A MODEL OF STOCHASTIC STRUCTURE OF DAILY PRECIPITATION OVER AN AREA

by
Clarence W. Richardson

July 1977

HYDROLOGY PAPERS
COLORADO STATE UNIVERSITY
Fort Collins, Colorado

91

A MODEL OF STOCHASTIC STRUCTURE OF DAILY PRECIPITATION OVER AN AREA

by
Clarence W. Richardson*

HYDROLOGY PAPERS
COLORADO STATE UNIVERSITY
FORT COLLINS, COLORADO

July 1977 No. 91

*Agricultural Engineer, Southern Region, Agricultural Research Service, USDA, Temple, Texas 76501.

TABLE OF CONTENTS

ABSTRACT

Daily precipitation over an area is studied by modeling the stochastic structure of the time-area daily precipitation process. Daily precipitation at a point is envisioned as a continuous random variable that has been truncated at zero. The zero daily precipitation amounts are considered negative amounts of unknown quantity. The square roots of daily precipitation at a point for a selected study region approximate a sample from a truncated univariate normal distribution. The multivariate normal distribution is used to describe the time-area variation of daily precipitation over an area.

The means and standard deviations of the normalized precipitation are periodic within the year. A method was developed for obtaining maximum likelihood estimates of daily means and standard deviations from the truncated samples. The periodic components of the means and standard deviations were described with Fourier series. The Fourier coefficients were related to position within the study region. Sequences of random components were obtained for each station in the study region by removing the periodic means and standard deviations. The sequences of random components were normally distributed with zero means and unity standard deviations and were dependent in time and space. The lag-one autocorrelation coefficients were found to approximate a regional constant. The lag-zero cross-correlation coefficients were found to be a function of inter-station distance.

Precipitation sequences were generated for two areas in the study region using the truncated multivariate normal distribution model. Parameters of the model were defined using the latitude and longitude of each station. The new sequences closely resembled the observed sequences in (1) the periodic daily means and standard deviations (2) the lag-one autocorrelation coefficients, (3) the lag-zero cross-correlation coefficients, (4) the Markov chain wet-dry transition probabilities, and (5) the means, standard deviations, and skewness coefficients of 28-day and annual precipitation.

ACKNOWLEDGMENTS

I gratefully acknowledge the assistance given me by my graduate committee. The guidance, suggestions, support, and patience of Drs. V. Yevjevich and D. A. Woolhiser are sincerely appreciated. Without their experience and knowledge I could not have completed this study. I also thankfully acknowledge the valuable comments and suggestions by Drs. M. M. Siddiqui, D. B. McWhorter, and N. S. Grigg.

The support given me throughout my graduate program by the Agricultural Research Service is acknowledged. Special thanks are extended to Dr. Earl Burnett, Director of the Blackland Conservation Research Center, Temple, Texas, for his encouragement and patience. I also acknowledge the assistance of D. A. Wright for computer programming and preparation of graphic material and Brenda Lytle for typing the manuscript.

I am especially thankful to my wife, Paulette; and my children, Paula, Wade, and Keith for their patience, kindness, and understanding during difficult times of my graduate studies.

Clarence W. Richardson
U.S. Department of Agriculture
Agricultural Research Service
Blackland Conservation Research Center
Temple, TX 76501

FOREWORD

Hydrologic time processes have been classified for practical purposes as continuous and intermittent. Most climatologic and hydrologic time processes are continuous series, meaning that there is a non-zero value of that variable at any time. Instantaneous precipitation, evaporation, sediment transport in rivers, some runoff (usually on small rivers with negligible underground or surface water storage) represent the typical hydrologic intermittent time series. For some times the observed values are zeros; for other times values are greater than zero. Though there may be a continuous flux of water molecules through the liquid-gaseous or solid-gaseous inter-phases on the continental areas, with a difference in the number of molecules passing in two directions, the original concept of precipitation variable was designed in such a way that the process of instantaneous or short-interval precipitation is intermittent.

In practice, many intermittent processes, with positive series values for some time intervals and zero values for the other time intervals, are observed as totals for given time intervals, usually counted in minutes, hours, days, or a longer interval. Therefore, a sequence of intervals with values greater than zero is interchanged with intervals of zero values. This is the way many observed or computed time series have been processed and their data published. A large amount of available data of this type makes it necessary to design methods most feasible for their investigation and mathematical description that would permit the simulation of these intermittent series by the data generation methods.

Because of spatial interrelation for most of the climatological variables, the resulting hydrologic variables such as precipitation, evaporation, sediment transport, runoff of small rivers, and similar variables may all have intermittent series that are also spatially dependent. Solutions of practical water resources problems require data on time series either at a point or at a set of points. When a point series is studied independently of time series at the other points, methods are already available for the description of these intermittent series in the form of mathematical models and the estimation of their parameters. The classical approach to univariate (or point), intermittent time series is to first describe the process by such random events and their time process as the sequence of zero and non-zero intervals. The difficulty in this approach arises from the fact that nearly all parameters, especially the interval mean, standard deviation and autocorrelation coefficients (and sometimes the skewness and kurtosis coefficients), are or may be periodic. To avoid the difficulty of this combination of periodicities and intermittency, an approach to analysis starts by dividing the annual cycle into the seasons and the daily cycle into its parts, with an assumption that all the parameters are constants inside these intervals. This assumption requires the break of cycles into a relatively large number of seasons or parts, in order to justify it.

When the problem of generating new samples by using the Monte Carlo (experimental statistical method) is posed in hydrology and water resources, with the generated data to preserve both the time and space properties of random variables involved, this problem becomes that of a mathematical description and that of the generation of new samples in case of periodic-stochastic, intermittent time series. Both the periodicity in parameters, and the fact that the non-zero values occur at some space points while the zero values are observed simultaneously at the other points, create difficulties in generating new samples of multi-point intermittent time series. Attempts have been made to apply the combinatorial analysis and Markov chains in order to generate simultaneously the series of 2-3 stations, by generating first their zero and non-zero intervals, and then by preserving both the space and time dependences within the non-zero intervals. Researchers following this approach have been able to simulate only 2-3 station series. For more than four stations, the combinatorial approach becomes so complex that it is then difficult to extend it to cases of five, six, and more intermittent time series.

The generation of multivariate time series, which are periodic, intermittent and also stochastically dependent both in time and space, can be best accomplished by using the approach of the multivariate normal distribution and the principal component analysis. It seems logical to proceed in that direction also for variables which have asymmetric probability distributions and periodic-stochastic, intermittent time series. When a multivariable process is found to be periodic-stochastic, intermittent, non-normal stochastic process, difficulties arise both in mathematical description and in generation of new multivariate samples. When it becomes feasible to study intermittency by assuming it to be a truncated process of a non-intermittent time series, by removing periodicities in parameters, and by transforming the original variables or their residuals into the normal variables, then the principal component analysis for the generation of new samples becomes a feasible and very desirable approach.

The Ph.D. dissertation by Jerson Kelman, entitled "Stochastic Modeling of Intermittent Daily Hydrologic Series" (1976), and the Ph.D. dissertation by Clarence Wade Richardson, entitled "A Model of Stochastic Structure of Daily Precipitation over an Area" (1976), represent attempts to mathematically model the multiseries processes and to generate the new multivariate samples of periodic-stochastic, intermittent time series of daily precipitation as asymmetrically distributed random variable. As shown by the first dissertation, also the non-intermittent daily runoff series may be conceived as two intermittent processes, with variables transformed to normal distributions. Daily series are selected as typical examples of the short-interval time series. The basic approach is then in postulating that an intermittent time series with short time interval is only a truncated process of a non-intermittent, discrete time series. Basically, it is assumed that the probability distribution of non-zero values of an intermittent time series is only a tail, or a part of, either a truncated normal distribution, or a truncated other distribution, such as gamma, lognormal and similar. Therefore, techniques become needed for estimation of properties of a non-intermittent process from a periodic-stochastic, intermittent process. Techniques are further needed for the transformation of original variables or of their stochastic residuals in such a way that the periodic-stochastic, intermittent process of an asymmetric variable becomes only the truncated part of a normal distribution in case of the non-normal distribution of variables. The above two doctoral theses, one more tilted toward the theoretical and the other more toward the practical side, are the attempts to implement the above concepts by postulating the mathematical models and by estimating parameters of non-intermittent time series from the original, intermittent series. Once the properties of the non-intermittent discrete time series are estimated for each point of a multi-point set of series, it then becomes feasible to approximate closely by transformations their multivariate non-normal distribution by a multivariate normal distribution. From it then the periodic parameters can be estimated by fitting a set of harmonics in the Fourier analysis, and the periodic parameters appropriately removed from the series. The remaining stationary stochastic components may be either dependent or independent time processes. For a dependent process, linear dependence models can be inferred and their parameters estimated. This permits the computation of the independent identically distributed residuals, as the time independent stochastic components (TISC-variables). Once the series have been reduced to a set of normal, time independent, identically distributed stochastic processes, their spatial lag-zero correlation matrix enables a transformation of this set of series to their principal components, as a new set of space and time independent normal process. To generate the new samples of multi-point series, the normal independent samples are generated for each point and the reversed procedure applied on these time and space normal independent processes. Further transformations of reverse order produce the periodic-stochastic, non-intermittent process at each point. They preserve then the space dependence, periodicity and time dependence. By equating each negative value with zero, the multivariate, periodic-stochastic truncated (or intermittent) normal process is simulated by a set of new samples. Variables are then transformed from normal to the corresponding non-normal distribution.

The writer of this Foreword is convinced that the approach outlined above, and studied in this paper, for the generation of new samples by using the Monte Carlo (or statistical experimental) sample generation method is a feasible, practical method to model a set of periodic-stochastic, intermittent, time and space dependent series.

The other problem investigated by Dr. Jerson Kelman is the difference process applicable to the non-intermittent discrete time series, such as the non-intermittent daily runoff series. It is assumed that whenever the flow increases for a river the response of the river basin is different from its response during the river flow decrease. Therefore, the process could be divided into two separate but interconnected intermittent processes: the positive intermittent process as a difference process during the runoff increase, and a negative intermittent process as another difference process during the runoff decrease. The two difference processes, each considered as an intermittent process, are then combined to become a non-intermittent process.

Further research into the application of the above concept of considering the intermittent processes at a set of points along a line, over an area or across a space as the truncated processes of the periodic-stochastic, non-intermittent processes, is needed to sharpen the practical aspects of this method for the generation of new series.

February 1977

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

LIST OF SYMBOLS

Symbol	Definition
A	$m \times m$ matrix
A_j	Fourier coefficients for j-th harmonic
a	Parameter used in defining p_{max} and p_{min}
B	$m \times m$ matrix
B_j	Fourier coefficient for j-th harmonic
C_j	Amplitude of j-th harmonic
$C_j^2/2$	Variance of j-th harmonic
c	Parameters used in defining p_{max} and p_{min}
d	Distance between station i and station j, miles
e_i	Expected frequency
f_i	Observed frequency
g	Fisher's parameter
I_0	Integral of standard normal distribution from γ to ∞
L	Likelihood function
L_1	Station latitude, degrees north
L_2	Station longitude, degrees west
$L_1(i,j)$	Latitude of the midpoint between station i and station j, degrees north
$L_2(i,j)$	Longitude of the midpoint between station i and station j, degrees west
M_0	Lag-zero covariance matrix
M_1	Lag-one cross-covariance matrix
m	Number of harmonics, rank of $y_{p,\tau}$ for determining plotting position, or number of class intervals of a sample
m_τ	Sample mean for day τ
N	Sample size
n	Number of stations in precipitation network
n_0	Number of observations equal to or greater than the truncation point in a truncated sample
n_1	Number of unmeasured observations in a truncated sample
$P(D)$	Probability of a dry day
$P(D/D)$	Probability of a dry day given a dry day on the previous day
$P(W)$	Probability of a wet day
$P(W/W)$	Probability of a wet day given a wet day on the previous day
$P(y)$	Plotting position
p	Number of elements in a Markov chain
p_m	Percent of the variance of a parameter explained by the first m harmonics

LIST OF SYMBOLS (continued)

Symbol	Definition
p_{min}	Lower critical percent explained variance for determining significant harmonics
p_{max}	Upper critical percent explained variance for determining significant harmonics
r_k	Sample autocorrelation coefficient for lag k
r'_k	Sample autocorrelation coefficient of truncated series for lag k
s_τ	Sample standard deviation for day τ
s_ν^2	Estimate of the variance of the parameter ν
w	Observation translated to left truncation point
x_i	Element in a discrete series
$x_{p,\tau}$	Daily precipitation at a given station with p the year and τ the day within the year
$y_{p,\tau}$	Normalized daily precipitation at a given station with p the year and τ the day within the year
y_0	Left truncation point for a truncated sample
z	Random variable
α	Parameter of the gamma distribution
α_i	Regression coefficients
β	Parameter of the gamma distribution
γ	Standardized value for left truncation point
$\varepsilon_{p,\tau}$	Random component of a precipitation series with p the year and τ the day within the year
ε_0^τ	Truncation point of $\varepsilon_{p,\tau}$ for day τ
θ_j	Phase angle of the j-th harmonic
μ_τ	Population mean for day τ
ν	An arbitrary parameter
$\xi_{p,\tau}$	Independent random component of precipitation series
ρ_k	Population autocorrelation coefficient for lag k
ρ'_k	Population autocorrelation coefficient of a truncated series for lag k
$\rho_1^{(n)}$	Lag-one autocorrelation coefficient for n-day precipitation
$\rho_k(i,j)$	Lag-k cross-correlation coefficient between random variables i and j
σ_τ	Population standard deviation for day τ
χ^2	Chi-square fit parameter
ω	Length of the basic period for a periodic parameter

CHAPTER I
Introduction

Precipitation over an area is a physical process that occurs continuously in time. The precipitation process is basically a random (or stochastic) phenomenon in that, given the present state, the future behavior of the process can be predicted only on a probability basis. The precipitation process also contains periodic components due to the seasonal variation of precipitation within the year. The instantaneous rate of precipitation (precipitation intensity) is the most basic descriptor of the precipitation process. Precipitation intensity at all points on the earth's surface has some value, either zero or greater than zero, at all points in time. Precipitation intensity, therefore, may be described as a continuous time-area stochastic process.

The continuous time-area precipitation process cannot be measured directly. In measuring precipitation, the process is usually made discrete in space by sampling the process at selected points (stations). The precipitation process at a point is usually described by a discrete time series of precipitation amounts for some time interval, like an hour, day, month or year. Therefore, most precipitation data are discrete in both time and space. If precipitation over an area is required, the precipitation amounts for the desired time interval at given points are weighted with some weighting technique to give precipitation over the area for the time interval.

1. Scope and Objectives of the Study

This study is concerned with the development of a model of the periodic-stochastic structure of daily precipitation over an area or watershed. One day was chosen as the length of the discrete time interval because daily time series contain more information about the precipitation process than monthly or annual time series, and long records of daily precipitation are available at many locations. Daily precipitation is sufficient for many water resource projects.

The model developed in this study is intended to be used to generate daily precipitation samples at selected points within an area. The parameters of a time-area daily precipitation model should be a function of position within a region. Therefore, the primary objectives of this study are (1) to develop a model capable of being used to generate daily precipitation samples at n arbitrary points within a region with the same time-area characteristics as observed samples, and (2) to regionalize the model parameters within a region so that the model can be used anywhere within the region.

2. Advantages of a Daily Precipitation Model

The stochastic generation of large samples of daily precipitation at stations within an area cannot add information to that contained in a historical set of data. However, a model capable of being used to generate daily precipitation over an area offers several advantages over historical daily precipitation data. These advantages include the following:

1. The information contained in a historic set of data can be more completely extracted by examining a large number of time-area precipitation patterns that are as likely to occur in the future as the observed pattern.

2. If the parameters of the model are determined as a function of position within the region, new samples can be generated at points where no data have been recorded.

3. Generated data can be immediately available in computer-compatible form for analysis of water resource projects. Historic data often must be laboriously extracted from publications.

3. Approach to the Problem

The basic approach used in this study was (1) to develop a model capable of describing the periodic and stochastic characteristics of daily precipitation over an area, (2) to use daily precipitation data at multiple points within a region to infer the model parameters and (3) to test the model by generating long sequences of daily precipitation data for several stations in an area and comparing generated data to historic data.

Daily precipitation series for all stations in an area contain many zero values. Most precipitation models developed for daily or shorter time intervals have been restricted to a single station and utilize a Markov chain model for describing the probabilities of occurrence or nonoccurrence of precipitation (Smith and Schreiber, 1973; and Pattison, 1965). These models cannot easily be generalized to describe the probabilities of rainfall at multiple points because the number of states involved becomes large and estimation of the transition probabilities becomes difficult. For example, if a simple Markov chain is assumed to describe the probability of wet or dry intervals at a single station, the transition matrix contains only four probabilities: (1) wet given wet on the previous interval, (2) wet given dry, (3) dry given wet, and (4) dry given dry. If the Markov chain is expanded to include n stations with a wet or dry state at each station, the number of elements in the transition matrix is given by

$$p = 2^{2n} \qquad (1\text{-}1)$$

where p is the number of elements in the transition matrix, and n is the number of stations in the network. With four stations, p becomes 256. A matrix of this size is unmanageable, and accurate estimation of all the transition probabilities from historic data would be practically impossible.

To avoid the problem involved with Markov chain approaches, a model for generating daily rainfall at multiple stations should treat daily rainfall as a continuous variable. This implies the use of a multivariate distribution with continuous rainfall amounts at discrete points in time and space. The only multivariate distribution for which a multivariate generation technique has been well-developed is the multivariate normal distribution (Matalas, 1967). The model proposed here is, therefore, based on the multivariate normal distribution.

The assumption that the stochastic structure of daily precipitation over an area may be described by a multivariate normal distribution means that the daily precipitation at each point in the area must be normally distributed or be capable of being transformed to a normal distribution. The general approach used in this study is to describe the distribution of daily precipitation at each station with a univariate normal distribution. The integral of the normal distribution from $-\infty$ to 0 is the probability of zero daily precipitation and the remainder of the distribution describes the distribution of rainfall amounts for days of measurable rainfall.

The daily precipitation data at each station are reduced to approximately second-order stationary, normally distributed random variables. These random variables are dependent in both time and space. The cross correlation coefficients between pairs of these random variables are used to describe the dependence in space. The serial correlation coefficients of the sequence of random variables for each station are used to describe the dependence in time.

4. Probability Distributions of Point Rainfall Amounts

Many researchers have attempted to fit a distribution to precipitation data or have transformed the data to obtain a fit by a particular distribution. The first attempt to fit a continuous probability distribution to precipitation data frequency curves was made by Slade (1936). He used a logarithmic transformation of annual rainfall amounts and fitted a normal distribution to the results. Thom (1940) considered the frequency of annual precipitation but fitted smooth distribution curves to the data, rather than using a specific distribution function. A similar procedure was used by Beer et. al. (1946) for monthly rainfall amounts. Whitcomb (1940) fitted a Pearson Type III curve to monthly precipitation.

Several transformations that would make precipitation data normally distributed have been studied. Thom (1957) used a gamma distribution to fit storm amounts and then transformed the gamma distribution to a normal distribution. Stidd (1953) and Beals (1954) suggested that precipitation amounts raised to a fractional power are normally distributed. Stidd used a cube root transformation to transform annual, monthly and daily rainfall values to a normal distribution. Beals found that daily rainfall amounts to the one-fourth power were normally distributed. Franz (1970) followed the fractional power concept of Stidd and Beals and used a non-linear estimation technique to determine the magnitude of the normalizing fractional exponent for hourly precipitation data. The exponents were found to range from 0.23 to 0.52 for the stations studied by Franz.

Markovic (1965) studied the distributions of annual precipitation for several precipitation stations in the Western United States and Southwestern Canada. The normal, two-parameter log-normal, three-parameter log-normal, two-parameter gamma, and three-parameter gamma probability distribution functions were fit to the observed data for each precipitation station. The five functions were found to be applicable with a difference in the number of cases, which passed a Chi-square test in fitting the observed annual precipitation frequency distributions.

5. Time Persistence in Precipitation

Precipitation amounts for some interval of time are usually not independent of preceding values. Meteorological conditions at one time tend to carry over, or persist, into later times. The most common technique of determining the degree of persistence in a time series is by autocorrelation. The autocorrelation coefficients give a measure of the degree of linear association of values in a time series that are k units apart. The autocorrelation coefficient ρ_k for an infinite discrete series $\{x_i\}_{i=1}^{\infty}$ is defined by

$$\rho_k = \frac{\text{cov}(x_i,\, x_{i+k})}{\text{var}\ x_i}, \qquad (1\text{-}2)$$

and may be estimated from a sample series by

$$r_k = \frac{\frac{1}{N-k}\sum_{i=1}^{N-k} x_i x_{i+k} - \frac{1}{(N-k)^2}\left(\sum_{i=1}^{N-k} x_i\right)\left(\sum_{i=1}^{N-k} x_{i+k}\right)}{\left[\frac{1}{N-k}\sum_{i=1}^{N-k} x_i^2 - \frac{1}{(N-k)^2}\left(\sum_{i=1}^{N-k} x_i\right)^2\right]^{1/2}\left[\frac{1}{N-k}\sum_{i=1}^{N-k} x_{i+k}^2 - \frac{1}{(N-k)^2}\left(\sum_{i=1}^{N-k} x_{i+k}\right)^2\right]^{1/2}}. \qquad (1\text{-}3)$$

where r_k, k=1,2...,N is the k-th autocorrelation coefficient; N is the total number of values in the discrete time series; k is the units of lag, and x_i is the value of the variable for the i-th position in the time series (Yevjevich, 1972a).

Intuitively, the degree of persistance should decrease as the length of the time interval of a discrete precipitation time series increases. Yearly precipitation amounts have consistently been found to be independent. Kotz and Neumann (1959) cited studies by Yule (1945) in which serial correlations of annual rainfall in Great Britain were found to be nonsignificant. Brittain (1961) also found annual rainfall amounts to be independent series. Monthly precipitation amounts have been found to be independent in most cases (Pattison, 1965; Yevjevich and Karplus, 1973; and Namias, 1952). Precipitation amounts for time intervals of 1 day or less have consistently displayed persistence. Feyerherm and Bark (1965) found dependence in daily precipitation. Pattison (1965) and Franz (1970) found hourly rainfall to be dependent in time.

6. Univariate Rainfall Generation Models

There have been many attempts to develop methods of generating new sequences of rainfall at a point. Rainfall for short time intervals, like a day or an hour, has been difficult to model because of the sequential persistence between rainfall amounts and because the time series are dominated by zero values (intermittent process). The occurrence or nonoccurrence of rainfall for short intervals, like an hour or a day, have normally been described by Markov chains. Gabriel and Neumann (1962) seemed to have been the first to successfully describe the occurrence or nonoccurrence of daily rainfall with a Markov chain model. Additional evidence of the feasibility of using a Markov chain to describe the occurrence of sequences of wet or dry days was given by Caskey (1963), Weiss (1964), and Hopkins and

Robillard (1964). However the findings of Newnham (1916), Jorgensen (1949), and Cooke (1953) demonstrated that the Markov chain was not universally successful.

Smith and Schreiber (1973) tested the hypothesis of sequential independence (Bernoulli model) versus a first-order Markov chain hypothesis for the occurrence of wet or dry days during the summer rainy season in southeastern Arizona. The Markov chain model was found to be significantly superior to the Bernoulli model in reproducing the distributions of wet and dry run-lengths, occurrence of the first wet day in the season, number of runs per season, and the total number of rainfall days per season.

7. Multivariate Generation Models

Fiering (1964) introduced the use of multivariate techniques for generating new sequences of a hydrologic process at several stations. He assumed that annual streamflow at each site was normally distributed or could be rendered normal by a suitable transformation. By computing the eigenvectors of the correlation matrix, he transformed the observed data into sequences of principal components, with sequences uncorrelated and independent. A single station model was then used to generate the sequences of principal components.

Matalas (1967) pointed out that the Fiering model fails to yield new multivariate sequences that resemble the multivariate historic sequences in terms of the lag-one serial correlation coefficients of each station. Matalas then presented a technique for generating multivariate sequences that resemble the historic sequences in terms of the means, standard deviations, lag-one serial correlation coefficients, and lag-zero cross-correlation coefficients. The basic equation was

$$x_{i+1} = Ax_i + B\zeta_{i+1} , \tag{1-4}$$

where x_{i+1} and x_i are vectors whose values are the hydrologic variable minus the means for times i and i+1 for m stations; A and B are m × m matrices, whose elements must be defined from the historic data; and ζ_{i+1} is a vector m of random components with zero means and unit variances, whose elements are independent of x_i. Matalas (1967) showed that the A and B matrices are defined by

$$A = M_1 M_0^{-1} \tag{1-5}$$

and

$$BB^T = M_0 - M_1 M_0^{-1} M_1^T \tag{1-6}$$

where M_0 is the lag-zero covariance matrix of the historic data; M_1 is the lag-one cross covariance matrix of the historic data, and the superscripts -1 and T denote the inverse and transpose of the matrix, respectively. Equation (1-5) may be solved by straightforward matrix operations. The solution of equation (1-6) for matrix B is more involved. Matalas (1967) pointed out that the principal components analysis technique could be used to solve for B. However, Young (1968) described a simple and direct solution for B. Inherent in the Matalas model is the assumption that the hydrologic process is a sample from a multivariate normal distribution or may be reduced to a multivariate normal process.

Young and Pisano (1968) presented a modification of the Matalas model for generating multiple site monthly streamflow. Monthly streamflow data were made to conform to a normal distribution by using a logarithmic or square root transformation. The seasonal mean and standard deviations were removed using a technique given by Yevjevich (1966) to yield second-order stationary residuals. The Matalas model was then used to generate new residuals and the cyclic patterns in the means and standard deviations were added. The inverse of the normality transformation was used to produce the new multisite monthly streamflow.

Nicks (1974) developed a technique for generating daily rainfall at several raingages in a watershed. The occurrence or nonoccurrence of rainfall on each day at some gage on the watershed was generated using the Markov chain approach. When a wet day was generated, the location of the maximum rainfall amount within the area was determined, assuming the maximum amount could occur with equal probability at any gage within the network. The maximum daily rainfall amount was then generated using a skewed normal distribution. The spatial pattern of rainfall over the watershed was then generated using a deterministic-probabilistic model relating rainfall at a given point to the maximum rainfall amount. The Markov chain method of generating wet-dry sequences for a large area and the method of generating maximum daily amounts was found to be highly satisfactory. Improvements in the method of generating spatial patterns of rainfall were found to be needed.

8. Regionalization of Hydrologic Parameters

With the relatively recent advent of generation of hydrologic series, attention has been given to methods of regionalizing the parameters required to generate new sequences of hydrologic variables. Benson and Matalas (1967) stated that the two major deficiencies in generation of hydrologic sequences are: (1) large errors in estimating statistical parameters due to errors in the original sample and (2) new sequences could not be generated for ungaged locations. To overcome these shortcomings a method was proposed that would use statistical parameters derived from generalized relationships with hydrologic characteristics of a drainage basin, rather than the sample statistics determined from a historic series of data at a site.

Yevjevich and Karplus (1973) analyzed the structure of monthly precipitation over an area, based on the concept that the process is composed of deterministic components due to the seasonal nature of precipitation and a stationary stochastic component. The parameters were found to follow regional trends. Models describing the periodicity and regional trends in parameters were then developed. When the periodicity and regional trends in monthly precipitation were removed, the remaining stationary stochastic components were found to be approximately time independent and distributed according to the three-parameter gamma probability distribution function. The stochastic components were highly cross-correlated and the lag-zero cross correlation was a function of interstation distance. The results of the study showed that precipitation data at several points in a region have more information on all parameters of a given point series than each individual series. The regionalization of the parameters of the process and of the coefficients of the periodic parameters significantly reduced the number of coefficients to be estimated.

CHAPTER II
Structural Model for Daily Precipitation

The objective of this study, as stated in the introduction, is to develop a structural model capable of being used to generate daily precipitation at arbitrary locations within a watershed. Such a model would describe the time-space variation of daily precipitation over an area. This chapter presents mathematical models describing the structure of the area-time daily precipitation process and models describing the generation process.

The model developed in this study to describe daily precipitation over an area was based on a multivariate normal distribution. The assumption of a multivariate normal distribution means that the marginal distributions (rainfall at a point) must be normally distributed. However, the transformation of point rainfall to a normally distributed random variable, as stated by Franz (1970), does not insure that the precipitation at several points is multivariate normal, because normal marginal distributions are a necessary but not a sufficient condition for a multivariate normal distribution. In this study, however, it is assumed that, if precipitation at all points in an area are transformed so that each conforms to a univariate normal distribution, the precipitation at all points can be described by a multivariate normal distribution.

1. *Outline of the Time-Area Precipitation Model*

There are two basic alternatives that may be used in developing the structure of a model of daily precipitation over an area based on a multivariate normal distribution. The primary difference in the two alternatives is that with one approach the periodic means and standard deviations are removed before the data are transformed to a normal distribution and with the other approach the data are transformed before the periodic means and standard deviations are removed. The steps involved in the two alternatives are outlined below. These steps, for each alternative, can vary depending on the outcome of the analysis for a particular set of data.

a. Alternative I. The first alternative is to remove the periodic means and periodic standard deviations before transforming the data. The steps in this alternative include the following:

1. Calculate, for each day of the year, the mean and standard deviation of daily precipitation given the occurrence of a wet day.

2. Remove the periodic means and standard deviations from the original data, considering only the nonzero daily precipitation data.

