf{X}_p\,\boldsymbol{\beta}$ regardless of how q is estimated. Table 1 also shows that the distribution of $\widehat{q}^{\mathrm{V}}$ is shifted left (toward 0) and more concentrated than that of $\widehat{q}^{\mathrm{ML}}$. This is clearest where the simulated levels of variation or the order of the constraint matrix is small. We highlight in particular that the distribution of $\widehat{q}^{\mathrm{V}}$ collapses to the smallest possible estimate (0.01) in 6 of the 9 cases reported in Table 1.

Biases of the ME are shown in the upper panel of Fig. 3 for simulations with direct estimate SEs of 5%. With constraint matrix $\mathbf{R}_1$, bias is low when q is estimated by ML, except around year 13 where there is rapid change in the population trend (see Fig. 2) and to a lesser extent in years 1 and 20 where estimation suffers from lack of data for years 0 and 21. Biases are at least as large when $\widehat{q}_s^{\mathrm{V}}$ are used, because small values of $\widehat{q}_s$ induce more shrinkage away from (unbiased) direct estimates and toward the implied trend. Biases increase when $\mathbf{R}_2$ and $\mathbf{R}_3$ constraints are applied, because linear and quadratic functions still provide poor approximations of the trend. Biases eventually fall as p increases because the polynomial $\mathbf{X}_p\,\boldsymbol{\beta}$ approaches a perfect interpolator of the unbiased direct estimates as p approaches n.

ME accuracy is summarized by the center panel of Fig. 3. For low-order constraint matrices such as $\mathbf{R}_1$, there are substantial gains in precision over direct estimates for most years. In fact, using ML estimation, rmses of $\widehat{\boldsymbol{b}}$ for years 5–10 average 48% those of the direct estimates, a gain equivalent to adding approximately 1670 plots to each of the direct estimates (i.e., $5\% \times 0.48 = 5\% \times \frac{\sqrt{500\text{ plots}}}{\sqrt{2170\text{ plots}}}$). Yet Fig. 3 also shows that bias (particularly around year 13) can overwhelm these gains in precision and lead to rmses larger than those of the direct estimates. When higher-order constraints are applied, rmse of the ME does not exceed that of the direct estimates because of the reduction in bias as p approaches n. Yet gains in precision are also reduced such that rmses for years 5–10, for example, are larger for $\mathbf{R}_{10}$ than for $\mathbf{R}_1$.

Finally, the combined effects of bias in $\widehat{\boldsymbol{b}}$ and in its model-based variance estimator are illustrated via 95% confidence interval failure rates in the lower panel of Fig. 3. Results suggest that pronounced biases in $\widehat{\boldsymbol{b}}$ induced by mismatch of population trend and implied regression structure around year 13 invalidate confidence intervals. For these simulations, this effect was largest for $\mathbf{R}_2$ and $\mathbf{R}_3$ or where $\widehat{q}_s$ was small (e.g., when van Deusen's method was used). For $\mathbf{R}_1$, confidence intervals were also quite conservative (too wide) when bias in $\widehat{\boldsymbol{b}}^{\mathrm{ML}}$ was small.

Annual estimates of forest carbon in Montana

To illustrate application of the ME, FIA plot measurements from Montana, USA, between 2003 and 2019 were obtained. The FIA program uses 10 annual panels in Montana, but owing to access and budget challenges not all plots in a panel are measured in the year assigned. Specifically, over this period in Montana, 4% of measurements were made outside of the prescribed panel year, and plot remeasurement intervals varied from 4 to 15 years. As such, direct estimates for individual years were obtained by aggregating measurements by year taken, rather than by panel. The number of plots contributing to direct estimates thus varied (from 399 to 498) over the 17 year period.

For each year t, live tree measurements were obtained for all sampled base-intensity plots spanning at least one forested condition. Total carbon (c_{it}) in trees above 12 cm in diameter at breast-height was computed for each plot measurement (i), as was the corresponding plot area (a_{it}) of accessible forestland conditions. Direct estimates of annual live tree carbon were then obtained as ratios of means (i.e., $y_t = \sum_i c_{it} / \sum_i a_{it}$). These equate to the standard FIA estimators given by Westfall et al. (2022, p. 10) for the case of a single post-stratum. The latter source also provides variance estimators, but not estimators of covariances between direct estimators for distinct periods. Derivation of analytic covariance estimators is complicated by deviations of plot measurement years from assigned panel years. Thus, a bootstrap approach was employed using with-replacement sampling of individual plot records (i.e., complete measurement sequences for individual plots), with variances and covariances among direct estimates obtained from 999 bootstrap estimates. Bootstrap and analytic estimates of standard errors aligned across years and differed by less than 3.2%. Thus, the bootstrapping results were used to fully specify the variance-covariance matrix of direct estimates $\widehat{\boldsymbol{\Sigma}}$, and to develop the 95% confidence intervals shown in Fig. 4.

