STOCHASTIC MODELING OF HYDROLOGIC, INTERMITTENT DAILY PROCESSES

by

Jerson Kelman

February 1977

HYDROLOGY PAPERS
COLORADO STATE UNIVERSITY
Fort Collins, Colorado

89

STOCHASTIC MODELING OF HYDROLOGIC, INTERMITTENT DAILY PROCESSES

by
Jerson Kelman

HYDROLOGY PAPERS
COLORADO STATE UNIVERSITY
FORT COLLINS, COLORADO 80523

February 1977 No. 89

TABLE OF CONTENTS

ACKNOWLEDGMENTS

The writer expresses the gratitude to his advisor and major professor, Dr. Vujica Yevjevich, Professor of Civil Engineering, for his help during the writer's graduate work. Thanks go to Dr. D. C. Boes, Associate Professor of Statistics, whose guidance and suggestions were very valuable, as well as to the other members of the Graduate Committee: Dr. D. A. Woolhiser, Affiliated Professor of Civil Engineering, and Dr. M. C. Bryson, Associated Professor of Statistics, for their suggestions in the critical review of the thesis manuscript.

The financial support to the writer during his studies from the Conselho Nacional de Pesquisas (National Research Council, Brazil) is acknowledged. The Acknowledgment goes also to US National Science Foundation, Grant ENG 74-17396 (Stochastic Water Resources Processes), from which the advising time by professors was covered, and some financial support for numerical computations and shaping of the thesis.

Jerson Kelman

ABSTRACT

A model for description and generation of new samples of intermittent daily precipitation series is developed. The basic assumption is that precipitation is a result of truncating a non-intermittent process. Classical methods for modeling the time dependence in this latter process can then be applied. The univariate non-intermittent process permits then an extension to multivariate case. Specific tests, related to stationarity and time independence of the process, are formulated. The model is tested on series of several precipitation stations in USA. Results have been found satisfactory.

Another model, in this case for the description and generation of new samples of daily streamflow, is also developed. The basic assumption is that the rising and falling limbs of discharge hydrographs can be modeled individually as two difference, intermittent processes, also physically different. The rising limb process is mainly due to factors external to watersheds. It is modeled similarly as the intermittent precipitation process. The falling limb is conceived as governed by regularities of water outflow from watersheds, with the watershed storage and outflow represented by two linear reservoirs. A sequence of recession flows is then a stochastic output from these two reservoirs. The model is tested for a case study. Results are satisfactory in reproducing the combined process.

FOREWORD

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

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

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

When the problem of generating new samples by using the Monte Carlo (experimental statistical method) is posed in hydrology and water resources, with the generated data to preserve both the time and space properties of random variables involved, this problem becomes that of a mathematical description and that of the generation of new samples in case of periodic-stochastic, intermittent time series. Both the periodicity in parameters, and the fact that the non-zero values occur at some space points while the zero values are not observed simultaneously at the other points, create difficulties in generating new samples of multi-point intermittent time series. Attempts have been made to apply the combinatorial analysis and Markov chains in order to generate simultaneously the series of 2-3 stations, by generating first their zero and non-zero intervals, and then by preserving both the space and time dependences within the non-zero intervals. Researchers following this approach have been able to simulate only 2-3 station series. For more than four stations, the combinatorial approach becomes so complex that it is then difficult to extend it to cases of five, six, and more intermittent time series.

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

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

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

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

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

February 1977

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

Chapter 1

Introduction

1-1 Needs to Simulate Hydrologic Processes

The need for generating hydrologic sequences in the study of complex water resources problems is recognized by many hydrologists. It does not mean that this so-called experimental (Monte Carlo) method needs to be applied in every or most hydrologic problems. One should use an anlytical solution whenever available rather than any other method. Unfortunately such explicit solutions are rare. Usually the way to extract probabilistic information about the performance of a system is to determine its response or output to a set of new hydrologic sequences obtained through simulation.

2-2 Objectives of the Study

This study is devoted to a development of a model for generating of sequences (samples) of daily precipitation and another model for daily streamflow sequences.

The precipitation is assumed to be a filtered realization of the first-order, linear, autoregressive stochastic process. Figure 1-1 illustrates this filtering procedure. It will be seen in the ensuing chapters that the resulting X_t-process is not only intermittent but also it possesses a mechanism that ensures the persistence in data. Furthermore, the set

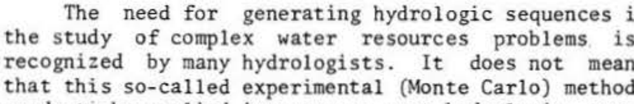


Figure 1-1. Representation of the Intermittent Model.

of positive outcomes can be accepted as drawn from a highly skewed marginal distribution. These characteristics are quite relevant to the time series studied herein. $\{\xi_t\}$ are independent random variables with standard normal distribution;

$\{Z_t\}$ being $Z_t = \mu + \rho(Z_{t-1}-\mu) + \sigma\sqrt{1-\rho^2}\,\xi_t$; $\{Y_t\}$ being $Y_t = Z_t I_{(0,\infty)}(Z_t)$; $\{X_t\}$ being $X_t = Y_t^{1/\alpha}$; $I_{(0,\infty)}(\cdot)$ is the indicator function; and μ, σ, ρ, α are parameters.

The streamflow record, $q(t)$, is analyzed according to its increments $q(t)-q(t-1)$. The positive sequences of these increments are modeled differently from the negative ones. This approach is intended to bring forth a model that takes into consideration the diversity of physical factors that produce streamflows. The positive increments, produced mainly by spells of surface and sub-surface flow, are characterized by a weak persistence. The negative increments are the consequences of the watershed retention and outflow process, and therefore have a strong persistence. The sequence of positive increments has the same form as the precipitation process, because the surface flow may be considered as a slightly filtered rainfall. The sequence of negative increments is obtained by assuming that the recession discharges are a stochastic output of two linear reservoirs. In this model the sequence of positive increments and the sequence of negative increments are, respectively, the realizations of *master* and *slave* stochastic processes.

The model is designed for rivers with runoff predominantly produced by rainfall. Care is recommended when it is used under different conditions, say when snowmelt is a significant input to streamflow. No attempt is made to route the rainfall excess to end up as the streamflow. In fact, these two processes are dealt with separately.

Time intervals shorter than one day are not dealt with in order to avoid the complexities resulting from the diurnal variations in the processes. Nevertheless, techniques are available in the literature for it by breaking down the daily values into hourly values. One of these *within-the-day* periodicities are found to be not significant, there is no conceptual impediment for the use of the models presented herein for modeling processes on a *fraction-of-the-day* time interval.

Meteorological factors related to the precipitation process, for example cloud type, temperature, winds, humidity, etc., are not considered. The observed record is examined merely as a realization of the stochastic process. No physical explanation of precipitation occurrence can be derived from the statistical description of the observations presented herein.

The precipitation model was conceived as reproducing (in a stochastic sense) processes with significant time persistence. No claim is made on the goodness of fit of this model to various types of precipitation.

1-3. Needs for the Use of Daily Series

The use of models to model and generate the annual and/or monthly sequences is already widespread in hydrology. In many situations, when a large scale project is involved, further refinement of the time scale becomes an exercise in futility. However, in many hydrologic studies the use of series of short time intervals is required. For example, Beard (1968) stated that "although fluctuations of flows within a month usually have minor influence on reservoir storage required for conservation purposes, such fluctuations are ordinarily crucial in the determination of reservoir space requirements for flood control." The optimization of a system involving a *run-of-the-river* hydroelectric power plant is another example. In fact, Pfaehler (1933), referring to duration curves, said that "...the monthly curves were used as a basis in arriving at the estimated power output, and checking the figures... by the use of daily streamflow records, the results thus obtained sometimes differed as much as 35% to the disadvantage of the project." In the 30's the difficulty in handling the prodigious amount of daily data justified using the monthly values. In the computer age this is no longer the case. It should be pointed out, however, that the shorter the time interval of the series to be studied, the more difficult it is to develop a generation scheme. It is understandable that time series with long intervals, say annual streamflow, *better behave* than those with the short interval, say the daily streamflow.

1-4 Needs to Model Precipitation Process of Short Intervals

The needs for developing a reliable model for daily streamflow is in general accepted. However, with respect to the precipitation process, a discussion on its needs seems to be appropriate. First, a precipitation time series is mostly homogeneous, a property not always found in streamflows. The latter is frequently affected by man-made structures, while climate is in general stationary.

Second, the generated rainfall samples can be used in deterministic models which route rainfall through the several phases of the land segment of the hydrologic cycle. These models implicitly assume that the stochasticity of the streamflow process is due only to rainfall and potential evapotranspiration, which is equivalent to stating that the stochasticity imbedded in the watershed is integrated in such a way that it yields mean values. These models may be used to predict modifications in the streamflow due to changes in the watershed (for examples, the urbanization) without modifying the generation model for precipitation. Whether the accuracy and the physical meaning of deterministic models is sufficiently high to assure the reliable samples is a question which cannot be answered in the text. Riviews of various philosophies in this approach, as well as a closer examination of some of these deterministic models, may be found in Flemming (1975) or Brown et al. (1974)

Third, many *black-box* techniques are available which connect runoff to rainfall. It is conceivable that due to the better quality and quantity of rainfall data one may choose to face the uncertainty in the transfer function, rather than generating new sequences of streamflow from the unreliable historic records. Regional studies may fit these conditions.

Finally, generated rainfall sequences may be important by themselves, and not merely to be used to produce streamflow sequences, as would be the case in water resources systems which involve the irrigation and urban drainage.

1-5 Outline of Chapter Contents

Chapter II gives a brief survey on the state of the art of modeling as related to the present study.

Chapter III presents the conceptual framework for modeling the intermittent processes. An intermittent process is such that there is a positive probability that an observation is equal to a constant. For example, daily rainfall is such a process, since there is a finite probability that at any given day no rainfall would occur, i.e., that the observation is equal to the constant, in this case the value of zero. Similarly, daily streamflows of small rivers may have zero flows between floods; therefore, they satisfy the definition of intermittent processes.

In Chapter IV the developed model is tested whether it reproduces the major statistics of the rainfall process, either univarite or multivariate. Periodic functions are used to account for the seasonal variation of parameters.

In Chapter V a model for daily streamflow is presented. It uses the dual approach: positive increments of streamflow are represented by an intermittent process which is different from the one related to the negative increments.

Chapter VI presents conclusions and recommendations for further studies.

Chapter II

Brief Review of Models For Daily Rainfall and Daily Streamflow

The first part of this chapter includes a description of mathematical models used for the precipitation process. The second part does the same thing for the streamflow process. No attempt is made to report about all the efforts and contributions made on this subject; only those that are designed for daily data and/or are relevant to the present study are mentioned. For a broader perspective on the topic of stochastic modeling on hydrology the interested reader might consult, for example, Yevjevich (1972), Lawrance and Kottegoda (1976), or Clarke (1973).

2-1 Models for the Precipitation Processes

One of the first thoughts on the subject. One should start with recognizing the fact that most records of daily rainfall have large numbers of zeros. They can be conceived as the realizations of a non-negative, intermittent stochastic process. As a first approach to the problem of modeling such a process, one might consider that a good fit would be obtained by a mixed distribution, with the probability mass concentration $p(0<p<1)$ at the origin of $x = 0$, and a continuous probability density distribution, $(1-p)f(x)$ for $x > 0$. Alternatively, one could lump all the values which are smaller than a small value δ^* (including the zero), and fit a continuous distribution for $x > 0$ in such a way that $P(X<\delta^*) = \int_0^{\delta^*} f(x)\,dx$ is close to the relative frequency in the interval $[0,\delta^*]$. Das (1955) made $\delta^* = 0.05$ inches and applied this method for the Sydney rainfall data from October 17 to November 7 for 94 years. He used a truncated gamma distribution to fit the values larger than δ^*, and obtained good results. Unfortunately, the approach is not powerful enough to satisfy the needs of stochastic hydrology. As it will become clear in the following subsections, a general model should have the capability to cope with these subjects: (i) the non-stationarity of the process, (ii) the time persistence of the process, (iii) the expansion from the univariate case to the multivariate (several rainfall stations), and (iv) the extreme events (the model should be able to reproduce the flood causing type of events).

Stationarity. It is conceivable that if Das (1955) wanted to model not 22 but, say, 100 days he could have analyzed five periods of 20 days each. This is the so-called *season* approach to the obviously non-stationary hydrologic time series with discrete series intervals shorter than the year. An abrupt transition between the last day of season i and the first day of season $i + 1$ may not be acceptable. Therefore, some hydrologists use the continuous variation of parameters along the seasons of the year, creating thus a smooth representation for changing parameters for non-stationary processes. The season can be as short as one day, though it may not be advisable because of an increase of uncertainty in the estimation. Most of the rainfall models use the seasonal approach.

Todorovic (1968), Verschuren (1968), and Todorovic and Yevjevich (1969), attempted at obtaining the explicit expressions for distributions of some functionals of a hypothesized continuous and instantaneous rainfall process, rather than develop a model oriented to generation of samples. In the latter reference the year was divided into 28 seasons, each 13 days long, and the following functionals were studied; (i) the number of complete storm events in a given time interval, (ii) the maximum number of storm events, with the total precipitation which does not exceed a given amount, (iii) the end times of storm events; (iv) the total precipitation for a given number of storm events; (v) the total precipitation for a specific storm event; and (vi) the total precipitation during a given time interval. A storm event was defined either as an uninterrupted period of rainfall, or as a day (or an hour) with rainfall. An assessment of the sensitivity of final results to these two interpretations of uninterrupted period was established. Assuming that the number of storm events in a given time interval (within a season, in order to assure the stationarity) was distributed as Poisson, made some simplifications possible. The result was that the desired probability distributions could actually be evaluated. It was demonstrated that all the functionals were dependent on the two parameters: Γ_1 = the number of storms in a time unit, and Γ_2 = the inverse of the average yield per storm at a given time of the year. The time variation of these two parameters was studied for four precipitation stations in the USA and periodic functions were fitted for the set of 28 points (one for each season), respectively of Γ_1 and Γ_2.

There is an alternative to the seasonal approach. One might consider the raw data as the combination of a deterministic and a stationary stochastic process. When the deterministic component in form of periodic parameters is identified, the hydrologist can isolate the remaining stochastic process, usually modeled by a linear autoregressive scheme. As is well known, there is always a random independent component in any autoregressive model. Therefore, the last task is to fit some probability distribution to this *noise*. Yevjevich (1972b) gives a thorough discussion of this method for the general application in hydrologic time series. However, the type of distribution of the *noise* for the daily rainfall process remains undefined. Adamowski and Smith (1972) assumed that the noise was normally distributed, without giving a justification. It seems that this approach does not work properly for hydrologic time series with short time intervals, although it may be satisfactory for longer time intervals, as for example a month or perhaps even a week. This assertion will be investigated when the streamflow models are dealt with.

Persistence. Wiser (1964) has shown dependence in daily precipitation for North Carolina gauging stations. He states that the dependence is quite a general phenomenon. The degree of dependence is smaller in monthly than in daily series, and also smaller for wet periods than for dry periods of these series. At some locations the dependence tends to a condition in which the information about only the previous day is required for its description.

Grace and Eagleson (1967) report that there is a definite persistence in rainfall values with time interval equal or shorter than the day. They developed a dependence model for the 10-minute rainfall increments by fitting the probability distributions to the length of time between storms and to the duration of each storm. A *storm* is defined as the sequence of observations separated from the others by a row of zeros longer than a certain *critical lag*. An alternating sequence of wet and dry periods could thus be generated. They divided storms into three classes: trace, moderate and peaked. For each class they fitted a linear regression to the storm depth given the storm duration. By fitting a probability distribution to the residuals of the above regression they were able to generate a sequence of storm events, with the total depth of each storm known. The question comes of how to distribute the total amount of precipitation in a given time interval in such a way that the serial dependence is preserved. They developed an interesting technique that is particularly relevant to the present study because it might be applied to transform the generated daily sequences into hourly ones. Suppose that there are n hundreths of an inch of rainfall to be distributed amongst k intervals. An equivalent problem is how to distribute n black balls contained in an urn amongst k boxes. The serial dependence is introduced by adding to the urn m red balls and allocating the balls to the boxes according to the following rule. The first black ball is allocated at random, say to box j. To box j is then given m_o red balls, boxes $j-1$ and $j+1$ are given m_1 of the remaining red balls, and so on. The next black ball is allocated in such a way that the probability of it falling in any given box is proportional to the number of red balls that it contains. Then the process is repeated again. The first and last box must be given at least one black ball in order to assure the duration of the storm. The values of m, m_o, m_1, ..., were selected by trial and error, comparing the correlation coefficients and probability distributions between the generated historic sequences.

Gabriel and Neumann (1962) studying the succession of wet and dry days for the mid-winter period in Tel Aviv, showed that a two state (wet and dry) Markov chain was a good model for representing this dichotomized process. This means that, at least for the situation analyzed by these authors, the probability that day $i+1$ will be wet (or dry) is clearly dependent upon the event which occurred on day i. They were not concerned with generating new rainfall sequences, but concluded by suggesting that a valuable information could be obtained if the amounts of precipitation were included in the analysis.

Green (1964) approached the same problem by assuming that the sequence of dry and wet periods could be modeled by an alternating renewal process, with exponential density functions for the lengths of dry runs and the lengths of wet runs. It was found that the results yielded by this non-Markovian approach were comparable, and sometimes even better, than those obtained by Gabriel and Neumann (1962).

Nicks (1974) used the two-state Markov chain to model the occurrence and non-occurrence of rain on each day for a whole region, rather than for a single point in space. For a wet day the rainfall was generated in two steps: (i) determine which station receives the maximum rainfall and generate its value; this is done by sampling from distributions fitted to the historical data, and (ii) determine the rainfall depth for each station, based on regression on the center of the storm type equations.

Todorovic and Woolhiser (1974) aimed at finding an explicit expression for the probability distribution of the total amount of precipitation, S_n, during a period of n days. Under the hypothesis that the total precipitation for k wet days in a period of n days long is independent of which of these k days were actually wet and which of the $(n-k)$ days were dry, they showed that

$$P(S_n \leq s) = P(N_n = 0) + \sum_{\nu=1}^{n} P(S_\nu^* \leq s)\ )(N_n = \nu),$$

where N_n is the number of wet days in an n-day period; and S_ν^* is the total amount of precipitation for ν wet days. $P(S_\nu^* \leq s)$ was evaluated assuming that $P(S_1^* \leq s) = 1-e^{-\lambda s}$, i.e., the amount of rainfall of a wet day is exponentially distributed. They further assumed that the rainfall depths on different wet days were independent; therefore, S_ν^* is the sum of ν independent exponentially distributed random variables and thus has the gamma distribution. $P(N_n = \nu)$ was evaluated under two hypothesis: (i) that there is no serial dependence in the sequence of wet and dry days, and consequently N_n is assumed binomially distributed, and (ii) the sequence of wet and dry days follows a two-state Markov chain and hence the results of Gabriel and Neumann (1962) are applicable. They found that the Markov-chain exponential model was superior to the binomial-exponential model. This is one more indication that precipitation cannot be treated as a succession of independent events.

Ison, Feyerhem and Bark (1971) also considered the sequences of wet and dry days as a Markov chain. The amount of rainfall in a sequence of n wet days was assumed to be gamma distributed with the scale parameter dependent on n. Therefore the results were similar in some respects to those of Todorovic and Woolhiser (1974).

