ZESZYTY NAUKOWE POLITECHNIKI ŚLĄSKIEJ 1987
Seria: MATEMATYKA-FIZYKA z. 56 GEOCHRONOMETRIA Nr 4 Nr kol. 912

Tomasz GOSLAR, Adam WALANUS
Radiocarbon Laboratory
Silesian Technical University, Gliwice

A MONTE CARLO STUDY OF THE PROBABILITY DISTRIBUTION OF "t" VALUE USED IN DENDROCHRONOLOGICAL DATING

Summary: A series of Monte-Carlo experiments has been carried out in order to investigate the distribution of "t" value used in dendrochronological studies. The results demonstrate that the autocorrelation existing in sequences of tree-ring indices leads to essential increase of the dispersion of distributions. The probability distribution of "t" value can be approximated with the modified Student's distribution with extended scale. In consequence, probability of random $t>t_0$ events is much greater than predicted basing on theoretical Student's distribution.

1. INTRODUCTION

Dendrochronological dating is based on the assumption that trees which grew under the same environmental conditions over the same period of time reveal similar tree-ring width patterns. Fitting two tree ring series to each other is called cross-dating. This can be done either visually or more effectively by using a computer and quantifying cross-correlation. One of the statistical measures used is the so-called t-value, which is related to the correlation coefficient and is calculated at any position of overlap of two compared tree-ring sequences (Baillie, Pilcher, 1973; Aniol, 1983). This procedure has been firstly introduced considering that if both data series are independent and normally distributed, the variable "t", defined as

$$t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}}, \tag{1}$$

where N denotes number of data pairs (number of years of overlap of the two tree-ring patterns) and r is the correlation coefficient, has a Student's probability distribution with k=N-2 degrees of freedom. Because of existing long-term trends, raw tree-ring widths are not suitable for correlation calculations and should be therefore subjected to appropriate preliminary standardization. One of the methods used to eliminate such trends consists in the conversion of tree-ring widths to the so called indices defined as

$$x_i = \ln \frac{5 \cdot d_i}{d_{i-2}+d_{i-1}+d_i+d_{i+1}+d_{i+2}} \quad (2)$$

The values of x_i obtained in such way are approximately normally distributed, what is not true for d_i. Thus the procedure of cross-dating is based on the rejection of the null hypothesis that the two compared tree-ring sequences are not correlated (i.e. do not descent from the same period of time). Under the null hypothesis the probability that "t" is greater than a certain value t_α is equal to $\alpha=1-F_k(t_\alpha)$, where $F_k(t_\alpha)$ denotes Student's distribution function with k degress of freedom; α is the level of significance. If we check the null hypothesis for certain α and obtain $t>t_\alpha$ we should reject it and accept the alternative hypothesis that the two compared tree-ring sequences descent from the same period of time.

The practice, however, seems not to confirm these theoretical predictions. For example, for ~42000 positions of overlap checked in our lab (Goslar, in print) for different oak-ring sequences (mean length of interval of overlap $N \approx 70$), $t>3.0$ had happened in 252 cases, and $t>3.5$ in 70 cases, i.e. $P(t>3.0) \approx 6 \cdot 10^{-3}$ and $P(t>3.5) \approx 1.7 \cdot 10^{-3}$ whereas the corresponding theoretical probabilities are equal $1-F_{70}(3.0) \approx 1.9 \cdot 10^{-3}$ and $1-F_{70}(3.5) \approx 4 \cdot 10^{-4}$. High "t" values were found therefore much more frequently than it should be expected from Student's probability distribution. In this article we attempt to explain the cause of this discrepancy.

2. THE PROPERTIES OF INDICES

There are two possible reasons why "t" value calculated after eq. (1) do not follow Student's distribution:

- the distribution of indices differs from normal;
- indices of single tree-ring sequence are not self-independent i.e. each index is a bit correlated with previous ones.

In order to check this, indices have been collected from all oaks measured in our lab. The experimental distribution of indices is shown in Fig. 1. As can be seen from Fig. 1, the empirical distribution is not normal. This feature is also confirmed by the χ^2 test. A similar shape of distribution is shown by indices calculated for 3 floating chronologies constructed in our lab (total ~630 years, ~5 trees/year) and for SGerman-Swiss Master Chronology (Becker, 1981). The only difference is a bit less dispersion, σ=0.064 for our chronologies and σ=0.047 for Becker's one. It seems that decrease of σ is an effect of averaging tree-ring widths for several trees.

