FLUCTUATIONS OF WET AND DRY YEARS, AN ANALYSIS BY VARIANCE SPECTRUM

by
Vujica Yevjevich

August 1977

HYDROLOGY PAPERS
COLORADO STATE UNIVERSITY
Fort Collins, Colorado

94

FLUCTUATIONS OF WET AND DRY YEARS, AN ANALYSIS BY VARIANCE SPECTRUM

by
Vujica Yevjevich

HYDROLOGY PAPERS
COLORADO STATE UNIVERSITY
FORT COLLINS, COLORADO 80523

August, 1977

No. 94

TABLE OF CONTENTS

ACKNOWLEDGMENT

The study of fluctuations of wet and dry years has been supported by various US National Science Foundation grants in the past. This study was supported by the current grant "Stochastic Water Resources Processes," NSF Grant ENG 74-17396. This support is gratefully acknowledged.

ABSTRACT

The spectral analysis is applied to annual series of precipitation and runoff. The precipitation series are divided in homogeneous P_1-precipitation (no significant changes occured in gauge positions, or no significant inconsistency in data), and non-homogeneous P_2-precipitation (changes occurred in gauge positions, with systematic errors or inconsistency). Runoff series are either the observed values as the Q_1-series, or they are reduced to the effective precipitation as the Q_2-series (precipitation minus evaporation, determined by changing the annual values for the annual difference in water stored in river basins). Data of annual precipitation and annual runoff of a large number of gauging stations in the United States are used, in the study, dividing them into six areas.

The techniques of spectral analysis used in the study are described in a condensed form. Average spectra are estimated for each of the four variables (P_1, P_2, Q_1, Q_2) and in turn for each of six areas, with the proper tolerance limits, for the 95 percent probability level, drawn around the expected values of average spectral densities of independent series.

Conclusions drawn are that the annual precipitation series are very close to independent time series. They are stationary series, at least temporary stationary for the length of time of the order of available series lengths of 50-150 years. Annual runoff and annual effective precipitation series are dependent series (with the average first serial correlation coefficient of the order of 0.10-0.20). They are stationary series, at least temporary stationary for the order of time length of 50-150 years. The first- and the second-order autoregressive models of series dependence seem sufficiently accurate for the use in practical problems.

PREFACE

The contemporaneous scientific and professional literature is full at present of various claims for the ongoing climatic changes. Some of their authors forecast the eventual forthcoming of the new ice age (therefore, they continue to speak about the present-day climate as the interglacial climate). Others claim that a warming trend is at hand due to the man's release both of the heat in using the various sources of energy, and, through the burning of focil fuels, also of the carbon dioxide with its green-house effect of heating the lower atmosphere. The concept of the increased carbon dioxide and the resulting warming effects is a sound approach in the analysis of man's influence on the earth's environments. Three factors, however, should be taken into account: (1) The tremendous potential of oceans to absorb the additional quantities of carbon dioxide; (2) The increase of production of the total green mass of modern agriculture in feeding the continuously increased population; and (3) The need for some heating on the earth for the purpose of compensating some expected, but relatively small, cooling in the Northern Hemisphere of the Earth, because of the future changes in distribution of solar radiation over the Earth (Milankovich's phenomena of long-range, almost-periodic fluctuations in solar energy distribution over the earth). Likely, the effects of the artificial heat and the carbon dioxide releases, plus the other man-made effects on atmospheric composition and its transparency for radiation and irradiation waves, are the most attractive short-range, middle-range and long-range objectives of monitoring changes and forecasting the future climatologic effect.

The practical water resources problems impose an interest for the immediate future, say for the next 100-200 years. This paper approaches the fluctuations of wet and dry years from the point of view what can be extracted from the best data of the near past, with the high probability that the future data will show, in the limits of the sampling variation, the same or very close to the same characteristics of climatic and hydrologic time processes as they were for the last 100-200 years. Some recent studies of dendrochronology may extend the past instrumental data up to several more centuries, but with the increased errors.

This study leads to the conclusion of an unusually high "stability" of properties of major processes, namely the stability in the fluctuations of wet and dry years of precipitation and runoff. Because the proofs of an approximate stability of phenomena, and the projection that the stability will likely continue for some time to come, are not as glamorous conclusions as the projection of an "ice age" or "heating up" of many Earth's environments. The writer hopes that the conclusions drawn in this study may give some confort to those in practical fields of endeavor, who plan systems and make decisions, drawn on the conclusions from the best data of the past, assuming that the near future will be similar to the past. Those who doubt this approach are invited to place themselves at the year 1890 (with some instrumentally obtained data of about 85 years long, available at that time), and project the behavior of those phenomena for the period 1890-1975. How surprised they would be at the accuracy of their projections, based on the temporary stationarity of annual precipitation and annual runoff data.

August 1977
Fort Collins, Colorado

Vujica Yevjevich
Professor of Civil Engineering and
Professor-in-Charge of Hydrology and Water Resources Program

Chapter I
INTRODUCTION

Investigations of wet and dry years of precipitation, runoff, and other basic hydrologic phenomena are as old as the earliest human water resource activities. Even the writers of the Bible incorporated statements on these fluctuations by referring symbolically to the seven wet and seven dry years. These continuous investigations have paralleled similar inquiries in geophysics on fluctuations of annual series of climatic variables, particularly temperature. From reliable records of measured values of certain hydrologic phenomena, inferences have been drawn in this paper regarding fluctuations of wet and dry years, using variance spectrum analysis. It is the purpose of this introduction to put these results in perspective. The study of fluctuations of wet and dry years is conceived in this text as being equivalent to the investigation of fluctuations of annual series of the most important hydrologic variables, here chosen to be precipitation and runoff.

1-1 Previous Papers of the Same General Title

Hydrology Paper No. 1, July 1963, under the title "Fluctuations of Wet and Dry Years" and the subtitle "Research Data Assembly and Mathematical Models" [1], dealt with the selection of samples of data, general mathematical models of physical processes which produce the time dependence in annual hydrologic time series, the compilation of sample data, discussion of problems related to errors and nonhomogeneity in data, and gave an appendix of data of annual series in modular coefficients of 140 selected runoff gauging stations from around the world.

Hydrology Paper No. 4, June 1964, with the same title as Paper No. 1 and the subtitle "Analysis by Serial Correlations" [2], dealt with annual precipitation and runoff series analyzed by the method of serial correlation or the autocorrelation technique. The research data consisted of the six sets of annual time series: (1) A set of 140 world-wide series of annual runoff (Q_1-series); (2) The same set of these 140 series but with the estimated annual effective precipitation (Q_2-series) defined as the annual precipitation minus the annual evaporation, and computed by adding to or subtracting from the annual runoff the annual change in the total water stored in the river basin at the end of each water year; (3) A set of annual runoff series (Q_1-series) of 446 river gauging stations in western North America; (4) The same set of 446 series but with the estimated annual effective precipitation (Q_2-series) defined by the same method as described under (2); (5) A set of annual precipitation series of 1141 gauging stations (P_1-series) in the same region of western North America as for the set of runoff series, with these series considered more or less homogeneous (not affected by man's activities) and/or consistent (not having significant systematic errors in the form of trends and jumps, introduced by methods of measuring precipitation and/or the change in the surroundings of gauges, which would affect the instrument's catch of precipitation); and (6) A set of 475 annual precipitation series (P_2-series) in western North America, which series have been inferred to be either nonhomogeneous (especially through the changes in station location and/or elevation during the time of observations) or inconsistent (produced by the effects of changing environment around the gauge sites during the observational period).

The analysis presented in Hydrology Paper No. 4 [2] led to the following basic conclusions:

(1) The hydrologic phenomenon of contemporaneously observed precipitation by instruments, particularly in the form of the total annual precipitation, is an excellent measure of eventual climatic changes on the scale of decades. The results show that this variable is approximately a time independent and stationary stochastic process for a period of the most reliable data. The time dependence, measured by the mean first serial correlation coefficient of annual precipitation (P_1-series), is of the order of $\bar{r}_1 = 0.028$ as the average for the 1141 stations for the period of simultaneous observations of 30 years (1931-1960). For all years of observations available at all stations of P_1-series, with the average series length of 54 years, the mean value is $\bar{r}_1 = 0.055$. These values imply that only the portions of 0.028^2 and 0.055^2 (or 0.079% and 0.305%) of the unit variances of standardized annual precipitation (P_1-series) are explained (or affected) by the previous year(s). In other words, only 0.08% and 0.30% of the total variation of annual P_1-precipitation in any year is explained, on the average, by the annual precipitation which has occurred in the previous year. This conclusion results from using both the average series lengths of 30 and 54 years. For all practical purposes, the annual precipitation (P_1-series) can be considered an independent time process for time spans of many decades.

(2) The 475 annual series of precipitation (P_2-series), inferred to be nonhomogeneous and/or inconsistent, show the $\bar{r}_1$ values to be somewhat larger than for the homogeneous and/or consistent annual precipitation (P_1-series), namely 0.053 for the 30-year period (1931-1960) and 0.071 for the average series length of 57 for all the available observations. Any inconsistency and nonhomogeneity, such as positive and negative jumps in the series mean, or linear and nonlinear trends if introduced into an independent or dependent time series will, on the average, increase the values of r_1. Differences between the averages, $\bar{r}_1$ for 475 P_2-stations and $\bar{r}_1$ for 1141 P_1-stations, for the average lengths of 30 and 54 years, are 0.025 and 0.016, respectively, which are 90 and 29 percent greater than the corresponding values for the series of 1141 stations, inferred to be approximately homogeneous and/or consistent. Regardless of the inference techniques used in separating 1614 annual precipitation series into 1141 approximately homogeneous and/or consistent, and 473 nonhomogeneous and/or inconsistent series, the probability is high that part of the positive serial correlation for the case of 1141 series, for both the 30 and 54-year lengths, may be due to some nonhomogeneity and inconsistency, present in nearly all the annual precipitation series. This is a reasonable conclusion because it is

self-evident that the longer a time series becomes, the larger the probability of introducing at least one form of inconsistency or nonhomogeneity. For practical problems related to water resources conservation, control, and development, the annual precipitation can be considered to be either an independent or an almost independent stationary stochastic time process, provided the effects of nonhomogeneity and inconsistency are properly taken into account on the time scales of many decades.

(3) The two Q_2-series of annual effective precipitation, worldwide series and series of North America [4], have higher average estimated first serial correlation coefficients than were found for the P_1-series of annual precipitation. For worldwide Q_2-series, whose average length is 55 years, $\bar{r}_1$ was estimated as 0.136. For Northern American Q_2-series, whose average length is 37 years, $\bar{r}_1$ turned out to be 0.181. Let's define {P} as the time series of annual precipitation, {E} as the time series of annual evapotranspiration, and $\{P_e\}$ as the time series of annual effective precipitation. Evidently $P_e = P - E$. Since {P} is an approximately independent time series and the values of $\bar{r}_1$ for both Q_2-series (P_e-series) are much greater than $\bar{r}_1$ for P_1-series, than {E}, regardless of its dependence on {P}, must be an auto-correlated and hence time dependent process. Because $P - E = R + \Delta W$, with ΔW the change in the total water stored in a river basin at the end of each water year and R the runoff, then $E = F(\Delta W)$, i.e. it is dependent on the change ΔW in the stored water available for evaporation, besides being dependent on the total precipitation. One may expect that the potential annual evaporation (evaporation where water is always available for full evaporation potential) should also be an independent annual process similar to the annual precipitation. Since more stored water means more water is available for evaporation, and since the stored water depends on the hydrologic history of previous time intervals, the effective annual evaporation must be a dependent process, similar to the dependence of basin water outputs, with precipitation the input and both these outputs dependent on the state of water storage of various river basins.

(4) Series of annual runoff are either independent processes, when negligible changes in the basin stored water occur at the end of each water year, or they are dependent processes when the storage at years' ends fluctuates in a relatively large range in comparison with the average annual runoff. Large variations in water carryovers from year to year may be considered as the principal physical factor which affects the time dependence of both the annual evaporation and the annual runoff. The two sets of annual runoff series used in investigations gave $\bar{r}_1 = 0.175$ as the average for the 140 worldwide selected runoff series, with the mean series length of 55 years, and $\bar{r}_1 = 0.197$ as the average for the 446 runoff series in western North America, with the mean series length of 37 years. Both sets showed that the average first serial correlation coefficient of annual runoff series is close to about $\bar{r}_1 = 0.20$.

1-2 Basic Scientific Controversies Related to Persistence in Hydrologic Time Processes

Dependence in hydrologic time series is often referred to as hydrologic persistence. Values of the process tend to persist in the sense that probabilities of high values following high values (and the converse, probabilities of low values following low values) tend to be higher than probabilities associated with the same high (or low) values of time independent hydrologic processes. Sometimes, the concepts of short-range, mid-range, and long-range persistences are used; rarely are the ranges of time intervals associated with these concepts adequately defined.

It is generally accepted by most geophysicists that basic climatic changes occur as long-range variations. The changes occur mainly as a result of astronomical causes, related to changes in the distribution of incoming solar energy over the earth's surface.

The Milankovich theory of astronomical causal factors shows regular, almost-periodic changes in the eccentricity of the earth's orbit (one complete oscillation in about 93,000 years), the tilt or obliquity of the ecliptic (one complete oscillation in about 41,000 years), and the precession of the equinoxes (21,000 years per one complete oscillation), [3, 4, 5]. These deterministic, astronomical movements produce long-range changes in the distribution of incoming solar energy over the earth's surface, even under the assumption (which is now in doubt) that no significant changes in the solar energy constant have occurred for the last couple of millions of years. The change in the seasonal distribution of energy over the earth's surface must result in changes of climate on the earth. When ice sheets grow over the continents, the ocean level recedes, the continental shelves become exposed with a resulting increase in the continental surface and a decrease in the ocean areas. This leads to changes in oceanic processes (such as currents, heat budget, evaporation, types of water mass exchanges, etc.). Similarly, changes occur in the atmospheric composition, circulation, climate and the basic hydrologic processes of precipitation, evaporation, and runoff. Only well-studied geophysical problems, examined jointly as paleo-oceanography, paleo-meteorology, paleo-geology, paleo-morphology, paleo-glaciology, paleo-hydrology, and other geophysical paleo-processes, could explain the real physical interactions between the astronomical, almost-periodic movements and the various geophysical processes in order to explain these long-range climatic changes.

Historic evidence, particularly from the last Pleistocene ice age, confirms that long-range climatic changes do occur on the earth. These changes undoubtedly affect the annual series of precipitation, evaporation, and runoff of various river basins. The main question from the hydrologic standpoint is, what are the rates of change with time of various parameters associated with these processes. It can be shown that the rate of change is so small for a time span of 3-4 centuries [6], say 150-200 years of the past and 150-200 years of the future , that the annual processes of precipitation, evaporation, and runoff may be safely considered to be *temporarily stationary* stochastic processes. The best available observational data on precipitation and runoff of the last 100-200 years show no significant change in the basic characteristics of annual hydrologic processes, particularly if account is taken of the unavoidable nonhomogeneities and/or

inconsistencies (and the sampling fluctuations of these characteristics), to be found in the data associated with any real geophysical stochastic process.

The question which highlights the basic controversy among climatologists, hydrologists, and other specialists in geophysics could be summarized as follows: are the climatic and hydrologic processes, taken on an annual basis, to be considered as *temporarily stationary* or *quasi-stationary*, with a slow rate of change of the basic characteristics of stochastic processes for the period limited to 150-200 years of the recent past, and by extrapolating this recent past for the period of 150-200 years of the near future? If This *temporary stationarity* was rejected, one would then expect that the extreme events of some distant past, especially of the post-glacial era, may occur today--suddenly--with the same probabilities as they occurred before; this is not a plausible hypothesis, as the following argument shows. The biblical Noah inundation may well have been an event produced by a combination of extreme precipitation and the simultaneous melting of accumulated mountain snow and ice in an era of general melting and retreating of ice sheets and ice glaciers of the Northern Hemisphere. While it is reasonable to expect the extreme precipitation event of the Noah type to occur from time to time somewhere in the world by change (such as the 40-days precipitation event in Tunisia in September 1969, or similar examples), the other basic condition of rapid melting of large quantities of accumulated snow and ice does not exist at present in most areas of interest and, therefore, this melting cannot be compounded with similar rare events as experienced in the recent past.

This controversy has an important and very practical implication for water resources planning and management: is it legitimate (and with a very high probability it is) to draw information about the characteristics of hydrologic processes and available water resources in the last 150 to 200 years from the best data on precipitation and runoff available in the world, and to expect approximately the same or very close characteristics to occur in the realization of these processes and in available water resources in the next 150 to 200 years? If this approach is not justified, should then the planners of future water resources systems use the opposite approach, namely to speculate with various climatic change theories (mostly supported by unreliable or at least questionable evidence), developing the inevitable conclusion that the hydrologic processes and their characteristics could suddenly or relatively rapidly evolve? In the extreme, these conclusions may imply that the climate could rapidly deteriorate into a new ice age in the northern parts of America, Europe, and Asia; however, this is very unlikely from the physical point of view as the following argument demonstrates.

A recent study [6] underlines the point that the buildup of ice sheets and large mountain glaciers is a relatively slow process, while the melting of those once created may be a relatively rapid process. This implies that the rate of change in the initiation phase of build-up of ice sheets and large glaciers is much slower than the rate of change during their disappearance. Therefore, a relatively fast rate of melting of the Pleistocene ice sheets in northern America and northern Europe cannot be taken as the potential rate of the buildup of a new ice sheet. Besides, the extrapolation of the Milankovich astronomical almost-periodic long-range fluctuations, as the predictable, deterministic astronomic process, shows that for the next 100,000 years little buildup of an ice sheet in the northern hemisphere can be expected, though some minor cooling should be expected to take place.

Considering an interval of time of about 350 years, say from 1800 through 2150 (175 years in the immediate past and 175 years in the immediate future), the following conclusions may be safely drawn for the investigations of long-range water resources problems, with a very high probability that these conclusions will be confirmed by future observations:

(1) Processes of annual precipitation, annual evaporation, annual effective precipitation on river basins, annual runoff from river basins, and similar and/or interconnected hydrologic processes may be considered as approximate *temporary stationary* stochastic processes, provided the systematic errors in observed data (inconsistency), the man-made changes and accidents in nature (nonhomogeneity in data), and the sampling fluctuations in realizations of these random processes, are properly taken into account.

(2) If the annual precipitation may be considered as an approximate, temporary stationary stochastic process in the interval of the past 150-200 years, it is a logical analogy to consider the annual evaporation also as an approximate, temporary stationary stochastic process.

