AN APPROACH TO DENDROCLIMATOLOGY:

Screening by Means of Multiple Regression Techniques

Harold C. Fritts

Assistant Professor of Dendrochronology

Laboratory of Tree-Ring Research

University of Arizona, Tucson¹

New techniques have recently been developed for statistically evaluating multivariate atmospheric processes for weather prediction (Malone 1959). These techniques employ a multiple regression process to screen from many variables, those that are most significantly related to a particular dependent variable. For the selected variables, coefficients for a linear equation predicting the dependent variable are calculated. It is then possible to evaluate from the equation and the individual coefficients, the relative importance of the variables and their inter-relationships. It is the objective of this paper to illustrate an application of these techniques to the problem of establishing dendroclimatic indices.

The materials selected for this study were increment cores taken along 3 radii of each of 5 mature white oaks on a south-facing slope and 5 on a north-facing slope in a ravine forest 2 miles south of Charleston, Illinois. In most samples annual rings were present for the years 1901-1958, and these rings were cross-dated and measured to the nearest .001 inch. The widths of the earlywood and latewood were also determined. The data were reduced to average values for the 5 trees of each slope. Weather bureau records of average monthly temperature and monthly precipitation for Charleston, 1901-1958, were employed; and estimations of monthly soil moisture storage and monthly evapo-transpiration deficits were calculated according to Thornthwaite and Mather (1957).

---

¹Presented at the Tree-Ring Research Coordination Conference, February 3 and 4, 1961, Tucson.

Several stepwise multiple regression analyses (Ralston 1960) were undertaken on both tree groups, analyzing in separate problems the factors associated with variation of (1) earlywood, (2) latewood, and (3) total ring width. Each of these was subdivided into three problems using the different calculations of available moisture: (1) monthly precipitation, (2) monthly soil moisture storage, and (3) monthly evapo-transpiration deficit. Moisture values for the individual months from May through September of the year preceding the formation of the growth ring and from April through August of the concurrent year were used as 10 possible independent variables. Average temperatures for the same months were included as 10 more variables. Each problem had additional variables as follows: total soil moisture surplus of the winter and spring preceding growth, the number of the year (02-58), number of the year squared, number of the year cubed, number of the year to the fourth power, width of the earlywood of the preceding year, width of the latewood of the preceding year, and for analysis of latewood only--width of earlywood during the concurrent year. The year number as a power series was included to account for trends in the growth rings not related to environmental factors. The inclusion of preceding growth values as independent variable was found to be necessary to take into account the autocorrelation which characterized the tree-ring series.

The procedure for each step-wise regression analysis was to add one variable at a time to the regression equation (most significant ones first) and to test for improvement of fit by means of the F ratio (Ralston 1960). When the F ratio became 1.0 or less, no further variables were added, and the predicted and residual values for the final equation were calculated. If the significance for a given variable dropped below 1.0 after it had been included in the equation, the variable was dropped and a new equation was calculated before additional variables were

added. The F level of 1.0 was thought more appropriate than higher levels, as experience showed that due to intercorrelation and interaction, some variables which were entered with low F levels were very significant when compared with their standard errors, or became significant when a following variable was added. The F level of 1.0 was also found to approximate the point where the multiple correlation was highest.

The standardized regression coefficients for 6 of these final equations for the trees on the south-facing slope are graphically presented in Figures 1 and 2. The height and direction of each bar is proportional to the relative contribution of the variable to variation in the total ring. The coefficients of correlation, coefficients of determination, and standard errors of estimate for 12 of the most significant problems are given in Table I. The analyses incorporating soil moisture storage as the moisture variable were least successful in accounting for ring variation and are not included.

