RESEARCH REPORT

INTERPRETATION OF CROSS CORRELATION BETWEEN TREE-RING SERIES

DAVID K. YAMAGUCHI

U.S. Geological Survey
David A. Johnston Cascades Volcano Observatory
5400 MacArthur Blvd.
Vancouver, Washington 98661

ABSTRACT

Correlation analysis assumes that individual observations are statistically independent. Since tree-ring indices are typically serially correlated, cross-correlation coefficients computed between standardized tree-ring series may be spurious and inflated. To obtain valid estimates of these coefficients, ARIMA time series models should be fit to standardized series before cross-correlation analysis. ARIMA modelling was used successfully to obtain an unambiguous match between a "floating" series and a master chronology using program CROS.

INTRODUCTION

In recent years, many computer programs for analyzing cross correlation have been developed specifically for use in dendrochronologic research. These programs fall into 3 categories: (1) crossdating programs that aid traditional skeleton plotting (Stokes and Smiley 1968) in assigning calendar dates to floating chronologies (Fritts 1963; Scott 1966; Parker 1967, 1971; Baillie and Pilcher 1973; Munro 1984); (2) programs for data quality control (Wendland 1975; Holmes 1983; Kickert et al. 1983); and (3) general-purpose correlation programs used to examine changes in cross correlation between chronologies over time (Stockton and Fritts 1971; Fritts 1976: 294) or distance (Fritts 1963; Cropper and Fritts 1982).

While these programs are useful research tools, care must be exercised when interpreting cross-correlation coefficients they compute, because tree-ring indices are typically serially correlated (Rose 1983; Monserud 1986; Landwehr and Matalas 1986). Cross-correlation coefficients computed between serially-correlated series may be spurious and inflated (Jenkins and Watts 1968: 338; Box and Newbold 1971; Chatfield 1975: 173; Chiu *et al.* 1981; Monserud 1986). This paper illustrates the problem by using a crossdating program (CROS; Baillie and Pilcher 1973) to date a floating chronology from the Pacific Northwest region of North America. CROSS was selected for illustrative purposes because of its simplicity; however, the concepts presented in this example apply equally to other dendrochronologic computer programs that employ cross correlation.

CROS

CROS "slides" one floating standardized tree-ring series past another standardized (master) series and calculates correlation coefficients, or more correctly, cross-correlation coefficients, between the two series at all possible positions of significant overlap. In short, it computes the standard cross-correlation function relating the two series. The output of CROS consists of a sequential listing of Student's t values, calculated for non-negative values of the standard cross-correlation function. The Student's t values permit interpretation of the cross-correlation coefficients. At each

potential "bark" date (date of the last ring in the floating series), the t value is simply the cross-correlation coefficient divided by its standard error:

$$t = r_{xy}(k)/s_{r_{xy}}(k), \tag{1}$$

where $r_{xy}(k)$ is the cross-correlation coefficient relating series x and y at lag k. The standard error of $r_{xy}(k)$ is calculated as:

$$s_{r_{xy}}(k) = \sqrt{\frac{1 - [r_{xy}(k)]^2}{N - 2}} \tag{2}$$

where N is the number of years of overlap shared by both series at that position (Zar 1984: 309). Under the assumption of serial independence within series x and y, the significance levels of the t values can be read from Student's t tables at $(N - 2)$ degrees of freedom (Zar 1984: 309). Generally, positions having t values greater than 3.2 and over 70 years of overlap are significant at the 0.001 level, and are considered likely matching positions.

Missing and intra-annual rings must be absent from the two series being compared because CROS treats both as continuous input data. While the latter requirement makes CROS impractical for many areas of the world (Holmes 1983), it is well suited for applications in western Oregon and Washington, where missing and intra-annual rings are rare.

DATA

In 1984, several logs were found imbedded in a pyroclastic-flow deposit west in the valley of the South Fork Toutle River of Mount St. Helens, Washington. Crossdating of a Douglas-fir (*Pseudotsuga menziesii*) stump rooted on the surface of this deposit with local ring patterns (Yamaguchi 1983, 1985) indicates that this flow was emplaced before A.D. 1668. Further, it is known that this deposit cannot have been emplaced before A.D. 1482, because it overlies a Mount St. Helens tephra set known to have fallen after this date from dendrochronologic dating of an older tephra east of the volcano (Yamaguchi 1985).