(3) The major time dependence in hydrologic annual series is produced by the complex geophysical processes of water storage in river basins, with their random fluctuations from year to year and periodic-stochastic fluctuations within the year. The ambiguity of the concept of hydrologic long-range persistence as related to the time scale of several decades, or a couple of centuries, as contrasted with the analysis of actual geophysical processes which create the time persistence, only confuses the issues, though it may serve particular objectives of supporting theories and mathematical models advanced for hydrologic persistence.

(4) The more ancient the data of observed (measured) or inferred hydrologic variables are, the more likely it is that they contain some systematic errors (inconsistency). The use of earlier, less reliable instruments and measuring techniques, and the ensuing changes in instrumentation and techniques of measurements, as well as the environmental impacts on observational stations, support the existence of inconsistency in various series.

(5) The longer a series, the greater is the probability of some nonhomogeneity being present in the data, produced either by man's activities or by accidents in nature.

(6) The probability that two sample means of two subseries of an observed series are identical is very small. Sampling fluctuations which leave visual impressions of trends, jumps, and light cyclicity are often erroneously treated as population trends, jumps, and cyclicities.

(7) Some mathematical models proposed for the description of time dependence in hydrologic annual series may often be the results, partly or fully, of inconsistencies, nonhomogeneities, and sampling fluctuations, rather than of the underlying true geophysical processes as derived from large sets of series from stations all around the world.

(8) The analysis of only a limited number of stations, particularly when these sample series contain inconsistency, nonhomogeneity, and evidently large sampling deviations in comparison with the adjacent stations, may well support a particular concept or mathematical model, even though it cannot be justified by the existing geophysical and/or historial evidence

about the reliability of available data for those stations.

This investigation of fluctuations of wet and dry years of annual series of hydrologic processes is thus committed to using large sets of series. In using large sets of series, the biases due to individual series are minimized (especially the bias contained in the form of extreme sample deviations), while inconsistency and nonhomogeneity in the data may be significantly reduced by some objective criteria of selecting the sets of series. In some cases, the bias may be reduced, because of the combined effect of opposing biases in a large number of series.

It is feasible to process a very large number of time series in the present age of large digital computers, treating them as space-time processes. The space variation is covered by a set of points in a geographical coordinate system and the time variation by the longest observed series, reasonably consistent and homogeneous. Results of investigations should be independent of particular characteristics of a limited number of series in a restricted area.

It is a common practice among researchers to abandon and not to report on results of investigations if the research data do not support either the approach taken or the hypotheses advanced. Mostly, the confirming results are reported in literature. A reasonable question may be, whether in some cases the confirming results in the assumed approach are nothing else than the extremes of sampling deviations, with small probabilities for them to occur again in future realizations at the same stations. The well-publicized Brueckner 35-year climatic cycle, developed for the first 70 years of hydrologic data in Europe, was not supported by the data of the next 40-50 years. If Brueckner had used a wide range of variables, and from several large regions of the world, it is likely that he would not have concluded that a regular 35-year cyclicity existed in his climatic and hydrologic processes as he did by using the European data only.

1-3 General Explanation of Long-Range Climatic and Hydrologic Persistence

It is an attractive and plausible approach, at least to the writer of this paper, to explain the long-range climatic and hydrologic changes of annual processes by the theory of a deterministic-stochastic climatic process. The deterministic part is produced by the processes following the Milankovich theory of astronomical movements. The stochastic part is explained by various random processes in the earth's environments, or by the geophysical processes.

The distribution of incoming solar energy at the upper atmosphere is determined uniquely at a given historic time by the astronomical movements of orbital eccentricity, tilt, or obliquity of the ecliptic, and precision of the equinoxes for whatever the solar constant may be at that time. Figure 1-1 shows periodicities of the major astronomical cycles which affect the geophysical processes. Because the computed annual series mask the higher frequency cycles up to the year, they are mainly affected by the lower frequencies associated with the Milankovich astronomical processes. The sunspot cyclicity is not a function of positions of celestial bodies, and is not discussed here (although it may introduce perturbations on the solar constant). Therefore, for a given state of tectonic plates of the earth's crust (say the positions and elevations of continents and continental shelves), and for given states of accumulated snow and ice over the various areas at a given historic time, the earth reacts in a given manner to the deterministic distributions in space and time of the incoming solar energy. The state of distributions of water, snow, ice, and volatiles in the atmosphere, oceans, and on the continents predetermines the general earth's response to these deterministic distributions of incoming solar energy. The average existing patterns of oceanic, atmospheric, and continental processes adjust to the incremental changes in distributions of this energy in an evolutive manner.

The stochastic part of the climatic process results from the versatile stochastic processes in oceans, on continents, in the continental crust, but particularly in the atmosphere. The major properties (probability distribution, time dependence) of the stochastic part of climatic variations may be more or less dependent on the average deterministic responses of the earth to the earth's distribution of the incoming solar energy in space and time. The random processes of various earth's environments are mutually dependent processes, some of them being the preceding, causal processes, with others the resulting, effect processes, or they may be simultaneous dependent processes.

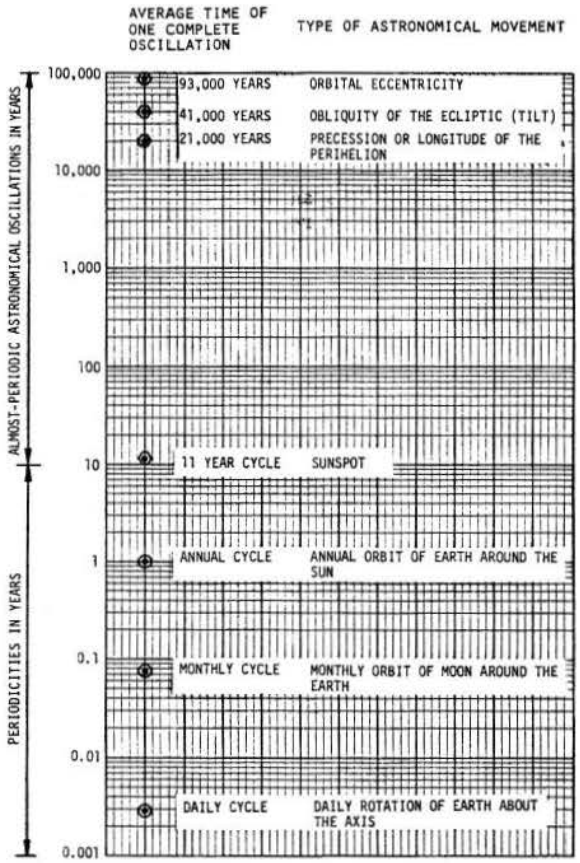

Fig. 1-1. Periodicities of Major Astronomical Cycles which Affect the Geophysical Processes.

Until evidence to the contrary comes to light, one may consider the solar constant as a real constant, with only the position of bodies in the solar system being the major factor in determining the space-time distribution of incoming solar energy over the earth. In a condensed way, this approach to the long-range climatic changes is based on the hypothesis that the astronomical movements produce the deterministic part in the averages and other parameters of the climatic

and hydrologic variables of the earth, while all types of geophysical random processes in all of the earth's environments (particularly in the atmosphere, with the air being a nonconservative fluid), represent the sources of the stochastic part of the climatic and hydrologic variables. The dependence in the stochastic part is the result of various feedback processes, due mainly to storage or depletion in the earth's environments of many physical equivalents of random variables (such as vapor, water, snow, ice, heat, volatiles, solid particles, chemicals, kinetic energy, etc.). The above general approach to climatic and hydrologic changes assumes also that man's activities and special accidents in nature (catastrophies) do produce the nonhomogeneities or nonstationarities in stochastic components in complex deterministic-stochastic series.

1-4 Selection of Investigation Method

The selection of an investigation method for the analysis of hydrologic time series may depend on whether a series is stationary or nonstationary. As stated in the preceding text, the annual series of most hydrologic time processes may be safely considered either *temporarily stationary* (say, approximately stationary for a couple of centuries) or *quasi-stationary* (the trend in the change of basic population parameters in the time span of 3-4 centuries may be neglected), or both.

In reference [2] the investigations of fluctuations of wet and dry years used the autocorrelation technique. In this paper the spectral analysis, or the variance density spectrum, is selected as the technique for investigation. One may question this latter selection by asking whether the use of spectral analysis provides any improvement, substantial difference or additional information in comparison with the use of autocorrelation technique. The question is logical because the Wiener-Khinchine equations, as shown in Chapter II, provide unique transformations between the autocorrelation function (estimated by the sample correlogram) and the spectral, variance density function (estimated either by the Fast Fourier Transforms and smoothed, or by smoothing and transforming the correlogram). Essentially, the two techniques should produce the same results as two equivalent methods in the investigation of hydrologic stationary processes. Three reasons have induced the writer to use the spectral analysis in this study of hydrologic annual series:

(1) Some specialists are more exposed to the spectral analysis technique than to the autocorrelation technique; they can better see and infer the type of stationary process in the frequency domain of spectral analysis than in the time-lag domain of autocorrelation analysis.

(2) Spectral density graphs are smoothed in two ways: (a) by using the smoothing functions (either in the time-lag domain by a smoothing function, or in the frequency domain by the kernel function), and (b) by using a large number of series in a region, averaging the correlograms and estimating the spectral densities at given frequencies, to obtain mean regional results. The inference from the averaged, regional spectra is expected to demonstrate more reliably the basic properties of any process studied than the case would be if only a limited number of spectral graphs of individual series were examined.

(3) New data have been accumulated since reference [2] appeared. The conclusions derived by the autocorrelation technique may be then either revised or reinforced by another technique regardless of the strong interrelationship between these two techniques. Furthermore, the United States is divided into six regions, in this new study, in order to investigate whether significant differences in results may be discerned among the regions in using these two techniques.

1-5 Objectives of Investigations by Spectral Analysis

The analysis of annual precipitation, annual effective precipitation and annual runoff by using the variance spectra technique has the following detailed objectives:

(a) To show whether these hydrologic series are independent or dependent stationary processes;

(b) To find the degree of time dependence when it is present in time series, analyzed for sets of series:

(c) To make inferences concerning the most appropriate mathematical models to be used for description of dependence for a set of time series;

(d) To compare the degree of time dependence in these series, especially how it increases from precipitation to runoff, and

(e) To investigate the self-stationarity of annual runoff time series from a regional point of view.

1-6 Continuous Variance Density Spectrum Versus the Discrete or Line Spectrum

Since previous studies have shown that annual series of precipitation, effective precipitation and runoff do not contain periodicities, the use of the line spectrum (periodogram) is not the most feasible technique to study the approximately stationary time processes. When the range of frequencies with significant variance densities of a stationary process must be estimated, the line spectrum is not the most appropriate technique because of bias and inefficiency in estimates. It is replaced by the technique of continuous variance spectrum. For this reason, the periodogram approach is not even attempted in these investigations.

Chapter II
TECHNIQUE OF SPECTRAL ANALYSIS

2-1 Mathematical Description of Time-Dependent Hydrologic Processes

A time-dependent hydrologic process is a stochastic process involving hydrologic variables. Sequences of observations on the variables which characterize the hydrologic processes are either continuous or discrete stochastic processes. The continuous stochastic processes found in hydrology (and some discrete ones) are nonstationary processes, mainly because of the periodicity found in basic parameters induced by the diurnal, monthly, and annual astronomic cyclicities. Various sources of trends and jumps also cause some series to be nonstationary processes. The term *process* is used here in the narrow sense of *stochastic process*. It is further assumed in these investigations that any deterministic dependence on time, such as known trends or built-in periodicity, have been removed from the process under consideration. The diurnal, monthly, and within-the-year periodicities in parameters disappear by integrating a process over intervals of a year. The discrete annual series of precipitation, effective precipitation, and runoff considered in this study are free of trends and low frequency periodicity. By contrast, any undetected nonhomogeneity in parameters of these series, produced by various factors as discussed in Chapter I, will cause nonstationarity, and it is this that should be detected by using variance spectrum analysis.

The remainder of this chapter contains a condensed presentation of the practical variance spectrum technique, adapted to the objectives of this paper.

2-2 Variance Density Spectrum

Fourier (periodogram) analysis of a time series tacitly assumes that the series is made up from sums of harmonics which have fixed (or almost fixed) periodicities. The continuous spectrum introduced by Wiener overcomes this prerequisite. Spectral analysis has been used effectively in many fields for the analysis of the structure of time series. It is based on the concept of the continuous spectrum, which is a relationship between variance densities and frequencies.

The population spectral (variance) densities, $\nu(\lambda)$, of a continuous series are obtained by the Fourier transform, for a given angular frequency, , of the corresponding population continuous autocorrelation function, $\rho(\tau)$, by using the Wiener-Khinchine equation:

$$\nu(\lambda) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \rho(\tau)\, e^{i\lambda\tau}\, d\tau = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \rho(\tau)\cos\lambda\tau\, d\tau \quad . \tag{2-1}$$

Similarly, for a discrete time series, the spectral density function is the Fourier transform of the discrete autocorrelations function, $\rho(k)$, and is again a function of λ:

$$\nu(\lambda) = \sum_{k=-\infty}^{+\infty} \rho(k)e^{i\lambda k} = \frac{1}{2\pi}\sum_{k=-\infty}^{+\infty} \rho(k)\cos\lambda k. \tag{2-2}$$

In the opposite transformation, the autocorrelation function is the Fourier transform of the spectral function, so that for the continuous case

$$\rho(\tau) = \int_{-\infty}^{\infty} \nu(\lambda)e^{i\tau\lambda}\, d\lambda = \int_{-\infty}^{+\infty} \nu(\lambda)\cos\tau\lambda\, d\lambda \tag{2-3}$$

and for the discrete case

$$\rho(k) = \int_{-\infty}^{\infty} \nu(\lambda)e^{ik\lambda}\, d\lambda = \int_{-\infty}^{+\infty} \nu(\lambda)\cos k\lambda\, d\lambda \quad . \tag{2-4}$$

It frequently occurs in *spectrum carpentry* that an appropriate moving average scheme is needed to smooth either the sample estimates $r(\tau)$ and $r(k)$ of the autocorrelation functions $\rho(\tau)$ and $\rho(k)$ (Eqs. 2-1 and 2-2), or to smooth the estimates $v(\lambda)$ of $\nu(\lambda)$ after the transformation has been made. For the continuous case, smoothing in the time domain gives

$$v(\lambda) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} D(\tau)r(\tau)e^{-i\lambda\tau}d\tau = \frac{1}{2\pi}\int_{-\infty}^{+\infty} r(\tau)D(\tau)\cos\lambda\tau\, d\tau \ , \tag{2-5}$$

where $D(\tau)$ = a smoothing function for the estimate $r(\tau)$ of $\rho(\tau)$ in the time domain. The transformation in the discrete case gives the following estimates of $v(\lambda)$

$$v(\lambda) = \frac{1}{2\pi}\left[1 + 2\sum_{k=1}^{\infty} r(k)\cos\lambda k\right] \ , \tag{2-6}$$

where $0 \le \lambda \le \pi$. The period ω, the frequency f and the angular frequency λ are related by $f = 1/\omega = \lambda/2\pi$. Now $E[v(\lambda)] = \nu(\lambda)$, [8], and the mean of $\nu(\lambda)$ is $1/\pi$ over the range of λ, as the expected value of all variance densities over $0 \le \lambda \le \pi$ in case of a standardized process (0,1). For the estimates of variance densities for cases in which the ordinary frequencies, f, are used, Eq. 2-6 becomes

$$g(f) = 2\left[1 + 2\sum_{k=1}^{\infty} r(k)\cos 2\pi fk\right] \ , \tag{2-7}$$

with $g(f)$ = the estimates of the variance densities $\gamma(f)$. Here, $0 \le f \le 0.50$, $f = 1/\omega$, $E[g(f)] = \gamma(f)$, and the mean of all $\gamma(f)$ values is two over the range from 0 to 0.50 for a standardized process (0,1).

Using Eqs. 2-6 and 2-7 for the estimation of variance densities, $\gamma(f)$, requires more computer time than the application of the Fast Fourier Transform, FFT, for estimating $g(f)$ directly and smoothing these estimated variance densities in the frequency domain. Both approaches--the Wiener-Khinchine transformations and Fast Fourier Transforms--lead to the same results, if the same smoothing function is used in the frequency domain (or the corresponding function in the time domain) and the same resolution (the number of variance density ordinates) is selected for the estimation of densities; the only difference is a saving of computer time by using FFT. If Eq. 2-6 or Eq. 2-7 are used, with the infinite limit of the sum replaced by a value $k = m$, with m determined by some criterion (either objective or subjective), and if a smoothing function $D(k)$ is introduced, Eq. 2-7 becomes

$$\hat{g}(f) = 2\left[1 + 2\sum_{k=1}^{m} D(k)r(k)\cos 2\pi fk\right] \quad . \tag{2-8}$$

This is a practical form of the equation for transforming the estimate $r(k)$ of the autocorrelation coefficients $\rho(k)$ into the estimates $\hat{g}(f)$ of the variance densities $\gamma(f)$. In practice, $\rho(k)$ are estimated by the sample values $r(k)$, and their transforms $g(f)$ are

smoothed either in the time domain by Eq. 2-8, or similarly in the frequency domain by the corresponding kernel function.

It should be stressed that the following terms: the moving average scheme, the moving average model, the smoothing function, the filtering function, the filter, the window function, and similar terms, are considered all synonymous for the purposes of this paper. The term *smoothing function* is used in this text.

2-3 Estimation of Variance Densities

The practical use of the continuous spectrum began in 1948 and 1949, mainly in the work of H. T. Budenbom and F. W. Tukey, when they initiated the analysis of radar trackings for the Bell Laboratories in the form of continuous power spectra. Fourier cosine transforms of the type of Eqs. 2-1 and 2-2 were used. Blackman and Tukey [8] suggested the use of a three-point smoothing function in the frequency domain to smooth the computed variance densities, with the symmetrical weights: (1/4, 1/2, and 1/4), to produce smooth and less biased estimates of the variance density spectrum.

