BIOMETRICS 38, 479–484
June, 1982

A Regression Application for Comparing Growth Potential of Environments at Different Points in the Growth Cycle

F. Thomas Lloyd

Forestry Sciences Laboratory, 2730 Savannah Highway,
Charleston, South Carolina 29407, U.S.A.

H. David Muse[1]

Southeastern Forest Experiment Station,
Asheville, North Carolina 28802, U.S.A.

and

William L. Hafley

Department of Forestry, North Carolina State University,
Raleigh, North Carolina 27650, U.S.A.

SUMMARY

Regression techniques are used to quantify a procedure for indexing the productive capacity of forested sites. The regression model parameters $\boldsymbol{\theta}' = (\gamma^*, \beta_1, \beta_2, \ldots, \beta_m)$ are defined in such a way that the value of the indexing parameter γ^* identifies individual growth-over-time response curves from a family of curves that have the parameters $\beta_1, \beta_2, \ldots, \beta_m$ in common. The β terms are estimated from historical cumulative growth data, while γ^* is estimated from these historical data together with r values of the response characteristic observed at a single point in the growth cycle of each new environment to which the prediction equations are to be applied. The formulation allows comparison of growth potential of environments containing equal-aged groups of experimental units that are at different points in the growth cycle, by means of an index defined as projected growth at a preselected common age.

1. Introduction

In forestry practice it is frequently of interest to develop an index of growth potential, called the 'site index', for a forested environment containing equal-aged trees of a single species. This species-specific site index is an estimate of the average height that tree species will attain by a given age in that environment. In order to compare growth potential of environments supporting trees of different ages, cumulative growth is projected to a common age. This procedure is especially useful where the growth cycle covers a long time span and instantaneous comparison of environment growth potential is needed. The purpose of this paper is to generalize the work of F. T. Lloyd, in a 1975 thesis from North Carolina State University, Raleigh, and Lloyd and Hafley (1977), in their development of (i) a family of species-specific site-indexing curves, (ii) a procedure for predicting site index for different environments containing different-aged groups of experimental units, and (iii) a measure of the precision of the site index predictor.

---

[1] Present address: Middle Tennessee State University, Department of Mathematics and Computer Science, Murfreesboro, Tennessee 37132, U.S.A.

Key words: Least squares; Linear regression; Growth curves; Prediction; Projection.

2. A Model for Indexing Environments

Let the set

$$\{(t_k, Y_{ijk}): i=1,2,\ldots,p;\ j=1,2,\ldots,s_i;\ k=1,2,\ldots,n\} \tag{1}$$

denote species-specific growth-curve data, where Y_{ijk} is an observation on the growth characteristic of interest for the jth experimental unit in the ith environment at the kth age, with common ages t_k $(k=1,2,\ldots,n)$ used for the repeated measurements on all experimental units. The model for these responses $(\mathbf{Y}_{(\nu n\times 1)}, \nu=\sum_{i=1}^{p} s_i)$ is

$$\mathbf{Y}=(\mathbf{X}_1\mathbf{X}_2)\begin{pmatrix}\boldsymbol{\gamma}\\ \boldsymbol{\beta}\end{pmatrix}+\boldsymbol{\varepsilon}, \tag{2}$$

where the p elements of the vector $\boldsymbol{\gamma}$ are environment-specific indexing parameters for the environments comprising (1), $\boldsymbol{\beta}$ is an m-vector of recurrent parameters, and $\boldsymbol{\varepsilon}$ is a νn-vector $\mathrm{N}(\mathbf{0}_{\nu n}, \mathbf{I}_\nu\otimes\boldsymbol{\Sigma})$. Where, $\mathbf{0}_{\nu n}$ is a νn-vector of zeros, $\mathbf{I}_\nu$ is the identity matrix of rank ν, $\otimes$ denotes the Kronecker product of two matrices, such that

