THE ANALYSIS OF TREE RING CHRONOLOGIES USING A MIXED LINEAR MODEL

O. Brian Allen  
Daniel A.J. Ryan  
David L. McLaughlin

Follow this and additional works at: http://newprairiepress.org/agstatconference

Part of the Agriculture Commons, and the Applied Statistics Commons

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License.

Recommended Citation


This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.
THE ANALYSIS OF TREE RING CHRONOLOGIES USING A MIXED LINEAR MODEL

O. Brian Allen and Daniel A.J. Ryan
Department of Mathematics and Statistics
University of Guelph

David L. McLaughlin
Ontario Ministry of the Environment

ABSTRACT

The analysis of a tree’s annual growth rings can provide a great deal of information about the environment in which the tree has grown. In this paper we propose statistical methodology for analysing the incremental growth of sugar maple sampled throughout southern and central Ontario, by the Ontario Ministry of the Environment. Two trees, ranging in age from 75 to 150 years, were sampled from each of 42 stands in 6 regions. The data were analysed using a mixed linear model, incorporating age of tree, region, year, a year by region interaction and average monthly air temperature and total seasonal precipitation for the current year and the previous year, as fixed effects. Stand and tree were regarded as random effects and the repeated annual growth measurements on a tree were assumed to follow a first order autoregressive process.

Keywords: specific volume increment, dendrochronology, mixed linear model, variance component, autoregressive process.

1. INTRODUCTION

The growth rings of a tree provide a historical record of the growth and general health of the tree each year since it began growing. The width of the rings are determined by the age of the tree, the tree’s genotype and the environment in which it has grown. This relationship between the ring widths and the environment has been exploited in a number of ways. By matching the ring widths from trees which grew in different but overlapping eras (referred to as crossdating), scientists have been able to develop an unbroken chronology of ring widths extending back thousands of years (Ferguson, 1969). This allows the investigator to make historical predictions about a growth limiting aspect of the environment (eg: rainfall) of the region in which these trees grew (Monserud, 1986). The chronology of ring widths can also be related to a collection of measured environmental factors (eg: precipitation, temperature, insect infestations, atmospheric pollutants), in an attempt to determine which factors are important in controlling growth or health of the tree (Bauce and Allen, 1991). The study methodology currently in the literature is generally restricted to even aged stands.
Since ring widths are also affected by the age of the tree, in order to look at the relationship between growth and environment it is usually necessary to adjust for age of the tree when the ring was produced. There have been a large number of proposed ways to standardize chronologies. These include fitting a growth function such as exponential (Fritts, 1976; Sundberg, 1974), gamma (Monserud, 1986), Weibull (Yang et al., 1978), reciprocal (Zahner et al., 1989), or a compound growth function (Warren and LeBlanc, 1990). The growth curve is either incorporated into a larger model involving the environmental factors of interest or the residuals from the growth curve model are fitted to the environmental factors. Other methods include such data driven techniques as taking the first differences of the log ring widths (Van Deusen, 1987), fitting a polynomial (Fritts, 1976; Van Deusen, 1991), or fitting a cubic smoothing spline (Cook and Peters, 1981). It is clear that there is no generally accepted method of removing growth effects due to age. Consequently, there is an ever present danger that the technique chosen for removing age related growth effects will either remove the environmental trend which is of interest or will not adequately remove the age trend (Cook, 1987; Fritts, 1976, pp. 267-268; Innes and Cook, 1989; Zahner et al., 1989). The ring width is often converted to a ring width index by dividing the ring width for each year by its predicted width, based on the growth model chosen. This is done, not only to remove the growth trend, but also to stabilize the error variance (Fritts, 1976).

It is widely recognized that the chronology of ring widths from a tree will be correlated. Fritts (1976) presents a variety of filtering techniques for dealing with this correlation. Monserud (1986), however, cautions that filtering a series can introduce spurious correlations into the series. When regressing growth on environmental variables, Fritts (1976, pp 344) recommended that prior growth at lags 1, 2 and 3 years be included as regression variables in order to deal with autoregressive errors. Cook (1987), assuming an AR(3) series, prewhitened the series prior to examining the relationship with climatic variables. More generally, the autocorrelation can be modelled as a ARMA process. Monserud (1986), after analysing 33 actual tree ring series, reported that the overall best representation is an ARMA(1,1) process.