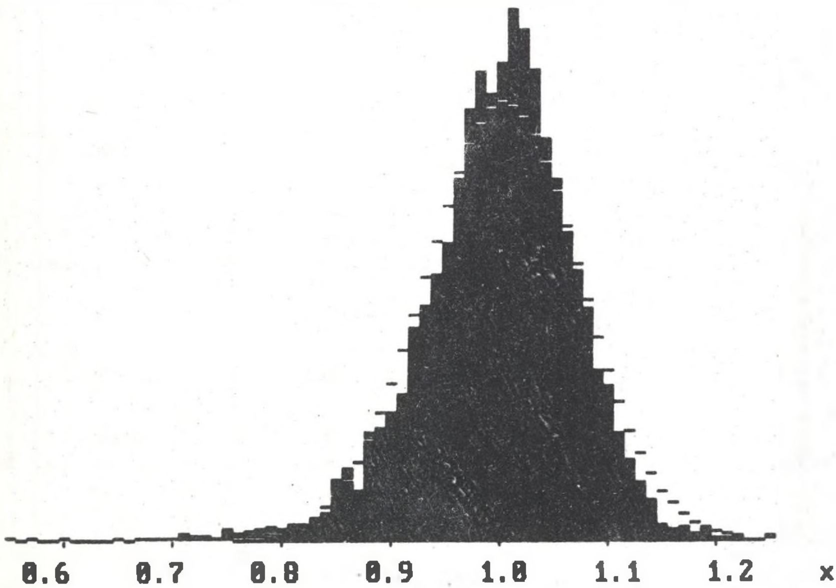

Fig. 1. The probability distribution of indices (eq (2)), for single oak trunks measured in Gliwice lab. The normal probability distribution with empirical mean ($\bar{x}$=.997) and dispersion (σ=.071) is also plotted.

Rys. 1. Rozkład prawdopodobieństwa indeksów (wz. (2)) dla pojedynczych pni dębów zmierzonych w Gliwicach. Dla porównania przedstawiono również rozkład normalny o średniej i dyspersji określonych na podstawie rozkładu doświadczalnego ($\bar{x}$=.997, σ=.071).

Considering the statistical independence of indices, the correlation coefficient between x_i and x_{i-l} has been calculated for all oak turnks already measured. The results for l=1,2,3 are summarized at Table 1. The autocorrelation is significant for l=1 and l=2, for l=3 results are inconsistent. The autocorrelation for l>3 seems to be insignificant. It should be mentioned that, as it can be expected from mathematical form of eq. (2), even the indices calculated for independently normally distributed random numbers show significant autocorrelation. Results of corresponding simulations indicate that autocorrelation coefficients for such indices are equal to c.a. -0.29, -0.34, 0.09 and 0.02 for l=1,2,3,4 respectively, and zero for l>4. Comparison of these figures with empirical values indicate that autocorrelation of indices results from the method of calculation as well as from some undetermined biological factors.

In order to check the influence of both properties described above, a series of Monte-Carlo experiments with computer simulation of cross-dating procedures has been carried out.

Table 1.

Autocorrelation coefficients for tree-ring indices sequences

	Total rings	$r(x_i-x_{i-1})$	$r(x_i-x_{i-2})$	$r(x_i-x_{i-3})$
Single trees	~6900	-0.134	-0.341	-0.052
Chrono-logies	~630	-0.153	-0.346	0.000
Becker (1981)	~3300	-0.181	-0.354	0.036

3. MONTE CARLO EXPERIMENTS

In the Monte Carlo experiments pairs of sequences of indices were generated. To obtain artifical indices, uniformly distributed random numbers generated by EMIX 86 XT microcomputer were transformed using method of inverse distribution function (Zieliński, 1979) according to experimentally determined probability distribution (Fig. 1). In the next step the resulting numbers were then recalculated in order to get appropriate autocorrelation. Some experiments have been performed on sequences with no autocorrelation and in one experiment normally distributed uncorrelated random numbers were used.

Next, the shorter sequence of length=q was passed over the longer one of length=p, and the "t" value was calculated for each position of overlap s from s=1 to p-q+1, so the number of years of overlap was constant and equal p. After checking all positions, new pair of sequences was generated, and the whole procedure was repeated.

4. RESULTS AND DISCUSSION

The parameters of obtained probability distributions of "t" value are listed in Table 2. The following features of those distributions should be pointed out:

1) The dispersion of distributions obtained for autocorrelated sequences is essentially greater in comparison to those with no autocorrelation. Dispersions of the latter ones seem to correspond to those of Student's.

Table 2.

