The Computation of Climatological Power Spectra

PAUL T. SCHICKEDANZ
Illinois State Water Survey, Urbana 61801

E. G. BOWEN[1]
Embassy of Australia, Washington, D.C. 20036

(Manuscript received 22 March 1976, in revised form 14 March 1977)

ABSTRACT

A technique for computing climatological power spectra based on the concept of utilizing non-integer values in the sine and cosine waveforms (NI technique) is developed and applied to climatological rainfall data. This technique provides a powerful alternative to the more common techniques used in the computation of climatological power spectra. The major advantage of this technique is the greatly improved resolution of wavelengths in the 5–25 year region, often a critical region of interest for climatologists. The technique produces spectral density values which are not necessarily independent; however, methods of specifying and then testing the departure from independence (orthogonality) are given. Furthermore, it is shown that the usual equations for the Fourier coefficients are special cases of the more general condition in which the spectral estimates include some degree of non-independence (i.e., lack of orthogonality). It is anticipated that this technique will have wide applicability in climatology, meteorology, hydrology and the other geophysical sciences.

1. Introduction

The computation of power spectra was systematized and standardized by the development of the Blackman-Tukey methods during the late 1950's. One of the major reasons for the wide acceptance of these methods was computational efficiency, since fewer computational operations are required with a truncated autocorrelation function (Otnes and Enochson, 1972). In the mid-1960's, high-speed algorithms for computation of power spectra (Fast Fourier Transform) received much attention. The Fast Fourier Transform (FFT) provided an efficient method of computing the finite Discrete Fourier Transform (DFT). The efficiency of the FFT made it practical to compute the spectrum directly in the frequency domain without the use of lag correlation functions (Otnes and Enochson, 1972). The introduction of the FFT has brought about a slow but steady transition from the method of lag correlation to the method of direct transformation from time to frequency via the Fourier coefficients.

However, the computation of the DFT for climatological data series creates several difficulties due to the constraint that the DFT must operate on sampled natural waveforms over relatively short, finite intervals. Furthermore, since the spacing of the DFT values is related solely to the reciprocal of the record length or period, many of the true peaks in the spectra are never sampled and the analysis is "forced" because of the discrete and fixed record length. This condition creates a serious resolution problem in various regions of the climatological spectra, particularly for wavelengths of 5–25 years.

Accordingly, an alternative method for computing the power spectra was proposed by Schickedanz and Bowen (1975). This method, identified as the Variable Record Length (VRL) technique, increases the resolution of spectra peaks by computing multiple sets of orthogonal Fourier coefficients with terms lying between the original harmonics (Schickedanz and Bowen, 1975). The VRL technique was found to provide a powerful alternative to the more common techniques used in the computation of climatological power spectra.

However, there are certain problems associated with the VRL technique. These include 1) augmentation and truncation of the data series, 2) the lack of a precise test of significance, and 3) the failure of the spectral estimates to sum to the variance of the original series. In this paper we turn our attention to another approach, the Non-Integer (NI) technique, which effectively resolves all of the above problems and, at the same time, provides a degree of resolution comparable to that obtained with the VRL technique. The NI method increases the resolution by modifying the usual Fourier equations so as to permit the computation of non-

[1] Present affiliation: Consultant, 5010 Maxwell Ave., West River, Md. 20881.

orthogonal (non-independent) coefficients using non-integer values.

2. Background assessment and the problem of resolution

We start with a climatological time series represented by a sequence $\{x_j\}$ of rainfall measurements. In the DFT method of power spectral estimation, the transformation from the time to the frequency domain is made by first obtaining the Fourier sine and cosine coefficients:

$$a_i=(2/n)\sum_{j=1}^{n} x_j \sin(2\pi ij/P), \quad i=1, 2, \ldots [n/2], \quad (1)$$

$$b_i=(2/n)\sum_{j=1}^{n} x_j \cos(2\pi ij/P), \quad i=1, 2, \ldots [n/2]-1, \quad (2)$$

$$b_{[n/2]}=(1/n)\sum_{j=1}^{n} x_j \cos(2\pi ij/P), \quad (3)$$

where n is the number of observations, P the record length or fundamental period (equal to $n\Delta x$ where Δx is the spacing of the observations in time), i the harmonic number and $[n/2]$ the integer part of $n/2$.

The raw spectral estimates are then computed by

$$S_i=C_i^2/2=(a_i^2+b_i^2)/2, \quad i=1, 2, \ldots [n/2]-1, \quad (4)$$

$$S_{[n/2]}=C_{[n/2]}^2=b_{[n/2]}^2, \quad (5)$$

where C_i is the amplitude of the ith harmonic.