The relationship between growth and climatic factors has most often been examined by regressing ring width on selected climate variables. In an attempt to reduce the number of regressors, principal components regression has been recommended (Fritts, 1976). In dynamic regression models, introduced by Visser (1986), the regression parameters, \( \alpha_t \), are assumed to vary over the chronology according to the first order Markov process

\[
\alpha_t = G_t \alpha_{t-1} + w_t
\]

where \( G_t \) is a known square matrix and \( w_t \) is an error vector with expectation 0 and variance \( W_t \). These parameters are estimated using the Kalman filter (Meinhold and Singpurwalla, 1983). A limitation of this procedure is that the matrices \( G_t \) and \( W_t \) need to be specified. Van Deusen (1989) recognized that modelling of growth,
autocorrelation and environmental variables should be incorporated into a single model, fit by generally accepted techniques such as maximum likelihood.

The width of the growth rings of a tree can be determined non-destructively, by removing one or more cores from the tree. However, this measure of growth has its limitations. Ring widths vary along the trunk of the tree. Under adverse growing conditions, the ring associated with a particular year may be missing altogether at the point on the trunk at which the core was taken. This could cause all older rings to be associated with the wrong year. In addition, if the core misses the center of the tree, the information from the oldest rings will not be available. For these reasons, a more comprehensive measure of growth has been developed. Termed the specific volume increment (SVI), it is an estimate of the ratio of the annual volume of wood produced and the cambial surface area which produced the volume (Duff and Nolan, 1957; Shea and Armson, 1972). The SVI requires cutting the tree and obtaining disks at intervals along the entire vertical stem of the tree.

2. SAMPLING DESIGN

In a recent study, the Ontario Ministry of the Environment performed a detailed stem analysis on the two co-dominant sugar maple trees from a plot in each of forty-two stands from six regions in southern or central Ontario. Five plots were chosen from each of the southern regions (St. Williams, Barrie, Peterborough) and nine plots from each of the northern regions (Ottawa, Algonquin Park and Sault Ste. Marie). The stands were chosen to be remote from point sources of air pollution and urban centers and to have an average age of tree of 75 to 150 years. For each tree, the SVI was determined for each year of growth. For the southern sites, the SVI of the two trees from each plot were combined by year, resulting in one chronology per plot. The individual chronologies were not available.

Mean monthly temperatures and total seasonal precipitation were obtained from the Canadian Climate Center, Atmospheric Environment Service for several weather stations in the immediate vicinity of the plots in each region, for each the years for which an SVI was recorded in the region. These data were not complete for any weather station, however. The missing data were imputed using neighbouring weather stations and other years. The following model was fit to the existing data from the weather stations in a region.

\[ w_{ij} = \mu + s_i + y_j + \epsilon_{ij} \]

where \( w_{ij} \) is the weather record for station \( i \) in year \( j \), \( s_i \) is the station effect and \( y_j \) is the year effect. A complete weather record is then constructed by estimating \( \mu + s_i + y_j \) for each year \( j \) where \( s_i \) is the average of the station effects. When records are available from all stations for a particular year, the above estimate results in the average of the records for that year from all stations. For the analyses which follow, only data after 1915 were used, since weather data before this time were largely missing.
3. THE STATISTICAL MODEL

Many factors may have affected the SVI produced by a tree in a given year: the region, the plot, the tree’s age, weather conditions in the year or the year before, soil fertility, insect and disease infestations, harvesting activities and atmospheric deposition. The last factor was not measured but has generally increased since about 1940 but more so in the south than in the north (Tang et al. 1986). The six regions form a gradient in amount of deposition, the highest being in the south and declining as one moves north and east.

The model fit to these data was

\[ Y_{ijk} = X_{ijk} \alpha + \kappa_i + \tau_{ij} + \epsilon_{ijk} \]

where \( i=1,...,42 \) (the number of plots), \( j=1,...,t_i \) (\( t_i \) is the number of trees in the \( i^{th} \) plot), \( k=1,...,r_{ij} \) (\( r_{ij} \) is the number of rings from the \( j^{th} \) tree in the \( i^{th} \) plot). \( \alpha \) is the 126x1 parameter vector associated with the fixed effects (parameterized in such a way that all elements are estimable). This includes 5 age of tree classes, 15 year classes, 6 areas, the year x area interaction, average monthly air temperature for each month of the year and the year previous and total precipitation for each season of the year and the year previous. \( \kappa_i, \tau_{ij} \) and \( \epsilon_{ijk} \) are the random effects with zero means corresponding to plot, tree and rings. It is assumed that \( \text{Var}(\kappa_i) = \sigma_{\kappa}^2 \), \( \text{Var}(\tau_{ij}) = \sigma_{\tau}^2 \), and \( \text{Var}(\epsilon_{ijk}) = \sigma_{\epsilon}^2 \).