The ME was implemented using the first-order constraint matrix $\mathbf{R}_1$ with variance scaling parameter estimated using both ML (R script provided in Supplementary Material) and the method proposed by van Deusen (1999). Both sets of estimates are superimposed on Fig. 4.

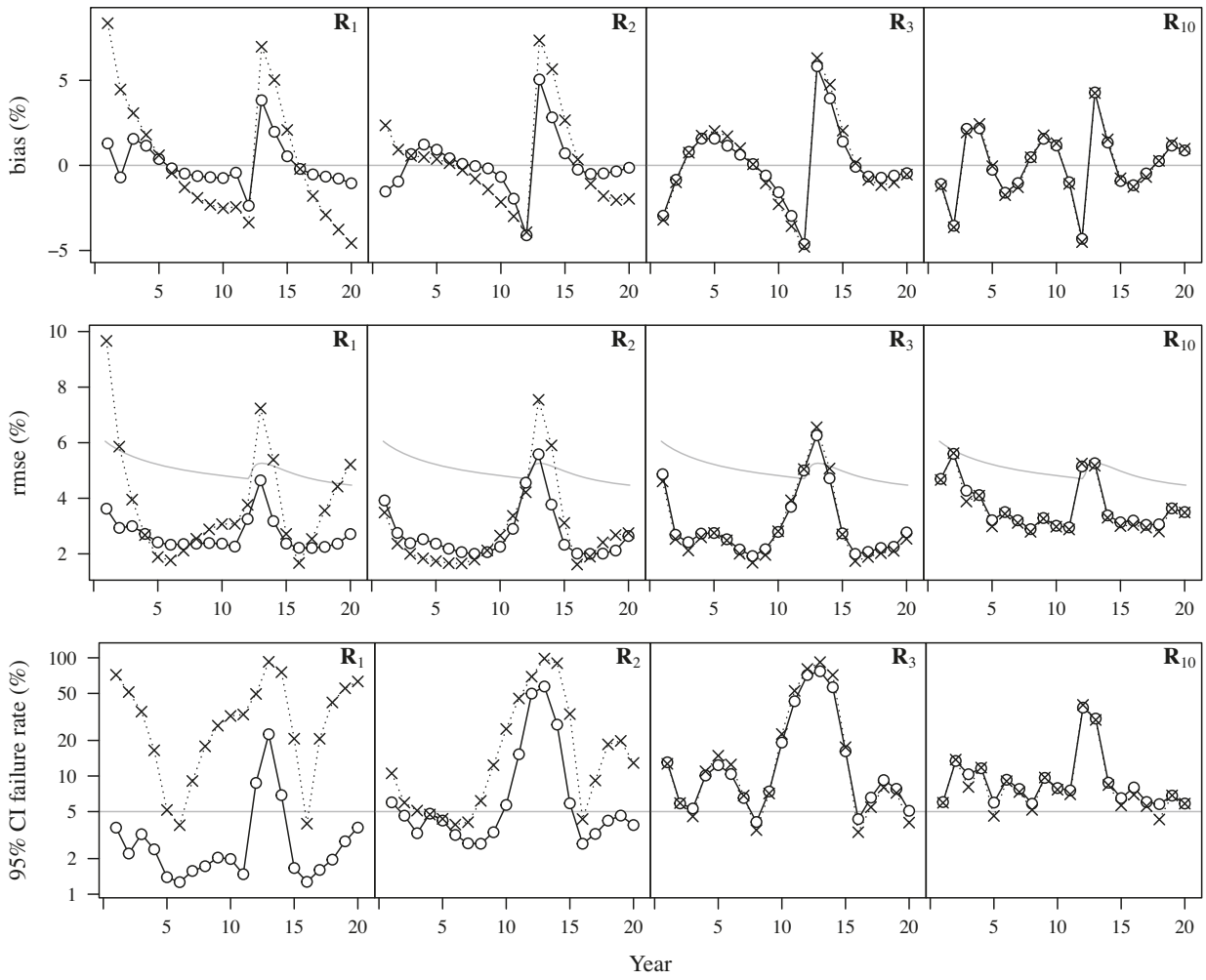
Direct estimates suggest a decline in tree carbon on forestlands across Montana from 2003 to 2010 (Fig. 4). Stocks then appear to be fairly level through 2016, with the suggestion of an increase beginning thereafter. However, direct estimates between 2010 and 2017 are quite variable, dropping approximately 14% from 2009 to 2010, then similarly rebounding upwards in 2011. Annual estimates obtained by the ME with $\widehat{q}^{\mathrm{V}} = 0.01$ trace an overall decline in carbon stocks but show a strong departure from direct estimates before 2007. The annual estimates obtained from the ME using $\widehat{q}^{\mathrm{ML}} = 0.24$ exhibit greater variation owing to larger (in absolute value) year-specific random effects. These in turn allow for more flexibility in adapting from the base constant-in-time model to the trends suggested by the direct estimates.

