STATISTICAL SIGNIFICANCE AND REPRODUCIBILITY OF TREE-RING RESPONSE FUNCTIONS

B. M. GRAY
T. M. L. WIGLEY

Climatic Research Unit
University of East Anglia
Norwich, U. K.

and

J. R. PILCHER

Palaeoecology Centre
The Queen's University of Belfast
Northern Ireland

ABSTRACT

This paper is concerned with the overall significance and reproducibility of the response function. A test of significance is devised which is based on the Binomial distribution. Combined with other tests, the method is then used to compare two different response functions to examine the reproducibility of climate-chronology response. Two approaches are used: the first compares two response functions covering the same period from the same site, based on independent chronologies of the same species; the second compares the response of a single chronology over two equal non-overlapping time periods. The results suggest that the response in the examples used is statistically reproducible on a site, and statistically stable over periods of time.

Der Beitrag befaßt sich mit der Gesamtsignifikanz und Reproduzierbarkeit von Response-Funktionen. Es wird ein Signifikanztest auf der Grundlage einer Biominalverteilung entwickelt. In Verbindung mit anderen Tests wird dieses Verfahren dann zum Vergleich von zwei unterschiedlichen Response-Funktionen eingesetzt, um die Reproduzierbarkeit der Klima-Chronologie-Response zu prüfen. Hierzu werden zwei Wege beschritten: — Es werden zwei Response-Funktionen verglichen, die sich auf dieselbe Periode desselben Standortes beziehen und auf unabhängigen Chronologien derselben Art beruhen; — es werden Response-Funktionen einer einzigen Chronologie verglichen, die sich auf zwei gleichlange, sich nicht überlappende Zeitabschnitte beziehen. Die Ergebnisse weisen daraufhin, daß die Response-Funktionen in den ausgewählten Beispielen für einen Standort statistisch reproduzierbar und über verschiedene Zeitabschnitte statistisch stabil sind.

Cet article concerne la signification générale et la reproductibilité des fonctions de réponses. Un test de signification basé sur la distribution binomiale a été mis au point. En combinaison avec d'autres tests, cette méthode est utilisée pour comparer deux fonctions de réponse, afin de vérifier la reproductibilité de la réponse climatique d'une chronologie. Deux approaches sont utilisées: la première compare deux fonctions-réponses couvrant la même période et basée sur deux chronologies indépendantes provenant d'une même espèce échantillonnée dans un mdme sites. La seconde compare les réponses d'une seule chronologie, calculées pour deux périodes qui ne se chevauchent pas. Les résultats suggèrent que dans les exemples utilisés, la réponse est statistiquement reproductible sur un site, et statistiquement stable dans le temps.

INTRODUCTION

Except in ideal cases where tree-rings respond dominantly to a *single* climate variable (such as growing season temperature, summer rainfall, etc.), the general tree-climate relationship is so complex that the climate 'signal' can only be extracted from tree-ring data by fairly sophisticated techniques. Of these, the *response function* is the most commonly used (Fritts et al. 1971). The response function relates ring widths from a single-site or area chronology to the previous growth (up to 3 years previously) and to monthly precipitation and temperature values:

$$W_i = a_1W_{i-1} + a_2W_{i-2} + a_3W_{i-3} + b_1T_1 + b_2T_2 \ldots + b_kT_k \ldots + b_{N/2}T_{N/2} + c_1R_1 + c_2R_2 \ldots + c_kR_k \ldots + c_{N/2}T_{N/2} \quad (1)$$

Here W_i is the width for year i, and T_k and R_k are monthly temperature and rainfall values for September, August, July, June, etc. back to N/2 months prior to the end of the growing season (N is usually 28).

The obvious way to calculate the response function (i.e. the coefficients b_k and c_k) is to use stepwise multiple linear regression. However, when calculated in this direct way, the coefficients tend to be unstable; they show considerable variation from step to step in the regression analysis because of the intercorrelations in the climate data. To overcome this, Fritts et al. (1971) employed principal component analysis, and replaced equation (1) by:

$$W_i = a_1W_{i-1} + a_2W_{i-2} + a_3W_{i-3} + d_1V_1 + d_2V_2 \ldots + d_kV_k \ldots + d_NV_N \quad (2)$$

where V_k are the eigenvectors of the temperature and rainfall: i.e. each V_k is a linear combination of the variables T_k ($k = 1, ..N/2$) and R_k ($k = 1, ..N/2$). This is not a completely consistent method, since, although the V_k are orthogonal, the previous growth terms are not. It does, however, appear to overcome the stability problem noted above. When the regression has terminated, the V_k are transformed back to the appropriate T_k and R_k. The transformed coefficients are then the elements b_k and c_k of the response function. This paper is concerned with the overall significance of the response function, and how it can be tested.

If all of the principal components are retained and the regression analysis is carried through to completion ($N + 3$ steps) the results (1) and (2) must be identical. There is a point when addition of a new variable in the regression equation does not increase the explained variance by a statistically significant amount. Because of this possibility it is likely that an "optimum" response function can be calculated by curtailing the number of steps, *m*, in the regression analysis after less than $N + 3$ steps. Because the regression used is stepwise, and variables may drop out at any stage, *m* is more accurately defined as the number of variables in the regression at the final step.

