State Water Survey Division
ATMOSPHERIC SCIENCES SECTION
AT THE
UNIVERSITY OF ILLINOIS

Illinois Institute of Natural Resources

PREDICTION OF RAINFALL TRENDS

by

James C. Neill
Principal Investigator

FINAL REPORT - PHASE II

to

Climate Dynamics Office
National Science Foundation

NSF Grant ATM 76-11379

June 1980

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF FIGURES IN THE APPENDIX

LIST OF TABLES

ACKNOWLEDGMENTS

The research on which this report is based was supported by NSF Grant No. ATM 76-11379. Additional support came from the State of Illinois.

The research was performed under the general guidance of Stanley A. Changnon, Jr., Head, Atmospheric Sciences Section, Illinois State Water Survey. The contribution of Dr. Paul T. Schickedanz, Atmospheric Sciences Section Meteorologist and Statistician, in the early phases of this research is acknowledged. The support and advice of Floyd A. Huff, Senior Meteorologist, Atmospheric Sciences Section, is greatly appreciated. Edward G. Bowen, a consultant, contributed as an advisor on the digital filtering and analysis aspects of the prediction process. Dr. W. K. Jenkins, University of Illinois Electrical Engineering Dept., a consultant, performed the digital filtering study described in the Appendix. David A. Brunkow of the Atmospheric Sciences Section staff assisted with computer programming. Pramella Reddy, Merry Johnson, and Rebecca Loh of the Survey staff performed the computer analysis required. Marvin Clevenger supervised the key punching of climatic data that was not previously available in machine readable form. Appreciation is expressed to John W. Brother, Jr., who supervised the preparation of diagrams in this report, and to Debbie K. Hayn for an excellent job of typing this report.

INTRODUCTION

Purpose and Scope

The research discussed in this final report under NSF Grant ATM 76-11379 has involved a 3-year, 2-phase program whose major objective has been to develop methods which will provide useful estimates of future variations in crop production due to weather. Research' was centered on the 5-state Corn Belt (Fig. 1) consisting of Iowa, Missouri, Illlinois, Indiana, and Ohio. The major objective of Phase I was to develop methods which will provide quantitative estimates of future time-space variations in crop production that are caused by natural fluctuations in agriculturally-relevant climatic factors Emphasis was concentrated on defining the effect of these natural weather fluctuations on corn and soybean yields for one to five consecutive growing seasons. Time-space relationships were expressed in terms of probability estimates for areas of various sizes and time periods. Results and discussion for Phase I are presented in a separate report entitled "Assessment of Effects and Predictability of Climate Fluctuations as Related to Agricultural Production by F. A. Huff and J. C. Neill."

The objective of Phase II research discussed in this report has been to investigate, develop, and test statistical methods for the predictions of trends in rainfall for one to three years in advance with emphasis on annual and agriculturally relevant seasonal predictions. Efforts have been concentrated on application of methods for computing climatological oscillations that occur with a considerable degree of regularity. For this purpose, spectrum analysis followed by bandpass and filtering has been the primary investigative procedure. The filtering technique was originally employed with considerable success by Bowen (1967 unpublished) in predicting

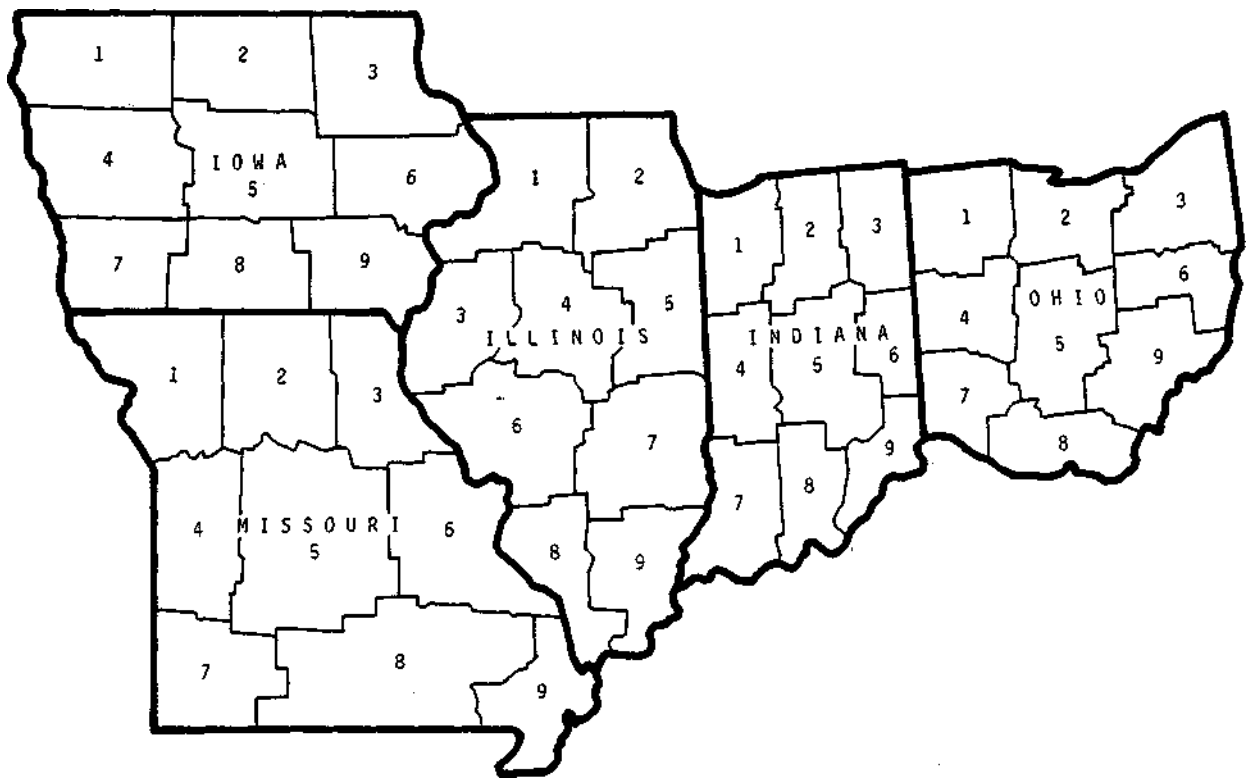

Figure 1. Research study area

annual precipitation trends in Australia. Since crop yields are strongly related to growing season precipitation, successful trend prediction for periods of one to three years in advance would provide a major component for crop planning and estimation of future yields. Therefore, emphasis was placed on prediction of seasonal and annual precipitation trends.

Part of the analysis process (noted above) involved the application of non-recursive symmetrical digital filters to determine wavelengths and amplitudes of underlying periodicities in a precipitation data series. As a part of the subsequent prediction processes, a periodicity with wavelength and amplitude determined near the end of the data series was projected into the future. A weakness of filtering in this application is due to the fact that half of the filter coefficients extend beyond the data series at the conclusion of the filtering process. This results in an unknown amount of "attenuation" of the amplitude near the end of the data record and raises concern about the validity of subsequent prediction results. Dr. W. K. Jenkins, University of Illinois Electrical Engineering Department, was employed as a consultant to study ways to lessen the severity of the filter attenuation problem. Results of his research are summarized in the appendix of this report.

Data Used in Studies

Average monthly temperature and monthly precipitation totals for each of the 45 crop reporting districts in the 5-state area for 1931 through 1975 were obtained from the Asheville Climatic Center of the National Weather Service. Average monthly temperature and monthly precipitation totals for each district were computed from station data for the period 1901-1930. Crop yield data were obtained through the agricultural departments of the several states.

SPECTRAL ANALYSIS PROCEDURES

Introduction

Tims series data analysis techniques were used to search for predictive power in seasonal and annual rainfall totals. The first step in predicting with time series analysis techniques involves a search for periodicities (non-random fluctuations or oscillations) in the historical rainfall data. Periodicities, if present, represent potential predictive power for anticipated variations in rainfall trends (up or down) during a future time period. The second step in the time series prediction approach used in this study involved the projection of periodicities (harmonics) into the future.

Nonrandom fluctuations in precipitation, temperature, and crop yield are assumed due to conditions (terrestrial, extraterrestrial, atmospheric, etc.) that influence variation over a large area (several adjacent districts) in a similar manner. Integration of precipitation experience over time and space, such as summing over a crop exposure season and averaging over a district or larger area, is necessary to reduce, or filter, the random component in individual storm events. Random components interfere with possible identification of periodicities.

A pertinent question the research faced was that of how much integration was required. It was, therefore, assumed that adjacent district average records should have similar temporal characteristics (trends) which would become evident in the spectra and correlations among district data sets within a state (or region) of similar size. Any fluctuations without some regional homogenity were considered random events. Thus, while individual events appear rather random in their occurrence in space and time, the underlying forces which produce them have to vary regionally from season-to-season if

periodicities are to be a reality in historical data. The establishment of periodicities in historical data provide the basis for trend predictions, assuming periodicities will continue into the future with similar wavelength and amplitude.

Mathematical Background

A climatological time series (x_t) can be expressed mathematically by:

$$x_t = \bar{x} + \sum_{i=1}^{N/2} C_i \cos\left(2\pi i(t - t_i)/T\right), \qquad (1)$$

where the x_t are a sequential series of precipitation measurements taken at times t = 1, 2, 3 ---_N, $\bar{x}$ is the average of the x_t series, i is an integerfrom 1 to N/2, the C_i are amplitudes of the N/2 cosine waves, the t_i are phase angles for each cosine wave, and T is the length of the observation period (T = N Δt equally spaced observations). The constant terms, x, C_i, and t_i, are to be determined by the method of "least squares." However, these constants cannot be readily determined with equation 1 in its' present form. The differences of cosine terms (cos $2\pi it$ - cos $2\pi it_i$) must be replaced by the trigonometric addition identity to obtain x_t as:

$$x_t = x + C_i \left[(\sin 2\pi it \sin 2\pi it_i + \cos 2\pi it \cos 2\pi it_i)/T\right] \qquad (2)$$

from which

$$x_t = x + \left[a_i \sin 2\pi it + b_i \cos 2\pi it)/T\right] \qquad (3)$$

is obtained where

$$a_i = C_i \sin 2\pi it_i$$

and

$$b_i = C_i \cos 2\pi it_i$$

Equation (3) is now entirely in terms known from the observed data except for the values of constants, x, a_i and b_i, which can be determined by utilizing the concept of least squares. Thus, the time series (x_t) is equal to the average plus the sum of N/2 harmonics (sine and cosine terms). More complete presentations of harmonic analysis are available from Diachisin (1957), Panofsky and Briar (1958), Schickedanz and Bowen (1977) and others.

Non-integer Spectral Algorithm

The search for periodicities was accomplished with a non-integer spectral analysis algorithm developed by Schickedanz and Bowen (1977). The algorithm is a set of computer instructions designed to test the "goodness of fit" of sine-cosine waveforms to historical data. This spectral analysis technique was employed because it is capable of performing a very thorough search for periodicities. The non-integer feature permits a search for wavelengths which are not a multiple, as well as for those wavelengths for which the observation period is a multiple. Wavelengths resolvable to 0.1 year (or 0.1 the time unit of data measurement) were searched for. Figure 2 shows a sample of the many sine-cosine waveforms which were correlated (using multiple regression techniques) with historical data in the search for periodicities.

The algorithm determines a multiple correlation coefficient for the dependent variable (annual precipitation totals in this case) with each waveform. Thus, the multiple correlation between the precipitation variable and the sine-cosine waveforms was the statistic employed to determine evidence of a spectral peak or periodicity. These correlations were plotted versus the period (wavelength) of the waveforms (Fig. 3a) for Illinois crop district number 5 (east central Illinois), as a visual representation of spectral outputs. The amplitude of each waveform is also plotted (Fig. 3b) for a

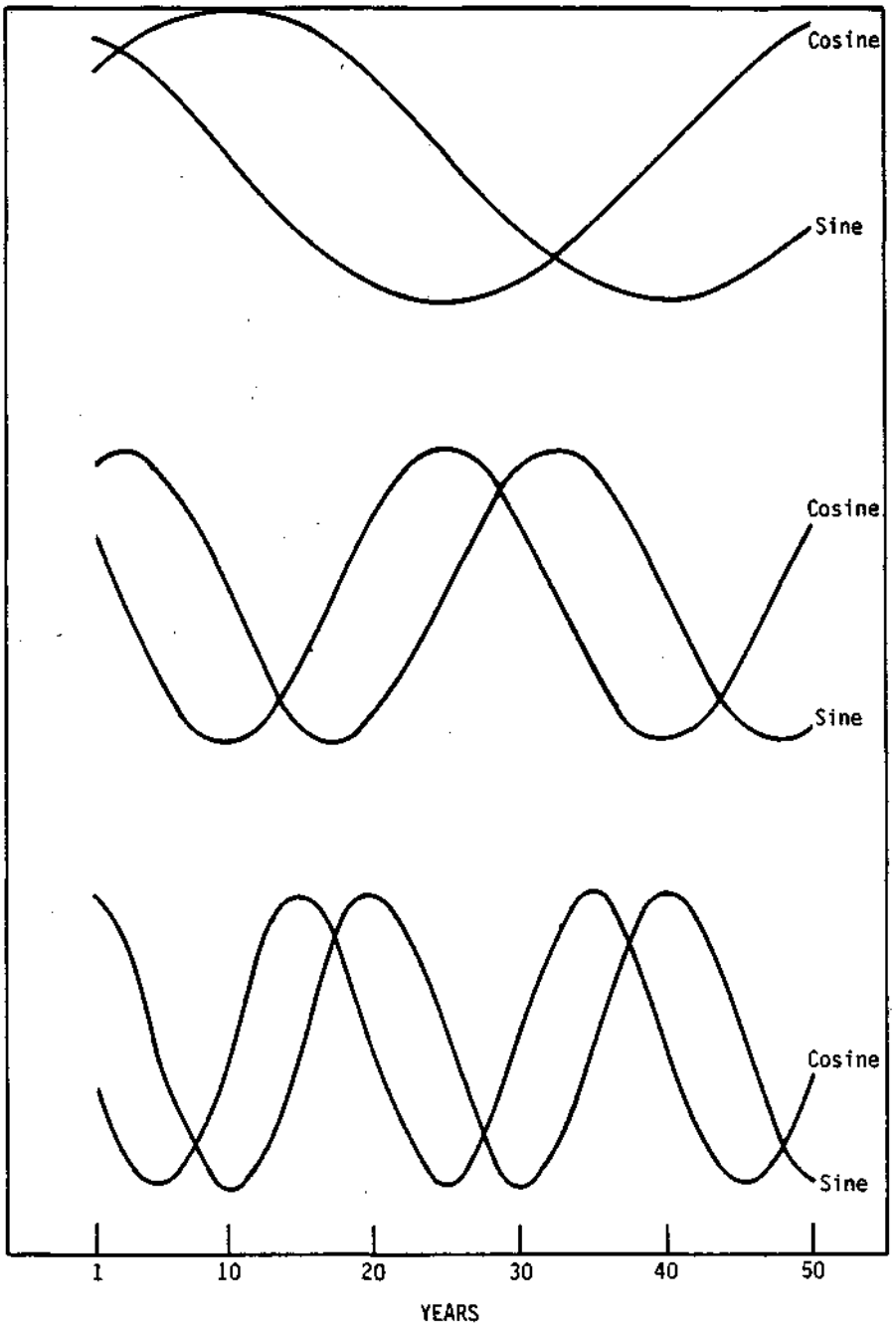

Figure 2. Example of sine-cosine waveforms used as predictor variables in multiple regression with rainfall.

visual representation of the association between amplitudes and associated correlations for the waveforms. Plots of the two statistics have the same form and either statistic can be used to depict the annual precipitation spectrum. The square of either the amplitude or the correlation provide a measure of the total variance explained by each waveform. The correlation (a commonly used measure of association) will be used in this report for the selection of waveforms to use as predictors. The amplitude of the waveforms will be directly involved in the prediction procedure.

Resolution of Periods

Only periods of 2 years to 10 years are displayed in Figures 3a and b. Periods shorter than 2 years are not resolvable with annual data. The resolution of periodicities longer than 10 years is questionable from base records no longer than 45 years. Experimentation (see a later section on length of observation requirements) with the non-integer program suggested a data record length of 4 to 5 times the period of oscillation is needed for accurate determination of the wavelength. Thus, the resolution of periodicities greater than 10 years is not considered reliable from observation periods of 45 years.

Statistical Significance of Periodicities

Examination of Illinois district 5 spectrum (Fig. 3a) reveals several peaks in the correlation trace between 2 and 10 years. Hence, a procedure was needed to select the more meaningful peaks (periodicities) and to reject those less meaningful. The 10 percent level of statistical significance was selected as a guideline (Fig. 3a) for accepting a periodicity for use in prediction. With a 45-year record (sample size), a correlation of 0.32 is

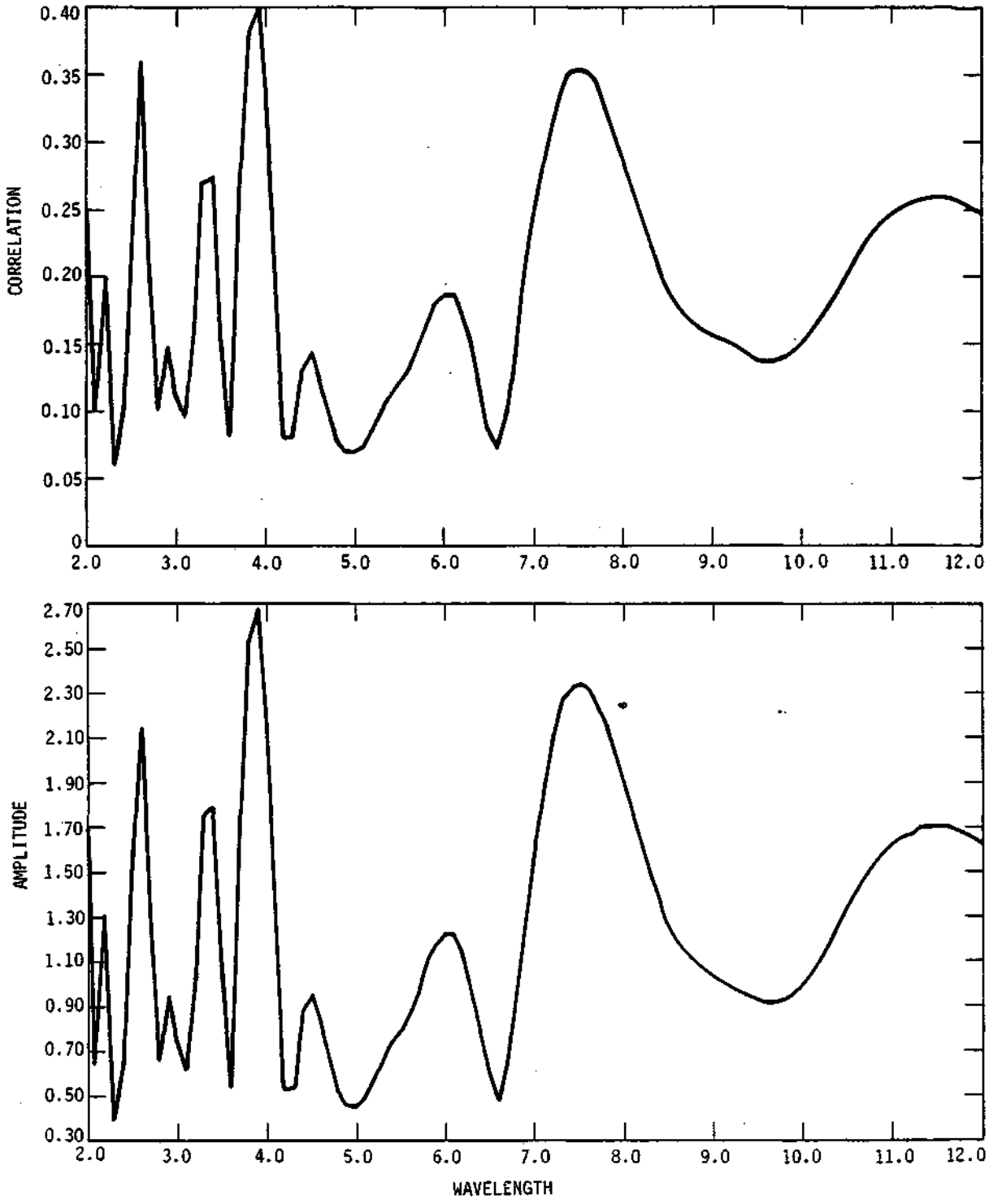

Figure 3. Multiple correlation and amplitude of harmonics with annual precipitation totals for Illinois crop reporting district No. 5, 1931-1975.

needed for significance at the 10 percent level (odds are 1 in 10 that the peak is real rather than chance variation). Another expression of the importance of a periodicity chosen in this manner is the fact that the square of the correlation represents that portion of the variance of the annual precipitation totals for the 45-year sampling (measurement) period. For example, a correlation of 0.30 accounts for 9 percent of the annual precipitation variation (variance). Illinois district 5 had three peaks (2.6, 3.9, and 7.4-year periodicities, Fig. 3a) which exceeded the rather moderate 10 percent level of statistical significance.

Spatial Spectral Variation

Annual precipitation totals for 1931-1975 from each of the 45 districts of the 5-state study area were "normalized" by dividing each by the district average of the analysis period. Normalization adjusted the amplitude of periodicities to the climatological base experienced within each district. Wavelength, phase point and the correlation associated with each periodicity were not affected by normalization. Normalized precipitation values were subjected to a spectrum analysis, 1) to search for significant periodicities, and 2) to determine the degree of areal coherence across the study area. Spectral coherence for one or more periodicities from adjacent districts is considered a very important condition for the establishment of a potential prediction technique with periodicities as predictors.

Computer outputs of spectral computations were examined for evidence of periodic variations which had similar frequency of occurrence in adjacent districts. Intuitively, spatial frequency coherence should strengthen the agrument for using periodic waveforms as predictors of precipitation for either individual districts or in larger (combination of adjacent districts)

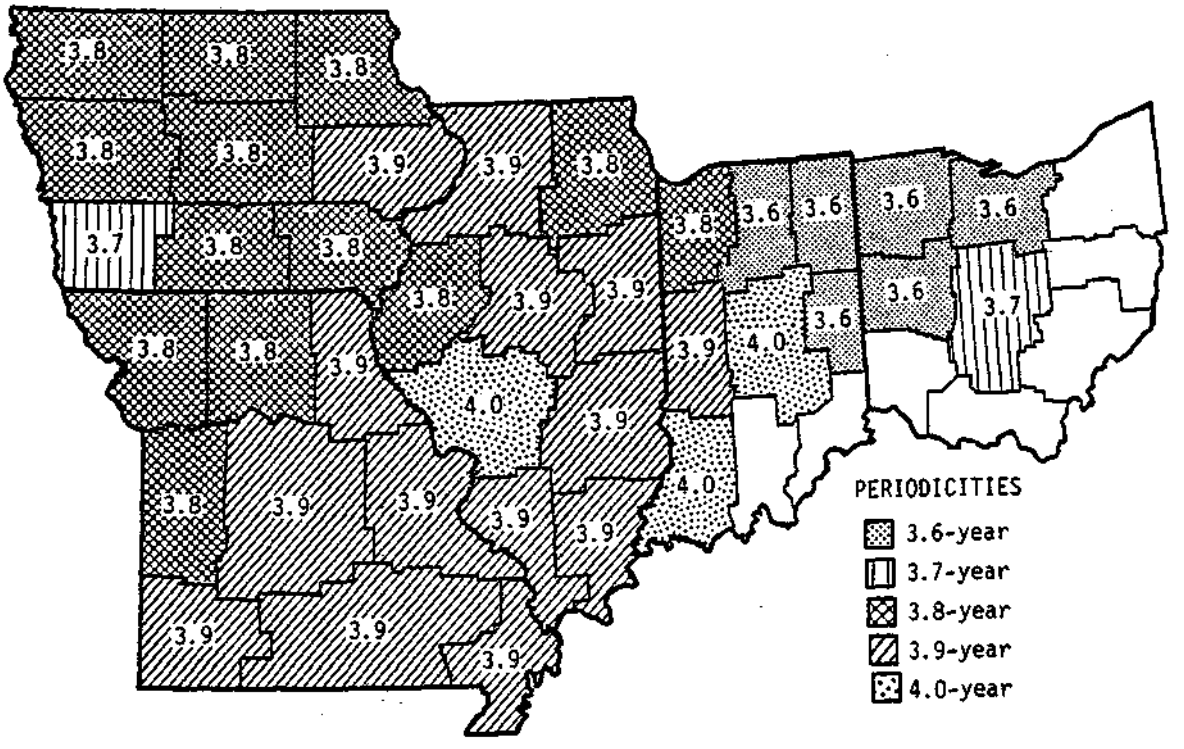

Figure 4a. Wavelengths (in years) for periodicities in the 3.6- to 4.0-year group from crop reporting district annual precipitation, 1931-1975.

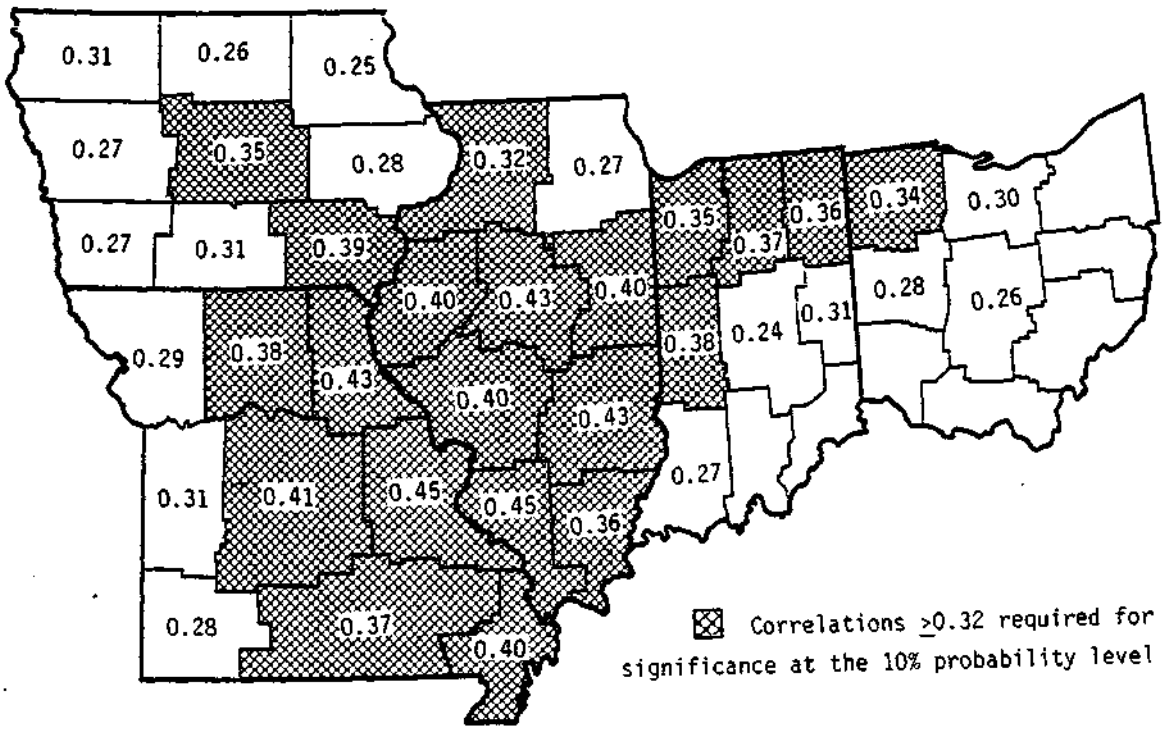

Figure 4b. Multiple correlation coefficients for crop reporting district annual precipitation with periodicities in tne 3.6- to 4.0-year group, 1931-1975.

areas. Spatial (areal) frequency coherence does not prove the precipitation experience over adjacent districts is varying with similar causative forces. However, such an areal condition strengthens the evidence that periodicities exist in the data (climatic) system.

Wavelength and Correlation for a 3.9-year Periodicity Group. Wavelengths for each of the 45 districts were sorted into groups (bands) and tabulated on the 5-state study map. Periodicities with wavelengths (periods) in the vicinity of 3.9 years are shown in Figure 4a. Many of the periodicities have correlation coefficients (shaded area of Fig. 4b) that equal or exceed the value (0.32) required for significance at the 10 percent level. Most of the periodicities with significant correlations are concentrated in Iowa, Missouri, Illinois, and northern Indiana in a southwest to northeast orientation across the 5-state area.

Wavelengths displayed in Figure 4a range from 3.6 to 4.0 years. Periodicities in this band were nonexistant in seven Indiana and Ohio districts. There are basically three regions (adjacent districts) with the same wavelengths (3.6, 3.8, and 3.9). The three regions are outlined by different shading in Figure 4a. The 3.6-year periodicities are is a group of six districts in northern and central Indiana and Ohio. A group of 11 adjacent districts with 3.8-year periodicities is located in Missouri, Iowa and Illinois. The largest group (14 districts with a 3.9-year wavelength) is located in the central part of the 5-state study area.

The above three areas account for all but 1) two 3.7-year periods on opposite sides of the study area, 2) two 3.8-year wavelengths in northeastern Illinois and northwestern Indiana, and 3) three nonadjacent districts with 4.0-year periods.

Amplitude and Phase for a 3.9-year Periodicity Group. The wavelength (period) of a periodicity and the strength of the periodicity association

(correlation) with the sample historical data are two very pertinent statistics in time series analysis (as discussed previously). Correlation and period were used to identify the existence of a periodicity. After a periodicity or periodicities have been identified they are used to 1) reconstruct the past and 2) to predict the future. For either usage, two other characteristics of a periodicity 1) the amplitude and 2) the phase point (phase angle) are required. The amplitude is the maximum departure of a wave from the data average (Fig. 5). Phase point is the time (temporal distance) from the time of the first observation (annual rainfall total) in a historical data sample to the time the periodicity reaches it first maximum amplitude (Fig. 5).

Since the amplitude and phase are also part of the prediction equation, they should also be included in the areal coherency evaluation. Amplitude and phase of the 3.9-year periodicity band are shown in Figures 6a and 6b, repsectively. The shaded area (Fig. 6a) outlines an area of relatively high amplitudes. Amplitude is a time series statistic which depicts the degree of climatic variability. An amplitude of 3.0 (an arbitrary choice) and greater is an indication of where the relative climatic (annual rainfall) variation was greatest during the historical record of 1931-1975.

Phase points associated with the 3.9-year periodicity group are shown in Figure 6b. Phase points can be either positive or negative. They vary between +1/2 the wavelength depending on where the data observation record started with reference to the progression of the wave. A positive sign indicates the maximum amplitude occurred after the start of the observation period (1931) whereas the negative sign indicates the maximum was before the year 1931 or prior to the end of the sample (1975).

Phase points of the 3.9-year group were predominately positive. Some of the variation in phase points in likely associated with variation in

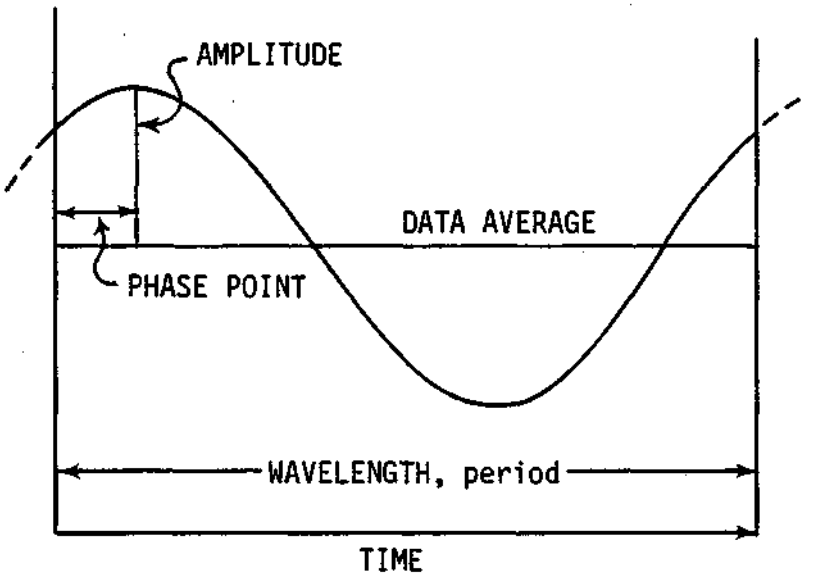

Figure 5. Sketch of an harmonic, y = sine x + cos x.