3. Transform the nonzero values of the new sequences to approximate a sample from a truncated normal distribution.

4. Examine the time dependence (autocorrelation) of the transformed sequence for each station.

5. Examine the space dependence (cross correlation) between sequences for pairs of stations.

These steps may be simplified, depending on the results of the analysis. For example, if the ratios of the means and standard deviations calculated in step 1 are found to be a constant during the year, only the periodic standard deviations need to be removed in step 2 by dividing each nonzero daily precipitation value by the standard deviation for the given day. The resulting sequence would contain only zero or positive values. If both the means and standard deviations were removed by subtracting the means and dividing by the standard deviations, the resulting sequence would contain negative values for days with nonzero precipitation amounts smaller than the mean.

b. Alternative II. The second alternative is to apply a normalizing transformation before inferring and removing the periodic means and standard deviations. The steps in this alternative are as follows:

1. Transform the nonzero data so that the data for each day of the year approximates a sample from a truncated normal distribution.

2. Determine the mean and standard deviation for each day of the year using the transformed nonzero daily precipitation data.

3. Remove the periodic means and standard deviations from the transformed data, considering only the nonzero data.

4. Examine the time dependence of the stationary sequence for each station.

5. Examine the space dependence between stationary sequences for pairs of stations.

Both alternatives were examined to determine which alternative would give the best model for the time-area daily precipitation process. The two alternatives will be examined with actual precipitation data in Chapter III. (Alternative II proved to be the most desirable method of analysis.) In the following section, the concepts, procedures, and mathematics of the time-area daily precipitation model, using Alternative II, are given.

The procedure used to model daily precipitation over an area is outlined with block diagrams in Figures 2-1 and 2-2. Figure 2-1 illustrates the concepts involved in the time-area precipitation model and indicates the procedure used to analyze the precipitation data and evaluate the model parameters. Figure 2-2 illustrates the procedure used to generate new sequences of precipitation over an area. Each step in the procedure is described in detail in later sections. The entire process is briefly outlined here.

In analyzing the data and determining the model parameters, daily precipitation for stations within the region under study must be obtained (Figure 2-1).

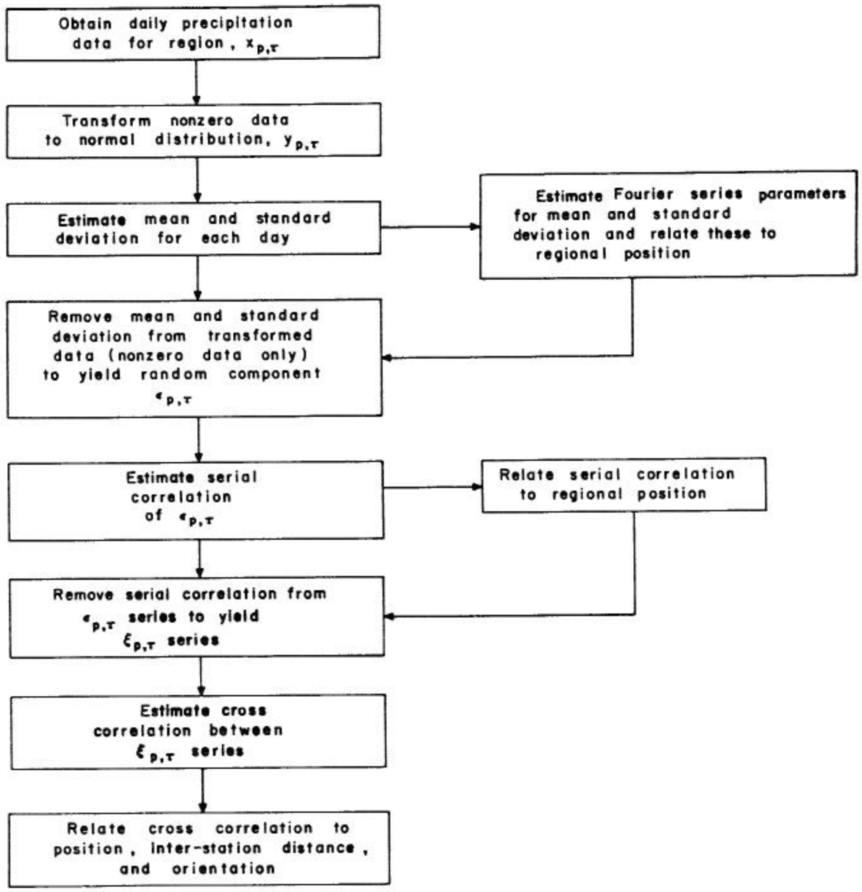

Figure 2-1. Diagram of precipitation data analysis procedure.

The sequence of daily precipitation amounts (zero and greater than zero) at a given station is considered as a sample from a continuous process containing both positive and negative values, but which has been truncated at zero so that all the negative values appear as zero. Since the model is based on a multivariate normal distribution, a normalizing transformation must be applied to the data for each station. The transformation is applied to the nonzero data only, since the zero values are assumed to be negative values of unknown magnitude. After the transformation, the data are assumed to be normally distributed, but with periodic means and standard deviations due to the seasonal nature of precipitation. The means and standard deviations are then determined for each day of the year at each station, the periodic component of the means and standard deviations are described with Fourier series, and the nonzero transformed data are standardized by removing the periodic means and standard deviations. The nonzero values of the new sequence for each station are then assumed to be a sample from a stationary, standard normal (mean of zero and variance of one) process. The serial correlation coefficients are determined for the stationary sequences for all stations, using only the nonzero values. The cross correlation between the stationary random components for the various stations are then determined. The model parameters (Fourier coefficients, serial correlation coefficients, and cross correlation coefficients) are related to position, inter-station distance, etc., so that the parameters can be determined for any arbitrary watershed position and precipitation station grid.

The generation procedure is the inverse of the data-analysis procedure and is illustrated in Figure 2-2. A hypothetical grid of precipitation stations is selected for the area where new precipitation sequences are desired. The model parameters are determined from the grid configuration and position in the region. Sequences of independent standard normal random numbers are generated for each station. Serial and cross correlation are added into the series by using equation (1-4) with the A and B matrices properly defined. The periodic mean and standard deviation are introduced into the sequence for each station. Each sequence is then truncated (negative values are set to zero) and the inverse of the normalizing transformation is applied to yield the new, generated precipitation series.

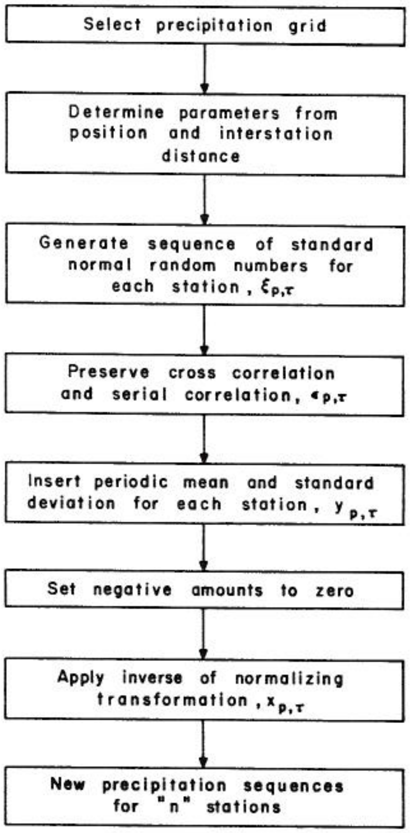

Figure 2-2. Diagram of procedure for generating new precipitation sequences for an area.

2. Mathematical Model of Daily Precipitation at a Point

Let the daily precipitation at a given station define the random variable $x_{p,\tau}$ with p the year and τ the day within the year. The $x_{p,\tau}$ series for most precipitation stations are dominated by zero amounts. Let $y_{p,\tau}$ be the daily precipitation series after application of a normalizing transformation. The transformation chosen for this study is the square root transformation. The reason for choosing the square root transformation will be illustrated in Chapter III by using the actual precipitation data. For the immediate purpose of developing the model structure, the square root transformation will be assumed adequate for transforming nonzero daily precipitation to an approximately normal distribution. The $y_{p,\tau}$ series is given by

$$y_{p,\tau} = (x_{p,\tau})^{1/2}. \qquad (2\text{-}1)$$

a. Approximation of daily precipitation with a truncated normal distribution. The probability density function of a continuous, normally-distributed random variable, z, is given by

$$f(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2} \qquad (2\text{-}2)$$

where μ is the expected value (or mean), and σ is the standard deviation of the distribution. The distribution of daily precipitation (either $x_{p,\tau}$ or $y_{p,\tau}$) is actually a mixed distribution containing both discrete and continuous variable values. For any given day, there is a finite probability of zero rainfall, while the distribution of rainfall amounts greater than zero must be described by a continuous probability density function. With the approach taken in this study, transformed daily precipitation, $y_{p,\tau}$, at a station is considered as a sample from a truncated normal distribution. The zero values may then be considered negative amounts of unknown quantity. The integral of the normal distribution from $-\infty$ to 0 is the probability of zero daily precipitation, and the remainder of the distribution describes the distribution of rainfall amounts for days with rainfall greater than zero. The concept is illustrated by the frequency function and observed frequency histogram shown in Figure 2-3. The area under the curve to the left of zero is the probability of the zero daily rainfall. Nonzero daily precipitation values less than 0.01 inch are recorded as traces and are treated as zeros in this study. The area under the curve between 0 and 0.1 (square root of 0.01) is the probability of a trace amount.

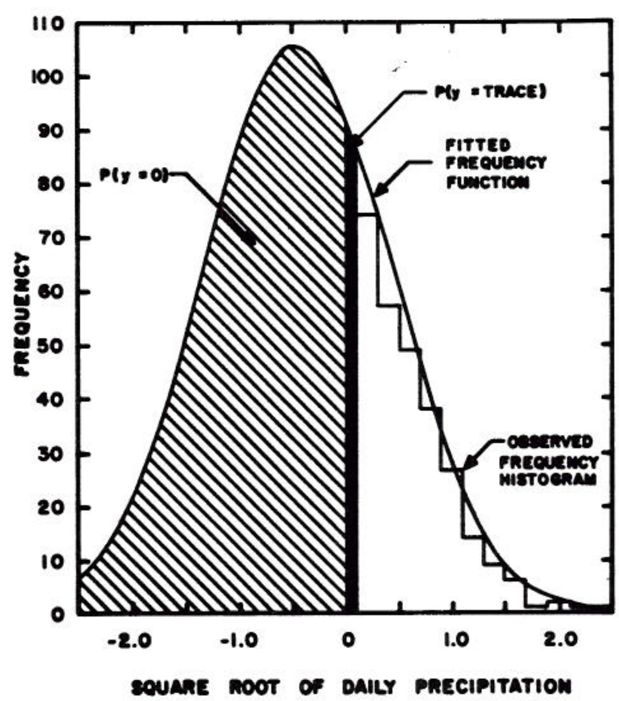

Figure 2-3. The truncated normal distribution of daily precipitation with a square root transformation.

b. Estimation of μ_τ *and* σ_τ. Daily precipitation at a given station exhibits a periodic component with a basic period of 1 year, as do practically all hydrologic time series with a time interval less than a year. This periodic nature causes the time series to be nonstationary. If the time series is to be treated as a stationary stochastic process, the periodic component must be detected and removed. Periodicity may be present in the mean, standard deviation, autocorrelation function, or higher order moments of a hydrologic time series. In this study, only the periodicity in the mean and standard deviation are considered.

Because of the periodicity of the mean and standard deviation, μ_τ and σ_τ for a given station must be estimated for each day, τ, of the year. The method of moments are normally used to estimate the mean and standard deviation of the normal distribution. However, the method of moments could not be used to estimate the μ_τ and σ_τ of the transformed daily precipitation data because the data were truncated at zero. A method given by Cohen (1950) for obtaining maximum likelihood estimates of the mean and variance of normal populations from truncated samples was adapted for estimating μ_τ and σ_τ. The method given by Cohen is summarized below for a singly truncated normal population with the number of measured and the number of unmeasured observations known.

Let y_0 designate the left truncation point, i.e., values less than y_0 cannot be measured and values equal to or greater than y_0 are measured. Let n_0 be the number of measured observations equal to or greater than y_0, and let n_1 be the number of unmeasured observations. For the specific case of daily precipitation at a point, y_0 is 0.10 (square root of 0.01 inch), n_0 is the number of days of measurable rainfall on a given day of the year, and n_1 is the number of days with zero rainfall on a given day. Translate the origin to the left truncation point by $w_i = y_i - y_0$ and let $\gamma = (y_0 - \mu)/\sigma$. The probability density function of w is given by

$$f(w) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\gamma + \frac{w}{\sigma})^2} . \qquad (2\text{-}3)$$

Define I_0 by

$$I_0 = \frac{1}{\sqrt{2\pi}} \int_\gamma^\infty e^{-t^2/2} dt. \qquad (2\text{-}4)$$

The likelihood function is given by

$$L = (1 - I_0)^{n_1} \prod_{i=1}^{n_0} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\gamma + \frac{w_i}{\sigma})^2} \qquad (2\text{-}5)$$

or

$$L = (1 - I_0)^{n_1} \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{n_0} e^{-\frac{1}{2}\sum_{i=1}^{n_0} (\gamma + \frac{w_i}{\sigma})^2} . \qquad (2\text{-}6)$$

Taking logarithms of equation (2-6) gives

$$\ln L = n_0 \ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2}\sum_{i=1}^{n_0} (\gamma + \frac{w_i}{\sigma})^2 + n_1 \ln(1 - I_0). \qquad (2\text{-}7)$$

Taking the partial derivatives of equation (2-7) with respect to γ and σ and equating to zero yield the maximum likelihood estimation equations

$$\left.\begin{aligned} \frac{\partial L}{\partial \gamma} &= n_1 \frac{\phi}{1 - I_0} - \sum_{i=1}^{n_0} (\gamma + \frac{w_i}{\sigma}) = 0 \\ \text{and} \quad \frac{\partial L}{\partial \sigma} &= -\frac{n_0}{\sigma} + \frac{1}{\sigma^2}\sum_{i=1}^{n_0} \left\{w_i(\gamma + \frac{w_i}{\sigma})\right\} = 0 \end{aligned}\right\} \qquad (2\text{-}8)$$

where

$$\phi = \frac{1}{\sqrt{2\pi}} e^{-\gamma^2/2}. \qquad (2\text{-}9)$$

Let

$$Y = \frac{n_1}{n_0} \frac{\phi}{(1 - I_0)} , \qquad (2\text{-}10)$$

then equations (2-8) may be written as

$$\left.\begin{aligned} &\sigma\left[Y - \gamma\right] - \eta_1 = 0 \\ \text{and} \quad &\sigma^2 - \sigma\gamma\eta_1 - \eta_2 = 0, \end{aligned}\right\} \qquad (2\text{-}11)$$

where η_1 and η_2 are the first and second moments about the truncation point y_0, or

$$\eta_k = \sum_{i=1}^{n_0} \frac{w_i^k}{n_0} . \qquad (2\text{-}12)$$

Eliminating σ from equations (2-11) yields

$$\frac{\eta_2}{\eta_1^2} = \frac{1}{Y - \gamma}\left(\frac{1}{Y - \gamma} - \gamma\right) . \qquad (2\text{-}13)$$

Equation (2-13) may be solved for γ by iterative techniques. The maximum likelihood estimate of σ may then be determined from either of the simultaneous equations given as equations (2-11). The estimate of μ is then given by

$$\mu = y_0 - \sigma\gamma. \qquad (2\text{-}14)$$

For daily precipitation at a point the maximum likelihood estimates of μ_τ and σ_τ for τ = 1,2, ..., 365 are determined by using n_0, n_1, η_1, and η_2 to solve equation (2-13) for γ. The value of σ_τ is then determined from equation (2-11), and μ_τ is determined from equation (2-14).

c. Determination of the stationary random component, $\varepsilon_{p,\tau}$. The stationary random component, $\varepsilon_{p,\tau}$, with the periodic mean and standard deviation removed is given by

$$\varepsilon_{p,\tau} = \frac{y_{p,\tau} - \mu_\tau}{\sigma_\tau}, \quad y_{p,\tau} \neq 0. \qquad (2\text{-}15)$$

The periodic movement of μ_τ and σ_τ may be described by using Fourier series representation. The periodic component of a statistic, ν_τ, with a basic period ω can be represented by

$$\nu_\tau = \nu_x + \sum_{j=1}^{m} (A_j \cos \frac{2\pi j\tau}{\omega} + B_j \sin \frac{2\pi j\tau}{\omega}) \qquad (2\text{-}16)$$

where ν_τ = the value of the parameter for the τ-th interval, ν_x = the mean of ω values of ν_τ, m = the number of harmonics, ω = the number of intervals in the basic period, A_j and B_j = the Fourier coefficients, and j is the harmonic index. The Fourier coefficients can be determined by

$$A_j = \frac{2}{\omega} \sum_{\tau=1}^{\omega} (\nu_\tau - \nu_x) \cos \frac{2\pi j\tau}{\omega} \qquad (2\text{-}17)$$

and

$$B_j = \frac{2}{\omega} \sum_{\tau=1}^{\omega} (\nu_\tau - \nu_x) \sin \frac{2\pi j\tau}{\omega}. \qquad (2\text{-}18)$$

It is often more convenient to express equation (2-16) in the form

$$\nu_\tau = \nu_x + \sum_{j=1}^{m} C_j \cos (\frac{2\pi j\tau}{\omega} + \theta_j) \qquad (2\text{-}19)$$

where

$$C_j^2 = A_j^2 + B_j^2 \text{ and } \theta_j = \text{arc tan} - \frac{B_j}{A_j}, \qquad (2\text{-}20)$$

while

$$A_j = C_j \cos \theta_j \text{ and } B_j = -C_j \sin \theta_j. \qquad (2\text{-}21)$$

The C_j values are the amplitudes of the various harmonics, and the θ_j values are the phase angles of the harmonics.

For daily precipitation at a station with ω = 365, there are 365 values of any periodic parameter, ν_τ. The maximum number of harmonics that may be used to describe the periodic movement of ν_τ is $(\omega - 1)/2$ or 182. However, the seasonal change in μ_τ or σ_τ is relatively slow for most daily precipitation series. Normally only a few harmonics are sufficient to describe the periodic movement of these parameters. The inclusion of too many harmonics to describe μ_τ or σ_τ only serves to perpetuate the sampling errors inherent in estimating μ_τ or σ_τ. Yevjevich and Karplus (1973) found that only one harmonic was required to describe the periodic movement of the mean and the standard deviation of monthly precipitation for stations in the central part of the United States. Further, they found that the same phase angle, θ_j, could be used to describe the phase of the harmonic of the mean or the standard deviation for a given station.

Equation (2-15) can be applied only to the nonzero values of the $y_{p,\tau}$ series because the zero amounts are assumed to be negative quantities of unknown magnitude. When equation (2-15) is applied to all positive values in a series, with μ_τ and σ_τ given by the Fourier series representation, the result is an $\varepsilon_{p,\tau}$ series with zero values that are unchanged from the $y_{p,\tau}$ series and positive values that are stationary in the mean and standard deviation. The entire $\varepsilon_{p,\tau}$ series, including the negative unknown amounts represented by the zeros, is assumed to be stationary in the mean and standard deviation with a mean of zero and a standard deviation of unity. The new variable $\varepsilon_{p,\tau}$ will, in general, be dependent in time, and there may be periodicities in the autocorrelation coefficients and in the higher-order moments.

d. Determination of the serial dependence of $\varepsilon_{p,\tau}$. Assuming that the serial dependence of $\varepsilon_{p,\tau}$ can be described by a linear autoregressive model, $\varepsilon_{p,\tau}$ is then given by

$$\varepsilon_{p,\tau} = \sum_{k=1}^{m} \alpha_{k,\tau} \varepsilon_{p,\tau-k} + \xi_{p,\tau} \qquad (2\text{-}22)$$

where $\alpha_{k,\tau}$ = the regression coefficients for lag k and $\xi_{p,\tau}$ = the independent stochastic component. Assuming a first-order autoregressive model, equation (2-22) reduces for $\alpha_{1,\tau} = \rho_{1,\tau}$ to

$$\varepsilon_{p,\tau} = \rho_{1,\tau} \varepsilon_{p,\tau-1} + \xi_{p,\tau}. \qquad (2\text{-}23)$$

If $\rho_{1,\tau}$ is periodic, the periodic movement can be described by Fourier series, as shown above. However, Yevjevich (1972b) stated, "Precipitation discrete series with time intervals as fractions of the year show clearly that their nonstationarity basically results from the periodicity in the mean and standard deviation . . ." For the further development of the model, $\rho_{1,\tau}$ will be assumed constant throughout the year. The periodicity of $\rho_{1,\tau}$ will be investigated in Chapter III.

The autocorrelation of zero amounts within the $\varepsilon_{p,\tau}$ series is meaningless. If the autocorrelation coefficients are computed for the $\varepsilon_{p,\tau}$ series, considering only the cases when both $\varepsilon_{p,\tau}$ and $\varepsilon_{p,\tau+k}$ are nonzero, the autocorrelation coefficients of the stochastic component of daily precipitation, $\varepsilon_{p,\tau}$, may be determined.

3. Model of the Dependence in Space of Daily Precipitation

The $\xi_{p,\tau}$ series for stations in a region are independent in sequence but dependent in space (or cross correlated). The linear cross correlation coefficient between the $\xi_{p,\tau}$ series at different stations may be used to express the degree of linear association between the series. Considering only the lag-zero cross correlation, the linear space dependence may be expressed by

$$\xi_{p,\tau}(i) = \rho_0(i,j)\xi_{p,\tau}(j) + \zeta_{p,\tau}(i), \qquad (2\text{-}24)$$

where $\xi_{p,\tau}(i)$ and $\xi_{p,\tau}(j)$ are the time-independent stationary stochastic components of equation (2-23) for stations i and j, $\rho_0(i,j)$ is the lag-zero cross correlation coefficient, and $\zeta_{p,\tau}(i)$ is a random component independent of $\xi_{p,\tau}(j)$. Yevjevich and Karplus (1973) pointed out that, in general, the cross correlation coefficient between the $\xi_{p,\tau}$ series of two stations is a function of position of one of the stations, the inter-station distance, and the orientation of the line connecting the two stations. It was shown that the effects of position and orientation were small as compared with that of the inter-station distance.

4. The Multivariate Generation Model

The multivariate generation procedure used here is a modification of that given by Matalas (1967). The matrix equations used by Matalas for generating multivariate data and for defining the matrices used in the generation procedure were given in equations (1-4), (1-5), and (1-6). The generation equation, as used in this study, may be written as

$$\varepsilon_{i+1} = A\varepsilon_i + B\zeta_{i+1}, \qquad (2\text{-}25)$$

where ζ_{i+1} is a vector of m random components; ε_{i+1} and ε_i are vectors whose values are the generated hydrologic series for m stations with the means removed; and A and B are m × m matrices, whose elements are defined in such a way that the new multivariate sequences preserve means, standard deviations, skewnesses, lag-one serial correlation coefficients, and lag-zero cross correlation coefficients of the population inferred from the historic multivariate sequences. The ε_i series generated with equation (2-25) are both serially correlated and cross correlated and correspond to the $\varepsilon_{p,\tau}$ series given by equation (2-23). The A and B matrices are determined from M_0 and M_1 as seen from equations (1-5) and (1-6). The M_0 and M_1 matrices, as defined by Matalas (1967), may be written

$$M_0 = \begin{bmatrix} \sigma_1^2 & \rho_0(1,2)\sigma_1\sigma_2 & \cdots & \rho_0(1,n)\sigma_1\sigma_n \\ \rho_0(2,1)\sigma_2\sigma_1 & \sigma_2^2 & \cdots & \rho_0(2,n)\sigma_2\sigma_n \\ \cdot & \cdot & \cdots & \cdot \\ \rho_0(n,1)\sigma_n\sigma_1 & \rho_0(n,2)\sigma_n\sigma_2 & \cdots & \sigma_n^2 \end{bmatrix} \qquad (2\text{-}26)$$

and

$$M_1 = \begin{bmatrix} \rho_1(1)\sigma_1^2 & \rho_1(1,2)\sigma_1\sigma_2 & \cdots & \rho_1(1,n)\sigma_1\sigma_n \\ \rho_1(2,1)\sigma_2\sigma_1 & \rho_1(2)\sigma_2^2 & \cdots & \rho_1(2,n)\sigma_2\sigma_n \\ \cdot & \cdot & \cdots & \cdot \\ \rho_1(n,1)\sigma_n\sigma_1 & \rho_1(n,2)\sigma_n\sigma_2 & \cdots & \rho_1(n)\sigma_n^2 \end{bmatrix} \qquad (2\text{-}27)$$

where $\rho_1(i)$ is the lag-one serial correlation coefficient of the $\varepsilon_{p,\tau}$ series for station i, $\rho_0(i,j)$ is the lag-zero cross correlation coefficient between $\xi_{p,\tau}$ series for stations i and j, and $\rho_1(i,j)$ is the lag-one cross correlation coefficient between $\xi_{p,\tau}$ series with the series for station j lagged one unit relative to the series for station i.

By mathematical definition, M_1 contains the lag-one cross-correlation coefficients. The values of $\rho_1(i,j)$ and $\rho_1(j,i)$ are not the same. Use of cross-correlation coefficients with lags greater or smaller than zero implies that one random variable precedes the other random variable in time. For daily precipitation, significant lag cross-correlation coefficients may occur as a result of frontal-type precipitation that deposits precipitation at one station before another station. However, the lag between the precipitation at two stations depends on the distance between the stations and the rate of movement of the precipitation front. Conceivably, the cross-correlation coefficient for lag-two (or some other lag) may be greater than that for lag-one. From a physical point of view, calculating and attempting to preserve the lag-one cross-correlation coefficient without considering other lags (both positive and negative) is an unsound approach.

The model was developed to preserve the lag-zero cross-correlation coefficients and the lag-one serial correlation coefficients of daily precipitation over an area. There is no physical reason to attempt to preserve the lag-one cross-correlation coefficients. However, use of the multivariate generation equation, given in equation (2-25), requires that some estimate of $\rho_1(i,j)$ be made. Matalas (1967) stated that, assuming a lag-one Markov process, the lag-one cross-correlation coefficient is the product of the lag-zero cross-correlation coefficient and the lag-one serial correlation coefficient of the variate, whose events occur at time i+1, or

$$\rho_1(i,j) = \rho_0(i,j) \cdot \rho_1(i). \qquad (2\text{-}28)$$

The appropriateness of equation (2-28) for defining $\rho_1(i,j)$ for daily precipitation is questionable. However, equation (2-28) was used to define the $\rho_1(i,j)$ values contained in the off-diagonal elements of M_1. The assumption of the relationship in equation (2-28) for defining $\rho_1(i,j)$ was made to permit the multivariate generation equation to be used to generate new sequences that preserve the lag-zero cross correlation coefficients and the lag-one serial correlation coefficients inferred from historic sequences, and does not imply that equation (2-28) is a valid description of the lag-one cross correlation coefficients for daily precipitation.

In this study, the daily precipitation series, $x_{p,\tau}$, have been normalized, using an appropriate transformation to give the $y_{p,\tau}$ series. The $y_{p,\tau}$ series have been standardized by removing the periodic means and periodic standard deviations, equation (2-15), resulting in $\varepsilon_{p,\tau}$ series that are assumed to be second-order stationary, with a mean of zero and a standard deviation of unity. Therefore, all the σ_i terms in equations (2-26) and (2-27) becomes unities. Using the fact that, for the $\xi_{p,\tau}$ series, $\rho_0(i,j) = \rho_0(j,i)$, the M_0 matrix reduces to the symmetric matrix

$$M_0 = \begin{bmatrix} 1 & \rho_0(1,2) & \cdots & \rho_0(1,n) \\ \rho_0(1,2) & 1 & \cdots & \rho_0(2,n) \\ \cdot & \cdot & \cdots & \cdot \\ \rho_0(1,n) & \rho_0(2,n) & \cdots & 1 \end{bmatrix} . \qquad (2\text{-}29)$$

Using equation (2-28), M_1 reduces to

$$M_1 = \begin{bmatrix} \rho_1(1) & \rho_0(1,2)\rho_1(1) & \cdots & \rho_0(1,n)\rho_1(1) \\ \rho_0(1,2)\rho_1(2) & \rho_1(2) & \cdots & \rho_0(2,n)\rho_1(2) \\ \cdot & \cdot & \cdots & \\ \rho_0(1,n)\rho_1(n) & \rho_0(2,n)\rho_1(n) & \cdots & \rho_1(n) \end{bmatrix} \qquad (2\text{-}30)$$

With these simplifications, M_0 is simply the lag-zero cross-correlation matrix of the $\xi_{p,\tau}$ series, and is symmetric with each element of the principal diagonal equal to unity. M_1 contains the lag-one serial correlations on the diagonal and the off-diagonal elements are the product of the lag-one serial correlations and the lag-zero cross correlations.

a. Relationship between ρ for the total distribution and ρ' for the truncated distribution. The A and B matrices may be determined from M_0 and M_1, using equations (1-5) and (1-6). The elements of M_0 and M_1 are defined from sample estimates of the lag-one serial correlation coefficients and the lag-zero cross-correlation coefficients. The estimates of the serial correlation and cross-correlation coefficients must be determined from the $\varepsilon_{p,\tau}$ and the $\xi_{p,\tau}$ series that are each approximately normally distributed with a mean of zero and a standard deviation of unity and have been truncated so that values less than the truncation point are unknown.

Both the lag-one serial correlation coefficients and the lag-zero cross-correlation coefficients are equivalent to the product moment correlation coefficient of a bivariate normal distribution. The sample estimates of the correlation coefficients must be obtained using only the data above the truncation point (nonzero daily precipitation). The estimate of the correlation coefficient of a bivariate normal obtained from a truncated distribution is less than that obtained from the total distribution. This point is illustrated graphically in Figure 2-4. In this example, a sample of 100 random variables from a bivariate standard normal distribution were generated. The correlation coefficient used in the generation procedure was 0.60. Let x_i and y_i represent the random variables generated using the bivariate normal distribution. The x_i and y_i values were plotted in Figure 2-4. The correlation coefficient of both the total sample and the sample with both distributions truncated at zero (both x_i and y_i nonzero) were calculated. The correlation coefficient for the 100 points of the total sample was 0.57. The correlation coefficient for the 28 points in the truncated sample was only 0.35.