It appears that a new class of models is at hand if one assumes that not only the dry-wet condition of day $i+1$ depends on the condition of day i, but also the *amount* of precipitation on day $i+1$ depends on the measured depth of the day i. This seems to be a reasonable assumption to make when one deals with precipitation of the frontal type. The obvious way to proceed is to divide the range of observations in $n > 2$ classes (states), rather than have only $n = 2$. An $n \times n$ transition matrix can then be estimated and the concept of Markov chains again applied. Pattison (1965) used this approach to model the hourly rainfall. However, the large probability that a state 0 (no rainfall) will follow the state 0 made the model incapable of reproducing the length of dry periods:

the generated sequences usually had *dry runs* longer than the historic series. To resolve this he broke the models into two parts. The first part, used for wet periods, was the same first-order n x n Markov chain. The second part, used for dry periods, was a sixth-order Markov chain in which each hour was classified only as wet or dry. For generation, the first-order chain was used when the hour i was wet. If the hour i was dry the model shifted to the sixth-order chain, i.e., the information was used on the wet or dry state of hours i-5, i-4, ..., i, in order to generate the new state at the hour i+1. If it happened for the (i+1)-th hour to be dry, the sixth-order chain was used again. If it was wet, the rainfall depth was sampled from a distribution fitted to the *first wet hour*, and the model would again switch to the first-order chain. In the wet period the actual rainfall depth would be obtained by sampling from a uniform distribution defined only for the interval under consideration.

Khanal and Hamrick (1974) used the n state Markov chain to model the daily rainfall. They report that "the problem that Pattison had with the inbetween sequence while synthesizing the hourly rainfall values, does not arise with the daily rainfall values." They considered the process stationary for each month, i.e., the year was divided in twelve seasons. The range for the daily rainfall depth was divided in 14 intervals. Therefore, 13 x 14 = 182 transition probabilities ought to be estimated for each season. They did not attempt to fit the analytical distributions to conditional probabilities of the transition matrix. Whenever a state was reached, the midpoint of the corresponding interval was assigned as the generated rainfall depth. Any Markov-chain approach suffers from the opposite effects between the need for a large number of states (for an increase of precision) and the explosion of the number of transition probabilities which must be estimated. Analytical distributions not always can be fitted to alleviate the problem.

The Multivariate Case. None of the techniques described so far appears to be apt for generalization in order to treat the multiple station case. Franz (1974) developed a model for the multivariate hourly rainfall. He tested it for a three-station network in northern California. The storm and interstorm events were modeled separately. A storm was taken to be a consecutive series of hours in which each hour had the rainfall recorded at one or more stations of the network. It was assumed that the data corresponding to storm periods can be transformed in such a way that it will appear as a sample from a multivariate normal distribution. Strictly speaking the set of transformed observations does not constitute a random sample because of the persistence in data and the lack of negative values. The persistence was included by treating the transformed series as a Markov model of lag one. The limited range was included by assuming that all the negative values have been set to zero before the sample was observed. The transformation used to normalize the marginal distributions was of the form $Y = a+bX^g$, where Y = the normal variable, X = the observed values, and a, b and g are the parameters. The estimation of these parameters was performed by fitting the above equation, by a least squares approach, to the pairs (x_i, y_i). For each x_i the value of y_i was obtained in such a way that $\int_{-\infty}^{y_i} \phi(t)dt = P(Y < y_i) = \hat{P}(X < x_i)$, where $\phi(\cdot)$ is the p.d.f. of the standard normal and $\hat{P}(X < x_i)$ stands for the sample c.d.f. evaluated at x_i. The covariances between the transformed variables of different stations were also found through fitting procedures. For storms the generation procedure followed the steps suggested by Matalas (1967). It was necessary, to take into account the stationarity considerations, to divide the year into four seasons. The interstorm model required that the year to be divided into 50 seasons, for each one of them an empirically defined distribution was found for the interstorm length. No single distribution could be fitted accurately to interstorms. It was concluded that empirical adjustments had to be used to obtain an acceptable level of performance.

One must not confuse the multivariate with the multidimensional models. The first category deals with the rainfall as point processes; with the observations of one station related to those of the other through a correlation structure, regardless of the distance between the stations. The second category deals with a process that is not only dependent upon the time but also upon the geographic location; a good introduction to multidimensional models can be found in Bras and Rodriguez (1975).

Extreme Events. The generated sequences should imitate the historical sequence, and not only for the average conditions but also for the situations in which floods or droughts are of the concern. In other words, the *tail events*, or the very large observations, should occur in the generated sequences with the same magnitudes, pattern and frequence as in the historical one. By the same token, the dry intervals should be correctly reproduced. An agricultural drought is related to the sequence of dry and wet runs of rainfall during the growing season for crops. It is surprising that very little attention has been given to these two factors by the builders of hydrologic models for rainfall. An exception is the work by Todorovic and Woolhiser (1976), who gave the distribution of the largest daily value of precipitation in the n-day period, for the same set of assumptions as advanced in their previous work (Todorovic and Woolhiser, 1974). Gupta and Duckstein (1975) concentrated on the problem of the maximum dry interval for a point rainfall process. They assumed, as many others did, that the number of wet days in an n-day period is Poisson distributed. They reported a good agreement between the theoretical and empirical distribution functions.

Other Reviews. Complementary reviews to the present one might be found in Todorovic and Woolhiser (1976); also in Rhenals, Rodriguez, and Schaake (1974).

2-2 Models for Streamflow Processes

Direct Approach. Likewise to precipitation, a daily streamflow model should be able to cope with four aspects: non-stationarity, persistence, multivariate case, and extremes. At first, one might try to approach the problem of how to model the process by using the same successful techniques employed in studying the hydrologic series with the longer time intervals, such as a month. For example, following Yevjevich (1972b) let the daily streamflow sequence be represented by $\{x_i\}$, where $i = 1, 2, \ldots, n$ (n = number of years). If m_τ and s_τ, $\tau = 1, 2, \ldots, 365$, are designated as daily means and daily standard deviations, respectively, the standardization of the process gives $\varepsilon_i = (x_i - m_\tau)/s_\tau$, in which ε_i is the new reduced variable. This process may be stationary and quite often well modeled by a linear autoregressive scheme. For the sake of simplicity, let us assume that a second-order model is appropriate, namely

$$\varepsilon_i = \alpha_1 \varepsilon_{i-1} + \alpha_2 \varepsilon_{i-2} + \sqrt{1 - \alpha_1^2 - \alpha_2^2 - 2\alpha_1\alpha_2\rho}\ \xi_i ,$$

where α_1, α_2, and ρ = the parameters and ξ_i = the random component, with mean zero and variance unity, independent and identically distributed over all τ positions. Quimpo (1967) applied this scheme to daily runoff records of the 17 rivers and found that indeed all the residual series satisfied the second-order autoregressive representation.

Tao (1973), using the same data as Quimpo (1967), made an extensive attempt to fit a distribution function to the random component, ξ_i. In his words "...no distribution was found to fit the frequency distribution of the daily variables, because of the sharp peak and high skewness of the empirical distributions." For longer time intervals, however, he was able to fit distributions with unusually high number of parameters. For example, for the 7-day variables the double-branch gamma function with six parameters was found most applicable. He also devoted attention to testing whether the distribution of the random component had or had not a heavy tail. This is somewhat surprising since the very important problem of tail behavior (for extreme events) is usually neglected by model builders. He concluded that the distributions of the studied variables did not possess heavy tails.

Kottegoda (1972) avoided the complexities of daily streamflow because "...the high variance of the flows, the unconventional probability distributions, and the failure of the simulation processes to transfer hydrograph characteristics of the historical flows." Instead he aimed to model the 5-day streamflow. For the ε_i-process he found the fourth-order autoregressive representation to be appropriate. The distribution for the random component, ξ_i, was searched among the Pearson system and Johnson type distribution functions. It was concluded that the best fits were obtained by using the Pearson Type III and Type VI, and the lognormal distribution functions.

Indirect Approach. Since the direct approach for generating daily sequences is unsuccessful most of the time, one alternative procedure is often used, namely the values are generated for longer time intervals, say a month or a week, and then *distributed* among the days. For example, Green (1973) used Kottegoda's (1972) model to generate sequences of 5-day average flows, and then split them into daily average flows using a sophisticated method of interpolation. Beard (1968) used a linear regression of the standard deviation of daily flow logarithms, within each month of record, upon the logarithm of the total flow for that month. The daily values were obtained considering $\{\varepsilon_i\}$ as a second-order autoregressive process, and s_τ as a linear function of the generated monthly streamflow, this one generated by some other model.

Further Comments. Many other attempts have been made to develop the daily runoff models. However, it seems fair to say that all of them have serious limitations. It is this writer's opinion, this condition can only be changed if hydrologists recognize that the high complexity of the process stems from the diversity of the factors that are lumped into the streamflow. The only hope for improvement is to embody into the stochastic models some knowledge about the physical processes that cause runoff.

Chapter III

MODEL FOR INTERMITTENT PROCESSES

In this chapter a model for intermittent processes is developed and proposed. Hopefully it will be a useful tool for hydrologists studying time series such as rainfall, overland flow, and the runoff of ephemeral rivers. In Chapter V it will be shown that the model is also often appropriate to represent the positive increments of streamflow. The model was conceived with the generation of new samples in mind. Therefore, an important objective in the model building stage was to obtain a simple-to-use scheme of generation, even for a multivariate case; and yet fulfill all the requirements specified in Chapter II. This does not imply that the estimation procedure is simple. As a matter of fact, quite the opposite comes out to be true.

3-1 The Conceptual Framework

Let us assume that a stochastic process follows a first-order autoregressive model. Furthermore, let us admit that the marginal distribution is normal, namely

$$Z_t = \mu + \rho(Z_{t-1}-\mu) + \sigma\sqrt{1-\rho^2}\ \xi_t$$

where $\xi_t \doteq N(0,1)$, and $Z_t \doteq N(\mu,\sigma^2)$. (3-1)

Obviously, the Z_t-process is far from resembling an intermittent record such as daily rainfall (for the sake of simplicity in this chapter only daily rainfall will be considered). Therefore, some filtering is necessary, at least to eliminate the negative values of Z_t.

Define a Y_t-process as:

$$Y_t = Z_t, \text{ if } Z_t > 0$$

$$Y_t = 0, \text{ if } Z_t \leq 0 \qquad (3\text{-}2)$$

A realization of the Y_t-process can be considered as a censored sample of Z_t. A censored sample is such sequence for which the values of the process that fall in a specified interval are not known. For example, all zero values in a realization of the Y_t-process represent negative but unknown observations of Z_t. In this case the censoring interval is $(-\infty,0)$. For this example, the resulting sample would be truncated, if the negative values of Z_t were not only censored but also deleted from the record. In this case even the number of negative outcomes would not be known.

It is clear that Y_t is an intermittent process, provided with a mechanism of persistence. It remains to be seen whether this mechanism is appropriate in modeling and whether the marginal distribution of the positive observations obtained through the Y_t model, namely $P(Y_t<y|Y_t>0)$ fits the sample distribution well. In fact this last condition is not satisfied, because quite often the marginal distributions, in case the positive observations of the processes are only studied, are characterized by a high skewness (higher than the one obtained by the truncated normal). Incidently, the *truncated normal* is the name given to the cumulative distribution function (c.d.f.)

$$P(Y<y) = \frac{\Phi[(y-\mu)/\sigma]}{\Phi[\mu/\sigma]}\ I_{(o,\infty)}(y) \qquad (3\text{-}3)$$

where $\Phi(\cdot)$ is the c.d.f. for the standard normal distribution. The positive values of Y_t might then be considered as a sample of this truncated normal distribution.

3-2 Need for a Power Transformation of the Truncated Normal

An examination of a typical case will help to explain why Y_t is not sufficient to represent the precipitation process. The histogram of the positive observations of daily rainfall at Austin for 70 years during the period May 1-June 1 is plotted in Figure 3-1. For comparison the probability density functions (P.d.f.) which correspond to the truncated normal, and to the exponential distributions are also plotted in Figure 3-1. The exponential distribution is included because it is often used to model the precipitation (see Chapter II). The p.d.f. of the exponential distribution is

$$f_X(x) = \psi e^{-\psi x} I_{(o,\infty)}(x) \qquad (3\text{-}4)$$

The parameter ψ is routinely estimated as the inverse of the arithmetic mean of the positive observations. For the Austin example $\hat{\psi} = 1.898$. The p.d.f. of the truncated normal distribution is

$$f_X(x) = \frac{1}{\Phi(\frac{\mu}{\sigma})\ \sqrt{2\pi}\sigma} \exp\{-\frac{1}{2}(\frac{x-\mu}{\sigma})\}\ I_{(o,\infty)}(x) \qquad (3\text{-}5)$$

The parameters μ and σ are in principle estimated following the procedure proposed by Cohen (1959). However, Cohen was mostly concerned with cases in which the number of censored elements is small compared with the total number of observations. In precipitation data there is a large number of zeros (censored observations). It turns out that graphs and tables supplied by Cohen are not sufficiently complete to handle this situation. Alternatively, an estimation procedure presented in Section 3-3 is employed, and as will be seen, it is a better approach, because it takes into consideration the serial dependence. For the moment it is sufficient to give the estimates $\hat{\mu}= 0.627$ and $\hat{\sigma}=0.951$. The exponential one-parameter distribution was fitted only to positive observations, while the two-parameter truncated normal was fitted to the censored sample, in which the number of zeros was important. Since the probability of a zero outcome depends on the ratio μ/σ, it can be said that both distributions, exponential and truncated normal, had one degree-of-freedom to fit the data.

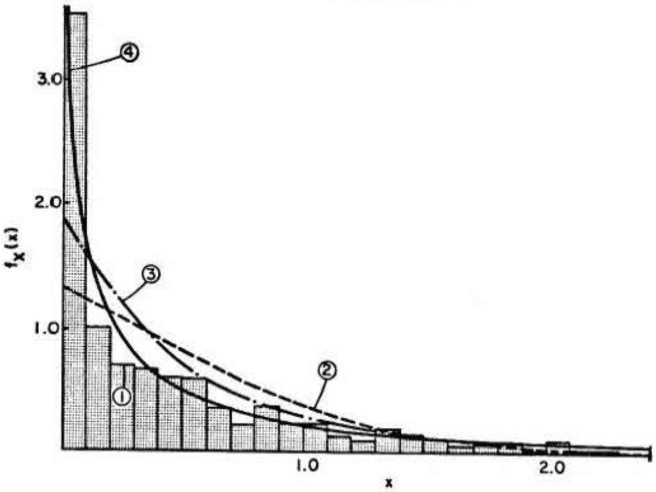


Fig. 3-1. Comparison in Fitting Three Probability Density Functions to the Frequency Histogram of Daily Rainfall at Austin for the Interval May 1-June 1: (1) Histogram for 70 Years of Data; (2) Fit of the Truncated Normal, Eq. (3-5); (3) Fit of the Negative Exponential, Eq. (3-4); and (4) Fit of the Power-Transformed Truncated Normal, Eq. (3-7).

The inspection of Figure 3-1 leads to the conclusion that none of the two distributions produces a good fit. The form of the histogram suggests that a better fit could be obtained by using a p.d.f. which is asymptotic to the vertical axis.

Suppose that the Y_t-process is filtered according to

$$X_t = Y_t^{1/\alpha} , \tag{3-6}$$

with α = a real number. In this case the marginal distribution of positive observations of the X_t-process is the power-transformed truncated normal distribution (p.t.t.n., for short), namely

$$f_X(x) = \frac{\alpha x^{\alpha-1}}{\Phi(\mu/\sigma)\sigma\sqrt{2\pi}} \exp\{-\frac{1}{2}(\frac{x^\alpha-\mu}{\sigma})\} \; I_{(o,\infty)}(x) \tag{3-7}$$

Notice that when $\alpha < 1$, $\lim_{x\to 0} f_x(x) = \infty$. From the procedure to be presented in Section 3-3, for the Austin rainfall example, the estimate is $\hat{\alpha} = 0.595$. The corresponding p.d.f. is plotted in Figure 3-1. From visual inspection, without any test, it is apparent that the p.t.t.n. does fit better the frequency histogram than the other two p.d.f.

3-3 The Estimation Procedure

Seek for the maximum likelihood estimates. Given a sample x_1, x_2, x_3, ... of an intermittent process, a method should be available for estimating the parameters $\mu, \sigma, \rho, \alpha$ (see Figure 1-1). Usually the available samples will be large. The maximum likelihood estimators possess several asymptotic properties. Some of these properties are essential to the analysis of data. Therefore it is natural to select the maximum likelihood estimation procedure for the intermittent process.

Let us approach the problem straightforwardly, but showing that some *tricks*, as presented later, are necessary. Suppose the time series displayed in Figure 3-2(a) is available. In general the likelihood function L is

$$L(\underline{\theta},\underline{x}) = f_{\underline{X}}(\underline{\theta};x_1,x_2,\ldots,x_m) , \tag{3-8}$$

where $\underline{\theta}$ is the parameter vector. The first difficulty arises from the fact that some of the x_i are zero, and therefore represent censored outcomes of the Z_t-process. For the realization shown in Figure 3-2(a), z_3, z_4, z_{m-6}, z_{m-3}, z_{m-2}, z_{m-1} are all censored. The second difficulty arises from the fact that x_i are not independent in sequence, so that Eq. 3-8 cannot be written as a product of marginals.

To show how the difficulties of censoring and dependence complicate the estimation procedure, it is sufficient to examine a simple case. Assume $\alpha = 1$ and the need to write the likelihood function, given the realization displayed in Figure 3-2(b).

The m values between x_o and x_{m+1} represent the unmeasured negative values. Assume the dependence follows a Markov or linear autoregressive model,

$$f_{X_i|X_{i-1},X_{i-2},\ldots}(x_i|x_{i-1},x_{i-2},\ldots) = f_{X_i|X_{i-1}}(x_i|x_{i-1}) \tag{3-9}$$

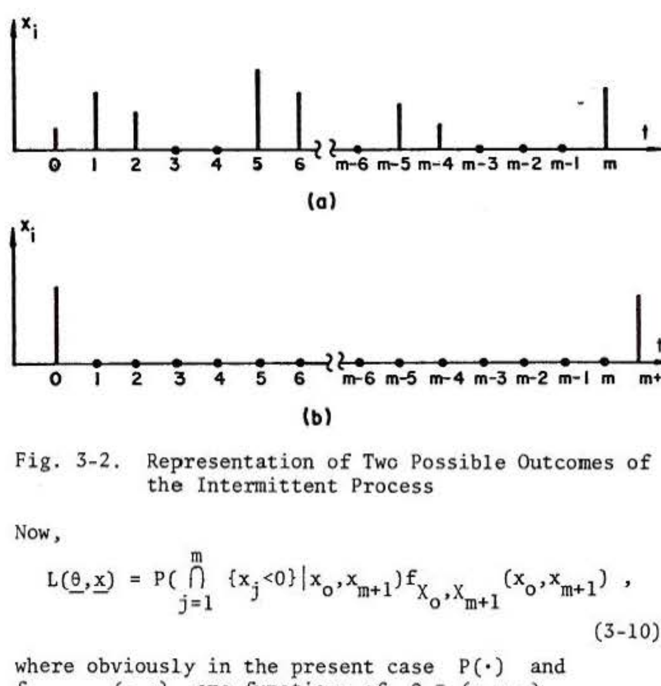


Fig. 3-2. Representation of Two Possible Outcomes of the Intermittent Process

Now,

$$L(\underline{\theta},\underline{x}) = P(\bigcap_{j=1}^{m} \{x_j<0\}|x_o,x_{m+1})f_{X_o,X_{m+1}}(x_o,x_{m+1}) , \tag{3-10}$$

where obviously in the present case $P(\cdot)$ and $f_{X_o,X_{m+1}}(\cdot,\cdot)$ are functions of $\underline{\theta} \equiv (\mu,\sigma,\rho)$.