Table 1. Medians (50%) and quartiles (25%, 75%) of the distributions of estimates of q made by maximum likelihood (ML) ($\widehat{q}^{\mathrm{ML}}$) and the method of van Deusen (1999)($\widehat{q}^{\mathrm{V}}$).

		$\mathbf{R}_1$			$\mathbf{R}_2$			$\mathbf{R}_3$		
SE	Estimator	25%	50%	75%	25%	50%	75%	25%	50%	75%
2.5%	$\widehat{q}^{\mathrm{V}}$	1.08	1.26	1.48	0.03	0.05	0.08	0.01	0.01	0.01
	$\widehat{q}^{\mathrm{ML}}$	2.07	2.34	2.64	0.90	1.28	1.83	0.38	0.71	1.63
5%	$\widehat{q}^{\mathrm{V}}$	0.02	0.04	0.09	0.01	0.01	0.01	0.01	0.01	0.01
	$\widehat{q}^{\mathrm{ML}}$	0.56	0.68	0.84	0.09	0.15	0.26	0.01	0.01	0.05
10%	$\widehat{q}^{\mathrm{V}}$	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	$\widehat{q}^{\mathrm{ML}}$	0.11	0.17	0.26	0.00	0.01	0.03	0.00	0.00	0.00

Note: Results broken out by simulated standard errors (SE) of the direct estimates and by constraint matrices $\mathbf{R}_p$.

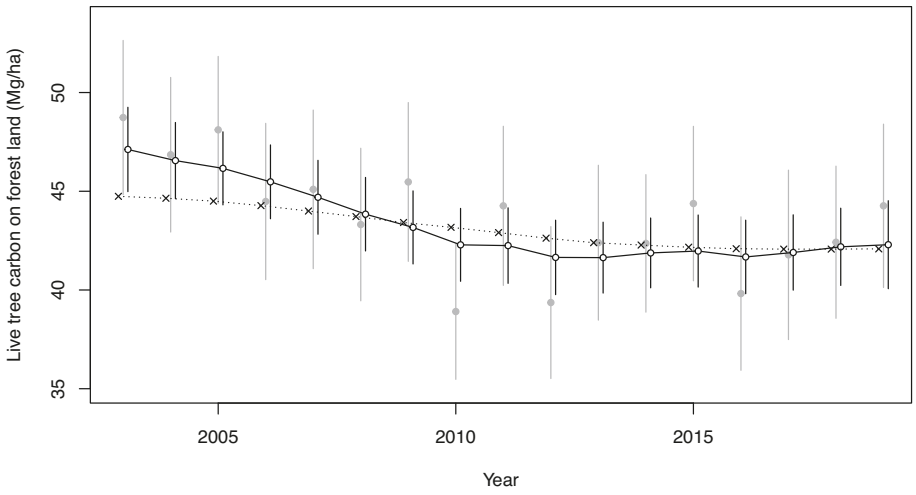
Fig. 3. Bias, root mean squared error (rmse), and 95% confidence interval (CI) failure rates by year at simulated standard error (SE) 5% for the mixed estimator across various constraint matrices $\mathbf{R}_p$. Results for maximum likelihood estimation of q are shown with open circles and solid lines; those where estimation is based on the method of van Deusen (1999) are shown with crosses and dotted lines. Grey lines are for reference: zero bias in the upper tier, the simulated SEs of direct estimates in the middle tier, and nominal 95% CI failure rate in the lower tier.



Discussion

van Deusen (1999) developed his ME from econometric formulations facilitating incorporation of nonsample information into generalized least squares estimation (Theil 1971, section 7.8). In this application, the nonsample information is the temporal constraint on $\boldsymbol{b}$. Yet the model is under-specified relative to modern regression constructions (e.g., what is $\mathbb{E}[\boldsymbol{b}]$ under model 1?), complicating interpretation. The LMM approach provides an equivalent but more accessible framework that could facilitate further

Fig. 4. Estimated live tree carbon on forest lands in Montana. Direct estimates (±1.96 estimated standard errors) are shown in grey; open circles and solid line denote empirical best linear unbiased predictors (EBLUPs) (±1.96 estimated standard errors) for $\mathbf{R}_1$ under maximum likelihood estimation; mixed estimator (ME) using van Deusen (1999) estimation procedure for $\mathbf{R}_1$ shown with crosses and dotted lines. EBLUPs and MEs are horizontally offset for visualization.



development, such as to applications where additional covariates are available. We also emphasize that LMM 6 coincides with the structure of the Fay–Herriot model commonly used in forestry applications of SAE (see Dettmann et al. 2022), connecting the ME with its focus on narrowly-defined temporal subpopulations to methods used in SAE that focus on limited spatial subpopulations.