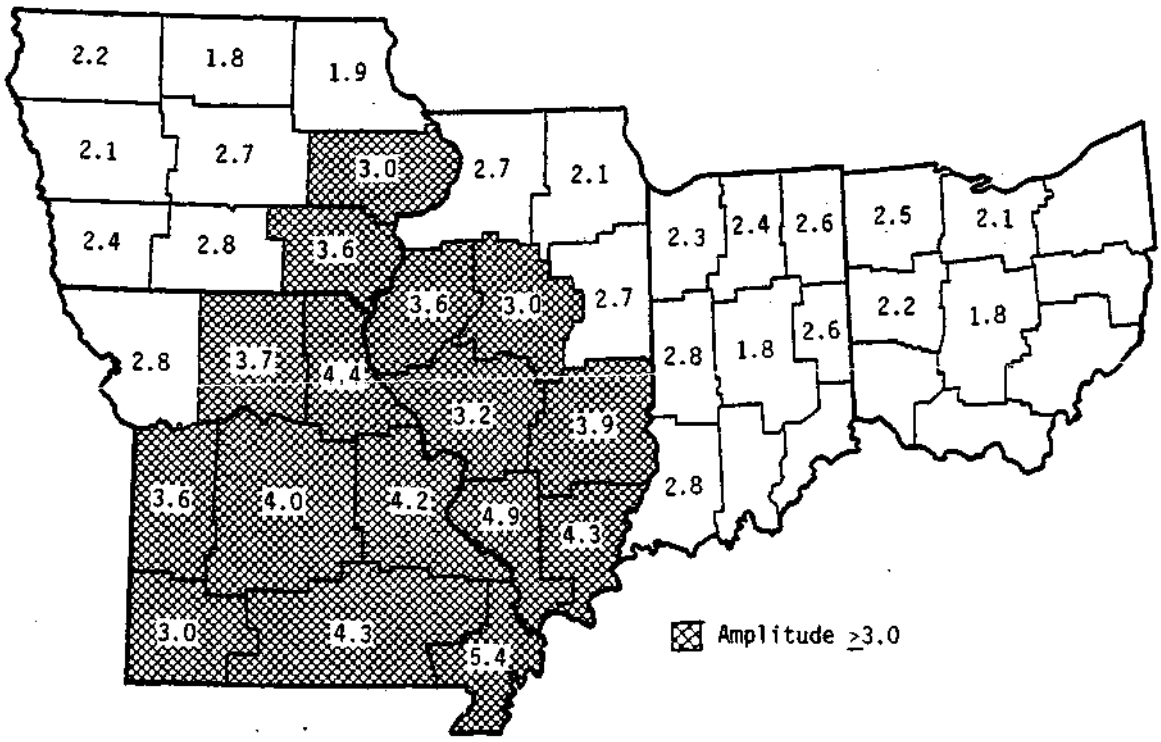

Figure 6a. Amplitudes for periodicities in the 3.6- to 4.0-year group for crop reporting district annual precipitation, 1931-1975.

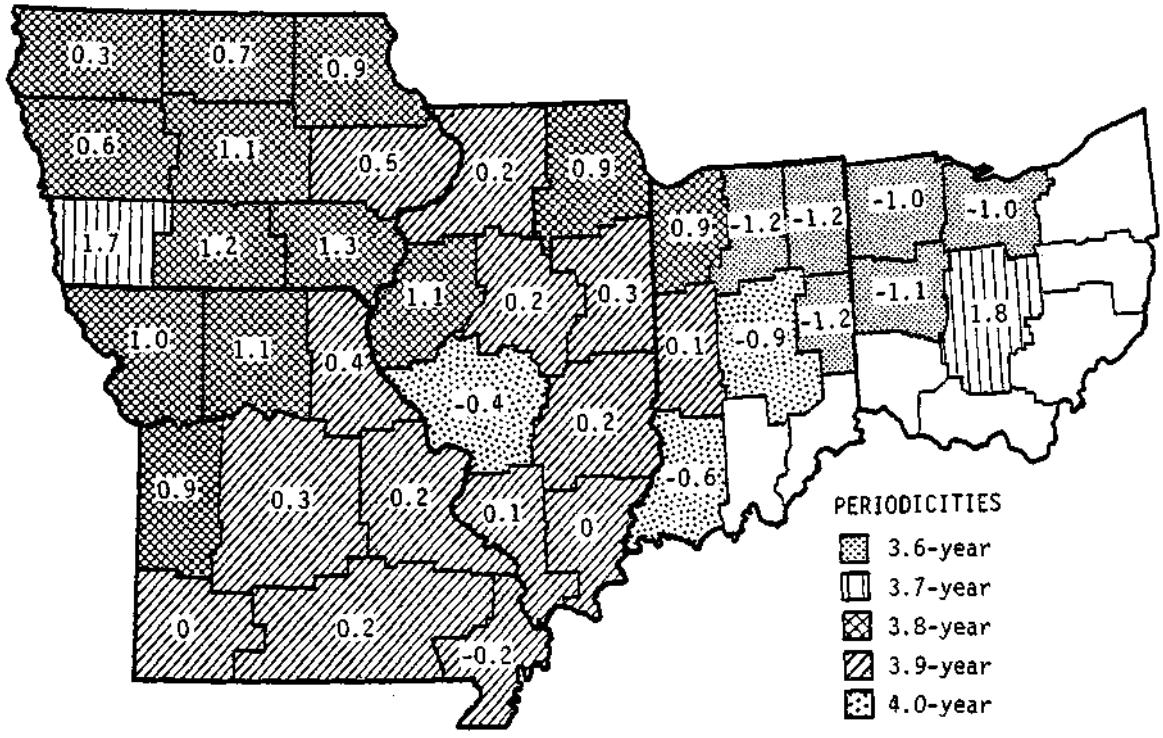

Figure 6b. Phase points (in years) for periodicities in the 3.6- to 4.0-year group from crop reporting district annual precipitation, 1931-1975.

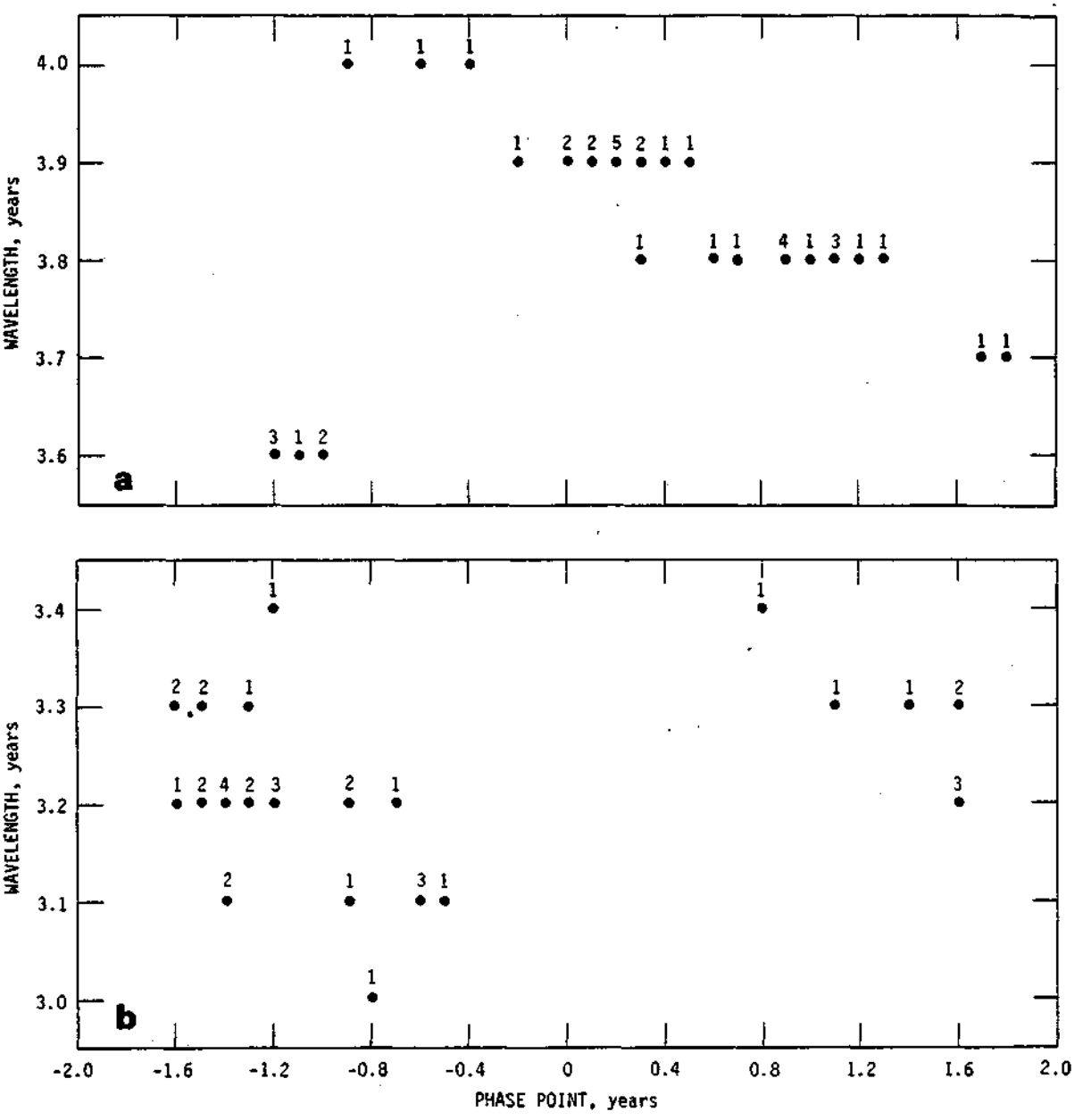

Figure 7. Scatter diagram of phase points versus wavelength within the (a) 3.6- to 4.0-year periodicity group and (b) the 3.0- to 3.4-year periodicity group for the 45-district area, 1931-1975.

wavelength within the group (3.6-4.0 years). The relation of phase points to wavelengths in the 45-district study area is shown in Figure 7a for the 3.9-year group. There was a tendency for phase point groups to shift from positive to negative as the wavelength changed from 3.7 to 4.0-years. Phase points for six 3.6-year periodicities were an exception to this tendency. All 3.6-year wavelengths had large negative phase points. These were all from adjacent districts located in northern and central Indiana and Ohio. This is the same area or group of adjacent districts which had 3.6-year wavelengths (Fig. 4a).

Two other larger groups of phase points are associated with the 3.8 and 3.9-year wavelengths (Fig. 6b). These two groups overlap in magnitude in one district (northwestern Iowa) which had a phase point of 0.30 year (Fig. 7a). A grouping of 14 phase points associated with the 3.9-year wavelength is located in the central portion of the 5-state area (Fig. 6b). Phase points for these 14 districts are small and positive with the exception of 1 which is a -0.2 in southeast Missouri. Their range is from -0.2 to 0.5 year, a variation of 0.7 year. The second largest group (13) of phase points are associated with 3.8-year periodicities. Eleven of these are in a contiguous area in the northwestern part of the 5-state area. The other two districts are adjacent to each other in northeast Illinois and northwest Indiana (Fig. 6b). The range of phase points for this group is relatively large (0.3 year to 1.3 years).

Wavelength and Correlation for a 3.2-year Periodicity Group. District periodicities with wavelengths in the vicinity of 3.2 (3.0 to 3.4) years are shown in Figure 8a). The 3.2-year periodicity is the most prominant feature (13 contiguous districts) of the group. Many of the periodicities have correlation coefficients (shaded area of Fig. 8b) that equal or exceed the

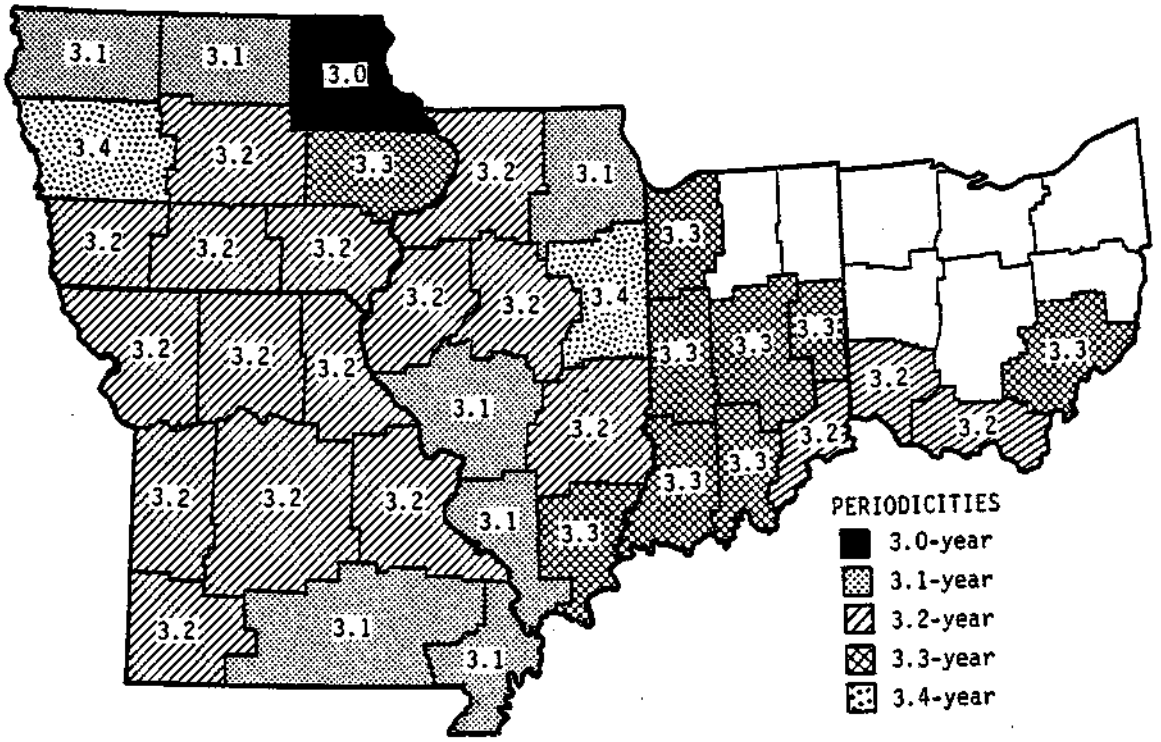

Figure 8a. Wavelengths for periodicities in the 3.0- to 3.4-year group from crop reporting district annual precipitation, 1931-1975.

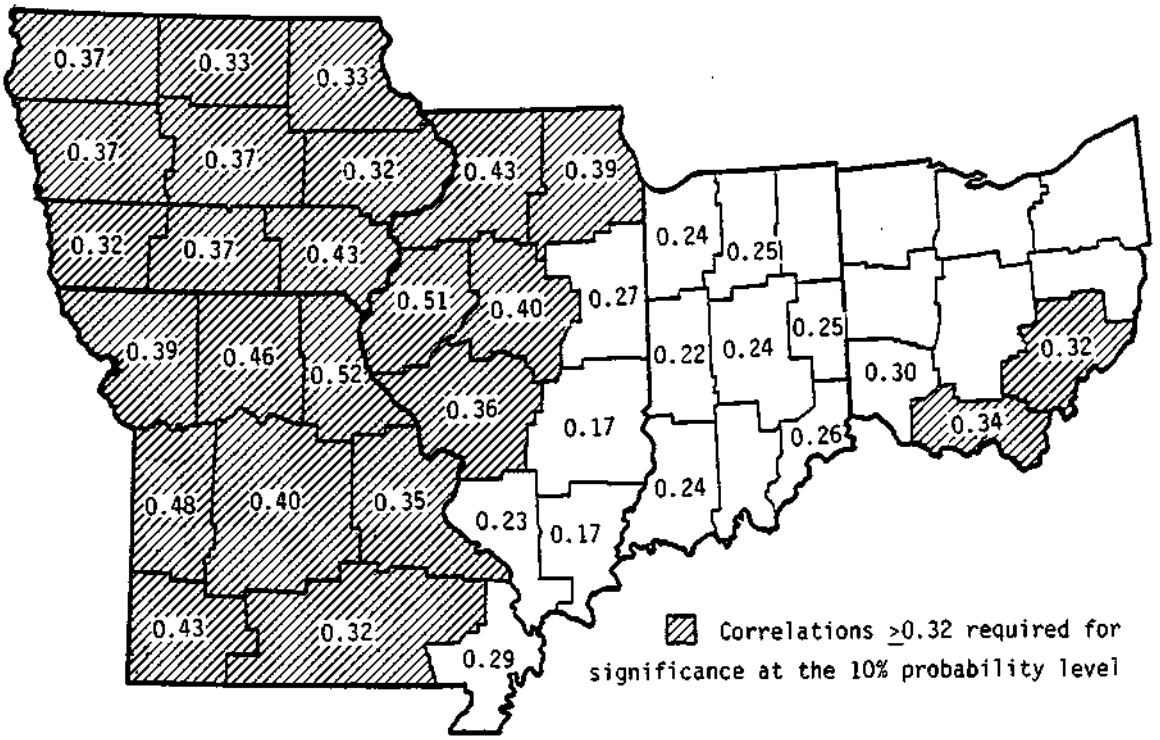

Figure 8b. Correlation coefficients of crop reporting district annual precipitation with periodicities in the 3.0- to 3.4-year group, 1931-1975.

value (0.32) required for significance at the 10 percent level. Most significant correlations for periodicities in this periodicity group are located in a contiguous group of districts in Illinois, Iowa, and Missouri (Fig. 8b).

Amplitude and Phase for a 3.2-year Periodicity Group. Amplitude and phase of this group are displayed in Figures 9a and 9b. Districts with amplitudes ≥ 3.0 are shaded in Figure 9a. The selection of 3.0 as a boundary for relatively high amplitude was an arbitrary choice as an aid to visual inspection of Figure 9a. As indicated earlier, amplitude, like correlation is a measure of the variance explained and therefore a measure of significance of a periodicity and its predictive power. However, since an amplitude of 3.0 or more was an arbitrary choice for visual aid in inspection, the shaded areas of Figure 4b and 9a do not correspond entirely.

Phase points of Figure 9b are predominately negative whereas there were many more positive values for this periodicity characteristic with the 3.9-year group. Phase points were plotted against wavelength (as was done in Figure 7a for the 3.9-year group) to check on a variation or relation of phase with wavelength (Fig. 7b). Figure 7b shows mainly a separation of negative and positive phase points but no real indication of any association of phase and wavelength. Geographically, negative phases are predominately in the western part of the the 5-state map and a few (8) positive phases points are scattered across the 5-state area (Fig. 9b).

Summary of Spatial Spectral Coherence

Five-state study area maps of spectral characteristics for the two most prominent periodicity groups (3.2 and 3.9 years) were prepared. These maps (Figs. 4a and 4b, 6a and 6b, 8a and 8b, 9a and 9b) contain distirct values for

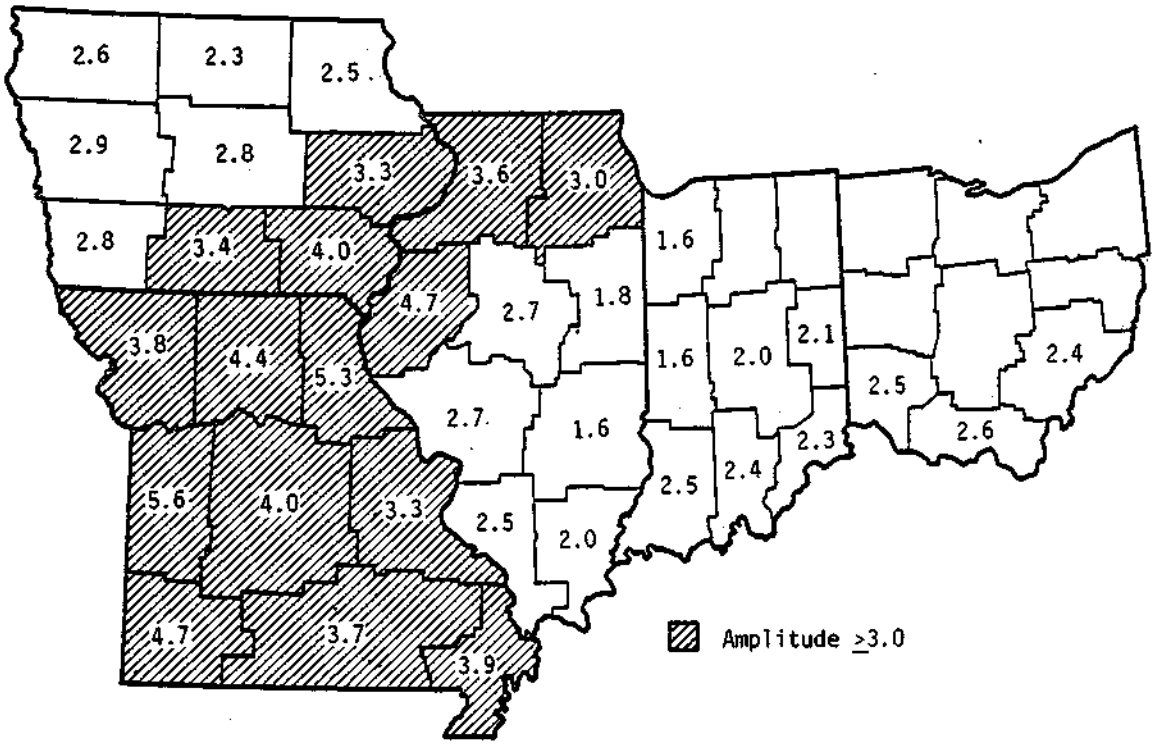

Figure 9a. Amplitudes of a 3.0- to 3.4-year periodicity group from crop reporting district annual precipitation, 1931-1975.

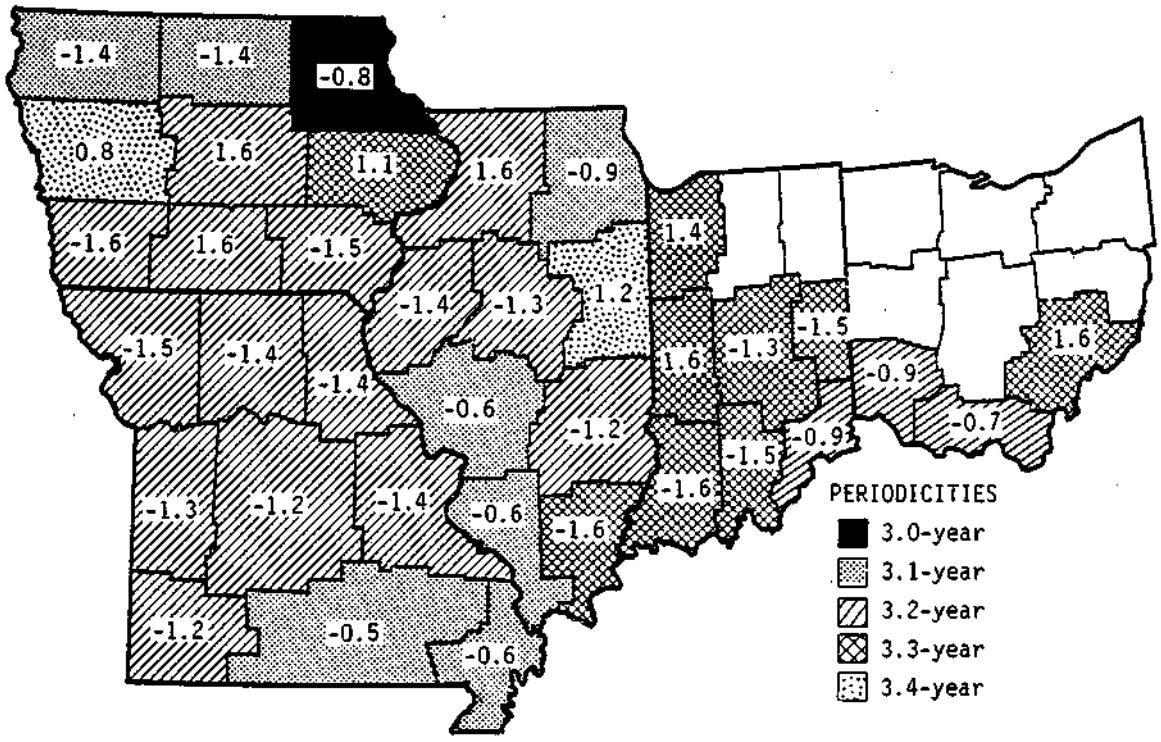

Figure 9b. Phase points (in years) for periodicities in the 3.0- to 3.4-year group from crop reporting district annual precipitation, 1931-1975.

wavelength, correlation, amplitude, and phase point. These maps indicate definite regions (contiguous districts) with the same or very similar significant wavelengths. Significance was established by the degree of correlation of a periodicity with the historical data.

Wavelength shown in Figures 4a and 8a are also summarized graphically in frequency diagrams in Figure 10. This presentation disregards the spatial variations discussed in prior sections. The figure simply shows a count of the number of district periodicities of the two groups (3.2 and 3.9) over the 45-district study area that were significant at the 10 percent level (correlations $\geq .32$). These frequency plots indicate which periodicities within each of the two groups contained the greatest potential predictive power over the 45-district area from a spectral analysis of the 1931-1975 anuual precipitation record.

Spatial spectral coherence was demonstrated for regions within the 5-state study area and two periodicities (3.2 and 3.9) with the most predictive power for use in computing annual precipitation were identified. Percentages of precipitation variance explained as determined by the sum of squared values of correlations in each district are shown in Figure 11. For example, 42 percent in district 5 in west central Illinois was determined by squaring 0.40 (Fig. 4b) and 0.51 (Fig. 8b) summing and multiplying by 100. Thus, the map indicates a measure of the predictive power over the 5-state area.

A much greater indication of predictive power than indicated above was desired for prediction purposes. Since other periodicities or periodicity groups with a display of coherence were not identified, the decision was made to base annual district predictions for 1973, 1974, and 1975 on all periodicities which were identified as significant from spectral analyses. The last three years (1973, 1974, and 1975) of the 1931 through 1975 record

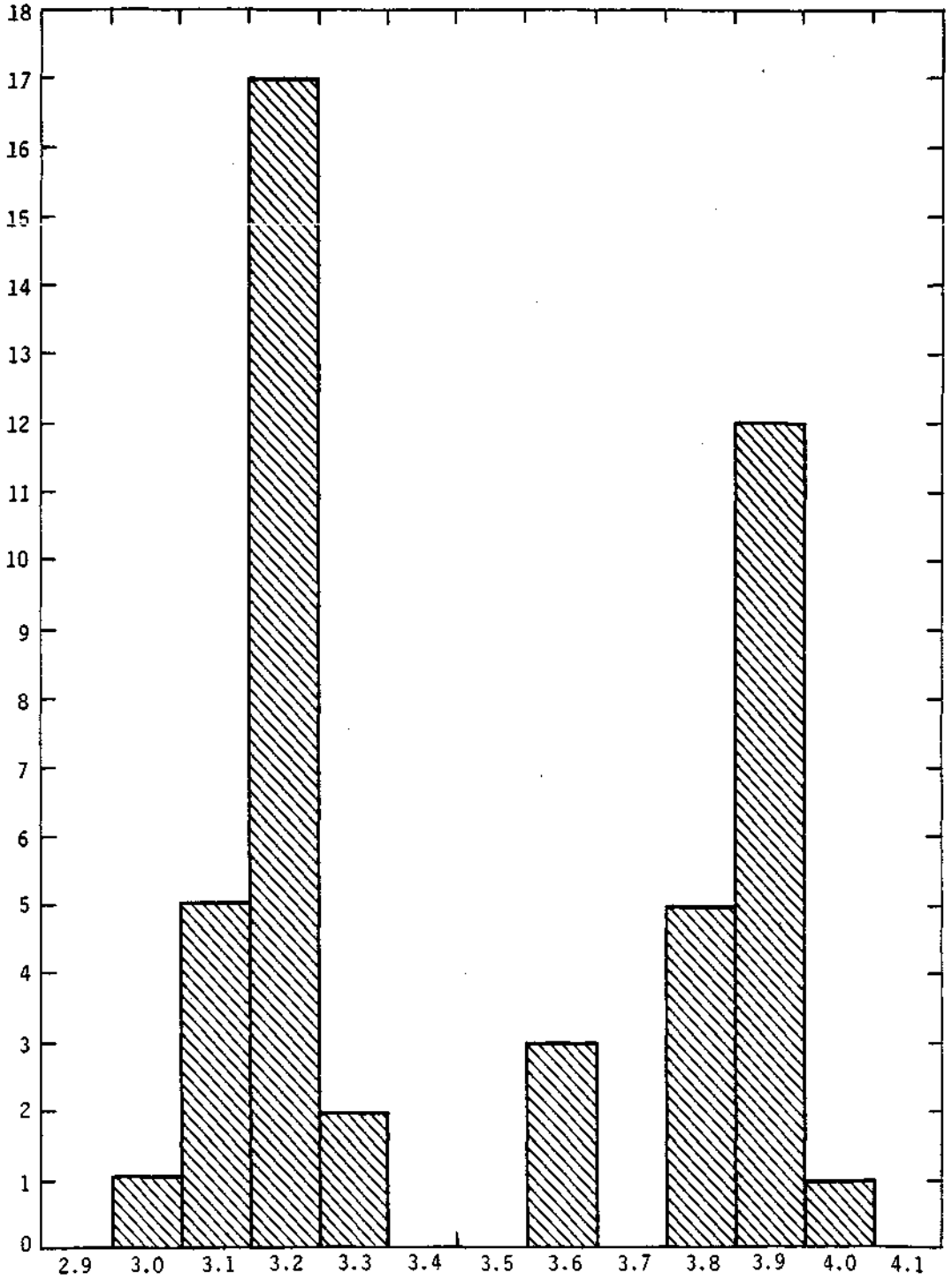

Figure 10. Frequency histograms of the number of districts having significant periodicities near 3.2 and 3.9 years.

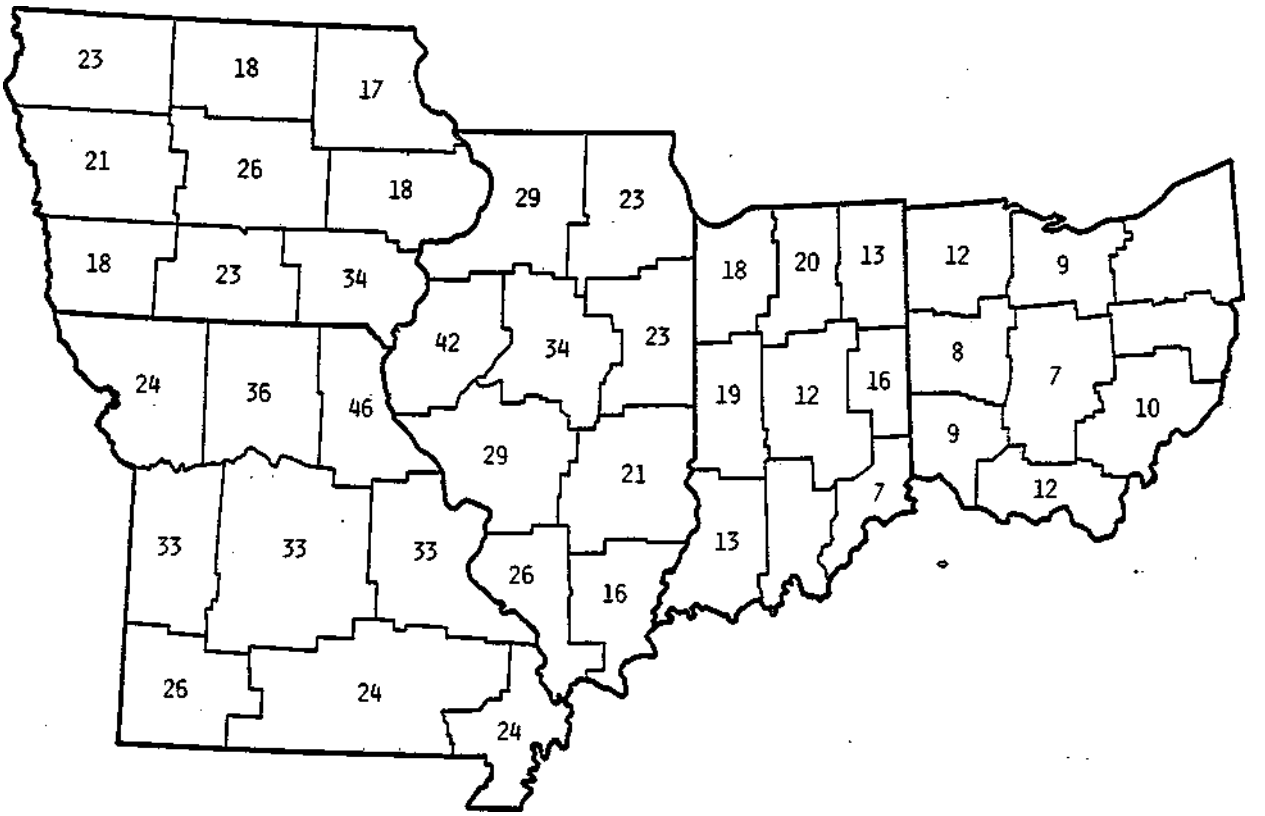

Figure 11. Percent of the 1931 through 1975 annual precipitation variance explained by periodicities in the 3.2- and 3.9-year groups.

were omitted from subsequent district spectral analyses and retained for prediction verification purposes. Also, the year 1931 was omitted from subsequent spectral analyses of annual data in order to provide the best coordination of annual analyses and predictions with subsequent seasonal analyses and predictions. Seasonal data began with the winter of 1931-1932 (December 1931, January 1932, and February .1932).