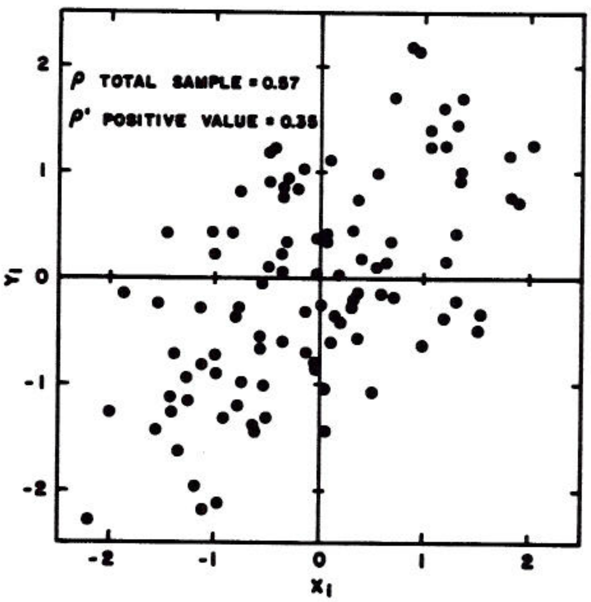

Figure 2-4. Comparison of the lag-one serial correlation coefficient for the total sample and for the positive values only for a generated N(0,1) process.

Regier and Hamden (1971) derived an expression relating the correlation, ρ', of a bivariate standard normal distribution that has been truncated at a given point, a, to the correlation, ρ, of the total, untruncated distribution. The results apply to both the lag-one serial correlation and lag-zero cross correlation of daily precipitation when the truncated normal distribution approach, which was described earlier, is used. The results were given by Regier and Hamden (1971) in tabular form and are plotted in Figure 2-5 for the range of truncation points that are applicable for daily precipitation in this study. The truncation point of the $\varepsilon_{p,\tau}$ and the $\xi_{p,\tau}$ series depends on the normalizing transformation that is used and the values of μ_τ and σ_τ, equation (2-15). Since μ_τ and σ_τ change throughout the year, the truncation point for a given series also changes during the year.

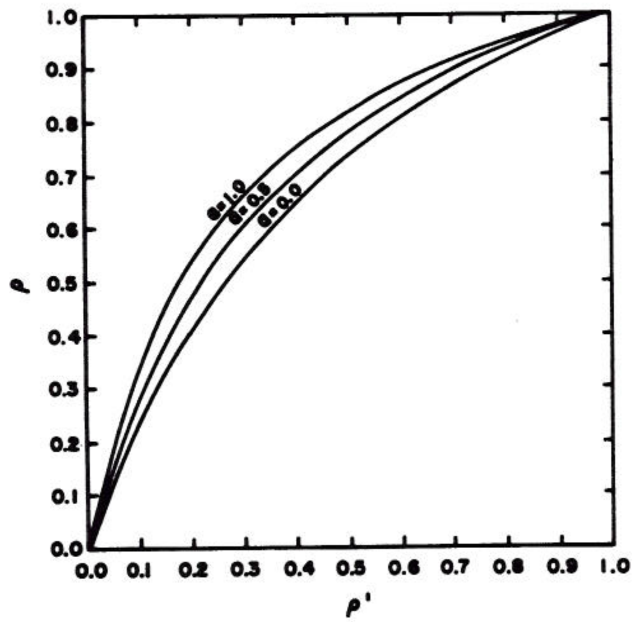

Figure 2-5. Relationships between the correlation, ρ', of a bivariate standard normal distribution that has been truncated at a given point, a, to the correlation, ρ, of the total, untruncated distribution.

In Chapter III, the estimates of the lag-one serial correlation coefficients for the total distribution, ρ, are obtained by applying the relationship given in Figure 2-5 to the ρ' values calculated from the truncated samples. The truncation point was calculated for each day of the year and each station. The average truncation point for each station was determined by weighting the truncation point for each day of the year by the number of nonzero observations for that day. The average truncation point for each station was used to convert ρ' to ρ. Similarly, the estimates of the lag-zero cross-correlation coefficients were obtained by applying the relationship in Figure 2-5 and the average truncation point for the two series to the cross-correlation coefficient calculated from the two truncated $\xi_{p,\tau}$ series.

b. Determination of the new $x_{p,\tau}$ *series.* After the $\varepsilon_{p,\tau}$ series are generated, the periodic means and periodic standard deviations are inserted into the generated sequence for each station to yield the $y_{p,\tau}$ series by

$$y_{p,\tau} = \sigma_\tau \varepsilon_{p,\tau} + \mu_\tau, \qquad (2\text{-}31)$$

where the μ_τ and σ_τ are given in the Fourier series form. The new $y_{p,\tau}$ series obtained at this point contain both positive and negative values. The negative values are assumed to represent days without measurable precipitation and are set to zero. The daily precipitation series are determined by employing the inverse of the normalizing transformation, namely

$$x_{p,\tau} = (y_{p,\tau})^2. \qquad (2\text{-}32)$$

Values of $x_{p,\tau}$ less than 0.01 are set to zero.

Using the procedure described, new $x_{p,\tau}$ sequences are generated that preserve the periodic means and periodic standard deviations, the lag-one serial correlation coefficients, and the lag-zero cross-correlation coefficients that are obtained from historic $x_{p,\tau}$ sequences. To be of value, the new $x_{p,\tau}$ sequences must also duplicate, within statistical limits, the mean and standard deviation of annual precipitation for each station, the probability of zero daily precipitation, and the other parameters that are inferred from the historic data.

CHAPTER III
Evaluation of the Daily Precipitation Model for a Selected Region

In this chapter, the parameters of the time-area daily precipitation process are evaluated for a specific region by using the model described in the previous chapter, and whether the model offers a valid description of the time-space variation of daily precipitation for the selected region is determined.

1. *The Study Region*

A region in central Texas was chosen to be used in evaluating the proposed model for generating daily precipitation at several stations in an area. The study region (Figure 3-1) lies between 31 degrees and 33 degrees North latitude and 95 and 99 degrees West longitude. Mean annual precipitation in the study region increases gradually from about 25 inches along the west side, to about 45 inches near the east side. There are no abrupt topographic features within the region to cause major changes in precipitation patterns. The elevation changes gradually from about 300 feet above mean sea level in the southeast portion of the region to about 1400 feet along the western edge.

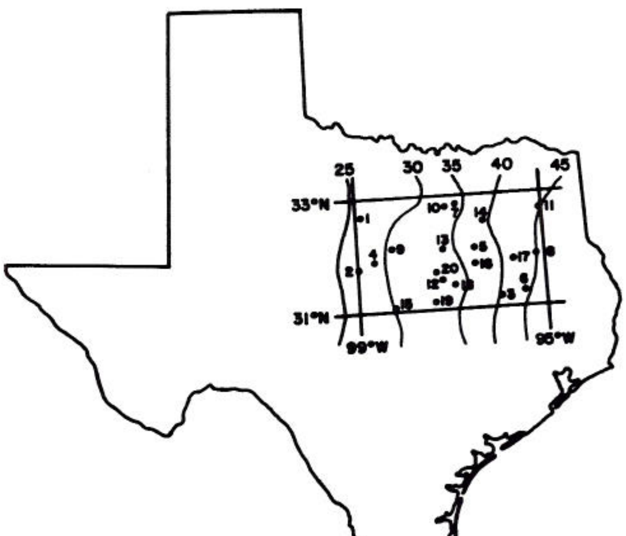

Figure 3-1. Location and station number of precipitation stations within the study region and the mean annual precipitation over the region.

The variation of precipitation within the year is complex over the region. The mean monthly precipitation for two stations is shown in Figure 3-2. The mean monthly precipitation pattern for Gilmer (station 1), located near the eastern edge of the study region, shows a peak of 5.50 inches in April and a smaller peak of 4.42 inches in December. At Brownwood (station 2), near the western edge of the region, all of the monthly means are smaller than those at Gilmer. The largest mean monthly precipitation at Brownwood is 4.50 inches in May. A secondary peak of 2.77 inches occurs in September. At both stations the smallest mean monthly precipitation occurs in August. The seasonal patterns of precipitation for these two stations are typical of that for the entire region. The precipitation within the region is characterized by a wet period in the late spring, followed by a dry period in mid to late summer. A second peak, of smaller magnitude than the first, occurs in the fall. The mean annual and mean monthly precipitation, in general, decrease from east to west.

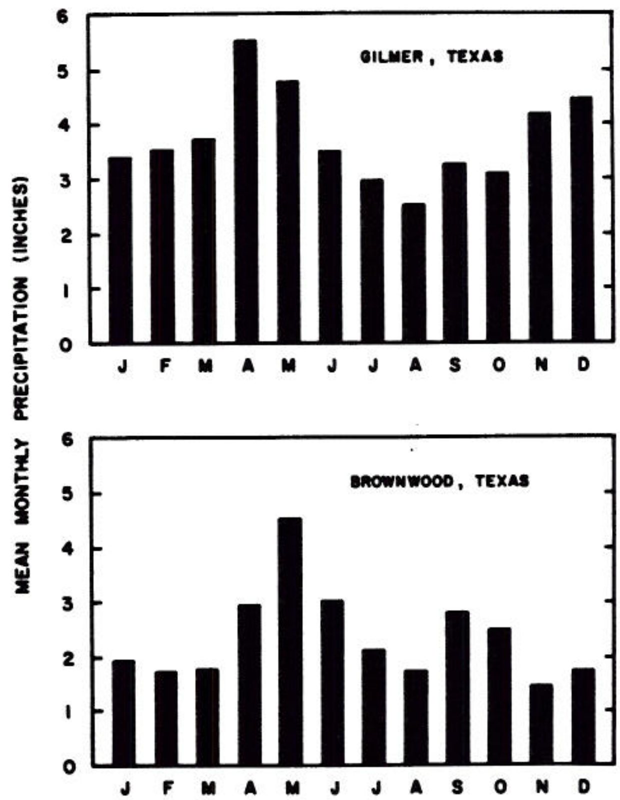

Figure 3-2. Mean monthly precipitation for Gilmer, near the eastern edge of the study area, and Brownwood, near the western edge of the study area.

Twenty precipitation stations within the study region were chosen for use in this investigation. Daily precipitation data for each station for the 40 years (1933-1972) were assembled and used in evaluating the model parameters. Data on each station are given in Table 3-1.

Table 3-1. Data on precipitation stations used in this study.

No.	Station Name	Index No.[1/]	Latitude (° North)	Longitude (° West)	Elevation (feet)
1	Breckenridge	1042	32.77	98.90	1185
2	Brownwood	1138	31.72	98.98	1435
3	Centerville	1596	31.27	95.98	330
4	Comanche	1914	31.90	98.60	1345
5	Corsicana	2019	32.08	96.46	425
6	Crockett	2114	31.30	95.45	347
7	Dallas	2244	32.85	96.85	481
8	Dialville	2444	31.87	95.27	620
9	Dublin	2598	32.10	98.33	1502
10	Fort Worth	3283	32.83	97.05	537
11	Gilmer	3546	32.73	94.98	390
12	Hewitt	4122	31.45	97.18	642
13	Hillsboro	4182	32.02	97.12	550
14	Kaufman	4705	32.58	96.32	438
15	Lampasas	5018	31.05	98.18	1024
16	Mexia	5869	31.68	96.48	529
17	Palestine	6757	31.78	95.65	600
18	Riesel	2/	31.45	96.88	560
19	Temple	8910	31.10	97.35	700
20	Waco	9419	31.62	97.22	500

1/ U.S. Weather Bureau index number.
2/ Operated by USDA Agricultural Research Service.

2. *Evaluation of Normalizing Transformations*

For the multivariate normal distribution model that was described in Chapter II to describe the time-area variation of daily precipitation, the precipitation at a given station must approximate a sample from a truncated univariate normal distribution. The general transformation that was investigated in this study for transforming daily precipitation to approximate a sample from a normal distribution was $y_{p,\tau} = x_{p,\tau}^{\alpha}$ where α is a parameter to be estimated. Values of α will generally be periodic within the year and will vary across the region. In this study, a transformation was desired that could be used for all periods of the year and all points in the region. Attention was centered on three simple values of α to determine if one of these values could be used to transform the daily precipitation data to approximate a sample from a normal distribution. The three transformations that were investigated were: (1) $y_{p,\tau} = x_{p,\tau}$ (no transformation), (2) $y_{p,\tau} = x_{p,\tau}^{1/2}$ (square root transformation), and (3) $y_{p,\tau} = x_{p,\tau}^{1/3}$ (cube root transformation).

Data for eight of the twenty stations within the study region were chosen to evaluate the three transformations. The eight stations were chosen so that all areas of the study region were represented. The location of the eight stations within the study region is shown in Figure 3-3.

For each precipitation station, a different estimate of the mean and standard deviation could be made for each day of the year. However, the μ_τ and σ_τ values for individual days are subject to large

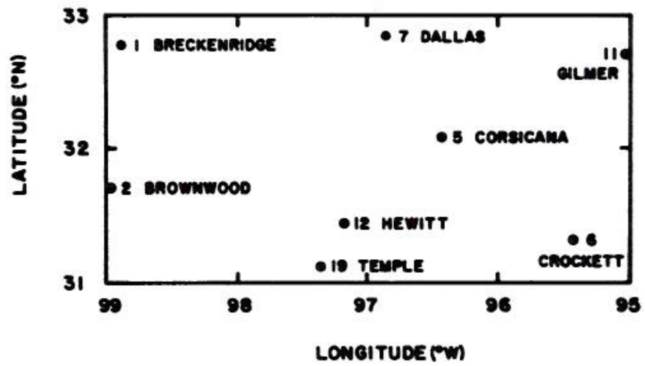

Figure 3-3. Location of eight test stations within the study region.

sampling errors due to relatively small sample sizes. For comparing the three cases, the year was divided into 13 periods of 28 days each. Since μ_τ and σ_τ change little from one day to the next, μ_τ and σ_τ were assumed constant for the 28 days of a given period. By grouping the data into 28-day periods, the sample sizes were larger and less sampling error was involved in the estimation of the mean and standard deviation for each period. Estimating μ_τ and σ_τ by 28-day periods caused an abrupt change in the two parameters from the last day of a given period to the first day of the next period. Such an abrupt change was clearly unreasonable. However, the approach of dividing the year into 28-day periods was used only for comparing the three transformations. Later in this chapter, after the appropriate transformation has been determined, μ_τ and σ_τ are estimated for each day of the year.

Maximum likelihood estimates, m_τ and s_τ, of the means and standard deviations were determined for each of the three cases using the truncated normal technique described in Chapter II. For each transformation, m_τ and s_τ were computed for each of the 13 periods for the eight test stations. A chi-square goodness-of-fit parameter was calculated for each period and each station to determine how well the distribution of the data with a given transformation approximates a normal distribution. The observed data were sorted into classes. The days with zero precipitation were included in one class, and the days with nonzero precipitation were sorted among nine other classes, depending on the amount of precipitation. The nine classes for days with nonzero precipitation represented about equal probabilities. The chi-square parameter was calculated by

$$\chi^2 = \sum_{i=1}^{m} \frac{(f_i - e_i)^2}{e_i}, \qquad (3\text{-}1)$$

where χ^2 = the chi-square parameter, f_i = the observed frequencies (number of observations in the class interval i), e_i = the expected frequencies, and m = the number of class intervals. The χ^2 computed with equation (3-1) was compared to a critical chi-square value, χ_0^2, to test the hypothesis that the sample was from a normal population. The χ^2 parameter will have a chi-square distribution with k-1 degrees of freedom, if the population parameters were not estimated from sample observations. If the population parameters were estimated from observations, the number of degrees of freedom is decreased by the number of parameters that were estimated. In this study, two parameters, μ_τ and σ_τ, were estimated. With 10 intervals and two estimated parameters the number of degrees of freedom was 7. The χ_0^2 value with 7 degrees of freedom and a 0.95 level of significance is 14.1. The hypothesis that the sample was from a normal population would be rejected if the χ^2 parameter for a particular sample and a given transformation was greater than 14.1. The hypothesis would be accepted if χ^2 was less than 14.1.

The values of the chi-square fit parameter are shown for each of the 13 periods and for three stations in Table 3-2. The results for the Crockett station showed that the square root transformation gave $\chi^2 < \chi_0^2$ for 11 of the 13 periods, and the cube root transformation gave $\chi^2 < \chi_0^2$ for only 2 periods. For the Hewitt data, the square root transformation resulted in $\chi^2 < \chi_0^2$ for 11 periods and the cube root transformation gave $\chi^2 < \chi_0^2$ for 8 periods. The Breckenridge results showed that the square root transformation resulted in $\chi^2 < \chi_0^2$ for all 13 periods, while the cube root transformation produced $\chi^2 < \chi_0^2$ for 3 periods. The χ^2 for no transformation (y = x) was much greater than χ_0^2 for each period and each station shown in Table 3-2.

The average chi-square values for the 13 periods are shown for each of the eight test stations and the three transformations in Table 3-3. The square root transformation resulted in $\chi^2 < \chi_0^2$ for six of the eight stations. The cube root transformation gave an average $\chi^2 < \chi_0^2$ for only one of the eight stations. The average χ^2 for no transformation was much greater for all stations than χ_0^2.

Typical cumulative probability distribution curves for data with the three transformations are shown in Figures 3-4, 3-5, and 3-6. The daily precipitation data for the third 28-day period of the year for the Gilmer station were used for all three figures. The $y_{p,\tau}$ data for each transformation were arranged in ascending order, and the empirical cumulative probabilities, or plotting positions, were computed by

$$P(y) = \frac{m}{N + 1}, \qquad (3\text{-}2)$$

where P(y) is the plotting position; m is the rank of the $y_{p,\tau}$ value; and N is the total number of days. For the data shown in Figures 3-4, 3-5, and 3-6, the period was 28 days long and the station series was 40 years long; therefore, N = 40 × 28 = 1120. The sample contained 844 days with zero precipitation and 276 days with measurable precipitation. The rank of the smallest $y_{p,\tau}$ greater than zero was 845, and the plotting position was P(y) = 845/1120 = 0.754. The P(y) and $y_{p,\tau}$ values for all $y_{p,\tau}$ greater than zero were plotted on normal probability paper and are shown for the three transformations in Figures 3-4, 3-5, and 3-6. The maximum likelihood estimates of μ_τ and σ_τ were used to plot the normal distribution cumulative probability line on the figures. The normal distribution line fit the square root data well throughout the range of the data (Figure 3-5). The χ^2 value was 9.4. The cube root data differed significantly from the normal distribution line for the larger precipitation amounts (Figure 3-6) and resulted in a χ^2 of 28.5. The precipitation data with no transformation was highly skewed and departed significantly from the normal distribution line throughout the range of the data (Figure 3-4). The χ^2 value for no transformation was 130.3.

These results indicated that, for the study region, the square root transformation is the best normalizing transformation of the three cases that were investigated. The sample χ^2 for the square root transformation was less than χ_0^2 for most of the samples. Therefore, the hypothesis that the samples with the square root transformation were from normal populations was accepted. For the analysis that will follow, the square root transformation will be assumed to be adequate for transforming the daily precipitation to a normal distribution.

3. *Characteristics of the Square Root Transformation*

The density function of daily precipitation at a point given the occurrence of a wet day is, generally, considered to be monotonically decreasing. Days with small rainfall amounts occur more frequently than days with large rainfall amounts. Several probability density functions have been used to describe the distribution of daily precipitation amounts for days with measurable rainfall. The most commonly used distributions are the exponential and gamma distributions.

If the distribution of daily precipitation for days with measurable rainfall is monotonically decreasing, it is appropriate to investigate the distribution of daily precipitation after the application of the square root transformation. Assume that the two-parameter gamma distribution function, given by

$$f(x) = \frac{x^{\alpha-1}\, e^{-x/\beta}}{\beta^{\alpha}\, \Gamma(\alpha)}, \qquad (3\text{-}3)$$

describes the distribution of daily precipitation, x, given the occurrence of nonzero daily rainfall. If

Table 3-2. Chi-square goodness-of-fit parameter for no transformation, square root transformation and cube root transformation for three precipitation stations in the study region.

Period	Crockett			Hewitt			Breckenridge		
	$y=x$	$y=x^{1/2}$	$y=x^{1/3}$	$y=x$	$y=x^{1/2}$	$y=x^{1/3}$	$y=x$	$y=x^{1/2}$	$y=x^{1/3}$
1	169.1	7.0[1/]	23.8	288.0	7.1[1/]	4.5[1/]	159.9	4.9[1/]	10.0[1/]
2	201.6	12.7[1/]	31.3	526.2	27.0	9.6[1/]	349.6	6.2[1/]	16.7
3	208.0	10.3[1/]	20.3	265.3	19.3	11.7[1/]	77.2	6.1[1/]	16.2
4	142.2	3.3[1/]	14.0[1/]	232.9	13.6[1/]	15.8	93.6	4.5[1/]	17.3
5	207.5	10.1[1/]	23.1	168.2	9.9[1/]	35.3	132.2	6.5[1/]	31.0
6	72.3	9.7[1/]	38.1	119.2	13.8[1/]	36.1	149.6	11.7[1/]	30.5
7	39.9	16.3	44.1	74.9	2.8[1/]	13.4[1/]	169.7	6.8[1/]	2.6[1/]
8	133.1	4.8[1/]	16.9	47.7	8.0[1/]	21.7	126.1	3.9[1/]	5.1[1/]
9	133.1	9.7[1/]	29.9	157.1	6.6[1/]	11.1[1/]	135.8	10.8[1/]	20.4
10	117.4	7.6[1/]	23.5	139.2	5.0[1/]	20.0	84.6	10.8[1/]	30.8
11	151.0	14.9	10.5[1/]	169.6	6.5[1/]	7.1[1/]	89.1	3.9[1/]	16.9
12	199.8	9.6[1/]	20.3	199.3	8.4[1/]	9.8[1/]	34.5	13.5[1/]	34.8
13	235.5	13.9[1/]	23.3	266.2	12.2[1/]	9.3[1/]	74.4	12.1[1/]	21.9
Avg.	154.7	10.0[1/]	24.6	204.1	10.8[1/]	15.8	128.9	7.8[1/]	19.6

1/ The χ^2 value is less than the χ_0^2 value of 14.1.

Table 3-3. Average Chi-square fit parameter for no transformation, square root transformation, and cube root transformation for eight stations in the study region.

No.	Station Name	$y=x$	$y=x^{1/2}$	$y=x^{1/3}$
1	Breckenridge	128.9	7.8[1/]	19.6
2	Brownwood	113.3	10.6[1/]	23.8
5	Corsicana	89.7	9.9[1/]	31.5
6	Crockett	154.7	10.0[1/]	24.6
7	Dallas	293.8	15.6	12.4[1/]
11	Gilmer	167.0	9.4[1/]	23.7
12	Hewitt	204.1	10.8[1/]	15.8
19	Temple	357.5	25.3	16.2
Ave.		188.4	12.3[1/]	20.9

1/ The χ^2 value is less than the χ_0^2 value of 14.1.

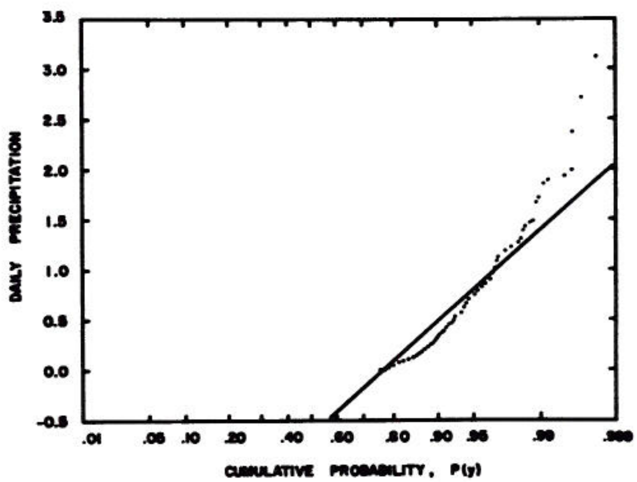

Figure 3-4. Cumulative probability distribution for daily precipitation (y = x) from the third 28-day period of the year at Gilmer, Texas.

the transformation $y = x^{1/2}$ is made, it may be shown that the distribution of y is

$$g(y) = \frac{2y^{2\alpha-1}\, e^{-y^2/\beta}}{\beta^{\alpha}\, \Gamma(\alpha)} . \qquad (3\text{-}4)$$

The two-parameter gamma distribution given in equation (3-3) was fit to a sample of daily precipitation data. The data for the fourth 28-day period of the year at the Gilmer station was used to estimate the parameters. Maximum likelihood estimates of the α and β parameters were obtained using the sample data. The estimated values of the parameters are $\alpha = 0.76$ and $\beta = 0.69$. The gamma function with the estimated parameters may be written

$$f(x) = 1.09\, x^{-0.24}\, e^{-1.45x} . \qquad (3\text{-}5)$$

The observed frequency histogram and the fitted gamma frequency function are shown in Figure 3-7. The gamma function is a good fit of the observed histogram.

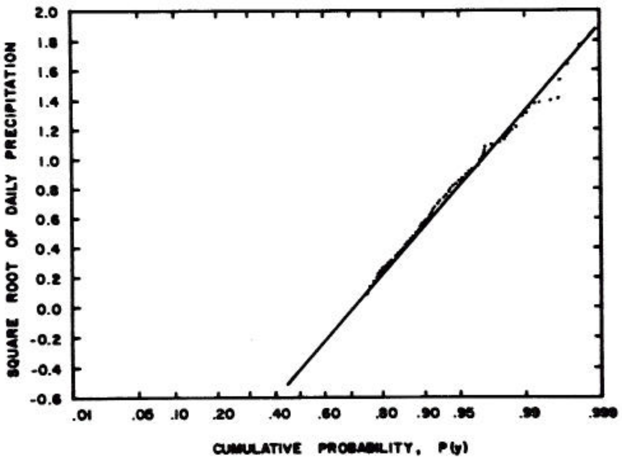

Figure 3-5. Cumulative probability distribution for the square root of daily precipitation ($y = x^{1/2}$) from the third 28-day period of the year at Gilmer, Texas.

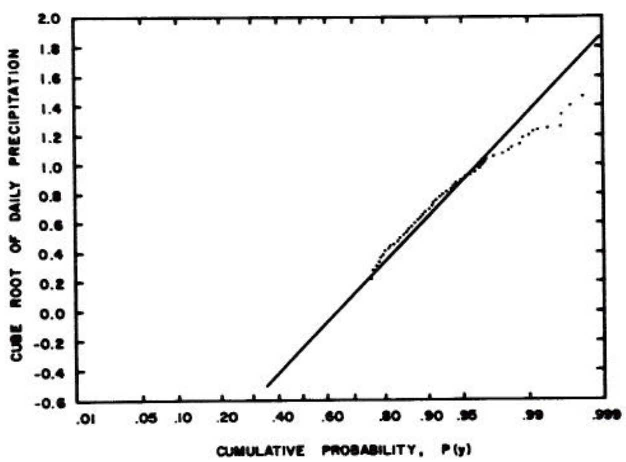

Figure 3-6. Cumulative probability distribution for the cube root of daily precipitation ($y = x^{1/3}$) from the third 28-day period of the year at Gilmer, Texas.

The estimated values of α and β, obtained from the sample of daily precipitation data, were substituted into equation (3-4). The fitted distribution of y, assuming x has a two-parameter gamma distribution, may be written

$$g(y) = 2.18\ y^{0.51}\ e^{-1.45y^2}. \qquad (3\text{-}6)$$

The observed frequency histogram of the square root of daily precipitation is shown in Figure 3-8. The fitted frequency function obtained from equation (3-6) is also shown in Figure 3-8. The frequency function has a value of zero for $y = 0$. There is no data for the observed histogram for the interval $0 < y < 0.1$, because values of y less than 0.1 ($x < 0.01$)

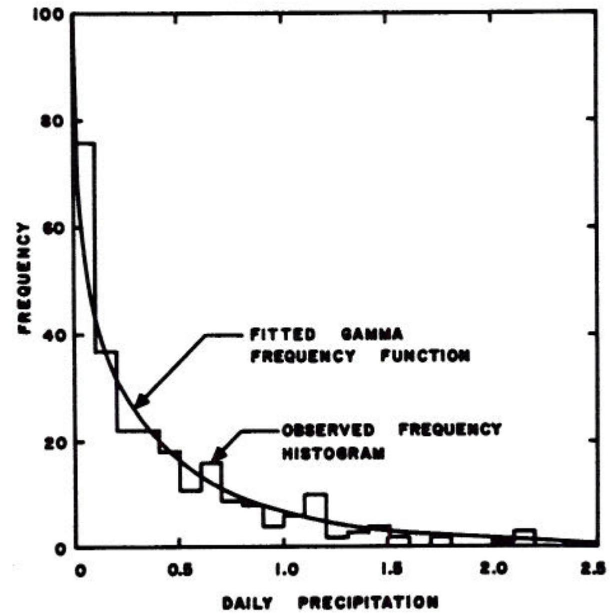

Figure 3-7. Daily precipitation data for the fourth 28-day period of the year at Gilmer, Texas and the fitted two-parameter gamma frequency function.

are recorded as zero. The observed frequency histogram has a mode at the interval $0.2 \leq y < 0.3$. The frequency function also has a mode, but the peak occurs at a y value of about 0.4. The frequency function is a good approximation of the observed histogram for values of y greater than 0.5.