The LMM framework developed here also facilitated derivation of ML estimators. Indeed, the most important contribution of this work may be the recognition that the estimation procedure recommended by van Deusen (1999) is not ML. That procedure does not carry the theoretical foundations of likelihood-based inference, skews estimates of q toward 0 (Table 1), and can therefore lead to larger biases in the ME when constraints are too inflexible to capture population trends. By contrast, the ML estimator is easily implemented and further connects the ME to standard statistical methodologies.

LMM structure 6 is not unique. Alternative random effects structures $\mathbf{Z}\boldsymbol{u}'$ can be specified using more complex design matrices $\mathbf{Z}$ provided $\boldsymbol{v} = \mathbf{R}_p\boldsymbol{u} = \mathbf{R}_p\mathbf{Z}\boldsymbol{u}'$. Also, there are connections between model 6 and the penalized spline framework of Eilers and Marx (1996), which employs the same contrast matrices $\mathbf{R}_p$ to constrain spline coefficients. Their approach, and more generally the ridge-regression structure of eq. 2*a*, further suggests the potential utility of cross validation strategies for estimating q.

van Deusen (2002) also used simulation to assess the performance of the ME. However, the temporal trends simulated were low-order polynomial and so the estimator's performance there is less generalizable. Though not reported above, we also evaluated the ME for polynomial trends. When the order of such trends was less than p the results were straightforward: the fixed effects components implied by $\mathbf{R}_p$ were capable of accurately capturing trends and distributions of $\widehat{q}$ were concentrated at 0. The simulation trend (Fig. 2) yielding the results described above is not necessarily realistic: the abrupt drop in level ahead of year 13 may be larger than what would occur in forested areas the size, say, Montana. Nonetheless, these simulations were designed so that the underlying trend would be poorly approximated by low-order polynomials, so as to reveal properties of the ME under model-misspecification.

Based on our simulations we recommend using low-order constraints and ML estimation to balance potentially large gains in precision against bias from model misspecification. As illustrated in Fig. 4, use of $\mathbf{R}_1$ and ML estimation allows the ME to deviate from the implied fixed effects structure. The ME adheres more to that structure as p increases or when $\widehat{q}$ is close to 0. Otherwise, more research is needed regarding variance estimation and inference. Substituting $\widehat{q}$ into eq. 3 provides feasible estimators of standard errors but these ignore the uncertainty associated with estimation of q and $\boldsymbol{\Sigma}$. Given the connections noted above, research in SAE may provide useful avenues for improved accuracy estimation for the ME in CFI.

Acknowledgements

The authors thank three reviewers who provided insightful critique and supported the advancement of this work. The findings and conclusions in this publication are those of the authors and should not be construed to represent any official USDA or U.S. government determination or policy.

Article information

History dates

Received: 3 July 2024
Accepted: 23 September 2024
Version of record online: 23 January 2025

Data availability
The data used in this research are publicly available for download from USDA Forest Service's FIA DataMart (www.fs.usda.gov/research/products/dataandtools/tools/fia-datamart).

Author information

Author ORCIDs
David L.R. Affleck https://orcid.org/0000-0001-8370-7631
George C. Gaines, III https://orcid.org/0000-0002-4341-6802

Author contributions
Conceptualization: DA, GG
Formal analysis: DA
Funding acquisition: GG
Investigation: DA
Methodology: DA, GG
Software: DA
Writing – original draft: DA, GG

Competing interests
The authors declare there are no competing interests.

Funding information
This research was supported by funding from the USDA Forest Service through a joint venture agreement with the University of Montana (22-JV-11221638-186).

Supplementary material

Supplementary data are available with the article at https://doi.org/10.1139/cjfr-2024-0180.

References

Bechtold, W.A., and Patterson, P.L. (*Editors*). 2005. The Enhanced Forest Inventory and Analysis Program—National sampling design and estimation procedures. General Technical Report SRS-80, U.S. Department of Agriculture, Forest Service, Southern Research Station, Asheville, NC.

Dettmann, G.T., Radtke, P.J., Coulston, J.W., Green, P.C., Wilson, B.T., and Moisen, G.G. 2022. Review and synthesis of estimation strategies to meet small area needs in forest inventory. Front. For. Glob. Change, **5**: 813569. doi:10.3389/ffgc.2022.813569.

Edgar, C.B., and Westfall, J.A. 2022. Timing and extent of forest disturbance in the Laurentian mixed forest. Front. For. Glob. Change, **5**: 963796. doi:10.3389/ffgc.2022.963796.

Eilers, P.H.C., and Marx, B.D. 1996. Flexible smoothing with B-splines and penalties. Stat. Sci. **11**: 89–121. doi:10.1214/ss/1038425655.