The three smoothing functions that were most commonly used in USA during the 1950's and 1960's are those proposed by Bartlett, Tukey, and Parzen. Smoothing in this investigation is made by Parzen's smoothing function in the time-lag domain, substituted into Eq. 2-8. Parzen's smoothing function is:

$$\left.\begin{aligned} D(k) &= 1 - 6\left(\frac{k}{m}\right)^2 + 6\left(\frac{k}{m}\right)^3 \quad ; \text{ for } k \le \frac{m}{2} \\ &= 2\left(1 - \frac{k}{m}\right)^3 \quad ; \qquad \text{for } \frac{m}{2} < k \le m \\ &= 0 \quad ; \qquad \text{for } k > m \quad . \end{aligned}\right\} \tag{2-9}$$

Smoothing by Eq. 2-9 accomplishes the following objectives: (1) it is simple, (2) it uses small computational time, (3) the resolution of distance between discrete spectral lines in estimating the continuous spectrum can be kept relatively large, or the inverse, the number of points at which the spectral estimates are made on the line $0 \le f \le 0.50$ can be relatively small; and (4) the bias and inefficiency of estimates $\hat{g}(f)$ of the spectral densities $\gamma(f)$ are relatively small [9]. The selection of m in this study is made as $m = N/4$, and the spacing of the values is unrelated to the choice of m. It was arbitrarily selected as $\Delta f = 0.05$, so that eleven ordinates of $\hat{g}(f)$ need to be estimated on the interval $0 \le f \le 0.50$.

2-4 Whitening

The concept of whitening the sample series is based on the hypothesis of a stochastically dependent model for a stationary process. The computed residuals of the model should then be time independent stochastic components (TISC) or independent identically distributed random variables (IIDRV). If the hypothesis of the model with its unbiased and most efficient estimates of parameters is correct, the residuals should pass the test of independence; the hypothesis of the dependence structure and/or model, with the corresponding estimates, should be accepted. The term *whitening* usually refers to normal random variables; however, the concept of whitening does not necessarily include the condition that the probability distribution of residuals must be normal. As long as residuals are time independent identically distributed random variables, with moderate skewness coefficients, they may be considered as white noise, and the series called a whitened series. The major advantage of the whitening concept is that the inference technique in spectral analysis about the hypothesized independent residuals is simpler than in the case of dependent processes. The expected variance densities of the whitened process are all equal. The case of white noise, Eqs. 2-7 and 2-9 yield $E[\hat{g}(f)] = \gamma(f) = 2.00$. The tolerance limits for the hypothesis of $\gamma(f) = 2.00$ for all frequencies are also equal for all f values except for the two extreme values of $\hat{g}(f)$, namely $\hat{g}(0)$ and $\hat{g}(0.50)$, because of the double sampling variance of the two estimates $\hat{g}(0)$ and $\hat{g}(0.50)$ in comparison with the variance of all other estimates of intermediate frequencies.

2-5 Variance Density Functions for the First and Second-Order Linear Autoregressive Models of Temporary, Stationary Annual Stochastic Processes

Denoting the time interval of discrete series as unity (in this study t = 1 year), the expected spectral function of a standardized process (i.e. with zero mean and unit variance) following the first-order autoregressive (Markov) model

$$x_i = \rho x_{i-1} + (1 - \rho^2)^{1/2} \varepsilon_i \tag{2-10}$$

is found to be:

$$\gamma(f) = \frac{2(1 - \rho^2)}{1 - 2\rho \cos 2\pi f + \rho^2} \quad , \tag{2-11}$$

with $\gamma(0) = 2(1+\rho)/(1-\rho)$, $\gamma(0.25) = 2(1-\rho^2)/(1+\rho^2)$, and $\gamma(0.50) = 2(1-\rho)/(1+\rho)$. The maximum ordinate is always $\gamma(0)$ at f = 0.

For the second-order autoregressive model, with both x and ε standardized variables (0,1),

$$x_i = \alpha_1 x_{i-1} + \alpha_2 x_{i-2} + \sigma \varepsilon_i \quad , \tag{2-12}$$

the spectrum function is

$$\gamma(f) = \frac{2(1+\alpha_2)(1-\alpha_1^2+\alpha_2^2-2\alpha_2)}{(1-\alpha_2)[1+\alpha_1^2+\alpha_2^2+2\alpha_2-2\alpha_1(1-\alpha_2)\cos 2\pi f - 4\alpha_2 \cos^2 2\pi f]} \tag{2-13}$$

with $\sigma^2 = (1+\alpha_2)(1-\alpha_1^2+\alpha_2^2-2\alpha_2)/(1-\alpha_2)$, and the maximum of $\gamma(f)$ at $f = f_o$ determined from $\cos 2\pi f_o = \alpha_1(1-\alpha_2)/4\alpha_2$.

2-6 Tolerance Limits of the Spectrum for Time Independent Identically Distributed Random Variables

The estimated spectral densities of standardized normal independent process are chi-square distributed with a proper number of degrees of freedom. This number depends on: (1) the length m of the correlogram, r_k with $k = 1,2,\ldots, m$, used in Eq. 2-7 in estimating the spectral densities, (2) the sample size N, and

(3) the smoothing function applied. This latter effect is usually given for any proposed smoothing function by its author(s), with EDF (effective degrees of freedom) expressed only as a function of N and m, for each smoothing scheme. For the smoothing function of Eq. 2-9, the suggested values [9] for ν are: EDF = 3.7 N/m for normal variables, and EDF = 4 N/m for non-normal variables; for m = N/4, EDF = 14.8 for normal variables and EDF = 16 for non-normal variables.

For the selected tolerance level α (say $\alpha = 0.05$, or $\alpha = 0.10$), the tolerance limits are

$$T_1 = \frac{2\chi^2_{\alpha/2}(EDF)}{EDF} \text{ and } T_2 = \frac{2\chi^2_{1-\alpha/2}(EDF)}{EDF} \qquad (2\text{-}14)$$

with $\chi^2_{\alpha/2}(\nu)(EDF)$ the value of χ^2 for given EDF at the left tail for the probability $\alpha/2$ and $\chi^2_{1-\alpha/2}(EDF)$ the value of χ^2 on the right tail for the probability $(1 - \alpha/2)$. Because $E[\chi^2(EDF)] = EDF$, while $E[\hat{g}(f)]=2$, the values $\chi^2_{\alpha/2}(EDF)$ and $\chi^2_{1-\alpha/2}(EDF)$, divided by EDF and multiplied by two, produce the necessary scale, so that the $E[\chi^2(EDF)] = E[g(f)] = 2.0$. Similarly, var $[\chi^2(EDF)] = 2$ EDF, so that var $\hat{g}(f) =$ var $[2\chi^2(EDF)/EDF] = 8/EDF$.

2-7 Tolerance Limits for the Spectrum of a Set of Spacially Dependent Series of a Region for Time Independent and Dependent, Identically Distributed Random Variables

For the study of a large number of station series of the same random variable in a given region, the following two methods may be used to investigate whether the observed series or their whitened series are time independent. One method consists of testing each series individually to discover whether it is independent, and then seeing whether the total number of stations for which the hypothesis of independence is accepted is greater than a critical tolerance number. In the opposite case, if the percentage of cases with the accepted hypothesis is equal to or greater than the tolerance level (given as the probability of accepted hypotheses for independent processes), the observed or whitened series of a region are accepted as time independent. The other method uses the mean spectral variance densities. These mean densities are computed for each discrete frequency for which variance densities are estimated from all the individual station spectra of a region. In this paper the estimated variance densities of all n series of a region and a given variable are averaged. The tolerance limits for the mean spectrum are then determined, taking into account the entire set of n series, their sample sizes, N_j, and their cross correlation and autocorrelation dependence.

In using either of the above two tests for independence, the relation of all the estimated variance densities to tolerance limits must be precisely defined. The approach that all the estimated variance densities should be confined within the tolerance limits should be viewed as a conservative approach, or as a too rigorous a criterion. If variance densities are estimated at m + 1 points of the spectrum, a criterion for accepting the independence hypothesis may be

$$m^*_\alpha \leq \alpha(m + 1) \quad , \qquad (2\text{-}15)$$

with m^*_α = the number of densities allowed to be outside the tolerance limits. The dependence between the densities $\hat{g}(f)$ and frequencies (as a result of smoothing by Eq. 2-9 in order to obtain less biased and more efficient estimates), speaks somewhat against any density being outside the tolerance limits. However, assuming that for a given smoothing function the effective number of degrees of freedom, EDF, of Eq. 2-14 for the distribution of estimates $\hat{g}(f)$ has been well determined, then a test using Eqs. 2-14 and 2-15 for individual series should be performed.

The first method of testing the regional validity of time independence by the number of series which pass the test is not attractive, because of high lag-zero cross correlations between station series. If by pure chance a station series in the center of a region has the sample statistics drawn from the tails of their distribution, a large number of surrounding stations should also show similar sampling deviations from the population mean of that statistic. In other words, the proportion of the number of individually tested regional station series, which pass the test of independence, is not a proper test statistic to use.

The second method, that of testing the averaged spectra of n station series, requires the determination of the effective number n_e of mutually independent station series as being equivalent to n interdependent series; evidently $n_e \leq n$. The variance of the estimated mean spectrum of n_e mutually independent series is equivalent to the variance of the estimated mean spectral densities of all the station series for n regionally dependent series.

For every statistic there is a different n_e value, because the variance of estimates vary from one statistic to another. As the estimated variance densities, $\hat{g}(f)$, are the Fourier transforms of the estimated autocorrelation coefficients, r_k, one can use the effective number n_e of r_k as being equivalent to the effective number of $\hat{g}(f)$. Therefore, the problem is to determine the number of n_e spacially independent series equivalent to n spacially dependent series. Because the autocorrelation coefficients for annual series converge fast from $r_1 \geq 0$ to $r_k \approx 0$, it is sufficient in general to study the effective number n_e of stations only for $r_1 > 0$.

In cases the available annual series of size N are autocorrelated, then a small size $N_e \leq N$ is the sample size equivalent to the time independent series for each statistic. Therefore, n series spacially dependent, each series of size N also time dependent, can be replaced by the number n_e of space independent series, each of size N_e of the time independent series. The variance of r_1, and correspondingly of $\hat{g}(f)$, of nN space-time dependent, observed annual values should be equal to the variance of r_1, and correspondingly of $\hat{g}(f)$, of $n_e N_e$ of space-time independent annual values.

To estimate $n_e N_e$, the r_1 values of all series are needed. Two methods are feasible for the estimation

of r_1 values: (1) To use the lag-zero cross correlation coefficient matrix of all the station series in each area for a given variable; and (2) To use the distribution of r_1 to find its variance, and from this variance the number $n_e N_e$ of space-time equivalent independent series for each area and each variable.

The second approach by using the distribution of r_1 of n regional series to determine n_e seems simpler than the first approach by using the cross correlation matrix of all the series. In using the matrix method, one must compute n(n - 1)/2 values of the lag-zero cross correlation coefficients and average them. In the second approach only the n values of the first serial correlation coefficients r_1 need to be estimated for each region. Both approaches are described herein to show how they should be used, although the second method is used only in presenting the computational results and in obtaining the tolerance limits for the average spectral densities in Chapter IV.

Correlation matrix approach. The simple average $\bar{r}_1$ is computed using

$$\bar{r}_1 = \frac{1}{n} \sum_{j=1}^{n} r_{1,j} \quad , \tag{2-16}$$

in which $r_{1,j}$ = the r_1-value of the j-th station of the sample size N_j. To take into account the different sample size, which determines the information content in the $r_{1,j}$-estimates, a weighted mean $\overline{r_1^*}$ may be used, with N_j - 1 as the weights, so that

$$\overline{r_1^*} = \frac{\sum_{j=1}^{n} (N_j - 1) r_{1,j}}{\sum_{j=1}^{n} (N_j - 1)} \tag{2-17}$$

Similarly, the variances of $r_{1,j}$'s are computed either by using the weights N_j - 1 and r_1^*, or by using the average sample size as N_a - 1 and $\bar{r}_1$, in the general equation

$$\text{var } \bar{r}_1 = \frac{1}{n_e} \text{ var } r_{1,j} \tag{2-18}$$

or

$$\text{var } \overline{r_1^*} = \frac{1}{n_e} \text{ var } r^*_{1,j} \quad . \tag{2-19}$$

The correlation coefficient ρ_{ij} between the first serial correlation coefficients, $r_1(x)$ and $r_1(y)$, of the two series x and y, are given by [2, page 11]

$$\rho_{ij} = \frac{N+2}{N-1} \rho_{xy}^2 \left\{ 1 + \frac{N(N-2)[r_1^2(x) + r_2^2(y)]}{2(N+2)\quad(N+4)} \right\}, \tag{2-20}$$

in which N = the sample size and ρ_{xy} = the lag-zero correlation coefficient between the x and y series. The correlation coefficient ρ_{ij} is estimated by the sample value r_{ij} obtained in replacing ρ_{xy} by r_{xy} in Eq. 2-20. The variance of $\bar{r}_1$ is then

$$\text{var } \bar{r}_1 = \frac{1}{n^2} \left[\sum_{j=1}^{n} \text{var } r_{1,j} + 2 \sum_{j=1}^{n} \sum_{i=j+1}^{n} \text{cov } r_{1,j}\, r_{1,i} \right], \tag{2-21}$$

which gives

$$\text{var } \bar{r}_1 = \frac{\text{var } r_1}{n} [1 + \bar{r}_{ij}(n - 1)] \quad , \tag{2-22}$$

with

$$\bar{r}_{ij} = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{i=1}^{n} r_{ij} \quad , \tag{2-23}$$

in which r_{ij} = the estimate of ρ_{ij} of Eq. 2-20. For var $\bar{r}_1$ = var r_1/n_e, Eq. 2-22 gives

$$n_e = \frac{n}{1 + \bar{r}_{ij}(n - 1)} \quad . \tag{2-24}$$

Approach by determining the variance of r_1 from its frequency distribution. The variance of r_1, estimated in the open-series approach, with the estimated mean and variance of a normal variable, is [2]

$$\text{var } r_1 = \frac{N_o^3 - 3N_o^2 + 4}{N^2(N_o^2 - 1)} \quad , \tag{2-25}$$

in which N_o = the length of a unique series which will have the same variance of r_1 as the n space-time dependent series of a region. From Eq. 2-25 a value N_o can be obtained which should always be greater than either N_a, the average size of n series, or N, the size of n series of equal length.

$$N_o = n_e N_e \quad . \tag{2-26}$$

with n_e = the equivalent number of independent series in the region, and N_e = the mean sample size of all n series, for N_e to correspond to independent series, then

$$\text{var } r_1 = \frac{(n_e N_e)^3 - 3(n_e N_e)^2 + 4}{(n_e N_e)^2 [(n_e N_e)^2 - 1]} \quad . \tag{2-27}$$

Given N_e as the effective average length of n series in an area, then Eq. 2-27 permits the computation of n_e, as the effective number of independent series in the region.

The value $N_o = n_e N_e$ represents the sample size of a unique independent series for the determination of tolerance limits for the average spectral variance densities.

A still simpler way to determine $n_e N_e$ is by using the Fisher z-transforms of r_1 values as

$$z = \frac{1}{2} \ln \frac{1 + r_1}{1 - r_1} \tag{2-28}$$

The n values of z give the variance of z by [2]

$$\text{var } z = \frac{1}{n_e N_e} \quad , \tag{2-29}$$

or

$$n_e = \frac{1}{N_e \text{ var } z} \quad . \tag{2-30}$$

This simple method is used in this paper to determine the $n_e N_e$ values for the 24 cases (four variables and six areas), with the assumption that r_1's are so small that the difference between N_e and either N_a (average sample size) or N (individual sample sizes) may be neglected. In general, an approximation to N_e for the first-order Markov model is

$$N_e = \frac{N(1 - r_1^2)}{(1 + r_1^2)} \quad . \tag{2-31}$$

For $r_1 = 0.20$, $N_e = 0.90$ N. For very small r_1 (say $r_1 \leq 0.10$, $N_e \approx N_a$; for $r_1 > 0.10$, Eq. 2-31 should be used to find N_e, and then n_e can be found from Eq. 2-30.

Determination of tolerance limits. The distribution of individual estimates, $\hat{g}(f)$, is chi-square with EDF = the number of effective degrees of freedom. The mean values of $\hat{g}(f)$, determined by

$$\overline{\hat{g}(f)} = \frac{1}{n} \sum_{i=1}^{n} \hat{g}_i(f) \quad , \tag{2-32}$$

may be approximated by the normal distribution, when EDF of $\sum \hat{g}_i(f)$ is at least 30. Then the variance of the mean of g(f) becomes

$$\text{var } \overline{\hat{g}(f)} = \frac{\text{var } \hat{g}(f)}{n_e} = \frac{8}{\text{EDF} \cdot n_e} \quad . \tag{2-33}$$

If m = N/4 is chosen, then EDF = 3.7 x 4.0 = 14.8 when the distribution is close to normal. If the Parzen's smoothing function, Eq. 2-9, is used then

$$\text{var } \overline{\hat{g}(f)} = \frac{0.54}{n_e} \quad , \tag{2-34}$$

with the 95 percent tolerance limits of standardized normal variables ($t \pm 1.96$) given by

$$g_{1,2}(f) = \gamma(f) \pm t\, s_g \quad , \tag{2-35}$$

or

$$g_{1,2}(f) = 2.00 \pm \frac{1.44}{\sqrt{n_e}} \quad . \tag{2-36}$$

Tolerance limits of Eq. 2-36 should be used in the analysis of the average spectral graphs in further investigations only for the central ordinates of estimated $\overline{\hat{g}(f)}$, while for their end ordinates of spectral density graphs the corrections (the larger tolerance limits) are:

$$g_{1,2}(f) = 2.00 \pm \frac{1.44\ \sqrt{2}}{\sqrt{n_e}} \quad , \tag{2-37}$$

because of a further loss of degrees of freedom in the estimation of end densities.

Chapter III
RESEARCH DATA ASSEMBLY

This chapter refers to the selection of variables in the study of fluctuations of wet and dry years, and particularly to the division of the United States (excluding Alaska and Hawaii) into six areas.

3-1 Selection of Variables

As shown in Introduction, four variables of annual series are investigated:

(1) P_1, the series of annual precipitation, inferred by the analysis of data to be consistent (no apparent systematic errors) and homogeneous (negligible man-made influences or natural accidental disruptions);

(2) Q_1, the series of annual runoff, selected by criteria described in papers of the same title [1,2] as this paper;

(3) P_2, the series of annual precipitation, inferred to be either inconsistent or nonhomogeneous; and

(4) Q_2, the effective annual precipitation, obtained from the Q_1 series by $P_1 - Q_2 = Q_1 + \Delta W$, or $Q_2 = P - Q_1 + \Delta W$, where ΔW = the change at the end of each water year in the total water stored in a river basin above the gauging station of Q_1 [1].