A disk (no. SFT-349) was cut from one of the Douglas-fir logs exhumed from the pyroclastic flow to see if its date of death could be determined by crossdating. The sampled log retains much of its bark and shows little evidence of pre-burial wood decay. Thus its date of death should coincide with or slightly antedate the eruption of Mount St. Helens (*ca.* A.D. 1668) that produced this pyroclastic flow.

Sample SFT-349 could not be dated by visual crossdating or skeleton plotting, however, because waterlogging had reduced its visible latewood-density variation, and its rings were relatively complacent. Computerized crossdating was therefore attempted using CROS. The 290 rings of this sample were measured to the nearest 0.01 mm, and standardized by fitting a negative-exponential curve to the data (Fritts 1976; Graybill 1979, 1982). A modified version of CROS, adapted to accept standardized data, was then used to cross correlate this series with an index chronology developed from 15 Douglas-fir at a site (Lava Beds) 32 km southeast of Mount St. Helens (length = 579 years; Brubaker 1980). No attempt was made to limit CROS to

the interval of known possible dates (A.D. 1482-1668), in order to independently test its ability to distinguish potential matching positions from non-matching ones across a long master chronology. Before cross correlation, both series were tested for normality following Zar (1984) to ensure that the indices in at least one series are normally distributed, as required, for using the t statistic to evaluate correlation coefficients (Zar 1984: 311).

CROS RESULTS USING RING-WIDTH INDICES

CROS identified 113 bark dates for sample SFT-349 as being "significant at the 0.001 level" over the A.D. 1411-2240 time interval. The future dates were obtained when SFT-349 was lagged positively against the master chronology. Minimum amounts of overlap for "significant" dates at the beginning and end of the master chronology were 87 and 75 years, respectively. Student's t values for these dates are shown in Figure 1.

These results suggest that 113 of the 830 potential bark dates examined are likely matching positions. Clearly, 113 is too large a number of likely bark dates because less than one likely date should have arisen by chance in a time interval of this length, and only one can be the correct date. Several additional features are evident in these results. The future bark dates (A.D. 2078-2195) identified by this analysis are obviously in error. Further, the 23 dates that occur during the A.D. 1668-1771 interval are inconsistent with stratigraphic evidence for an earlier date of death for tree SFT-349. Where did CROS go awry?
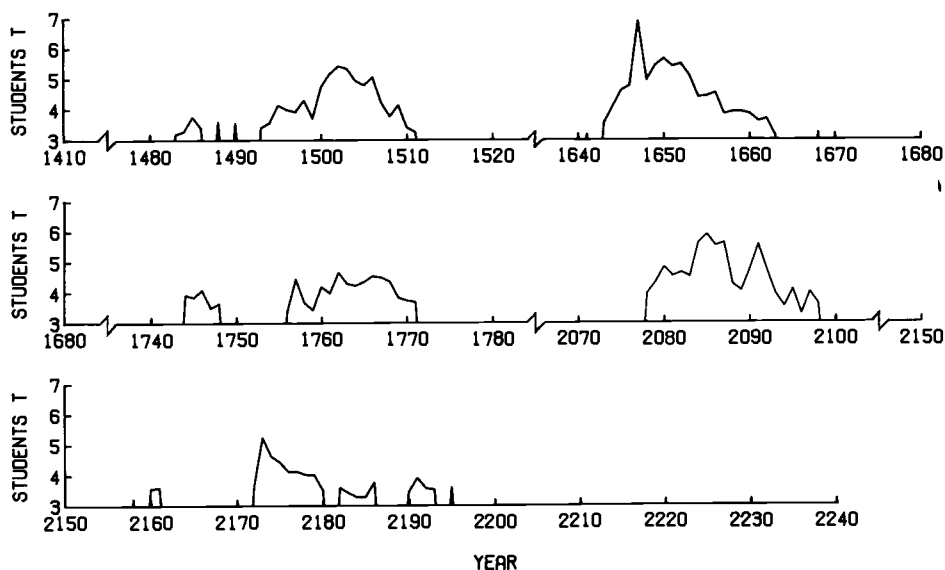

Figure 1. Student's t values for sample SFT-349 bark dates identified by CROS as "significant at the 0.001 level" when ring-width indices were used as input data.