$$\mathbf{A}\otimes\mathbf{B}=\begin{bmatrix} a_{11} & \cdots & a_{1n}\\ \vdots & & \vdots\\ a_{m1} & \cdots & a_{mn}\end{bmatrix}\otimes\mathbf{B}=\begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B}\\ \vdots & & \vdots\\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B}\end{bmatrix},$$

and $\boldsymbol{\Sigma}$ is an $n\times n$ variance–covariance matrix associated with each vector of repeated measurements on an experimental unit, in which $\operatorname{var}(\varepsilon_{ijk})=\sigma^2$ and $\operatorname{cov}(\varepsilon_{ijh}, \varepsilon_{ijk})=\sigma_{hk}$. The leftmost $\nu n\times p$ partition of the design matrix is given by

$$\mathbf{X}_1=\text{block diag}\,(\mathbf{J}_{s_1}, \mathbf{J}_{s_2}, \ldots, \mathbf{J}_{s_p})\otimes\mathbf{F}, \tag{3}$$

where $\mathbf{J}_u=$ a u-vector of ones,

$$\text{block diag}\,(\mathbf{B},\mathbf{C})=\begin{bmatrix}\mathbf{B} & \mathbf{0}\\ \mathbf{0} & \mathbf{C}\end{bmatrix},$$

and $\mathbf{F}'=[f(t_1), f(t_2), \ldots, f(t_n)]$. The rightmost $\nu n\times m$ partition of the design matrix is

$$\mathbf{X}_2=\mathbf{J}_\nu\otimes\mathbf{G}, \tag{4}$$

where

$$\mathbf{G}=\begin{bmatrix} g_1(t_1) & \cdots & g_m(t_1)\\ \vdots & & \vdots\\ g_1(t_n) & \cdots & g_m(t_n)\end{bmatrix}.$$

In this research, the gain in precision from using weighted least squares (Rao, 1967) was marginal, so, to keep estimation procedures relatively simple and easy to apply in practice, the standard least squares estimators were used for $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\beta}}$, namely,

$$\begin{bmatrix}\hat{\boldsymbol{\gamma}}\\ \hat{\boldsymbol{\beta}}\end{bmatrix}=\begin{bmatrix}\mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{X}_2\\ \mathbf{X}_2'\mathbf{X}_1 & \mathbf{X}_2'\mathbf{X}_2\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{X}_1'\mathbf{Y}\\ \mathbf{X}_2'\mathbf{Y}\end{bmatrix}. \tag{5}$$

From (3) and (4), the estimators of $\boldsymbol{\beta}$ and each element of $\boldsymbol{\gamma}$ in (5) can be expressed in terms of $\mathbf{F}$ and $\mathbf{G}$, where

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \nu^{-1}(\mathbf{G}'\mathbf{A}\mathbf{G})^{-1}\mathbf{G}'\mathbf{A}\sum_{i=1}^{p}\sum_{j=1}^{s_i}\mathbf{Y}_{ij}, \\ \hat{\gamma}_i &= \mathbf{F}'\left(s_i^{-1}\sum_{j=1}^{s_i}\mathbf{Y}_{ij}-\mathbf{G}\hat{\boldsymbol{\beta}}\right)\Big/\sum_{k=1}^{n}\{f(t_k)\}^2, \\ \mathbf{A} &= \mathbf{I}_n-\mathbf{F}\mathbf{F}'\Big/\sum_{k=1}^{n}\{f(t_k)\}^2, \end{aligned} \tag{6}$$

and $\mathbf{Y}_{ij}$ denotes the n-vector of repeated measurements for the (i, j)th experimental unit. Further details are included in the Appendix.

