AREA-TIME STRUCTURE OF THE MONTHLY PRECIPITATION PROCESS

by
Vujica Yevjevich
and
Alan K. Karplus

AUGUST 1973

HYDROLOGY PAPERS
COLORADO STATE UNIVERSITY
Fort Collins, Colorado

64

AREA-TIME STRUCTURE OF THE MONTHLY PRECIPITATION PROCESS

by
Vujica Yevjevich*
and
Alan K. Karplus**

HYDROLOGY PAPERS
COLORADO STATE UNIVERSITY
FORT COLLINS, COLORADO

August 1973

No. 64

*Professor of Civil Engineering, Colorado State University, Fort Collins, Colorado.
**Post-Doctoral Research Associate, Department of Civil Engineering, Colorado State University, Fort Collins, Colorado.

TABLE OF CONTENTS

ABSTRACT

The structural analysis of area-time hydrologic process of monthly precipitation is based on the concept that these processes are composed of deterministic components specified by periodic parameters and a stationary stochastic component, with the coefficients of the periodic parameters following regional trends.

Sets of monthly precipitation series at stations within two regions are used as examples to demonstrate the structural composition. Region I, located in eastern North Dakota, South Dakota, and Minnesota has 41 monthly precipitation stations; Region II, located in eastern Nebraska has 29 monthly precipitation stations. Both have such precipitation characteristics so that the above concept of structural analysis may be incorporated.

Mathematical models for the periodicity and trends in parameters are inferred, with five regional constants and three regression coefficients for each of the two regions. When the periodicity and regional trends in parameters are removed, the stationary stochastic components of monthly precipitation series are found to be approximately time independent processes. In addition they follow closely the identical three-parameter gamma probability distribution. The independent stationary stochastic components of monthly precipitation are highly cross correlated, with the lag-zero cross correlation coefficient found to be primarily a function of the interstation distance. This correlation coefficient may be more generally specified as a function of the stations location within the region, the interstation distance, the azimuth of the straight line connecting the stations, and the time of the year.

The developed methodology permits the generation of new samples consisting of a set of time series for a region, either at the observed station points or at any new grid of points.

ACKNOWLEDGMENTS

The material in this paper includes also a portion of Ph.D. dissertation submitted by A. Karplus to Colorado State University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The research leading to the dissertation was financially sponsored by the U. S. National Science Foundation under the grant GK-11564 (Large Continental Droughts), with V. Yevjevich as the principal investigator. This financial support is gratefully acknowledged by both writers. The research leading to this paper has been conducted within the Hydrology and Water Resources Graduate and Research Program, Department of Civil Engineering, Colorado State University. Thanks is also extended to several reviewers of the manuscript.

CHAPTER I

INTRODUCTION

This introductory chapter describes the approach used in studying the regional extraction of information on various hydrologic parameters and in describing the hydrologic processes, particularly the area-time properties of monthly precipitation. The objectives of this study are defined. Also as an introduction to subsequent chapters, the procedures used in the investigations are outlined and the related contributions in literature are cited.

1.1. Basic Approach for Regional Extraction of Information on Hydrologic Parameters.

The basic geophysical random processes found in nature are four-dimensional space-time processes, $x\{x,y,z,t\}$; x, y, z are the space coordinates of any point, and t is the time. When a surface (i.e., area, region, river basin) is considered, the process becomes an area-time process, $x\{x,y,t;\ z_d\}$; z_d is well defined for the surface either as a constant or as a function of x and y . A line-time process is represented by $x\{x,t;\ y_d;\ z_d\}$; y_d and z_d are functions only of x . Finally, the point-time process is defined as $x\{t;\ x_d,\ y_d,\ z_d\}$, with the precise coordinates of the point, x_d, y_d, z_d .

Statistical parameters, which are usually used to describe the point-time, line-time, area-time, and space-time hydrologic processes and which must be estimated from the data, include the following parameters based on moments: mean, variance (standard deviation), coefficient of variation, autocorrelation coefficients, skewness and kurtosis coefficients, the amplitudes and phases of the harmonics of periodic parameters, the lag cross correlation coefficients, etc. The regression coefficients of polynomials used to describe the trends in these parameters must also be estimated from data. The classical approach used in observing natural multi-dimensional processes is to define the space by a set of points and to observe the time varying processes at these points. The processes are then analyzed as a set of point-time processes. If the individual point-time processes are observed at discrete times, or are averaged or totaled over time intervals, instead of being recorded continuously, then the variable values are available at a set of points and at a set of discrete times or time intervals. The continuous processes have then been approximated by N discrete observations in time for each of the M discrete points. For example, in the case of the area-time processes, the M regional stations (points) and the N time intervals result in a total of MN observations.

Hydrologic area-time processes are usually dependent both in area and in time. Thus a dependent area-time process with a total of MN observations has information on population characteristics equivalent to an effective number of points, $M_e \leq M$, and an effective sample size, $N_e \leq N$, of an equivalent area-time stochastic process which is independent both in area and in time. Since the product $M_e N_e$ is usually much smaller than MN , the station-time approach to study the area-time dependent hydrologic continuous processes (such as the classical station-year approach in hydrology) has much less information than the product MN for independent processes. However, it should be noted that $M_e N_e$ is usually greater than N .

Two among several problems may be relevant in hydrologic investigations of area-time processes:

(1) To improve the reliability of estimated hydrologic parameters at a given i-th point of observation by using all observations at M points, including the point of interest. The basic hypothesis is that the time series data of $M-1$ points around the given i-th point have additional information which can improve the estimates of models and parameters of the i-th point-time series beyond that information contained only in the N observations at that point.

(2) To estimate the basic hydrologic parameters of time series at m points other than the observed M points, with these m points being inside or very close to the area of the M points. This is equivalent to transferring information inside an area from the M points with observed time series to the m new points having no observations.

The problems of improving the reliability in estimates of models and parameters at given points are numerous for the line-time, area-time, and space-time processes. The case of area-time processes is used in this study as an example of investigating the improvement of model and parameter estimations. This is done because a great many hydrologic stochastic processes are referred to an area (such as river basin, region, lake or reservoir surface, infiltration surface, etc). Precipitation, evaporation, effective precipitation (precipitation minus evaporation), unit area runoff, infiltration, unit area erosion, temperature, snow depth, unit area snow melt, and other similar variables are typical area-time hydrologic stochastic processes. The monthly random area-time precipitation process is used as a representative process for the further investigation. Because precipitation is a periodic-stochastic time process for time units of less than one year, the monthly discrete time series of precipitation at M points in a region provide both an example and testing data for the following analysis.

The problems studied are now defined as follows:

(1) To determine the best method for estimating the parameters of $x\{t;\ x_d,\ y_d,\ z_d\}$ process, the premise is that a set of point-time series, $x\{x_i,\ y_i,\ t;\ z_d\}$, $i = 1,2,\ldots,M$, has more information than a particular point-time series, $x\{t;\ x_d,\ y_d,\ z_d\}$. Therefore, estimates of the parameters of $x\{t;\ x_d,\ y_d,\ z_d\}$ using all $x\{x_i,\ y_i,\ t;\ z_d\}$ observed samples should be more reliable (unbiased, and more efficient, i.e. having lower sampling variances), than using only the observed sample of the $x\{t;\ x_d,\ y_d,\ z_d\}$ point-time process.

(2) The premise is that the estimates of parameters at a set of points, of the process $x\{x_j,\ y_j,\ t;\ z_d\}$, with $j=1,2,\ldots,M$, of the general area-time process $x\{x,\ y,\ t;\ z_d\}$, by using the observed data of individual point-time series, $x\{x_i,\ y_i,\ t;\ z_d\}$, $i=1,2,\ldots,M$, may or may not represent the same points in the area defined.

The underlying assumption for the solution of the above two problems is that the mathematical models used to describe approximately the individual point-time processes are the same for all M(or M+m) point-time processes. However, their parameters vary over the area with a well defined trend which may be described by functions. In general, this is valid for any point in the area defined by the M observational points. In other words, the mathematical description models, and their parameters, of all periodic-stochastic point-time processes in the area are jointly inferred by using all data of the M available point-time series.

A classical concept in climatology and hydrology is the concept of regional homogeneity or non-homogeneity. The usual understanding of this concept is that various basic parameters change only slightly over a region; thus the region can be considered homogeneous for that parameter and for a given climatologic or hydrologic random variable. Since all parameters of any random process continuously change to some extent along a line, in an area, or in a space, the homogeneity would become an objective concept only if the rates of change, $\partial\nu/\partial x$, $\partial\nu/\partial y$, and $\partial\nu/\partial z$, of a parameter ν, are prescribed in advance as to the rates of change tolerable within the definition of homogeneity. Integrated over the range of x, y, or z, these rates would give the limits of parameter fluctuations along a line, across an area, or over a space, for which the term homogeneity could be used.

The approach to be used in this study does not divide regions into hydrologically or climatologically homogeneous or non-homogeneous regions. It does, however, divide the regions into those for which simple first or second-order polynomical functions can fit well the areal changes in the parameters studied, and those which have large changes so that higher-order orthogonal polynomials (the third, the fourth and higher-order polynomials) must be used to fit well the regional variation of the parameters studied. Usually, the gently rolling and flat continental areas will have a regional change of parameters which may be well described by the first-order (small areas) and the second-order polynomials (large areas). A proper statistical inference must be made to support any conclusion concerning the goodness of fit. Highly rolling and mountainous terrains would require third or higher order polynomials. Because an increase in the order of the polynomial requires more points for an accurate estimate of a larger number of polynomial coefficients, the accuracy of coefficients estimates usually decreases with an increase of the polynomial order. This is true because usually the number of polynomial coefficients increases and the number of observation points decreases with the increase of orographic complexity. When this complexity is very large, the idea of determining a deterministic regional trend surface for the parameters may become untenable because of the low accuracy in the estimation of the polynomial coefficients. A change in the approach may then be necessary by using another method of transferring the information on parameters from the (M-1) - points to the i-th point, or from the M points of observations to m points inside the same area which have none or a limited number of observations.

All hydrologic parameters may be assumed either to follow a deterministic, periodic time process with the periods equal to known astronomical cycles or to be constants in time (properties of stationary processes of a given order). However, both the average values of these periodic parameters and the coefficients of their periodic functions follow deterministic trend functions.

The regional extraction of information concerning hydrologic parameters offers the following advantages:

(1) The mathematical description models of time series structure are more adequate when developed from and tested on a large set of point-time series;

(2) The regional estimation of time series parameters gives more reliable estimates;

(3) The models and parameters can be obtained at all points of interest, rather than only at the points of observations, which are usually selected on the basis of various convenience factors;

(4) The future points of interest for the area-time hydrologic processes may not be known at the present time; however, the reliable estimation methods using regional extraction of information enable the future reconstruction of both models and parameters of the time series at these points of interest;

(5) The study of the regional extent of extreme events, such as floods and droughts, may be made much simpler and more accurate if related to a systematic grid of points, from either the analytical or the data generation methods of solution, than if related only to points of the observed time series;

(6) The use of regional extraction of information will result in a more accurate generation of a set of point-time series, both at the observed and at any new grid of points.

In cases for which either the use of the historic (observed) discrete point-time series, or the use of analytical methods to solve problems cannot yield reliable results for probability related problems, the data generation method (Monte Carlo or experimental statistical method) may be the only approach for finding reliable answers. This approach would then consist of three stages: (a) the regional extraction of information for models and parameters; (b) the generation of a set of samples of a given size at a set of points, either at the observed points or at any new grid of points; and (c) the use of these generated samples for finding the area-time properties of random variables related to the problems being studied (such as finding the most reliable characteristics of regional droughts). However, the studies of this paper are limited to the first subject, namely the regional extraction of information. This is a prerequisite for the reliable solutions to the latter two subjects.

Two approaches appear feasible for the joint regional estimation of parameters at a set of points: (a) by fitting a trend function through the set of parameter estimates of individual point-time series in such a way that the deviations between the function and the individual estimates may be inferred to be only the sampling variations, with the trend function fitted by some appropriate technique; and (b) by developing procedures of joint estimation of parameters at the selected points using all the data from a set of observed points, and without the fitting of trend functions. The first approach may be divided into two alternatives: the use of equal weights for all points in a least-square fitting procedure for the selected trend functions; and the use of different weights for the point estimates, because of unequal variances of the point estimates.

The development of fast digital computers presents a relatively economical approach to the regional extraction of information on hydrologic parameters. An optimum is sought between the total available information in data, the reliability (accuracy, efficiency) in the extracted information, and the cost of extraction. Computers have permitted the use of much more sophisticated statistical methods in estimating models and their parameters than the precomputer computational devices permitted.

The underlying approach for investigations is illustrated by the following example: the precipitation of a region is measured by rain gages at a set of points, with the observations totaled into monthly time series. The study is concerned with the regional variation of the basic parameters of these monthly precipitation series, as deterministic regional population properties, and the annual periodicity in many of the parameters of the time series. Once the deterministic models of space and time variations have been inferred as mathematical functions with estimated coefficients, these deterministic variations can be removed from the series, with the remaining stochastic components of monthly series well approximated by a second-order or higher-order stationary stochastic process. The particular investigation is related to the regional dependence among the standardized, approximately independent, second-order stationary stochastic variables of monthly precipitation. Because these variables are standardized, and their higher-order moments and boundaries are approximately equal for all months, they may have the same probability distribution and be mutually dependent but identically distributed random variables. It is expected that the correlation coefficients between the pairs of these variables would follow a well defined decay function dependent on the distance, in addition to occasionally being dependent on the orientation, season, and position of the pairs of stations.

This study is further directed toward demonstrating the application of the simplified, but appropriate, regional models to two regions having a relatively small rate of change in parameters with longitude and latitude. A comparison of the basic statistics for the stochastic variables of the two regions is made to demonstrate that only sampling variations are involved in the distribution characteristics for these standardized, approximately independent, identically distributed stationary stochastic components of monthly precipitation series.

1.2. General Objectives of Investigations on Regional Extraction of Information. One of the general objectives of these investigations relates to those water resources problems which depend on both the area and the time characteristics of the problem. Such a problem is the evaluation of the area-time probabilistic characteristics of hydrologic droughts. For droughts many descriptors are relevant, particularly such properties as the areal extent, total water deficit, duration, maximum unit-time deficit, and shape of uninterrupted water deficiencies. To find the joint probability distribution of a set of drought descriptors, for series samples of a given size, and considering all descriptors as random variables, the most reliable regional information for mathematical models and parameters of the drought defining hydrologic random variables should be extracted. The joint distribution of drought descriptors cannot be more accurate than the accuracy in estimating the mathematical models and their parameters in the structural analysis of this area-time hydrologic stochastic process.

Atmospheric circulation determines the basic characteristics of precipitation and evaporation in a region. In general, the further a region is from the oceanic sources of moisture and the smaller the orographic effect on precipitation and evaporation, the smaller is the total annual precipitation. In cases of small precipitation, evaporation usually takes a large portion of the precipitation, which results in a smaller runoff. This situation is accompanied by a larger variability of precipitation and runoff, both across the region and in time.

Because of definite climatic patterns and river basin characteristics, it can be expected that the basic hydrologic parameters exhibit some deterministic regional characteristics. At present the regional variations of hydrologic properties are mainly determined statistically from the observed data. In the future one should expect other inputs, such as those resulting from studies on atmospheric circulation, patterns of deposition of moisture, and on evaporation.

In the absence of solid physical information on the regional variation of the basic hydrologic parameters concerning precipitation, another objective of this investigation is to utilize as much as possible the statistical methods of extracting the maximum information on the regional variation of hydrologic parameters. This variation can be expected to be in the form of a deterministic regional trend surface function with an increase of its complexity as the area increases. The distance from the sources of moisture may change significantly, and the patterns of atmospheric circulation and orographic effects may also change as the area increases. It can be expected that the goodness of fit of a hypothesized mathematical trend surface function to point estimates of a hydrologic parameter will decrease, on the average, as the size of the region increases.

A criticism can be made that the efforts in various water resources analyses have been directed more towards understanding the structural composition of hydrologic point-time processes than towards understanding the line-time, area-time, and space-time processes. However, an expansion of the interrelated water resource uses across a region, the regional variations of the basic hydrologic parameters deserve as much attention as the analysis of their time properties.

Therefore, the objective of this paper is to study the regional properties of hydrologic parameters by using trend functions fitted to their point estimates. The general joint estimates of a parameter at a set of points in a region by techniques other than by fitting trend functions are not considered in this paper.

1.3. Procedures Used in Investigations. In general cases of area-time hydrologic stochastic processes, the hydrologic parameters usually vary over the basic astronomical time periods in a complex periodic process, and across an area in a complex deterministic trend surface. Fitting the periodic changes in parameters by a limited number of harmonics in the Fourier analysis of periodic processes and the area variation in parameters by a trend function, usually as a polynomial with a limited number of terms, poses the problem of determining the cut-off points as to the number of harmonics and the number of polynomial terms. When properties of underlying processes are simple, such as in the case of time independent, but area dependent, multivariate normal processes, or both time and area dependent multivariate normal processes with relatively simple time and area dependence models, the proper statistical inference techniques can be designed to decide upon these cut-off points. However, the analytical approach for designing these inference techniques for more complex processes may itself be of questionable reliability. Therefore, the dilemma of selecting the cut-off points may be resolved in such a way that the individual harmonics and the individual polynomial terms beyond the cut-off point mainly represent sampling errors. The neglected harmonics and polynomial terms should then contribute only relatively small percentages to the explanation of either the total periodic variation or the total regional deterministic variation in estimates of the parameter under consideration.

A significant part of periodic estimates of a parameter may represent sampling errors. This can be demonstrated in estimating parameters by increasing the sample size in steps, and each time comparing the deviations from each smooth fitted periodic function having a limited number of harmonics. The variance of these deviations decreases with an increase of the sample size N . The eventual convergence of this variance to approximately zero, as N goes to infinity, justifies the use of a small number of harmonics for the population periodic function of a parameter. Similarly, the variance of deviations of the point estimates of a parameter about fitted polynomials, having a limited number of terms, decreases with an increase of the sample size N . This variance may also be assumed to converge to approximately zero when N goes to infinity, justifying the use of a deterministic polynomial function with a small number of terms, as a good approximation to the regional population trend surface.

Periodic movements and regional surface trends of various hydrologic parameters may be close in many cases to be proportional, or simply related, to the periodicity and trend surface functions of the mean. This property may significantly decrease the number of parameters and coefficients which must be estimated. A simplification in the description of the area-time monthly precipitation process can be accomplished in five ways: (a) by neglecting the harmonics of periodic parameters having relatively small amplitudes; (b) by neglecting the small magnitude terms, with relatively small regression coefficients, in deterministic regional polynomial trends; (c) by using the proportionality, or simple relations existing between the periodic functions of basic parameters; (d) by using the proportionality, or simple relations, existing between the trend surface functions of basic parameters; and (e) by properly inferring that some parameters and coefficients are not significantly different from a constant either over the time period of 12 months or over the region under study. These will be the major procedures used in this paper to economize on the total number of parameters and/or coefficients to be estimated for the area-time monthly precipitation process.

1.4. Regional Hydrologic Information. The regionalizing of parameters of a random phenomenon has interested hydrologists for a long time. The drawing of smoothed isolines for a set of point estimates of a parameter is equivalent to fitting a complex trend surface. However, this approach does not use a test of goodness of fit of these smoothed isolines to the estimated values. Assigning weights to point estimates, in determining the overall regional averages of parameters, has also been of constant interest to hydrologists. Recently, Amorocho and Brandstetter [1] studied the problem of estimating regional description parameters based on the density of a network of precipitation stations. Alternative methods of weighting the estimates of parameters of precipitation stations on irregularly spaced networks were developed by Whitmore, Van Eeden, and Harvey [2]. Several computer approaches to the computation of mean areal depth of precipitation are given by Akin [3] using available historic records. Solomon, Denouviller, Chart, Wooley, and Cadou [4] employ a technique whereby a dense square grid is placed over a region and grid points assigned within local boundaries previously selected. Once the grid is assigned, various hydrologic parameters can be rapidly determined for the area by the use of a computer.

Instead of using the subjective approach of smoothing the interpolated isolines between a set of point estimates of a parameter, an objective approach is preferred. Fitting the trend surface functions and testing the goodness of fit represents an objective approach to regional estimation of hydrologic parameters. When the polynomials are used, and the procedure is to fit and test, in a sequence, the trends of the first-, second-, or higher-order terms, the approach is objective only if the testing criteria are prescribed.

The problem of fitting the trend functions to the line, area and space variations of basic hydrologic parameters is also very old. In addition to the mean areal depth of precipitation, the manner in which the mean of a point-time series changes over an area has been of continuous interest. This type of problem is currently found in nearly all geophysical disciplines. In geology, the character of the earth's crust is studied by surface mapping, and estimates of the thickness of ore deposits and similar thicknesses are made. Krumbein [5] has used irregularly spaced sampling points to study the polynomial functions of best fit for use as trend surfaces. Chidley and Keys [6] have also investigated the use of trend surface functions. Mandelbaum [7] provides a criterion of fitting these functions, based on the step-by-step variation in the explained variance by the fitted polynomials with an increase of the number of polynomial terms. Trend surfaces have been used in the extension of rainfall records by Unwin [8]. In the case of gravitational data, Simpson [9] points out that "There is some question whether or not the removal of regional effects from gravity may be effectively accomplished by the method of least squares. One of the basic assumptions in the application is that the regional be a relatively low order effect." Oldham and Sutherland [10] have looked for regional trends in experimental data using orthogonal polynomials. The latter refers to equal spaced sampling points on a cartesian grid. These authors suggest that no unique solution should be expected for the terminology "...'regional effect' is in itself ambiguous." In discussing mapped data analyzed into "trend" and "residuals", Grant [11] reinforces this position, and points out that, based on labor and cost considerations, polynomial fitting has objective advantages over smoothing and gridding methods.

The dependence among the precipitation point-time series in a region is usually studied by using the linear correlation for these series taken pairwise. Then the correlation coefficients are expected to be related to the distance and the azimuth angle of orientation between stations. Sometimes, all these pairwise correlation coefficients of the M precipitation stations in a region are presented in the form of an MxM square correlation matrix. Hutchinson [12] uses a second-order polynomial for fitting the decay of the zero lag cross-correlation coefficients with the interstation distance. Huff and Shipp [13] present areal correlation coefficients for storm, monthly and seasonal rainfall, with no mathematical functional relations sought using the distance or azimuth. Steinitz, Huss, Manes, and Alperson [14], in evaluating the optimum station network size for pressure and wind fields, arrive at a nonlinear exponential decay function, $\rho = A \exp(Bd) + 1 - A$, with ρ, the zero lag cross-correlation coefficient between the series of two stations, A and B, coefficients to be estimated from the data, and d, the interstation distance. Alexeyev [15] uses a similar correlation mathematical model in the form $\rho = 1/(1+Ad)$. This relationship consists of the first two terms in the power expansion of the previous model. Hendrick and Comer [16] evaluated the space correlation for daily precipitation in Vermont by using the model $\rho = A + Bd - dC \sin(220^{o} - 2\phi)$, with A, B and C, coefficients to be estimated, and ρ, d and ϕ, the correlation coefficient, distance and azimuth, respectively. Cornish, Hill, and Evans [17] used the z-transform for ρ, and considered a linear plus a sinusoidal regression with distance and time of the year for the precipitation in Australia. Caffey [18] evaluated the correlation patterns for annual precipitation series in the United States by applying a weighted exponential decay function for ρ as a function of the interstation distance and the azimuth angle.

Since one objective of this study is to solve the problems of estimating regional drought characteristics in probability terms, a brief review of references related to the regional aspects of drought is given here. The evaluation of hazards caused by large continental droughts has long been of interest. Tannehill [19] and Campbell [20] made early studies of the continental drought phenomenon. The former gives a vivid history of the early United States in its fight to endure droughts, while the latter discusses a similar situation in Australia. Yevjevich, Saldarriaga, and Millan [21, 22, 23] present the concept of large continental drought in a stochastic hydrologic setting, while Barger and Thom [24], and Hershfield [25] showed how water shortages affected local farmers. The regional drought aspects have been evaluated subjectively in Australia by Foley [26] in 1957 and again in a more recent symposium, and by Maher [27] in 1967. Some rigor is needed in these methods for the objective definition and evaluation of droughts.

Agreement continues to persist that (i) drought must be defined in terms of water use; (ii) precipitation is one of the best single drought determining random processes; and (iii) a probabilistic area-time approach is needed for objective drought investigation and evaluation.

CHAPTER II

AREA-TIME MATHEMATICAL MODELS FOR PRECIPITATION PARAMETERS

This chapter presents the mathematical models for the structure of both the point-time series and the areal variation of basic parameters combined as a joint area-time structure necessary to reduce the M regional point-time series of monthly precipitation to M standardized, identically distributed, second-order or third-order stationary stochastic variables, independent in sequence but dependent among themselves. These stochastic components of the precipitation time series are then cross correlated for the study of the areal dependence, and regional models are investigated to determine the relationship of the simple lag-zero cross correlation coefficients with distance between the corresponding pairs of stations.

2.1. Mathematic Model for Time Structure of Monthly Precipitation. Define the random variable values $x_{p,\tau}$ as the precipitation for a given station i $(i=1,2,\ldots,M)$, with M the number of stations studied in a region, $p=1,2,\ldots,n$, the sequence of years, n the sample size expressed in years, $\tau=1,2,\ldots,\omega$, the sequence of intervals within the year, and ω the number of intervals of discrete series in a year (for the monthly precipitation, $\omega=12$). Also define $\varepsilon_{p,\tau}$ as the standardized random variable with the periodicity of the mean and the standard deviation removed from the station series as

$$\varepsilon_{p,\tau} = \frac{x_{p,\tau} - \mu_\tau}{\sigma_\tau} \quad , \qquad 2.1$$

in which μ_τ is the periodic mean and σ_τ is the periodic standard deviation for the station series over the interval ω within the year.

The new variable $\varepsilon_{p,\tau}$ may be a second-order or third-order stationary random process; however, in the case of stationarity, it is either independent or dependent in sequence. Generally, the variable $\varepsilon_{p,\tau}$ may have periodicities in the autocorrelation coefficients and in the third- and fourth-order parameters. The general autoregressive linear model with periodic second-order parameters may be expressed as

$$\varepsilon_{p,\tau} = \sum_{k=1}^{m} \alpha_{k,\tau-k}\varepsilon_{p,\tau-k} + S_{m,\tau}\xi_{p,\tau} \quad , \qquad 2.2$$

$$\text{with} \quad S_{m,\tau} = \left[1 - \sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_{i,\tau-i}\alpha_{j,\tau-j}\rho_{|i-j|,\tau-\iota}\right]^{1/2},$$

where $\iota=\max(i,j)$, m is the order of the model, $\alpha_{k,\tau}$ is the k-th periodic autoregressive coefficient of $\varepsilon_{p,\tau}$ which is a function of the periodic autocorrelation coefficients $\rho_{k,\tau}$, and $\xi_{p,\tau}$ is a second-order stationary and independent stochastic variable.

The periodic parameters μ_τ, σ_τ, and $\alpha_{k,\tau}$ in Eqs. 2.1 and 2.2 are symbolized by ν_τ. The mathematical description of the periodic variation of ν_τ is represented in the Fourier series analysis by

$$\nu_\tau = \nu + \sum_{j=1}^{h(\nu)} C_j(\nu)\cos[\lambda j\tau+\theta_j(\nu)] \quad , \qquad 2.3$$

in which ν is the average value of ν_τ, $C_j(\nu)$ the amplitude, $\theta_j(\nu)$ the phase, j indexes the sequence of harmonics $(j=1,2,\ldots,h)$, and $h(\nu)$ denotes the total number of significant harmonics to be inferred, while $\lambda=2\pi/\omega$ is the basic frequency of the periodic process.

The general mathematical model of the time structure of $x_{p,\tau}$ then becomes

$$x_{p,\tau} = \mu + \sum_{j=1}^{h(\mu)} C_j(\mu)\cos[\lambda j\tau+\theta_j(\mu)] + \left\{\sigma + \sum_{j=1}^{h(\sigma)} C_j(\sigma)\cos[\lambda j\tau+\theta_j(\sigma)]\right\} \cdot \left\{\sum_{k=1}^{m}\left(\alpha_k + \sum_{j=1}^{h(\alpha_k)} C_j(\alpha_k)\cos[\lambda j(t-k)+\theta_j(\alpha_k)]\right)\varepsilon_{p,\tau-k} + S_{m,\tau}\xi_{p,\tau}\right\} \quad , \qquad 2.4$$

in which the arguments μ, σ, S and α_k for C_j and θ_j refer to parameters in the Fourier series description of the periodicity with $h(\mu)$, $h(\sigma)$ and $h(\alpha_k)$ the corresponding numbers of significant harmonics.

An assumption may be made that each periodic parameter has the same number of significant harmonics with the selected value of h so that $h \ll \frac{\omega}{2}$. This is equivalent to stating that the same harmonics are present in each parameter, but with different amplitudes and phases. This may be justified if it can be physically inferred that the precipitation producing factors induce the same type of harmonics. Based on statistical inference techniques used this may not be true; however, it may be true for the effects of the physical process in generating precipitation. If a different number and different harmonics are inferred as significant, the h value becomes $h(\mu)$, $h(\sigma)$ and $h(\alpha_k)$'s, as shown in Eq. 2.4.

When Eq. 2.4 is applied to monthly precipitation with $\omega=12$, the maximum number of harmonics for all periodic parameters is $\omega/2$ or 6. However, it is shown by various studies that for a monthly precipitation process, one, two, or a maximum of three harmonics are sufficient for each periodic parameter. Furthermore, other simplifications may be introduced and supported by regional tests.

A simple model for the monthly precipitation series could be obtained under the following conditions:

(a) Only the first harmonic having a period of 12 months has an amplitude significantly greater than for the non-periodic process;

(b) The autocorrelation coefficients for the precipitation series in general are not periodic, and for monthly series the α_k values, especially α_1, are not significantly different from zero, so that $\alpha_{k,\tau}=0$ and $S_{m,\tau}=1$, with $\varepsilon_{p,\tau}=\xi_{p,\tau}$.

The simplified model, with the periodic μ_τ and σ_τ, and with the above hypotheses of time series structure for monthly precipitation, is

$$x_{p,\tau} = \mu + C_1(\mu)\cos[\lambda\tau+\theta_1(\mu)] + \{\sigma + C_1(\sigma)\cos[\lambda\tau+\theta_1(\sigma)]\}\ \xi_{p,\tau} \quad . \qquad 2.5$$

The $\xi_{p,\tau}$ series is then an independent, standardized, stationary random variable at any station. When $\xi_{p,\tau}$ is computed from $x_{p,\tau}$ by Eq. 2.5, the subscript p,τ may be replaced by i; that is $\xi_{p,\tau}$ by ξ_i, with $i = \omega(p-1) + \tau$.