PREDICTION AND PREDICTION METHODOLOGY

Two prediction methods were used in the precipitation phase of the study. The two methods are referred to as "band pass" and "filters." Both methods utilize the periodicity results of a spectral analysis of historical data. Periodicities, which explained a statistically significant portion of the variance of the data in each district, were used as predictors in each method. However, the manner in which the two methods use the significant periodicities differs.

Bandpass Method

Description of Method. The bandpass method utilizes an empirical equation for extrapolating a periodicity into future years. The manner in which the bandpass method obtains a predicted annual rainfall value is described with the aid of Figure 12. This figure illustrates the determination of an annual precipitation total for the prediction years 1972 through 1974 using 4 significant periodicities, 10.3, 5.8, 3.8, and 2.7 years shown in Figure 12.

The method requires, 1) the amplitude, 2) period (wavelength), 3) phase point, 4) harmonic number, and 5) the number of years (the number of years in the analysis period plus the number of years to be predicted) as input. The first four of these quantities come from a spectral analysis. In theory, an

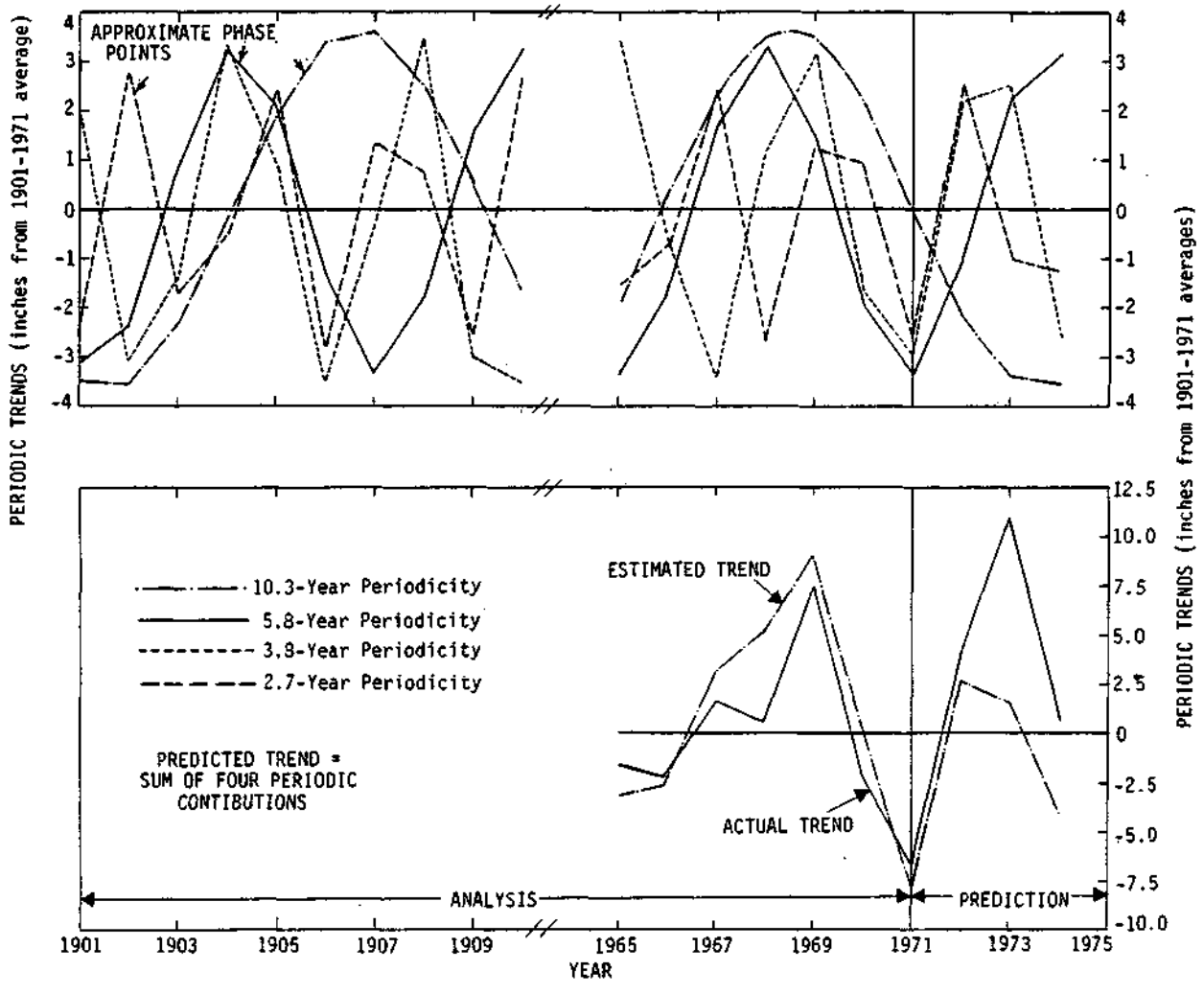

Figure 12. Bandpass predictions for 1972-1974 for Illinois district No. 8 based on 1901-1971 data.

estimated annual precipitation curve is a smooth and regular oscillation. However, with annual observations, the bandpass formula can only determine values at one year intervals as shown by the irregular lines of Figure 12.

Corresponding annual computed precipitation values (contributions) for each periodicity (Figure 12) were added algebraically. Each of these sums can be added to the average precipitation for the data analysis period (1901-1971) to obtain a reconstruction of the prediction basis data and to obtain three independent annual precipitation predictions. However, only trends are presented in Figure 12.

These summations provide an estimated precipitation trend for each year of the data analysis and a predicted precipitation for each of 3 years into the future. This prediction is illustrated in Figure 12 for Illinois district 8. The figure also shows a portion of the actual precipitation experience for the district. The period from 1901 through 1971 was the data analysis period (prediction basis period) used in a search for the significant periodicities which were used as predictors. The 1972 through 1974 data were saved for verification of the predictions. A good trend comparison is evident for the predicted values.

Annual District Predictions from Annual Input. The bandpass method was used to compute annual precipitation trend (up or down from previous year) predictions for all nine Illinois districts. For these predictions, the annual precipitation record for 1932 through 1972 was used as the analysis base with 1973-1975 retained for comparison with predictions (Table 1). Trend predictions for 8 of 9 districts were correct for the first prediction year (1973), a very good skill score. However, the accuracy of trend predictions for the two succeeding years (1973 and 1974) averaged about the same as chance expectation. Within districts, results ranged from 1 to 3 (a perfect score)

Table 1. Yearly precipitation trend predictions from bandpass method compared with actual and predicted percent of actual for 1973 through 1S75, Illinois, using 1932-1972 for prediction basis.

	Trends, UP (U) or DOWN (D) (Predicted on the left and actual on the right) Year				(Predicted/Actual) x 100 Year		
Districts	1973	1974	1975	percent correct	1973	1974	1975
1(NW)	D/D	U/U	D/D	100	54	110	88
2(NE)	D/D	U/U	D/D	100	69	102	74
3(WC)	U/U	U/D	U/D	33	70	88	103
4(C)	U/U	D/D	D/D	100	85	87	85
5(EC)	D/D	U/U	D/D	67	82	89	77
6(SW)	U/U	U/D	U/D	33	71	84	93
7(SE)	D/U	D/U	D/D	33	91	80	76
8(SSW)	U/U	D/D	D/U	67	92	97	89
9(SSE)	U/U	D/D	D/U	67	100	95	73
Year%	89	56	56	67			

correct predictions. The right hand side of Table 1 shows the percent the predicted was of the actual precipitation experienced. Periodicities with wavelengths shown in Table 2 for Illinois were used as predictors.

Filtering Method

Description of the Filtering Method. The filtering prediction method investigated involved mathematical filters. The term "filtering" is from electrical engineering and has the same general connotation and usage here. A filter is designed to suppress most periodicities (or frequencies) except the designated frequency it is designed to "pass." All filters used in this study were symmetrical around a vertical central value of 17 points to the left and 17 points to the right of center (Fig. 13). In the vertical, the 35 points vary above and below zero. Points above the zero line are positive, in the algebraic sense, and those below zero are negative. Values which are computed during a filtering operation will be referred to as a filter output.

The procedure for determining a filter output is described with the aid of Figure 14. Figure 14a is a time series plot of annual precipitation for the years, 1932 through 1972. A filter (Fig. 14b) for a 3.9-year periodicity (a periodicity determined from the spectral analysis) was applied to the data for the period 1932 through 1972 in Illinois district 7. A plot of the digital values (X symbols) of the 3.9-year filter is shown in Figure 14c. In this illustration, the precipitation value for the year 1932 is positioned over the center filter value. The individual 3.9-year filter output value for 1932 is plotted directly below in Figure 14c. The digital value of the filter output with the data in this position, with respect to the filter, is the sum of products of the corresponding 35 filter and precipitation values. The entire filter output shown in Figure 14c is obtained by matching each precipitation

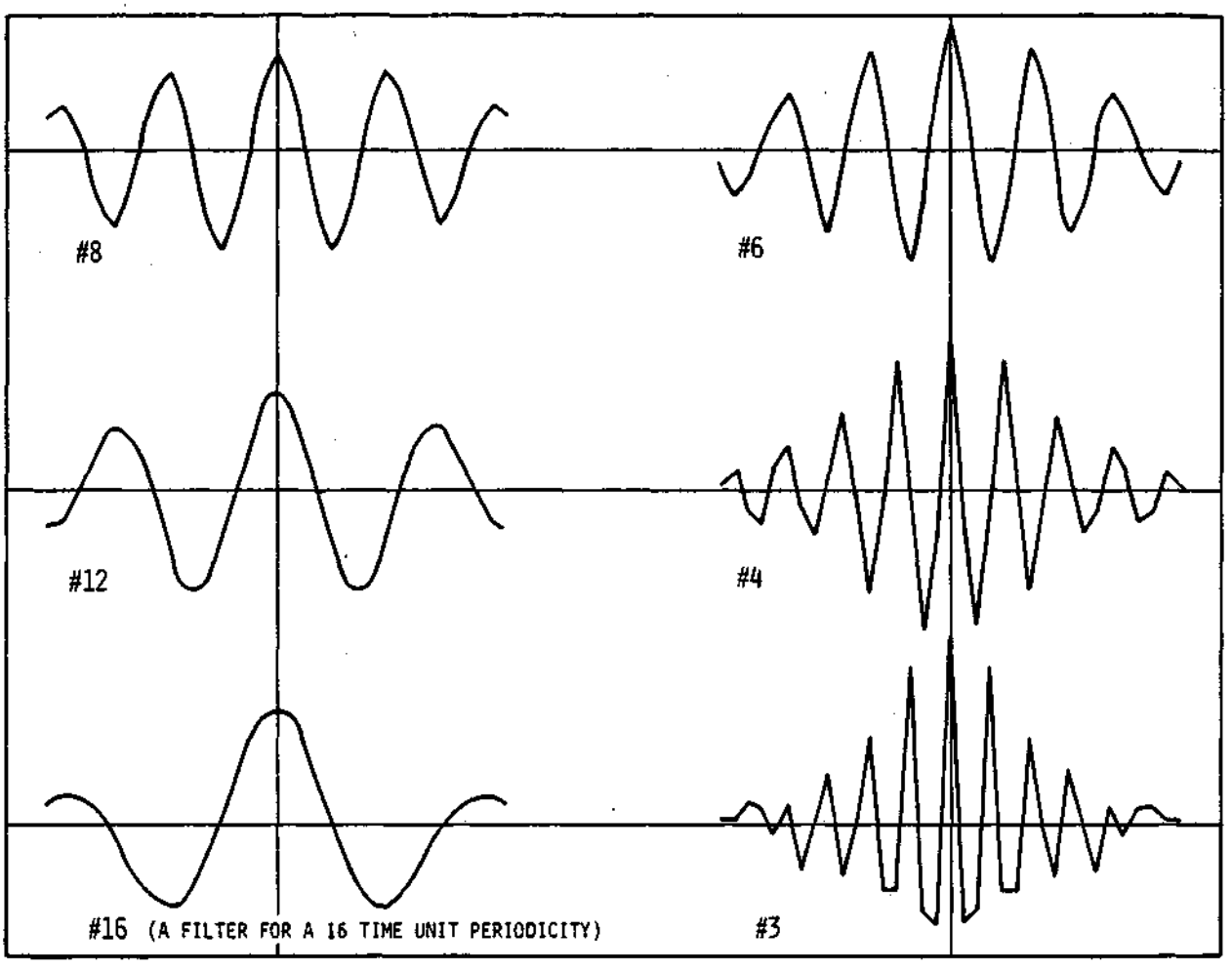

Figure 13. Examples of numerical filters.

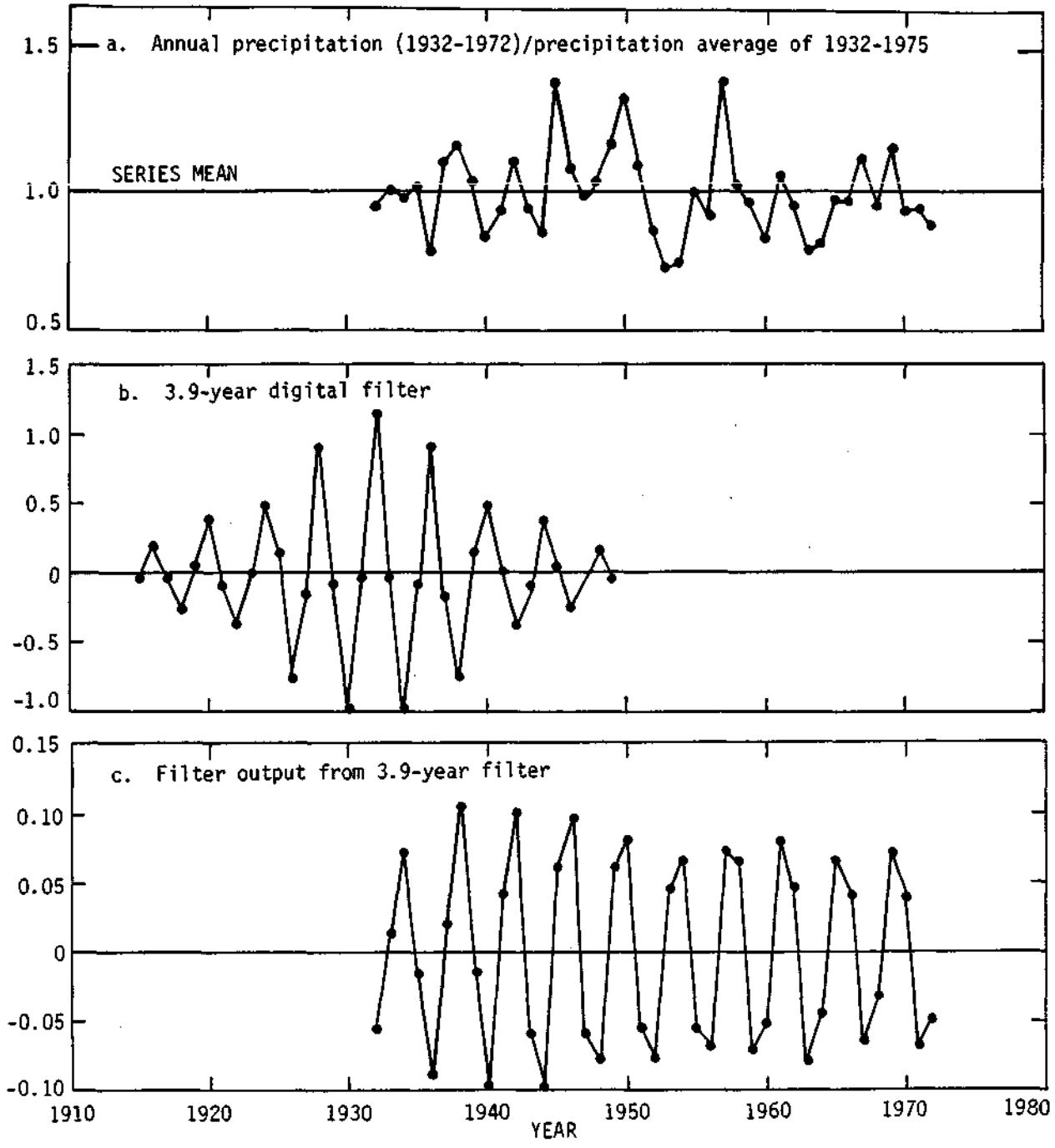

Figure 14. Illustration of a filtering operation using Illinois district 7 annual precipitation and a 3.9-year digital filter.

Table 2. Wavelengths of significant Periodicities used for annual district precipitation trend predictions from a 1932-1972, 5-state data base.

District & Location in State	Wavelength (Years) 2.0-2.4	2.5-2.9	3.0-3.4	3.5-3.9	4.0-4.9	5.0-5.9	6.0-6.9	8.0-8.9	9.0-9.9	10.0-10.9
Illinois										
1(NW)		2.7	3.2		4.8					
2(NE)	2.3		3.2							
3(WC)			3.2	3.9	4.8					
4(C)			3.2	3.9	4.8.					
5(EC)		2.6		3.9						
6(SW)				3.9						10.4
7(SE)						5.9			9.9	
8(SSW)				3.9					9.8	
9(SSE)				3.9		5.9				
Indiana										
1(NW)	2.3			3.8						
2(NC)				3.7		5.8				
3(NE)				3.7		5.9			9.5	
4(WC)				3.8						10.1
5(C)	2.4						6.3		9.8	
6(EC)	2.4						6.1		9.9	
7(SW)							6.1		9.8	
8(SC)							6.1		9.5	
9(SE)							6.0		9.3	

Table 2 Continued

District & Location in State	Wavelength (Years) 2.0-2.4	2.5-2.9	3.0-3.4	3.5-3.9	4.0-4.9	5.0-5.9	6.0-6.9	8.0-8.9	9.0-9.9	10.0-10.9
Iowa										
1(NW)		2.7	3.1							
2(NC)		2.7	(3.0)		4.6					
			(3.4)							
3(NE)	2.1	2.7	3.0							
4(WC)		2.7	3.4							
5(C)		2.7	3.3							
6(EC)		2.7						8.0		
7(SW)			3.3		4.8					
8(SC)		2.7	3.2		4.8					
9(SE)		2.7	3.2							
Missouri										
1(NW)			3.3							
2(NC)	2.0		3.2							
3(NE)			3.2	3.8						
4(WC)			3.3							
5(C)			3.3	3.9						
6(EC)		2.7				5.7			9.0	
7(SW)			3.3							
8(SC)				3.9						
9(SE)				3.9		5.7		8.3		

Table 2 Continued

District & Location in State	Wavelength (Years) 2.0-2.4	2.5-2.9	3.0-3.4	3.5-3.9	4.0-4.9	5.0-5.9	6.0-6.9	8.0-8.9	9.0-9.9	10.0-10.9
Ohio										
1(NW)						5.8			9.6	
.2(NC)	2.4	2.7								10.0
3(NE)		2.7			4.6					10.1
4(WC)	2.4								9.9	
5(C)	(2.0)									10.1
	(2.4)									
6(EC)	2.4	2.7								
7(SW)	2.4		3.2						9.3	
8(SC)	2.0		3.1			5.5				
9(SE)	(2.0)	2.7								
	(2.3)									

data point with the center filter value, and repeating the 35 multiply and sum computations. A filter output is similar to a moving correlation as the filter is systematically moved along the data. In effect, the precipitation time series serves as an input to the filter, and the output is a new time series which becomes an input for the next step of the prediction process.

The entire filter prediction process was performed with a computer program package which was executed by an operator through a computer terminal with a cathode ray presentation and an on-line plotter. The operator requested a precipitation series from the computer data bank and called in the proper filter from a bank of 50 available filters. The filter output was computed as described above and plotted on the computer terminal screen (Fig. 15). The operator then selected a minimum of three points (+ signs) at the peaks and valleys of the filter output and near the end of an output record. These chosen points formed the basis for determining a sine wave which was projected back through filter output points and projected forward for the desired number of years. At the time, the operator made a visual inspection of the "goodness-of-fit" of the sine wave to the data. The operator either retained (in the computer system) the estimated annual points on the sine wave, or elected to adjust the three basic points to try to obtain a better fitting sine wave.

A sine wave fit to the filter output of the 3.9-year periodicity is shown in Figure 15. Points A, B, and C, represent the contribution of the 3.9-year periodicity to the predictions for years 1973 through 1975. If only one periodicity is used, these point values would be added to a mean value of the 1932-1972 period to finalize the yearly predictions. However, more than one significant periodicity was usually involved. In this case, the sum of

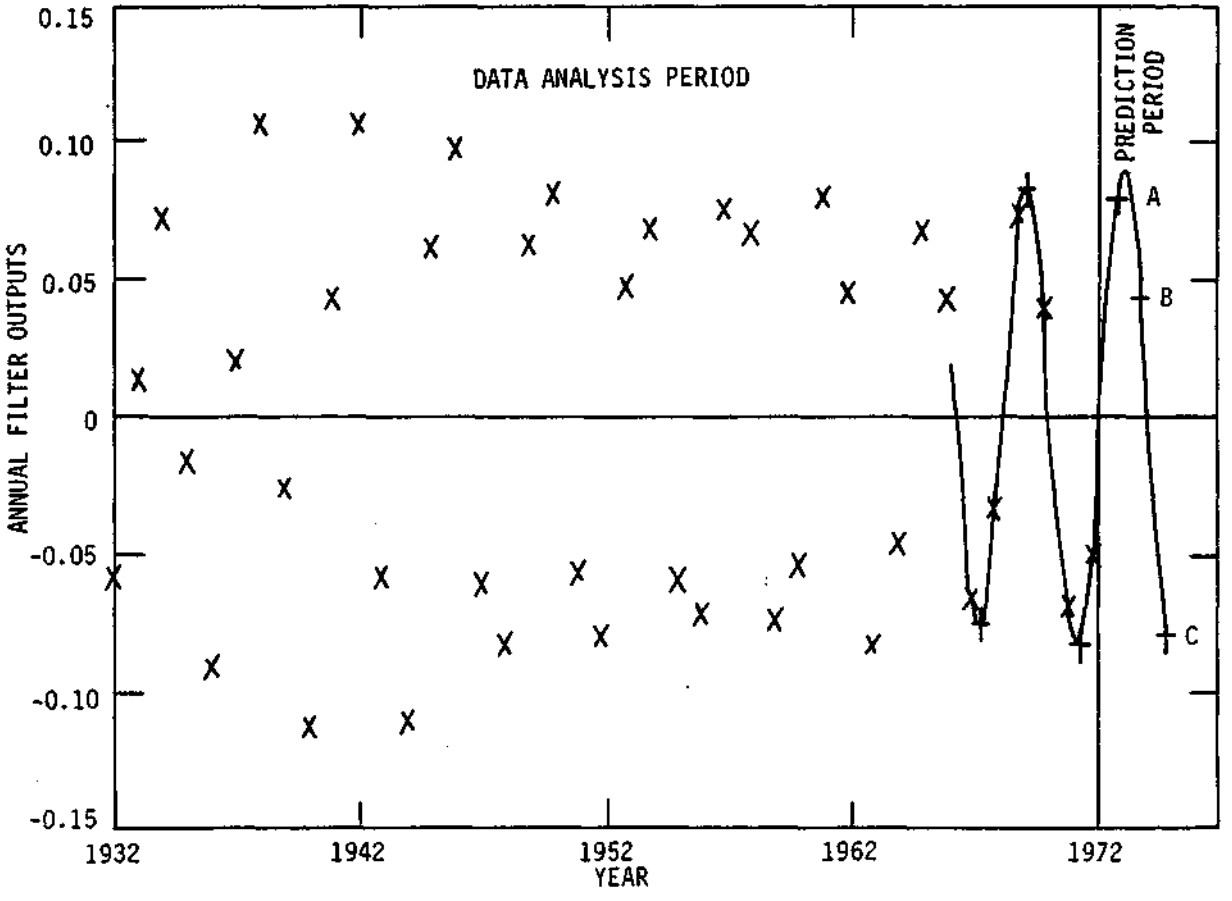

Figure 15. Output from a 3.9-year filter and subsequent sine wave fit and predictions for Illinois district 7.

contributions (Figs. 15, 16, and 17) for 10.0, 6.0, and 3.9-year periodicities formed the yearly predictions (Fig. 18). Predictions in Figure 18 are normalized values. Actual precipitation values can be obtained by multiplying the 1932-1972 average to the normalized values.

Annual District Predictions from Annual Input. Annual predictions were computed for the 45 districts in the manner described above. All significant periodicities (determined by spectral analysis) in each district were used (Table 2). Trend and quantitative prediction results are tabulated in Table 3. Except for first year (1973) prediction results in Illinois, predicted trend skill was about the same as that for chance expectation.

Annual District Predictions from Seasonal Input. Thus, far in this study, spectral analysis and predictions for annual trends were performed on annual precipitation totals for various crop districts. It was presumed that predictors for shorter periods of time (such as seasons) would be more accurate than those for annual. Presumably, more accurate seasonal predictions could be summed to get more accurate annuals. Therefore, 3-month seasonal totals were determined for each of four seasons beginning with the winter season (December, January, and February). This particular 3-month seasonal sequence was chosen to obtain an agriculturally relevant June-July-August seasonal total for the Midwest. The 1932 through 1972 (beginning with December 1931 and ending with November 1972) was chosen for the analysis basis in order to have the seasonal analysis coincide (within one month) of the analysis basis used for previous annual predictions. Each seasonal total was "normalized" by dividing it by the average of 1932-1972 totals for the respective season. Normalization produced seasonal values which varied above and below 1.

Normalized seasonal precipitation values were analyzed spectrally to determine significant periodicities for each district in Illinois and

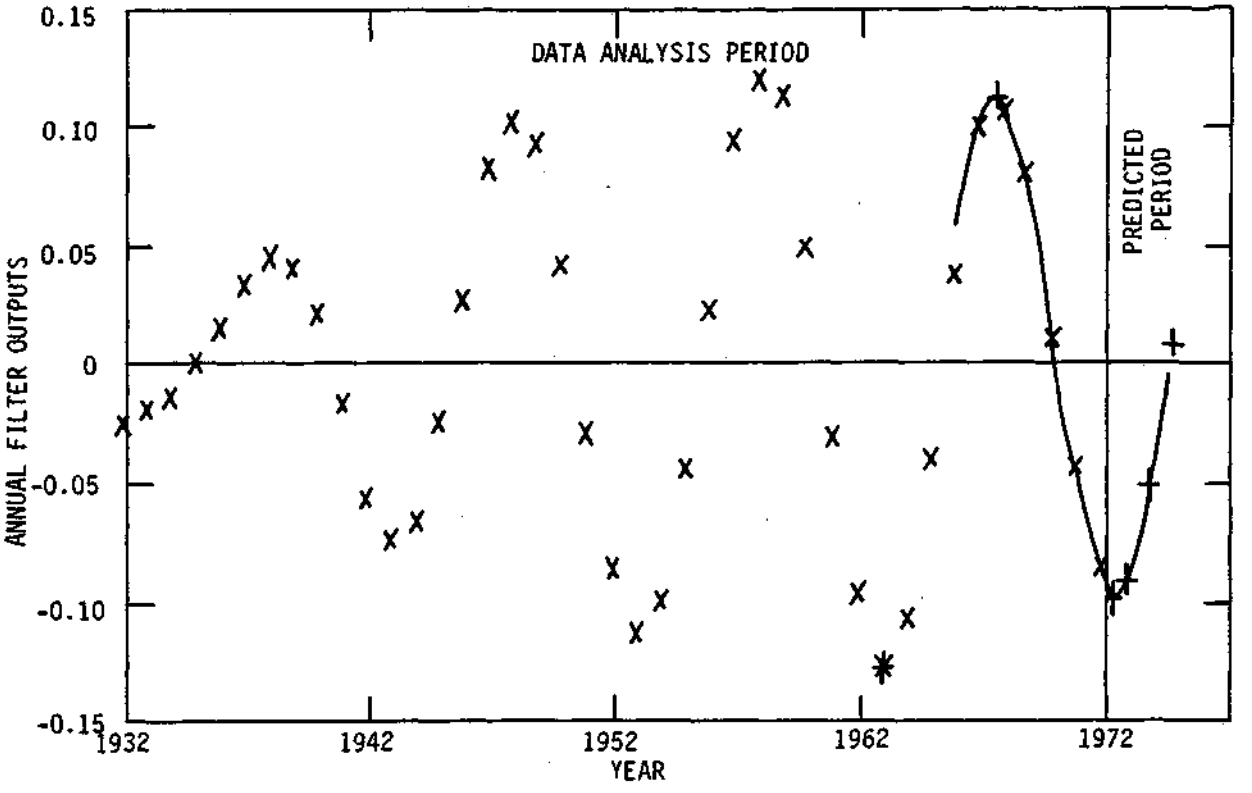

Figure 16. Output from a 6.0-year filter and subsequent sine wave fit and predictions for Illinois district 7.

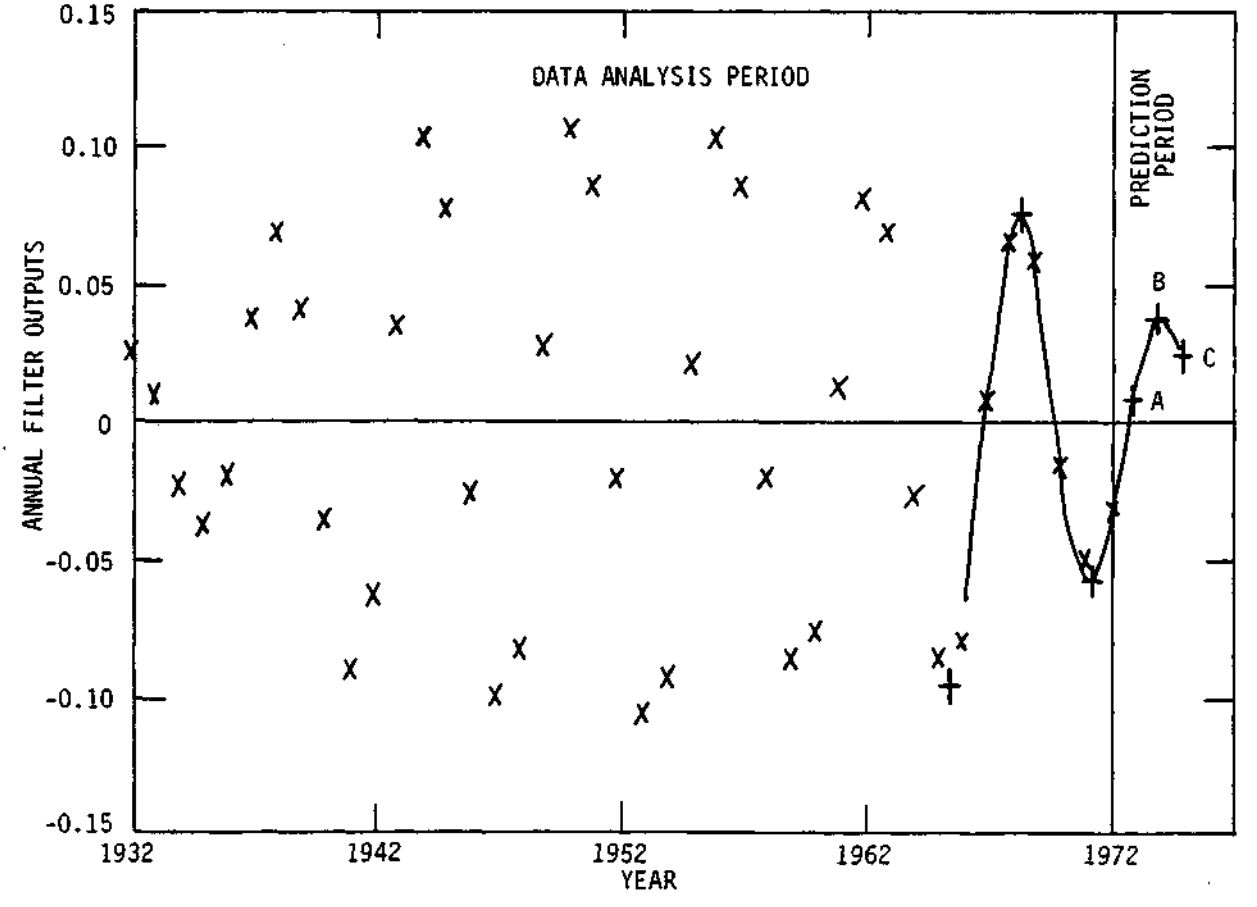

Figure 17. Output from a 10.0-year filter and subsequent sine wave fit and predictions for Illinois district 7.