The sine and cosine functions are orthogonal, i.e.,

$$\left.\begin{aligned}\sum_{j=1}^{n}\sin(2\pi ij/P)\sin(2\pi kj/P)=0\\ \sum_{j=1}^{n}\cos(2\pi ij/P)\cos(2\pi kj/P)=0\end{aligned}\right\}, \quad i\neq k,$$

where i, j and k are integer values. It can also be shown that

$$\sum_{j=1}^{n}\sin(2\pi ij/P)\cos(2\pi ij/P)=0$$

as long as i and j are integers $\leqslant[n/2]$. Since the averages of the sine and cosine functions are zero on the interval $0\leqslant j\leqslant P$, the covariances between differing sine functions on the interval $0\leqslant j\leqslant P$ and the covariances between the differing cosine functions on the interval $0\leqslant j\leqslant P$ are zero unless $i=k$. Similarly, the covariances between the sine and cosine functions are zero on the interval $0\leqslant j\leqslant P$ as long as i and j are integers. Therefore, orthogonality implies independence in the statistical sense, and the two terms are used synonymously in this paper.

The direct transform method as described in (1)–(5) is rapidly replacing the method of the truncated autocorrelation function (lagged products) since the introduction of the Fast Fourier Transform (FFT) algorithm (Cooley and Tukey, 1965). With the FFT algorithm, the direct transform method is much faster and more accurate than the lagged products method. The computation time is proportional to $n \ln n$ for the FFT and to n^2 for the older method (Endlich *et al.*, 1969). Since the relative advantage of the FFT increases for long records, it should be used instead of (1)–(5) whenever relatively long records are being computed. Furthermore, the practice of "prewhitening" (high-pass filtering) is seldom needed or used with the direct transform method (Endlich *et al.*, 1969).

It is highly desirable to have phase estimates in addition to spectral estimates. The phase angle ϕ is given by

$$\phi=\arctan(a_i/b_i). \quad (6)$$

Furthermore, it is often useful to express the original sequence $\{x_j\}$ by

$$x_j=\bar{x}+\sum_{i=1}^{[n/2]}[a_i\sin(2\pi ij/P)+b_i\cos(2\pi ij/P)]. \quad (7)$$

For ease in interpretation, the sines and cosines corresponding to a given harmonic can be combined into a single term (Panofsky and Brier, 1963). The combined term is given by

$$C_i\cos[2\pi i(j-j_i)/P]=a_i\sin(2\pi ij/P)+b_i\cos(2\pi ij/P), \quad (8)$$

where j_i is the point at which the ith harmonic has a maximum and is given by

$$j_i=\phi P/(2\pi i). \quad (9)$$

Thus, x_j can be expressed as

$$x_j=\bar{x}+\sum_{i=1}^{[n/2]} C_i\cos[2\pi i(j-j_i)/P]. \quad (10)$$

The rectangular window (boxcar function) is often used to obtain a smoothed version of the spectrum computed with a DFT. The rectangular window is applied to the data by averaging l neighboring components and the resulting estimates follow a χ^2/df distribution with 2 degrees of freedom (df), where l is the number of estimates averaged (Otnes and Enochson, 1972). Whereas the frequency resolution (bandwidth) was equal to $1/P$ before smoothing, the effective bandwidth (B_e) is now equal to l/P.

The spacing of the DFT values obtained on the frequency scale via the above relationships is given by

$$F=1/P=1/n\Delta x. \quad (11)$$

Furthermore, the spectral values are equally spaced and occur at $1/P, 2/P, 3/P, \ldots, [n/2]/P$. However, on

the wavelength scale, the values are unequally spaced and occur at $P/1$, $P/2$, $P/3$, ..., $P/[n/2]$. For example, if P is equal to 100 years, the spectral values occur at wavelengths of 100, 50, 33, 25, 20, 16.7, ... and 2 years. The spectral estimates become progressively closer together as the wavelength decreases, and the resolution is reasonably good for wavelengths <5 years. However, at the longer wavelengths there is a void of spectral estimates in the regions between 100 and 50, 50 and 33, 25 and 20, and 20 and 16.7 years; thus, important information concerning periodicities of interest is lost. For example, spectral peaks of 21 or 18.6 years (which may be important periodicities in precipitation data) would never be detected unless the record length was an exact multiple of 21 or 18.6.

As mentioned previously, a rectangular window is applied to the raw spectral estimates to obtain a smoothed version of the spectrum. This obviously decreases the resolution even more, for now the smoothed estimate represents the average power over the effective bandwidth. It is indeed a fortuitous event when a spectral estimate occurs in the vicinity of a spectral peak with a narrow B_e. The obvious solution to the resolution problem is to increase the number of spectral estimates; this can be done using either the VRL technique (Schickedanz and Bowen, 1975) or the NI technique which is discussed here.

3. The NI technique

In the usual methods of spectral estimation via the DFT, integer values are used for the harmonic number (i) within the sine and cosine arguments of (1)–(3). This "forces" the analysis in that the sine and cosine cycles must begin and terminate at the end points of the record. Furthermore, only $[n/2]$ spectral estimates can be obtained using integer values, and many of the peaks in the spectrum are never detected. If non-integer values are used for the harmonic number, one can obtain as many sine and cosine functions as needed to adequately sample the spectra. Unfortunately, (1)–(3) are valid only for integer values of the harmonic number.