To obtain the spurious and inflated "significant" Student's t values in Figure 1, the cross-correlation coefficients calculated by CROS must have been inflated, because N in equation (1) is fixed for each cross-correlated position. This example illustrates the type of error that occurs when tree-ring series being cross-correlated violate the

assumption of serial independence. Only by cross-correlating autocorrelated series could the "significant" Student's t values in Figure 1 rise and fall in clusters over time, a pattern indicative of similar levels of cross correlation at adjoining matching positions.

The degree of serial correlation in data from sample SFT-349 is shown in a correlogram, which plots sample autocorrelation coefficients at increasing lags for its ring-width and index series (Figure 2). The correlogram of the Lava Beds site index chronology is virtually identical to that of Figure 2 (Series B).
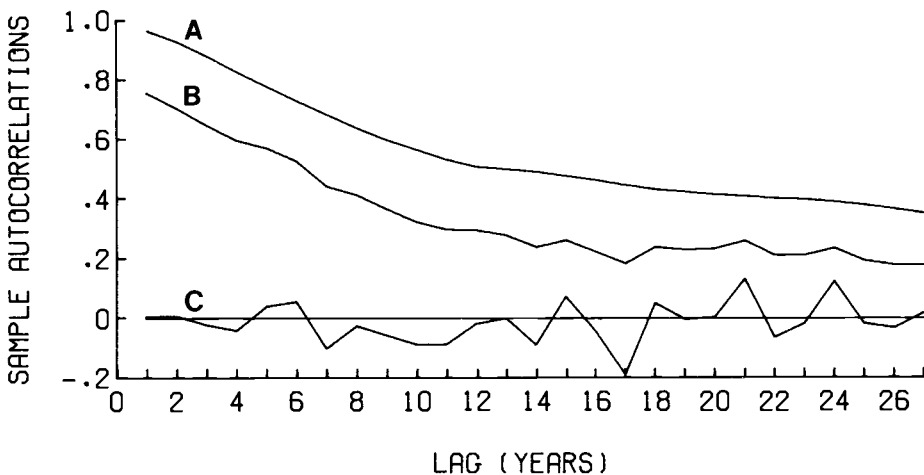

Figure 2. Correlograms for time series from sample SFT-349 (A) ring widths; (B) ring-width indices; (C) ARIMA (0, 1, 1) residuals.

Comparison of Series A and B in Figure 2 shows how standardizing reduces the amount of autocorrelation in a ring-width series as described by Fritts (1976). However, Figure 2 also shows that the ring-width indices of sample SFT-349 retain significant autocorrelation after standardization. To date, most dendrochronologists have given such remaining autocorrelation only limited attention when evaluating cross correlation relating tree-ring series (Fritts 1976: 324; Cropper and Fritts 1982; Munro 1984).

One way to circumvent this problem is to fit autoregressive intergrated moving average (ARIMA) models (Chatfield 1975; Box and Jenkins 1976; O'Donovan 1983) to standardized tree-ring series to remove autocorrelation from them, or "prewhiten" them, before cross correlation. This approach is well established (Jenkins and Watts 1968: 340; Chatfield 1975: 173; Haugh 1976; Haugh and Box 1977; Granger and Newbold 1977: 232; Pierce and Huagh 1977; Chiu et al. 1981; Monserud 1986). Its application to dendrochronologic cross correlation is demonstrated below.

CROS RESULTS AFTER ARIMA MODELING

ARIMA time series models were individually fit to the standardized SFT-349 series and the Lava Beds site chronology using Minitab (Ryan et al. 1982, 1985), following the procedures outlined by Chatfield (1975), Box and Jenkins (1976), and O'Donovan (1983).