Therefore,

$$L(\underline{\theta},\underline{x}) = f_X(x_o)f_{X_{m+1}|X_o}(x_{m+1}|x_o) \underbrace{\int_{-\infty}^{o}\int_{-\infty}^{o}\ldots\int_{-\infty}^{o}}_{m} f_{\underline{X}}(x_1,x_2,\ldots,x_m|x_o,x_{m+1}) \quad (dx_1, dx_2, \ldots, dx_m) \tag{3-11}$$

where $f_{\underline{X}}(\underline{x})$ is the p.d.f. for a multivariate normal,

$$f_{\underline{X}}(x_1, x_2, \ldots x_m | x_o, x_{m+1})$$

$$= (2\pi)^{-m/2} \Sigma^{-1/2} \exp[-\tfrac{1}{2}(\underline{x}-\underline{\mu})' \Sigma^{-1} (\underline{x}-\underline{\mu})] \tag{3-12}$$

$\underline{\mu}' = (\mu_1, \mu_2, \ldots, \mu_m)$ is the mean vector

$$\mu_i = \frac{1}{1-\rho^{2m+2}} [(\rho^i - \rho^{2m+2-i})x_o + (\rho^{m+1-i} - \rho^{m+1+i})x_{m+1}$$

$$+ \mu(1-\rho^{2m+2} - \rho^{m+1-i} + \rho^{m+1+i} - \rho^i + \rho^{2m+2-i})]$$

and $\Sigma = (\sigma_{ij})$, i=1, m and j=1, m, is the covariance matrix,

$$\sigma_{ij} = \frac{\sigma^2}{(1-\rho^{2m+2})} (\rho^{j-i} - \rho^{2m+2-i-j} + \rho^{2m+2+i-j} - \rho^{i+j}) \ , \ i \leq j \tag{3-13}$$

In general the analytical solution for the m-fold integral of Eq. (3-11) is not available, except for $m \leq 3$. Even for $m = 3$ the expression is very cumbersome. An alternative is to use the tetrachoric series expansion suggested by Kendal et al. (1963), used and extended by Saldarriaga and Yevjevich (1970). Nevertheless, according to Kendall (vol. 1, p. 351), the technique "though convergent, converges too slowly to be of general use."

This example shows that the straightforward approach of evaluating the likelihood function is, in this particular case, untractable.

The iterative algorithm for the univariate case. Several attempts were made to find the approximate solutions to this estimation problem. Unfortunately, none has worked satisfactorily. As an alternative, a solution to an approximate problem was searched for, rather than looking for an approximate solution to the correct problem.

The approximate problem is to find the estimates for the parameters assuming the pairs of values (X_1, X_2), (X_3, X_4), (X_5, X_6), ..., to be independent. The experience obtained on generated series, i.e., in such situations that *population* parameters were known, supports the results obtained under the above simplifying approximation.

The estimation problem reduces to the evaluation of parameters of a bivariate distribution. Suppose a sample (x_t, x_{t+1}), t=1,3,5,...,n, to be available. Define the three events as:

$$A_{1t} = \{X_t = 0, X_{t+1} = 0\}$$

$$A_{2t} = \{X_t = x_t, X_{t+1} = y_t\}$$

$$A_{3t} = \{X_t = z_t, X_{t+1} = 0\} \text{ or } \{X_t = 0, X_{t+1} = z_t\}$$

$$0 < x_t, y_t, z_t$$

Assume further for the sample given that each of the events A_{1t}, A_{2t}, A_{3t} occurs respectively n_1, n_2 and n_3 times $(n - n_1+n_2+n_3)$. The likelihood function is then

$$L(\mu,\sigma,\rho,\alpha) = \frac{n!}{n_1! n_2! n_3!} [P(A_1)]^{n_1} \prod_{t=1}^{n_2} P(A_{2t}) \prod_{t=1}^{n_3} P(A_{3t}) \tag{3-14}$$

for $(U,V)' \simeq N[\binom{\mu}{\mu}, \begin{pmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{pmatrix}]$

$$P(A_1) = P(U < 0, V < 0) \tag{3-15}$$

The random variables $(U-\mu)/\sigma$ and $(V-\mu)/\sigma$ might be expressed respectively by

$$\left.\begin{array}{l} \sqrt{\rho}\, W_1 + \sqrt{1-\rho}\, W_2 \\ \text{and} \\ \sqrt{\rho}\, W_1 + \sqrt{1-\rho}\, W_3 \end{array}\right\} \tag{3-16}$$

with W_1, W_2, W_3 the independent standard normal variables. From (3-15) and (3-16 it results

$$P(A_1) = \int_{-\infty}^{\infty} \phi(t) [\Phi(\frac{-\mu-\sqrt{\rho}\sigma t}{\sigma\sqrt{1-\rho}})]^2 \, dt \tag{3-17}$$

Similarly,

$$P(A_{2t}) = f_{U,V}(x_t^\alpha, y_t^\alpha) du\, dv \tag{3-18}$$

where

$$\left.\begin{array}{l} f_{U,V}(u,v) = \dfrac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp\{\dfrac{Q}{2(1-\rho^2)}\} \\ \text{with} \\ Q = (\frac{u-\mu}{\sigma})^2 - 2\rho(\frac{u-\mu}{\sigma})(\frac{v-\mu}{\sigma}) + (\frac{v-\mu}{\sigma})^2 \end{array}\right\} \tag{3-19}$$

Now

$$f_{U,V}(x_t^\alpha, y_t^\alpha)\, du dv = J[\frac{u,v}{x,y}]\, f_{U,V}(x_t^\alpha, y_t^\alpha)\, dx dy$$

where the Jacobian is

$$J[\frac{u,v}{x,y}] = \begin{vmatrix} \partial u/\partial x & \partial u/\partial y \\ \partial v/\partial x & \partial v/\partial y \end{vmatrix} = \begin{vmatrix} \alpha x_t^{\alpha-1} & 0 \\ 0 & \alpha x_t^{\alpha-1} \end{vmatrix} = \alpha^2 (x_t y_t)^{\alpha-1}$$

Hence, Eq. (3-18) becomes

$$P(A_{2t}) = \alpha^2 (x_t y_t)^{\alpha-1} f_{U,V}(x_t^\alpha, y_t^\alpha) dx\, dy \tag{3-20}$$

Finally,

$$P(A_{3t}) = f_U(z_t^\alpha)\, du \int_{-\infty}^{0} f_{V|U}(v|u = z_t^\alpha) dv$$

$$= f_U(z_t^\alpha)\, du \int_{-\infty}^{0} \frac{1}{\sigma\sqrt{1-\rho^2}} \phi[\frac{v-\rho z_t^\alpha - \mu(1-\rho)}{\sigma\sqrt{1-\rho^2}}]\, dv \ ,$$

or

$$P(A_{3t}) = \frac{\alpha z_t^{\alpha-1}}{\sigma} \phi[\frac{z_t^\alpha - \mu}{\sigma}]\, \Phi\, [\frac{-\rho z_t^\alpha - \mu(1-\rho)}{\sigma\sqrt{1-\rho^2}}]\, dz \tag{3-21}$$

From Eqs. (3-14), (3-17), (3-20), and (3-21), and after dropping the subscripts,

$\log L =$

$$= LL(\mu,\sigma,\rho,\alpha) = C + n_1 \log \int_{-\infty}^{\infty} \phi(t)\left[\Phi\left(\frac{-\mu-\sqrt{\rho}\sigma t}{\sigma\sqrt{1-\rho}}\right)\right]^2 dt$$

$$+ (2n_2 + n_3) \log \frac{\alpha}{\sigma} - n_2\left[\frac{\log(1-\rho^2)}{2} + \frac{\mu^2}{(1+\rho)\sigma^2}\right]$$

$$+ \sum^{n_2} \left[(\alpha-1)\log(xy) + \frac{2\mu(1-\rho)(x^\alpha+y^\alpha)-(x^{2\alpha}+y^{2\alpha})+2\rho(xy)^\alpha}{2(1-\rho^2)\sigma^2}\right]$$

$$+ \sum^{n_3} \left[(\alpha-1)\log z + \log\Phi\left[\frac{-\rho z^\alpha-\mu(1-\rho)}{\sigma\sqrt{1-\rho^2}}\right] + \log\phi\left(\frac{z^\alpha-\mu}{\sigma}\right)\right] \tag{3-22}$$

where C is a constant.

The estimate $\underline{\hat{\theta}} \equiv (\hat{\mu},\hat{\sigma},\hat{\rho},\hat{\alpha})$ ought to be found in such a way that the likelihood function, or its logarithm, becomes the maximum for $\underline{\hat{\theta}}$. When the objective function LL is concave, as is the case here, it is enough to search for a local maximum, since this will be a global maximum. Recall that a necessary condition for a local optimum is that the first derivatives become zeros. Therefore, the estimation problem is equivalent to finding the point $\underline{\hat{\theta}}$ for which the first derivatives of $LL(\underline{\theta})$ are simultaneously equal to zero. This can be accomplished numerically, through the Newton-Raphson algorithm, by

$$\underline{\theta}_{NEW} = \underline{\theta}_{OLD} - (H^{-1}\underline{D})\underline{\theta}_{OLD} \tag{3-23}$$

where H is the Hessian matrix corresponding to the LL-function,

namely

$$H = \begin{bmatrix} \partial^2 LL/\partial\mu^2 & \partial^2 LL/\partial\mu\partial\sigma & \partial^2 LL/\partial\mu\partial\rho & \partial^2 LL/\partial\mu\partial\alpha \\ & \partial^2 LL/\partial\sigma^2 & \partial^2 LL/\partial\sigma\partial\rho & \partial^2 LL/\partial\sigma\partial\alpha \\ & & \partial^2 LL/\partial\rho^2 & \partial^2 LL/\partial\rho\partial\alpha \\ & & & \partial^2 LL/\partial\alpha^2 \end{bmatrix} \tag{3-24}$$

and $\underline{D}' = (\partial LL/\partial\mu, \partial LL/\partial\sigma, \partial LL/\partial\rho, \partial LL/\partial\alpha)$. The first and second derivatives of LL, needed to evaluate Eq. (3-23), are given in Appendix A.

Extension to the Multivariate Case. In generating samples of several dependent station series, the cross correlations ought to be preserved. For the sake of simplicity only the lag-zero cross correlation will be considered. Yet, the estimation problem becomes greatly complicated because of the increase on the dimensionality of the parameter space. In order to avoid the use of the objective functions with too many variables, the following two-step procedure is proposed in dealing with the multivariate cases: (i) To find for each station the parameters $\mu(j)$, $\rho(j)$, and $\alpha(j)$, according to algorithm of Eq. (3-23); for ℓ the number of stations, $j = 1,2,\ldots,\ell$; and (ii) Find each lag-zero cross correlation coefficient, $\rho(j,k)$, $1 \leq j < k \leq \ell$, using only the data of station series j and k.

The estimation procedure for the multivariate case can, in principle, follow the approach used for the univariate case, namely the maximization of the likelihood function. However, if all the positive observations of station series j are raised on the $\hat{\alpha}(j)$ power, the problem is reduced to the question as to how to estimate the correlation coefficient of a standard bivariate normal distribution, with a censored sample. The censoring is done in such a way that observations respectively of the two variables to the left of $-\hat{\mu}(j)/\hat{\sigma}(j)$ and $-\hat{\mu}(k)/\hat{\sigma}(k)$ are not available. For truncated samples (rather than censored), when only the observations (events) type A_2 are available, the problem was solved by Rosenbaum (1961) and by Regier et al. (1971). Often one might be satisfied with the use of expressions derived by the *truncated approach*, even at the cost of losing some information, because they are easier to use. This is the course of action herein chosen. Rosenbaum used the method of moments and obtained a particularly simple expression, adopted for the purposes of this study as,

$$- \left[\frac{\hat{\mu}(j)}{\hat{\sigma}(j)} + \frac{\hat{\mu}(k)}{\hat{\sigma}(k)}\right] \hat{\rho}^2(j,k)$$

$$+ \left\{\left[\frac{\hat{\mu}(j)}{\hat{\sigma}(j)} + \frac{\hat{\mu}(k)}{\hat{\sigma}(k)}\right] \hat{m}(j,k) + \frac{\hat{\mu}(j)\hat{\mu}(k)}{\hat{\sigma}(j)\hat{\sigma}(k)} [\hat{m}_1(j)$$

$$+ \hat{m}_1(k)]\right\}\hat{\rho}(j,k) + \left[\frac{\hat{\mu}(j)}{\hat{\sigma}(j)} + \frac{\hat{\mu}(k)}{\hat{\sigma}(k)}\right] - \frac{\hat{\mu}(j)\hat{\mu}(k)}{\hat{\sigma}(j)\hat{\sigma}(k)} [\hat{m}_1(j)$$

$$+ \hat{m}_1(k)] - \frac{\hat{\mu}(j)}{\hat{\sigma}(j)} \hat{m}_2(j) - \frac{\hat{\mu}(k)}{\hat{\sigma}(k)} \hat{m}_2(k)$$

where

$$\hat{m}_1(j) = \frac{\sum^{n_2} x^{\hat{\alpha}(j)} - \hat{\mu}(j)n_2}{n_2\,\hat{\sigma}(j)}, \quad \hat{m}_2(j) = \sum^{n_2}\left[\frac{[x^{\hat{\alpha}(j)} - \hat{\mu}(j)]^2}{n_2\,\hat{\sigma}^2(j)}\right],$$

$$\hat{m}_1(k) = \frac{\sum^{n_2} y^{\hat{\alpha}(k)} - \hat{\mu}(k)n_2}{n_2\,\hat{\sigma}(k)}, \quad \hat{m}_2(k) = \sum^{n_2}\left[\frac{[y^{\hat{\alpha}(k)} - \hat{\mu}(k)]^2}{n_2\,\hat{\sigma}^2(k)}\right],$$

$$\hat{m}(j,k) = \sum^{n_2}\left[\frac{(x^{\hat{\alpha}(j)} - \hat{\mu}(j))(y^{\hat{\alpha}(k)} - \hat{\mu}(k))}{n_2\hat{\sigma}(j)\hat{\sigma}(k)}\right], \tag{3-25}$$

and $\hat{\rho}(j,k)$ is the only unknown. Once again it is emphasized that the expression of Eq. (3-25) is to be used for the data of days with non-zero observations occur in both stations under consideration. All the remaining information is neglected. Because of the sample variation Eq. (3-25) may not have the real roots.

3-4 The Asymptotic Covariance Matrix for the Estimators

Method of Obtaining the Covariance Matrix. The estimation of a parameter is not always sufficient. Sometimes it is necessary to find how the system under study reacts to variations in parameter values. This is the so-called sensitivity analysis. Its use results from the recognition that an estimate $\underline{\hat{\theta}}$ is one observation of a random variable and as such is subject to sampling departure from the unknown population value $\underline{\theta}$. Then the question to deal with is the variation of the parameter vector. One would expect that the different parameters will have different *reasonable* ranges of variation, according to the confidence one has on the accuracy of its estimation. Actually, the higher reliability of an estimate, the narrower is its *reasonable* range. The measure of reliability of an estimate most often used is the variance of the corresponding estimator. Similarly, covariances between the estimators give measures of their dependence, which help in the decision how the parameters should be simultaneously changed in the sensitivity analysis. Hence, it is highly needed to calculate the covariance matrix of the estimators. It happens that maximum likelihood estimators asymptotically follow the multivariate normal distribution. Furthermore, the asymptotic covariance matrix can be expressed by $\{-E(H)\}^{-1}$, where H is the Hessian matrix given by Eq. (3-24) and $E(\cdot)$ is the expectation operator. This is useful for the problems herein analyzed due to the fact that most of the samples are of large size.

In order to find the asymptotic covariance matrix, the first step is to evaluate the expected value for each one of the second derivatives which appear in Appendix A. An inspection of these expressions shows that this task is difficult. As a result numerical approximations are used. Fortunately these approximations have no significant effect on the accuracy of the results, as it will be seen in an ensuing example.

The values of n_1, n_2, and n_3 which appear in equations of Appendix A, are the actual observations of random variables. In order to evaluate the asymptotic covariance maxtrix, they should be substituted into their corresponding expected values; namely

$$E(N_1) = \frac{n}{2} P(U < 0,\ V < 0) = \frac{nI(0,0,2)}{2} \tag{3-26}$$

$$E(N_2) = \frac{n}{2} P(U > 0,\ V > 0) = \frac{n}{2}\,[1-2I(0,0,1)+I(0,0,2)] \tag{3-27}$$

$$E(N_3) = \frac{n}{2} P(U>0,\ V<0 \ \text{ or } \ U<0,\ V>0) = n[I(0,0,1)-I(0,0,2)] \tag{3-28}$$

where $I(i,j,k)$ is as defined in Appendix A. Similarly, $T(\nu;i,j,k)$, also as defined in Appendix A, should be substituted by the corresponding expected values. These expected values are not always available in closed form. However, the most frequent occurrence of $(T(\nu;i,j,k)$ is for $i=k=0$; and for these, explicit solutions can be derived. The following five results are helpful:

$$\int_\eta^\infty \int_\eta^\infty x\phi(x,y;\rho)\ dxdy = \phi(\eta)\Phi(\delta)(1+\rho) \tag{3-29}$$

$$\int_\eta^\infty \int_\eta^\infty x^2\phi(x,y;\rho)\ dxdy = [1-2I(0,0,1)+I(0,0,2)] + \eta\phi(\eta)\Phi(\delta)(1+\rho^2) + \rho\sqrt{1-\rho^2}\ \phi(\eta)\phi(\delta) \tag{3-30}$$

$$\int_\eta^\infty \int_\eta^\infty xy\phi(x,y;\rho)\ dxdy = \rho[1-2I(0,0,1)+I(0,0,2)] + 2\rho\eta\phi(\eta)\Phi(\delta) + \sqrt{1-\rho^2}\ \phi(\eta)\phi(\delta) \tag{3-31}$$

$$\int_{-\infty}^\eta \int_\eta^\infty x\phi(x,y;\rho)\ dxdy = \phi(\eta)[1-\Phi(\delta)(1+\rho)] \tag{3-32}$$

$$\int_{-\infty}^\eta \int_\eta^\infty x^2\phi(x,y;\rho)\ dxdy = I(0,0,1) - I(0,0,2) + \eta\phi(\eta)[1-\Phi(\delta)(1+\rho^2)] - \rho\sqrt{1-\rho^2}\ \phi(\eta)\phi(\delta) \tag{3-33}$$

where

$$\phi(x,y;\rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\{\frac{-1}{2(1-\rho^2)}(x^2-2\rho xy+y^2)\} \tag{3-34}$$

and

$$\eta = \frac{-\mu}{\sigma},\ \delta = \frac{\mu}{\sigma}\sqrt{\frac{1-\rho}{1+\rho}} \tag{3-35}$$

Equations (3-29) to (3-31) are given, but in a modified form by Rosenbaum (1961). Equations (3-32) and (3-33) can be evaluated in a straightforward but tedious way. These expressions could be also found by using the moment generating function of the truncated multinormal distribution, as given by Tallis (1961). Then,

$$E[T(x,0,1,0)] = E[T(y,0,1,0)] = E(N_2E(X^\alpha)) = \frac{n}{2} P(U>0,V>0)E(X^\alpha) \tag{3-36}$$

