AREA-DEFICIT-INTENSITY CHARACTERISTICS OF DROUGHTS

by

Norio Tase

November 1976

HYDROLOGY PAPERS
COLORADO STATE UNIVERSITY
Fort Collins, Colorado

87

AREA-DEFICIT-INTENSITY CHARACTERISTICS OF DROUGHTS

by
Norio Tase

HYDROLOGY PAPERS
COLORADO STATE UNIVERSITY
FORT COLLINS, COLORADO 80523

November 1976

No.87

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

The author wishes to express the gratitude to his adviser and major professor, Dr. V. Yevjevich, Professor of Civil Engineering, for his guidance and help during the author's graduate work and research. Special thanks are extended to Dr. D. C. Boes, Associate Professor of Statistics, for his suggestions and guidance. Thanks are also expressed to other members of the Graduate Committee, Dr. D. A. Woolhiser of the Agricultural Research Service and Dr. M. M. Siddiqui, Professor of Statistics.

The financial supports during his graduate studies, first from the Japan Society for the Promotion of Science as a fellowship, and later from Colorado State University in the form of graduate research assistantship, under the U. S. National Science Foundation Grant ENG 74-17396, with Dr. V. Yevjevich as principal investigator are gratefully acknowledged.

The acknowledgement for their encouragement and help goes to Dr. Soki Yamamoto and Mr. Yuichi Suzuki. Fellow Ph.D. graduate students, Kedar Mutreja, Jerson Kelman and Douglas Vargas have given an opportunity to the writer for a mutually beneficial exchange of views and research results in various discussions.

ABSTRACT

Under the concept that monthly precipitation series over an area are composed of deterministic components specified by periodic parameters and a stationary stochastic component, a mathematical model of area-time process of monthly precipitation, especially of the stationary stochastic component, using the Upper Great Plains in the U.S.A. as an example of the model, is developed. The independent identically distributed variables are obtained from the transformed stochastic component. Their regional dependence structure is given by an exponential decay function with the interstation distance. By using this model, new samples of time series over the area at a new grid of 80 points are generated in order to investigate area-deficit-intensity characteristics of droughts.

The deficit area, the total areal deficit, and the maximum deficit intensity are defined as primary indices of drought characteristics. The basic parameters of their frequency distributions and of mutual relationships are analyzed for various truncation levels of drought definitions. The areal drought characteristics are modeled and their parameters defined by three basic indices.

Probabilities of areal coverage of droughts are further investigated by applying the theory of runs, the theory of recurrent events, and by similar approaches. Probabilities of specific areas covered by droughts of given properties are also investigated by considering the effects of the size and the shape of an area.

Run properties of a simple, periodic-stochastic process are investigated analytically. Moments of negative run-sums are found by considering the negative run-length and the onset time. Some other techniques are discussed in comparison with the use of run properties in evaluating drought characteristics of periodic-stochastic processes.

FOREWARD

Droughts are characterized by several properties. In general, mostly droughts of point processes have been investigated, meaning droughts at a given point on the earth's surface are investigated by using time series of variables which determine the drought phenomenon. From these time series serveral indices have been used for drought descriptions, such as the total deficit of water, its maximum deficit intensity, shape, duration or any other characteristics of drought runs. When droughts are investigated for its distributions over a region, investigations become much more complex. Two area concepts are then necessary, namely the fixed region with its size and shape must be defined, and probabilities must be found for a part of this region to be covered by the drought of given point characteristics. Therefore, drought area coverage inside a fixed region, studied simultaneously with the size and time characteristics of droughts, represent a realistic approach to analysis of drought properties by using probability theory, mathematical statistics and stochastic processes.

Two problems have been emphasized by Dr. Norio Tase in his Ph.D. dissertation work in studying droughts. First, it was necessary to select a variable which describes the drought area coverage. Second, it was necessary to select drought characteristics which will be studied simultaneously with the area coverage. When the region to be studied for drought occurrence is large, variables which determine drought conditions must be relatively simple. It is most appropriate for agricultural droughts to use either the soil moisture variable, or the total moisture available in soil for plants in function of their water requirements. However, this simple approach requires data which usually are not available, or must be computed indirectly from other variables; therefore, a simplification was needed by selecting the monthly precipitation as the basic variable in defining droughts. The concept is based on the principle that long historical developments of agriculture in an area have already adjusted mainly to mean values of monthly precipitation, so that the variation around the monthly means and not the mean monthly precipitation themselves determine drought characteristics. The varitions in the form of the periodic standard deviation of monthly precipitation should be included in one way or another to simplify and make uniform drought investigations for a region. The standardized monthly precipitation, equivalent to precipitation of each month decreased by the mean monthly precipitation and divided by its standard deviation, is used as the basic random variable. The new standardized variable is then the same all over a large region.

The selection of the area-time parameters for description of droughts must be simplified. In the study by Dr. N. Tase only three parameters are selected for investigations: area covered by a drought inside the fixed region, total water deficit below the level which defines drought conditions, and maximum intensity or deficit. It was difficult to include the drought duration as a simple parameter with the three above parameters, because the duration changes in length from point to point over a region and does not coincide over the drought area. By simplifying the selection of parameters, the major objective was to obtain a general idea on probabilities of large droughts covering extensive areas. Because of importance of food production in Great Plains of the United States, a large, fixed region inside the Great Plains was selected as an example to show the properties of these drought probabilities. Monthly precipitation series are treated by the already standard technique in studying the area-time periodic-stochastic processes within the Graduate and Research, Hydrology and Water Resources Program of the Department of Civil Engineering at Colorado State University. To simplify the investigation, a relatively limited number of precipitation station series over this large region is selected.

In the real case of forecasting drought occurrences in probability terms for a large region, all the available information should be condensed in form of mathematical models and their estimated parameters, and not only in form of a limited number of station series of a given, same sample size. To obtain best estimates of models and their parameters, all observations over that region should be included in practical cases. Models represent the time structure of monthly precipitation and their estimated parameters are presented in form of their changes over the region. Once the time independent stochastic components (TISC) of monthly precipitation have been determined for all the stations, their interstation dependence in form of lag-zero cross-correlation coefficients can be determined as a model relating these coefficients to station position, distance and orientation. By condensating all the information on monthly percipitation over a large region in form of mathematical models, the generation of new samples of monthly precipitation process over that region becomes feasible and independent of observation points. To simplify this generation, it is feasible to cover the region of drought investigation by a square grid of points, each point being associated with a well defined unit area. In other words, the use of sample generation method for the investigation of droughts properties can be separated from the observation points. This is important because the observation points were selected basically by two criteria in the past, as points at which the observations could be easily organized, with the constraint of available funds for observations.

Because of difficulties for the application of analytical method in the investigation of area-defficit-intensity characteristics of droughts, the experimental (Monte Carlo) or sample generation method was used exclusively in Dr. N. Tase's study in order to estimate these characteristics. In general, one can start with the analytical method by trying to obtain close solutions for simplified cases of drought problems. Then, these simple results serve as the guide to the approach by generating samples over region in order to investigate the more complex drought problems. Or, in the opposite case, one can start with the experimental, sample generation method, by investigating the characteristics of droughts over a large region, and then--as a second phase--apply the analytical method for obtaining the generalized solutions in the close forms. This second approach, in its first phase of the application of experimental method, has been followed in this study. It is expected that the results presented would stimulate specialists in stochastic processes and mathematical statistics to theoretically investigate the joint distributions of drought characteristics, especially including the drought area coverage.

The study by Dr. Norio Tase gives relationships between the three selected drought characteristics as well as probability of these characteristics, either as marginal distributions or as joint distributions. Furthermore, the study shows that the shape of a region, especially of small region, is also an important factor for drought area coverage. However, the larger the region the lesser becomes the effect of the shape and the more important becomes the surface of that region.

In studying the effect of periodicity of periodic-stochastic processes on drought characteristics it was shown that periodicity is one of the major obstacles for extensive studies of drought characteristics by the analytical method. However, the effect of periodicity in parameters can be studied by generating many time series over a region, in preserving not only the time periodic-stochastic character of series but also their regional dependence among the time independent stochastic components. Because periodicities in parameters involve a large number of coefficients, especially Fourier coefficients of harmonics, it would be difficult to relate the various drought characteristics to all these coefficients. This fact then requires the regional studies only, by generating new samples as closely as possible of the area-time processes of controling random variables, and by properly defining what are the droughts for periodic-stochastic processes of water supply and water demand. By generating new samples of these processes, the experimental method produces estimates of marginal probabilities or joint probabilities of drought characteristics.

This study is a part of a continous effort in the Hydrology and Water Resources Program of Department of Civil Engineering at Colorado State University in the analysis of various aspects of droughts. Basically, first their physical aspects are investigated, and then studies are broadened to economic and social aspects.

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

November, 1976
Fort Collins, Colorado

LIST OF SYMBOLS

SYMBOL	DEFINITION
a	Constant or coefficient of polynomial function
a_*	Normalized deficit area
A	Deficit area
$\underline{A}$	Diagonal matrix
b	Constant
$\underline{B}$	Diagonal matrix
$C_{j,i}(\nu)$, $C_j(\nu)$, C	Amplitude at station i, for harmonic j, and for ν
$\mathrm{Cov}(\cdot,\cdot)$	Covariance
d_{ij}, d	Interstation distance
D, D_s	Total areal deficit of ξ
D_a	Total areal deficit of χ
e	Exponent
$E(\cdot)$	Expected value
$f(\cdot)$	Function sign or probability density function
$F(\cdot)$	Cumulative distribution function
ΔF	Critical value of Kolmogorov-Smirnov statistic
$g(\cdot)$	Function sign
G_n	Jacobi polynomial of degree n
$h(\nu)$	Number of significant harmonics in ν
i	Counter
i_*	Normalized deficit intensity
I	Maximum deficit intensity
$\mathcal{I}$	Indicator function
j	Counter
k	Counter
$K(\cdot)$	Cumulant generating function
ℓ	Semigrid interval
L	Grid interval
m	Counter or order of autoregressive model and polynomial function
m_i, $m_{\tau,i}$, m_τ	Monthly mean at station i
$\overline{m}$, $\overline{m}_\tau$	Mean of m_τ over τ
$\hat{m}_\tau$	Estimated monthly mean
M	Positive run-length
$\mathcal{M}$	Number of stations
$M(\cdot)$	Moment generating function
M_{max}	Longest drought duration
n	Number of small squares inside a grid or degree of Jacobi polynomial
$\mathcal{n}$	Number of years of data
N	Negative run-length
$\mathcal{N}$	Sample size
O	Remaining expansion error
p	Counter for year
$\mathcal{p}$	Probability level
$P(\cdot)$	Probability
$\mathcal{q}$	Probability level
Q	Random variable
r_{ij}, r	Sample cross correlation coefficient
$\overline{r}$	Mean areal correlation coefficient
r_k	Sample k-th serial correlation coefficient
R^2	Explained variance
R_0	Lag zero cross correlation matrix
R_1	Lag one serial correlation matrix
$s(\cdot)$	Sample standard deviation
s_i, $s_{\tau,i}$, s_τ	Monthly standard deviation at station i
$\overline{s}$, $\overline{s}_\tau$	Mean of s_τ over τ
$\hat{s}_\tau$	Estimated monthly standard deviation
$\overline{s}_i$	Mean of s_i over area
S	Positive run-sum
$\mathcal{t}$	Time of season
T	Onset time of drought
v_i	Sample value of parameter ν at station i
$\mathrm{Var}(\cdot)$	Variance
W, $W_{N,T}$	Negative run-sum
X	Random variable (deficit area)
X^*	Deficit area
$\mathcal{X}$	Longitude coordinate

SYMBOL	DEFINITION
x_0	Reference for x
Y	Random variable (total areal deficit)
y	Latitute coordinate
y_0	Reference for y
z	Fisher's z variable
Z	Random variable (maximum deficit intensity)
Z^*	Deficit intensity
Z^*_{min}	Minimum deficit intensity
α	Parameter of beta distribution
α_k	k-th autoregressive coefficient
β	Regression coefficient or parameter of beta distribution
ε	Residual
ζ	Transformed ξ variable
$\theta_{j,i}(\nu), \theta_j(\nu), \theta$	Angular phase for harmonic j and for ν parameter
κ^*_m	m-th cumulant of ξ^*
λ	Basic frequency
μ_τ	Monthly population mean
$\bar{\mu}$	Mean of μ_τ over τ
μ_k	k-th moment
ν_τ	Fourier series representation for a periodic function
$\bar{\nu}$	Mean of ν_τ
$\hat{\nu}$	Estimate of ν
$\xi_{p,\tau}, \xi_j$	Second order stationary stochastic variable
$\bar{\xi}$	Mean of ξ over area
ξ_0	Truncation level of ξ
ξ_{min}	Minimum value of ξ
$\underline{\xi}$	Matrix of ξ
ξ^*	Truncated series of ξ
ρ, ρ_{ij}	Population cross correlation coefficient
ρ_k	Population k-th autocorrelation coefficient
σ_τ	Monthly population standard deviation
$\bar{\sigma}$	Mean of σ_τ over τ
τ	Counter for month
υ	Standard normal variable
$\underline{\upsilon}$	Matrix of υ
ϕ, ϕ_{ij}	Orientation between stations i and j
ϕ_{max}	Major axis
$\chi, \chi_{p,\tau}$	Monthly precipitation series
$\chi_0, \chi_{0,\tau}$	Truncation level of χ
χ^2	Chi-square value
$\psi(\cdot)$	Function sign
ω	Second order independent stationary stochastic variable

Chapter 1
INTRODUCTION

1.1. General on Droughts

Droughts and floods are extremes in the fluctuation of various hydrologic phenomenon. Generally human settlements have been river valley-oriented since time immemorial. This has attracted the attention of people to flood problems more than to drought problems, because flood damages to society are much more visible and sudden in comparison with drought damages. In modern times, this situation has been changed due to the following reasons: (1) the pressure on limited water resources by an increase of population and the standard of living, especially in big cities, required attention to water shortage or drought problems; (2) the specialization of regions as it concerns the use and allocation of water resources, such as the granary region of the Great Plains in the United States, makes a region's role especially important. Thus, crop failures in such regions may heavily affect not only the national but also the world economy. With an increase of the world population, the food problems become more serious day by day. Therefore, reduction or failure in grain production for several years in an important region, such as in the wheat belt of the United States, would make a great impact on the world total food supply. Drought is one of the main causes of food supply deficits.

Drought problems are a critical aspect of water resources conservation, development, and control at present. Continued pressure on limited water supplies will make drought problems much more serious in the future. Therefore, intensive and systematic inveatigations on drought problems are urgent and necessary.

The definition of drought is a controversial subject. The difference between drought and water shortage is also vague. Every water user may have his own concept of drought, and furthermore, that concept may change with conditions of operation. In agriculture, drought means a shortage of moisture in the root zone of crops. To a hydrologist, it means below average water levels in streams, reservoirs, groundwater, lakes, etc. In an economic sense, drought means a water shortage which affects or disturbs the established production. Although these concepts are based on different viewpoints, they basically depend upon the effects of prolonged or unusual weather conditions. This study is only concerned with the hydrologic and/or meteorologic drought concepts. The writer contends that an evaluation of the hydrologic or meteorologic drought, defined by an objective way, permits each water user to apply such measures as to determine the effect-relationship in which it has an interest. For a more accurate estimation of drought effects, the definition of drought must be tailored to a particular problem. For an analysis of hydrologic droughts in this study, monthly precipitation phenomenon is taken into account, as a primary water supply.

1.2. Major Problems Needing Studies

Two main drought problems need solutions. First, the problem of the areal coverage, or the extent of a drought, relates to the scale and the shape of droughts and their probability of occurrence. It has not been studies because the precise definition of the areal coverage by a drought and the analyses of areal extent are not simple to attack. For regions within a large area related to each other in many aspects, the areal extent of a drought should be studied for a good planning of water resources development and of alleviating drought effects over the large areas, such as regional water exchange (Takeuchi, 1974).

The second problem is related to difficulties involved with evaluating drought characteristics for the periodic-stochastic time processes such as the daily or monthly precipitation or runoff series. Compared with the stationary series such as annual precipitation or runoff, in the periodic-stochastic series the time position or season is a very important factor in evaluating the drought characteristics such as its duration, magnitude, intensity, etc. This means that in case of periodic-stochastic processes it is difficult to find and/or define the basic drought characteristics such as the negative run-length and the negative run-sum, which are useful characteristics of describing droughts of stationary processes such as annual precipitation or runoff.

1.3. Objectives of the Study

Since the fundamental causes of drought in the form of physical factors of atmospheric circulation are still not well understood, the practical method of studying droughts is to consider their properties as random variables and to use the statistics and observed time series in order to estimate these characteristics.

The first objective of this study was to find experimentally the general characteristics of hydrologic droughts over a large area after developing the mathematical models of area-time processes of monthly precipitation for the case of the Upper Great Plains in the United States. An areal structure of droughts is also studies.

The second objective was to study probabilities of droughts covering a specific area, such as a state within the Great Plains, in considering the effects of the size and shape of this area on probabilities obtained.

Since there are not many investigations on the areal coverage of droughts, this study is related to several general aspects of the areal drought coverage. To document concepts and present the ideas for further studies on large droughts, as many figures and tables are given as was considered necessary or warranted.

The third study objective was to discuss some feasible methods of analyzing droughts of periodic-stochastic processes, by finding some basic properties of negative runs of these processes.

1.4. Procedures Used

The procedures used in developing the mathematical model of area-time stochastic process of monthly precipitation are presented with the model applied to the Upper Great Plains in Chapter III.

In Chapter IV, the determination of the grid system and grid interval is studied with the generation of a new series at new, systematic grid points. The generated series based on the model and the new grid system, are tested statistically to verify that they simulated the basic processes well. Using the generated series, the characteristics of large area droughts are studied in Chapter V. The three variable: the deficit area, the total areal deficit, and the maximum deficit intensity, are defined as the basic characteristics of regional droughts. Their basic properties studied are probability distributions and mutual relationships. The areal structure of droughts is also described in this chapter. Probabilities of areal coverage by droughts are investigated in considering both the size and the shape of an area.

In Chapter VI, the analysis of droughts of periodic-stochastic processes is discussed. The basic properties of negative runs of the periodic-stochastic processes are studied analytically and compared with those of the stationary processes.

Chapter 2
REVIEW OF LITERATURE

2.1. Drought Definition and Studies

The definitions of hydrologic or meteorologic droughts have already been discussed for a long time. Hoyt (1938) stated that drought conditions might prevail when the annual precipitation was as low as 85 percent of the mean. McGuire and Palmer (1957) defined the drought as condition of monthly or annual precipitation less than some particular percentage of normal. Thomas (1962) used the definition that drought was a meteorologic phenomenon and occurred during a period when precipitation is less than the long-term average. Yevjevich (1967) defined a hydrologic drought as the deficiency in water supply on the earth's surface and used the runs as the basic concept for an objective definition of droughts. Drought investigations until 1968 have been presented in the form of annotated references by Palmer and Denny (1971), which can give a good insight to problems and approaches.

The classical approach to drought problems was to find the probability of the instantaneous smallest value on the basis of the theory of extremes (Gumbel, 1963). This approach does not tell anything about the duration and areal coverage of droughts. Unlike flood problems, the duration and areal coverage are very important in drought problems.

Figure 2.1 represents a discrete series of a variable X. By selecting an arbitrary truncation level x_0, two new truncated series of positive and negative deviations are obtained. The number or length of consecutive negative deviations preceded and followed by positive deviations is defined as a negative run-length, which may be associated with the concept of the duration of a drought. The sum or integral of all negative deviations over such a run-length is defined as the negative run-sum. The ratio of the negative run-sum and the negative run-length is defined as the negative run-intensity (Yevjevich, 1967). The negative run-sum and run-intensity can be associated with the severity of a drought.

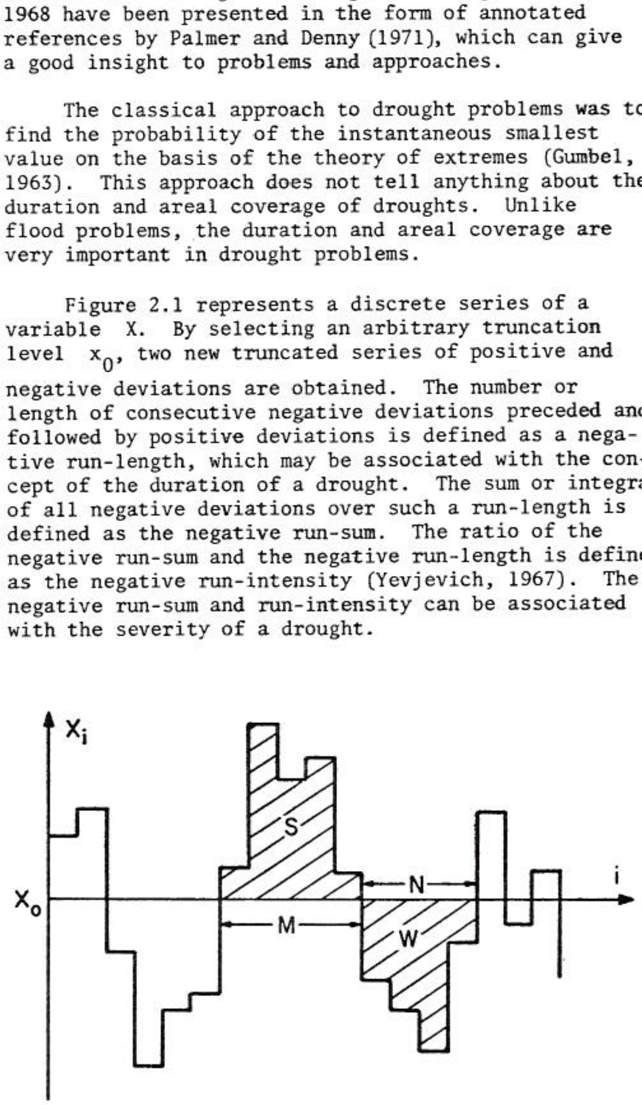


Fig. 2.1. Definition of the Positive Run-Length, M, the Negative Run-Length, N, the Positive Run-Sum, S, and the Negative Run-Sum, W, for a Discrete Series, X_i.

Several theoretical and experimental studies of runs related to drought problems are available. The run-length has been more widely investigated. Saldarriaga and Yevjevich (1970) summarized the exact properties of distributions of run-length for univariate independent random variables, which showed that the run-length properties are free of underlying distribution of input processes. They further studied the properties of run-length for univariate dependent random variables, especially defined by the first-order autoregressive model.

The study of run-sums is very complex theoretically. Only for the univariate independent normal process, the exact properties of run-sums were found by Downer et al. (1967). The exact properties of run-sums of normal dependent or non-normal independent and dependent processes have not been developed.

The application of the theory of runs to a univariate stationary process is useful because it gives the main drought characteristics, such as the probability of occurrence of duration and severity, except the probability of areal coverage. Millan and Yevjevich (1971) studied the probability of historic hydrologic droughts by using the longest negative run-length and the largest negative run-sum as basic parameters of samples of a given size for given probability of the truncation level, the autoregressive coefficients, and the skewness coefficients. Guerrero-Salazar (1973) further studied probabilities of the longest negative run-length and the largest negative run-sum for both univariate and bivariate processes, analytically and experimentally. For bivariate process, Llamas and Siddiqui (1969) studied several basic properties. The overall summary of runs is given by Guerrero-Salazar and Yevjevich (1975).

The application of run properties of univariate and bivariate stationary processes to drought investigations is limited to processes such as annual precipitation or annual runoff series, where the assumption of a stationary process is sufficiently accurate. The short interval processes such as monthly, weekly, and daily precipitation series are periodic-stochastic processes. Hence the above analysis is not directly applicable to such processes, and the problem of developing the techniques to study the periodic-stochastic processes needs attention.

Based on the water budget of the soil, Palmer (1965) used the difference between the actual precipitation and the computed precipitation which is required for the average climate of the area to evaluate drought severity in space and time. Since many factors such as runoff and evapotranspiration are estimated, an application of this method to a large area is very difficult. Herbst et al. (1966) developed a technique for the evaluation of drought only from monthly precipitation. The technique determines the duration and intensity of droughts and their months of onset and termination. It can also compare the intensity of droughts irrespective of their seasonal occurrence.

Few investigations on areal coverage of droughts have been carried out. Even a descriptive method of areal characteristics of drought has not been well developed, and little has been done on applying quantitative or statistical methods to drought coverage. Pinkayan (1966) studied the probability of occurrence of wet and dry years over a large area. He used the conditional probability mathematical functions to describe the occurrence of wet and dry years over the area. He concluded that the occurrence of wet and dry years between two stations up to a distance of 1000 miles is dependent. Gibbs and Maher (1967) analyzed the areal extent of past droughts in Australia by classifying the annual precipitation with the decile range. As a crude index of drought, the first decile range of calendar year rainfall is used to find the return periods of droughts covering certain percentage of the continent. Many investigators, such as Spar (1967) used a kind of precipitation or runoff distribution to discuss the drought phenomenon, without analyzing it quantitatively.

2.2. Models of Monthly Precipitation Series

Roesner and Yevjevich (1966) studied the time structures of monthly precipitation series for 219 stations in the Western United States. They concluded that the monthly precipitation series is composed of deterministic periodic parameters and a nearly independent stochastic component. The periodic component can be described by a Fourier series, mainly with a harmonic of the 12-month cycle.

The spatial extension or smoothing of the time structure parameters of the point series has been investigated by using surface-fitting techniques. In particular, polynomial functions of the space coordinates are usually used (Amorocho and Brandstetter, 1967). These techniques or surface trend analysis are extensively used in geology (Krumbein, 1959, 1963; Mandelbraum, 1963) for separating the relatively large-scale systematic changes in mapped data from essentially non-systematic small scale variations due to local effects or errors.

Spatial or regional structures of monthly precipitation are studied by using cross correlation coefficients (Stenhouse and Cornish, 1958; Huff and Shipp, 1969). Usually the cross correlation coefficients are expressed by various functions of the interstation distance and the orientation of the line connecting the two stations, as well as some other factors. Stenhouse and Cornish (1958) showed that the decay rate of the cross correlation coefficients with the distance and the axis of the maximal correlation are changing month by month. Yevjevich and Karplus (1973) studied the regional dependence structure of the stochastic component of monthly precipitation by using the cross correlation coefficients with ten different functions of the interstation distance and the orientation.

Karplus (1972) studied the structures of area-time hydrologic process of monthly precipitation based on the concept that the process consists of deterministic components specified by periodic parameters and a stationary stochastic component, with the coefficients of the periodic parameters following regional trends. As a result of the application of this concept, the area-time process of monthly precipitation is sufficiently described by the four mathematical models:

(1) Periodic functions for the periodic parameters of the mean and the standard deviation;

(2) Regional trend planes;

(3) Three-parameter gamma probability distribution function of the identically distributed, time independent stationary stochastic component; and

(4) The regional dependence function for the stochastic component.

2.3. Multivariate Data Generation and Grid System

The data generation or the experimental Monte Carlo method has been used in hydrology for some time. The method of generating hydrologic series over an area, or multivariate series, preserves the properties of historic sequences in terms of the sample means, standard deviations, lag-one or lag-one and lag-two serial correlation coefficients, and the lag-zero cross correlation coefficients as a second-order stationary process (Fiering, 1964; Matalas, 1967). The Matalas model and its modified model are used for generation of annual, seasonal, monthly, and daily precipitation (Schaake et al., 1972).

Karplus (1972) applied the method of principal components, using only the first several statistically significant principal components, to the generation of new samples of monthly precipitation and studied the effects of the number of principal components used on the generated series.

Millan and Yevjevich (1971) showed that the experimental method could give very good results by comparing the exact solution with the solution obtained by the experimental method.