The best-fitting models for both series are ARIMA (0, 1, 1) models, i.e., first-order moving average models fit to first-order differenced data. These are most commonly expressed as:

$$X_t - X_{t-1} = Z_t - \theta Z_{t-1}, \qquad (3)$$

where X_t is the ring-width index for year t, 0 is a constant estimated from individual series, and Z_t is the stochastic error at year t. Note that the left side of (3) is the first-order differencing model of Fritts (1976). A more meaningful form of (3) is:

$$X_t = (1 - \theta)X_{t-1} + P(1 - \theta)X_{t-2} + \theta^2(1 - \theta)X_{t-3} + \ldots + Z_t, \quad (4)$$

which states that individual ring-width indices can be modeled as the weighted sum of preceding indices. For example, the Minitab-estimated θ value for sample SFT-349's ring-width indices is 0.495. Substituting this value into (4), a weight of about 0.1 is obtained for this series after 3 years. This indicates that after differencing to transform the data into a stationary series, this tree retained a "memory" of about 3 years; i.e., its growth during any given year was serially correlated with its growth during the preceding 3 years. The effectiveness of ARIMA model-fitting in removing the "memory" from this standardized series can be seen in the correlogram of its ARIMA residual series, which are the differences between observed and modeled indices (i.e., the Z_t in equation (4); Figure 2, series C).

CROS was then run between these residuals and those of the Lava Beds chronology. A single bark date — A.D. 1647 — was found to have a t value significant at the 0.001 level (t = 5.05, compared to its prior inflated value of 6.91). This date is consistent with stratigraphic bracketing of the date of death of tree SFT-349. Further, when this date is assigned to this tree's last-formed ring, it is supported by both graphical (Figure 3) and visual crossdating.

DISCUSSION

This example illustrates that spurious and inflated cross-correlation coefficients arise when they are computed between autocorrelated tree-ring series. Autocorrelation is a common feature of tree-ring data from most regions (Rose 1983; Monserud 1986; Landwehr and Matalas 1986). Tree-ring studies whose conclusions rest on "significant" cross-correlation coefficients are therefore suspect. One example is the extensive use of CROS to date floating oak chronologies in western Europe (Baillie et al. 1985), because chronologies from this region show strong autocorrelation (Munro 1984). To illustrate, the Scotland oak chronology (Baillie 1977) has a first-order autocorrelation coefficient of 0.544, yet has been cross correlated with many floating chronologies (Baillie et al. 1985).

The present study also shows that fitting ARIMA models to tree-ring series improves the effectiveness of their analysis by cross correlation. This approach is theoretically sounder than high-pass filtering approach of Munro (1984) (see Monserud 1986) and can be implemented just as easily. Further, it is preferable to the traditional method of adjusting for autocorrelation after cross correlation by substituting a lower "effective sample size" for N in equation (1) (Fritts 1976; Cropper and Fritts 1982). The latter method is not recommended on statistical grounds because it makes assumptions about the type of autocorrelation present in series
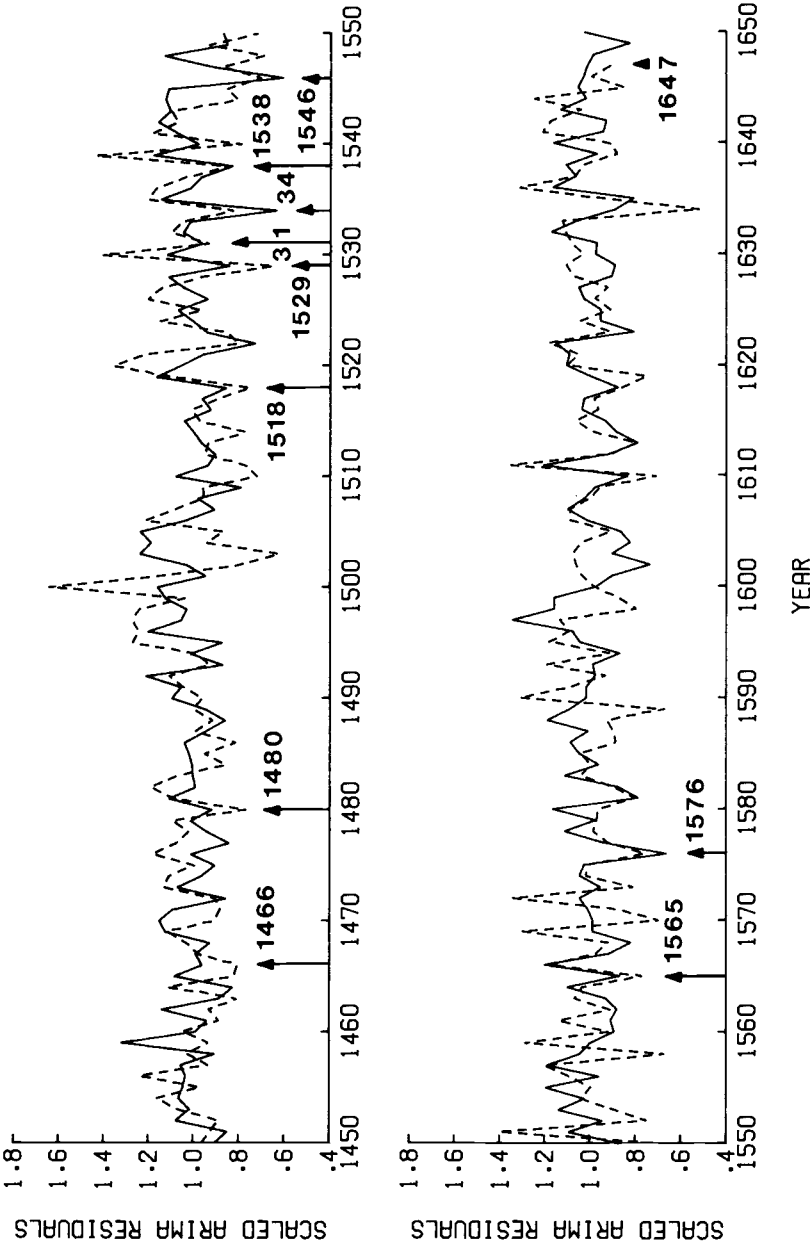
Figure 3. Crossdating between SFT-349 ARIMA residuals (dashed line) and those of the Lava Beds site chronology (solid line) at A.D. 1647. Residuals have been scaled for comparative purposes. Labeled narrow rings are distinctive features of local Douglas-fir ring patterns (Yamaguchi 1983, 1985).