The choice of *m* can be made in several ways. The classic method is to test the entering variable with an F-ratio test. For example, the regression may be terminated when $F < 1.0$, and when the explained variance due to the addition of a variable is less than the remaining error variance. Another criterion would be to halt the regression at the point where the maximum number of statistically significant b_k and c_k occur. Yet another criterion for *m* is to chose the step at which the variance of the chronology explained by climate is a maximum. A test of overall statistical significance is required for all methods of choosing *m*.

A second test of statistical significance needs to be devised when two response functions are compared. Comparisons can arise in two ways: on comparing two response functions from the same site or area based on independent chronologies covering the same time period; and in comparing the response of a single chronology over two equal non-overlapping time periods.

OVERALL SIGNIFICANCE OF A SINGLE RESPONSE FUNCTION

In presenting response function results the coefficients b_k for temperature and c_k for rainfall are displayed graphically, with error bars which usually represent 95% confidence limits. These imply that there is only a 5% chance that the true value of the coefficient lies outside the limits shown. The previous growth coefficients a_j are also delimited this way. If the confidence limits for any coefficient exclude zero, then that coefficient is significantly different from zero at the 0.05 level: that is to say, the probability that any one coefficient is non-zero (at the level $\alpha = 0.05$) by chance is < 0.05. The probability, P, that n coefficients in a total number of response coefficients N will be significantly non-zero at the 0.05 level is determined by the Binomial distribution and is:

$$P = \frac{N!}{(N-n)\,!n!} (.05)^n\,(.95)^{N-n} \tag{3}$$

The results of this distribution, assuming that the coefficients are independent, are given in Table 1.

Equation 3 is, of course, an idealised approximation to the problem, since in practice the elements in the response function may not be independent. Since only *k* of the N original variables can be considered to be independent (where *k* is the number of significant eigenvectors in the principal component analysis), we might reasonably assume that only *k* of the possible regression coefficients are independent. The determination of *k* is not trivial. Guiot (1981) has shown that a reasonable criterion of significance is the condition that PVP > 1, where PVP is the product of the eigenvalues in descending order of rank. Since the determinant of the correlation matrix of orthogonal variables is equal to 1, the last significant eigenvector is considered to be that whose eigenvalue brings the PVP to as near a value of 1 as possible. Typically, this would be eigenvector 22 for the cases considered in this paper.

A reduction in the number of effectively independent elements in the response function is analagous to the non-independence of the trials in our theory. If the trials were tosses of a single coin, the length of runs of either outcome (tails or heads) would be greater than that expected by chance. It is not at all clear how the introduction of this apparent autocorrelation in the results of the trials would affect either the mean or the variance of the results. In the response functions discussed below, the effect is, in any case, small and will be ignored.

The null hypothesis to be tested is that the response function is zero. The alternative hypothesis is that the coefficients of response functions are not all zero, that is, that the function is significantly non-zero. We wish to accept this hypothesis at the 0.05 level. Table 1 shows that, even if there were no climate signal, it is highly probably that 1, 2 or 3 coefficients will be significantly non-zero. For N = 24 or 28, at least 4 coefficients must be non-zero at the 0.05 level to get an overall result which is significant at the 0.05 level, while at least 5 coefficients must be non-zero at the 0.05 level to get an overall result which is significant at the 0.01 level (see Table 1).

Table 1. Probability of n significant coefficients occurring by chance.

Number of coefficients significantly non-zero at the 0.05 level (n)	Probability of occurrence by chance			
	N = 24	N = 28	N = 12	N = 14
0	.292	.238	.540	.488
1	.369	.350	.341	.359
2	.223	.249	.099	.123
3	.086	.114	.017*	.026*
4	.024*	.037*	.002**	.004**
5	.005**	.009**		
6	.001**	.002**		

*Significant at the 0.05 level. **Significant at the 0.01 level.

This simple theory can be illustrated using the response function of the Schleswig-Holstein Chronology (Germany, 54° 20'N, 9° 10'E) from Eckstein and Schmidt (1974). Only 26% of the ring width variance is explained by climate in this response function. No values of the F level are quoted, and it is not known at what step the regression terminated. However, examination of the diagram (Figure 6 in *their* paper), shows 15 coefficients to be significantly different from zero. The overall climatic response is therefore highly significant using the Binomial test, contrary to the conclusions given in the original paper.

However, the above discussion of the overall significance of the response function does not take into account the numerical values of the individual coefficients. Suppose a single coefficient is very large and the 95% confidence limits are a long way from the zero line on the diagram. Such a coefficient would be highly significant, with a probability level α well below 0.05. The probability of such a result by chance is $N\alpha(1-\alpha)^{N-1}$. For $N = 28$, and only this one very significant coefficient, the overall response function will be significant at the 0.05 level if $\alpha < 0.00187$ and at the 0.01 level if $\alpha < 0.00036$. Thus a response function may be significant either if a few elements have error bars a long way away from the zero line, or if many elements have error bars only just clear of the zero line.

COMPARING TWO RESPONSE FUNCTIONS

Suppose two response functions, F_1 and F_2, are to be compared to see whether they are significantly different. Three tests are given here because no single test could be found which was completely satisfactory. The third test is thought to be the most satisfactory. Although the analyses are restricted to temperature and rainfall response, they can clearly be extended to include the prior growth coefficients.