2.2. Regional Structural Models for Basic Hydrologic Parameters. Let the hypothesis be that the regional variation of any parameter ν can be obtained from the M point estimates v_i $(i=1,2,\ldots,M)$, and is well described in the form of a trend surface function

$$\nu = \Psi(X,Y) \quad , \qquad 2.6$$

with X and Y the coordinates (latitude and longitude) of point positions. In sampling the population function $\Psi(X,Y)$ by a limited number of station points and a limited number of observations for each point during n years, the estimate of both the type of the function $\Psi(X,Y)$ and its coefficients by a sample fitted surface $f(X,Y)$ requires a regression equation such as

$$\hat{\nu} = f(X,Y) + \zeta \quad , \qquad 2.7$$

in which ζ represents the sampling deviations and the differences between the true regional surface function and the fitted function. The larger the number of points and the larger the sample size $N=n\omega$, the smaller should be the variance of ζ, and the better are the estimates of the function and its parameters. By accepting $f(X,Y)$ as the best estimate of $\Psi(X,Y)$, and by removing the trends in all parameters, the random variable ζ is contained within the range of the individual $\xi_{p,\tau}$ series.

Because $\Psi(X,Y)$ is a continuous function, it can always be expanded in a power series form. When a polynomial in X and Y of the order t is selected, Eq. 2.6 becomes

$$\nu = \beta_1 + \beta_2 X + \beta_3 Y + \beta_4 X^2 + \beta_5 Y^2 + \beta_6 XY + \ldots + \beta_k X^t + \beta_{k+1} X^{t-1} Y + \ldots + \beta_{k+t} Y^t + O(X,Y), \qquad 2.8$$

with β_j, $j=1,2,\ldots,k+t$, being regression coefficients to be estimated by the method of least squares for the regional parameter $\hat{\nu}$ from its m estimates v_i, and $O(X,Y)$ being the remaining expansion error.

Under the assumption that a first-order bivariate regression linear trend is adequate in order to account for the major variation over the region of any parameter ν of the monthly precipitation process $x_{p,\tau}$, Eq. 2.8 becomes a simple plane,

$$\nu = \beta_1 + \beta_2 X + \beta_3 Y + O(X,Y) \quad , \qquad 2.9$$

with $O(X,Y)$ an error term.

The boundaries of trend surfaces are greatly affected by the estimates v_i of those stations located near the edges of a region. These estimates may introduce undesirable values of v_i at these edges, as a kind of distortion, when the coefficients β_j of Eqs. 2.8 or 2.9 are estimated by a least squares method. The proper approach in estimating the β_j coefficients is not only for them to be as close as possible to the population values but also to have a minimum of distortion on the trend surfaces at the boundaries. To minimize the boundary effects, the trend surfaces may be fitted to a larger region having more stations, rather than the region under study with M stations. The β_j coefficients of Eqs. 2.8 or 2.9 are then estimated for all stations but are applied only to the small interior region defined by M stations.

To evaluate the fitted function $\hat{\nu} = f(X,Y)$ of $\nu = \Psi(X,Y)$, estimated from the v_i values, the residuals $e_i = v_i - \hat{\nu}$ should be investigated. The standard error of the fitted $\hat{\nu}$ can be used for this evaluation. However, a regional plot of e_i values and an interpretation of isolines of e_i usually will show clusters of subregions with positive and negative e_i values. These "cells" of positive and negative residuals, separated by isolines of $e_i=0$, are the results of areal dependence in the e_i values. Because of a very high correlation between the underlying stochastic stationary process $\{\xi_i\}$ at close points, the estimates v_i of any parameter ν around its true function $\nu = \Psi(X,Y)$ must also be areally dependent. Therefore, large "cells" of e_i with opposite signs, which are located in different sectors of the region under study, may imply that some minor surface trends have not been taken into account by the estimated function $\hat{\nu} = f(X,Y)$. The larger a region, the greater is the probability of having cells distributed randomly over the entire region, and the smaller is the likelihood of inferring by visual inspection of the e_i isolines, that there is an unaccounted trend remaining.

To further simplify the analysis of regional patterns in hydrologic parameters in general, and in parameters of monthly precipitation in particular, some other hypotheses are worth testing. Let the hypothesis be that the variation of the estimated point means $\bar{x}_i$ over the region, with $i=1,2,\ldots,M$, accounts for the variation in all other parameters of Eq. 2.5 such that the surface trends in population values of other parameters are proportional to the surface trend in the population mean. Then Eq. 2.5 becomes

$$x_{p,\tau} = \mu\{1 + \frac{C_1(\mu)}{\mu}\cos[\lambda\tau+\theta_1(\mu)]\} + \mu\{\frac{\sigma}{\mu} + \frac{C_1(\sigma)}{\mu}\cos[\lambda\tau+\theta_1(\sigma)]\}\ \xi_{p,\tau} \ , \qquad 2.10$$

with the ratios $C_1(\mu)/\mu$, σ/μ, and $C_1(\sigma)/\mu$ assumed to be the regional constants, and μ varying as a trend plane $\mu = \beta_1 + \beta_2 X + \beta_3 Y$. If the simple models of Eqs. 2.5 and 2.9 are rejected by proper statistical tests, then the above hypothesis for the model of Eq. 2.10 and the more complex models of Eqs. 2.4 and 2.8 can be used. The phase angles which may be assumed not to vary over the region can be tested as such. As a consequence, three ratios of parameters and two parameters can be studied as they change over the region, namely: $C_1(\mu)/\mu$, σ/μ, $C_1(\sigma)/\mu$, $\theta_1(\mu)$, and $\theta_1(\sigma)$. The hypotheses to be tested are that these three ratios and two parameters are not significantly different from regional constants, being subject only to sampling variation over the region because of the limited number of points m and the limited sample size of n years for the point-time series.

This sampling variation can also be evaluated by observing how the isolines of these three ratios and two parameters vary over the region, whereas the ratio can be evaluated by studying its variation in comparison with the variation of isolines of the numerator of the ratio. If no marked areal trend can be rightfully inferred, the hypothesis can then be accepted for this ratio as a regional constant. Also, the comparison of the original statistics of the numerator of the ratio and the ratio itself may be useful. For example, if the ratio is truly a regional constant, its variance should be small, and much smaller when compared to the variance of the numerator.

2.3. Removal of Time Periodicity and Regional Trends in Parameters. In order to remove both the deterministic periodicity and the deterministic regional trends of parameters, a combined area-time structural analysis is needed. The basic premise is that once the deterministic area-time components in the parameters of the basic random variable of precipitation $x_{p,\tau}$ have been estimated (both the mathematical models and their coefficients) and have been removed from all the point-time series, a second-order stationary independent area-time stochastic process would remain. When this is shown to be the case, then $\{\xi_i\}$ is the basic stochastic process to be studied. To accomplish this, it is first necessary to estimate a minimum set of parameters for Eqs. 2.9 and 2.10, namely the three regression coefficients of Eq. 2.9 and the five regional constant parameters of Eq. 2.10. The stochastic process $\{\xi_i\}$ can then be considered as a multivariate, identically distributed, stationary area-time process, areally dependent but time independent. In other terms, the discrete point series at various stations are mutually dependent, identically distributed time independent variables. To justify this reduction to stochasticity, in this simple area-time process of monthly precipitation, the three main conditions must be satisfied:

(a) All point ξ_i series are approximately standardized variables (with the expected mean of zero and the variance of unity);

(b) They have approximately the same expected lower boundary; and

(c) The skewness coefficients of M point-time series cannot be distinguished statistically from a constant.

2.4. Analysis of Area-Time Stationary Stochastic Component of the Monthly Precipitation Area-Time Process. Once the inferred periodic and trend functions of basic parameters have been used to reduce the set of monthly precipitation time series to a stationary area-time process, the following tests are pertinent:

(a) Test that the process is independent in time;

(b) Test that the point-time series in the region are identically distributed variables; and

(c) Tests for the type of dependence among series, with a development of the regional mathematical dependence model.

The time independence of $\xi_{p,\tau}$ is tested by using the correlograms of individual sample time series of the $\xi_{p,\tau}$ variable, and the average correlogram, with the tolerance limits for independent series drawn about the expected correlograms. The hypothesis of identically distributed $\xi_{p,\tau}$ for all M points in the region is tested by comparing their distribution parameters as estimated from the observed individual M time series. When these two basic tests show that the $\xi_{p,\tau}$ series are time independent, identically distributed variables, the series are designated by ξ_i, with $i=1,2,\ldots,N$, and $N=\omega n$, and the investigation of regional dependence can be undertaken.

The present problem is the investigation of a stationary area-time process, represented by M point-time series, independent in time but dependent areally. It is usually carried out in regard to the areal dependence, using the linear correlation matrix. This matrix is a diagonally symmetric, square matrix with elements $\rho_{i,j}$ and a major diagonal having all elements equal to unity. The $\rho_{i,j}$ correlation coefficient is the simple linear lag zero correlation coefficient between the i-th and j-th series.

Let $d_{i,j}$ be defined as the interstational distance between the series of the i-th and j-th stations, and $\phi_{i,j}$ be the corresponding azimuth of the straight line connecting these two stations. Then

$$d_{i,j} = [(X_i-X_j)^2 + (Y_i-Y_j)^2]^{\frac{1}{2}} \quad , \qquad 2.11$$

and

$$\phi_{i,j} = \tan^{-1}\left[\frac{X_i-X_j}{Y_i-Y_j}\right] \quad , \qquad 2.12$$

for X_i, Y_i and X_j, Y_j, the respective coordinates for stations i and j.

The hypothesis is that the relationship between the correlation coefficients and the four parameters X, Y, d and ϕ is a continuous, positive definite function of the form

$$\rho = \Psi(X,Y,d,\phi) \quad , \qquad 2.13$$

with ρ any $\rho_{i,j}$ value, and X, Y, d and ϕ the corresponding values of X_i, Y_i, $d_{i,j}$ and $\phi_{i,j}$. The X and Y, as the longitude and latitude of the station position, imply that the relation $\rho=f(d,\phi)$ may change from one point to another inside the region. Generally, the simple linear correlation coefficient between the ξ series at stations i and j is a function of absolute position of one of the two stations, the interstation distance, and the orientation of the line connecting the two stations.

The approximation of the unknown population function $\Psi(X,Y,d,\phi)$ is made by a selected function $f(X,Y,d,\phi)$ with its parameters estimated from a limited number of points and a time series of sample size N at each point. This results in

$$\hat{\rho} = f(X,Y,d,\phi) + \zeta \quad , \qquad 2.14$$

in which ζ are the deviations resulting both from the use of an inappropriate function, and from the sampling errors in estimating its coefficients due to the sampling errors in the individual $r_{i,j}$ correlation coefficients used as estimates of $\rho_{i,j}$. The larger the sample size N and the larger the number M of points, the smaller should be the variance of ζ.

It is difficult to study, test, and infer the effect of the position (X,Y) on $\rho_{i,j}$ for given d and ϕ. For that purpose, much more information is needed. Generally, the effect of (X,Y) position is significant when the region is extremely uneven topographically, or when the precipitation over the different subregions varies. However, for a topographically homogeneous region, the assumption that $\rho_{i,j}$ is independent of the absolute position seems to be justified.

It can easily be shown that the $\rho_{i,j}$ values are functions not only of the distance $d_{i,j}$ but also of the orientation of the connecting straight line, measured by $\phi_{i,j}$. However, the effect of $\phi_{i,j}$ in most cases is much smaller than the effect of $d_{i,j}$; thus Eq. 2.14 can be simplified to $\rho=f(d)$ only as a first general approximation.

When there is some evidence that $\rho_{i,j}$, for given $d_{i,j}$ and $\phi_{i,j}$, varies in the region with X and Y, proper statistical tests are needed to support the acceptance of the hypothesis that $\rho_{i,j}$ is independent of absolute position. In a region with a very dense network of precipitation stations, this hypothesis can be meaningfully tested by subdividing the region and testing whether the functions of $\Psi(d,\phi)$, estimated for each subregion, deviate significantly among themselves. If the hypothesis of $\rho_{i,j}$ being independent of $\phi_{i,j}$ is advanced, then the frequency distribution of $r_{i,j}$ is tested for their circular distribution.

For this study of the $\Psi(d,\phi)$ function, a definite relationship of particular properties is required in advance. The range of ρ for the function should be between 1 and 0 for all values of d. For $d=0$ by definition $\rho=1$, and for $d=\infty$, ρ should be zero, because for two widely spaced precipitation stations, the variables ξ_i and ξ_j should be independent. Functions which satisfy these conditions are available in the literature.

Several functions relating the estimated interstation correlation coefficients $\rho_{i,j}$ with the interstation distance $d_{i,j}$ and the azimuth $\phi_{i,j}$ have been selected for this study. They are listed in Table 2.1. Details regarding these functions are discussed in Chapters 3 and 4, where the application of these models to actual data is made. Models I and II of Table 2.1 follow from the work by Caffey [18] but have the disadvantage that the intercept is not $\rho=1$ for $d=0$. Model II is the first term of Model I, with ρ independent of the orientation between stations. Model III is used by Steinitz et. al. [14], and has bias introduced by the parameter A because ρ converges to 1-A as d goes to infinity, which is contrary to the hypothesis of $\rho=0$ for $d=\infty$. This model can be used despite ρ converging to 1-A, because:

(a) The limits of the region studied may be insufficient to provide the large d values needed for showing how ρ behaves at the right extreme;

(b) Model III admits the lower bound of $\rho=0$ for $A=1$, which is the anticipated value for A when d becomes a large value, with B negative; and

(c) The intercept of $\rho=1$ for $d=0$ is of greater importance than $\rho=0$ for the very large values of d.

Model IV of Table 2.1 is similar to Model III. It has the linear term Cd which could possibly give a negative ρ for large d. Whether ρ will be negative depends on the limits of the observed values of r and d. However, the obtained relationship $\rho=f(d)$ is only valid for estimating the coefficients in the relation equation within the range of the observed r and d.

TABLE 2.1

Regional Dependence Mathematical Models for the Stationary Stochastic Components of Monthly Precipitation Series

Model Number	Function
I	$\rho = A \exp(Bd + Cd \cos 2\theta + Dd \sin 2\theta)$
II	$\rho = A \exp(Bd)$
III	$\rho = A \exp(Bd) + 1-A$
IV	$\rho = A \exp(Bd) + Cd + 1-A$
V	$\rho = A + Bd + Cd^2 + Dd^3 + Ed^4 + Fd^5$
VI	$\rho = (1 + Ad)^{-1}$
VII	$\rho = (1 + Ad)^{-n}$
VIII	$\rho = \exp(Ad)$
IX	$\rho = \exp(Ad)/(1 + Bd)$
X	$\rho = (1 + Ad)^{-0.50}$

Model V is a fifth order polynomial but is included in this study for comparative purposes. This model also has the property that $\rho=A$ for $d=0$, and ρ may not be defined for d infinite. The remaining models fulfill the boundary conditions of $\rho=1$ for $d=0$ and $\rho=0$ for $d=\infty$.

Model VI is a simple one-parameter function, which lends itself to a linear transformation. Such a transformation takes the form of

$$\frac{1}{\rho} = 1 + Ad \quad . \qquad 2.15$$

For $d=0$, $\rho=1$, and for d very large or approaching infinity, ρ approaches zero, as should be.

Model VII does not require the exponent n to be 1 and as a result inhibits simple linearization. Model VIII is a simple exponential decay. It can be linearized, as can Model II, but with one less coefficient. Model IX which is a combination of Models VI and VIII has the advantages of the exponential decay for small d and the inverse relationship to d for large d, in cases where the exponential decay may otherwise force ρ to zero too rapidly. Model X forces the exponent $n=0.50$ of Model VII in an attempt to modify Model VI.

Models of Table 2.1 represent the simplified relations between the interstation lag-zero cross correlation coefficient and the interstation distance. Nothing restricts d except that it is the distance between the i-th and j-th stations. A matrix of d, with the elements being the $d_{i,j}$ values for a given set of M stations, can be obtained for regions with existing stations, or for regions with stations specified as the intersections of a cartesian grid. In any case, the dimension M and elements $d_{i,j}$ determine the matrix of correlation coefficients through the use of a selected relationship, $\rho=\Psi(d)$, defining the space dependence of the ξ_i process, if ρ has been determined to be independent of ϕ.

APPLICATION OF MODELS TO REGION I

A joint area-time model of the important regional parameters describing the monthly precipitation series over a set of points is presented in this Chapter. This model is designed for a region which covers parts of North Dakota, South Dakota, and the state of Minnesota. Some parameters are studied as regional constants, while the others vary over the region. The models of Chapter II allow the reduction of the random part of monthly precipitation into an ensemble of stationary, identically distributed stochastic variables, independent in time and dependent among themselves.

The demonstration of the application of models in Region I, with a large number of station series, starts with a reduction of the applicable area. Applicable stations are limited to those in a subregion having an area internal to the total area. This internal area contains only about one-half of all stations, while all other stations are used only to minimize the bias of the model at the boundaries of the reduced area.

3.1. Data Assembly for Region I. Region I was selected as an area having a relatively simple variation of the basic parameters. This required mild topographical variations over the region. Mountain ranges, where sudden changes of precipitation parameters are found, were excluded from this study.

The area was selected in such a way as to satisfy the criteria used in the Hydrology Data Unit of the Hydrology and Water Resources Program of the Department of Civil Engineering at Colorado State University:

(1) A minimum continuous series of 40 years of monthly precipitation;

(2) Allow for a change in station location during the period of observation of less than one mile in horizontal direction and less than 100 feet in elevation; and

(3) No more than three years of missing data estimated by using data of adjacent stations for any one series during the period of observation.

Region I in the North Central Continental United States covers an area of 53,300 square miles. It lies between 92.50 and 100.00 degrees west longitude, and 43.75 and 47.75 degrees north latitude.

Seventy-seven precipitation stations were selected for use in investigations, each with 40 years of monthly values (N=480 values) for the period 1931-1970. The position of Region I within the U.S.A. is shown in Fig. 3.1. The locations of the 77 stations are given in Fig. 3.2., upper graph, with the coordinates origin at 100.00 degrees west longitude and 43.75 degrees north latitude. This origin is used for many of the graphs which demonstrate the analysis of the regional distribution of parameters. Table 1 in the Appendix gives the station identity number, which is identical to the U.S. Weather Bureau index number, station name, degrees west longitude, degrees north latitude, feet above the mean sea level, the 40-year monthly mean and the 40-year monthly standard deviation. The index number is prefixed with 21 for Minnesota, 32 for North Dakota, 39 for South Dakota and 47 for Wisconsin (one border station included). No station has more than 5.62 percent of its monthly values estimated by data from the adjacent stations, and 30 stations had no missing monthly values. Of the 36,960 monthly values, 0.62 percent are estimated by using the normal ratio method, Clark and O'Connor [28].

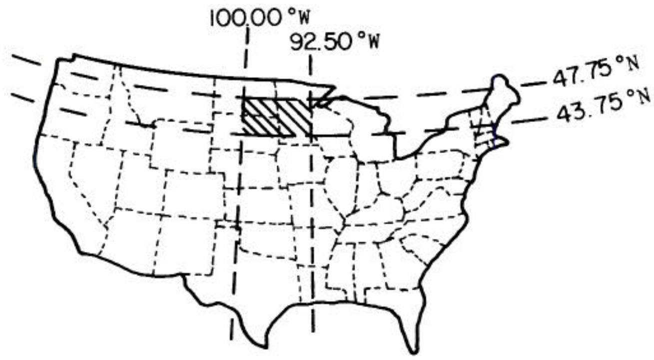

Fig. 3.1. Location within U.S.A. of Region I, used as the example for the regionalization of parameters of monthly precipitation series.

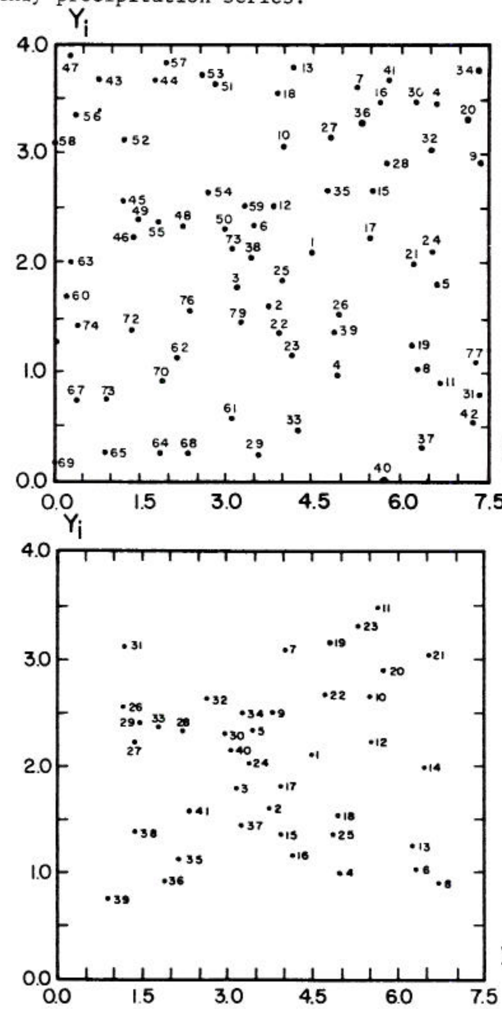

Fig. 3.2. Positions of 77 stations of Region I, upper chart; the positions of 41 stations of Region I, lower chart.

An area with only 41 stations was placed internal to the area of the 77 stations. Figure 3.2., lower graph, shows these stations, and Table 2 in Appendix indicates the station number and percent of area of the Thiessen polygons for each of the 41 stations. The objectives in selecting an internal area having only 41 stations are: (1) to minimize the boundary distortions in fitting the regional trend functions for parameters, and (2) to provide a closed network for the outer bounds in the Thiessen polygon net for the 41 stations by using the entire 77 stations.

General precipitation features given in Fig. 3.3, show the average of monthly means, $\overline{x}_i$, of the 77 stations. It changes almost linearly from the western boundary (1.5 inches) to the eastern border (2.5 inches), with a slight south-north decrease. The general standard deviation, s_i , of monthly series varies in a like manner, increasing from the western border (1.5 inches) to the eastern border (2.0 inches).

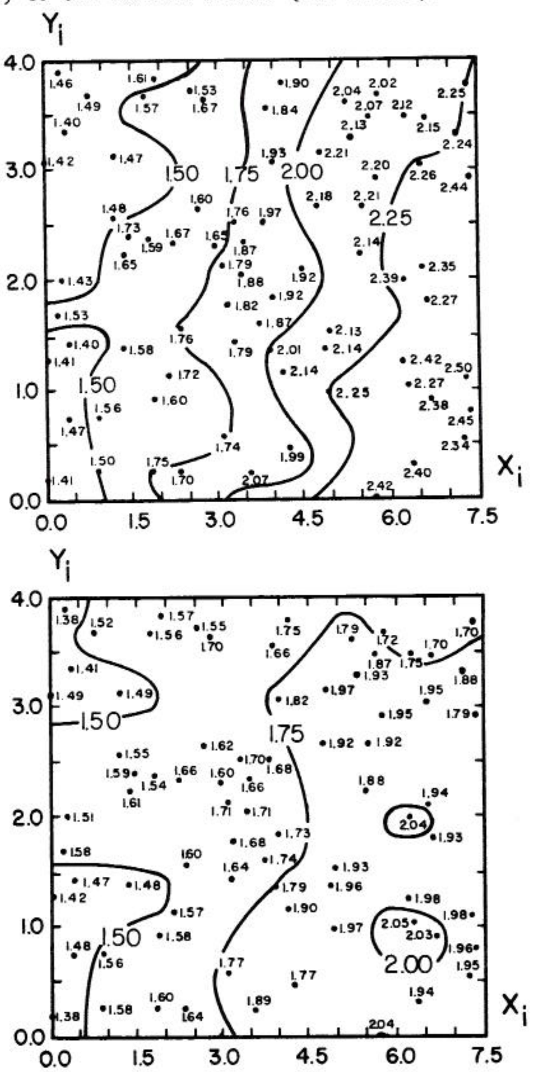

Fig. 3.3. Isolines of the 40-year general monthly mean $\overline{x}_i$, upper graph, and the 40-year monthly standard deviation s_i , lower graph.

Major storms usually enter the region from the southwest and move through the region toward the northeast central zone. The eastern sector receives cold polar air from the north during most of the year. For prolonged periods moist warm air is supplied from the south, mainly from the Gulf of Mexico.

3.2. <u>Time Variation of Parameters</u>. The stochastic component $\xi_{p,\tau}$, assumed to be a second-order stationary and ergodic process, is obtained from the monthly precipitation series $x_{p,\tau}$ at any station i , $i=1,2,\ldots,M$, with $p=1,2,\ldots,n$, and $\tau=1,2,\ldots,\omega$, M the number of station, n the number of years, and ω the number of intervals within the year. In the case of Region I, M=77 , n=40 and ω=12 for the large area, and M=41 , n=40 and ω=12 for the subregion.

The hypothesis is that the monthly means follow the periodic cycle of the year, represented by the Fourier series of Eq. 2.3. For the estimated sample monthly means, m_τ , the fitted Eq. 2.3 is then of the form

$$\mu_\tau = \overline{x} + \sum_{j=1}^{h(\mu)} C_j(\mu)\cos[\lambda j\tau+\theta_j(\mu)] \quad , \qquad 3.1$$

in which $\overline{x}$ (or $\overline{m}_\tau$) is the average of ω values of m_τ. The ratio of the variance $s^2(\mu_\tau)$ of the fitted μ_τ to the variance $s^2(m_\tau)$ of the estimated monthly means m_τ, is used to select the cut-off point in determining the significant $h(\mu)$ harmonics, since the ratio increases with an increase of the number of harmonics $h(\mu)$.

Equation 2.3 is applied similarly to that done for Eq. 3.1 and μ_τ to obtain the fitted periodic function σ_τ for the estimated monthly standard deviations s_τ . This gives

$$\sigma_\tau = \overline{s}_\tau + \sum_{j=1}^{h(\sigma)} C_j(\sigma)\cos[\lambda j\tau+\theta_j(\sigma)] \quad , \qquad 3.2$$

with $\overline{s}_\tau$ the mean of the ω values of s_τ . Once $h(\mu)$ and $h(\sigma)$ have been inferred, the differences $(m_\tau-\mu_\tau)$ and $(s_\tau-\sigma_\tau)$ are considered the random sampling variations. This means that the annual cycle in basic parameters m_τ and s_τ is accounted for with only $h(\mu)$ and $h(\sigma)$ significant harmonics, respectively.

For the m_τ and s_τ series studied, $h(\mu)=h(\sigma)=1$ is hypothesized and tested by statistical analysis of all M time series. Table 3.1 presents the percent of variance of both the monthly mean, m_τ, and the monthly standard deviation, s_τ , explained by the fitted 12-month harmonic of μ_τ and σ_τ for the area with the entire 77 stations. From 83.1 to 98.3 percent of the variance, or on the average 92.6 percent, is explained by the 12-month harmonic in the case of m_τ , and from 75.5 to 98.4 percent of the variance, or on the average 91.1 percent, is explained by this first harmonic in the case of s_τ . The second (6-month) harmonic explains only an average of 3.42 percent of the variance of m_τ , and 2.66 percent of s_τ , while the third (4-month) harmonic has the corresponding figures of 1.02 and 2.21 percent. The average unexplained variances of m_τ and s_τ by the first harmonic, 7.4 and 8.9 percent, respectively, mean that on the average the remaining five harmonics (second through sixth) would explain only 1.5 and 1.8 percent, respectively. The differences in the order of magnitude of contribution between the first and all other five harmonics are large; thus a decision to retain only the first harmonic in μ_τ and σ_τ seems justified. The gain in the explained variance due to the second and the third harmonics is not commensurate with the additional work required or the accuracy of the additional parameters required.

To test the significance of harmonics, the distributions of amplitudes of harmonics should be known under the assumption that both m_τ and s_τ are periodic. A criterion that at least 90 percent of the

TABLE 3.1

Percent of Variance of Monthly Means m_τ and Monthly Standard Deviations s_τ , Explained by the First Harmonic (12-month harmonic) of μ_τ and σ_τ . (Underlined figures denote extremes).

Station Number	For Monthly Mean m_τ	For Monthly Standard Deviation, s_τ	Station Number	For Monthly Mean m_τ	For Monthly Standard Deviation, s_τ
1	.932	.928	40	.938	.976
2	.974	.878	41	.965	.941
3	.939	.933	42	.964	.889
4	.951	.960	43	.877	.935
5	.950	.871	44	.928	.917
6	.934	.956	45	.907	.960
7	.947	.894	46	.903	.968
8	.936	.933	47	.901	.955
9	.959	.945	48	.913	.964
10	.934	.904	49	.880	.915
11	.950	.847	50	.879	.958
12	.901	.949	51	.933	.933
13	.924	.806	52	.914	.924
14	.976	.984	53	.905	.968
15	.952	.914	54	.937	.934
16	.960	.905	55	.907	.899
17	.952	.915	56	.936	.919
18	.923	.908	57	.955	.942
19	.942	.889	58	.859	.891
20	.983	.947	59	.914	.950
21	.941	.849	60	.841	.848
22	.944	.904	61	.895	.948
23	.930	.947	62	.946	.905
24	.907	.783	63	.875	.933
25	.967	.909	64	.927	.818
26	.935	.870	65	.934	.958
27	.935	.938	66	.868	.895
28	.945	.909	67	.894	.936
29	.917	.911	68	.926	.938
30	.983	.982	69	.877	.755
31	.933	.868	70	.923	.911
32	.971	.980	71	.902	.882
33	.937	.818	72	.857	.800
34	.977	.965	73	.869	.833
35	.940	.960	74	.831	.815
36	.969	.947	75	.926	.969
37	.956	.937	76	.917	.959
38	.914	.956	77	.933	.761
39	.920	.873			
			Average	0.926	0.911

variance of m_τ for 80 percent of the station series for n=40 be explained by μ_τ , and a criterion that at least 85 percent of the variance of s_τ for 80 percent of the station series for n=40 be explained by σ_τ , are simple objective approaches to determine $h(\mu)$ and $h(\sigma)$, respectively. By this criterion, all m_τ and s_τ parameters of the 77 stations in Region I may be adequately approximated by only the 12-month harmonic of the μ_τ and σ_τ periodic functions. The Fisher g-test is used and is premised on having a normal population distribution of residuals about μ_τ and σ_τ . Under this hypothesis the critical g values for μ_τ and for 12 values are 0.683 and 0.788 at the five and one percent levels for the first harmonic, and 0.367 and 0.399 at the five and one percent levels for the second harmonic, respectively. At both the five and one percent levels, the 12-month harmonic has an amplitude significantly greater than zero in fitting μ_τ and σ_τ functions to m_τ and s_τ in all but station numbers 24, 69 and 77 for the one percent level. At the five percent level, 34 stations have the second harmonics significant in the case of m_τ and 19 stations in the case of s_τ . The hypothesis of normal distribution of deviations of m_τ or s_τ from μ_τ and σ_τ may be well satisfied. However, these deviations $m_\tau - \mu_\tau$ and $s_\tau - \sigma_\tau$ are mutually dependent, and because of the periodicity in μ_τ and σ_τ , the variances of these deviations are also periodic. Therefore, the Fisher's test is not applicable, though it gives the approximate results. However, in 64 of the 77 stations, or 83.2 percent, 90 percent of the variance of m_τ is explained by only the 12-month harmonic of μ_τ , while for 65 of the 77 stations, or 84.3 percent, 85 percent of the variance of s_τ is explained by only the 12-month harmonic in σ_τ .