Recall that U and V follow the bivariate normal distribution, with equal marginal distribution $N(\mu,\sigma^2)$ and the correlation coefficient ρ. Therefore,

$$E(X^\alpha) = E(U|U>0,V>0) = [P(U>0,V>0)]^{-1} \int_\eta^\infty \int_\eta^\infty (x\sigma+\mu)\phi(x,y;\rho)\ dxdy \tag{3-37}$$

From Eqs. (3-27), (3-29) and (3-37)

$$E(X^\alpha) = \mu + \frac{\sigma\phi(\eta)\Phi(\delta)(1+\rho)}{[1-2I(0,0,1)+I(0,0,2)]} \tag{3-38}$$

From Eqs. (3-36) and (3-38)

$$E[T(x,0,1,0)] = E[T(y,0,1,0)] = \frac{n}{2}\,[\sigma\phi(\eta)\Phi(\delta)(1+\rho)+\mu[1-2I(0,0,1)+I(0,0,2)]] \tag{3-39}$$

Similarly,

$$E[T(x,0,2,0)] = E[T(y,0,2,0)] = \frac{n}{2} P(U>0,V>0)E(X^{2\alpha}) \tag{3-40}$$

and

$$E(X^{2\alpha}) = E(U^2|U>0,V>0)$$

$$= [P(U>0,V>0)]^{-1} \int_{\eta}^{\infty}\int_{\eta}^{\infty} (x\sigma+\mu)\phi(x,y;\rho)\ dxdy \tag{3-41}$$

From Eqs. (3-27), (3-29), (3-30), and (3-41),

$$E(X^{2\alpha}) =$$

$$(\mu^2 + \sigma^2) + \frac{-\mu\sigma(\rho^2-2p-1)\phi(\eta)\Phi(\delta)+\sigma^2\rho\sqrt{1-\rho^2}\phi(\eta)\phi(\delta)}{[1-2I(0,0,1)+I(0,0,2)]} \tag{3-42}$$

From Eqs. (3-40) and (3-42),

$$E[T(x,0,2,0)] = E[T(y,0,2,0)] =$$

$$\frac{n}{2}\ [(\mu^2+\sigma^2)[1-2I(0,0,1)+I(0,0,2)]$$

$$- \mu\sigma(\rho^2-2\rho+1)\phi(\eta)\Phi(\delta) + \sigma^2\rho\sqrt{1-\rho^2}\ \phi(\eta)\phi(\delta)] \tag{3-43}$$

Similarly,

$$E[T(xy,0,1,0)] = \frac{n}{2}\ P(U>0,V>0)\ \ E[(XY)^{\alpha}] \tag{3-44}$$

and

$$E[(XY)^{\alpha}] = E(UV|U>0,V>0) =$$

$$[P(U>0,V>0)]^{-1} \int_{\eta}^{\infty}\int_{\eta}^{\infty} (x\sigma+\mu)(y\sigma+\mu)\phi(x,y;\rho)\ dxdy \tag{3-45}$$

From Eqs. (3-27), (3-29), (3-31), and (3-45),

$$E[(XY)^{\alpha}] = \mu^2 + \sigma^2\rho + \frac{2\mu\sigma\phi(\eta)\Phi(\delta) + \sigma^2\sqrt{1-\rho^2}\phi(\eta)\phi(\delta)}{[1-2I(0,0,1)+I(0,0,2)]} \tag{3-46}$$

From Eqs. (3-44) and (3-46),

$$E[T(xy,0,1,0)] =$$

$$\frac{n}{2}\ [(\mu^2+\sigma^2\rho)[1-2I(0,0,1)+I(0,0,2)]+2\mu\sigma\phi(\eta)\Phi(\delta)$$

$$+ \sigma^2\ \sqrt{1-\rho^2}\ \phi(\eta)\phi(\delta)] \tag{3-47}$$

Similarly

$$E[T(z,0,1,0)] = nP(U>0,V<0 \text{ or } U<0,V>0)\ E[Z^{\alpha}] \tag{3-48}$$

and

$$E[Z^{\alpha}] = E(U|U>0,V<0) =$$

$$[P(U>0,V<0)]^{-1} \int_{-\infty}^{\eta}\int_{\eta}^{\infty} (x\sigma+\mu)\phi(x,y;\rho)\ dxdy \tag{3-49}$$

From Eqs. (3-28), (3-30), and (3-49),

$$E[Z^{\alpha}] = \mu + \frac{\sigma\phi(\eta)[1-\Phi(\delta)(1+\rho)]}{[I(0,0,1)-I(0,0,2)]} \tag{3-50}$$

From Eqs. (3-48) and (3-50),

$$E[T(z,0,1,0)] =$$

$$n[\mu[I(0,0,1)-I(0,0,2)]+\sigma\phi(\eta)[1-\Phi(\delta)(1+\rho)]] \tag{3-51}$$

Similarly

$$E[T(z,0,2,0)] = nP(U>0,V<0 \text{ or } U<0,V>0)E(Z^{\alpha}) \tag{3-52}$$

and

$$E(Z^{2\alpha}) = E(U^2|U>0,V<0) =$$

$$= [P(U>0,V<0)]^{-1} \int_{-\infty}^{\eta}\int_{\eta}^{\infty} (x\sigma+\mu)^2\phi(x,y;\rho)\ dxdy \tag{3-53}$$

From Eqs. (3-28), (3-32), (3-33), and (3-53),

$$E[Z^{2\alpha}] =$$

$$\mu^2+\sigma^2+ \frac{\mu\sigma\phi(\eta)[1+\Phi(\delta)(\rho^2-2\rho+1)]-\sigma^2\rho\sqrt{1-\rho^2}\phi(\eta)\phi(\delta)}{[I(0,0,1)-I(0,0,2)]} \tag{3-54}$$

From Eqs. (3-52) and (3-54),

$$E[T(z,0,2,0)] = n[(\mu^2+\sigma^2)[I(0,0,1)-I(0,0,2)]$$

$$+\mu\sigma\phi(\eta)[1+\Phi(\delta)(\rho^2-2\rho+1]-\sigma^2\rho\sqrt{1-\rho^2}\phi(\eta)\sigma(\delta)] \tag{3-55}$$

The values of $E[T(\nu;i,j,k)]$ for $i \geq 1$ and/or $k \geq 1$ can be found in an approximate way. If two random variables R and S follow a functional relationship, $R = g(S)$, then the Taylor series expansion may be used giving

$$R \cong g[E(S)]+[S-E(S)]g'[E(S)]+ \frac{[S-E(S)]^2}{2}\ g''[E(S)] \tag{3-56}$$

where the terms of the order higher than two were neglected.

Taking the expected value on both sides of Eq. (3-56) then,

$$E[R] \cong g[E(S)] + \frac{g''[E(S)]}{2}\ \mathrm{var}(S) \tag{3-57}$$

where $\mathrm{var}(\cdot)$ stands for the variance operator. Therefore,

$$E[T(\nu;i,j,k)] \cong \frac{nP^*}{2}\,[g[E(S)] + \frac{g''[E(S)]}{2}\,\mathrm{var}(S)] \tag{3-58}$$

where P^* and S are given in Table 3-1, according to the meaning of ν,

Table 3-1. Values of P^* and S of Eq. (3-58)

ν	P^*	S
x	$P(U>0,V>0) = [1-2I(0,0,1)+I(0,0,2)]$	X^α
y	$P(U>0,V>0) = [1-2I(0,0,1)+I(0,0,2)]$	Y^α
xy	$P(U>0,V>0) = [1-2I(0,0,1)+I(0,0,2)]$	$(XY)^\alpha$
z	$P(U>0,V<0 \text{ or } U<0,V>0) = 2[I(0,0,1)-I(0,0,2)]$	Z^α

and

$$g(x) = [h(x)]^i\, x^j (\frac{\log x}{\alpha})^k \tag{3-59}$$

where

$$\eta(x) = \frac{\phi[-\rho x-\mu(1-\rho)/(\sigma\sqrt{1-\rho^2})]}{\Phi[-\rho x-\mu(1-\rho)/(\sigma\sqrt{1-\rho^2})]} \tag{3-60}$$

From Eqs. (3-59) and (3-60)

$$g''(x) =$$

$$[h(x)]^i \{\frac{\rho i x^{j-1}}{\sigma\sqrt{1-\rho^2}} (\frac{\log x}{\alpha})^{k-1} [x(\frac{\log x}{\alpha}) \frac{\rho}{\sigma\sqrt{1-\rho^2}} [i(h(x)$$

$$- \frac{\rho x+\mu(1-\rho)}{\sigma\sqrt{1-\rho^2}})^2 - 1 + h(x)(h(x) - \frac{\rho x+\mu(1-\rho)}{\sigma\sqrt{1-\rho^2}})$$

$$+ 2(j(\frac{\log x}{\alpha}) + \frac{k}{\alpha})(h(x) - \frac{\rho x+\mu(1-\rho)}{\sigma\sqrt{1-\rho^2}})]$$

$$+ \frac{x^{j-2}(\log x)^{k-2}}{\alpha^k} (k(k-1) + \log x(jk+(j-1)(j\log x+k)))\} \tag{3-61}$$

Equations (3-59) and (3-61) should be substituted into Eq. (3-58). The expected values of S, for each case, are given by Eqs. (3-38), (3-42), (3-46), and (3-50), respectively.

From Eqs. (3-38) and (3-42) it follows

$$\mathrm{var}(X^\alpha) = \sigma^2 + \frac{\sigma\phi(\eta)}{[1-2I(0,0,1)+I(0,0,2)]} [\sigma\rho\sqrt{1-\rho^2}\,\phi(\delta) - \mu(1+\rho^2)\Phi(\delta)] - [\frac{\sigma(1+\rho)\phi(\eta)\Phi(\delta)}{[1-2I(0,0,1)+I(0,0,2)]}]^2 \tag{3-62}$$

From Eqs. (3-50) and (3-54) it follows

$$\mathrm{var}(Z^\alpha) = \sigma^2 + \frac{\sigma\phi(\eta)}{[I(0,0,1)-I(0,0,2)]} [\mu[\Phi(\delta)(1-\rho^2)-1] - \sigma\rho\sqrt{1-\rho^2}\,\phi(\delta)] - [\frac{\sigma\phi(\eta)[1-\Phi(\delta)(1+\rho)]}{I(0,0,1)-I(0,0,2)}]^2 \tag{3-63}$$

The derivation of $\mathrm{var}[XY^\alpha]$ would require the evaluation of the fourth-order moments. In order to avoid this complication, the Taylor expansion is applied once more, yielding:

$$\mathrm{var}[(XY)^\alpha] \cong 2[E(X^\alpha)]^2\{\mathrm{var}(X^\alpha)E[(XY)^\alpha]-[E(X^\alpha)]^2\} \tag{3-64}$$

It could be shown that for $i=k=0$, Eq. (3-58) is identical to Eqs. (3-39), (3-43), (3-47), (3-51), and (3-55), respectively. In other words, the Taylor expansion yields the exact results for these particular cases.

The approximations used to calculate the asymptotic covariance matrix might cast doubt upon the accuracy of results. Fortunately, the experience shows that the procedure is worthy, and this can be best expounded through the example to follow.

Example. 195 new samples were generated with $\mu=-0.25$, $\sigma=1.00$, $\rho=0.40$, and $\alpha=0.60$. These parameters are fairly typical for the precipitation process. Algorithm of Eq. (3-23) was applied to each generated time series, resulting in a sample of 195 observations for the estimator vector (random vector). To better understand the role of the sample length, the time series were grouped into three classes respectively of lengths 500, 1000 and 2000. Therefore, there are three sets, each of 65 observations of a random vector, as shown in Tables 3-2, 3-3, and 3-4. For reasons that will become clear in the following text, each set was further divided into three subsets, each with 15, 20, and 30 samples, respectively. The information on each set is condensed in Tables 3-5, 3-6, and 3-7. The results are:

(a) First row gives the population parameters, which were used for the generation. Second row gives the means of the 65 estimates. The comparisons between the means of estimates and the population values suggest that the estimation procedure is unbiased.

(b) The use of asymptotic expressions for the estimation of the covariance matrix is subject to the following two sources of errors: *The sample size might not be large enough; Numerical approximations are used to evaluate the expressions.*

Third row gives the asymptotic standard deviation (of the estimators) evaluated at the *correct point*, i.e., calculated at the population parameter vector.

Table 3-2. Estimates of the Parameters for the 65 Samples of Length M = 500

#	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\rho}$	$\hat{\alpha}$
1	-0.2302	0.8749	0.2881	0.6437
2	-0.1719	0.9607	0.3413	0.5211
3	-0.2759	0.9735	0.3298	0.6023
4	-0.2266	0.9783	0.4109	0.5758
5	-0.2441	0.9578	0.3374	0.5741
6	-0.2357	1.0118	0.3404	0.6303
7	-0.3444	1.0565	0.4300	0.6102
8	-0.1593	0.9306	0.3985	0.6236
9	-0.2909	1.0311	0.3335	0.6669
10	-0.2600	0.9809	0.4298	0.6088
11	-0.3438	0.9525	0.3120	0.6514
12	-0.3425	1.0335	0.3410	0.6083
13	-0.4058	1.1929	0.3858	0.6038
14	-0.2514	0.9290	0.3741	0.6208
15	-0.2304	0.9668	0.3217	0.6537
16	-0.2662	1.0195	0.3852	0.5727
17	-0.1820	1.0791	0.5279	0.6051
18	-0.4205	1.0374	0.4347	0.6322
19	-0.1237	0.9439	0.3788	0.5672
20	-0.3160	0.9303	0.3381	0.5957
21	-0.1948	1.0011	0.4531	0.6093
22	-0.2491	0.9729	0.5949	0.5970
23	-0.2853	1.0883	0.5568	0.6170
24	-0.1895	0.9308	0.3920	0.5887
25	-0.1111	0.9387	0.3764	0.5760
26	-0.2329	1.0209	0.3050	0.6132
27	-0.3029	1.0547	0.3178	0.6221
28	-0.3450	1.1022	0.3755	0.6461
29	-0.2322	0.9766	0.3874	0.5795
30	-0.2310	0.9573	0.3878	0.5916
31	-0.2242	0.9876	0.4261	0.6144
32	-0.3805	1.1171	0.4342	0.6190
33	-0.2351	1.0069	0.4804	0.5928
34	-0.1441	0.9139	0.4034	0.6168
35	-0.4101	1.1717	0.4285	0.6257
36	-0.2405	0.9212	0.2813	0.5583
37	-0.1584	0.9177	0.4213	0.6138
38	-0.2564	0.9541	0.3116	0.5757
39	-0.2018	0.9411	0.3861	0.5591
40	-0.2105	1.0194	0.4379	0.6378
41	-0.3412	0.9594	0.3690	0.5881
42	-0.2012	0.9446	0.4141	0.5793
43	-0.2944	1.0262	0.2329	0.5788
44	-0.2138	0.9794	0.3250	0.5241
45	-0.3821	1.0580	0.3818	0.6248
46	-0.2494	0.9404	0.4270	0.6041
47	-0.2662	0.9807	0.3807	0.6407
48	-0.2575	1.0168	0.5128	0.5920
49	-0.3500	1.0880	0.2582	0.6436
50	-0.1693	0.8910	0.3020	0.5516
51	-0.3085	0.9383	0.4135	0.5734
52	-0.4103	1.0576	0.3354	0.6157
53	-0.2909	0.9464	0.3319	0.6179
54	-0.1803	0.9750	0.5094	0.6455
55	-0.2178	0.9587	0.3899	0.5701
56	-0.0908	0.8829	0.3727	0.5309
57	-0.1972	0.9003	0.3193	0.5973
58	-0.1922	0.9755	0.4238	0.5938
59	-0.2583	1.0037	0.4282	0.6046
60	-0.3284	0.9580	0.3296	0.6193
61	-0.2885	1.0799	0.4560	0.6363
62	-0.2973	1.0078	0.4916	0.5830
63	-0.2259	1.0403	0.3246	0.5860
64	-0.2813	1.1096	0.5248	0.6191
65	-0.1627	0.8997	0.3119	0.5936

Table 3-3. Estimates of the Parameters for the 65 Samples of Length m = 1000

#	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\rho}$	$\hat{\alpha}$
1	-0.1994	0.9185	0.3552	0.5828
2	-0.2290	0.9491	0.3512	0.5686
3	-0.2846	0.9992	0.4130	0.6186
4	-0.2457	0.9823	0.3237	0.5785
5	-0.2971	1.0184	0.3549	0.5660
6	-0.2576	0.9614	0.4050	0.6269
7	-0.3034	1.0523	0.3906	0.6162
8	-0.2369	0.9145	0.3594	0.5614
9	-0.3435	0.9969	0.3335	0.6153
10	-0.1983	0.9653	0.4536	0.6042
11	-0.1428	0.8921	0.3497	0.5605
12	-0.2245	0.9891	0.4263	0.5990
13	-0.3120	1.0229	0.4037	0.6253
14	-0.2623	1.0252	0.4053	0.5830
15	-0.2125	0.9956	0.4266	0.6035
16	-0.2195	0.9338	0.3756	0.5598
17	-0.1956	1.0994	0.4631	0.6510
18	-0.2750	1.0702	0.5016	0.5978
19	-0.2338	1.0453	0.3580	0.6148

#	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\rho}$	$\hat{\alpha}$
20	-0.2007	0.9410	0.3514	0.5997
21	-0.3178	0.9596	0.2694	0.5876
22	-0.2514	1.0134	0.4406	0.5673
23	-0.3110	0.9482	0.2869	0.5697
24	-0.2403	0.9890	0.4038	0.5845
25	-0.1969	0.9887	0.4778	0.6472
26	-0.2926	0.9636	0.3815	0.5546
27	-0.3147	1.0596	0.3925	0.6311
28	-0.1884	1.0128	0.4133	0.5869
29	-0.1940	1.0047	0.4366	0.5835
30	-0.3312	1.0908	0.3697	0.6406
31	-0.1910	1.0012	0.4329	0.6374
32	-0.2309	1.0266	0.3990	0.6116
33	-0.2674	1.0467	0.3723	0.6311
34	-0.1694	0.9800	0.3815	0.5968
35	-0.3590	1.0002	0.3126	0.5588
36	-0.3116	0.9845	0.4008	0.5877
37	-0.2714	0.9508	0.3604	0.6167
38	-0.1471	0.9631	0.3883	0.5706
39	-0.2384	0.9633	0.3280	0.5771
40	-0.2770	1.0877	0.4409	0.6199
41	-0.1706	0.9654	0.4326	0.5611
42	-0.2562	0.9670	0.3177	0.5880
43	-0.2701	0.9774	0.3751	0.6234
44	-0.2636	0.9632	0.3867	0.6118
45	-0.2254	1.0500	0.4608	0.5871
46	-0.2636	0.9924	0.4117	0.5950
47	-0.2572	0.9659	0.4026	0.6007
48	-0.2668	1.0293	0.5745	0.6036
49	-0.1504	0.9349	0.3846	0.5813
50	-0.2671	1.0372	0.3113	0.6174
51	-0.2879	1.0369	0.3842	0.6084
52	-0.2275	0.9722	0.4073	0.6028
53	-0.3026	1.0553	0.4558	0.6043
54	-0.2565	1.0262	0.4158	0.6098
55	-0.2195	0.9338	0.3756	0.5598
56	-0.1956	1.0994	0.4631	0.6510
57	-0.2750	1.0702	0.5016	0.5978
58	-0.2338	1.0453	0.3580	0.6148
59	-0.2007	0.9410	0.3514	0.5997
60	-0.3178	0.9596	0.2694	0.5876
61	-0.2514	1.0134	0.4406	0.5673
62	-0.3110	0.9482	0.2869	0.5697
63	-0.2403	0.9890	0.4038	0.5845
64	-0.1969	0.9887	0.4778	0.6472
65	-0.2926	0.9636	0.3815	0.5546