The VRL technique is a solution to the resolution problem in which the record length or fundamental period (P) is varied in (1)–(3). The NI technique presented in this paper is a solution to the resolution problem in which non-integer values can be used for the harmonic number (i) to obtain as many spectral estimates as needed. Also, the sine and cosine cycles will no longer terminate at the end points of the record, so the analysis will no longer be "forced," i.e., the sine and cosine components will not be limited to frequencies that are whole-number multiples of the fundamental frequency. General equations will be presented for the computation of a_i and b_i in which i can be a non-integer value, and it will be shown that when integer values are used, (1)–(3) become special cases of the more general equations. It is noted that i is no longer strictly a harmonic number since a harmonic is defined as a whole-number multiple of the fundamental.

We first consider the following sine and cosine functions XS_{ij} and XC_{ij}:

$$XS_{ij}=\sin(2\pi ij/P), \quad j=1, 2, \ldots n, \tag{12}$$

$$XC_{ij}=\cos(2\pi ij/P), \quad j=1, 2, \ldots n, \tag{13}$$

where i can assume both integer and non-integer values, and j assumes integer values which correspond to the data points in the original data sequence $\{x_j\}$. For a given value of i, one can determine the values of the XS_{ij} and XC_{ij} functions for any x_j on the interval $0\leqslant j\leqslant n$.

There are always i cycles within the record length of $\{x_j\}$ for a given value of i, and the period of the given waveform is P/i. For integer values (harmonic values), the cycles *always terminate at the endpoints of the record length*, but for non-integer (non-harmonic) values, the cycles *never terminate at the end points of the record length.* Accordingly, one can superimpose as many waveforms (both harmonic and non-harmonic) as desired on the record length of the $\{x_j\}$.

The spectral estimates for a given waveform i are now determined by simply performing a multiple regression of the actual data sequence $\{x_j\}$ on the sine and cosine waveforms associated with the given value of i. The required multiple regression equation is given by

$$\hat{x}_j=x_{ij}=a+bs_iXS_{ij}+bc_iXC_{ij}, \quad j=1, 2, \ldots n, \tag{14}$$

where a is the intercept, and bs_i and bc_i are the partial regression coefficients. The multiple correlation coefficient R_i and the correlation coefficient r_i between the variables XS_{ij} and XC_{ij} are now determined for a given value of i. The coefficient of determination R_i^2 expresses the fractional amount of variance explained by the sine and cosine waveforms. Hence R_i^2 is approximately equal to the usual normalized spectral density associated with the ith harmonic number. If i is an integer value, R_i^2 is identical to the usual normalized spectral density.

The non-normalized spectral density C_i^2 is given by

$$C_i^2=2R_i^2S_x^2(n-1)/n, \tag{15}$$

where S_x^2 is the variance of the $\{x_j\}$. Also for any value of i,

$$C_i^2+\delta_i^2=(bs_i^2+bc_i^2), \tag{16}$$

where δ_i^2 is defined to be the power differential. The power differential represents the fractional amount of variance by which a non-independent spectral estimate differs from the value it would have if it were independent. It follows that δ_i^2 will be positive when the lack of independence produces an overestimate of the spectral density, $\delta_i^2>0$, it will be negative when the

lack of independence produces an underestimate of the spectral density, $\delta_i^2<0$. Thus, δ_i^2 is a precise estimate of the departure of the spectral estimate from orthogonality (independence). It follows that whenever r_i (the partial correlation coefficient) is zero, or small and insignificant, $C_i^2=bs_i^2+bc_i^2$.

We now proceed to show that the partial regression coefficient bs_i is equal to a_i of (1) and that the partial regression coefficient bc_i is equal to b_i of (2) when only integer values are used in (14) (i.e., when the spectral estimates are orthogonal or independent). Let the normal equations required for the estimate of bs_i and bc_i be given as follows:

$$bs_i\sum_{j=1}^{n}(XS_{ij}-\overline{XS}_{i\cdot})^2+bc_i\sum_{j=1}^{n}(XS_{ij}-\overline{XS}_{i\cdot})(XC_{ij}-\overline{XC}_{i\cdot})=\sum_{j=1}^{n}(x_j-\bar{x}\cdot)(XS_{ij}-\overline{XS}_{i\cdot}), \quad (17)$$

$$bs_i\sum_{j=1}^{n}(XC_{ij}-\overline{XC}_{i\cdot})(XS_{ij}-\overline{XS}_{i\cdot})+bc_i\sum_{j=1}^{n}(XC_{ij}-\overline{XC}_{i\cdot})^2=\sum_{j=1}^{n}(x_j-\bar{x}\cdot)(XC_{ij}-\overline{XC}_{i\cdot}), \quad (18)$$

where

$$\left.\begin{aligned}\overline{XS}_{i\cdot}&=\sum_{j=1}^{n}XS_{ij}/n\\ \overline{XC}_{i\cdot}&=\sum_{j=1}^{n}XC_{ij}/n\\ \bar{x}\cdot&=\sum_{j=1}^{n}x_j/n\end{aligned}\right\}$$

and the intercept is given by

$$a=\bar{x}\cdot-bs_i\overline{XS}_i-bc_i\overline{XC}_i. \quad (19)$$

If the variables XS_{ij} and XC_{ij} are intercorrelated, (17) and (18) are solved simultaneously for the regression coefficients bs_i and bc_i. However, if $r_i=0$, then

$$\sum_{j=1}^{n}(XS_{ij}-\overline{XS}_{i\cdot})(XC_{ij}-\overline{XC}_{i\cdot})=0$$

and bs_i and bc_i can be solved without solving a set of simultaneous equations, i.e.,

$$bs_i=\sum_{j=1}^{n}(x_j-\bar{x}\cdot)(XS_{ij}-\overline{XS}_{i\cdot})/\sum_{j=1}^{n}(XS_{ij}-\overline{XS}_{i\cdot})^2, \quad (20)$$

$$bc_i=\sum_{j=1}^{n}(x_j-\bar{x}\cdot)(XC_{ij}-\overline{XC}_{i\cdot})/\sum_{j=1}^{n}(XC_{ij}-\overline{XC}_{i\cdot})^2. \quad (21)$$