Figure 3.4 gives the correlogram and the graph of spectral densities for the first station out of the 77 station series given in Table 1 of the Appendix. The correlogram is very close to be the 12-month cosine function, while the spectrum has a significant spike only at the 12-month frequency. Similar spectra can be shown for nearly all 77 series. These spectra also support the conclusion that the first harmonic in the m_τ series, computed from the $x_{p,\tau}$ series, is the most important, and the higher-order harmonics (the second through the sixth) can be neglected.

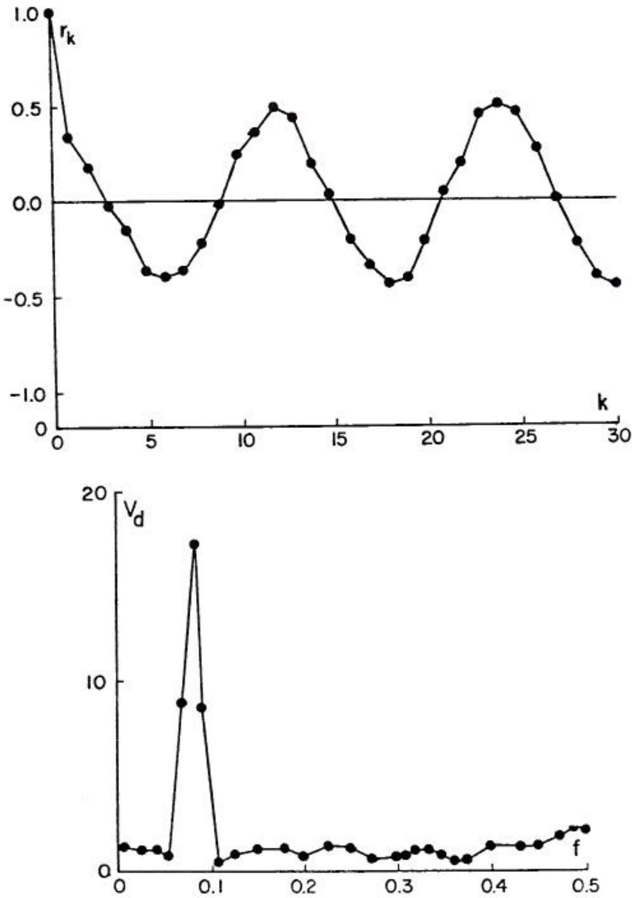

Fig. 3.4. Correlogram (upper graph) and spectrum (lower graph) of the first precipitation series (station 1) of Table 1 in the Appendix.

The values of $h(\mu)=h(\sigma)=1$ satisfy the objective of obtaining a minimum number of parameters using a minimum number of significant harmonics, and in nearly all cases the 12-month harmonic is required to account for the annual cycle in monthly means and monthly standard deviations. Therefore, only the first harmonic is necessary for both the μ_τ and σ_τ functions in Region I.

3.3. Regional Variation in Parameters. Because only the 12-month harmonic is selected to account for the annual cycle in the basic parameters of $x_{p,\tau}$ for all stations, interest is focused on how the basic parameters, represented by $\bar{x}$ (or $\bar{m}_\tau$) , $C_1(\mu)$, $\theta_1(\mu)$, $\bar{s}_\tau$, $C_1(\sigma)$, and $\theta_1(\sigma)$, vary over the Region I.

Of major interest is the general sample monthly mean, $\bar{x}$ or $\bar{m}_\tau$. Figure 3.3 shows its overall variation within Region I for the 40-year observation period. Because of a marked west to east increase and a slight south to north decrease, the simple linear plane trend for $\bar{x}$ over Region I is investigated first instead of starting with a quadratic or higher-order polynomial function. The hypothesis of regional variation in $\bar{x}_i$ is then

$$\mu_i = \beta_1 + \beta_2 X_i + \beta_3 Y_i \quad , \qquad 3.3$$

in which μ_i denotes the fitted regional means $\bar{x}_i$ and β_1 , β_2 and β_3 are the regression coefficients to be estimated. X_i , the longitude coordinate, is referenced to X_o=100.00 degrees west longitude as the zero abscissa, and Y_i , the latitude coordinate, is referenced to Y_o=43.75 degrees north latitude as the zero ordinate. Table 3.2 presents the general statistics for the sample series of 77 stations. The parameter $\bar{x}_i$ has the general average 1.902, the variance 0.107, the relatively small skewness coefficient 0.091, and the excess coefficient -1.243. Equation 3.3, by using the simple least-square approach for estimating β_1 , β_2 , and β_3 , becomes

$$\mu_i = 1.45870 + 0.14095X_i - 0.04452Y_i \quad . \qquad 3.4$$

With the statistically significant regression coefficients as shown in Eq. 3.4, this trend plane gives an explained variance of 94.8 percent for the regional variation of $\bar{x}_i$. This high percentage of explained variance, using such a simple model as Eq. 3.4 and the simple least-square estimation method, is as much a surprising result as that obtained in explaining the variation of m_τ and σ_τ by only the first harmonic.

Figure 3.5 presents the differences between the estimated sample means $\bar{x}_i$ and the fitted model of means μ_i by Eq. 3.4. The isolines of differences are drawn with a stress given to the zero isoline. The differences show that the sample means, $\bar{x}_i$, are larger than the model means μ_i , in the middle, southwestern, and northeastern sectors, with a maximum value of 0.13 inches. The negative differences dominate the northwest, eastcentral, and southeastern sectors, with a minimum value of -0.20 inches.

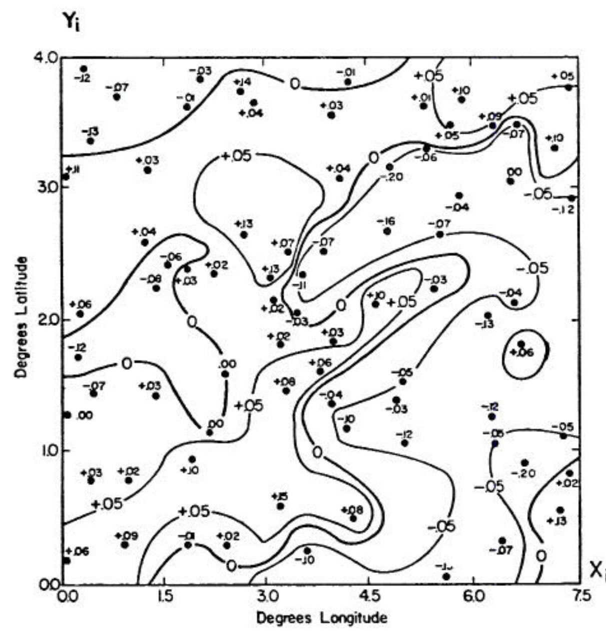

Fig. 3.5. Differences, and isolines of differences of the estimated means and the fitted mean by simple least-square approach. The point denotes the station location.

TABLE 3.2

Average Statistics for Regional Parameters
for the Set of 77 Station Series

Regional Parameter	Average	Variance	Standard Deviation	Coefficient of Variation	Skewness Coefficient	Excess Coefficient	Chi-Square of Normal Fit
$\overline{x}_i$	1.902	.107	.327	.172	.091	-1.243	
$C_{1,i}(\mu)$	1.537	.026	.162	.105	-.027	- .800	
$\theta_{1,i}(\mu)$	2.840	.011	.107	.038	.086	- .455	14.31
$\overline{s}_{\tau,i}$	1.185	.021	.147	.124	.128	- .958	
$C_{1,i}(\sigma)$	.772	.009	.099	.128	.078	- .380	
$\theta_{1,i}(\sigma)$	2.705	.025	.160	.059	.082	- .097	6.02
$C_{1,i}(\mu)/\overline{x}_i$	.818	.005	.074	.090	.292	.373	20.71
$\overline{s}_{\tau,i}/\overline{x}_i$	.629	.002	.048	.076	-.043	- .057	3.90
$C_{1,i}(\sigma)/\overline{x}_i$	.412	.007	.055	.134	.074	- .737	7.32

The fit by Eq. 3.4 was made by using the least square method, giving each $\overline{x}_i$ value the same weight. Figure 3.3, lower graph, shows that the standard deviation s_i increases similarly as $\overline{x}_i$ does. A more proper approach in estimating β_1, β_2 and β_3 is by giving weights $1/s_i$ to the deviations $(\overline{x}_i-\mu_i)$. This makes the relative deviations $(\overline{x}_i-\mu_i)/s_i$ have a more homogeneous variance than was found in the case of using only the $(x_i-\mu_i)$-deviations for the least square fitting. The weighted deviations can represent an improvement in the estimation of models and coefficients of the type of Eq. 3.3. For example, by using this weighted approach for the case of $\overline{x}_i$, $\hat{\beta}_1=1.4518$, $\hat{\beta}_2=0.1394$, and $\hat{\beta}_3=-0.0405$, with an explained variance for $\overline{x}_i$ of 98.04 percent instead of the 94.8 percent shown for the equal-weight approach. Figure 3.6 shows the $(\overline{x}_i-\mu_i)$-deviations and their isolines for the case of weighted least-square approach.

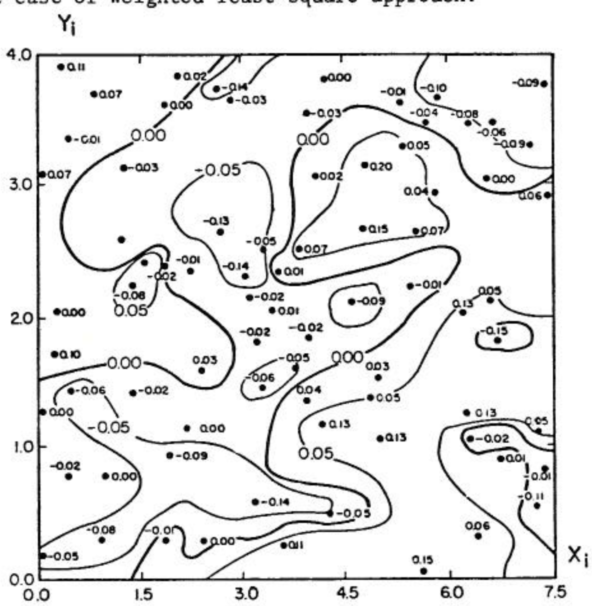

Fig. 3.6. The same points and isolines as in Fig. 3.5 except that the weighted least-square estimation approach is used for standardized deviations.

The clustering of positive and negative differences $(\overline{x}_i-\mu_i)$ into cells, as shown in Figs. 3.5 and 3.6, may be the result, as postulated earlier, of the areal dependence amongst the station series. In spite of the individual differences, these cells may be inferred as being randomly distributed over the region, separated by the zero isoline. No particularly strong trend seems present in the areal distribution of the cells. Therefore, the remaining 5.2 or 1.96 percent of variation in $\overline{x}_i$ unaccounted for by the model is attributable to sampling variation, and the mean $\overline{x}_i$ changes approximately linearly over Region I.

The five other regional parameters, $C_{1,i}(\mu)$, $\theta_{1,i}(\mu)$, $\overline{s}_{\tau,i}$, $C_{1,i}(\sigma)$, and $\theta_{1,i}(\sigma)$, are also hypothesized to have linear regional variation. Their general statistics are given in Table 3.2. The isolines of the amplitude $C_{1,i}(\mu)$, of the 12-month harmonic in the mean, are shown in the upper graph of Fig. 3.7. These isolines suggest an east-west trend with the lowest values in the west (1.3 isoline) and the highest values in the east (1.7 isoline). On the first glance, the variation does not seem linear as was found for $\overline{x}_i$. Under the hypothesis of fitting a trend plane for the regional variation in $C_{1,i}(\mu)$, the 77.1 percent of its variance is explained by

$$C_{1,i}(\mu) = 1.33470 + 0.06329X_i - 0.01816Y_i \quad . \quad 3.5$$

Because $C_{1,i}(\mu)$ is a function of the second moment of variable values, the expected sampling variation of deviations about a fitted trend plane should be greater than for the mean $\overline{x}_i$. The amplitude $C_{1,i}(\mu)$ has a regional mean 1.537, variance 0.026, and skewness coefficient -0.027. It is expected that the weighted least-square fit would increase the explained variance of $C_{1,i}(\mu)$.

Because the trend planes are fitted to both $\overline{x}_i$ and $C_{1,i}(\mu)$ as the first approximations, another hypothesis is advanced, namely that the ratio $C_{1,i}(\mu)/\overline{x}_i$ is a regional constant. This is equivalent to stating that the regional variation in $C_{1,i}(\mu)$ is accounted for by the regional variation in $\overline{x}_i$. It is expected that the variation in this ratio about its mean value would be highly reduced. The mean for 77 regional values of this ratio is 0.818, the variance is 0.005 and the skewness coefficient 0.292. Because of the very small value of the variance of $C_{1,i}(\mu)/\bar{x}_i$, this ratio may not be significantly different from its mean value. In other words, the inference is that the variation of the ratio about its mean is only the result of the sampling variation. The mean ratio is then considered to be a regional constant. Figure 3.7, lower graph, shows how this ratio changes over Region I. The 0.90 and 0.80 isolines are north-south oriented, with the regional variation highly reduced. The variance of $C_{1,i}(\mu)/\bar{x}_i$ of 0.005 is only 18 percent of the variance of 0.026 of $C_{1,i}(\mu)$. The 22.9 percent of the variance 0.026 of $C_{1,i}(\mu)$, unexplained by Eq. 3.5, is 0.006; this is greater than the variance of this ratio. In other words, the use of this ratio has a smaller sampling variance than the use of the trend plane of Eq. 3.5 for the amplitude $C_{1,i}(\mu)$.

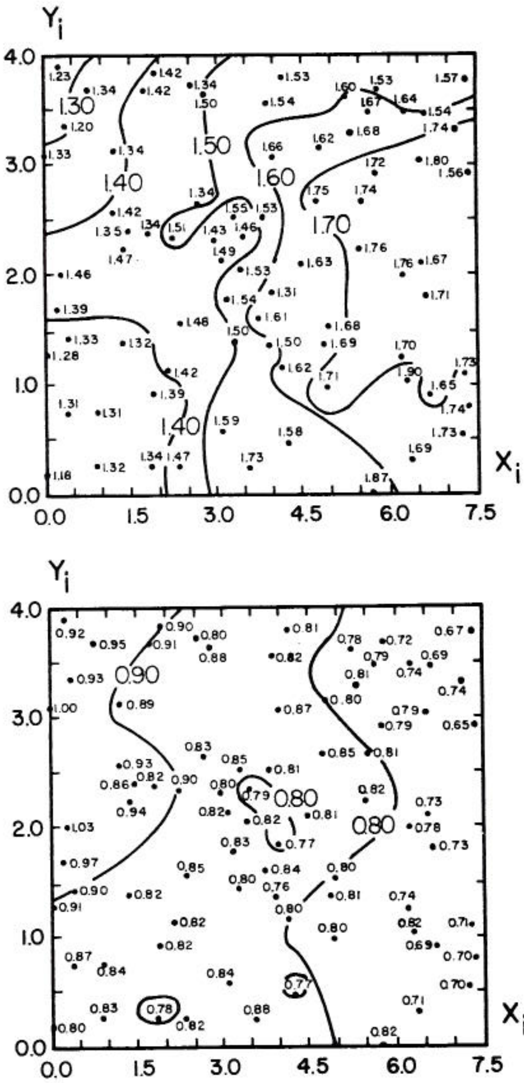

Fig. 3.7. The isolines of the amplitude $C_{1,i}(\mu)$ for the mean μ_τ , upper graph; and the isolines of the ratio $C_{1,i}(\mu)/\bar{x}_i$, lower graph.

Dividing the ratio into equal frequency class intervals, and testing with chi-square statistic for the difference between this ratio and its average value of 0.818, for a normal distribution of the ratio with the mean 0.818 and the variance 0.005, the chi-square is 20.717. This relatively large computed chi-square is mainly obtained because the normal distribution for the ratio might not be the best hypothesis; twice as many ratios fall in the central class interval containing the mean 0.818 in case of normal distribution than in the other class intervals. Because of the high dependence of the ratio over the region, the chi-square test may not be the best test. Therefore, for all practical purposes this ratio may be considered a regional constant.

Figure 3.8, upper graph, gives the isolines of $\bar{s}_{\tau,i}$, the mean of the 12 values of s_τ standard deviations for each of the 77 individual $x_{p,\tau}$ series, while the lower graph gives the isolines of $\bar{s}_{\tau,i}/\bar{x}_i$. Equation 3.2 gives the values for $\bar{s}_{\tau,i}$. There is a difference between the s_i estimates of Fig. 3.3, lower graph, and the $\bar{s}_{\tau,i}$ individual means of s_τ of Fig. 3.8, upper graph; the $\bar{s}_{\tau,i}$ values are smaller than the s_i values. Because of periodicities in the standard deviations, and the use of Eq. 3.2, the regionalization is done for $\bar{s}_{\tau,i}$ rather than for s_i . There is a general increase in $\bar{s}_{\tau,i}$ values from the west to the east, as for the previously studied parameters. The mean for the 77 values of $\bar{s}_{\tau,i}$ is 1.185, the variance 0.021, and the coefficient of skewness 0.128. However, a uniform change from the isoline 1.0 to the isoline 1.2 in Fig. 3.8 suggests that a trend plane may also well represent the regional variation of $s_{\tau,i}$. With this hypothesis the plane is

$$\sigma_i = 1.03990 + 0.05794X_i - 0.03277Y_i \quad . \quad 3.6$$

Equation 3.6 explains 81.6 percent of the variance of $\bar{s}_{\tau,i}$.

Considering the ratio $\bar{s}_{\tau,i}/\bar{x}_i$ as a new regional parameter, which has a mean 0.629, variance 0.002, and skewness coefficient -0.043 for the 77 station series, this reduced variance of 0.002 as compared with the variance 0.021 for $\bar{s}_{\tau,i}$ is about 10 percent. In this case, using the trend plane of Eq. 3.6, the unaccounted variation of $\bar{s}_{\tau,i}$ is approximately 0.039, or about two times the variance of this ratio. By using the 77 estimates of $\bar{s}_{\tau,i}/\bar{x}_i$, and considering them to be normally distributed about their mean, 0.629, as a result of random sampling variation only, the computed chi-square for 11 equal probability class intervals is 3.90. This value is not significant at the 95 percent level because the critical chi-square is 16.9, regardless of the fact that the 77 ratios are mutually dependent values. The ratio $\bar{s}_{\tau,i}/\bar{x}_i$ is selected as a regional constant of 0.629. The isolines and point values of this ratio are also shown in Fig. 3.8, lower graph. Its variation, with the 0.65 isoline in the west and the 0.60 isoline in the east, is now east-west oriented; however, this supports the hypothesis that $\bar{s}_{\tau,i}$ and $\bar{x}_i$ are close to being proportional parameters.

The isolines of the amplitude of 12-month harmonic of the monthly standard deviation are shown in Fig. 3.9, upper graph. The regional variation of $C_{1,i}(\sigma)$ has a general west-east increase. Consequently, for the hypothesis of a trend plane,

$$C_{1,i}(\sigma) = 0.65824 + 0.02697X_i + 0.00534Y_i \quad . \; 3.7$$

The explained variance in $C_{1,i}(\sigma)$ by Eq. 3.7 is only 37.6 percent. The very small regression coefficient for Y_i of 0.00534 shows very small variation in

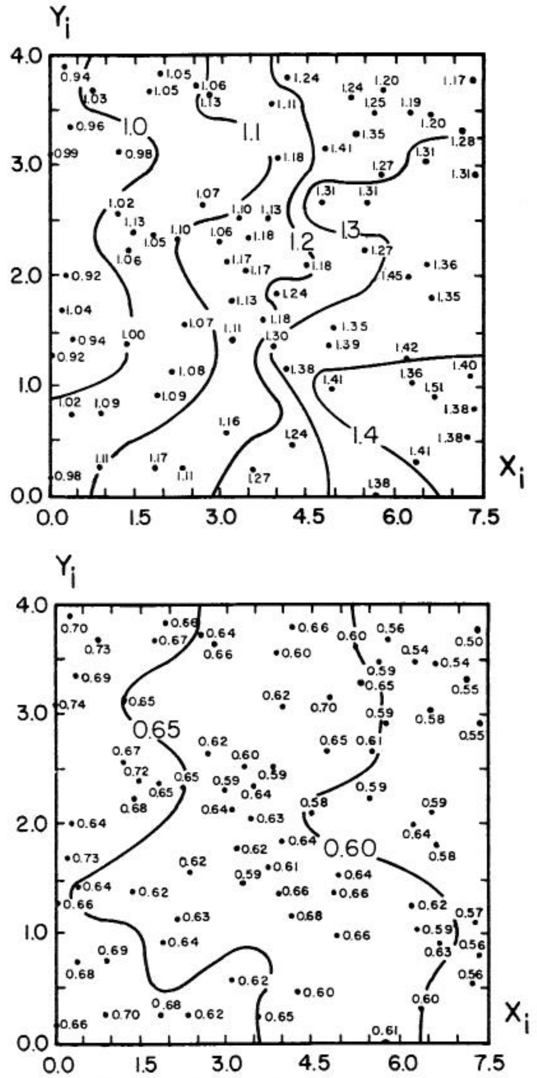

Fig. 3.8. The isolines of the parameter $\bar{s}_{\tau,i}$, the mean of 12 values of s_τ standard deviation, upper graph; and the isolines of the ratio $\bar{s}_{\tau,i}/\bar{x}_i$, lower graph.

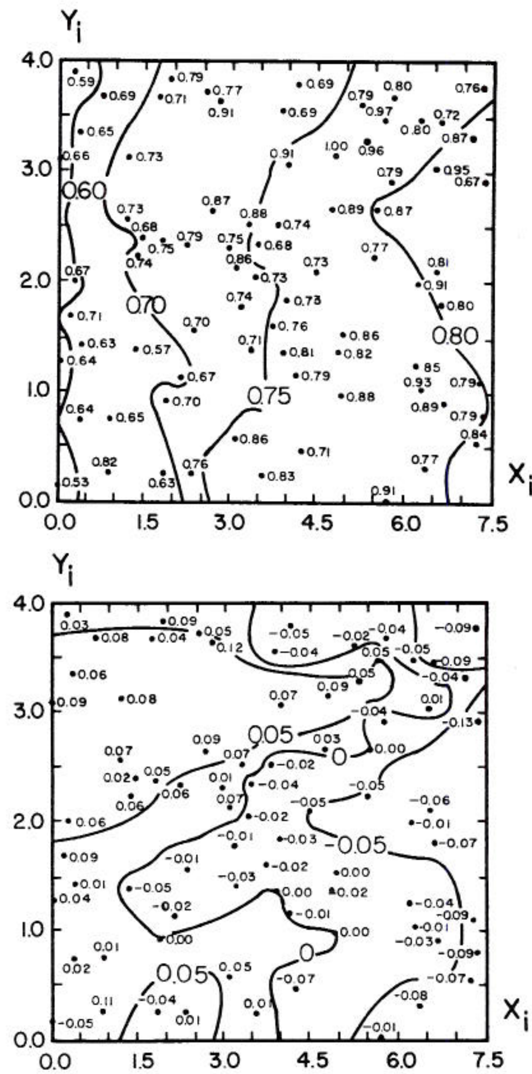

Fig. 3.9. The isolines of the amplitude $C_{1,i}(\sigma)$ of 12-month harmonic of the monthly standard deviation, upper graph and the isolines of deviations of the ratio $C_{1,i}(\sigma)/\bar{x}_i$ from the mean ratio of 0.412, lower graph.

$C_{1,i}(\sigma)$ in north-south direction. The fit by Eq. 3.7 seems relatively poor. The mean of 77 values of $C_{1,i}(\sigma)$ is 0.772, the variance 0.009 and the skewness coefficient 0.078. The unexplained variance of $C_{1,i}(\sigma)$ by Eq. 3.7 is 0.0056. The ratio $C_{1,i}(\sigma)/\bar{x}_i$ has a mean of 0.412, variance 0.007 and skewness coefficient 0.074. The isolines of deviations of this ratio from its mean value are given in Fig. 3.9, lower graph. Only the major isolines are drawn. Though the variance of this ratio of 0.007 is relatively high in comparison with the unexplained variance of $C_{1,i}(\sigma)$ by Eq. 3.7 of 0.0056, or nearly 20 percent greater, this ratio can still be considered as approximately a constant of 0.412, and its variation over the region may also be considered as random sampling variation. Assuming the ratio is normally distributed about 0.412 with variance 0.007, the chi-square is 7.3. Because this chi-square is relatively small, the ratio is inferred to be a regional constant.

The isolines and point values of the phase angles, $\theta_{1,i}(\mu)$ for the periodic μ_τ , and $\theta_{1,i}(\sigma)$ for the periodic σ_τ , are shown in the upper and lower graphs of Fig. 3.10, respectively. For $\theta_{1,i}(\mu)$ the isoline values range from 2.7 to 3.0, and for $\theta_{1,i}(\sigma)$ from 2.5 to 2.9, both increasing in the east-west direction.

The trend planes for these two parameters are

$$\theta_{1,i}(\mu) = 3.05755 - 0.03715X_i - 0.03608Y_i , \quad 3.8$$

and

$$\theta_{1,i}(\sigma) = 2.98795 - 0.05317X_i - 0.03815Y_i . \quad 3.9$$

The variation in $\theta_{1,i}(\mu)$ is hypothesized to result only from sampling variation, normally distributed with mean 2.840 and variance 0.011, and a chi-square of 14.31 is obtained, which is not significant at the 95 percent significance level for 11 equal probability class intervals. Therefore, the parameter $\theta_{1,i}(\mu)$ is practically a regional constant having a value of 2.840. The regional variation in $\theta_{1,i}(\sigma)$ may also be hypothesized to be the result of a sampling variation. Thus the parameter would be a regional constant of 2.705. Assuming a normal distribution of $\theta_{1,i}(\sigma)$, with mean 2.705 and variance 0.025, a chi-square of 6.02 is obtained, which is not significant at the 95 percent level. The east-west pattern in $\theta_{1,i}(\mu)$ and $\theta_{1,i}(\sigma)$ may, therefore, be the result of regional dependence of sampling variation rather than a population characteristic. Because there is no physical reason why the phases for the periodic mean and the periodic variance should be different, a further simplification is obtained by setting $\theta_{1,i}(\mu)=\theta_{1,i}(\sigma)$. In that case either $\theta_1=2.840$ or the average $\theta_1=2.772$ are used. Regardless of the advantages of this unique θ_1 value, the two distinct constants of 2.840 and 2.705 will be used here in further analysis of monthly precipitation.

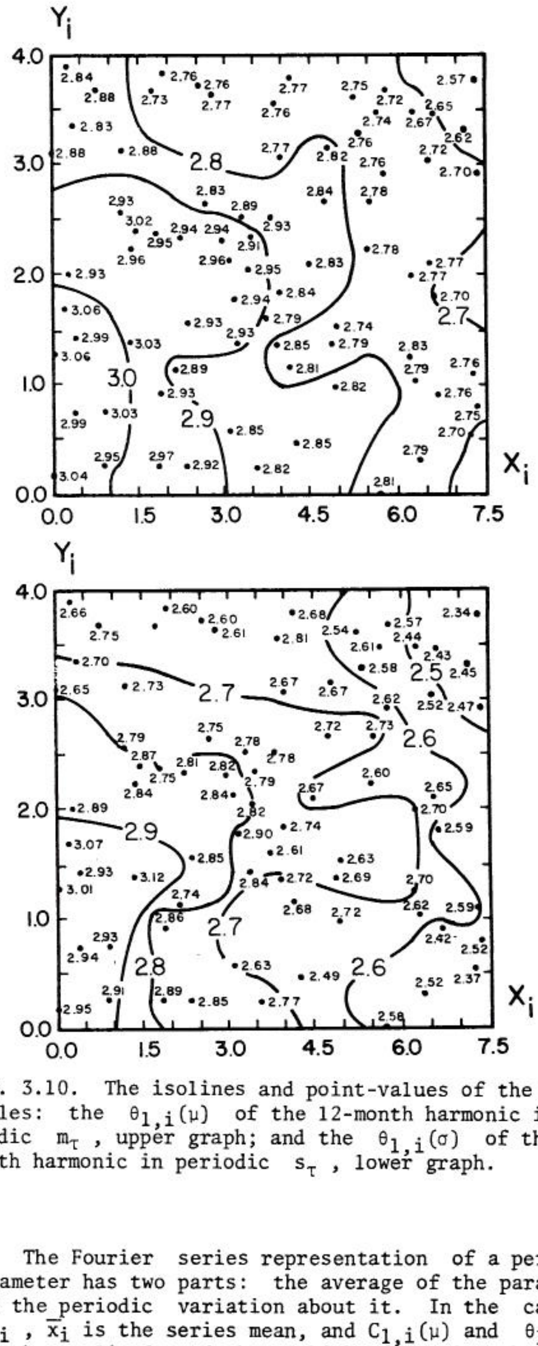

Fig. 3.10. The isolines and point-values of the phase angles: the $\theta_{1,i}(\mu)$ of the 12-month harmonic in periodic m_τ , upper graph; and the $\theta_{1,i}(\sigma)$ of the 12-month harmonic in periodic s_τ , lower graph.

The Fourier series representation of a periodic parameter has two parts: the average of the parameter and the periodic variation about it. In the case of $m_{\tau,i}$, $\overline{x}_i$ is the series mean, and $C_{1,i}(\mu)$ and $\theta_{1,i}(\mu)$ are the amplitude and phase which determine the periodic variation. The regional range in the isolines of $\overline{x}_i$ is approximately 0.75 inches while for $C_{1,i}(\mu)$ it is about half that amount, or 0.40 inches. The coefficient of variation of $\overline{x}_i$ is 0.172, which is 1.63 times that for $C_{1,i}(\mu)$, while the mean of $\overline{x}_i$ is 1.24 times that for $C_{1,i}(\mu)$. The regional variability in $\overline{x}_i$ is greater than in $C_{1,i}(\mu)$. The fact that the major regional variation exists in the mean $\overline{x}_i$ rather than in the periodic fluctuation about the mean, measured by $C_{1,i}(\mu)$, suggests that the trend plane which accounts for the variation in $\overline{x}_i$ is the most important aspect in the regionalization of precipitation parameters.