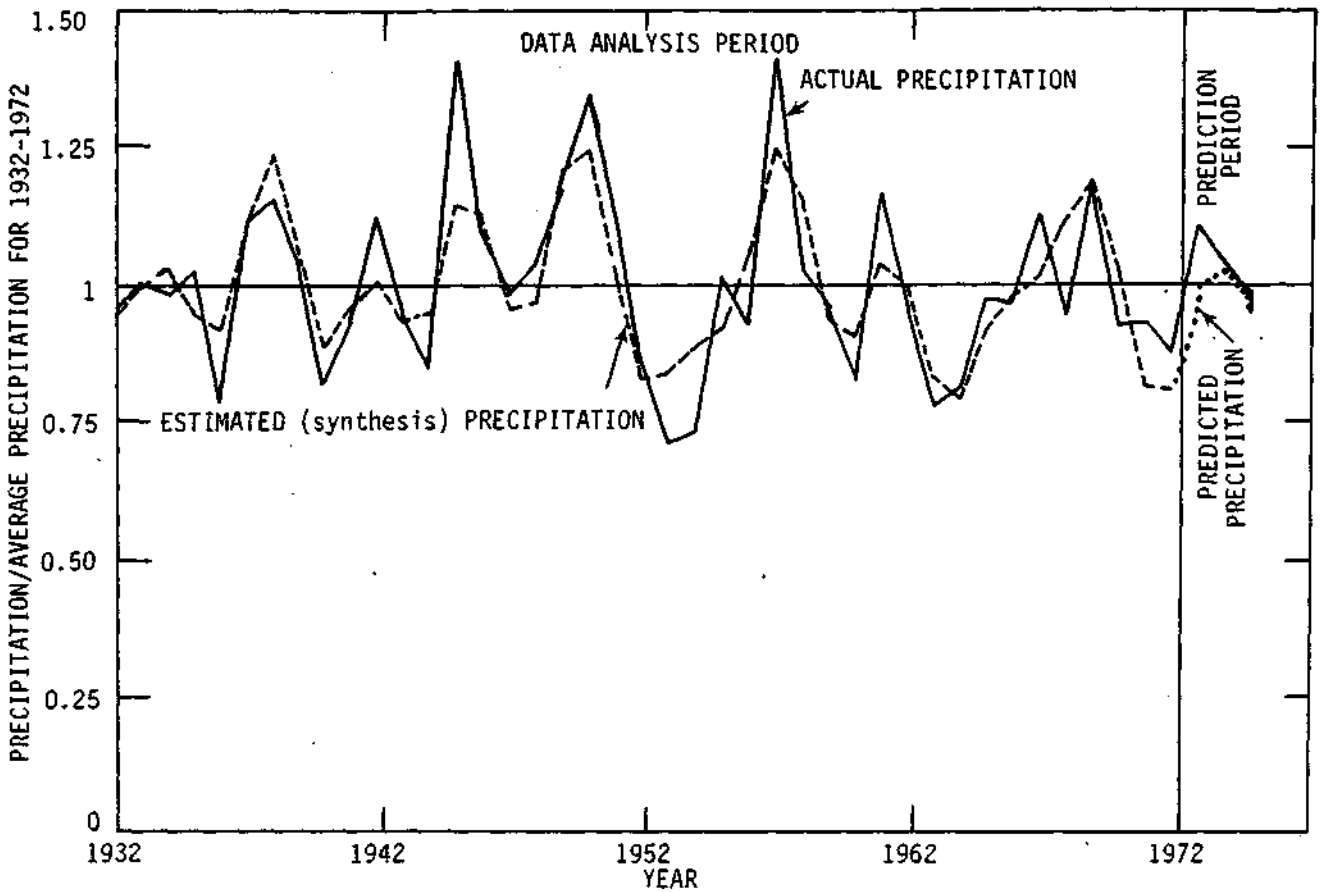

Figure 18. Actual precipitation synthesis of actual, 1932-1972, and predictions for 1973-1975 for Illinois district 7. Determined with 10.0-, 6.0-, and 3.9-year filters.

Table 3. Comparison of yearly predicted precipitation trends (1973-75) with actual trends using filter method and 1932-1972 annual data base.

	Trends, UP (U) or DOWN (D) (Predicted on the left and actual on the right)				(Predicted/Actual) x 100		
	Year				Year		
Districts	1973	1974	1975	percent correct	1973	1974	1975
				Illinois			
1(NW)	D/D	D/U	U/D	33	71	85	124
2(NE)	D/D	U/U	D/D	100	66	104	103
3(WC)	U/U	D/D	U/D	67	71	70	118
4(C)	U/U	D/D	D/D	100	88	94	86
5(EC)	D/D	U/D	D/D	67	74	91	81
6(SW)	U/U	U/D	U/D	33	73	82	100
7(SE)	U/U	D/U	D/D	67	87	84	84
8(SSW)	U/U	D/D	U/U	100	98	82	81
9(SSE)	U/U	D/D	D/U	67	98	100	85
%Correct	100	56	56	70			
				Indiana			
1(NW)	D/D	U/U	U/U	100	81	96	92
2(NC)	U/D	U/D	D/U	0	102	110	. 86
3(NE)	U/D	U/D	U/U	33	108	112	100
4(WC)	D/U	U/U	D/D	67	80	81	91
5(C)	U/U	U/D	D/D	67	90	96	92
6(EC)	D/D	U/D	D/U	33	94	117	95
7(SW)	U/U	D/U	U/U	67	92	85	87
8(SC)	U/U	D/D	D/U	67	93	89	77
9(SE)	D/U	D/D	D/D	67	82	92	8P
%Correct	56	44	67	56			

Table 3 Continued

Districts	Trends, UP (U) or DOWN (D) (Predicted on the left and actual on the right) Year 1973	1974	1975	percent correct	(Predicted/Actual) x 100 Year 1973	1974	1975
				Iowa			
1(NW)	D/U	D/D	U/U	67	82	95	107
2(NC)	U/U	D/D	D/U	67	82	108	79
3(NE)	D/D	U/D	D/D	67	80	101	110
4(WC)	D/U	U/D	D/U	0	46	135	96
5(C)	D/U	U/D	D/D	33	59	91	75
6(EC)	D/U	D/U	U/D	33	83	76	126
7(SW)	D/U	U/D	D/U	0	53	139	93
8(SC)	D/U	U/D	D/D	33	52	99	92
9(SE)	D/U	U/D	D/D	33	66	100	126
%Correct	33	22	56	37			
				Missouri			
1(NW)	D/U	U/D	U/U	33	59	98	113
2(NC)	D/U	U/D	U/D	0	68	97	140
3(NE)	U/U	D/D	D/D	100	84	84	91
4(WC)	U/U	D/D	U/D	67	72	72	111
5(C)	U/U	D/D	D/D	100	81	89	94
6(EC)	D/U	U/D	U/U	33	75	96	98
7(SW)	No significant periodicities						
8(SC)	D/U	D/D	D/D	67	67	76	74
9(SE)	D/U	U/D	D/U	0	45	133	35
%Correct	38	50	63	50			
				Ohio			
1(NW)	D/D	D/D	D/U	67	101	116	80
2(NC)	D/D	U/D	D/U	33	92	107	94
3(NE)	D/D	D/U	D/U	33	93	84	76
4(WC)	D/D	U/D	D/U	33	75	109	88
5(C)	D/D	U/D	U/U	67	78	104	105
6(EC)	U/D	D/U	D/U	0	92	86	79
7(SW)	D/U	U/D	U/U	33	76	93	94
8(SC)	D/U	U/D	U/U	33	80	85	85
9(SE)	D/U	U/U	D/U	33	82	84	74
%Correct	56	22	33	37			

Indiana. The filtering prediction technique described previously was performed with significant seasonal periodicities for each of the 18 districts. Predicted precipitation trends and actual trends are presented in Table 4 for 1973, 1974, and 1975. In the case of Illinois the overall prediction results improved from 70% correct (Table 3) to 74% correct (Table 4) when seasonal input was used in place of annual. The 3-year prediction sequencies (1973-1975) for 6 of the 9 districts were the same as those for annual input (Table 3). In Indiana, seasonal and annual inputs produced the same trend predictions. An increase from 70-74% which was due to predicting one more district year trend correctly with seasonal input was not considered significantly better than using annual input for annual predictions in Illinoia and Indiana. Illinois and Indiana were considered representative of the 5-state area. Therefore, the comparison between annual trend predictions from both seasonal and annual was not done for the other 3 states.

Seasonal District Predictions from Seasonal Input. Development and testing of methods for predicting seasonal precipitation totals as well as annual totals was proposed for this research. The same analysis techniques (non-integer spectral, and filtering) were applied to seasonal precipitation totals in the same manner as was done for annual amounts. Spectral and filtering algorithms analyze any precipitation total as a unit of information. Their outputs provide spectra in terms of seasonal periodicity characteristics (wavelengths, amplitude, and phase point) when seasonal totals are used as input.

For this analysis, seasonal totals were computed from monthly totals for winter, spring, summer, and fall, where the winter season started with December of 1931 and ended with November 1972. Starting the analysis year in

Table 4. Comparison of yearly predicted precipitation trends (1973-1975) with actual trends using filter method and 1932-1972 seasonal data base.

Trends, UP (U) and DOWN (D)
(Predicted on the left and actual on the right)

	Illinois					Indiana				
	Year					Year				
Districts	1973	1974	1975	Correct		Districts	1973	1974	1975	Correct
1(NW)	D/D	U/U	U/D	67		1(NW)	D/D	U/U	U/U	100*
2(NE)	D/D	U/U	D/D	100*		2(NC)	U/D	U/D	D/U	0*
3(WC)	D/U	D/D	U/D	33		3(NE)	U/D	U/D	U/U	33*
4(C)	U/U	D/D	D/D	100*		4(WC)	D/U	U/U	D/D	67*
5(EC)	D/D	U/D	D/D	67*		5(C)	U/U	U/D	D/D	67*
6(SW)	U/U	U/D	U/D	33*		6(EC)	D/D	U/D	D/U	33*
7(SE)	U/U	U/U	D/D	100		7(SW)	U/U	D/U	U/U	67*
8(SSW)	U/U	D/D	U/U	100*		8(SC)	U/U'	D/D	D/U	67*
9(SSE)	U/U	D/D	D/U	67*		9(SE)	D/U	D/D	D/D	67*
%Correct	89	78	56	74		Correct	56	44	67	56

*Predictions were the same as those from annual input (Table 3).

December provided a 3-month seasonal sequence which included a June-July-August seasonal which is agriculturally relevant in our 5-state study area.

Predicted and actual up-down trends are shown in Table 5 for the eight seasons of 1973 through 1974. Predictive skill for the first season should be better than for succeeding seasons. This result was definitely indicated for the nine Indiana districts. In Illinois, first season predictions were also better than chance expectation. First season predictions for the 18 Illinois and Indiana districts combined were 83 percent (15/18) correct. However, there was not a definite tendency for predictions to decrease in accuracy for subsequent seasons. All spring predictions in 1973 (second season) for both states were incorrect (18 down predictions for 18 acutal ups). The 18 actual spring ups followed 18 actual winter downs. Periodicities used were not sensitive to this change in seasonal trend. Later seasonal predictions (winter 1974, for example) were better (14/18) than predictions for the preceding spring, summer, and fall of 1973. However, it is the opinion of the author that improvement in accuracy following inaccurate predictions is not as authentic as prediction accuracy for previous seasons.

Areal Spectral Analysis and Prediction

The report thus far has dealt with analyses and predictions for districts as the areal sampling unit. Both seasonal and annual time units were involved in the precipitation amounts. The optimum or appropriate amount of integration and smoothing of climatic data in time and space prior to spectral analysis is not well understood. It is reasonable to assume that forces which may influence climate in a periodic manner will exert their effect over a large area. The area may well be much larger than a crop district. The periodic

Table 5. Seasonal precipitation trend predictions compared with actual for 1973 and 1974, Illinois and Indiana using filter method and 1932-1972 seasonal totals for the prediction basis.

Trends, UP(U) or DOWN (D) with predicted on the left and actual on the right

	1973				1974				
Districts	Winter	Spring	Summer	Fall	Winter	Spring	Summer	Fall	Correct
				Illinois					
1(NW)	D/D	D/U	U/D	U/U	U/U	U/D	D/D	D/D	63
2(NE)	D/D	D/U	U/D	U/U	U/U	U/D	D/D	D/D	63
3(WC)	U/D	D/U	D/D	U/U	U/U	U/D	D/D	D/D	63
4(C)	D/D	D/U	U/D	U/D	D/U	D/D	U/D	D/D	38
5(EC)	D/D	D/U	U/U	U/D	D/U	U/D	U/D	D/D	38
6(SW)	U/D	D/U	D/D	D/D	D/U	U/D	U/D	D/D	38
7(SE)	D/D	D/U	U/U	D/D	U/U	D/D	D/D	U/D	63
8(SSW)	D/D	D/U	U/D	D/U	D/D	U/D	D/D	U/D	50
9(SSE)	U/D	D/U	D/D	D/U	D/D	D/D	D/U	U/D	38
%Correct	67	0	56	56	67	33	56	78	52
				Indiana					
1(NW)	D/D	D/U	U/D	U/D	U/U	D/D	D/D	U/U	63
2(NC)	D/D	D/U	U/D	D/D	U/U	D/D	D/D	U/U	75
3(NE)	D/D	D/U	U/U	D/D	U/U	D/D	U/D	U/U	75
4(WC)	D/D	D/U	U/U	D/D	U/U	D/D	D/D	U/D	75
5(C)	D/D	D/U	U/U	D/D	U/U	D/D	U/D	D/D	75
6(EC)	D/D	D/U	U/U	D/D	U/U	D/D	D/U	U/D	63
7(SW)	D/D	D/U	U/D	D/U	D/D	U/U	D/U	U/U	50
8(SC)	D/D	D/U	U/D	D/U	D/D	D/U	D/D	U/U	50
9(SE)	D/D	D/U	U/U	D/U	U/D	D/U	D/D	U/U	63
%Correct	100	0	56	56	89	78	56	78	65

signal may also be relatively weak among other more local and non-periodic factors. Consequently, it is reasonable to study spectra of average precipitation amounts over several contiguous districts. Averaging over areas larger than the basic district areal unit should smooth the data in a manner that will reduce the effect of local and more random factors, and allow spectral analysis procedures a better opportunity to determine evidence for periodic influences.

The selection of a region within the 5-state study area for spectral analysis was influenced by previous areal coherence analysis (see section on spatial spectral variation). Periodicities with similar wavelengths in adjacent districts were used as a basis for areal integration. An experimental region of 30 contiguous districts with 3.8 to 4.0-year periodicities in Iowa, Missouri, Illinois, and Indiana (Fig. 4a) was chosen for further study.

Areal Precipitation Trend Predictions. A 30-district regional average was computed from normalized district monthly precipitation. Normalization was accomplished by dividing each district monthly precipitation total (average of station totals within the district) by the 1931-1971 district average. Spectral analysis of this set of 41 annual averages was performed. Periodicities of 3.8 and a 3.3 years were the most significant periodicities. Other periodicities (Table 2) were not significant at the 10% level. Data smoothing by averaging over districts apparently reduced the number of significant periodicities to the two periodicities (3.8 and 3.3) with the most areal coherence (Figs. 4a and 8a).

Trend predictions were computed for each of the 4 years (1972-1975) withheld from the spectral analysis. Only the 3.8-year periodicity was used as a single predictor for the first prediction trial. The 30-district area annual

trend predictions were correct for each of the 4 years. When the 3.3-year was used with the 3.8, predicted trends for the first two year (1972 and 1973) were correct but the 1974 and 1975 predicted trends were incorrect.

The above areal trend predictions were encouraging. Further analyses with larger and smaller groups of contiguous districts based on spectral coherence and correlation should be done.

Temporal Spectral Variation. Considerable attention was given to the areal coherence of district periodicities in a previous section (Areal Spectral Variation) of this report. Temporal as well as spatial coherence is considered pertinent to a prediction approach that uses periodicities, since predictions will be made on the assumption that the periodic functions of the past continue into the future.

The existence of temporal spectral variation was suggested by differences in significant periodicities determined from 1931-1975 and 1932-1972 data records. Differences may be seen by an examination of Figures 4a and 4b and Table 2. For example, Illinois district 1 (northwest Illinois) had a significant 3.9-year periodicity in the 1931-1975 record. This periodicity was not significant at the same significance level (10%) in the 1932-1972 analysis.

Temporal variation was examined in more detail by computing spectral outputs for 20-year samples within the 1931 to 1975 record. The analysis was performed on the 30-district annual area averages discussed in the previous section. The 20-year analyses began with the annual precipitation record for 1931-1950 and proceeded through the 45-year record in a "sliding fashion" with 20-year records determined by dropping the first year and picking up the year following year 20 of the previous sequence. A 20-year record was used in the moving spectral analyses in order 1) to search for temporal variation in 2- through 5-year periodicities, and 2) because 20 years is considered the

necessary record length (see Length of Observation Section) from which periodicities of up to 5 years in wavelength can be determined with reasonable accuracy by the non-integer, spectral analysis algorithum.

A tabulation of the strongest (most significant) periodicity determined from each 20-year analysis of the 30-district area is presented in Table 6. The periods (wavelengths) in Table 6 basically represent variations within the 3.2 and 3.9-year periodicity groups discussed in previous areal coherence sections. Wavelengths for the 26 samples varied from 3.1 to 3.9 years (a difference of 0.8 year) during the overall 45-year data record. The dominate periodicity for the 20-year records was shortest in wavelength during the middle of the 45-year data reocrd and longest during more recent 20-year periods. The maximum change from one 20-year record to the next was 0.3 year. Generally, the change was either 0.0 or 0.1 year.

A previous spectral analysis of the 1931-1971 record had yielded a periodicity of 3.8 years for wavelength with a correlation of 0.40 and an amplitude of 2.46 (last row of Table 6). These values are not averages of corresponding statistics of the twenty-two (1931-1971) 20-year analyses. The average of the 22 correlations (1931-1950 through 1952-1971) is 0.51 as compared to 0.40 for the 41-year analysis. Similarly, the average of the 22 wavelengths (periods) and amplitudes are 3.1 compared to 3.8 and 3.51 compared to 2.46, respectively, for the 41-year analysis. Thus, averages for measures of association (correlation and amplitude) of the 20-year analyses are larger than the correlation and amplitude obtained from the 41-year record. However, the wavelength average is smaller than the wavelength from an analysis of the encompassing 41-year record. Variation in the phase was anticipated, since the starting point of the 20-year analyses moved along the periodicity or periodicities in the data under investigation.

Table 6. Moving spectral outputs from 20-year records of annual precipitation totals over a 30-district region in Iowa, Missouri, Illinois, and Indiana, 1931-1975.

Sample No.	Years	R*	Period	Amplitude	Phase
1	1931-50	.60	3.6	3.10	0.99
2	1932-51	.55	3.3	3.15	0.54
3	1933-52	.57	3.2	3.34	-0.22
4	1934-53	.58	3.2	3.80	-1.14
5	1935-54	.58	3.3	3.66	0.77
6	1936-55	.49	3.3	3.09	-0.32
7	1937-56	.53	3.2	3.37	-0.92
8	1938-57	.56	3.2	3.73	1.25
9	1939-58	.55	3.1	3.63	0.58
10	1940-59	.55	3.1	3.63	-0.40
11	1941-60	.46-	3.1	2.94	-1.34
12	1942-61	.47-	3.1	3.16	0.87
13	1943-62	.48-	3.1	3.25	-0.13
14	1944-63	.48-	3.2	3.46	-1.38
15	1945-64	.43-	3.3	3.16	0.51
16	1946-65	.55	3.6	3.77	-1.77
17	1947-66	.58	3.5	4.01	1.17
18	1948-67	.56	3.5	3.87	0.11
19	1949-68	.55	3.4	3.84	-0.54
20	1950-69	.63	3.5	4.41	1.69
21	1951-70	.53	3.4	3.74	0.89
22	1952-71	.48-	3.7	3.13	-0.81
23	1953-72	.50	3.7	3.28	-1.83
24	1954-73	.67	3.9	4.77	0.29
25	1955-74	.67	3.9	4.80	-0.61
26	1956-75	.60	3.9	4.16	-1.62
	1931-71	.40	3.8	2.46	0.76

*A multiple correlation (R) of 0.49 is required for significance at the 10% level for 20-year samples.

-Not significant at the 10% level.

The wavelengths of the last four lines of Table 6 probably indicate at least a partial explanation for the four correct annual areal predictions (1972-1975) discussed in the previous section of this report. The last five 20-year records had dominate 3.7 to 3.9-year periodicities. The 3.8 used in the filter prediction process is practically an average of the most significant wavelengths of the last six 20-year records. Thus, predictions were based on a relatively significant and existent periodicity.

On the basis of 20-year analyses through the data basis period ending in 1971, a 3.4 (the dominate and significant) instead of a 3.8-year periodicity would probably have been used as a predictor. However, it is unlikely that the four annual predictions for 1972-1975 would have been appreciably different. A discussion regarding possible influence of small temporal variation in wavelength on prediction is included in the summary and recommendations.

Bandpass Versus Filter Rationale

Now that both bandpass and filtering prediction techniques have been described, differences in their rationales can be discussed and more easily understood. Both techniques are based on sample estimates of the three basic characteristics of cyclical variation; namely, amplitude, period, and phase point (often referred to as phase angle).

Bandpass utilizes amplitude, period and phase point of significant periodicities determined during a spectral analysis. All three cyclical characteristics are based directly on the whole data analysis sample. Amplitude and period are based on a spectral analysis of the whole sample, and the phase point determined is located near the beginning of a data record (Fig. 5).

The filtering technique utilizes the same significant cyclical tendencies but involves an intermediate step, namely that of determining a filter output from which the amplitude, period, and phase point sample estimates are determined. Amplitude, period, and phase for each significant periodicity are determined from approximately the last two peaks and valleys of a filter output. Therfore, a sample amplitude and period may be different than those utilized by bandpass, since a filter output indicates the waveform often varies in amplitude and period during the period of record (note Figure 17 and Table 2) of previous sections. The phase point, in case of the filter method, is located (or determined) at the last peak amplitude of the data record, instead of near the beginning in the bandpass approach.

Thus, in the filter approach, emphasis is placed on the most recent and up to-date data knowledge. The filtering approach and its associated manner of determining the amplitude, period, and phase are essentially an attempt to circumvent problems due to variable amplitudes and periods, and phase changes (waveforms damping out and reforming) which may have occurred earlier in the historical record. It is apparent that the filtering approach places a greater emphasis on periodic occurrences just before the prediction period starts, whereas bandpass is based more on representative periodic occurrences over long sample records. However, the success of any predictive system that uses cyclical tendencies is determined by whether the chosen periodicities reoccur during the forecast period.

Weakness of Filtering Method

Filtering has an inherent weakness related to its application for getting an accurate output at the beginning and ending of a time series record. Referring back to Figure 13, it is evident that as the filter is moved to the right along the data, 17 filter points will have passed beyond the right hand

end (1972) of the data when the filtering process is completed. A dampening or some type of modification of the output waveform begins as the right end of the filter progresses beyond the observed data. In order for the filtering computation to run to completion, the data record is extended with 17 values equal to the mean of the data analysis period. Filtering theory for this type of application has not been developed to the state of being able to avoid this "end effect."

The anticipated "attenuation" on a filter output is illustrated through the use of some experimental data which were generated from four known harmonics (sine-cosine waves) with periods (wavelengths) of 4, 6, 12, and 24 years and with amplitudes of 3.0, 2.0, 1.0, and 0.5, respectively. The resulting experimental data series is presented in Figure 19for a record length of 97 years. Filter outputs for 6-year and 12-year filters (two of the four cycles known to be in the data series) are shown in Figures 20 and 21, respectively. A "rounding" or dampening of the amplitude of both filter outputs is evident at the beginning and ending of each series. This "end effect" occurred as a result of the left half of the filter being applied to the data series mean value at the start of filtering, and the right half being applied to the mean as it moved past the data end on the right (latest year). The end effect is only critical on the right since this part of a filter output is emphasized in phase, amplitude, and period determinations.

The end effect pattern (Fig. 20) for the 6-year filter output for the 97-year data series is very similar at the beginning and the end. An examination of the end effect pattern for the 12-year filter output (Fig. 21) reveals some differences in dampening at the first of the output series, as compared to that at the last. An even greater difference between the beginning and ending effects occurred when the last two of the input data

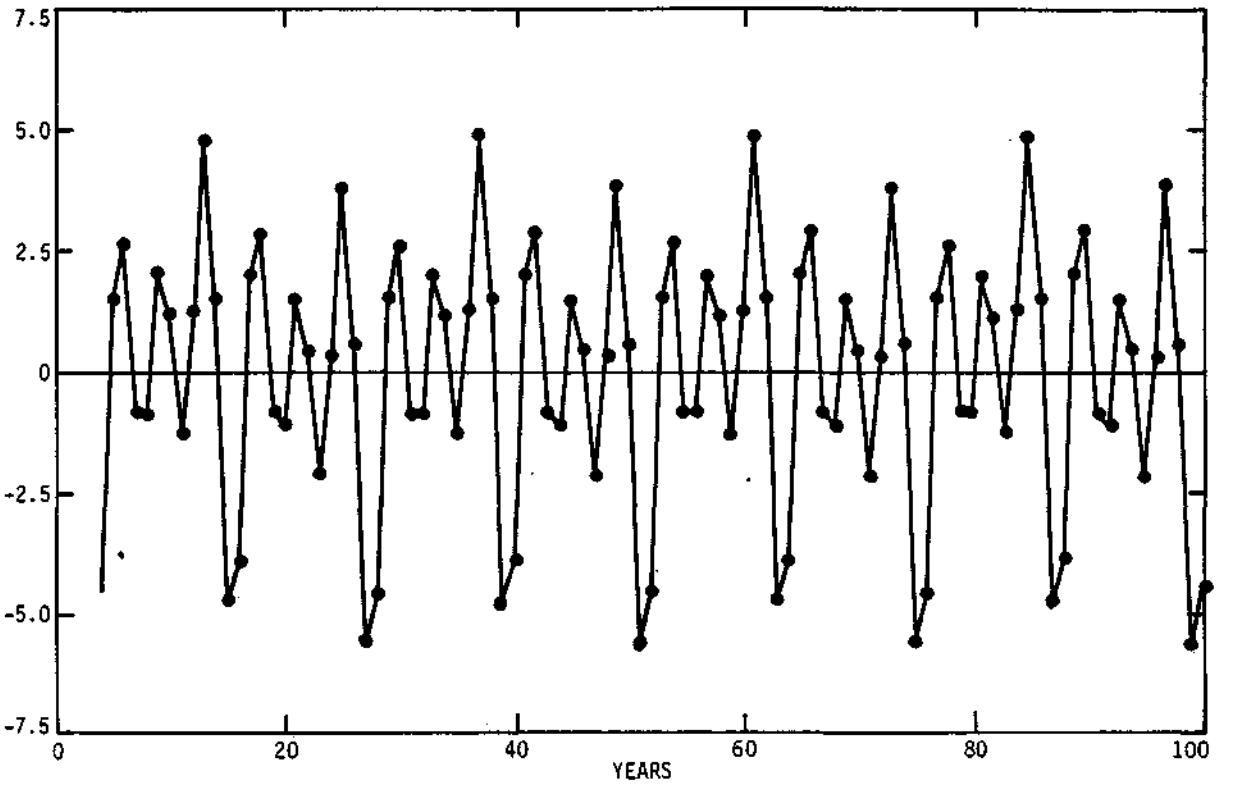

Figure 19. Experimental data series generated from 4-, 6-, 12-, and 24-year cycle with amplitudes of 3.0, 2.0, 1.0, and 0.5, respectively.

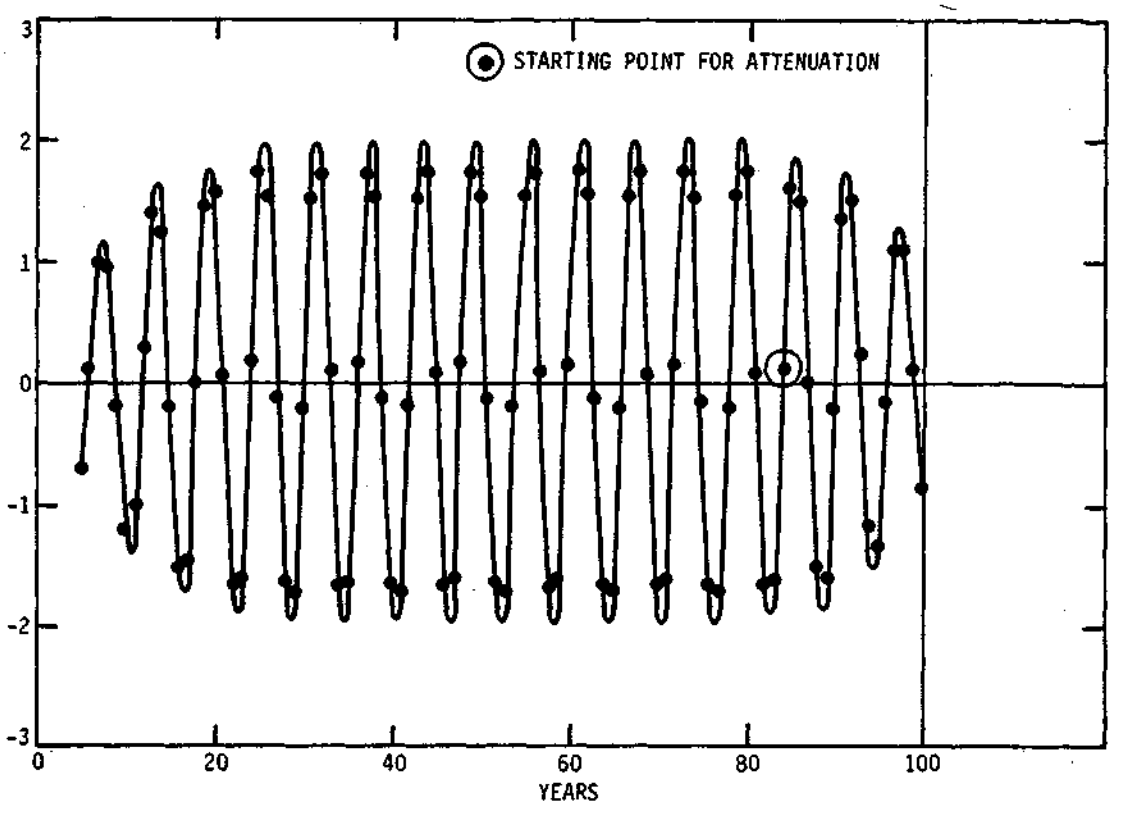

Figure 20. Filter output of 6-year digital filter applied to experimental data of figure 19.

