ZESZYTY NAUKOWE POLITECHNIKI ŚLĄSKIEJ 1989

Seria: MATEMATYKA-FIZYKA z. 57 GEOCHRONOMETRIA Nr 5 Nr kol. 898

Adam WALANUS

Radiocarbon Laboratory
Silesian Technical University, Gliwice

RUNNING PHASE ANALYSIS AND ITS APPLICATION TO CYCLE SEARCHING IN LONG SERIES OF UNCERTAIN DATA

Summary: The purpose of the described method is to reveal the presence of the cyclicity in a long series of uncertain experimental data in some realistic conditions when the period has only approximately known value. Additionally the cyclicity may appear only in part of the series or in a few parts with a variable phase. The method is based on the idea of Fourier Analysis, performed, however, over few cycles in a running manner. The square wave is used instead of sine function. Data must not be equally spaced. The method is fast and needs no advanced mathematics.

1. THE FOURIER ANALYSIS AS A BASIS OF THE METHOD

High precision of the Fourier Analysis (FA) results in a few drawbacks which may be very important in analysing imprecise natural data. If, for example, evident cycles occur in the analysed series but only in a small part of it and the remaining part of the series consists of white noise then the result of FA may be insignificant. Moreover, if the really existing cycles cover the whole series but with varying phase then the result of Fourier analysis will be also insignificant. As the most disadvantageous case it may be quoted an example of the presence of distinct and precise periodicity in the whole long series but with only one fault, i.e. the gap in the middle of the series. If the duration of this gap is equal to the half of the period (the reason for such a feature may be sampling error) then the result of application of Fourier analysis in such a case will be zero for this value of the period.

The second important idea involved in the introducing of described method is connected with the concept of running statistics. Let us consider a series of data consisting of n elements. The running statistics s_i is defined as any function of subseries of k consecutive elements of this series, starting from the ith element

$$s_i = f(x_i, x_{i+1}, \dots, x_{i+k}), \qquad (1)$$

where x_j is the j-th element of the series, and k is the length of subseries. In a special case s_i may be the result of FA. For i=1 and k=n-1 there is only one value of s. For k << n, what is the usual case in application of running statistics, there is n-k ≈ n values of s (i.e. n results of FA). So large number of results seems to be completely unsuitable for drawing any definite conclusions.

In order to avoid difficulties inherent in both the classical formulation of the Fourier analysis and the method of running statistics it is proposed to search for a periodicity in a long series of uncertain data by application of the Fourier analysis for a priori chosen value of period in a running way over a subseries covering time span of few periods. Two arguments of a general nature may be quoted to support the proposed approach. Firstly, from statistical point of view the task of proving the presence of a definite feature in a given data set is quite different than the task of proving the presence of anything. Secondly, an a priori knowledge about the data is usually available.

In the classical Fourier analysis the analysed series of data is compared with the sine and cosine functions in order to approximate the series with the sum of sine functions with different amplitudes, periods and phases. In the present method it is proposed to use the digital square wave instead of the analog sine function, as is shown in Fig. 1. The general argument which may be quoted for using the ±1 function is that the computers which are commonly used today are almost exclusively digital devices. It may appear supprising that the cheapest radio receiver is able to perform FA much faster than the fastest computer. Additionally using of ±1 function instead of sine and cosine considerably saves computing time. The main argument is the simplification of statistical testing of the presence of the cycle.

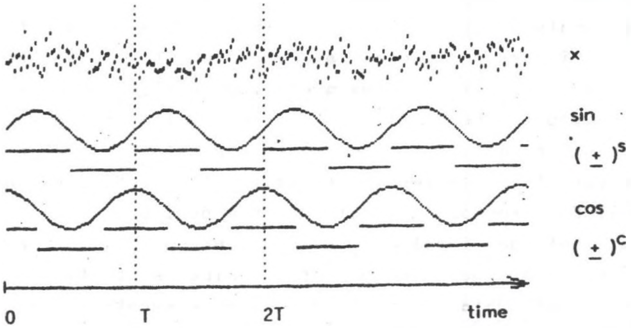

Fig. 1. Data, analog and digital periodical functions.