Table 3-4. Estimates of the Parameters for the 65 Samples of Length m = 2000

#	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\rho}$	$\hat{\alpha}$
1	-0.2798	0.9871	0.4132	0.5893
2	-0.2807	0.9679	0.3903	0.6135
3	-0.2860	1.0601	0.3994	0.6358
4	-0.1781	0.9373	0.3878	0.5575
5	-0.2397	0.9399	0.4120	0.6144
6	-0.2404	0.9813	0.4078	0.5737
7	-0.2539	1.0350	0.4375	0.6001
8	-0.2398	1.0024	0.3973	0.5833
9	-0.2632	0.9669	0.3378	0.5941
10	-0.2710	0.9877	0.3955	0.6059
11	-0.1973	1.0585	0.4012	0.6031
12	-0.2250	1.0115	0.4223	0.6428
13	-0.2507	0.9813	0.4156	0.6042
14	-0.2446	1.0393	0.3431	0.6314
15	-0.2098	1.0282	0.3780	0.6211
16	-0.2665	1.0363	0.4329	0.6058
17	-0.1800	0.9774	0.3818	0.5918
18	-0.2763	0.9697	0.3484	0.5639
19	-0.2519	0.9824	0.3906	0.5787
20	-0.2877	1.0716	0.4234	0.6232
21	-0.2363	0.9893	0.3806	0.6131
22	-0.3075	1.0444	0.4738	0.6028
23	-0.2800	1.0344	0.4545	0.6187
24	-0.2640	1.0405	0.4078	0.6241
25	-0.2174	1.0239	0.4144	0.6123
26	-0.2047	1.0120	0.4448	0.6042
27	-0.1648	0.9113	0.3783	0.5762
28	-0.2912	1.0458	0.4054	0.6323
29	-0.2597	0.9838	0.4487	0.5999
30	-0.2359	1.0320	0.4682	0.5891
31	-0.2078	0.9792	0.3219	0.6075
32	-0.2907	1.0167	0.3683	0.5986
33	-0.3311	1.0178	0.4242	0.6125
34	-0.2092	1.0219	0.4035	0.5910
35	-0.2314	0.9450	0.3638	0.5823
36	-0.2714	0.9428	0.4588	0.5819
37	-0.2755	1.0344	0.4193	0.6419
38	-0.2091	1.0273	0.4256	0.5960
39	-0.2407	0.9791	0.3961	0.6022
40	-0.1990	0.9139	0.4154	0.6080
41	-0.2118	0.9867	0.3389	0.6028
42	-0.2866	1.0574	0.4426	0.5928
43	-0.2985	1.0535	0.4070	0.5908

#	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\rho}$	$\hat{\alpha}$
44	-0.2370	0.9866	0.4479	0.6101
45	-0.2330	1.0120	0.4052	0.5796
46	-0.2351	0.9692	0.4181	0.6064
47	-0.1640	1.0001	0.4372	0.6074
48	-0.2779	1.0216	0.4321	0.6167
49	-0.2849	1.0090	0.3936	0.6190
50	-0.2550	0.9897	0.3904	0.5874
51	-0.3345	0.9977	0.4233	0.5998
52	-0.3041	0.9794	0.3861	0.5805
53	-0.2555	0.9906	0.4416	0.5882
54	-0.2975	0.9478	0.3842	0.5873
55	-0.2869	1.0068	0.4211	0.5938
56	-0.2071	0.9699	0.4367	0.5786
57	-0.2764	1.0136	0.4004	0.5798
58	-0.3019	1.0016	0.3794	0.5828
59	-0.2298	1.0352	0.4219	0.6026
60	-0.2424	1.0124	0.4400	0.5983
61	-0.2815	1.0137	0.3987	0.6075
62	-0.2284	0.9914	0.4049	0.5914
63	-0.2650	0.9768	0.4321	0.5911
64	-0.2442	1.0567	0.4508	0.6426
65	-0.2935	1.0390	0.4878	0.6673

Table 3-5. Characteristics of Sample Estimates with Sample Length m = 500

Row	n^*		θ: μ	σ	ρ	α
1		θ_p	-0.250	1.000	0.400	0.600
2		$\bar{\theta}^*$	-0.256	0.992	0.387	0.602
3		$\sigma_\infty(\theta_p)$	0.071	0.063	0.073	0.040
4		σ_θ^*	0.074	0.066	0.069	0.030
5	15	$\bar{\theta}$	-0.268	0.989	0.358	0.613
6	15	θ_{15}	-0.230	0.967	0.322	0.654
7	15	σ_θ	0.068	0.073	0.044	0.037
8	15	$\sigma_\infty(\theta_{15})$	0.066	0.058	0.077	0.043
9	20	$\bar{\theta}$	-0.254	1.013	0.419	0.604
10	20	θ_{35}	-0.410	1.172	0.429	0.626
11	20	σ_θ	0.086	0.071	0.076	0.021
12	20	$\sigma_\infty(\theta_{35})$	0.076	0.084	0.066	0.054
13	30	$\bar{\theta}$	-0.251	0.979	0.380	0.595
14	30	θ_{65}	-0.163	0.900	0.312	0.594
15	30	σ_θ	0.070	0.060	0.077	0.032
16	30	$\sigma_\infty(\theta_{65})$	0.079	0.068	0.092	0.039

n^* = the number of generated samples

θ = the dummy parameter (may represent any one of μ, σ, ρ, or α)

$\bar{\theta}$ = $(\sum^{n} \theta_i)/n^*$ (θ_i is the estimate θ for the ith sample)

$\bar{\theta}^*$ = $\bar{\theta}$ for the lumped set of 65 observations

θ_p = the population parameter

σ_θ = $\sqrt{[\sum^{n}(\theta_i - \theta_p)^2]/n^*}$ (the sample standard deviation)

σ_θ^* = σ_θ for the lumped set of 65 samples

$\sigma_\infty(\theta)$ = the asymptotic standard deviation, evaluated at θ.

Table 3-6. Characteristics of Sample Estimates with Sample Length m = 1000

Row	n^*		θ: μ	σ	ρ	α
1		θ_p	-0.250	1.000	0.400	0.600
2		$\bar{\theta}^*$	-0.249	0.996	0.392	0.598
3		$\sigma_\infty(\theta_p)$	0.050	0.045	0.051	0.028
4		σ_θ^*	0.049	0.047	0.056	0.026
5	15	$\bar{\theta}$	-0.250	0.979	0.383	0.594
6	15	θ_{15}	-0.213	0.996	0.427	0.604
7	15	σ_θ	0.051	0.049	0.041	0.024
8	15	$\sigma_\infty(\theta_{15})$	0.049	0.045	0.049	0.028
9	20	$\bar{\theta}$	-0.249	1.009	0.391	0.601
10	20	θ_{35}	-0.359	1.000	0.313	0.559
11	20	σ_θ	0.055	0.047	0.059	0.030
12	20	$\sigma_\infty(\theta_{35})$	0.052	0.043	0.059	0.028
13	30	$\bar{\theta}$	-0.248	0.996	0.398	0.597
14	30	θ_{65}	-0.293	0.964	0.382	0.555
15	30	σ_θ	0.044	0.046	0.063	0.024
16	30	$\sigma_\infty(\theta_{65})$	0.049	0.041	0.053	0.026

n^* = the number of generated samples

θ = the dummy parameter (may represent any one of μ, σ, ρ, or α)

$\bar{\theta}$ = $(\sum^{n} \theta_i)/n^*$ (θ_i is the estimate θ for the ith sample)

$\bar{\theta}^*$ = $\bar{\theta}$ for the lumped set of 65 observations

θ_p = the population parameter

σ_θ = $\sqrt{[\sum^{n}(\theta_i - \theta_p)^2]/n^*}$ (the sample standard deviation)

σ_θ^* = σ_θ for the lumped set of 65 samples

$\sigma_\infty(\theta)$ = the asymptotic standard deviation, evaluated at θ.

Table 3-7. Characteristics of Sample Estimates with Sample Length m = 2000

Row	n^*		θ: μ	σ	ρ	α
1		θ_p	-0.250	1.000	0.400	0.600
2		$\bar{\theta}^*$	-0.252	1.002	0.410	0.602
3		$\sigma_\infty(\theta_p)$	0.035	0.031	0.036	0.020
4		σ_θ^*	0.038	0.036	0.032	0.020
5	15	$\bar{\theta}$	-0.244	0.999	0.396	0.605
6	15	θ_{15}	-0.210	1.028	0.378	0.621
7	15	σ_θ	0.031	0.038	0.026	0.023
8	15	$\sigma_\infty(\theta_{15})$	0.037	0.035	0.037	0.020
9	20	$\bar{\theta}$	-0.250	1.007	0.407	0.601
10	20	θ_{35}	-0.231	0.945	0.364	0.582
11	20	σ_θ	0.044	0.039	0.041	0.018
12	20	$\sigma_\infty(\theta_{35})$	0.036	0.031	0.040	0.019

13	30	$\overline{\theta}$	-0.258	1.001	0.418	0.601
14		θ_{65}	-0.293	1.039	0.488	0.667
15		σ_θ	0.038	0.034	0.029	0.020
16		$\sigma_\infty(\theta_{65})$	0.038	0.035	0.034	0.022

n^* = the number of generated samples

θ = the dummy parameter (may represent any one of μ, σ, ρ, or α)

$\overline{\theta}$ = $(\sum^n \theta_i)/n^*$ (θ_i is the estimate θ for the ith sample)

$\overline{\theta}^*$ = $\overline{\theta}$ for the lumped set of 65 observations

θ_p = the population parameter

σ_θ = $\sqrt{[\sum^n(\theta_i - \theta_p)^2]/n^*}$ (the sample standard deviation)

σ_θ^* = σ_θ for the lumped set of 65 samples

$\sigma_\infty(\theta)$ = the asymptotic standard deviation, evaluated at θ.

Fourth row gives the sample standard deviations calculated from the 65 outcomes. In order not to increase the complexity of the analysis the covariances between the estimators have not been investigated. Comparing the third and fourth rows, the approximation between corresponding values seems good.

(c) It must be emphasized that the standard deviations of the estimators do not appear in the fourth row, but rather the corresponding estimates obtained from a sample of 65 items. Therefore, there is a new source of error, as far as this comparison is concerned, namely: *The number of generated samples may not be sufficient to produce accurate estimates of standard deviations of the estimators.*

To clarify this issue each set was subdivided into three subsets, with unequal number of samples in each (15, 20, 30). The means and sample standard deviation were then calculated. These values are shown in rows 5, 9, 13, and 7, 11, 15, respectively. The second row is the result of a weighted average of the corresponding values of rows 5, 9, 13. The same is true of the fourth row, with regard to rows 7, 11, 15.

(d) An important question to be addressed is whether the asymptotic expressions can be used reliably in any real case. Usually only one historic record is available, which yields the estimates of the parameters. Not knowing the population values, the logical thing to do is to evaluate the asymptotic expressions at these estimated points. Therefore, a fourth source of error is: *The uncertainty about the parameter values at which the asymptotic expressions are evaluated.*

This question was investigated assuming that only the 15th, 35th, and 65th samples of each of the three sets were available. The corresponding estimates are found in rows 6, 10, 14 (and also in Tables 3-2, 3-3, and 3-4). The asymptotic standard deviations evaluated at these points are given in rows 7, 11, and 15. The feasibility of the proposed procedure may be evaluated by comparing rows 7, 11, and 15 to the fourth row. The evaluation is approximate because only nine cases were examined.

(e) Though the accuracy of results increases with an increase of the sample length, it seems that the procedure may be applied to samples as short as 500. Likely even shorter samples will yield satisfactory results.

(f) For the sample lengths of n_1 and n_2, the relationship

$$\frac{\sigma_\infty(\theta,n_1)}{\sigma_\infty(\theta,n_2)} = \sqrt{\frac{n_2}{n_1}} \tag{3-65}$$

holds, where $\sigma_\infty(\theta,n)$ is the asymptotic standard deviation evaluated at θ, for a time series of length n.

Thus, for instance

$$\frac{\sigma_\infty(\theta_p,500)}{\sigma_\infty(\theta_p,2000)} = \sqrt{\frac{2000}{500}} = 2.00$$

and

$$\frac{\sigma_\infty(\theta_p,1000)}{\sigma_\infty(\theta_p,2000)} = \frac{\sigma_\infty(\theta_p,500)}{\sigma_\infty(\theta_p,1000)} = \sqrt{2} = 1.41$$

which can be easily verified by comparing the third rows of Tables 3-5, 3-6, and 3-7.

3-5 Tests to be Performed

The material so far presented had implicitly assumed a number of hypotheses, for example that there is time persistence in the data and that the process is stationary. Actually, this last condition will be relaxed in Chapter IV. In what follows, tests for the two above mentioned hypotheses are developed.

Test of Serial Independence. One might wonder whether the model assumed for the continuous process, namely the first-order-Markov, may be excessively sophisticated for the problem at hand. This can be put in another way, whether it is possible that the continuous process is in fact serially independent, therefore with ρ=0. If this is the case, any positive value estimated for ρ would be due to sample fluctuations. Hence, a test of the null hypothesis that ρ=0 may be appropriate.

Let $\overline{\underline{\Theta}}$ be the four-dimensional parameter space, namely $\overline{\underline{\Theta}} = \{(\mu,\sigma,\rho,\alpha);\ -\infty < \mu < \infty,\ 0 < \sigma,\ 0 \le \rho \le 1,\ -\infty < \alpha < \infty\}$. Let us define the three-dimensional parameter subspace by $\overline{\underline{\Theta}}_o = \{(\mu,\sigma,\rho,\alpha);\ -\infty < \mu < \infty,\ 0 < \sigma,\ \rho = 0,\ -\infty < \alpha < \infty\}$ We want to test the null hypothesis H_o: $\theta = (\mu,\sigma,\rho,\alpha) \in \overline{\underline{\Theta}}_o$ versus the alternative hypothesis H_A: $\theta = (\mu,\sigma,\rho,\alpha) \in \overline{\underline{\Theta}} - \overline{\underline{\Theta}}_o$. The generalized likelihood-ratio, denoted by λ is defined to be

$$\lambda = \frac{\sup_{\theta \in \overline{\underline{\Theta}}_o} L}{\sup_{\theta \in \Theta} L} \tag{3-66}$$

with sup(·) meaning the supremum.

Notice that λ is a function only of the observations and therefore is a statistic. When the observations are replaced by their corresponding random variables then λ is itself a random variable. It is know, for example from Mood et al. (1974), that for large sample $-2 \log \lambda$ is approximately distributed as chi-square with one degree of freedom, for this particular case.

Recalling that LL = log L, we have, from Eq. (3-66)

$$2[\sup_{\theta\varepsilon\overline{\underline{\Theta}}} LL - \sup_{\theta\varepsilon\overline{\underline{\Theta}}_o} LL] \sim \chi^2(1) \qquad (3\text{-}67)$$

$\sup_{\theta\varepsilon\overline{\underline{\Theta}}} LL$ is the value of the log-likelihood function evaluated at the estimated vector. Let

$$LL^* = \sup_{\theta\varepsilon\overline{\underline{\Theta}}} LL \qquad (3\text{-}68)$$

Therefore, one should reject the null hypothesis, for the size of the test equal to γ, if

$$2(LL^* - \sup_{\theta\varepsilon\overline{\underline{\Theta}}_o} LL) > \chi^2_{1-\gamma}(1) \ , \qquad (3\text{-}69)$$

where γ is the probability that a wrong decision is reached, if the null hypothesis is rejected (Type I error).

Test of Stationarity. The stationarity assumption made for each season is perhaps the hypothesis that may raise most of the doubts. Actually, this is an intermediate step toward a more realistic representation of the usually non-stationarity processes of hydrology; each season yields a parameter vector and the fitting of periodic functions to this number of points, which equals the number of seasons, will give a non-stationary representation of the whole process. Hence, the stationarity assumption ought to be seen as an approximation used in an estimation procedure designed for a non-stationary model. Nevertheless, one should expect the period of the year to be divided into seasons in such a way that the non-stationarity inside each season is kept at a *low level*. A test for the stationarity of each season is, therefore, in order.

A way of testing for the stationarity is again through the use of the generalized likelihood ratio. Let the season be split into two subseasons, A and B. If the process was indeed stationary, one could expect the estimates for the subseason A to be *close* to those for the subseason B. Different season splitting criteria represent different alternative hypotheses. From the several ways of splitting a season, the following two schematic representations may be most convenient:

Alternative I

A is the first half and B is the second half of the season. See Figure 3-3(a).

Alternative II

A is made of the first and last quarters, and B of the second and third quarters of the season. See Figure 3-3(b).

It is easily verified that Alternative I would work satisfactorily whenever a parameter Θ varies with time in the way shown in Figure 3-3(c). However it would not be appropriate for a situation like the one displayed in Figure 3-3(d); despite the fact that Θ is not constant with time, still $\hat{\Theta}_A$ and $\hat{\Theta}_B$ might turn out to be statiscally equal. By a similar reasoning it can be shown that Alternative II is appropriate for situations like the one displayed in Figure 3-3(d) and not appropriate for situations exemplified by Figure 3-3(c)

For safe tests, both alternatives should be used. The parameter space can be reshaped in the following way:

$\Theta_1 = \mu_A - \mu_B, \quad \Theta_2 = \sigma_A - \sigma_B, \quad \Theta_3 = \rho_A - \rho_B, \quad \Theta_4 = \alpha_A - \alpha_B,$

$\Theta_5 = \mu_A \quad , \quad \Theta_6 = \sigma_A \quad , \quad \Theta_7 = \rho_A \quad , \quad \Theta_8 = \alpha_A \quad .$

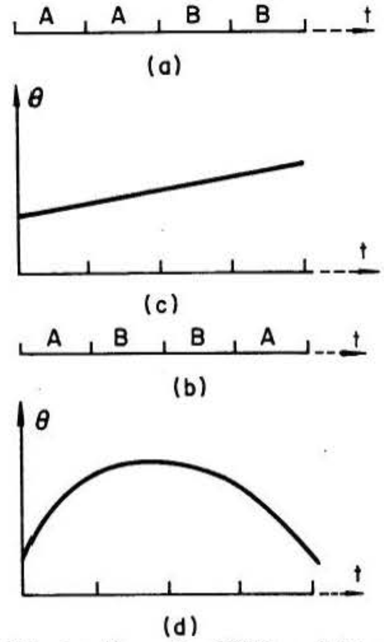


Fig. 3-3. Illustration of splitting Criteria for Testing for the Stationarity.

Let $\overline{\underline{\Theta}}$ be the eight-dimensional parameter space $\overline{\underline{\Theta}} = \{(\Theta_i), -\infty < \Theta_i < \infty$ for $i=1,2,3,4,5,8;\ 0 < \Theta_6;\ 0 < \Theta_7 < 1\}$. Define the four-dimensional parameter space $\overline{\underline{\Theta}}_o = \{(\Theta_i);\ \Theta_i = 0$ for $i=1,2,3,4,\ -\infty < \Theta_5,\ \Theta_8 < \infty;\ 0 < \Theta_6;\ 0 < \Theta_7 < 1\}$. The null hypothesis is then H_o: $\Theta\varepsilon\overline{\underline{\Theta}}_o$, versus the alternative H_A: $\Theta\varepsilon\overline{\underline{\Theta}}_o$.