Correlation. The null hypothesis is that the two response functions are different, that is, that the correlation coefficient between elements in F_1 and elements in F_2 is zero. The null hypothesis can be tested by examining the statistical significance of the correlation coefficient in the standard way using, as sample size, the number of elements compared. If the temperature elements are tested, followed by the rainfall elements, significant correlations indicate the similarity of the two functions in temperature response, rainfall response or both. This test does not consider whether the individual elements are significant or not.

Coin-Tossing Analogues. The probability is considered of two patterns arising by chance, each having a certain number of matching elements. Two kinds of matching elements are distinguished, those where the error bars do not cross the zero axis and those where the error bars do; i.e. significant and non-significant elements. This problem is then equivalent to tossing two loaded coins and calculating the number of matched heads or tails that arise in N throws (Hoel 1965). The null hypothesis is that the two response functions are different: i.e. that the number of matches could have arisen by chance. This is contrasted with the alternative hypothesis that this number is significantly greater than expected by chance.

Suppose that two response functions with N coefficients each are calculated from independent samples. Let the number of significant coefficients be m_1 and m_2 and the number of non-significant coefficients be $(N-m_1)$ and $(N-m_2)$ respectively. The probability, P_1, of having one significant coefficient of the same sign coincident in the two samples is:

$$P_1 = \frac{m_1 m_2}{2N^2} \tag{4}$$

The probability P_2 of one coincident nonsignificant coefficient of either sign is:

$$P_2 = \frac{(N-m_1)(N-m_2)}{N^2} \tag{5}$$

The probability P_3 that a pair of coefficients will consist of one significant and one non-significant coefficient is then:

$$P_3 = 1 - P_1 - P_2 \tag{6}$$

The joint probability of having the two samples with n_1 coincident significant and n_2 coincident non-significant coefficients by chance is then:

$$P = (P_1)^{n_3} (P_2)^{n_2} (P_3)^{n_3} \frac{N!}{n_1!\, n_2!\, n_3!} \tag{7}$$

Where $n_3 = N - n_1 - n_2$ is the number of non-matched elements.

In this approach the position of the error bars relative to the zero line is important and depends on the regression step; the numerical value of each coefficient is not considered.

Element Matching. Each element of the response function F_1 is tested against the corresponding element of F_2. If the 95% error bars overlap, then a match is designated. The number of non-matches is counted, and the significance tested by exactly the same theory used in determining the overall significance of the response function. The null hypothesis is that the two functions F_1 and F_2 are the same. If the confidence limits of any element do not overlap, then those coefficients are significantly different at the .05 level. The probability of any coefficient in F_1 not matching, by chance, the similar coefficient in F_2 at the $\alpha = 0.05$ level is then < 0.05. The probability of n coefficients mismatching by chance in N comparisons is given by equation (3). In this test, the numerical size of the error bars is dependent on the step in the multiple regression at which they were calculated. Generally, early steps in the regression have many degrees of freedom and the 95% error limits are small. As the regression proceeds, the error bars become larger and fewer non-matches are found. The choice of *m* therefore influences the apparent stability or reproducibility as determined by this method.

For maximum matching, we wish the results of the first and last tests to be high, and the middle one to be low.

TESTS OF SITE REPRODUCIBILITY

An example of the method can be used to examine the reproducibility of a response function on a site. Clearly, in practice one would examine the correlation coefficient between the chronologies rather than compare the response functions. Nevertheless, the example serves to illustrate the method.

Site reproducibility was tested using the Rostrevor, Northern Ireland oak chronology (54° 6'N, 6° 12'W) from Pilcher (1976). A sample of one core/tree from 11 trees was taken from the original chronology, and formed into a separate chronology (ROSTRE). A new sample of 11 different trees (1 core/tree) was collected from the same site and processed by Pilcher to form the NNROST chronology. We first tested the individual response functions (Figure 1) for each chronology. Both the temperature responses and the rainfall responses were found to be statistically significant at the 1% level using the method described in this paper.

Comparing the chronologies directly, we found for 1920-1969 a correlation coefficient of 0.83. To compare the response functions, we used the three tests outlined above. The correlation test had the following results:

Temperature: $r = .74$, significant at the 1% level
Rainfall: $r = -.12$, not significant

The values for calculating the significance level of the coin-tossing comparison are:

$m_1 = 13$ $n_1 = 5$ (3 in temp., 2 in rainfall)
$m_2 = 14$ $n_2 = 8$ (5 in temp., 3 in rainfall)
$N = 28$ $n_3 = 15$ (6 in temp., 9 in rainfall)

This test finds the response functions to be similar at the 2% level of significance (i.e. the probability that the response functions would be so similar by chance is 2%). Four differences were found in the element matching test, all in rainfall, indicating that the response functions were significantly different at the 5% level. Obviously the difference arises only in the rainfall response.

The relatively low correlation between the chronologies is thus reflected in the response functions. The overall conclusion from the three tests is that the response of the site to temperature is highly reproducible, but the response to rainfall is probably influenced by the microclimate on the site.