We now note that the mean values of $\sin(2\pi ij/P)$ and $\cos(2\pi ij/P)$ are equal to zero on the interval $0\leqslant j\leqslant n$ when integer values are used for i and, therefore, $\overline{XS}_{i\cdot}=\overline{XC}_{i\cdot}=0$. We further note that in (10) (which applies to integer values), $\bar{x}$ is added to the sine and cosine terms in order for the sequence to form the $\{x_j\}$. Accordingly, the sine and cosine terms fluctuate about a mean of zero ($\bar{x}\cdot=0$). Thus, (20) and (21) reduce to

$$bs_i=\sum_{j=1}^{n}x_j\sin(2\pi ij/P)/\sum_{j=1}^{n}XS_{ij}^2, \quad (22)$$

$$bs_j=\sum_{j=1}^{n}x_j\cos(2\pi ij/P)/\sum_{j=1}^{n}XC_{ij}^2. \quad (23)$$

The summation terms

$$\sum_{j=1}^{n}XS_{ij}^2 \quad \text{and} \quad \sum_{j=1}^{n}XC_{ij}^2$$

are equal to $n\,\overline{XS}^2$ and $n\,\overline{XC}^2$, respectively. Furthermore, the mean values of $[\sin(2\pi ij/P)]^2$ and $[\cos(2\pi ij/P)]^2$ are equal to $\frac{1}{2}$ over the interval $0\leqslant j\leqslant n$ (Panofsky and Brier, 1963) and, consequently,

$$\sum_{j=1}^{n}XS_{ij}^2 \quad \text{and} \quad \sum_{j=1}^{n}XC_{ij}^2$$

reduce to $n/2$. Hence (22) and (23) reduce to

$$bs_i=a_i=(2/n)\sum_{j=1}^{n}\sin(2\pi ij/P), \quad (24)$$

$$bs_j=b_i=(2/n)\sum_{j=1}^{n}\cos(2\pi ij/P). \quad (25)$$

Clearly, then, the usual equations for the Fourier coefficients [(1) and (2)] are special cases of the more general cases represented by (17) and (18). The difference between the general case and the special case is due to the lack of independence of the spectral estimates. This difference can be specified exactly in terms of the power differential δ_i^2 and r_i. Regardless of whether the spectral estimates are independent, R_i^2 is the correct value for the normalized spectral density.

The significance of R_i, r_i, bs_i and bc_i can be assessed by the usual tests employed in multiple regression and correlation analyses. Consequently, R_i^2 can be tested for significance by the F test. The value of F (Steele and Torrie, 1960) can be expressed as

$$F=\frac{[R_i^2/k]}{[(1-R_i^2)/(n-k-1)]}, \quad (26)$$

where k is equal to the number of independent variables ($k=2$ in this case). The F table (Snedecor and Cochran, 1967) can then be used to obtain the F_α, the probability of obtaining a multiple correlation coefficient greater than that observed from random sampling of a population in which the true $R^2=0$.

There are tables (Steele and Torrie, 1960) for the 0.05 and 0.01 levels of significance by which this test can be applied directly if one knows R_i and the degrees of freedom $(n-k-1)$.

The partial correlation coefficient r_i can be tested for significance by the t test. The value of t (Steele and Torrie, 1960) can be expressed as

$$t(n-k-1)=\frac{r_i(n-k-1)^{\frac{1}{2}}}{(1-r_i^2)^{\frac{1}{2}}}. \tag{27}$$

The t table can then be used to obtain t_α, the probability of obtaining a partial correlation coefficient greater than that observed by random sampling from a population in which the true $r=0$. Again, there are tables (Steele and Torrie, 1960) for the 0.05 and 0.01 levels of significance by which this test can be applied directly if one knows r_i and the degrees of freedom. Tests of significance for the partial regression coefficients bs_i and bc_i are given by Snedecor and Cochran (1967).

If r_i is insignificant, then the spectral estimates are independent and the relationships pertaining to independent spectral estimates can be used, including (10). Since bs_i and bc_i are equal to a_i and b_i, respectively, for orthogonal conditions, the test provided by Snedecor and Cochran (1967) for partial regression coefficients can be used to test the significance of a_i and b_i. If r_i is significant, then the spectral estimates are non-independent, and the test for partial regression coefficients is only applicable to bs_i and bc_i.