Data selected for this study covers the continental USA except Alaska. Basically all the precipitation and runoff stations of longest record satisfy the prescribed selection criteria [1]. Sample sizes vary from N = 35 to N = 150 for annual precipitation P_1 and P_2 series, and N = 30 to N = 97 for annual runoff Q_1 series and effective precipitation Q_2 series. The criteria used in selecting the P_1 and Q_1 series for the Western United States in Reference [1] were extended for the selection of series in the Eastern United States.

3-2 Selection of Six Investigation Areas

The maps of the average annual precipitation, mean annual lake evaporation and the average annual runoff, published in the Water Atlas of the United States, by the Water Information Center, Inc., were used to deliniate the six areas according to precipitation, evaporation, and runoff characteristics. The basic criterion in separating these areas was to have within each area the approximately similar climatic conditions, though the orographic local differences made it difficult to carry out this criterion consistently for each area. Figures 3-1, 3-2, and 3-3 show these six areas.

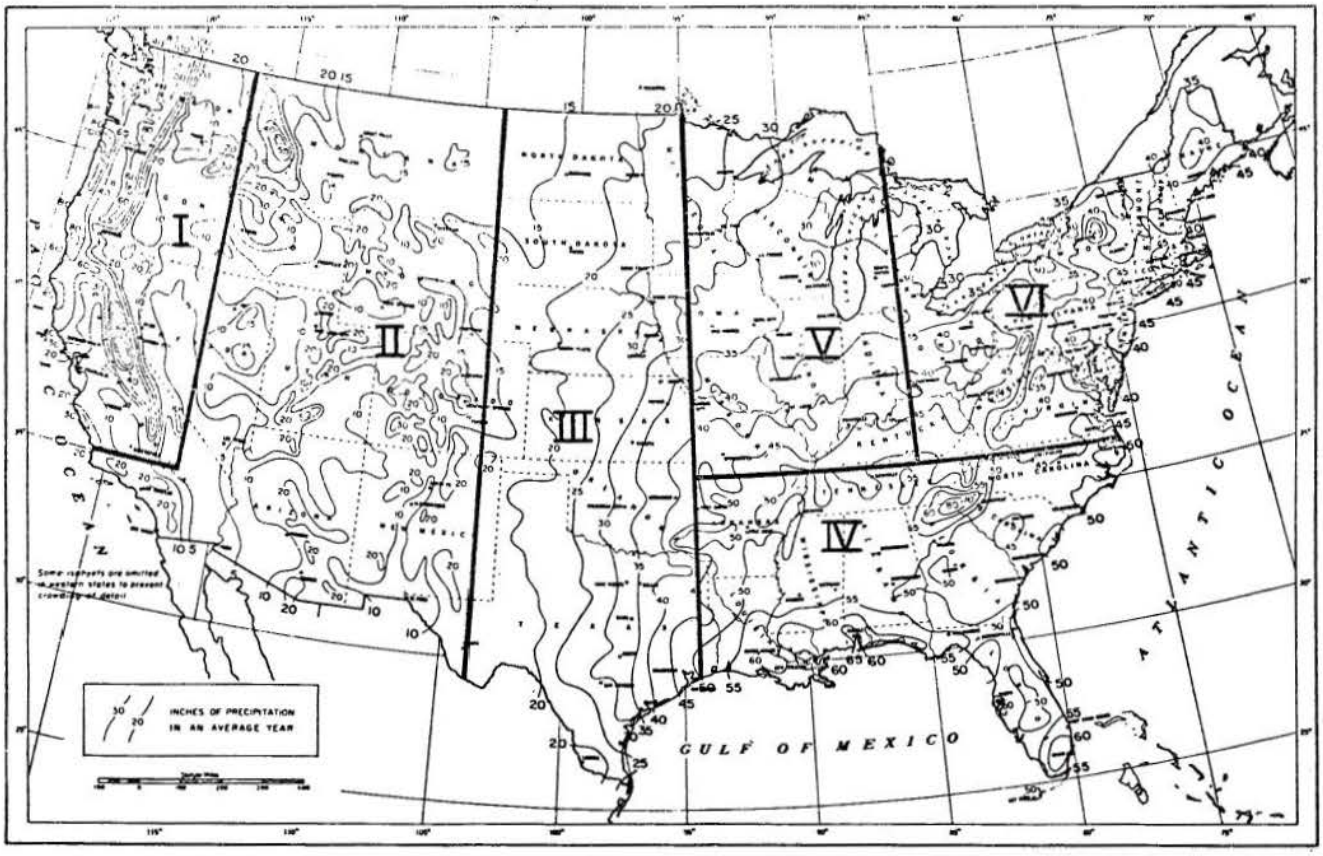

Fig. 3-1. Average Annual Precipitation for the United States Based on 40-year Period. (After U.S. Department of Agriculture, "Climates of the United States.")

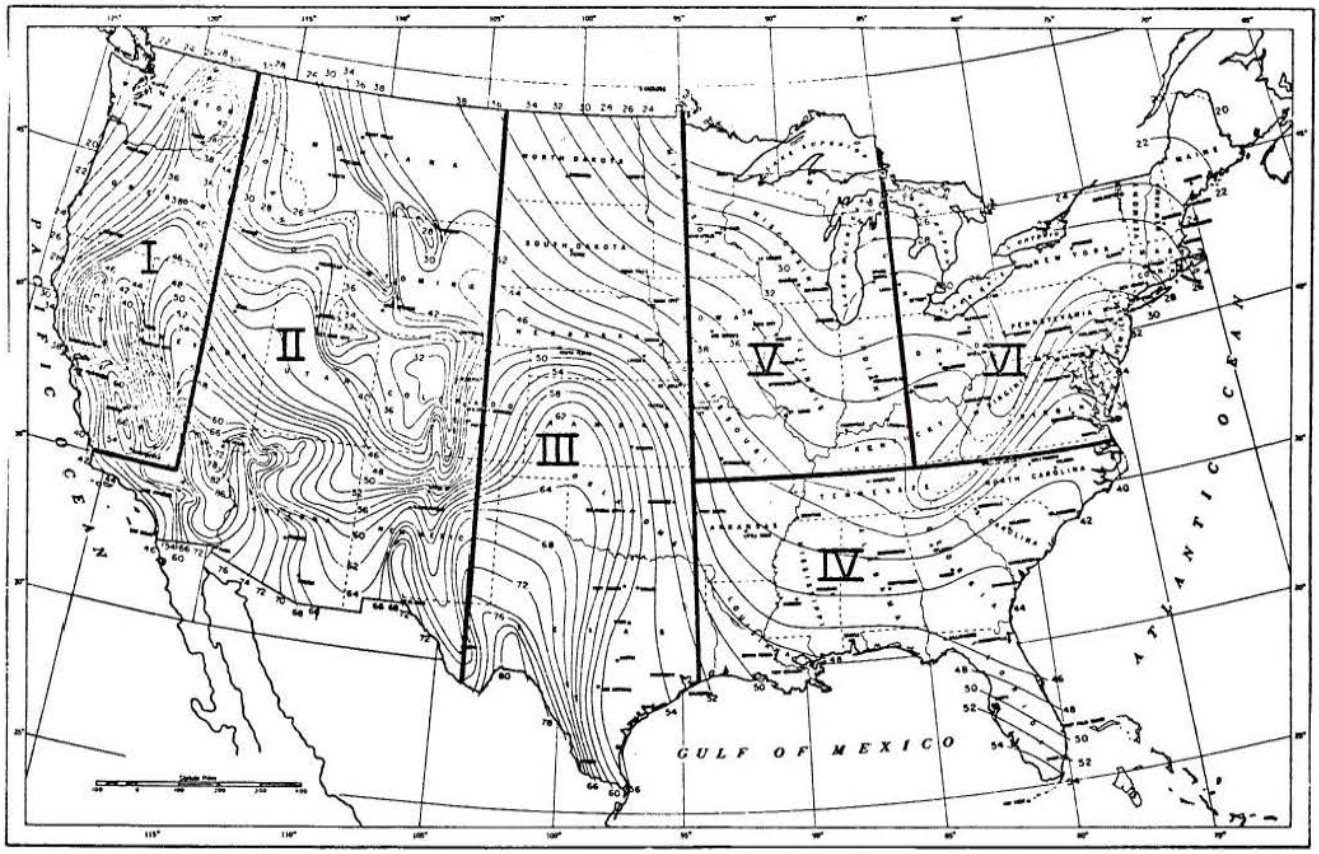

Fig. 3-2. Mean Annual Lake Evaporation for the United States. Lines Show Mean Annual Lake (free-water) Evaporation in Inches Based on Period 1946-1955. (After "U.S. Weather Bureau Technical Paper 37.")

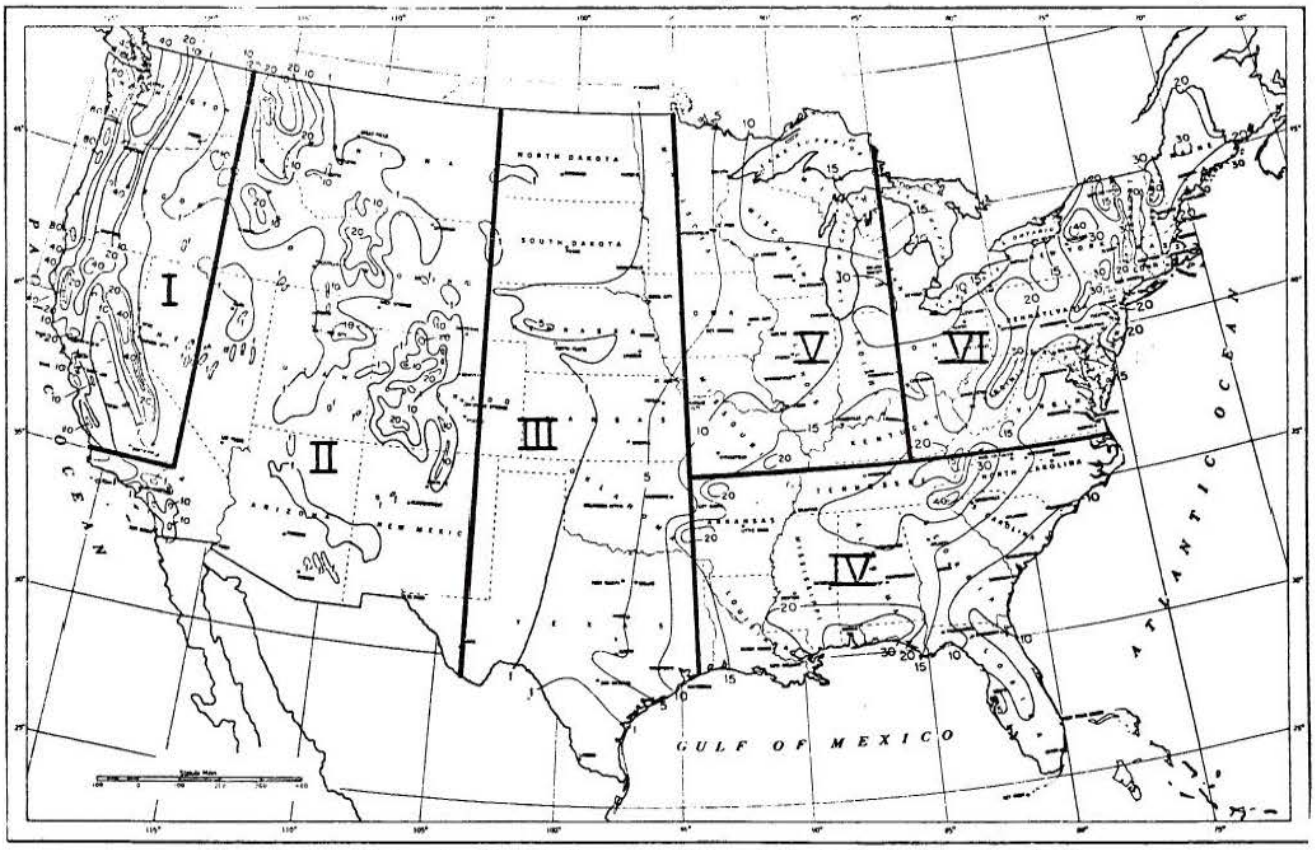

Fig. 3-3. Average Annual Runoff for the United States. (After U.S. Geological Survey.) Lines Show Average Annual Runoff in Inches. The 5-, 15-, and 30-in. Runoff Lines have been Omitted in Western United States to Prevent Crowding of Map Detail.

3-3 Brief Description of Areas

Area I covers USA from the West Coast to longitude 117°, and from latitude 35° to the Canadian border. Area II covers the western part of the US, from longitude 117° to longitude 104°, and from Mexico to Canada (basically the dry areas west of the Rocky Mountains, the Rocky Mountains, and part of the plateau east of them). Area III covers the USA between the longitudes 104° and 94° (basically the central part of USA along the broad valleys of the Mississippi and Missouri Rivers and some of their major west tributaries). Area IV covers the Southeast of USA between longitude 94° and the Atlantic Ocean, and the Gulf of Mexico and latitude 36.5°. Area V covers the north-central part of USA between longitudes 94° and 85° and between latitude 36.5° and the Canadian border. Area VI covers the northeastern part of USA, east of longitude 85° and north of latitude 36.5°.

The basic characteristics of annual precipitation (mean, minimum, and maximum) and annual runoff (minimum and maximum) for the six areas given in Tables 3-1 and 3-2, respectively.

Table 3-1. Characteristics of Annual Precipitation of Six Areas

Area No.	Mean Annual Precipitation in Inches	Minimum In.	Minimum Location	Maximum In.	Maximum Location
I	23.8 in.	5	Southeast	140	Northwest
II	14.4 in.	5	Southwest	50	Northwest
III	23.6 in.	15	Northwest	45	Southeast
IV	51.1 in.	45	North & South	80	North
V	34.9 in.	25	Northwest	50	South
VI	41.1 in.	30	Northwest	50	North, East & South

Table 3-2. Characteristics of Annual Runoff of Six Areas

Area No.	Minimum Inches	Minimum Location	Maximum Inches	Maximum Location
I	1	Central	80	West Coast
II	1	Central	40	West
III	1	Central	20	East
IV	5	South	40	Northeast
V	5	West	20	South
VI	10	Northwest	40	Northeast

3-4 Reasons for Using Areas

The first objective of this paper is to investigate whether it can be reasonably inferred that basic hydrologic variables of annual precipitation and annual runoff are *temporary stationary stochastic processes*. *Temporary stationarity* is conceived of in this text as stationarity in processes extending only about 150-200 years both into the past and into the future from the present. It is not considered valid for longer periods. By investigating six different areas of USA, rather than the total USA area, it is felt that a better insight could be obtained as to whether differences between the areas can be ascribed only to the inevitable sampling fluctuations resulting from the limited sample sizes of the series. Furthermore, it is hoped to demonstrate that the differences in climate, such as humid, semi-humid, semi-arid and arid climates, do not significantly affect the basic conclusions about this temporary stationarity.

The second objective of the study is hopefully to demonstrate that the annual precipitation process is an independent, temporarily stationary process, or very close to it, in the above sense of *temporary stationarity*. The division of USA into six areas, approximately based on the general climate, should answer the question whether the type of climate influences the degree of closeness to the independent, temporarily stationary process. Several other aspects of hydrologic stochastic processes of annual values may be investigated by comparing their properties for stations inside large but adjacent areas.

3-5 Preparation of Data on Tapes

Annual series data for each area and for each of the variables: the homogeneous precipitation (P_1), the runoff (Q_1), the nonhomogeneous precipitation (P_2), and runoff corrected for carryover (Q_2) as the effective precipitation, were taken from the existing magnetic tapes at Colorado State University for each of these variables, and separately for the West and the East of the United States. The data was first examined by Special Programs to check whether there had been any change in the location of any station. Then all stations were classified into their six geographical areas, as described in Section 3-3 of this chapter, and series of four variables were recorded on six new magnetic tapes, one for each area, with a sequence of four segments: P_1, Q_1, P_2, Q_2 on each tape. The number of annual series in each area for each variable is given in Table 3-3.

Table 3-3. Number of Stations Series for Each Variable and Each Area

Variable	AREA I	II	III	IV	V	VI
P_1	239	380	444	222	231	343
Q_1	166	156	85	78	88	175
P_2	77	132	204	175	155	176
Q_2	166	156	85	79	89	173

Most of the data were updated to the year 1965, and a few corrections in location were made.

Chapter IV
ANALYSIS OF COMPUTATIONAL RESULTS

The computational results are analyzed in this Chapter for the average variance density spectra of investigated four series P_1, P_2, Q_1, Q_2, and for the six areas in the United States. The variance densities for eleven frequencies (0, 0.05, 0.10, ..., 0.45, and 0.50) are computed for each individual series. Then, the n variance densities for the n areal series, for each frequency, are averaged to produce the 24 average variance density spectra (four variables multiplied by six areas).

Before the average spectra are analyzed for each of the four variables (P_1, P_2, Q_1, Q_2), and in turn for each of the six areas (Nos. I, II, III, IV, V, VI), the auto-correlation and cross-correlation properties of these 24 cases are presented and discussed. This was considered necessary for determining the product ($n_e N_e$) of the number of equivalent, spacially independent stations (n_e), multiplied by the effective time independent sample size (N_e), for each area and in turn for each variable. The values n_e and N_e then serve for the various computations but particularly of tolerance limits for the average variance density spectra.