COMPARISONS TO TEST STABILITY IN TIME

One of the basic assumptions underlying dendroclimatological work is that the tree shows a constant response throughout its life to the microhabitat in which it grows, that is, that the response function should be stable in time. Factors which might change the response function from one period to another could be derived either from changes in climate or from changes in tree growth. In order to test time stability, a single chronology may be divided into two non-overlapping time periods and the response functions calculated for each. The similarities and differences between the elements of the response function may then be compared by the three methods described. In chosing time periods for comparison, these must be dictated by the span of the available chronologies, and not by any characteristics noticeable in the climatic data.

Three oak chronologies were chosen. The first is an indexed site chronology from Maentwrog, in North Wales (Hughes et al. 1978), where only a short length of climate data is available. The second is a long non-indexed area chronology covering Hereford and Cumberland (V. Giertz-Siebenlist, pers. comm.). The third series is an indexed site chronology collected by Pilcher from Fontainebleau near Paris. The chronology

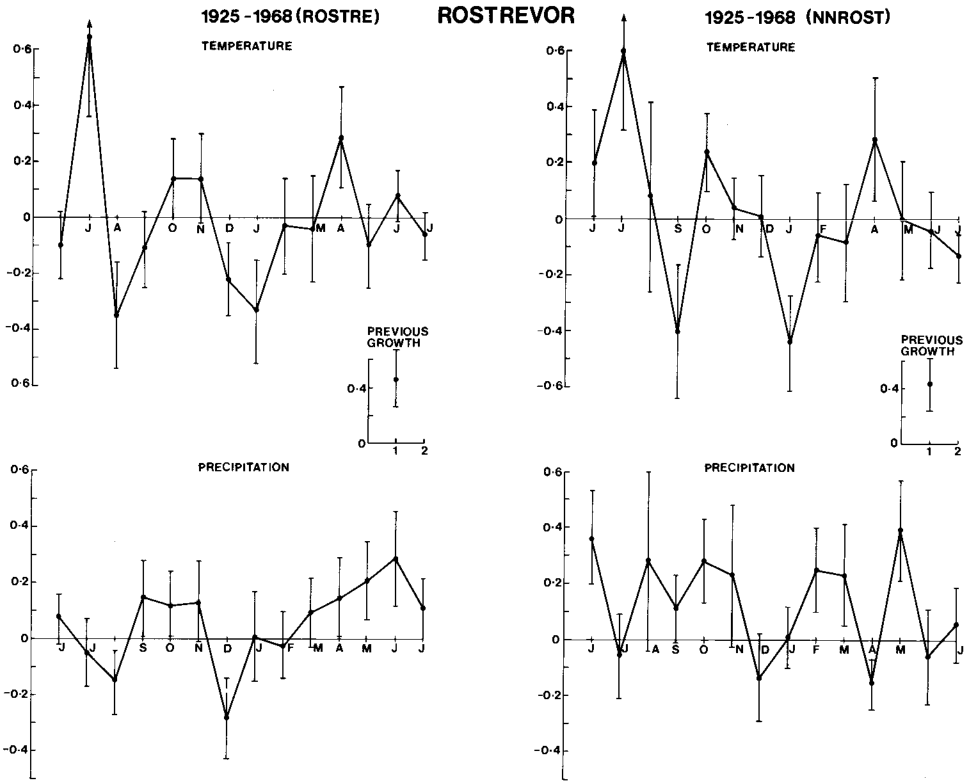

Figure 1. Response functions for Rostrevor.

Table 2. Comparisons of response functions.

CHRONOLOGY	PERIOD	% Climate	% Growth	% Total	No. of Sig. Elements		No. of Years	No. of Steps
					Temp	Rain	(Y)	(m)
ROSTRE	1925-1968	57	21	78	5	7	44	12
NNROST	1925-1968	64	18	82	8	7	44	13
	1907-1940	42	20	62	2	1	34	9
Maentwrog	1941-1974	44	36	80	4	2	34	11
	1755-1839	24	49	73	4	5	85	18
	1855-1949	15	40	55	3	4	95	16
Fontainebleau	1800-1869	39	19	58	4	3	70	16
		20	27	47	7	4	70	9
FONT 1	1870-1939	49	5	54	3	4	70	14
		41	6	47	5	6	70	9
FONT 2	1870-1969	33	25	58	7	3	100	15

variance explained by climate is higher in the first and third series (Table 2). When these six individual response functions were tested using the Binomial test, all were significant at the 0.05 level or better, except Maentwrog (1947-1974).

Maentwrog. The chronology for Maentwrog was divided into two sections, 1907-1940, and 1941-1974. To preserve a reasonable number of degrees of freedom with only 34 observations, the climate year considered was reduced to October of the year i − 1 through to September of year *i* (N = 24). Response functions were calculated for the two periods, (Figure 2), and the properties listed in Table 2.