The fitted truncated normal distribution is also shown in Figure 3-8. The normal distribution does not have a mode within the range of Figure 3-8. Since the probability of zero daily precipitation is greater than 0.5 the peak of the normal distribution occurs at a y value less than zero. The truncated normal frequency function is greater, for values of y less than 0.2, than both the observed histogram and the function determined by applying the square root transformation to the gamma distribution. Both frequency functions are good approximations of the observed frequency histogram for values of y greater than 0.5.

If the truncated normal distribution shown in Figure 3-8 is used to generate new samples of the square root of daily precipitation, more values less than 0.2 would be generated than were observed. Generated values less than 0.1 are considered a trace and are set to zero. The values between 0.1 and 0.2 become daily precipitation amounts of 0.01 to 0.04 inch, after the inverse of the transformation is applied. The generated sample would contain more days with precipitation between 0.01 and 0.04 inch than the observed sample.

The observed frequency histograms shown in Figures 3-7 and 3-8 are typical of the distribution of daily precipitation for all stations in the study

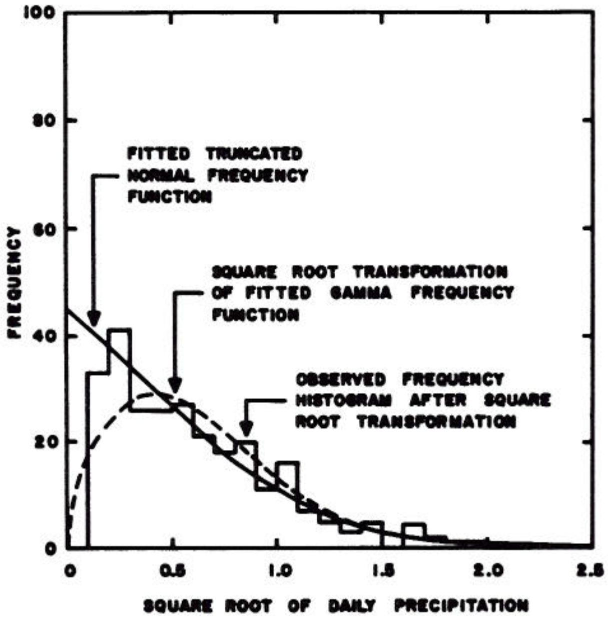

Figure 3-8. Square root of daily precipitation data for the fourth 28-day period of the year at Gilmer, Texas, the square root transformation of the fitted gamma frequency function, and the fitted truncated normal frequency function.

region. The frequency of occurrence of daily precipitation of a given amount decreases as the amount of daily precipitation increases. When the square root transformation is applied the distribution has a mode. The mode occurs at a small value of the square root of daily precipitation. The mode occurs in the interval $0.2 \leq y < 0.3$ for some samples, like that in Figure 3-8. For other samples, the mode occurs in the interval $0.1 \leq y < 0.2$ and does not seem to be a mode because values below 0.1 are considered zero (Figure 2-3). The truncated normal distribution used in this study does not describe the mode observed in the distribution of the square root of daily precipitation. Use of the truncated normal distribution to generate new samples should result in samples that contain too many days with small rainfall amounts. However, days with very small precipitation amounts are unimportant for most hydrologic purposes.

4. Evaluation of Alternative I and Alternative II

Two basic alternatives for developing the time-area daily precipitation model were outlined in Chapter II. The first alternative is to infer and remove the periodic means and standard deviations before transforming the data to approximate a sample from a normal distribution. The second alternative is to apply the normalizing transformation before inferring the means and standard deviations.

There are two primary factors that cause daily precipitation values to be periodic within the year. First, the probability of the occurrence of a wet day may be periodic. Second, the parameters of the distribution of daily precipitation given the occurrence of a wet day may also be periodic. If the probability of the occurrence of a wet day is periodic within the year but the distribution of precipitation amounts given the occurrence of a wet day is nonperiodic, the resulting daily precipitation is periodic due to the frequency of occurrence of daily precipitation. Similarly, if the probability of a wet day is nonperiodic but the distribution of precipitation amounts given a wet day is periodic, the resulting daily precipitation is periodic. In many cases both the probability of a wet day and the distribution of precipitation amounts given a wet day are periodic. The periodic characteristics of these two factors, as will be shown later, affect the selection of the alternative to use in developing the model.

Data from the Gilmer station were used to evaluate the two alternatives. The probability of a wet day, P(W), was calculated for each day of the year and is shown in Figure 3-9. The P(W) values were highly variable but displayed periodicity within the year. A Fourier series with two harmonics explained 14.2 percent of the variance of the 365 values of P(W). The Fourier series representation, using two harmonics, is also shown in Figure 3-9.

a. Alternative I. The means and standard deviations of daily precipitation given the occurrence of a wet day were calculated for each day of the year. The results are shown in Figure 3-10. Fourier series with two harmonics explained about 10 percent of the variance of the mean, μ_τ, and 6 percent of the variance of the standard deviation, σ_τ. The periodic movement of μ_τ and σ_τ is not in phase with that given in Figure 3-9 for P(W). The means and standard deviations have harmonics with about the same phase. The value of μ_τ/σ_τ was approximately constant (0.97) for all days of the year.

The periodic means and standard deviations may be removed from the precipitation data by

$$z_{p,\tau} = \frac{x_{p,\tau} - \mu_\tau}{\sigma_\tau}, \quad x_{p,\tau} \neq 0, \qquad (3\text{-}7)$$

or

$$z_{p,\tau} = \frac{x_{p,\tau}}{\sigma_\tau} - \frac{\mu_\tau}{\sigma_\tau}, \quad x_{p,\tau} \neq 0. \qquad (3\text{-}8)$$

Use of equation (3-8) would result in both positive and negative values of $z_{p,\tau}$ for values of $x_{p,\tau}$ greater than zero. Since μ_τ/σ_τ was approximately constant, the periodicity in the mean and standard deviation may be removed by

$$z_{p,\tau} = \frac{x_{p,\tau}}{\sigma_\tau}, \quad x_{p,\tau} \neq 0. \qquad (3\text{-}9)$$

Equation (3-9) was used to remove the periodicity in the means and standard deviations of the daily precipitation data for the Gilmer station. The σ_τ values were determined from the Fourier series

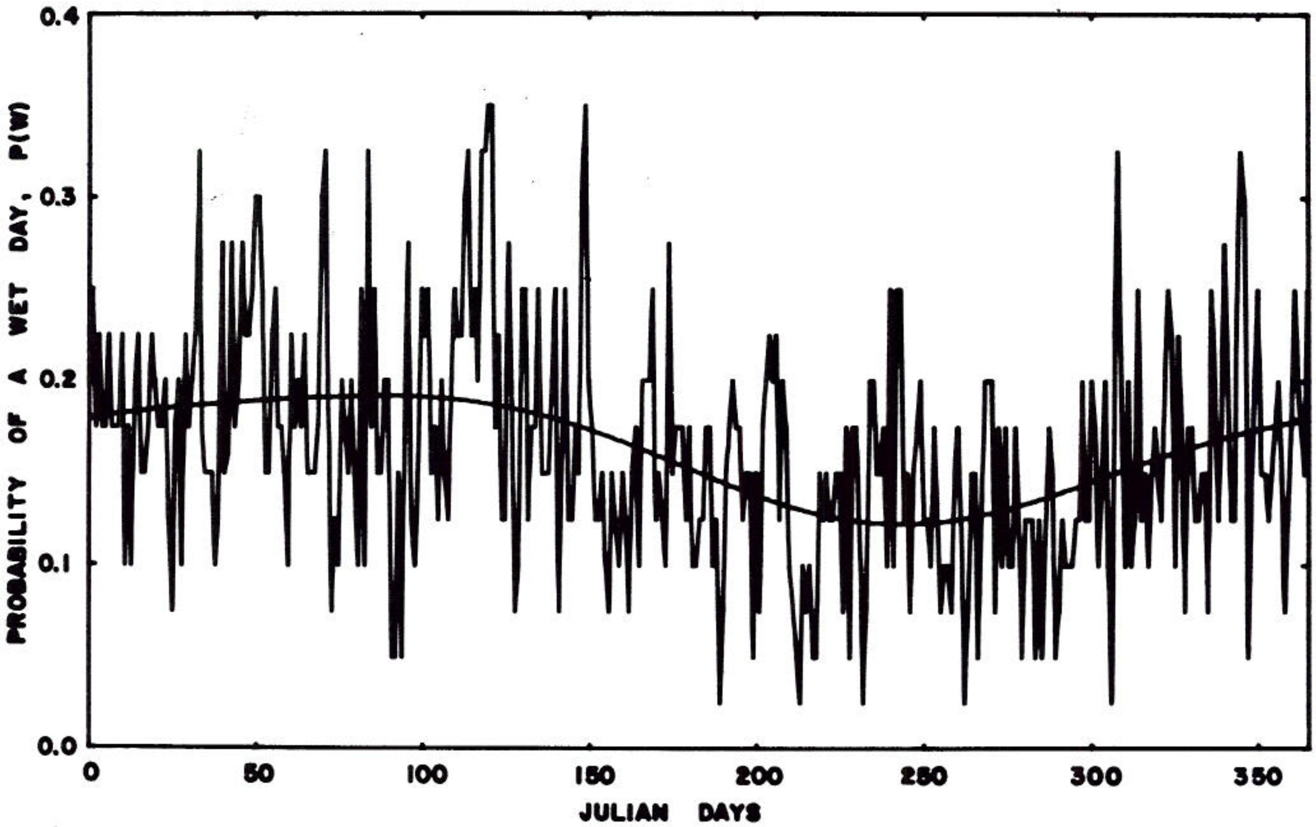

Figure 3-9. Daily values of the probability of a wet day and the fitted periodic component with two harmonics, Gilmer, Texas.

representation. The $z_{p,\tau}$ data were transformed to approximate a sample from a truncated normal distribution by

$$y_{p,\tau} = (z_{p,\tau})^{1/2}. \tag{3-10}$$

The means and standard deviations of the $y_{p,\tau}$ series were determined for each day of the year using the truncated normal estimation technique. The results are shown in Figure 3-11. The standard deviations are approximately constant for the year. The means show a periodic movement similar to that of P(W) in Figure 3-9. With the truncated normal distribution concept that is used in this study, the probability of a day with zero precipitation, P(D), is given by the integral of the normal distribution from $-\infty$ to the truncation point. The probability of a wet day is given by

$$P(W) = 1 - P(D) . \tag{3-11}$$

Since P(W) is periodic within the year and the standard deviation of the truncated normal distribution is constant, the mean of the truncated normal distribution must be periodic to properly define P(W). Therefore, with Alternative I separate descriptions are required for the periodicity of P(W) and the periodicity of the parameters of the distribution of daily precipitation given the occurrence of a wet day.

b. Alternative II. The square root transformation was applied to all nonzero daily precipitation data for the Gilmer station. The mean, μ_τ, and the standard deviation, σ_τ, were estimated for all 365 days of the year, using the truncated normal estimation technique. The results are given in Figure 3-12. The estimates of both μ_τ and σ_τ have periodic components and large random sampling fluctuations. Fourier series with two harmonics were fit to the estimates of both μ_τ and σ_τ.

All of the estimated μ_τ values are negative because P(D) is greater than 0.50 for each day of the year. P(D) for any day is given by the integral of the normal distribution from $-\infty$ to the truncation point. The distribution of daily precipitation, given the occurrence of a wet day, is given by the remainder of the distribution (Figure 2-3). With Alternative II both the periodicity of P(W) and the periodicity of the parameters of the distribution of daily precipitation given the occurrence of a wet day are defined by the Fourier series of μ_τ and σ_τ shown in Figure 3-12.

Figure 3-10 shows that, with Alternative I, the means and the standard deviations of daily precipitation given the occurrence of a wet day are directly proportional. Days with high means also had high standard deviations. The values of μ_τ/σ_τ were shown

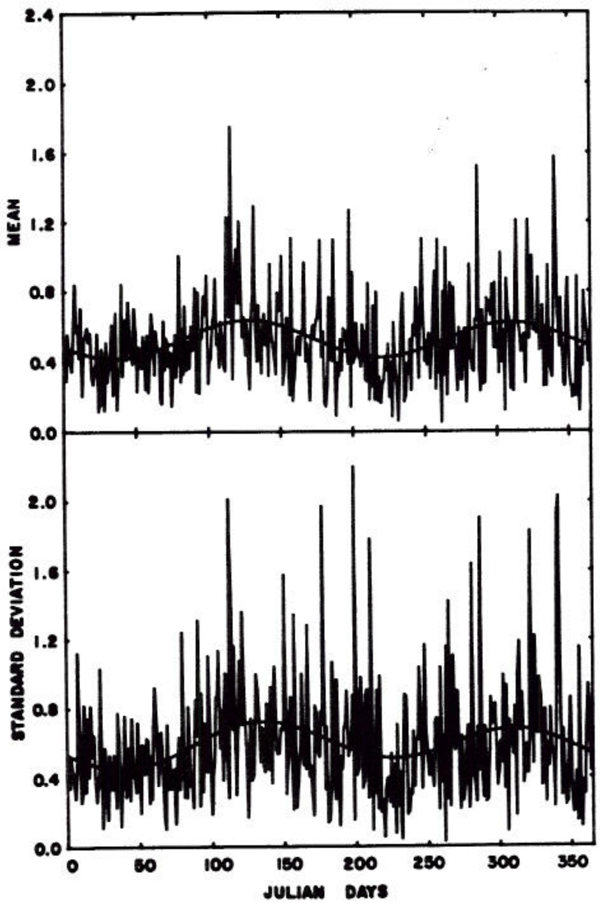

Figure 3-10. The means and standard deviations of daily precipitation given a wet day and fitted periodic components with two harmonics, Gilmer, Texas.

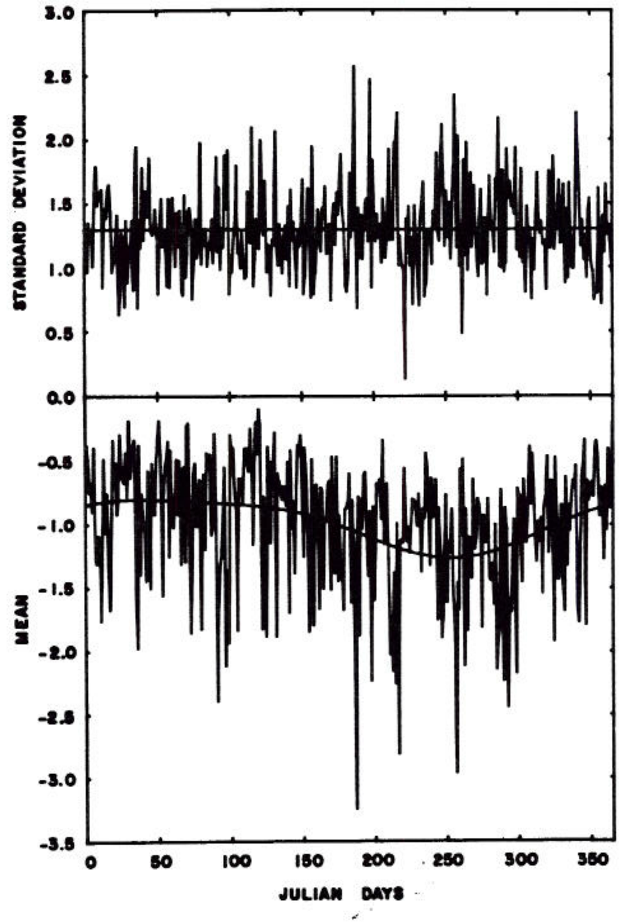

Figure 3-11. The means and standard deviations of the truncated normal distribution after removal of the means and standard deviations given a wet day, Gilmer, Texas.

to be approximately constant for all days of the year. However, Figure 3-12 shows that with Alternative II the standard deviations tend to be large when the means are small. Since the means are negative and the standard deviations are positive, the values of μ_τ/σ_τ are negative. For μ_τ/σ_τ to be approximately constant σ_τ must increase as μ_τ decreases.

Alternative II was chosen for developing the time-area daily precipitation model because the periodicity of both the probability of a wet day and the distribution of precipitation given a wet day could be described by describing the periodicity of the mean and standard deviation of the truncated normal after applying the square root transformation. Alternative II will be used in the remainder of this study.

5. *Determination of Harmonics in μ_τ and σ_τ*

The estimates of μ_τ and σ_τ, $\tau = 1, 2, \ldots, 365$, contain both periodic components and random sampling components. The random sampling fluctuations are particularly large in this study because the large number of days with zero precipitation on a given day of the year causes the nonzero sample size to be small. Values of m_τ and s_τ (the sample estimates of μ_τ and σ_τ, respectively) were determined using the maximum likelihood technique for all 365 days of the year and for each station in the study region. The variability of m_τ and s_τ was illustrated in Figure 3-12 for a selected station within the region. The random fluctuations of m_τ and s_τ cause greater departures from the overall means than the amplitudes of the periodic components. This fact makes difficult the determination of the number of harmonics required to describe the periodic movement of μ_τ and σ_τ. The inclusion of too many harmonics will perpetuate part of the sampling error involved in estimating the parameters. However, retaining too few harmonics will result in an inaccurate description of the periodic nature of the physical process.

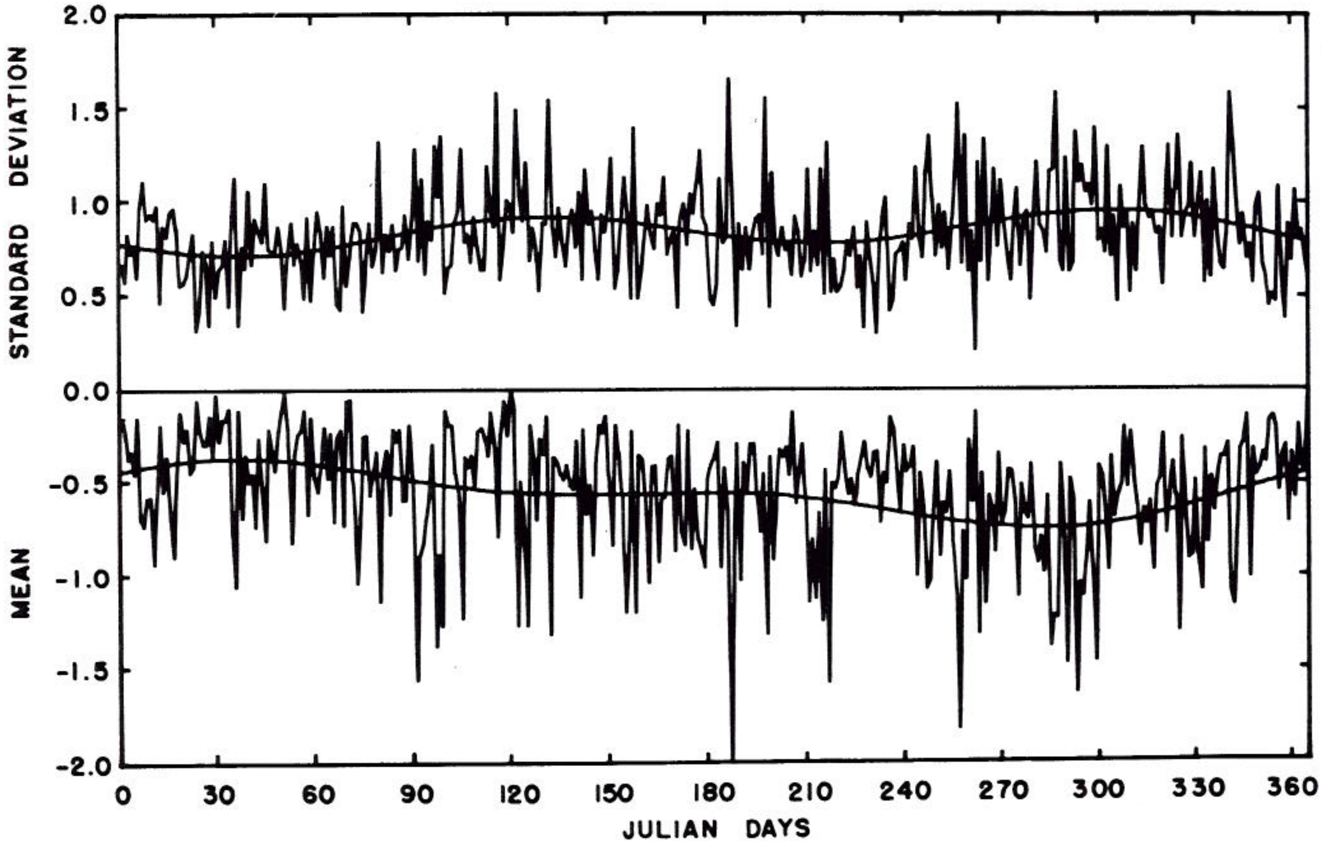

Figure 3-12. Daily values of μ_τ and σ_τ and fitted periodic components with two harmonics, Gilmer, Texas.

Yevjevich (1972b) outlined three approaches for determining the significant harmonics for describing the periodicity of parameters. The three approaches include: (1) Fisher's approach, (2) an approach using the break point in a cumulative periodogram, and (3) an empirical approach using the first m harmonics that are required to explain a critical percent of the variation of the parameter.

a. Fisher's approach. Fisher's test is the classical method of testing each harmonic of a Fourier series description of a parameter for significance. The test is based on the variance of individual harmonics, $C_j^2/2$. If a given $C_j^2/2$ value is not greater than a critical $C_c^2/2$ value of a normal independent stochastic process, the j-th harmonic is considered insignificant. The Fisher parameter for the harmonic of maximum magnitude is

$$g = \frac{C_{max}^2}{2s_\nu^2} , \qquad (3\text{-}12)$$

where s_ν^2 is the estimate of the variance of the parameter ν. For the second and higher harmonics, the parameter is given by

$$g = \frac{C_j^2}{2s_\nu^2 - \sum_{i=1}^{j-1} C_i^2} . \qquad (3\text{-}13)$$

If the g value for a specific harmonic is greater than a critical value, g_c, the harmonic is significant. A table of g_c values for different sample sizes and probability levels is given by Yevjevich (1972b).

Fisher's test is very difficult to apply in many cases. This is particularly true when attempting to determine the number of significant harmonics of a parameter in a very complex hydrologic time series with large random sampling components. Frequently, Fisher's test will indicate no significant harmonics in a hydrological parameter that has obvious seasonal fluctuation. In this study, Fisher's test was rejected because of the difficulty in applying it, and the other two approaches given by Yevjevich were investigated to determine an appropriate test procedure for the number of significant harmonics of μ_τ and σ_τ.

b. Periodogram break point technique. The portion of the variance of a parameter that is explained by the first m harmonics may be determined by

$$p_m = \sum_{j=1}^{m} \frac{C_j^2/2}{s_\nu^2} \qquad (3\text{-}14)$$

where s_ν^2 is the estimate of the variance of the 365 values of the parameter. The cumulative periodogram method was described by Yevjevich (1972b) as being based on the concept that p_m as a function of m is composed of two parts: (1) a fast rising part associated with the periodic component, and (2) a slow rising part associated with the random component. Yevjevich suggested that the two parts be approximated by smooth curves (or straight lines) intersecting at a point. The number of harmonics corresponding to the intersection point is taken to be the number of significant harmonics.

Values of m_τ and s_τ, τ = 1, 2, ..., 365, were computed for selected $y_{p,\tau}$ series, using the truncated normal technique described earlier. A Fourier series with 182 harmonics was fitted to each set of m_τ and s_τ values. The $C_j^2/2$ values were computed and the portion of the variance of the parameters explained by the first m harmonics was determined using equation (3-14). Periodograms of m_τ and s_τ were plotted for the selected $y_{p,\tau}$ series. The periodograms for two stations are shown in Figures 3-13 and 3-14. The periodograms contain the fast rising part for the low order harmonics and the slow rising part for the higher order harmonics. However, the transition from the fast rising part to the slow rising part is gradual for the four curves given in Figures 3-13 and 3-14. These periodograms are typical of those obtained for all stations in the study region. An objective decision about the number of significant harmonics in μ_τ and σ_τ, based on these periodograms, was practically impossible.

c. Critical explained variance approach. The third method given by Yevjevich for determining the number of significant harmonics of a hydrologic parameter is an empirical method based on determining the number of harmonics required to explain a critical portion of the variance of the parameter. Experience has shown that for most hydrologic series harmonics beyond the sixth are very rarely shown to be significant.

With the critical explained variance approach, the portion of the variance explained by the first j harmonics are computed by equation (3-14). The part of the variation explained by the first six harmonics is given by p_6. Two critical p values, p_{min} and p_{max}, are defined. If $p_6 \leq p_{min}$, no significant harmonics exist in the parameter. If $p_{min} < p_6 \leq p_{max}$, all six harmonics of the parameter are significant. If $p_6 > p_{max}$, less than six harmonics are significant. When less than six harmonics are found to be significant, the $C_j^2/2$ values are arranged in descending order. The explained variance for each harmonic is determined and summed. The harmonics that are required to first give a p_j value greater

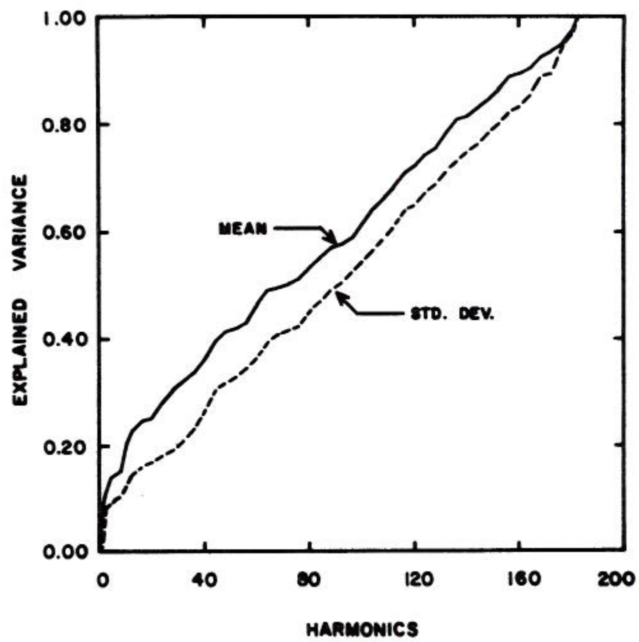

Figure 3-13. Periodograms of the portion of the variance of the mean, m_τ, and the variance of the standard deviation, s_τ, for the $y_{p,\tau}$ series explained by the first m harmonics, Gilmer, Texas.

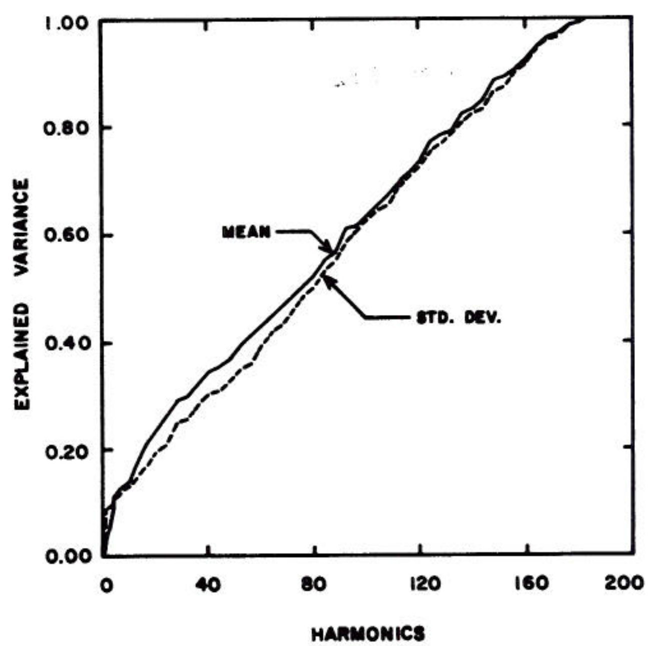

Figure 3-14. Periodograms of the portion of the variance of the mean, m_τ, and the variance of the standard deviation, s_τ, of the $y_{p,\tau}$ series explained by the first m harmonics, Breckenridge, Texas.

than p_{max} are considered the significant harmonics.

Empirical expressions for p_{min} and p_{max}, given by Yevjevich (1972b) are

$$p_{min} = a\sqrt{\frac{\omega}{cn}} \qquad (3\text{-}15)$$

and

$$p_{max} = 1 - p_{min} \qquad (3\text{-}16)$$

where ω is the length of the basic period, c is the highest moment used in defining the parameter, n is the series length in years, and a is an empirical constant. The suggested value for the constant is a = 0.033.

For the 40 years of daily precipitation that were used in this study, ω = 365 and n = 40. For m_τ the value of c is 1, and for s_τ the value of c is 2. Using the constant a = 0.033, as suggested by Yevjevich, the critical p values are

$$p_{min}(m_\tau) = 0.099 \qquad (3\text{-}17)$$

$$p_{max}(m_\tau) = 0.901 \qquad (3\text{-}18)$$

$$p_{min}(s_\tau) = 0.070 \qquad (3\text{-}19)$$

$$p_{max}(s_\tau) = 0.930. \qquad (3\text{-}20)$$

The Fourier coefficients for the first six harmonics of m_τ and s_τ were calculated for each of the 20 stations in the study region. The portion of the variance of each parameter explained by the first six harmonics was determined using equation (3-14). The results are given in Table 3-4. For 14 of the 20 stations, $p_6(m_\tau)$ was between the two critical p values for the mean. All 20 of the $p_6(s_\tau)$ values were between the two critical standard deviation p values. Therefore, at this point in the analysis, all six harmonics of the mean and the standard deviation were assumed significant for stations within the study region.