With the advent of computers, introduction of the grid system facilitates storing, processing, and retrieving of a large amount of information (Solomon et al., 1968). Yevjevich and Karplus (1973) recommended the use of a systematic grid of points across an area to solve problems related to area-time processes such as droughts, especially by generating multivariate series, based on their mathematical models.

The problem of finding an appropriate grid system and grid interval has been studied indirectly in relation with the problem of areal representativeness of point information or network design of observatories (Linsley and Kohler, 1951; Huff and Neill, 1957; Steinitz et al., 1971; Rodriguez-Iturbe and Mejia, 1974; etc.).

Chapter 3
MATHEMATICAL MODEL OF MONTHLY PRECIPITATION OVER A LARGE AREA

The basic hydrologic processes such as precipitation and river run-off are four dimensional space-time random processes and usually dependent both in time and in area. When a surface is considered, the processes are reduced to three dimensional area-time processes. The series of the amount of monthly precipitation at ground level can be considered as an area-time process as long as the regional topography is fairly homogeneous. To develop mathematical models of area-time process such as monthly precipitation series, Yevjevich and Karplus (1973) gave the basic assumption that the process consists of deterministic components specified by periodic parameters and a stationary stochastic component. The coefficients of the periodic parameters follow some regional trends and the stationary stochastic component has regional characteristics such as regional dependence. Generally, the more stations used and the longer their data, the better is the area-time information of the models. The following discussions are mainly based on Yevjevich and Karplus (1973) and the more detailed procedures can be found in their work.

3.1. Deterministic and Stochastic Components

Two ways of separating deterministic and stationary stochastic components can be considered. The differences between these two methods come from how the monthly means and standard deviations are estimated. The first method is called a non-parametric method. A stationary stochastic component series ξ for a given station is defined by a standardized series of monthly precipitation series χ. That is, the observed monthly means and standard deviations are considered as the deterministic components and are removed from the monthly precipitation series. Therefore, the stochastic component is given by

$$\xi_{p,\tau} = \frac{\chi_{p,\tau} - m_\tau}{s_\tau} \tag{3.1}$$

where p and τ denote the sequence of years and the month within the year, respectively, and m_τ and s_τ are the estimated sample means and standard deviations of the monthly precipitation, respectively. The new variable $\xi_{p,\tau}$ can be a second-order stationary process and can be replaced by ξ_j, that is, $\xi_j = \xi_{p,\tau}$ in which $j = 12(p-1) + \tau$. Now ξ_j is the basic stochastic process to be studied. The mathematical model of area-time structures of the stochastic series may be developed from all series of ξ_j. Models for area-time structures of the deterministic components of the monthly precipitation series can be made on the basis of m_τ and s_τ. Finally, both models can be combined together.

On the other hand, the second method is parametric, as used by Yevjevich and Karplus (1973). It is based on the assumption that once the deterministic area-time components in the parameters of the basic random variable of the monthly precipitation $\xi_{p,\tau}$ have been estimated or modeled and removed from all the point-time series, a second-order stationary area-time stochastic process $\xi_{p,\tau}$ would remain. That is, $\xi_{p,\tau}$ is given by

$$\xi_{p,\tau} = \frac{\chi_{p,\tau} - \hat{m}_\tau}{\hat{s}_\tau} \tag{3.2}$$

where $\hat{m}_\tau$ and $\hat{s}_\tau$ are estimated means and standard deviations by some area-time models, which are discussed later.

These two procedures are compared in Figure 3.1. In the second method, the population means and standard deviations of the monthly precipitation are estimated by using areal information besides the point values. If the model of the deterministic area-time components is derived well, the second method is more useful. However, modeling of the deterministic area-time components for a large area is difficult. When the stochastic component is of primary concern in an investigation and the area studied is fairly large, the first method is easier to apply.

Although there are some differences between the two methods in separating the deterministic and stochastic components, the analysis and modeling procedures are basically the same. The following discussion will be done in this chapter. First, the mathematical model for the time structure of the monthly precipitation series at a given point is found. The model of extending the basic parameters of the deterministic components at a point to regional structures is followed. After discussing the analysis of the area-time stationary stochastic component of the monthly precipitation, the mathematical model of the area-time process of the monthly precipitation is applied to the Upper Great Plains, as a case study.

3.2. Mathematical Model for Time Structure of Monthly Precipitation

Define the random variable $\chi_{p,\tau}$ as the monthly precipitation for a given station i $(i=1,2,\ldots,M)$, with M the number of stations in a region, $p = 1,2,\ldots,n$, the sequence of years, n the sample size expressed in years, $\tau = 1,2,\ldots,12$ month within the year. Also define $\xi_{p,\tau}$ as the standardized random variable with the periodicity of the mean and the standard deviation removed from the station series as

$$\xi_{p,\tau} = \frac{\chi_{p,\tau} - \mu_\tau}{\sigma_\tau} \tag{3.3}$$

where μ_τ is the periodic monthly mean and σ_τ is the periodic monthly standard deviation for the station series.

The variable $\xi_{p,\tau}$ can be assumed as second-order stationary and it is either independent or dependent in sequence. In the case of dependence, the general m-th order autoregressive linear dependence model is usually used. The m-th order autoregressive linear model is given in general by

$$\xi_{p,\tau} = \sum_{k=1}^{m} \alpha_{k,\tau-k}\xi_{p,\tau-k}$$
$$+ \left(1 - \sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_{i,\tau-i}\alpha_{j,\tau-j}\rho_{|i-j|,\tau-k}\right)^{1/2} \omega_{p,\tau} \tag{3.4}$$

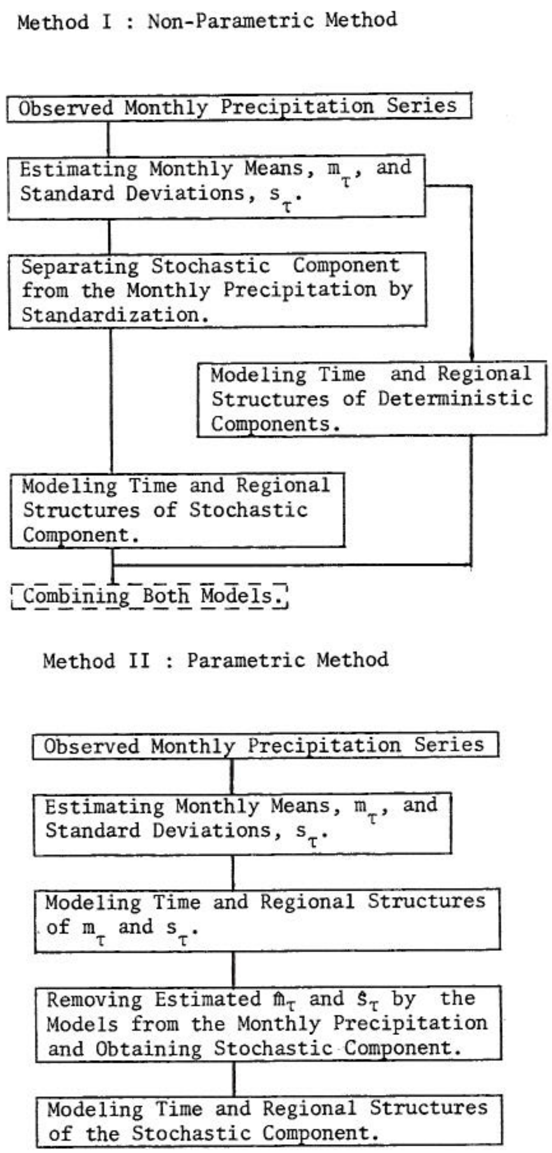


Fig. 3.1. Comparison of Two Methods of Separating Deterministic and Stochastic Components.

in which $k = i$ if $i < j$ and $k = j$ if $i > j$ with $\alpha_{k,\tau-k}$ the autoregressive coefficient at the position $\tau-k$, which are dependent on the autocorrelation coefficients, $\rho_{k,\tau-k}$, and ω is a second-order stationary and independent stochastic variable. In hydrology, the first three linear models have been used by various investigators (Yevjevich, 1964; Roesner and Yevjevich, 1966).

The periodic parameters μ_τ and σ_τ in Eq. (3.3) are symbolized by ν_τ. The mathematical description of the periodic variation of ν_τ is represented in the Fourier series analysis by

$$\nu_\tau = \bar{\nu} + \sum_{j=1}^{h(\nu)} C_j(\nu)\cos[\lambda j\tau + \theta_j(\nu)] \tag{3.5}$$

where $\bar{\nu}$ is the average value of ν_τ, $C_j(\nu)$ the amplitude, $\theta_j(\nu)$ the angular phase, j indexes the sequence of harmonics ($j=1,2,\ldots,h$), $h(\nu)$ denotes the total number of significant harmonics, and λ is the basic frequency of the periodic process.

The general mathematical model of the time structure of $x_{p,\tau}$ is expressed as

$$x_{p,\tau} = \{\bar{\mu} + \sum_{j=1}^{h(\mu)} C_j(\mu)\cos[\lambda j\tau + \theta_j(\mu)]\}$$
$$+ \{\bar{\sigma} + \sum_{j=1}^{h(\sigma)} C_j(\sigma)\cos[\lambda j\tau + \theta_j(\sigma)]\}\,\xi_{p,\tau} \tag{3.6}$$

where the symbols μ_τ and σ_τ for C_j and θ_j correspond to ν_τ in Eq. (3.5). In case of monthly precipitation, the maximum number of harmonics for all periodic parameters is six. However, it is shown by several studies that for monthly precipitation series one, two, or a maximum of three harmonics are sufficient for each periodic parameter.

The ratio of the variance $s^2(\nu_\tau)$ of the fitted ν_τ to the variance of the estimated values such as m_τ and s_τ, is used to select the cutoff point in determining the significant $h(\nu)$ harmonics, since the ratio increases with an increase of the number of harmonics $h(\nu)$.

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

(1) The first harmonic with a period of 12 months alone explains most of the variance; and

(2) ξ is an independent, random variable.

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

$$x_{p,\tau} = \bar{\mu} + C_1(\mu)\cos[\lambda\tau + \theta_1(\mu)]$$
$$+ \{\bar{\sigma} + C_1(\sigma)\cos[\lambda\tau + \theta_1(\sigma)]\}\xi_{p,\tau} \tag{3.7}$$

The $\xi_{p,\tau}$ series is then an independent, stationary random variable at any station.

3.3. Regional Structure Model for Basic Hydrologic Parameters

Let the hypothesis be that the regional variation of any parameter can be obtained from the M point estimates v_i ($i=1,2,\ldots,M$), and be well defined in the form of a trend surface function

$$\nu = \psi(X, Y) \tag{3.8}$$

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

$$\hat{\nu} = f(X, Y) + \varepsilon \tag{3.9}$$

in which ε represents the sampling errors and the difference between the true regional surface function and the fitted function. However, $f(X,Y)$ is usually accepted as the best estimate of $\psi(X,Y)$.

Since $\psi(X,Y)$ is a continuous function, it can always be expanded in a power series form. By taking a polynomial in X and Y of the m-th order, Eq. (3.8) becomes

$$\nu = \beta_1 + \beta_2 X + \beta_3 Y + \beta_4 X^2 + \beta_5 XY + \beta_6 Y^2 + \cdots + \beta_k X^t + \beta_{k+1} X^{m-1} Y + \cdots + \beta_{k+m} Y^m + O(X,Y) \quad (3.10)$$

where β_j, $j = 1,2,\ldots,k + m$, are regression coefficients to be estimated by the least-squares method and $O(X,Y)$ is the remaining expansion error.

The boundaries of the trend surface are greatly affected by the estimates v_i of those stations located near the edge of a region. These estimates may introduce undesirable values of v_i at these edges, such as negative means or standard deviations, when the coefficients β_j of Eq. (3.10) are estimated by the least-squares method. To minimize the boundary effects, the trend surface may be fitted to a larger region having more stations, rather than the region under study with M stations. The β_j coefficients of Eq. (3.10) are estimated for all stations but are applied only to the small interior region defined by M stations.

To evaluate the fitted function $\hat{\nu} = f(X,Y)$ of $\nu = \psi(X,Y)$, the residuals $\varepsilon_i = v_i - \hat{\nu}$ should be analyzed. If the estimate is not good, the areal distribution of the residuals may have some patterns over a region.

3.4. Separation of Deterministic and Stochastic Components of Monthly Precipitation

In order to separate deterministic and stochastic components, and to obtain a second-order stationary independent stochastic process ξ, two methods can be considered as mentioned earlier. As the first method is called a non-parametric method, Eq. (3.1) is directly applied for all the stations to obtain an approximately second-order stationary series ξ, by using the observed means m_τ and standard deviations s_τ of the monthly precipitation $x_{p,\tau}$. The second method is based on the assumption that once the deterministic area-time components in the parameters of the monthly precipitation $x_{p,\tau}$ have been established by Eqs. (3.6), (3.7), and (3.10) and removed from all point-time series by Eq. (3.2), an approximately second-order stationary independent area-time stochastic process ξ would remain.

The stochastic process ξ, either by the first or the second method, can be then considered as a multivariate, identically distributed, stationary area-time process, which is time independent but areally dependent. In other words, the point series at stations over an area are mutually dependent, identically distributed time independent variables.

3.5. Analysis of Area-Time Stationary Stochastic Component of the Monthly Precipitation Series

Once the deterministic component of the monthly precipitation series are separated by the first or second method, the following conditions should be checked:

(1) Each series is time independent;
(2) The point-time series in the region are identically distributed variables; and
(3) The type of dependence among series is in the form of the mathematical regional dependence model.

The time independence of ξ is tested by using the correlograms of individual sample time series of ξ. The hypothesis of identically distributed ξ for all stations in the region is tested by comparing their distribution or distribution parameters as estimated from the observed individual time series.

Once the ξ-series can be assumed as time independent, identically distributed variable, the analysis of regional dependence can be undertaken. Generally the areal dependence is studied by using the linear correlation coefficient ρ_{ij} among stations. The correlation coefficient between the series at station i and j may be a function of the absolute position of the station (X,Y), the interstation distance d_{ij}, the orientation of the line connecting the two stations ϕ_{ij}, and time with the year t. This relation is expressed by

$$\rho = \psi(X,Y,d,\phi,t) \quad (3.11)$$

For a fairly topographically, hydrologically, and meteorologically homogeneous region, the correlation coefficients may be approximated by a function of only the interstation distance and the orientation. That is, Eq. (3.11) is reduced to

$$\rho = f(d,\phi) \quad . \quad (3.12)$$

Several functions relating the estimated interstation correlation coefficient with the interstation distance and the orientation have been studied (Caffey, 1965; Stenhouse and Cornish, 1958; Karplus, 1972).

Under the consideration that the range of ρ for the function should be between zero and unity for all values of d, and that for $d = 0$ by the definition $\rho = 1$, and for $d = \infty$, ρ should be zero, the following two functions are used in this study:

$$\rho = \exp(\beta_1 d) \quad (3.13)$$

and

$$\rho = \exp[(\beta_1 + \beta_2 \cos 2\phi + \beta_3 \sin 2\phi)d] \quad . \quad (3.14)$$

The first model is a simple regional dependence relation between the lag-zero cross correlation coefficient and interstation distance. In this model, the dependence structure is considered to be isotropic and a simple exponential decay. On the other hand, the second model relates the lag-zero cross correlation coefficient of any pair of stations to their distance and orientation. Characteristics of this function are that the rate of the decrease of the correlation coefficient with distance from the station varies with the direction and the slope is symmetrical about the station for a given axis. The direction for the least rate of the decrease of the correlation coefficient, which is called the major axis (Caffey, 1965), is given by

$$\phi_{max} = \frac{1}{2} \tan^{-1}(\beta_3/\beta_2) \quad (3.15)$$

where ϕ_{max} is measured from the reference axis in degrees, counter-clockwise from the east in this study. The ratio of the rates of change of the correlation coefficients along the major and the minor axes shows the degree of ellipticity.

3.6. Application of the Models to the Upper Great Plains in U.S.A.

The Upper Great Plains in the United States is chosen to show how the mathematical models of monthly precipitation are applied and to analyze drought characteristics. The Upper Great Plains is an important agricultural region for production of wheat, corn, and livestock. It is considered to be fairly homogeneous topographically.

Study Area. In the area studied as shown in Fig. 3.2, seventy-nine stations (M=79) with 30 years of monthly values (N=360) for the period 1931-1960 are selected for use in this investigation. These available data are assumed to be statistically homogeneous and are selected to avoid climatic effects of the Rocky Mountains, the Ozark Mountains, and the Great Lakes. The locations of the 79 stations are shown in Fig. 3.2.

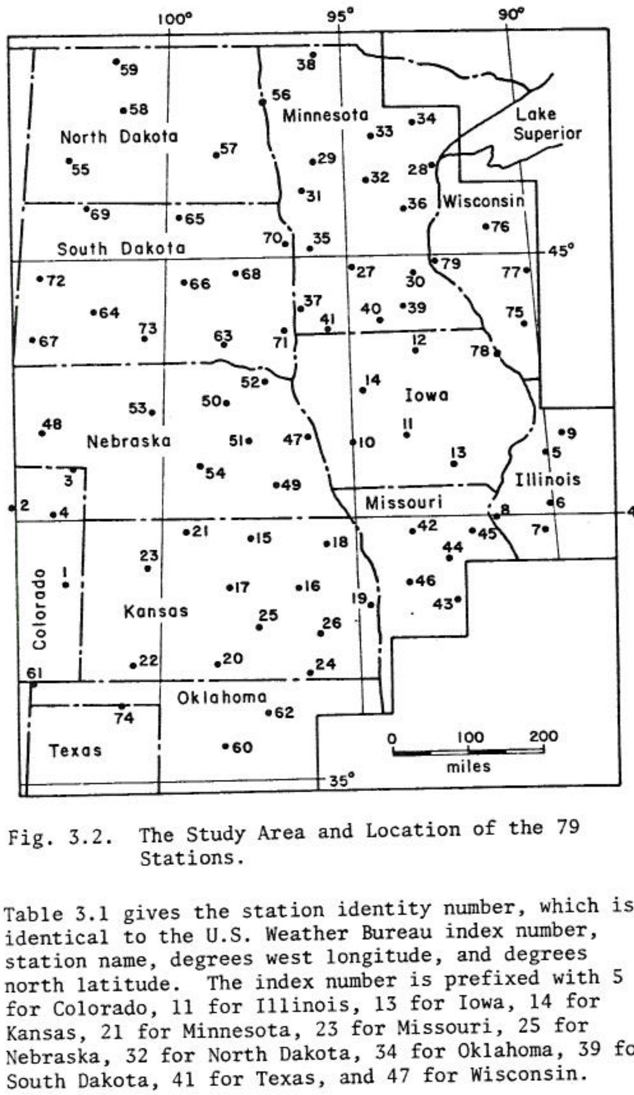


Fig. 3.2. The Study Area and Location of the 79 Stations.

Table 3.1 gives the station identity number, which is identical to the U.S. Weather Bureau index number, station name, degrees west longitude, and degrees north latitude. The index number is prefixed with 5 for Colorado, 11 for Illinois, 13 for Iowa, 14 for Kansas, 21 for Minnesota, 23 for Missouri, 25 for Nebraska, 32 for North Dakota, 34 for Oklahoma, 39 for South Dakota, 41 for Texas, and 47 for Wisconsin.

Looking to the average monthly mean and standard deviation of the 79 stations in Figs. 3.3 and 3.4, it is clear that they have almost the same pattern. They decrease from the southeast to the northwest, though the pattern of the average monthly standard deviation is more complex, as would be expected for a parameter related to the second central moment.

Time Variation of Parameters. The monthly means for each station follow the periodic cycle of the year, which can be described by the Fourier series of Eq. (3.5). For the estimated monthly means μ_τ, Eq. (3.5) takes the form

$$\mu_\tau = \bar{m}_\tau + \sum_{j=1}^{h(\mu)} C_j(\mu)\cos[\lambda j\tau + \theta_j(\mu)] \tag{3.16}$$

where $\bar{m}_\tau$ is the average of the monthly means m_τ.

Similarly for the estimated monthly standard deviations σ_τ, Eq. (3.5) is given by

$$\sigma_\tau = \bar{s}_\tau + \sum_{j=1}^{h(\sigma)} C_j(\sigma)\cos[\lambda j\tau + \theta_j(\sigma)] \tag{3.17}$$

in which $\bar{s}_\tau$ is the average of the monthly standard deviations s_τ. Once $h(\mu)$ and $h(\sigma)$ have been inferred by the method mentioned in Section 3.2, the differences $(m_\tau - \mu_\tau)$ and $(s_\tau - \sigma_\tau)$ are considered to be random sampling variations. That is, the annual cycle of the parameters m_τ and s_τ is considered with only $h(\mu)$ and $h(\sigma)$ significant harmonics, respectively.

For the m_τ and s_τ series, $h(\mu) = h(\sigma) = 1$ is hypothesized and tested by statistical analysis for each of the 79 stations. Table 3.2 presents the estimated values of $\bar{m}_\tau$, $C_1(\mu)$, $\theta_1(\mu)$, $\bar{s}_\tau$, $C_1(\sigma)$, $\theta_1(\sigma)$, and the percent of variance of both m_τ and s_τ, explained by the fitted 12-month harmonic of μ_τ and σ_τ for the 79 stations. From 70.21 to 97.86 percent of the variance, or on the average 88.26 percent, is explained by the fitted 12-month harmonic in the case of m_τ, and from 49.53 to 97.84 percent of the variance, or on the average 83.90 percent, is explained by the first harmonic in the case of s_τ. The average explained variances of m_τ and s_τ by the second harmonic or 6-month harmonic are 5.01 and 5.02 percent, respectively. The second harmonic is not significant in comparison with the first harmonic.

For 72 out of the 79 stations, or 91.14 percent, more than 80 percent of the variance of m_τ is explained by the first harmonic of μ_τ only. While in 71 out of the 79 stations, or 89.87 percent, more than 70 percent of the variance of s_τ is explained by the first harmonic of σ_τ. Since the explained variance by the first harmonic is large for the majority of the stations, the hypothesis of $h(\mu) = h(\sigma) = 1$ seems to be acceptable.

The above hypothesis can be justified by studying the correlograms of the stations 25, 29, and 52 given in Fig. 3.5. It may be noticed that the correlograms are very close to the 12-month cosine function, which indicates that the first harmonic for m_τ is the most important and all the other harmonics could be neglected. Similar correlograms can be found for the other stations also. Thus, the values of $h(\mu) = h(\sigma) = 1$ satisfy the objective in obtaining the minimum number of parameters in using only the most significant harmonic, and in this case only the 12-month harmonic, as required in considering the annual cycle of the monthly means and the monthly standard deviations.

Regional Variation in Parameters. Since only the 12-month harmonic is selected to be considered as the

Table 3.1. Monthly Precipitation Stations Used for Investigation.

Station Number	Index Number	Station Name	Degrees North Lat.	Degrees West Long.	Station Number	Index Number	Station Name	Degrees North Lat.	Degrees West Long.
1	5.1564	Cheyenne Wells	38.82	102.35	41	21.9166	Worthington	43.62	95.60
2	5.3038	Fort Morgan	40.25	103.80	42	23.1580	Chillicothe 25	39.75	93.55
3	5.4413	Julesburg	41.00	102.25	43	23.2503	Eldon	38.35	92.58
4	5.9295	Yuma	40.12	102.73	44	23.2823	Fayette	39.15	92.68
5	11.3335	Galva	41.17	90.03	45	23.7720	Shelbina	39.68	92.05
6	11.3930	Havana	40.30	90.05	46	23.8712	Warrensburg	38.77	93.73
7	11.4442	Jacksonville	39.73	90.23	47	25.0930	Blair	41.55	96.13
8	11.7067	Quincy	39.95	91.40	48	25.1145	Bridge Port	41.67	103.10
9	11.8916	Walnut	41.57	89.58	49	25.2020	Crete	40.62	96.95
10	13.0364	Atlantic 1 NE	41.42	95.00	50	25.2805	Ewing	42.25	98.35
11	13.2208	Des Moines WB City	41.58	93.62	51	25.3185	Genoa	41.45	97.73
12	13.5230	Mason City 3 N	43.18	93.20	52	25.3630	Hartington	42.62	97.27
13	13.6391	Ottumwa	41.00	92.43	53	25.6970	Purdum	42.07	100.25
14	13.7161	Rockwell City	42.40	94.62	54	25.7040	Ravenna	41.03	98.92
15	14.1769	Concordia WB City	39.57	97.67	55	32.2188	Dickenson Expt. Stat.	46.88	102.80
16	14.1866	Council Grove	38.67	96.50	56	32.3621	Grand Forks U.	47.92	97.08
17	14.2459	Ellsworth	38.73	98.23	57	32.4418	Jamestown St. Hosp.	46.88	98.68
18	14.3759	Holton	39.47	95.73	58	32.5638	Max	47.82	101.30
19	14.4421	La Cygne	38.35	94.77	59	32.6025	Mohall	48.77	101.52
20	14.5173	Medicine Lodge	37.27	98.58	60	34.3497	Geary	35.63	98.32
21	14.6374	Phillipsburg	39.77	99.32	61	34.4766	Kenton	36.92	102.97
22	14.6427	Plains	37.27	100.58	62	34.7012	Perry	36.28	97.28
23	14.6637	Quinter	39.07	100.23	63	39.0296	Armour	43.32	98.35
24	14.7305	Sedan	37.12	96.17	64	39.1972	Cotton Wood	43.97	101.87
25	14.7313	Sedgwick	37.92	97.43	65	39.2797	Eureka	45.77	99.62
26	14.8186	Toronto	37.80	95.95	66	39.3832	Highmore 1 W	44.52	99.47
27	21.0783	Bird Island	44.77	94.90	67	39.4007	Hot Springs	43.43	103.47
28	21.1630	Cloquet For. Res. Cent	46.68	92.50	68	39.4661	Ladelle 7 NE	44.68	98.00
29	21.2142	Detroit Lakes 1 NNE	46.83	95.85	69	39.4864	Lemmon	45.93	102.17
30	21.2737	Farmington 3 NW	44.67	93.18	70	39.5536	Milbank	45.22	96.63
31	21.2768	Fergus Falls	46.28	96.07	71	39.7667	Sioux Falls WB AP	43.57	96.73
32	21.3411	Gull Lake Dam	46.42	94.35	72	39.8552	Vale	44.62	103.40
33	21.4652	Leech Lake Dam	47.25	94.22	73	39.9442	Wood	43.50	100.48
34	21.5020	Mahonig Mine	47.47	92.98	74	41.6950	Perryton	36.40	100.82
35	21.5400	Milan	45.12	95.93	75	47.3654	Hillsboro	43.65	90.33
36	21.5615	Mora	45.88	93.30	76	47.4391	Ladysmith	45.47	91.08
37	21.6565	Pipe Stone	44.00	96.30	77	47.5120	Marshfield Expt. Farm	44.65	90.13
38	21.7087	Rseau Power Plant	48.85	95.77	78	47.6827	Prairie Du Chien	43.05	91.17
39	21.8692	Waseca Expt. Farm	44.07	93.52	79	47.7226	River Falls	44.87	92.62
40	21.9046	Winnebago	43.77	94.17					

annual cycle in the basic parameters of the monthly precipitation for all the stations, the next investigation is focused on how the basic parameters of $\bar{m}_\tau$, $C_1(\mu)$, $\theta_1(\mu)$, $\bar{s}_\tau$, $C_1(\sigma)$, and $\theta_1(\sigma)$ vary over the area studied.