The generalized likelihood ratio is then given by Eq. (3-66). From the definition of $\overline{\underline{\Theta}}_o$, $\sup_{\theta\varepsilon\overline{\underline{\Theta}}_o} LL = LL^*$, where LL^* is defined by Eq. (3-68). For large sample $-2 \log \lambda$ is distributed as chi-square, this time with four degrees of freedom. Therefore, one should reject the hypothesis, for a test of size γ, if

$$2(\sup_{\theta\varepsilon\overline{\underline{\Theta}}} LL - LL^*) > \chi^2_{1-\gamma}(4) \qquad (3\text{-}70)$$

3-6 Generation of New Samples

Once the parameters are estimated, the generation of new samples can be accomplished by following the stepwise procedure illustrated in Figure 1-1.

The standard normal noise, ξ_t, can be obtained by using several *canned* computer subroutines. However, for the multivariate case some care must be paid in generating $\xi_{t,j}$; $j = 1,2,\ldots,\ell$ because the variables may not be independent, with j as the station subscript. A way of doing this is by the use of:

$$\underline{\xi}_t = \pi \underline{\eta}_t \ , \tag{3-71}$$

in which π is a $\ell x \ell$ matrix and $\underline{\eta}_t$ is a $\ell x1$ vector of independent standard normal deviations. Then

$$\text{cov}(\underline{\xi}_t) = \text{cov}(\pi \underline{\eta}_t) = \pi \ \text{cov}(\underline{\eta}_t) \ \pi' \tag{3-72}$$

where cov(·) means the covariance matrix of the argument vector. Because

$$\text{cov}(\eta_t) = I_\ell \ , \tag{3-73}$$

where I_ℓ is the $\ell x \ell$ identity matrix, then from Eqs. (3-72) and (3-73)

$$\text{cov}(\underline{\xi}_t) = \pi\pi' \tag{3-74}$$

On the other hand, the linear autoregressive equations for stations j and k are:

$$Z_{t,j} = \mu(j) + \rho(j)(Z_{t-1,j} - \mu(j)) + \sigma(j) \sqrt{1-\rho^2(j)} \ \xi_{t,j} \tag{3-75}$$

and

$$Z_{t,k} = \mu(k) + \rho(k)(Z_{t-1,k} - \mu(k)) + \sigma(k) \sqrt{1-\rho^2(k)} \ \xi_{t,k} \tag{3-76}$$

Multiplying Eqs. (3-75) and (3-76) and finding the expected values, then

$$\text{cov}(\xi_{t,j};\xi_{t,k}) = \frac{\rho(j,k)(1-\rho(j)\rho(k))}{\sqrt{1-\rho^2(j)(1-\rho^2(k))}} \tag{3-77}$$

where $\rho(j)$ and $\rho(k)$ are the serial correlation coefficients respectively for stations j and k, and $\rho(j,k)$ is the lag-zero cross correlation between the two station series. From Eqs. (3-74) and (3-77) one may conclude that the (j,k)-element of matrix $\pi\pi'$, $j \neq k$, is given by Eq. (3-77). The diagonal elements are unities. Several methods are available for finding the matrix π when $\pi\pi'$ is known; Young (1968) gives a straightforward one.

Chapter IV

.APPLICATION OF INTERMITTENT PROCESSES MODEL TO PRECIPITATION DAILY SERIES

In this chapter the model previously presented is applied and tested to daily precipitation data. The approach undertaken is to show, by a few examples, that the model has sufficient merit to find place among the techniques already used in hydrologic practice. No attempt is made to test the method on a large number of station series. The emphasis is on demonstrating the reliability of the methodology, rather than an exhaustive examination of the stochastic characteristics of hundreds of precipitation series.

4-1 Data Selection

Choosing a particular precipitation record to be one of the cases studied in this chapter has been conditioned by the two requirements:

(i) The climatological description of the station location should be easily available; and

(ii) The stations should be sufficiently apart to possess different climatological conditions. However, a few stations should be sufficiently close in order to display some dependence, in this way serving as an illustration for the multivariate case for which the model is also applicable.

The first requirement was satisfied by imposing that a station would only qualify if it had been selected to receive a detailed description in WIC (1974). This publication gives a summary of climatological data of a large number of precipitation stations in USA, furnished on a state by state basis. Among those, only a few are chosen to receive a complete description. The stations herein selected for study belong to this second category. They are given in Table 4-1.

Table 4-1. List of Stations Used for the Study

Station	Period of Record	Location LATIT.	Location LONG.	Elevation (ft.)	Average Precip. (in.)	Annual Days w/ Precip.
Columbia (MO)	1951-1968	38°58'	92°22'	778	33.66	107
Kansas City (MO)	1946-1968	39°07'	94°36'	742	33.04	98
Springfield (MO)	1946-1968	37°14'	93°23'	1268	38.46	106
Raleigh-Durham (NC)	1951-1971	35°52'	78°47'	434	41.35	113
Austin (TX)	1898-1967	30°18'	97°42'	597	33.02	81
Rapid City (SD)	1951-1968	44°02'	103°03'	3165	16.39	93
Flagstaff (AZ)	1953-1970	35°08'	111°40'	6993	19.82	72
Seattle-Tacoma (WA)	1950-1970	47°27'	122°18'	386	39.95	164

The periods of record given in Table 4-1 were selected on the basis of the availability of data. They do not necessarily coincide with the periods in the WIC (1974) publication. Figure 4-1, with the locations of eight stations shows that the second requirement is also satisfied, namely that the stations are scattered throughout USA, with the exception of the three stations located in the State of Missouri, used to illustrate the multivariate case.

To avoid an excessive number of tables and graphs throughout this chapter, the detailed results pertinent to the station of Columbia (MO) will be only given in the text. The results related to the other stations will be referred to and introduced in a summarized way. However, Appendices B1 to B7, give the detailed outputs corresponding respectively to each one of the remaining stations. They are presented in the same order as the one employed in the text for the station of Columbia (MO).

Figure 4-1. Location of Stations of Daily Precipitation Series Used in this Study.

4-2 Application of the Model for the Stationary Case

A possible application of the model may be in generating the new samples related to a specific short interval of time during the year, say a particular month. For this case one is tempted to assume the stationarity in the data, therefore enabling the use of the model in the form developed in Chapter III. To study the applicability of the model for this case the data of each station series is divided in twelve seasons. The tests developed in Chapter III are then applied to each season of each station series yielding a detailed picture of how the data conforms to the hypotheses of the model. The seasons have alternating lengths of 32 and 28 days, adding up to a total of 360 days. The selection of these lengths stems from the fact that the stationarity tests developed in Section 3.5 require the number of daily observations to be the multiples of 8 and 4, respectively.

For each season and each station, the following procedure was used in the analysis:

(i) To estimate the parameters μ, σ, ρ and α of the marginal distributions, using all the data available;

(ii) To find the approximate covariance matrix of the estimators by using the asymptotic expressions developed in Section 3-4;

(iii) To test the goodness of fit of the marginal distribution by using the χ^2 statistic;

Table 4-2. Results Obtained in Case the Year is Divided in Twelve Seasons, for the Columbia Station

PERIOD	PARAMETERS				ASYMPTOTIC COVARIANCE MATRIX ($X\ 10^{-6}$)				T.S.1	T.S.2	T.S.3	T.S.4	T.S.5
	μ	σ	ρ	α					(d.f.)	(1d.f.)	(4d.f.)	(4d.f.)	N(0,1)
001-032	-.4109	.5475	.3848	.6121	2988	-2071	- 582	865	8.979	17.305	10.294	4.403	-6.902
						2383	950	-1302	(4)				
							7832	- 206					
								2398					
033-060	-.3291	.5370	.3584	.6655	2382	-1555	- 416	607	7.015	15.023	6.195		-3.237
						2048	827	1227	(5)				
							7934	- 148					
								2760					
061-092	-.2170	.5383	.1928	.6249	1347	- 767	- 113	140	10.857	5.633	7.601	3.191	-5.925
						1235	320	- 648	(6)				
							6754	- 25					
								1676					
093-120	-.1947	.5578	.2295	.7106	1557	- 832	- 196	101	10.427	7.316	1.028		-2.940
						1381	455	- 752	(7)				
							7166	- 22					
								2362					
121-152	-.3110	.7080	.3169	.7143	2547	-1329	- 264	- 4	16.416	15.243	2.587	5.804	-1.528
						1885	682	- 421	(9)				
							6146	- 40					
								2274					
153-180	-.2939	.7182	.1900	.6052	2743	-1460	- 171	- 37	6.570	4.521	2.297		-4.870
						2076	472	- 346	(9)				
							7771	- 12					
								1784					
181-212	-.3846	.7701	.2641	.6353	3274	-1818	- 313	- 2	7.825	9.650	2.242	7.865	-4.185
						2352	675	- 334	(9)				
							6980	- 22					
								1876					
213-240	-.5526	.7812	.2158	.6304	5566	-3465	- 427	346	7.823	4.374	4.990		-4.469
						3771	777	- 726	(7)				
							10608	- 62					
								2597					
241-272	-.6331	.9114	.3958	.6065	6681	-3916	- 771	- 64	5.637	19.562	5.601	4.058	-4.325
						4312	1348	- 169	(9)				
							7176	- 67					
								2093					
273-300	-.6799	.8538	.3706	.6254	8188	-5006	- 959	348	6.093	13.139	5.042		-3.186
						5115	1533	- 679	(7)				
							9684	- 122					
								2879					
301-332	-.5193	.6632	.2948	.6576	4355	-2864	- 551	788	10.577	8.961	1.449	3.572	-4.172
						3048	891	-1171	(5)				
							9407	- 148					
								2739					
333-360	-.3301	.5219	.3451	.6527	2365	-1595	- 424	676	4.687	13.269	0.730		-4.137
						2068	813	-1276	(4)				
							8330	- 158					
								2696					

(iv) To test the hypothesis that there is no serial dependence in the data (in which case a much simpler model would do the job...)

(v) To test the hypothesis of stationarity for alternative II (only for the seasons of length 32); and

(vi) To investigate which family of distributions, light or heavy tail, best fits the data.

The results obtained are presented in Table 4-2 (see also the Appendices B1 to B7 for the corresponding tables concerning the seven other stations). The remainder of this section is devoted to the description and comments related to these tables.

The Seasons. The first column of tables gives the beginning and the end, in days of each season. January 1 is day one. The 12th season ends on the 360th day of the year; therefore, five or six days (in leap years) are neglected for each year. This is certainly irrelvant for the objectives of this investigation, namely to evaluate the applicability of the model.

The Estimates and the Covariance Matrix. The next four columns give the estimates of μ,σ,ρ and α, respectively. The next four columns give the asymptotic covariance matrix of the estimators, assuming them arranged in the above order (μ,σ,ρ and α). As stated, previously these results are helpful in designing sensitivity studies. For example, assume that a generated sample will be used to perform a hydrologic routing, with the resulting output hydrograph at a location of particular interest. Suppose further that the hydrograph will be used for the design of a flood control structure. An important information is how the final *product* say the height of a dam or its cost, will be modified when *reasonable* variations are imposed on parameters of the generation model. What is a *reasonable* variation depends on the subjective criteria of an engineer. Regardless of this subjectivity, one would not expect that the *true* parameter would lie, say five standard deviations (of the pertient estimator) away from the estimated value. What is suggested is that the approximate covariance matrix may be useful in establishing the variations of parameters, which will be found *reasonable* by most people. Logically sensitivity analysis is a procedure to be applied on a case-by case basis. It seems worthwhile to point out that the first step in any sensitivity study will likely be to construct confidence intervals around each of the estimates, or even better, a confidence region. An approximate procedure to do it is to assume that each estimator is normally distributed, (which is true for large samples), and find intervals, rather than regions. In this case the approximate limits of the confidence intervals are obtained by adding to and subtracting from each estimate a quantity that is equal to the appropriate quantile of the standard normal distribution multiplied by the corresponding standard deviation. The last one can be obtained from the asymptotic expressions. For instance, for the Columbia data, period from 1 to 32, $\hat{\mu} = -0.4109$, std $(\mu) = \sqrt{2988 \times 10^{-6}} = 0.0547$. Therefore, an approximate 95 percent confidence interval for μ is given by $(-0.4109 + 1.960 \times 0.0547)$, so that the limits are for μ: -0.5180 and -0.3038; similarly for σ: 0.4518 and 0.6432; for $\rho = 0.2113$ and 0.5583; and for α: 0.5161 and 0.7081. The same precedure could be repeated for each of the four parameters, for each season of each station.

Goodness of Fit. Column 10, headed by the label T.S.1 (symbol for the test statistic No. 1), gives important information, namely the chi-square goodness of fit statistic. Inside the parenthesis are displayed the number of degrees of freedom, which depend on how the data have been arranged into groups. The good performance of the model with respect to reproducing the marginal distribution for each season-station is remarkable. Indeed none of the twelve seasons in which the Columbia data was divided has yielded a statistic that would lead to the rejection of the marginal fit, at a 5 percent significance level. (Appendix C gives the critical values of the chi-square distribution, at the 5 percent and the 1 percent significance levels). As it concerns the other stations (see Appendix B) the following results are given.

Table 4-3. Seasons for Which the Marginal Fit Were Rejected at the Five Percent Significance Level.

Station	Period	Chi-square Statistic	Degrees of freedom
Kansas City (MO)	061-092	21.586	9
Springfield (MO)	061-092	16.871	8
	333-360	15.226	7
Raleigh-Durham (NC)	032-060	24.353	9*
	301-332	19.083	8
Austin (TX)	001-032	34.282	11*
	033-060	27.749	13
	061-092	23.074	13
	333-360	32.065	14*
Rapid City (SD)	301-332	7.061	2
Flagstaff (AZ)	213-240	12.934	6
	241-272	14.631	6
Seattle-Tacoma (WA)	333-360	25.928	10*

*Rejected also at a 1 percent significance level

Therefore, out of 12 x 8 = 96 season-stations, 13 had the hypothesis of correct fit of the marginal distribution rejected at the 5 percent significance level. At the 1 percent significance level only four cases, those marked by an asterisk in Table 4-3 are rejected. No null hypothesis stating the *universality* of the model applications is tested here. If this was the case, and if the studied time series were spatially and serially uncorrelated, then one would expect to have no more than 5 season-stations rejected at a 5 percent level or no more than 1 at a 1 percent level. The purpose of this particular study is rather to identify cases for which it is not avisable to apply the model. For instance, an examination of Table 4-3 reveals that the four seasons that roughly span from December to March for the Austin station should not be modeled by this approach.

Test of Serial Independence. The next test statistic to be examined, T.S.2 of Table 4-2 and Appendix B, is described in section 3-5. It tests H_o: $\rho = 0$ against H_A: $\rho \neq 0$. If the null hypothesis holds, then the test statistic has approximately a chi-square distribution with one degree of freedom. Hence, one can reject the hypothesis, say at 5 percent significance level, whenever the test statistic takes a value greater than 3.84. For all except two of the 96 cases the null hypotheses were rejected. The only exceptions occurred for the 11th season of Rapid City station, where $\chi^2 = 3.81$, and 4th season of Flagstaff, with $\chi^2 = 3.79$. These two cases may be results of pure chance variations.

This overwhelming rejection of the hypothesis of serial independence in the analysed precipitation series makes one wonder about the reality of several models, as described in Chapter II, that neglect the time dependence in daily precipitation

Tests of Stationarity. The next two test statistics, T.S.3 and T.S.4 in Table 4-2 and Appendix B are related to the question of stationarity. One would not use the *season approach* in generating new samples if the data of a particular season shows some evidence of non-stationarity. In Section 3-5 two test statistics were developed to test the null hypothesis H_o: the process is stationary, versus two alternative hypotheses: Alternative I, with the parameters varying with time in a symmetrical manner around the center of the season. Under the null hypothesis both test statistics have approximately a chi-square distribution with four degrees of freedom. Hence, the hypothesis should be rejected at 5 percent significance level whenever the test statistic is greater than 9.49.

The test with the Alternative II was only applied for the seasons composed of 32 days, because the splitting procedure requires a sample which is a multiple of eight.

For the Columbia station data, only one season has the null hypothesis rejected. This happened for the first season and only against Alternative I. With regard to the other stations, very few rejections occurred either, as it is demonstrated by Table 4-4.

Hence the hypothesis of stationarity was rejected for 11 seasons, at the 5 percent significance level. Some of these rejections might be due to randomness in data, rather than to a deficiency of the model. Conversely the test might have been accepted for some of the other seasons due to chance and not to the adequacy of the model. However if one had to select *suspicious* seasons, as far as the application of the model is concerned, then the eleven cases which had the stationarity hypothesis rejected would lead the list. In the ensuing sections, it will be shown that alternative procedures may allow the use of the model for all year around with satisfactory results.

Table 4-4. Seasons for Which the Hypothesis of Stationarity was Rejected at the Five Percent Significance Level (critical value = 9.49)

Station	Period	Test Statistic	Alternative
Columbia (MO)	001-032	10.29	I
Kansas City (MO)	001-032	10.57	I
Rapid City (SD)	061-092	11.45	II
	121-152	13.27	II
Flagstaff (AZ)	121-152	11.25	I
	181-212	39.62	I
	241-272	14.23	I
Seattle-Tacoma (WA)	033-060	13.06	I
	061-092	11.18	I
	121-152	11.64	I
	241-272	19.30	I

Tests for Extreme Events. The last test statistic to appear in Table 4-2 is not directly related to the model, but rather is an evaluation of the characteristics of the data. Do sample frequencies need to be fitted by a light or a heavy tail probability distribution? The importance of this question stems from the singular role played by extreme events in the hydrologic design. The use of some light tail distribution, when the data require some heavy tail distribution, can lead to serious mistakes.

Bryson (1974) classified a distribution with c.d.f. $F_X(x)$ as heavy tailed whenever $g(x) =$

$$= \frac{1}{\bar{F}_X(x)} \int_x^\infty \bar{F}_X(t)dt$$ is an increasing function of x,

with $\bar{F}_X(x) = 1 - F_X(x)$. If $g(x)$ is constant, namely $g(x) = \frac{1}{\psi}$, $\forall x$, then $F_X(x) = 1 - e^{-\psi x}$, i.e. the exponential distribution is the benchmark between light and heavy tail distributions. It is important to underline that the family of distributions frequently used in hydrology has the exponential tail.

Holander and Proschan (1975) developed a procedure for testing the null hypothesis, namely H_o: the distribution has exponential tail. They proposed the test statistic (modified by a constant factor here)

$$V^* = \sqrt{\frac{210}{n^5}} \frac{\sum_{i=1}^{n} C_{in} R_i}{\sum_{i=1}^{n} R_i} \tag{4-1}$$

where

$$C_{in} = \frac{4}{3} i^3 - 4ni^2 + 3n^2 i - \frac{n^3}{2} \tag{4-2}$$

with $R_1, R_2, \ldots, R_n$ = the order statistics. They also proved that under the null hypothesis the test statistic follows the standard normal distribution. Lighter than exponential distributions will tend to have large V^* values and conversely heavier than exponential distributions will tend to have small (negative) V^* values. Therefore, the exponential and light tail hypothesis ought to be rejected, at the 5 percent significance level, whenever the observed test statistic is smaller than -1.645. Checking the column T.S.5 of Table 4-2, one realizes that this happened for all the 12 seasons but the 5th. Table 4-5 lists all the instances where the null hypothesis failed to be rejected.