The correlation test results were:

Temperature: r = .55, significant at the 10% level
Rainfall: r = -.01

Out of the 24 elements calculated for the coin-tossing analogue, we have (recall that m_1 and m_2 are the number of significant coefficients in the two response functions being compared, while n_1 is the number of matched significant coefficients, and n_2 is the number of matched non-significant coefficients):

m_1 = 7 n_1 = 1 (1 in temp., 0 in rainfall)
m_2 = 5 n_2 = 13 (4 in temp., 9 in rainfall)
N = 24 n_3 = 10 (7 in temp., 3 in rainfall)

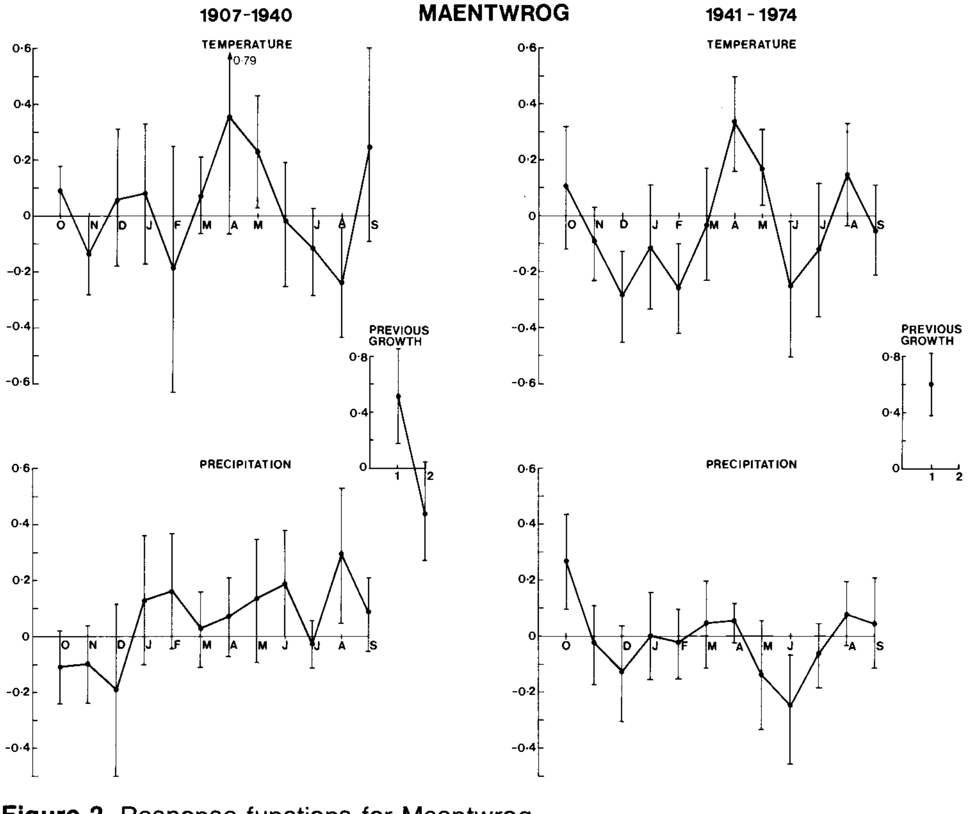

Figure 2. Response functions for Maentwrog.

The probability of such a match occurring by chance is 0.0266, so the two samples are similar at the 3% significance level. In applying the element matching test, three non-matches (1 in temp., 2 in rainfall) were found, indicating a difference which is significant at the 7% level.

The overall conclusion is that the response function is reproducible in time, more so for the temperature elements than the rainfall. In this case, it is relevant to point out that there is a considerable difference in the variance explained by the previous growth factors (Table 2). Since these are non-orthogonal to the climate elements, some of the differences in the rainfall elements may be attributed to a redistribution of variance between climate and previous growth in the two samples.

Hereford. The Hereford-Cumberland chronology was also tested for reproducibility in time. It was divided into two periods, 1755-1839 and 1855-1949. Response functions were calculated using the central England temperature series and the England and Wales rainfall series (Figure 3). The value used for N was 28.

In the correlation test, the temperature result again indicated a higher reproducibility.

Temperature: r = .50, significant at the 10% level
Rainfall: r = .17

Some difference is seen in the previous growth factor in the two samples as in the Maentwrog chronology.