Rys. 1. Dane oraz funkcje okresowe (analogowa i cyfrowa).

2. THE RUNNING PHASE ANALYSIS

Because the phase of the cycle possibly existing in data is unknown, two square functions are to be used. They are analogous to sine and cosine functions in FA (see Fig. 1). The data x_i are summed over the fragment of series with the + or - sign. Let us denote

$$s = \sum_j (\pm)^s_j x_j \qquad (2a)$$

$$c = \sum_j (\pm)^c_j x_j \qquad (2b)$$

where $(\pm)^s_j$ and $(\pm)^c_j$ are equal to +1 or -1 according to Fig. 1. It may be easily proved that the function

$$a = \mathrm{ABS}(s) + \mathrm{ABS}(c) \qquad (3)$$

is independent of the phase and is proportional to the amplitude of oscillation present in the analysed part of data series. The value of the phase p may be obtained through the following algorithm:

$$\text{IF}\quad s>0 \quad \text{AND} \quad c>0 \quad \text{THEN} \quad p = -\frac{1}{4}\, T\, \frac{c}{c+s} \qquad (4a)$$

$$\text{IF}\quad s>0 \quad \text{AND} \quad c<0 \quad \text{THEN} \quad p = \frac{1}{4}\, T\, \frac{c}{c-s} \qquad (4b)$$

$$\text{IF}\quad s<0 \quad \text{AND} \quad c>0 \quad \text{THEN} \quad p = -\frac{1}{4}\, T\, \frac{c-2s}{c+s} \qquad (4c)$$

$$\text{IF}\quad s<0 \quad \text{AND} \quad c<0 \quad \text{THEN} \quad p = \frac{1}{4}\, T\, \frac{c+2s}{c+s} \qquad (4d)$$

where T is the period of the square wave.

Let us denote by s_i and c_i the values of s and c are calculated over the ith cycle

$$s_i = \sum_{j=(i-1)*T+1}^{i*T} (\pm)^s_j x_j \ , \qquad (5a)$$

$$c_i = \sum_{j=(i-1)*T+1}^{i*T} (\pm)^c_j x_j \ . \qquad (5b)$$

Calculation of s_i and c_i is repeated for all cycles of the square wave; the number of pairs (s_i, c_i) obtained in this way is equal to $m = \mathrm{INT}(n/T)$. From Eq. (3) it is possible to obtain the value of amplitude a_i of the ith cycle, and from Eqs. (4a-d) the corresponding value of phase p_i. In the next step the very informative plot of a_i and p_i values may be obtained. However, for noisy data appropriate smoothing by calculating running averages of s_i and c_i is necessary.

In analogy to filters the range of averaging of s_i and c_i may be denoted by the quality factor q, equal to the number of cycles in the moving window of the filter

$$s_i^q = \sum_{j=i}^{i+q} s_j , \tag{6a}$$

$$c_i^q = \sum_{j=i}^{i+q} c_j \tag{6b}$$

The values of a_i^q and p_i^q obtained from s_i^q and c_i^q may be treated as a results of FA performed over the fragment of the series from (i-1)*T to (i+q-1)*T. The values of a_i^q and a_{i+1}^q are not independent the resulting plot of a^q is smooth.

Choice of q is arbitrary but should depend on expected behavior of searched cyclicity with period T in the analysed series of data. Values q=1, 2, 3 may lead to spourius and accidental results. On the other hand, the value q=10 gives safe results, but may overaverage short sequences of few periods which are existing in the data.

3. EXAMPLE

The annualy laminated sediments of the Gościąż Lake consist of about 13,000 couplets (Pazdur et al. 1987). The cores are of course not continous, so the material gives numerous sequences consisting of n ≈ 2,000 data points. The laminae thicknesses were measured in Gliwice Radiocarbon Laboratory (Goslar et al., 1989) with the purpose of correlation of different cores. The laminae are sometimes not very distinctive, it is probable that two layers may be taken as one or one as two. In consequence the time scale, i.e. the number of laminae may be imprecise. The results of running phase analysis are shown in Fig. 2. The upper part of each plot shows values of amplitude a calculated according to Eq.(3), in the lower part the phase p (estimated from Eqs. (4a-d)) is plotted within boundaries ±T/2. One pixel in the horizontal scale is equivalent to one period T. Jumps of the phase may indicate errors in laminae counting. Slope of the phase plot indicates that actual period is either greater (positive slope) or smaller (negative slope) then the assumed value T. The value of slope enables to obtain in a simple way a correct value of T.