Therefore only in 10 cases out of 96 cases the null hypothesis failed to be rejected at the 5 percent significance level. This could lead to the conclusion that in the majority of cases daily precipitation data ought to be fitted by a heavy tail distribution. However, it should be pointed out that the test assumes time particular application. Furthermore, the distribution of the test statistic is only asymptotically known. In other words, the above test might not perform the task it is supposed to do. Even for a correct test, insight on the *degree of heaviness* of the precipitation distributions would be useful. An alternate procedure for checking the tails seems to be in order.

Table 4-5. Seasons for Which the Hypothesis of Light Tail Failed to Be Rejected

Station	Period	Test Statistic
Columbia (MO)	121-152	-1.528
Kansas City (MO)	153-180	-1.213
Raleigh-Durham (NC)	033-060	0.571
	333-360	0.002
Flagstaff (AZ)	033-060	-1.592
	093-120	-0.697
	121-152	-0.741
Seattle-Tacoma (WA)	001-032	-1.627
	061-092	-1.325
	273-300	-1.378

Bryson (1974) suggested that a graph of log $\overline{F}_X(x)$ versus x could give a visual perception of the tail behaviour of the probability distribution function. In fact, the procedure serves the purpose of amplifying the tail characteristics of a distribution. Obviously in case of exponential probability distribution this graph will plot as a straight line. According to Bryson, "...the graph of a distribution with an exponential tail, such as the gamma will approach such linearity for large x. (Heavy tail) distributions then, will be characterized by graphs that do not approach such linearity and which remain too high. Unfortunately, it is difficult to be more precise. This property means that the graph will tend to be concave for large x..."

Figure 4-2 displays the graphs of log $\overline{F}_X(x)$ versus x for several power-transformed-truncated-normal-distributions with different power parameters α, and for $\mu = o$. For comparison other distributions were also plotted, such as Pareto, exponential, and kappa. The kappa distribution was used by Mielke (1973) for fitting precipitation data. The plot was done in such a way that $\overline{F}(10) = 0.05$ for all cases. It is apparent that the smaller α the heavier is the tail. For $\alpha = 1$, which corresponds to the truncated normal, the tail is lighter than the exponential.

An excess number of figures is avoided by considering only the sequence of four seasons (120 days) that have the highest average precipitation. This seems to be an appropriate criterion due to the problem being investigated, namely extreme or flood type events. Figure 4-3 displays the four graphs pertinent to Columbia station. The same graphs related to the other stations are given in Appendix B. In general, no concavity of the curves, either the fitted (continuous) or the observed (dashed), is evident. Rather, they resemble very much straight lines, which indicates that the distributions with exponential type tail are not precluded from fitting precipitation data. This statement is in contradiction with the previous conclusion about the test made on tails. The important point, however, is that even if data require indeed the heavy tail distributions, the *degree of heaviness* would be very low. Based on the above study, one can say that the issue of heavy or light tail does not seem to be very relevant to the application of the precipitation model developed in this study.

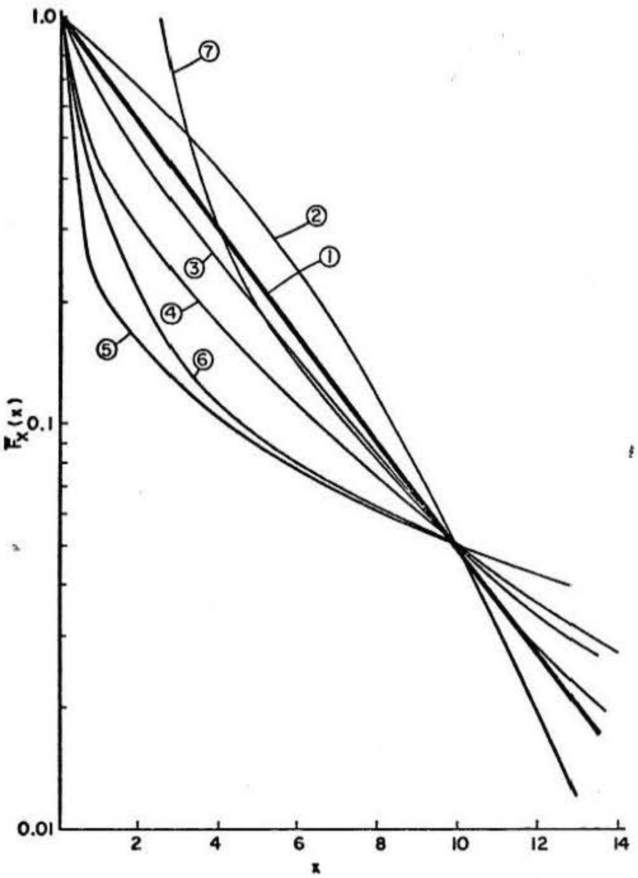


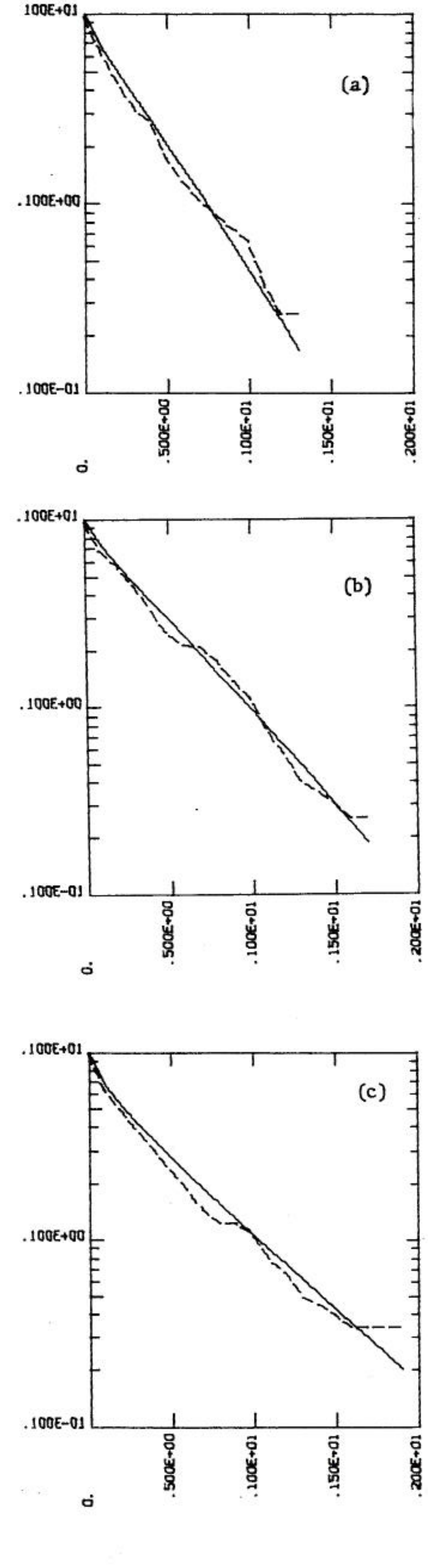
Fig. 4-2. Graphs of log $\overline{F}_X(x)$ Versus x for Selected Distributions: (1) Exponential; (2) Truncated Normal; (3) Power-Transformed Truncated Normal, $\alpha = 0.6$; (4) Power-Transformed Truncated Normal, $\alpha = 0.4$; (5) Power-Transformed Truncated Normal, $\alpha = 2$; (6) Kappa; (7) Paretto.

4-3 An Example of a Multivariate Application

A simple illustration is given here to show the use of the model in a multivariate case. Suppose that one wants to produce the new samples of precipitation data for the stations of Columbia, Kansas City, and Springfield simultaneously by preserving the areal dependence among them. Assume further that only the most rainy month for the region is of interest. This is in June, roughly corresponding to the 6th season of the classification used in the last section (period 153-180 or June 2-June 29). Therefore the marginal parameters can be found respectively from Table 4-2, and Appendices B-1 and B-2, respectively.

Next step is to find each of the three lag-zero cross correlation coefficients between station series by using Eq. (3-24). The results are summarized in Figure 4-4.

One hundred trivariate samples each for the month of June, were generated simultaneously according to the procedure explained in Section 3-6. Out of many



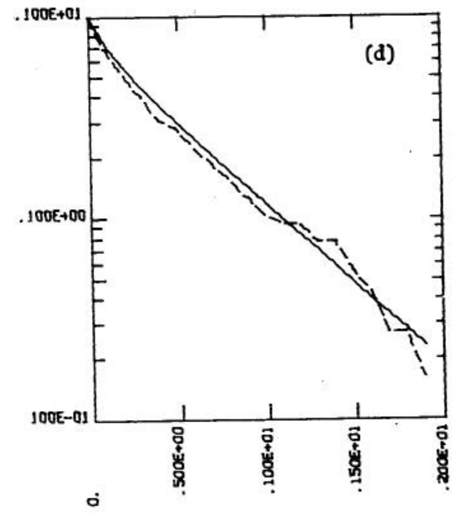
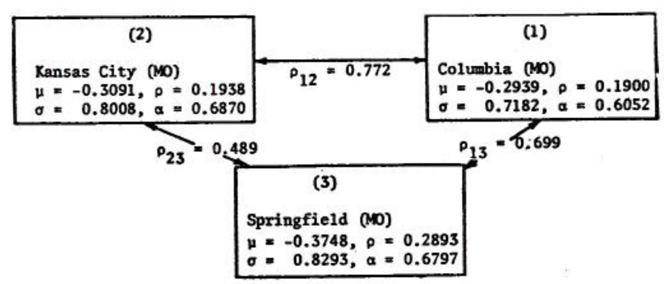


Fig. 4-3. Plot of $\log \bar{F}_X(x)$ Versus x for Selected Distributions: (a) Period 93-120; (b) Period 121-152; (c) Period 153-180; (d) Period 181-212.

Fig. 4-4. Representation of Parameters Needed for Generation of Daily Precipitation Series for the Month of June.



ways of comparing the historic and the generated series, it was decided to focus attention on the joint positive runs. A joint positive run is defined as a succession of days for which the precipitation is observed at all three stations, preceded and followed by days for which at least at one of the stations no precipitation occurred. For each joint positive run, the two variables of interest are: (i) The length, defined by $L_2 - L_1 + 1$, and (ii) the joint run-sum, defined as

$$\sum_{i=1}^{3} \sum_{j=L_1}^{L_2} X_{i,j}$$ where X_{ij} is the amount of precipitation at the ith station in the jth day, L_1 = the first day of the joint positive run, and L_2 = the last day of the joint positive run. These two variables were selected with the solution of flood problems in mind. Table 4-6 gives the absolute frequencies of run-lengths for the historic and generated series.

Table 4-6. Absolute Frequency of the Joint Positive Run-Lengths

Sample	Run-length 1	2	3	4	5	6	Total
Historic	31	8	5	1	0	0	45
Generated	190	31	19	4	0	1	245

Whether the two samples of Table 4-6 can be considered as drawn from the same population is of crucial importance in the evaluation of the model. A way of answering this question is by using the test of equality of two multinomial distributions. The reader is referred to Mood et al. (1974) where a description of the test is given (pages 448-452). It is sufficient to state here that the sample space is divided in $k + 1$ subsets and the null hypothesis states that Ho: $p_{1j} = p_{2j}$, $j = 1, 2, \ldots, k + 1$, where p_{1j} = the probability that an outcome of the first population belongs to the jth subset, and p_{2j} = the same as p_{1j} but in regard to the second population. For the above data the division in three subsets ($k = 2$) seems convenient, namely: (i) run of length 1; (ii) run of length 2; and (iii) run of length > 2.

It can be shown that

$$T.S.6 = \sum_{i=1}^{2} \sum_{j=1}^{k+1} \frac{(G_{ij} - g_i(G_{1j} + G_{2j})/(g_1 + g_2))^2}{g_i(G_{1j} + G_{2j})/(g_1 + g_2)} \qquad (4\text{-}3)$$

has a limiting chi-square distribution with k degrees of freedom, where g_1 = the total number of observations for the first population (in the present case, 45); g_2 = the same as g_1, but for the second population (245); G_{1j} = number of outcomes in class j, from the first population; and G_{2j} = the same as G_{1j}, but from the second population.

The use of Eq. (4-3) with the data of Table 4-6 yields a value of T.S.6 = 1.58. Since the 95 percent quantile of the chi-square distribution with two degrees of freedom is 5.99, the null hypothesis cannot be rejected at the 5 percent significance level.

With regard to the joint run-sums, again the test whether the two samples (not given in tables) were drawn from the same population if performed. Since this variable is continuous, the two-sample-Smirnov test seems more suitable than the multinomial one. For a description of that test see Bradley (1968). Here it is sufficient to state that under the null hypothesis of equality of the two distributions, the random variable

$$T.S.7 = \max_x |S_1(x) - S_2(x)| \qquad (4\text{-}4)$$

has some distribution which 95 percent quantile is given approximately by

$$1.358 \sqrt{\frac{g_1 + g_2}{g_1 g_2}} \qquad (4\text{-}5)$$

where $S_1(x)$ is the sample c.d.f. of the historic sample and $S_2(x)$ is its counterpart for the generated sample.

The application of Eqs. (4-4) and (4-5) to data gives the values of 0.1868 and 0.2202, respectively. Therefore, the hypothesis stating that the two samples can be considered as drawn from the same population should be accepted at the 5 percent significance level.

In synthesis, the application of the model to the multivariate case is satisfactory for the example used. This is a positive indication about the feasibility of using the model.

4-4 The Non-Stationary Approach to Analysis of Intermittent Stochastic Processes

If the model is applied to a long period of time, say to the whole year, the season approach is no more feasible. First, one might consider the possibility of generating new samples by using a succession of seasons. For example, one could divide the year in twelve seasons (as in Section 4-2), and assume stationarity inside each season, though each season would be stochastically different from the others. However if, say, April 30 is the last day of a season, and May 1 is the first day of the next season, then according to described procedure the daily precipitation process is expected to undergo an abrupt transition in parameters between these two days. This is not a realistic approach.

Another alternative is to assume that any parameter of the model is a periodic function of time, rather than a constant. In this case, the question is how to estimate the time functions $\mu_\tau, \sigma_\tau, \rho_\tau, \alpha_\tau$, and $\rho_\tau(i,j)$. The *season approach* can be used as an intermediate step for solving the problem, namely split the year into seasons and for each season estimate a set of parameters under the hypothesis of stationarity. The time variation of each parameter can be represented by a bar graph, as in Figure 4-5, where Θ represents any of the above referred parameters. Next fit the bar graph with some smooth function.

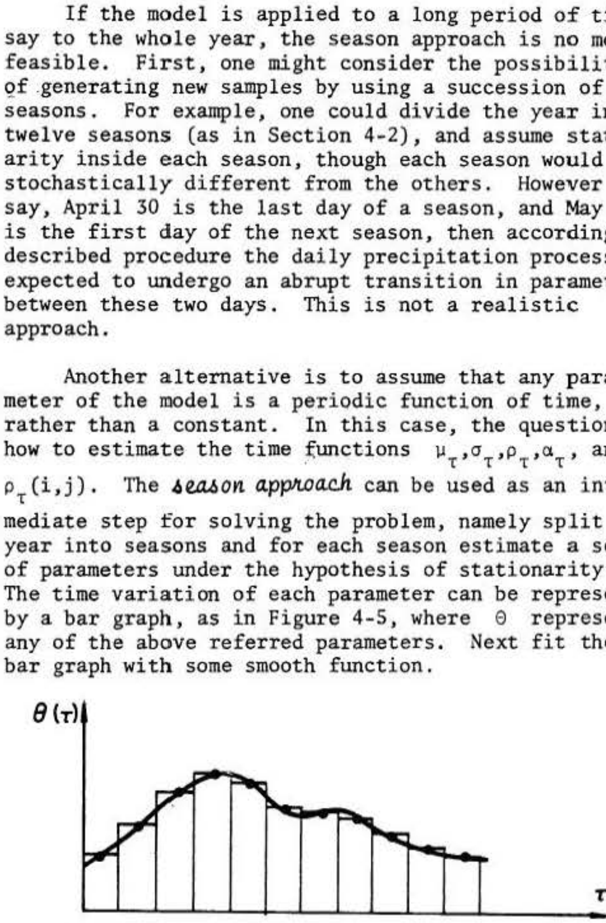


Fig. 4-5. Fitting a Smooth Function to Values Obtained by Season Approach.

The most convenient way of expressing Θ_τ is through an analytical expression. Smoothing techniques like the moving-average scheme are not appropriate. Usually the periodic parameters of hydrologic time series are fitted by Fourier series, with the present study following this technique.

The periodic function Θ_τ, $\tau = 1,2\ldots, \Omega$ is difined by

$$\Theta_\tau = \bar{\Theta} + \sum_{j=1}^{k/2} [\hat{a}_j \cos(\tfrac{2\pi j\tau}{\Omega}) + \hat{b}_j \sin(\tfrac{2\pi j\tau}{\Omega})] m(j) \qquad (4\text{-}6)$$

with

$$\hat{a}_j = \frac{2}{k} \sum_{\tau=1}^{k} \theta_\tau \cos \frac{2\pi j \tau}{k} \tag{4-7}$$

and,

$$\hat{b}_j = \frac{2}{k} \sum_{\tau=1}^{k} \theta_\tau \sin \frac{2\pi j \tau}{k} \tag{4-8}$$

and,

$$\bar{\theta} = \frac{\sum_{\tau=1}^{k} \theta_\tau}{k} \tag{4-9}$$

with θ_τ the individual estimates of θ along the values $\tau = 1, 2, \ldots, k$: k = the number of seasons in which the year is divided (for simplicity assumed to be an even number); τ = the day index; Ω = the number of days; and $m(j) = \{$1, if the jth harmonic is considered significant; 0, otherwise.

A clarification is necessary about what is meant by *significant harmonic*. The use of Eqs. (4-7) and (4-8) yields only the estimates of the *true* parameters a_j and b_j. Hence even if a_j and b_j were both equal to zero, meaning that there is no periodic signal with frequency j/k, still $\hat{a}_j$ and $\hat{b}_j$ are likely to be different from zero. In these conditions, if the jth harmonic is accepted as a legitimate periodic component, then a spuriors periodicity would be included in the formation of θ_τ. The question is then how to decide whether a_j and b_j are significantly different from zero, i.e. whether the jth harmonic is significant. A way of accomplishing this is by the classical Fisher's test for a process composed of the sum of a harmonic and a normal independent process. This test, as well as some empirical procedures, are described in detail by Yevjevich (1972b). A difficulty with the Fisher's approach for this kind of problem, besides those pointed out by Yevjevich, is that when fitting the periodic functions one is much more worried about the possibility of committing the Type II error than the Type I error. The Fisher's test only controls the Type I error. In other words, while smoothing a *step function* one is very much concerned with missing some periodic signal that should be included but does not care so much when the case is opposite, namely of a harmonic being wrongly considered as significant.

As mentioned before, Yevjevich (1972b) suggests a couple of empirical procedures for testing the significance of harmonics. Another empirical methodology was found to be convenient in the present study.

The mean square of deviations of $\theta_\tau, \tau = 1, 2, \ldots, k$, from the mean $\bar{\theta}$ is given by

$$\sum_{j=1}^{k/2} \left(\frac{\hat{a}_j + \hat{b}_j}{2} \right)^2 = \sum_{j=1}^{k/2} \text{var } h_j \tag{4-10}$$

where (var h_j) is called variance of the jth harmonic.