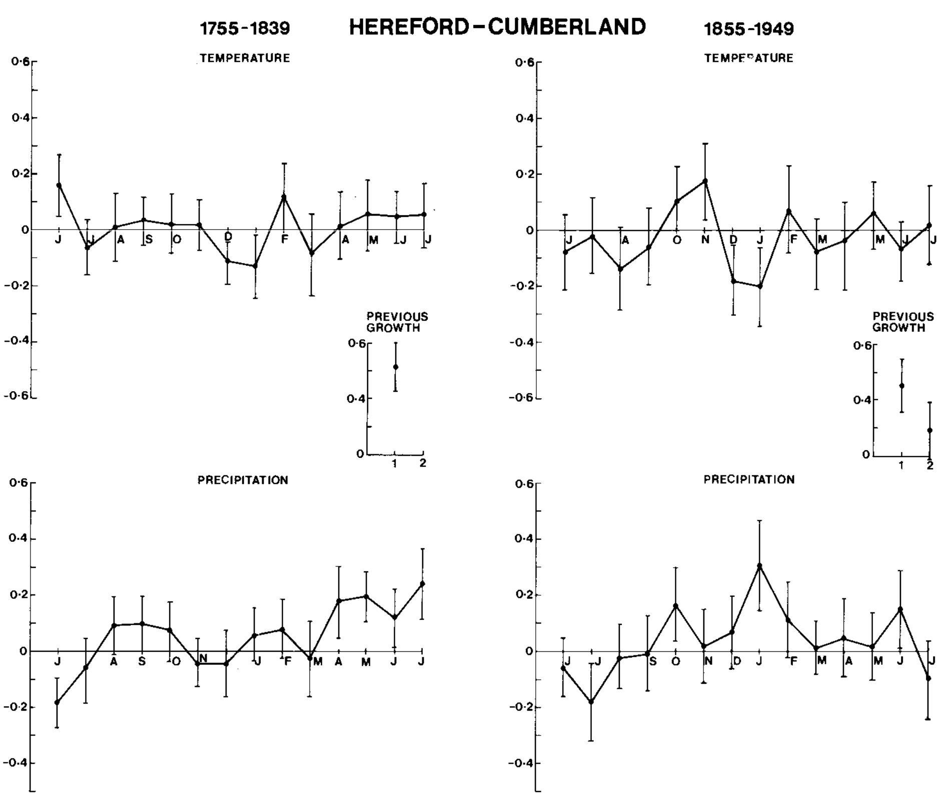

Figure 3. Response functions for Hereford — Cumberland.

The values obtained for the coin tossing test were:

$m_1 = 8$ $n_1 = 3$ (2 in temp., 1 in rainfall)
$m_2 = 10$ $n_2 = 12$ (7 in temp., 5 in rainfall)
$N = 28$ $n_3 = 13$ (5 in temp., 8 in rainfall)

Despite the low variance explained by climate in this area chronology, and inhomogeneities in the rainfall data, the two response functions were similar at the 2% level of significance. Only one non-match was found in the element matching test, which means that the null hypothesis of similarity cannot be rejected: i.e. this result indicates no significant difference between the two samples. Overall, two out of the three tests gave satisfactory results, the only significant difference being in the rainfall coefficients as indicated by the correlation test.

Fontainebleau. The third chronology from Fontainebleau is a particularly long modern site chronology which extends from 1540 to 1979. Climatic data exist for Paris from 1770 onwards, but the early part of the record may be unreliable. Two periods 1800-1869 and 1870-1939 were chosen. As far as is known the site was selectively felled in a regular manner between 1800 and 1939. The modern period from 1940 onwards was discarded because of interference at the chronology site, which might well have affected the response function.

The results of the comparison tests are as follows. For the correlation test:

Temperature: $r = .22$
Rainfall: $r = .12$

Neither result is significant. (Figures 4 and 5.)

The coin-tossing test was done twice, for two different sets of m, the number of variables in the regression. These were obtained by curtailing the stepwise regression at different stages.

Comparing $m = 9$, $m = 9$: we have $m_1 = 16$ $n_1 = 5$
$m_2 = 14$ $n_2 = 7$
$N = 28$ $n_3 = 16$

Comparing $m = 16$, $m = 14$: we have $m_1 = 7$ $n_1 = 3$
$m_2 = 7$ $n_2 = 16$
$N = 28$ $n_3 = 9$

Both these sets of results show that the null hypothesis that the response functions are different has to be rejected at the 3% level; i.e. the response functions are significantly the same.

The element matching test was also done for the different values of m. For $m = 16, 14$, there were no differences in matching. For $m = 9,9$, there were four differences. Such a mismatch could arise by chance only 4% of the time. In other words, the result of this test depends crucially on the choices of m; for the larger m the response functions are similar, for the smaller m they are dissimilar (at the 4% level). In fact, overall the Fontainebleau results are equivocal. One possible interpretation of

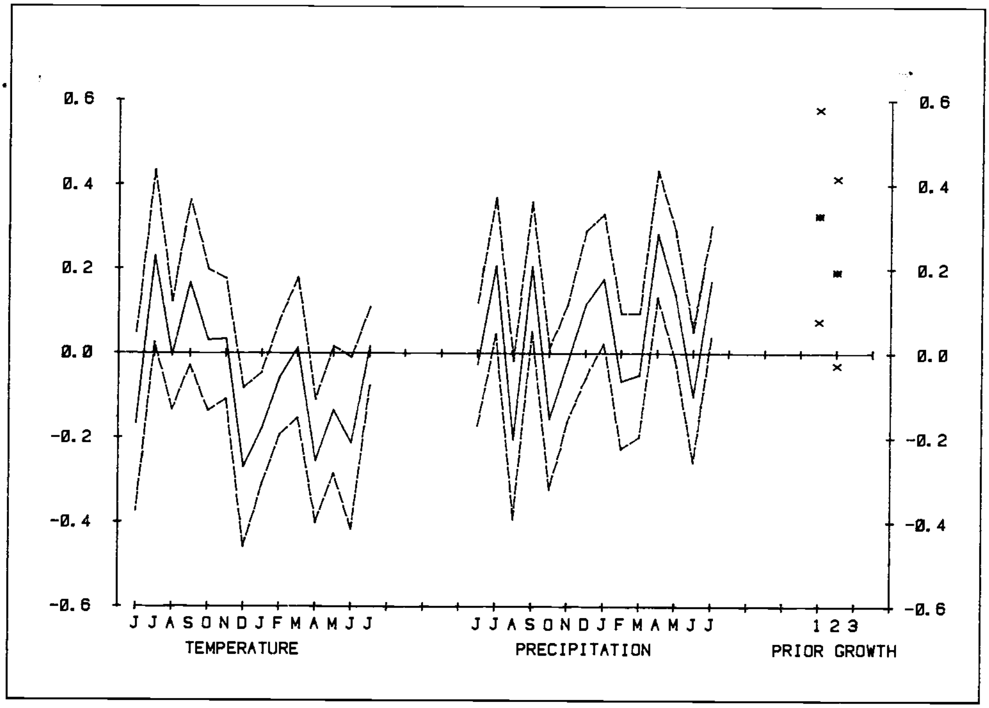

Figure 4. Response function for Fontainebleau, A.D. 1800-1869.