As mentioned previously, (10) is no longer valid for non-independent spectral estimates, and the various waveforms will not sum to the original sample variance. In order to form the original $\{x_j\}$ or to make a valid extension of the combined waveforms into future years, it is necessary for the waveforms to sum to the proper variance. We now present a method for obtaining the proper summation of waveforms for the non-independent case. It will be shown that this method reduces to (10) when the spectral estimates are independent (i.e., r_i is insignificant).

We assume that 1, 2, 3, ..., m *non-independent spectral estimates* ($2m$ waveforms) were determined to be significant at the α level of significance. We can then determine a relationship for the addition of a set of non-integer waveforms by performing a multiple regression of the original data $\{x_j\}$ on the $2m$ waveforms. The required regression equation is given by

$$\hat{x}_j=ar+bs_1XS_{1j}+bc_1XC_{1j}+bs_2XS_{2j}+bc_2XC_{2j} +\cdots bs_mXC_{mj}+bc_mXC_{mj}, \tag{28}$$

where 1, 2, 3 ..., m pertain to non-independent spectral esimates and

ar is the intercept

bs_1, bc_1, ..., bs_m, bc_m are the partial regression coefficients

$$XS_{1j}=\sin(2\pi 1j/P) \quad XC_{2j}=\cos(2\pi 2j/P)$$
$$XC_{1j}=\cos(2\pi 1j/P) \quad XS_{mj}=\sin(2\pi mj/P)$$
$$XS_{2j}=\sin(2\pi 2j/P) \quad XC_{mj}=\cos(2\pi mj/P).$$

The normal equations required for the estimates of the partial regression coefficients are

$$bs_1\sum_{j=1}^{n}xs_{1j}^2+bc_1\sum_{j=1}^{n}(xs_{1j}xc_{1j}) +bs_2\sum_{j=1}^{n}(xs_{1j}xs_{2j})+bc_2\sum_{j=1}^{n}(xs_{1j}xc_{2j}) +\cdots+bs_m\sum_{j=1}^{n}(xs_{1j}xs_{mj}) +bc_m\sum_{j=1}^{n}(xs_{1j}xc_{mj})=\sum_{j=1}^{n}(x_j-\bar{x}\cdot)xs_{1j}, \tag{29}$$

$$bs_1\sum_{j=1}^{n}(xc_{1j}xs_{1j})+bc_1\sum_{j=1}^{n}xc_{1j}^2 +bs_2\sum_{j=1}^{n}(xc_{1j}xs_{2j})+bc_2\sum_{j=1}^{n}(xc_{1j}xc_{2j}) +\cdots+bs_m\sum_{j=1}^{n}(xc_{1j}xs_{mj}) +bc_m\sum_{j=1}^{n}(xc_{1j}xc_{mj})=\sum_{j=1}^{n}(x_j-\bar{x}\cdot)xc_{1j}, \tag{30}$$

$$\vdots$$

$$bs_1\sum_{j=1}^{n}(xs_{mj}xs_{1j})+bc_1\sum_{j=1}^{n}(xs_{mj}xc_{1j}) +bs_2\sum_{j=1}^{n}(xs_{mj}xs_{2j})+bc_2\sum_{j=1}^{n}(xs_{mj}xc_{2j}) +\cdots+bs_m\sum_{j=1}^{n}xs_{mj}^2 +bc_m\sum_{j=1}^{n}(xs_{mj}xc_{mj})=\sum_{j=1}^{n}(x_j-\bar{x}\cdot)xs_{mj}, \tag{31}$$

$$bs_1\sum_{j=1}^{n}(xc_{mj}xs_{1j})+bc_1\sum_{j=1}^{n}(xc_{mj}xc_{1j}) +bs_2\sum_{j=1}^{n}(xc_{mj}xs_{2j})+bc_2\sum_{j=1}^{n}(xc_{mj}xc_{2j}) +\cdots+bs_m\sum_{j=1}^{n}(xc_{mj}xs_{mj}) +bc_m\sum_{j=1}^{n}xc_{mj}^2=\sum_{j=1}^{n}(x_j-\bar{x}\cdot)xc_{mj}, \tag{32}$$

where xs_{1j}, xc_{1j}, xs_{2j}, xc_{2j}, ..., xs_{mj}, xc_{mj} are departures from their respective means. The intercept is given by

$$ar=\bar{x}\cdot-bs_i\overline{XS}_{1j}-bc_i\overline{XC}_{1j}\ldots-bs_m\overline{XS}_{mj}-bc_m\overline{XC}_{mj}. \tag{33}$$

Hence, for non-independent spectral estimates, the normal equations (29)–(32) must be solved simultaneously for the partial regression coefficients. This requires a new solution of a set of simultaneous equations each time a different set of wave forms is summed. Once the regression coefficients and the intercept have been estimated, (28) can be used to form an estimate of the original $\{x_j\}$. An extension of the $\{x_j\}$ into future years can be made by simply substituting into (28) the j values for data points beyond the record length of the original $\{x_j\}$.