Figures 3.2 and 3.3 show that the trend surfaces of $\bar{m}_\tau$ and $\bar{s}_\tau$ can be well defined by a low-order polynomial function of the longitude and latitude. However, Figs. 3.6 through 3.9 indicate that the trend surfaces of $C_1(\mu)$, $\theta_1(\mu)$, $C_1(\sigma)$, and $\theta_1(\sigma)$ may be too complex to be defined easily by a low-order polynomial function. In this study, the longitude X_i is referenced to X_0 = 95.00 degrees west longitude as the zero abscissa, and the latitude Y_i is referenced to Y_0 = 42.00 degrees north latitude as the zero ordinate.

Since a low-order polynomial function explaining a high variance of the regional variation is preferable for further investigations, the polynomial function with more than the fifth-order was considered too complicated to be carried out. Since the usual techniques of step-wise multiple regression analysis are used, elimination of the terms in the regression equation for which the regression coefficients are not significant, and/or those for which the simple correlation coefficients are low, produce the incomplete polynomial equations.

Table 3.3 presents the percent of variance of $\bar{m}_{\tau,i}$, $C_{1,i}(\mu)$, $\theta_{1,i}(\mu)$, $\bar{s}_{\tau,i}$, $C_{1,i}(\sigma)$, and $\theta_{1,i}(\sigma)$, explained by the fitted polynomial functions with various orders up to the fifth. These equations are:

$$\bar{m}_{\tau,i} = 2.4114 - 0.0991Y - 0.1203X + 0.0026Y^2 - 0.0046X^2 + 0.0105XY \tag{3.18}$$

$$\bar{s}_{\tau,i} = 1.4909 - 0.0489X - 0.0945Y - 0.0042X^2 + 0.0051Y^2 + 0.0081XY \tag{3.19}$$

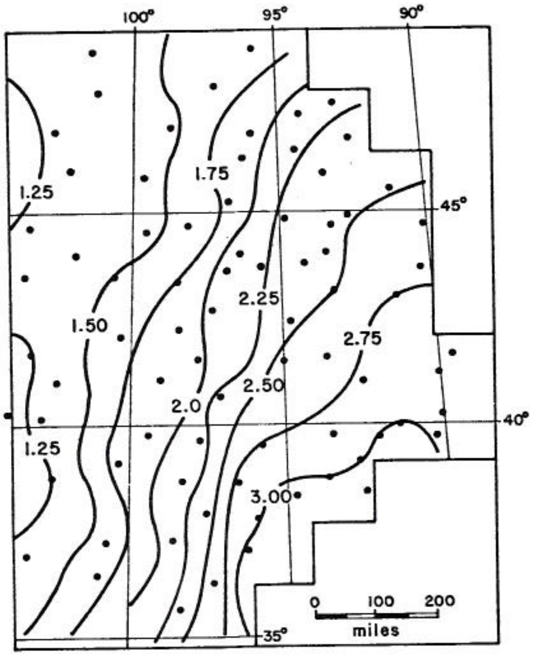


Fig. 3.3. Isolines of the 30-Year General Monthly Mean, $\overline{m}$.

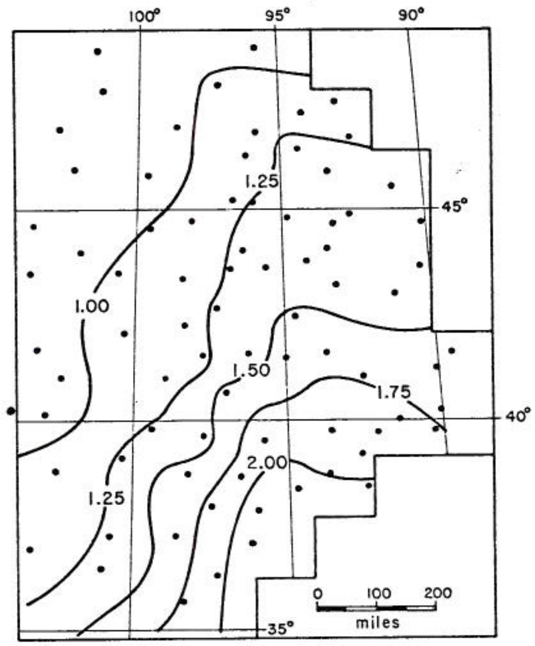


Fig. 3.4. Isolines of the 30-Year General Monthly Standard Deviation, $\overline{s}$.

Table 3.2. Monthly Average Value, $\overline{m}_\tau$, of Means, Amplitude $C_1(\mu)$ and Phase $\theta_1(\mu)$ of the First Harmonic of Monthly Means, Monthly Average Value, $\overline{s}_\tau$, of Standard Deviations, Amplitude $C_1(\sigma)$ and Phase $\theta_1(\sigma)$ of the First Harmonic of Monthly Standard Deviations, and the Percent of Variance of m_τ and s_τ Explained by the First Harmonic of m_τ and s_τ.

Station Number	$\overline{m}_\tau$	$C_1(\mu)$	$\theta_1(\mu)$	%	$\overline{s}_\tau$	$C_1(\sigma)$	$\theta_1(\sigma)$	%
1	1.247	1.182	2.886	93.36	1.050	0.699	2.679	87.36
2	1.072	0.925	3.074	85.15	0.745	0.521	2.722	89.96
3	1.359	1.213	3.099	91.95	0.896	0.605	2.952	94.19
4	1.394	1.220	3.041	94.47	0.948	0.659	2.675	96.90
5	2.827	1.230	2.961	86.44	1.557	0.529	2.263	50.25
6	2.865	1.160	3.014	88.66	1.707	0.636	2.481	78.05
7	2.901	1.004	3.014	74.39	1.764	0.651	2.559	73.94
8	2.997	1.338	2.856	91.32	1.887	0.915	2.524	90.63
9	2.842	1.275	2.818	86.90	1.596	0.635	2.252	61.37
10	2.598	1.843	2.763	91.15	1.603	0.915	2.549	91.03
11	2.563	1.503	2.862	86.72	1.679	0.865	2.424	92.71
12	2.490	1.844	2.766	92.07	1.424	0.894	2.664	80.77
13	2.863	1.594	2.858	84.93	1.762	0.979	2.374	84.14
14	2.447	1.648	2.874	91.27	1.522	0.860	2.735	86.87
15	2.116	1.584	2.885	89.87	1.477	0.964	2.639	95.93
16	2.799	1.962	2.833	92.27	1.892	1.045	2.539	90.15
17	2.183	1.482	2.904	90.06	1.527	0.913	2.635	95.48
18	2.827	1.954	2.815	88.41	1.841	1.102	2.625	90.46
19	3.267	1.874	2.886	89.65	2.247	1.291	2.479	79.68
20	2.088	1.369	2.956	81.31	1.607	0.824	2.752	70.47
21	1.860	1.497	2.949	96.14	1.311	0.819	2.860	86.20
22	1.562	1.125	2.953	86.48	1.269	0.627	2.900	64.03
23	1.796	1.485	3.023	92.92	1.452	1.008	2.846	80.99
24	3.172	1.774	2.837	81.92	2.250	1.270	2.476	74.99
25	2.444	1.612	2.866	90.42	1.760	1.004	2.732	89.10
26	2.989	1.822	2.988	80.95	2.111	1.146	2.746	82.53
27	2.277	1.746	2.837	91.34	1.377	0.916	2.738	97.74
28	2.427	1.615	2.723	94.72	1.249	0.718	2.591	96.86
29	1.963	1.672	2.773	90.79	1.205	0.951	2.746	88.62
30	2.372	1.724	2.755	92.09	1.461	0.915	2.444	82.37
31	1.949	1.541	2.927	88.64	1.087	0.716	2.880	90.66
32	2.207	1.760	2.781	91.98	1.272	0.900	2.811	91.45
33	2.059	1.717	2.741	95.28	1.198	0.988	2.664	85.98
34	2.037	1.595	2.069	97.86	1.075	0.755	2.510	94.84
35	1.969	1.480	2.841	88.80	1.263	0.811	2.763	83.69
36	2.347	1.720	2.762	86.54	1.320	0.880	2.755	81.60
37	2.038	1.731	2.818	89.80	1.224	0.817	2.755	93.48
38	1.634	1.393	2.560	96.47	0.952	0.670	2.392	83.41
39	2.321	1.605	2.848	91.21	1.335	0.763	2.623	93.60
40	2.340	1.812	2.852	89.21	1.300	0.860	2.611	94.23
41	2.206	1.781	2.819	91.33	1.250	0.758	2.658	83.81
42	2.821	1.643	2.815	88.51	1.829	0.985	2.427	78.10
43	3.312	1.281	2.804	76.98	2.113	0.891	2.393	67.59
44	2.971	1.291	2.929	87.32	1.810	0.796	2.459	84.18
45	3.012	1.390	2.903	89.91	1.902	0.946	2.321	91.74
46	3.011	1.378	2.851	75.25	1.958	0.821	2.493	72.07
47	2.210	1.505	2.905	92.90	1.411	0.637	2.781	82.03
48	1.239	1.180	3.071	85.18	0.848	0.711	2.864	80.54
49	2.353	1.778	2.912	88.75	1.550	0.909	2.733	83.18
50	1.872	1.608	2.954	92.55	1.197	0.850	2.731	96.40
51	1.957	1.607	2.994	87.50	1.227	0.770	2.646	96.36
52	2.062	1.402	3.029	92.57	1.276	0.684	2.631	86.24
53	1.703	1.350	3.077	95.43	1.132	0.655	2.932	87.75
54	1.818	1.647	3.007	89.80	1.107	0.679	2.827	91.06
55	1.285	1.198	2.954	74.72	0.894	0.666	2.735	76.33
56	1.676	1.409	2.675	91.72	1.092	0.851	2.421	92.51
57	1.480	1.291	2.883	88.32	0.915	0.605	2.734	83.32
58	1.406	1.358	2.888	80.70	0.868	0.670	2.749	92.10
59	1.387	1.281	2.853	82.42	0.962	0.756	2.761	75.52
60	2.315	1.219	2.986	70.21	1.808	0.770	2.775	53.56
61	1.282	1.059	2.835	90.84	1.172	0.811	2.778	76.80
62	2.652	1.415	3.016	77.91	1.982	0.815	2.737	49.53
63	1.768	1.384	2.938	85.37	1.096	0.601	2.620	83.61
64	1.219	1.074	3.207	81.05	0.879	0.491	3.039	79.81
65	1.422	1.441	2.900	86.32	0.887	0.624	2.861	89.63
66	1.424	1.186	3.008	85.83	1.005	0.578	2.899	91.83
67	1.338	1.243	3.171	90.92	0.984	0.846	3.140	91.59
68	1.581	1.388	2.896	90.97	1.022	0.686	2.756	94.90
69	1.277	1.154	3.019	76.28	0.868	0.637	2.959	67.65
70	1.762	1.431	2.925	86.07	1.092	0.686	2.928	81.16
71	2.096	1.628	2.896	90.76	1.252	0.725	2.554	86.38
72	1.287	1.155	3.117	81.49	0.971	0.827	3.064	80.55
73	1.529	1.207	3.153	86.89	1.110	0.602	2.927	83.18
74	1.700	1.310	2.866	87.18	1.287	0.697	2.726	69.49
75	2.642	1.536	2.661	93.11	1.396	0.772	2.345	89.49
76	2.480	1.808	2.668	96.78	1.437	0.992	2.650	97.54
77	2.606	1.600	2.672	89.15	1.376	0.740	2.492	72.61
78	2.756	1.734	2.734	92.80	1.467	0.912	2.426	89.78
79	2.468	1.746	2.779	93.56	1.366	0.823	2.659	76.10
Mean	2.138	1.478	2.891	88.26	1.367	0.802	2.677	83.90
S.d.	0.592	0.245	0.128	5.63	0.364	0.161	0.187	10.58

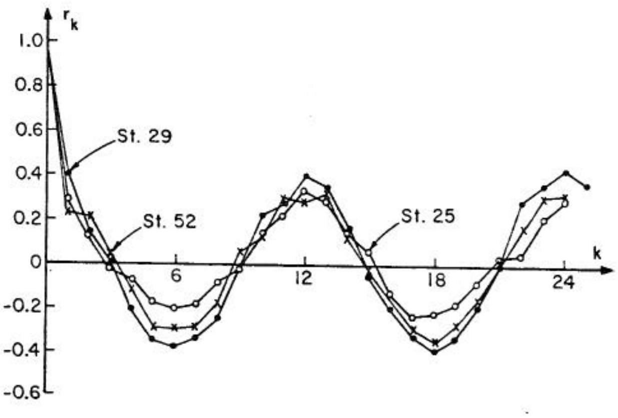


Fig. 3.5. Correlograms of the Monthly Precipitation Series for Stations 25, 29, and 52.

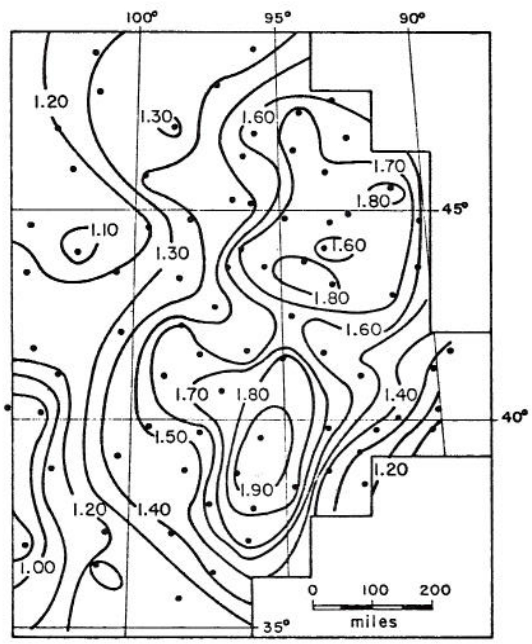


Fig. 3.6. Isolines of the Amplitude, $C_1(\mu)$, of the First Harmonic of Monthly Means.

$$C_{1,i}(\mu) = 1.7538 - 0.0225X - 0.0173Y - 0.0155X^2 - 0.0092Y^2 - 0.0173XY + 0.0011X^3 + 0.0025X^2Y + 0.0012XY^2 + 0.0011Y^3 \quad (3.20)$$

$$\theta_{1,i}(\mu) = 2.8450 + 0.0314X - 0.0020Y + 0.0037X^2 + 0.0069XY - 0.0004X^3 - 0.0004X^2Y - 0.0009XY^2 - 0.0008Y^3 \quad (3.21)$$

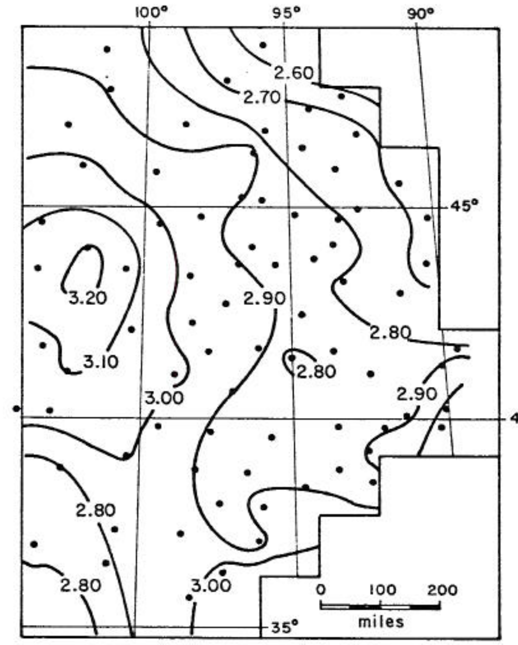


Fig. 3.7. Isolines of the Angular Phase, $\theta_1(\mu)$, of the First Harmonic of Monthly Means.

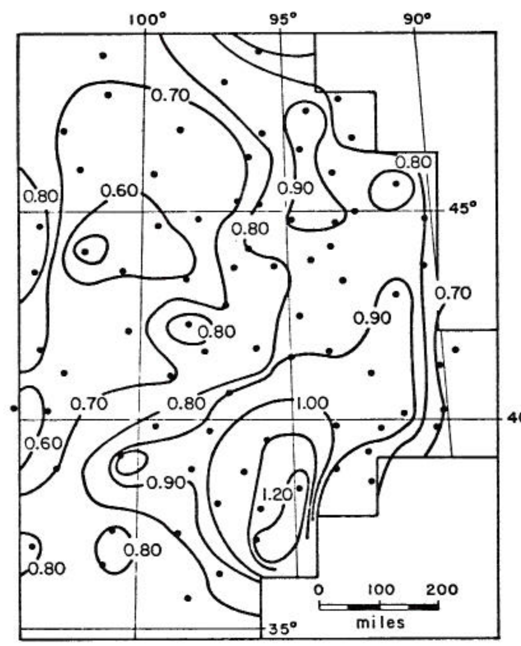


Fig. 3.8. Isolines of the Amplitude, $C_1(\sigma)$, of the First Harmonic of Monthly Standard Deviations.

$$C_{1,i}(\sigma) = 0.8349 - 0.0250X - 0.0605Y - 0.0055X^2 - 0.0094XY + 0.0210Y^2 + 0.0008X^3 + 0.0025X^2Y - 0.0013XY^2 - 0.0010Y^3 - 0.0006Y^4 + 0.0001X^3Y^2 - 0.0001X^2Y^3 + 0.0001XY^4 \quad (3.22)$$

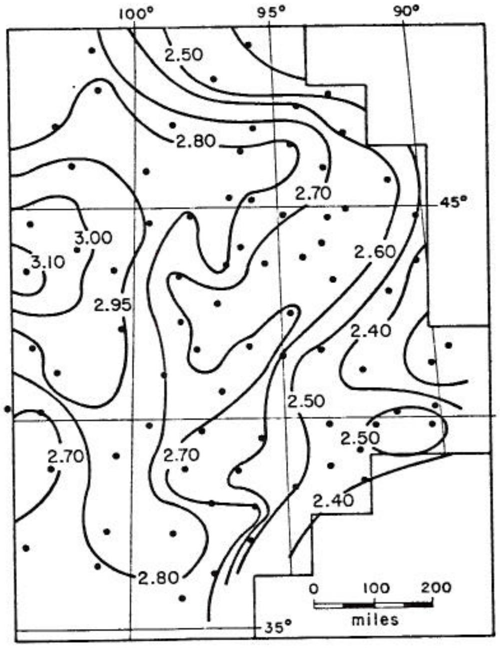


Fig. 3.9. Isolines of the Angular Phase, $\theta_1(\sigma)$, of the First Harmonic of Monthly Standard Deviations.

Table 3.3. Percentage of the Explained Variance of Trend Surfaces by the Polynomial Functions with Various Orders.

Parameter	Order of Polynomial Function 1st	2nd	3rd	4th	5th
$\bar{m}_\tau$	87.9	93.3			
$C_1(\mu)$			78.4		
$\theta_1(\mu)$			81.8		
$\bar{s}_\tau$	78.9	90.6			
$C_1(\sigma)$			57.4		70.0
$\theta_1(\sigma)$			71.7		

$$\theta_{1,i}(\sigma) = 2.6239 + 0.0501X + 0.0384Y + 0.0002X^2 + 0.0009XY + 0.0007Y^2 - 0.0002X^3 + 0.0001X^2Y - 0.0007XY^2 - 0.0014Y^3 \quad (3.23)$$

The fitted functions do not explain substantially the variance of the parameters except for $\bar{m}_\tau$ and $\bar{s}_\tau$. It may be noticed that for the $C_{1,i}(\sigma)$ even the fifth-order polynomial function does not explain the variance well. This means that the regionalization of parameters of periodic variations is very difficult for a large area, and also that the separation of the stochastic component from the monthly precipitation series by the parametric method may not be successful.

Yevjevich and Karplus (1973) had the same difficulties over the area used, which was much smaller than the area used in this study. They tried to infer that the ratios of $C_1(\mu)/\bar{m}_\tau$, $\bar{s}_\tau/\bar{m}_\tau$, and $C_1(\sigma)/\bar{m}_\tau$ were constants over the area. The other method to resolve this difficulty in fitting is to divide the total area into small subarea, and to find a different model for each subarea.

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

$$\xi_{p,\tau} = \frac{x_{p,\tau} - \bar{m}_\tau - C_1(\mu)\cos[\frac{\pi}{6} + \theta_1(\mu)]}{\bar{s}_\tau + C_1(\sigma)\cos[\frac{\pi}{6} + \theta_1(\sigma)]} \quad . \quad (3.24)$$

Separation of Stochastic Component from Monthly Precipitation Series. Since the mathematical model of area-time process of the monthly precipitation is not estimated well, the stochastic component given by Eq. (3.24) cannot be assumed to be a second-order stationary process. Therefore, the non-parametric method of separating the stochastic component from monthly precipitation series is used. The stochastic component is defined by

$$\xi_{p,\tau} = \frac{x_{p,\tau} - m_\tau}{s_\tau} \quad (3.25)$$

with the supscript p,τ then replaced by j, $\xi_{p,\tau}$ by ξ_j, where $j = 12(p-1) + \tau$. The next study is to analyze properties of the stochastic component ξ_j.

Test of Time Independence of Stationary Stochastic Component. In order to test whether ξ is an independent stationary stochastic process, the correlogram of each series ξ of the 79 monthly precipitation series is tested for significant departures. Some of the correlograms are shown in Fig. 3.10. The test is carried out with the 95 percent tolerance limits for the correlogram of an independent series. Only the first twenty lags of correlograms were checked. The 95 percent tolerance limits, r_u and r_ℓ, for an independent series are given by

$$r_{u,\ell} = \frac{-1 \pm t\sqrt{N - k - 2}}{N - k - 1} \quad (3.26)$$

with k = the lag, $t = 1.96$ being the deviate from the standard normal distribution for a two-tail test that $\rho_k = 0$ for $k > 0$ for the ξ-series, and N = the sample size. Table 3.4 presents the number of r_k values for the lags 1 through 20, which are outside the tolerance limits for the 79 stations. The percentage of the total number of the serial correlation coefficients which are outside the 95 percent tolerance limits is 6.27, which is a little higher than the expected value of five percent. Among the serial correlation coefficients, r_k, which are often outside the tolerance limits, only the first serial correlation coefficient r_1 is prominent. This fact is important because the r_1 is affected by some small dependence in the monthly precipitation series due to the dependence in the monthly precipitation series due to the inherent in meteorological processes from day to day. However, only 13 out of 79 stations have significant differences of the first correlation coefficient from $\rho_1 = 0$. The stochastic component of

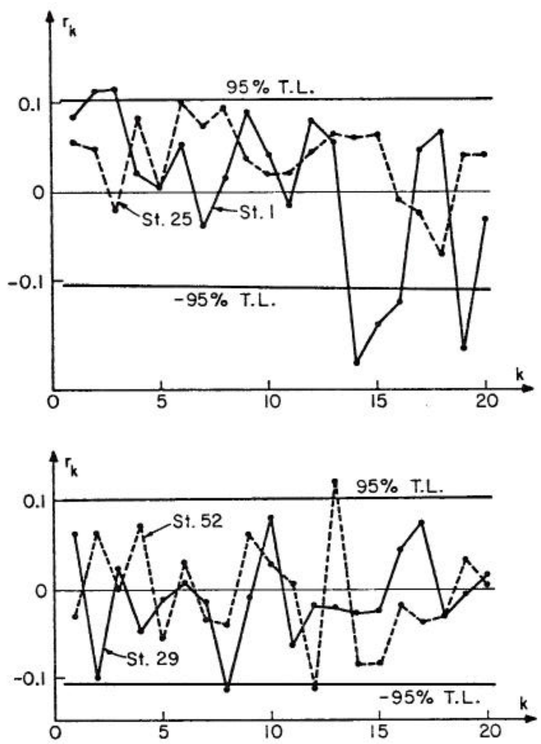


Fig. 3.10. Correlograms of the Stochastic Component of Monthly Precipitation Series of Stations 1, 25, 29, and 52.

Table 3.4. Numbers of Serial Correlation Coefficients for the 79 Series of ξ Outside the 95 Percent Tolerance Limits of Independent Series.

Lags	Number of Stations Outside 95% T. L.	Percentage
1	13	16.46
2	5	6.33
3	3	3.80
4	4	5.06
5	2	2.53
6	2	2.53
7	3	3.80
8	6	7.59
9	6	7.59
10	4	5.06
11	6	7.59
12	3	3.80
13	7	8.86
14	3	3.80
15	1	1.27
16	6	7.59
17	8	10.13
18	5	6.33
19	8	10.13
20	4	5.06
Total	99	6.27

the monthly precipitation series over the Upper Great Plains can be considered to be approximately the time independent variable, also supported by other studies, such as Roesner and Yevjevich (1966) and Karplus (1972).

Analysis of Independent Identically Distributed Stochastic Component. For the further mathematical description of monthly precipitation series, it is necessary to determine the probability distribution function of the ξ-series. It is more important to transform the distribution of the stochastic component ξ into a normal distribution than to determine the exact distribution of the stochastic component itself, because the normal distribution is convenient for the generation of multivariate stationary series, and the ξ has usually a positively skewed distribution. The problem is what transformation should be used to transform ξ into a normal distribution. Generally logarithmic, square-root, cube-root, and similar transformations are used (Stidd, 1953; Kirby, 1972). Suzuki (1968) developed a way of finding the root which can be taken in transformation to obtain a symmetric (not necessarily normal) distribution from an asymmetrical distribution. The exponent "e" required to normalize the distribution is found from the following two equations:

$$Q^e_{(0.16N)} + Q^e_{(0.84N)} - 2Q^e_{(0.5N)} = 0 \quad , \qquad (3.27)$$

and

$$Q^e_{(0.25N)} + Q^e_{(0.75N)} - 2Q^e_{(0.5N)} = 0 \qquad (3.28)$$

where N = the sample size and $Q_{(\cdot)}$ = an ascending ordered value of the concerned variable. The exponent e, satisfying the above two equations, can be found by Newton's approximation method. The simple way of finding the exponent, suggested by Suzuki (1968), is used in this study. The approximated values of the e are given by

$$e_1 = \log 2/[\log Q_{(0.84N)} - \log Q_{(0.5N)}] \qquad (3.29)$$

and

$$e_2 = \log 2/[\log Q_{(0.75N)} - \log Q_{(0.5N)}] \quad , \qquad (3.30)$$

while the difference terms are given by

$$\Delta e_1 = [Q_{(0.16N)}/Q_{(0.84N)}]^{e_1}/[\log Q_{(0.5N)} - \log Q_{(0.84N)}] \qquad (3.31)$$

and

$$\Delta e_2 = [Q_{(0.25N)}/Q_{(0.75N)}]^{e_2}/[\log Q_{(0.5N)} - \log Q_{(0.75N)}] \ . \qquad (3.32)$$

Finally, the estimate e is given by

$$e = (e_1 + \Delta e_1 + e_2 + \Delta e_2)/2 \ . \qquad (3.33)$$

Since the stochastic component of monthly precipitation series is standardized, thus having negative values, a direct application of the above method to ξ series is not feasible. A rate of transformation is the same in both directions with regard to the mean or zero and the skewed shape is not transformed well by direct application. Therefore, $(\xi + c)$-series is used in this transformation. The c may be a parameter rather than a constant. However, it is not easy to find simultaneously the best estimates of these two parameters. In this study, c is considered a constant. To obtain a general idea of the value of e, the minimum values of ξ for each station are taken as constants. In this approach, the average value of the exponent was 0.348, with the standard deviation of 0.164 and the range of values from -0.078 to 0.778. Since the average value of the estimated exponent is close to 0.333 and no special pattern of the areal distribution can be found, the cube-root transformation is chosen to make the distribution of the stochastic component normal. The cube-root transformation was used for the same purpose by Karplus (1972).