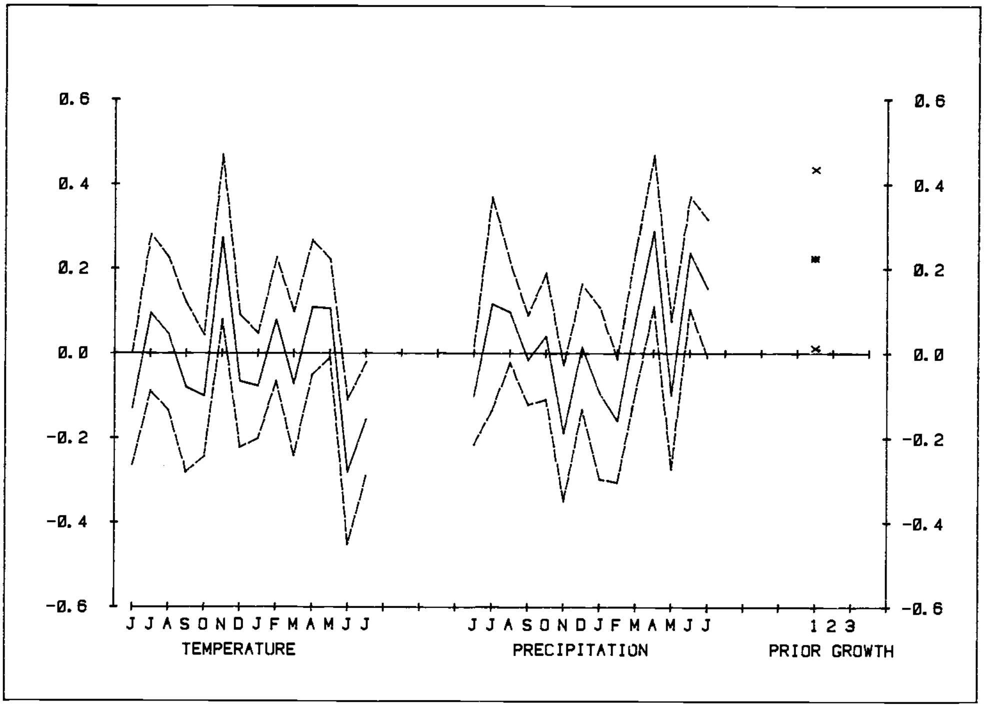

Figure 5. Response function for Fontainebleau, A.D. 1870-1939.

this is that there is a real difference in the response functions which could be explained by an age effect.

The majority of the 25 trees in the second portion of the chronology (1870-1939) were over 170 years old. The first portion of the chronology (1800-1869) also has 25 trees but they are of younger age. In order to identify an age factor, a second chronology was built from 20 trees not more than 50 years old, growing at the same site and covering the period 1870-1969 (Figure 6).

This chronology of younger trees was then compared with the original chronology of older trees. Table 2 clearly shows the difference in the prior growth factor in the two chronologies.

The correlation test gives r = .64 for temperature, significant at the 1% level, and $r = -.27$ for rainfall. The results of the coin-tossing test are:

for m = 16,15: $m_1 = 10$ $n_1 = 4$
$m_2 = 10$ $n_2 = 12$
$N = 28$ $n_3 = 12$

indicating a similarity at the .1% significance level. Three mismatches are found in the element matching test, which could arise by chance 12% of the time. Thus, the overall conclusion is that the response function for the young chronology is very similar to the response function for the old chronology. Effectively, this has been a second study of site reproducibility. The hypothesis of an age effect influencing the response function must therefore be rejected.

If the lack of convincing stability is not due to an age effect in the tree, clearly we must examine the climate data. The Paris temperature record shows a steady rise in