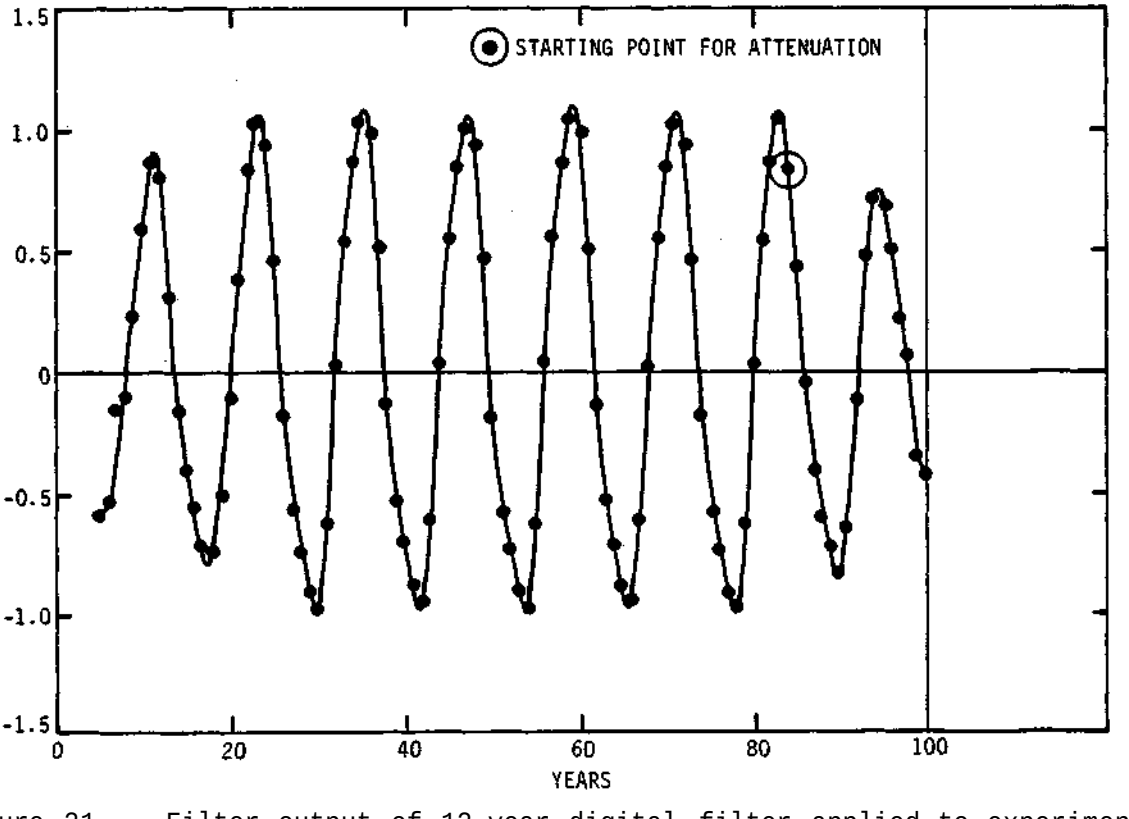

Figure 21. Filter output of 12-year digital filter applied to experimental data of figure 19.

points were removed before filtering (Figs. 22 and 23). Thus, the end effect pattern is influenced by the data input series values and their periodicity characteristics near the end of the series. The influence could also be modified by the random component in a natural data series. The end effect for filter outputs of geophysical, meteorological, agricultural and other types of data is not simple to evaluate.

Filtering Correction Constants

As discussed in the previous section, an end effect problem exists in the filtering method. Some way of making a correction is needed. It was, therefore, assumed that the end effect is one of dampening. Correction factors or constant multipliers were determined as described in the next section to provide a way of expanding the dampened amplitudes. These multipliers are tabulated in Table 7. The multipliers start with a value of unity (no correction) and increase as each of the 17 right side digital filter points are moved past the last data (year) input value. They are designed to provide an increasing amount of adjustment as the anticipated attenuation increases, when more and more filter points are moved past the data. Therefore, the multipliers are computed values which will produce average adjustments. Some adjustments will be correct, some amplitudes will be over corrected, and still others will not be corrected enough. Over a number of adjusted filter outputs, the average adjustment should be correct. Two examples of the resulting application of these factors are shown in Figures 24 and 25 for the 6-year and the 12-year filter outputs from the experimental data of Figure 19. An overcorrection of the dampening occurred in both of these examples. The adjustment factors of Table 7 were used in the 45 district predictions for comparisons with bandpass results.

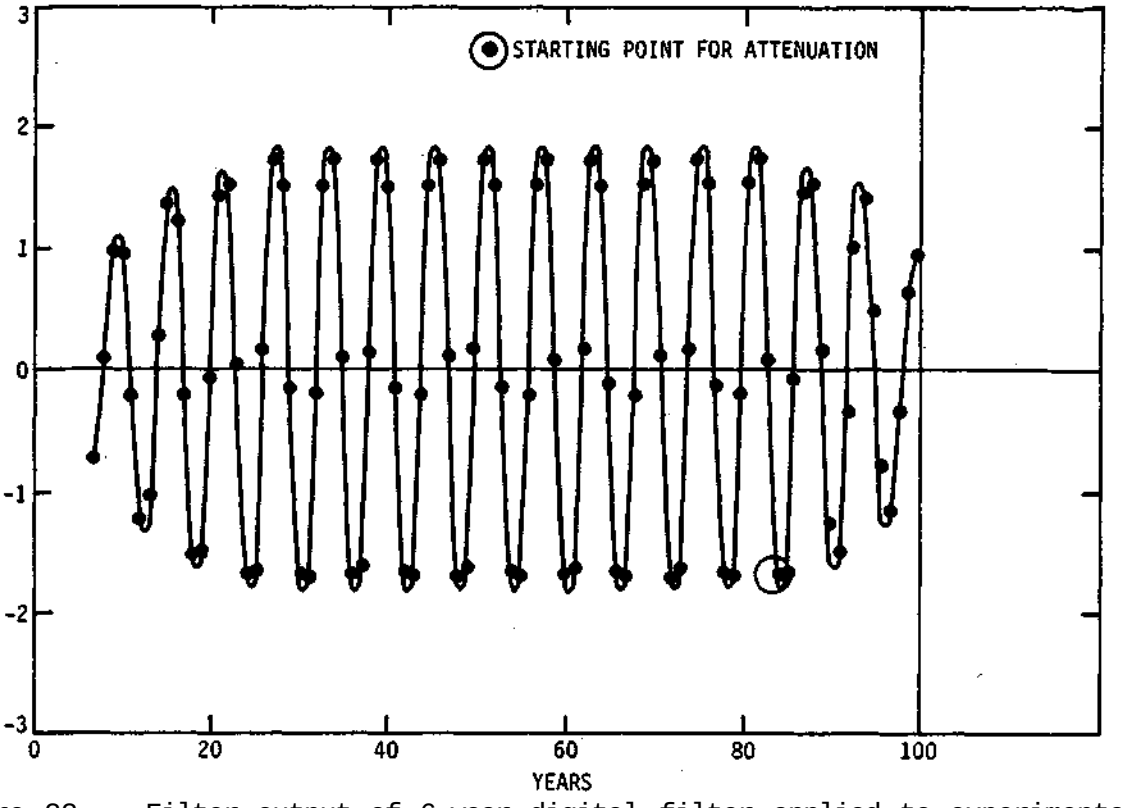

Figure 22. Filter output of 6-year digital filter applied to experimental data of figure 19 with last two data points removed before filtering.

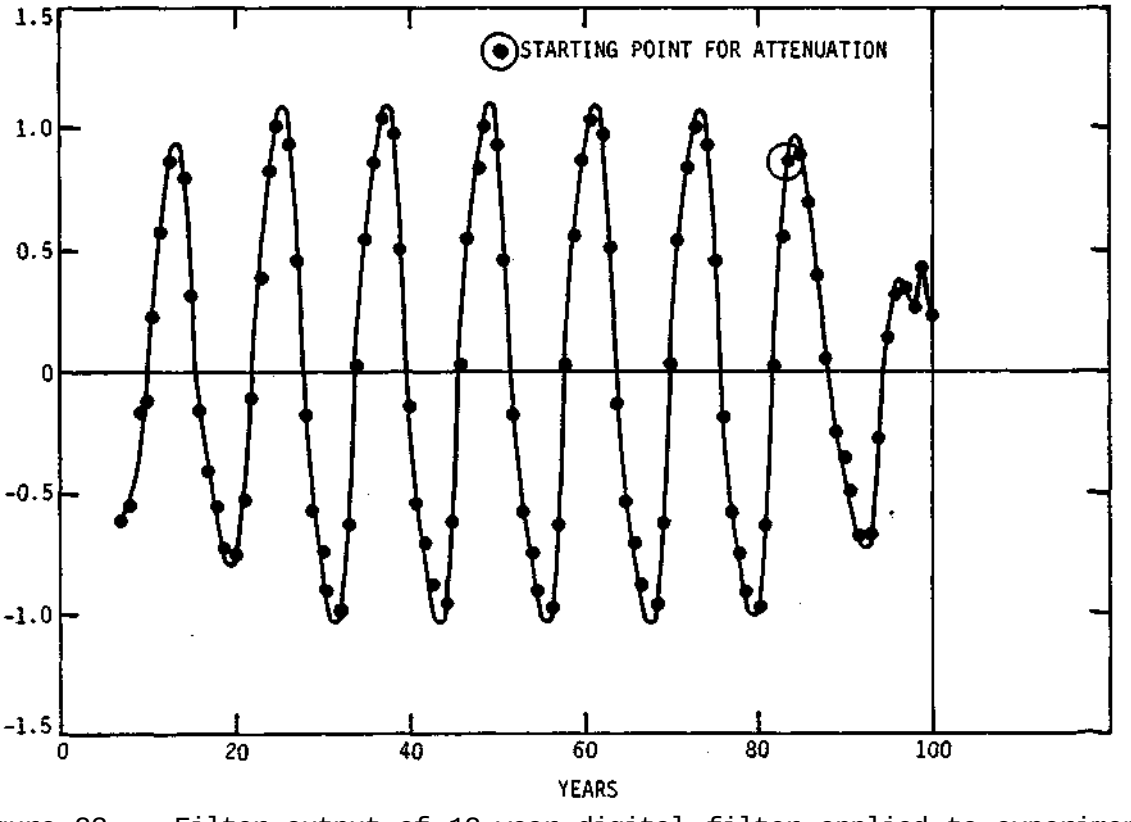

Figure 23. Filter output of 12-year digital filter applied to experimental data of figure 19 with last two data points removed before filtering.

Table 7. Multipliers for filter attenuation correction.

Number	Multiplier
1	1.00
2	1.01
3	1.02
4	1.04
5	1.06
6	1.09
7	1.12
8	1.16
9	1.20
10	1.24
11	1.29
12	1.35
13	1.41
14	1.49
15	1.58
16	1.68
17	1.80

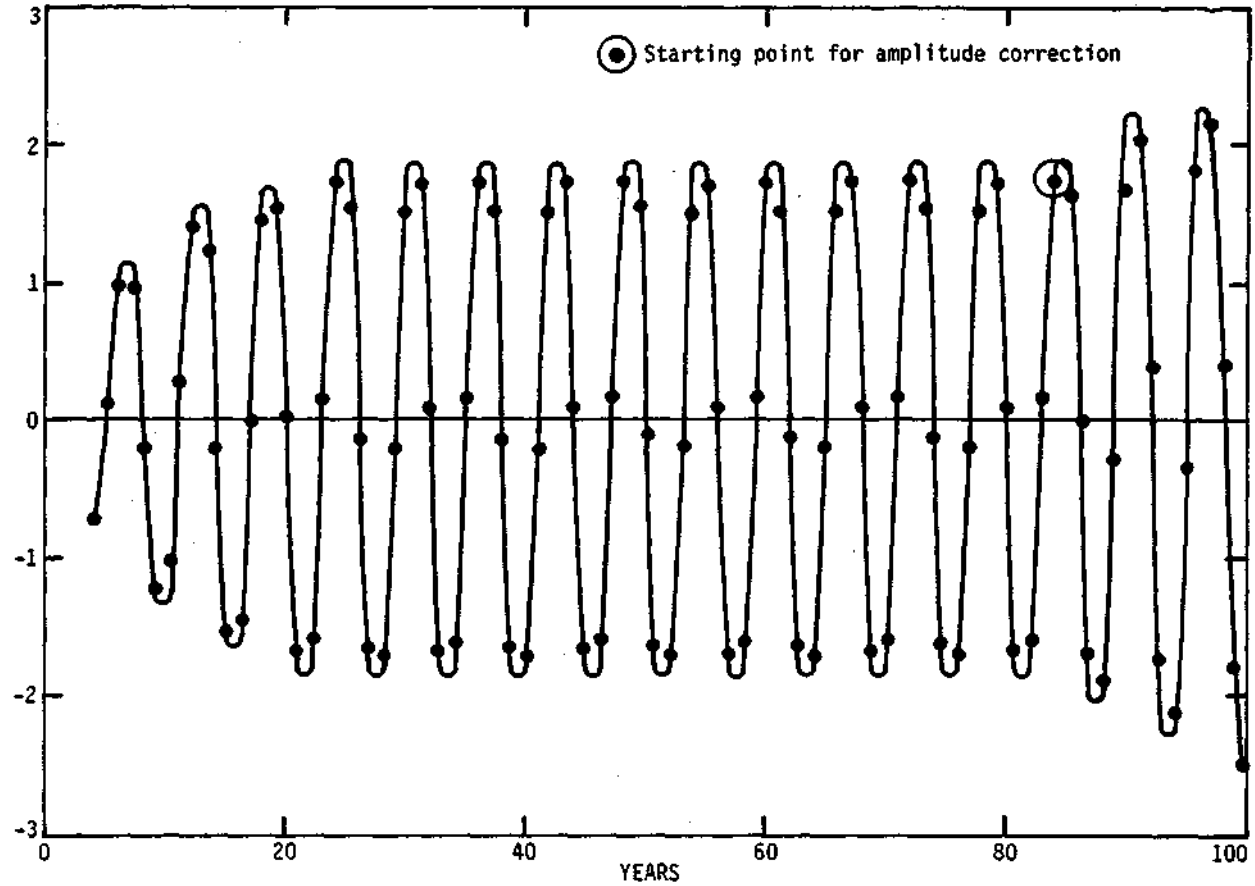

Figure 24. Amplitude correction multipliers applied to the last 17 points of the 6-year filter output of figure 20.

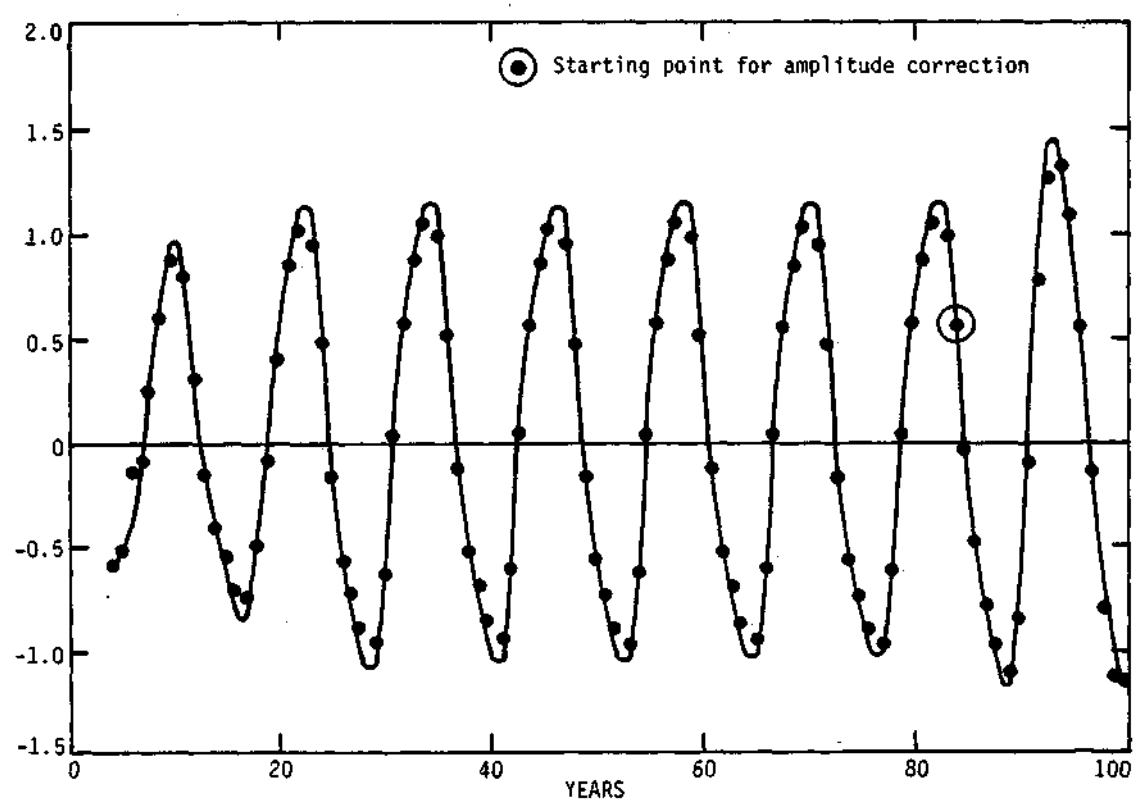

Figure 25. Amplitude correction multipliers applied to the last 17 points of the 12-year filter output of figure 21.

Calculation of "End Effect" Correction Constants

1. Select a filter, for example, the one for 12 units (years).
2. Construct a test series of unit amplitude which is a pure sine series with a wavelength of 12 years. The sine series should be 5 or 6 cycles long or longer. The series used for the current discussion was 8 cycles in length plus one year.
3. Add 17 zero values (the mean of the series) to the series to allow the filtering process to proceed to completion; that is, the filter is moved along the data until the central value of the filter is on the last point of the sine series. A filter output value is produced for each data point of the sine series. Figure 26 shows the original sine series and the 17 zero values extended on the right.
4. Figure 27 is a plot of the filter output. The test series was reproduced exactly through the center wavelengths (between point A, 17 points from the left and point B, 17 from the end on the right, when the filter is entirely within the original series. An increasing amount of dampening (reduction in amplitude) is evident throughout the last 17 points of the filter output as an increasing number of filter values were applied to the zeros.
5. The amplitude corrections required to bring the dampened amplitudes back up to that of the original series are the ratios of the original amplitude (unity) to each of 17 points (years) on an "attenuation line" (Fig. 27). The ratios increase from 1.0 (no correction) to about 1.8 on the extreme right.
6. The attenuation line will vary with each filter. . Consequently, the process was repeated with several different filters. Averages of the ratios were computed to obtain the multipliers shown in Table 7 of this report.

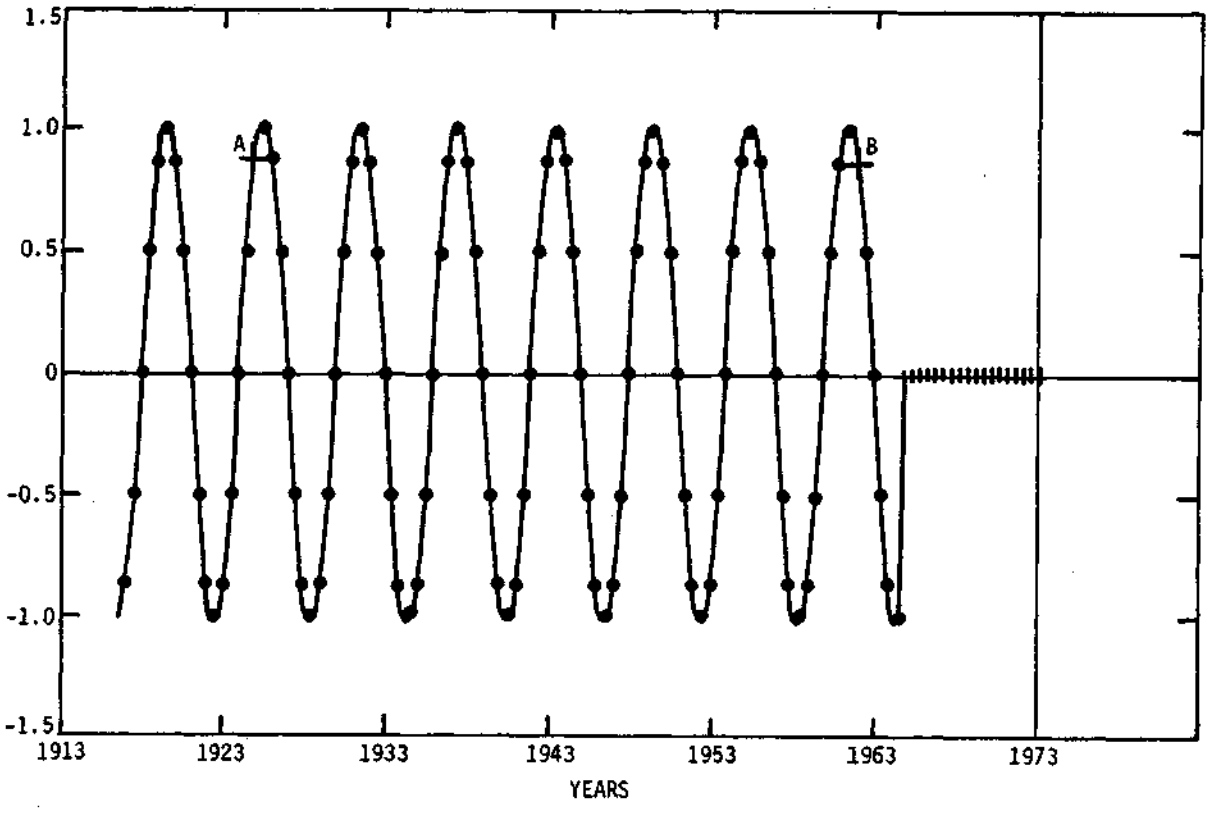

Figure 26. A sine series generated with unit amplitude and 12-year period (wavelength) and 17 zeros on the right.

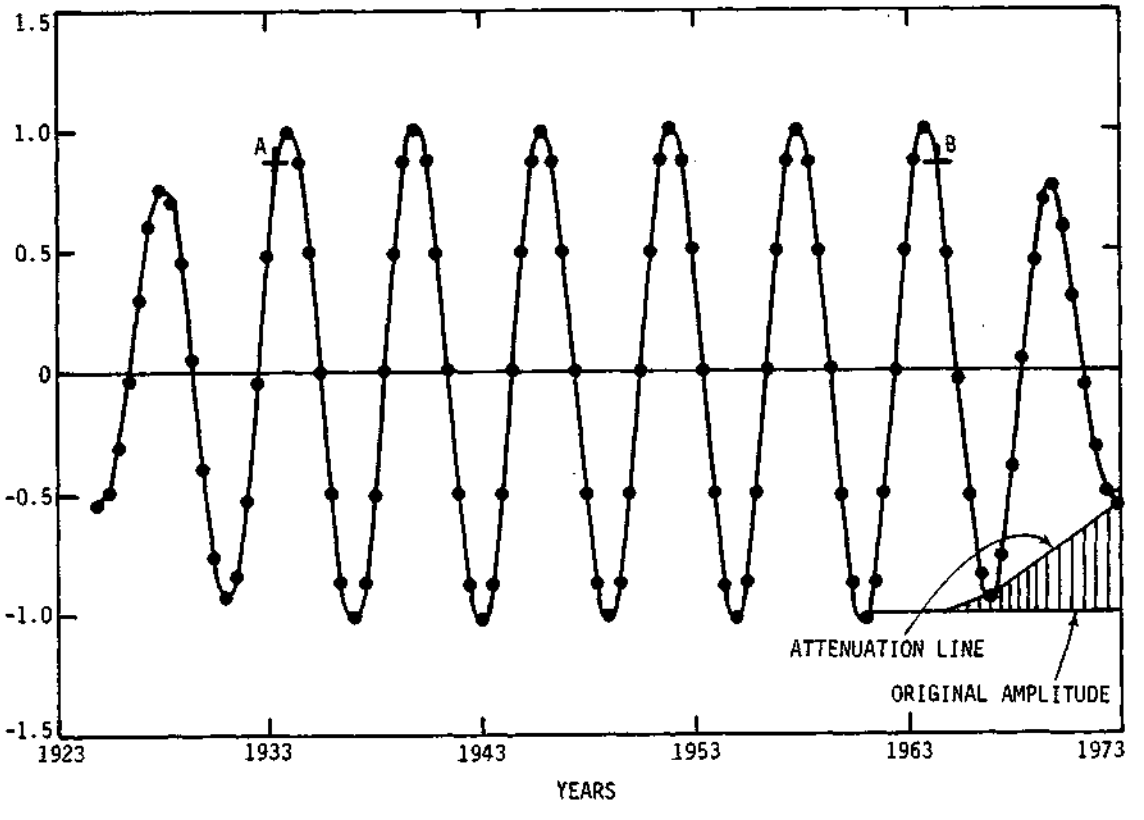

Figure 27. Filter output of figure 26 using a 12-year digital filter on the generated series of figure 26.

Length of Observation Requirements

Early in the study, the author felt that an observation record considerably longer than the wavelength of a periodicity was necessary for the non-integer algorithm to accurately determine the characteristics (period, amplitude, and phase) of a periodicity. The 97-point experimental data set of Figure 19 was considered a good set for a study of length of requirements for record. It contained periodicities of known wavelengths (4, 6, 12, and 24), amplitudes (3.0, 2.0, 1.0, and 0.5, respectively) and was reasonably long (97 points or years). A moving spectral version of the non-integer program was used to determine sample spectral characteristics for each known periodicity using various record lengths. Sample spectral characteristics of a portion of this experimental study are presented in Table 8.

Tabulations shown in Table 8 depict the variation in spectral characteristics which was determined from records of 24 and 48 years (time units). Several features are prominent in the table. These include: 1) correlations of the harmonics with the data values is proportional to the "built in" amplitude; 2) correlation, period (wavelength), amplitude, and phase vary as the analysis moves (sequentially) along the data record through samples 1 to 12; and 3) absolute magnitude of spectral variation decreased as the sample record length increased relative to the wavelength of a periodicity. Variation in phase point was expected since the starting point of each sequential sample was moved along each underlying wave causing the phase point (distance to first maximum amplitude) to vary with respect to starting observation.

There was very little variation in correlation, period, and amplitude when 24-year samples were used to search for the underlying 4-year periodicity. The periods varied from 3.9 years to 4.1 years and the amplitude ranged from

Table 8. Moving spectral analyses of experimental data prepared from sine waves with wavelengths of 4, 6, 12, and 24 years and amplitudes of 3.0, 2.0, 1.0, and 0.5, respectively.

Sample size = 24 years

Sample No.	R*	Period	Amp-litude	Phase Point	R	Period	Amp-litude	Phase Point	R	Period	Amp-litude	Phase Point
	4-year wave				6-year wave				12-year wave			
1	.83	3.9	3.14	-1.3	.53	6.0	2.0	-2.5	.27	12.5	1.02	-5.8
2	.79	4.0	3.00	1.3	.55	6.2	2.1	2.1	.28	13.0	1.04	5.0
3	.79	4.0	3.00	0.3	.53	6.1	2.0	1.3	.27	11.5	1.01	5.5
4	.79	4.0	3.00	-0.7	.53	6.0	2.0	0.5	.34	10.5	1.27	5.2
5	.80	4.1	3.03	-2.0	.54	5.8	2.1	0.0	.31	10.7	1.16	4.1
6	.79	4.0	3.00	1.3	.57	5.7	2.2	-0.7	.28	11.1	1.05	2.8
7	.80	4.1	3.01	0.0	.53	6.0	2.0	-2.5	.30	10.5	1.12	2.6
8	.79	4.0	3.00	-0.7	.56	5.7	2.1	2.8	.30	10.5	1.13	1.7
9	.80	4.1	3.03	-2.0	.53	6.0	2.0	1.5	.33	10.7	1.26	0.5
10	.79	4.0	3.00	1.3	.53	5.9	2.0	0.6	.30	10.8	1.14	-0.7
11	.80	3.9	3.02	0.6	.54	6.1	2.0	-0.7	.33	14.1	1.23	-5.8
12	.80	3.9	3.00	-0.3	.56	6.3	2.1	-2.2	.39	14.1	1.46	-6.8

Sample size = 48 years

	6-year wave				12-year wave				24-year wave			
1	.53	6.0	2.00	-2.5	.27	12.2	1.00	-5.5	.14	26.6	.54	9.0
2	.53	6.1	2.00	2.1	.27	12.2	1.01	5.6	.20	29.6	.74	4.6
3	.53	6.0	2.00	1.5	.27	11.9	1.00	5.2	.17	28.6	.63	4.5
4	.53	6.0	2.00	0.5	.29	11.4	1.08	5.1	.13	24.6	.50	6.6
5	.53	6.0	2.00	-0.5	.28	11.6	1.05	3.6	.14	26.2	.53	4.2
6	.54	5.9	2.02	-1.0	.27	11.8	1.01	2.4	.15	27.6	.57	2.0
7	.53	6.0	2.00	-2.5	.27	11.6	1.03	1.8	.14	25.3	.52	2.6
8	.53	5.9	2.01	2.8	.27	11.6	1.03	0.7	.13	24.4	.50	2.5
9	.53	6.0	2.00	1.5	.28	11.4	1.07	0.2	.14	25.9	.52	0.2
10	.53	6.0	2.00	0.5	.28	11.6	1.04	-1.2	.14	25.6	.51	-0.7
11	.53	6.0	2.00	-0.5.	.28	12.6	1.07	-4.4	.15	28.2	.56	-6.5
12	.54	6.0	2.00	-1.9	.31	12.8	1.16	-5.8	.17	30.3	.64	-11.1

*R is the multiple correlation of the sine-cosine waves with the data values.

3.00 to 3.14 (all ≥ 3.0) for the 12 samples shown in Table 8. A few amplitudes were slightly less than 3.0 for sample numbers 13 through 24 (not shown in the table). All periods and amplitudes from 48-year sample records (not shown) were the same as the amplitude and period of the underlying 4-year wave.

The 6-year underlying wave had an amplitude of 2.0. Amplitudes from the 24-year samples varied from 2.0 to 2.2 in the first 12 samples. The period varied from 5.7 to 6.3 (a variation of ±0.3). Increasing the sample size to 48 reduced the amplitude variation to ±0.1.

Amplitudes determined for the 12-year periodicity from the 24-year samples ranged from 1.01 to 1.46 (Table 8). They were never less than 1.00 (the underlying amplitude) for sample determinations from 13 through 24 (not shown in Table 8). The periods varied from a low of 10.5 to a high of 14.1 around the underlying values of 12.0. Increasing the sample size to 48 (4 times the 12-year wavelength) reduced the period deviation from the underlying 12 to approximately 50 percent of that found when 24-year samples were used.

The lower right hand portion of Table 8 shows spectral estimates of a 24-year wave with an amplitude of 0.5. The sample period was 48 years or twice the wavelength of the 24-year underlying periodicity. Correlations were small due to the low amplitude of 0.5. Waves with small amplitudes, among waves with larger amplitudes, produce a relatively small portion of the total variation in the data sample. The identity of waves with small amplitudes is difficult to establish in real observational data sets which include random variation as well as other possible periodic variation. Periods deviated as much as + 6.3 year from the underlying period of 24 years. Periods as short as 20 years occurred in some sequential samples from 13 through 24 (not shown in Table 8).

The above discussion involving record length necessary for the non-integer algorithm to accurately determine spectral characteristics was based on a generated data set which included specific cyclical (deterministic) component. Record length influences were also examined with actual precipitation data. Spectral analyses were done for each of the nine Illinois districts for two precipitation data records, 1931-1971 (41 years) and 1901-1971 (71 years). The two records differ in length by 30 years and the last 41 years are common to both records. Wavelengths and correlations from the two data records for each district are shown in Figures 28a and b for periodicities in the vicinity of 20 years and in Figures 28c and d for periodicities in the vicinity of 10 years. The shorter, 1931-1941 year, record is four times longer than a 10-year periodic oscillation and twice that of a 20-year periodicity. The 71-year record is approximately 7 and 3.5 times the 10 and 20-year oscillations, respectively.

The wavelength presented in Figure 28a and b from the two records are considerably different for each district. Since the data are observed data with essentially unknown deterministic components in contrast to generated data, it is not possible to attribute the differences to record length entirely. The two data records, although the same in 30 of the 71 years, are expected to contain different information content (means, variances, auto-correlations). Moving spectral analyses (Table 8) and filter outputs (Figures 15 and 16) are evidence that spectral characteristics vary temporarily.