Eskelson, B.N.I., Temesgen, H., and Barrett, T.M. 2009. Estimating current forest attributes from paneled inventory datausing plot-level imputation: a study from the Pacific Northwest. For. Sci. **55**: 64–71. doi:10.1093/forestscience/55.1.64.

Grafström, A., and Matei, A. 2018. Spatially balanced sampling of continuous populations. Scand. J. Stat. **45**: 792–805. doi:10.1111/sjos.12322.

Hou, Z., Domke, G.M., Russell, M.B., Coulston, J.W., Nelson, M.D., Xu, Q., and McRoberts, R.E. 2021. Updating annual state- and county-level forest inventory estimates with data assimilation and FIA data. For. Ecol. Manage. **483**: 118777. doi:10.1016/j.foreco.2020.118777.

Johnson, D.S., Williams, M.S., and Czaplewski, R.L. 2003. Comparison of estimators for rolling samples using forest inventory and analysis data. For. Sci. **49**: 50–63. doi:10.1093/forestscience/49.1.50.

McRoberts, R.E. 2001. Imputation and model-based updating techniques for annual forest inventories. For. Sci. **47**: 322–330. doi:10.1093/forestscience/47.3.322.

R Core Team. 2024. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rao, J.N.K., and Molina, I. 2015. Small area estimation. 2nd ed. Wiley, Hoboken, NJ.

Robinson, G.K. 1991. That BLUP is a good thing: the estimation of random effects. Stat. Sci. **6**: 15–51.

Searle, S.R., Casella, G., and McCulloch, C.E. 1992. Variance components. Wiley, New York.

Theil, H. 1971. Principles of econometrics. Wiley, New York.

van Deusen, P.C. 1999. Modeling trends with annual survey data. Can. J. For. Res. **26**: 1824–1828.

van Deusen, P.C. 2002. Comparison of some annual forest inventory estimators. Can. J. For. Res. **32**: 1992–1995. doi:10.1139/x02-115.

Westfall, J.A., Coulston, J.W., Moisen, G.G., and Andersen, H.E. (*Editors*). 2022. Sampling and estimation documentation for the Enhanced Forest Inventory and Analysis Program: 2022. General Technical Report NRS-207, U.S. Department of Agriculture, Forest Service, Northern Research Station, Madison, WI. doi:10.2737/NRS-GTR-207.

Appendix A. Constraint matrices

Consider a function $g(x)$ observed at n values of x separated by a unit amount (i.e., $x_{t+1} = x_t + 1$). The numerical approximation to the first derivative of $g(x)$ across this span is

$$d_1(x_t) = g(x_t) - g(x_{t-1})$$

for $t = 2, 3, \ldots, n$. In matrix notation, these are given by

$$\begin{bmatrix} -1 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & -1 & 1 \end{bmatrix} \boldsymbol{g} = \mathbf{R}_1 \boldsymbol{g}$$

where $\boldsymbol{g} = \left[g(x_1)\ g(x_2)\ \cdots\ g(x_n) \right]^T$ is the $n \times 1$ vector of observations and $\mathbf{R}_1$ is the first-order constraint matrix (of dimension $(n-1) \times n$) described by van Deusen (1999). By extension, numerical approximations to the second derivatives of $g(x)$ are

$$\begin{aligned} d_2(x_t) &= d_1(x_t) - d_1(x_{t-1}) \\ &= g(x_t) - 2g(x_{t-1}) + g(x_{t-2}) \end{aligned}$$

for $t = 3, 4, \ldots, n$, or in matrix notation

$$\begin{bmatrix} 1 & -2 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & -2 & 1 \end{bmatrix} \boldsymbol{g} = \mathbf{R}_2 \boldsymbol{g}$$

The third derivatives are similarly approximated as $d_2(x_t) - d_2(x_{t-1})$ $(t = 4, 5, \ldots, n)$ and can be obtained via the $(n-3) \times n$ matrix $\mathbf{R}_3$ having rows of the form $[\cdots\ 0\ -1\ 3\ -3\ 1\ 0\ \cdots]$. This logic can be carried forward to the pth order derivative and corresponding constraint matrix $\mathbf{R}_p$ of size $(n-p) \times n$.