Estimates of some characteristics of probability distributions of "t" value obtained in Monte-Carlo simulations. (μ_1 - 1-th central moment, $\sigma_0=\sqrt{k/(k-2)}$ - theoretical dispersion of Student's distribution).

length p	length q	auto-corr.	mean $\bar{t}$	increase of dispersion σ/σ_0	assymetry $\gamma_1=\frac{\mu_3}{\sigma^3}$	excess $\gamma_2=\frac{\mu_4}{\sigma^4}-3$
300	1000	Yes	.000	1.1261	0.008	0.026
100	1000	Yes	.000	1.1175	0.034	0.073
100	1000	No	.000	0.9986	0.032	0.080
20	1000	Yes	.001	1.0869	0.030	0.503
20	1000	No	.001	0.9432	0.032	0.455
20	100	Yes	.001	1.0896	0.034	0.460
20	20	Yes	.000	1.0840	0.030	0.443
20	1000	Yes[1]	.001	1.0924	0.034	0.500
20	1000[2]	Yes	.001	1.0878	0.038	0.470
20	1000[3]	No	.001	0.9919	0.006	0.483
Theoretical characteristics of Student's probability distribution						
	k= 298		0	1	0	0.020
	k= 98		0	1	0	0.064
	k= 18		0	1	0	0.120

[1] - distribution of indices and autocorrelation like for Becker's chronology (see Table 1); [2] - distribution of indices like for Becker's chronology; [3] - normal probability distribution;

2) The excess of all distributions seems to be close to that of Student's, irrespectively whether the data sequences were autocorrelated or not. It seems that the excess is greater in the case when a shorter sequence is passed over a longer one than in the case when both sequences are of approximately the same length (compare 20/1000...20/20 in Table 2). However, it is difficult to say which case better reflects the reality. Significance of this effect, however, seems to be very low, especially for great number of years of overlap.

3) It seems that there is no dependence between distribution of "t" value and the dispersion of indices distribution. There is nearly no difference, whether single trunk sequences or scales are compared.

4) All distributions reveal a slight positive assymetry, which seems to be insignificant only in the distribution obtained for normally distributed numbers. Thus it appears, that the resulting assymetry is caused by the assymetry of indices distribution (Fig. 1). The significance of resulting assymetry, especially for not very high "t" values, seems to be very little.

Taking into account the features described above, we can finally state that the main difference between the theoretical Student's probability distribution and that observed for tree trunks lies in greater dispersion caused by the autocorrelation of indices. Thus, the distribution of "t" value can be approximated with the Student's distribution with the scaling factor σ/σ_0

$$p(t) = A \cdot \left[1 + \frac{1}{k} \cdot \left[\frac{t}{\sigma/\sigma_0}\right]^2\right]^{-\left[\frac{k+1}{2}\right]} \qquad (3)$$

where A denotes the normalization factor. Exemplary comparison is shown in Fig. 2. This approximation seems to be fairly good for low values of t. For high t (say t>4), however, it seems to be better to calculate the probability $P(t>t_0)$ directly from the obtained distributions. Calculated values of probability P for some selected values of t_0 and k are given in Table 3.

It should be mentioned that similar experiments could be applied in the case, where any other method of standardization of tree-ring widths (e.g. so-called "Wuchswert") is used. It may be expected that the results would be qualitatively rather similar.

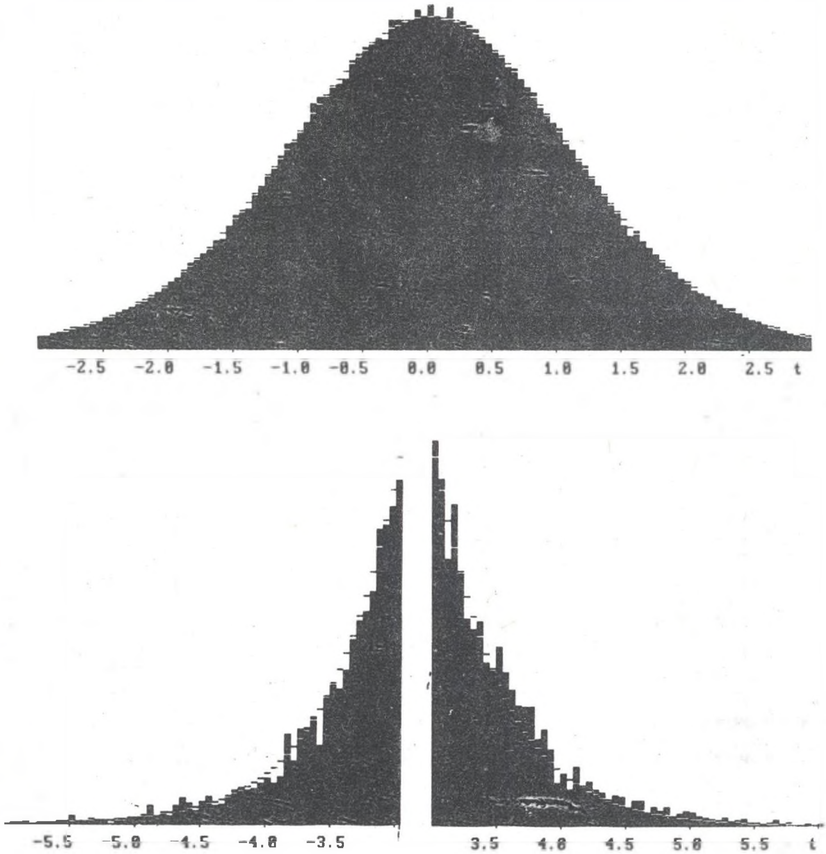

Fig. 2. Comparison of obtained 20/1000 distribution with modified Student's distribution (χ^2_{15}=17.65). The tails of distribution are presented separately with extended vertical scale.