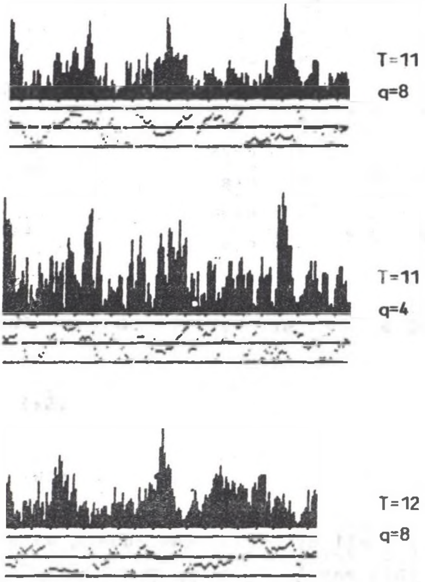

Fig. 2. Results of RPA for series of n = 1911 varves.

Rys. 2. Wyniki obliczeń RPA dla serii n = 1911 lamin.

4. BY EYE AND STATISTICAL VERIFICATION OF THE PRESENCE OF CYCLES

Many books and papers have been devoted to the subjectivity of scientific inquiery (cf Levi 1980). The personal a priori knowledge is inavoidably engaged in even most refined statistical test. Subjective impression is no less important then yes or no answer of exact test. The importance of readily appreciable picture cannot be exaggerated.

In Fig. 2 the fragments of series with high amplitude of oscillations and stable phase may be found. If the values of the period T and the phase p are known, the cycles may be plotted together with raw data. Examples are given in Fig. 3. Understanding of this picture does not need any mathematical knowledge. It should be only mentioned that the data are smoothed by running averaging with time constant t ≈ T/3. Application of low-pass filter does not impose any artificial periodicity. However, picture perception as a rule overestimates fast changes, and, in consequence, leads to subjective overestimation of periods from the band covering T.

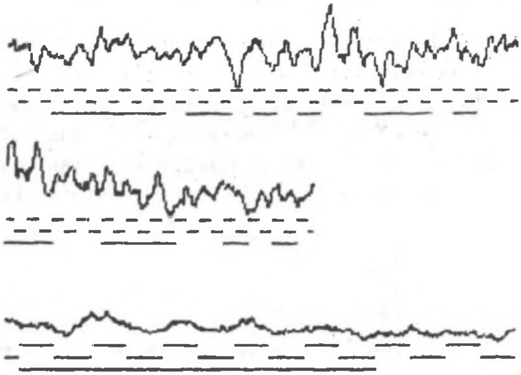

Fig. 3. Plot of same data as in Fig. 2 smoothed by running average with indicated square wave with appropriate values of the period T and phase p. Horizontal lines indicate cycles where significant agreement is observed between data and the wave predicted from Fig. 2.

Rys. 3. Wykres tych samych danych co na rys. 2 wygładzonych metodą średniej bieżącej wraz z przebiegiem fali prostokątnej o odpowiednio dobranych wartościach okresu T i fazy p. Linie poziome wskazują cykle, w których występuje istotna zgodność między danymi pomiarowymi a falą przewidywaną na podstawie rys. 2.

For all periods separately, Student "t" test is applied. The "+" and "-" parts of the period are compared by means of t statistics. The significance level was chosen equal to 0.25. In Fig. 3 the cycles with positive test result are underlined.

Because the phase of periodic function was not chosen at random the calculation of expected number of positive test results is not a trivial application of the binomial distribution. To solve this problem the Monte Carlo experiment was performed. The results are as follow: for the total number of cycles in the subsequence equal to 10 the number of "yes" test results hardly exceeds 6, and for 20 periods this limiting value is about 10-11.