6. Regionalization of Amplitudes and Phases in Harmonics of μ_τ and σ_τ

For the time-area model of daily precipitation to be useful at any part of the study region, the amplitude and phase coefficients of harmonics must be related to position within the region. The general mean of the sample daily means, $\bar{m}_\tau$, and the amplitude and phase angle, $C_j(m)$, j = 1, 2, ..., 6, and $\theta_j(m)$, j = 1, 2, ..., 6, of each of the first six harmonics of the daily means are given in Table 3-5 for all 20 stations. Similarly, the general mean of the sample daily standard deviations, $\bar{s}_\tau$, and the amplitude and phase angle of the daily standard deviations, $C_j(s)$, j = 1, 2, ..., 6, and $\theta_j(s)$, j = 1, 2, ..., 6, are given in Table 3-6.

It was thought that the m_τ and s_τ values for a given station would be proportional, resulting in a constant coefficient of variation throughout the year. If such were the case, the parameters

Table 3-4. Ratio of variances of the means and standard deviations for a Fourier series with six harmonics to the total variances of m_τ and s_τ.

No.	Station Name	$p_6(m_\tau)$	$p_6(s_\tau)$
1	Breckenridge	0.128	0.140
2	Brownwood	0.075	0.108
3	Centerville	0.112	0.081
4	Comanche	0.097	0.093
5	Corsicana	0.189	0.150
6	Crockett	0.092	0.102
7	Dallas	0.112	0.121
8	Dialville	0.139	0.100
9	Dublin	0.100	0.100
10	Fort Worth	0.128	0.109
11	Gilmer	0.146	0.122
12	Hewitt	0.103	0.089
13	Hillsboro	0.082	0.114
14	Kaufman	0.133	0.100
15	Lampasas	0.118	0.086
16	Mexia	0.149	0.086
17	Palestine	0.152	0.130
18	Riesel	0.095	0.072
19	Temple	0.127	0.096
20	Waco	0.061	0.079

describing the periodic movement of the standard deviation would not need to be related to position and fewer model parameters would be required. However, for the stations in the study region the periodic means and periodic standard deviations were slightly out of phase (Figure 3-12). This was tested by comparing $\theta_1(m)$ and $\theta_1(s)$ values for each station. If m_τ and s_τ for a given station are in phase, the difference in $\theta_1(m)$ and $\theta_1(s)$ should be zero. The differences between $\theta_1(m)$ and $\theta_1(s)$ for all 20 stations are shown in Table 3-7. The value of $\theta_1(m)$ is consistently less than $\theta_1(s)$. The difference was significant at the 1% level. The periodic coefficient of variation, caused by the difference in phase of m_τ and s_τ, made it necessary to develop equations relating the amplitude and phase of the standard deviation to position, as well as those for the mean.

Yevjevich and Karplus (1973) found a linear model adequate for relating precipitation amplitude and phase to position for two regions. Figure 3-1 illustrates an approximately linear variation of mean annual precipitation with position in the region used in this study. This suggests that perhaps the amplitude and phase of m_τ and s_τ for this region also vary linearly across the region. A linear model of the form

$$\nu = \alpha_1 + \alpha_2 L_1 + \alpha_3 L_2 \qquad (3\text{-}21)$$

where L_1 = the station latitude and L_2 = the station longitude was assumed to describe the regional variation of each parameter given in Tables 3-5 and 3-6. The equation coefficients (α_1, α_2, and α_3) were evaluated for each of the 26 parameters, using stepwise multiple linear regression. Each coefficient was examined to determine if it differed significantly

Table 3-5. Amplitudes and phases of the first six harmonics fitted to the 365 daily means for each station.

Station Number	Mean		Harmonic, j 1	2	3	4	5	6
1	-.668	C_j	0.215	-0.113	0.111	-0.135	0.054	0.042
		θ_j	-0.421	-1.391	-0.688	-0.562	1.457	-0.622
2	-.723	C_j	0.131	-0.073	0.071	-0.097	0.058	0.144
		θ_j	-1.026	-1.242	0.507	0.027	-0.804	1.397
3	-.612	C_j	0.206	0.133	0.068	-0.120	-0.089	0.070
		θ_j	-0.977	-0.439	-0.879	-0.628	0.383	-0.854
4	-.714	C_j	0.152	-0.075	0.063	-0.161	-0.096	0.107
		θ_j	-1.005	-1.546	-1.403	-0.168	0.152	0.215
5	-.611	C_j	0.241	0.127	0.069	-0.077	-0.120	0.047
		θ_j	-1.063	0.231	-1.143	0.099	-0.652	-1.549
6	-.692	C_j	0.129	0.233	0.045	-0.047	-0.100	0.046
		θ_j	-0.895	-0.764	-0.926	-1.052	0.609	-1.469
7	-.705	C_j	0.238	0.098	0.076	-0.115	-0.061	0.103
		θ_j	-0.994	-0.228	0.314	0.095	-1.337	-0.171
8	-.535	C_j	0.192	0.091	-0.046	-0.057	-0.083	0.072
		θ_j	-0.899	-0.291	0.977	-1.231	-1.350	-0.378
9	-.584	C_j	0.152	0.040	0.019	-0.127	0.042	0.091
		θ_j	-1.436	0.623	-1.235	-0.037	1.476	0.364
10	-.652	C_j	0.269	0.083	0.046	-0.059	-0.051	0.054
		θ_j	-1.058	0.967	-0.407	-0.129	-0.785	-0.695
11	-.587	C_j	0.197	0.128	0.036	-0.061	-0.020	0.107
		θ_j	-1.035	-0.806	0.293	-0.271	-0.173	0.371
12	-.740	C_j	0.205	0.082	0.122	-0.074	-0.060	0.096
		θ_j	-1.041	0.316	-1.105	-0.195	-0.769	-1.160
13	-.688	C_j	0.186	0.074	0.022	-0.101	0.038	0.039
		θ_j	-1.005	-0.417	-0.994	0.041	0.841	-1.453
14	-.633	C_j	0.271	0.044	0.085	-0.089	-0.023	0.032
		θ_j	-0.856	0.714	-0.196	-0.271	-1.090	0.732
15	-.652	C_j	0.117	0.047	0.159	-0.121	-0.023	0.021
		θ_j	-0.686	0.429	-0.623	-0.319	-0.115	0.196
16	-.588	C_j	0.249	0.031	0.109	-0.051	-0.060	0.081
		θ_j	-1.055	0.663	-1.199	-0.705	-0.059	-0.273
17	-.606	C_j	0.187	0.172	0.083	-0.160	-0.063	0.098
		θ_j	-0.714	-0.672	-0.664	-0.415	0.006	-0.745
18	-.639	C_j	0.181	0.070	0.096	-0.052	-0.041	0.109
		θ_j	-0.830	0.143	-1.463	0.147	-0.369	0.026
19	-.689	C_j	0.208	0.031	0.111	-0.143	-0.033	0.079
		θ_j	-1.109	-1.419	-0.650	-0.052	-1.269	-1.396
20	-.667	C_j	0.148	0.066	0.069	-0.103	-0.018	0.067
		θ_j	-1.005	-0.319	-0.896	0.095	-0.777	-1.118

Table 3-6. Amplitudes and phases of the first six harmonics fitted to the 365 daily standard deviations for each station.

Station Number	Mean		Harmonic, j 1	2	3	4	5	6
1	.795	C_j	-0.190	-0.019	-0.019	0.063	-0.036	-0.014
		θ_j	-0.299	0.819	-0.229	-0.525	0.795	-0.027
2	.838	C_j	-0.118	-0.022	0.079	0.050	-0.047	-0.091
		θ_j	-0.457	-0.561	-1.561	0.161	-1.047	1.460
3	.857	C_j	-0.090	-0.123	0.020	0.056	0.033	-0.030
		θ_j	-0.922	-0.758	1.529	-0.666	0.326	-1.108
4	.843	C_j	-0.118	-0.036	0.039	0.080	0.042	-0.061
		θ_j	-0.575	-0.229	-0.499	-0.171	0.392	0.526
5	.841	C_j	-0.107	-0.119	-0.023	0.028	0.054	0.040
		θ_j	-0.917	-0.399	1.027	0.398	-0.876	0.995
6	.927	C_j	-0.041	-0.175	0.018	-0.019	0.029	0.021
		θ_j	-1.007	-0.945	0.354	1.066	0.665	0.761
7	.875	C_j	-0.127	-0.126	-0.057	0.049	-0.027	-0.052
		θ_j	-0.723	-0.601	1.227	0.170	1.083	-0.339
8	.834	C_j	-0.073	-0.089	0.052	-0.031	-0.054	-0.155
		θ_j	-0.675	-0.883	0.515	1.398	1.273	0.167
9	.781	C_j	-0.103	-0.073	0.048	0.045	-0.020	-0.040
		θ_j	-0.904	-0.635	-0.747	-0.009	0.610	0.733
10	.825	C_j	-0.155	-0.065	0.032	0.013	0.008	-0.026
		θ_j	-0.717	-0.252	-1.336	-0.043	-0.327	-1.168
11	.872	C_j	-0.078	-0.127	0.006	0.023	-0.024	-0.042
		θ_j	-0.926	-1.084	-1.297	-0.700	-0.363	0.961
12	.876	C_j	-0.127	-0.090	0.041	0.015	-0.019	-0.049
		θ_j	-0.760	-0.332	1.253	0.157	1.168	-1.557
13	.882	C_j	-0.090	-0.125	0.036	0.058	-0.040	0.044
		θ_j	-0.607	-0.881	-1.391	0.024	0.281	1.223
14	.863	C_j	-0.140	-0.072	-0.036	0.037	-0.020	0.013
		θ_j	-0.677	-0.770	0.517	-0.350	0.085	-0.576
15	.820	C_j	-0.100	-0.071	-0.041	0.043	0.023	-0.003
		θ_j	-0.321	-0.496	-0.374	-0.287	1.512	1.325
16	.826	C_j	-0.118	-0.055	0.034	0.018	-0.002	-0.022
		θ_j	-0.911	-0.982	1.456	-0.641	-0.300	-0.375
17	.851	C_j	-0.078	-0.150	-0.028	0.081	0.013	-0.031
		θ_j	-0.510	-0.944	-1.137	-0.413	-0.009	-0.987
18	.838	C_j	-0.104	-0.082	0.040	0.011	-0.008	-0.041
		θ_j	-0.506	-0.778	1.262	0.435	0.295	0.033
19	.855	C_j	-0.123	-0.073	-0.025	0.059	-0.017	0.040
		θ_j	-0.871	-1.342	-0.535	-0.190	0.339	1.561
20	.829	C_j	-0.080	-0.105	0.010	0.052	-0.025	0.035
		θ_j	-0.621	-0.777	-0.685	0.208	-0.281	1.484

from zero. A standard t-test was used to test each coefficient. Each term in the equation with a nonsignificant coefficient was dropped from the equation. The equations for each parameter, with only the significant coefficients shown, are given in Table 3-8. If the coefficients of both L_1 and L_2 for a particular parameter were found to be not significantly different from zero, the parameter was taken to be a regional constant. In such cases, the values of the parameter for each of the 20 stations were averaged, and the average value is given in Table 3-8 as a regional constant. Each amplitude that was found to be a regional constant was tested to determine if it differed significantly from zero. If it was not significantly different from zero, it was assumed zero. The corresponding phase angle then became meaningless and was neglected.

The mean annual precipitation in the study region increases from west to east with very little north-south change. This east-west variation is also characteristic of most of the parameters given in Table 3-8. Nine of the parameters are a function of only L_2, indicating that those nine parameters change significantly with longitude (east-west position) but not with latitude. Only one parameter was significantly related to only L_1 (latitude), while two parameters were related to both L_1 and L_2. The east-

Table 3-7. Comparison of the phase angle of the first harmonic of the mean and the standard deviation.

No.	Station Name	$\theta_1(m)$	$\theta_1(s)$	$D = \theta_1(m) - \theta_1(s)$
1	Breckenridge	-0.421	-0.300	-0.121
2	Brownwood	-1.026	-0.457	-0.569
3	Centerville	-0.978	-0.922	-0.056
4	Comanche	-1.005	-0.575	-0.430
5	Corsicana	-1.063	-0.917	-0.146
6	Crockett	-0.895	-1.007	0.112
7	Dallas	-0.994	-0.723	-0.271
8	Dialville	-0.899	-0.675	-0.224
9	Dublin	-1.436	-0.904	-0.532
10	Fort Worth	-1.058	-0.717	-0.341
11	Gilmer	-1.035	-0.926	-0.109
12	Hewitt	-1.041	-0.760	-0.281
13	Hillsboro	-1.005	-0.607	-0.398
14	Kaufman	-0.856	-0.677	-0.179
15	Lampasas	-0.686	-0.321	-0.365
16	Mexia	-1.055	-0.911	-0.144
17	Palestine	-0.714	-0.510	-0.204
18	Riesel	-0.830	-0.506	-0.324
19	Temple	-1.110	-0.871	-0.239
20	Waco	-1.005	-0.621	-0.384
				$\bar{D}$ = -0.260
				s_D = 0.165
				t = -7.069[1/]

1/ $\bar{D}$ is significantly different from zero at 1 percent level.

Table 3-8. Equations relating amplitudes and phases to position.

Equation	Correlation Coefficient
$\bar{m}_\tau = 1.704 - 0.024L_2$	0.528
$C_1(m) = -1.201 + 0.044L_1$	0.559
$\theta_1(m) = -0.956$	--
$C_2(m) = 5.671 - 0.058L_2$	0.836
$\theta_2(m) = -0.272$	--
$C_3(m) = 0.071$	--
$\theta_3(m) = -0.619$	--
$C_4(m) = 1.377 - 0.015L_2$	0.489
$\theta_4(m) = -14.799 + 0.150L_2$	0.459
$C_5(m) = -2.496 + 0.025L_2$	0.577
$\theta_5(m) = -0.231$	--
$C_6(m) = 0.075$	--
$\theta_6(m) = -0.420$	--
$\bar{s}_\tau = 2.304 - 0.015L_2$	0.554
$C_1(s) = 2.366 - 0.028L_1 - 0.016L_2$	0.769
$\theta_1(s) = -10.185 + 0.098L_2$	0.557
$C_2(s) = -2.660 + 0.027L_2$	0.758
$\theta_2(s) = -32.647 + 0.316L_1 + 0.226L_2$	0.722
$C_3(s) = 0.000$	--
$\theta_3(s) =$ --	--
$C_4(s) = -1.194 + 0.013L_2$	0.511
$\theta_4(s) = 0.001$	--
$C_5(s) = 0.000$	--
$\theta_5(s) =$ --	--
$C_6(s) = -0.023$	--
$\theta_6(s) = 0.255$	--

west trend of $\bar{m}_\tau$ and $\bar{s}_\tau$ is illustrated in Figures 3-15 and 3-16.

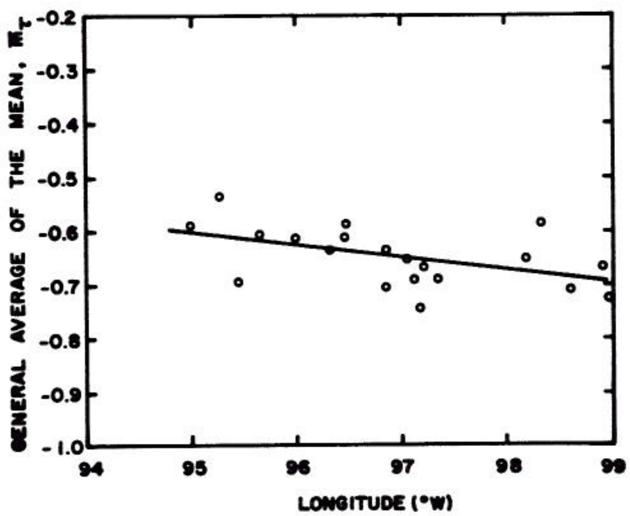

Figure 3-15. Values of the general mean, $\bar{m}_\tau$, for all stations as a function of longitude.

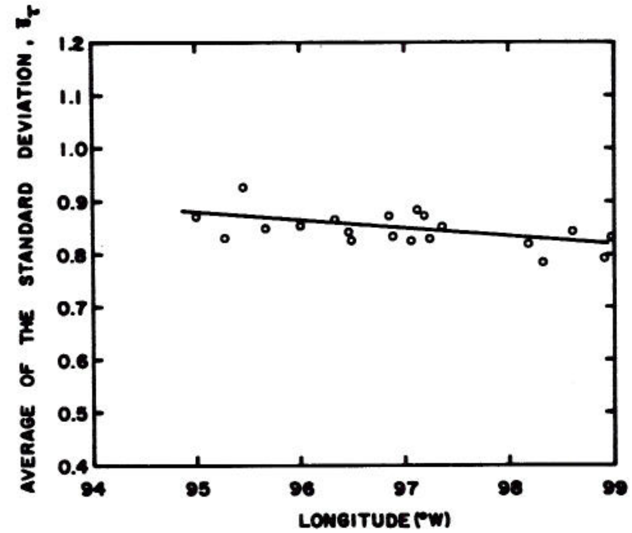

Figure 3-16. Values of the general standard deviation, $\bar{s}_\tau$, for all stations as a function of longitude.

For the mean, all six harmonics had significant amplitudes; however, the amplitude and phase angles of the third (4 month) and the sixth (2 month) harmonics were regional constants. The third and fifth harmonics of the daily standard deviations were not significantly different from zero, while the amplitude and phase of the sixth harmonic of the standard deviations were regional constants.

The regional description of the Fourier series representation of the periodic mean and standard deviation is illustrated for two stations in Figure 3-17. The Fourier series representations are shown using the coefficients for the first six harmonics obtained from the station data (Table 3-5 and 3-6) and the coefficients obtained for the station using the regionalization equations given in Table 3-8. The two stations represent those on the east and west sides of the study region, respectively. The representation of m_τ and s_τ, obtained using the coefficients from the station m_τ and s_τ values, contain more pronounced harmonics than that obtained from the regionalization equations. The pronounced harmonics are due largely to sampling error. The representation of m_τ and s_τ, obtained using the regionalization equations, are a more realistic approximation of the periodicity of m_τ and s_τ.

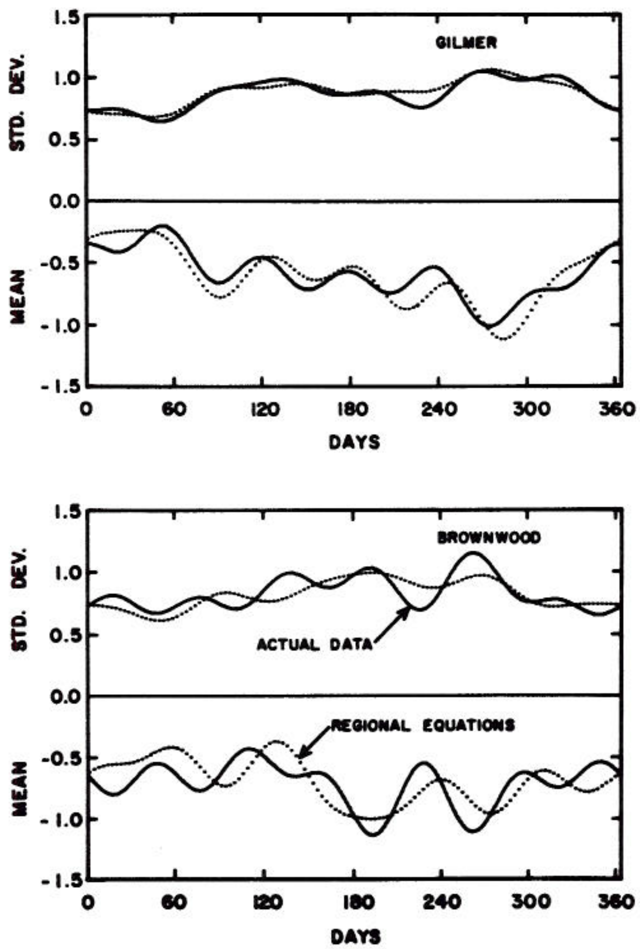

Figure 3-17. Fourier series of m_τ and s_τ for two stations using the coefficients from actual data and the coefficients from the regional equations.

7. Test of Stationarity of $\varepsilon_{p,\tau}$ Series

The stationary random component, $\varepsilon_{p,\tau}$, for each station was determined by removing the periodic mean and standard deviation. The equation used to determine $\varepsilon_{p,\tau}$ was

$$\varepsilon_{p,\tau} = \frac{y_{p,\tau} - \left\{\bar{m}_\tau + \sum_{j=1}^{6} C_j(m) \cos\left[\frac{2\pi jt}{365} + \theta_j(m)\right]\right\}}{\bar{s}_\tau + \sum_{j=1}^{6} C_j(m) \cos\left[\frac{2\pi jt}{365} + \theta_j(s)\right]} \qquad (3\text{-}22)$$

Equation (3-22) was applied only to the nonzero values of the $y_{p,\tau}$ series. The zero values were not changed. The Fourier coefficients for equation (3-22) were determined for each station using the latitude and longitude of the station and the equations given in Table 3-8. The resulting $\varepsilon_{p,\tau}$ series for each station should be second-order stationary (no harmonics in the mean and standard deviation) with a mean of zero and a standard deviation of unity.

The $\varepsilon_{p,\tau}$ series for each of the eight test stations given in Table 3-3 were used to determine if the $\varepsilon_{p,\tau}$ series were stationary in the means and standard deviations with zero means and unity standard deviations. Values of $m_\tau(\varepsilon)$ and $s_\tau(\varepsilon)$ (mean and standard deviation of the $\varepsilon_{p,\tau}$ series for day τ) for all 365 days of the year were determined from the $\varepsilon_{p,\tau}$ series, using the maximum likelihood estimation technique for a truncated normal distribution. The truncation point of the $\varepsilon_{p,\tau}$ series must be known to estimate $m_\tau(\varepsilon)$ and $s_\tau(\varepsilon)$. The truncation point of the $y_{p,\tau}$ series (y_0) was constant for all days of the year and all stations at a value of 0.10 (square root of 0.01 inch of precipitation). However, the truncation point of the $\varepsilon_{p,\tau}$ series was variable depending on the day of the year and the location of the station. The truncation point was determined by

$$\varepsilon_o^\tau = \frac{y_0 - m_\tau}{s_\tau} \qquad (3\text{-}23)$$

or

$$\varepsilon_o^\tau = \frac{0.10 - \left\{\bar{m}_\tau + \sum_{j=1}^{6} C_j(m) \cos\left[\frac{2\pi jt}{365} + \theta_j(m)\right]\right\}}{\bar{s}_\tau + \sum_{j=1}^{6} C_j(m) \cos\left[\frac{2\pi jt}{365} + \theta_j(m)\right]} \qquad (3\text{-}24)$$

where ε_o^τ denotes the truncation point for day τ for the $\varepsilon_{p,\tau}$ series at a given station. The ε_o^τ values for the stations in the study region ranged from 0.48 to 1.19, depending on the location of the station and the day of the year.

The amplitudes and phases of the harmonics of the $m_\tau(\varepsilon)$ and $s_\tau(\varepsilon)$ values for the eight stations were calculated, and the portion of the variance of each explained by the first six harmonics was determined. If the $\varepsilon_{p,\tau}$ series were stationary in the mean and standard deviation the portion of the variance of $m_\tau(\varepsilon)$ and the portion of the variance of $s_\tau(\varepsilon)$ explained by the first six harmonics should be less than the $p_{min}(m_\tau)$ and $p_{min}(s_\tau)$ values given in equations (3-17) and (3-19), respectively. The portion of the variance of $m_\tau(\varepsilon)$ and the portion of the variance of $s_\tau(\varepsilon)$ explained by the first six harmonics for the eight test stations are shown in Table 3-9. The general mean, $\bar{m}_\tau(\varepsilon)$, of $m_\tau(\varepsilon)$ and the general mean, $\bar{s}_\tau(\varepsilon)$, of $s_\tau(\varepsilon)$ are also given in Table 3-9.

Table 3-9. The general mean of $m_\tau(\varepsilon)$, the general mean of $s_\tau(\varepsilon)$, and the variance of $m_\tau(\varepsilon)$ and $s_\tau(\varepsilon)$ explained by the first six harmonics for eight stations.

		General means		Explained variance	
No.	Station Name	$\bar{m}_\tau(\varepsilon)$	$\bar{s}_\tau(\varepsilon)$	$m_\tau(\varepsilon)$	$s_\tau(\varepsilon)$
1	Breckenridge	0.004	0.978	0.064	0.032
2	Brownwood	-0.062	1.037	0.058	0.050
5	Corsicana	-0.034	1.014	0.064	0.067
6	Crockett	-0.099	1.069	0.032	0.048
7	Dallas	-0.091	1.039	0.019	0.032
11	Gilmer	0.009	0.995	0.063	0.020
12	Hewitt	-0.088	1.031	0.035	0.038
19	Temple	-0.043	1.017	0.042	0.035
	Average	-0.051	1.023		

The value of $\bar{m}_\tau(\varepsilon)$ for each of the eight stations was near zero. The average value was -0.051. Similarly, the $\bar{s}_\tau(\varepsilon)$ values were near unity with an average for the eight stations of 1.023. The portion of the variance of $m_\tau(\varepsilon)$ explained by the first six harmonics was less than the $p_{min}(m_\tau)$ value of 0.099 for all eight stations. The portion of the variance of $s_\tau(\varepsilon)$ explained by the first six harmonics was also less than the $p_{min}(s_\tau)$ value of 0.070 for all eight stations. Therefore, the $\varepsilon_{p,\tau}$ series for the 20 stations in the region, produced by using equation (3-22), were stationary in the mean and standard deviation with a mean of zero and a standard deviation of unity. Since the $y_{p,\tau}$ values for any fixed τ were shown to be almost normally distributed, the $\varepsilon_{p,\tau}$ values for any fixed τ are also normally distributed.

8. Autocorrelation of the $\varepsilon_{p,\tau}$ Series

The $\varepsilon_{p,\tau}$ series are approximately stationary in the mean and standard deviation with a mean of zero, a standard deviation of unity, and a variable truncation point. The $\varepsilon_{p,\tau}$ series may be either dependent or independent in sequence. In general, $\varepsilon_{p,\tau}$ will be dependent in sequence because of the persistence that exists in daily precipitation. The autocorrelation coefficients, ρ_k', of the truncated series may be estimated using equation (1-3) and considering only cases when $\varepsilon_{p,\tau}$ and $\varepsilon_{p,\tau+k}$ are nonzeros. The autocorrelation coefficients, ρ_k, for the total, untruncated series may then be approximated using the relationships given in Figure 2-5.

a. Periodicity of ρ_k. The ρ_k values may be either periodic or constant throughout the year. The sample estimates, r_k, of ρ_k for lag one and lag two were investigated for periodicity. The daily r_1 and r_2 values were calculated by 7-day periods of the year for selected $\varepsilon_{p,\tau}$ series. A Fourier series with 26 harmonics was fitted to the 52 values of r_1 and the 52 values of r_2. The percent of the variance of r_1 and r_2 explained by the harmonics was determined. The periodograms were then plotted. The periodograms for two stations are shown in Figures 3-18 and 3-19. The periodograms of r_1 and r_2 for both stations are near the straight line that is expected for a nonperiodic parameter. For further analysis of the autocorrelation of $\varepsilon_{p,\tau}$, ρ_k will be assumed nonperiodic, and only one estimate of ρ_k will be determined for a given $\varepsilon_{p,\tau}$ series.

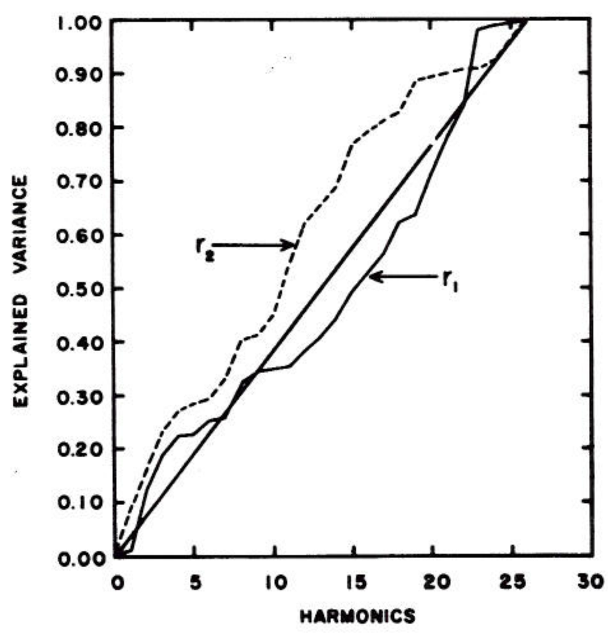

Figure 3-18. Periodograms of the autocorrelation coefficients for lag-one and lag-two of daily precipitation by 7-day periods, Breckenridge, Texas.

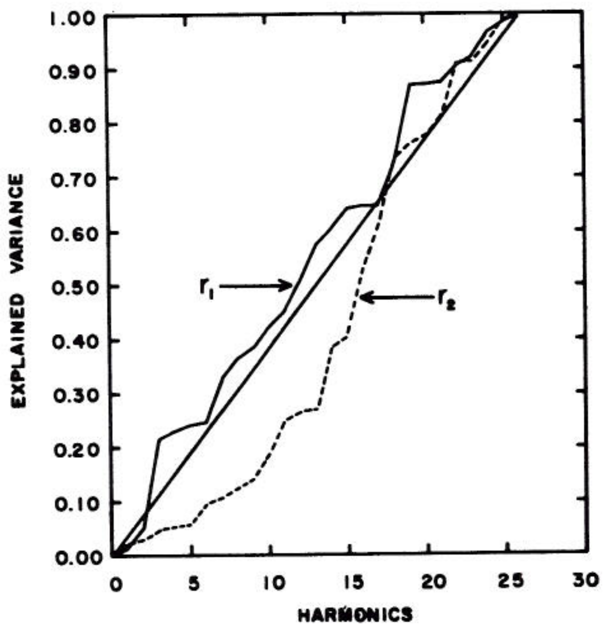

Figure 3-19. Periodograms of the autocorrelation coefficients for lag-one and lag-two of daily precipitation by 7-day periods, Corsicana, Texas.