The new series ζ, whose values are all or almost all positive, is defined by adding a certain constant to the stochastic component and taking the cube-root, that is,

$$\zeta = (\xi + c)^{1/3} \quad . \tag{3.34}$$

The absolute minimum values of ξ for each station, denoted by $|\xi_{min}|$, and their average value of 1.7653 are taken as constants. The goodness of fit of the normal distribution to the ζ-series is determined. The chi-square values for the 79 fits are computed using ten class intervals of equal probability with seven degrees of freedom at the 95 percent tolerance limit. For the $|\xi_{min}|$ as the constant, 60 out of the 79 stations were accepted to be normally distributed, but with different means and standard deviations. For the average minimum, or 1.7653, only 49 stations were accepted. The distribution of the ζ-series over the area may be approximated by the normal distribution. Distributions are not identical for each station, however. Therefore, further steps are needed to obtain time independent, but regionally dependent, identically distributed normal variable.

On the basis of the above analysis, the ξ-series are assumed to be defined by the standard normal variable υ, as

$$\xi = (1.1604 + 0.2381\upsilon)^3 - 1.7653 \quad . \tag{3.35}$$

The three constants, 1.1604, 0.2381, and 1.7653, are the averages of the means and standard deviations of the 79 series of ζ, for $|\xi_{min}|$ as the constant, and the average value of $|\xi_{min}|$, respectively. If the relation given by Eq. (3.35) is accepted for the area, the ξ-series can be assumed independent identically distributed random variable. The expected value and the expected variance of ξ can be calculated analytically when υ is a standard normal variable. For a standard normal distribution, the k-th moment μ_k is given by

$$\mu_k = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^k e^{-x^2/2} dx \quad . \tag{3.36}$$

When k is an odd number, $x^k e^{-x^2/2}$ in Eq. (3.36) is an odd function, and its integration from $-\infty$ to ∞ gives zeros. For k even, Eq. (3.36) becomes

$$\mu_k = 1 \cdot 3 \cdot 5 \cdots (k-1) \; , \qquad (k=2,4,6,\cdots) \tag{3.37}$$

Since the first six moments of a standard normal variable are $E(\upsilon) = E(\upsilon^3) = E(\upsilon^5) = 0$, $E(\upsilon^2) = 1$, $E(\upsilon^4) = 3$, and $E(\upsilon^6) = 15$, the expected value and variance of ξ are

$$E(\xi) = E[(1.1604 + 0.2381\upsilon)^3 - 1.7653] = -0.0054 \tag{3.38}$$

and

$$\mathrm{Var}(\xi) = E\{[\xi - E(\xi)]^2\} = 1.0809 \quad . \tag{3.39}$$

After the distribution function of ξ is calculated from the distribution function of the standard normal distribution, the comparison of the observed frequency distribution with the assumed distribution function of ξ is carried out by the Kolmogorov-Smirnov test at the five percent level of significance, though it is a weak test. Some comparisons are shown in Fig. 3.11. Only two out of the 79 tests give significant test statistics. That is, the maximum differences of probabilities of distribution curves, $(\Delta F)_{max}$, for these two stations are bigger than the critical value of Kolmogorov-Smirnov statistic of $\Delta F = 0.0717$, for the sample size $N = 360$. Thus the 79 series of ξ of the area can be considered as independent identically distributed random variable, defined by Eq. (3.35).

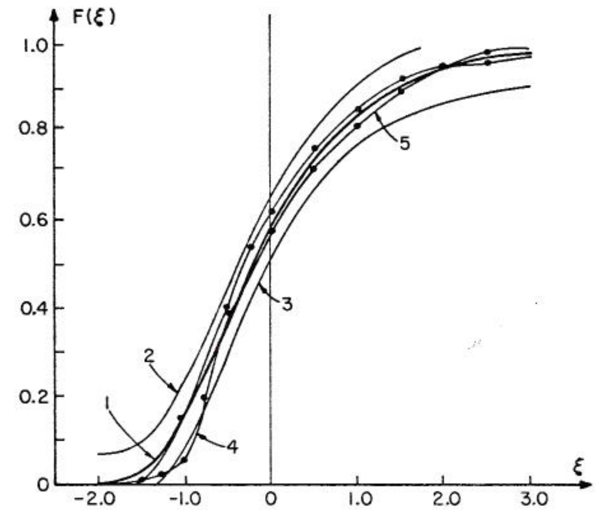


Fig. 3.11. Theoretical Distribution of ξ (1), the Corresponding 95 Percent Tolerance Limits (2) and (3) for the Use of Kolmogorov-Smirnov Statistic Tests, and Two Observed Frequency Distributions of Station 1 (4) and Station 52 (5).

Regional Dependence Structure for Stochastic Component. The regional dependence of the identically distributed series of stationary stochastic variables is analyzed by using the lag-zero cross correlation coefficients as related to the interstation distance d and the orientation ϕ. The two models discussed in Section 3.5 are studied for the standardized stochastic component ξ and the transformed series ζ, which are defined in the previous Section. Figures 3.12, 3.13, and 3.14 present relations of the (r,d)-point and figures of isolines of the interstation cross correlation coefficients from a given center station. The correlation structure is not isotropic as shown in Fig. 3.13. In the Upper Great Plains, the decay of correlation coefficient for the north-south direction is higher than that for the east-west direction. Those figures support the two models

$$r = \exp(\beta_1 d) \tag{3.40}$$

and

$$r = \exp[(\beta_1 + \beta_2 \cos 2\phi + \beta_3 \sin 2\phi) d] \quad . \tag{3.41}$$

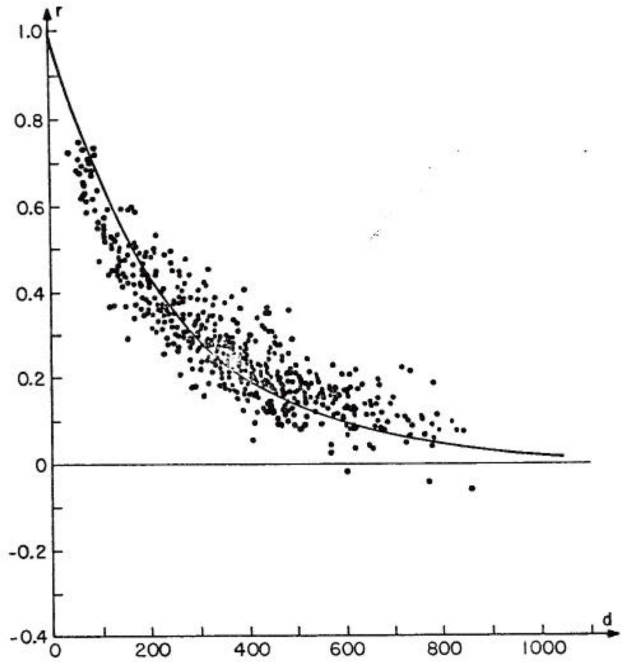


Fig. 3.12. Lag Zero Cross Correlation Coefficient, r, Versus the Interstation Distance, d, and the Fitted Function $r = \exp(-0.00418d)$ for the Stochastic Component of Monthly Precipitation.

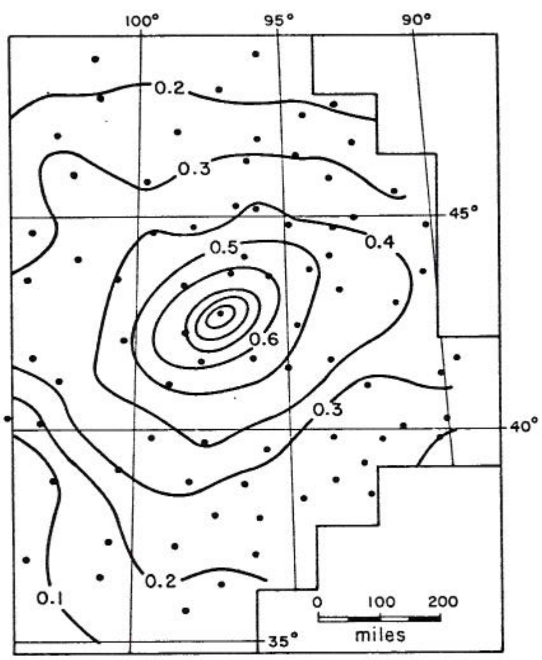


Fig. 3.13. Isocorrelation Patterns Based on Relationship Between Station 52 and All Other Stations for $(\xi - \xi_{min})^{1/3}$.

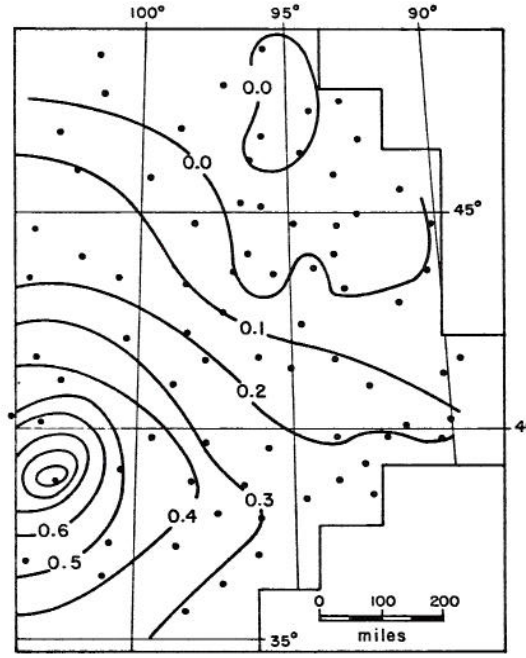


Fig. 3.14. Isocorrelation Patterns Based on Relationship Between Station 1 and All Other Stations for the Series ξ.

The unknown coefficients in Eqs. (3.40) and (3.41) are estimated by using a nonlinear least-squares fitting procedure, which is available in the Biomedical Computer Programs (Dixon, 1969). The 3081 values of r, d, and ϕ are used to estimate the coefficients. The results are summarized in Table 3.5. For the different series, the two types of regression explain over 80 percent of the variance of r. There are little differences between the estimated coefficients of the original and the transformed series, though the maximum orientation of the original series is a little bigger than those of the transformed series. The transformation of the series does not affect the correlation structure of the original series, which is the same conclusion as that made in the study by Huff and Shipp (1969). Since the differences of the explained variances by Eq. (3.41) from those by Eq. (3.40) are only three percent higher, thus introducing the complication and the additional labor by using Eq. (3.41), the increased accuracy does not seem worthwhile. Therefore, the model given by Eq. (3.40) is used in consequent investigations.

Under the circumstances, the regional dependence structure of the series υ is defined by

$$r = \exp(-0.00402d) \tag{3.42}$$

However, the regional dependence structure of the original series ξ is defined by

$$r = \exp(-0.00418d) \quad . \tag{3.43}$$

Table 3.5. Fitting Models $\rho = f(d,\phi)$ for the Regional Dependence for the Sample Size N=3081.

Series	Regression Equation	Explained Variance	ϕ_{max}	Ellipticity
ξ	$r = \exp(-0.00418d)$	80.6%		
$(\xi - \xi_{min})^{1/3}$	$r = \exp(-0.00402d)$	80.3%		
$(\xi + 1.7653)^{1/3}$	$r = \exp(-0.00410d)$	79.3%		
ξ	$r = \exp[(-0.00416 + 0.00048\cos2\phi + 0.00023\sin2\phi)d]$	83.1%	12.98	0.773
$(\xi - \xi_{min})^{1/3}$	$r = \exp[(-0.00402 + 0.00055\cos2\phi + 0.00006\sin2\phi)d]$	83.4%	3.05	0.758
$(\xi - \xi_{min})^{1/3}$	$r = 0.922\exp(-0.00404d) + 0.078$	82.7%		

Chapter 4
MULTIVARIATE DATA GENERATION AT A NEW GRID OF POINTS

The data generation or the experimental Monte Carlo method is useful when problems are too complicated to be analyzed analytically. In hydrologic problems associated with area and time or space and time, the main concern is for the areal or spatial distribution and the variation in time. Since the present observational points are located irregularly over an area, representing different area sizes, the estimates of areal distributions are made by drawing the isohyetal maps and measuring enclosed areas by a planimeter. This is time consuming work, subject to some errors. The generation of new samples of data at a new systematic grid of points can help to solve the above problem, by using the models as discussed in Chapter III, since the models condense the area-time information. Hence, the solution of the problems involved may be dissociated from the set of observational stations.

4.1. Multivariate Generation Method

To generate hydrologic series at points over an area, or generate multivariate sample series, the cross correlation coefficients between stations over the area must be considered as well as the series means, variances, skewness coefficients, and serial correlation coefficients (Fiering, 1964; Matalas, 1967). A rather simple way of generating multivariate series is based on a multivariate second-order stationary generating process, which is discussed by Matalas (1967) and Young and Pisano (1968).

The multivariate generation of samples of the stochastic component of the monthly precipitation, or ξ, is the subject of this study, and its model is then

$$\underline{\xi}_{j+1} = \underline{A}\underline{\xi}_j + \underline{B}\underline{\upsilon}_{j+1} \tag{4.1}$$

with j the month, $\underline{A}$ and $\underline{B}$ the $(M \times M)$ diagonal matrices, $\underline{\xi}_{j+1}$ a $(M \times 1)$ matrix. The variable ξ should be considered normally distributed with mean zero and variance unity. When ξ is not normally distributed, transformations such as logarithmic and cube-root are made to approach a normal distribution as closely as possible. Therefore, ξ and υ follow the standard normal distribution, and $E(\xi) = 0$, $E(\upsilon) = 0$, $\mathrm{Var}\ (\xi) = 1$, and $\mathrm{Var}\ (\upsilon) = 1$. The matrices $\underline{A}$ and $\underline{B}$ are given by

$$\underline{A} = R_1 R_0^{-1} \tag{4.2}$$

and

$$\underline{B}\ \underline{B}^T = R_0 - R_1 R_0^{-1} R_1^T \tag{4.3}$$

where R_0 = the lag-zero cross correlation matrix of ξ, R_1 = the lag-one serial correlation matrix, R_0^{-1} = the inverse of R_0, and $\underline{B}^T$ and R_1^T = the transposes of $\underline{B}$ and R_1, respectively. Equations (4.2) and (4.3) define the coefficients of matrices $\underline{A}$ and $\underline{B}$. The solution for $\underline{B}$ can be obtained either by orthogonalization or recursive scheme technique, or by principal component technique (Matalas, 1967; Young and Pisano, 1968).

In general, the stochastic component of the monthly precipitation can be assumed to be independent, then Eqs. (4.2) and (4.3) are reduced to the following forms

$$\underline{A} = 0 \tag{4.4}$$

and

$$\underline{B}\ \underline{B}^T = R_0\ . \tag{4.5}$$

For the generation at a new grid point instead of the original stations, the lag-zero cross correlation matrix R_0 can be defined by regional dependence model such as Eq. (3.42).

4.2. Determination of Grid System

One of the important problems in using the grid system is how to define the system, especially the grid interval which determines, to a large extent, the accuracy of the representation.

The objective and general method of designing the grid network has not been established. Determination of the grid system may depend upon the phenomenon studied, the scale of area studied, the needed accuracy of results, and economic restrictions. Among several kinds of grid systems such as square, rectangular, and triangular ones, a square grid system is exclusively considered in this study because the correlation structure in the Upper Great Plains is assumed to be isotropic. If the regional dependence structure is anisotropic or elliptic like that of Eq. (3.14), a rectangular grid system with the major and minor axes of the ellipse as the sides of the rectangles may be better to give the equal weights to information available for all directions.

Considering the scale of the area, the number of the original stations used in modeling, and a capacity of the available computer, the number of the grid points in this study was selected to be around one hundred, giving a grid interval of about one hundred miles. On the basis of these considerations, the grid interval is decided by the criterion that a correlation coefficient r between the value at a grid point and the average value of the area represented by the grid point is higher than a certain value. From purely statistical grounds, one might plausibly choose an acceptable value of the coefficient of determination such as $r^2 = 0.8$ and 0.9. The procedure of obtaining this correlation coefficient is as follows. A $(L \times L)$ - square area consisting of $(\ell \times \ell)$ smaller squares, as shown in Fig. 4.1, is considered to be an area represented by a grid point located in the middle area. Assuming that each small area $(\ell \times \ell)$ is homogeneous and the point in the middle of each area represents that area properly, the mean value of the big square area is given by the average of values of these small squares. Let ξ_i and ξ_c denote the value of the small square i and the value of the representative point c, respectively. The number of the small squares is denoted by n. Then the average value $\bar{\xi}$ is given by

$$\bar{\xi} = \frac{1}{n} \sum_{i=1}^{n} \xi_i\ . \tag{4.6}$$

Since ξ_i is standardized, the mean, variance, and covariance are given by $E(\xi_i) = 0$, $\mathrm{Var}\ (\xi_i) = 1$,

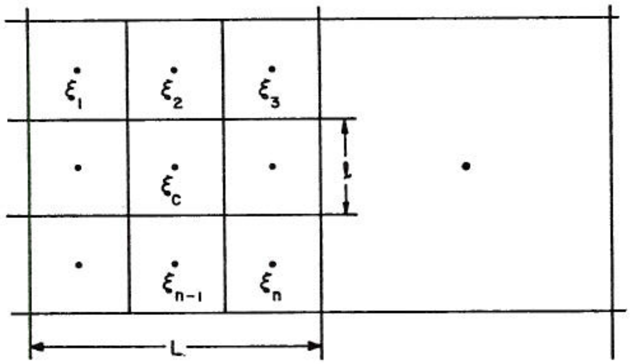


Fig. 4.1. Graph for Determining the Grid Interval L on the Basis of the Correlation Coefficient of ξ_c and the Average Value $\bar{\xi} = \frac{1}{n}\sum_{i=1}^{n} \xi_i$, where ξ_i Represents Values of the Small (ℓ x ℓ) Squares.

and Cov $(\xi_i,\xi_j) = r_{ij}$, where r_{ij} is the cross correlation coefficient given by Eq. (3.43). The mean and variance of $\bar{\xi}$ and the covariance of $\bar{\xi}$ and ξ_c are given by

$$E(\bar{\xi}) = 0 \quad , \tag{4.7}$$

$$\mathrm{Var}(\bar{\xi}) = \mathrm{Var}[\frac{1}{n}(\xi_1 + \xi_2 + \cdots + \xi_n)]$$

$$= \frac{1}{n} + \frac{2}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} r_{ij} \quad , \tag{4.8}$$

and

$$\mathrm{Cov}(\xi_c,\bar{\xi}) = \mathrm{Cov}[\xi_c,\frac{1}{n}(\xi_1 + \xi_2 + \cdots + \xi_n)]$$

$$= \frac{1}{n}\sum_{i=1}^{n} r_{ci} \quad . \tag{4.9}$$

Then the correlation coefficient of ξ_c and $\bar{\xi}$ is

$$r = \mathrm{Cov}(\xi_c,\bar{\xi})/[\mathrm{Var}(\xi_c)\mathrm{Var}(\bar{\xi})]^{1/2}$$

$$= \mathrm{Cov}(\xi_c,\bar{\xi})/[\mathrm{Var}(\bar{\xi})]^{1/2} \quad . \tag{4.10}$$

Because the value of ten miles for ℓ can be reasonable for the monthly precipitation, the correlation coefficients are calculated for a several set of n and L with ℓ = 10 miles. The results are shown in Table 4.1. When $r^2 = 0.9$ is taken as the criterion to define the grid system, the grid interval of 100 miles is chosen from Table 4.1. For $r^2 = 0.8$, the grid interval of 200 miles is chosen. Considering these values as well as the previously described conditions, the 100-mile grid interval is chosen in this study.

To cover the study area, a (10 x 8) square grid system with 100-mile grid interval is chosen for the generation of new samples and the investigation of drought characteristics. Each station represents an area of 10,000 square miles, which is called here a *unit-area*. The grid system covers a total of 800,000 square miles, or 80 unit-areas. Figure 4.2 shows the grid system over the area studied and the location of the new eightly stations or points.

Table 4.1. Correlation Coefficients of the Areal Representative Value ξ_c and its Areal Average $\bar{\xi}$ for Various Grid Interval L Consisting of n small (ℓ x ℓ) Square Miles Areas.

n	ℓ	L	r	r^2
9	10.0	30.0	0.9861	0.9723
25	10.0	50.0	0.9759	0.9524
49	10.0	70.0	0.9661	0.9334
81	10.0	90.0	0.9563	0.9145
121	10.0	110.0	0.9467	0.8962
169	10.0	130.0	0.9372	0.8783
225	10.0	150.0	0.9275	0.8602
289	10.0	170.0	0.9181	0.8428
361	10.0	190.0	0.9086	0.8255

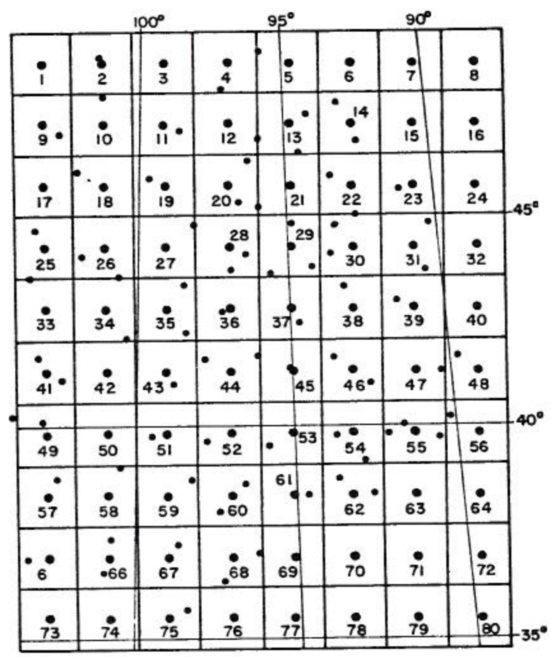


Fig. 4.2. Grid System (Big Points) with 100-Mile Grid Interval over the Upper Great Plains and Observed Stations (Small Points) with Their Series Used for Objectives of Modeling.

4.3. Checking the Generated Samples

The sample series of ξ, each 1200 months long, are generated at the selected 80 grid points by preserving the mean, variance, and lag-zero cross correlation coefficient, as defined by Eqs. (3.35) and (3.43). The mean, standard deviation, skewness coefficient, excess coefficient, and the minimum value for the generated samples of the 80 stations, each sample 500 months long, are given in Table 4.2. The generated samples could be accepted without a rigorous statistical test, because the multivariate generation of

Table 4.2. Statistics of Generated Samples Each of N = 500 Months, with Extremes Underlined, and Values Outside the 95% Tolerance Limits with the Sign*.

Station	Mean	Standard Deviation	Skewness Coefficient	Excess Coefficient	Minimum
1	-0.0092	1.0143	1.0753	1.6518	-1.7459
2	-0.0556	1.0130	1.4441	4.8783	-1.7498
3	-0.0099	1.0080	0.9464	0.9979	-1.7480
4	-0.0218	1.0780	1.3290	2.5068	-1.6758
5	-0.0106	1.0752	1.2288	1.8722	-1.7073
6	0.0145	1.1113*	1.4363	3.0546	-1.7081
7	-0.0162	1.0263	1.3321	3.0162	-1.7243
8	-0.0525	0.9432*	1.2666	3.4078	-1.6223
9	0.0006	1.0162	1.0291	1.1354	-1.6939
10	0.0064	1.0314	1.0755	1.7220	-1.7123
11	0.0451	1.0398	0.8527	0.7601	-1.6607
12	-0.0157	1.0187	1.2112	2.0770	-1.7117
13	-0.0107	0.9891	0.9321	1.1784	-1.6105
14	-0.0323	1.0259	0.9910	1.1056	-1.6380
15	-0.0102	1.0659	1.5958	4.4568	-1.6138
16	0.0127	1.0977	1.4840	3.9648	-1.6379
17	0.0283	1.0554	1.1003	1.7790	-1.7053
18	0.0401	1.0719	1.1637	2.7489	-1.7148
19	0.0191	1.0403	1.1568	2.2538	-1.7135
20	0.0456	1.0661	1.1845	3.7411	-1.7138
21	0.0605	1.0292	1.0108	2.2834	-1.6235
22	0.0214	1.0007	0.9786	1.1286	-1.6970
23	0.0205	1.0432	1.2424	2.1849	-1.7383
24	0.0583	1.1234*	1.4594	3.5526	-1.5929
25	0.0403	1.0983	1.2557	2.2994	-1.6675
26	0.0554	1.1430*	2.0074	8.7175	-1.7003
27	0.0540	1.1430*	1.4288	3.1851	-1.7147
28	0.0880*	1.0569	0.8362	0.4613	-1.7646
29	0.0775	1.0114	0.9030	1.2285	-1.7639
30	0.1219*	1.0469	0.9191	1.0467	-1.6655
31	0.0487	1.0651	1.1486	2.0287	-1.6767
32	0.0121	1.0797	1.3614	2.4110	-1.6554
33	0.0730	1.0951	1.0044	1.0339	-1.7539
34	0.0824	1.1138*	1.3459	2.8359	-1.7206
35	0.0727	1.0092	0.9388	1.2501	-1.7368
36	0.1050*	1.0695	1.0547	1.3829	-1.7222
37	0.1128*	1.0806	1.0994	1.6346	-1.7590
38	0.1192*	1.0577	0.9681	0.9507	-1.7440
39	0.0275	0.9962	1.0742	1.4130	-1.6443
40	-0.0226	1.0248	1.5394	3.6085	-1.5218
41	0.0624	1.0913	1.1138	1.4827	-1.7219
42	0.1078*	1.0294	0.9604	1.1568	-1.7631
43	0.1260*	0.9931	0.7483	0.5507	-1.6633
44	0.0898*	1.0584	0.9789	1.0333	-1.7560
45	0.0733	1.1176*	1.3619	2.3754	-1.5759
46	0.0854	1.0734	1.3244	2.4630	-1.6270
47	0.0528	1.0439	1.3168	2.8399	-1.6480
48	-0.0339	0.9573	1.3622	3.5096	-1.6764
49	0.0884*	1.0732	0.9714	0.9643	-1.6662
50	0.0453	1.0260	0.8745	0.4671	-1.7059
51	0.0409	1.0396	0.8772	0.6285	-1.7155
52	0.0263	0.9980	0.8687	0.7992	-1.7032
53	0.0207	1.0919	1.4936	3.3643	-1.7480
54	0.0453	1.0799	1.2616	2.8006	-1.6501
55	0.0203	1.0671	1.1695	1.6651	-1.6326
56	-0.0071	0.9952	1.1775	2.2717	-1.7160
57	0.0246	1.0744	1.2268	2.1626	-1.6512
58	-0.0302	0.9802	0.8889	0.9994	-1.6672
59	-0.0019	0.9477*	0.8000	0.8283	-1.5826
60	-0.0101	0.9895	0.8964	1.0881	-1.6943
61	-0.0191	1.0921	1.3049	2.3381	-1.6223
62	0.0075	1.0575	0.9993	0.7510	-1.6900
63	0.0205	1.0621	1.1519	1.4333	-1.6302
64	0.0154	1.0019	1.0248	1.1351	-1.5703
65	0.0144	1.0408	1.3217	2.8499	-1.6368
66	-0.0195	1.0425	1.2620	2.3821	-1.7068
67	0.0099	1.0317	1.0599	1.2130	-1.6868
68	-0.0226	1.0046	0.9950	1.1571	-1.6305
69	-0.0191	1.0222	1.1140	1.7290	-1.6511
70	-0.0099	0.9974	1.0198	1.3643	-1.7067
71	-0.0096	1.0353	1.0398	0.9857	-1.6963
72	0.0304	1.0652	1.3349	2.4434	-1.6549
73	0.0158	1.0350	0.9232	0.9390	-1.6832
74	0.0075	1.0763	1.2237	1.8539	-1.7577
75	-0.0066	1.0972	1.4046	2.7600	-1.6779
76	0.0165	1.0941	1.4824	3.4468	-1.7132
77	-0.0135	1.0044	1.2262	3.0616	-1.6279
78	-0.0047	1.0231	1.0212	1.2794	-1.7146
79	0.0124	1.0768	1.1739	2.0796	-1.7410
80	0.0199	1.0466	1.3887	2.9847	-1.6796
Mean	0.0259	1.0464	1.1569	2.0561	-1.6838
S.D.	0.0424	0.0421	0.2202	1.2476	0.0513
Observed Mean	0.0000	1.0000	0.9737	0.8878	-1.7653

samples has been used for a long time. However, since the method was applied to generate many samples of point series over a large area, simple statistical tests were carried out to check whether the generated samples preserved the inferred statistical properties. For testing the preservation of the mean, standard deviation, serial correlation coefficient, and cross correlation coefficient, tolerance limits at the 95 percent level with the sample size N = 500 were used.