Applying a similar analysis to $\overline{s}_{\tau,i}$, the monthly standard deviation, the range of regional variation in $\overline{s}_{\tau,i}$ is about 0.40 inches, while in $C_{1,i}(\sigma)$, the first harmonic amplitude of $\overline{s}_{\tau,i}$, it is 0.20 inches. The regional mean of $\overline{s}_{\tau,i}$ is 1.18; this is 1.53 times greater than the mean of $C_{1,i}(\sigma)$ which is 0.772. However, the regional coefficients of variation for $\overline{s}_{\tau,i}$ and $C_{1,i}(\sigma)$ are approximately equal, namely 0.124 and 0.128. Again the major regional variation is in the mean $\overline{s}_{\tau,i}$ rather than in the regional variation of the amplitude of the first harmonic. The linear trend planes, which explain variances of 81.6 percent for $\overline{s}_{\tau,i}$ and 37.6 percent for $C_{1,i}(\sigma)$, support the conclusion that the parameter of periodic variation is the more difficult of the two to regionalize. This is due to a greater regional random sampling variation.

With the above regional models for various parameters, the second-order stationary component, $\xi_{p,\tau}$, for any station series, with both the numerator and denominator divided by $\overline{x}$, becomes

$$\xi_{p,\tau} = \frac{\dfrac{x_{p,\tau}}{\overline{x}} - 1 - \dfrac{C_1(\mu)}{\overline{x}} \cos[\lambda\tau+\theta_1(\mu)]}{\dfrac{s_\tau}{\overline{x}} + \dfrac{C_1(\sigma)}{\overline{x}} \cos[\lambda\tau+\theta_1(\sigma)]} \qquad 3.10$$

Replacing $\overline{x}$ by Eq. 3.4, and the other five parameters $C_1(\mu)/\overline{x}$, $\theta_1(\mu)$, $\overline{s}_\tau/\overline{x}$, $C_1(\sigma)/\overline{x}$, and $\theta_1(\sigma)$ by their values, 0.8187, 2.8402, 0.6293, 0.4120, and 2.7053, respectively, then Eq. 3.10, for a given i-th station, having coordinates X_i and Y_i , becomes

$$\xi_{p,\tau} = \frac{\dfrac{x_{p,\tau}}{1.45870 + 0.14095X_i - 0.04452Y_i} - 1 - 0.8187\cos\left(\dfrac{\pi\tau}{6} + 2.8402\right)}{0.6293 + 0.4120\cos\left(\dfrac{\pi\tau}{6} + 2.7053\right)} \qquad 3.11$$

Three approaches for selecting the sample parameters were considered for determining $\xi_{p,\tau}$, as a stationary stochastic component. The first approach is made using $\xi_{p,\tau}$ of Eq. 3.11 as Model A. The second approach is similar to Model A, except that, in place of the constant ratio $C_{1,i}(\sigma)/\overline{x}$, the constant value 0.772 is used for $C_{1,i}(\sigma)$ in the form of Model B,

$$\xi_{p,\tau} = \frac{\left\{\dfrac{x_{i,p,\tau}}{\overline{x}} - 1 - \dfrac{C_1(\mu)}{\overline{x}} \cos[\lambda\tau+\theta_1(\mu)]\right\} \overline{x}}{\left(\dfrac{\overline{s}_\tau}{\overline{x}}\right) \overline{x} + C_1(\sigma)\cos[\lambda\tau+\theta_1(\sigma)]} \quad , \qquad 3.12$$

which still has five regional contants, but with a linear trend surface for $\overline{x}$. The third approach, Model C, is in computing the $\xi_{p,\tau}$ series utilizing the linear regression equations in the form of linear surface trends, fitted to each of the six regional parameters, $\overline{x}$, $C_1(\mu)$, $\theta_1(\mu)$, $\overline{s}_\tau$, $C_1(\sigma)$, and $\theta_1(\sigma)$. In the case of Model C, using the best estimates of regression coefficients for parameters, a reduction in the total number of parameters is made from 77 x 12 x 2 = 1848 parameters in the case of the use of the nonparametric approach, or the use of all sample values of m_τ and s_τ at the 77 stations, to a maximum of 6 x 3 = 18 regression coefficients of the six most important parameters. For Models A and B the total number of regional regression coefficients to be estimated is reduced to three (β_1, β_2, and β_3) for $\overline{x}$, and the five regional constants, for a total of eight.

The economy in the number of estimates and the simplification in models are the major characteristics of the above regionalization of parameters of monthly precipitation. More complex models are not justified due to time and regional sampling errors. Equation 3.11 is used in reducing $x_{p,\tau}$ to $\xi_{p,\tau}=\xi_i$ random variables. However, further improvements may be obtained in the estimated coefficients of this equation by using the weighted least-square fits for determining the parameters of both the periodic and regional models.

3.4. Testing Sequential Independence of Stationary Stochastic Components. To determine how closely ξ_i, where ξ_i is equal to $\xi_{p,\tau}$, approximates an independent stationary stochastic process, as a result of the simplifying assumptions for the deterministic periodic time and regional trend variations in the basic parameters, the correlogram of each ξ_i series of the 77 monthly precipitation series is tested for significant departures. This is done on the 95 percent probability level, from the correlogram of an independent series. Only the first 20 lags of correlograms are used. The 95 percent tolerance levels, r_u and r_ℓ, for an independent series are computed by

$$r_{u,\ell}(95\%) = \frac{-1 \pm t\sqrt{N-k-2}}{N-k-1} \quad , \qquad 3.13$$

with k lag, t=±1.96 being the deviates from the standard normal distribution for a two-tail test that $\rho_k=0$ for $k>0$ for the ξ_i series, and N sample size. Table 3.3 presents the number of r_k values for the lags 1 through 20 which are outside the tolerance limits for the 77 region stations and the 41 subregion stations of monthly precipitation series. The serial correlation coefficients r_k for the lags 1, 5, 8, 11, 12 and 19 in Table 3.3 are often outside the tolerance limits and for more than 4 series (5 percent of 77). The r_k values of other lags also have some tendency to be outside the tolerance limits. The same observations are valid for the sample of 41 series in the subregion.

TABLE 3.3

Number of Serial Correlation Coefficients of ξ_i Outside the Tolerance Limits (T.L.) of an Independent Series

Lags	Number of Station Series	
	77	41
1	31	14
2	4	2
3	4	3
4	5	1
5	11	7
6	3	1
7	6	1
8	18	8
9	7	4
10	1	0
11	14	1
12	48	24
13	3	0
14	2	2
15	3	2
16	1	0
17	2	2
18	5	1
19	12	1
20	8	2
Total outside T.L.	188	76
Number without values of r_1 and r_{12}	109	38
Total numbered r_k computed	1540	820
Percent outside	12.2	9.3
Percent without values of r_1 and r_{12}	7.1	4.6

The first serial correlation coefficient, r_1, is affected both by some dependence in the monthly precipitation values due to the dependence inherent in hydrometeorological synoptic situations from day to day, and by an incomplete removal of harmonics of the periodic parameters, as the serial correlation coefficient r_{12} demonstrates. By neglecting the numbers of r_1 and r_{12} values outside the tolerance limits, the new percentages of r_k values outside these limits are also given in Table 3.3. In that case, for the 41 station series, only 4.6 percent of r_k values are outside the tolerance limits, instead of 9.3 percent as in the case where r_1 and r_{12} are included.

For example, the correlograms of two station series (nos. 8 and 22) out of the 77 series are presented for ξ_i in Fig. 3.11. For station no. 8 several r_k values are outside (but close to) the tolerance limits, while station no. 22 has a correlogram relatively well confined within the 95 percent tolerance limits.

The correlation coefficient $\text{corr}(r_{1,i}, r_{1,j})$ between the first serial correlation coefficients $r_{1,i}$ and $r_{1,j}$ of the i-th and j-th series of ξ_i is given by [29]

$$\text{corr}(r_{1,i}, r_{1,j}) = \frac{N+2}{N-1}\hat{\rho}^2_{i,j}\left[1 + \frac{N(N-2)\ (r^2_{1,i} + r^2_{1,j})}{2(N+2)(N+4)}\right] \qquad 3.14$$

with $\hat{\rho}_{i,j}$ the estimates of the lag-zero correlation coefficients of the ξ_i series for the i-th and j-th stations. Because the average, $\bar{r}_1$, of all series is about 0.08, its square is about 0.0064 and the term in the brackets of Eq. 3.14 is, therefore, very close to unity. Even if one would use the individual values of $r_{1,i}$ and $r_{1,j}$ instead of using the average, $\bar{r}_1$, to replace $r_{1,i}$ and $r_{1,j}$ in Eq. 3.14, the term in the bracket is still close to unity. Therefore, Eq. 3.14 may be accurately replaced by

$$\text{corr}(r_{1,i}, r_{1,j}) = \frac{N+2}{N-1}\hat{\rho}^2_{i,j} \quad . \qquad 3.15$$

The effective number of stations in the region for $\text{corr}(r_{1,i}, r_{1,j})$, assuming the ξ_i are mutually uncorrelated series,

$$M_e = \frac{M}{1 + \overline{\text{corr}(r_{1,i}, r_{1,j})}(M-1)} \quad , \qquad 3.16$$

with $\overline{\text{corr}(r_{1,i}, r_{1,j})}$ estimated by $(N+2)\overline{\hat{\rho}^2_{i,j}}/(N-1)$.

The average $\overline{\hat{\rho}^2_{i,j}}$ for the 77 station ξ_i series is 0.264, and the estimate of $\overline{\text{corr}(r_{1,i}, r_{1,j})}$ is 0.266. Then Eq. 3.16 gives $M_e=3.63$ for M=77 . Similarly, for M=41 , the values are $\overline{\hat{\rho}^2_{i,j}}=0.348$, the estimate of $\overline{\text{corr}(r_{1,i}, r_{1,j})}$ is 0.351, and $M_e=2.72$.

Figure 3.12 gives the average correlogram, $\bar{r}_k = f(k)$ for the ξ_i series of 77 stations, with the tolerance limits $(r_k)_u$ and $(r_k)_\ell$ at the 95 percent level, computed by

$$(\bar{r}_k)_{u,\ell} = E\bar{r}_k \pm t\left[\frac{\text{var } r_k}{M_e}\right]^{\frac{1}{2}} = -\frac{1}{N-k+1} \pm t\frac{\sqrt{N-k}}{(N-k+1)\sqrt{M_e}} \quad . \qquad 3.17$$

Similarly, Fig. 3.13 gives the average correlogram for the 41 subregion stations. These two average correlograms lead to the following conclusions:

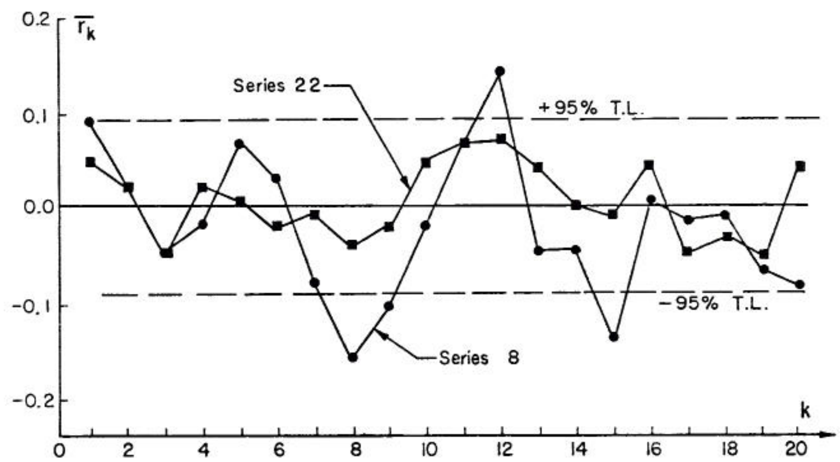

Fig. 3.11. Correlograms of ξ_i stochastic components with 95 percent tolerance limits for Station series nos. 8 and 22.

(1) In both cases, 77 and 41 stations, the $\overline{r}_1$ and $\overline{r}_{12}$ are outside but close to the tolerance limits;

(2) All other $\overline{r}_k$'s are either on or within the tolerance limits;

(3) The 95 percent probability level of tolerance limits implies that only one $\overline{r}_k$ value out of twenty should be outside the limits; in these cases there are two instead of one value outside the limits;

(4) The squares of $\overline{r}_1$ and $\overline{r}_{12}$ are about one percent (0.1^2), which has little effect on the time dependence of the ξ_i series; for all practical purposes, they can be neglected.

Thus the ξ_i series may be considered as time independent processes.

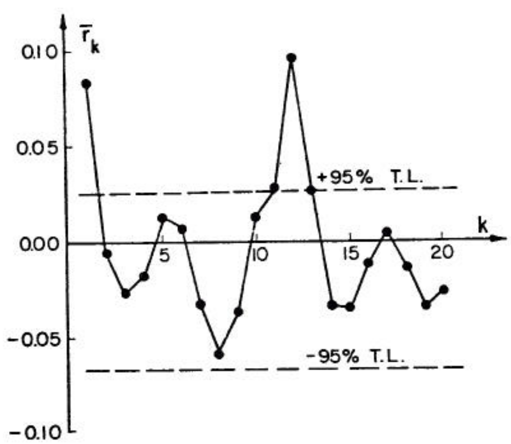

Fig. 3.12. The average correlogram of the 77 series of ξ_i variables (Region I), with the corresponding 95 percent tolerance limits.

The average of the means of the ξ_i independent stochastic components of the 77 precipitation series is 0.033 and the variance of the means is 0.007. The means of approximately standard ξ_i variables can be considered the same. The average of the 77 sample standard deviations of ξ_i is 1.067, with a variance of 0.006.

The selected regional model of Eq. 3.11 accomplishes the following:

(1) It removes the series periodicities and the regional trend in the basic parameters;

(2) It provides an approximately standard (0,1), stationary, sequentially uncorrelated set of stochastic variables; and

(3) It reduces the number of periodicity and trend parameters to only eight for the monthly precipitation series observed in Region I.

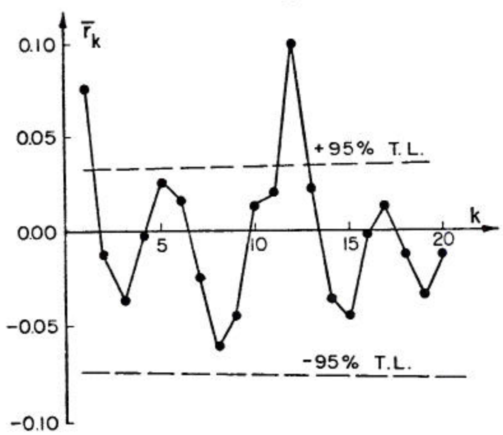

Fig. 3.13. The average correlogram of the 41 series of ξ_i variables (subregion to Region I), with the corresponding 95 percent tolerance limits.

3.5. Analysis of Independent Identically Distributed Stochastic Components. In structural analysis and mathematical description of monthly precipitation series it is useful to determine the form of the probability distribution function for the ξ_i series. The assumptions made and tests performed, in removing the periodicity and regional trends in the major parameters, lead to the conclusion that the computed ξ_i series is approximately an independent, standardized stochastic variable, which has absorbed into its frequency distributions the major time and areal random sampling variations of the parameters. By standardizing a set of non-normal random variables, differences can appear only in the autocovariances and the third-order and fourth-order parameters. For the ξ_i series, the conclusion of time independence (at least by the series being linearly uncorrelated) implies that they are second-order stationary processes. Since their probability distributions can have different asymmetry and kurtosis, the basic problem is to investigate whether the various ξ series have skewness and excess coefficients which are significantly different from being constants.

The differences between the estimated constants for the regional ratios of parameters in Eq. 3.11 and the estimates from trend functions of the parameters evaluated at the observed points have a tendency to be greater as the distance from the geographical centroid of the region increases. This can be inferred from Figs. 3.7 through 3.10, in which the largest differences appear to be located at the edges of Region I. However, when only the 41 stations of the interior subregion are used, the differences at these new boundaries are reduced. This fact reinforces the assumption of identically distributed, second-order stationary, ξ_i series. Regardless of this decreasing effect at the boundaries for the 41 stations, the analysis of identically distributed ξ_i stationary variables is performed here for all 77 stations of Region I.

Using a standard computer program, tests were performed for each ξ_i to determine whether the frequency distribution curve can be well fitted by one of three basic probability distribution functions: the normal distribution, the three-parameter lognormal distribution, and the three-parameter gamma distribution. Because the ξ_i variables have been standardized, $E(\xi_i)=0$ and $E(\text{var}\xi_i)=1$, to fit the normal function, the values of the skewness and excess coefficients are crucial. In fitting a bounded three-parameter function, either lognormal or gamma, the variation in the estimated lower boundaries is decisive in determining the fit of these functions to the actual distribution of these 77 station ξ_i series.

The variables ξ_i of 77 stations, with $i=1,2,\ldots,N$ and $N=\omega n=480$, were found to fit the three-parameter gamma probability function,

$$f(\xi) = \frac{1}{\beta^{\alpha}\,\Gamma(\alpha)} (\xi-\gamma)^{\alpha-1}\exp(-\xi/\beta) \quad , \qquad 3.18$$

with α the shape parameter, β the scale parameter, and γ the lower boundary. The chi-square values for the 77 fits were computed using 20 class intervals of equal probability with 17 degrees of freedom and were tested at the 95 percent significance level. For the 77 stations, 7.7 percent were found to be unacceptable as three-parameter gamma distributed variables, while for the 41 station subregion, only stations 15 and 26, or 5 percent of the 41 stations, have the computed chi-squares greater than the critical value 27.6.

Table 3.4 shows the basic statistics and the computed chi-square values for fitting the three distribution functions to the 41 series of the subregion. In case of the normal function, none of the 41 stations have computed chi-squares less than the critical value. For the three-parameter lognormal function, 21 series have unacceptable chi-squares. All series are positively skewed, and the regional average skewness coefficient for the 41 series is 1.155. The averages of the maximum likelihood estimates of the three-parameter gamma distribution function are as follows: (1) the lower boundary $\hat{\gamma}=-1.694$, with a variance of 0.003; (2) the shape parameter $\hat{\alpha}=2.620$, with a variance of 0.090; and (3) the scale parameter $\hat{\beta}=0.668$, with a variance of 0.004. In addition, the average mean for the 41 variables is 0.042, with a range in the means of 0.379, from -0.127 to 0.252, and the average variance is 1.136, with a range in the variances of 0.704, from 0.914 to 1.616. Although the ξ_i variables are considered to be standardized, they do not have a mean of zero and an exact variance of unity, due to the simplifications made in the time and areal deterministic functions describing the variations in the parameters. However, the ensemble of the ξ_i variables is treated as if it had mean zero and variance unity.

Considering the regional variation in the lower boundary to be a sampling variation, a fit of the normal function to this variation is tested and made, with a mean of -1.694 and a standard deviation of 0.056, regardless that the theoretical distribution is not normal. The normal function has a chi-square of 1.72 using five class intervals for the 41 values of $\hat{\gamma}$, which is not significant in comparison with the critical chi-square of 7.81 at the 95 percent level with three degrees of freedom. Therefore, all ξ_i series may be considered identically distributed random variables, (considering the third-order stationarity), with all parameters of the three-parameter gamma probability distribution function being regional constants. The three sample parameters are each estimated for the three-parameter gamma distribution function as the average of the estimates for the 41 series.

The 41 series of ξ_i variables in the subregion may then be considered as independent, standardized, third-order stationary, identically distributed random variables, with a three-parameter gamma probability density function. Table 3.5 gives the mean, variance, standard deviation and skewness coefficient for the mean, and the variance, standard deviation, skewness coefficient, and minimum observed value of the 41 computed ξ_i series, as well as the lower boundary, shape, and scale parameters of the 41 fitted three-parameter gamma distribution functions. The average of the means, 0.042, with a variance of 0.008, is close to zero, while the average of the variances is 1.135, with a variance of 0.022, and is close to unity. The average of the 41 standard deviations is 1.063, with a variance of 0.002, while the average skewness coefficient is 1.156, with a variance of 0.033. The small variations of the $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ estimates about the average values of 2.620, 0.668, and -1.694, as well as the small variations of the above general parameters, imply that the ξ_i variables computed by Eq. 3.11 are from identically distributed, third-order stationary stochastic process.

3.6. <u>Regional Correlation Structure for Identically Distributed Stochastic Variables</u>. The linear regional dependence between the identically distributed ensemble of stationary stochastic variables is investigated for the lag-zero cross correlation coefficients as related to the interstation distance, d, and its azimuth, ϕ , or its plane normal angle, θ . Ten models are studied as given in Table 2.1. The first model relates the lag-zero cross correlation coefficient of any pair of stations to their distance and plane normal angle, following Caffey [18]. The next nine models are the investigations for a simple dependence between the lag-zero cross correlation coefficient and the distance.

All of the ten models chosen are nonlinear in d, because the graphical plot of the (r,d)-points shows a nonlinear relationship. The unknown coefficients are estimated by using a nonlinear least-squares fitting procedure, Dixon [30]. Models III, IV, and V are applied to all the correlation coefficients of the correlation matrix (actually only the upper triangle of

Table 3.4

Fitting Probability Distribution Functions to Frequency Distribution of ξ_i Series Obtained by Model A

No. in 77 seq.	No. in 41 seq.	NORMAL FUNCTION Mean	NORMAL FUNCTION Standard Deviation	NORMAL FUNCTION χ^2	LOGNORMAL FUNCTION Lower Boundary	LOGNORMAL FUNCTION Mean	LOGNORMAL FUNCTION Standard Deviation	LOGNORMAL FUNCTION χ^{2*}	GAMMA FUNCTION Lower Boundary	GAMMA FUNCTION alpha	GAMMA FUNCTION beta	GAMMA FUNCTION χ^2	Minimum Observed Value	Variance	Skewness Coefficient
1	1	-.073	.997	102.30	-2.067	.573	.488	18.00	-1.612	2.315	.665	20.16	-1.575	.996	1.252
2	2	-.063	1.011	82.20	-2.147	.628	.463	14.58	-1.650	2.490	.637	21.00	-1.611	1.004	1.391
3	3	-.000	1.018	111.91	-1.690	.320	.696	46.58	-1.687	2.585	.652	27.33	-1.650	1.038	.933
4	4	.152	1.092	93.16	-2.095	.697	.480	13.91	-1.560	2.318	.739	15.08	-1.520	1.194	1.087
6	5	.072	1.111	85.75	-2.227	.723	.472	9.91	-1.659	2.339	.740	8.00	-1.623	1.236	1.352
8	6	-.020	.958	76.08	-2.330	.752	.416	14.83	-1.626	2.502	.642	17.75	-1.586	.919	.809
10	7	.029	1.031	83.50	-2.289	.747	.435	9.58	-1.703	2.722	.636	14.41	-1.662	1.065	1.139
11	8	.109	1.120	103.25	-1.726	.413	.665	47.00	-1.722	2.719	.673	22.00	-1.690	1.255	1.296
12	9	.137	1.047	80.41	-1.741	.454	.645	37.91	-1.742	3.001	.626	17.33	-1.704	1.097	.934
15	10	.097	1.021	82.33	-2.430	.849	.396	22.25	-1.787	3.361	.560	21.75	-1.621	1.044	1.027
16	11	.003	.979	58.50	-2.450	.820	.396	26.50	-1.655	2.590	.640	24.50	-1.616	.959	.940
17	12	.008	.994	77.25	-1.765	.403	.621	35.33	-1.763	3.108	.570	23.83	-1.729	.990	1.027
19	13	.193	1.080	61.00	-1.768	.504	.631	30.66	-1.769	3.108	.631	9.91	-1.731	1.169	.926
21	14	.171	1.099	72.25	-1.746	.467	.660	42.41	-1.747	2.884	.665	14.66	-1.708	1.211	1.103
22	15	.103	1.112	109.08	-1.727	.413	.659	29.75	-1.724	2.762	.661	34.66	-1.690	1.239	1.254
23	16	.169	1.163	100.33	-1.709	.427	.682	18.66	-1.704	2.601	.720	10.75	-1.672	1.354	1.143
25	17	.025	1.081	87.33	-2.173	.677	.470	14.58	-1.669	2.482	.682	16.33	-1.631	1.170	1.503
26	18	.055	1.068	97.66	-1.752	.408	.650	31.41	-1.750	2.869	.629	12.00	-1.714	1.144	1.204
27	19	.252	1.197	86.75	-1.673	.436	.724	42.08	-1.669	2.420	.794	19.91	-1.634	1.435	1.076
28	20	.072	.997	72.83	-1.662	.361	.677	44.08	-1.663	2.794	.621	22.16	-1.622	.995	.959
32	21	.031	.955	69.66	-2.395	.813	.384	22.91	-1.751	3.469	.514	21.91	-1.555	.914	1.160
35	22	.161	1.089	67.50	-2.358	.833	.428	18.75	-1.625	2.462	.725	14.33	-1.585	1.187	1.003
36	23	.088	1.075	54.66	-1.797	.453	.653	37.50	-1.797	2.918	.646	11.91	-1.759	1.159	1.006
38	24	.047	1.066	104.33	-1.656	.315	.716	38.91	-1.651	2.428	.699	18.75	-1.617	1.139	1.093
39	25	.067	1.113	97.66	-1.676	.344	.702	31.58	-1.671	2.492	.698	16.33	-1.637	1.240	1.529
45	26	-.083	1.087	146.08	-1.751	.301	.688	25.25	-1.744	2.994	.666	29.66	-1.712	1.184	1.432
46	27	.082	1.104	100.41	-2.112	.664	.499	18.83	-1.601	2.177	.773	19.00	-1.564	1.221	1.060
48	28	-.028	1.064	111.50	-1.749	.334	.695	35.08	-1.744	2.526	.679	19.75	-1.712	1.354	1.041
49	29	.235	1.270	139.41	-1.646	.391	.748	31.00	-1.638	2.201	.851	21.75	-1.607	1.616	1.284
50	30	-.109	.996	78.58	-2.264	.666	.454	8.66	-1.702	2.412	.660	11.00	-1.662	.993	1.126
52	31	-.049	1.047	101.41	-2.224	.668	.470	19.50	-1.713	2.451	.679	15.58	-1.674	1.097	1.234
54	32	-.127	.980	102.41	-2.140	.589	.472	10.50	-1.709	2.594	.610	12.75	-1.670	.962	1.269
55	33	.005	1.063	103.41	-2.050	.596	.502	7.41	-1.628	2.357	.693	17.33	-1.593	1.131	1.307
59	34	-.045	.993	94.41	-1.617	.231	.722	36.83	-1.612	2.391	.655	13.08	-1.578	.987	1.065
62	35	-.079	1.070	123.91	-1.812	.356	.664	24.25	-1.808	2.713	.637	24.41	-1.687	1.060	.986
70	36	-.088	1.044	92.91	-1.701	.245	.751	31.16	-1.698	2.284	.705	11.50	-1.661	1.091	1.179
71	37	-.047	.984	67.91	-1.731	.332	.666	35.91	-1.730	2.794	.602	13.00	-1.693	.970	1.024
72	38	.007	1.099	91.41	-2.216	.686	.478	5.41	-1.677	2.794	.741	10.75	-1.635	1.210	1.353
73	39	.016	1.182	107.75	-1.784	.370	.711	30.58	-1.779	2.423	.741	21.66	-1.745	1.400	1.631
75	40	.010	1.039	83.83	-1.695	.333	.685	28.75	-1.692	2.630	.647	9.66	-1.656	1.082	1.168
76	41	.046	1.026	88.33	-1.739	.399	.648	35.33	-1.738	2.929	.609	22.91	-1.702	1.046	1.058

* ____, χ^2 larger than the χ^2 critical

the matrix due to symmetry) and for the following: (1) the raw data, or the $x_{p,\tau}$ series; (2) the series $\varepsilon_{p,\tau}$ computed by the non-parametric standardization method $(x_{p,\tau}-m_\tau)/s_\tau$ for each station; and (3) the ξ_i series obtained by Eq. 3.11. The other models are applied only to the correlation matrix for the ξ_i series.

TABLE 3.5

Four Statistics of the Various Parameters of the 41 ξ_i Stochastic Variables

Parameter	Average	Variance	Standard deviation	Skewness coefficient
Mean	.042	.008	.092	.306
Variance	1.135	.022	.150	.972
Standard deviation	1.063	.002	.068	.743
Skewness coefficient	1.156	.033	.181	.633
Minimum value observed	-1.651	.003	.056	.175
Shape parameter	2.620	.090	.301	.972
Scale parameter	.668	.004	.063	.414
Lower boundary	-1.694	.003	.056	.247

The nonlinear least-squares fit to the (r,d)-points of the function

$$\rho = f(d;\ A,B,\ldots,F) \quad , \qquad 3.19$$

is made by using the Gauss-Newton iteration on A,B,...,F given the initial values for these coefficients. The functions used contain from 1 to 6 coefficients. Their initial computational values are given in Table 3.6. The mean square error is computed at each iteration until the tolerance of less than 10^{-5} is satisfied for the change in the mean square error between iterations. However, this may not always assure optimum coefficient values for the model; the algorithm used selects the parameter which provides a maximum reduction in the error sums of squares, Dixon [30].

Model I, when applied to $\rho=f(d,\theta)$, has an explained variance of 90.0 percent; this is 30 percentage points above Caffey's [18] method, which gives 60 percent of the explained variance for the annual precipitation series. The expression of Model I when compared with Model II offers very little in the way of the increased explained variance, namely only 2.6 percentage points, with twice as many parameters required. Model I is computationally more cumbersome because of having two independent variables. Model II fits the data well but does not satisfy the condition of $\rho=1$ at $d=0$. The fitted $\rho=f(d)$, for this model, accounts for 87.4 percent of the variation of the ξ_i correlation matrix. The coefficients for Models I and II, as well as for the other models, are given in Table 3.7.

Models III, IV, and V, fitted to the set of (r,d)-points taking two series at a time, are applied to the two samples series, of 41 and of 77 stations. Model III is computationally convenient, since it requires about half the computer time of Model IV and about one-tenth of computer time of Model V to estimate the model coefficients for the ξ_i sample series of 41 stations, with 861 terms, or 41 variances and 820 lag-zero cross correlation coefficients. The preliminary trials suggest that the azimuth ϕ , or the plane angle θ , is not an important parameter in describing the changes in the r parameter over the region. The inferred model, as the estimate of $\rho=f(d)$, is

$$r = 0.613e^{-0.011d} + 0.387 \quad , \qquad 3.20$$

with an explained variance for r of 94.2 percent. This model is selected as the regional, mathematical condensation of information contained in the correlation matrix for the 41 series of the ξ_i variable. Only two coefficients must be estimated, A and B .