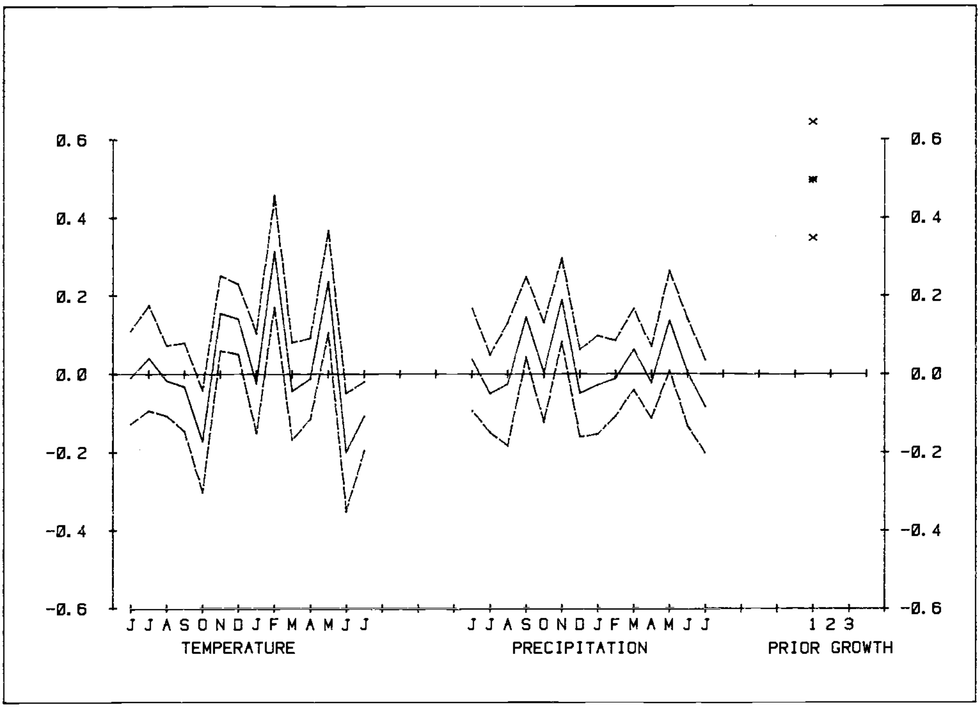

Figure 6. Response function for Fontainebleau, A.D. 1870-1969.

Table 3. Results of comparisons of response functions.

Chronology	Significance level of Correlation Test		**Coin-Tossing** Probability of being so similar by chance	**Element-Matching** Probability of being so different by chance
	T	R		
ROSTREVOR (site stability)	1%	-	.02	.05
MAENTWROG (time stability)	10%	-	.03	.07
HEREFORD (time stability)	10%	-	.02	.35
FONTAINEBLEAU (time stability)	- -	- -	(m = 9,9) .03 (m = 16,14) .03	.04 .24
FONTAINEBLEAU (age effects) (site stability)	1%	-	.001	.12

decadal values between 1880 and 1960. Comparison with central England temperature data (which are homogeneous) suggests that there is an urban heat island effect in the Paris data over and above a common long-term variation associated with larger scale climatic change. This may be the factor which prevents the Fontainebleau results from showing a convincingly stable pattern. The basic reason for this is that Paris temperatures in recent decades are not related to temperature at Fontainebleau in a consistent way.

CONCLUSIONS

The use of the Binomial distribution has been proposed as an indicator of the overall significance of a response function. By including considerations of the size of the elements and the number of degrees of freedom, three tests have been suggested to establish the significance of the similarity or difference between two response functions. These tests have been used to examine reproducibility at a site, and the stability of a response function in time. The results are summarized in Table 3.

In examining site stability we have found very convincing evidence of stability, especially with respect to temperature response, at a site in northern Ireland. Time stability was tested at three different sites. The first, Maentwrog, provided good evidence of stability, especially for temperature response, although the element matching test was not entirely consistent with the other tests. For the second site, Hereford, all tests indicated stability of the response function. For the third site, Fontainebleau, the results for a large number of regressions were stable over time; but when fewer regressors were used the results of different tests were not in complete agreement. In all three cases where the probability of the response functions being similar by chance is small (cointossing test), the probability of the response functions being *different* by chance (element matching test) is larger. The correlation test results indicate that temperature response is more stable than rainfall response.

In an attempt to explain the results of the small-m Fontainebleau study we examined the question of age stability and found overwhelming evidence that there is no significant age effect on tree-climate response.

None of the comparison tests is totally satisfactory. Our Binomial test for individual response functions (which only examines significant coefficients) indicates that all of the response functions considered are significant overall at the 5% level (and 8 out of 9 are significant at the 1% level). Our stability tests (which are based on a comparison of significant *and* non-significant coefficients) are not quite as convincing. This may be important in the test of site reproducibility; in testing time stability it is possible that real, but small, changes in response may arise from site effects. It is hoped that alternative methods will be proposed and tested by other authors to improve on this first attempt to show overall significance and stability of tree-climate response functions.

REFERENCES

Eckstein, D. and B. Schmidt
1974 Dendroklimatologische Untersuchungen an Stieleichen aus dem maritimen klimagebiet Schleswig-Holsteins. *Angew. Botanik* 48: 371-83.

Fritts, H. C., T. J. Blasing, B. P. Hayden, and J. E. Kutzbach
1971 Multivariate techniques for specifying tree-growth and climate relationships and for reconstructing anomalies in paleoclimate. *Journal of Applied Meteorology* 10(5) 845-64.

Guiot, J.
1981 Analyse mathématique des données géophysiques (Applications à la dendroclimatologie). Doctoral dissertation, Université Catholique de Louvain.

Hoel, P. G.
1965 *Introduction to mathematical statistics.* John Wiley & Sons, New York.

Hughes, M. K., B. Gray, J. Pilcher, M. Baillie, and P. Leggett
1978 Climatic signals in British Isles tree-ring chronologies. *Nature* 272: 605-06.

Pilcher, J. R.
1976 A statistical oak chronology from the North of Ireland. *Tree-Ring Bulletin* 36: 21-27.