For the mean, the 95 percent tolerance limits, m_u and m_ℓ, are given by

$$m_{u,\ell} = \mu \pm 1.96\sigma/\sqrt{500} \tag{4.11}$$

For $\mu = -0.0054$ and $\sigma = 1.0397$, then $m_u = 0.0857$ and $m_\ell = -0.0965$. In case of the standard deviation, the 95 percent tolerance interval is defined by

$$\frac{\sigma^2\chi^2_{0.025}}{499} < s^2 < \frac{\sigma^2\chi^2_{0.975}}{499} \tag{4.12}$$

For $\sigma = 1.0397$, $\chi^2_{0.025} = 439.0701$, and $\chi^2_{0.975} = 562.7223$, then Eq. (4.12) gives

$$0.9375 < s < 1.0615 \tag{4.13}$$

The 95 percent tolerance limits for the serial correlation coefficients with the lag k are given by Eq. (3.26), namely

$$r_{u,\ell} = \frac{-1 \pm 1.96\sqrt{498-k}}{499 - k} \quad . \tag{4.14}$$

The tolerance limits for the zero lag cross correlation coefficient as related to distance d, defined by Eq. (3.43), was found by transforming the estimated ρ values into the Fisher's z variable. The 95 percent tolerance limits for z are

$$z_{u,\ell} = z_\rho \pm 1.96\sigma_z \ , \tag{4.15}$$

where $z_\rho = \tanh^{-1}\rho$, $\sigma_z = 1/\sqrt{497}$, and $\rho = \exp(-0.00418d)$. Then these 95 percent tolerance limits are converted back into $r_{u,\ell}(d)$, by

$$r_{u,\ell}(d) = \tanh z_{u,\ell} \quad . \tag{4.16}$$

Using these above tolerance limits, the statistics are tested. For the means, nine out of 80 stations, or 12.5 percent of stations, showed the values outside the 95 percent tolerance limits. For the standard deviation, eight out of 80 stations, or ten percent of total stations, were outside the tolerance limits. Though for both statistics the numbers of stations outside the 95 percent tolerance limits were somewhat larger than four (5 percent of 80 stations), the deviations are not very large.

Table 4.3 shows the number of stations, whose serial correlation coefficients are outside the 95 percent tolerance limits. Though the serial correlation coefficient for the lag three is often outside the tolerance limits, the generated samples as a whole were considered to be time independent, because the total percentage of the stations with the r_k outside the tolerance limits was small, 3.88 percent. Figure 4.3 presents some randomly selected cross correlation coefficients of the generated samples versus the interstation distance, with the original regional dependence function and its corresponding 95 percent tolerance limits. The points outside the tolerance limits are few. Therefore, the regional dependence structure was preserved in the generated samples. The time and regional structures were well preserved.

Table 4.3. Number of Serial Correlation Coefficient of the 80 Generated Samples, Which are Outside the 95 Percent Tolerance Limits of the Independent Series.

Lags	Number of Stations Outside 95% T. L.
1	2
2	3
3	12 (15.0%)
4	2
5	2
6	7
7	2
8	1
9	1
10	0
11	2
12	3
13	1
14	4
15	3
16	3
17	1
18	7
19	3
20	3
Total	62 (3.88%)

The average of the skewness coefficients for the generated samples was greater than that of the original series. The excess coefficients for the generated samples were very scattered, with their average value

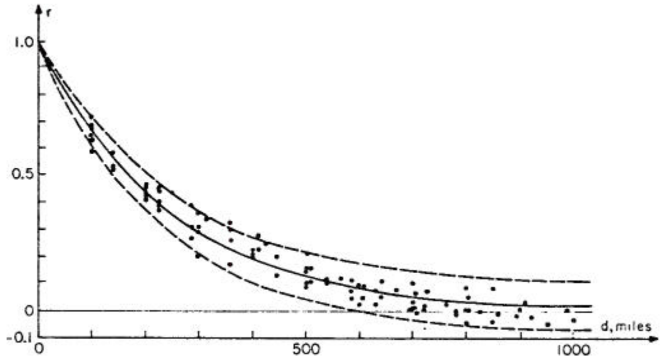


Fig. 4.3. Randomly Selected Cross Correlation Coefficients of Generated Samples Versus the Interstation Distance, with the Original Regional Dependence Function and its Corresponding 95 Percent Tolerance Limits.

much greater than the average value of the original series. The difference between the minima of the generated and original samples was not large, but their standard deviation was different, which may come from the model of the stochastic component defined by Eq. (3.35). Border effects on generation seem to exist in locations of occurrences of extreme values. That is, the extreme values seem to occur around the border more often than in the middle of the area. However, a rigorous statistical test was not carried out.

In view of these tests and considerations, the generated samples for the stochastic component of the monthly precipitation series can be accepted as approximately a time independent, second-order stationary process, with the regional dependence defined by Eq. (3.43).

Chapter 5
EXPERIMENTAL METHOD OF ANALYSIS OF AREAL DROUGHT CHARACTERISTICS

The monthly precipitation is a periodic-stochastic process, because each month has a different mean and a different standard deviation. Therefore, the time position inside the year is an important factor in the analysis of discrete time series with their interval a fraction of the year. The stochastic component inside the monthly precipitation series is approximately a second-order stationary, identically distributed stochastic process in most cases. In this Chapter, only the stochastic component is analyzed in order to derive general characteristics of regional droughts. Drought characteristics of monthly precipitation are discussed in Chapter VI.

The generated samples at the 80 station points of the i.i.d. process of ξ, with the length of 1200 monthly values for each sample, are based on the models developed in Chapter III. These samples are then used to investigate the areal drought characteristics. As shown in Fig. 4.2, these 80 stations of the selected grid cover an area larger than the area of the Upper Great Plains, from whose observational stations the models were inferred. Since the coverage of this larger area does not seem to influence the general characteristics obtained on droughts of the Upper Great Plains, the generated series at the 80 stations are used without further modification.

5.1. Definition of Droughts and Development of Indices of Drought Characteristics

In this study, a drought is defined by using the differences between the water supply and water demand series. The water supply series are defined by the stochastic component in the monthly precipitation, which is assumed to be a time independent stationary stochastic process. The water demand series are given either by the mean of the stochastic component or by its probability quantiles as

$$q = F(\xi_0) = P(\xi \leq \xi_0) \quad . \tag{5.1}$$

To analyze the general drought characteristics, the mean of the stochastic component may be used as the truncation level. For a standardized random input process, the water demand is then $\xi_0 = 0$. The other truncation levels, such as $\xi_0 = F^{-1}(0.5)$, are used for comparisons, with all the truncation levels used as water demands being time invariant. In practice, truncation levels at the mean or the median are important as the benchmark levels.

Indices expressing drought characteristics should be defined before an analysis. Some of them are listed up by Yevjevich (1967) and Kates (1971), such as (1) magnitude, (2) duration, (3) areal coverage, (4) intensity (maximum, average), (5) spatial distribution, (6) drought initiation or termination, etc. The capability for drought prediction on a medium or long range is very limited if not zero. By using a statistical approach, the first four characteristics of drought are investigated in this chapter, particularly the areal coverage and drought severity over that area, with their probabilities and time durations. First, the areal drought characteristics for a month are studied without any consideration of drought duration. Then, the time factor of areal drought characteristics is investigated.

For the areal drought characteristics, the three indices selected are: the deficit area, A, the total areal deficit, D, and the maximum deficit intensity, I. For the truncation level, ξ_0, these indices as random variables are defined as

$$A = \sum_{i=1}^{80} I_{(\xi \leq \xi_0)}(\xi_i) \quad , \tag{5.2}$$

$$D = \sum_{i=1}^{80} (\xi_0 - \xi_i) I_{(\xi \leq \xi_0)}(\xi_i) \quad , \tag{5.3}$$

and

$$I = \xi_0 - \min(\xi_1, \xi_2, \cdots, \xi_{80}, \xi_0) \tag{5.4}$$

where $I_{(\xi \leq \xi_0)}(\xi_i)$ is an indicator function defined by

$$\begin{aligned} I_{(\xi \leq \xi_0)}(\xi_i) &= 1 \qquad \text{if } \xi_i \leq \xi_0 \\ &= 0 \qquad \text{if } \xi_i > \xi_0 \quad , \end{aligned} \tag{5.5}$$

and i = the station number. According to definitions, the deficit area does not express how the deficits are distributed over the whole area. The total areal deficit, therefore, does not take into account the spatial distribution of the deficit. Though these variables do not present the complete information on regional droughts, they are primary factors in drought investigations for a large area. Other viewpoints can be applied in further investigations on the areal aspects of droughts.

The indices defined by Eqs. (5.2) through (5.5) can be considered as non-negative, bounded random variables. The deficit area, A, is bounded by zero and 80. Since the probability of $P(\xi \leq -1.7653)$ is less than 0.0001, the highest maximum deficit intensity at the truncation level ξ_0 may be conceived as $(1.7653 + \xi_0)$. Therefore, the maximum deficit intensity, I, may be considered as bounded by zero and $(1.7653 + \xi_0)$. A product of the highest maximum deficit intensity and the whole area, that is, $80(1.7653 + \xi_0)$, gives the maximum possible total areal deficit. The total areal deficit, D, is bounded by zero and $80(1.7653 + \xi_0)$. Although the deficit area is defined as a discrete variable by Eq. (5.2), the deficit area should be a continuous random variable.

5.2. Statistical Analyses of Drought Characteristics

The basic statistics of the three drought characteristics (the deficit area, the total areal deficit, and the maximum deficit intensity) are computed from the 1200-month generated samples for different truncation levels. The results are presented in Table 5.1. The variation of basic statistics of distributions of the deficit area and the total areal deficit, as functions of the truncation level, are shown in Fig. 5.1. The distribution statistics of the deficit area are expressed in percentage of the total whole area (or of 80 stations). The mean of the deficit area increases with the probability of the truncation level, as expected. The

Table 5.1. Basic Statistics of the Deficit Area, the Total Areal Deficit, and the Maximum Deficit Intensity for the Sample Size N = 1200.

		Truncation Level			
	q	0.4	0.5	0.58	0.65
	ξ_0	-0.4340	-0.2028	0.0000	0.1987
Deficit Area in Percent	Mean	38.9000	49.0330	57.0187	64.0198
	St. Dev.	18.8658	19.4985	19.0549	18.2022
	Maximum	92.90	96.25	98.75	100.00
	Minimum	1.25	3.75	7.50	7.50
	Skewness	0.3763	0.1201	-0.6731	-0.2909
	Excess	-0.4868	-0.6544	-0.6731	-0.5786
Total Areal Deficit	Mean	14.8555	23.0262	31.6495	41.2836
	St. Dev.	10.2524	13.6123	16.5516	19.3148
	Maximum	63.5489	80.5565	96.0696	111.5682
	Minimum	0.0374	0.7310	1.9529	4.2244
	Skewness	1.1588	0.9028	0.7240	0.4679
	Excess	1.4293	0.6572	0.2606	-0.0259
Maximum Deficit Intensity	Mean	0.9756	1.2074	1.4102	1.6089
	St. Dev.	0.2080	0.2080	0.2080	0.2080
	Maximum	1.3311	1.4623	1.7651	1.9638
	Minimum	0.0214	0.2527	0.4554	0.6542
	Skewness	-0.9033	-0.9033	-0.9033	-0.9033
	Excess	1.1296	1.1296	1.1296	1.1296

standard deviation of the deficit area is symmetrical about the point q = 0.5. The highest variation of the deficit area is found for the truncation level at the median. The mean of the total areal deficit increases exponentially with an increase of the truncation level.

A comparison of the maximum and minimum percentages of the deficit area for various truncation levels indicates that at the truncation levels of q = 0.5 and q = 0.58, or at the median and the mean, respectively, the deficit never covers the whole area completely. On the other hand, at least a small deficit area always occurs within the whole area. Though this does not prove that the probabilities of the two extremes, of the whole area completely covered by a drought and no part of the whole area covered by a drought, are zeros, these two extremes very rarely occur for such a large region, as used in this study, for the truncation levels of the median and the mean. The fact that there is always at least a small deficit area covered by a drought within the area studied, is a specific feature of droughts. The three variables have time independent sequences, as expected, because the original stochastic component series is time independent.

The probability distributions are fitted to the frequency distributions of the three variables for the truncation level of ξ_0 = 0 (q = 0.58). The beta distribution function is used to these frequency distributions of the three variables, because they are assumed to be bounded at both tails, as specified in the previous section. The bounds for the three variables are given in Table 5.2. The weak point in the fit of the beta probability distribution function to the frequency distribution of these three variables is that the probability densities at both bounds are zeros, though the probabilities for the bounds may

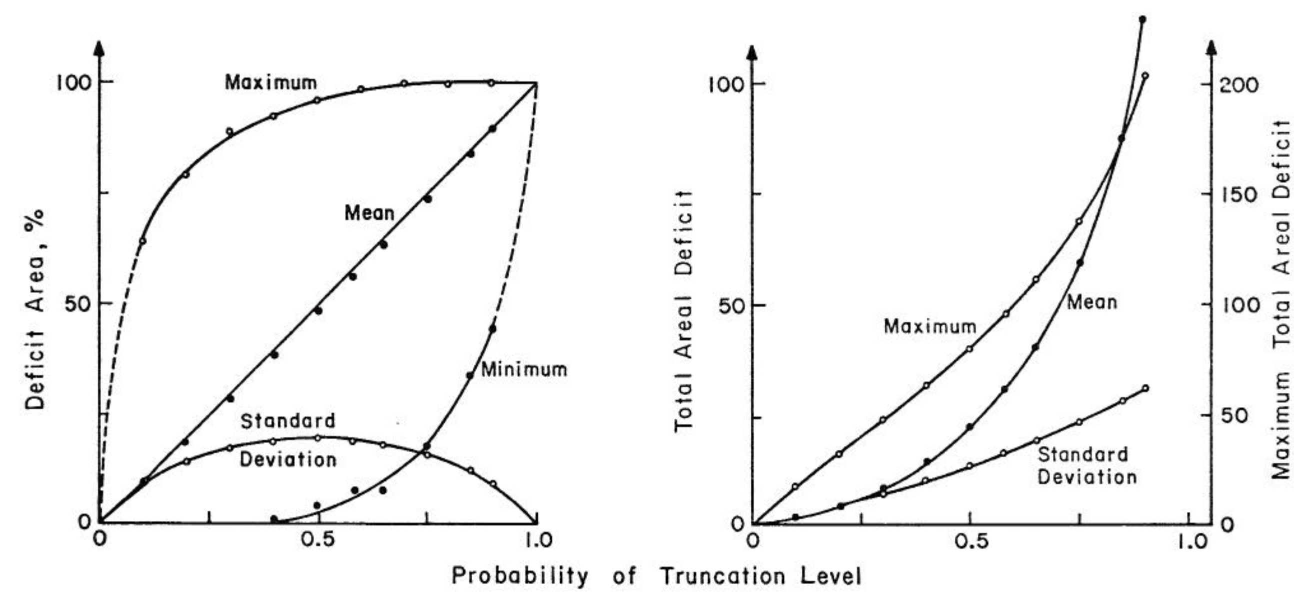


Fig. 5.1. Variations of Basic Statistics of the Deficit Area and the Total Areal Deficit in Function of the Probability of Truncation Level for the Sample Size N=600.

Table 5.2. Boundary Values of the Deficit Area, A, the Total Areal Deficit, D, and the Maximum Deficit Intensity, I, for the Truncation Level of $\xi_0 = 0$.

Variables	Lower Bound	Upper Bound
A	0	80.0
D	0	141.224
I	0	1.7653

not be zeros. Since probabilities of both extremes are close to zero, as mentioned earlier, and the objectives of this study were not to analyze these extremes, this weak point is not decisive so that the use of the beta distribution may be accepted. When this weak point is not negligible, a mixed distribution could be used.

The general form of the beta distribution function is

$$f_Y(y) = \frac{1}{B(\alpha,\beta)} \frac{(y-a)^{\alpha-1}(b-y)^{\beta-1}}{(b-a)^{\alpha+\beta-1}} \quad , \qquad (5.6)$$

with $\alpha>0$, $\beta>0$, and $B(\alpha,\beta)$ = the beta function of α and β. By using the transformation

$$X = (Y - a)/(b - a) \quad , \qquad (5.7)$$

the standard form of the beta distribution with parameters α and β becomes

$$f_X(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1 - x)^{\beta-1} \; , \; (0 < x < 1). \qquad (5.8)$$

Johnson and Kotz (1970b) gives estimates $\hat{\alpha}$ and $\hat{\beta}$, as the first approximation as

$$\hat{\alpha} = \frac{1}{2}\Big(1 - \prod_{i=1}^{n} x_i^{1/n}\Big)\Big/\Big[1 - \prod_{i=1}^{n} x_i^{1/n} - \prod_{i=1}^{n} (1 - x_i)^{1/n}\Big] \; , \qquad (5.9)$$

and

$$\hat{\beta} = \frac{1}{2}\Big[1 - \prod_{i=1}^{n} (1 - x_i)^{1/n}\Big]\Big/\Big[1 - \prod_{i=1}^{n} x_i^{1/n} - \prod_{i=1}^{n} (1 - x_i)^{1/n}\Big]. \qquad (5.10)$$

Using the transformation of Eq. (5.7) with the values in Table 5.2, the estimates $\hat{\alpha}$ and $\hat{\beta}$ for the three variables are made by using Eqs. (5.9) and (5.10). The goodnesses of the fit of the beta distribution to the computed frequency distributions of the deficit area, the total areal deficit, and the maximum deficit intensity, are tested by using the chi-square statistic of ten class intervals at the five percent level of significance. At this significance, the critical value of chi-square statistic with the number of degrees of freedom of seven is 14.1 The estimates $\hat{\alpha}$ and $\hat{\beta}$ and the test are as follows.

For the deficit area, the estimates $\hat{\alpha}$ and $\hat{\beta}$ are 3.417 and 2.559, respectively, with the beta density function

$$f(x) = 27.6128\, x^{2.417}\,(1 - x)^{1.559} \quad . \qquad (5.11)$$

Its chi-square statistic is 12.80, which is less than the critical value of 14.1.

For the total areal deficit, $\hat{\alpha}$ and $\hat{\beta}$ estimates are 2.760 and 9.566, respectively, with the beta density function

$$f(y) = 403.4121\, y^{1.760}\,(1 - y)^{8.566} \quad , \qquad (5.12)$$

with the chi-square statistic of 9.33.

For the maximum deficit intensity, the estimated density function with the parameters 8.314 and 2.092 is

$$f(z) = 91.6392\, z^{7.314}\,(1 - z)^{1.092} \qquad (5.13)$$

and the chi-square statistic of 11.70, being smaller than the critical value.

Since the chi-square statistics for all the three variables are smaller than the critical value at the five percent level of significance, the fitted distributions may be accepted. Comparisons of fitted distribution functions and the frequency distributions of these three variables are shown in Fig. 5.2.

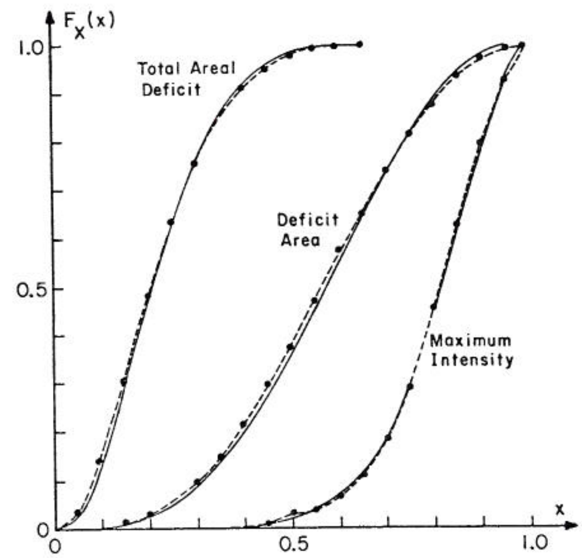


Fig. 5.2. Relative Frequency Distribution (Broken Lines) and the Fitted Beta Distribution Functions (Solid Lines) for the Deficit Area, the Total Areal Deficit, and the Maximum Deficit Intensity.

The cross correlation and regression analysis are carried out to find relationships among the deficit area, A, the total areal deficit, D, and the maximum deficit intensity, I. The lag zero cross correlation coefficients among the pairs of the three variables are given in Table 5.3, and relationships are presented in Figs. 5.3 through 5.6. It is seen that the pairs of the three variables are highly correlated, especially the correlation coefficient of the deficit area and the total areal deficit is high, namely 0.934. Since Fig. 5.3 shows that the relation is non-linear (quadratic or exponential) rather than linear, two types of regression functions are fitted, namely

Table 5.3. Pairwise Cross Correlation Coefficients for the Deficit Area, A, the Total Areal Deficit, D, the Maximum Deficit Intensity, I, and Their Derived Variables, for the Sample Size N=600.

	A	D	I	AI	D/A	I/D	DI	I/A
A	1.000							
D	0.936	1.000						
I	0.660	0.719	1.000					
AI	0.973	0.963	0.789	1.000				
D/A	0.741	0.900	0.797	0.821	1.000			
I/D	-0.796	-0.710	-0.553	-0.744	-0.648	1.000		
DI	0.903	0.991	0.765	0.958	0.911	-0.665	1.000	
I/A	-0.730	-0.581	-0.260	-0.628	-0.366	0.877	-0.522	1.000

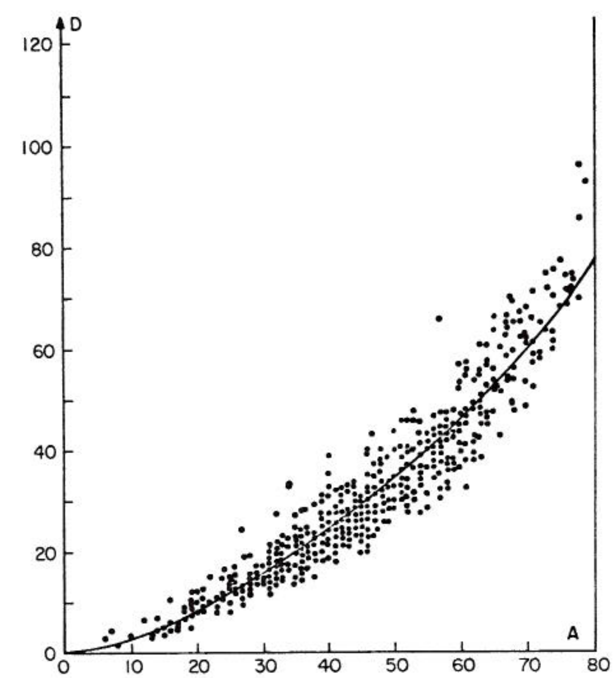


Fig. 5.3. Relationship of the Total Areal Deficit, D, to the Deficit Area, A, with the Inferred Regression Equation or $D = 0.2479A + 0.0088A^2$.

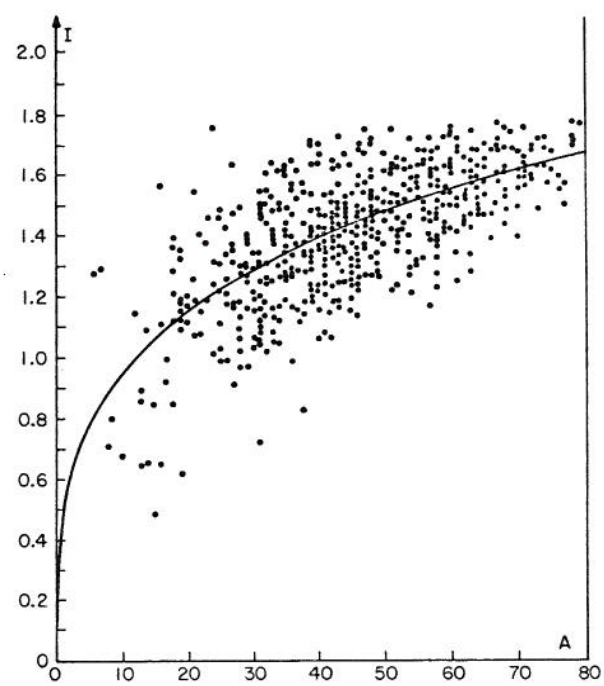


Fig. 5.4. Relationship of the Maximum Deficit Intensity, I, to the Deficit Area, A, with the Inferred Regression Equation or $I = 0.5131A^{0.2682}$.

$$D = 0.2479A + 0.0088A^2 , \tag{5.14}$$

with the explained variance of $R^2 = 0.9078$, and

$$D = 0.0558A^{1.6433} \tag{5.15}$$

with the explained variance of $R^2 = 0.9031$. As the two equations explain the variance of the total areal deficit well, the unexplained variance may result from the neglect of the shape or the spatial distribution of the deficits as well as from the sampling errors.

The relation between either the deficit area or the total areal deficit and the maximum deficit intensity, as shown by Figs. 5.4 and 5.5, are not easily inferred, though the maximum deficit intensity increases both with an increase of the deficit area and with an increase of the total areal deficit. Using the power function relationships, the fits give

$$I = 0.5131A^{0.2682} \tag{5.16}$$

and

$$I = 0.7265D^{0.1994} \tag{5.17}$$

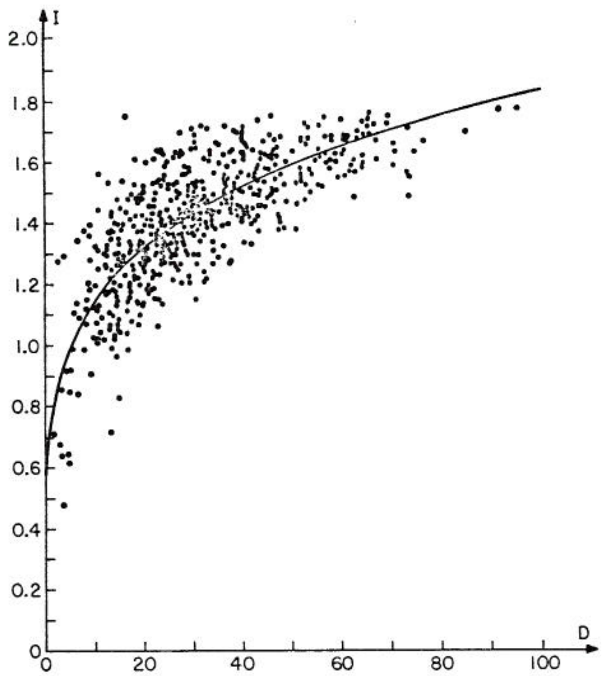


Fig. 5.5. Relationship of the Maximum Deficit Intensity, I, to the Total Areal Deficit, D, with the Inferred Regression Equation or $I = 0.7265D^{0.1994}$.