The experimental unit residual variance–covariance matrix is

$$\mathbf{C}=\sum_{i=1}^{p}\sum_{j=1}^{s_i}(\mathbf{Y}_{ij}-\hat{\mathbf{Y}}_{ij})(\mathbf{Y}_{ij}-\hat{\mathbf{Y}}_{ij})'/(\nu-m-p),$$

and the elements of the estimate of $\boldsymbol{\Sigma}$ are

$$\hat{\sigma}^2=\sum_{k=1}^{n}c_{kk}/n \tag{7}$$

for the diagonal and

$$\hat{\sigma}_{hk}=\hat{\sigma}^2c_{hk}/(c_{hh}c_{kk})^{\frac{1}{2}} \tag{8}$$

for the off-diagonal. Note that (8) preserves the simple correlation relationships defined in $\mathbf{C}$, thereby assuring the positive semi-definite property of $\hat{\boldsymbol{\Sigma}}$.

3. The Index Predictor

Suppose that there exist sufficient prior data in the form of (1) to permit the development of the model (2) for the growth characteristic of interest. Then the resulting estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ as defined by (6), (7) and (8) can be used in the prediction of a growth-over-time response curve for a new environment without the need to wait for the experimental units to attain maturity; this assumes, of course, that the model and prior estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ remain valid for the new environment. The index predictor is completed by estimating an environment-specific indexing parameter, γ^*, and then using the resulting response curve to predict cumulative growth at a preselected age.

The model for the response, Y^*, in the new environment is

$$Y^*=\gamma^*f(t)+\mathbf{g}'(t)\boldsymbol{\beta}+\varepsilon^*, \tag{9}$$

where γ^* is the environment-specific indexing parameter, $\mathbf{g}'(t)=[g_1(t), g_2(t), \ldots, g_m(t)]$, and the ε^* are NID $(0, \sigma^2)$. The index predictor is

$$\hat{Y}_I^*=\hat{\gamma}^*f(t_I)+\mathbf{g}'(t_I)\hat{\boldsymbol{\beta}}, \tag{10}$$

where t_I is the common age for comparison of response curves, $\hat{\boldsymbol{\beta}}$ is given by (6), and the estimator of γ^* remains to be derived.

Suppose that the set of observations $\{(t_\tau, Y_l^*),\ l=1, 2, \ldots, r\}$ is obtained by measurement of the response characteristic, Y_l^*, on r equal-aged experimental units at time $t=t_\tau$. Then, using the prior estimator of $\boldsymbol{\beta}$ given in (6), we find from (9) that

$$\bar{Y}_\tau^*=\gamma^*f(t_\tau)+\mathbf{g}'(t_\tau)\hat{\boldsymbol{\beta}}+\bar{\varepsilon}^*.$$

This gives the estimator

$$\hat{\gamma}^* = \{\bar{Y}^*_\tau - \mathbf{g}'(t_\tau)\hat{\boldsymbol{\beta}}\}/f(t_\tau), \tag{11}$$

where $\bar{Y}^*_\tau$ is the mean of the Y^*_l. Substitution of (11) into (10) yields the index predictor

$$\hat{Y}^*_I = \bar{Y}^*_\tau f(t_I)/f(t_\tau) + \mathbf{D}'\hat{\boldsymbol{\beta}}, \tag{12}$$

where

$$\mathbf{D}' = \mathbf{g}'(t_I) - \mathbf{g}'(t_\tau)f(t_I)/f(t_\tau).$$

A shortcoming of past site-index prediction research in forestry has been the inability of each approach to support evaluation of the precision of the site-index estimator. This approach presents no such problems since

$$\operatorname{var}(\hat{Y}^*_I) = \{f(t_I)/f(t_\tau)\}^2\sigma^2/r + \mathbf{D}'\boldsymbol{\Sigma}_\beta\mathbf{D}, \tag{13}$$

where

$$\boldsymbol{\Sigma}_\beta = (\mathbf{G}'\mathbf{A}\mathbf{G})^{-1}\mathbf{G}'\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}\mathbf{G}(\mathbf{G}'\mathbf{A}\mathbf{G})^{-1},$$

and at least $\nu - m - p$ degrees of freedom are available for estimation of $\operatorname{var}(\hat{Y}^*_I)$. Consequently, using (12) and (13) and the appropriate percentile for the t distribution with $\nu - m - p$ degrees of freedom, we find a conservative $100(1-\alpha)\%$ confidence interval for Y^*_I to be

$$\hat{Y}^*_I \pm t_{(\frac{1}{2}\alpha, \nu-m-p)}\{\operatorname{var}_{\text{est}}(\hat{Y}^*_I)\}^{\frac{1}{2}}. \tag{14}$$