4-1 Determination of the Effective Numbers of n_e and N_e

Frequency histograms of the Fisher z_1 transforms of the estimated first serial correlation coefficient, r_1, are given in Fig. 4-1. For each of the four

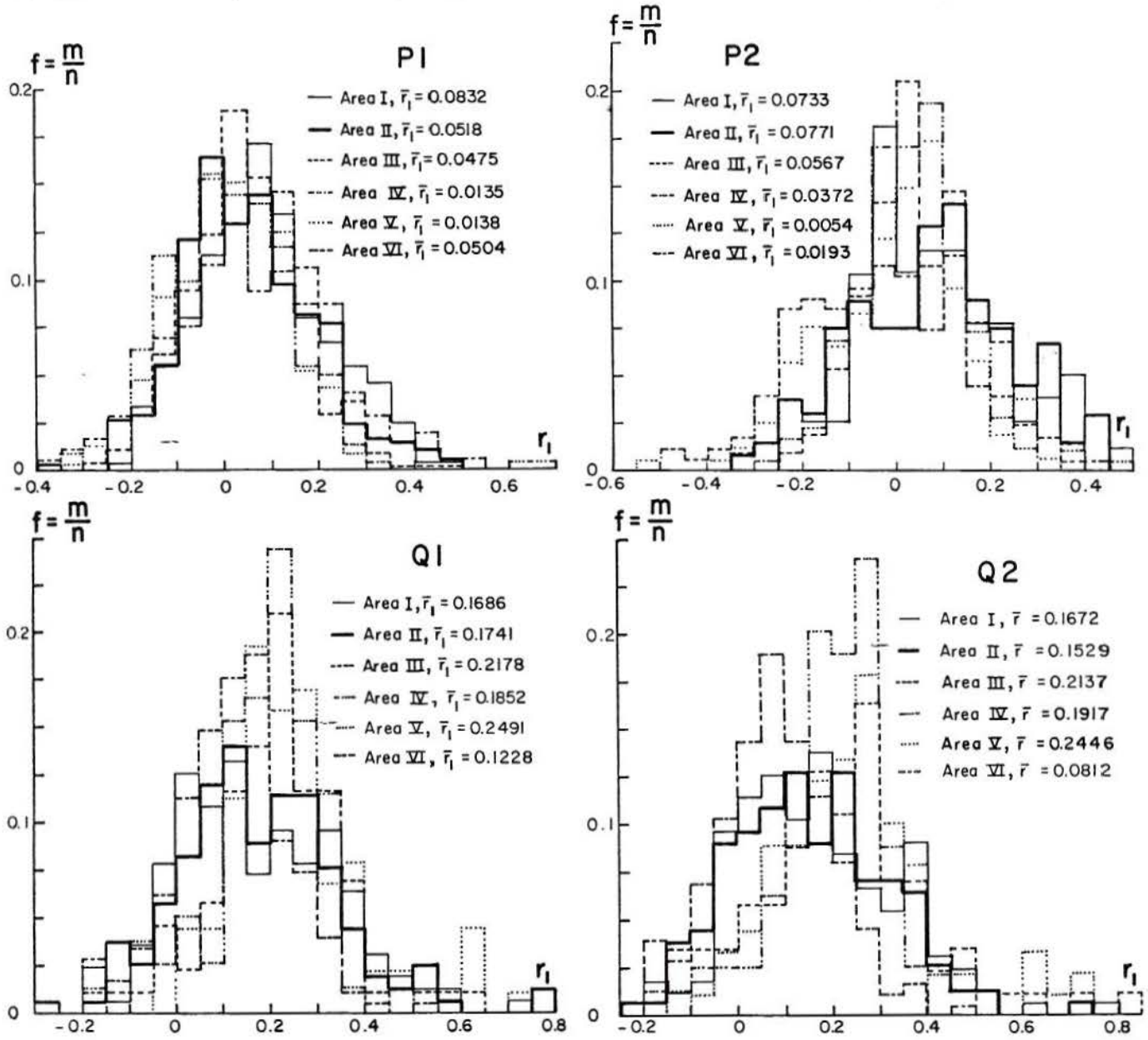

Fig. 4-1. Frequency distribution histograms of the first serial correlation coefficients, r_1, of the four variables P_1, P_2, Q_1, Q_2 (the four graphs), and in turn a separate frequency histogram for each of the six areas of the United States (I, II, III, IV, V, VI).

variables P_1, P_2, Q_1 Q_2, there are six histograms, one for each area. The r_1-values are first estimated, then transformed into the Fisher z_1-variable values by using Eq. 2-28.

Two approaches can be used in practice to determine the standard deviations of frequency distributions of the z_1-variables:

(1) A direct computation of var z_1 from the n values of each variable: P_1, P_2, Q_1, Q_2, and in turn for each of the six areas. This approach is used in this study to compute $s_z = \sqrt{\text{var } z_1}$.

(2) An indirect computation by the graphical estimation, in plotting the frequency curves of z_1 in Cartesian-probability scales, in drawing by a visual inspection through the plotted points, the straight lines and in finding the standard deviation s_z. This approach is not used in this study, however. Since the z_1-variables are normally distributed, the straight line fits to the plotted frequency distribution points (usually to the points between 10% and 90% of frequencies) enable the estimation of the standard deviations s_z of the z_1-transforms for the 24 variables. For the probabilities on the straight lines of 84.13% and 15.87%, the differences between their z_1 values give $2s_z$, with var $z_1 = s_z^2$.

By using Eq. 2-29 , $n_e N_e$ is obtained for a given var z_1. For $N_e = N_a$ (this is the case for the average sample size of n series, if r_1 is very small, say $r_1 < 0.10$), then $n_e = (n_e N_e)/N_a$. For $r_1 > 0.10$, Eq. 2-31 gives N_e for $N = N_a$, and then n_e is computed. Table 4-1 presents the results for the four variables (P_1, P_2, Q_1, Q_2) and in turn for each of the six areas. These estimates are then used to determine the tolerance limits of $\bar{z}_1$.

4-2 Tolerance Limits for the $\bar{z}_1$-Transforms of the Average First Serial Correlation Coefficients

With the r_1-values transformed into the Fisher z_1-values, with z_1 normally distributed, the expected value of $\bar{z}_1$ and the upper and lower tolerance limits are determined for the estimates of the $\bar{z}_1$-variables.

The expected value of $\bar{r}_1$ is

$$E\bar{r}_1 = -\frac{1}{n_e N_e} , \tag{4-1}$$

with the expected value of $\bar{z}_1$ resulting as

$$E\bar{z}_1 = \frac{1}{2} \ln \frac{1 + E\bar{r}_1}{1 - E\bar{r}_1} \tag{4-2}$$

and the upper and lower tolerance limits for the 95% probability level computed by

$$UTL_z = E\bar{z}_1 + \frac{1.96\, s_z}{n_e} \tag{4-3}$$

and

$$LTL_z = E\bar{z}_1 - \frac{1.96\, s_z}{n_e} \tag{4-4}$$

The four values, $E\bar{r}_1$, $E\bar{z}_1$, UTL_z, and LTL_z, for each of the four variables (P_1, P_2, Q_1, Q_2) and each of the six areas, are given in Table 4-1.

4-3 Results from the Study of the Fisher $\bar{z}_1$-Transforms of the Average First Serial Coefficients, $\bar{r}_1$

The basic results drawn from the numbers in Table 4-1 are presented in Fig. 4-2. The values of $\bar{z}_1$ (the Fisher z_1-transform of the first serial correlation coefficient, r_1) for the four variables (P_1, P_2, Q_1, Q_2) and for the six areas of the United States, are compared with the upper tolerance limits for $\bar{z}_1$, given as UTL_z , for the 95% tolerance probability level. For these 24 cases the positive values of the tolerance intervals are shaded in order to emphasize the differences between the computed $\bar{z}_1$ and UTL_z-values.

Both the homogeneous annual precipitation series (P_1) and the non-homogeneous annual precipitation series (P_2) show the $\bar{z}_1$-values to be mostly above the upper tolerance limits of $\bar{z}_1$, with four out of 12 values either close to these limits (two values of P_1) or below these limits (two values of P_2).

Both the annual runoff series (Q_1) and the effective annual precipitation series (Q_2), obtained directly from the Q_1-series, have in all the 12 cases (two times six areas) the $\bar{z}_1$-values located significantly above the upper tolerance limits, UTL_z .

The P_1- and P_2-series are close to be independent time processes, with the maximum $\bar{z}_1$ being 0.0832 for Area I of the P_1-variable, and the next two highest values of $\bar{z}_1$ being 0.0734 and 0.0772 for Areas I and II of the P_2-variable. The positive values of $\bar{z}_1$ may be variously explained. Among the most important factors are:

(1) Annual values are obtained by cutting the precipitation process at a given date, with the daily precipitation of the previous and the succeeding days to that date being time dependent process.

(2) Inconsistency and non-homogeneity in data may be unavoidable, and as such imbedded in the annual series. Various changes in observations (definition of trace values of precipitation, changes in the vertical and horizontal gauge positions, in instruments, in the methods of determining the catch of precipitation, in gauge surroundings, etc.) produce these non-homogenety and inconsistency (systematic errors). They tend to increase their $\bar{r}_1$ values.

The Q_1- and Q_2-series are, on the average, dependent time series. The basic reasons for the positive auto-correlation, with the $\bar{z}_1$-values ranging between 0.0814 (Area VI of Q_2) to 0.2545 (Area V of Q_1), are:

(1) Changes in water storage capacities in river basins from year to year, so that evaporation,

Table 4-1. Properties of the Average First Serial Cross-Correlation Coefficient, $\bar{r}_1$, and its transform, $\bar{z}_1$, for Four Variables (P_1, P_2, Q_1, Q_2) and Six Areas of USA

Variable	Parameter	AREA WITHIN USA					
		I	II	III	IV	V	VI
P_1	n	239	380	444	222	231	345
	N_e	58.1	54.4	60.9	62.7	60.8	57.5
	$\bar{r}_1$	0.0832	0.0518	0.0475	0.0135	0.0138	0.0504
	$\bar{z}_1$	0.0834	0.0519	0.0475	0.0135	0.0138	0.0504
	$n_e N_e$	285.8	494.2	467.6	437.0	388.5	508.5
	s_z	0.05946	0.04512	0.04640	0.04800	0.05093	0.04448
	n_e	4.92	9.08	7.68	6.97	6.39	8.84
	$E\bar{r}_1$	-0.00351	-0.00202	-0.00213	-0.00229	-0.00257	-0.00197
	$E\bar{z}_1$	-0.00350	-0.00202	-0.00214	-0.00229	-0.00257	-0.00197
	UTL_z	+0.01825	+0.00772	+0.00970	+0.01121	+0.01305	+0.00790
	LTL_z	-0.02525	-0.01176	-0.01398	-0.01579	-0.01819	-0.01183
P_2	n	77	132	204	175	155	176
	N_e	56.4	52.3	59.1	59.9	61.2	63.4
	$\bar{r}_1$	0.0733	0.0771	0.0567	0.0372	-0.0054	-0.0193
	$\bar{z}_1$	0.0734	0.0772	0.0568	0.0372	-0.0054	-0.0193
	$n_e N_e$	227.6	470.0	352.5	364.6	377.3	514.2
	s_z	0.06673	0.04627	0.05349	0.05259	0.05169	0.04423
	n_e	4.04	8.99	5.96	6.09	6.16	8.11
	$E\bar{r}_1$	-0.00439	-0.00213	-0.00284	-0.00274	-0.00265	-0.00194
	$E\bar{z}_1$	-0.00439	-0.00213	-0.00284	-0.00274	-0.00265	-0.00194
	UTL_z	+0.02798	+0.00796	+0.01475	+0.01418	+0.01380	+0.00874
	LTL_z	-0.03677	-0.01221	-0.02043	-0.01867	-0.01910	-0.01263
Q_1	n	166	156	85	78	88	175
	N_e	40.4	41.8	38.0	37.5	39.5	43.4
	$\bar{r}_1$	0.1686	0.1741	0.2178	0.1852	0.2491	0.1228
	$\bar{z}_1$	0.1702	0.1759	0.2214	0.1874	0.2545	0.1234
	$n_e N_e$	130.9	247.1	214.7	146.1	133.1	183.1
	s_z	0.08843	0.06401	0.06873	0.08361	0.08768	0.07452
	n_e	3.24	5.91	5.65	3.89	3.37	4.22
	$E\bar{r}_1$	-0.00764	-0.00405	-0.00466	-0.00684	-0.00751	-0.00546
	$E\bar{z}_1$	-0.00764	-0.00405	-0.00466	-0.00684	-0.00751	-0.00546
	UTL_z	+0.04586	+0.01718	+0.01919	+0.03528	+0.04348	+0.02915
	UTL_z	-0.06113	-0.02528	-0.02850	-0.04897	-0.05851	-0.04007
Q_2	n	166	156	85	79	89	173
	N_e	36.1	38.0	33.6	33.1	35.3	39.2
	$\bar{r}_1$	0.1672	0.1529	0.2137	0.1917	0.2446	0.0812
	$\bar{z}_1$	0.1688	0.1541	0.2171	0.1941	0.2497	0.0814
	$n_e N_e$	108.4	214.3	182.1	126.8	119.3	173.6
	s_z	0.09738	0.06879	0.07472	0.08988	0.09273	0.07656
	n_e	3.00	5.64	5.42	3.83	3.38	4.43
	$E\bar{r}_1$	-0.00923	-0.00467	-0.00549	-0.00789	-0.00838	-0.00576
	$E\bar{z}_1$	-0.00923	-0.00467	-0.00549	-0.00789	-0.00838	-0.00576
	UTL_z	+0.05440	+0.01924	+0.02153	+0.03811	+0.04539	+0.02811
	LTL_z	-0.07285	-0.02857	-0.03251	-0.05388	-0.06215	-0.03963

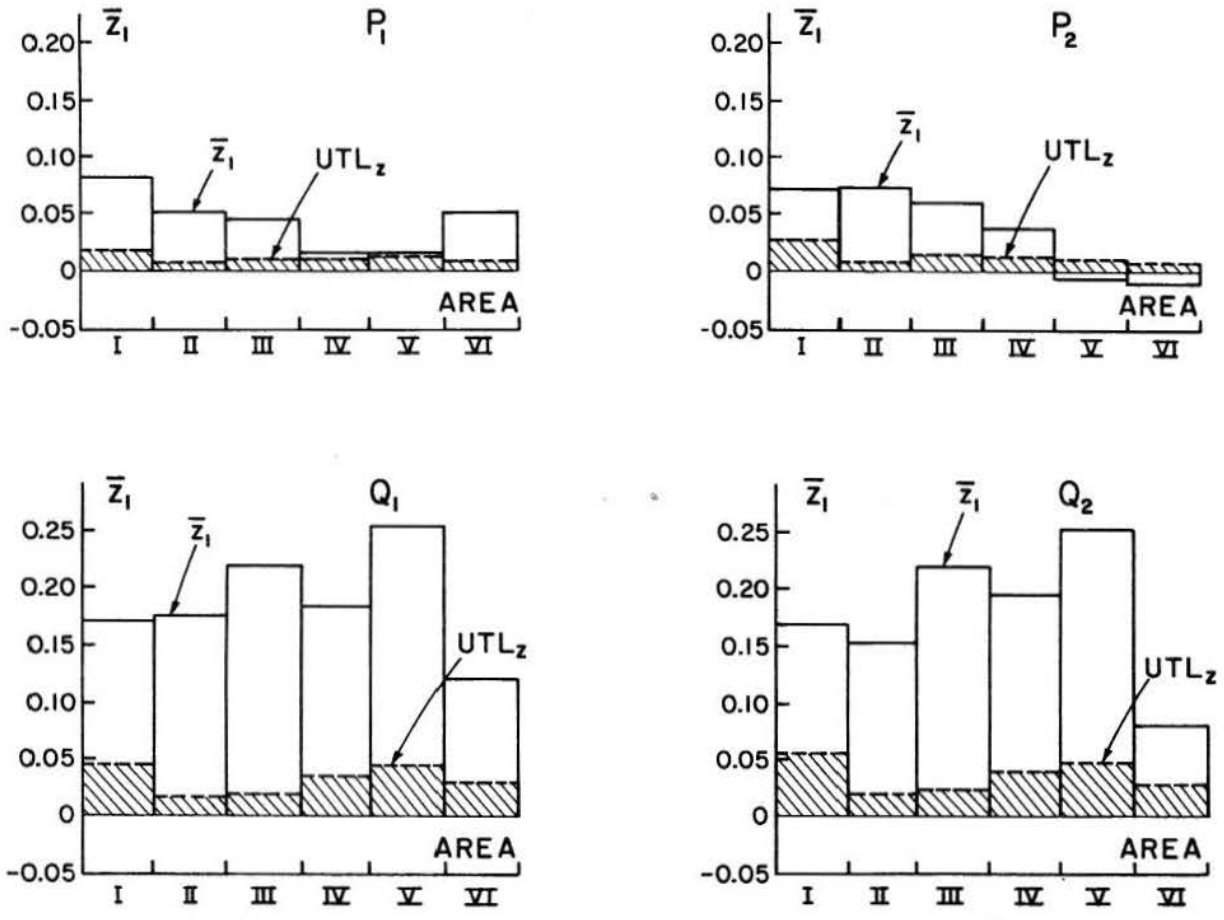

Fig. 4-2. The Average Fisher $\bar{z}_1$-Transforms, for the Estimates of the Average First Serial Correlation Coefficients (heavy lines) of the Four Variables (four graphs) and the Six Areas of the United States, Compared with the Upper Tolerance Limits (dashed lines) at 95% Tolerance Probability Level.

evapotranspiration and runoff process of any year depend on the history of water storage and depletion in previous years.

(2) Same factors, similar as for the precipitation, namely the effect of the year beginning date cutting a continuous dependent process , and of inconsistency and non-homogeneity (man-made changes mostly).

4-4 Computation of Tolerance Limits for the Average Regional Variance Densities for the Time Independent Normal Processes.

By using the n_e-values taken from Table 4-1, the tolerance limits, $g_{1,2}(f)$, of the average spectral densities, $\overline{\hat{g}(f)}$, are computed by Eq. 2-36 for the spectral densities Nos. 2 through 10 (central ordinates of the spectra), and by Eq. 2-37 for the densities Nos. 1 and 11 (the end ordinates of the spectra). Equations 2-36 and 2-37 are applicable under the following conditions:

(1) The $\overline{\hat{g}(f)}$-values are normally distributed around $Eg(f) = 2.00$.

(2) The estimates of the sampling variance of $\hat{g}(f)$, by using the expression: $\text{var}\ \hat{g} = 8/EDF$, with EDF = the effective number of degrees of freedom determined by the approximate Parzen's approach in using $m = N/4$ of the estimated correlogram r_k-values in the computation of the $\hat{g}(f)$-values, are unbiased and sufficiently efficient.

(3) That Eq. 2-37 is a good approximation for the estimates of the end spectral densities.

Since n_e varies between 3.00 (Q_2-variable, Area I) and 9.08 (P_1-variable, Area II) in Table 4-1, and since EDF = 14.8 for the Parzen's smoothing function of estimates of spectral densities, then (EDF n_e) is between 44.4 and 135. Therefore, the normal distribution is applicable for the distribution of $\overline{\hat{g}(f)}$, and for determining its tolerance limits at the 95 percent probability level.