Wavelengths in the vicinity of 10 years (Figure 28c and d) generally indicate smaller differences (percentages) between the two data records. Smaller differences could be attributed to record length relative to the wavelength under study. Wavelengths for the approximate 10-year periodicities are more uniform from the longer data record. When the 1901-1930 data were

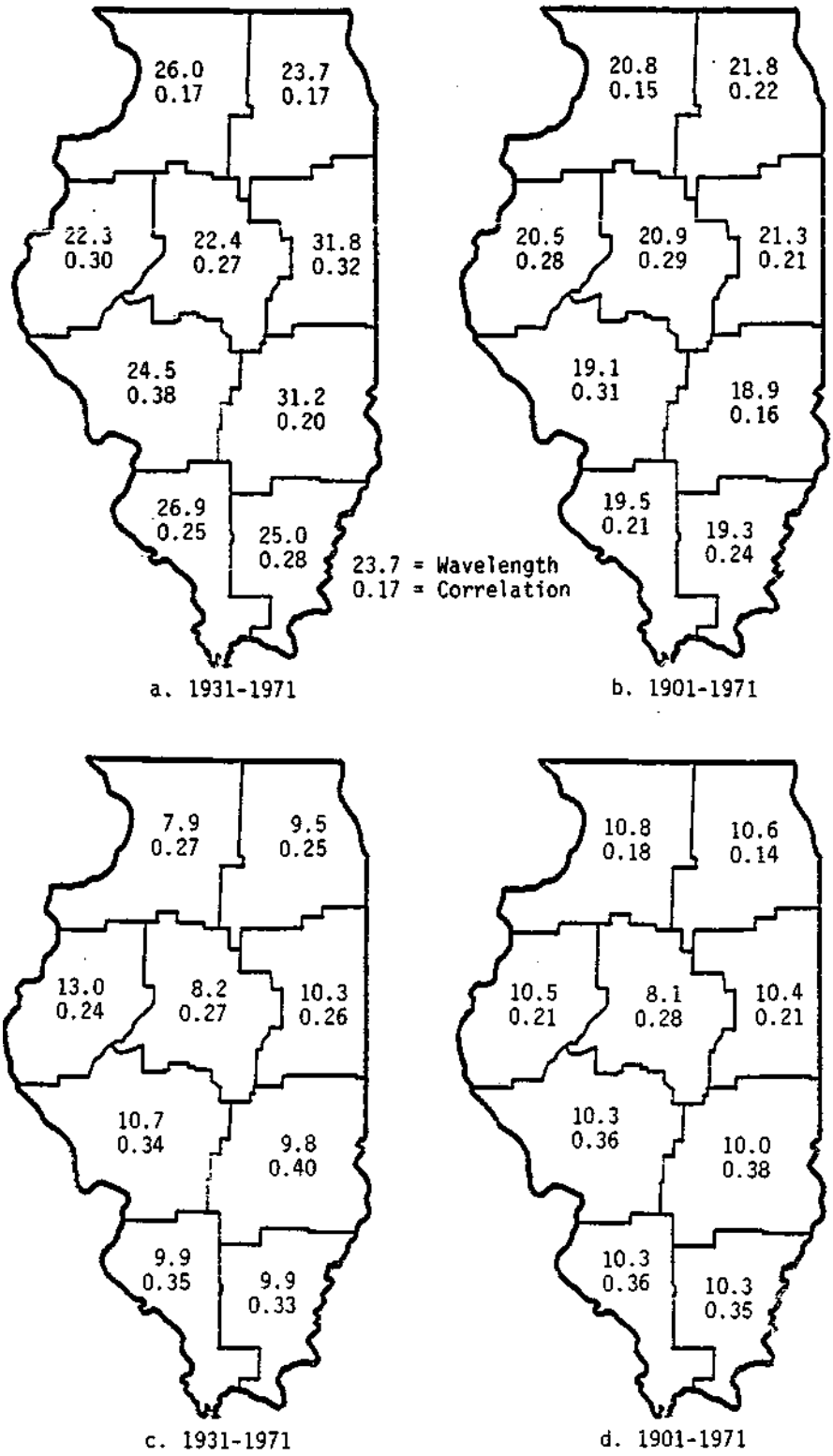

Figure 28. Annual precipitation periodicities from 1931-1971 and 1901-1971 in Illinois.

incltided, wavelength (10-year) changes were both up and down relative to those from the shorter 1931-1971 record. Wavelengths for the 20-year which all decreased with the additional record length. This result is not clearly understood but may result from the varying location of phase point relative to the time of the data observations. Wavelengths for the different sample results from generated data (Table 8) varied up and down around the actual underlying wavelengths.

Although wavelength consistency for a 3.9-year periodicity (a periodicity often identified in Illinois and surrounding districts) is not shown in Figure 28, it can be stated that 3.9-year wavelengths for Illinois districts did not vary from those for 1931-1971 when 1901 to 1930 records were included. Thus, on the basis of the results discussed here on both generated and observed data, the author recommends that a record length of at least four times the wavelength should be used to establish the characteristics of a periodicity.

SUMMARY AND RECOMMENDATIONS

Research for this project was designed to develop and test methods of predicting short term (1-3 years) annual and seasonal precipitation trends. Basic data used in the study were monthly precipitation averages for each of 45 crop reporting districts in a 5-state area (Illinois, Indiana, Iowa, Missouri, and Ohio) for the period 1931-1975. Since the research was agriculturally oriented, the seasonal totals used in seasonal analyses were December-February, March-May, June-August, and September-November, in order to include a total for the agriculturally relevant June-August season. Annual

totals were based on the 12 calendar months, January through December. Successfull prediction of precipitation trends (up or down) of even one season to one year in advance would be extremely useful for agricultural and water supply planning.

Time series analysis procedures were employed for the prediction study. The first step involved a search of the historical precipitation data for significant periodicities (non-random fluctuations). Significant periodicities were expressed mathematically as harmonics (sine-cosine waves) and used as predictor variables. Two variations (a bandpass method and a filtering method) were tested initially for their utility in the use of these predictors.

A summary of 1973-1975 annual prediction trend results for the bandpass method for Illinois districts are presented in Table 1. Eight of nine district trends (up or down) were correct for the first year (1973). Trend predictions for 1974 and 1975 were correct for only 56 percent of the cases, a result that is about the same as chance expectation.

Greater emphasis was given to a digital filtering technique which had been used by Bowen (unpublished research) for the prediction of precipitation trends in Australia. Annual trend prediction results from the filtering method are shown in Table 3 for each of the 45 districts in the 5-state area. Except for 1973 in Illinois, prediction trend skill was about the same as that for chance expectation.

The filtering method was also tried with Illinois and Indiana seasonal data input. Annual predicted district trends were obtained by accumulating the four seasonal predicted amounts. Fifteen of the 18 up or down trend sequences of these predicted annual amounts (Table 4) were the same as those obtained directly from annual input (Table 3). The overall average number of

correct trend predictions for the nine Illinois districts is 70% (Table 3) from annual input and 74% (Table 4) from seasonal input. The accuracy for Indiana is the same for both seasonal and annual input. . Therefore, according to this comparison of annual trend predictions from the filter method there was no significant difference between using annual and/or seasonal input.

Seasonal precipitation trend predictions from the filtering method were compared with actual precipitation trends for 1973 and 1974 in Illinois and Indiana (Table 5). District seasonal trend predictions for the first season (Winter, 1973) were above chance expectation (67% for Illinois and 100% for Indiana). Predictions for the following spring, summer, and fall seasons of 1973 exhibited no predictive skill. The 1974 winter season prediction accuracy was again better than chance expectation. Spring and summer prediction accuracy was variable from above to below chance. Predictions for the fall season were 78% accurate.

Development of successful techniques for use in predicting weather and its accumulative and average features (which we think of as climate) has been one of man's ambitions for generations. Many techniques have been tried. For the research described in this report, time series statistical techniques involving spectral analysis, bandpass, and filtering of historical data series were used. These techniques are current statistical tools in use. . However, it is a fact that neither bandpass nor filtering techniques and other statistical procedures can succeed without, 1) proper identification of underlying periodicities (nonrandom fluctuations) in historical data of variables people want to predict, and 2) the underlying periodicities repeat their influence in the future.

Spectral analyses performed for this project have pointed out either a difficulty in identification of periodicities and their characteristics or that periodicities may not exist to the degree envisioned. Periodicities

found in precipitation data were generally not as consistent (coherent) as desired in either time or space. Unless a periodicity can be identified in adjacent districts, its presence in the data of a single district must be considered a random event. Spatial variation was evident in Table 2 and Figures 4a and 8a. Temporal variation was demonstrated in Table 6 and Figure 28. However, the question of what constitutes the same and/or a different periodicity from district to district needs clarification. Likewise, more knowledge is needed on what temporal variation in computed wavelength and amplitudes is really significant in the statistical sense and of practical importance in predicting.

A partial demonstration of how small differences in wavelength of a periodicity can change a prediction was inspired by Table 6 of this report. The wavelength of a 30-district area periodicity varied slowly within a range of 3.1 to 3.9 years as spectra were computed for 20-year samples along the 45-year annual precipitation (1931-1975). Another question is, do period changes from 0.1 to 0.5 year represent changes of practical importance in predictions? Six computed sine-cosine curves are shown in Figure 29. Each curve had the same amplitude and phase but the period was varied from 3.1 to 3.6 years. These curves all drawn for yearly computed values are very similar from their beginning to the sixth year. All up or down trends are in the same direction during the first 5 years. Between the 6th and 7th years, it is clear that the 3.1- and 3.2-year periodicities have changed to an upward trend, the 3.3-year is leveling off, and the 3.4 to 3.6-year perioditicies are still definitely downward. Thus, periodicities with wavelength differences as small as 0.1 and 0.2 year soon become out of phase and produce very different contributions to a prediction process at a specified time. In case, of the bandpass method, the phase angle or computed starting point is near the start

of a data record. By the time any of these curves passed along or through the data basis (analysis) period, say 20 years in this example, their projections into the predictions period would be quite different. For example, the 3.6-year periodicity would contribute about -1.4 precipitation units to be subtracted from a mean value, whereas the 3.2-year periodicity would contribute about +1.4 precipitation units to be added to a mean.

The point being stressed with the illustration in Figure 29 is that sample wavelength variation of as much as 0.5 year may in the above range be tolerated for up to 5 years after the phase angle (first maximum amplitude following the beginning of data record). Therefore, it may be important that phase points be fixed (determined) just prior to the prediction period, i.e., end of the data record rather than at the beginning. It is apparent from Table 6 that sample periodicities determined from historical data are not constant in wavelength and amplitude; instead, they are almost constantly varying, perhaps due to random variations in the data sample and other unknown causes.

Suggestions for future study are as follows:

1) Perform an extremely thorough spatial and temporal spectral analysis on the best data records available to gain a thorough understanding of periodic tendencies in the data sample. Moving spectral analysis of average data for groups of contiguous crop districts should be employed. Follow each moving spectral analysis with short term predictions in an effort to obtain numerous examples for skill determinations.
2) Devise a method of determining a phase point at the end of the historical data record or sample. The filtering approach the author experimented with in this report attempts to accomplish this but it

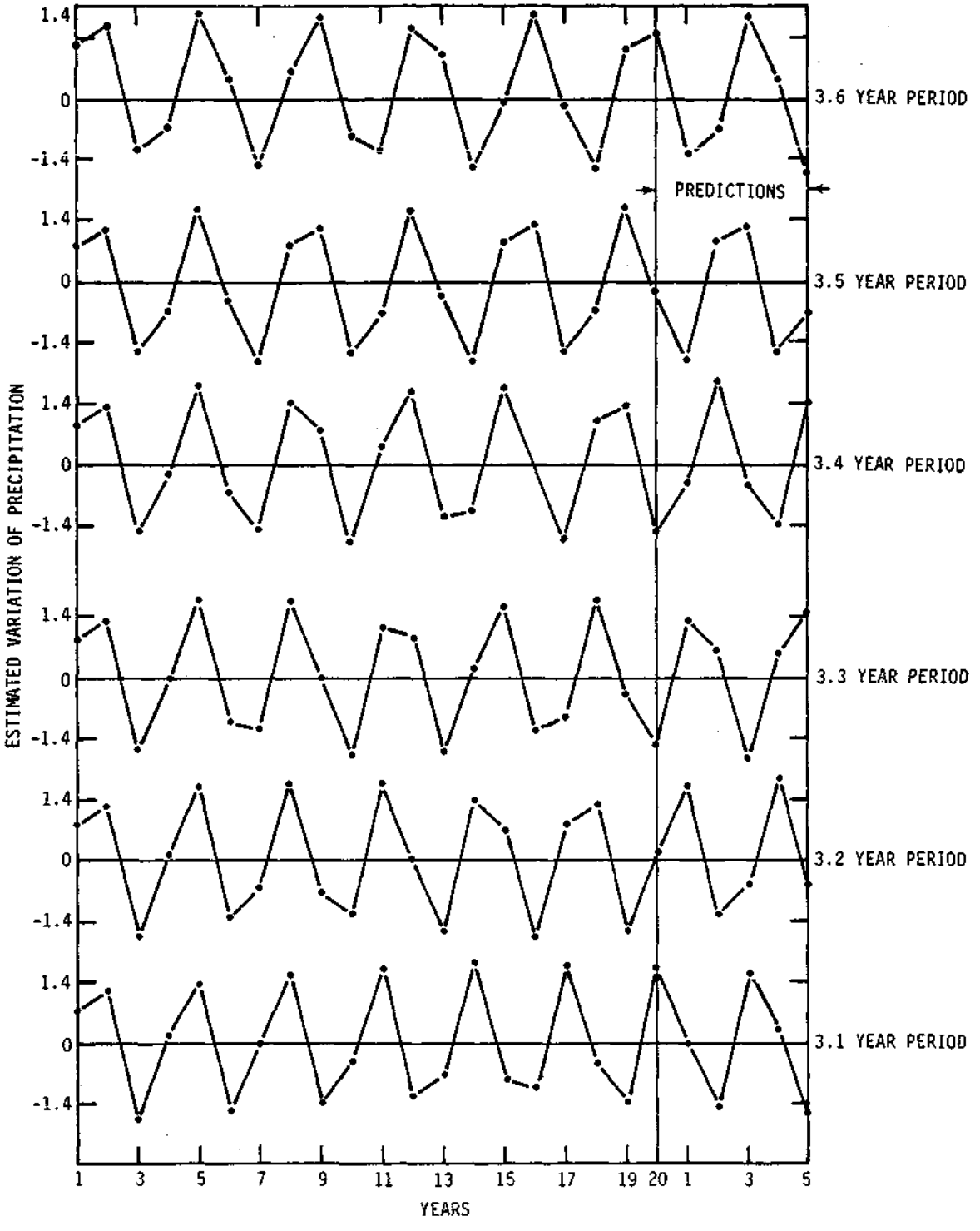

Figure 29. Six sine-cosine waveforms with the same amplitude and phase point and with different periods.

has amplitude "end effect" difficulties. Intuitively, it seems important to have the periodicity accepted as a predictor variable in synchronization with the very recent part of the data sample.

3) If more than one predictor periodicity variable is being used, each should be as much as possible, tied in with recent oscillations in the data analysis sample.

In summary, it is possible to reconstruct the past with only a few periodicities identified in the data. These periodicities generally are documented by rather low correlation coefficients. Those used in predictions had correlations in the range of 0.30 to 0.40. This represents an explanation, in the variance sense, of 9 to 16 percent of the historical sample. The periodicities are, therefore, not strong and their amplitudes are not large. Consequently, the predictive power of each is nominal.

Despite various problems encountered in this research, the results are sufficiently encouraging to recommend further search and documentation of precipitation periodicities in the Midwest and other regions. Likewise, associations and physical explanations of periodicities should be sought.

REFERENCE

Schickedanz, P. T., and E. G. Bowen, 1977: The computation of climatological power spectra. J. Appl. Meteor., 16(4), 359-367.

APPENDIX

Digital Filtering in Time Series Analysis for Annual Rainfall Prediction

Report on Consulting Activities
Prepared for the Illinois State Water Survey

by

W. K. Jenkins

1. Introduction

Since the patterns of annual rainfall have a strong influence on annual crop production, there has been a prevailing interest in techniques that can accurately predict rainfall by extracting behavior patterns from past records. It is well established that in Illinois (and other midwestern states) the average rainfall during the month of July is one of the most significant predictors of the fall corn yield. The work that is described in this report constituted an effort to develop techniques that will accurately predict the average July rainfall for several years into the future. It is apparent that a data sequence consisting of annual recordings of a given monthly average is inherently a discrete-time process, and for this reason, discrete-time processing techniques are required.

There are two fundamental schools of thought on the approach needed to properly analyze a finite length discrete-time sequence. The first technique uses spectral analysis in an attempt to identify the dominant sinusoidal components in the data record. These components can then be used to synthesize an approximation to the record that is extrapolated into the future. The FFT (Fast Fourier Transform) spectral analysis approach implicity assumes that the sequence being analyzed is periodic, and that the finite length of recorded data defines one period of the periodic data sequence. Therefore, if the sequence components have periods longer than the observation interval, the FFT spectral analysis will be unable to characterize the sequence properly, and the accuracy of the predictions will be limited.

The second basic approach is to assume that the data sequence is a sample sequence from a stochastic process. In this case the random nature of the process is considered to be dominant, and we do not attempt to find dominant sinusoidal components. Well known techniques such as Wiener and Kalman

filtering can be applied to generate an approximate reproduction of the data sequence that represents a minimum mean square fit [1]. Once the parameters of the stochastic model are established by exercising the filter on the data record, the prediction filter is then incremented into the future in an effort to predict future behavior.

It is important to establish at the onset that since the data sequence has been observed through a relatively short window (~ 60 yrs.), the exact nature of the sequence for all time cannot be known. Therefore it is a value judgement as to which of these basic approaches to pursue. For annual rainfall prediction, scientists at the Illinois State Water Survey decided that the spectral analysis model is most appropriate for this problem. This choice was guided by the premise that there is an oscillatory component in rainfall records which may be associated with an approximate 20-year sun spot activity. There are also periodic components due to periodic behavior in the relative orbital mechanics of the earth-sun system which may influence climate. Proceeding on the belief that periodicities exist in the data, and on the faith that periodic behavior is dominant, the prediction techniques throughout this work were based on spectral analysis.

Section 2 presents the FFT spectral analysis and describes how the spectral analysis can be accomplished by digital filtering. Both nonrecursive and recursive digital filters are discussed in an effort to clarify the differences and similarities between these two classes. Section 3 presents the design procedure used to obtain the recursive filters. Section 4 summarizes results of the predictions using the digital filters. Several computer programs were developed during the course of this work.

2. Spectral Analysis for Rainfall Predicton

Let x(n), n = 0, ..., N - 1, be the finite length sequence of recorded data (for example, annual July average rainfall). The discrete Fourier transform (DFT) X(K) is given by

$$X(K) = \sum_{n=0}^{N-1} x(n)\, e^{-j \frac{2\pi}{N} nk} \qquad K = 0, \ldots, N - 1 . \qquad (1)$$

The DFT can be interpreted as a set of N bandpass filters, where the Kth filter in the set has a radian center frequency at $w_k = \frac{2\pi k}{N}$. The N filters are evenly spaced throughout the usable frequency range, i.e. between w = 0 and w = 2π. In order to see the characteristic of an individual filter, it is convenient to consider the k = 0 filter, the output of which is

$$X(0) = \sum_{n=0}^{N-1} x(n) . \qquad (2)$$

This is equivalent to a symmetrical nonrecursive filter that has N coefficients, all of which take on a value of unity. The frequency domain characteristic of the Oth filter is shown in Figure 1. The other filters for K = 1, ..., N - 1, have similar response characteristics, although the response peak for the Kth filter is centered at $w_K = 2\pi K/N$.

Several undesirable features of the standard FFT analysis can be observed from the characteristics in Figure 1. Frequency resolution can be improved by increasing the value of N. This increases the total number of filters and causes a corresponding narrowing of the filter mainlobes. However an increase in N does not change the shape of the response, i.e., it is still a sin(WN/2) / sin(W/2) form. Also, an increase in N causes the density of the

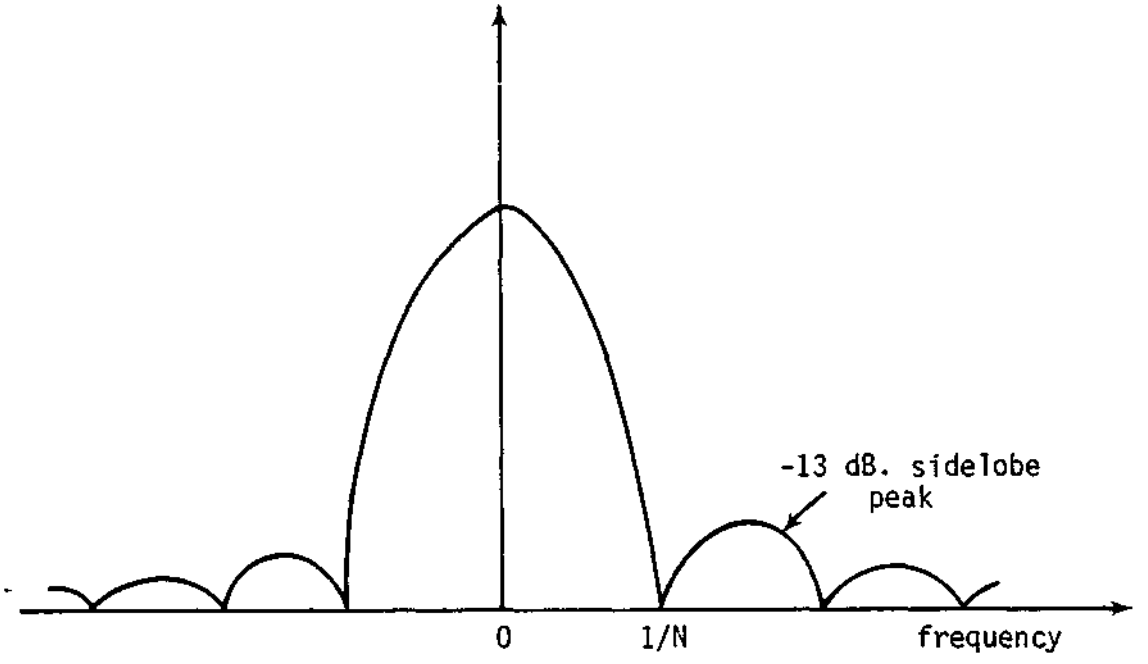

Figure 1. Frequency response of the k = o filter in the Fast Fourier Transform.

filters to be uniformly increased. It is not possible to move the filters closer together in specific regions, while leaving them sparsely separated in others. The FFT program used in this work is a modified FFT routine that is capable of computing spectral samples at noninteger values of the index K. Since a description of this technique is available in Ref. [2], it will not be discussed further here.

Another difficulty occurs in FFT spectral analysis because the discrete Fourier transform processes the data as if the finite length record consisting of N samples is one period of a periodic sequence. Since the sequence is, in fact, not truly periodic with this exact period, the periodic treatment has the effect of introducing artificial discontinuities in the processed sequence. The artificial discontinuities cause artifacts to appear in the spectrum. For example, the discontinuity may introduce high frequency energy that is not really present in the data. This problem is often controlled by applying a window (for example, a Hanning window) to smooth the transitions at the ends of the record, thereby eliminating the discontinuities. The windowing causes some distortion of the original data near the ends of the data record, but this distortion is usually less of a problem than the original discontinuity. However, for the rainfall prediction problem, it was decided that a window could not be used because critical information at the end of the record would be destroyed.

An FFT analysis was used to identify dominant frequency components in the data. Once a given component was identified a digital filter tuned to the spectral peak was used to filter the component from the original data. Instead of filtering with the restricted response characteristics of the FFT filters, various digital filters, each with their own characteristic response curves, were used to remove the dominant frequency components. Both nonrecursive and recursive digital filters were applied for this purpose.

Due to the fact that N can be made arbitrarily large and we can position frequency samples as close together in the frequency domain as we wish, there is a temptation to conclude that frequency resolution can be made arbitrarily small by increasing N. THIS IS NOT TRUE. Frequency resolution is fundamentally limited by the length of the interval over which the data are observed. The frequency resolution is proportional to 2/T, where T is the length of the observation interval (~ 60 years). This means that we cannot resolve frequencies if they are closer together than $\Delta f = 0.333$, no matter how large N is made. The only way to improve this limit would be to collect data over a longer period of time.

2.1 Nonrecursive Digital Filtering

The first digital filters used in this study were zero-phase symmetric nonrecursive filters characterized by the following expression,

$$y(n) = \sum_{k=-\frac{N-1}{2}}^{\frac{N-1}{2}} h(k)\, x(n-k) \qquad (3)$$

(assume N odd)

where N is the filter length, x(n) is the original rainfall record, and y(n) is the filtered data. The frequency response of the filter is given by Equ. (4).

$$H(e^{jw}) = \sum_{k=0}^{N-1} h(k)\, e^{-jwk} \qquad (4)$$

For example, the frequency response of a 4 year filter is shown in Figure 2. This filter, which has a length of 35, has a sharp response at the center frequency and relatively low sidelobes outside the passband. Since the phase response of this type of nonrecursive filter is zero, there can be no

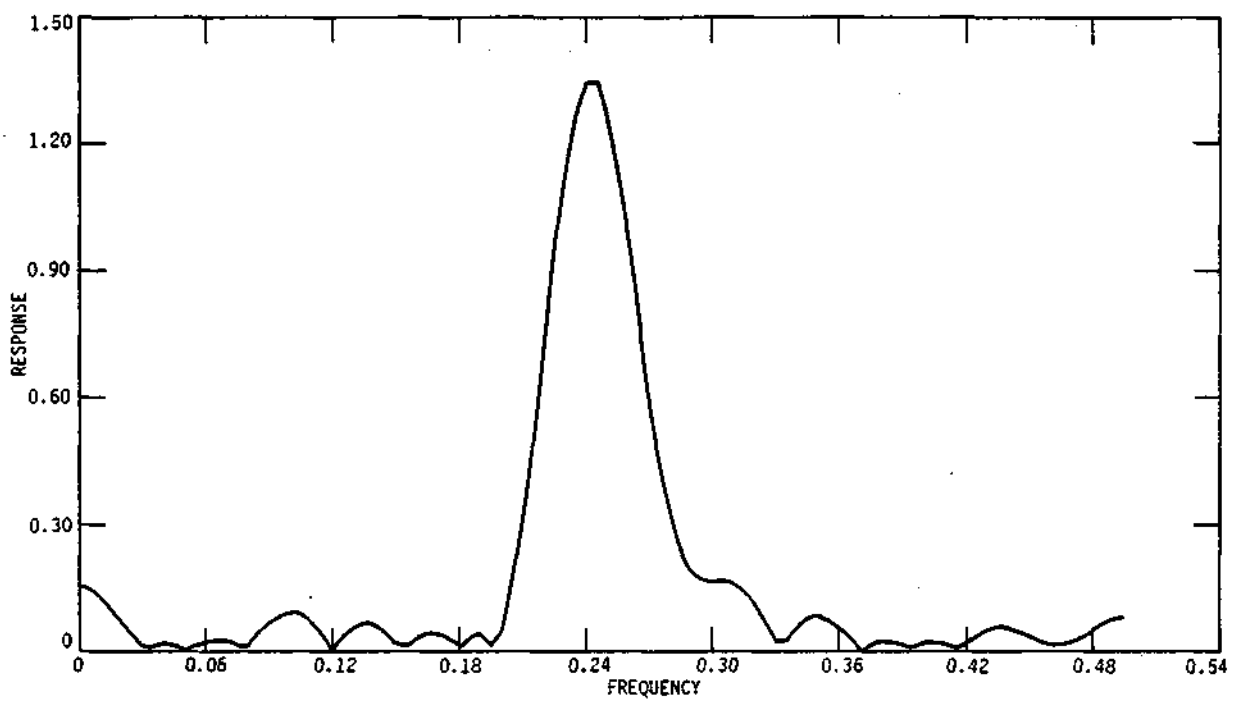

Figure 2. Magnitude of the frequency of a 4-year symmetrical nonrecursive filter.

distortion of the filtered signal due to phase nonlinearity. From the standpoint of this criterion, the nonrescursive symmetrical filters are ideal.

The major difficulty with the symmetrical nonrecursive filter in this application is that, since the filter coefficients extend beyond the data record for the last 17 points of filtered data, there is a transient response introduced just at the critical time when the dominant component should be observed. This transient is caused by the sudden discontinuity at the end of the recorded data. Attempts were made to lessen the severity of the discontinuity by filling the 17 positions of the array immediately beyond the recorded data with the mean value of the previous data. This procedure improved the performance of the nonrecursive filtering significantly, but the essential problem of the filter "falling off" the real data continued to be a problem that raised concern about the validity of the predicted results.

2.2 Recursive Digital Filtering

The transient response at the end of the filtered rainfall sequence was caused by the two-sided weighting of a symmetrical nonrecursive filter. In an attempt to solve this problem, a class of one sided recursive digital filters was tried. Butterworth and Chebychev bandpass filters were designed by means of the bilinear-z transformation for response functions with orders 1, 2, and 4. There were two problems that occurred where the recursive filters were exercized on the data. Since the recursive filters have a nonlinear phase characteristic, it is inevitable that some distortion occurs due to nonlinear phase distortion. The second problem was that on a short data record of only 65-70 samples, the start-up transient did not have sufficient time to decay away. Therefore, at the time the prediction was to occur the periodic behavior of the data was obscured by the transient response. This phenomenon

was particularly noticeable in the 4th order filters. Figures 2 and 3 show the magnitude and response phase of a 4th order bandpass Chebychev filter. The phase response of this filter is approximately linear throughout the passband, although noticeable nonlinear distortion occurs near the band edges. The phase response is not important in the stopband because all components there are attenuated.

A recursive filter is characterized by a difference equation in the time domain.

$$y(n) = \sum_{k=0}^{N} a_k x(n-k) - \sum_{k=1}^{N} b_k y(n-k) \qquad (5)$$

The complex frequency response is given as a ratio of two polynomials.

$$H(e^{jw}) = \frac{\sum_{k=0}^{N} a_k e^{-jnw}}{1 + \sum_{k=1}^{N} b_k e^{-jnw}} \qquad (6)$$

Equation (5) illustrates that only present and past data samples are required to compute y(n), so that recursive filters do not "fall off" the data prior to reaching the end of the recording. The denominator in $H(e^{jw})$ is frequency sensitive, thereby allowing very sharp passband cutoff in the frequency domain. The poles of $H(e^{jw})$ contribute to the ringing behavior that was mentioned earlier.

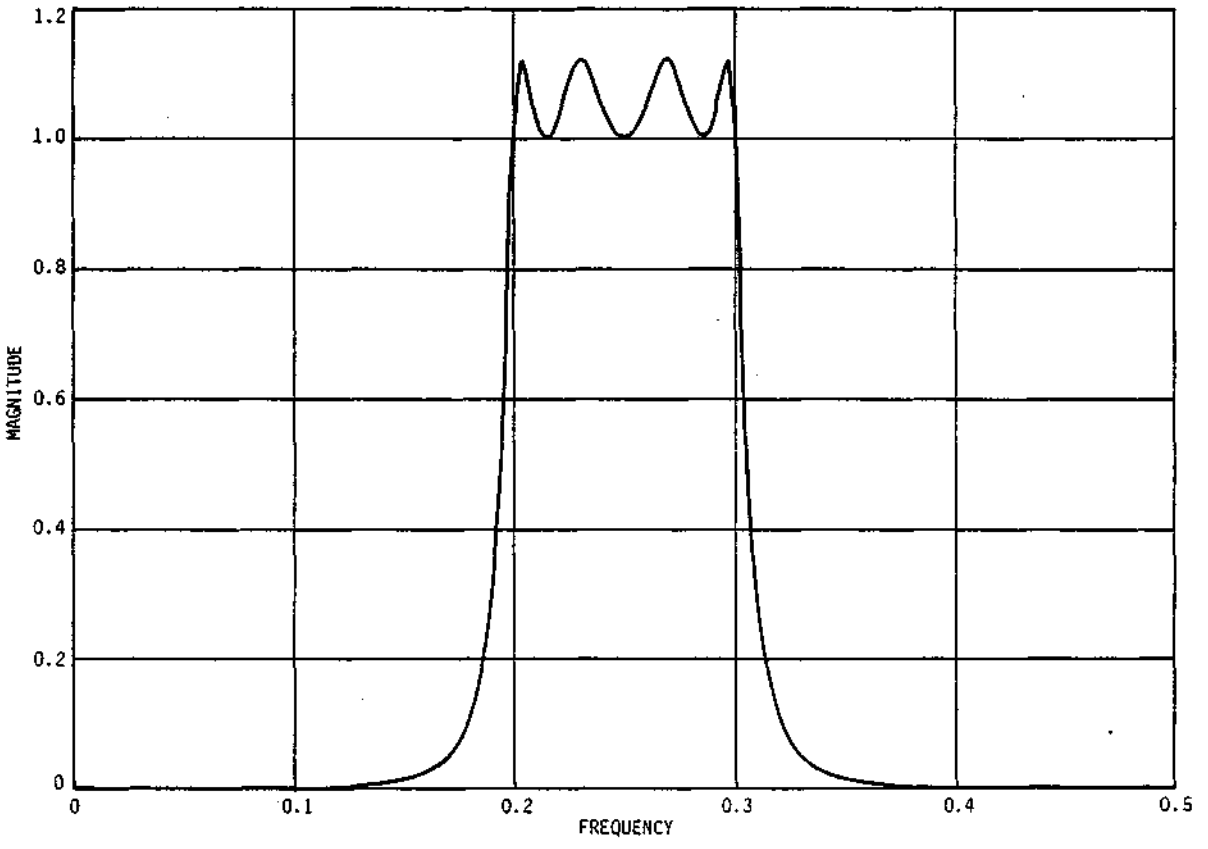

Figure 3. Magnitude of the frequency response of a 4th order Chebychev recursive filter (bandpass).