The interstation distances may also be expressed as the upper triangle of a symmetric matrix with the distances obtained from the station coordinates. The d values range from 12 to 312 miles for Region I. The constant (1-A) in Model III is introduced because for d=0 the value of r should be equal to one. The use of the coefficients (1-A) in this model is unrealistic for large distances of continental dimensions, because r=0.387 for $d=\infty$. In general, for d approaching infinity, r should approach zero, although there are other opinions, Steinitz et. al. [14]. For the 77 sample stations, Model III is selected as the best of the three models, with the explained variance for r of 87.4 percent for the 3003 values of r and d , with d ranging from 4 to 424 miles. The A coefficient is larger than for the 41 stations (0.722 in contrast to 0.613), showing that A increases as the range of d increases. It can be expected that A goes to unity as d goes to infinity.

TABLE 3.6

Initial Computational Values of Coefficients in the Relation $\rho=f(d)$ of Equation 3.19 with One to Six Terms

Number of Coefficients	A	B	C	D	E	F	n
1	0.001						
2	0.500	-0.010					
2		1.000					- .050
3	1.000	-0.001	-0.080				
4	0.900	-0.005	1.000	1.000			
6	1.000	-0.006	$0.3x10^{-4}$	$-1.0x10^{-8}$	$2.5x10^{-10}$	$-1.0x10^{-10}$	

Table 3.7

Fitting Models $\rho=f(d)$ for the Dependence Among the ξ_i Series

Model	Regression Equation	Series	Computer Time seconds	Number of Iterations*	Mean of r	Variance of r	Standard Deviation of r	Mean Square Error	Sample Size	Number of Stations	Percent of Explained Variance R^2
1	$r=A\exp(Bd + Cd\cos 2\theta + Dd\sin 2\theta)$ A=0.903 C=0.0005 B=-0.004 D=0.0002	ξ_i	28.63	10	.598	.021	.145	.002	861	41	90.0
II	$r=Ae^{Bd}$ A=0.889 B=-0.003	ξ_i	8.22	9	.598	.021	.145	.002	861	41	87.4
III	$r=Ae^{Bd} + 1-A$ A=0.613 B=-0.011	ξ_i	8.68	7	.598	.021	.145	.001	861	41	94.2
IV	$r=Ae^{Bd}+Cd + 1-A$ A=0.279 B=-0.033 C=-0.0014	ξ_i	15.27	11	.598	.021	.145	.001	861	41	91.4
V	Six term polynomial A=1.000 $D=-10^{-8}$ $B=-6\times10^{-3}$ $E=-10^{-10}$ $C=2.15\times10^{-5}$ $F=1.011\times10^{-13}$	ξ_i	85.75	7	.598	.021	.145	.002	861	41	94.0
III	$r=Ae^{Bd} + 1-A$ A=0.722 B=-0.008	ξ_i	33.10	8	.508	.024	.158	.003	3003	77	87.4
IV	$r=Ae^{Bd} + Cd + 1-A$ A=0.366 B=-0.019 C=-0.001	ξ_i	48.96	10	.508	.024	.158	.002	3003	77	88.7
V	Six term polynomial A=1.00 $D=-10^{-8}$ $B=-6\times10^{-3}$ $E=-10^{-10}$ $C=2.05\times10^{-5}$ $F=1.507\times10^{-13}$	ξ_i	148.30	7	.508	.024	.158	.002	3003	77	88.0
IV	$r=Ae^{Bd} + Cd + 1-A$ A=0.246 B=-0.0242 C=-0.0006 A=0.315 B=-0.0281 C=-0.0021 A=0.279 B=-0.0333 C=-0.0014	$\chi_{p,\tau}$ $\varepsilon_{p,\tau}$ ξ_i	13.36 10.30 15.27	12 10 11	.721 .580 .598	.009 .022 .021	.097 .150 .145	.001 .001 .001	861	41	88.4 92.2 91.4
VI	$r=(1+Ad)^{-1}$ A=0.006	ξ_i	9.16	8	.598	.021	.145	.002	861	41	89.8
VII	$r=(1+Ad)^{-n}$ A=0.016 n=0.534	ξ_i	21.64	10	.598	.021	.145	.001	861	41	90.8
VIII	$r=e^{Ad}$ A=-0.006	ξ_i	7.71	7	.598	.021	.145	.004	861	41	78.8
IX	$r=e^{Ad}(1+Bd)^{-1}$ A=0.001 B=0.009	ξ_i	16.60	9	.598	.021	.145	.002	861	41	90.4
X	$r=(1+Ad)^{-\frac{1}{2}}$ A=0.018	ξ_i	23.17	9	.598	.021	.145	.001	861	41	91.6

*The figure is the number of iterations, up to 20, needed to converge to a tolerance of less than 10^{-5} change in the mean square error per iteration.

Figure 3.14 presents the (r,d)-points and the fitted r=f(d) functions of the 41 series for the $x_{p,\tau}$ series by using Model IV in graph (1). The same is done both for the $\varepsilon_{p,\tau}$ series obtained by nonparametric standardization methods and by using Model IV in graph (2), and for the ξ_i series obtained by using Model III in graph (3). The bias on the extreme right due to the constant (1-A) can be seen in all three cases, forcing the fitted function upwards at large distances (say at 200-300 miles). Another important feature of these fitted functions is the range of the dispersion of observed points around the function as measured by R^2 , the explained variance of r . These are shown in the last column of Table 3.7. The two graphs of Fig. 3.15 present a comparison of the fitted r=f(d) functions for the (r,d)-points of the 41 station series. The first three curves represent Model IV, applied to the (r,d)-points, as follows: for the $x_{p,\tau}$ series with an 88.4 percent explained variance of r , and $\bar{r}$=0.721 , line (1); for the ξ_i series obtained by Eq. 3.11 with a 91.4 percent explained variance of r , and $\bar{r}$=0.598 , line (2); and for the nonparametrically obtained $\varepsilon_{p,\tau}$ series, with a 92.2 percent explained variance of r , and $\bar{r}$=0.580 , line (3). If the small differences in percentages of explained variances are disregarded, line (2) is close but above line (3), with both lines (2) and (3) significantly below line (1). This shows that the removal of the periodic and regional trend variations in parameters significantly decreases the vertical position of the r=f(d) curve. Both Eq. 3.11 and the nonparametric methods of removing the periodicity and trends in parameters give very close r=f(d) functions for the ξ_i series.

Lines (4), (5), and (6) of Fig. 3.15 represent the r=f(d) functions for Models V, III, and IV fitted to the (r,d)-points, with Eq. 3.11 used in producing the ξ_i series. They explain 94.0, 94.2, and 91.4 percent of the variance of r , respectively. Considering the explained variance attributed to each fitted model to be about the same, these three curves appear to be similar, except for large values of d .

Models VI, VII, VIII, and IX of Tables 2.1 and 3.7, shown as lines (6), (7), (8), and (9) in the two graphs of Fig. 3.16, meet the requirements of ρ=1 for d=0 , and ρ=0 for d=∞ . Models VI and VII, which are computationally more expensive than Model III, have the explained variances of 89.8 and 90.8 percent, respectively. Model VII has two parameters, requiring more than twice the computational time to converge to constants with a 10^{-5} tolerance in the mean square error. The coefficients of Model VI, given by Alexeyev [15] as $r=(1 + 0.006)^{-1}$, fall within the ranges quoted, namely 0.005<A<0.10, and A=0.006. Model VII is a modification of Model VI to improve the symmetry of fit about the (r,d)-points for the values of 0<d<150 miles and especially for the values of 0<d<75 miles. Model VII, which is

$$r = (1 + 0.016d)^{-0.534} \quad , \qquad 3.21$$

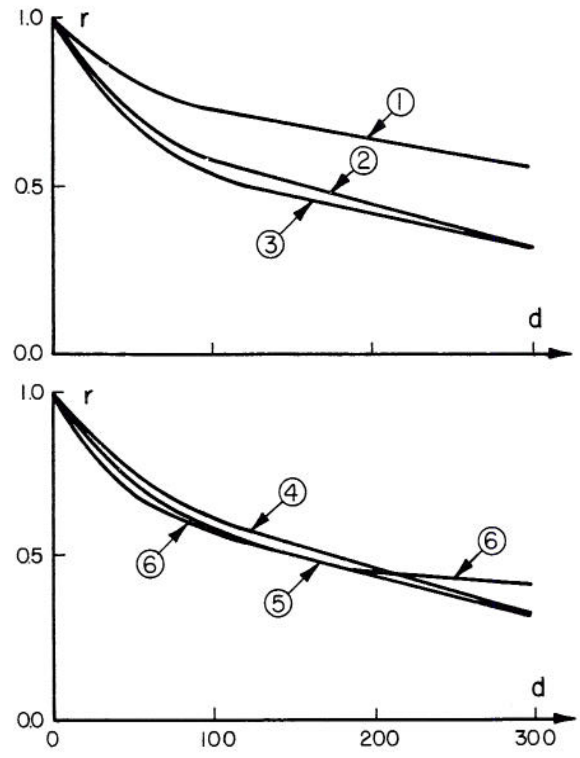

Fig. 3.15. Comparisons of the fitted r=f(d) functions:
(1) Model IV, $x_{p,\tau}$ series (4) Model V, $\xi_{p,\tau}$ series
(2) Model IV, $\varepsilon_{p,\tau}$ series (5) Model III, $\xi_{p,\tau}$ series
(3) Model IV, $\xi_{p,\tau}$ series (6) Model IV, $\xi_{p,\tau}$ series

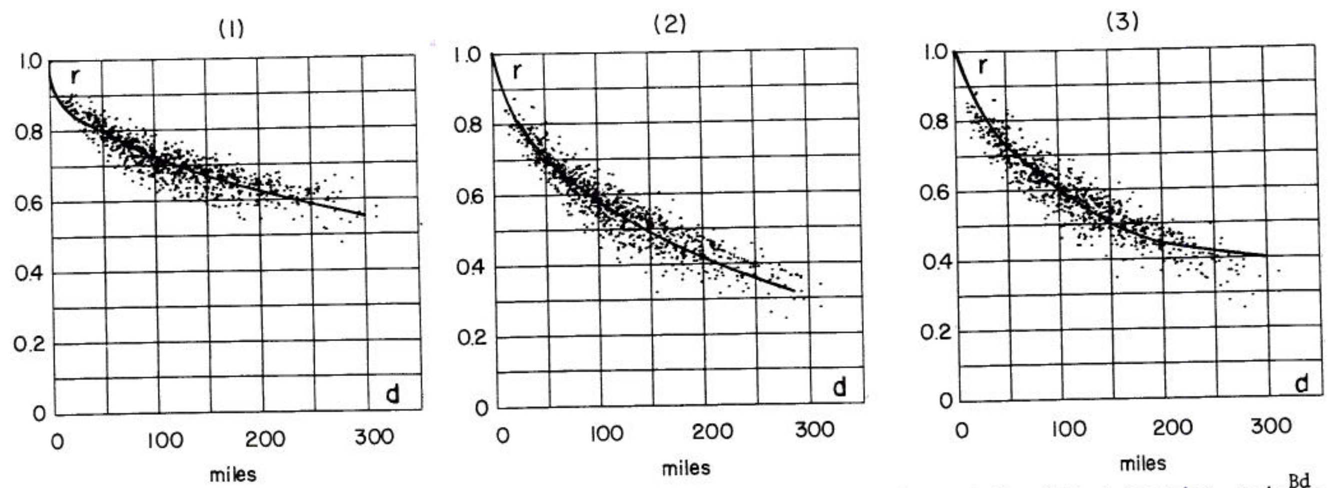

Fig. 3.14. Correlation coefficient, r , versus the interstation distance, d , and the fitted function $r=Ae^{Bd}+Cd+1-A$ (solid line); (1) for the historical $x_{p,\tau}$ series, A=0.315, B=-0.0242, C=-0.00062, (2) for the nonparametrically determined $\varepsilon_{p,\tau}$ series, A=0.315, B=-0.0281, C=-0.00217, and $r=Ae^{Bd}+1-A$, and (3) for the regional ξ_i series obtained by Model III, A=01613, B=-0.011.

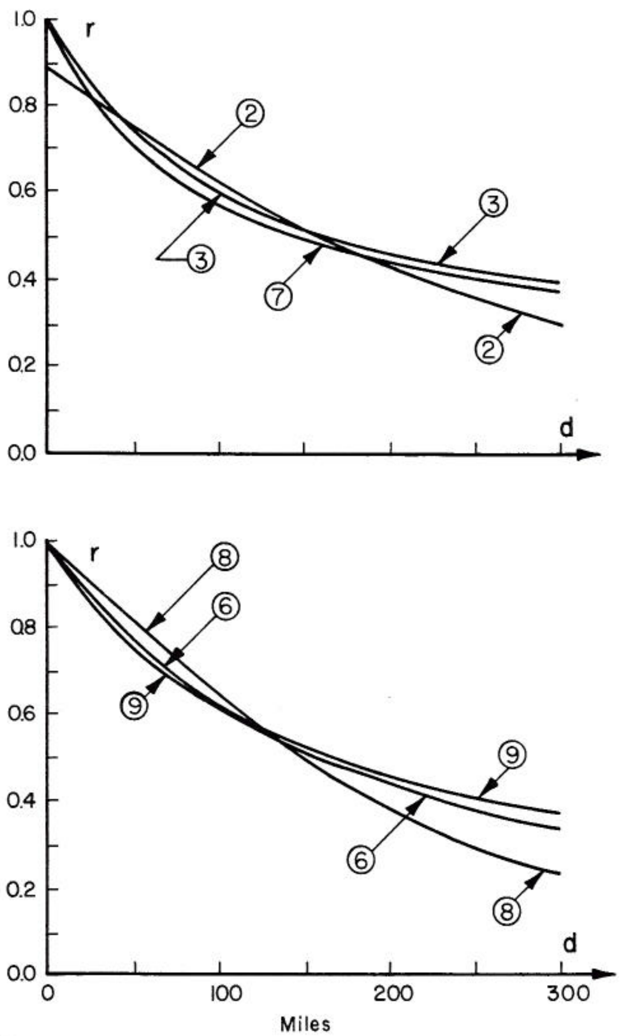

Fig. 3.16. Comparison of the fitted r=f(d) function for the ξ_i series for different models of Table 2.1 and Table 3.7. Line numbers coincide with the model numbers in these two tables.

accounts for 90.8 percent of the variation in r , and the exponent is close to -0.50; perhaps this value of -0.50 would be superior to the use of minus one as used in Model VI. The selection of the exponent between 0.50 and 1.00 is not pursued further in this paper. However, when the exponent is 0.50, defining the model as Model X,

$$r = (1 + 0.018d)^{-0.50} \quad , \qquad 3.22$$

which accounts for 91.6 percent of the variation in r. The fitted Model VII, line (7) of Fig. 3.16, decreases initially faster than line (6), but is close to line (3) and the line of the models using the (1-A) term and the six-term polynomial. Model VIII, which is a simple exponential function, explains 78.8 percent of the variation in r . Combining Model VIII with Model VI, as a gamma-type function, designated as Model IX, explains 90.4 percent of the variation in r . Model IX produces a more rapid drop for small values of d than does Model VI.

There is very little difference between the best fitted models except for the symmetry of the (r,d)-points about the fitted r=f(d) function. Model VII or Model X are best in this respect, but each requires twice as much computer time as Model VI.

For the following reasons it is difficult to study the distribution of r values about the fitted function:

(1) The distribution of r is not readily obtainable for the ξ_i variables of the three-parameter gamma distribution;

(2) The dependence structure $\rho=f(d)$ should be introduced in deriving the sampling distributions f(r) for given d ;

(3) The bound of $\rho=1$ requires all tolerance limits to converge to $\rho=1$ for d=0 .

Perhaps a good approach to placing tolerance limits on the r=f(d) functions is to transform the r values into the Fisher's z variables, with the standard deviation of z defined as

$$\sigma_z = \frac{1}{\sqrt{N-3}} \qquad 3.23$$

in an interval Δd . For each d , the model $\rho=f(d)$ gives the estimated population value. Assuming z to be normally distributed, $N(z_\rho, \sigma_z)$, the tolerance limits, z_u and z_ℓ , at the 95 percent level are

$$z_{u,\ell} = z_\rho \pm 1.96\sigma_z \qquad 3.24$$

This z transformation takes care of the fact that f(d) is bounded by $\rho=+1$, because z for $\rho=1$ becomes infinite.

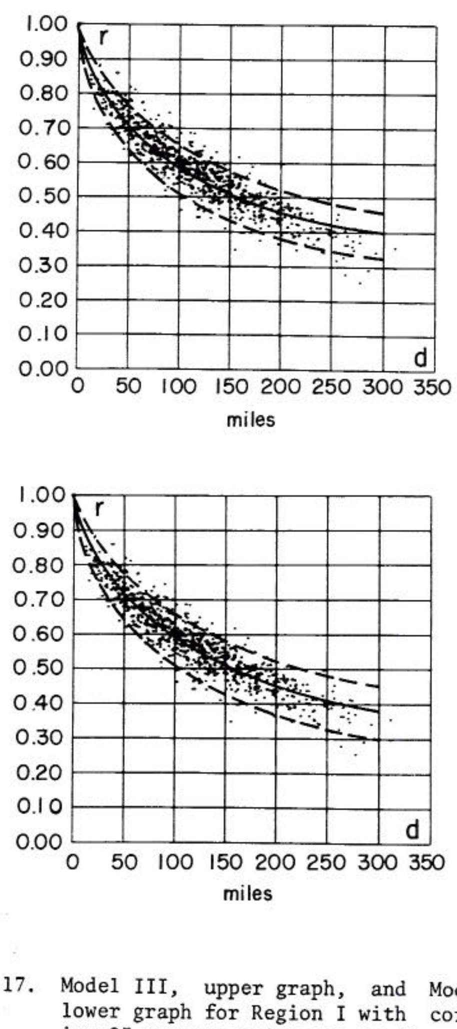

Fig. 3.17. Model III, upper graph, and Model VII , lower graph for Region I with corresponding 95 percent tolerance limits.

Another approach is to consider the standard error of the estimate of the fitted $\rho=f(d)$ function. This standard error of the estimate, s_p , is a measure of the size of residuals (the standard error or differences between the observed and the estimated values of r). Then the residuals may be assumed to be normally distributed with the mean of zero and the variance s_p^2 for each interval Δd about a given d .

Model III and Model VII are selected to illustrate the computation of the tolerance limits for the ξ_i variables. This information is shown in Table 3.8. Table 3.8 also gives the percentage of points outside these limits in comparison with the expected percentage for the assumptions used in deriving either the distribution of r about $\rho=f(d)$ or the distribution of the residuals $(r-\rho)$. Figure 3.17 gives the tolerance limits for the $\rho=f(d)$ functions of ξ_i , namely for Model III (upper graph) and Model VII (lower graph), with the 95 percent tolerance limits computed by Eq. 3.24.

Table 3.8

Tolerance Limits for $\rho=f(d)$ Models for the ξ_i Series for 41 Stations

Interstation Distance Miles	Model	$\rho=f(d)$ Values	Eq. 3.24 Tolerance limits		Percentage of points outside	Standard Error of Estimate	Percentage of points outside	
			r_1	r_2	r_1, r_2	s_p	$\pm 1s_p$	$\pm 2s_p$
0		1.000	--	--				
50		740	.777	.696				
100	III	.590	.645	.528	18.1	0.047	23.3	5.5
200		.454	.522	.380				
300		.409	.484	.332				
0		1.000	--	--				
50		.724	.764	.679	14.2	0.037	29.0	8.5
100	VII	.592	.646	.529				
200		.457	.522	.380				
300		.384	.458	.304				

CHAPTER IV

APPLICATION OF MODELS TO REGION II

Region II is in the North Central Continental United States covering an area of 20,000 square miles. It lies between 96.00 and 99.50 degrees west longitude, and 40.00 and 42.50 degrees north latitude. This region demonstrates that a smaller number of monthly precipitation series from gauging stations in a topographically homogeneous area can effectively be used for regionalizing the basic parameters of series, without using the subregion approach as it was done for Region I. Each of the resulting stochastic ξ_i variables obtained from the regionalization of the basic parameters, by accounting for the annual cyclicity and regional trends in these parameters, are shown to follow the same three-parameter gamma probability distribution function. In addition, the linear correlation structure of the stochastic ξ_i variables can be modeled with a similar $\rho=f(d)$ relationship for Region II, as was previously shown for Region I, with ρ the lag-zero cross correlation coefficient and d the interstation distance between any two stations of monthly precipitation series.

Since Region II is used to demonstrate the procedures developed, it is not necessary to reiterate all the details as presented in Chapter III.

4.1. Data Assembly for Region II. Region II which has relatively simple variation of the basic parameters over the area satisfies the following three criteria: (1) a minimum continuous series of 30 years of monthly precipitation; (2) a change in station location during the 30-year period of less than one mile in horizontal direction and less than 100 feet in elevation; and (3) no more than three years of missing data for any one series estimated by using data from adjacent stations.

Twenty-nine precipitation stations are selected for investigation, each with 30 years of monthly values (N=360 values) for the period 1931-1960. The position within the U.S.A. of Region II is shown in Fig. 4.1. The locations for these 29 stations are given in Fig. 4.2., with the coordinates origin at 99.50 degrees west longitude and 40.00 degrees north latitude. This origin is used for all subsequent graphs in the analysis of the regional distribution of parameters. Table 3 in the Appendix gives the information for Region II similar to that given for Region I. The identity number is prefixed with 25 for Nebraska.

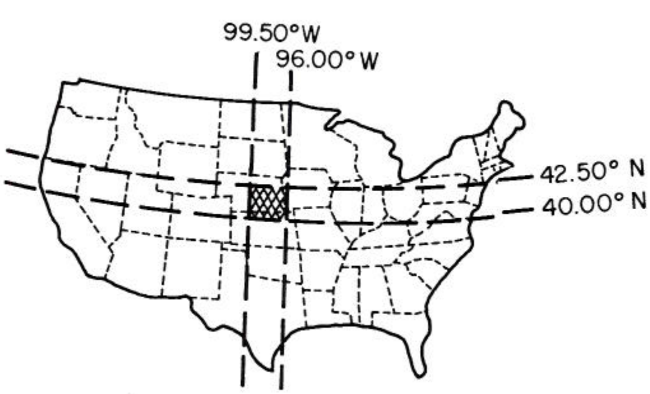

Fig. 4.1. Location within U.S.A. of Region II used as the second example for the regionalization of parameters of monthly precipitation series.

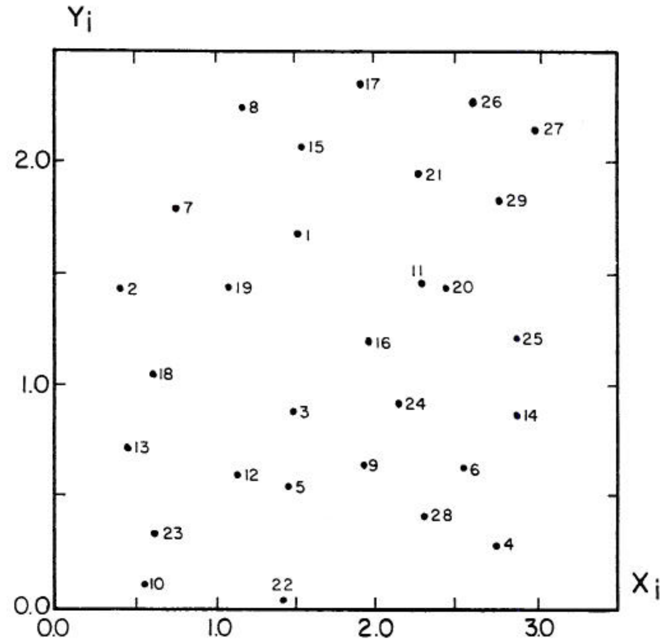

Fig. 4.2. Positions of 29 stations, with numbers for the Nebraska study Region. II.

Climatic features are given in Fig. 4.3. in the form of the variation of the general monthly mean, $\bar{x}_i$, for the individual stations $(i=1,2,\ldots,M)$, with M being the total number of stations. The series means change over the Region II almost linearly, Fig. 4.3 upper graph, from the Northwest corner with a value of 1.8 inches to the Southeast corner having a value of 2.2 inches, with a southwest-northeast orientation of isolines. The series standard deviation, Fig. 4.3 lower graph, varies in a similar manner with values of 1.8 in the Northwest and 2.0 in the Southeast, again with the isolines oriented approximately southwest-northeast.

4.2. Time Variation of Parameters. The stochastic component $\xi_{p,\tau}$ is obtained from the monthly precipitation series $x_{i,p,\tau}$, at any station i, $i=1,2,\ldots,M$, with $p=1,2,\ldots,n$, and $\tau=1,2,\ldots,\omega$, M being the number of stations, n the number of years, and ω the number of intervals within the year. In the case of Region II, $M=29$, $n=30$, and $\omega=12$.

The periodicity in the monthly means are approximated by the Fourier series of Eq. 2.3. For the estimated sample monthly means, m_τ, the fitted Eq. 2.3 is then of the form of Eq. 3.1. Similarly, Eq. 2.3 applies to σ_τ resulting in a fitted periodic function for the estimated monthly standard deviations. Once $h(\mu)$ and $h(\sigma)$ have been inferred, the differences $(m_\tau-\mu_\tau)$ and $(s_\tau-\sigma_\tau)$ are considered random sampling variations.

For the m_τ and s_τ series first studied, $h(\mu)=h(\sigma)=1$ is hypothesized and tested by statistical analysis of all M time series. Table 4.1 presents the percent of the variance of the monthly mean, m_τ, and of the monthly standard deviation, s_τ, explained by the fitted 12-month harmonic of μ_τ and σ_τ for all 29 stations of Region II. From 87.2 to 93.5 percent of the variance, or 90.6 percent, on the average, is

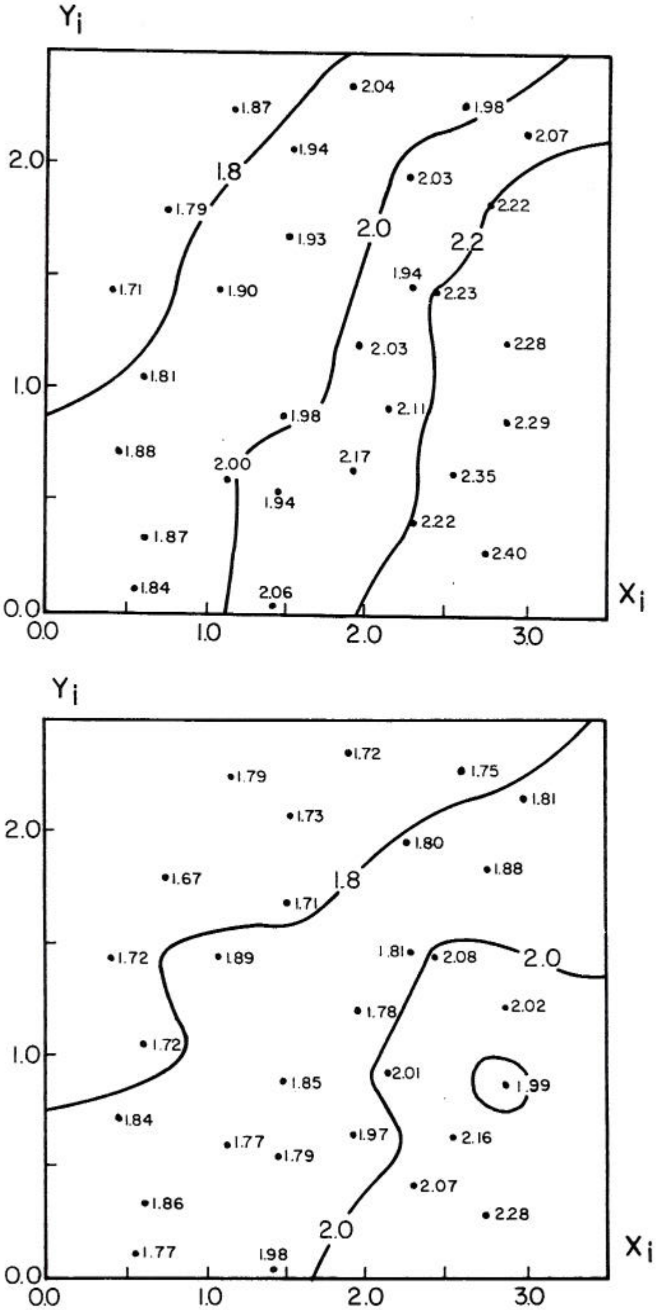

Fig. 4.3. Isolines of the 30-year general monthly mean $\overline{x}_i$, upper graph, and the 30-year monthly standard deviation s_i, lower graph.

explained by the single 12-month harmonic in the case of m_τ. From 76.0 to 97.0 percent of the variance, or on the average 90.1 percent, is explained by this first harmonic in the case of s_τ. The second (6-month) harmonic explains only an additional 3.08 percent of the variance of m_τ, and 2.73 percent for s_τ, while the third (4-month) harmonic has the corresponding figures of 2.44 and 1.70 percent. The average unexplained variances of m_τ and s_τ by the first harmonic are 9.4 and 9.9 percent, respectively. This means that the remaining five harmonics (second through sixth) would allow only 1.8 or 2.0 percent, on the average, respectively. The differences in the order of magnitude of contributions by the first and all the other five harmonics are such that a decision to retain only the first harmonic in μ_τ and σ_τ is justified.

To test for the significance of harmonics in m_τ and s_τ, the same criteria are used for Region II as for Region I; that is at least 90 percent of the variance of m_τ for 80 percent of the station series is explained by μ_τ, and that at least 85 percent of the variance of s_τ for 80 percent of the station series is explained by σ_τ. If this criteria is met, the m_τ and s_τ parameters for the 29 stations in Region II may be adequately approximated using only the 12-month harmonic in the μ_τ and σ_τ periodic functions. For the first harmonic in μ_τ, only 17 of the 29 series (55 percent) exceed the 90 percent level; however, if the level is lowered to 88 percent, 28 of the 29 station series (96 percent) are included. This is considered sufficient for accepting the hypothesis. For the monthly standard deviation, s_τ, 24 of the 29 station series (82 percent) exceed the 85 percent cut-off level, and the hypothesis is accepted.

Figure 4.4. gives the correlogram and the spectral graph for the first of the 29 stations given in Table 3 of Appendix. The correlogram closely resembles the 12-month cosine function, while the spectrum has a significant spike only at the 12-month frequency. Similar graphs can be shown for nearly all of the 29 series. These graphs also support the conclusion that the first harmonic in the m_τ series, computed from the $x_{p,\tau}$ series, is the most important, and that the higher-order harmonics (the second through the sixth) can be neglected.