It is apparent from the figures that the factors related to variation in earlywood are not the same, nor do they act in the same manner, as those affecting the latewood. Earlywood is not only a function of the environment during early season but is highly related to conditions and growth of the previous season, while latewood is directly related to the earlywood and to environmental conditions in June and July of the current season. The major exception to this is an inhibiting effect of high moisture during May of the previous year. As one would expect, the analyses employing total ring width represent a synthesis of earlywood and latewood relationships. A consideration of the physiological implications of these data is beyond the scope of this paper but will be considered in a subsequent publication.

The percent variance accounted for (coefficient of determination) was highest when precipitation was used in analyzing earlywood and when evapo-transpiration

was used in analyzing latewood. Portions of the power series of year number were frequently included to remove trends apparent in the raw data. Median cross tests of the final residuals showed them to be randomly distributed throughout the 57 year period which indicates that the autocorrelation had been largely removed (Quenouille 1952) and the error estimates were appropriate.

If our objective is to establish a dendroclimatic index using growth rings in white oak, then it is evident from the above analyses that latewood most closely reflects the environment of the current June and July, but the influence of the earlywood on latewood and influence of trends should be removed. In order to simplify the equation, the variation in latewood was reanalyzed using the following as possible independent variables: x_1 - average temperature for June of the current year, x_2 - evapo-transpiration deficit for July of the current year, x_3 - the product of x_1 and x_2 , x_4 - year number, x_5 - year number squared, x_6 - year number cubed, x_7 - earlywood of the current year, and x_8 - the product of x_4 and x_7 . The best fitting equations were: Latewood for trees on the south-facing slope = $-.548x_1 - .287x_3 + 3.226x_6 + 2.090x_7 + 43.087$. Latewood for trees on the north-facing slope = $-.389x_1 - .178x_3 - 2.902x_4 + 8.059x_6 + 1.145x_7 + 38.462$. All coefficients were significant at the .99 level and most of them significant at the .999 level. The multiple correlations and standard errors of estimate are included under the heading "simplified equation" in Table I.

It should be noted that the stepwise multiple regression procedure not only gives an objective means of screening all possible variables, but also provides a means of establishing a simplified predictive equation with a minimum number of variables. The variance accounted for in this simple equation is only 5% less than the original more inclusive equation.

In the simplified equations previously given, the selection of x_3 instead of x_2 indicates that evapo-transpiration deficit is more limiting in July if temperatures

in June are high. The selection of x_4 and x_6 indicates that significant trends were in the original data but were removed by the analysis. Ordinarily when the tree ring series is sufficiently long, this can be accomplished beforehand by fitting a trend line and calculating an index (Schulman 1956).

The significant point here is that the above equations can be used to remove the effect of earlywood and trends on the variation in latewood thus enabling one to obtain an index which closely approximates the inhibiting effects of high temperatures in June and high evapo-transpiration deficits in July. In Figure 3 this procedure is illustrated: 1 and 2 are the average widths in latewood for the two groups of white oak, 3 and 4 are the indices with effect of trend and relation to earlywood removed, 6 and 7 are the environmental variables x_1 and x_3 and 5 is the sum of variables x_1 and x_3 inverted for comparison with the tree-ring indices. The high correlation between the index and the sum of the two environmental variables is evident. The only major discrepancy occurs during 1902-08, and this may have resulted from either interpolation of data for the several short cores which lacked rings, or from abnormally rapid growth at the turn of the century which was perhaps due to cutting and release in several of the younger trees.

In conclusion, the study suggests that promising dendroclimatological information may be obtained by the following procedures.

1. Sample cores from a number of trees of a given species and group according to sites.
2. Distinguish between earlywood and latewood where possible, and consider these along with total ring width as dependent variables.
3. Allow for removal of trends either by calculating the usual index or by standard statistical techniques.
4. Consider the effect of both precipitation and temperature during the current as well as previous year. Allow for flexibility in analysis by using

monthly or other periodic climatic values rather than yearly averages.

5. Explore other measures of climatic variation such as the evapo-transpiration deficit.

6. Include as an independent variable the previous growth so that autocorrelation in the tree-ring series can be removed.