Table 4-2. Tolerance Limits for the Averages of the Estimated Spectral Densities, with UTL_{2-10} and LTL_{2-10}, the Upper and Lower Limits for Central Ordinates, and $UTL_{1,11}$ and $LTL_{1,11}$, the Upper and Lower Limits for the End Ordinates

Variable	Parameter	AREA WITHIN USA					
		I	II	III	IV	V	VI
P_1	UTL_{2-10}	2.649	2.478	2.520	2.545	2.570	2.484
	LTL_{2-10}	1.351	1.522	1.480	1.455	1.430	1.516
	$UTL_{1,11}$	2.918	2.676	2.735	2.771	2.805	2.685
	$LTL_{1,11}$	1.082	1.324	1.265	1.229	1.195	1.315
P_2	UTL_{2-10}	2.716	2.480	2.590	2.583	2.580	2.506
	LTL_{2-10}	1.284	1.520	1.410	1.417	1.420	1.494
	$UTL_{1,11}$	3.013	2.679	2.834	2.825	2.820	2.715
	$LTL_{1,11}$	0.987	1.321	1.166	1.175	1.180	1.285
Q_1	UTL_{2-10}	2.800	2.592	2.606	2.730	2.784	2.701
	LTL_{2-10}	1.200	1.408	1.394	1.270	1.216	1.299
	$UTL_{1,11}$	3.131	2.838	2.857	3.032	3.109	2.991
	$LTL_{1,11}$	0.869	1.162	1.143	0.968	0.891	1.009
Q_2	UTL_{2-10}	2.831	2.606	2.619	2.736	2.783	2.684
	LTL_{2-10}	1.169	1.394	1.381	1.264	1.217	1.316
	$UTL_{1,11}$	3.176	2.857	2.875	3.040	3.107	2.967
	$LTL_{1,11}$	0.824	1.143	1.125	0.960	0.893	1.033

Table 4-2 presents the upper and lower tolerance limits at the 95 percent probability level for the average value $\overline{\hat{g}(f)}$, of the estimated spectral densities, $\hat{g}(f)$, for each f. The UTL_{2-10} and LTL_{2-10} and the limits of the estimated central spectral densities (namely the ordinates Nos. 2, 3, ..., 10), while $UTL_{1,11}$ and $LTL_{1,11}$ are the limits for the end ordinates Nos. 1 and 11.

Figure 4-3 through 4-6 give the averages of the estimated spectral densities for the four variables: P_1, P_2, Q_1, Q_2, and in turn for each of them of the six areas of the United States, as the left graphs of figures, respectively. Each graph contains the tolerance limits at the 95 percent probability level, obtained by Eqs. 2-36 and 2-37. For the Q_1- and Q_2-variables, the curves of the fitted autoregressive models are also plotted, either as the first-order model, AR(1), of Eq. 2-10, or as both AR(1) and the second-order model, AR(2), of Eq. 2-12. Simultaneously, the right graphs of Figs. 4-3 through 4-6 give the corresponding correlograms for up to k = 7, for each of the 24 spectral graphs given on the left side of these four figures.

4-5 Results from the Study of the Average Spectra

The spectra of Figs. 4-3 and 4-4 clearly show that nearly all the $\overline{\hat{g}(f)}$-values, as the averages of n estimated g(f)-values, from the n station series of P_1 and P_2, and for each of the six areas, are confined within the tolerance limits. No spectrum, except that of P_2 for Area II, has density values outside the tolerance limits. It can be concluded from this spectral analysis that the annual homogeneous precipitation series, P_1, and even the annual-nonhomogeneous series, P_2, are independent time series, or at least very close to be independent.

In the analysis of $\bar{z}_1$, only the r_1-estimates of the n stations for each area and each variable were used, while the $\hat{g}(f)$ values are obtained from the N/4 serial correlation coefficients, r_k. Giving a proper emphasis on the spectral analysis, and taking into account the fact that the tolerance limits are wider for the end ordinates of estimated spectra than for the central ordinates, the annual precipitation series may be safely considered as an independent, stationary stochastic process. This is valid at least for the periods of time equal to the lengths of historic samples. No evidence exists in all the spectra of any periodicity, either on the frequency of the sunspot cycle of about 11.3 years, or its double cycle of 22.6 years, or any other cycle. It is evident that neither the inconsistency (nearly always present in a small degree in annual precipitation series) nor the nonhomogeneity, nor the year end dependence of precipitation process, create a sufficient dependence in annual precipitation series to question the general conclusion of the series independence and its stochastic stationarity.

Table 4-3. Coefficient of Determination, R^2, for the Relationships of the Estimated First Serial Correlation Coefficient, r_1, and the Estimates of Mean, $\bar{x}$, Standard Deviation, s_x, Coefficient of Variation, C_v, as well as Relationships Between these Three Estimates

Variable	Relationship vs = versus	R^2 - VALUES FOR AREA I	II	III	IV	V	VI
P_1	r_1 vs $\bar{x}$	0.0798	0.0020	0.0105	0.0191	0.0839	0.0236
	r_1 vs $\bar{x}, s_x$	0.1132	0.0624	0.0145	0.2356	0.0840	0.0350
	r_1 vs $\bar{x}, C_v$	0.1246	0.0352	0.0168	0.2396	0.0839	0.0352
	$\bar{x}$ vs s_x	0.7496	0.4693	0.7588	0.4012	0.7100	0.3661
	$\bar{x}$ vs C_v	0.2404	0.1271	0.0628	0.0187	0.0833	0.0520
P_2	r_1 vs $\bar{x}$	0.1618	0.0001	0.0007	0.0091	0.0489	0.0002
	r_1 vs $\bar{x}, s_x$	0.1686	0.0086	0.0150	0.0946	0.1676	0.0614
	r_1 vs $\bar{x}, C_v$	0.1699	0.0184	0.0184	0.0897	0.1675	0.0633
	$\bar{x}$ vs s_x	0.8301	0.6458	0.7524	0.3459	0.6697	0.3959
	$\bar{x}$ vs C_v	0.3379	0.2071	0.1660	0.0008	0.0606	0.0470
Q_1	r_1 vs $\bar{x}$	0.0068	0.0000	0.0002	0.0053	0.0003	0.0052
	r_1 vs $\bar{x}, s_x$	0.0276	0.0184	0.0128	0.0215	0.0006	0.0055
	r_1 vs $\bar{x}, C_v$	0.0754	0.0309	0.0694	0.0091	0.3335	0.0155
	$\bar{x}$ vs s_x	0.5965	0.6219	0.7284	0.7218	0.6527	0.6986
	$\bar{x}$ vs C_v	0.0270	0.0400	0.1622	0.0000	0.0176	0.2233
Q_2	r_1 vs $\bar{x}$	0.0051	0.0000	0.0018	0.0040	0.0000	0.0058
	r_1 vs $\bar{x}, s_x$	0.0250	0.0185	0.0061	0.0131	0.0154	0.0059
	r_1 vs $\bar{x}, C_v$	0.1679	0.0838	0.0240	0.0042	0.3008	0.0322
	$\bar{x}$ vs s_x	0.6120	0.6212	0.7493	0.7236	0.6475	0.6977
	$\bar{x}$ vs C_v	0.0739	0.0406	0.1850	0.0000	0.0276	0.0285

The spectra of Figs. 4-5 and 4-6 clearly show that they cross the tolerance limits of independent time series for nearly all the areas, with their average spectral density values, $\overline{\hat{g}(f)}$, obtained by averaging the n estimated $\hat{g}(f)$-values for these n stations of the Q_1- and Q_2-variables. These crossings exist for most of the six areas. The spectra of $\overline{\hat{g}(f)}$ lead to the conclusion that the two series of annual runoff and annual effective precipitation are the *red noise* series (with the averages of estimated densities decreasing with an increase of the frequency). In most cases, the fits of the corresponding first-order autoregressive dependence model , AR(1), or in some cases also of the second-order model, AR(2), show to be the attractive mathematical models for the population processes. The correlograms smoothed by the Parzen function produce also the very smooth spectra of the *red noise* type, as opposed to the *white noise* type of spectra for the annual precipitation series.

For all practical purposes, the annual runoff series, as well as the effective annual precipitation (precipitation minus evaporation) series are stationary but dependent stochastic processes. At least, they can be considered as *temporary stationary* for the sample sizes of available data. These Q_1- and Q_2-series do not show any significant periodicity of the sunspot cycle (11.3 years), or its double cycle (22.6 years), or any other cycle.

4-6 Relationships Among the Estimates of Parameters of Four Variables for Six Areas

Table 4-3 gives the coefficient of determination (R^2), either for the simple correlation (for two estimates of parameters and linear simple correlation), or for the multiple correlation (for three estimates of parameters and linear multiple correlation). First, the estimates of the first serial correlation coefficient, r_1, for each series P_1, P_2, Q_1, Q_2, and for each area, are linearly correlated with $\bar{x}$ (the mean of the historic series), then with $\bar{x}$ and s_x (s_x = the standard deviation of historic series), and finally with $\bar{x}$ and C_v (C_v = the coefficient of variation of historic series). The R^2-values are given in Table 4-3 in the first three rows for each of the four variables (P_1, P_2, Q_1, Q_2). Then, the estimates of the mean ($\bar{x}$) are correlated with the estimates of the standard deviation s_x, and then with C_v, and R^2 are given in the forth and fifth rows of each variable in Table 4-3.

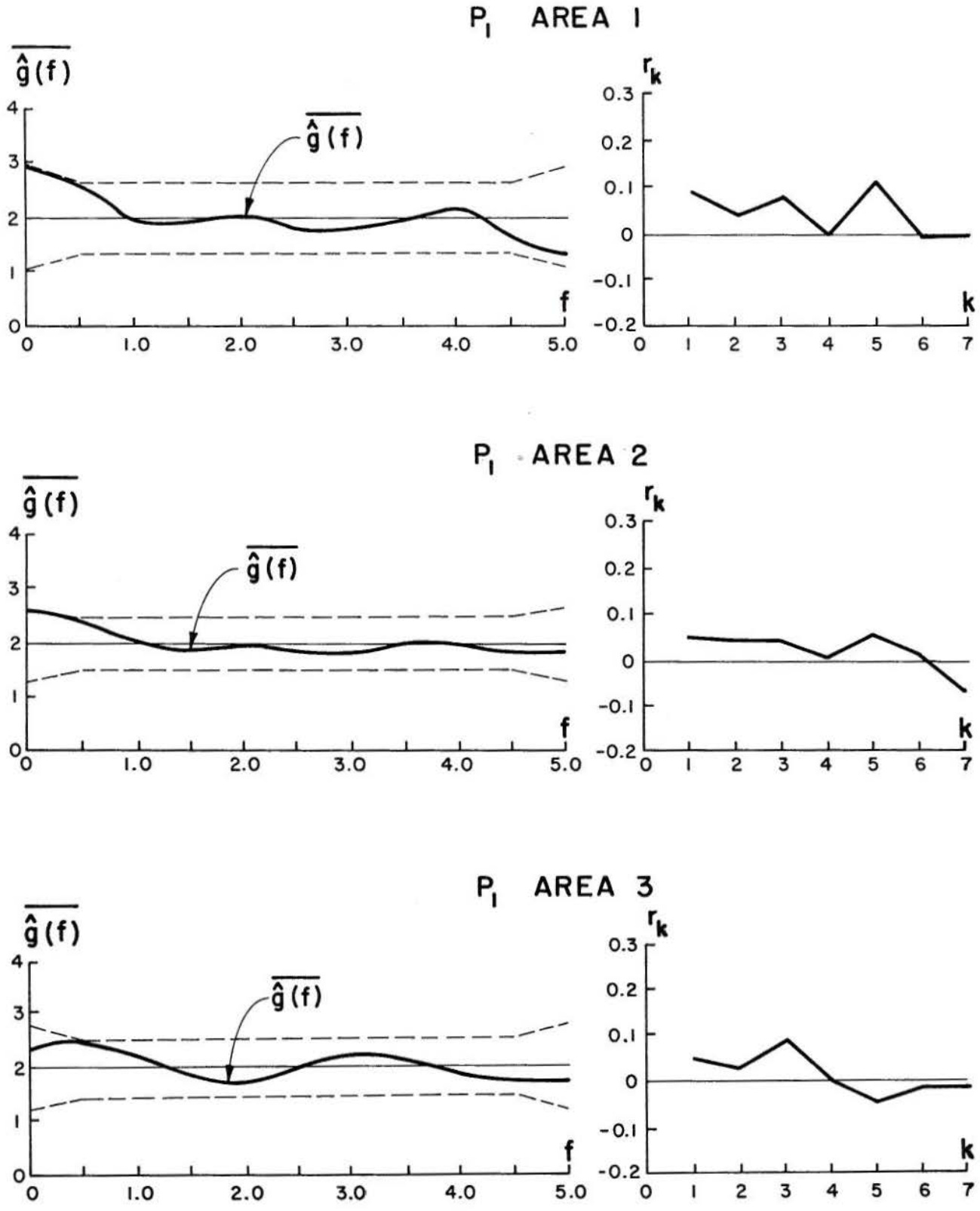

Figure 4-3. Average Spectra (The Full Lines of the Left Graphs) with the 95% Tolerance Limits (The Dashed Lines of the Left Graphs), and the Average Correlograms (The Right Graphs), for the Six Areas (1, 2, 3, 4, 5, 6) of the United States for the n-Series in Each Area of the Annual Homogeneous P_1- Precipitation Series.

First Part of Fig. 4-3. Areas 1, 2, 3.

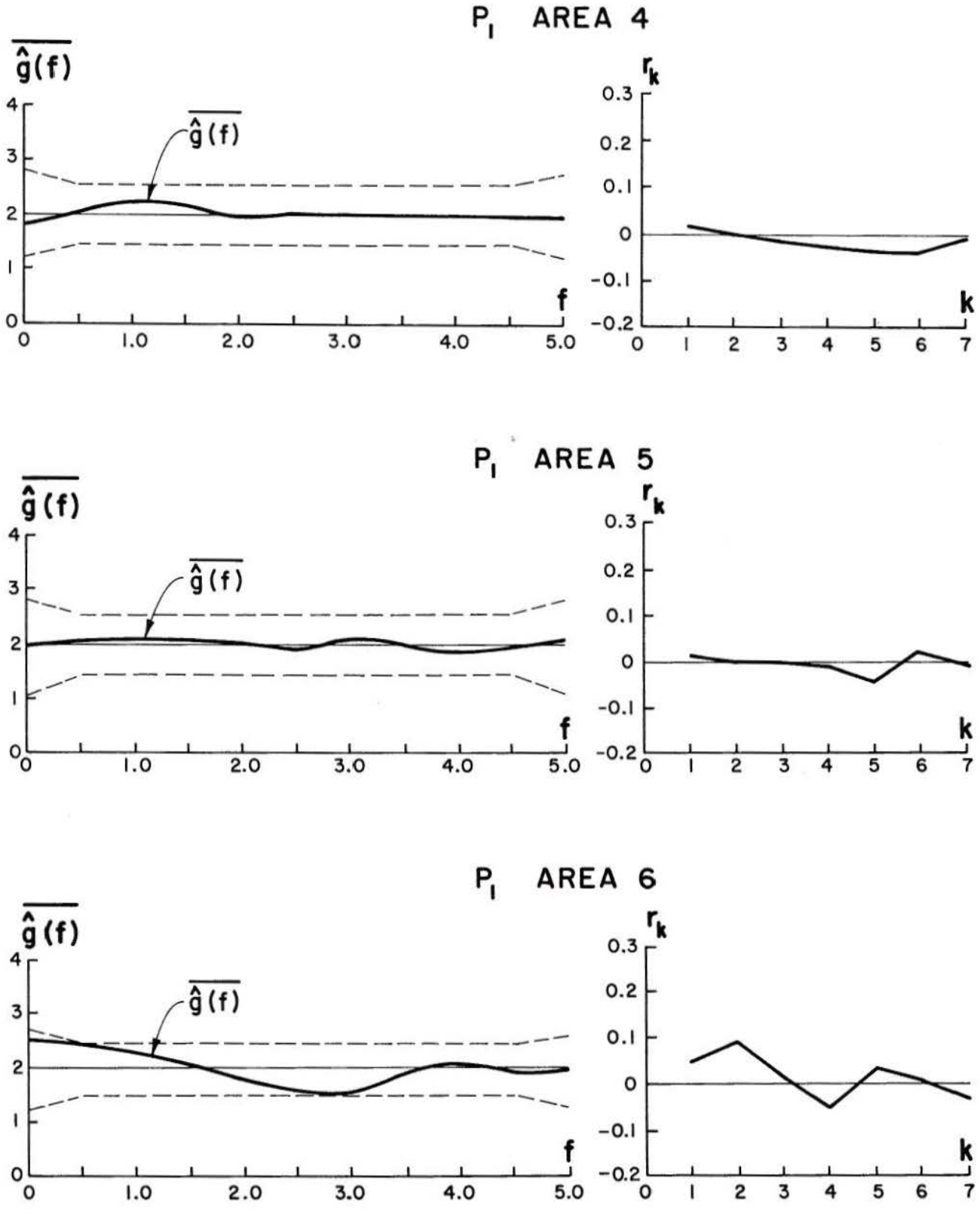

Figure 4-3. Continued. Second Part of Fig. 4-3, Areas 4, 5, 6.