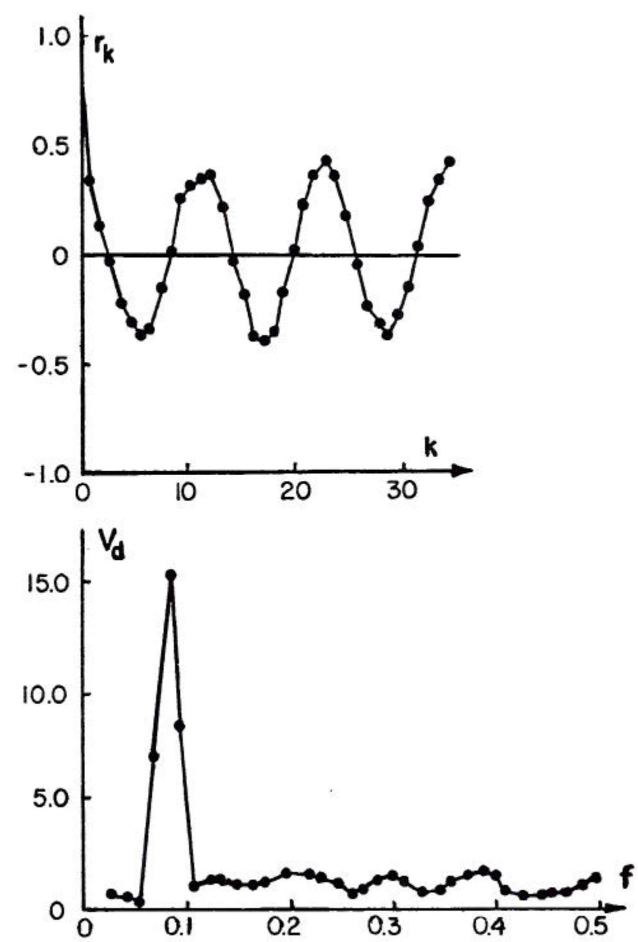

Fig. 4.4. Correlogram (upper graph) and spectrum (lower graph) of the first precipitation series (station 1) of Table 3 in the Appendix.

Table 4.1

Percent Variance of Monthly Means, m_τ, and Monthly Standard Deviations, s_τ, Explained by the First Harmonic (12-Month Harmonic) of μ_τ and σ_τ. (Underline denotes extreme values)

Station Number	For Monthly Mean m_τ	For Monthly Standard Deviation s_τ
1	90.4	89.1
2	91.9	93.8
3	90.7	97.0
4	92.4	89.0
5	90.9	93.0
6	88.7	83.3
7	88.9	87.4
8	92.5	96.3
9	91.5	92.8
10	93.1	92.2
11	88.6	96.7
12	91.0	92.9
13	88.4	76.0
14	91.9	95.5
15	89.0	95.4
16	91.6	89.5
17	89.7	80.6
18	89.6	90.9
19	89.3	87.3
20	89.8	81.7
21	93.1	95.0
22	88.9	90.3
23	87.2	78.3
24	91.8	89.9
25	89.8	92.5
26	90.9	93.2
27	93.2	93.9
28	91.8	87.3
29	93.5	92.9
Average	90.6	90.1

The values of $h(\mu)=h(\sigma)=1$ satisfy the objective of obtaining a minimum number of parameters with a necessary minimum number of significant harmonics, and in all cases the 12-month harmonic is required to account for the annual cycle in the monthly means and monthly standard deviations. Therefore, only the first harmonic is used for both the μ_τ and σ_τ functions in Region II.

4.3. Regional Variation in Parameters. Because only the 12-month harmonic is selected to account for the annual cycle in the basic parameters of $x_{p,\tau}$ for all stations, attention is focused on how the basic parameters, represented by x, $C_1(\mu)$, $\theta_1(\mu)$, $\bar{s}_\tau$, $C_1(\sigma)$, and $\theta_1(\sigma)$, vary over the Region II.

Of major interest is the general sample monthly mean, $\bar{x}$ or $\bar{m}_\tau$. Figure 4.3. shows its overall variation over Region II using the 30-year observation period. Because the marked southeast to northwest decrease is normal to the isolines, which are also close to being parallel, first the simple linear plane trend for $\bar{x}$ over Region II is investigated following Eq. 3.3. The coordinate X_i is referenced to 99.50 degrees west longitude as the zero abscissa, and the coordinate Y_i is referenced to 40.00 degrees north latitude as the zero ordinate. Table 4.2 presents the general statistics for the sample series of 29 stations. The parameter $\bar{x}_i$ has the general average 2.034, variance 0.031, skewness coefficient 0.407, and excess coefficient -0.588. Equation 3.3, by using the simple least-square approach of estimating β_1, β_2, and β_3, becomes

$$\mu_i = 1.8055 + 0.2014X_i - 0.1015Y_i \quad . \qquad 4.1$$

With the statistically significant regression coefficients shown in Eq. 4.1, this trend plane gives an explained variance which is 88.4 percent of the regional variation of $\bar{x}_i$. This high percentage of explained variance of $\bar{x}_i$ by such a simple model, Eq. 4.1, is as surprising as being able to explain the variation of m_τ and σ_τ by the first harmonic. The lower explained variance by Eq. 4.1 in comparison with Eq. 3.4 can be explained by the reduced sample size, 30 years for Region II as compared with 40 years for Region I.

Figure 4.5 presents the differences between the estimated sample means $\bar{x}_i$ and the fitted model means μ_i given by Eq. 4.1. The isolines of differences are drawn with a stress on the zero isoline. The differences show that the sample means, $\bar{x}_i$, are larger than the model means μ_i, in the middle, northwestern and eastern sectors, with a maximum value of 0.10 inches. The negative differences dominate the central sectors, with a minimum value of -0.12 inches in the northeast.

The fit by Eq. 4.1 is made by using the least square method and giving each $\bar{x}_i$ value the same weight. Figure 4.3, lower graph, shows that the standard deviation s_i increases as $\bar{x}_i$ increases. A more proper approach in estimating β_1, β_2, and β_3 is to give weights $1/s_i$ to the deviations $(\bar{x}_i-\mu_i)$, thus using the relative deviations $(\bar{x}_i-\mu_i)/s_i$ which have a more homogeneous variance than using only the $(\bar{x}_i-\mu_i)$-deviations in the least square fitting method. The weighted deviations can be an improvement in the estimation of models and coefficients of the type contained in Eq. 3.3. for example, using this weighted approach for the case of $\bar{x}_i$, the new Eq. 4.1 has

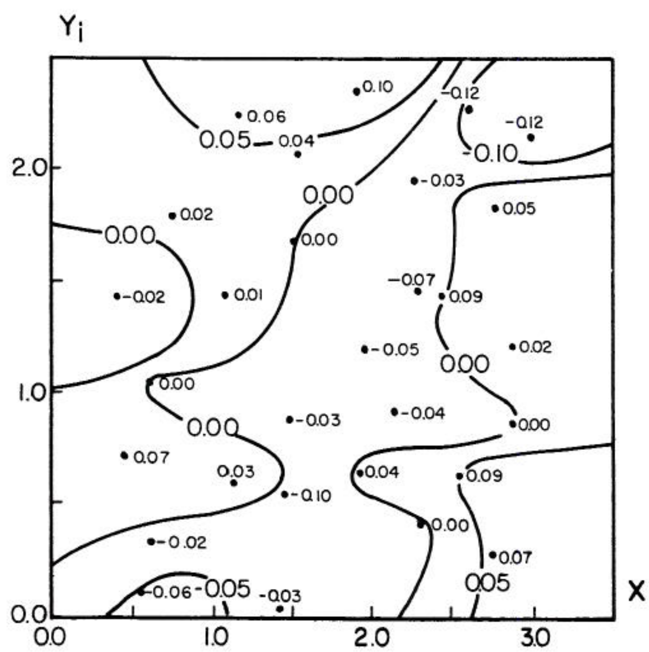

Fig. 4.5. Differences and isolines of differences of the estimated means and the fitted mean by simple least-square approach.

Table 4.2

Average Statistics for Regional Parameters for Set of 29 Station Series of Region II

Regional Parameter	Avg.	Variance	Standard Deviation	Coefficient of Variation	Coefficient of Skewness	Coefficient of Excess
$\bar{x}_i$	2.034	.031	.178	.087	.407	- .588
$C_{1,i}(\mu)$	1.618	.011	.107	.066	.742	- .028
$\theta_{1,i}(\mu)$	2.963	.005	.074	.025	-.593	- .126
$\bar{s}_{\tau,i}$	1.326	.016	.127	.096	.781	- .080
$C_{1,i}(\sigma)$	.824	.012	.111	.134	.875	1.613
$\theta_{1,i}(\sigma)$	2.784	.005	.075	.027	-.631	1.684
$C_{1,i}(\mu)/\bar{x}_i$	.798	.002	.054	.068	.273	- .625
$\bar{s}_{\tau,i}/\bar{x}_i$	.652	.0008	.029	.045	.118	-1.043
$C_{1,i}(\sigma)/\bar{x}_i$	.406	.002	.050	.125	.140	- .730

$\hat{\beta}_1$=1.8060, $\hat{\beta}_2$=0.1959, and $\hat{\beta}_3$=-0.0965, with an explained variance for $\bar{x}_i$ of 96.4 percent instead of the 88.4 percent shown above for the equal-weight least-square approach. Figure 4.6. shows the $(\bar{x}_i-\mu_i)$ -deviations and their isolines using the weighted least-square approach.

A clustering of the positive and negative differences $(\bar{x}_i-\mu_i)$ into cells, as shown in Figs. 4.5. and 4.6, is the result, as postulated earlier, of the areal dependence amongst the station series. These cells, in contrast to the individual differences, may be inferred as being randomly distributed over the region, separated by the zero isoline. No particularly strong trend seems present in the areal distribution of the cells, at least as inferred by a visual inspection. Therefore, it is attractive to conclude that the remaining 11.2 or 3.6 percent of the variation in $\bar{x}_i$, unaccounted for by Eq. 4.1, is attributable to sampling variation, or that the values of the mean $\bar{x}_i$ change in an approximately linear manner over Region II.

The five other regional parameters, $C_{1,i}(\mu)$, $\theta_{1,i}(\mu)$, $\bar{s}_{\tau,i}$, $C_{1,i}(\sigma)$, and $\theta_{1,i}(\sigma)$, are also hypothesized to have a linear regional variation. Their general statistics are given in Table 4.2. The isolines of the amplitude, $C_{1,i}(\mu)$, of the 12-month harmonic in the mean are shown in Fig. 4.7., upper graph. These isolines suggest a northwest to southeast trend with the lowest values in the northwest (1.5 isoline) and the highest values in the southeast (1.7 isoline). At first glance, the variation does not seem to be linear. The amplitude $C_{1,i}(\mu)$ has a regional mean 1.618, variance 0.011 and skewness coefficient 0.742.

Following the approach used for Region I, the hypothesis is advanced, that the ratio $C_{1,i}(\mu)/\bar{x}_i$ is a regional constant. The mean for the 29 individual values of this ratio is 0.798, the variance is 0.002, and the skewness coefficient is 0.273. Because of the very small value of the variance of $C_{1,i}(\mu)/\bar{x}_i$, this ratio may be considered as not being significantly different from its mean value. Figure 4.7, lower graph, shows how this ratio changes over Region II. The 0.75, 0.80, and 0.85 isolines are north-south oriented, with the regional variation highly reduced. The variance of $C_{1,i}(\mu)/\bar{x}_i$, 0.002, is only 18 percent of the variance of $C_{1,i}(\mu)$, 0.011, the same percent as was found for Region I.

The upper graph of Fig. 4.8. gives the isolines of $\bar{s}_{\tau,i}$, the mean of the 12 values of standard

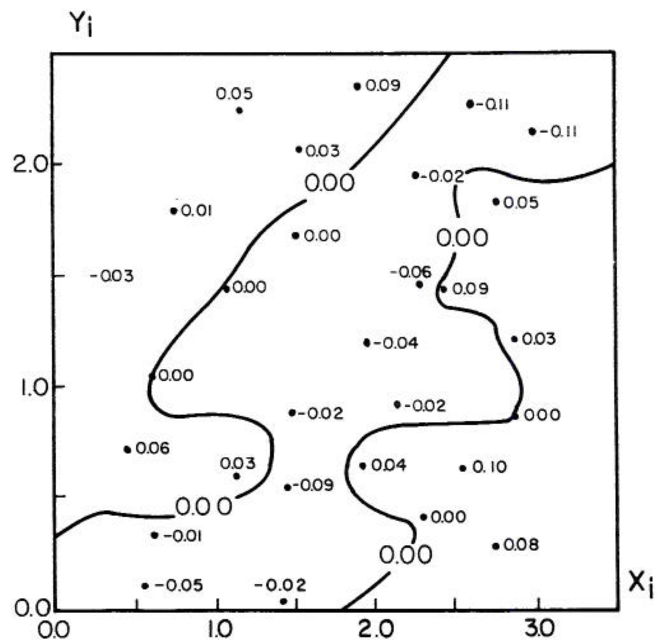

Fig. 4.6. The same points and isolines as in Fig. 4.5. except that the weighted least-square estimation approach is used for standardized deviations.

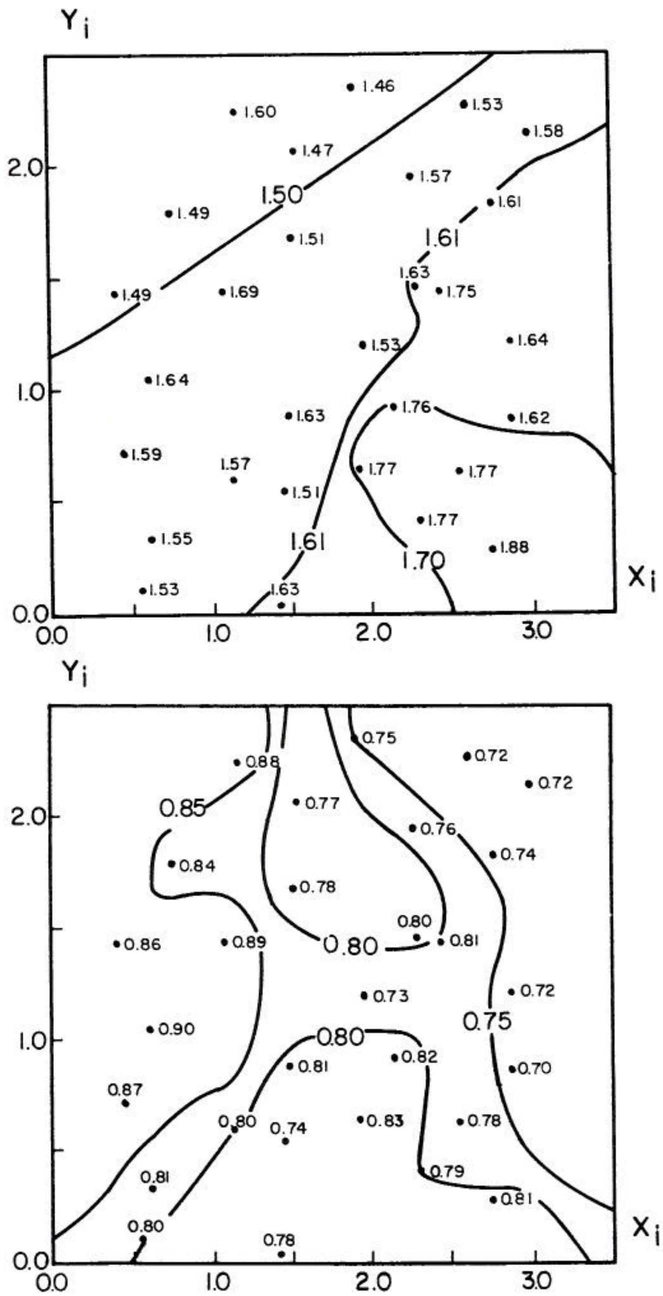

Fig. 4.7. The isolines of the amplitude $C_{1,i}(\mu)$ for the mean μ_τ, upper graph; and the isolines for the ratio $C_{1,i}(\mu)/\bar{x}_i$, lower graph.

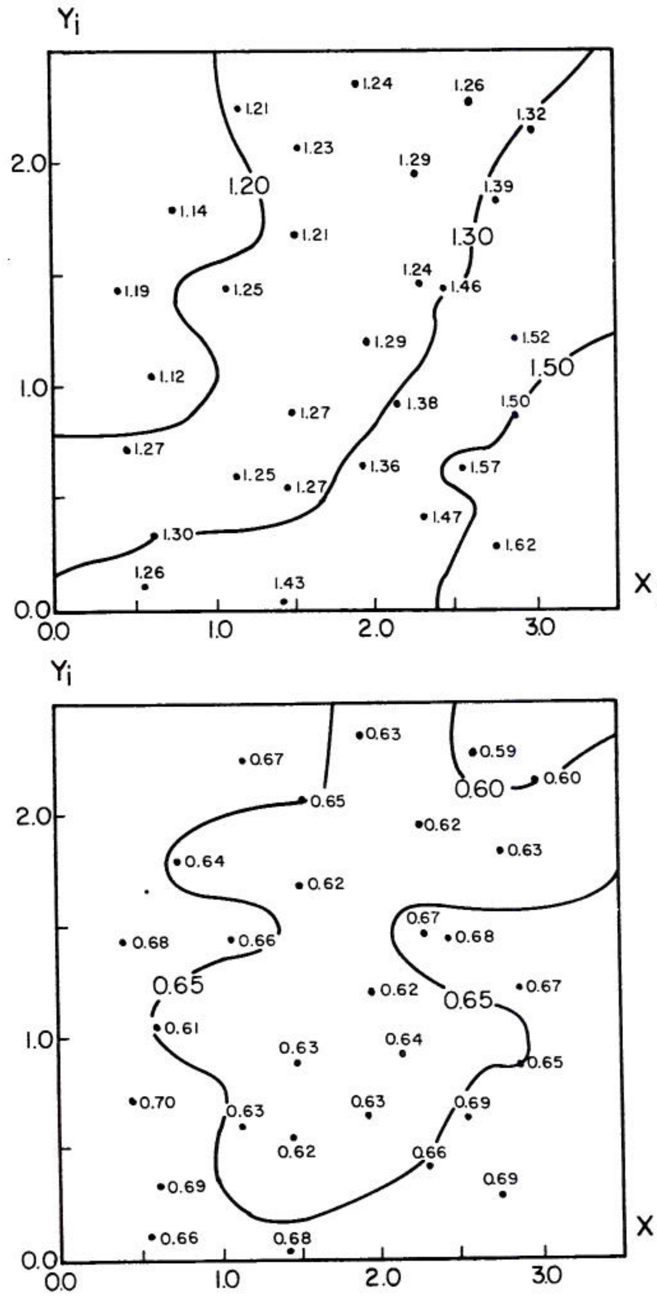

Fig. 4.8. The isolines of the parameter $\bar{s}_{\tau,i}$, the mean of 12 values of s_τ standard deviations, upper graph; and the isolines of the ratio $\bar{s}_{\tau,i}/\bar{x}_i$, lower graph.

deviations s_τ for each of the 29 individual $x_{i,p,\tau}$ series, while the lower graph gives the isolines of $\bar{s}_{\tau,i}/\bar{x}_i$. There is a difference between the s_i estimates shown in Fig. 4.3., lower graph, and the $\bar{s}_{\tau,i}$ individual means of s_τ shown in Fig. 4.8., upper graph; with the $\bar{s}_{\tau,i}$ values smaller than the s_i values. Because of periodicities in the standard deviations and the use of Eq. 3.2, the regionalization is done for $\bar{s}_{\tau,i}$ but not for s_i. There is a general increase in $\bar{s}_{\tau,i}$ values from the northwest to the southeast as for the previously studied parameters. The mean for the 29 values of $\bar{s}_{\tau,i}$ is 1.326, variance 0.016, and coefficient of skewness 0.781. The uniform change from the 1.5 isoline to the 1.2 isoline in Fig. 4.8. suggests that a trend plane may also well represent the regional variation of $\bar{s}_{\tau,i}$.

Considering the ratio $\bar{s}_{\tau,i}/\bar{x}_i$ as a new regional parameter, it has a mean equal to 0.652, with a variance of 0.0008, and a skewness coefficient of 0.118. For the set of 29 station series, this reduced variance of 0.0008 when compared with the variance 0.016 of $\bar{s}_{\tau,i}$ is only about 5 percent.

The ratio $\bar{s}_{\tau,i}/\bar{x}_i$ is selected as a regional constant equal to 0.652. The isolines and point values for this ratio are also shown in Fig. 4.8., lower graph. Its small variation, with the 0.65 isoline winding throughout the area supports the hypothesis that $\bar{s}_{\tau,i}$ and $\bar{x}_i$ closely approximate the proportional parameters.

The isolines for the amplitude of the 12-month harmonic of the monthly standard deviation are shown in the upper graph of Fig. 4.9. The mean of the 29 individual values of $C_{1,i}(\mu)$ is 0.824, with variance 0.012 and skewness coefficient 0.875. The ratio $C_{1,i}(\sigma)/\bar{x}_i$ has a mean of 0.406, variance 0.002 and skewness coefficient 0.140. The isolines of the ratio are given in the lower graph of Fig. 4.9., and show that the hypothesis of the ratio being a constant is well founded.

The isolines and point values for the phase angles, $\theta_{1,i}(\mu)$ for the periodic μ_τ, and $\theta_{1,i}(\sigma)$ for the periodic σ_τ, are shown in Fig. 4.10., the upper and lower graphs of Fig. 4.10., respectively. For $\theta_{1,i}(\mu)$ the isoline values range from 2.80 to 3.05 and for $\theta_{1,i}(\sigma)$ from 2.7 to 2.8, with increases from the southeast to northwest corners of Region II.

The variation in $\theta_{1,i}(\mu)$ is considered to result only from sampling variation. The parameter $\theta_{1,i}(\mu)$, with mean 2.963 and variance 0.005, is selected as a regional constant equal to 2.963. The regional variation in $\theta_{1,i}(\sigma)$ may also be considered the sampling variation, and the parameter may be considered a regional constant equal to 2.784. The southeast-northwest pattern in both $\theta_{1,i}(\mu)$ and $\theta_{1,i}(\sigma)$ is quite likely the result of regional dependence in sampling variations rather than a population characteristic. As stated for Region I, there is no physical reason why the phases for the periodic mean and the periodic standard deviation should be different. A

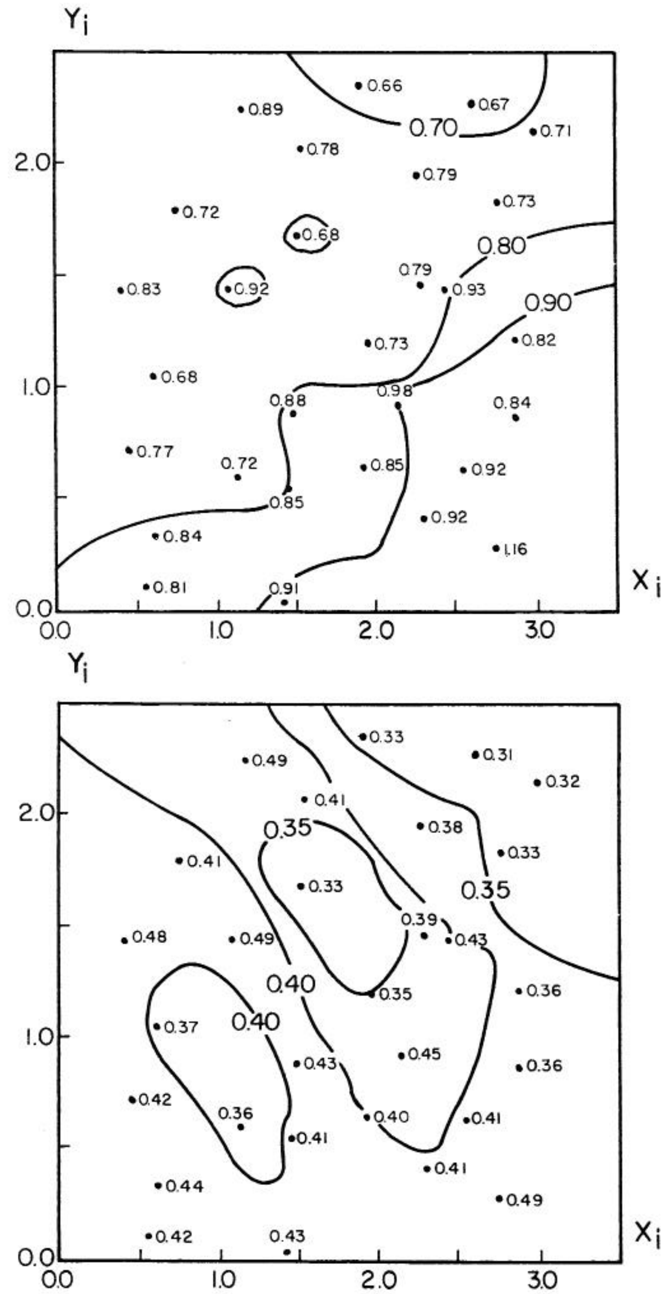

Fig. 4.9. The isolines of the amplitude $C_{1,i}(\sigma)$ of 12-month harmonic of the monthly standard deviation, upper graph; and the isolines of the ratio $C_{1,i}(\sigma)/\bar{x}_i$, lower graph.

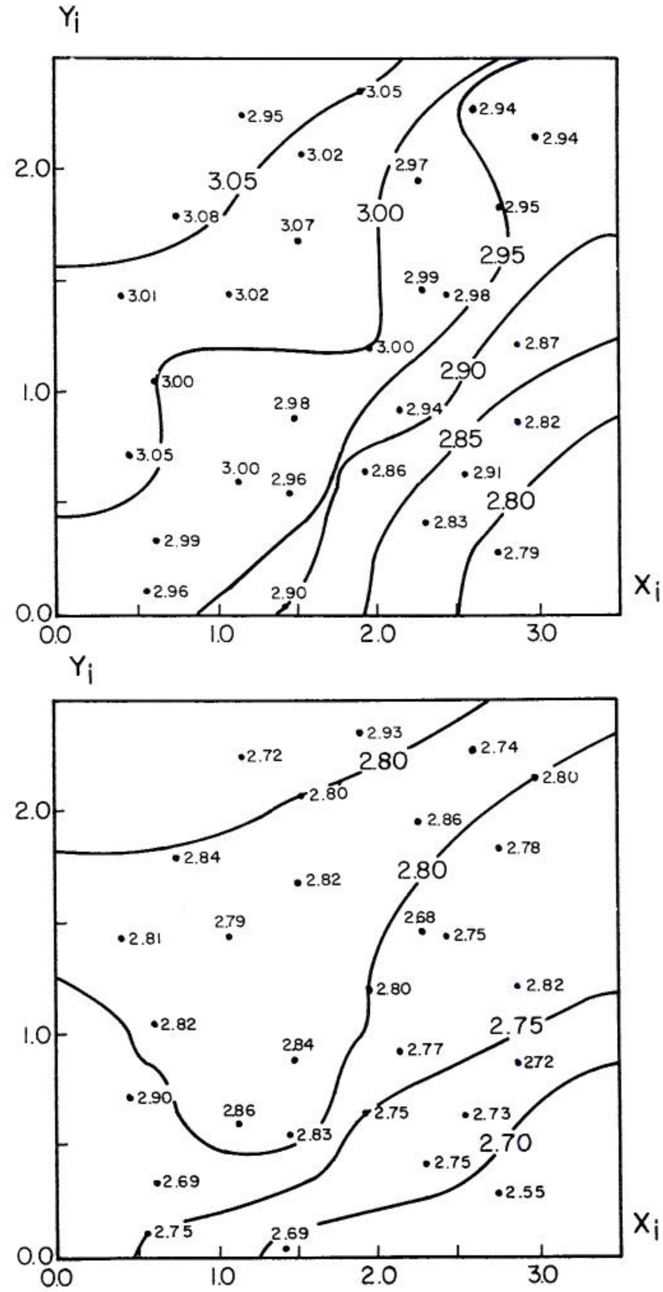

Fig. 4.10. The isolines and point values of the phase angles: the $\theta_{1,i}(\mu)$ of the 12-month harmonic in periodic m_τ , upper graph; and the $\theta_{1,i}(\sigma)$ of the 12-month harmonic in periodic s_τ , lower graph.

further simplification may be obtained by putting $\theta_{1,i}(\mu)=\theta_{1,i}(\sigma)$. In that case either $\theta_1=2.963$ or the average $\bar{\theta}_1=2.873$ is used. Regardless of the advantages of this unique $\bar{\theta}_1$ value, both constants, 2.963 and 2.784, will be used in the further analysis of monthly precipitation.

The regional range in the isolines of $\bar{x}_i$ is approximately 0.40 inches, while for $C_{1,i}(\mu)$ it is about half that, or 0.20 inches. The coefficient of variation of $\bar{x}_i$ is 0.087, which is 3.5 times that for $C_{1,i}(\mu)$, while the mean of $\bar{x}_i$ is 1.27 times that for $C_{1,i}(\mu)$. The regional variability is then much greater in $\bar{x}_i$ than in $C_{1,i}(\mu)$. The fact that the major regional variation exists in the mean $\bar{x}_i$ rather than in the periodic fluctuation about the mean, measured by $C_{1,i}(\mu)$, suggests that the trend plane which accounts for the variation in $\bar{x}_i$ is the most important aspect of the regionalization of precipitation parameters.

Applying a similar analysis to the monthly standard deviation $\bar{s}_{\tau,i}$, the range of regional variation in $\bar{s}_{\tau,i}$ is about 0.30 inches, while in $C_{1,i}(\sigma)$ it is 0.20 inches. The regional mean of $\bar{s}_{\tau,i}$ is 1.32; this is 1.61 times the mean of $C_{1,i}(\sigma)$, which is 0.824. However, the regional coefficient of variation for $\bar{s}_{\tau,i}$ is 0.096 which is only 71 percent of that for $C_{1,i}(\sigma)$.

Again the major regional variation is in the mean $\bar{s}_{\tau,i}$ rather than in its periodic fluctuation. This supports the conclusion that the parameter of periodic variation is more difficult to regionalize. Although it has a greater consistency over the region, it also a greater regional random sampling variation.

With the above regional models for various parameters, the second-order stationary $\xi_{p,\tau}$ component for any station series is

$$\xi_{p,\tau}=\frac{\dfrac{x_{p,\tau}}{1.8055+.2014X_i-0.1015Y_i}-1-0.798\cos\left(\dfrac{\pi\tau}{6}+2.963\right)}{0.652+0.406\cos\left(\dfrac{\pi\tau}{6}+2.784\right)} \qquad 4.2$$

This was obtained by following Eq. 3.11 with μ_τ from Eq. 4.1 replacing $\bar{x}_i$, and five other parameters $C_1(\mu)/\bar{x}$, $\theta_1(\mu)$, $s_\tau/\bar{x}$, $C_1(\sigma)/\bar{x}$, and $\theta_1(\sigma)$ of Region II being replaced by their constant values, 0.798, 2.963, 0.652, 0.406, and 2.784, respectively.