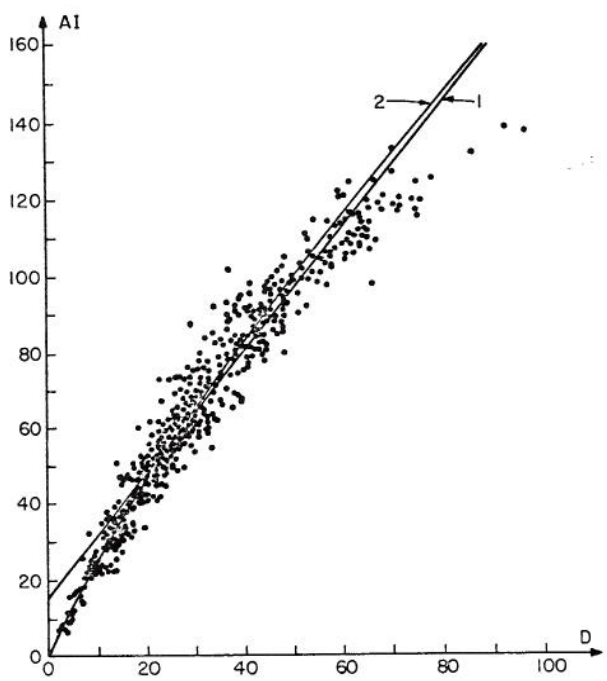


Fig. 5.6. Relationship of the Product AI to the Total Areal Deficit, D, with the Inferred Regression Equations: (1) $AI = 14.7749 + 1.6306D$ and (2) $AI = 3.6489D^{0.8457}$.

with the explained variances 0.4480 and 0.6054, respectively.

The correlation coefficient between D and AI suggests a high relationship among the variables, as shown by Fig. 5.6. The simple linear regressions between D and AI are given by

$$D = -6.1420 + 0.5684AI \tag{5.18}$$

and

$$AI = 14.7749 + 1.6306D \quad , \tag{5.19}$$

with the explained variances of 0.9266 and 0.9274, respectively. Figure 5.6 implies a power relation, as a better function to fit,

$$AI = 3.6489D^{0.8457} \quad , \tag{5.20}$$

with the explained variance of 0.9482, and with the difference from Eq. (5.19) of 0.0208, or about 2.1 percent.

Since the regression functions represent only the expected value of the dependent variable for a given value of the independent variable, the joint probability distribution may be of interest also, especially the joint probability distribution of the deficit area and the total areal deficit may be useful in practice. The joint cumulative frequency values of the deficit area, given in percentage of the whole area, and the total areal deficit are given in Table 5.4.

5.3. Trivariate Distribution

Instead of studying the joint distributions of A and I or D and I, the trivariate distribution for these three variables is studied. Joint frequency distribution of the deficit area, the total areal deficit, and the maximum deficit intensity is fitted by using the beta distribution functions with the Jacobi polynomials, because the marginal distributions for the three variables are assumed to be beta distributions as per Eqs. (5.11) through (5.13). The product of the three beta distribution functions with the Jacobi polynomials is used as an approximation to the joint, trivariate probability density function, expressed by

$$f_{X,Y,Z}(x,y,z) = f_X(x)f_Y(y)f_Z(z) \sum_i \sum_j \sum_k a_{ijk} \cdot G_i(\alpha_1,\beta_1;x)G_j(\alpha_2,\beta_2;y)G_k(\alpha_3,\beta_3;z) \tag{5.21}$$

with a_{ijk} the coefficients and $G_i(\alpha,\beta;x)$ the Jacobi polynomials of degree i with two parameters α and β.

A Jacobi polynomial of degree n, $G_n(\alpha,\beta;x)$, when expanded in the power series of x, becomes

$$G_n(\alpha,\beta;x) = \frac{\Gamma(\beta+n)}{\Gamma(\alpha+2n)} \cdot \sum_{m=0}^{n} (-1)^m \binom{n}{m} \frac{\Gamma(\alpha+2n-m)}{\Gamma(\beta+n-m)} x^{n-m} \quad . \tag{5.22}$$

For $\alpha-\beta>-1$ and $\beta>0$, the polynomials $G_n(\alpha,\beta;x)$ for $(n = 1,2,\ldots)$ form an orthogonal system in $(0,1)$, with the weight function $g(x) = x^{\beta-1}(1-x)^{\alpha-\beta}$, so that

$$\int_0^1 G_i(\alpha,\beta;x)G_n(\alpha,\beta;x)g(x)dx = 0 \quad \text{if } i \neq n \tag{5.23}$$

$$= d_n^2 \quad \text{if } i = n \; , \tag{5.24}$$

Table 5.4. Joint Cumulative Frequency of the Deficit Area and the Total Areal Deficit.

		Deficit Area (%)									
		10	20	30	40	50	60	70	80	90	100
Total Areal Deficit	100										1.0000
	90									$\uparrow$	0.9975
	80								$\uparrow$	0.9625	0.9933
	70							$\uparrow$	0.8592	0.9558	0.9767
	60						$\uparrow$	0.7142	0.8567	0.9267	0.9342
	50					$\uparrow$	0.5492	0.7125	0.8300	0.8550	$\rightarrow$
	40				$\uparrow$	0.3492	0.5467	0.6717	0.7150	0.7167	$\rightarrow$
	30			$\uparrow$	0.1933	0.3475	0.4983	0.5275	0.5283	$\rightarrow$	
	20	$\uparrow$	$\uparrow$	0.0842	0.1867	0.2550	0.2650	$\rightarrow$			
	10	0.0017	0.0192	0.0542	0.0583	$\rightarrow$					

where
$$d_n^2 = \frac{n!\Gamma(n+\alpha)\Gamma(n+\beta)\Gamma(n+\alpha-\beta+1)}{(2n+\alpha)\Gamma^2(2n+\alpha)} \quad . \qquad (5.25)$$

Coefficients a_{ijk} can be estimated by taking the expected value of $G_i(\alpha_1,\beta_1;x)\ G_j(\alpha_2,\beta_2;y)\ G_K(\alpha_3,\beta_3;z)$, that is

$$E[G_i(\alpha_1,\beta_1;x)G_j(\alpha_2,\beta_2;y)G_k(\alpha_3,\beta_3;z)]$$
$$= \int_0^1\int_0^1\int_0^1 G_i(\alpha_1,\beta_1;x)G_j(\alpha_2,\beta_2;y)G_k(\alpha_3,\beta_3;z)f(x,y,z)\,dxdydz\ . \qquad (5.26)$$

Replacing $f(x,y,z)$ by its value of Eq. (5.21), and considering that the postulated marginal distribution are beta functions

$$f_X(x) = \frac{1}{B(\beta_1,\alpha_1-\beta_1+1)}\, x^{\beta_1-1}(1-x)^{\alpha_1-\beta_1} \quad , \qquad (5.27)$$

$$f_Y(y) = \frac{1}{B(\beta_2,\alpha_2-\beta_2+1)}\, y^{\beta_2-1}(1-y)^{\alpha_2-\beta_2} \quad , \qquad (5.28)$$

and

$$f_Z(z) = \frac{1}{B(\beta_3,\alpha_3-\beta_3+1)}\, z^{\beta_3-1}(1-z)^{\alpha_3-\beta_3} \quad , \qquad (5.29)$$

then Eq. (5.26) becomes

$$E[G_i(\alpha_1,\beta_1;x)G_j(\alpha_2,\beta_2;y)G_k(\alpha_3,\beta_3;z)]$$
$$= \sum_i\sum_j\sum_k a_{ijk}\int_0^1 G_i(\alpha_1,\beta_1;x)G_\ell(\alpha_1,\beta_1;x)f(x)dx\int_0^1 G_j(\alpha_2,\beta_2;y)\cdot$$
$$G_m(\alpha_2,\beta_2;y)dy\int_0^1 G_k(\alpha_3,\beta_3;z)G_n(\alpha_3,\beta_3;z)dz \quad . \qquad (5.30)$$

Taking further into account Eqs. (5.23) and (5.24), the expected value of Eq. (5.26) becomes

$$E[G_i(\alpha_1,\beta_1;x)G_j(\alpha_2,\beta_2;y)G_k(\alpha_3,\beta_3;z)]$$
$$= \frac{a_{ijk}}{B(\beta_1,\alpha_1-\beta_1+1)B(\beta_2,\alpha_2-\beta_2+1)B(\beta_3,\alpha_3-\beta_3+1)}\, d_i^2 d_j^2 d_k^2 \quad . \qquad (5.31)$$

Therefore, the a_{ijk} coefficients can be expressed as

$$a_{ijk} = \frac{B(\beta_1,\alpha_1-\beta_1+1)B(\beta_2,\alpha_2-\beta_2+1)B(\beta_3,\alpha_3-\beta_3+1)}{d_i^2 d_j^2 d_k^2}\ .$$
$$E[G_i(\alpha_1,\beta_1;x)G_j(\alpha_2,\beta_2;y)G_k(\alpha_3,\beta_3;z)] \quad . \qquad (5.32)$$

The a_{ijk} coefficients are then obtained for the selected values of i, j, and k by Eq. (5.32). Using only up to the values $i+j+k=3$, and simplifying Eq. (5.32), then

$$a_{000} = 1;\quad a_{100} = a_{010} = a_{001} = 0;$$

$$a_{200} = a_{020} = a_{002} = 0;$$

$$a_{110} = \frac{E(XY) - E(X)E(Y)}{Var(X)Var(Y)}\ ;$$

$$a_{210} = \frac{d_0^2(x)d_0^2(y)}{d_2^2(x)d_1^2(y)}\left\{ E(X^2Y) - \frac{2(\beta_1+1)}{\alpha_1+3}[E(XY) - E(X)E(Y)] - E(X^2)E(Y) \right\};$$

$$a_{111} = [E(XYZ) - E(X)E(YZ) - E(Y)E(XZ) - E(Z)E(XY) + 2E(X)E(Y)E(Z)]/[Var(X)Var(Y)Var(Z)]\ ;\ \text{and}$$

$$a_{300} = \frac{d_0^2(x)}{d_3^2(x)} \{ E(X^3) - 3\frac{\beta_1+2}{\alpha_1+5} E(X^2) + 3\frac{(\beta_1+2)(\beta_1+1)}{(\alpha_1+5)(\alpha_1+4)} E(X) - \frac{(\beta_1+2)(\beta_1+1)\beta_1}{(\alpha_1+5)(\alpha_1+4)(\alpha_1+3)} \} \quad . \tag{5.33}$$

The other coefficients can be obtained by interchanging the variables and subscripts. Computing $E(X)$, $E(Y)$, $E(Z)$, $E(X^2)$, $E(Y^2)$, $E(Z^2)$, $E(X^3)$, $E(Y^3)$, $E(Z^3)$, $E(XY)$, $E(YZ)$, $E(XZ)$, $E(X^2Y)$, $E(Y^2Z)$, $E(Z^2X)$, $E(X^2Z)$, $E(Y^2X)$, $E(Z^2Y)$, and $E(XYZ)$ from the 1200 months long samples, the coefficients a_{ijk} can be obtained. Using these estimates of the a_{ijk} coefficients and expanding Eq. (5.21) in function of up to the values $i+j+k=3$, the beta distribution function with the Jacobi polynomials becomes

$$\begin{aligned} f_{X,Y,Z}(x,y,z) &= 1021401.514x^{2.417}(1-x)^{1.559} \\ & y^{1.769}(1-y)^{8.566} \quad z^{7.314}(1-z)^{1.092} \; (8.1901x^3 \\ & + 2.4360y^3 - 15.2448z^3 + 53.4116x^2y - 14.5653x^2z \\ & - 54.6704xy^2 - 111.3110y^2z + 4.8804xz^2 + 81.6306yz^2 \\ & - 6.7835xyz - 13.6763x^2 + 117.9596y^2 + 11.6667z^2 \\ & + 16.8549xy - 7.8702yz + 39.8227zx - 27.5350x \\ & - 72.1956y - 30.5372z + 28.8814) \end{aligned} \tag{5.34}$$

The use of this function for the triviate frequency distribution of deficit area, total areal deficit, and maximum deficit intensity, did not show a good fit because the probability densities often take negative values. The main reason for it is that these three variables are highly mutually correlated. Therefore, the use of the product of the three marginal distributions might not be a valid assumption for the trivariate distribution. Since the convergence of the polynomials is slow due to a high dependence among the components, taking more values of $i+j+k$, up to 4 or 5, does not improve the goodness of fit.

The method of transforming the three variables into the normal variables and/or into mutually independent variables may show a better fit. However, it was not tried to find this trivariate distribution in this study.

5.4. Model for the Areal Drought Structure

It is useful to generalize the areal drought structure as in the form of depth-area relationship of thunderstorms or showers (Woolhiser and Schwalen, 1960; Court, 1961). The analogous idea to the depth-area relationship is applied herein to define the areal drought structure, though the deficit area, A, does not always consist of one deficit cell but may be composed of several deficit cells. To obtain a general idea of the areal drought structure, the dimensionless deficit-area relationship is first investigated. The deficit intensity, Z_i^*, at a point i, is denoted by the absolute value of a negative deviate from a certain truncation level, that is,

$$Z_i^* = \xi_0 - \xi_i \quad , \tag{5.35}$$

where ξ_i = the stochastic component of monthly precipitation. Under the deficit area, A, and the maximum deficit intensity, I, the dimensionless expression of the area-intensity relationship may be in the form

$$i_* = f(a_*) \quad , \tag{5.36}$$

where $i_* = Z_i^*/I$ with $0 \le i_* \le 1$, $a_* = X_i^*/A$ with $0 \le a_* \le 1$, and X_i^* is the area where $Z_i^* \le Z^*$ and $0 < X_i^* < A$. To find a functional form of Eq. (5.36), some relations obtained from the generated samples are presented in Fig. 5.7. This figure shows that the dimensionless area-intensity relationship is very close to be linear. This means that the deficit intensity may be relatively uniformly distributed over the area. Figure 5.6 and Eq. (5.19) also suggest the uniform distribution of the deficit intensity because the relationship between D and AI is close to $D = AI/2$ for a uniform distribution. This linear relationship is given by

$$i_* = 1 - a_* \quad . \tag{5.37}$$

As the first approximation, the distribution of the deficit intensity over the deficit area can be considered to be uniformly distributed.

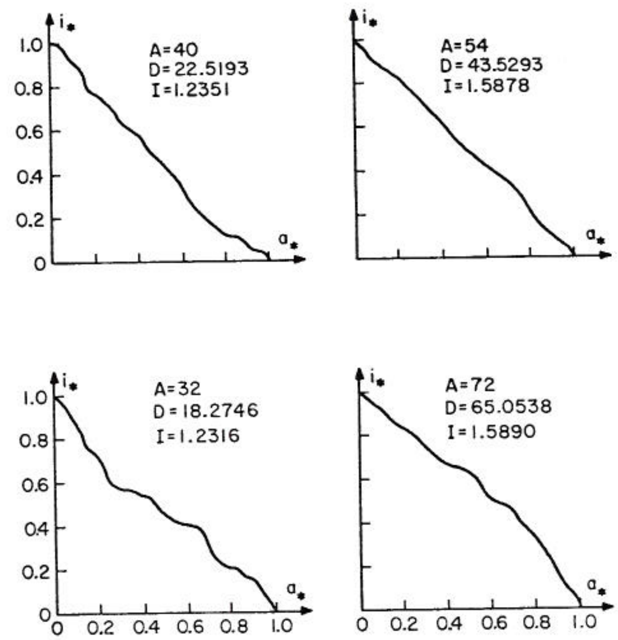


Fig. 5.7. Dimensionless Area-Intensity Relationships of Drought, Obtained from the Generated Samples.

For more flexibility in this relationship and for explaining the concave or convex shapes of Fig. 5.7, a parameter is added to Eq. (5.37), expressed by

$$i_* = (1 - a_*)^b \quad . \tag{5.38}$$

This dimensionless expression becomes

$$Z_i^* = IA^{-b}(A - X_i^*)^b \quad . \tag{5.39}$$

The conceptual relationship is then shown as in Fig. 5.8. The total areal deficit, D, is given by integrating Z^* over X^* from zero to A,

$$D = \int_0^A IA^{-b} (A - X^*)^b \, dx^*$$

$$= \frac{AI}{1 + b} \quad , \tag{5.40}$$

then the parameter b is defined by

$$b = \frac{AI}{D} - 1 \quad . \tag{5.41}$$

With the parameter b computed from A, I, and D of generated samples, the goodness of fit of this model for the areal drought structure is tested. The 600 samples at three different truncation levels, q = 0.58, 0.50, and 0.65 are used. For all the truncation levels, the average of the explained variance by the model is greater than 95 percent. Comparisons between the data and model are shown in Fig. 5.9. The average b value for the 600 samples at the truncation level q = 0.58 is 1.2292, with the standard deviation 0.3467. Generally the larger the deficit area or the total areal deficit, the smaller is the parameter b. This comes from the conditions that the deficit area and the maximum deficit intensity are limited so that the deficit intensity function should be convex upwards in order to produce a large total areal deficit.

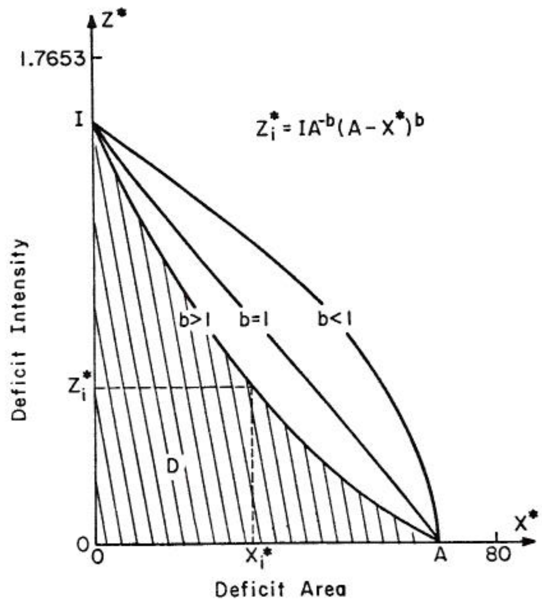


Fig. 5.8. Graph for Modeling the Relationship of Deficit Intensity to Deficit Area.

Once the deficit area, the total areal deficit, and the maximum deficit intensity are computed, the model defines the areal drought structure. The model may be more useful if the joint distribution of the three variables is obtained. However, the model does not indicate how the deficit area or how the deficit intensity is distributed areally.

In this model, the minimum deficit intensity, Z^*_{min}, is assumed to be zero. However, the minimum

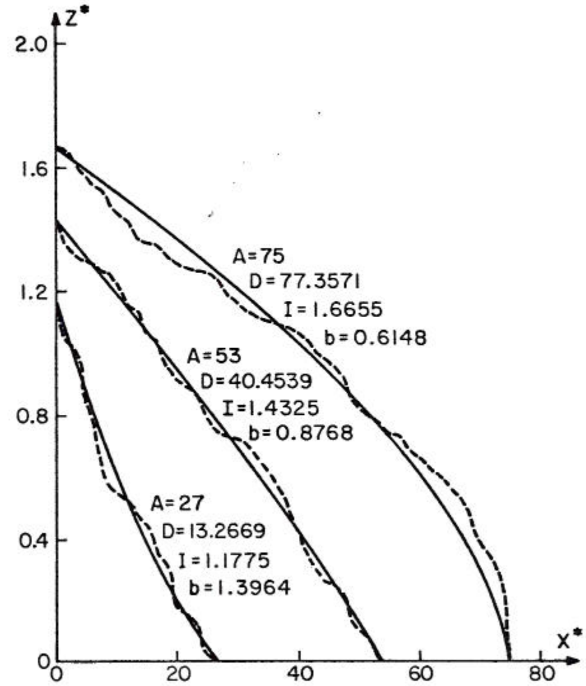


Fig. 5.9. Comparisons of the Model with the Results Obtained from the Generated Samples: Model Values are Given by Solid Lines, and Computed Values by Broken Lines.

deficit intensity can be larger than zero when a concerned region is small and a drought covers the whole region. In this case, the above model is not valid and some modifications are needed. By using deficit intensity defined as $(Z^* - Z^*_{min})$ instead of Z^*, the above areal drought structure model may be applicable.

5.5. Probability of Areal Coverage by Droughts

The areal coverage by droughts is further investigated without considering the corresponding total areal deficit.

The probability density function of the deficit area, A, over the whole area is given by Eq. (5.11). Defining a random variable by X = A/80 and using Eq. (5.11), the probability of more than the $100x_0$ percent of the whole area to be covered by a drought, with the truncation level $\xi_0 = 0$, is given by

$$P(X > x_0) = 1 - P(X \le x_0)$$

$$= 1 - F(x_0)$$

$$= 1 - \int_0^{x_0} 27.6128x^{2.417}(1-x)^{1.559}dx \quad . \tag{5.42}$$

This probability can be found easily from Fig. 5.2. For example, the probability of more than the 50 percent of the whole area to be covered by a drought in a certain month is 0.65 at the truncation level $\xi_0 = 0$. This probability is relatively high, though the drought does cover each time a different area within the whole area.

Since the deficit area is a time independent variable, some information on the time duration of areal coverage can be obtained by using the theory of runs and the recurrence of random events. For an independent process, the expected positive run-length, M, is given by

$$E(M) = 1/q \tag{5.43}$$

with $q = F(x_0) = P(X \leq x_0)$. Therefore, the expected duration of a drought which covers more than the $100x_0$ percent of the whole area can be found. Table 5.5 shows some probabilities. For $x_0 = 0.5$ in the deficit area given by the truncation level $\xi_0 = 0$, or $q = F(0.5) = 0.35$, the expected duration of that drought from Eq. (5.43) is 2.857 months. This expected duration of about three months seems to be long enough to affect many water users, such as crop producers.

Table 5.5. The Expected Duration of Drought Covering More than $100x_0$ Percent of the Whole Area

x_0	0.4	0.5	0.6	0.7
For $\xi_0 = 0$	5.113	2.857	1.876	1.393
For $\xi_0 = -0.2028$	2.984	1.935	1.438	1.185

The other interesting property is the longest drought duration. Millan and Yevjevich (1971) applied the theory of recurrent events to find probabilities of the longest negative run-length of a given sample size N, on the basis of the formula given by Feller (1957). The probability of the longest drought duration in a given sample size N is, as an approximation,

$$F_N(\ell) = \frac{1 - pw}{[\ell + 2 - (\ell+1)w]q} \frac{1}{w^{N+1}} \tag{5.44}$$

with $q = F(x_0) = 1 - p$ and

$$w = 1 + pq^{\ell} + (1+\ell)(pq^{\ell})^2 + (1+\ell)^2(pq^{\ell})^3 + \cdots .$$

Probabilities of the longest drought duration during a given period covering more than the $100x_0$ percent of the whole area are calculated by Eq. (5.44). For 60-, 600-, and 1200-month periods, the expected longest drought durations $E(M_{max})$, with droughts covering more than the $100x_0$ percent of the whole area are computed for the deficit area of the truncation level $\xi_0 = 0$ and -0.2028 and given in Table 5.6. For the 1200-month period of the deficit area series based on the generated series ξ at the truncation level $\xi_0 = 0$, the observed longest duration of the drought covering more than the 50 percent of the whole area is 13 months, which is a little lower than the expected value, or 14.5 months. During 600 months or 50 years, the expected longest duration of the drought which covers more than a half of the whole area is more than one year. This duration is fairly long.

The conditional distributions of the deficit area, A, given that a certain point has or has not a deficit are investigated analytically. The probability density function of the deficit area defined by A/80 is given by Eq. (5.11), namely

$$f_X(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

$$= 27.6128\, x^{2.417}(1-x)^{1.559} \tag{5.45}$$

Table 5.6. The Expected Longest Duration of Droughts Covering More Than 50 Percent of the Whole Area in 60-, 600-, 1200-Month Periods for $\xi_0 = 0$ and -0.2038.

Sample Size in Months	Longest Drought Duration for $\xi_0 = 0$ Expected	Observed	for $\xi_0 = -0.2028$ Expected	Observed
60	7.866	-	5.007	-
600	13.077	11.5	8.181	8.0
1200	14.505	13.0	9.134	10.0

The probability that a certain point has a deficit at the truncation level ξ_0 is defined by

$$P(\xi \leq \xi_0) = q \ . \tag{5.46}$$

Assuming that every point is equally likely to experience a drought, the conditional probability that a certain point has a deficit given that the deficit area x exists in the whole area is

$$P(\xi \leq \xi_0 | X=x) = x \ . \tag{5.47}$$

Then the joint probability $P(\xi \leq \xi_0, X \leq x)$ is defined by

$$\begin{aligned} &P(\xi \leq \xi_0, X \leq x) \\ &= P(\xi \leq \xi_0 | X \leq x) P(X \leq x) \\ &= \int_0^x x f_X(x)\,dx \ . \end{aligned} \tag{5.48}$$

Therefore, the conditional probability of the deficit given that a certain point has a deficit is given by

$$P(X \leq x | \xi \leq \xi_0)$$

$$= P(X \leq x, \xi \leq \xi_0)/P(\xi \leq \xi_0)$$

$$= \frac{1}{q}\int_0^x x f_X(x)\,dx$$

$$= \int_0^x 47.6083\, x^{3.417}(1-x)^{1.559}dx \tag{5.49}$$

$$= \frac{1}{B(\alpha+1,\beta)} \int_0^x x^{\alpha}(1-x)^{\beta-1} \ . \tag{5.50}$$

Because this is also a gamma distribution, the conditional density function is

$$f_{X|\xi\leq\xi_0}(x|\xi\leq\xi_0) = \frac{1}{B(\alpha+1,\beta)} x^{\alpha}(1-x)^{\beta-1} \ . \tag{5.51}$$

Similarly, the conditional probability of the deficit area given that a certain point has no deficit is

$$P(X \le x | \xi > \xi_0)$$

$$= P(X \le x,\ \xi > \xi_0)/P(\xi > \xi_0)$$

$$= \frac{1}{1-q} \int_0^x (1-x) f_X(x) dx$$

$$= \int_0^x 65.7448\ x^{2.417} (1-x)^{2.559} dx \qquad (5.52)$$

$$= \frac{1}{B(\alpha,\beta+1)} \int_0^x x^{\alpha-1} (1-x)^{\beta} dx \qquad (5.53)$$

Therefore, the density function is

$$f_{X|\xi>\xi_0}(x|\xi>\xi_0) = \frac{1}{B(\alpha,\beta+1)} x^{\alpha-1}(1-x)^{\beta} \qquad (5.54)$$

The two conditional probability density functions of the deficit area, with the two corresponding conditional frequency curves, are shown in Fig. 5.10, in comparison with the probability density function of the deficit area. A quantitative comparison of the three probability density functions may be possible by applying the Fisher's information content, which is defined as the reciprocal of the variance of the concerned variable or parameter. The conditional distributions clearly have more information than the original distribution. As the extension of this concept, the conditional distribution of the deficit area given that more than a station, selected randomly or systematically, have deficits and/or no deficits can be analyzed by an experimental method. A number

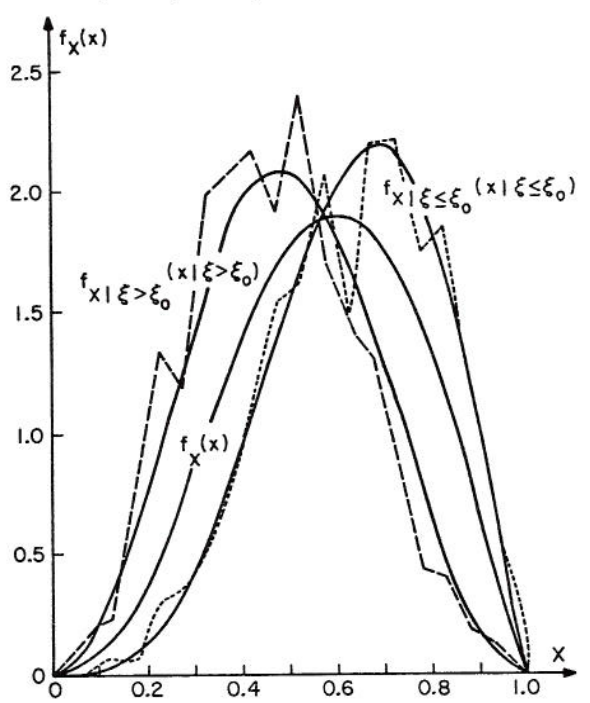


Fig. 5.10. The Probability Density Function of the Deficit Area and the Two Conditional Probability Functions of the Deficit Area Given That a Certain Point Has or Has not a Deficit (Solid Lines), with the Two Sample Conditional Frequency Curves (Broken Lines).

of selected stations and their locations will affect additional information content. Their effects are worthy of studying as concerning to the network design or some other problems.