3. Mathematical Foundations of Digital Filter Design for Rainfall Prediction

3.1 Symmetric Nonrecursive Filters

The purpose of this section is to provide a mathematical basis to complement the document entitled "Notes on the Design of Numerical Filters," by E. G. Bowen, unpublished. The filters are characterized as lowpass (smoothing) filters or bandpass filters. The lowpass filter is a special case of a bandpass filter, where the passband is centered at the origin. For purposes of discussion it is convenient to discuss these separately.

3.1.1 Lowpass Filter

The simplest symmetric nonrecursive lowpass filter is defined by the following weight factors (coefficients).

$$h_R(n) = \begin{cases} 1.0 & -(N-1)/2 \leq n \leq (N-1)/2 \\ 0.0 & \text{elsewhere.} \end{cases} \qquad (7)$$

If u(n) is the recorded data, then the filtered sequence is described by Equ. (8)

$$y(n) = \sum_{K=-\frac{(N-1)}{2}}^{\frac{(N+1)}{2}} h_R(K)\, u(n-K) \qquad (8)$$

This filter is called a rectangular window, as shown in Figure 4. The corresponding frequency is given by Equ. (9) and shown in Figure 5.

$$H_R(e^{j2\pi f}) = \sum_{k=-\infty}^{\infty} h_R(K)e^{-j2\pi kF} = \frac{\text{Sin}\ (2\pi fN/2)}{\text{Sin}\ (2\pi F/2)} \qquad (9)$$

Since $H_R(f)$ is real, the phase is either 0 or Π, depending on the sign. The filter passes frequency components in the band $-1/N \leq f \leq 1/N$ and approximately rejects other components. In the passband the phase is strictly zero because the frequency response is real and positive.

The rectangular symmetric non-recursive filter is only approximately lowpass in performance. There is leakage of frequency components $|f| > 1/N$ due to the nonzero sidelobes on the frequency response. Also, the passband frequencies near the band edge will be attenuated severely. The passband edge at 1/N can be made very narrow if the filter length N is made large. The non-recursive filters used in the rainfall prediction study used N = 35 so the response was very sharp and narrow.

When the sidelobes of the rectangular filter are unacceptable, a non-rectangular window can be applied to reduce the peak sidelobes. An example is the Hanning window defined by Equ. (10) and shown in Figure 6.

$$h_H(n) = \begin{cases} \frac{1}{2} + \frac{1}{2}\ \cos\frac{2\pi n}{N}\ , & -\frac{(N-1)}{2} \leq n \leq \frac{(N-1)}{2} \\ 0\ , \text{ elsewhere} \end{cases} \qquad (10)$$

The frequency response of the Hanning windowed filter is related to the rectangular response according to Equ. (11). The response characteristic is shown in Figure 7.

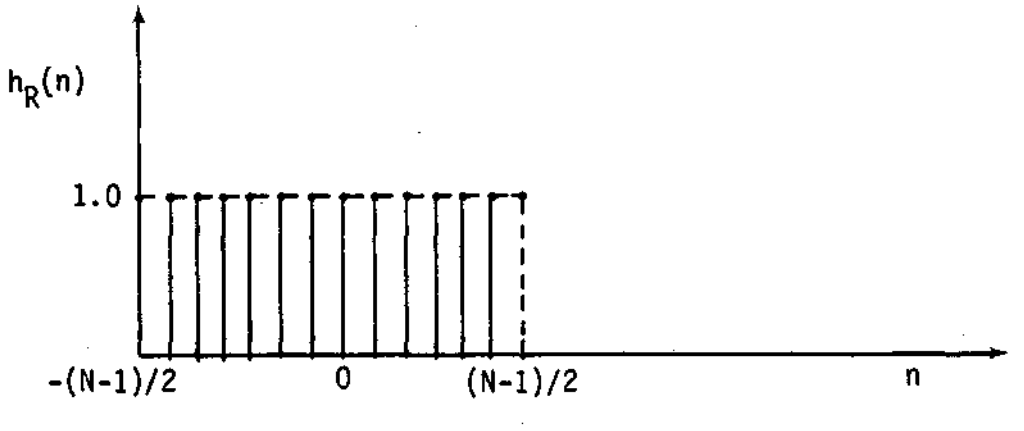

Figure 4. Unit pulse of a rectangular lowpass filter.

$$H_H(e^{j2\pi f}) = \frac{1}{2} H_R(e^{j2\pi f}) + \frac{1}{2} H_R(e^{j2\pi(f-1/N)}) + \frac{1}{2} H_R(e^{j2\pi(f+1/N)}) \quad (11)$$

The Hanning spectral sidelobes are significantly lower than sidelobes for the rectangular window (reduced from 0.223 to 0.0178 relative to a unity mainlobe peak), although the width of the mainlobe is doubled. Therefore, the Hanning filter does not allow much leakage from higher frequencies, but it is not as sharp and selective in the passband as the rectangular filter.

$$h_{RB}(n) = h_R(n) \cos(2\pi f_o n) \quad (12)$$

3.1.2 Bandpass Filter

The passband center frequency can be moved to a nonzero value of f_o by modulating the rectangular window by a cosine waveform of frequency f_o.

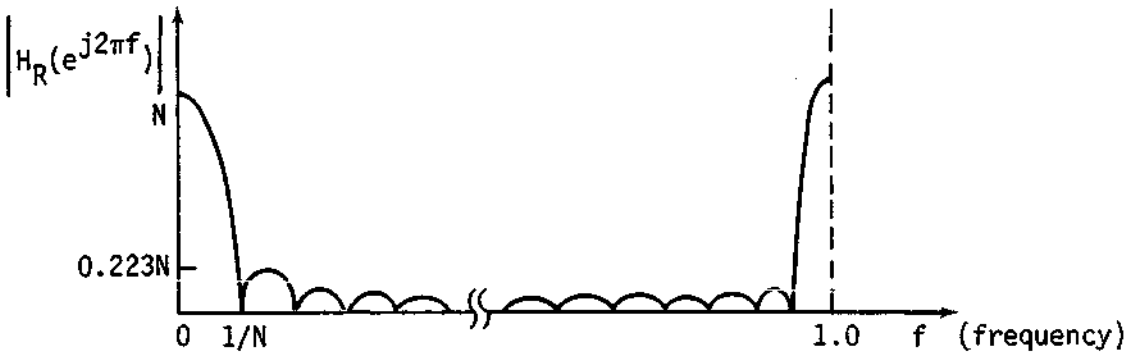

Figure 5. Frequency response of the rectangular lowpass filter.

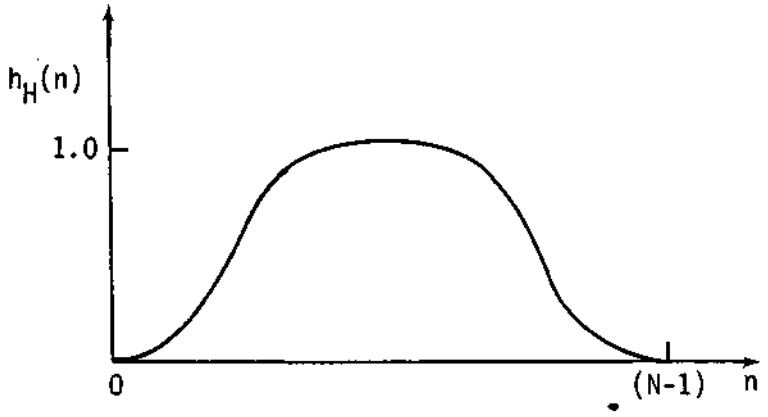

Figure 6. Hanning window.

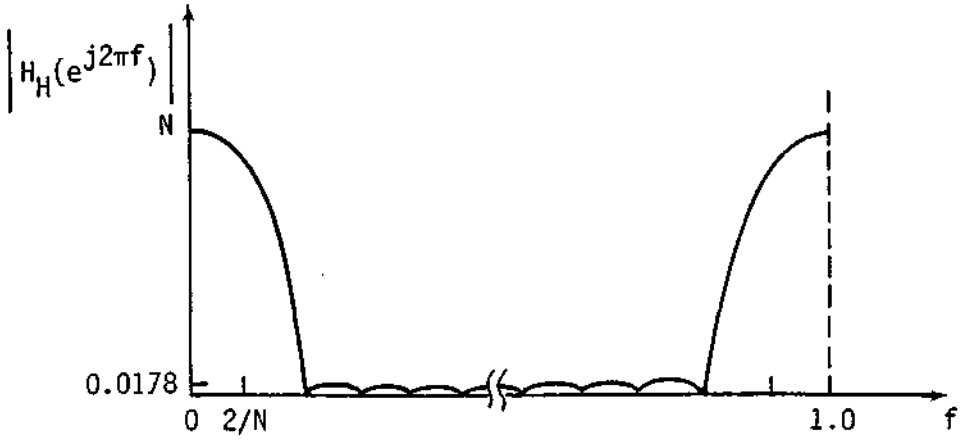

Figure 7. Magnitude of the Hanning spectrum.

The resulting frequency response is

$$H(e^{j2\pi f}) = \frac{1}{2} H_R(e^{j2\pi(f + 1/N_o)}) + \frac{1}{2} H_R(e^{j2\pi(f-1/N_o)}). \qquad (13)$$

The unit pulse response and the corresponding frequency response are shown in Figures 8 and 9, respectively. The behavior of the rectangular bandpass filter is similar to the lowpass case, except that the passband is positioned so that frequencies in the ranges $-f_o - 1/N \leq f \leq -f_o + 1/N$ and $f_o - 1/N \leq f \leq f_o + 1/N$ are passed. The filter coefficients are found by sampling the windowed cosine curve at N discrete integer values.

A Hanning bandpass filter is produced by modulating a Hanning window that was described by the lowpass filter. The Hanning response curve is translated to be centered about -fo and fo. The filter coefficients are given by

$$h(n) = h_R(n) \{1/2 + 1/2 \cos 2\pi fn\} \cos(2\pi f_o n). \qquad (14)$$

Other windows, such as the Hamming, Bartlett, Blackman, and Kaiser windows, can be used in a similar way [3].

3.2 Recursive Filters

Recursive filters used in this study were designed by transforming 1st, 2nd, and 4th order Butterworth and Chebychev analog prototypes into digital bandpass filters. The general design procedure is outlined below:

A. The transfer function H(s) of the analog prototype is obtained from filter design tables.

B. H(s) is transformed into a digital filter by the bilinear-z transformation. A discrete-time lowpass prototype $\hat{H}_{LP}(z)$ results that retains the characteristics of the analog protype.

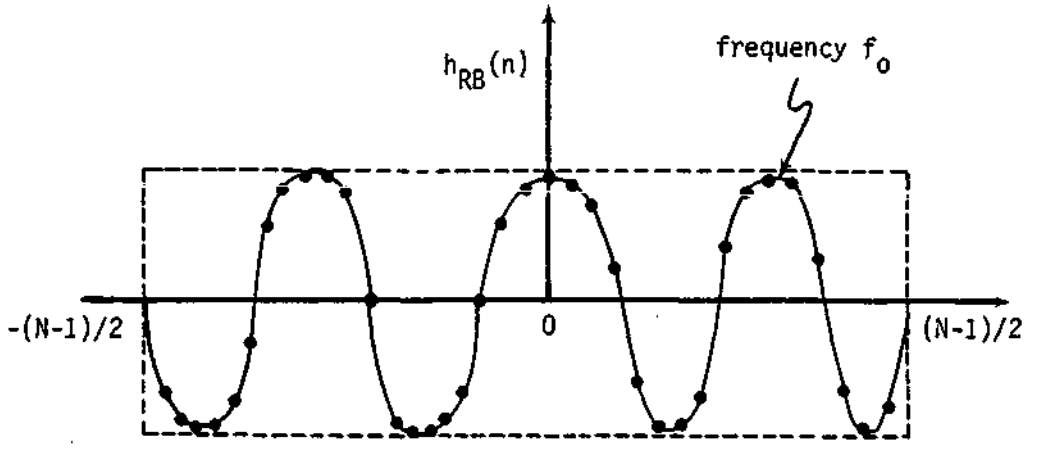

Figure 8. Unit pulse response for a rectangular bandpass filter.

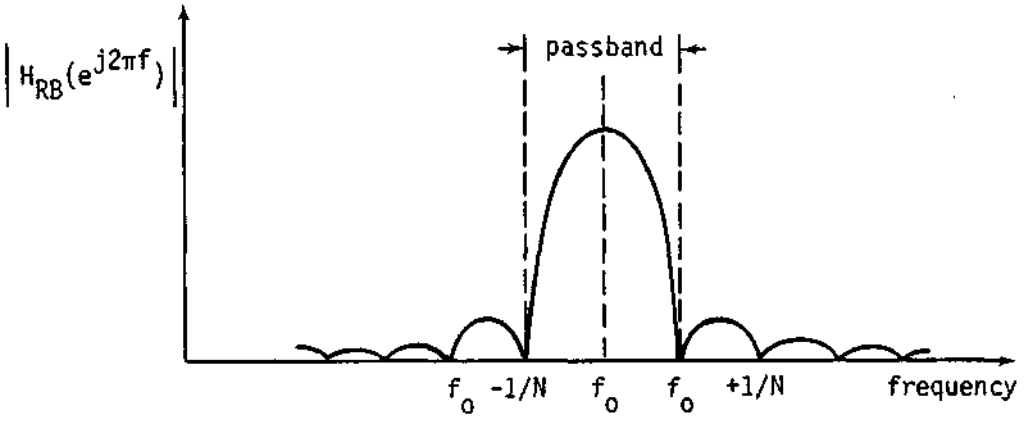

Figure 9. Magnitude response for a rectangular bandpass filter.

C. $\hat{H}_{LP}(z)$ is transformed into a discrete-time bandpass filter by applying a lowpass-to-bandpass spectral transformation. This results in a bandpass filter characterized by $\hat{H}_{BP}(z)$.

D. $\hat{H}_{BP}(z)$ is then transformed back to the (discrete) time domain to produce the constant coefficient difference equation that characterizes the filter.

A number of computer programs were written to compute the coefficients of the final bandpass filter and to evaluate and plot the frequency response. The band edges and center frequency are specified as the inputs for a given filter order. (Separate programs were written for 1st, 2nd, and 4th order filters.) Since these programs are available at the Illinois State Water Survey, the details of the programs will not be repeated here. However, the design procedure implemented in the programs will now be illustrated for the 4th order Chebychev class of filters.

The frequency response of the analog lowpass prototype is expressed in the following form,

$$H(s) = \frac{b_o}{s^4 + b_3 s^3 + b_2 s^2 + b_1 s + b_o}$$

where the b_i's are found in filter design handbooks [5]. For example, a 4th order Chebychev prototype with a 1dB passband ripple and a ripple cutoff frequency $w_c = 1$ is defined by the following b_i's [5].

$b_0 = 0.2756276$

$b_1 = 0.7426194$

$b_2 = 1.4539248$

$b_3 = 0.9528114$

The bilinear - z transformation is then used to convert H(s) into $\hat{H}_{LP}(z)$.

$$\hat{H}_{LP}(z) = H(s) \Big|_{S = \frac{2}{T}\left[\frac{1 - z^{-1}}{1 + z^{-1}}\right]} \tag{16}$$

For annual rainfall prediction, a sampling period of T = 1 year was used because only the July average rainfall was used. Next, the lowpass-to-bandpass spectral transformation was applied to transform $\hat{H}_{LP}(z)$ into $\hat{H}_{BP}(\hat{z})$

$$\hat{H}_{BP}(\hat{z}) = \hat{H}_{LP}(z) \Big|_{z = T[\hat{z}]}$$

$$\text{where } T[\hat{z}] = \left[\frac{\hat{z}^{-2} - \frac{2\alpha K}{K+1}\hat{z}^{-1} + \frac{K-1}{K+1}}{\frac{(K-1)}{(K+1)}\hat{z}^{-2} - \frac{2\alpha K}{K+1}\hat{z}^{-1} + 1}\right]$$

$$\alpha = \left[\frac{\cos\left(\frac{W2 + W1}{2}\right)}{\cos\left(\frac{W2 - W1}{2}\right)}\right]$$

$$\text{and } K = \cot\left[\frac{W2 - W1}{2}\right] \tan\frac{\theta_2}{2} .$$

In the above expressions, W1 and W2 are the lower and upper ripple bandedges, θ_p is ripple cutoff frequency of $\hat{H}_{LP}(z)$, which is related to the cutoff frequency Ω_a of the analog prototype by

$$\Omega_a = \frac{2}{T} \tan \frac{\theta_p}{2} . \tag{18}$$

The algebraic manipulations for the above design are simplified if the analog prototype is frequency scaled so that Ω_a is given by

$$\Omega_a \equiv \frac{2}{T} \tan \left\{ \frac{W2 - W1}{2} \right\} . \tag{19}$$

Then $\theta_p = W2 - W1$, $K = 1$ and $[\hat{z}]$ s considerably simplified.

The filter to be designed is described to the programs by specifying the b_i's and the lower and upper bandedges W1 and W2, respectively. The program generates the filter coefficients (C_i's and d_i's) that specify the frequency response and the characteristic difference equation.

$$\hat{H}_{BP}(\hat{z}) = \left[\frac{C_o + C_1 \hat{z}^{-1} + C_2 \hat{z}^{-2}}{1 + d_1 \hat{z}^{-1} + d_2 \hat{z}^{-2}} \frac{+ C_3 \hat{z}^{-3} + C_4 \hat{z}^{-4}}{+ d_3 \hat{z}^{-3} + d_4 \hat{z}^{-4}} \right] \tag{20}$$

$$y(n) = \sum_{k=0}^{4} C_k x((n-k)T) - \sum_{k=1}^{4} d_k y ((n-k) T) \tag{21}$$

The characteristic equation (21) is then used to recursively filter the data sequence x(nT).

Examples of recursive filter responses designed by this procedure are shown in Figures 3, 10, 11, 12, and 13. The phase response in all cases is nonlinear, but follows an approximate linear relationship in the passband similar to the 4th order Chebychev filter of Figure 14.

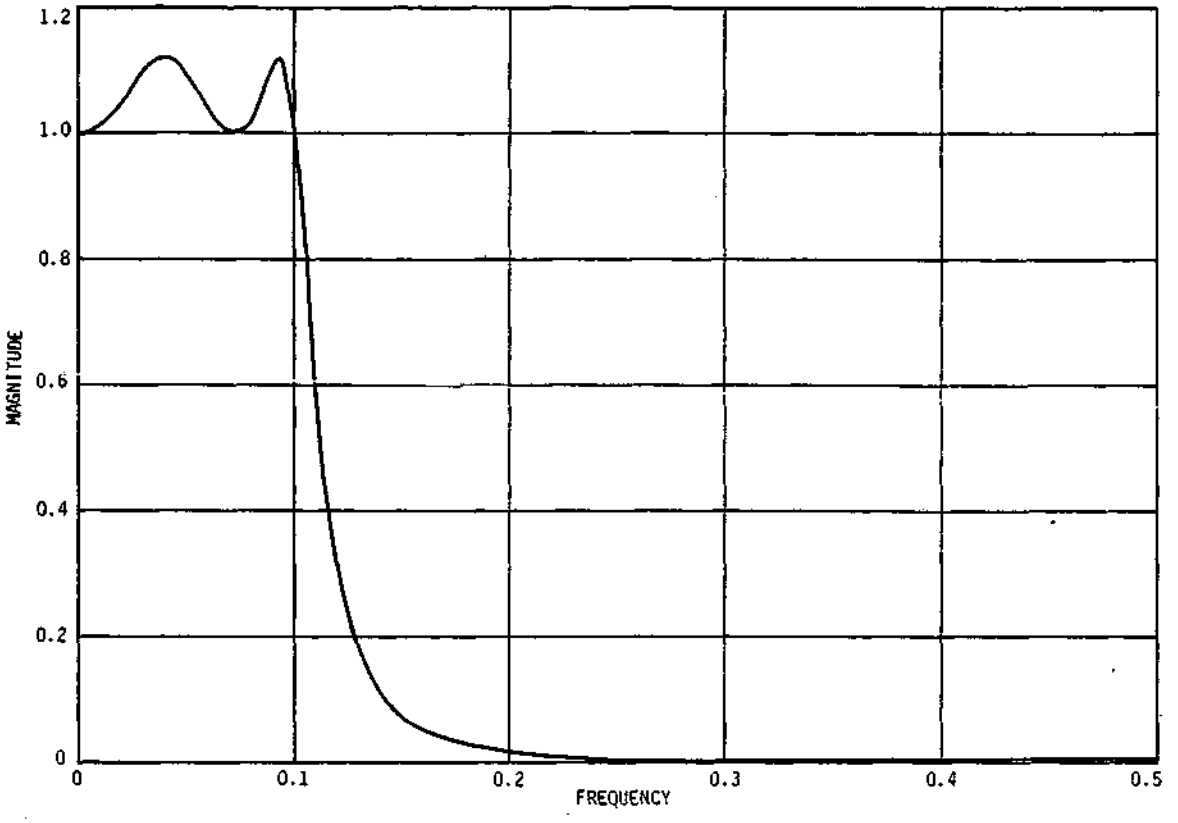

Figure 10. Fourth order lowpass recursive Chebychev magnitude response.

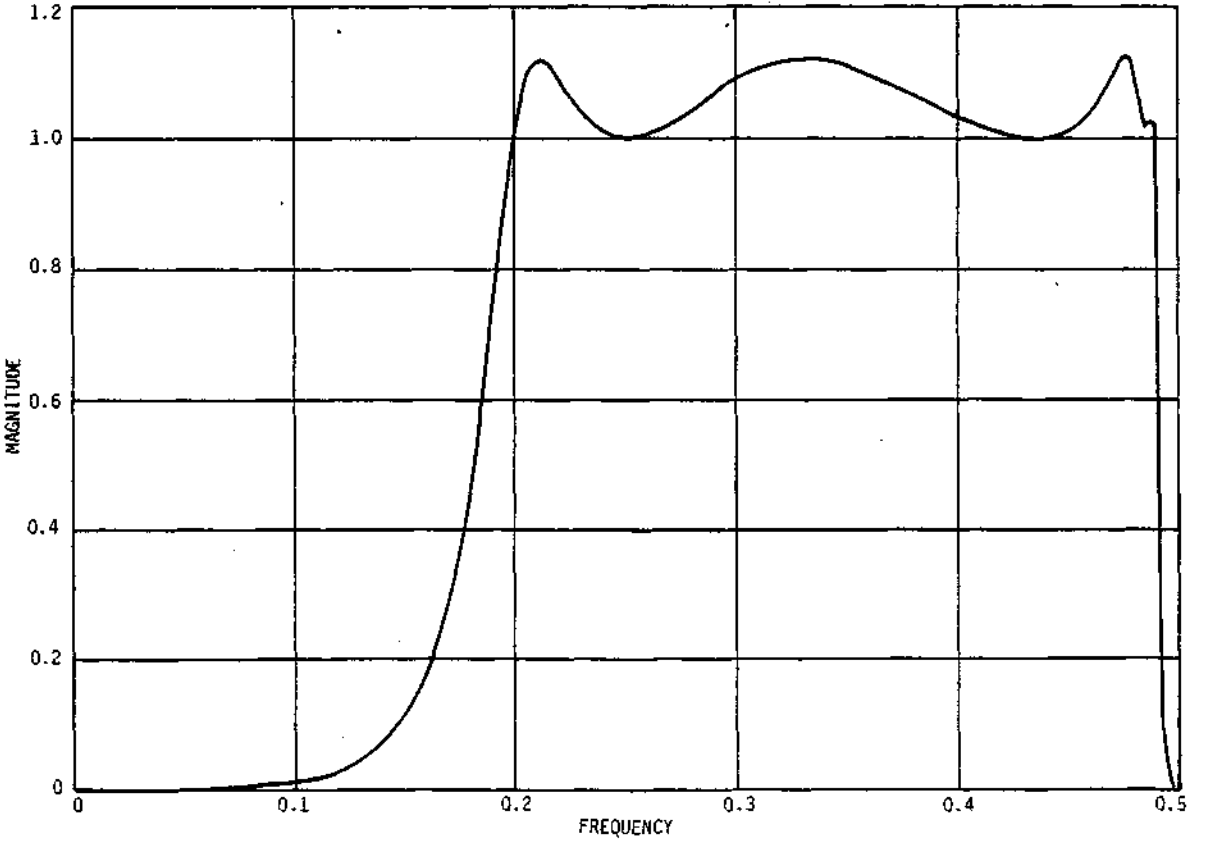

Figure 11. Response of a 4th order Chebychev highpass filter.

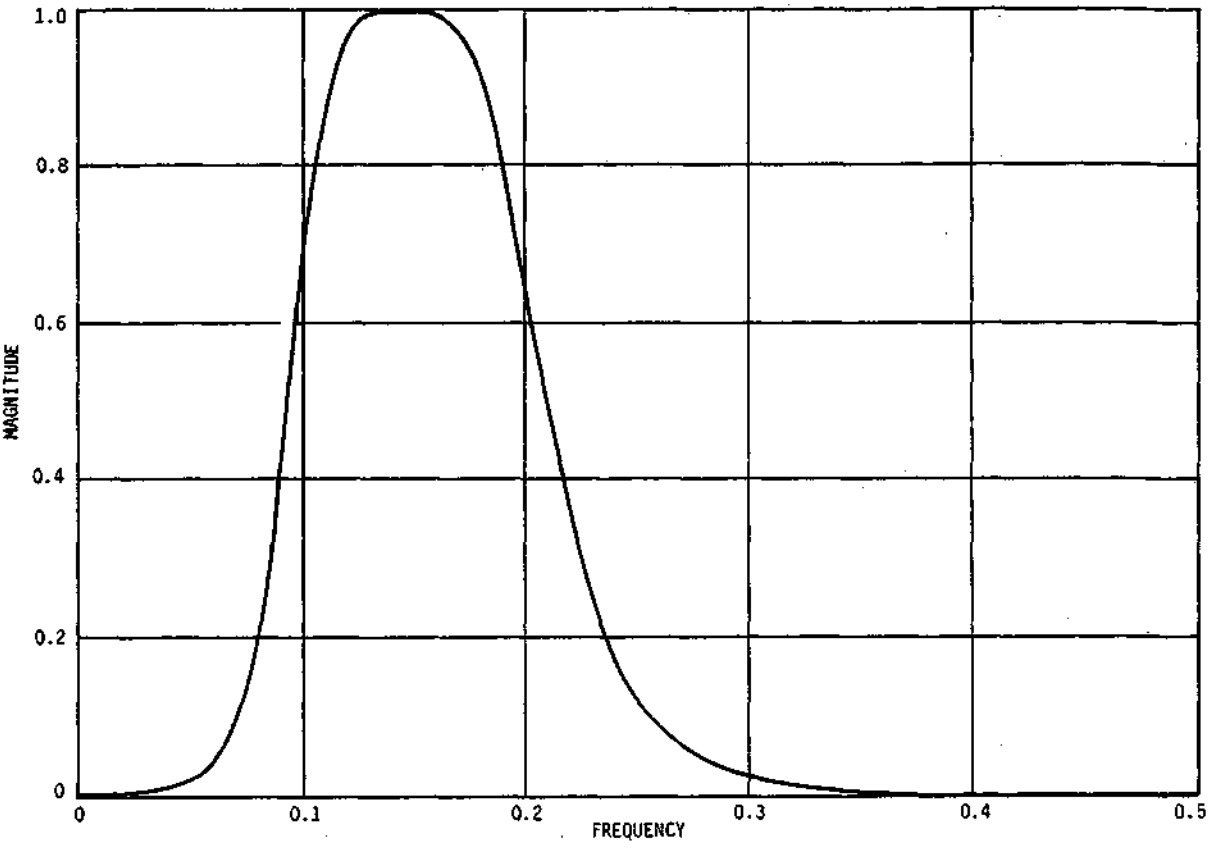

Figure 12. Response of a 1st order Chebychev bandpass filter.

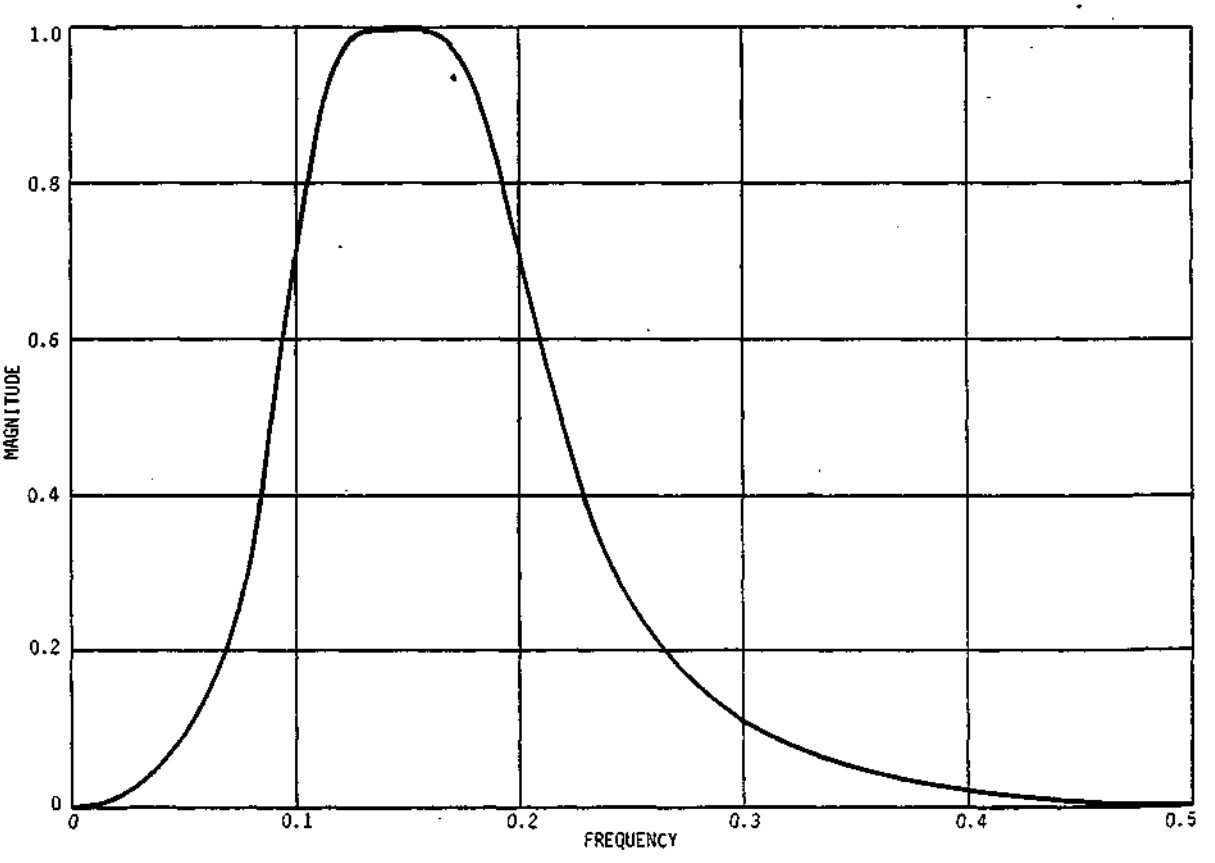

Figure 13. Response of a 2nd order Butterworth bandpass filter.