4.4. Testing Sequential Independence of Stationary Stochastic Components. To determine how closely $\xi_i=\xi_{p,\tau}$ approximates an independent stationary stochastic process as a result of the simplifying assumptions made for the deterministic periodic time and regional trend variations in the basic parameters, the correlogram of each ξ_i series from the 29 monthly precipitation series of Region II is tested to determine whether it significantly departs, on the 95 percent probability level, from the correlogram of an independent series. Only the first 20 lags are used in the correlograms, with the tolerance limits set at the 95 percent level as computed by Eq. 3.13.

Table 4.3 presents the number of r_k values for the lags 1 through 20 which are outside the tolerance limits. The serial correlation coefficients r_k for the lags 1, 8, and 13 in Table 4.3 are often outside the tolerance limits.

Table 4.3

Number of Serial Correlation Coefficients of ξ_i Outside the 95 Percent Tolerance Limits

Lag	Number of Deviations
1	9
2	4
3	0
4	1
5	1
6	0
7	0
8	6
9	2
10	0
11	4
12	0
13	5
14	3
15	4
16	1
17	3
18	1
19	1
20	2
Total r_k computed	580
Deviations	47
Percent of Deviations	8.10
Total Deviations less those for r_1 and r_{13}	33
Percent	5.69

The first serial correlation, r_1, is affected in Region II, as it was in Region I, by some physically justified dependence in the monthly precipitation, as well as by the non-removal of harmonics other than the first one in the periodic parameters. By neglecting the numbers of r_1 and r_{13} values outside the tolerance limits, the new percentages of r_k values outside these limits are also given in Table. 4.3. In that case, and for the 29 station series, only 5.69 percent of r_k values outside the tolerance limits, instead of 8.1 percent in the case where r_1 and r_{13} are included.

Using Eq. 3.16, the effective number of stations for Region II is 2.24 for $\overline{\hat{\rho}^2_{i,j}}=0.421$. Applying Eq. 3.17 for the average correlogram of Fig. 4.11 supports the premise that the ξ_i series may be considered a time independent process.

4.5. Analysis of Independent Identically Distributed Stochastic Components. In the structural analysis and mathematical description of monthly precipitation series it is useful to determine the form of probability distribution function of the ξ_i series. Tests were performed for each ξ_i to determine whether the frequency distribution curve can be well fitted by the normal function, the three-parameter lognormal function, and the three-parameter gamma function. Because $E(\xi_i)=0$ and $E(\text{var}\,\xi_i)=1$, the values of the skewness and excess coefficients are crucial. In fitting the bounded three-parameter function, either lognormal or gamma, the variation in estimated lower

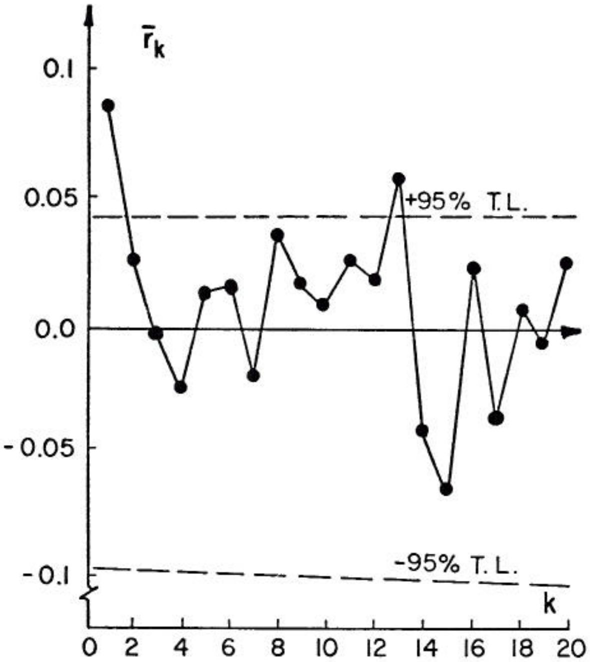

Fig. 4.11. The average correlogram of ξ_i variables (Region II) with corresponding 95 percent tolerance limits.

boundaries is decisive for the identical distributions of these 29 station ξ_i series.

The variables ξ_i of the 29 stations, with $i=1,2,\ldots,N$ and $N=\omega n=360$, were found to follow the three-parameter gamma probability function of Eq. 3.18 well. The chi-square values of the 29 fits were computed using 20 class intervals of equal probability resulting in 17 degrees of freedom, and were tested at the 95 percent significance level. For the 29 stations, 9.67 percent were not acceptable as three-parameter gamma distributed variables. These three stations, nos. 5, 24, and 27 of Table 3 of the Appendix, had computed chi-squares of 28.00, 29.77 and 28.00, respectively, these are significantly higher than the critical chi-square of 27.6 at the 95 percent level; however, they are very close to the critical value.

Table 4.4 shows the basic statistics and the computed chi-square values in fitting the three distribution functions to the 29 series of Region II. In the case of the normal function, none of the 29 stations had computed chi-squares less than the critical. For the three-parameter lognormal function, 17 series have unacceptable chi-squares. All series are positively skewed, and the regional average skewness coefficient for the 29 series is 1.173. The mean parameters of the maximum likelihood estimates of the three-parameter gamma distribution function are as follows: (1) the lower boundary $\hat{\gamma}=-1.683$, with a variance of 0.0004; (2) the shape parameter $\hat{\alpha}=2.616$, with a variance of 0.059; and (3) the scale parameter $\hat{\beta}=0.665$, with a variance of 0.001. In addition, the average mean for the 41 variables is 0.022, with a range in the means equal to 0.191, from -0.048 to 0.143. The average variance is 1.091 with a range in the variances equal to 1.306, from .336 to 1.306. Although the ξ_i variables do not have mean zero and variance unity exactly, the ensemble of the ξ_i variables is treated as if it had mean zero and variance unity.

Considering the regional variation in the lower boundary as a sampling variation, a fit of the normal function to this variation, with mean -1.683 and standard deviation 0.067, has a chi-square of 2.68 using five class intervals for the 29 values of $\hat{\gamma}$; this is not significant in comparison with the critical chi-square value of 7.81 at the 95 percent level, with three degrees of freedom. The ξ_i series of Region II may be also considered as identically distributed random variables, as was done for Region I. The 29 series of ξ_i variables in Region II may then be inferred as independent, standardized, third-order stationary, identically distributed random variables, following the three-parameter gamma probability density function. Table 4.5 gives the mean, variance, standard deviation, and skewness coefficient for the 29 individual series' means, variances, standard deviations, skewness coefficients, minimum observed values of the ξ_i series; as well as the lower boundary, shape and scale parameters of the 29 fitted three-parameter gamma distribution functions. The average of means is 0.022, with a variance of 0.002, and is close to zero, while the average of variances is 1.091, with a variance of 0.008, and is close to unity. The average of the 29 standard deviations is 1.043, with a variance of 0.001, while the average of the skewness coefficients is 1.173, with a variance of 0.033. The small variations of the $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ estimates about the average values of 2.616, 0.655, and -1.683, as well as the small variations of the above general parameters, imply that the ξ_i variables computed by Eq. 3.11 are a identically distributed, third-order stationary stochastic process.

Table 4.5

Average Statistics of the Parameters of the 29 ξ_i Variables for Nebraska

Statistic	Average	Variance	Standard Deviation	Skewness Coefficient
Mean	0.022	0.002	0.053	0.411
Variance	1.091	0.008	0.093	0.972
Standard Deviation	1.043	0.001	0.044	0.575
Skewness Coefficient	1.173	0.033	0.182	0.055
Minimum Observed Value	-1.631	0.003	0.057	0.410
Shape Parameter	2.616	0.059	0.248	-0.045
Scale Parameter	0.655	0.001	0.043	0.090
Lower Boundary	-1.683	0.004	0.067	0.151

4.6. Regional Correlation Structure for Identically Distributed Stochastic Components. The linear regional dependence between the identically distributed ensemble of stationary stochastic variables in Region II is here investigated for the lag-zero cross correlation coefficients as related to the

Table 4.4

Fitting Benchmark Probability Distribution Functions to Frequency Distributions of ξ_i Series for Region II

Station Number	NORMAL FUNCTION Mean	Standard Deviation	χ^2	LOGNORMAL FUNCTION Lower Boundary	Mean	Standard Deviation	χ^2	GAMMA FUNCTION Lower Boundary	alpha	beta	χ^2	Minimum Observed Value	Variance	Skewness Coefficient	Excess Coefficient
1	.041	1.033	90.22	-1.588	.266	.727	41.55	-1.613	2.404	.688	18.11	-1.579	1.068	.997	.953
2	-.048	1.046	74.88	-1.724	.328	.647	22.77	-1.749	2.767	.615	13.88	-1.716	1.095	1.451	3.343
3	-.016	.985	78.22	-1.732	.369	.626	22.22	-1.761	3.042	.574	16.00	-1.724	.970	.962	.641
4	.063	1.091	69.00	-1.671	.342	.718	35.55	-1.786	2.982	.620	14.55	-1.654	1.190	1.352	2.832
5	-.042	.989	81.77	-1.687	.309	.663	29.11	-1.714	2.787	.599	28.00	-1.677	.978	1.111	1.322
6	.104	1.143	69.66	-1.682	.373	.701	23.88	-1.711	2.560	.709	13.88	-1.674	1.306	1.473	3.601
7	.030	1.021	60.00	-1.687	.357	.648	35.22	-1.713	2.839	.614	20.88	-1.679	1.044	1.099	1.537
8	.030	1.017	69.22	-1.503	.178	.785	36.44	-1.528	2.192	.711	16.66	-1.493	1.035	1.103	1.495
9	.030	1.032	99.77	-1.570	.241	.744	34.88	-1.595	2.352	.691	15.55	-1.560	1.066	1.088	1.326
10	-.048	1.016	73.66	-1.701	.301	.682	23.22	-1.727	2.626	.639	12.88	-1.692	1.033	1.011	.958
11	-.035	.990	75.66	-2.237	.695	.441	26.33	-1.662	2.564	.634	25.33	-1.543	.981	1.088	1.571
12	.051	1.028	56.88	-1.682	.365	.655	33.55	-1.709	2.831	.622	25.22	-1.673	1.058	1.180	2.504
13	.039	1.109	86.66	-1.604	.274	.710	24.55	-1.626	2.369	.703	8.66	-1.597	1.231	1.455	2.818
14	.058	1.074	76.44	-1.654	.332	.706	27.77	-1.682	2.598	.670	13.11	-1.642	1.153	1.324	2.811
15	.077	1.028	71.22	-1.607	.299	.755	41.44	-1.739	2.910	.624	15.33	-1.597	1.058	.747	-.047
16	.023	1.036	73.22	-1.667	.329	.677	36.44	-1.693	2.682	.640	20.77	-1.658	1.075	1.291	2.930
17	.143	1.050	77.00	-1.660	.412	.638	36.00	-1.688	2.915	.628	21.88	-1.653	1.104	.971	1.011
18	-.041	.987	71.66	-1.734	.348	.647	26.11	-1.762	2.925	.588	15.77	-1.724	.974	1.016	1.026
19	-.019	1.037	99.22	-1.622	.264	.689	23.22	-1.626	2.465	.651	16.22	-1.592	1.076	1.323	2.183
20	.098	1.101	77.66	-1.618	.322	.721	35.88	-1.644	2.448	.712	9.11	-1.608	1.212	1.268	2.492
21	.023	1.009	64.55	-2.275	.742	.432	18.33	-1.628	2.409	.685	11.00	-1.589	1.019	.989	1.026
22	-.001	1.086	80.88	-1.676	.284	.754	34.33	-1.697	2.316	.732	8.44	-1.659	1.180	1.115	1.257
23	-.005	1.090	82.88	-1.670	.296	.710	29.66	-1.670	2.386	.697	18.88	-1.633	1.188	1.456	2.868
24	-.019	1.014	70.44	-1.566	.190	.790	48.00	-1.576	2.151	.724	29.77	-1.537	1.028	1.098	1.460
25	.083	1.127	73.55	-2.384	.810	.438	18.22	-1.822	2.902	.657	15.22	-1.674	1.270	1.438	4.156
26	-.040	.997	76.44	-1.657	.284	.670	18.55	-1.682	2.677	.613	15.33	-1.649	.995	1.135	1.581
27	-.035	.994	82.00	-1.611	.244	.697	39.88	-1.633	2.504	.638	28.00	-1.602	.989	1.129	1.583
28	.013	1.066	73.22	-1.662	.291	.751	23.55	-1.764	2.753	.645	9.22	-1.639	1.136	1.104	1.490
29	.102	1.072	77.00	-1.612	.332	.700	43.66	-1.626	2.524	.684	26.77	-1.588	1.149	1.252	2.481

interstation distance d . This is similar to that done for Region I, with the simplification that the azimuth, or plane normal angle is not assumed to contribute a significant amount to the explained variance of r for the fitted $\rho=f(d)$ function. Models VI, VII, VIII, and X are selected for study utilizing the approach of Eq. 3.19 and the initial values of coefficients in the functions, as given in Table 4.6. The range of interstation distance for Region II is from 15 to 192 miles. The findings for the model, $\rho=f(d)$, for Region II are similar to those for Region I. Model VII was found to be somewhat superior to Model VI, with the explained variance of 92.1 and 91.0 percent for the variation in r, respectively. Forcing the exponent to 0.50 in place of the 0.571 used in Model VII, R^2 is increased by only 0.10 percent, showing that the choice of the $\rho=f(d)$ model is not critical. Figure 4.12. shows the fit of Model VII with the tolerance limits computed by Eq. 3.24; 12.9 percent of the points fall outside these limits instead of the desired five percent.

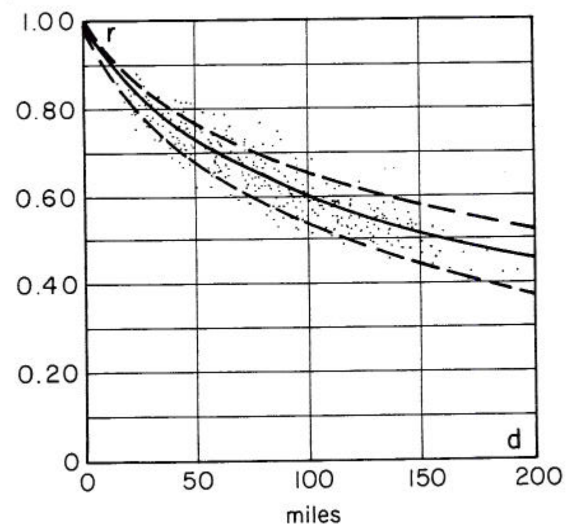

Fig. 4.12. Model VII for Region II and ξ_i variables with 95 percent tolerance limits.

Table 4.6

Fitting Model of $\rho=f(d)$ to Relationship of r to d .

Model	Regression Equation	Series	Computer Time sec.	Number of Iterations	Mean of r	Variance of r	Std. of r	Mean ξ_i Error	Sample Size	No. of Points	% of Explained Variance, R^2
VI	$r=(1+Ad)^{-1}$ A=0.007	ξ_i	4.70	12	.665	.017	.132	.001	29	435	91.0
VII	$r=(1+Ad)^{-n}$ A=0.015 n=+0.571	ξ_i	12.25	10	.665	.017	.132	.001	29	435	92.1
VIII	$r=\exp(Ad)$ A=-0.005	ξ_i	4.76	7	.665	.017	.132	.002	29	435	85.8
X	$r=(1+Ad)^{-\frac{1}{2}}$ A=0.018	ξ_i	14.17	9	.665	.017	.132	.001	29	435	92.2

CHAPTER V

COMPARISONS AND CONCLUSIONS

This chapter contains two sections, the comparison of the results of the structural analysis of area-time hydrologic processes for obtaining independent, stationary and identically distributed stochastic variables of monthly precipitation for two topographically homogeneous regions in the central United States, and the resulting conclusions. The comparison shows that the stochastic ξ_i variables are the same regardless of the region once the basic parameters of the monthly precipitation have been regionalized and the specific regional characteristics removed. For the three-parameter gamma probability distribution of ξ_i, with the same parameters over all of the regions, the cross correlation structure is a function of position, interstation distance, orientation of station-connecting lines, and lastly of the time interval of the year.

5.1. Comparisons of Regionalized Parameters. The major parameters, estimated by the sample statistics, of the ξ_i variables, of the 41 stations in Region I and of the 29 stations in Region II are compared. For each region the mean, variance, standard deviation, skewness coefficient, the lowest observed value, and the shape, scale, and lower boundary parameters of the three-parameter gamma probability distribution function are either made and/or found to be statistically identical for all the ξ_i variables. Table 5.1 presents the basic statistics of these eight parameters, namely the average, variance, standard deviation and skewness coefficient for Region I, Region II, and composite values.

Because the variables ξ_i are approximately standardized for both regions, and because the lowest monthly precipitation values are zero, the mean, variance, and the lower boundary should be equal for all series. Table 5.1 shows that the parameters are approximately equal for Region I and Region II. The parameters are also computed for the composite samples of 41 and 29 stations. This total sample has properties similar to the individual samples, except for the skewness coefficients, which vary over a larger range.

The comparison of the ξ_i variables of Region I and II is made because this enables ξ_i to be the same variable not only for these two regions but for all others, provided that proper inferences are made for the time periodicity and areal trends in the basic parameters, coefficients and/or their ratios. In other words, the ξ_i variables are time independent and three-parameter gamma distributed with $\hat{\alpha}=2.618$, $\hat{\beta}=0.663$, and $\hat{\gamma}=1.690$, independent of both the region and the seasonal variation in the case of monthly precipitation.

Further improvements in the basic models leading to the ξ_i variables can be accomplished as follows:

(a) By including more harmonics in the periodic functions, and by better discriminating the significant amplitudes by advanced test procedures;

(b) By using more terms, when proven to be significant, in the trend surface functions for parameters

Table 5.1

Comparison of Basic Statistics of Parameters of ξ_i Variables for Region I and II.

Parameter	Area	Sample Size	Basic Statistics: average	variance	standard deviation	skewness coefficient
Mean	Region I	41	0.042	0.008	0.092	0.306
	Region II	29	0.022	0.002	0.053	0.411
	Composite	70	0.034	0.006	0.078	0.505
Variance	Region I	41	1.135	0.022	0.150	0.972
	Region II	29	1.091	0.008	0.093	0.636
	Composite	70	1.117	0.017	0.131	1.132
Standard Deviation	Region I	41	1.063	0.004	0.068	0.743
	Region II	29	1.043	0.001	0.044	0.575
	Composite	70	1.055	0.003	0.060	0.903
Skewness Coefficient	Region I	41	1.156	0.033	0.181	0.633
	Region II	29	1.173	0.033	0.182	0.055
	Composite	70	1.163	0.032	0.180	0.386
Lowest Observed Value	Region I	41	-1.651	0.003	0.056	0.175
	Region II	29	-1.631	0.003	0.057	0.410
	Composite	70	-1.642	0.003	0.057	0.267
$\hat{\alpha}$	Region I	41	2.620	0.090	0.301	0.972
	Region II	29	2.616	0.059	0.244	-0.045
	Composite	70	2.618	0.076	0.277	0.698
$\hat{\beta}$	Region I	41	0.668	0.004	0.063	0.414
	Region II	29	0.655	0.001	0.043	0.090
	Composite	70	0.663	0.003	0.056	0.475
$\hat{\gamma}$	Region I	41	-1.694	0.003	0.056	0.247
	Region II	29	-1.683	0.004	0.067	0.151
	Composite	70	-1.690	0.003	0.060	0.239

when expanded in a power series;

(c) By finding the most appropriate functions for relationships between the basic parameters, in order to decrease the total number of coefficients to be estimated; and

(d) By assuming that the mean of the first serial correlation coefficients, $\bar{r}_1$, of the ξ_i monthly precipitation series, is a positive value, and using the model $\xi_i = \bar{r}_1 \xi_{i-1} + \eta_i$, with η_i a less time dependent variable than ξ_i, and consequently studying η_i as the basic variable for analysis.

Instead of fitting the three-parameter gamma probability function to ξ_i, ξ_i may be transformed to approximately symmetrical random variables, ν_i, by the cube-root transformation or by a similar approach. Then, a fit of the normal function would make ν_i identically distributed normal variables, with mean zero and variance unity.

5.2. Correlation Structure Among ξ_i Variables. The comparison is made for Region I and II for the lag-zero cross correlation functions, $\rho=f(d)$, for Models VI, VII, VIII, and X, as shown in Table 5.2.

Table 5.2

Comparisons of Correlation Structural Models of ξ_i Variables for Region I and II

Model No.	Model Equation	Region	Coefficients A	Coefficients n	Explained Variance of r R^2
VI	$\rho=(1+Ad)^{-1}$	I	0.006		89.8
		II	0.007		91.0
VII	$\rho=(1+Ad)^{-n}$	I	0.016	0.534	90.8
		II	0.015	0.571	92.1
VIII	$\rho=e^{-Ad}$	I	0.006		78.8
		II	0.005		85.8
X	$\rho=(1+Ad)^{\frac{1}{2}}$	I	0.018		91.6
		II	0.018		92.2

This Table 5.2 demonstrates that for all four models the difference either in the model coefficients (A, or A and n) or in R^2 are relatively small for these two regions. In Model VII the value of A is higher when n is small and vice versa.

To improve the description of the cross correlation structure, the following steps may be feasible:

(a) To test models with more coefficients in order to increase the explained variance;

(b) To test whether the other independent variables, such as X, Y, ϕ, and τ, in the general model $\rho=f(d,X,Y,\phi,\tau)$ contribute significantly to the explained variance in the variation of r, as an estimate of ρ.

To determine whether the type of the model, as well as the number of independent variables can contribute significantly to better mathematically describing the cross correlation structure for the ξ_i variables, two procedures should be applied: first, develop good testing techniques for the inferences of both the models and the effects of various independent variables; and second, test the models and independent variables on a large number of regions, so that the models and the order of magnitude of the effects of individual independent variables (d,X,Y,ϕ,τ) may be well inferred and used to determine the best generally applicable approach. This will be similar to studying the probability distribution functions of the best fit by testing them in many regions and for a large number of variables of the same physical magnitude.

5.3. Conclusions. The major parameters of the line-time, area-time, and space-time hydrologic stochastic processes are periodic in time and have trends along the line, across an area, or over a space. The basic approach in this study was to investigate a typical area-time random process as an example for inferring and removing the area-time variations in the basic parameters. The coefficients of any of the periodic parameters may also vary along a line, across an area, or over a space. The random variables observed at the set of points can be utilized to produce a set of stationary, but cross correlated time series. The sets of monthly precipitation series over the two regions are used to demonstrate the methodology applied in the structural analysis, by showing that the removal of periodicities and trends in the basic parameters produces a cross correlated stationary process.

The results of investigations in previous chapters are summarized as follows:

A. Periodic Parameters.

A.1. All basic parameters of an observed line-time, area-time, or space-time stochastic hydrologic process may be periodic. The mean and the variance (or the standard deviation) are periodic for all time series with an interval of less than a year. The autocovariances (or the autocorrelation coefficients), the third-order parameters (skewness coefficients and the dimensionless three-terms autoproducts), and the higher-order parameters may also be periodic.

A.2. For the area-time stochastic monthly precipitation process, the mean and standard deviation are always periodic, except in very rare cases of even distribution of the mean monthly precipitation over the year.

A.3. The monthly periodic parameters may be described by a very small number of harmonics (say the 12-month, 6-month, and possibly 3-month harmonics). Of the two regions studied, the 12-month harmonic explains a major portion of the periodic variations in both the monthly mean and the monthly standard deviation.

A.4. Since the number of coefficients to be estimated in the parameter periodic function is $2j+1$, with j the number of significant harmonics, (amplitudes are significantly greater in these harmonics for the periodic parameter than they would be for a non-periodic parameter). The smaller the number of significant harmonics, the smaller is the number of coefficients to be estimated. For the monthly mean precipitation of the two regions, only three coefficients were needed: the general mean, and the amplitude and phase of the 12-month harmonic for the periodic mean. Similarly, only three coefficients were needed for the periodic standard deviation.

A.5. Since the above two parameters and the four coefficients of the periodic mean and standard deviation vary across any region, the regional trend surfaces are required. In case other harmonics (6-month, 4-month,...) are inferred to be periodic, this increases the number of coefficients to be regionalized by four for each additional harmonic in the two basic parameters of the mean and the standard deviation. Therefore, an economy in the number of periodicity coefficients is required both from the computational point of view and because of the large sampling variations in coefficients.

A.6. For the two regions studied it was found that the 12-month harmonic explains large portions of the variations in the periodic mean and periodic standard deviation. The five remaining harmonics each explain a relatively small portion of the variations. Therefore, the five harmonics (other than the 12-month) are inferred to be non-significant and to be the result only of sampling variation of residuals about the 12-month harmonic. For a further refinement of the periodic function, the 6-month harmonic may be added. However, a very small gain in the explained variance in the estimates of the periodic mean and standard deviation most likely would not justify the additional work required in the regionalization. This additional work would entail fitting trend surfaces to the four additional coefficients, two amplitudes, and two phases. The simple 12-month cosine function for the periodic mean and the periodic standard deviation, as found for the two regions used here as examples, can only be expected in exceptional cases, and may not be a general feature for most monthly precipitation series.

A.7. Since the variance of monthly precipitation is a periodic parameter over the year, the estimates of the periodicity in monthly means and monthly standard deviation by the least-square method with equal weights for the residuals, or the linear approach in estimating the Fourier coefficients A_j and B_j (and from them the amplitude C_j and the phase θ_j), does not give the most efficient estimates. The use of weighted residuals in the least-square approach, with the appropriate weights as functions of the variance of each estimated parameter, may produce estimates of C_j and θ_j which explain a larger portion of the variances in the periodic mean and standard deviation.

A.8. Improvements in the mathematical models for periodicity in parameters can only be done by a better discrimination of significant harmonics using a better method of separating them from the sampling errors of residuals around the periodic function. To accomplish these improvements, three approaches are feasible: (a) make the logarithmic or cubic-root transformation of the original monthly precipitation values so that the sampling variations of residuals about the parameter periodic functions are close to normal; (b) test the residuals for time dependence, and if the dependence is inferred, reduce the length of dependent residuals to an equivalent length of independent normal residuals; and (c) use the Fisher's test, based on an underlying normal independent process, to discriminate all significant harmonics.

A.9 The phases of harmonics which describe the periodic parameters of monthly precipitation must change over the long distances, because the periodicity in parameters of major climatic variables also change over these distances. Significant peaks and lows of periodic parameters slowly change their positions inside the year with the distance. This is equivalent of changing the phases with the distance. By using the appropriate regional models of the change with longitude and latitude in phases of harmonics of periodic mean and standard deviation, even if their rates of change are very small, could further improve the reliability in accounting for the regional variation in coefficients of significant harmonics.

B. Regional Trends in Parameters.

B.1. All basic parameters, and coefficients of the periodic parameters, vary according to some continuous line, area, or space functions. All of them can be expanded in a power series; thus as many terms should be used in fitting these functions to the estimated set of parameter or coefficient values as the appropriate inference procedures require. It is convenient to start with a linear (line, area, space) trend and then increase the accuracy of fitting by going to second-order or higher-order terms.

B.2. For the two regions studied, with given sets of points of monthly precipitation series, it was found that the simple trend planes, with the parameters (general mean and general average monthly standard deviation) and the coefficients of periodic functions of these parameters being linear functions of the point coordinates, explain very large portions of the variation in these parameters and coefficients (the two amplitudes and two phases of the 12-month harmonic of these two parameters). The second-order terms of coordinates (X^2, Y^2, and XY) explain relatively small portions of the total regional variation of these two parameters and four coefficients. This leads to the conclusion that the second and higher-order terms in the regional trend surface are mainly the results of the sampling variation of residuals about the trend planes.

B.3. It was also found that the ratio of the general average monthly standard deviation to the general mean monthly precipitation, or the general coefficient of variation, is not significantly different from a constant for the two regions. In addition the two other ratios, the amplitude of the 12-month harmonic in the monthly mean divided by the general mean, and the amplitude of the 12-month harmonic of the monthly standard deviation divided by the general mean, are also two regional constants. Furthermore, the phases of the two 12-month harmonics are inferred to be constants. In this approach, the six regional trend planes, each requiring three trend regression coefficients, of a total of 18 coefficients to be estimated, are reduced to one regional trend plane with three regression coefficients, and four or five other constants (three ratios, and one or two phases). Altogether 7 or 8 coefficients are to be estimated. This is a significant reduction in the number of required estimates. In the case of using all six harmonics ($j=6$) for the periodic mean and periodic standard deviation, a total of $2(2j+1)=26$ coefficients is required. Assuming further that each of these 26 coefficients required six regression coefficients of a second-order polynomial trend, a total of 156 coefficients must be estimated. The various simplifications tested for this study have reduced this number to only seven or eight coefficients necessary to be estimated from all the data available.

B.4. The average phases of the 12-month harmonic in the periodic mean and periodic standard deviation were not found to be identical. Because of the observed difference in phases, the pertinent question may be asked whether there is actually a shift in the phase or whether the observed difference was due to sampling variations. It would be difficult to find physical reasons why the harmonics in the basic parameters of a periodic-stochastic hydrologic process would have different phases. One conclusion is that the difference between the phase of 12-month harmonic in the two basic periodic parameters is the result of sampling variation rather than the result of some systematic patterns. Many more regions should be studied before convincing evidence can be generated for a population difference in phases.

B.5. To decrease the adverse boundary effects of the end points in a fitted trend, the trend regression equation may be estimated by a larger number of points and the resulting equation applied only to the interior points of a line, area, or space. For monthly precipitation in Region I, 77 points are used for determining the trend plane regression function; then this function is used for the 41 points of the interior subregion.

B.6. Improvements in fitting the trend functions can be accomplished in a number of ways; (a) by using a larger number of terms in the power series of these functions; (b) by weighting the residuals for the function of the trend in the variances; (c) by better testing the significance of regression coefficients; (d) by adding physical aspects to selecting the trend functions, namely, studying how the topography and other physical factors should affect the shape of the trend function; (e) by using a longer line, larger area, or greater space in fitting the trend functions, and then applying them to an interior shorter line, smaller area, or narrower space; and (f) by a more rigorous testing of proportionality (or other relationships) between the coefficients or the secondary parameters and the major parameters, so that either ratios and some coefficients may be inferred to be constants or to have some simple relationships, with the result being a reduced number of coefficients to be estimated.