If the waveforms were determined from integer values (i.e., 1, 2, 3, ..., *m now pertain to independent spectral estimates*), all of the cross-product terms i.e.,

$$\sum_{j=1}^{n}(xs_{1j}xc_{1j}), \text{ etc.}$$

on the left side of the normal equations are zero and the normal equations reduce to

$$bs_1\sum_{j=1}^{n}xs_{1j}^2=\sum_{j=1}^{n}(x_j-\bar{x}\cdot)xs_{1j}, \tag{34}$$

$$bc_1\sum_{j=1}^{n}xc_{1j}^2=\sum_{j=1}^{n}(x_j-\bar{x}\cdot)xc_{1j}, \tag{35}$$

$$\vdots$$

$$bs_m\sum_{j=1}^{n}xs_{mj}^2=\sum_{j=1}^{n}(x_j-\bar{x}\cdot)xs_{mj}, \tag{36}$$

$$bc_m\sum_{j=1}^{n}xc_{mj}^2=\sum_{j=1}^{n}(x_j-\bar{x}\cdot)xc_{mj}. \tag{37}$$

In a manner similar to that employed with (17)–(24), it can be shown that (34)–(37) reduce to

$$bs_1=(2/n)\sum_{j=1}^{n}\sin(2\pi 1j/P), \tag{38}$$

$$bc_1=(2/n)\sum_{j=1}^{n}\cos(2\pi 1j/P), \tag{39}$$

$$\vdots$$

$$bs_m=(2/n)\sum_{j=1}^{n}\sin(2\pi mj/P), \tag{40}$$

$$bc_m=(2/n)\sum_{j=1}^{n}\cos(2\pi mj/P). \tag{41}$$

Thus, (38)–(41) are independent of each other and new coefficients can be estimated by simply summing the appropriate sine and cosine functions. It is not necessary to solve a new set of simultaneous equations each time a new harmonic is added to the set of harmonics being summed. In fact, bs_1, bc_1, ..., bs_m, bc_m can be estimated by (1) and (2), and (28) is identical to (10). The choice of whether to use (28) and (29)–(33) instead of (1), (2) and (10) is based on the significance or lack of significance of r_i for all values of i used in the calculation of the spectrum. If r_i is insignificant for all values of i, then the equations for the independent case should be used when summing harmonics, since they are easier to compute and are also less time-consuming.

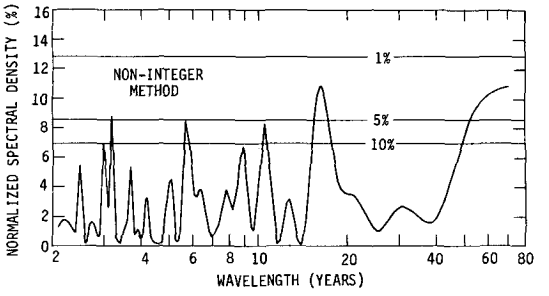

FIG. 1. The July rainfall spectrum for Amarillo, Tex. (1895–1964), as estimated by the Non-Integer (NI) spectral technique. All wavelengths >35 years are a measure of persistence and can be attributed to long-term trends or to very long cycles.

The large amount of resolution obtained is the chief advantage of applying the NI technique. It is recommended that the non-integer spectral estimates be determined at equal intervals of wavelength rather than frequency, a departure from the usual practice. It is further recommended that the spacing of the estimates be on the order of 0.1 data point. Hence, a spectrum for a 100 year record length of annual precipitation values will have a total of 1000 spectral estimates, and the wavelengths will be determined to the nearest 0.1 year.

Another advantage of the NI technique over the VRL technique is that a smoothed spectrum is produced so that further smoothing is neither essential nor desirable. One does not need "average power over a spectral band" because the large degree of resolution permits an approximation to a continuous spectrum, and this "approximation" is very smooth (see Fig. 1).

The NI method has further advantages in that 1) a valid test of significance is available and significance is easy to assess, and 2) a valid method of combining the non-independent waveforms is available so that an appropriate estimate can be made of the original data series or some future projection.

4. Application of the NI technique to climatological data series

The results of applying the NI technique to the July rainfall series of record length 70, 1895–1964, for

Amarillo, Tex. are shown in Fig. 1. The NI spectrum was calculated by determining bs_i and bc_i with (17)–(18) and then determining R_i^2, the normalized spectral density, by the usual methods of multiple regression and correlation. The non-normalized spectral density was computed by (15) and the power differential was computed by (16). The partial correlation coefficient r_i between the sine and cosine waveforms was also determined for each spectral estimate to assess which estimates were independent. This coefficient was small and insignificant for each spectral estimate. This indicated that, for practical purposes, the spectral estimates possess the property of orthogonality (independence). The spectral estimates were determined for each 0.1 data point, which provided 700 estimates of the spectra with wavelengths resolvable to 0.1 year. The computational time was 24.2 s of CPU time on the IBM 360-75 system.

There are significant peaks in the spectrum at wavelengths of 3.0, 3.2, 5.7, 10.5, 16.3 and 70 years. The presence of persistence (red noise) in the spectra is depicted by the significant spectra peak at the long wavelength of 70 years. Since all wavelengths greater than 35 years consist of only one cycle, it is difficult to determine whether the persistence can be attributed to a long cycle or to a long-term trend. For wavelengths <35 years, two or more cycles are present in the sample. As the number of cycles increases toward shorter wavelengths, the degree of confidence in the stability of the spectral peaks also increases. *Nevertheless, it should be noted that the NI technique is a powerful method for detecting long-term trends or cycles.* Furthermore, the order of trend (i.e., first order, second order, etc.) need not be assumed prior to the analysis since the sine and cosine waveforms will not only detect the trend or cycle but will also specify the mathematical form. For trends and cycles at the longer wavelengths, the sine and cosine waveforms will approximate a low-term trend with curvature. It is noted that many rainfall spectra do not possess persistence, and this lack of persistence is indicated by the June spectra on Fig. 2.