5.6. Probabilities of Specific Area Covered by Drought

In previous analysis, the deficit area inside the whole area consisting of 80 unit-areas or 800,000 square miles has been considered. In practice, however, one may be more interested in finding the probability of drought covering a small, specific area, or sub-area inside the whole area. For example, what is the probability of drought covering the State of Kansas, the wheat region, or the corn belt of the Midwest? Because of regional dependence, the probability of the deficit area depends upon the size and shape of this small, specific sub-area.

First, the probability of a drought covering a whole sub-area is investigated by considering the effect of the shape of the sub-areas. Intuitively, it is clear that the smaller a sub-area, the higher is the probability of the drought covering this whole sub-area. An elongated sub-area should have a smaller chance for a drought to cover the whole sub-area, than would be the case of a circle or square type sub-area of the same surface. The analysis is carried out quantitatively by using the generated samples. Figure 5.11 shows the experimental results of two cases, with probabilities of drought covering the sub-areas, from one to 32 unit-areas, by changing the area shape, in different ways. The probability of a drought covering the whole sub-area of 16 unit-areas or a 400 x 400 square miles square shape in case (1) is 0.0793, which is higher than 0.0583 for case (2), for which the same 16 unit-areas are arranged as a 200 x 800 square miles rectangular area. Figure 5.11 further indicates that the probability of a drought covering the whole sub-area decreases rapidly with an increase of the area. The probability of the drought coverage is very sensitive to the shape of the sub-area; this is especially the case for the sub-area with about eight unit-area. However, the effect of the shape of the sub-area decreases with an increase of the area. For a large area, the shape does not seem to be a decisive factor as long as a particularly irregular sub-area is not considered. Since probabilities are estimated by sample frequencies, and since these frequencies have small variations, the average frequencies as estimates of probabilities are used to evaluate the effect of the area size and the area shape on probabilities of areal coverage by droughts. The assumption used is that any geometrically identical sub-area would have the same probability regardless of sub-area position. Table 5.7 summarizes these average estimates of probabilities of the drought covering the whole sub-areas of various sizes and shapes. Table 5.7 indicates that the size of an area seems to be more important than its shape as it concerns the probability of drought coverage, as long as the shape is not of a very elongated type. Although the estimates of probabilities in Table 5.7 may be in error due to small sample sizes and the assumption, these values give a general picture of drought probabilities for a specific area. For example, the probability of the drought covering the whole state of Kansas, of 82,264 square miles, approximated by 200 x 400 miles rectangle, is about 17.5 percent and 11.7 percent for each month at the truncation levels of the mean and the median, respectively.

In order to evaluate the shape factor more quantitatively, a mean areal correlation coefficient

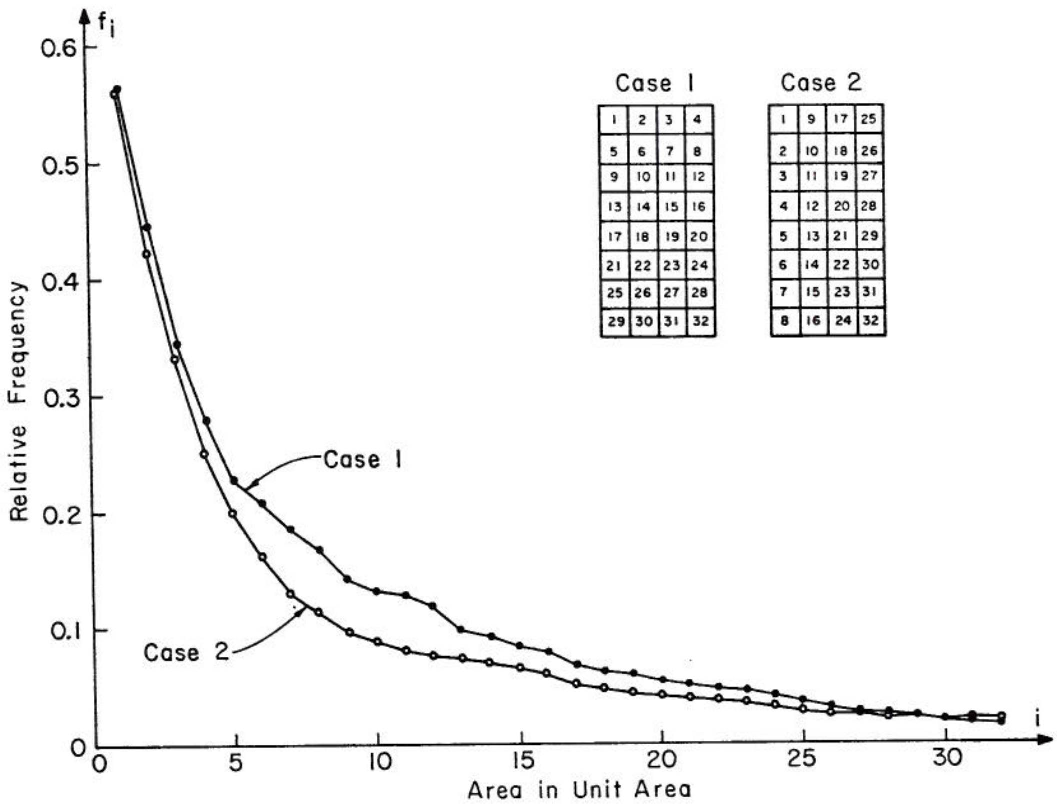


Fig. 5.11. Relative Frequencies of Droughts Versus the Whole First i Unit-Area Covered by the Droughts, Obtained by the Two Schemes Illustrated by Cases 1 and 2.

Table 5.7. Estimates of Probabilities of the Drought Covering the Whole Subarea with Various Sizes and Shapes for the Truncation Levels of the Mean and the Median.

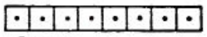
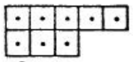
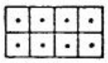
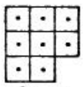
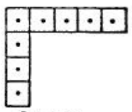
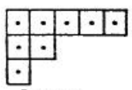
Subarea Size in Unit-Area	Shape	Probabilities of Drought: Nos. of Samples Used	Truncation Level=Mean $\xi_0=0$	Nos. of Samples Used	Truncation Level=Median $\xi_0=-0.2029$
1	1 x 1	14	0.5698		
2	1 x 2	14	0.4367		
3	1 x 3	11	0.3459		
4	1 x 4	9	0.2759		
4	2 x 2	3	0.3092		
5	1 x 5	4	0.2171		
6	2 x 3	4	0.2320	2	0.1654
8	1 x 8	3	0.1150		
8	2 x 4	8	0.1748	4	0.1167
9	3 x 3	6	0.1609		
10	2 x 5	1	0.1343		
12	3 x 4	11	0.1155	4	0.0736
15	3 x 5	1	0.0793		
16	2 x 8	3	0.0569		
16	4 x 4	14	0.0766	5	0.0437
18	3 x 6	1	0.0600		
20	4 x 5	3	0.0545		
21	3 x 7	1	0.0451		
24	3 x 8	3	0.0300		
24	4 x 6	3	0.0394		
25	5 x 5	3	0.0369		
27	3 x 9	1	0.0208		
28	4 x 7	3	0.0242		
30	5 x 6	3	0.0239		
32	4 x 8	6	0.0164	2	0.0059
36	6 x 6	2	0.0175		
49	7 x 7	2	0.0050		
64	8 x 8	2	0.0013		

is introduced. It is defined by

$$\bar{r} = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} r_{ij} \quad . \qquad (5.55)$$

Figure 5.12 compares the probabilities of the drought covering the whole sub-area of eight unit-areas with different shapes as a function of the mean areal correlation coefficient. For contiguous areas, probabilities of drought are linear functions of the mean areal correlation coefficient. A gradient of the line decreases with a decrease of the probability of truncation levels. This means that effects of shape of the sub-area decreases with a decrease of the truncation levels. Figure 5.13 shows the probabilities of the drought covering the whole sub-area with different sizes and shapes as a function of the mean areal correlation coefficient. Three probabilities for $\bar{r} = 0$ are calculated from a binomial distribution. For contiguous areas which are surrounded by broken lines, probabilities of the drought are approximated by a linear function of the mean areal correlation coefficient. The gradient of the line for eight unit-areas seems steepest, that is, the shape factor is most sensitive to the probabilities. With an increase of the area, the gradient of lines decreases.

Probability distributions of the deficit area within sub-areas of various sizes are also investigated. The sub-areas considered are of 2 x 3, 2 x 4, 3 x 4, 4 x 4, and 4 x 8 unit-areas. Besides, a sub-area of 63 unit-areas, as the original study area, is also considered. For the truncation levels at the mean and the median, frequency curves of the deficit

Shape of Area and Mean Areal Correlation Coefficient	Probability of Droughts at Truncation Level of · q=0.58	q=0.50	q=0.40
 $\bar{r}$=0.356	0.1150	0.0733	0.0369
 $\bar{r}$=0.481	0.1627		
 $\bar{r}$=0.511	0.1748	0.1167	
 $\bar{r}$=0.531	0.1774		
 $\bar{r}$=0.395	0.1304		
 $\bar{r}$=0.448	0.1444	0.0954	0.0523

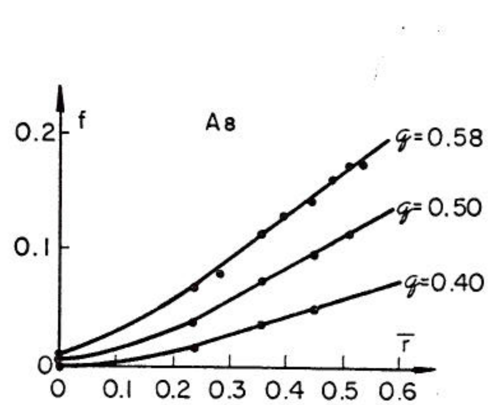


Fig. 5.12. Probabilities of Drought Covering the Whole 8 Unit-Areas (80,000 sq. miles) with Six Different Shapes as a Function of the Mean Areal Correlation Coefficient.

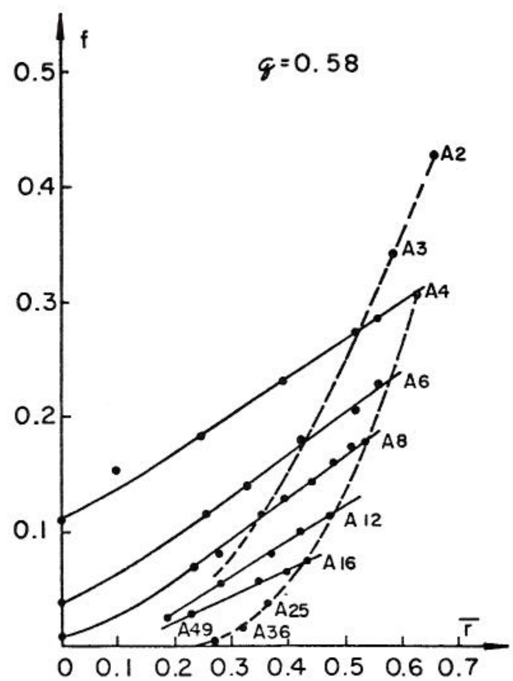


Fig. 5.13. Probabilities of Drought Covering the Whole Subareas of Different Sizes and Shapes, as a Function of the Mean Areal Correlation Coefficient.

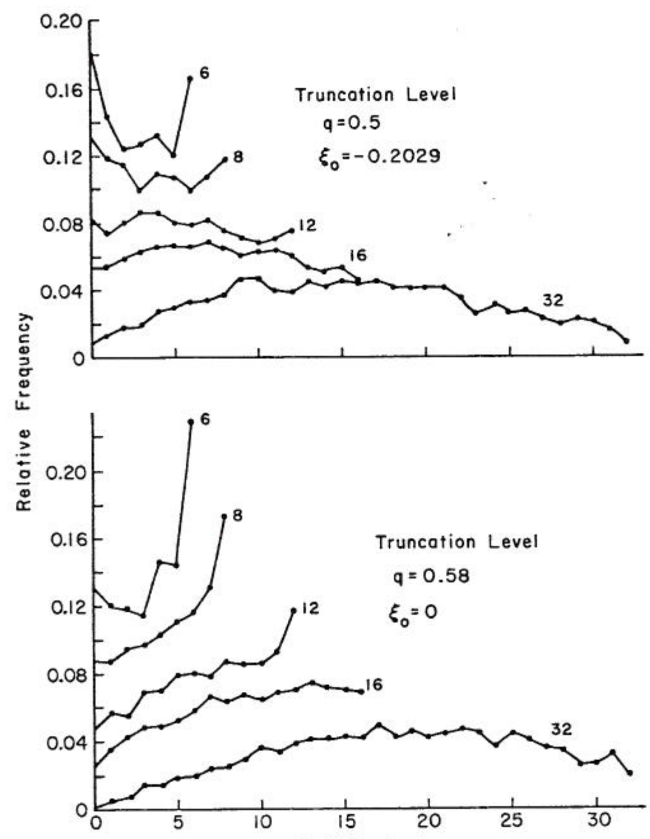


Fig. 5.14. Relative Frequency Curves of Deficit Area Within the Subareas of the 6, 8, 12, 16, and 32 Unit-Areas.

area within these sub-areas are obtained from some samples, and shown in Fig. 5.14. Their cumulative frequency distributions are shown in Fig. 5.15, as the reduced variables in percentages of the whole sub-area, for comparison.

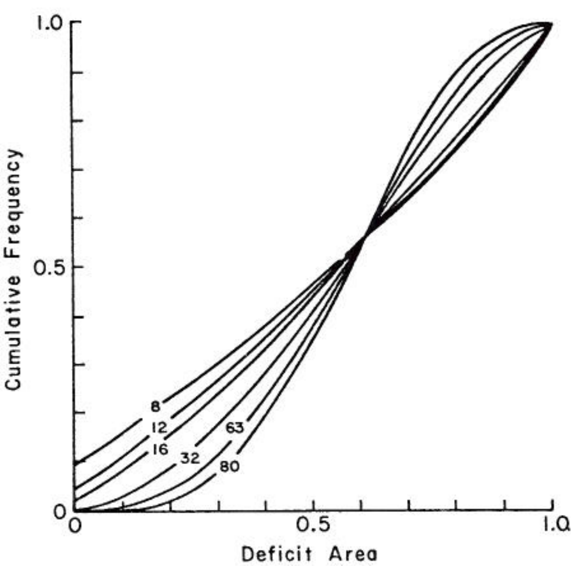


Fig. 5.15. Cumulative Frequency Curves of Percent Deficit Area Within the Subareas of the 8, 12, 16, 32, 63, and 80 Unit-Areas at the Truncation Level of the Mean.

For the 2 x 3 and 2 x 4 unit-areas, probabilities of both extremes, that is, the whole sub-area covered by a drought and no part of the sub-area covered by a drought, are very high. Their frequency curves are either U-shaped or J-shaped. These are intuitively obvious because of the regional dependence. In the case of the truncation level of the mean, the probability of no part of the area covered by a drought is decreasing faster than for the other extreme with an increase of the area. For the truncation level of the median, the opposite trend seems valid. The distribution of the deficit area is affected by the truncation level, especially for the small areas, as shown by Fig. 5.14. The probability of the deficit area for a given sub-area with the area less than 80 unit-areas can be approximated from Fig. 5.15.

5.7. Conversion of the Total Areal Deficit of Stationary Stochastic Series into the Total Areal Deficit of Periodic-Stochastic Series

The stationary stochastic series, ξ, are standardized, (or dimensionless), with the regional and temporal variations in the means and the standard deviations removed. Thus the drought characteristics, such as the total areal deficit and the maximum deficit intensity, do not represent the absolute values or the actual amount of deficits. However, these values are needed for planning drought control measures. Therefore, the method of converting the total areal deficit of the standardized series, ξ, over an area into the total areal deficit of the original monthly precipitation series over the same area must be available. In this section, only the regional variations in parameters are investigated, while their temporal (or periodic) variations are discussed in the next chapter.

Let D_s denote the total areal deficit of the series ξ, and D_a the total areal deficit of the original series χ, or the actual total areal deficit. They are defined as

$$D_s = \sum_{i=1}^{80} (\xi_0 - \xi_i) I_{(\xi \le \xi_0)}(\xi_i) \tag{5.56}$$

and

$$D_a = \sum_{i=1}^{80} (\chi_0 - \chi_i) I_{(\chi \le \chi_0)}(\chi_i)$$

$$= \sum_{i=1}^{80} s_i (\xi_0 - \xi_i) I_{(\xi \le \xi_0)}(\xi_i) \ , \tag{5.57}$$

where ξ_0 and χ_0 are the truncation levels for the series ξ and χ, respectively, s_i is the standard deviation at station i, and $I_{(\cdot)}(\cdot)$ is an indicator function. These definitions are valid as long as the relationship between the two truncation levels ξ_0 and χ_0 is defined by

$$\chi_0 = m_i + s_i \xi_0 \tag{5.58}$$

with m_i and s_i the mean and standard deviation at station i, respectively.

The actual total areal deficit, D_a, can be approximated by a function of the total areal deficit, D_s, the mean and standard deviation of the regional standard deviation, $\bar{s}_i$ and $s(s_i)$, and the deficit area, A. This relationship is expressed as

$$D_a = f[D_s, \bar{s}_i, s(s_i), A] \ . \tag{5.59}$$

In the case the regional variation of the standard deviation is small in comparison with the mean, Eq. (5.59) can be reduced to

$$D_a = f(D_s, \bar{s}_i, A) \ . \tag{5.60}$$

Figure 5.16 shows a relationship between D_s and D_a, with the regional standard deviations for the 80 stations given by Eq. (3.19), and their mean and standard deviation of 1.4360 and 0.4747, respectively. The regression analysis gives the relationship

$$D_a = 0.1060 + 1.4374 \, D_s \ , \tag{5.61}$$

with the explained variance of 95.84 percent. This equation is further approximated by

$$D_a = 0.1060 + 1.4374 \, D_s$$

$$\approx 1.4360 \, D_s = \bar{s}_i D_s \ . \tag{5.62}$$

This indicates that the actual total areal deficit D_a of the original series can be well approximated by the product of the total areal deficit D_s of the standardized stationary series and the mean of the regional standard deviations, $\bar{s}_i$. However, Fig. 5.17 showing a relationship between the ratio of D_a to D_s and the deficit area, A, indicates that, for a small deficit area, the mean of the regional standard

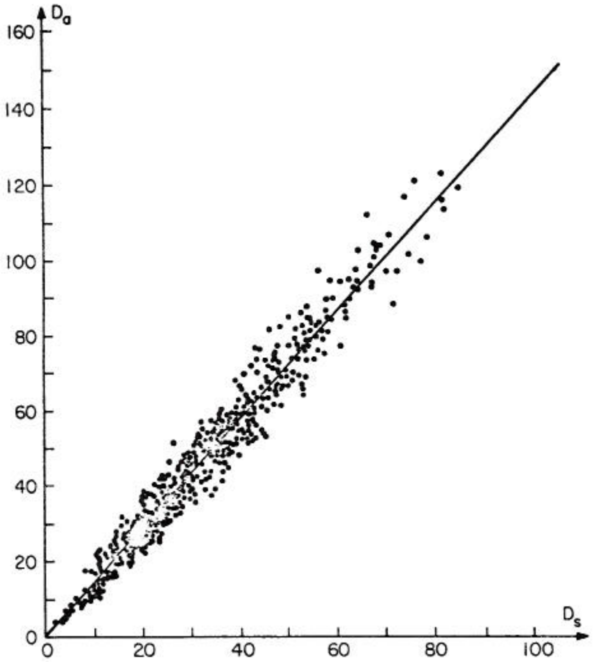


Fig. 5.16. Relationship Between the Total Areal Deficit D_a of Periodic-Stochastic Series and the Total Areal Deficit D_s of the Stationary Series, with the Inferred Regression Equation or $D_a = 0.1060 + 1.4374D_s$.

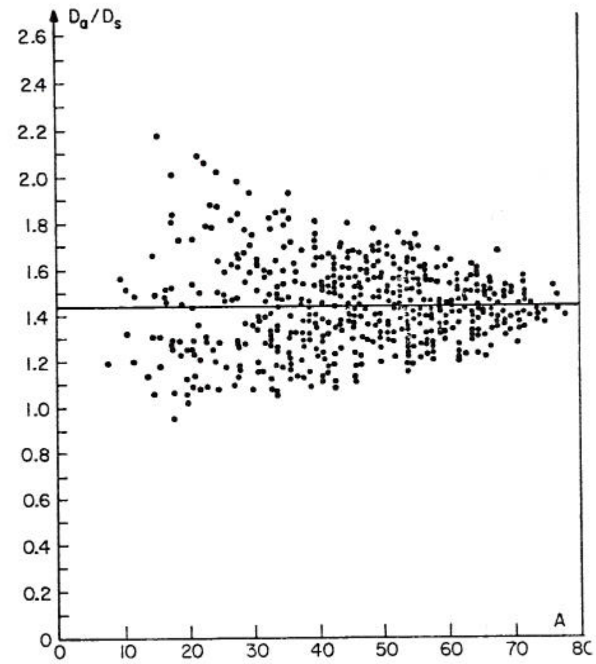


Fig. 5.17. Relationship Between the Ratio D_a/D_s and the Deficit Area A, with the Equation $D_a/D_s = 1.4360$.

deviations for that area rather than the mean of all the standard deviations over the whole area should be taken for determining $\overline{s}_i$. Since the larger deficit areas are mostly of interest, the above approximation to the determination of the actual total areal deficit, by using the product $\overline{s}_i D_s$, has a practical significance.

Chapter 6
DROUGHT ANALYSIS OF PERIODIC — STOCHATIC PROCESSES

Drought characteristics of the stochastic component of the monthly precipitation were studied in the previous chapter. This stochastic component of the monthly precipitation is a stationary process for good, consistent data, while the monthly precipitation is a periodic-stochastic process. The drought analysis of periodic-stochastic processes is much more complex than the drought analysis of stationary stochastic processes. Each month has a different mean and a different standard deviation. Therefore, the time position is one of the important factors in evaluating drought characteristics. The main reason of analyzing drought characteristics of periodic-stochastic series is to find fluctuations of deficits within a year and at specific months such as a growing season rather than to find long-term fluctuations. Drought characteristics depend upon the hydrologic phenomenon such as precipitation and river runoff which is used for drought definition and upon the objectives of drought analysis. They make difficult a generalization of drought characteristics of periodic-stochastic processes.

6.1. Run Properties of Periodic-Stochastic Processes

Although run properties of univariate periodic-stochastic processes may not give the best parameters for drought analysis, they are studied analytically for some simple cases in order to compare them with those of univariate stationary stochastic processes and to find out possibilities of using run properties in evaluation of drought characteristics of periodic-stochastic processes.

The following assumptions, considered to be close to real processes of nature such as for the monthly precipitation, are used for a simple study. The monthly precipitation process, χ, as an example of periodic-stochastic processes, is assumed to be composed as

$$\chi_{p,\tau} = \mu_\tau + \sigma_\tau \xi_{p,\tau} \tag{6.1}$$

where p and τ denote the sequence of years and the month within the year, respectively, μ_τ and σ_τ are the means and standard deviations of monthly precipitation, respectively, and ξ is a time independent stationary stochastic component. Next, the standard deviation σ_τ is assumed to be composed only of the 12-month harmonic

$$\sigma_\tau = \bar{\sigma} + C\cos(\frac{\pi}{6}\tau + \theta) \ , \tag{6.2}$$

with $\bar{\sigma}$ = the average value of σ_τ, C = the amplitude, and θ = the phase. The truncation level $\chi_{0,\tau}$ (or the water demand series) is also assumed to be a periodic function

$$\chi_{0,\tau} = \mu_\tau + \sigma_\tau \xi_0 \ , \tag{6.3}$$

with ξ_0 = the truncation level for ξ corresponding to χ_0. Under these conditions, negative runs as various random variables are defined as: N = the negative run-length, T = the onset time of a negative run, $W_{N,T}$ = the negative run-sum with the run-length N and the onset time T, ξ_0 = the truncation level of $q = P(\xi \leq \xi_0) = F(\xi_0)$ or $p = 1-q$, $E(\xi^*) = \bar{\xi}^*$ = the expected value of the truncated series, ξ^*, and $\mathrm{Var}(\xi^*)$ = the variance of the truncated series. The distribution of the negative run-length of an independent process is a geometric distribution (Downer et al., 1967), with the expected value and the variance given by

$$E(N) = 1/p \tag{6.4}$$

and

$$\mathrm{Var}(N) = q/p^2 \ . \tag{6.5}$$

Since each month is assumed to have the same probability for the negative run to start, the probability of the onset time can be defined by

$$P(T=t) = 1/12 \ , \qquad (t=1,2,\cdots,12) \ . \tag{6.6}$$

Under these basic assumptions, run properties of univariate periodic-stochastic processes can be obtained. As long as the truncation level of χ is defined by Eq. (6.3), the distribution of the negative run-length of the monthly precipitation $\chi_{p,\tau}$ is a geometric distribution, with the expected value and the variance given by Eqs. (6.4) and (6.5), respectively. The expected value, variance, and conditional expected values and variances of the negative run-sum of $W_{N,T}$ can be also derived analytically.