The estimators of γ^*, $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are precisely the standard least squares estimates produced by the augmented model

$$\begin{bmatrix}\mathbf{Y}^*\\ \mathbf{Y}\end{bmatrix} = \begin{bmatrix}\mathbf{J}_r f(t_\tau) & \mathbf{0} & \mathbf{J}_r\mathbf{g}'(t_\tau)\\ \mathbf{0} & \mathbf{X}_1 & \mathbf{X}_2\end{bmatrix}\begin{bmatrix}\gamma^*\\ \boldsymbol{\gamma}\\ \boldsymbol{\beta}\end{bmatrix} + \begin{bmatrix}\boldsymbol{\varepsilon}^*\\ \boldsymbol{\varepsilon}\end{bmatrix},$$

which suggests a relatively efficient utilization of the data. The previously mentioned thesis by Lloyd on the height growth of trees indicates a minimal loss of efficiency due to the use of standard least squares estimation over generalized least squares estimation. Also, the supplementary data could be used to update the estimate of $\boldsymbol{\Sigma}$; however, for the sample sizes anticipated, the improvement was thought to be negligible.

4. Example

In this research, height measurements from loblolly pine trees, collected by the U.S. Forest Service and originally studied by Trousdell, Beck and Lloyd (1974), were re-analyzed according to the model (2). These data used for (1) involved $p = 19$ environments (plots), each of which supported stands of trees at least 50 years old, with $s_i = 2$ experimental units (trees) selected per environment. The subgroup from which the sample trees came was limited to those presently receiving full sunlight from above and having no prior periods of slow growth as evidenced by the absence of very narrow growth rings. The height-over-age data were obtained by felling the sample trees, cutting them into sections, taking increment borings at the top of each section to determine age at that height, and, when necessary, splitting the sections to determine past terminal height by examining the pith. The data were used to develop height-over-age data at five-year intervals by linear interpolation when necessary from age 10 through age 55, which means that $n = 10$ repeated measurements were established for each tree.

Lloyd's investigation of the data resulted in choices of the $f(t)$ and $g(t)$ functions such that the model (2) took the following form for the (i, j, k)th component of the observation vector $\mathbf{Y}$:

$$Y_{ijk} = \gamma_i t_k(1-0.01t_k) + \beta_1 t_k^2 + \beta_2 t_k^3 + \varepsilon_{ijk}.$$

Subsequent analysis of the data produced the estimates $\hat{\beta}_1 = 0.00172718$, $\hat{\beta}_2 = 0.000101827$, $\hat{\sigma}^2 = 16.4511$ and

$$\hat{\mathbf{\Sigma}}_\beta = \begin{bmatrix} 3.5359\times 10^{-4} & -4.9125\times 10^{-6} \\ -4.9125\times 10^{-6} & 6.9534\times 10^{-8} \end{bmatrix}.$$

Suppose now that data from a new environment yields a mean height from five randomly selected ten-year-old trees of $\bar{Y}_\tau^* = 33.2$ feet. Use of $t_I = 25$ as the index age in (12) produces the index prediction $\hat{Y}_I^* = 71.8$. The variance of the index estimator obtained by substituting $\hat{\sigma}^2$ and $\hat{\mathbf{\Sigma}}_\beta$ into (13) is $\text{var}_{\text{est}}(\hat{Y}_I^*) = 33.0$, which means that a conservative 95% confidence interval for the index, with $\nu - p - m = 17$ degrees of freedom, is 71.8 ± 12.0 feet at the index age $t_I = 25$.

Acknowledgements

We are grateful to Kenneth B. Trousdell and O. Gordon Langdon for the data used in this analysis, and to the U.S. Forest Service for the financial support of the training program from which this research resulted. We also want to thank the anonymous reviewer whose constructive criticisms improved the clarity of this paper.