(Mitchell 1963; Mitchell et al. 1966). Further, calculation of the "effective sample size" relies on sample autocorrelation coefficients, which themselves are subject to sampling error (P. J. Bartlein, written commun. 1985). These facts argue for adopting the ARIMA "prewhitening" approach for future applications of cross correlation in dendrochronology.

ACKNOWLEDGEMENTS

REFERENCES

Baillie, M. G. L.
    1977   An oak chronology for south central Scotland. *Tree-Ring Bulletin* 37: 33-44.
Baillie, M. G. L. and J. R. Pilcher
    1973   A simple crossdating program for tree-ring research. *Tree-Ring Bulletin* 33: 7-14.
Baillie, M. G. L., J. Hillam, K. R. Briffa, and D. M. Brown
    1985   Re-dating the English art-historical tree-ring chronologies. *Nature* 315: 317-319.
Box, G. E. P. and P. Newbold
    1971   Some comments on a paper of Coen, Gomme and Kendall. *Journal of the Royal Statistical Society, Series A* 134: 229-240.
Box, G. E. P. and G. M. Jenkins
    1976   *Time Series Analysis, Forecasting and Control.* Revised edition, Holden-Day, San Francisco.
Brubaker, L. B.
    1980   Spatial patterns of tree growth anomalies in the Pacific Northwest. *Ecology* 61: 798-807.
Chatfield, C.
    1975   *The analysis of time series.* Chapman and Hall, London.
Chiu, W., A. Lo, and D. H. Weidler, Jr.
    1981   A study of the possible statistical relationship between the tropical Pacific sea surface temperature and atmospheric circulation. *Monthly Weather Review* 109: 1013-1020.
Cropper, J. P. and H. C. Fritts
    1982   Density of tree-ring grids in western North America. *Tree-Ring Bulletin* 42: 3-9.
Fritts, H. C.
    1963   Computer programs for tree-ring research. *Tree-Ring Bulletin* 25: 2-7.
    1976   *Tree Rings and Climate.* Academic Press, New York.
Granger, C. W. J. and P. Newbold
    1977   *Forecasting Economic Time Series.* Academic Press, New York.
Graybill, D. A.
    1979   Revised computer programs for tree-ring research. *Tree-Ring Bulletin* 39: 77-82.
    1982   Chronology development and analysis. In *Climate From Tree Rings,* edited by M. K. Hughes, P. M. Kelly, J. R. Pilcher, and V. C. LaMarche, Jr., pp. 21-28. Cambridge University Press, Cambridge.
Haugh, L. D.
    1976   Checking the independence of two covariance-stationary time series: a univariate residual cross-correlation approach. *Journal of the American Statistical Association* 71: 378-385.
Haugh, L. D. and G. E. P. Box
    1977   Identification of dynamic regression (distributed lag) models connecting two time series. *Journal of the American Statistical Association* 72: 121-130.
Holmes, R. L.
    1983   Computer-assisted quality control in tree-ring dating and measurement. *Tree-Ring Bulletin* 43: 69-78.
Jenkins, G. M. and D. G. Watts
    1968   *Spectral Analysis and its Applications.* Holden-Day, San Francisco.
Kickert, R. N., B. Herren-Gemmill, R. Arkley, and R. E. Thomson
    1983   CURVES and ALIGN: some new computerized methods for the analysis of tree-ring data. *Tree-Ring Bulletin* 43: 79-88.
Landwehr, J. M. and N. C. Matalas
    1986   On the nature of persistence in dendrochronologic records with implications for hydrology. *Journal of Hydrology* 86: 239-277.