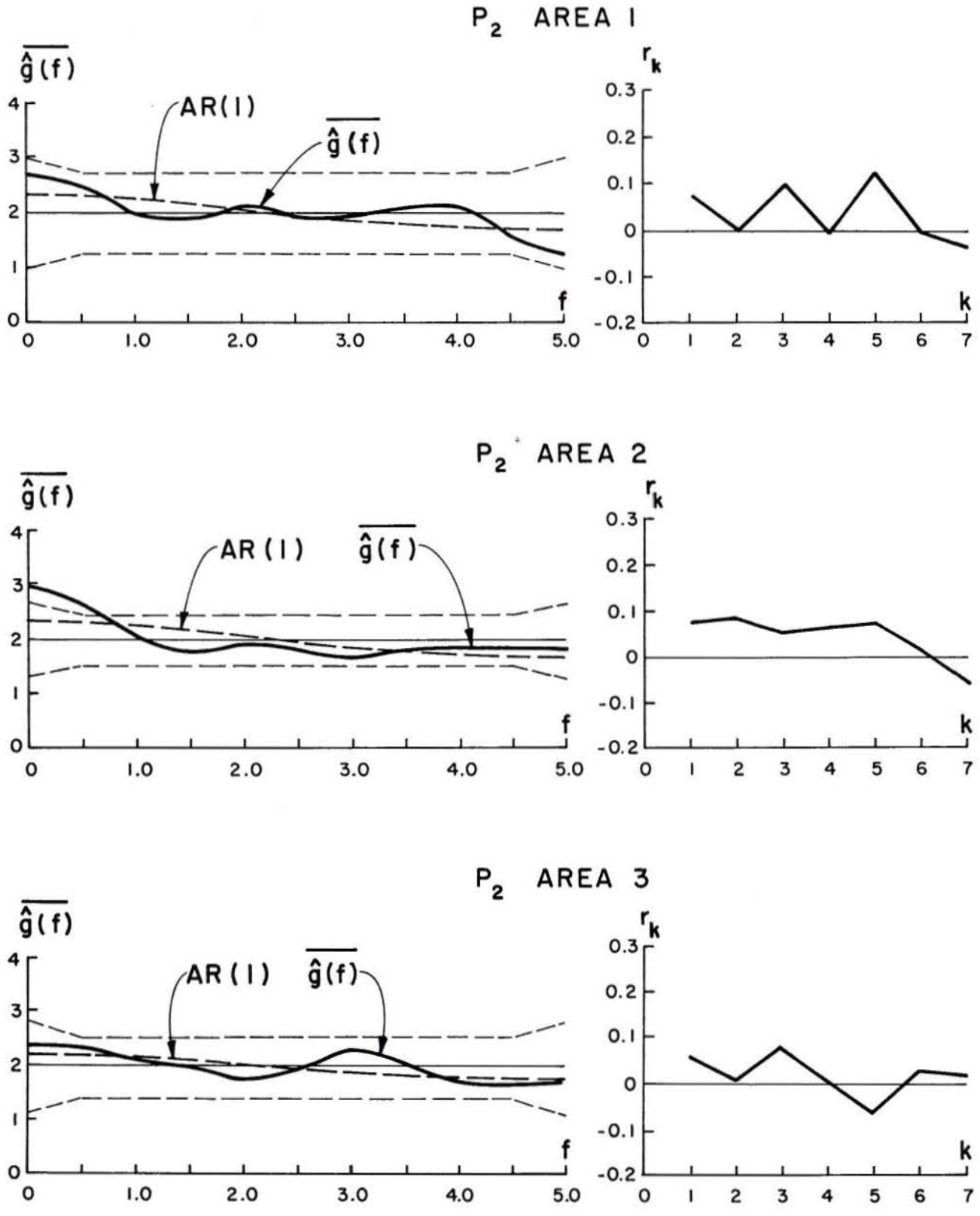

Figure 4-4. Average Spectra (The Full Lines of the Left Graphs), with the 95% Tolerance Limits (The Light Dashed Lines of the Left Graphs), the Fitted First-Order, AR(1), Autoregressive Model, (The Heavier Dashed Lines of the Left Graphs), and the Average Correlograms (The Right Graphs) for the Six Areas (1, 2, 3, 4, 5, 6) of the United States for the n-Series in Each Area of the Annual Non-Homogeneous P_2-Precipitation Series.

First Part of Fig. 4-4, Areas 1, 2, 3.

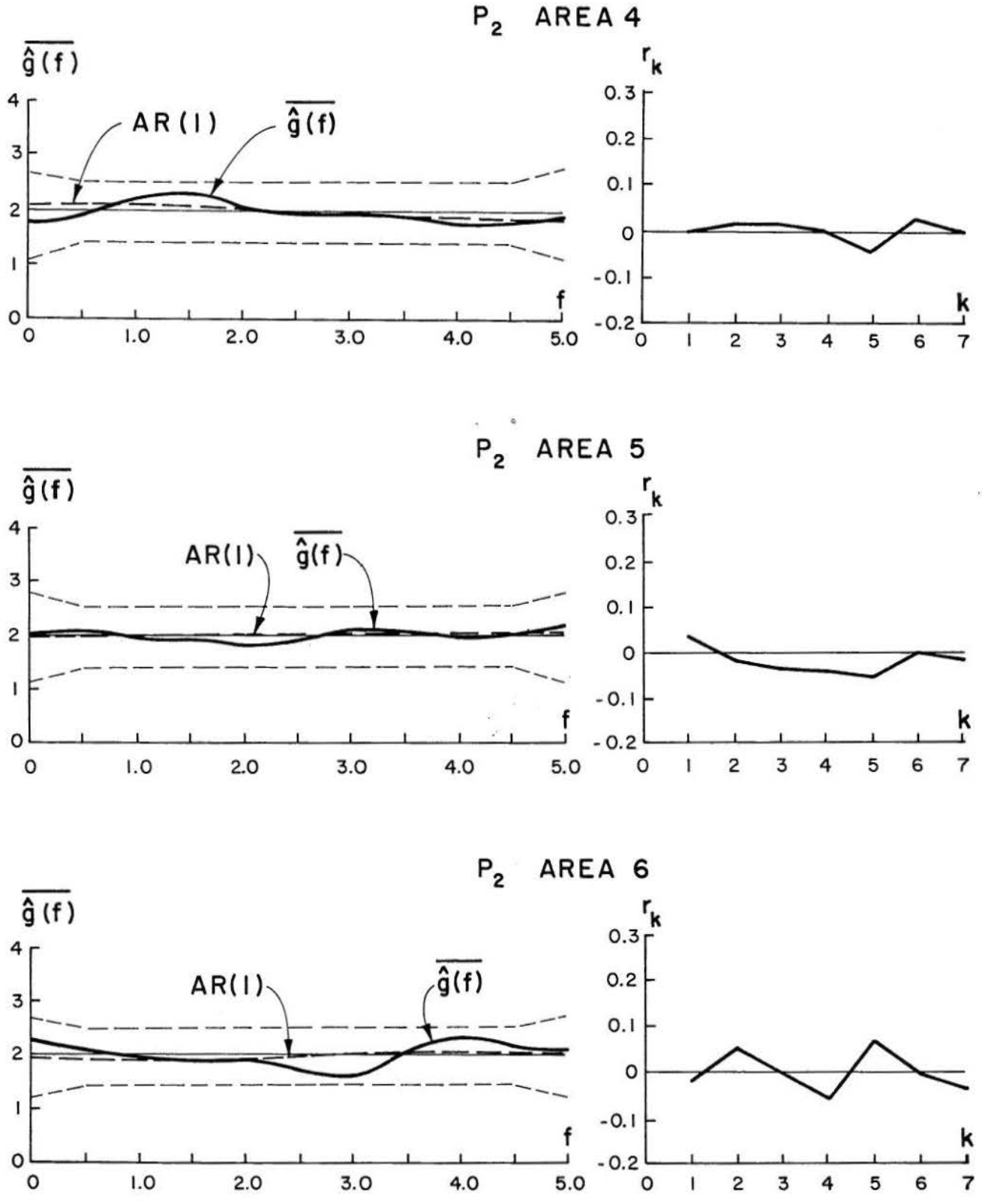

Figure 4-4. Continued. Second Part of Fig. 4-4, Areas 4, 5, 6.

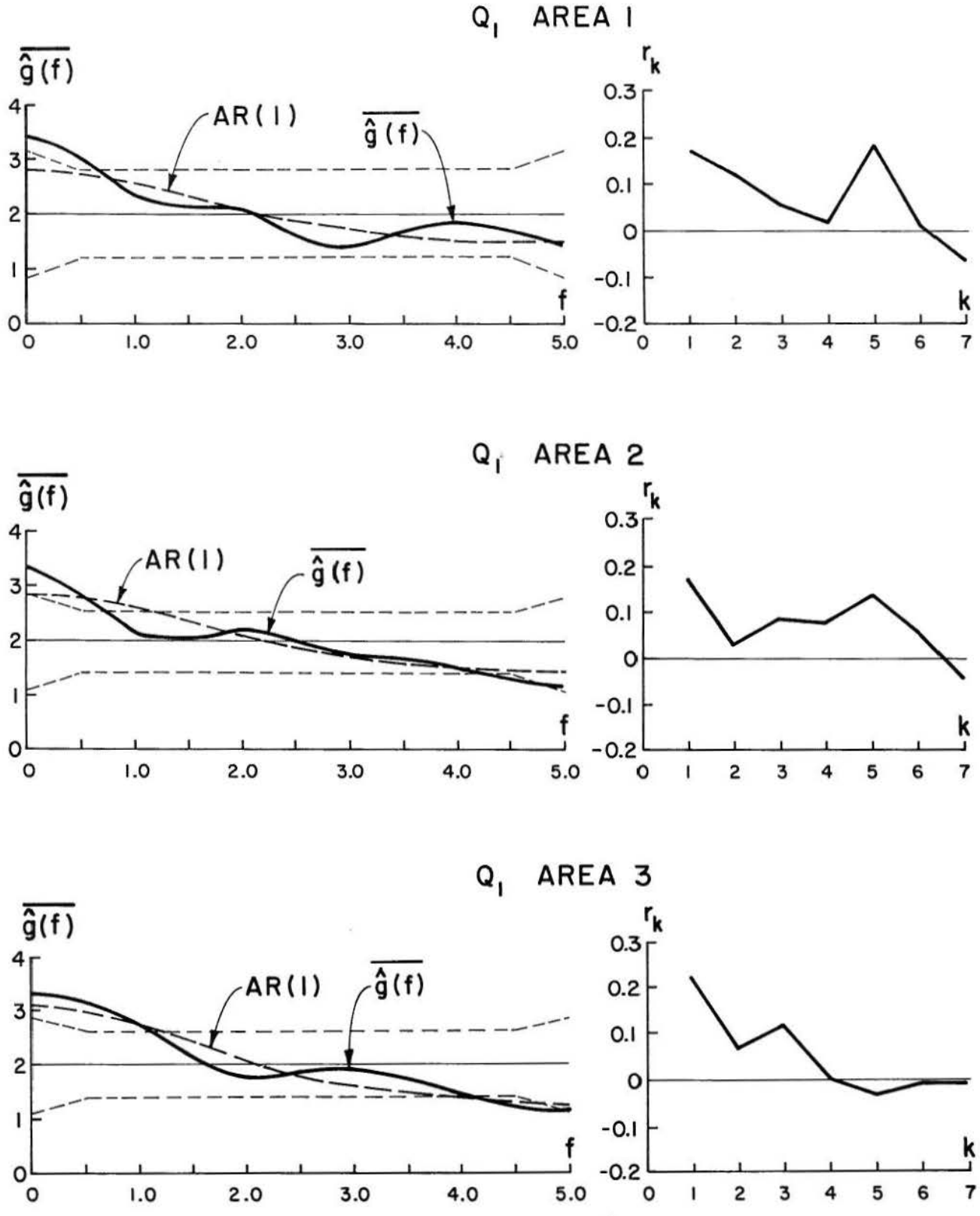

Figure 4-5. Average Spectra (The Full Lines of the Left Graphs) with the 95% Tolerance Limits (The Light Dashed Lines of the Left Graphs), the Fitted First-Order, AR(1), and the Second-Order, AR(2), Autoregressive Models (The Heavier Dashed, or Heavy Point Lines of the Left Graphs, Respectively), and the Average Correlograms (The Right Graphs for the Six Areas (1, 2, 3, 4, 5, 6) of the United States for the n-Series in Each Area of the Annual Q_1-Runoff Series.

First Part of Fig. 4-5, Areas 1, 2, 3.

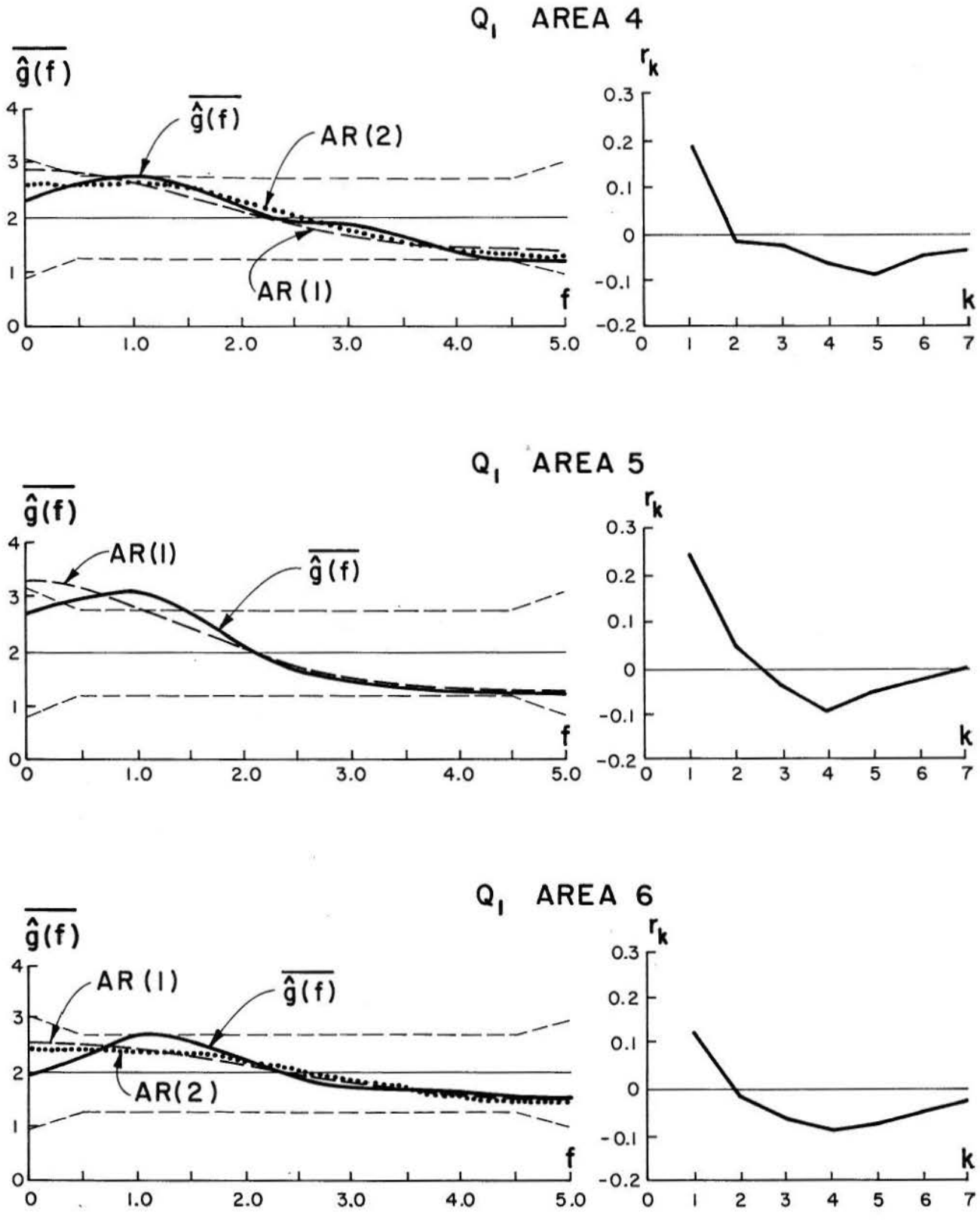

Figure 4-5. Continued. Second Part of Fig. 4-5, Areas 4, 5, 6.

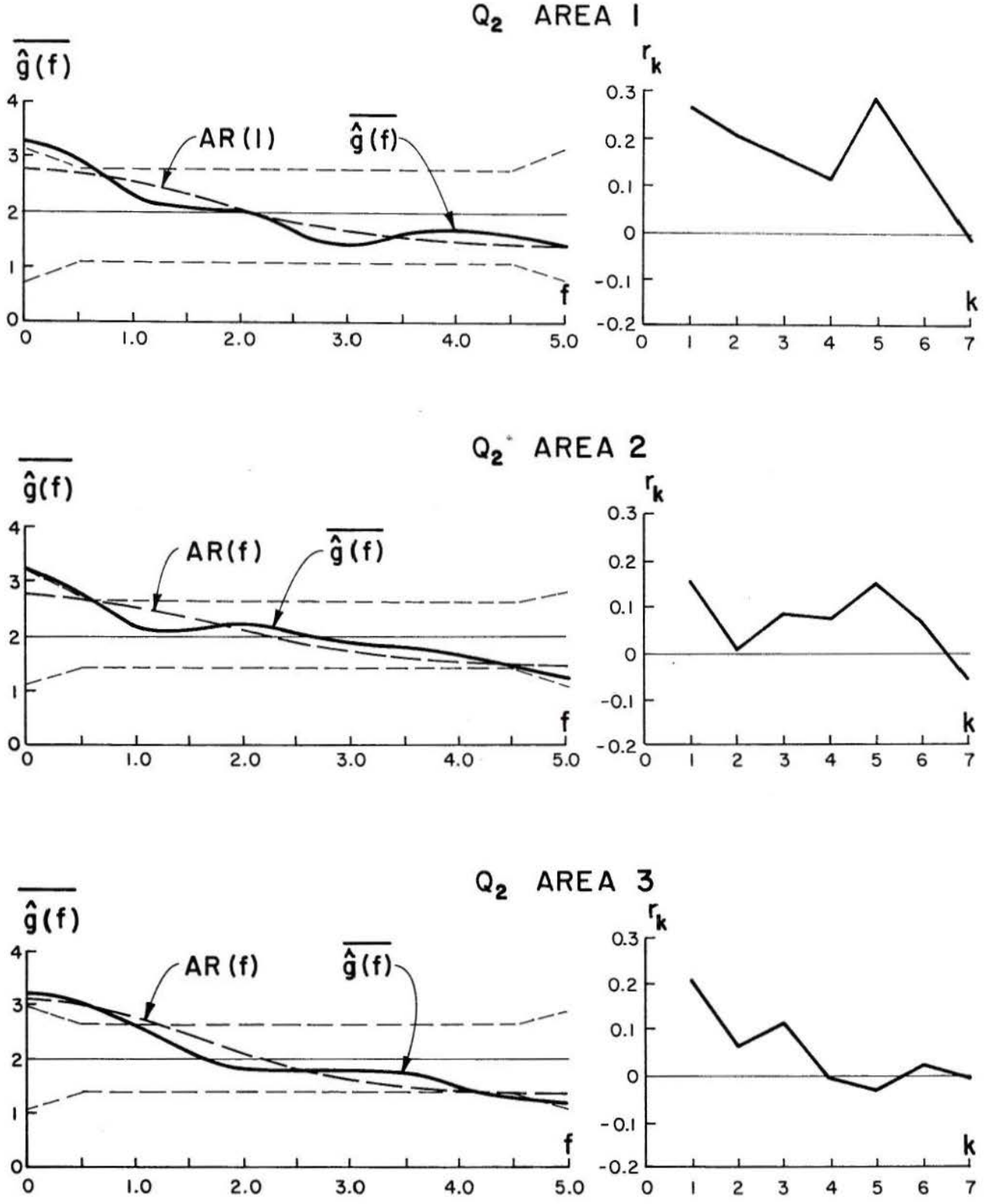

Figure 4-6. Average Spectra (The Full Lines of the Left Graphs) with the 95% Tolerance Limits (The Light Dashed Lines of the Left Graphs), the Fitted-First Order, AR(1), and the Second-Order, AR(2), Autoregressive Models (The Heavier Dashed, or Heavy Point Lines of the Left Graphs, Respectively), and the Average Correlograms (The Right Graphs for the Six Areas (1, 2, 3, 4, 5, 6) of the United States for the n-Series in Each Area of the Annual Q_2-Effective Precipitation Series.