Further, we assume that the \( \epsilon_{ijk} \), \( k=1,...,r_{ij} \), follow a first order autoregressive process; that is, \( \text{Cov}(\epsilon_{ijk}, \epsilon_{ijk'}) = \sigma_{\epsilon}^2 \rho \left| k-k' \right| \). The remaining random effects are assumed independent. Thus

\[ \text{Cov}(y_{ijk}, y_{ijk'}) = \sigma_{\kappa}^2 + \sigma_{\tau}^2 + \sigma_{\epsilon}^2 \left| k-k' \right|, \]

and

\[ \text{Cov}(y_{ijk}, y_{ij'k'}) = \sigma_{\kappa}^2, \text{ if } j \neq j', \]

\[ \text{Cov}(y_{ijk}, y_{i'j'k'}) = 0, \text{ if } i \neq i'. \]

Then, the covariance matrix of \( y \) (whose elements are the \( y_{ijk} \)) is

\[ V = \bigoplus_i \bigoplus_j \{ \text{Cov}(y_i) \} = \bigoplus_i \bigoplus_j \left( \sigma_{\kappa}^2 S_{ij} + \sigma_{\tau}^2 J_{ij} J'_{ij} \right) + \bigoplus_i \left( \sigma_{\epsilon}^2 \left( J_{ij} J'_{ij} \right) \right), \]

\[ = \sigma_{\epsilon}^2 V_1 + \sigma_{\tau}^2 V_2 + \sigma_{\kappa}^2 V_3 \]

where \( \bigoplus_i \bigoplus_j \) represents the matrix direct sum over \( i, j \) (Searle 1982),
\( J_{n_1}, J_{n_2} \) are \((n_1 \times 1)\) and \((n_2 \times 1)\) vectors of 1's and

\[
S_{ij} = \begin{pmatrix}
1 & \rho & \ldots & \rho^{n_{ij}-1} \\
\rho & 1 & \ldots & \rho^{n_{ij}-2} \\
\vdots & \vdots & \ddots & \vdots \\
\rho^{n_{ij}-1} & \rho^{n_{ij}-2} & \ldots & 1
\end{pmatrix}
\]

4. FITTING THE STATISTICAL MODEL

The fixed effects in the model were estimated using the generalized least squares or maximum likelihood estimator

\[
\hat{a} = (X' V^{-1} X)^{-1} X' V^{-1} y
\]  \hspace{1cm} (1)

where the \( x_{ijk} \) form the rows of \( X \) and the \( y_{ijk} \) form the rows of \( y \). The variance parameters were estimated by maximizing the residual likelihood

\[
-\frac{1}{2} \left[ \log(\det(V)) + \log(\det(X' V^{-1} X)) + (y' V^{-1} M y) \right]
\]  \hspace{1cm} (2)

where

\[
M = I - X (X' V^{-1} X)^{-1} X' V^{-1}
\]

This is done iteratively in two steps. \( V \) is linear in the variance components \( \rho \)

\[= (\sigma^2_{e}, \sigma^2_{c}, \sigma^2) \]. Thus for the current value of \( \rho \), the variance components are estimated by
the Anderson equations

\[ Zp = W \]  

(3)

where \( Z \) is the 3 x 3 matrix whose \((g,h)\)th element is

\[ z_{gh} = \text{trace} \left( V^{-1} M V_g M' V^{-1} V_h \right), \]  

(4)

\( w \) is the 3 x 1 vector with \(g\)th element

\[ w_g = y' V^{-1} M V_g M' V^{-1} y \]  

(5)

Using the current values of \( p \), the value of \( p \) which maximizes the residual likelihood is determined by direct search. We iterate on these two steps until the change in the residual likelihood is less than \(10^{-5}\). The fixed effects are then estimated using the estimator above. The data set is large, with 75 to 150 SVI measurements from each of the 84 trees sampled. Consequently, substantial algebraic simplification was necessary before the model could be fitted.