Mitchell, J. M., Jr.
    1963   Some practical considerations in the analysis of geophysical time series. Unpublished MS available from its author, National Oceanic and Atmospheric Administration, Silver Spring, Maryland.

Mitchell, J. M., Jr., B. Dzerdzeevskii, H. Flohn, W. L. Hofmeyr, H. H. Lamb, K. N. Rao, and C. C. Wallén
    1966   Climatic change. *World Meteorological Organization Technical Note* No. 79, WMO No. 195 TP100: 1-79.

Monserud, R. A.
    1986   Time series analyses of tree-ring chronologies. *Forest Science* 32: in press.

Munro, M. A. R.
    1984   An improved algorithm for crossdating tree-ring series. *Tree-Ring Bulletin* 44: 17-27.

O'Donovan, T. M.
    1983   *Short term forecasting.* John Wiley and Sons, New York.

Parker, M. L.
    1967   Dendrochronology of Point of Pines. M.A. thesis, University of Arizona, Tucson.
    1971   Dendrochronological techniques used by the Geological Survey of Canada. *Geological Survey of Canada Paper* 71-25.

Pierce, D. A. and L. D. Haugh
    1977   Causality in temporal systems: characterizations and a survey. Journal of Econometrics 5: 265-293.

Rose, M. R.
    1983   Time domain characteristics of tree-ring chronologies and eigenvector amplitude series from western North America. *Laboratory of Tree-Ring Research Technical Note No. 25,* University of Arizona, Tucson.

Ryan, T. A., Jr., B. L. Joiner, and B. F. Ryan
    1982   *Minitab 82 reference manual.* Minitab Project, Statistics Department, Pennsylvania State University, University Park, Pennsylvania.

Ryan, B. F., B. L. Joiner, and T. A. Ryan, Jr.
    1985   *Minitab handbook.* Second edition, Duxbury Press, Boston, Massachusetts.

Scott, S. D.
    1966   Dendrochronology in Mexico. *Papers of the Laboratory of Tree-Ring Research 2,* University of Arizona, Tucson.

Stockton, C. W. and H. C. Fritts
    1971   Conditional probability of occurrence for variations in climate based on width of annual tree rings in Arizona. *Tree-Ring Bulletin* 31: 3-24.

Stokes, M. A. and T. L. Smiley
    1968   *An introduction to tree-ring dating.* University of Chicago Press, Chicago.

Wendland, W. M.
    1975   An objective method to identify missing or false rings. *Tree-Ring Bulletin* 35: 41-47.

Yamaguchi, D. K.
    1983   New tree-ring dates for recent eruptions of Mount St. Helens. *Quaternary Research* 20: 246-250.
    1985   Tree-ring evidence for a two-year interval between recent prehistoric explosive eruptions of Mount St. Helens. *Geology* 13: 554-557.

Zar, J. H.
    1984   *Biostatistical analysis.* 2nd edition, Prentice-Hall, New Jersey.