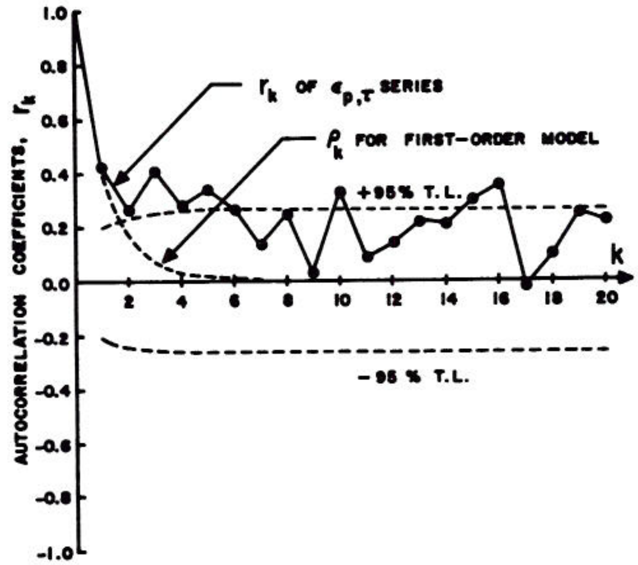

Figure 3-20. Correlogram for Fort Worth, Texas, showing r_k of the $\epsilon_{p,\tau}$ series, the 95-percent tolerance limits, and ρ_k for the first-order linear model.

b. Values of ρ_k. Only the lag-one autocorrelation coefficients are needed to define the elements of the M_1 matrix used in the multivariate generation equation, equation (2-25). However, to investigate the autocorrelation structure of the daily precipitation series, the r_k' values of the truncated series were computed up to lag 20 for several $\epsilon_{p,\tau}$ series. Again, only the cases when $\epsilon_{p,\tau}$ and $\epsilon_{p,\tau+k}$ were both nonzeros were considered. The r_k values for the untruncated series were determined from Figure 2-5 using the r_k' values and the average truncation point for the station. A typical correlogram is given in Figure 3-20.

If the first-order autoregressive model is applicable, the autocorrelation coefficient for lag k is determined from the lag-one autocorrelation coefficient by

$$\rho_k = \rho_1{}^k \quad . \qquad (3\text{-}25)$$

The ρ_k values for a first-order model, with ρ_1 estimated by r_1, are also shown in Figure 3-20. The 95 percent tolerance limits for r_k' were computed by

$$r_k'(95\%) = \frac{-1 \pm 1.96\sqrt{N_k - 2}}{N_k - 1} \qquad (3\text{-}26)$$

where $r_k'(95\%)$ are the 95 percent tolerance limits for lag k, 1.96 is the standard normal deviate for a two-tail test at the 95 percent significance level, and N_k is the number of cases when both $\epsilon_{p,\tau}$ and $\epsilon_{p,\tau+k}$ were nonzeros. The $r_k(95\%)$ values were determined from the $r_k'(95\%)$ values using Figure 2-5. The 95 percent tolerance limits for $r_k(95\%)$ are also shown in Figure 3-20. Coefficients falling between the tolerance limits are not significantly different from zero. The values of r_k are greater than the ρ_k values for a first-order model for most values of k.

c. Areal variation of ρ_1. The lag-one autocorrelation coefficients of the $\epsilon_{p,\tau}$ series were calculated for all 20 stations in the study region and are given in Table 3-10. The r_1 values vary from 0.229 to 0.532. The r_1 values could be related to position within the region. The linear model given in equation (3-21) was used to determine whether the r_1 values for the 20 stations are related to the latitude and longitude of the stations. The multiple correlation coefficient was only 0.034, and the coefficients of both latitude and longitude were not significantly different from zero. The r_1 values are plotted as a function of station longitude in Figure 3-21. There is no indication of a regional trend in the r_1 values. The variation of r_1 is considered to be sampling variation. Therefore, the lag-one autocorrelation coefficient of the total, untruncated $\epsilon_{p,\tau}$ series is practically a regional constant having a value equal to the average value for the 20 stations, $r_1 = 0.385$.

Table 3-10. Lag-one autocorrelation coefficients of the total, untruncated $\varepsilon_{p,\tau}$ series for the 20 stations in the study region.

No.	Station Name	Autocorrelation coefficient, r_1
1	Breckenridge	0.396
2	Brownwood	0.340
3	Centerville	0.294
4	Comanche	0.433
5	Corsicana	0.395
6	Crockett	0.532
7	Dallas	0.403
8	Dialville	0.313
9	Dublin	0.459
10	Fort Worth	0.426
11	Gilmer	0.343
12	Hewitt	0.335
13	Hillsboro	0.351
14	Kaufman	0.420
15	Lampasas	0.430
16	Mexia	0.333
17	Palestine	0.417
18	Riesel	0.464
19	Temple	0.221
20	Waco	0.396
	Avg.	0.385
	Std. Dev.	0.070

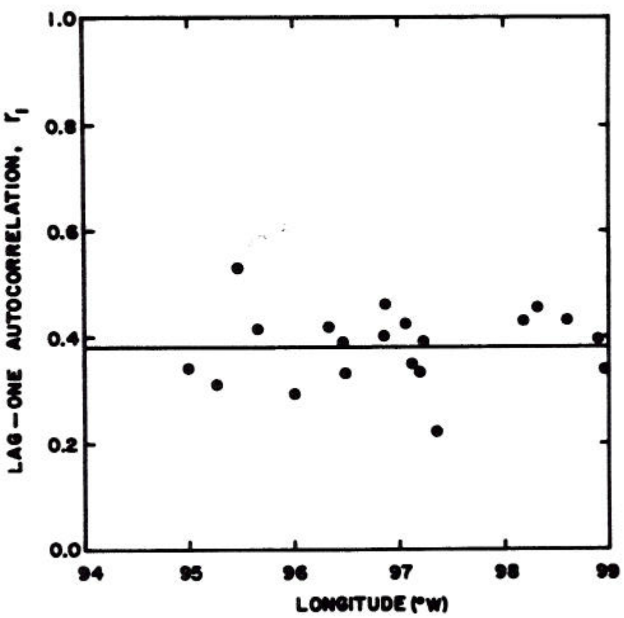

Figure 3-21. Lag-one autocorrelation coefficients of the 20 $\varepsilon_{p,\tau}$ series as a function of station latitude.

9. *Cross Correlation of the $\varepsilon_{p,\tau}$ Series*

The $\varepsilon_{p,\tau}$ series for stations in the region are dependent in space as well as in time. The lag-zero linear cross-correlation coefficients, $\rho_0(i,j)$, express for stations i and j the dependence in space between $\xi_{p,\tau}$ series that are independent in sequence. Assuming the serial dependence of the $\varepsilon_{p,\tau}$ series can be described by a first-order autoregressive model, the $\xi_{p,\tau}$ series are given by

$$\xi_{p,\tau} = \varepsilon_{p,\tau} - \rho_1 \varepsilon_{p,\tau-1} \; . \qquad (3\text{-}27)$$

Equation (3-27) can be applied only when both $\varepsilon_{p,\tau}$ and $\varepsilon_{p,\tau-1}$ are nonzeros. Many of the nonzero $\varepsilon_{p,\tau}$ values are preceded by a zero ($\varepsilon_{p,\tau-1} = 0$). In these cases, the serial dependence in the $\varepsilon_{p,\tau}$ series cannot be removed to yield the corresponding $\xi_{p,\tau}$ values.

Since the independent $\xi_{p,\tau}$ series for daily precipitation could not be determined because of the zeros in the $\varepsilon_{p,\tau}$ series, it was necessary to estimate the lag-zero cross-correlation coefficients from the $\varepsilon_{p,\tau}$ series. The relationship between the cross-correlation coefficients of the serially dependent $\varepsilon_{p,\tau}$ series and the cross-correlation coefficients of the independent $\xi_{p,\tau}$ series was not known. Bivariate samples of $\varepsilon_{p,\tau}$ series were generated that were normally distributed with a mean of zero and a standard deviation of one, dependent in sequence with a first-order autoregressive model, and cross correlated. Forty-four samples were generated with 1000 pairs of $\varepsilon_{p,\tau}$ values in each sample and with a range of lag-one serial-correlation coefficients and lag-zero cross-correlation coefficients. The cross-correlation coefficient was calculated for each generated bivariate sample of $\varepsilon_{p,\tau}$. The serial correlations of the generated $\varepsilon_{p,\tau}$ series were removed, using equation (3-27). The cross-correlation coefficients of the resulting bivariate $\xi_{p,\tau}$ series were calculated. The relationship between the cross-correlation coefficients of the $\varepsilon_{p,\tau}$ series and the cross-correlation coefficients of the corresponding $\xi_{p,\tau}$ series are shown in Figure 3-22. All of the points in Figure 3-22 are near the line of equal values. Apparently, for a bivariate standard normal process with a first-order autoregressive model the cross-correlation coefficients of the independent $\xi_{p,\tau}$ series are equal to the cross-correlation coefficients of the dependent $\varepsilon_{p,\tau}$ series.

The lag-zero cross-correlation coefficients, $\rho_0(i,j)$, were estimated from the $\varepsilon_{p,\tau}$ series for each station in the region. As stated in Chapter II, the $\rho_0(i,j)$ values could be related to position within the region, distance between stations i and j, and orientation of the line connecting stations i and j. Since the $\varepsilon_{p,\tau}$ series contain numerous zeros, the sample estimate of the cross-correlation coefficients of the truncated series, $r_0'(i,j)$, may be estimated considering only cases where $\varepsilon_{p,\tau}$ for both station i and station j are nonzeros. The sample estimates of the cross-correlation coefficients for the untruncated series may be estimated from $r_0'(i,j)$, using Figure 2-5 and the average truncation point for the two stations.

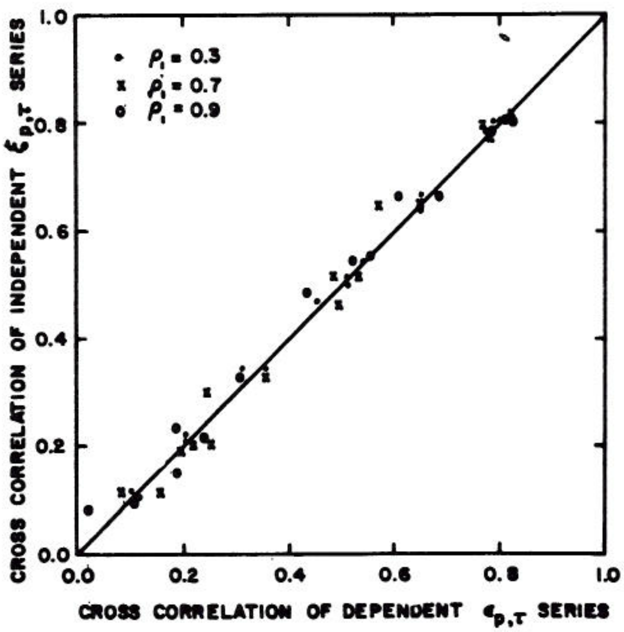

Figure 3-22. Relationship between the cross-correlation coefficients of serially dependent $\epsilon_{p,\tau}$ series and the cross-correlation coefficients of independent $\xi_{p,\tau}$ series.

With the 20 stations used in this study, there are 190 combinations for which the cross-correlation coefficients may be determined. Sixty-four combinations were selected for the determination of $r_0(i,j)$. The cross-correlation coefficients of the truncated series, $r_0'(i,j)$, were computed for the 64 combinations, using the $\epsilon_{p,\tau}$ data for each pair of stations. The $r_0(i,j)$ values were determined by using Figure 2-5 (Table 3-11).

Four parameters that describe the relative position of stations i and j and the position within the region were determined for each pair of stations. The parameters are: (1) latitude of the midpoint between stations i and j in degrees north, $L_1(i,j)$; (2) longitude of the midpoint between stations i and j in degrees west, $L_2(i,j)$; (3) distance between stations i and j in miles, d; and (4) azimuth about the midpoint in degrees, β. The four parameters are illustrated by the definition sketch given in Figure 3-23. The values of the four parameters for the 64 pairs of stations are given in Table 3-11.

To determine the important variables for describing the variation of $r_0(i,j)$, the simple linear correlation coefficient, r, was calculated for $r_0(i,j)$ against each of the four independent variables given in Table 3-11. The r values were:

Table 3-11. Lag-zero cross correlation, location of midpoint, inter-station distance, and azimuth for 64 pairs of stations in the study region.

Stations	Cross Correlation $r_0(i,j)$	Midpoint $L_1(i,j)$ (°N)	Midpoint $L_2(i,j)$ (°W)	Distance d (miles)	Azimuth (degrees)
1- 2	0.729	32.24	98.94	72.7	3.9
1- 3	0.419	32.02	97.44	209.9	119.6
1- 4	0.708	32.33	98.75	63.0	162.7
1- 9	0.718	32.43	98.61	58.4	142.4
1-10	0.657	32.80	97.97	115.7	87.9
1-12	0.511	32.11	98.04	141.0	130.3
1-15	0.512	31.91	98.54	127.1	159.3
1-17	0.492	32.27	97.27	214.3	108.6
1-18	0.591	32.11	97.89	155.7	125.8
2- 4	0.821	31.81	98.79	26.8	62.4
2- 8	0.554	31.79	97.13	232.1	87.4
2- 9	0.874	31.91	98.65	48.4	57.1
2-10	0.623	32.27	98.01	142.9	57.5
2-12	0.669	31.58	98.08	114.0	99.4
2-14	0.696	32.15	97.65	176.5	70.3
2-18	0.579	31.58	97.93	132.6	98.1
3- 5	0.799	31.68	96.22	63.5	151.8
3- 6	0.858	31.29	95.71	33.2	86.4
3-11	0.729	32.00	95.48	118.7	31.8
3-12	0.666	31.36	96.58	76.0	99.4
3-14	0.747	31.93	96.15	93.0	166.8
3-16	0.832	31.47	96.23	42.2	132.2
4- 5	0.584	31.99	97.53	134.3	84.7
4- 9	0.826	32.00	98.46	21.8	50.7
4-12	0.771	31.67	97.89	94.0	109.3
4-15	0.697	31.47	98.39	64.3	155.9
4-17	0.725	31.84	97.13	184.6	92.6
4-20	0.731	31.76	97.91	88.4	102.6
5- 6	0.725	31.69	95.96	83.0	130.5
5- 8	0.732	31.97	95.87	75.8	101.0
5- 9	0.710	32.09	97.40	116.9	90.7
5-10	0.501	32.46	96.76	63.6	144.6
5-11	0.739	32.40	95.72	102.8	64.1
5-14	0.864	32.33	96.39	35.6	14.2
5-17	0.640	31.93	96.05	54.7	112.3
6- 8	0.765	31.58	95.36	41.0	15.9
6-14	0.683	31.94	95.88	103.8	148.4
7- 8	0.579	32.36	96.06	119.7	124.4
7-10	0.916	32.84	96.95	12.6	83.7
8- 9	0.586	31.99	96.80	191.9	94.8
8-12	0.748	31.66	96.22	122.9	76.3
8-17	0.877	31.82	95.46	24.5	75.3
8-20	0.626	31.74	96.24	123.1	81.9
9-10	0.674	32.46	97.69	94.6	57.8
9-11	0.655	32.41	96.65	213.9	78.3
9-13	0,813	32.06	97.72	75.8	94.2
9-14	0.768	32.34	97.32	129.9	75.2
9-18	0.629	31.77	97.60	101.1	116.4
10-11	0.579	32.78	96.01	129.6	93.1
11-12	0.707	32.09	96.08	163.5	57.3
11-13	0.702	32.38	96.05	142.5	69.9
11-14	0.740	32.65	95.65	84.4	83.0
11-18	0.688	32.09	95.93	148.1	53.3
12-17	0.811	31.61	96.41	98.3	76.6
12-18	0.911	31.45	97.03	18.7	90.0
12-20	0.915	31.53	97.20	12.0	168.0
13-16	0.802	31.85	96.80	46.4	120.4
13-20	0.797	31.82	97.17	28.3	12.7
14-19	0.655	31.84	96.84	120.8	32.2
15-18	0.696	31.25	97.53	85.8	71.2
15-19	0.796	31.07	97.76	52.0	86.2
16-17	0.693	31.73	96.06	52.3	82.4
18-19	0.830	31.27	97.12	38.1	50.5
19-20	0.816	31.36	97.29	36.8	12.7

Independent variable	r
$L_1(i,j)$	-0.281
$L_2(i,j)$	-0.116
d	-0.732
β	-0.273

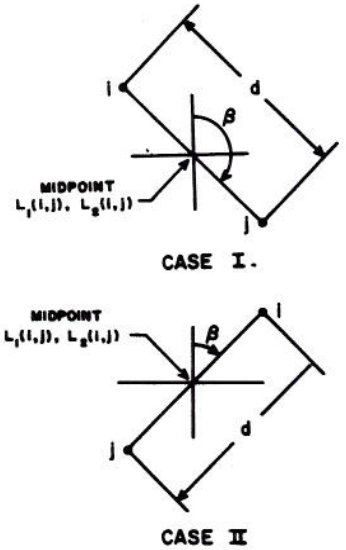

Figure 3-23. Definition of the four parameters used to describe the relative positions of stations i and j and the position of the midpoint within the region. The parameters are: (1) latitude of the midpoint, $L_1(i,j)$; (2) longitude of the midpoint, $L_2(i,j)$; (3) inter-station distance, d; (4) azimuth, β.

The parameters describing position of the midpoint and orientation of the line connecting stations i and j are not as highly correlated with $r_0(i,j)$ as inter-station distance. The correlation coefficient for $L_1(i,j)$, $L_2(i,j)$, and β are not significant at the 1 percent level, while the correlation coefficient for d is significant at the 1 percent level. Therefore, $\rho_0(i,j)$ will be assumed independent of $L_1(i,j)$, $L_2(i,j)$ and β, and a relationship between $\rho_0(i,j)$ and d will be determined.

One of the functions investigated by Yevjevich and Karplus (1973) was chosen for use in this study. The form of the equation is given by

$$\rho_0(i,j) = (1 + \alpha_1 d)^{\alpha_2} \qquad (3\text{-}28)$$

where α_1 and α_2 are coefficients that must be determined. The values of $\rho_0(i,j)$ from equation (3-28) have the desirable characteristics of $\rho_0(i,j) = 1$ for $d = 0$ and $\rho_0(i,j) = 0$ for $d = \infty$ for values of α_2 less than zero. The lag-zero cross-correlation coefficient should be zero for $d = \infty$ because the random components of daily precipitation for two widely spaced stations should be independent.

A non-linear optimization technique that minimizes the error sum-of-squares for an arbitrary function and a given data set was used to determine α_1 and α_2 for equation (3-28), using the $r_0(i,j)$ and d data given in Table 3-11. The resulting equation is

$$\rho_0(i,j) = (1 + 0.0028d)^{-1.44} \qquad (3\text{-}29)$$

Equation (3-29) explained 44 percent of the variance of the $r_0(i,j)$ data. The data and the fitted equation are plotted in Figure 3-24.

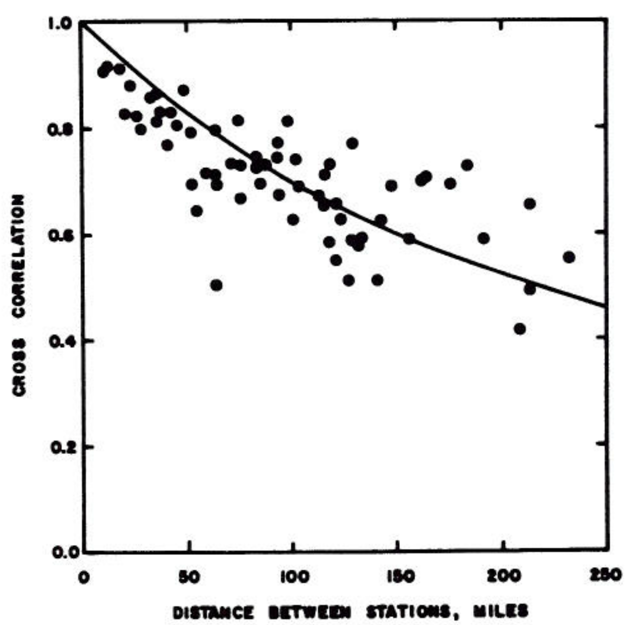

Figure 3-24. Relationship between lag-zero cross correlation of the $\varepsilon_{p,\tau}$ series and inter-station distance for 64 pairs of stations in the study region.

CHAPTER IV
Simulation of Daily Precipitation Over an Area

The purpose of this chapter is to test the model and regional description of parameters by generating daily precipitation sequences and comparing these new sequences to observed sequences.

There are many characteristics of daily precipitation that one could attempt to preserve in generating new sequences. It is not within the scope of this study to test for resemblance between new and observed sequences in all of these characteristics. However, preservation of certain important characteristics is necessary for the new sequences to be useful for water resource applications.

The time-area precipitation model was developed to preserve several stochastic-deterministic characteristics of an assumed truncated multivariate normal process. The characteristics that the model was developed to preserve include: (1) the lag-one autocorrelation coefficient of the random component for each station, (2) the lag-zero cross-correlation coefficient between the independent random components of each station, and (3) the periodic means and standard deviations of the normalized daily precipitation values. The model and the generation procedure dictate that, within the accuracy of the regional description of the parameters, these characteristics will be preserved for any station or group of stations within the study region.

If the model reasonably describes the daily precipitation process over an area, the new sequences must closely resemble observed sequences in terms of several other important characteristics, even though the characteristics were not modeled directly. For example, the distribution of annual precipitation amounts did not enter into the formulation of the model. However, the ability of the model to generate new sequences that preserve the distribution of annual precipitation at any point in the region is essential to a realistic description of the precipitation regime of the region. Similarly, the distribution of precipitation amounts for each month was not modeled directly but must be preserved in new sequences to describe the seasonal characteristics of precipitation. Parameters expressing the dependence of precipitation in time and space, other than the parameters used in developing the model, should also be preserved in the new sequences. With these general criteria in mind, several statistical parameters were selected for comparing new sequences to observed sequences. These parameters are by no means an exhaustive list of the characteristics that should be preserved in generating new precipitation sequences. However, the parameters that were chosen permit comparing new sequences to observed sequences in terms of the distribution of precipitation amounts, dependence in time, and dependence in space, using parameters that were not used in developing the daily precipitation model. The parameters that are used to compare observed and generated sequences are: (1) distribution of annual precipitation values for each station, (2) distribution of monthly (or 28-day) precipitation values for each station, (3) the probability of occurrence of a wet day for any day or season of the year, (4) the cross correlation of annual precipitation values among stations, and (5) the cross correlation of monthly (or 28-day) precipitation amounts among stations.

Two areas within the study region were chosen for use in testing the time-area daily precipitation model. The locations of the two areas are shown in Figure 4-1. In practical application these areas could be watersheds or other entities, for which daily precipitation data over an area are needed. New daily precipitation sequences were generated for specific locations within each test area. Only the latitude and longitude of each station were used to define the model parameters. The new sequences were compared with the observed sequences in terms of the above characteristics.

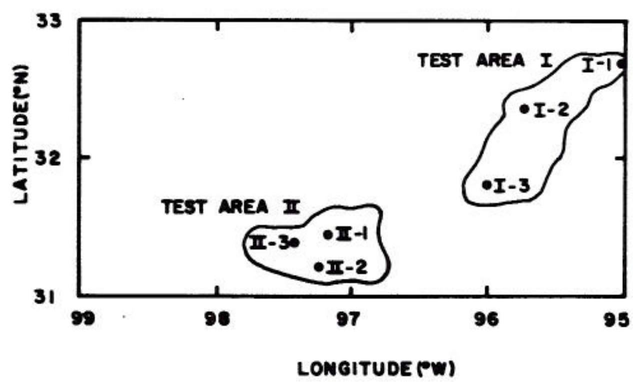

Figure 4-1. Location of two areas chosen for testing the time-area daily precipitation model.

1. *Test Area I*

Test Area I is located in the northeast part of the study region. Three precipitation stations were chosen for which daily precipitation data would be generated. Station I-1 was chosen to coincide with the Gilmer station that was used in developing the model parameters. The other two stations were chosen to coincide with precipitation stations that were not used in developing the model parameters. The three test stations were chosen at sites of actual precipitation stations so that the new sequences could be compared with observed sequences. A description of the precipitation stations in Test Area I is given in Table 4-1. The inter-station distances for the stations ranged from 31.1 miles to 89.8 miles.

Fifty years of daily precipitation data were generated for the three stations in Test Area I. The μ_τ and σ_τ values that were used in generating the data were described by Fourier series, using the regional description of the Fourier coefficients given in Table 3-8. The lag-one autocorrelation coefficient of the random component was shown in Chapter III to be a regional constant with a value of

Table 4-1. Description of precipitation stations in Test Area I.

No.	Station Name	Index No.	Latitude (deg. north)	Longitude (deg. west)	Available data (years)
I-1	Gilmer	3546	32.73	94.98	1933-72
I-2	Lindale	5228	32.45	95.37	1932-64
I-3	Long Lake	5327	31.48	95.78	1933-72

Stations	Inter-station distance (miles)
Gilmer-Lindale	31.1
Gilmer-Long Lake	89.8
Lindale-Long Lake	60.9

0.385; therefore, r_1 for each station was assumed to be 0.385. The lag-zero cross-correlation coefficients were computed using the inter-station distances and equation (3-28).

a. Values of m_τ and s_τ for the new sequences. The new daily precipitation data were analyzed in the same manner as the observed data to determine if the generation procedure was producing sequences with the desired m_τ and s_τ. The generated data for the three stations in Test Area I were normalized using the square root transformation. The mean and standard deviation of the transformed data were calculated for each day of the year, using the maximum likelihood estimation technique for a truncated normal distribution. The Fourier coefficients were calculated for the first six harmonics of the daily values of m_τ and s_τ. The Fourier series representation of m_τ and s_τ determined for the generated data are shown in Figures 4-2, 4-3, and 4-4 for the three stations in Test Area I. The m_τ and s_τ values that were obtained from the regionalization equations in Table 3-8 and used in the generation program are also shown in the figures. The Fourier series representation of the m_τ and s_τ values from the generated data approximate the values that were input to the generation procedure. The harmonics of m_τ and s_τ obtained from the generated data are more pronounced than the harmonics of m_τ and s_τ using the regional equations due to sampling error.

A Fourier series with six harmonics explained 20.8 percent of the variance of m_τ and 14.1 percent of the variance for s_τ from the generated data for Gilmer. Using the observed data for Gilmer, six harmonics explained 14.6 percent of the variance of m_τ and 12.2 percent of the variance of s_τ. The sampling random components of m_τ and s_τ from the generated data were about the same as that from the observed data. The generation model seems to produce, within the accuracy of the regional equations, new sequences with m_τ and s_τ having the same periodic characteristics and random variations as the observed data.

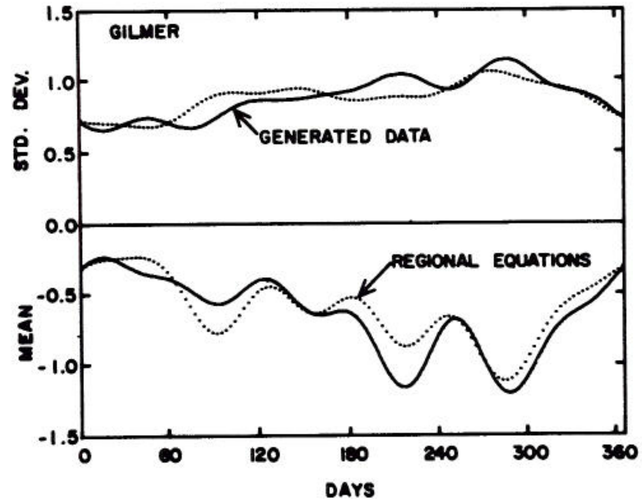

Figure 4-2. Fourier series representation of m_τ and s_τ determined from the generated data and obtained from the regional equations for Gilmer, Texas.

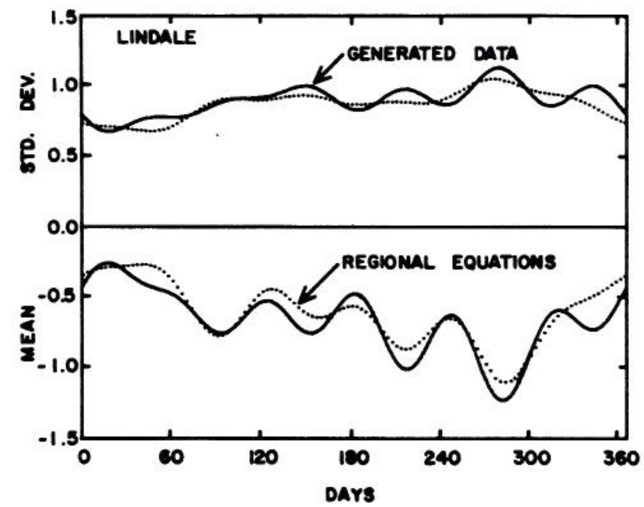

Figure 4-3. Fourier series representation of m_τ and s_τ determined from the generated data and obtained from the regional equations for Lindale, Texas.

b. Autocorrelation of $\varepsilon_{p,\tau}$ from generated data. The random component, $\varepsilon_{p,\tau}$, of the new sequences for each of the three stations was determined by removing the periodic m_τ and s_τ from the transformed data. The Fourier series description of m_τ and s_τ shown above were used to define m_τ and s_τ values for each day. The $\varepsilon_{p,\tau}$ values were determined using equation (3-22). The lag-one autocorrelation coefficients, r_1', of the truncated new series were estimated using

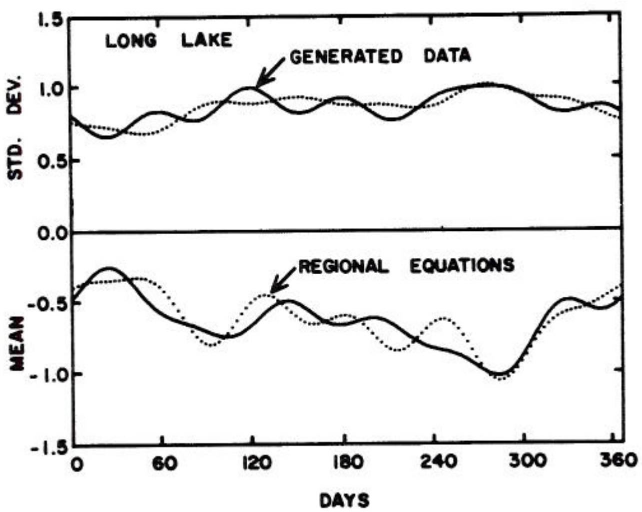

Figure 4-4. Fourier series representation of m_τ and s_τ determined from the generated data and obtained from the regional equations for Long Lake, Texas.

equation (1-3) considering only cases when $\varepsilon_{p,\tau}$ and $\varepsilon_{p,\tau+1}$ were nonzeros. The autocorrelation coefficients, r_1, for the total, untruncated series were approximated by using r_1' and Figure 2-5. The r_1 values of the generated sequences for each station are given in Table 4-2. The three values are not significantly different from the assumed regional constant of 0.385.