7. Use a stepwise multiple regression routine for screening the variables and force into the equation all variables which improve the multiple R (F 1.0).

8. Evaluate the relationships in terms of the existing knowledge of tree physiology, attempting to reconcile differences by further experimentation and measurements.

9. Establish a mathematical equation using only the factors found most significant in the screening process and compute the climatic index from long term tree-ring series.

- - - - -

The study was sponsored in part by the American Philosophical Society, Johnson Fund. Extensive computing time and assistance was provided by the I.B.M. Corporation, Argonne National Laboratory, and The University of Arizona Numerical Analysis Laboratory.

References Cited

- Malone, T. F. 1959. Multivariate Statistical Analysis of Atmospheric Processes. Final Report Contract Number AF19(604)-3877. Travelers Weather Research Center, The Travelers Insurance Company, Hartford, Connecticut. 89pp.
- Quenouille, M. H. 1952. Associated Measurements. Butterworths Scientific Publications, London. 242pp.
- Schulman, E. 1956. Dendroclimatic Changes in Semiarid America. University of Arizona Press, Tucson, Arizona. 142pp.
- Thornthwaite, C. W. and J. R. Mather. 1957. Instructions and Tables for Computing Potential Evapotranspiration and the Water Balance. Drexel Institute of Technology, Publications in Climatology 10:(3)311pp.

Table I

Coefficients of correlation (R), coefficients of determination (R^2), and standard errors of estimate (S.E.E.) obtained by means of stepwise multiple regression analysis of tree-rings in white oak.

<u>Site</u>	<u>Dependent variable</u>	<u>Precipitation as moisture variable</u>			<u>Evapo-transpiration deficit as moisture variable</u>		
		<u>R</u>	<u>R²</u>	<u>S.E.E.</u>	<u>R</u>	<u>R²</u>	<u>S.E.E.</u>
South-facing slope	Earlywood	.893	.797	+06.5	.871	.759	+07.1
	Latewood	.835	.697	+34.2	.884	.781	+29.1
	Total wood	.830	.689	+38.8	.842	.708	+37.6
	Simplified equation Latewood				.855	.731	+32.2
North-facing slope	Earlywood	.914	.836	+05.1	.857	.735	+06.4
	Latewood	.894	.800	+23.7	.903	.816	+23.2
	Total wood	.892	.795	+28.3	.880	.775	+29.7
	Simplified equation Latewood				.879	.772	+25.8

Figures 1 and 2:

Bar graph of standardized regression coefficients¹ for final stepwise multiple regression equations. Each of the two figures includes 3 problems with the dependent variable of earlywood, latewood, or total ring width. Sign is indicated by direction of bar, while a missing bar indicates that the coefficient was of low significance. Independent variables are previous season's earlywood (PEW), previous season's latewood (PLW), current season's earlywood (CEW)², average monthly temperatures for May through September of the previous season, average monthly temperatures for April through August of the current season, and soil moisture surplus during the previous winter (SMS). In figure 1, precipitation for May through September of the previous season and April through August of the current season was used; in figure 2, evapo-transpiration deficit was used in place of precipitation. Dots indicate significance levels; one dot = .95, two dots = .99, and three dots = .999.

Figure 3

A comparison of tree-ring indices and correlated environmental variables:

1. average width (in.001 inch) of latewood in 5 white oaks on south-facing slope,
  2. average width of latewood in 5 white oaks on north-facing slope, 3 and 4.
- corresponding tree-ring indices to temperature in June and evapo-transpiration deficit in July after effects of earlywood and trend have been removed, 5. sum of 6 and 7 inverted for comparison with tree-ring indices, 6. evapo-transpiration deficit in July times average temperature in June, and 7. average temperature in June.

---

¹Standardized regression coefficient is $b \sigma_x / \sigma_y$ where y is the total ring width.

²Used only in the analysis of latewood.

3