Next correction, which should be taken into account is connected with the fact that analysed subsequence was not chosen at random from the long series. It is still possible to perform the Monte Carlo experiment which would be general enough to include this question, but involved a priori knowledge concerning the homogeneity of the whole sequence would be impossible to quantify. Moreover, the analysed long series is usually not independent on other series analysed in the past or to be analysed.

5. RETURN TO FOURIER ANALYSIS

If one wishes to check the presence of any period in long data series the running phase analysis still may be useful. For every T from the spectrum of periods in interest, and for given q, the sum of a_i may be calculated according to Eq.(3). On the plot of a sum versus T one may search for the peaks. It should be mentioned that the step on logarithmic scale of T is dependent on q, not on the lenght of the series as in usual FA,

$$\frac{T_{next}}{T_{prev}} - 1 \le \frac{1}{q} \quad . \qquad (7)$$

However, because $q \ll n$, the precision is much less in running phase analysis than in FA.

ACKNOWLEDGEMENTS

This study is supported by grant from the Central Research Project CPBP 01.06.

REFERENCES

Goslar T., 1987, Dendrochronological studies in the Gliwice Radiocarbon Laboratory, equipment, first results; Ann. Acad. Sci. Fenn., Ser. IIIA, vol. 145, p. 97-104.

Kendall M. G., Stuart A., 1970, The advanced theory of statistics; vol. 3, Design and analysis, and time - series; Griffin, London.

Levi I., 1980, The enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance; MIT Press, Cambridge-London.

Pazdur M. F., Awsiuk R., Goslar T., Pazdur A., Walanus A., Wicik B., Więckowski K., 1987, Calibrated radiocarbon chronology of annually laminated sediments from the Gościąż Lake; Zesz. Nauk. Pol. Śląskiej, Ser. Mat.-Fiz., z. 56, Geochronometria No. 4, p. 69-83.

Recenzent: Prof. dr S. W. Alexandrowicz

Wpłynęło do Redakcji: 10 kwietnia 1988 r.

BIEŻĄCA ANALIZA FAZOWA I JEJ ZASTOSOWANIE DO POSZUKIWANIA CYKLICZNOŚCI W DŁUGICH CIĄGACH NIEPEWNYCH DANYCH POMIAROWYCH

Streszczenie

Opisana w artykule metoda przeznaczona jest do wykrywania obecności cykliczności w ługich ciągach niepewnych danych eksperymentalnych z uwzględnieniem realistycznych warunków gdy wartość okresu znana jest jedynie w przybliżeniu. Ponadto cykliczność może występować jedynie w części ciągu danych, lub też w kilku jego częściach lecz z różniącymi się fazami. Opisana metoda opiera się na koncepcji analizy Fouriera, przeprowadzanej w sposób bieżący na odcinku obejmującym kilka cykli. Zamiast funkcji sinusoidalnych wprowadzono falę prostokątną, co znacznie przyspiesza obliczenia. Nie jest konieczne równomierne rozmieszczenie danych w analizowanej sekwencji.

ПОСЛЕДОВАТЕЛЬНЫЙ ФАЗОВЫЙ АНАЛИЗ И ЕГО ПРИМЕНЕНИЕ ДЛЯ ИССЛЕДОВАНИЯ ЦИКЛИЧНОСТЕЙ В ДЛИННЫХ ПОСЛЕДОВАТЕЛЬНОСТЯХ НЕНАДЕЖНЫХ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

Резюме

В докладе предложен метод для обнаружения цикличностей в длинной последовательности ненадежных экспериментальных данных. Метод наиболее пригоден в случае, когда значение периода искомого цикла известно только приблизительно. Кроме того показано, что цикличность может проявляться либо в части исследуемой последовательности либо в нескольких частях с отличными значениями фазы. Предложенный метод основан на общем представлении спектрального анализа Фурье примененного последовательно для интервалов с длиной в несколько циклов. Исходные данные могут быть и неравномерно расположены. Очень проста математическая процедура приводит к высокой скорости вычислений.