Table 4-2. Lag-one autocorrelation coefficients of the new sequences for the three stations in Test Area I.

No.	Station	r_1'	r_1
I-1	Gilmer	0.106	0.343
I-2	Lindale	0.121	0.378
I-3	Long Lake	0.139	0.417
		Average	0.378

c. Cross correlation of $\varepsilon_{p,\tau}$ *from generated data.* The lag-zero cross-correlation coefficients of the dependent $\varepsilon_{p,\tau}$ series for each of the three pairs of stations in Test Area I were examined. The cross-correlation coefficients for the truncated series, $r_0'(i,j)$, were estimated considering only cases where $\varepsilon_{p,\tau}$ for both stations were nonzeros. The estimates of the cross correlation coefficients for the untruncated series were determined by using $r_0'(i,j)$ and Figure 2-5. The cross-correlation coefficients, determined from the generated data, are compared in Table 4-3 with those calculated with equation (3-28) for use in generating the data. For two of the three pairs of stations, the $r_0(i,j)$ values determined from the generated data were not significantly different (5 percent level) from the $r_0(i,j)$ values obtained by equation (3-28). Therefore, it may be concluded that the generation model produces new sequences with the desired lag-zero cross-correlation coefficients of the random components for the stations within the area.

Table 4-3. Lag-zero cross correlation for the stations in Test Area I from (1) the regional model, equation (3-29), and (2) the generated data.

Stations	Distance (miles)	Lag-zero cross-correlation coefficient: Equation 3-29	Generated data
Gilmer-Lindale	31.1	0.887	0.905[1/]
Gilmer-Long Lake	89.8	0.724	0.767
Lindale-Long Lake	60.9	0.797	0.829

1/ The cross-correlation coefficient of the generated data is significantly different (5 percent level) from the cross correlation coefficient obtained by equation (3-29).

d. Distribution of 28-day and annual totals. The underlying assumption of this study is that the time-area structure of the square root of daily precipitation may be approximated by a truncated multivariate normal distribution. No assumption is made about the distribution of precipitation for time intervals longer than 1 day. However, if the assumption of a truncated multivariate normal distribution for daily precipitation is a good description of the physical process, the sequences generated using the model should produce precipitation amounts for intervals longer than 1 day that are distributed in a similar way to that of the observed data. Interval lengths of 28 days and 365 days (1 year) were chosen to compare the distributions of the generated data with the observed data for intervals longer than 1 day. The total precipitation for each 28-day period of the year and the annual totals were determined for both the observed and generated sequences for the three stations in Test Area I. Three parameters were chosen to describe the distributions of the 28-day and annual totals: (1) the mean, (2) the standard deviation, and (3) the skewness coefficient. The values of these three parameters for both the observed and generated sequences were calculated for each 28-day period of the year and for the annual totals. The results are shown for the three stations in Tables 4-4, 4-5, and 4-6. The means, standard deviations, and skewness coefficients from the generated data compared well with those from the observed data. The means and the standard deviations for each 28-day period and for the 1 year were tested to determine if the parameters for the generated data differed significantly from the parameters for the observed data. A standard t-test for the equivalence of two means from normal populations was used to determine if the means were significantly different. The distributions were not normal, as shown by the positive skewness coefficients. However, most of the skewness coefficients were relatively small so that the assumption of normal populations for the purpose of testing the hypothesis that the means were equal did not introduce any great error. An F-ratio test for the equivalence of two variances from normal populations was used to determine if the standard deviations were significantly different.

Table 4-4. Means, standard deviations, and skewness coefficients of observed and generated 28-day and annual precipitation, Gilmer, Texas.

Period	Mean Observed (inches)	Mean Generated (inches)	Standard deviation Observed (inches)	Standard deviation Generated (inches)	Skewness coefficient Observed	Skewness coefficient Generated
1	3.06	3.18	1.98	2.24	0.35	1.39
2	3.55	3.51	1.70	3.01	0.17	1.56
3	3.10	2.15 1/	1.78	1.86	0.70	2.07
4	3.49	2.74	2.21	2.10	0.89	0.91
5	5.41	5.23	3.19	4.39	0.86	1.54
6	3.60	3.60	2.70	2.88	0.88	2.27
7	2.84	3.91	2.17	4.03	0.78	2.18
8	2.46	1.89	2.71	1.60 2/	1.53	1.13
9	2.82	2.49	2.23	1.75	2.06	0.73
10	2.67	3.49	2.05	3.53	0.71	1.78
11	3.07	2.36	2.25	2.65	0.89	1.39
12	3.82	4.29	2.43	3.65	0.61	1.84
13	3.82	3.44	2.69	2.60	1.05	0.89
Annual	43.71	42.45	10.82	10.41	0.44	0.49

1/ The mean of the generated data is significantly different from the mean of the observed data at the 5 percent level.
2/ The standard deviation of the generated data is significantly different from the standard deviation of the observed data at the 5 percent level.

Table 4-5. Means, standard deviations, and skewness coefficients of observed and generated 28-day and annual precipitation, Lindale, Texas.

Period	Mean Observed (inches)	Mean Generated (inches)	Standard deviation Observed (inches)	Standard deviation Generated (inches)	Skewness coefficient Observed	Skewness coefficient Generated
1	3.36	3.01	2.43	2.02	1.00	0.87
2	3.53	3.19	1.74	2.88	0.59	1.99
3	3.22	2.20 1/	1.97	1.93	1.35	2.98
4	3.72	2.80	2.85	2.47	1.78	1.31
5	5.46	5.18	3.14	4.64	0.70	2.16
6	3.78	4.00	2.57	3.23	1.26	2.06
7	2.79	3.16	2.13	3.15	1.16	2.26
8	2.32	2.15	2.15	2.35	1.60	2.03
9	2.76	2.61	2.48	2.10	2.93	1.01
10	2.57	3.69	1.92	3.56	0.60	1.35
11	3.21	2.48	2.47	2.97	0.97	1.66
12	4.02	4.16	2.47	3.26	0.52	2.04
13	3.87	3.54	2.65	2.41	0.45	0.61
Annual	44.91	42.31	10.63	11.39	0.33	0.39

1/ The mean of the generated data is significantly different from the mean of the observed data at the 5 percent level.

Table 4-6. Means, standard deviations, and skewness coefficients of observed and generated 28-day and annual precipitation, Long Lake, Texas.

Period	Mean Observed (inches)	Mean Generated (inches)	Standard deviation Observed (inches)	Standard deviation Generated (inches)	Skewness coefficient Observed	Skewness coefficient Generated
1	3.09	3.11	1.94	1.87	0.32	1.54
2	3.13	2.97	1.64	2.33	0.02	1.17
3	2.64	2.34	2.06	2.08	0.78	1.95
4	3.34	2.94	2.04	2.89	0.30	2.43
5	4.80	4.89	3.97	4.06	1.68	2.46
6	3.15	3.98	1.75	3.97	0.65	2.41
7	2.50	2.96	2.36	2.80	1.76	1.76
8	1.82	2.23	2.48	2.60	2.66	2.62
9	2.48	2.16	1.71	1.91	0.63	2.04
10	2.81	3.50	2.32	3.27	2.01	1.53
11	3.08	2.11	2.89	2.26	1.94	1.60
12	3.80	3.72	3.55	3.71	2.94	2.31
13	3.25	3.52	1.71	2.77	0.51	0.58
Annual	40.09	40.50	10.51	11.07	0.79	0.15

The mean of the generated data differed significantly (5 percent level) from the mean of the observed data for only one 28-day period for two of the three stations. The means of the generated data for period 3 for Gilmer and Lindale were significantly less than that from the observed data. The hypothesis that the mean precipitation amounts from the observed data and the means from the generated data were from the same population was accepted for all other periods for the three stations. The mean annual precipitation amounts from the generated data were very close to that from the observed data for all three stations. None of the differences in annual means were significant at the 5 percent level.

The standard deviations of the 28-day totals for the generated data differed significantly from the standard deviations of the observed data for only one period for Gilmer. None of the standard deviations of the generated data for Lindale or Long Lake differed significantly from the standard deviations of the observed data.

The periodic means and standard deviations of the 28-day totals from both the observed and generated data are illustrated in Figures 4-5, 4-6, and 4-7. The seasonal patterns of the means and the standard deviations of the generated data corresponded closely with that of the observed data. The largest 28-day mean occurred during period 5 for all stations for both the observed data and the generated data. A smaller peak occurred during periods 12 and 13 for both observed and generated data. The smallest mean 28-day precipitation occurred during period 8. The seasonal pattern of standard deviations were similar to that for the means, except that the peaks were not as pronounced.

e. Probability of a wet day. The probability of a wet day or a dry day at a point is often described by a Markov chain. The Markov chain approach requires the definition of the probabilities of a wet day, given a wet day on the previous day, P(W/W), and a dry day given a dry day on the previous day, P(D/D). The other two transition probabilities, P(D/W) and P(W/D), may be defined from the first two

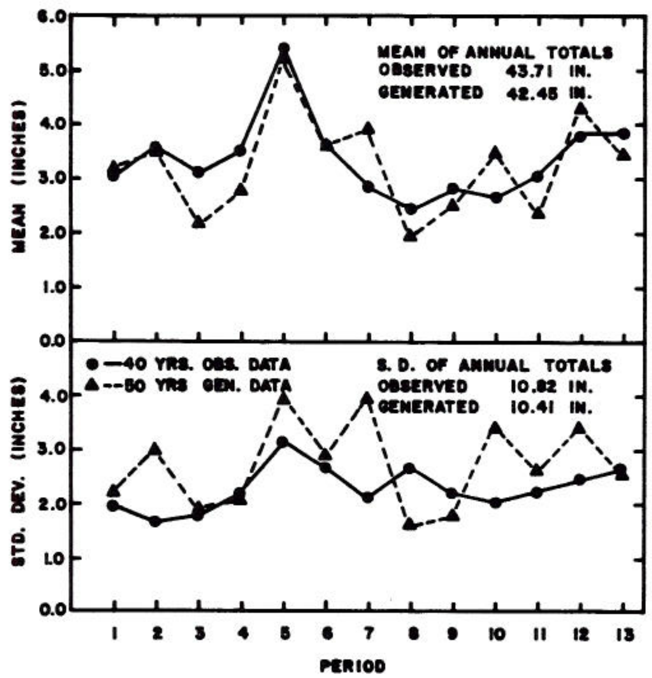

Figure 4-5. Means and standard deviations of the 28-day totals from 40 years of observed data and 50 years of generated data, Gilmer, Texas.

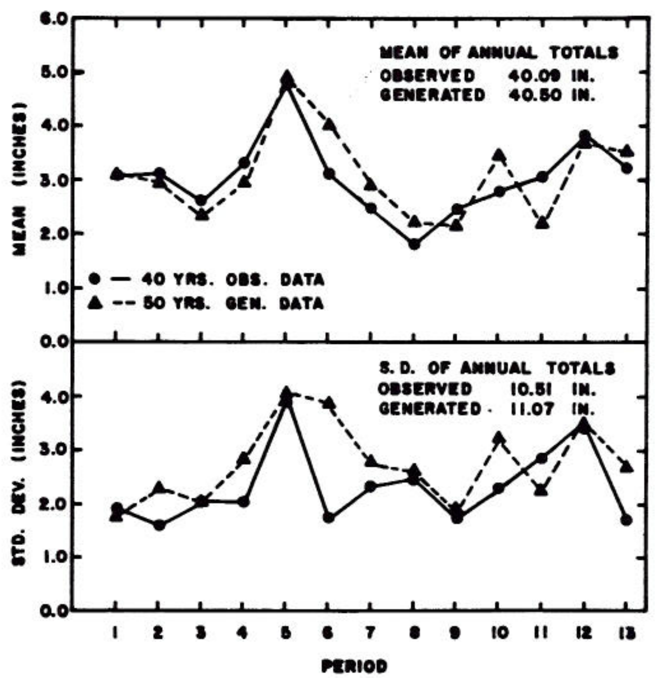

Figure 4-7. Means and standard deviations of the 28-day totals from 40 years of observed data and 50 years of generated data, Long Lake, Texas.

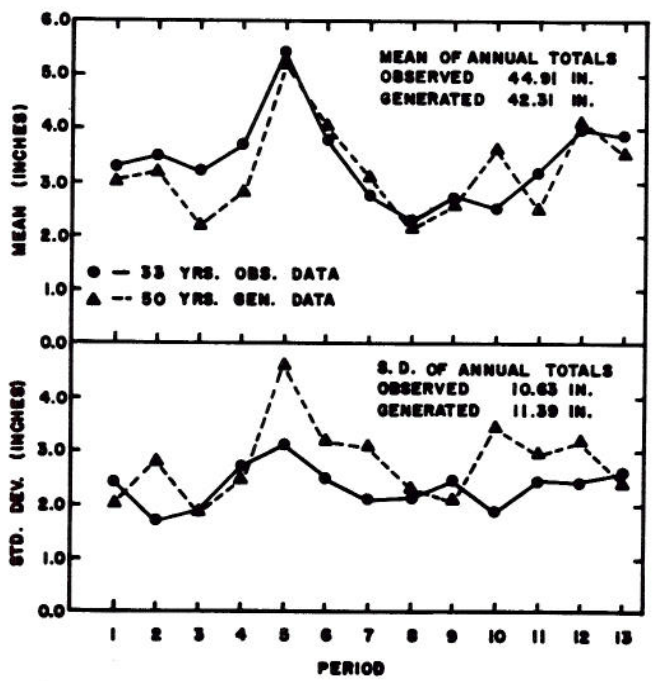

Figure 4-6. Means and standard deviations of the 28-day totals from 33 years of observed data and 50 years of generated data, Lindale, Texas.

probabilities. With the time-area model used in this study, the transition probabilities are not modeled directly. However, if the assumed first order Markov model with a constant lag-one autocorrelation coefficient adequately describes the dependence structure, the generated data should have approximately the same transition probabilities as the observed data.

The generated and observed data for the three stations in Test Area I were analyzed to determine sample estimates of P(W/W) and P(D/D) for each day of the year. The probabilities were grouped by 28-day periods and the average P(W/W) and P(D/D) were determined for each station and each period. The average P(W/W) and P(D/D) values for each 28-day period are given in Tables 4-7, 4-8, and 4-9 for both the generated data and the observed data. The transition probabilities were tested, using the normal approximation to the binomial distribution, to determine if the probabilities obtained from the generated data differed significantly from that determined from the observed data. In general, both P(W/W) and P(D/D) from the generated data were good approximations of that from the observed data. Several of the probabilities from the generated data differed significantly (5 percent level) from that of the observed data. There seemed to be a tendency for P(D/D) from the generated data to be less than P(D/D) from the observed data. However, the probabilities from the observed and generated data were remarkably close considering these probabilities were not modeled directly.

Table 4-7. Markov chain transition probabilities of observed and generated data for 13 28-day periods, Gilmer, Texas.

Period	P(W/W)		P(D/D)	
	Observed	Generated	Observed	Generated
1	0.472	0.449	0.820	0.760 1/
2	0.468	0.477	0.793	0.757 1/
3	0.367	0.355	0.790	0.837 1/
4	0.343	0.359	0.793	0.834 1/
5	0.461	0.479	0.790	0.795
6	0.428	0.372	0.826	0.816
7	0.396	0.439	0.849	0.830
8	0.384	0.333 1/	0.855	0.870
9	0.393	0.447	0.841	0.836
10	0.426	0.323 1/	0.875	0.856
11	0.337	0.305	0.857	0.896 1/
12	0.451	0.438	0.851	0.799 1/
13	0.465	0.442	0.824	0.767 1/

1/ The transition probability from the generated data is significantly different from the observed data at the 5 percent level.

Table 4-8. Markov chain transition probabilities of observed and generated data for 13 28-day periods, Lindale, Texas.

Period	P(W/W)		P(D/D)	
	Observed	Generated	Observed	Generated
1	0.501	0.448 1/	0.809	0.770 1/
2	0.553	0.454 1/	0.770	0.784
3	0.408	0.360	0.780	0.841 1/
4	0.428	0.343 1/	0.792	0.824
5	0.437	0.495 1/	0.768	0.813 1/
6	0.486	0.436	0.809	0.822
7	0.431	0.390	0.862	0.817 1/
8	0.384	0.341	0.852	0.883 1/
9	0.392	0.373	0.841	0.849
10	0.435	0.361 1/	0.872	0.849
11	0.385	0.310 1/	0.857	0.890 1/
12	0.520	0.385 1/	0.842	0.787 1/
13	0.543	0.451 1/	0.819	0.784 1/

1/ The transition probability from the generated data is significantly different from the observed data at the 5 percent level.

f. Cross correlation of 28-day and annual totals. The time-area daily precipitation model developed in this study uses the lag-zero cross-correlation coefficients of the $\varepsilon_{p,\tau}$ series to describe the dependence of daily precipitation in space. It was illustrated above that the cross-correlation coefficients of the daily generated data were not significantly different from that for the observed data. The purpose of this section is to compare the space dependence of the generated and observed data for intervals longer than 1 day.

The cross-correlation coefficients of the 28-day totals and the annual totals were calculated for each of the three pairs of stations in Test Area I using both the generated and the observed data. The results are given in Tables 4-10, 4-11, and 4-12. The cross-correlation coefficients of the 28-day totals of the generated data for the Gilmer and Lindale stations (Table 4-10) were not significantly

Table 4-9. Markov chain transition probabilities of observed and generated data for 13 28-day periods, Long Lake, Texas.

Period	P(W/W)		P(D/D)	
	Observed	Generated	Observed	Generated
1	0.435	0.420	0.833	0.797 1/
2	0.444	0.405	0.801	0.776
3	0.303	0.358 1/	0.821	0.845
4	0.283	0.366 1/	0.841	0.848
5	0.378	0.442	0.832	0.800 1/
6	0.381	0.401	0.855	0.834
7	0.355	0.376	0.892	0.831 1/
8	0.315	0.347	0.912	0.873 1/
9	0.289	0.323	0.866	0.858
10	0.418	0.435	0.885	0.853 1/
11	0.295	0.279	0.880	0.892
12	0.411	0.341	0.862	0.829 1/
13	0.404	0.463	0.841	0.826

1/ The transition probability from the generated data is significantly different from the observed data at the 5 percent level.

different (5 percent level) from the cross-correlation coefficients of the 28-day observed data for 12 of the 13 28-day periods. The cross-correlation coefficient of the annual totals of the generated data was not significantly different from the cross-correlation coefficient of the observed annual totals. Similarly, the cross-correlation coefficients of the 28-day totals of the generated data for the Gilmer and Long Lake stations (Table 4-11) were not significantly different from the cross-correlation coefficients of the observed data for 12 of the 13 28-day periods or for the annual totals. Only two of the cross-correlation coefficients of the 28-day totals of the generated data were significantly different from the observed data for the Lindale and Long Lake stations (Table 4-12).

Table 4-10. Cross correlation coefficients of observed and generated 28-day and annual precipitation, Gilmer and Lindale, Texas.

Period	Cross-correlation coefficient	
	Observed data	Generated data
1	0.920	0.825
2	0.912	0.859
3	0.908	0.862
4	0.907	0.788
5	0.894	0.878
6	0.838	0.885
7	0.667	0.903 1/
8	0.702	0.731
9	0.751	0.758
10	0.744	0.794
11	0.843	0.727
12	0.898	0.852
13	0.913	0.851
Average	0.838	0.824
Annual	0.906	0.812

1/ The cross-correlation coefficient of the generated data is significantly different from the cross-correlation coefficient of the observed data at the 5 percent level.

Table 4-11. Cross correlation coefficients of observed and generated 28-day and annual precipitation, Gilmer and Long Lake, Texas.

Period	Cross-correlation coefficient	
	Observed data	Generated data
1	0.768	0.612
2	0.773	0.719
3	0.600	0.627
4	0.712	0.604
5	0.538	0.775 1/
6	0.488	0.625
7	0.655	0.634
8	0.502	0.583
9	0.658	0.439
10	0.556	0.445
11	0.788	0.701
12	0.714	0.735
13	0.726	0.555
Average	0.652	0.620
Annual	0.760	0.625

1/ The cross-correlation coefficient of the generated data is significantly different from the cross-correlation coefficient of the observed data at the 5 percent level.

Table 4-12. Cross correlation coefficients of observed and generated 28-day and annual precipitation, Lindale and Long Lake, Texas.

Period	Cross-correlation coefficient	
	Observed data	Generated data
1	0.763	0.708
2	0.723	0.786
3	0.619	0.704
4	0.730	0.660
5	0.771	0.818
6	0.481	0.755 1/
7	0.599	0.748
8	0.278	0.863 1/
9	0.572	0.603
10	0.828	0.727
11	0.735	0.763
12	0.763	0.711
13	0.797	0.684
Average	0.666	0.733
Annual	0.817	0.756

1/ The cross-correlation coefficient of the generated data is significantly different from the cross-correlation coefficient of the observed data at the 5 percent level.

Most of the cross-correlation coefficients of the 28-day totals and annual totals from the generated data were not significantly different from the 28-day totals and annual totals of the observed data. The average of the 13 28-day cross-correlation coefficients and the annual cross-correlation coefficients are plotted with respect to inter-station distance in Figure 4-8. The decrease in cross-correlation coefficients with increasing inter-station distance is about the same for the generated and observed data. The time-area daily precipitation model does not directly preserve the cross-correlation coefficients for intervals longer than 1 day. However, by preserving the lag-zero cross-correlation coefficients of the random components the model produces new series with cross correlation coefficients for intervals of 28 days or longer that are close to the cross-correlation coefficients of the observed data.

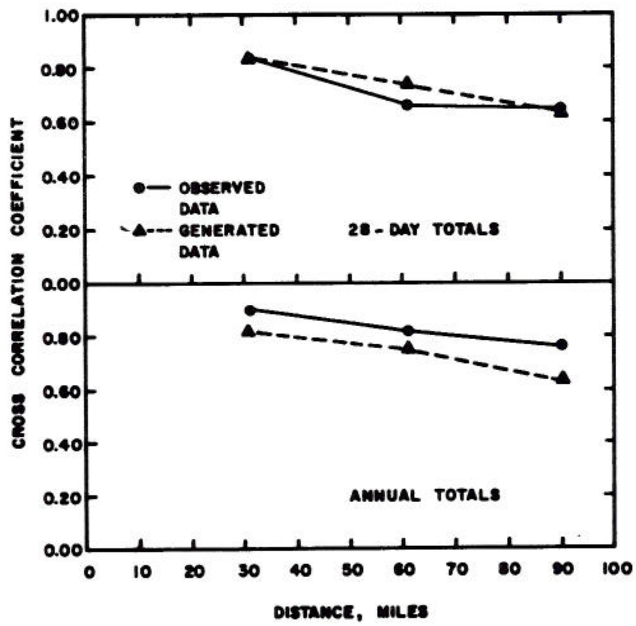

Figure 4-8. Cross correlation coefficients of 28-day totals and annual totals as a function of inter-station distance.

2. Test Area II

Test Area II is in the south-central part of the study region (see Figure 4-1). Three precipitation stations were again chosen for which daily precipitation data would be generated. A description of the three precipitation stations in Test Area II is given in Table 4-13. Station II-1 (Hewitt) is one of the stations used in developing the model parameters. Stations II-2 (Troy) and II-3 (McGregor) were not used in developing the model parameters. Only 20 years of observed data were available for stations II-2 and II-3. The three stations in Test Area II are closer together than the three stations in Test Area I. The inter-station distances for the stations ranged from 15.1 miles to 18.8 miles.

The time-area daily precipitation model was used to generate 50 years of precipitation data for the three stations in Test Area II. The generated data were compared with the observed data. The values of m_τ and s_τ for the new sequences were not computed for Test Area II because of the results for Test Area I (Figures 4-2, 4-3, and 4-4) indicated that the generation model will reproduce the m_τ and s_τ values

Table 4-13. Description of precipitation stations in Test Area II.

No.	Station name	Index No.	Latitude (deg. north)	Longitude (deg. west)	Available data (years)
II-1	Hewitt	4122	31.45	97.18	1933-72
II-2	Troy	9153	31.20	97.30	1953-72
II-3	McGregor	5757	31.43	97.42	1953-72

Stations	Inter-station distance (miles)
Hewitt-Troy	18.8
Hewitt-McGregor	15.1
Troy-McGregor	17.6

obtained from the regionalization equations in Table 3-8. Similarly, the autocorrelation and cross correlation of the $\varepsilon_{p,\tau}$ series from the Test Area II generated data were not investigated. The Test Area I results indicated that the model would produce new sequences with lag-one autocorrelation coefficients near the assumed regional constant. The Test Area I results also demonstrated that the daily precipitation model would generate sequences with lag-zero cross-correlation coefficients near that given by equation (3-29). Since the autocorrelation and cross-correlation coefficients are not a function of regional position, the results apply to Test Area II as well as Test Area I.

The Test Area II generated data were compared with the observed data in terms of the distribution of 28-day and annual totals, the probabilities of wet or dry days, and the cross-correlation coefficients of 28-day and annual totals.

a. Distribution of 28-day and annual totals. The total precipitation for each 28-day period and the annual totals were determined for both the observed and generated data for the three stations in Test Area II. The means, standard deviations, and skewness coefficients were calculated for each 28-day period and for the annual totals, using both the observed and the generated data for each station. The results are shown in Tables 4-14, 4-15, and 4-16 and in Figures 4-9, 4-10, and 4-11. The means and standard deviations of the 28-day totals and the annual totals were tested to determine whether the parameters for the generated data differed significantly from those of the observed data.

None of the three mean annual precipitation amounts from the generated data differed significantly from the mean annual precipitation from the observed data. Only one of the three annual standard deviations differed significantly from the standard deviations of the observed data. The mean annual precipitation for stations in Test Area II is about 10 inches less than that for stations in Test Area I. The time-area daily precipitation model accurately accounts for the differences in mean annual precipitation with position in the region.

The seasonal patterns of the 28-day means and standard deviations from the generated data were, in general, a good representation of the means and standard deviations of the observed data. Most of the means and standard deviations of the 28-day totals from the generated data were not significantly different from that of the observed data. The generated data contained only one 28-day mean for Hewitt, two means for Troy, and one mean for McGregor that were significantly different from the means of the observed data. Similarly, the generated data contained only two 28-day standard deviations for Hewitt, one standard deviation for Troy, and three

Table 4-14. Means, standard deviations, and skewness coefficients of observed and generated 28-day and annual precipitation, Hewitt, Texas.

Period	Mean: Observed (inches)	Mean: Generated (inches)	Standard deviation: Observed (inches)	Standard deviation: Generated (inches)	Skewness coefficient: Observed	Skewness coefficient: Generated
1	1.77	2.48	1.34	2.11	0.78	2.28
2	2.32	2.36	1.43	1.44	0.27	0.69
3	1.87	2.37	1.43	2.08	0.71	1.73
4	2.57	2.20	2.12	1.93	1.79	1.41
5	4.48	3.67	3.62	2.39 1/	1.26	0.74
6	3.01	2.63	1.83	2.48	0.38	2.86
7	1.77	2.77 2/	2.01	2.07	1.72	0.71
8	1.57	1.91	1.96	2.14	2.38	2.37
9	2.86	2.35	2.97	2.03 1/	1.90	0.85
10	2.76	2.84	2.64	2.27	1.48	0.97
11	2.67	2.07	2.10	2.23	0.92	1.67
12	2.38	2.35	2.12	1.88	1.38	0.98
13	1.99	2.17	1.47	2.09	0.90	2.54
Annual	32.17	32.17	8.81	7.84	0.41	0.59

1/ The standard deviation of the generated data is significantly different from the standard deviation of the observed data at the 5 percent level.

2/ The mean of the generated data is significantly different from the mean of the observed data at the 5 percent level.

Table 4-15. Means, standard deviations, and skewness coefficients of observed and generated 28-day and annual precipitation, Troy, Texas.

Period	Mean: Observed (inches)	Mean: Generated (inches)	Standard deviation: Observed (inches)	Standard deviation: Generated (inches)	Skewness coefficient: Observed	Skewness coefficient: Generated
1	1.75	2.20	1.70	2.20	1.64	3.62
2	2.65	2.00	1.50	1.47	0.40	0.84
3	1.69	2.16	1.27	1.67	1.09	1.11
4	2.18	1.87	1.47	1.64	0.25	1.48
5	5.44	3.64 1/	3.67	2.17 2/	0.74	0.39
6	2.51	2.79	1.84	2.69	0.84	2.08
7	1.67	2.20	2.05	2.05	1.41	1.34
8	1.77	2.06	2.03	2.14	2.18	1.41
9	2.42	2.74	1.89	2.43	1.36	2.02
10	3.70	2.87	2.82	2.46	0.87	1.35
11	3.31	2.01 1/	2.96	2.11	1.50	1.83
12	2.37	1.97	1.39	1.36	-0.02	0.47
13	1.92	2.08	1.52	1.86	1.14	1.34
Annual	33.50	30.66	8.88	7.11	-0.14	0.68

1/ The mean of the generated data is significantly different from the mean of the observed data at the 5 percent level.