Since the negative run-sum is defined by

$$W_{N,T} = \sum_{i=T}^{T+N-1} (\chi_{0,i} - \chi_i)$$

$$= \sum_{i=T}^{T+N-1} \sigma_i(\xi_0 - \xi_i) = \sum_{i=T}^{T+N-1} \sigma_i(\xi_i^* - \xi_0) \ , \tag{6.7}$$

then by taking into account Eq. (6.2) the conditional expected value of the negative run-sum for given $N=n$ and $T=t$ is

$$E(W_{N,T}|N=n,T=t)$$

$$=E[\sigma_t(\xi_0 - \xi_t) + \sigma_{t+1}(\xi_0 - \xi_{t+1}) + \cdots$$

$$+ \sigma_{t+n-1}(\xi_0 - \xi_{t+n-1})]$$

$$= [\, n\bar{\sigma} + C\sum_{i=t}^{t+n-1} \cos(\frac{\pi}{6}i + \theta)](\bar{\xi}^* - \xi_0) \ . \tag{6.8}$$

The conditional expected value of $W_{N,T}$ for given $N=n$ is given by the expected value of Eq. (6.8) with regard to T. Therefore, it is given by

$$E(W_{N,T}|N=n)$$

$$=E[E(W_{N,T}|N=n,T)] \ . \tag{6.9}$$

Taking into account $\frac{1}{12}\sum_{t=1}^{12}\sum_{i=t}^{n+t-1}\cos(\frac{\pi}{6}i+\theta)=0$,

Eq. (6.9) becomes

$$E(W_{N,T}|N=n) = n\bar{\sigma}(\bar{\xi}^* - \xi_0) \quad . \qquad (6.10)$$

The expected value of $W_{N,T}$ is given by the expected value of Eq. (6.10) with regard to N, that is,

$$E(W_{N,T})$$

$$=E[E(W_{N,T}|N)] = E[N\bar{\sigma}(\bar{\xi}^* - \xi_0)]$$

$$= \bar{\sigma}(\bar{\xi}^* - \xi_0)E(N) \qquad (6.11)$$

Since $E(N)$ is given by Eq. (6.4), the expected value of the negative run-sum is given by

$$E(W_{N,T}) = \frac{1}{p}\bar{\sigma}(\bar{\xi}^* - \xi_0) \qquad (6.12)$$

Since ξ is a time independent variable, the conditional variance of the negative run-sum for given N=n and T=t is given by the sum of variance of each term,

$$\mathrm{Var}(W_{N,T}|N=n,T=t)$$

$$= \mathrm{Var}[\sigma_t(\xi_0 - \xi_t)] + \mathrm{Var}[\sigma_{t+1}(\xi_0 - \xi_{t+1})] + \ldots$$

$$\ldots + \mathrm{Var}[\sigma_{t+n-1}(\xi_0 - \xi_{t+n-1})]$$

$$= \sigma_t^2\mathrm{Var}(\xi^*) + \sigma_{t+1}^2\mathrm{Var}(\xi^*) + \ldots + \sigma_{t+n-1}^2\mathrm{Var}(\xi^*)$$

$$= [n\bar{\sigma}^2 + 2\bar{\sigma}C\sum_{i=t}^{t+n-1}\cos(\frac{\pi}{6}i+\theta)$$

$$+ C^2\sum_{i=t}^{t+n-1}\cos^2(\frac{\pi}{6}i+\theta)]\mathrm{Var}(\xi^*). \qquad (6.13)$$

Applying the following relation (Thomas, 1971, p. 103)

$$\mathrm{Var}(Y) = \mathrm{Var}[E(Y|X)] + E[\mathrm{Var}(Y|X)] \quad , \qquad (6.14)$$

the conditional variance of the negative run-sum for given N=n is

$$\mathrm{Var}(W_{N,T}|N=n)$$

$$= \mathrm{Var}[E(W_{N,T}|N=n,T)] + E[\mathrm{Var}(W_{N,T}|N=n,T)]$$

$$= \mathrm{Var}\{[n\bar{\sigma} + C\sum_{i=T}^{T+n-1}\cos(\frac{\pi}{6}i+\theta)](\bar{\xi}^* - \xi_0)\} + E\{[n\bar{\sigma}^2$$

$$+ 2\bar{\sigma}C\sum_{i=T}^{T+n-1}\cos(\frac{\pi}{6}i+\theta) + C^2\sum_{i=T}^{T+n-1}\cos^2(\frac{\pi}{6}i+\theta)]\mathrm{Var}(\xi^*)\} \,.$$

$$(6.15)$$

Taking $1/12\sum_{t=1}^{12}\sum_{i=t}^{n+t-1}\cos(\frac{\pi}{6}i+\theta) = 0$ and $1/12\sum_{t=1}^{12}\sum_{i=t}^{n+t-1}\cos^2(\frac{\pi}{6}i+\theta) = n/2$, Eq. (6.15) becomes

$$\mathrm{Var}(W_{N,T}|N=n)$$

$$= [\frac{n}{2}C^2 + C^2\sum_{i=1}^{n}(n-i)\cos\frac{\pi}{6}i](\bar{\xi}^* - \xi_0)^2$$

$$+ (n\bar{\sigma}^2 + \frac{n}{2}C^2)\mathrm{Var}(\xi^*) \quad . \qquad (6.16)$$

Using Eq. (6.14), the variance of the negative run-sum is

$$\mathrm{Var}(W_{N,T})$$

$$= \mathrm{Var}[E(W_{N,T}|N)] + E[\mathrm{Var}(W_{N,T}|N)]$$

$$= \mathrm{Var}[N\bar{\sigma}(\bar{\xi}^* - \xi_0)] + E\{(N\bar{\sigma}^2 + \frac{N}{2}C^2)\mathrm{Var}(\xi^*)$$

$$+ [\frac{N}{2}C^2 + C^2\sum_{i=1}^{N}(N-i)\cos\frac{\pi}{6}i](\bar{\xi}^* - \xi_0)^2\} \qquad (6.17)$$

Finally, taking account of Eqs. (6.4) and (6.5),

$$\mathrm{Var}(W_{N,T}) = \frac{q}{p^2}\bar{\sigma}^2(\bar{\xi}^* - \xi_0)^2 + \frac{1}{p}\bar{\sigma}^2\mathrm{Var}(\xi^*)$$

$$+ \frac{C^2}{2}[\mathrm{Var}(\xi^*) + (\bar{\xi}^* - \xi_0)^2]$$

$$+ C^2(\bar{\xi}^* - \xi_0)^2E[\sum_{i=1}^{N}(N-i)\cos\frac{\pi}{6}i] \qquad (6.18)$$

The conditional moments of the negative run-sum for given N=n and T=t can be found by the moment generating function and the cumulant generating function, as used by Downer et al. (1967) to find run properties of stationary processes. The following notations are adopted: M_w and K_w are the moment and the cumulant generating function of $W_{N,T}$ (N=n, T=t), respectively, $K_w = \log M_w$, K^* is the cumulant generating function of ξ^*, and κ_m^* is the m-th cumulant of ξ^*. Thus

$$K_w(u) = \log M_w(u) = \log[E(e^{uW})] \qquad (6.19)$$

$$K^*(u) = \sum_{m=1}^{\infty}\kappa_m^*\frac{u^m}{m!} \quad . \qquad (6.20)$$

In particular,

$$\kappa_1^* = E(\xi^*),\ \kappa_2^* = \mathrm{Var}(\xi^*),\ \text{and } \kappa_3^* = E[(\xi^* - \bar{\xi}^*)^3] \,. \qquad (6.21)$$

The cumulant generating function of the conditional negative run-sum for given N=n and T=t is given by

$$K_w(u) = \sum_{i=t}^{t+n-1}K^*(\sigma_i u) \qquad (6.22)$$

where σ_i is defined by Eq. (6.2). The first two moments are the same as those given by Eqs. (6.8) and (6.13).

Looking at these results, the conditional expected value and variance of the negative run-sum for given $N=n$, the expected value and the variance of the negative run-sum do not depend upon the phase θ.

The expected value and variance of the negative run-sum of a stationary series with the mean zero and the standard deviation $\bar{\sigma}$ are given by Downer et al. (1967),

$$E(W) = \frac{1}{p}\bar{\sigma}(\bar{\xi}^* - \xi_0) \tag{6.23}$$

and

$$\mathrm{Var}(W) = \frac{q}{p^2}\bar{\sigma}^2(\bar{\xi}^* - \xi_0)^2 + \frac{1}{p}\bar{\sigma}^2\mathrm{Var}(\xi^*) \tag{6.24}$$

The expected values of the negative run-sum of the stationary process and the periodic-stochastic process are the same. The variance V of the negative run-sum of the periodic-stochastic process consists of two parts, namely the variance V_s due to the stationary process, which is given by Eq. (6.24), and the variance V_p due to the periodicity in C. This is expressed by

$$V = V_s + V_p \ . \tag{6.25}$$

For a given series ξ, the variance V depends on the truncation level ξ_0, the average of standard deviations $\bar{\sigma}$, and the amplitude C. A relationship of the ratio of V_s to V to the ratio of C to $\bar{\sigma}$ is considered. This relationship as shown in Table 6.1 indicates that the variance of the negative run-sum of periodic-stochastic processes is mainly due to their stationary part, especially for the ratio of C to $\bar{\sigma}$ less than 0.5.

The distribution of the negative run-sum, $W_{N,T}$, which cannot be obtained analytically, depends upon the distribution of ξ, the truncation level ξ_0, the average standard deviation $\bar{\sigma}$, and the amplitude C. The method used by Llamas and Siddiqui (1969) and the fitting of a theoretical distribution to experimentally obtained data may be a feasible way of finding the distribution of the negative run-sum.

Since in evaluating droughts of periodic-stochastic series of the monthly precipitation, deficit during a short period within the annual cycle, for example two or three month period including the specific months, may be of interest besides the longest duration and the largest deficit, the conditional distribution of the negative run-sum for given $N=n$ and $T=t$ should be useful in practice. Since the moments of the conditional distribution may be given by Eq. (6.22) for a distribution of ξ, and truncation level ξ_0, an average standard deviation $\bar{\sigma}$, and an amplitude C, the conditional distribution can be estimated from the Pearson system of distributions on the basis of the first four moments (Johnson and Kotz, 1970). Once the conditional distributions are determined, it is possible to calculate their various probabilities.

Table 6.1. Relationship of the Ratio of the Variance V_s Due to the Stationarity to the Variance V to the Ratio of the Amplitude C to the Average Standard Deviation $\bar{\sigma}$.

	Case 1	Case 2
Distribution of ξ	Normal	Eq. (3.35)
ξ_0	0.0	0.0
q	0.5	0.58
$E(\xi^*)$	0.7979	0.6988 *
$\mathrm{Var}(\xi^*)$	0.3634	0.1699 *
V_s	$2.0000\bar{\sigma}^2$	$2.0101\bar{\sigma}^2$
V_p	$1.6069C^2$	$1.3650C^2$
$C/\bar{\sigma}$	$V_s / (V_s + V_p)$	
0.1	0.992	0.993
0.2	0.969	0.974
0.3	0.933	0.942
0.4	0.886	0.902
0.5	0.833	0.855
0.6	0.776	0.804
0.7	0.718	0.750
0.8	0.660	0.697
0.9	0.606	0.645
1.0	0.555	0.596

* Obtained experimentally.

6.2. Discussion on Drought Analyses of Periodic-Stochastic Processes

The techniques for evaluating drought characteristics of periodic-stochastic series are not well developed. For doing it, several promising methods are outlined herein. First, the use of the theory of runs, as applied in the previous section, is a promising method. The good aspect of this approach is that the definitions of run-length, run-sum, and run-intensity are clear and these properties may be studied analytically except when they are very complex. At the same time, this aspect has a weak point. That is, objective definitions are based on the analytical approach so that the physical meanings of the series involved may be ignored. Definitions of runs do not take into account the previous moisture conditions. The previous moisture state, or the previous run-sum, which affects the following conditions in physical and practical terms of the next run-sum, must be considered in evaluating the effect of the present or the next state. This point may become critical with a decrease of time interval of the series. Therefore, run properties without any modifications may not be sufficient to analyze the drought characteristics of periodic-stochastic series such as monthly or weekly precipitation series. However, for the secondary series, which takes the carry-over effect of an original series and therefore is a time dependent series, the run properties may be useful for studying drought characteristics.

Another possible technique in treating the droughts of periodic-stochastic processes is the use of partial sum series or of its cumulative series. The concepts of deficit and surplus, the maximum deficit or the maximum surplus, the range, and the maximum range, are connected to the general theory of water storage. Therefore, these concepts may be useful and proper for the drought analysis of runoff series. They might be applicable for a rather long-range evaluation of droughts than for a short term such as a few months. Guerrero-Salazar and Yevjevich (1975) used the drought-magnitude and drought-duration criteria on the basis of these cumulative series. Their results of a case study seem to support the expectation that the criteria are useful for an analysis of long-term droughts. To analyze the short-term periods, other concepts must be added.

Still another alternative may be in deriving the series by using the water budget, and to investigate this series in the evaluation of drought characteristics. Palmer (1965) developed such a method of calculating the deficit and the surplus index, called the Palmer index. Herbst et al. (1966) developed a rather simple method, based on monthly rainfall data. The method derives the effective rainfall in considering the water carry-over from month to month. Though this method includes subjective procedures or criteria, it is useful and may have a possibility for further developments. An application of the theory of runs to these secondary series may help the investigation of drought characteristics to be objective and quantitative.

A comparison of these various techniques for the same original data would be worth of trying in order to find their advantageous or disadvantageous aspects.

Chapter 7
CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER STUDY

Under the inferred model of the stochastic component of monthly precipitation over the Upper Great Plains, generated samples in a systematic grid were used to analyze drought characteristics. The deficit area, the total areal deficit, and the maximum deficit intensity were used as the primary drought indices. Some other areal aspects of droughts were also investigated. Drought characteristics of periodic-stochastic processes were analyzed by using the theory of runs.

The study leads to the following conclusions:

(1) The extreme that the whole area is completely covered by a drought occurs very rarely in a large area, such as that used in this study, for the truncation levels at the median and the mean, though probabilities of droughts over the area are shown to be relatively high.

(2) Once the deficit area, the total areal deficit, and the maximum deficit intensity were computed, the areal drought structure was defined by Eqs. (5.39) through (5.41).

(3) The probability of a drought covering a whole subarea is more affected by the size than by the shape of the subarea. For a small area, the shape of the area affects probabilities of droughts.

(4) Probability distributions of the deficit area within a subarea are affected by the truncation levels, especially for a small subarea, as well as by the size and the shape of the subarea.

(5) The total areal deficit of monthly precipitation series can be obtained as a product of an areally averaged standard deviation and the total areal deficit of stochastic component series of monthly precipitation.

(6) Under simple assumptions, run properties of univariate periodic-stochastic processes were obtained analytically. The conditional moments of the negative run-sum for given run-length and the onset time, given by Eqs. (6.8), (6.13), and (6.22), are expected to be useful in practice.

(7) The variance of the negative run-sum, consisting of the variance resulting from the stationary stochastic component and the variance resulting from the periodicity, is explained predominately by the variance which results from the stationary stochastic component.

Suggestions and recommendatons for further studies:

(1) Improvements in the regionalization of important parameters of monthly precipitation series for a large area are needed in order to accurately describe the drought characteristics of monthly precipitation series over a large area.

(2) In generating new samples, at a systematic grid of points over a large area, the border effects should be investigated. Some properties of generated series may be biased due to these effects, and the generation method may not be as effective for the solution of some problems as expected.

(3) Systematic and quantitative analysis of observed data over a rather large area is very much needed for the study of drought characteristics, especially of areal coverage of droughts.

(4) The areal drought structure should be further investigated by analyzing each deficit cell (instead of using these combined multi-cell). Analysis of characteristics of each deficit cell, such as its size, shape, deficit intensity, and duration, would give the physical explanation to process modeling.

(5) There is a need for developing advanced techniques of evaluating the drought characteristics of periodic-stochastic processes. By advancing concepts of such techniques, by comparing their results, and by finding positive and negative aspects, general and objective techniques of evaluating drought characteristics of periodic-stochastic processes would be produced.

REFERENCES

Amorocho, J. and A. Brandstetter, 1967, The representation of storm precipitation fields near ground level, Jour. of Geophysical Research, Vol. 72, No. 4, pp. 1143-1164.

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

Court, A., 1961, Area-depth rainfall formulas, Jour. of Geophysical Research, Vol. 66, No. 6, pp. 1823-1831.

Dixon, W.J., 1969, Biomedical computer programs, X-Series Supplement, Univ. of California Publications in Automatic Computation, No. 3, Univ. of California Press, Los Angeles, California.

Downer, R., M.M. Siddiqui, and V. Yevjevich, 1967, Application of runs to hydrologic droughts, Proceedings of the International Hydrology Symposium, Fort Collins, Colorado, pp. 496-505.

Feller, W., 1957, An introduction to probability theory and its applications, Vol. 1, Wiley and Sons, New York.

Fiering, M.B., 1964, Multivariate technique for synthetic hydrology, Jour. of Hydraulic Div., ASCE, Vol. 90, HY 5, pp. 43-60.

Gibbs, W.J. and J.V. Maher, 1967, Rainfall deciles as drought indicators, Bull. No. 48, Bureau of Meteorology, Melbourne.

Guerrero-Salazar, P.L.A., 1973, Statistical modeling of droughts, Unpublished Ph.D. Dissertation, Colorado State University, Fort Collins, Colorado.

Guerrero-Salazar, P. and V. Yevjevich, 1975, Analysis of drought characteristics by the theory of runs, Hydrology Paper No. 80, Colorado State University, Fort Collins, Colorado.

Gumbel, E.J., 1963, Statistical forecast of droughts, Bull. IASH, Vol. VIII, No. 1, pp. 5-23.

Herbst, P.H., D.B. Bredenkamp, and H.M.G. Barker, 1966, A technique for the evaluation of drought from rainfall data, Jour. of Hydrology, Vol. IV, No. 3, pp. 264-272.

Hoyt, J.C., 1938, Drought of 1936 with discussion of the significance of drought in relation to climate, U.S. Geological Survey Water Supply Paper 820, 62 p.

Huff, F.A. and J.C. Neill, 1957, Areal representativeness of point rainfall, Trans., AGU, Vol. 38, No. 3, pp. 341-345.

Huff, F.A. and W.L. Shipp, 1969, Spatial correlations of storm, monthly, and seasonal precipitation, Jour. of Applied Meteorology, Vol. 8, No. 4, pp. 542-550.

Johnson, N.L. and S. Kotz, 1970a, Distributions in statistics: Continuous univariate distribution 1, Houghton Mifflin Company, Boston, Massachusetts, 300 p.

Johnson, N.L. and S. Kotz, 1970b, Distributions in statistics: Continuous univariate distribution 2, Houghton Mifflin Company, Boston, Massachusetts, 306 p.

Karplus, A.K., 1972, Simulation of precipitation series in a region, Unpublished Ph.D. Dissertation, Colorado State University, Fort Collins, Colorado.

Kates, R.W., 1971, Natural hazard in human ecological perspective: hypotheses and models, Economic Geography, Vol. 47, No. 3, pp. 438-451.

Kirby, W., 1972, Computer-oriented Wilson-Hilferty transformation that preserves the first three moments and the lower bound of the Pearson type 3 distribution, Water Resources Research, Vol. 8, No. 5, pp. 1251-1254.

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

Krumbein, W.C., 1963, Confidence intervals on low-order polynomial trend surfaces, Jour. of Geophysical Research, Vol. 68, No. 20, pp. 5869-5878.

Linsley, R.K. and M.A. Kohler, 1951, Variations in storm rainfall over small areas, Trans., AGU, Vol. 32, No. 2, pp. 245-250.

Llamas, J. and M.M. Siddiqui, 1969, Runs of precipitation series, Hydrology Paper No. 33, Colorado State University, Fort Collins, Colorado.

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

Matalas, N.C., 1967, Mathematical assessment of synthetic hydrology, Water Resources Research, Vol. 3, No. 4, pp. 937-945.

McGuire, J.K. and W.C. Palmer, 1957, The 1957 drought in the Eastern United States, Monthly Weather Review, Vol. 85, No. 9, pp. 305-314.

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

Palmer, W.C., 1965, Meteorological drought, Research Paper No. 45, United States Weather Bureau, 58 p.

Palmer, W.C. and L.M. Denny, 1971, Drought bibliography, NOAA Technical Memorandom EDS 20, Environmental Data Service, Silver Spring, Maryland, 237 p.

Pinkayan, S., 1966, Conditional probabilities of occurrence of wet and dry years over a large continental area, Hydrology Paper No. 12, Colorado State University, Fort Collins, Colorado.

Rodriguez-Iturbe, I. and J.M. Mejia, 1974, The design of rainfall networks in time and space, Water Resources Research, Vol. 10, No. 4, pp. 713-728.

Roesner, L.A. and V. Yevjevich, 1966, Mathematical models for time series of monthly precipitation and monthly runoff, Hydrology Paper No. 15, Colorado State University, Fort Collins, Colorado.

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

Schaake, J.C., Jr., M.J. Ganslaw, J.W. Fothergill, and T.E. Harbaugh, 1972, Multivariate rainfall generator for annual, seasonal, monthly and daily events, Proc. of International Symposium on Modeling Techniques in Water Resources System, Vol. 2, pp. 436-460.

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

Spar, J., 1968, The drought in the Northeast, 1962-1966: Introductory remarks, Proc. of the Conference on the Drought in the Northeastern United States, New York University, Sterling Forest, New York, pp. 1-31.

Steinitz, G., A. Huss, A. Manes, R. Sinai and Z. Alperson, 1971, Optimum station network in the tropics, Jour. of Applied Meteorology, Vol. 10, No. 3, pp. 364-369.

Stenhouse, N.S. and E.A. Cornish, 1958, Inter-station correlations of monthly rainfall in South Australia, Div. of Mathematical Statistics Technical Paper No. 5, Commonwealth Scientific and Industrial Research Organization, Australia, pp. 1-22.

Stidd, C.K., 1953, Cube-root normal precipitation distributions, Trans., AGU, Vol. 34, No. 1, pp. 31-35.

Suzuki, E., 1968, Statistical study on the amount of precipitation (part 15) - Normalization procedures of the frequency distribution of the short range amount of precipitation, Jour. of Meteorological Research, Vol. 14, No. 12, pp. 38-49.

Thomas, H.E., 1962, The meteorologic phenomenon of drought in the Southwest, U.S. Geological Survey Professional Paper 372-A.

Thomas, J.B., 1971, An introduction to applied probability and random process, John Wiley and Sons, Inc., New York, 338 p.

Takeuchi, K., 1974, Regional water exchange for drought alleviation, Hydrology Paper No. 70, Colorado State University, Fort Collins, Colorado.

Woolhiser, D.A. and H.C. Schwalen, 1960, Area-depth-frequency relations for thunderstorm rainfall in southern Arizona, Tech. Paper 527, Agricultural Engineering Dept., Univ. of Arizona, Tucson, Arizona.

Yevjevich, V., 1964, Fluctuations of wet and dry years Part II Analysis by serial correlation, Hydrology Paper No. 4, Colorado State University, Fort Collins, Colorado.

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

Yevjevich, V. and A.K. Karplus, 1974, Area-time structure of the monthly precipitation process, Hydrology Paper No. 64, Colorado State University, Fort Collins, Colorado.

Young, G.K. and W.C. Pisano, 1968, Operational hydrology using residuals, Jour. of Hyd. Div., ASCE, Vol. 94, HY 5, pp. 909-923.

Key Words: Drought, Areal Droughts, Drought Deficit, Drought Intensity, Monthly Precipitation.

Abstract: Under the concept that monthly precipitation series over an area are composed of deterministic components specified by periodic parameters and a stationary stochastic component, a mathematical model is developed of the area-time process of monthly precipitation, especially of the stationary stochastic component, using the Upper Great Plains in USA as an example. The independent indentically distributed variables are obtained from the transformed stochastic component. Their regional dependence structure is given by an exponential decay function of the interstation distance. By using this model, new samples of time series over the area at a new grid of 80 points were generated in order to investigate area-deficit intensity drought characteristics.

The deficit area, the total areal deficit, and the maximum deficit intensity are used as indices of drought characteristics. Basic

Key Words: Drought, Areal Droughts, Drought Deficit, Drought Intensity, Monthly Precipitation.

Abstract: Under the concept that monthly precipitation series over an area are composed of deterministic components specified by periodic parameters and a stationary stochastic component, a mathematical model is developed of the area-time process of monthly precipitation, especially of the stationary stochastic component, using the Upper Great Plains in USA as an example. The independent indentically distributed variables are obtained from the transformed stochastic component. Their regional dependence structure is given by an exponential decay function of the interstation distance. By using this model, new samples of time series over the area at a new grid of 80 points were generated in order to investigate area-deficit intensity drought characteristics.

The deficit area, the total areal deficit, and the maximum deficit intensity are used as indices of drought characteristics. Basic

Key Words: Drought, Areal Droughts, Drought Deficit, Drought Intensity, Monthly Precipitation.

Abstract: Under the concept that monthly precipitation series over an area are composed of deterministic components specified by periodic parameters and a stationary stochastic component, a mathematical model is developed of the area-time process of monthly precipitation, especially of the stationary stochastic component, using the Upper Great Plains in USA as an example. The independent indentically distributed variables are obtained from the transformed stochastic component. Their regional dependence structure is given by an exponential decay function of the interstation distance. By using this model, new samples of time series over the area at a new grid of 80 points were generated in order to investigate area-deficit intensity drought characteristics.

The deficit area, the total areal deficit, and the maximum deficit intensity are used as indices of drought characteristics. Basic

Key Words: Drought, Areal Droughts, Drought Deficit, Drought Intensity, Monthly Precipitation.

Abstract: Under the concept that monthly precipitation series over an area are composed of deterministic components specified by periodic parameters and a stationary stochastic component, a mathematical model is developed of the area-time process of monthly precipitation, especially of the stationary stochastic component, using the Upper Great Plains in USA as an example. The independent indentically distributed variables are obtained from the transformed stochastic component. Their regional dependence structure is given by an exponential decay function of the interstation distance. By using this model, new samples of time series over the area at a new grid of 80 points were generated in order to investigate area-deficit intensity drought characteristics.

The deficit area, the total areal deficit, and the maximum deficit intensity are used as indices of drought characteristics. Basic

parameters of their frequency distributions and of mutual relationships are analyzed for various truncation levels in drought definitions.

Probabilities of areal coverage of droughts are further investigated by applying the theory of runs, the theory of recurrent events, and similar approaches. Probabilities of specific areas covered by droughts of given properties are investigated by considering the effects of the size and the shape of an area.

Run properties of a simple, periodic-stochastic process are investigated analytically. Moments of negative run-sums are found by considering the negative run-length and the onset time. Other techniques are discussed in comparison with the use of run properties in evaluating drought characteristics of periodic-stochastic processes.

Reference: Tase, Norio, Colorado State University, Hydrology Paper No. 87 (November 1976), Area-Deficit-Intensity Characteristics of Droughts.

parameters of their frequency distributions and of mutual relationships are analyzed for various truncation levels in drought definitions.

Probabilities of areal coverage of droughts are further investigated by applying the theory of runs, the theory of recurrent events, and similar approaches. Probabilities of specific areas covered by droughts of given properties are investigated by considering the effects of the size and the shape of an area.

Run properties of a simple, periodic-stochastic process are investigated analytically. Moments of negative run-sums are found by considering the negative run-length and the onset time. Other techniques are discussed in comparison with the use of run properties in evaluating drought characteristics of periodic-stochastic processes.

Reference: Tase, Norio, Colorado State University, Hydrology Paper No. 87 (November 1976), Area-Deficit-Intensity Characteristics of Droughts.

parameters of their frequency distributions and of mutual relationships are analyzed for various truncation levels in drought definitions.

Probabilities of areal coverage of droughts are further investigated by applying the theory of runs, the theory of recurrent events, and similar approaches. Probabilities of specific areas covered by droughts of given properties are investigated by considering the effects of the size and the shape of an area.

Run properties of a simple, periodic-stochastic process are investigated analytically. Moments of negative run-sums are found by considering the negative run-length and the onset time. Other techniques are discussed in comparison with the use of run properties in evaluating drought characteristics of periodic-stochastic processes.

Reference: Tase, Norio, Colorado State University, Hydrology Paper No. 87 (November 1976), Area-Deficit-Intensity Characteristics of Droughts.

parameters of their frequency distributions and of mutual relationships are analyzed for various truncation levels in drought definitions.

Probabilities of areal coverage of droughts are further investigated by applying the theory of runs, the theory of recurrent events, and similar approaches. Probabilities of specific areas covered by droughts of given properties are investigated by considering the effects of the size and the shape of an area.

Run properties of a simple, periodic-stochastic process are investigated analytically. Moments of negative run-sums are found by considering the negative run-length and the onset time. Other techniques are discussed in comparison with the use of run properties in evaluating drought characteristics of periodic-stochastic processes.

Reference: Tase, Norio, Colorado State University, Hydrology Paper No. 87 (November 1976), Area-Deficit-Intensity Characteristics of Droughts.

LIST OF PREVIOUS 25 PAPERS

No. 62 Techniques for Modeling Reservoir Salinity, by John Hendrick, August 1973.

No. 63 Mechanics of Soil Erosion From Overland Flow Generated by Simulated Rainfall, by Mustafa Kilinc and Everett V. Richardson, September 1973.

No. 64 Area-Time Structure of the Monthly Precipitation Process, by V. Yevjevich and Alan K. Karplus, August 1973.

No. 65 Almost-Periodic, Stochastic Process of Long-Term Climatic Changes, by William Q. Chin and Vujica Yevjevich, March 1974.

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 the 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.