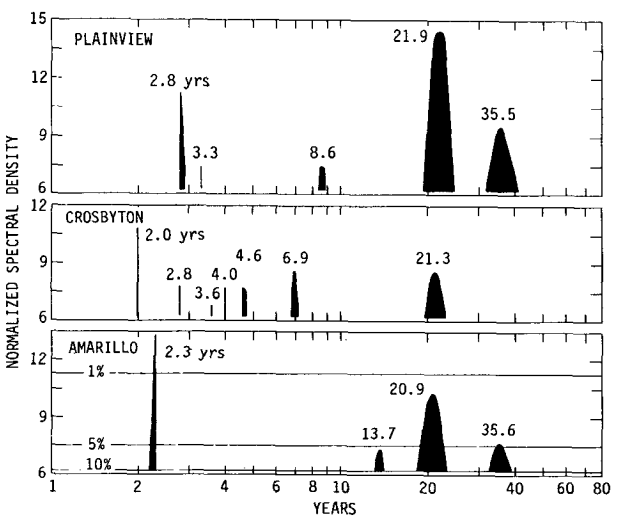

FIG. 2. A comparison of June rainfall spectra for three stations in the Texas High Plains over the period 1895–1974.

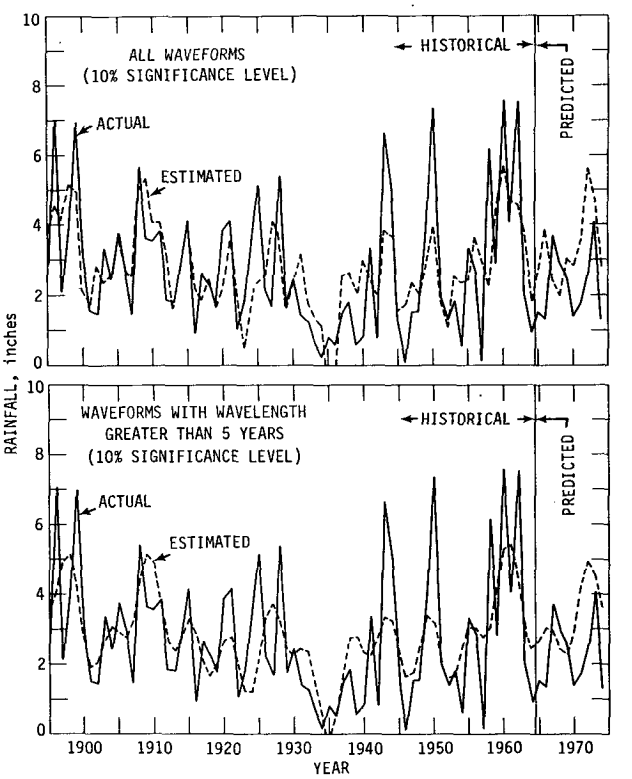

FIG. 3. The estimated and actual rainfall series for historical and predicted periods during July at Amarillo.

The significance and stability of the peaks may also be considered by comparing their stability for different stations over the same period of time. June rainfall spectra for three stations lying along a 120 mi north-south line in the Texas High Plains were computed and the spectral estimates exceeding the 10% level of significance are shown on Fig. 2. The striking aspect of these spectra is the consistent behavior of the major spectral peaks in the range of 20.9 to 21.9 years. In order to detect these peaks with the FFT or the lagged-products methods sample sizes that are exact multiples of 20.9, 21.3 and 21.9 would have been required.

Another way to test the reality of the spectral peaks in the actual data series (historical period) is to extend the waveforms into an independent period (future period). Such an extension was made by first selecting the amplitudes, phases and i values associated with those spectral peaks significant at the 10% level in the July rainfall series (Fig. 1). The projected rainfall series for the historical period (1895–1964) and the predicted period (1965–1974) was made by using (28) and (29)–(33). [It is noted that the projection could have been made with (1), (2) and (10) in this case

since the spectral estimates, for all practical purposes, were independent.]

Two projections were made: the first consisted of all of the significant waveforms, whereas the second consisted of only those waveforms with wavelengths >5 years. The projected rainfall series were compared to the actual rainfall series during the historical and predicted periods (see Fig. 3). The general agreement between the actual and projected rainfall series during the historical and predicted periods illustrates their stability and significance.

A final point needs to be made with regard to the significance of the spectral estimates. Some authors (Mitchell, 1966; Madden and Julian, 1971) carefully distinguish between the choice of testing a single spectral wavelength before the data are analyzed (*a priori*) and testing high peaks observed after the data have been analyzed (*a posteriori*). It is indicated that the confidence bands for the *a posteriori* case are much broader than those for the *a priori* case to allow for the fact that some high peaks will occur by chance. This would suggest that the confidence levels shown on Fig. 1 should be much higher, and it is conceivable that none of the spectral peaks would be high enough to equal or exceed the higher (*a posteriori*) confidence levels.