First Part of Fig. 4-6, Areas 1, 2, 3.

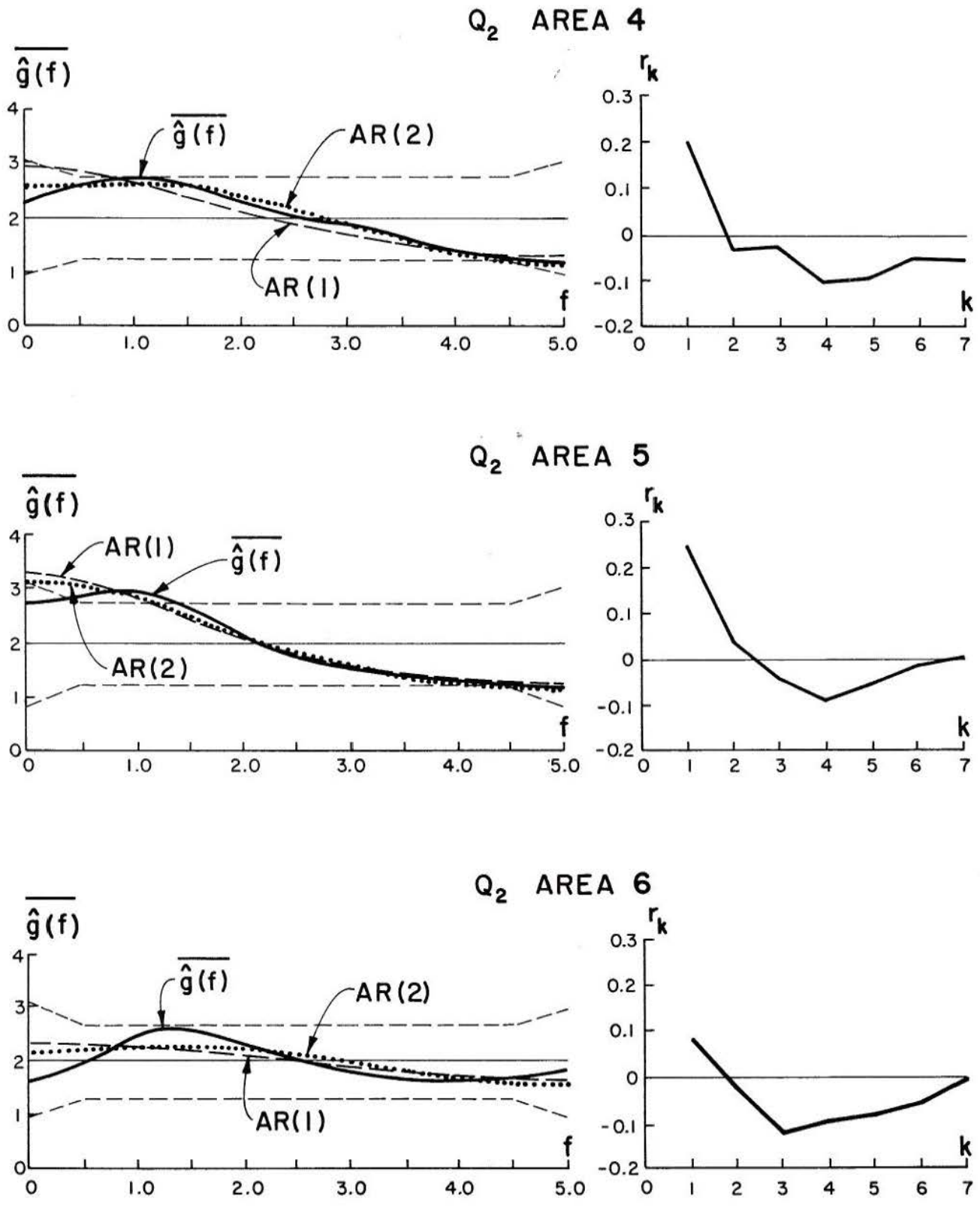

Figure 4-6. Continued. Second Part of Fig. 4-6, Areas 4, 5, 6.

The relationships of r_1 versus $\bar{x}$, or r_1 versus $\bar{x}$ and s_x, or r_1 versus $\bar{x}$ and C_v, show the R^2-values (or the explained variance by the linear correlation) to range between zero and 24% for P_1, between zero and 17% for P_2, between zero and 8% (except one value of 33%) for Q_1, and between zero and 8% (except one value of 17% and another of 30%) for Q_2, in each of four cases (P_1, P_2, Q_1, Q_2), for a total of 18 values of R^2.

The only significant correlation found is between $\bar{x}$ and s_x (the estimates of the mean and standard deviation are correlated), because whenever $\bar{x}$ is large, also s_x is large, or the opposite. However, when $\bar{x}$ is correlated with $C_v = s_x/\bar{x}$, the R^2-values are much smaller then those for s_x.

The R^2-values can not be considered as significant, except for the $\bar{x}$ versus s_x relationships. The estimates of the first serial correlation coefficient, r_1, may be considered as independent of the estimates of the mean, $\bar{x}$, the standard deviation, s_x, and the coefficient of variation, C_v. However, the relationships between the estimates $\bar{x}$ and s_x are very significant. This conclusion is supported by many computations of the past.

4-7 A Retrospective View at the Reliability of Results

The Parzen smoothing function has a bias which depends upon the second derivative of the actual spectrum. This derivative is very small in a region where the spectrum is nearly a straight line and very large near a sharp peak. Since the spectral estimates of the series of annual precipitation are serially correlated only to a very small degree (close to zero), these estimates represent the spectra which are likely to be close to a horizontal line. Thus, the Parzen smoothing function has a small bias most of the time. Even for the dependence in the annual runoff series of the order of $r_1 = 0.10 - 0.20$, the bias is small. The Parzen smoothing function is choosen for the use in this study since it yields no negative values of spectral estimates.

It is generally known (particularly for the first-order autoregressive processes [10]) that the increase of the correlogram truncation value m in computing the spectral densities will increase the variance of the estimates and decrease the number of degrees of freedom of their distribution. It results in a wider tolerance interval for the estimates, or a loss of confidence in the estimates. Furthermore, as it is generally known, an increase in the correlogram truncation value m will decrease the bias of the estimate. Most of the time, and for a single spectral estimate, the truncation value m is chosen as an average value for which, hopefully, neither the tolerance interval nor the bias are too large. For the purpose of the analysis in this paper, the bias in the spectral estimates was kept to a minimum. However, the variance of the estimates is large, but this is not considered important for the major assumption, namely that each of the n spectral estimates for the n area station series are considered as estimates of the same spectral density of a given frequency. Thus, by combining the n estimates into the average estimate, $\overline{\hat{g}(f)}$, the variance of this average estimate is greatly reduced (especially for the large number of station series of n = 200-400, with the equivalent n_e = 3-9 for the space independent series). To reduce the bias in the estimates, the truncation value m is taken as high as feasible without jeopardizing the covariance estimates. The truncation value m is taken as one fourth (1/4) of the sample size, N. To be consistent from sample to sample, this ratio is kept constant for all the series, although the absolute truncation value m changes from sample to sample depending upon the sample size.

Chapter V
CONCLUSIONS

The analysis of a large number of station time series of annual precipitation, annual runoff, and annual effective precipitation, leads to these basic conclusions:

(1) For all practical purposes, the method of spectral analysis shows that the annual precipitation series in the United States are time independent, stationary stochastic processes; at least, they are *temporarily stationary* processes for the periods of time of the order of lengths of the available historic series.

(2) The annual runoff series, and from them derived annual effective precipitation series (annual precipitation minus annual evaporation and annual evapotranspiration) in the United States, studied by the spectral analysis, are time dependent, stationary stochastic processes; at least, they are *temporarily stationary* processes for the periods of time of the order of lengths of the available historic series.

(3) The order of the magnitude of the average first serial correlation coefficient for the annual runoff series for many stations of large areas is somewhere between 0.10 and 0.20.

(4) The major factor responsible for the time dependence of annual series of runoff and effective precipitation is the change from year to year in the water volume stored in river basins.

(5) No significant spectral densities are found for any periodicity, particularly for the sunspot cycle of 11.3 years or its double value of 22.6 years, in annual series either of precipitation or runoff.

(6) It is safe to project the expectations of immediate future, say for the future lengths of times at least equal to the lengths of historic series of precipitation or runoff, that the expected future series of annual precipitation will be very close to independent, stationary stochastic processes, and those of annual runoff very close to dependent, stationary stochastic processes.

(7) For practical purposes, the simple first-order or second-order autoregressive dependence models seem to be sufficiently accurate mathematical description for the annual runoff series.

(8) By using a very large number of annual time series all over the United States, the time-space study of these processes for data of a relatively short periods of about 30-90 years for various time series, should compensate in some degree for the lack of very long, instrumentally obtained data on precipitation and runoff. What is not available in time, it may be somewhat compensated with the information in space, in order to project the near future with a sufficient reliability.

REFERENCES

1. Yevjevich, V., Fluctuations of Wet and Dry Years, Part I, Research Data Assembly and Mathematical Models, Colorado State University Series of Hydrology Papers, No. 1, July 1963.

2. Yevjevich, V., Fluctuations of Wet and Dry Years, Part II, Analysis by Serial Correlation, Colorado State University Series of Hydrology Papers, No. 4, June 1964.

3. Milankovitch, M., Canon of Insolation and the Ice-age Problem, translated from German by the Israel Program for Scientific Translations, Jerusalem 1969: U.S. Department of Commerce, Springfield, Va., 484 pages, 1941.

4. Vernekar, A. C., Long-period Global Variations of Incoming Solar Radiation, Meteorological Monographs, U.S. Weather Bureau, Vol. 12, No. 34, 21 pages, 1972.

5. Broecker, W. S., and J. van Donk, Insolation Changes, Ice Volumes, and the O-18 Record in Deep-sea Cores, Reviews of Geophysics and Space Physics, Vol. 8, p. 169-198, 1970.

6. Chin, W. Q., and V. Yevjevich, Almost-Periodic, Stochastic Theory of Climatic Changes, Colorado State University Series of Hydrology Papers, No. 65, March 1974.

7. Wiener, Nobert, Extrapolation, Interpolation, and Smoothing of Stationary Time Series (with Engineering Applications), the M.I.T. Press, Cambridge, Mass., 1949.

8. Blackman, R. B., and F. W. Tukey, The Measurement of Power Spectra, Dover Publications, New York 1958.

9. Parzen, Emanuel, Time Series Analysis Papers, Holden-Day, Inc., San Francisco, 1967.

10. Box, G. E. P. and G. M. Jenkins, Time Series Analysis, Forecasting and Control, Holden-Day, Inc., pp. 553, 1970.

Key Words: Wet and Dry Years, Annual Precipitation, Annual Runoff, Time Series Analysis.

Abstract: Spectral analysis is applied to annual series of precipitation and runoff. Precipitation series are divided in homogeneous P_1-precipitation and nonhomogeneous P_2-precipitation. Runoff series are either the observed Q_1-series, or they are reduced to the effective precipitation Q_2-series (precipitation minus evaporation). Data of annual precipitation and annual runoff of a large number of gauging stations in the United States divided in six areas are used in this study.

Key Words: Wet and Dry Years, Annual Precipitation, Annual Runoff, Time Series Analysis.

Abstract: Spectral analysis is applied to annual series of precipitation and runoff. Precipitation series are divided in homogeneous P_1-precipitation and nonhomogeneous P_2-precipitation. Runoff series are either the observed Q_1-series, or they are reduced to the effective precipitation Q_2-series (precipitation minus evaporation). Data of annual precipitation and annual runoff of a large number of gauging stations in the United States divided in six areas are used in this study.

Key Words: Wet and Dry Years, Annual Precipitation, Annual Runoff, Time Series Analysis.

Abstract: Spectral analysis is applied to annual series of precipitation and runoff. Precipitation series are divided in homogeneous P_1-precipitation and nonhomogeneous P_2-precipitation. Runoff series are either the observed Q_1-series, or they are reduced to the effective precipitation Q_2-series (precipitation minus evaporation). Data of annual precipitation and annual runoff of a large number of gauging stations in the United States divided in six areas are used in this study.

Key Words: Wet and Dry Years, Annual Precipitation, Annual Runoff, Time Series Analysis.

Abstract: Spectral analysis is applied to annual series of precipitation and runoff. Precipitation series are divided in homogeneous P_1-precipitation and nonhomogeneous P_2-precipitation. Runoff series are either the observed Q_1-series, or they are reduced to the effective precipitation Q_2-series (precipitation minus evaporation). Data of annual precipitation and annual runoff of a large number of gauging stations in the United States divided in six areas are used in this study.

Techniques of spectral analysis are described. Average spectra are estimated for four variables (P_1, P_2, Q_1, Q_2) and six areas, with the 95% tolerance limits.

The annual precipitation series are very close to independent time series, at least temporary stationary for the length of 50-150 years. Annual runoff and annual effective precipitation series are dependent series, at least temporary stationary for the order of length of 50-150 years. The first and the second order autoregressive models seem sufficiently accurate for practical use.

Reference: Yevjevich, Vujica; Colorado State University, Hydrology Paper No. 94 (August 1977), Fluctuations of Wet and Dry Years, An Analysis by Variance Spectrum.

Techniques of spectral analysis are described. Average spectra are estimated for four variables (P_1, P_2, Q_1, Q_2) and six areas, with the 95% tolerance limits.

The annual precipitation series are very close to independent time series, at least temporary stationary for the length of 50-150 years. Annual runoff and annual effective precipitation series are dependent series, at least temporary stationary for the order of length of 50-150 years. The first and the second order autoregressive models seem sufficiently accurate for practical use.

Reference: Yevjevich, Vujica; Colorado State University, Hydrology Paper No. 94 (August 1977), Fluctuations of Wet and Dry Years, An Analysis by Variance Spectrum.

Techniques of spectral analysis are described. Average spectra are estimated for four variables (P_1, P_2, Q_1, Q_2) and six areas, with the 95% tolerance limits.

The annual precipitation series are very close to independent time series, at least temporary stationary for the length of 50-150 years. Annual runoff and annual effective precipitation series are dependent series, at least temporary stationary for the order of length of 50-150 years. The first and the second order autoregressive models seem sufficiently accurate for practical use.

Reference: Yevjevich, Vujica; Colorado State University, Hydrology Paper No. 94 (August 1977), Fluctuations of Wet and Dry Years, An Analysis by Variance Spectrum.

Techniques of spectral analysis are described. Average spectra are estimated for four variables (P_1, P_2, Q_1, Q_2) and six areas, with the 95% tolerance limits.

The annual precipitation series are very close to independent time series, at least temporary stationary for the length of 50-150 years. Annual runoff and annual effective precipitation series are dependent series, at least temporary stationary for the order of length of 50-150 years. The first and the second order autoregressive models seem sufficiently accurate for practical use.

Reference: Yevjevich, Vujica; Colorado State University, Hydrology Paper No. 94 (August 1977), Fluctuations of Wet and Dry Years, An Analysis by Variance Spectrum.

LIST OF PREVIOUS 25 PAPERS

No. 69 Efficient Sequential Optimization in Water Resources, by Thomas E. Croley II, September 1974.

No. 70 Regional Water Exchange for Drought Alleviation, by Kuniyoshi Takeuchi, November 1974.

No. 71 Determination of Urban Watershed Response Time, by E. F. Schulz, December 1974.

No. 72 generation of Hydrologic Samples, Case Study of the Great Lakes, by V. Yevjevich, May 1975.

No. 73 Extraction of Information on Inorganic Water Quality, by William L. Lane, August 1975.

No. 74 Numerical Model of Flow in Stream-Aquifer System, by Catherine E. Kraeger Rovey, August 1975.

No. 75 Dispersion of Mass in Open-Channel Flow, by William W. Sayre, August 1975.

No. 76 Analysis and Synthesis of Flood Control Measures, by Kon Chin Tai, September 1975.

No. 77 Methodology for the Selection and Timing of Water Resources Projects to Promote National Economic Development, by Wendim-Agegnehu Lemma, August 1975.

No. 78 Two-Dimensional Mass Dispersion in Rivers, by Forrest M. Holly, Jr., September 1975.

No. 79 Range and Deficit Analysis Using Markov Chains, by Francisco Gomide, October 1975.

No. 80 Analysis of Drought Characteristics by the Theory of Run, by Pedro Guerrero-Salazar and Vujica Yevjevich, October 1975.

No. 81 Influence of Simplifications in Watershed Geometry in Simulation of Surface Runoff, by L. J. Lane, D. A. Woolhiser and V. Yevjevich, January 1976.

No. 82 Distributions of Hydrologic Independent Stochastic Components, by Pen-chih Tao, V. Yevjevich and N. Kottegoda, January 1976.

No. 83 Optimal Operation of Physically Coupled Surface and Underground Storage Capacities, by Dragoslav Isailovic, January 1976.

No. 84 A Salinity Management Strategy for Stream-Aquifer Systems, by Otto J. Helweg, and J. Labadie, February 1976.

No. 85 Urban Drainage and Flood Control Projects Economic, Legal and Financial Aspects, by Neil S. Grigg, Leslie H. Botham, Leonard Rice, W. J. Shoemaker, and L. Scott Tucker, February 1976.

No. 86 Reservoir Capacity for Periodic-Stochastic Input and Periodic Output, by Kedar Nath Mutreja, September 1976.

No. 87 Area-Deficit-Intensity Characteristics of Droughts, by Norio Tase, October 1976.

No. 88 Effect of Misestimating Harmonics in Periodic Hydrologic Parameters, by K. L. Bullard, V. Yevjevich, and N. Kottegoda, November 1976.

No. 89 Stochastic Modeling of Hydrologic, Intermittent Daily Processes, by Jerson Kelman, February 1977.

No. 90 Experimental Study of Drainage Basin Evolution and Its Hydrologic Implications, by Randolph S. Parker, June 1977.

No. 91 A Model of Stochastic Structure of Daily Precipitation Series Over an Area, by C. W. Richardson, July 1977.

No. 92 Effects of Forest and Agricultural Land Use on Flood Unit Hydrographs, by Wiroj Sangvaree and Vujica Yevjevich, July 1977.

No. 93 A Distributed Kinematic Model of Upland Watersheds, by Edward W. Rovey, David A. Woolhiser and Roger E. Smith, July 1977.