Appendix B. Linear mixed model equivalence

From standard linear mixed model theory (e.g., Robinson 1991), the best linear unbiased estimator of $\mathbf{X}_p\boldsymbol{\beta}$ in eq. 6 is

$$\begin{aligned}\mathbf{X}_p\widehat{\boldsymbol{\beta}} &= \mathbf{X}_p\left[\mathbf{X}_p^T\mathbf{V}^{-1}\mathbf{X}_p\right]^{-1}\mathbf{X}_p^T\mathbf{V}^{-1}\mathbf{y}\\ &= \mathbf{V}\mathbf{V}^{-1}\mathbf{X}_p\left[\mathbf{X}_p^T\mathbf{V}^{-1}\mathbf{X}_p\right]^{-1}\mathbf{X}_p^T\mathbf{V}^{-1}\mathbf{y} - \mathbf{V}\mathbf{V}^{-1}\mathbf{y} + \mathbf{V}\mathbf{V}^{-1}\mathbf{y}\\ &= -\mathbf{V}\left(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}_p\left[\mathbf{X}_p^T\mathbf{V}^{-1}\mathbf{X}_p\right]^{-1}\mathbf{X}_p^T\mathbf{V}^{-1}\right)\mathbf{y} + \mathbf{V}\mathbf{V}^{-1}\mathbf{y}\\ &= -\mathbf{V}\left(\mathbf{R}_p^T\left(\mathbf{R}_p\mathbf{V}\mathbf{R}_p^T\right)^{-1}\mathbf{R}_p\right)\mathbf{y} + \mathbf{y}\end{aligned}$$

where the last equality follows from the facts that $\mathbf{R}_p$ has full row rank and $\mathbf{R}_p\mathbf{X}_p = \mathbf{0}$, as well as from the identity

$$\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}_p\left(\mathbf{X}_p^T\mathbf{V}^{-1}\mathbf{X}_p\right)^{-1}\mathbf{X}_p^T\mathbf{V}^{-1} = \mathbf{R}_p^T\left(\mathbf{R}_p\mathbf{V}\mathbf{R}_p^T\right)^{-1}\mathbf{R}_p$$

from Searle et al. (1992, p. 452). Accordingly,

$$\text{(A1)}\quad \mathbf{X}_p\widehat{\boldsymbol{\beta}} = \left(\mathbf{I} - \mathbf{V}\mathbf{R}_p^T\left(\mathbf{R}_p\mathbf{V}\mathbf{R}_p^T\right)^{-1}\mathbf{R}_p\right)\mathbf{y}$$

Here note that if $q = 0$ then model 6 contains no random effects, $\mathbf{V}$ collapses to $\boldsymbol{\Sigma}$, and $\mathbf{X}_p\widehat{\boldsymbol{\beta}}$ consequently reduces to

$$\mathbf{X}_p\widehat{\boldsymbol{\beta}} = \left(\mathbf{I} - \boldsymbol{\Sigma}\mathbf{R}_p^T\left(\mathbf{R}_p\boldsymbol{\Sigma}\mathbf{R}_p^T\right)^{-1}\mathbf{R}_p\right)\mathbf{y}$$

which is the same as van Deusen's estimator of $\boldsymbol{b}$ for $\widehat{q} = 0$ in eq. 2*b*. On the other hand, if $q > 0$ then standard LMM theory (Robinson 1991) provides the best linear unbiased predictor (BLUP) of $\boldsymbol{u}$ in eq. 6 as

$$\widehat{\boldsymbol{u}} = (q\,\mathbf{G})\,\mathbf{V}^{-1}\left(\mathbf{y} - \mathbf{X}_p\widehat{\boldsymbol{\beta}}\right)$$

Substituting $\mathbf{X}_p\widehat{\boldsymbol{\beta}}$ from eq. A1 gives

$$\begin{aligned}\text{(A2)}\quad \widehat{\boldsymbol{u}} &= q\,\mathbf{G}\mathbf{V}^{-1}\left(\mathbf{V}\mathbf{R}_p^T\left(\mathbf{R}_p\mathbf{V}\mathbf{R}_p^T\right)^{-1}\mathbf{R}_p\right)\mathbf{y}\\ &= q\,\mathbf{G}\mathbf{R}_p^T\left(\mathbf{R}_p\mathbf{V}\mathbf{R}_p^T\right)^{-1}\mathbf{R}_p\mathbf{y}\end{aligned}$$

Now applying both eqs. A1 and A2, the BLUP of $\mathbf{X}_p\boldsymbol{\beta} + \boldsymbol{u}$ is

$$\begin{aligned}\mathbf{X}_p\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{u}} &= \left(\mathbf{I} - \mathbf{V}\mathbf{R}_p^T\left(\mathbf{R}_p\mathbf{V}\mathbf{R}_p^T\right)^{-1}\mathbf{R}_p\right)\\ &\quad\times\mathbf{y} + q\,\mathbf{G}\mathbf{R}_p^T\left(\mathbf{R}_p\mathbf{V}\mathbf{R}_p^T\right)^{-1}\mathbf{R}_p\mathbf{y}\\ &= \left[\mathbf{I} - \boldsymbol{\Sigma}\mathbf{R}_p^T\left(\mathbf{R}_p\mathbf{V}\mathbf{R}_p^T\right)^{-1}\mathbf{R}_p - q\,\mathbf{G}\mathbf{R}_p^T\left(\mathbf{R}_p\mathbf{V}\mathbf{R}_p^T\right)^{-1}\mathbf{R}_p\right.\\ &\quad\left.+q\,\mathbf{G}\mathbf{R}_p^T\left(\mathbf{R}_p\mathbf{V}\mathbf{R}_p^T\right)^{-1}\mathbf{R}_p\right]\mathbf{y}\end{aligned}$$