Résumé

Des techniques de régression sont utilisées pour quantifier une procédure d'indexation de la capacité de production de sites forestiers. Les paramètres du modèle de régression $\theta' = (\gamma^*, \beta_1, \beta_2, \ldots, \beta_m)$ sont définis de telle sorte que la valeur du paramètre d'indexation γ^* identifie des courbes de réponse individuelles de croissance par rapport au temps à partir d'une famille de courbes qui ont les paramètres $\beta_1, \beta_2, \ldots, \beta_m$ en commun. On estime les β à partir de données historiques de croissance cumulée, tandis que γ^* est estimée à partir des mêmes données auxquelles on ajoute r valeurs de la caractéristique de la réponse observée à un seul point du cycle de croissance de chaque environnement nouveau auquel les équations de prédiction doivent être appliquées. La formulation permet la comparaison de croissance potentielle d'environnements contenant des groupes de même âge des unités expérimentales; ces dernières sont à différents points du cycle de croissance par l'intermédiaire d'un index définie comme la projection de la croissance à un âge commun sélectionné à l'avance.

References

Graybill, F. A. (1969). *Introduction to Matrices with Applications in Statistics.* Belmont: Wadsworth.

Lloyd, F. T. and Hafley, W. L. (1977). Precision and the probability of misclassification in site index estimation. *Forest Science* **23,** 493–499.

Rao, C. R. (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol.* I, L. M. LeCam and J. Neyman (eds), 355–372. Berkeley: University of California Press.

Trousdell, K. B., Beck, D. E. and Lloyd, F. T. (1974). *Site Index for Loblolly Pine in the Atlantic Coastal Plain of the Carolinas and Virginia.* Research Paper SE-115. Asheville, North Carolina: Southeastern Forest Experiment Station, U.S.D.A. Forest Service.

Received February 1981; *revised August* 1981

Appendix

If from (5) we let

$$\begin{bmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{X}_2 \\ \mathbf{X}_2'\mathbf{X}_1 & \mathbf{X}_2'\mathbf{X}_2 \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}_{11}\mathbf{A}_{12} \\ \mathbf{A}_{21}\mathbf{A}_{22} \end{bmatrix},$$

then, from Graybill (1969, p. 165),

$$\mathbf{A}_{11} = (\mathbf{X}_1'\mathbf{X}_1)^{-1} + (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2(\mathbf{X}_2'\mathbf{P}\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1},$$
$$\mathbf{A}_{12} = -(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2(\mathbf{X}_2'\mathbf{P}\mathbf{X}_2)^{-1},$$
$$\mathbf{A}_{21} = -(\mathbf{X}_2'\mathbf{P}\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1},$$

and

$$\mathbf{A}_{22} = (\mathbf{X}_2'\mathbf{P}\mathbf{X}_2)^{-1},$$

where $P = \mathbf{I}_{vn} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'$. The least squares estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{A}_{21}\mathbf{X}_1' + \mathbf{A}_{22}\mathbf{X}_2')\mathbf{Y} \\ &= (\mathbf{X}_2'\mathbf{P}\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{P}\mathbf{Y}, \qquad \text{(A.1)}\end{aligned}$$

and

$$\begin{aligned}\hat{\boldsymbol{\gamma}} &= (\mathbf{A}_{11}\mathbf{X}_1' + \mathbf{A}_{12}\mathbf{X}_2')\mathbf{Y} \\ &= (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'(\mathbf{Y} - \mathbf{X}_2\hat{\boldsymbol{\beta}}). \qquad \text{(A.2)}\end{aligned}$$

Substitution of the direct product expressions of $\mathbf{X}_1$ and $\mathbf{X}_2$ from (3) and (4) yields estimators expressed in the form of (6).

Reprinted from
BIOMETRICS COPYRIGHT © 1982
THE BIOMETRIC SOCIETY, Vol. 38, No. 2, June 1982