Rys. 2. Porównanie otrzymanego rozkładu prawdopodobieństwa 20/1000 ze zmodyfikowanym rozkładem Studenta (χ^2_{15}=17.65). Ogony rozkładu przedstawiono oddzielnie w zwiększonej skali pionowej.

5. CONCLUSIONS

Described Monte Carlo experiments indicate that dispersion of probability distribution of "t" value is significantly greater than the theoretical dispersion of the Student's probability distribution. The existing autocorrelation of indices causes that the probability of random appearance of high "t" value is essentially greater than that predicted from the Student's distribution. A comparison of results shown in Table 3

Table 3.

Probabilities $P(t>t_0)$ estimated from Monte-Carlo simulations.

	Number of years of overlap		
t_0	20	100	300
3.0	0.008	0.005	0.005
3.5	0.003	0.0015	0.0012
4.0	0.0012	0.0004	0.0003
4.5	0.0005	0.0001	0.00005
5.0	0.0002	0.000025	0.00001

with the observed frequencies of events $t>t_0$ (see introduction) seems to confirm the justice of carried simulations. Hence it appears, that values given at Table 3 can be used in testing the null hypothesis described in introduction. But, after all, we must remember that all numbers quoted in Table 3 represent only the probability, and even if the "t" value obtained in a specific case is high enough, it is possible, that the correlation is really random. The "t" value can be used as a powerful tool of researcher, but a final decision will ever depend on his own judgement and experience.

ACKNOWLEDGMENTS

This study has been supported by the Central Research Project CPBP 01.06.

REFERENCES

Aniol R. W., 1983, Tree-ring analysis using CATRAS; Dendrochronologia; v. 1, 45-53

Baillie M. G. L., Pilcher J. R., 1973, A simple crossdating program for tree-ring research; Tree-ring Bull., v. 33, 7-14

Becker B., 1981, Fällungsdaten römischer Bauhölzer anhand einer 2350jährigen süddeutschen Eichen-Jahrringchronologie; Fundber. Bad.-Württ., B. 6, 369-386

Goslar T., in print, Dendrochronological studies in Gliwice Radiocarbon Laboratory, equipment, first results; Ann. Acad. Sci. Fenn.

Zieliński R., 1979, Generatory liczb losowych; WNT, Warszawa

Wpłynęło do Redakcji 3 marca 1987 r

BADANIE METODĄ MONTE CARLO ROZKŁADU PRAWDOPODOBIEŃSTWA WARTOŚCI "t" UŻYWANEJ W DATOWANIACH DENDROCHRONOLOGICZNYCH

S t r e s z c z e n i e

W celu zbadania rozkładu prawdopodobieństwa wartości "t" używanej w badaniach dendrochronologicznych przeprowadzono serię symulacji metodą Monte-Carlo. Istniejąca autokorelacja w sekwencjach indeksów powoduje istotne zwiększenie dyspersji rozkładu. Rozkład prawdopodobieństwa wartości "t" może być przybliżony przez zmodyfikowany rozkład Studenta o rozszerzonej skali. W efekcie, prawdopodobieństwo przypadkowego wystąpienia zdarzenia $t>t_0$ jest znacznie większe od przewidywanego na podstawie teoretycznego rozkładu Studenta.

ИССЛЕДОВАНИЕ МЕТОДОМ МОНТЕ КАРЛО РАССПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ ПАРАМЕТРА "t" ИСПОЛЬЗУЕМОГО В ДЕНДРОХРОНОЛОРИИ

Резюме

Для исследования свойств расспределения вероятностей параметра "t", применяемого в дендрохронологических изучениях, были проведены эксперименты методом Монте Карло. Используя результаты дендрохронологических измерений, проведеных до сих пор в гливицкой лаборатории, найдено что в секвенции индексов выступает существенна автокорреляция. Эта автокореляция является причиной истинного увеличения дисперсии расспределения вероятностей переменной "t". Обнаружено, что расспределение вероятностей переменной "t" может быть считано приблизительно равным модифицированному расспределению Студента с расширенной шкалой. Вследствие того вероятность случайного выступления события $t>t_0$ значительно больше вероятности этого события предвидыбаемой расспределением Студента.