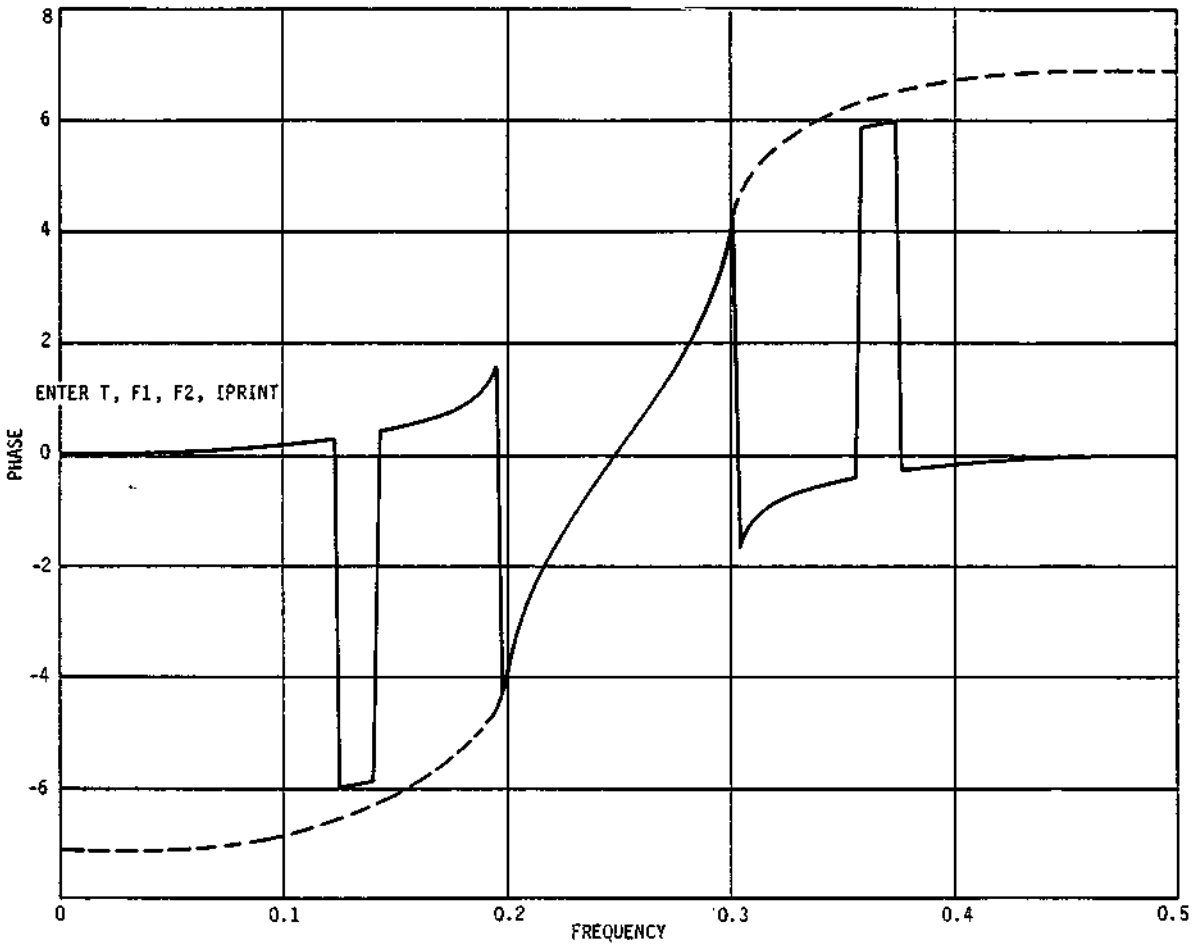

Figure 14. Phase response for a 4th order Chebychev recursive filter (band-pass).

4. Summary of Results

The 1st, 2nd, and 4th order Butterworth and Chebychev recursive filters described in Section 3 were used to replace the original symmetrical nonrecursive for the extraction of important sinusoidal components in the annual rainfall data record. Figure 15 shows the output of a 15 year non-recursive filter with 35 coefficients when the input signal consists of a 15 year and 8 year component. Figure 16 shows the output of a 4th order Chebychev recursive filter with a bandwidth of $f_2 - f_1 = 0.01$. Figure 16 illustrates the start-up transient in the nonrecursive filter. After 60 iterations the recursive filter has not yet reached steady state. Figures 17 and 18 show the recursive filter with the bandwidth widened to .015 and 0.05, respectively. The wider bandwidth in Figure 17 causes the filter to approximately reach steady state. The excessively wide bandwidth in Figure 18 is allowing a considerable amount of the 8 year component through, which is apparent in the distorted sinusoidal output. Figures 19 and 20 show the response of an 8 year recursive filter with bandwidths of 0.015 and 0.05, respectively.

Figure 21 shows the actual recorded data (solid line) and the reconstructed data (dashed line) using nonrecursive symmetric filters, L (long-wave ≥ 30.0) B (10.8) 15.0, 8.0, 4.0, and 2.5-year filters. Figure 22 shows a similar prediction in which recursive Chebychev filters were used for 12.7 years and 16.0 years (each with a bandwidth of 0.05). The prediction was made by combining these two filters with the L, 8.0, 4.0, and 2.5 yr. filters. The predicted results in these two cases are very similar, and neither shows a very faithful representation of the correct data. It is encouraging, however, that many of the general trends in the data (peaks and valleys) have been reproduced at approximately the correct times.

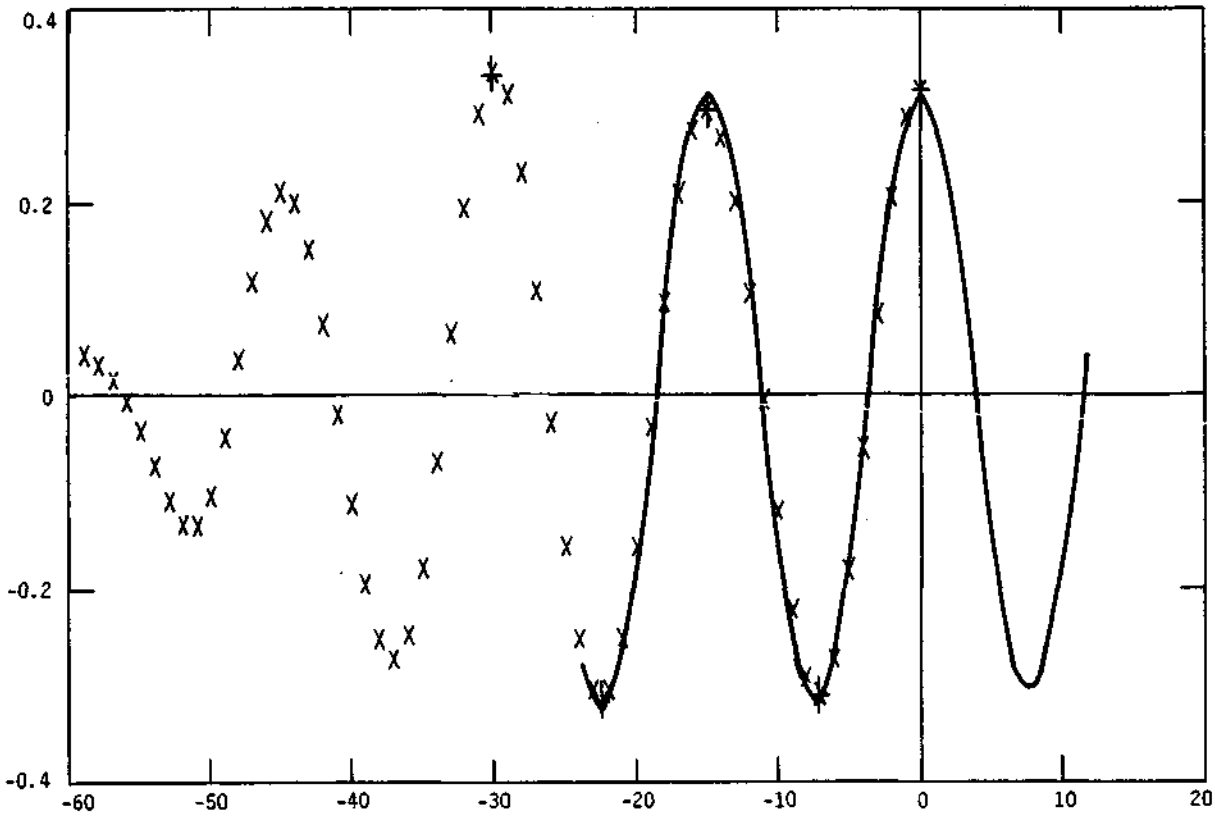

Figure 15. Output of the 15-year filter with an 8-year and 15-year component in the input (nonrecursive).

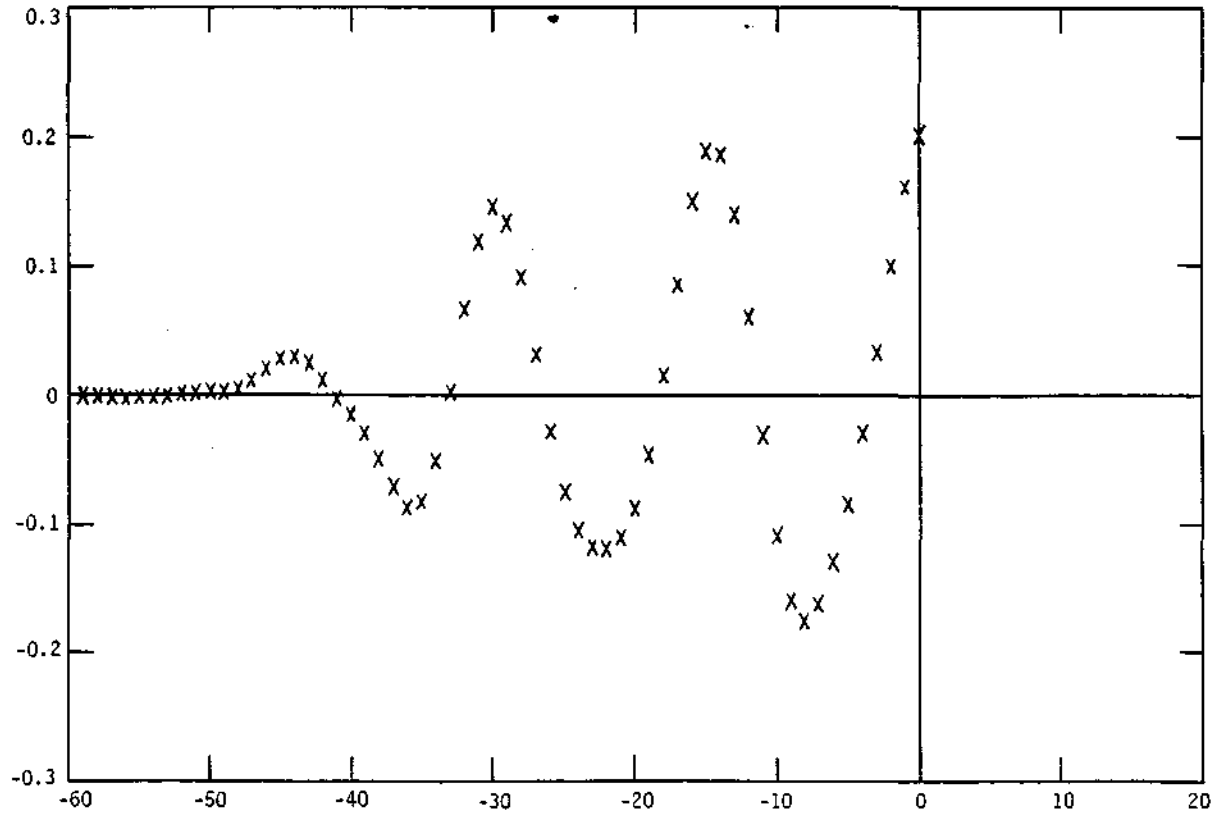

Figure 16. Output of the 15-year recursive filter (w2 - w1 = 0.01) for 8 and 15-year input components.

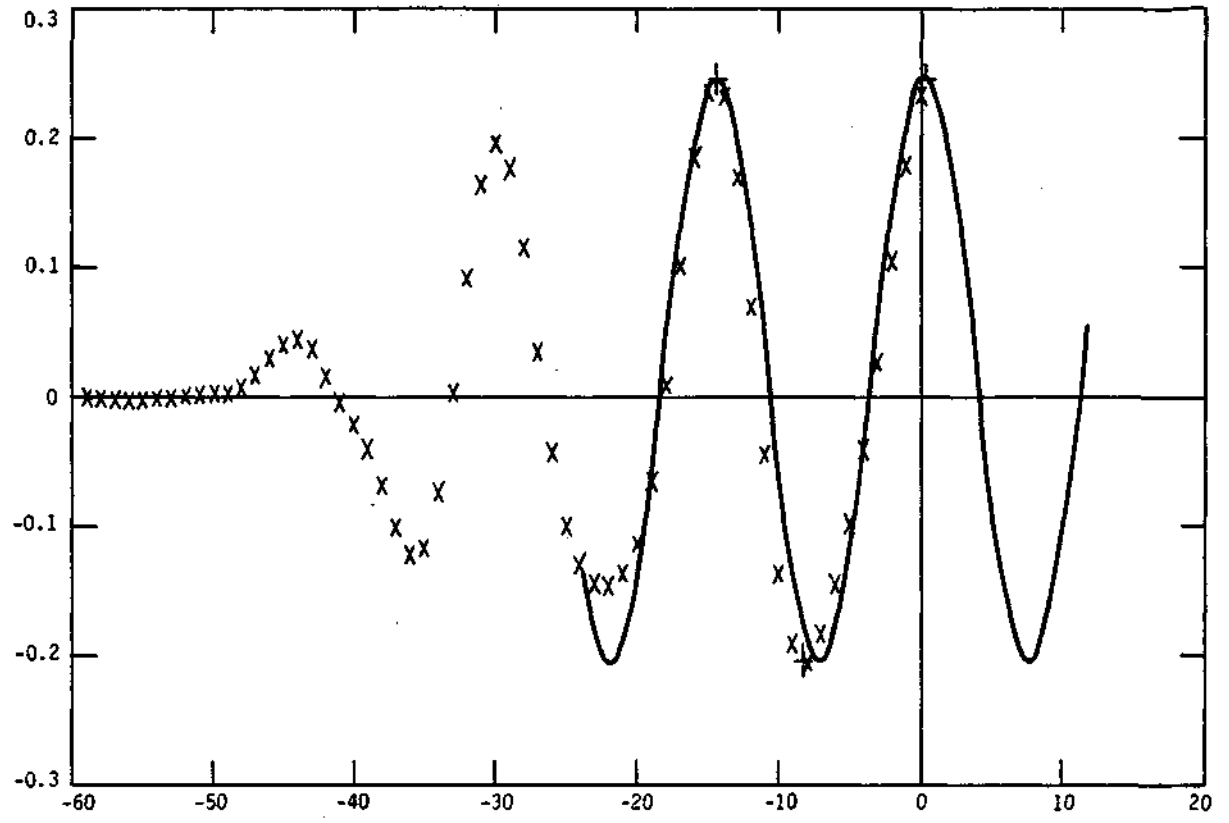

Figure 17. Output of the 15-year recursive filter (w2 - w1 = 0.015) for 8 and 15-year input components.

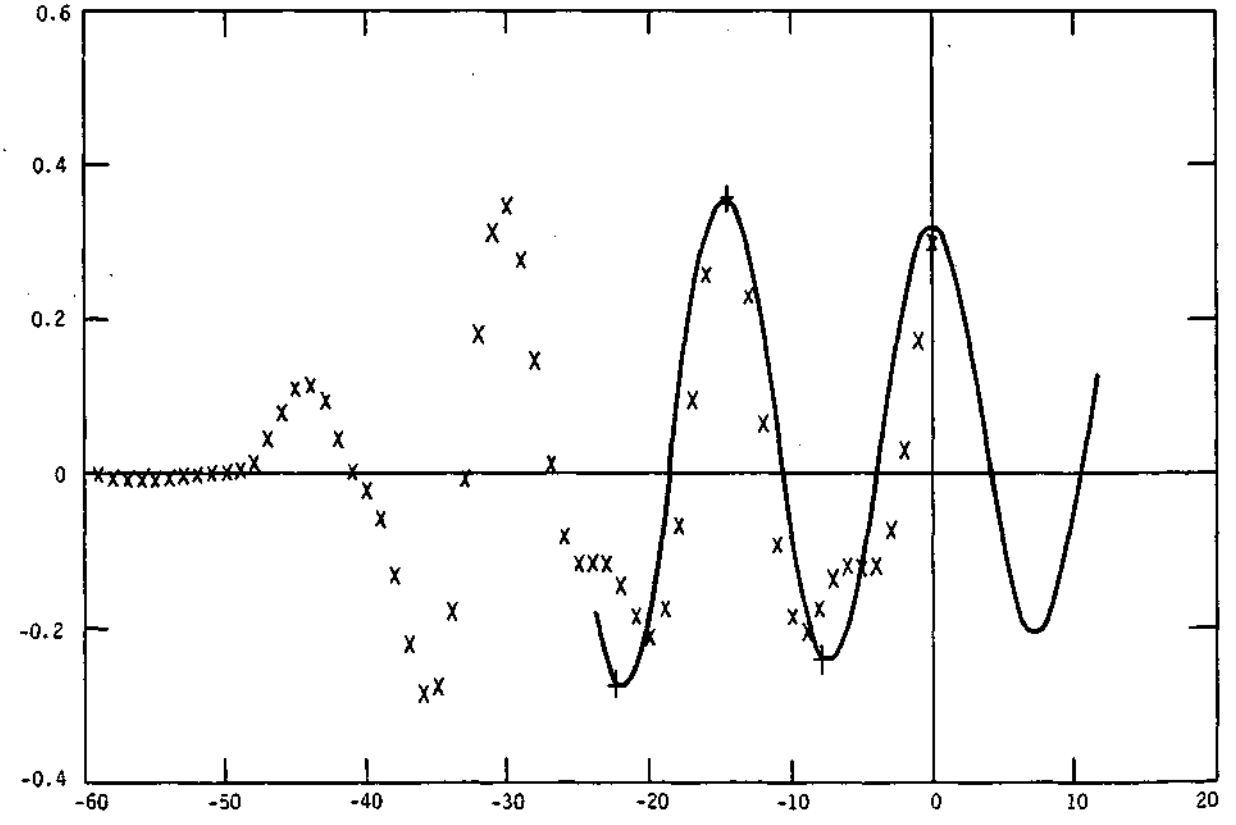

Figure 18. Output of the 15-year recursive filter (w2 - w1 = 0.05) with 8 and 15-year input components.

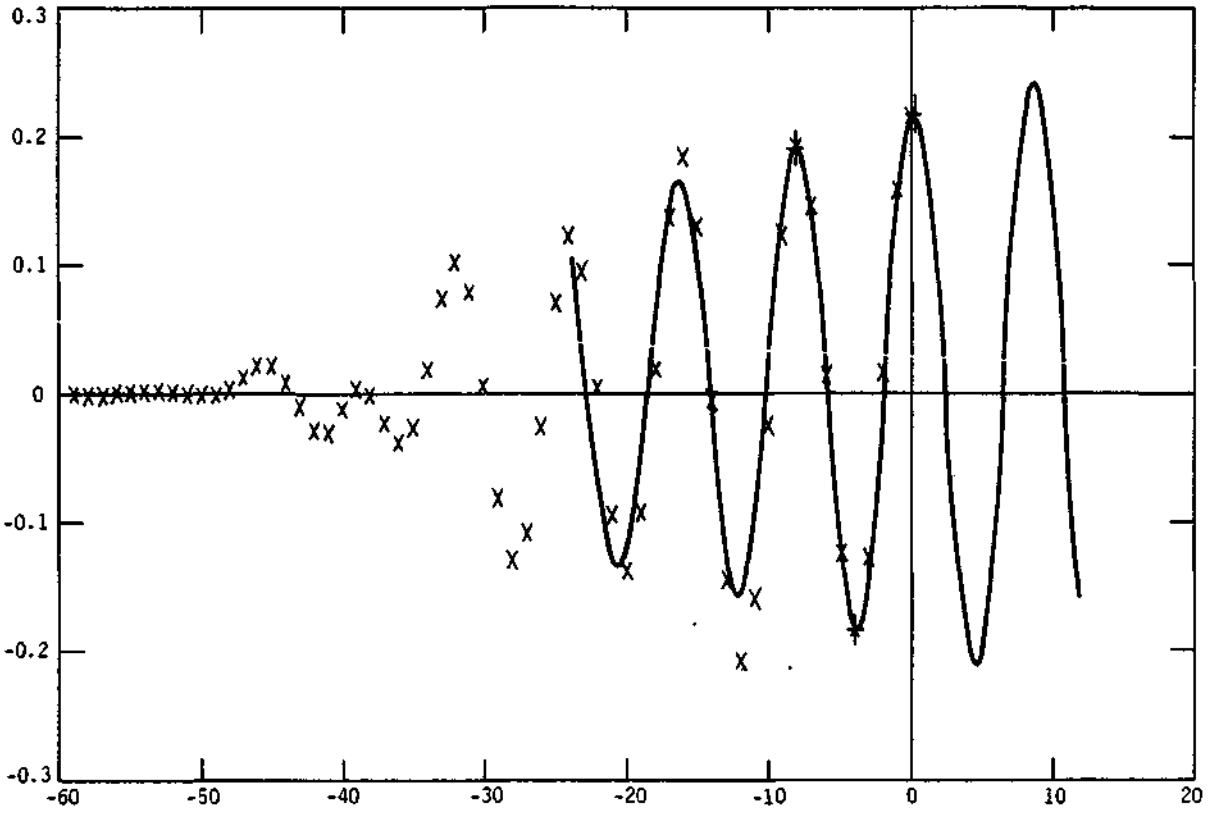

Figure 19. Output of the 8-year filter with 8 and 15-year input components.

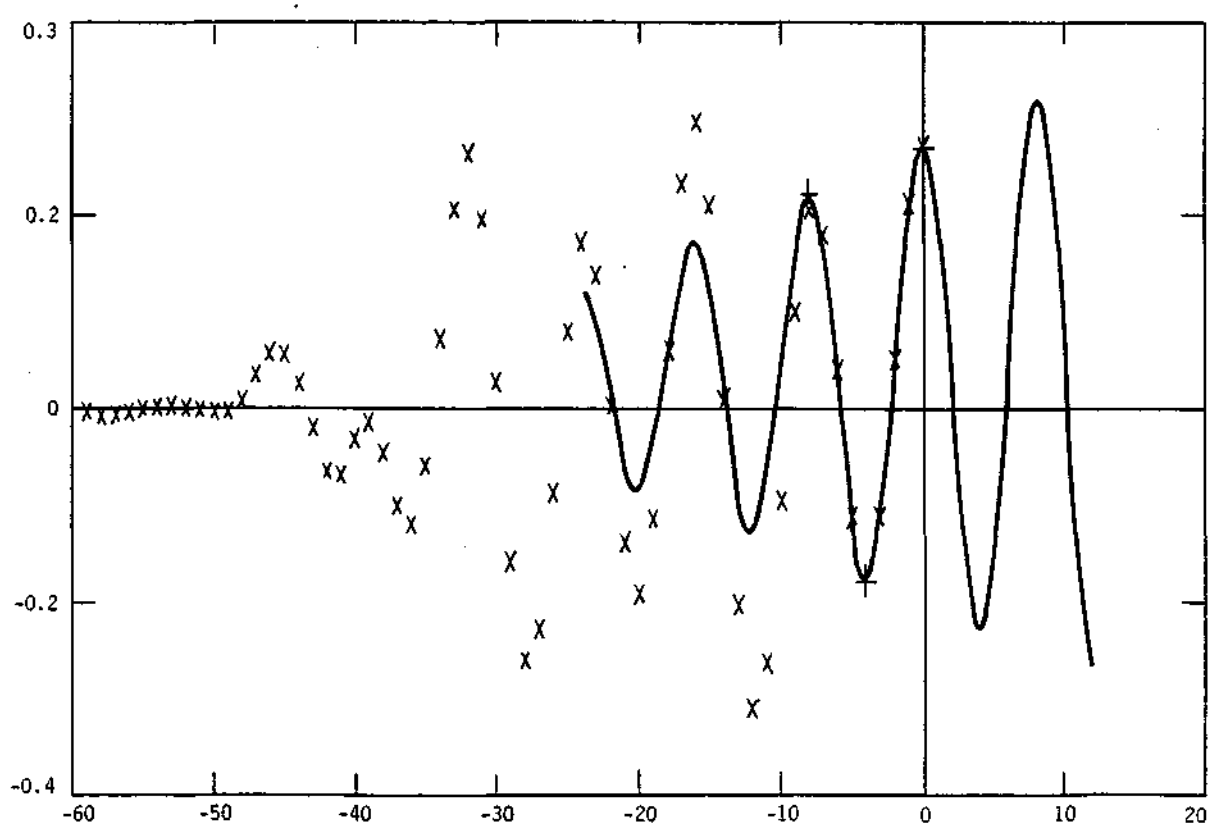

Figure 20. Output of the 8-year filter (recursive with w2 - w1 = 0.05) for 8 and 15-year input components.

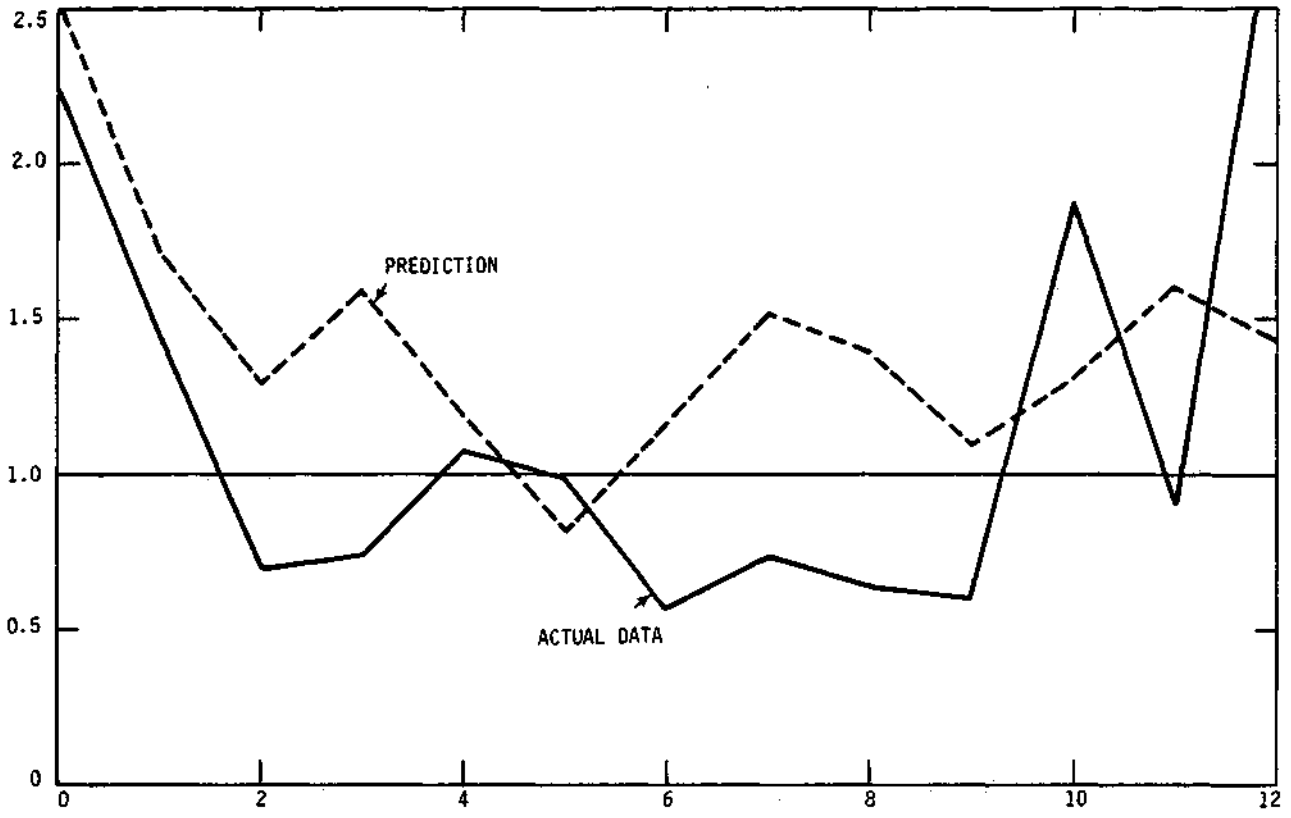

Figure 21. Actual recorded data and prediction using 30.0, 10.8, 15.0, 8.0, 4.0, and 2.5-year nonrecursive symmetrical filters.

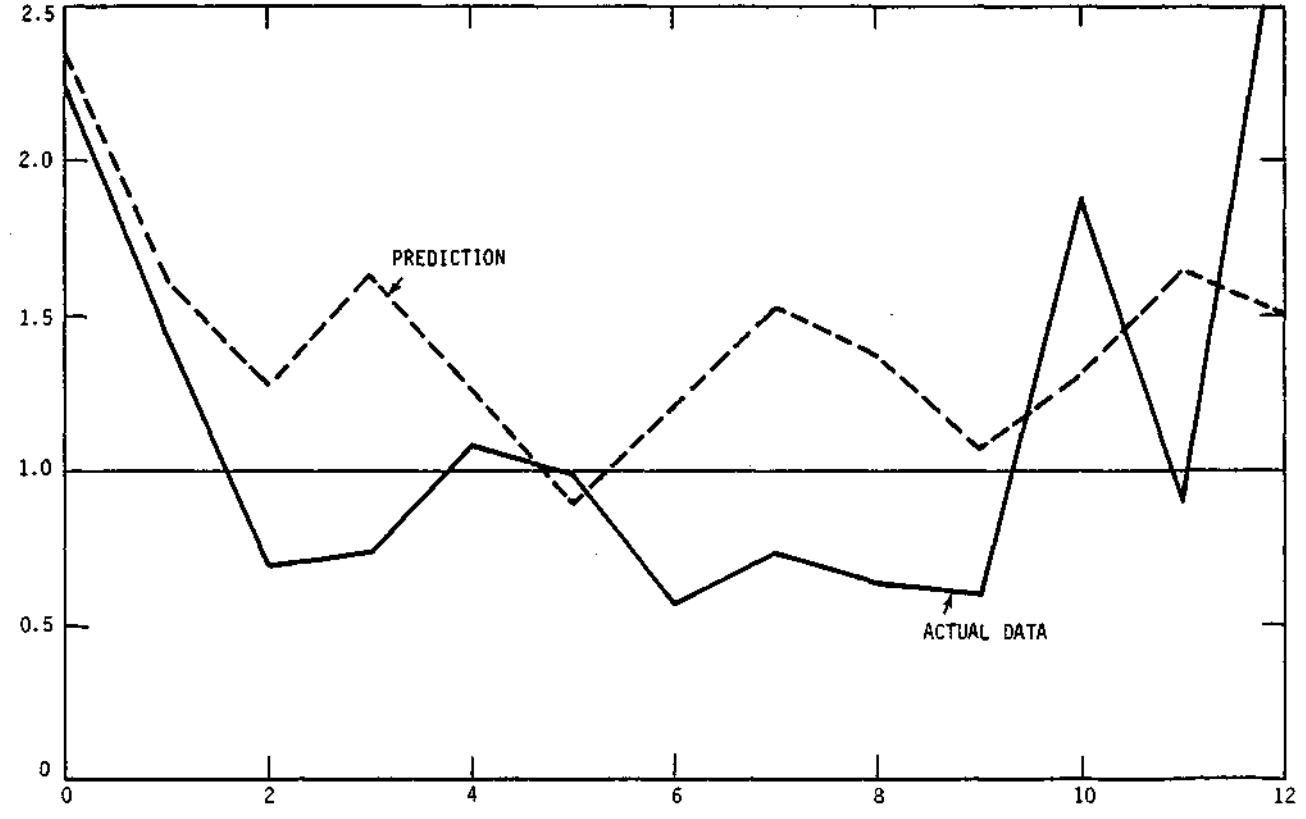

Figure 22. Actual recorded data and prediction using 12.7 and 15.0-year recursive Chebychev filters.

At the time this work was completed, there was no definitive improvement in the predictions when the recursive filters were used. It appears that any improved performance that results because the recursive filter does not "fall off" the data record is masked by transient response and nonlinear phase distortion. However, since there are many types of recursive filters that were not evaluated, it should be concluded that at this stage in the investigation the results are inconclusive. Certainly it could be justified to continue with the nonrecursive filters, since up to this time, they appear to perform as well as the others.

REFERENCES

[1] G. E. P. Box and G. M. Jenkins, Time Series Analysis Forecasting and Control, Holden-Day, Inc., San Francisco, 1970.

[2] P. T. Schickedanz and E. G. Bowen, "The Computation of Climatological Power Spectra," J.A.M. - 5315, April 1978.

[3] A. V. Oppenheim and R. W. Schafer, Digital Signal Processing, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1975.

[4] R. W. Hamming, Digital Filters, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1977.

[5] L. Weinberg, Network Analysis and Synthesis, McGraw-Hill Book Company, Inc., NY, 1962.