The elements of \( V^{-1} \) can be obtained analytically, in terms of \( \sigma^2_k, \sigma^2_t, \sigma^2 \) and \( p \), using the identity (Harville, 1977)

\[ V^{-1} = R^{-1} - R^{-1} Z D (I + Z' R^{-1} Z D)^{-1} Z' R^{-1}, \]

where \( R = \sigma^2 \bigoplus_1^p \bigoplus_1^n S_{ii}, \) \( D = (\sigma^2_k I_a) \bigoplus (\sigma^2_t I_b), \) \( Z = (Z_1 \mid Z_2) \),

\[ Z_1 = \bigoplus_1^n J_n, \quad Z_2 = \bigoplus_1^p \bigoplus_1^n J_{n,i} \] and \( I_n \) the identity matrix of dimension \( n \). The matrix \( R \) can be inverted analytically since

\[ S_{ij}^{-1} = (1-p^2)^{-1} \]

Furthermore, \((I + Z' R^{-1} Z D)\) can be inverted analytically using the form for the inverse of a partitioned matrix (Searle 1982, page 258). \( Z_{gh} \), given in (4), can be rewritten as
trace(V_{-1}V_{g}V_{-1}V_{h}) - 2 \text{trace}[(X'V^{-1}X)^{-1}(X'V^{-1}V_{g}V^{-1}V_{h}V^{-1}X)]

+ \text{trace}[(X'V^{-1}X)^{-1}(X'V^{-1}V_{g}V^{-1}X)(X'V^{-1}X)^{-1}(X'V^{-1}V_{h}V^{-1}X)],

and \( w_{g} \), given in (5), can be rewritten as

\[ y'V_{g}V^{-1}V_{g}V^{-1}y - 2y'V_{g}V^{-1}Xa' + a'(X'V^{-1}V_{g}V^{-1}X)a. \]

The third term of the residual likelihood, (2), can be rewritten as

\[ y'V^{-1}y - a'(X'V^{-1}X)a. \]

Thus, expressions (1), (4), (5) and the latter two terms of (2) can be written in terms of the matrix expressions \((X'V^{-1}X),(X'V^{-1}V_{g}V^{-1}y),(y'V^{-1}y),(y'V^{-1}V_{g}V^{-1}y),(y'V^{-1}V_{g}V^{-1}X),(X'V^{-1}V_{g}V^{-1}X)\), and \((X'V^{-1}V_{g}V^{-1}V_{h}V^{-1}X)\). These expressions are of dimension \( q \times q \) or \( q \times 1 \) and thus are easily numerically manipulated on the computer. For each of these expressions, its elements can be written algebraically in terms of \( \sigma_{g}, \sigma_{r}, \sigma^{2}, \rho \) and the elements of \( X \) and \( y \). They are not reproduced here, in the interest of space.

Finally, from Harville (1977), equation (5.1),

\[ \text{det}(V) = \text{det}(R).\text{det}(I + Z'R^{1}D). \]

Now

\[ \text{det}(S_{ij}) = (1 - \rho^{2})^{m_{ij}-1} \]

and \((I + Z'R^{1}D)\) can be written

\[
\begin{pmatrix}
  b_{11} & b_{12} \\
  b_{21} & b_{22}
\end{pmatrix}
\]

where each of the submatrices has a diagonal form. Then

\[ \text{det}(I + Z'R^{1}D) = \text{det}(b_{22}).\text{det}(b_{11} - b_{12}b_{22}^{-1}b_{21}) \]

(Searle, 1982, page 258). The two determinants on the right are easily evaluated algebraically since they are of diagonal matrices.

Finally, with these algebraic simplifications, expressions (1) through (5) may be evaluated numerically, using any software which allows standard matrix operations.
In fact, in the southern regions, the SVI of the two trees sampled from each plot were combined by year. In principle, the model can be adjusted to account for the fact that each observation is really an average of two SVI. If the ages associated with the two SVI fall in different age classes, the relevant row of \( x_{ijk} \) will contain a 0.5 in the two columns associated with the age classes rather than a single 1.0. Furthermore, in the covariance matrix of the responses for the southern regions, \( \sigma^2 \) is replaced by 0.5 \( \sigma^2 \) and \( \sigma^2 \) is replaced by 0.5 \( \sigma^2 \). For computational reasons these modifications were not implemented. However, in most cases the two age classes of the combined SVI would be the same, in any case. Furthermore, failing to adjust the covariance matrix will affect the precision of the estimates of the fixed effects but will not produce biased estimates.

SAS Institute Inc. has, in version 6.07 of SAS/STAT®, included the procedure PROC MIXED. The data were subsequently fit to the model given above, using this procedure, although it was not available when the data were initially analysed.

Results based on the fitting of the model given in this paper will appear elsewhere.

SAS/STAT® is a registered trademark of SAS Institute Inc.

5. REFERENCES


Harville, D.A. 1977. Maximum likelihood approaches to variance component