2/ The standard deviation of the generated data is significantly different from the standard deviation of the observed data at the 5 percent level.

Table 4-16. Means, standard deviations, and skewness coefficients of observed and generated 28-day and annual precipitation, McGregor, Texas.

Period	Mean: Observed (inches)	Mean: Generated (inches)	Standard deviation: Observed (inches)	Standard deviation: Generated (inches)	Skewness coefficient: Observed	Skewness coefficient: Generated
1	1.87	2.31	1.86	2.22	1.59	3.26
2	2.38	2.07	1.44	1.37	0.29	0.64
3	1.65	2.45	1.39	1.82	0.99	0.67
4	2.41	2.16	2.33	1.94	2.90	1.41
5	5.14	3.65	4.41	2.45 1/	1.45	1.01
6	3.14	2.78	3.30	2.97	2.20	2.15
7	2.02	2.57	2.62	2.56	1.69	1.48
8	2.20	2.14	3.44	2.29 1/	3.41	1.29
9	2.33	2.48	1.41	1.82	0.27	0.55
10	3.29	2.79	2.47	2.18	0.66	1.30
11	3.22	1.66 2/	2.38	1.63 1/	0.60	1.27
12	2.02	2.14	1.34	1.70	0.01	0.83
13	1.61	2.24	1.36	2.48	1.07	1.67
Annual	33.35	31.51	10.89	7.57 1/	0.37	0.62

1/ The standard deviation of the generated data is significantly different from the standard deviation of the observed data at the 5 percent level.

2/ The mean of the generated data is significantly different from the mean of the observed data at the 5 percent level.

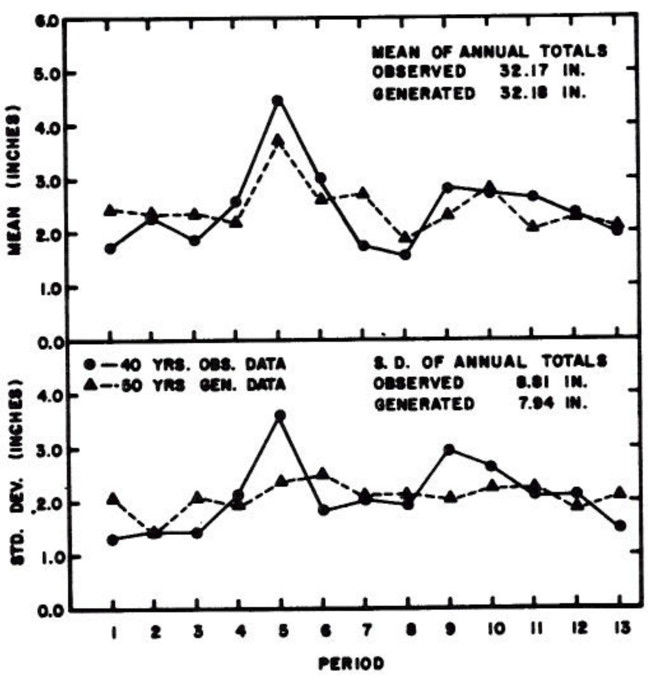

Figure 4-9. Means and standard deviations of the 28-day totals from 40 years of observed data and 50 years of generated data, Hewitt, Texas.

standard deviations for McGregor that were significantly different from the standard deviations of the observed data.

b. Probability of a wet day. The Markov chain transition probabilities were computed from the

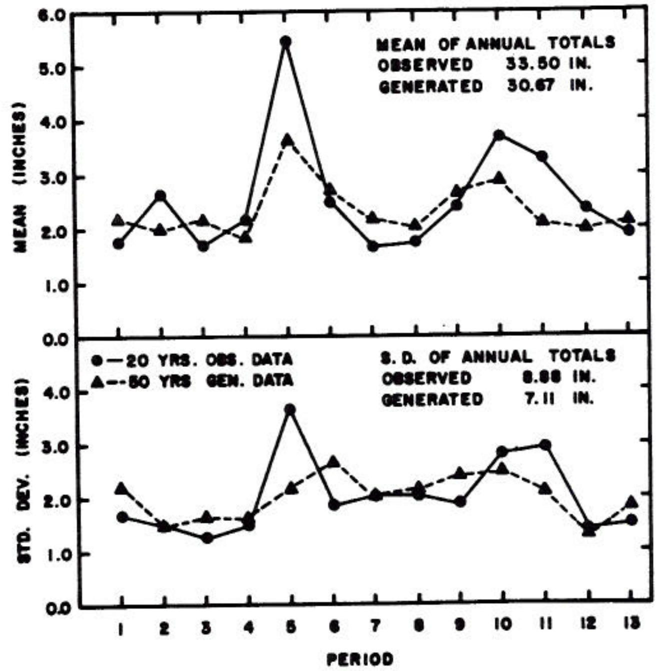

Figure 4-10. Means and standard deviations of the 28-day totals from 20 years of observed data and 50 years of generated data, Troy, Texas.

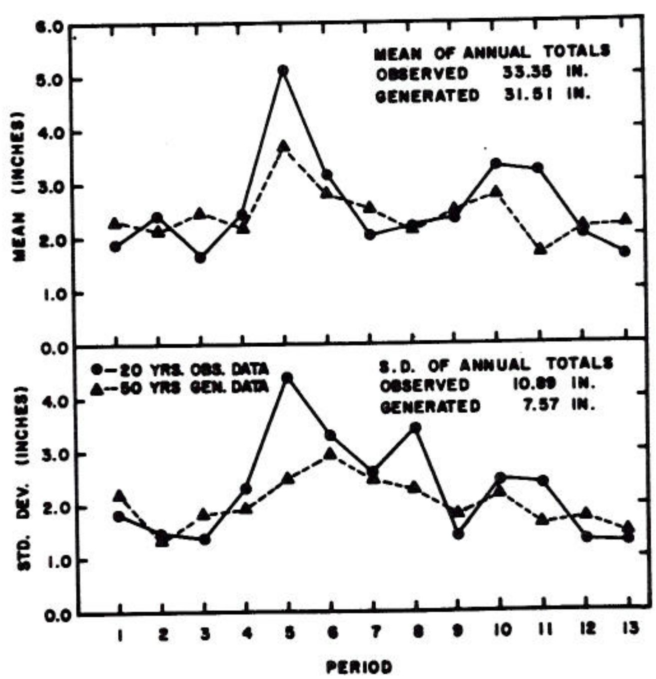

Figure 4-11. Means and standard deviations of the 28-day totals from 20 years of observed data and 50 years of generated data, McGregor, Texas.

observed and the generated data for the three stations in Test Area II. The average P(W/W) and P(D/D) for each 28-day period are given in Tables 4-17, 4-18, and 4-19. The transition probabilities obtained from the generated data were tested to determine whether they differed significantly from the probabilities obtained from the observed data. The agreement between P(W/W) and P(D/D) from the generated data and P(W/W) and P(D/D) from the observed data for Test Area II was about the same as that for Test Area I. Several of the probabilities from the generated data were significantly different from that of the observed data. In most cases P(D/D) values for generated data were less than P(D/D) from the observed data, but both P(D/D) and P(W/W) from the generated data were good approximations of that from the observed data.

Table 4-17. Markov chain transition probabilities of observed and generated data for 13 28-day periods, Hewitt, Texas.

Period	P(W/W) Observed	P(W/W) Generated	P(D/D) Observed	P(D/D) Generated
1	0.447	0.396 1/	0.840	0.825
2	0.479	0.351 1/	0.807	0.796
3	0.316	0.354	0.832	0.836
4	0.384	0.352	0.845	0.848
5	0.445	0.428	0.805	0.795
6	0.438	0.353 1/	0.843	0.822
7	0.385	0.342	0.900	0.847 1/
8	0.363	0.348	0.916	0.878 1/
9	0.362	0.330	0.874	0.852 1/
10	0.467	0.262 1/	0.883	0.851 1/
11	0.404	0.329 1/	0.877	0.876
12	0.455	0.325 1/	0.870	0.814 1/
13	0.465	0.376 1/	0.865	0.845

1/ The transition probability from the generated data is significantly different from the observed data at the 5 percent level.

Table 4-18. Markov chain transition probabilities of observed and generated data for 13 28-day periods, McGregor, Texas.

Period	P(W/W) Observed	P(W/W) Generated	P(D/D) Observed	P(D/D) Generated
1	0.461	0.387	0.879	0.822 1/
2	0.405	0.347	0.840	0.817
3	0.258	0.349	0.879	0.821 1/
4	0.325	0.336	0.876	0.849
5	0.442	0.435	0.815	0.803
6	0.292	0.374	0.860	0.831
7	0.410	0.353	0.930	0.864 1/
8	0.327	0.331	0.916	0.878 1/
9	0.328	0.368	0.874	0.836 1/
10	0.414	0.245 1/	0.861	0.848
11	0.295	0.311	0.878	0.881
12	0.461	0.348	0.891	0.839 1/
13	0.391	0.341	0.871	0.856

1/ The transition probability from the generated data is significantly different from the observed data at the 5 percent level.

Table 4-19. Markov chain transition probabilities of observed and generated data for 13 28-day periods, Troy, Texas.

Period	P(W/W) Observed	P(W/W) Generated	P(D/D) Observed	P(D/D) Generated
1	0.341	0.414 1/	0.854	0.830
2	0.391	0.368	0.823	0.807
3	0.215	0.349 1/	0.845	0.821
4	0.332	0.371	0.878	0.843 1/
5	0.430	0.437	0.825	0.797
6	0.311	0.321	0.877	0.840 1/
7	0.271	0.317	0.903	0.865 1/
8	0.476	0.313	0.913	0.875 1/
9	0.356	0.351	0.869	0.828 1/
10	0.375	0.340	0.857	0.851
11	0.337	0.291 1/	0.891	0.875
12	0.337	0.297 1/	0.885	0.839 1/
13	0.389	0.365	0.885	0.846 1/

1/ The transition probability from the generated data is significantly different from the observed data at the 5 percent level.

c. Cross correlation of 28-day and annual totals. The cross correlation coefficients of the 28-day totals and the annual totals were calculated for both the observed and generated data for each of the three pair of stations in Test Area II. The results are given in Tables 4-20, 4-21, and 4-22. The cross-correlation coefficients obtained from the generated data were tested to determine whether they differed significantly from the cross-correlation coefficients obtained from the observed data. The agreement between the cross correlation coefficients from the generated data and that from the observed data for Test Area II is similar to that for Test Area I. The average of the 13 28-day cross-correlation coefficients from the generated data were close to that of the observed data for all three pairs of stations. The annual cross-correlation coefficients from the generated data were less than that from the observed data for two pairs of stations and greater than that from the observed data for one pair of stations. These results reinforce the conclusion drawn from the Test Area I results that for intervals of 28-days or longer the cross-correlation coefficients of the generated data are close to the cross-correlation coefficients of the observed data.

Table 4-20. Cross correlation coefficients of observed and generated 28-day and annual precipitation, Hewitt and Troy, Texas.

Period	Cross-correlation coefficient: Observed data	Cross-correlation coefficient: Generated data
1	0.883	0.934
2	0.900	0.844
3	0.924	0.873
4	0.650	0.838
5	0.878	0.803
6	0.797	0.803
7	0.927	0.818
8	0.847	0.887
9	0.598	0.829
10	0.810	0.818
11	0.793	0.917 1/
12	0.942	0.796 1/
13	0.937	0.826
Average	0.837	0.845
Annual	0.919	0.845

1/ The cross-correlation coefficient of the generated data is significantly different from the cross-correlation coefficient of the observed data at the 5 percent level.

Table 4-21. Cross correlation coefficients of observed and generated 28-day and annual precipitation, Hewitt and McGregor, Texas.

Period	Cross-correlation coefficient: Observed data	Cross-correlation coefficient: Generated data
1	0.969	0.940
2	0.880	0.838 1/
3	0.974	0.835 1/
4	0.964	0.888 1/
5	0.964	0.735 1/
6	0.559	0.854 1/
7	0.917	0.889
8	0.932	0.892 1/
9	0.623	0.877 1/
10	0.836	0.869
11	0.944	0.881
12	0.888	0.853
13	0.954	0.915
Average	0.878	0.867
Annual	0.888	0.892

1/ The cross-correlation coefficient of the generated data is significantly different from the cross-correlation coefficient of the observed data at the 5 percent level.

Table 4-22. Cross correlation coefficients of observed and generated 28-day and annual precipitation, Troy and McGregor, Texas.

Period	Cross-correlation coefficient: Observed data	Cross-correlation coefficient: Generated data
1	0.897	0.950
2	0.902	0.821
3	0.917	0.853
4	0.627	0.839
5	0.841	0.830
6	0.798	0.876
7	0.833	0.829
8	0.878	0.900
9	0.615	0.756
10	0.794	0.829
11	0.753	0.862 1/
12	0.912	0.726 1/
13	0.869	0.823
Average	0.818	0.838
Annual	0.907	0.815

1/ The cross-correlation coefficient of the generated data is significantly different from the cross-correlation coefficient of the observed data at the 5 percent level.

CHAPTER V
Discussion and Conclusions

Precipitation over an area, observed at discrete points in space, is the result of the interaction of many atmospheric variables. The number of variables and the complexity of the processes are too great to attempt a purely deterministic description of the phenomena. This study is an effort to gain an understanding of the time-area daily precipitation process by modeling the stochastic structure of the process. Hopefully, the model will also be of practical value for generating new time-area daily precipitation sequences for water resource applications.

The stochastic structure of daily precipitation over an area was modeled by using a multivariate normal distribution. The multivariate normal approach had previously been applied to hydrologic series that did not contain zeros, like continuous streamflow or monthly precipitation. This study is an attempt to apply the approach to an intermittent process, like daily precipitation, that contains many zero values.

The parameters of the time-area daily precipitation model were determined as a function of position within a region. New precipitation series were generated for two areas within the region. The new series were similar to the observed series in many of the important time-area characteristics. The model successfully preserved: (1) the periodic means and standard deviations of the normalized daily precipitation, (2) the lag-one autocorrelation coefficients of the daily random components for each station, (3) the lag-zero cross-correlation coefficients between the daily random components for each pair of stations, and (4) the means, standard deviations, and skewness coefficients of 28-day and annual precipitation. The model was not developed to preserve the Markov chain wet-dry transition probabilities or the cross-correlation structure among stations for intervals longer than 1 day. The transition probabilities and the cross-correlation coefficients of 28-day totals and annual totals from generated sequences were often significantly different from the transition probabilities and cross-correlation coefficients from the observed data. These differences probably indicated that the model failed to accurately describe the time and space dependence structure of daily precipitation. However, both the Markov chain transition probabilities and the cross-correlation structure for intervals of 28 days or longer were close enough to that from the observed data so as not to impair the usefulness of the new sequences for most applications.

Basically, this study has shown that the truncated multivariate normal distribution is a useful model of daily precipitation over an area, if precipitation at a point can be transformed to approximate a truncated normal distribution. Specific conclusions resulting from this study are:

1. The square root proved to be adequate for transforming daily precipitation to conform to a truncated normal distribution for stations in the study region. A different transformation may be required in other regions.

2. Maximum likelihood estimates of the means and standard deviations of the truncated samples of normalized daily precipitation may be obtained using the method given by Cohen (1950).

3. The seasonal nature of precipitation can be described using a Fourier series representation of the means and standard deviations of transformed daily precipitation at a point. Six harmonics were assumed to describe the complex periodic movement of the means and standard deviations of the truncated normal distribution for the stations in the study region. The relatively large number of harmonics were required because of periodicity in both the probability of a wet day and the distribution of precipitation amounts given the occurrence of a wet day. The number of Fourier coefficients to be estimated for each parameter is $2m+1$, where m is the number of harmonics. The relatively large number of harmonics for stations in the study region greatly increases the number of coefficients. For most regions, the precipitation pattern during the year is less complex than that for the study region, and fewer harmonics would be required to describe the periodic movement of the means and standard deviations.

4. The harmonics of the means and standard deviations were not in phase for the stations in the study region. If the phase of the harmonics of the means and standard deviations had been the same, fewer coefficients would have been required. Sampling error may have caused part of the difference in phase for the two parameters. However, the periodic variation in precipitation could not be preserved by assuming that the harmonics of the means and standard deviations had the same phase.

5. The regional precipitation trend may be described by relating the coefficients of the periodic parameters to position within the region. The amplitudes and phases of harmonics of the daily means and standard deviations were related to station latitude and longitude, using a simple linear equation. Annual precipitation in the study region increased from west to east with little north-south change. This east-west trend was reflected in the equations relating the coefficients to position in the region. Most of the coefficients were a function of longitude and were independent of latitude. Some of the coefficients were regional constants.

6. The random component of daily precipitation for each station, determined by removing the periodic means and standard deviations was approximately stationary in the mean and variance with a mean of zero, a variance of unity, and a variable lower limit (truncation point). The means and standard deviations used in obtaining the random components were determined using the Fourier series representations with the coefficients given by the regional trend equations.

7. The lag-one autocorrelation coefficient of the random component of daily precipitation was approximately a regional constant with a value of 0.385. The autocorrelation coefficients for lags

greater than 1 day were greater than that given by a first-order Markov model. However, only the lag-one autocorrelation coefficient was used for generating new sequences of daily precipitation for an area. This simplication of the time dependence structure may explain why P(D/D) from the generated data tended to be less than P(D/D) from the observed data.

8. The cross-correlation coefficients between the random components of daily precipitation for stations in the study region were a function of inter-station distance and independent of regional position and orientation (azimuth).

This study raised several questions that could not be answered within the scope of the study. Some of the topics for future research include:

1. Determine general normalizing transformations applicable for daily precipitation for any region using the truncated normal concept. The square root transformation was satisfactory for the region used in this study. Different transformations may be required for other regions.

2. Develop methods of discerning the significant harmonics of periodic parameters for which the sampling error is large with respect to the amplitudes of the harmonics. The sampling errors of the daily means and standard deviations in this study were greater than the amplitudes of the harmonics. No suitable method was available to accurately define the number of harmonics for the two parameters.

3. Investigate the autocorrelation structure of the random components of daily precipitation. The results of this study showed greater persistence in daily precipitation than that given by a first-order Markov model.

LITERATURE CITED

Beals, G. A., "Specification of Daily Precipitation Through Synoptic Climatology," Meteorology M. S. Thesis, Massachusetts Institute of Technology, 1954. Cited by Grace and Eagleson (Reference 13).

Beer, A., Drummond, A. J., and Furth, R., "Sequences of Wet and Dry Months and the Theory of Probability," Quart. Journ. Roy. Met. Soc., Vol. 72, pp. 74-86, 1946.

Benson, M. A. and Matalas, N. C., "Snythetic Hydrology Based on Regional Statistical Parameters," Water Resources Research, Vol. 3, No. 4, pp. 931-935, 1967.

Brittan, M. R., "Past and Probable Future Variations in Stream Flow in the Upper Colorado River," University of Colorado Bureau of Economic Research, October 1961.

Caskey, J. E., Jr., "A Markov Chain Model for the Probability of Precipitation Occurrence in Intervals of Various Lengths," Monthly Weather Review, Vol. 91, pp. 298-301, 1963.

Cohen, A. C., "Estimating the Mean and Variance of Normal Populations from Singly Truncated and Doubly Truncated Samples," Annals of Math. Statist., Vol. 21, pp. 557-569, 1950.

Cooke, D. S., "The Duration of Wet and Dry Spells at Moncton, New Brunswick," Quart. Journ. Roy. Met. Soc., Vol. 79, No. 342, pp. 536-538, 1953.

Feyerherm, A. M. and Bark, L. D., "Statistical Methods for Persistent Precipitation Patterns," Journ. Appl. Met., Vol. 4, pp. 320-328, 1965.

Fiering, M. B., "A Multivariate Model for Synthesis of Streamflow," Journ. of Hyd. Div., ASCE, Vol. 90, pp. 43-60, 1964.

Franz, D. D., "Hourly Rainfall Synthesis for a Network of Stations," Technical Report No. 126, Dept. of Civil Engineering, Stanford University, 1970.

Gabriel, K. R. and Neumann, J., "A Markov Chain Model for Daily Rainfall Occurrence at Tel Aviv," Quart. Journ. Roy. Met. Soc., Vol. 88, pp. 90-95, 1962.

Hopkins, J. W. and Robillard, P., "Some Statistics of Daily Rainfall Occurrence for the Canadian Prairie Provinces," Journ. Appl. Met., Vol. 3, pp. 600-602, 1964.

Jorgensen, D. L., "Persistence of Rain and No-Rain Periods During the Winter of San Francisco," Monthly Weather Review, Vol. 77, pp. 303-307, 1949.

Kotz, S. and Neumann, J., "Autocorrelation in Precipitation Amounts," Journ. Met., Vol. 16, pp. 683-685, 1959.

La Riviere, R. F., "Study of Persistence in Hydrological Data," M. S. Thesis, Massachusetts Institute of Technology, September 1965.

Markovic, R. D., "Probability Functions of Best Fit to Distributions of Annual Precipitation and Runoff," Colorado State University Hydrology Paper No. 8, 1965.

Matalas, N. C., "Mathematical Assessment of Synthetic Hydrology," Water Resources Research, Vol. 3, No. 4, pp. 937-945, 1967.

Namias, J., "The Annual Course of Month-to-Month Persistence in Climatic Anomalies," Bull. Amer. Met. Soc., Vol. 33, No. 7, pp. 279-285, 1952.

Nicks, A. D., "Stochastic Generation of the Occurrence, Pattern, and Location of Maximum Amount of Daily Rainfall," Proceedings of the Symposium on Statistical Hydrology, Tucson, Arizona, August 31-September 2, 1971, U. S. Dept. of Agriculture, Agricultural Research Service, Misc. Pub. No. 1275, pp. 154-171, 1974.

Pattison, A., "Synthesis of Hourly Rainfall Data," Water Resources Research, Vol. 1, pp. 489-498, 1965.

Regier, M. H. and Hamdan, M. A., "Correlation in a Bivariate Normal Distribution with Truncation in Both Variables," Austral. J. Statist., Vol. 13, No. 2, pp. 77-83, 1971.

Slade, J. J., Jr., "An Asymmetric Probability Function," Trans. ASCE, Vol. 101, p. 35, 1936.

Smith, R. E. and Schreiber, H. A., "Point Processes of Seasonal Thunderstorm Rainfall 1. Distribution of Rainfall Events," Water Resources Research, Vol. 9, No. 4, pp. 871-884, 1973.

Stidd, C. K., "Cube-Root Normal Precipitation Distributions," Trans. Am. Geophys. Union, Vol. 34, No. 1, pp. 31-35, 1953.

Thom, H. C. S., "On the Statistical Analysis of Rainfall Data," Trans. Am. Geophy. Union, Vol. 21, Part II, p. 490, 1940.

Thom, H. C. S., "A Statistical Method of Evaluating Augmentation of Precipitation by Cloud Seeding," Technical Report No. 1, Final Report of the Advisory Committee on Weather Control, Washington, D. C., pp. 5-25, 1957.

Weiss, L. L., "Sequences of Wet and Dry Days Described by a Markov Chain Probability Model," Monthly Weather Review, Vol. 92, pp. 169-176, 1964.

Whitcomb, Margaret, "A Statistical Study of Rainfall Data," M. S. Thesis, Massachusetts Institute of Technology, 1940.

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

Yevjevich, V., "Structural Analysis of Hydrologic Time Series," Colorado State University Hydrology Paper No. 56, 1972b.

Yevjevich, V., "Stochastic Problems in Design of Reservoirs," in Water Research, ed. Kneese, A. V. and Smith, S. C., Johns Hopkins Press, Baltimore, 1966.

Yevjevich, V. and Karplus, A. K., "Area-Time Structure of the Monthly Precipitation Process," Colorado State University Hydrology Paper No. 64, 1973.

Young, G. K., "Discussion of 'Mathematical Assessment of Synthetic Hydrology' by N. C. Matalas," Water Resources Research, Vol. 4, No. 3, pp. 681-682, 1968.

Young, G. K, and Pisano, W. C., "Operational Hydrology Using Residuals," Journ. of Hyd. Div., ASCE, Vol. 94, HY4, pp. 909-923, 1968.

Yule, G. U., "On a Method of Studying Time-Series Based on Their Internal Correlations," J. of Roy. Stat. Soc., Vol. 108, pp. 208-225, 1945.

sequences of standardized random components were normally distributed with zero means and unity standard deviations and were dependent in time and space. The lag-one autocorrelation coefficients were found to be a regional constant, as a function of inter-station distance.

Precipitation sequences were generated for two areas in the study region using the truncated multivariate normal distribution model. Parameters of the model were defined using the latitude and longitude of each station. The new sequences closely resembled the observed sequences.

References: Richardson, W. Clarence; Colorado State University, Hydrology Paper No. 91 (July 1977), A Model of Stochastic Structure of Daily Precipitation Over an Area.

sequences of standardized random components were normally distributed with zero means and unity standard deviations and were dependent in time and space. The lag-one autocorrelation coefficients were found to be a regional constant, as a function of inter-station distance.

Precipitation sequences were generated for two areas in the study region using the truncated multivariate normal distribution model. Parameters of the model were defined using the latitude and longitude of each station. The new sequences closely resembled the observed sequences.

References: Richardson, W. Clarence; Colorado State University, Hydrology Paper No. 91 (July 1977), A Model of Stochastic Structure of Daily Precipitation Over an Area.

sequences of standardized random components were normally distributed with zero means and unity standard deviations and were dependent in time and space. The lag-one autocorrelation coefficients were found to be a regional constant, as a function of inter-station distance.

Precipitation sequences were generated for two areas in the study region using the truncated multivariate normal distribution model. Parameters of the model were defined using the latitude and longitude of each station. The new sequences closely resembled the observed sequences.

References: Richardson, W. Clarence; Colorado State University, Hydrology Paper No. 91 (July 1977), A Model of Stochastic Structure of Daily Precipitation Over an Area.

sequences of standardized random components were normally distributed with zero means and unity standard deviations and were dependent in time and space. The lag-one autocorrelation coefficients were found to be a regional constant, as a function of inter-station distance.

Precipitation sequences were generated for two areas in the study region using the truncated multivariate normal distribution model. Parameters of the model were defined using the latitude and longitude of each station. The new sequences closely resembled the observed sequences.

References: Richardson, W. Clarence; Colorado State University, Hydrology Paper No. 91 (July 1977), A Model of Stochastic Structure of Daily Precipitation Over an Area.

Key Words: Precipitation, Time Series, Intermittent Processes, Daily Series, Generation of Samples

Abstract: Daily precipitation over an area is modeled as the time-area process. Daily precipitation at a point is envisioned as a continuous variable that has been truncated at zero. The zero precipitation is considered negative unknown quantity. The square roots of daily precipitation are approximated by a truncated univariate normal distribution. The multivariate normal distribution is used to describe the time-area variation of daily precipiation over an area.

A method was developed for obtaining maximum likelihood estimates of daily means and standard deviations from the truncated samples. The periodic means and standard deviations are described by Fourier series. The Fourier coefficients were related to position within the study region. The

Key Words: Precipitation, Time Series, Intermittent Processes, Daily Series, Generation of Samples

Abstract: Daily precipitation over an area is modeled as the time-area process. Daily precipitation at a point is envisioned as a continuous variable that has been truncated at zero. The zero precipitation is considered negative unknown quantity. The square roots of daily precipitation are approximated by a truncated univariate normal distribution. The multivariate normal distribution is used to describe the time-area variation of daily precipiation over an area.

A method was developed for obtaining maximum likelihood estimates of daily means and standard deviations from the truncated samples. The periodic means and standard deviations are described by Fourier series. The Fourier coefficients were related to position within the study region. The

Key Words: Precipitation, Time Series, Intermittent Processes, Daily Series, Generation of Samples

Abstract: Daily precipitation over an area is modeled as the time-area process. Daily precipitation at a point is envisioned as a continuous variable that has been truncated at zero. The zero precipitation is considered negative unknown quantity. The square roots of daily precipitation are approximated by a truncated univariate normal distribution. The multivariate normal distribution is used to describe the time-area variation of daily precipiation over an area.

A method was developed for obtaining maximum likelihood estimates of daily means and standard deviations from the truncated samples. The periodic means and standard deviations are described by Fourier series. The Fourier coefficients were related to position within the study region. The

Key Words: Precipitation, Time Series, Intermittent Processes, Daily Series, Generation of Samples

Abstract: Daily precipitation over an area is modeled as the time-area process. Daily precipitation at a point is envisioned as a continuous variable that has been truncated at zero. The zero precipitation is considered negative unknown quantity. The square roots of daily precipitation are approximated by a truncated univariate normal distribution. The multivariate normal distribution is used to describe the time-area variation of daily precipiation over an area.

A method was developed for obtaining maximum likelihood estimates of daily means and standard deviations from the truncated samples. The periodic means and standard deviations are described by Fourier series. The Fourier coefficients were related to position within the study region. The

LIST OF PREVIOUS 25 PAPERS

No. 66 Hydrologic Effects of Patch Cutting of Lodgepole Pine, by Thomas L. Dietrich and James R. Meiman, April 1974.

No. 67 Economic Value of Sediment Discharge Data, by Sven Jacobi and Everett V. Richardson, April 1974.

No. 68 Stochastic Analysis of Groundwater Level Time Series in Western United States, by Albert G. Law, May 1974.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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