It is noted, however, that this distinction has been applied to spectra with resolution similar to the spectra estimated with the FFT method. It is not obvious that this distinction could or should be applied to the VRL and NI spectra, which have a much higher degree of resolution. Furthermore, many of the spectral peaks (in particular, those of shorter wavelengths) have been noted by other investigators in studies of temperature, precipitation and pressure data (Brier, 1968; Hay, 1970; Schonwiese, 1971; Velyazo, 1972). In addition, the other spectral peaks appear to be associated with certain physical, lunar, solar and terrestrial factors. (These factors and their possible influence on precipitation and temperature will be the subject of a forthcoming paper.) The important point to note here is that there is sufficient evidence to suggest that the *a priori* confidence band is the appropriate one to apply to the NI technique.

5. Concluding discussion

A method for computing climatological power spectra using the Non-Integer (NI) technique has been developed and applied to historical rainfall data. This technique provides a powerful alternative to the more common techniques used in the computation of climatological power spectra. The major advantage of the technique is the greatly improved resolution of wavelengths in the 5–25 year region, often a critical region of interest for climatologists. In addition, there are several other advantages: 1) a smooth spectrum is produced such that further smoothing is neither essential nor desirable, 2) a valid test of significance is available and significance is easy to assess, and 3) a valid method of combining the non-independent waveforms is available so that an appropriate estimate of the original data series can be made.

Furthermore, it was shown that the usual equations for the Fourier coefficients are special cases of the more general condition in which the spectral estimates include some degree of correlation (lack of orthogonality). A method is presented in which the departure of the spectral estimates from orthogonality (or independence) can be specified (i.e., the power differential). In addition, a test is available for assessing the significance of the departure from orthogonality.

Acknowledgments. This research was performed under the general direction of Stanley A. Changnon, Jr., Head of the Atmospheric Sciences Section of the Illinois State Water Survey. The authors would like to express their appreciation to Marion B. Busch for her pertinent comments and suggestions as well as her role in the development of the required computer programs. Further appreciation is expressed to Dr. Robert E. Bohrer, University of Illinois, and to Mr. Glenn W. Brier, Colorado State University, who first suggested to the authors the possibility of using non-integer values in the calculation of the sine and cosine waveforms. In addition, appreciation is expressed to Dr. Richard H. Jones, University of Colorado Medical Center, who called the authors' attention to the distinction between *a priori* and *a posteriori* testing of spectral peaks. Finally, appreciation is expressed to Eugene A. Mueller of the Water Survey staff for his review of this paper. This research was supported under Grants GI-43871 and AEN 73-07796 sponsored by the Weather Modification Program, RANN, National Science Foundation and ATM 76-11379 sponsored by Climate Dynamics, National Science Foundation.

REFERENCES

Brier, G. W., 1968: Long-range prediction of the zonal westerlies and some problems in data analysis. *Rev. Geophys.*, **6**, 525–551.

Cooley, J. W., and J. W. Tukey, 1965: An algorithm for the machine calculation of complex Fourier series. *Math. Comput.*, **19**, 297–301.

Endlich, R. M., R. C. Singleton and J. W. Kaufman, 1969: Spectral analysis of detailed vertical wind speed profiles. *J. Atmos. Sci.*, **26**, 1030–1041.

Hay, R. F. M., 1970: Further analysis of monthly mean pressure patterns near the British Isles, 1874–1968. *Meteor. Mag.*, **99**, 189–198.

Madden, R. A., and P. R. Julian, 1971: Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific. *J. Atmos. Sci.*, **28**, 702–708.

Mitchell, J. M., Jr., 1966: Climatic change. WMO Tech. Note No. 79, 79 pp.

Otnes, R. K., and L. Enochson, 1972: *Digital Time Series Analysis.* Wiley-Interscience, 467 pp.

Panofsky, H. A., and G. W. Brier, 1963: *Some Applications of Statistics to Meteorology.* Pennsylvania State University Press, 224 pp.

Schickedanz, P. T., and E. G. Bowen, 1975: Computation of climatological power spectra using variable record lengths. *Preprints 4th Conf. Probability and Statistics*, Tallahassee, Fla., Amer. Meteor. Soc., 104–109.

Schonwiese, C. D., 1971: Synoptic-statistical methods of forecasting the mean monthly air temperature and the quantity of precipitation during the winter months for the Far East. *Meteor. Gidrol.*, **11**, 32–37.

Snedecor, G. W., and W. G. Cochran, 1967: *Statistical Methods*, 6th ed. The Iowa State University Press, 593 pp.

Steele, R. G. D., and J. H. Torrie, 1960: *Principles and Procedures of Statistics*. McGraw-Hill, 479 pp.

Velyazo, V. A., 1972: Intrayear recurrence of atmospheric circulation forms and changes of monthly pressure in 22-year cycles. *Arktich. Antarktich. Nauchno-Issled. Inst., Tr.*, No. 313, 209–224.