using the definition of $\mathbf{V}$ in eq. 7. From here

$$\begin{aligned}\mathbf{X}_p\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{u}} &= \left[\mathbf{I} - \boldsymbol{\Sigma}\mathbf{R}_p^T\left(\mathbf{R}_p\mathbf{V}\mathbf{R}_p^T\right)^{-1}\mathbf{R}_p\right]\mathbf{y}\\ &= \left[\boldsymbol{\Sigma} - \boldsymbol{\Sigma}\mathbf{R}_p^T\left(\mathbf{R}_p\boldsymbol{\Sigma}\mathbf{R}_p^T + q\,\mathbf{R}_p\mathbf{G}\mathbf{R}_p^T\right)^{-1}\mathbf{R}_p\boldsymbol{\Sigma}\right]\boldsymbol{\Sigma}^{-1}\mathbf{y}\\ &= \left[\boldsymbol{\Sigma} - \boldsymbol{\Sigma}\mathbf{R}_p^T\left(\mathbf{R}_p\boldsymbol{\Sigma}\mathbf{R}_p^T + q\,\boldsymbol{\Omega}\right)^{-1}\mathbf{R}_p\boldsymbol{\Sigma}\right]\boldsymbol{\Sigma}^{-1}\mathbf{y}\end{aligned}$$

where the last equality uses the following result

$$\begin{aligned}\text{(A3)}\quad \mathbf{R}_p\mathbf{G}\mathbf{R}_p^T &= \mathbf{R}_p\left(\mathbf{R}_p^T\boldsymbol{\Omega}^{-1}\mathbf{R}_p\right)^{+}\mathbf{R}_p^T\\ &= \boldsymbol{\Omega}\left(\mathbf{R}_p\mathbf{R}_p^T\right)^{-1}\mathbf{R}_p\mathbf{R}_p^T\boldsymbol{\Omega}^{-1}\mathbf{R}_p\left(\mathbf{R}_p^T\boldsymbol{\Omega}^{-1}\mathbf{R}_p\right)^{+}\\ &\quad\times\mathbf{R}_p^T\boldsymbol{\Omega}^{-1}\mathbf{R}_p\mathbf{R}_p^T\left(\mathbf{R}_p\mathbf{R}_p^T\right)^{-1}\boldsymbol{\Omega}\\ &= \boldsymbol{\Omega}\left(\mathbf{R}_p\mathbf{R}_p^T\right)^{-1}\mathbf{R}_p\mathbf{R}_p^T\boldsymbol{\Omega}^{-1}\mathbf{R}_p\mathbf{R}_p^T\left(\mathbf{R}_p\mathbf{R}_p^T\right)^{-1}\boldsymbol{\Omega}\\ &= \boldsymbol{\Omega}\end{aligned}$$

Finally, since $q > 0$, we can apply the Woodbury identity

$$\left(\mathbf{A} + \mathbf{C}\mathbf{B}\mathbf{C}^T\right)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{C}\left(\mathbf{C}^T\mathbf{A}^{-1}\mathbf{C} + \mathbf{B}^{-1}\right)^{-1}\mathbf{C}^T\mathbf{A}^{-1}$$

where $\mathbf{A} = \boldsymbol{\Sigma}^{-1}$, $\mathbf{B} = \frac{1}{q}\boldsymbol{\Omega}^{-1}$, and $\mathbf{C} = \mathbf{R}_p^T$ (see e.g., Robinson 1991, p. 20), to obtain

$$\mathbf{X}_p\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{u}} = \left[\boldsymbol{\Sigma}^{-1} + \frac{1}{q}\mathbf{R}_p^T\boldsymbol{\Omega}^{-1}\mathbf{R}_p\right]^{-1}\boldsymbol{\Sigma}^{-1}\mathbf{y}$$

which is the same as the ME for $\widehat{q} > 0$ in eq. 2*a*. Thus LMM formulation 6 and its BLUP $\mathbf{X}_p\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{u}}$ are equivalent to van Deusen's model 1 and ME 2.