C. Analysis of Stationary Stochastic Process.

C.1. When the periodicity in basic parameters and the regional trends in parameters and/or coefficients are inferred and removed by the appropriate mathematical models, a second-order or higher-order stationary process is obtained. The set of series of this process then contains all randomness of this process, except the sampling variations in the small number of estimated coefficients (7-8 coefficients for the area-time stochastic monthly precipitation process of the two regions studied). In other words, the sampling variation around the functions of periodic parameters and around the regional trend functions are retained in the stationary stochastic components, instead of being removed as in the case of the nonparametric approach for removing time periodicity and regional (line, space) trends.

C.2. In the case of removal of the general mean and the general variance from the stochastic components by standardization, all the point time series have approximately mean zero and variance unity. Then, because the basic hydrologic variables have a lower boundary of zero, the set of stochastic components are nearly identically distributed variables. For the case of the monthly precipitation in the two sample regions, three-parameter gamma probability distribution function excellently fits the frequency curves derived from the set of stationary stochastic time series. As previously shown for each region or for their composite, the three parameters are constants.

C.3. If the periodicity in parameters and the trends in coefficients are well approximated, they should produce the same, identically distributed stochastic components with the probability function and its estimated parameters the same for a given type of hydrologic stochastic process. For monthly precipitation, this is the three-parameter gamma function, with $\hat{\alpha}=2.618$, $\hat{\beta}=0.663$, and $\hat{\gamma}=-1.690$ (shape, scale and lower boundary parameters). These three parameters can be improved: (a) by including more regions in the study; (b) by better estimating the number of harmonics and their coefficients for periodic parameters; and (c) by better inferring the order of the polynomial and its regression coefficients used as trend functions for the hydrologic parameters.

C.4. The time dependence structure of the stationary stochastic components require either the determination of the time dependence model, if the stationary process is dependent, or proof that it is time independent.

C.5. For the two sample regions the stationary stochastic components are approximately time independent processes. The first serial correlation coefficients rarely exceeds 0.10. There must be a small time dependence in the stochastic component of monthly precipitation, because the stochastic components of daily precipitation series usually have the first serial correlation coefficient of the order of 0.10-0.30. However, a sum of 30 values of stochastic components of the daily precipitation should exhibit only small time dependence. For practical purposes, the stationary stochastic components of monthly precipitation series may be considered time independent. To improve this independence, one may use the first autoregressive model and the average first serial correlation coefficient of all series in a region to better approximate the time independent portion of the stationary identically distributed stochastic components.

D. Cross Correlation Structure of Stationary Components.

D.1. The basic approach in studying the dependence between the stationary identically distributed stochastic components is: (a) by selecting any two points (i and j) at a time from a set of points, and (b) by using the lag-zero linear cross correlation coefficient, ρ, as a measure of dependence. In defining the two points in an area-time random process, the four coordinates (X_i, Y_i, X_j, Y_j) are needed. However, this correlation may also depend on the time interval τ (say, for precipitation, the correlation may be greater for winter frontal precipitation than for summer thunderstorm precipitation). The function $\rho=f(X_i,Y_i,X_j,Y_j,\tau)$ may be replaced by $\rho=f(d,X,Y,\phi,\tau)$, with X,Y determining the location of a point (either X_i, Y_i, or X_j, Y_j, or their mid-points), d the distance and ϕ the azimuth of the straightline which connects the points.

D.2. For the monthly precipitation stochastic components, ρ seems not to depend on the absolute position (X,Y), azimuth (ϕ) and the time interval (τ), but only on the distance. Despite these simplifying assumptions, $r=f(d)$ as a fitted estimate of $\rho=f(d)$ explains a large portion of the variation of r. It should be noted that the inclusion of the azimuth adds a very small percentage to the explained variance.

D.3. Any rigorous approach to the study of dependence among a set of stationary stochastic components, either along a line, across an area, or over a space, should be accompanied by the development of reliable tests for determining the effects of the position, azimuth and the time interval on the population variation in ρ. This may add to the reliability of estimates for models of $\rho=f(d,X,Y,\phi,\tau)$.

D.4. Because the stochastic components may be non-normal, or highly skewed, a transformation to a symmetrical or even a normally distributed set of time independent stationary variables may better justify the linear lag-zero cross correlation approach to the study of the dependence among these variables.

D.5. The investigations of the relationship $\rho = f(d)$ and of the residuals $(r-\rho)$ for the studied monthly precipitation series of the two regions, show that these residuals are distributed as would be expected from sampling variation alone; thus the simple model $\rho=f(d)$ is justified for all practical purposes.

D.6. The selection of the mathematical function for $\rho=f(d)$ is based mainly on a survey of literature concerning models previously used or tested on similar variables having dependence as a function of distance. Basically, the boundary conditions for $d=0$ and $d=\infty$ are considered, as well as the explained variance of r by the model and the model's estimated coefficients for the two regions studied. Unfortunately, there is no evidence in the literature which can physically justify the use of particular mathematical models.

D.7. The testing of a large number of $\rho=f(d)$, or $\rho=f(d,X,Y,\phi,\tau)$ types of models over many regions, and for interrelated variables (such as solar radiation, temperature, humidity, wind velocities, evaporation, precipitation, etc.) may empirically lead to the best practical models. This study may be a modest contribution toward that goal.

D.8. The problem of whether $\rho=f(d)$ is a positive function slowly converging to zero as d goes to infinity is a controversial subject. Because the sample r values can be negative for large d values, and even of the order of -0.30 to -0.40, there is an open question whether these negative values are the sampling variation about $\rho=0$, or the population ρ values are also negative. Powerful statistical inference techniques and large quantity of data from all over the world are needed to give a reliable answer to this question.

D.9. By inferring the type of ρ-function, either along a line, across an area, or over a space, the study of these line-time, area-time, and space-time periodic-stochastic hydrologic processes is dissociated from the set of observed point-time series.

D.10. The curve $\rho=f(d)$ for the periodic-stochastic $x_{p,\tau}$ discrete series with the interval less than a year in a region are usually above the corresponding $\rho=f(d)$ curve for the stationary stochastic ξ_i series. The reason is that the periodicity in parameters of $x_{p,\tau}$ is mostly in-phase over a small region. This increases the correlation coefficient ρ for a given distance d. Only in cases where the region is sufficiently large that there is an out-of-phase periodicity in parameters between the extremes of the region will it be possible for the $\rho=f(d)$ curve to be lower for $x_{p,\tau}$ than for ξ_i. Since the $\rho=f(d)$ curves for ξ_i are usually below the corresponding curves of $x_{p,\tau}$, the average cross correlation coefficient for a set of points in a region is smaller for ξ_i than for $x_{p,\tau}$.

E. General Conclusions.

E.1. The results of this study support the basic hypothesis, namely that a set of point-time series has more information on all parameters of a given point series than each individual series.

E.2. The regionalization of basic parameters of time series and of coefficients of periodic parameters reduces significantly the number of coefficients to be estimated.

E.3. In the case of monthly precipitation series, the area-time process is sufficiently described by four types of mathematical models:

(a) Periodic functions for periodic parameters;

(b) Regional trend planes;

(c) Three-parameter gamma probability distribution function of identically distributed, time independent stationary stochastic components; and

(d) The regional dependence function for these stochastic components. In the simplest case, there were only two periodic functions (for the monthly mean and the monthly standard deviation), one regional trend plane (for the general mean), one gamma function, and the simple relation of the lag-zero cross correlation coefficient to the distance. Only five equations and the estimation of 13 or 14 parameter-coefficients was necessary.

E.4. Starting with the simplest four mathematical models in the structural analysis of area-time periodic-stochastic process of monthly precipitation, with only 13 or 14 estimates, both the number (or also the type) of mathematical models and their terms can be increased by also continuously increasing the number of parameters and/or coefficients to be estimated. The procedures used in this paper have shown how to make these increases in the number of models, the number of model terms, and the number of estimates; and how to progress from a simple to a very sophisticated approach. The statistical inference and testing procedures are crucial in determining the cut-off points, that is, those points at which one should cease adding more models, more terms, and more coefficients.

E.5. The models and estimates of coefficients represent a condensation of all area-time information, so that the solution of practical problems may be dissociated from the set of observed point-time series. A systematic grid of points across an area (or along a line, or over a space) may facilitate the solution of problems related to these processes. If the grid points are fixed as to their mutual distances and if the probability distribution function of the stochastic components is always the same for all regions, as well as having their dependence function $\rho=f(d)$ the same for all regions studied, a set of point-time series for this grid may be permanently generated for each time sample size, and used for many regions. It is then necessary to add only the particular periodic and trend functions specific to each individual region.

E.6. A reduction of the area-time periodic-stochastic process to a stationary and ergodic stationary process greatly facilitates the solution of various practical problems.

E.7. In case the stationary stochastic components are also time dependent, a fifth type of mathematical model is needed, namely the time dependence model with additional coefficients to be estimated.

E.8. Because the multivariate normal independent processes are easiest to study and use in generating new samples, the approach in these investigations permits a reduction of the non-normal multivariate stochastic stationary process to a multivariate normal process. Then, principal component analysis enables this multivariate dependent normal process along a line, across an area, or over a space, to be replaced by an equivalent independent uncorrelated multivariate normal process.

APPENDIX 1

Precipitation Stations Selected

Region I

Table 1
m=77 Stations

m	Index Number	Station Name		Degrees West Long.	Degrees North Lat.	Elev. Above MSL (ft.)	12 mo. 40 yr. Mean	12 mo. 40 yr. STD
1	21.0112	Alexandria FAA Airport		95.383	45.867	1421	1.915	1.749
2	21.0287	Artichoke Lake		96.133	45.367	1075	1.871	1.738
3	21.0541	Beardsley		96.717	45.550	1090	1.815	1.675
4	21.0783	Bird Island		94.900	44.767	1089	2.253	1.968
5	21.1227	Cambridge St. Hospital		93.233	45.567	1000	2.266	1.926
6	21.1245	Campbell		96.400	46.100	975	1.872	1.662
7	21.1374	Cass Lake		94.633	47.383	1320	2.040	1.790
8	21.1465	Chaska		93.583	44.800	726	2.274	2.052
9	21.1630	Cloquet For Res Center		92.500	46.683	1265	2.440	1.789
10	21.2142	Detroit Lakes	1NNE	95.850	46.833	1375	1.926	1.818
11	21.2737	Farmington	3NW	93.183	44.667	902	2.379	2.030
12	21.2768	Fergus Falls		96.067	46.283	1210	1.972	1.682
13	21.2916	Fosston Power Plant		95.733	47.567	1289	1.898	1.748
14	21.3303	Grand Rapids NC School		93.500	47.250	1281	2.148	1.696
15	21.3411	Gull Lake Dam		94.350	46.417	1215	2.209	1.919
16	21.4652	Leech Lake Dam		94.217	47.250	1301	2.070	1.872
17	21.4793	Little Falls	1N	94.350	45.983	1120	2.138	1.878
18	21.5012	Manhomen	1W	95.983	47.317	1203	1.836	1.661
19	21.5136	Maple Plain		93.650	45.017	1030	2.421	1.975
20	21.5298	Meadowlands		92.733	47.067	1270	2.236	1.880
21	21.5392	Milaca		93.650	45.750	1080	2.392	2.043
22	21.5400	Milan		95.933	45.177	1005	2.007	1.794
23	21.5563	Montevideo	1SW	95.717	44.933	900	2.138	1.895
24	21.5615	Mora		93.300	45.883	1001	2.351	1.941
25	21.5638	Morris WC School		95.917	45.583	1130	1.923	1.727
26	21.5842	New London		94.933	45.300	1215	2.130	1.929
27	21.6360	Park Rapids		95.067	46.917	1434	2.214	1.966
28	21.6547	Pine River Dam		94.117	46.667	1251	2.198	1.865
29	21.6565	Pipestone		96.300	44.000	1735	2.071	1.894
30	21.6612	Pokegama		93.583	47.250	1280	2.122	1.754
31	21.6817	Red Wing		92.550	44.567	680	2.446	1.962
32	21.7460	Sandy Lake Dam Libby		93.317	46.800	1234	2.259	1.948
33	21.8323	Tracy Power Plant		95.617	44.233	1403	1.992	1.769
34	21.8543	Virginia 01MC Lab		92.533	47.533	1445	2.245	1.701
35	21.8579	Wadena		95.133	46.433	1350	2.175	1.919
36	21.8618	Walker Ah Gwan Ching		94.583	47.067	1407	2.129	1.928
37	21.8692	Waseca Experiment Farm		93.517	44.067	1153	2.403	1.938
38	21.8907	Wheaton		96.483	45.800	1018	1.876	1.707
39	21.9004	Willmar State Hospital		95.017	45.133	1133	2.143	1.955
40	21.9046	Winnebago		94.167	43.767	1110	2.420	2.038
41	21.9059	Winnibigoshish Dam		94.050	47.433	1315	2.023	1.723
42	21.9249	Zumbroto		92.667	44.300	985	2.340	1.952
43	32.1360	Carrington		99.133	47.450	1586	1.493	1.522
44	32.1766	Cooperstown		98.117	47.433	1428	1.573	1.557
45	32.2482	Edgeley Exp Farm		98.700	46.333	1568	1.484	1.550
46	32.2605	Ellendale		98.517	46.000	1460	1.654	1.607
47	32.2949	Fessenden		99.617	47.650	1620	1.462	1.377
48	32.3117	Foreman		97.650	46.100	1249	1.667	1.659
49	32.3287	Fullerton		98.417	46.167	1439	1.725	1.591
50	32.3908	Hankinson RR Station		96.900	46.067	1068	1.647	1.598
51	32.4203	Hillsboro		97.067	47.400	901	1.674	1.701
52	32.4418	Jamestown State Hospital		98.683	46.883	1457	1.469	1.489
53	32.5660	Mayville		97.317	47.500	975	1.526	1.551
54	32.5754	McLeod	3E	97.233	46.400	1075	1.596	1.616
55	32.6620	Oakes		98.083	46.133	1320	1.593	1.542
56	32.7047	Pettibone		99.517	47.117	1855	1.400	1.414
57	32.7986	Sharon		97.900	47.600	1516	1.614	1.571
58	32.8366	Steele		99.917	46.850	1857	1.415	1.493
59	32.9100	Whapton State School		96.600	46.267	960	1.764	1.704
60	39.0834	Bowdle		99.650	45.450	1993	1.532	1.583
61	39.1076	Brookings	1NE	96.767	44.333	1650	1.738	1.766
62	39.1739	Clark		97.733	44.883	1779	1.721	1.569
63	39.2797	Eureka		99.617	45.767	1884	1.429	1.510
64	39.3029	Forestburg	3NE	98.067	44.033	1231	1.727	1.596
65	39.3217	Gannvalley	1ESE	98.983	44.033	1750	1.496	1.580
66	39.3294	Gettysburg		99.950	45.017	2080	1.409	1.416
67	39.3832	Highmore	1W	99.467	44.517	1890	1.470	1.482
68	39.4037	Howard		97.517	44.017	1565	1.780	1.638
69	39.4516	Kennebec		99.867	43.917	1700	1.405	1.382
70	39.4661	La Delle	7NE	98.000	44.683	1396	1.602	1.575
71	39.5536	Milbank		96.633	45.217	1145	1.792	1.653
72	39.5456	Mellette		98.500	45.150	1290	1.581	1.483
73	39.5561	Miller		98.983	44.517	1587	1.559	1.553
74	39.6282	Onaka		99.450	45.183	1600	1.401	1.471
75	39.8652	Victor	5NE	96.800	45.900	1100	1.788	1.707
76	39.9004	Webster		97.533	45.333	1865	1.762	1.603
77	47.7226	River Falls		92.617	44.867	900	2.495	1.979

APPENDIX 2

Precipitation Stations Selected

Region I

Table 2
m=41 Stations

No. of 77	Thiessen Weight in Percent	No. of 41	Index Number	Station Name		Degrees West Long.	Degrees North Lat.	Elev. Above MSL (ft)
1	3.07	1	21.0112	Alexandria FAA Airport		95.383	45.867	1421
2	1.13	2	21.0287	Artichoke Lake		96.133	45.367	1075
3	1.34	3	21.0541	Beardsley		96.717	45.550	1090
4	4.43	4	21.0783	Bird Island		94.900	44.767	1089
6	.96	5	21.1245	Campbell		96.400	46.100	975
8	2.54	6	21.1465	Chaska		93.583	44.800	726
10	3.56	7	21.2142	Detroit Lakes	1NNE	95.850	46.833	1375
11	2.00	8	21.2737	Farmington	3NW	93.183	44.667	902
12	2.16	9	21.2768	Fergus Falls		96.067	46.283	1210
15	2.08	10	21.3441	Gull Lake Dam		94.350	46.417	1215
16	.96	11	21.4652	Leech Lake Dam		94.217	47.250	1301
17	3.39	12	21.4793	Little Falls	1N	94.350	45.983	1120
19	3.07	13	21.5136	Maple Plain		93.650	45.017	1030
21	2.53	14	21.5392	Milaca		93.650	45.750	1080
22	1.57	15	21.5400	Milan		95.933	45.117	1005
23	2.60	16	21.5563	Montevideo		95.717	44.933	900
25	2.00	17	21.5638	Morris WC School	1SW	95.917	45.583	1130
26	3.21	18	21.5842	New London		94.933	45.300	1215
27	2.30	19	21.6360	Park Rapids		95.067	46.917	1434
28	2.33	20	21.6547	Pine River Dam		94.117	46.667	1251
32	3.18	21	21.7460	Sandy Lake Dam Libby		93.317	46.800	1234
35	2.90	22	21.8579	Wadena		95.133	46.433	1350
36	1.40	23	21.8618	Walker Ah Gwan Ching		94.583	47.067	1407
38	1.07	24	21.8907	Wheaton		96.483	45.800	1018
39	1.59	25	21.9004	Willmar State Hosp		95.017	45.133	1133
45	3.27	26	32.2482	Edgeley Exp Farm		98.700	46.333	1568
46	3.17	27	32.2605	Ellendale		98.517	46.000	1460
48	2.28	28	32.3117	Foreman		97.650	46.100	1249
49	.72	29	32.3287	Fullerton		98.417	46.167	1439
50	1.08	30	32.3908	Hankinson RR Station		96.900	46.067	1068
52	4.00	31	32.4418	Jamestown State Hosp		98.683	46.883	1457
54	4.24	32	32.5754	McLeod	3E	97.233	46.400	1075
55	1.92	33	32.6620	Oakes		98.083	46.133	1320
59	1.80	34	32.9100	Whapton State School		96.600	46.267	960
62	2.79	35	39.1739	Clark		97.733	44.883	1779
70	2.88	36	39.4661	La Delle	7NE	98.000	44.683	1396
71	2.95	37	39.5536	Milbank		96.633	45.217	1145
72	3.87	38	39.5456	Mellette		98.500	45.150	1290
73	2.86	39	39.5561	Miller		98.983	44.517	1587
75	1.10	40	39.8652	Victor	5NE	96.800	45.900	1100
76	3.68	41	39.9004	Webster		97.533	45.333	1865

APPENDIX 3

Precipitation Stations Selected

Region II

Table 3
m=29 Stations

m	Index Number		Degrees West Long.	Degrees North Lat.	Elev. Above MSL (ft.)	30 yr. Monthly Mean	30 yr. Monthly STD
1	25.0070	Albion	98.00	41.68	1760	1.932	1.718
2	25.0320	Arcadia	99.13	41.42	2186	1.711	1.721
3	25.0445	Aurora	98.02	40.87	1792	1.982	1.856
4	25.0620	Beatrice 1	96.75	40.27	1235	2.405	2.287
5	25.1680	Clay Center	98.05	40.53	1778	1.940	1.795
6	25.2020	Crete	96.95	40.62	1360	2.355	2.164
7	25.2770	Ericson 6WNW	98.78	41.80	2095	1.792	1.675
8	25.2805	Ewing	98.35	42.25	1888	1.872	1.797
9	25.2840	Fairmont	97.58	40.63	1641	2.171	1.972
10	25.3035	Franklin	98.95	40.10	1883	1.844	1.777
11	25.3185	Genoa 2W	97.73	41.45	1590	1.949	1.818
12	25.3660	Hastings	98.38	40.58	1932	2.005	1.779
13	25.4335	Kearney	99.08	40.70	2146	1.884	1.840
14	25.4790	Lincoln Agro Farm	96.62	40.85	1195	2.290	1.996
15	25.6135	Oakdale	97.97	42.07	1705	1.941	1.738
16	25.6375	Osceola	97.55	41.18	1645	2.032	1.784
17	25.6395	Osmond	97.60	42.35	1650	2.047	1.722
18	25.7040	Ravenna	98.92	41.03	1995	1.818	1.728
19	25.7515	Saint Paul	98.45	41.20	1796	1.901	1.899
20	25.7640	Schuyler	97.07	41.43	1750	2.236	2.080
21	25.8110	Stanton	97.23	41.95	1472	2.036	1.809
22	25.8320	Superior	98.07	40.02	1578	2.061	1.986
23	25.8735	Upland	98.90	40.32	2158	1.876	1.863
24	25.8745	Utica	97.35	40.90	1582	2.112	2.016
25	25.8905	Wahoo	96.63	41.20	1221	2.286	2.029
26	25.8915	Wakefield	96.87	42.27	1413	1.987	1.751
27	25.8935	Walthill	96.50	42.15	1207	2.070	1.816
28	25.9150	Western	97.20	40.40	1460	2.227	2.074
29	25.9200	West Point	96.72	41.83	1310	2.226	1.886

REFERENCES

1. Amorocho, J. and Brandstetter, A., "The representation of storm precipitation fields near ground level," Journal of Geophysical Research, Vol. 72, No. 4, 1967, pp. 1145-1164.

2. Whitmore, J.S., Van Eeden, F.J. and Harvey, K.J., "Assessment of average annual rainfall over large catchments," Inter-African Conference on Hydrology,"Pub. 66, Nairobi, 1961, pp. 100-107.

3. Akin, J.E., "Calculation of mean areal depth of precipitation," Jour. of Hydrology, Vol. 12, 1971, pp. 363-376.

4. Solomon, S.I., Denouviller, J.P., Chart, E.J. , Wolley, J.A. and Cadou, C., "The use of a square grid system for computer estimation of precipitation, temperature and runoff," Water Resources Research, Vol. 4, No. 5, 1968, pp. 919-929.

5. Krumbein, W.C., "Trend surface analysis of contour -type maps with irregular control-point spacing," Jour. of Geophysical Research, Vol. 64, No. 7, 1959, pp. 823-834.

6. Chidley, T.R.E. and Keys, K.M., "A rapid method of computing areal rainfall," Jour. of Hydrology, Vol. 12, 1967, pp. 15-24.

7. Mandelbaum, H., "Statistical and geological implications of trend mapping with nonorthogonal polynomials," Jour. of Geophysical Research, Vol. 68, No. 2, 1963, pp. 505-519.

8. Unwin, D.J., "The areal extension of rainfall records: an alternative model," Jour. of Hydrology, Vol. 7, 1969, pp. 404-414.

9. Simpson, S.M. Jr., "Least square polynomial fitting to gravity trend data and density plotting by digital computer," Geophysics, Vol. 5, No. 19, 1954, pp. 255-269.

10. Oldham, C.H.G. and Sutherland, D.B., "Orthogonal polynomials; their use in estimating the regional effect," Geophysics, Vol. 20, No. 2, 1955, pp. 295-306.

11. Grant, F., "A problem in the analysis of geophysical data," Geophysics, Vol., 22, No. 2, 1957, pp. 309-344.

12. Hutchinson, P., "Estimation of rainfall in sparsely gaged areas," Bulletin of the International Association of Scientific Hydrology, Vol. 14, No. 1, 1969, pp. 101-119.

13. Huff, F.A. and Shipp, W.L., "Spatial correlations of storm, monthly and seasonal precipitation," Jour. of Applied Meteorology, Vol. 8, 1969, p. 542-550.

14. Steinitz, G., Huss, A., Manes, A., Sinai, R. and Alperson, Z., "Optimum station network in the tropics," Jour. of Applied Meteorology, Vol. 10, 1971, pp. 364-369.

15. Alexeyev, G.A., "Optimal spatial interpolation and its errors in case of various distances between observational points on watershed centers," Proc. International Symposium on Mathematical Models in Hydrology, Warsaw, July 1971, pp. 561-577.

16. Hendrick, R.L. and Comer, G.H., "Space variation of precipitation and implications for network design," Jour. of Hydrology, Vol. 10, No. 2, 1970, pp. 151-163.

17. Cornish, E.A., Hill, G.W. and Evans, M.J., "Interstation correlations of rainfall in southern Australia," Division of Mathematical Statistics Technical Paper No. 10, Commonwealth Scientific and Industrial Research Organization, Melbourne, Australia, 1961.

18. Caffey, J.E., "Inter-station correlations in annual precipitation and in annual effective precipitation," Colorado State University Hydrology Paper No. 6, 1965, Fort Collins, Colorado.

19. Tannehill, I.R.,"Drought - It's Causes and Effects," Princeton University Press, Princeton, New Jersey, 1947.

20. Campbell, D.," Drought: Causes - Effects - Solutions, F.W. Chesire, Melbourne, Australia, 1968.

21. Yevjevich, V., "An objective approach to definitions and investigations of continental hydrologic droughts," Colorado State University Hydrology Paper No. 23, 1967, Fort Collins, Colo.

22. Saldarriaga, J. and Yevjevich, V., Application of run-lengths to hydrologic series," Colorado State University Hydrology Paper No. 40, 1970, Fort Collins, Colorado.

23. Millan, J. and Yevjevich, V., "Probabilities of observed droughts," Colorado State University Hydrology Paper No. 50, Fort Collins, Colorado, June 1971.

24. Barger, G.L. and Thom, H.C.S., "Evaluation of drought hazard," Agronomy Journal, Vol. 41, No. 11, 1949, pp. 515-526.

25. Hershfield, D.M., "Generalizing dry-day frequency data," American Water Works Association Jour., Vol. 62, No. 1, 1970.

26. Foley, J.C., Droughts in Australia, Commonwealth Bureau of Meteorology, Bulletin 43, Melbourne, Australia, 1957, p. 281.

27. Maher, J.V., "Drought assessment by statistical analysis of rainfall," Anzaas Symposium on Drought, Melbourne, Australia, 1967, pp. 57-71.

28. Clark, R.A. and O'Connor, G.E., "Applications of climatology and meteorology to hydrologic simulation," Water Resources Institute, Texas A & M University, College Station, Texas, June 1971.

29. Yevjevich V., Stochastic Processes in Hydrology, Water Resources Publications, Fort Collins, Colorado, 1972.

30. Dixon, W.J., "Biomedical Computer Programs," X-Series Supplement, University of California Publications in Automatic Computation, No. 3, University of California Press, Los Angeles, California, 1969.

KEY WORDS: Regionalization of parameters, structure of area-time processes, monthly precipitation, periodicity in parameters, regional dependence.

ABSTRACT: The structural analysis of area-time hydrologic process of monthly precipitation specifies the deterministic components in form of periodic parameters and a stationary stochastic component. The parameters and coefficients of these components have regional trends.

Monthly precipitation series of two regions demonstrate the structural composition. Region I of North Dakota, South Dakota, and Minnesota with 41 monthly precipitation stations and Region II, of Nebraska with 29 stations are used as examples.

Mathematical models for periodicity and trends in parameters are developed with five regional constants and three regression coefficients. With periodicity and regional trends in parameters removed, the stationary stochastic components of monthly precipitation series are found to be

KEY WORDS: Regionalization of parameters, structure of area-time processes, monthly precipitation, periodicity in parameters, regional dependence.

ABSTRACT: The structural analysis of area-time hydrologic process of monthly precipitation specifies the deterministic components in form of periodic parameters and a stationary stochastic component. The parameters and coefficients of these components have regional trends.

Monthly precipitation series of two regions demonstrate the structural composition. Region I of North Dakota, South Dakota, and Minnesota with 41 monthly precipitation stations and Region II, of Nebraska with 29 stations are used as examples.

Mathematical models for periodicity and trends in parameters are developed with five regional constants and three regression coefficients. With periodicity and regional trends in parameters removed, the stationary stochastic components of monthly precipitation series are found to be

KEY WORDS: Regionalization of parameters, structure of area-time processes, monthly precipitation, periodicity in parameters, regional dependence.

ABSTRACT: The structural analysis of area-time hydrologic process of monthly precipitation specifies the deterministic components in form of periodic parameters and a stationary stochastic component. The parameters and coefficients of these components have regional trends.

Monthly precipitation series of two regions demonstrate the structural composition. Region I of North Dakota, South Dakota, and Minnesota with 41 monthly precipitation stations and Region II, of Nebraska with 29 stations are used as examples.

Mathematical models for periodicity and trends in parameters are developed with five regional constants and three regression coefficients. With periodicity and regional trends in parameters removed, the stationary stochastic components of monthly precipitation series are found to be

KEY WORDS: Regionalization of parameters, structure of area-time processes, monthly precipitation, periodicity in parameters, regional dependence.

ABSTRACT: The structural analysis of area-time hydrologic process of monthly precipitation specifies the deterministic components in form of periodic parameters and a stationary stochastic component. The parameters and coefficients of these components have regional trends.

Monthly precipitation series of two regions demonstrate the structural composition. Region I of North Dakota, South Dakota, and Minnesota with 41 monthly precipitation stations and Region II, of Nebraska with 29 stations are used as examples.

Mathematical models for periodicity and trends in parameters are developed with five regional constants and three regression coefficients. With periodicity and regional trends in parameters removed, the stationary stochastic components of monthly precipitation series are found to be

approximately time independent processes, with the identical three-parameter gamma probability distribution. These independent stochastic components are highly cross correlated. The lag-zero cross correlation coefficient is found to be primarily a function of the interstation distance. The developed methodology permits the generation of new samples consisting of a set of time series for a region, either at the observed station points or at any new grid of points.

approximately time independent processes, with the identical three-parameter gamma probability distribution. These independent stochastic components are highly cross correlated. The lag-zero cross correlation coefficient is found to be primarily a function of the interstation distance. The developed methodology permits the generation of new samples consisting of a set of time series for a region, either at the observed station points or at any new grid of points.

approximately time independent processes, with the identical three-parameter gamma probability distribution. These independent stochastic components are highly cross correlated. The lag-zero cross correlation coefficient is found to be primarily a function of the interstation distance. The developed methodology permits the generation of new samples consisting of a set of time series for a region, either at the observed station points or at any new grid of points.

approximately time independent processes, with the identical three-parameter gamma probability distribution. These independent stochastic components are highly cross correlated. The lag-zero cross correlation coefficient is found to be primarily a function of the interstation distance. The developed methodology permits the generation of new samples consisting of a set of time series for a region, either at the observed station points or at any new grid of points.