Appendix C. Profiled log-likelihood

The log-likelihood for q and $\boldsymbol{\beta}$ under model 6 is given by eq. 8 and defined for any $q \geqslant 0$. Substituting in the right hand side of expression A1 for $\mathbf{X}_p\boldsymbol{\beta}$ provides a profiled log-likelihood that must be maximized only for q:

$$\begin{aligned}\text{(A4)}\quad \mathcal{L}^*(q;\mathbf{y}) &\propto -\ln|\boldsymbol{\Sigma} + q\,\mathbf{G}| - \mathbf{y}^T\mathbf{R}_p^T\left(\mathbf{R}_p\mathbf{V}\mathbf{R}_p^T\right)^{-1}\mathbf{R}_p\mathbf{y}\\ &\propto -\ln|\boldsymbol{\Sigma} + q\,\mathbf{G}| - \mathbf{y}^T\mathbf{R}_p^T\left(\mathbf{R}_p\boldsymbol{\Sigma}\mathbf{R}_p^T + q\,\boldsymbol{\Omega}\right)^{-1}\mathbf{R}_p\mathbf{y}\end{aligned}$$

using result A3. This is a nonlinear function of q and thus the profile ML estimate of q must be found using numerical optimatization.

By contrast, the objective function 4 specified by van Deusen (1999, 2002) is defined only for $q > 0$. A profiled version of this function can be obtained by substituting the right hand side of expression 2*a* for $\boldsymbol{b}$. To that end, apply the substitution first to the next-to-last term in eq. 4

$$\begin{aligned}\text{(A5)}\quad &(\mathbf{y} - \boldsymbol{b})^T\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{b})\\ &= \mathbf{y}^T\left(\mathbf{I} - \mathbf{M}\boldsymbol{\Sigma}^{-1}\right)^T\left(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{M}\boldsymbol{\Sigma}^{-1}\right)\mathbf{y}\\ &= \mathbf{y}^T\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Sigma} - 2\mathbf{M} + \mathbf{M}\boldsymbol{\Sigma}^{-1}\mathbf{M}\right)\boldsymbol{\Sigma}^{-1}\mathbf{y}\end{aligned}$$

where $\mathbf{M} = \left[\boldsymbol{\Sigma}^{-1} + \frac{1}{q}\mathbf{R}_p^T\boldsymbol{\Omega}^{-1}\mathbf{R}_p\right]^{-1}$. Next, applying the same substitution to the last term in eq. 4 yields

$$\text{(A6)}\quad \frac{1}{q}\boldsymbol{b}^T\mathbf{R}_p^T\boldsymbol{\Omega}^{-1}\mathbf{R}_p\boldsymbol{b} = \frac{1}{q}\mathbf{y}^T\boldsymbol{\Sigma}^{-1}\mathbf{M}\mathbf{R}_p^T\boldsymbol{\Omega}^{-1}\mathbf{R}_p\mathbf{M}\boldsymbol{\Sigma}^{-1}\mathbf{y}$$

Combining eqs. A5 and A6 gives

$$\begin{aligned}
&\boldsymbol{y}^T\boldsymbol{\Sigma}^{-1}\left[\boldsymbol{\Sigma}-2\mathbf{M}+\mathbf{M}\left(\boldsymbol{\Sigma}^{-1}+\frac{1}{q}\mathbf{R}_p^T\boldsymbol{\Omega}^{-1}\mathbf{R}_p\right)\mathbf{M}\right]\boldsymbol{\Sigma}^{-1}\boldsymbol{y}\\
&=\boldsymbol{y}^T\boldsymbol{\Sigma}^{-1}\left[\boldsymbol{\Sigma}-\mathbf{M}\right]\boldsymbol{\Sigma}^{-1}\boldsymbol{y}\\
&=\boldsymbol{y}^T\boldsymbol{\Sigma}^{-1}\left[\boldsymbol{\Sigma}-\boldsymbol{\Sigma}+\boldsymbol{\Sigma}\mathbf{R}_p^T\left(\mathbf{R}_p\boldsymbol{\Sigma}\mathbf{R}_p^T+q\,\boldsymbol{\Omega}\right)^{-1}\mathbf{R}_p\boldsymbol{\Sigma}\right]\boldsymbol{\Sigma}^{-1}\boldsymbol{y}\\
&=\boldsymbol{y}^T\mathbf{R}_p^T\left(\mathbf{R}_p\mathbf{V}\mathbf{R}_p^T\right)^{-1}\mathbf{R}_p\boldsymbol{y}
\end{aligned}$$

where the next-to-last equality uses the Woodbury identity and assumes $q > 0$. Entering this last expression into eq. 4 provides a profiled version as

$$\begin{aligned}
\mathcal{V}^*(q;\boldsymbol{y}) \propto &-(n-p)\ln q-\ln|\boldsymbol{\Omega}|-\ln|\boldsymbol{\Sigma}|\\
&-\boldsymbol{y}^T\mathbf{R}_p^T\left(\mathbf{R}_p\boldsymbol{\Sigma}\mathbf{R}_p^T+q\,\boldsymbol{\Omega}\right)^{-1}\mathbf{R}_p\boldsymbol{b}
\end{aligned}$$

similarly defined only for $q > 0$ and still containing the first term $-(n-p)\ln q$.