Under the hypothesis that all the $k/2$ harmonics are not significant, the expected value of the variance of any harmonic is given by

$$E(\text{var } h_j) = \frac{2}{k} \sum_{j=1}^{k/2} \text{var } h_j = \overline{\text{var}}\ h \tag{4-11}$$

If the above hypothesis is not true the tendency will be to have at least one of var h_j well above $\overline{\text{var}}$ h. The suggested empirical procedure is to reject the hypothesis whenever

$$\max(\text{var } h_j) > \overline{\text{var}}\ h, \tag{4-12}$$

where c is a constant greater than one. In the cases analysed in the ensuing text the value of $c = 3$ was found appropriate.

The hypothesis being rejected means that the harmonic h_i with the highest variance ought to be considered significant. Next step is to redefine

$$\overline{\text{var}}\ h = \frac{2}{k-2} \left[\sum_{j=1}^{k/2} (\text{var } h_j) - \text{var } h_i \right] \tag{4-13}$$

(the ith harmonic is excluded) and check whether

$$\max_{j,\ j \neq i} (\text{var } h_j) > c\ \overline{\text{var}}\ h,$$

and so on. As long as the hypothesis is not rejected the procedure is continued.

The precipitation data for the eight stations studied were divided into 26 seasons ($k = 26$), 14 days each, adding up to 364 days. This is a rather arbitrary selection; one could choose the season length as short as two days. For each station-season the parameters μ, σ, α, and ρ are estimated according to the procedure described in Chapter III. No test as applied in Section 4-2 was repeated. Then the above technique was applied to each of the four parameters of the eight stations. After some studies, it was found that sometimes the harmonics corresponding to high frequencies were inferred significant, which somehow violates the general experience. Because of that a further criterion was added to the method. A harmonic was only considered able to be significant if its order was lower than six. The results are summarized in Table 4-7. Graphs of the functions $\mu_\tau, \sigma_\tau, \rho_\tau$, and α_τ for the Columbia station, obtained by using Eq. 4-6, with parameters given in Table 4-7, are plotted in Figures 4-6 and 4-7. The similar graphs, corresponding to the other seven stations are given in Appendix B.

A matter of interest is to check how the empirical procedure for determining the significant harmonics compares with the Fisher's test. The later was applied to all the 8 x 4 = 32 cases and the difference between the number of harmonics indicated for each case, by the two techniques is shown in Table 4-8.

In general the empirical method yields a greater or equal number of significant harmonics as compared with the Fisher's test. This means that the probability of rejecting a significant harmonic is lower if the empirical approach is used rather than vice-versa.

Therefore the empirical technique has the property for which it was developed.

The number of parameters estimated from data and used for generation is an important information. One can say that the reliability of any schematic representation of data varies inversely with the number of parameters estimated from data. Table 4-7 shows all the estimates required for generating new series for each one of the stations. Table 4-9 displays the number of parameters used for each station; this information is extracted from Table 4-7.

Taking into consideration that daily precipitation is being modeled, it seems fair to say that the necessary number of parameters is remarkably low. However, the central issue is whether the model is capable of producing the results of practical significance. This is investigated in the next section.

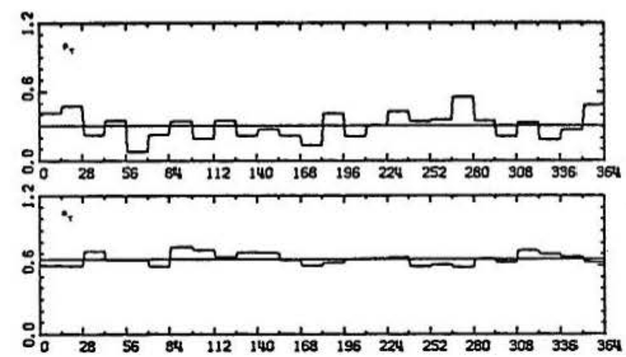


Fig. 4-7. The Periodic ρ_τ and α_τ for Daily Values of the Columbia Precipitation Station.

Table 4-7. Significant Harmonics of Four Parameters

	μ				σ				ρ				α			
Station	$\bar{\mu}$	j	$\hat{a}_j$	$\hat{b}_j$	$\bar{\sigma}$	j	$\hat{a}_j$	$\hat{b}_j$	$\bar{\rho}$	j	$\hat{a}_j$	$\hat{b}_j$	$\bar{\alpha}$	j	$\hat{a}_j$	$\hat{b}_j$
Columbia	.403666	1	.031489	-.190090	.673970	1	-.100808	-.144878	.306145				.653941			
Kansas City	.461488	1	.075859	-.191089	.721135	1	-.113136	-.148063	.320545	4	.057535	.067766	.661580	6	.022762	-.027159
						2	-.052782	-.009277		1	.069217	-.025740				
Springfield	.415715	1	.030349	-.140684	.724957	1	-.072608	-.088316	.314639				.620299	1	-.038077	.010293
Raleigh	.441735	1	.029385	-.176852	.751143	1	-.027255	-.124785	.306876				.695255			
						3	.017370	.074833								
Austin	.713762	1	-.150029	-.173400	.875042	1	-.119636	-.084729	.425415	5	-.047476	-.016918	.586908			
		3	-.042823	-.056385		2	-.042252	-.068506								
		5	.033895	.040210												
		6	-.010444	-.046761												
Rapid City	.241580	1	-.010936	-.127151	.388333	1	-.175118	-.048592	.319249				.702596	1	.110888	-.004344
		2	-.095697	-.007476		2	-.043872	-.034110								
		4	.013093	.063780												
Flagstaff	.569254	2	.088979	-.198420	.631429	4	.100050	-.022900	.488464				.724391			
		3	-.078080	.147733		1	.078021	-.043959								
		4	.138501	-.046148												
Seattle	.069268	1	-.240519	-.089106	.421739	1	.040363	-.039487	.426829	2	-.041110	.082934	.685665			
		2	-.010185	.060887												
		5	.024331	-.022304												
		4	-.030524	.005607												

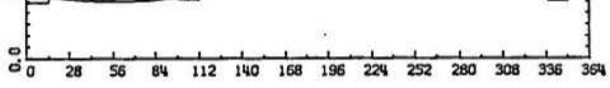


Fig. 4-6. The Peridoic μ_τ and σ_τ for Daily Values of the Columbia Precipitation Station.

4-5 Further Tests of the Model

The practical use of the model ultimately depends on its capacity to generate new series that correctly reproduce, in a stochastic sense, the historical series. In order to study this subject, samples of 50 years

Table 4-8. Number of Harmonics Obtained by Empirical Procedure Reduced for Number of Harmonics Obtained by Fisher's Test

	μ	σ	ρ	α
Columbia	0	0	0	0
Kansas City	0	1	2	1
Springfield	0	0	1	1
Raleigh-Durham	0	1	0	0
Austin	3	0	1	0
Rapid City	0	1	-1	0
Flagstaff	2	2	0	0
Seattle-Tacoma	2	0	0	0

Table 4-9. Total Number of Parameters Used for Generation of New Series

Columbia	8	Austin	18
Kansas City	16	Rapid City	16
Springfield	10	Flagstaff	14
Raleigh-Durham	10	Seattle-Tacoma	16

length were generated for each station using periodic parameters, as explained in Section 4-4. The periodic functions used for each station are given in Table 4-7. The objective is to compare whether the generated and the histroic series can be considered as drawn from the same population.

Likely, a practitioner will be satisfied with the model performance if the sample distributions, historic and synthetics of some functional of the process are similar to each other. In more specific terms, an engineer would be interested in a practical case in some random variable which is derived from the original process (called a functional). Which functional is selected depends on the problem the model user is facing. Eleven functionals are chosen to be investigated in this study. It is expected, that they adequately cover all the practical aspects an engineer might be interested in. Figure 4-8 represents a hypothetical year of record, for which only three *storms* have occurred. It helps the definitions of these functionals.

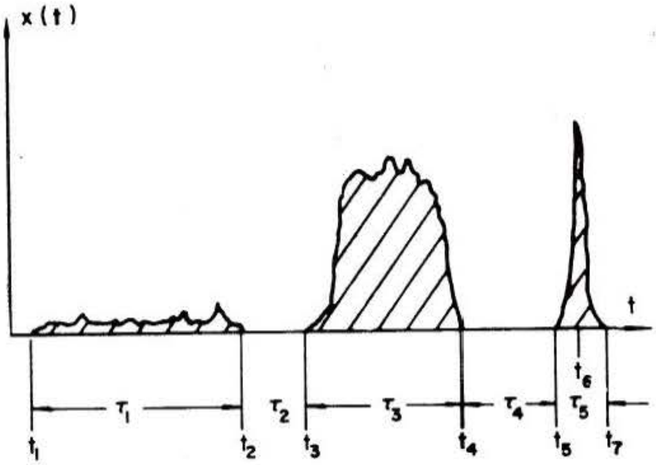


Fig. 4-8. Definition of Functionals.

The various functionals are named and defined as follows:

(i) The positive run-length as the length of a succession of days for which some precipitation is observed, preceded and followed by days with no precipitation registered, such as τ_1, τ_3, and τ_5. A run that starts in year j and ends in year $j + 1$ is counted as it had happened in year $j + 1$.

(ii) The negative run-lengths as the length of a succession of days for which no precipitation is observed, preceded and followed by days with some precipatation registered, such as τ_2, τ_4.

(iii) The longest positive run-length, as the length of the longest positive run in a year, such as τ_1.

(iv) The longest negative run-length as the length of the longest negative run in a year, such as τ_4.

(v) The number of total runs, as the number of complete pairs of positive and negative run-lengths in a sample, such as the two total runs, (τ_1, τ_2) and (τ_3, τ_4).

(vi) The time of occurrence of the longest positive run-length, as the time when the longest positive run begins, such as t_1.

(vii) The time of occurrence of the longest negative run-length as the time when the longest negative run begins, such as t_4.

(viii) The time of occurrence of the largest run-sum, as the time when the largest run-sum begins (see the definition of the next functional) such as t_3.

(ix) The maximum run-sum, as the highest amount of precipitation corresponding to a positive run-sum, such as $\int_{t_3}^{t_4} x(t)dt$.

(x) The annual total, as the total precipitation in a year, such as $\int_0^{t_7} x(t)dt$.

(xi) The daily maximum, as the maximum amount of precipitation registered for a single day, such as $x(t_6)$.

The null hypothesis to be tested, for any functional is that the two samples, historic and generated, are from the same underlying population distribution.

The test of above hypothesis depends on the functional. For the continuous variables, functional (ix) through (xi), the Smirnov two-sample test can be applied. For the discrete variables with outcomes well spread, the Smirnov two-sample test can be also applied through in an approximate way; this is the case for functionals (iv) through (viii). However, the functionals (i) through (iii) are discrete and have their outcomes clustered around few possibilities so that for this group the test of equality of two multinomial distributions is most appropriate. Both tests were referred to in Section 4-3.

Table 4-10 shows some of the results related to the functionals (i) through (iii), that were obtained from the, historic and generated series. The sample mean and standard deviation are given respectively as $\overline{x}$ and s. The degree of homogeneity between the sample distributions is measured by T.S.6, which is defined by Eq. (4-3). Under the null hypothesis this test statistic has a limiting chi-square distribution. The degrees of freedom, for each case, are shown inside parentheses, under the appropriate heading. Appendix C gives the critical values of the chi-square distribution at the 5 percent and at the 1 percent significance levels, respectively.

The values marked with asterisks are those that lead to the rejection of the null hypothesis at the 5 percent significance level. It can be seen that the generated series did quite well with respect to the maximum positive run-length inasmuch as none of the stations had the hypothesis of homogeneity rejected. Even with respect to the negative run-length the performance is good, with only two rejections: Raleigh-Durham and Flagstaff. However when it comes to the positive run-length the results are bad: for all the stations but one the hypothesis is rejected. This is a strong indication of the incapability of the model to reproduce this particular functional. An inspection

Table 4-10. Comparisons of Sample Distribution for Functionals (i), (ii), (iii), of Historic and Generated Series

	POSITIVE RUN-LENGTH			NEGATIVE RUN-LENGTH			MAX POS. RUN-LENGTH		
	$\bar{x}$	s	χ^2	$\bar{x}$	s	χ^2	$\bar{x}$	s	χ^2
	(HIST) (GEN)	(HIST) (GEN)	(d.f.)	(HIST) (GEN)	(HIST) (GEN)	(d.f.)	(HIST) (GEN)	(HIST) (GEN)	(d.f.)
Columbia	1.73	1.07	*24.29	4.42	4.00	7.71	5.61	1.83	3.99
	1.71	1.15	(5)	4.39	4.04	(15)	6.08	1.65	(3)
Kansas City	1.71	1.02	*11.29	4.72	4.56	26.40	5.22	1.06	2.37
	1.68	1.06	(5)	4.79	4.89	(17)	5.54	1.17	(3)
Springfield	1.81	1.10	*43.16	4.45	3.95	14.96	6.00	1.38	1.34
	1.74	1.16	(6)	4.42	4.25	(16)	6.24	1.30	(2)
Raleigh-Durham	1.82	1.13	*38.28	4.46	3.89	*46.77	6.00	1.38	1.12
	1.74	1.54	(6)	4.39	3.98	(14)	6.02	1.62	(2)
Austin	1.76	1.17	10.88	6.46	6.75	31.78	5.56	1.82	2.43
	1.69	1.13	(7)	6.28	6.35	(27)	5.64	1.73	(4)
Rapid City	1.81	1.19	*11.73	4.98	5.33	8.26	6.39	1.70	1.68
	1.80	1.31	(5)	4.89	5.09	(15)	6.90	1.84	(2)
Flagstaff	1.98	1.45	*30.17	7.69	9.13	*36.50	7.00	3.40	0.61
	1.87	1.43	(6)	7.31	8.77	(17)	6.96	1.79	(2)
Seattle-Tacoma	3.02	3.02	*19.78	3.80	4.64	14.43	15.10	5.75	2.01
	2.76	2.44	(11)	3.52	4.08	(12)	12.32	3.22	(1)

*The null hypothesis is rejected at the five percent significance level.

of sample distributions of the positive run-lenth helps to explain this case. See Table 4-11.

Table 4-11. Sample Distributions of Positive Run-Lengths

		P (J=1)	$\hat{P}$ (J=2)	$\hat{P}$ (J≤2)
Columbia	HIST	.5488	.2908	.8396
	GEN	.6017	.2235	.8251
Kansas City	HIST	.5488	.2882	.8370
	GEN	.5944	.2442	.8386
Springfield	HIST	.5011	.3079	.8090
	GEN	.5924	.2223	.8146
Raleigh-Durham	HIST	.5008	.3125	.8133
	GEN	.5892	.2269	.8162
Austin	HIST	.5648	.2535	.8183
	GEN	.6053	.2298	.8350
Rapid City	HIST	.5358	.2674	.8132
	GEN	.5828	.2270	.8098
Flagstaff	HIST	.4815	.2733	.7548
	GEN	.5750	.2231	.7981
Seattle-Tacoma	HIST	.3696	.2268	.5964
	GEN	.4045	.2173	.6214

In the Table 4-11 "J" stands for the positive run-length. Attention is called to the fact that the frequencies of runs of length one to the historic series are always lower than their equivalents obtained from the generated series. The situation is reversed when it comes to the runs of length two. These discrepancies are the main source for high outcomes of χ^2 as listed in Table 4-10. In Table 4-11 one can see also that the frequencies of runs of lengths shorter or equal to two for both, the historic and the generated series are very close.

It can be inferred from the above that the generated series fail consistently reproducing the distributions of positive run-length, because a part of the run of one should be the run of two. Whether this is a serious drawback of the model depends on the application to each particular case. The manipulation of the historic series is done in such a way that all daily precipitation that do not reach a minimum amount are considered as zero. Since for the generated series the same procedure is not applied, this might partially explain the problem.

The Smirnov two-sample test was used for testing the null hypothesis for the functionals (iv) through (xi). Equation (4-4) was used to find T.S.7 and Eq. (4-5) was used to compute the critical value, d_{cr}. g_1 corresponds to the number of years of the historic data, as given in Table 4-1; g_2 corresponds to the number of years in the generated series, which was set to 50 for all the stations. Tables 4-12, 4-13, and 4-14 show the results. As usual, an asterisk was used to mark the cases for which the null hypothesis is rejected at the 5 percent significance level. An examination of results in Tables 4-12, 4-13, and 4-14 indicate that the model can be trusted as a working technique. The worst discrepancy came from the daily maximum. However, this is more than compensated by the excellent results related to the maximum run-sum, as both functionals are related to flood proplems with the latter being far more important to flood designs than the former.

Table 4-12. Comparisons of Sample Distribution for Functionals (iv) and (v), of Historic and Generated Series

	MAX NEG RUN-LENGTH			NUMBER OF TOTAL RUNS			
	$\bar{x}$ (HIST) (GEN)	s (HIST) (GEN)	T.S.7	$\bar{x}$ (HIST) (GEN)	s (HIST) (GEN)	T.S.7	d_{cr}
Columbia	18.56 20.58	5.62 5.59	.2778	59.17 59.68	4.72 4.61	.1533	.3738
Kansas City	22.43 24.86	5.34 7.04	.1530	56.57 56.26	5.91 5.33	.1191	.3427
Springfield	19.26 20.60	3.98 6.22	.0878	58.04 59.12	6.05 6.03	.1696	.3427
Raleigh-Durham	19.62 18.80	5.78 3.64	.1933	57.90 59.40	3.84 6.25	.2143	.3536
Austin	30.09 28.86	10.36 13.40	.2000	44.19 45.70	6.81 4.28	.2029	.2518
Rapid City	27.89 25.36	7.36 6.87	.1844	53.61 54.46	7.36 6.87	.1400	.3738
Flagstaff	39.94 42.38	10.97 14.06	.1689	37.61 39.62	4.26 4.14	.3556	.3738
Seattle-Tacoma	24.67 22.98	9.69 6.98	.1505	53.33 57.88	6.15 4.62	*.4086	.3536

*The null hypothesis is rejected at the five percent significance level.

Table 4-13. Comparisons of Sample Distribution for Functionals (vi), (vii), (viii), of Historic and Generated Series

	TIME OF LONG. POS. RUN			TIME OF LONG. NEG. RUN			TIME OF LARGEST RUN-SUM			
	$\bar{x}$ (HIST) (GEN)	s (HIST) (GEN)	T.S.7	$\bar{x}$ (HIST) (GEN)	s (HIST) (GEN)	T.S.7	$\bar{x}$ (HIST) (GEN)	s (HIST) (GEN)	T.S.7	d_{cr}
Columbia	144.67 128.14	81.15 79.28	.1644	216.94 236.54	127.48 110.83	.1778	184.72 196.30	49.77 66.25	.2778	.3738
Kansas City	167.65 134.62	76.52 96.05	*.3461	190.17 222.16	131.60 137.59	.2061	195.48 206.56	64.39 69.21	.1557	.3427
Springfield	143.52 142.36	99.65 90.70	.1574	228.00 220.76	106.73 107.73	.1191	185.65 177.16	92.34 86.45	.1922	.3427
Raleigh-Durham	131.29 131.52	87.73 94.45	.2010	229.62 220.54	100.74 92.23	.2067	233.86 184.60	52.76 81.50	*.3895	.3536
Austin	143.07 152.28	109.28 116.58	.1057	186.60 179.22	87.04 83.73	.1314	187.34 186.30	88.32 90.26	.1114	.2518
Rapid City	109.78 134.06	69.66 55.98	.3244	249.28 187.64	115.56 131.15	.2689	170.67 163.52	40.51 35.05	.1844	.3738
Flagstaff	130.22 189.52	109.47 95.25	*.4578	225.00 147.28	97.23 61.27	*.4356	217.56 173.94	95.88 106.13	.2733	.3738
Seattle-Tacoma	171.71 144.96	146.12 136.74	.2210	195.86 202.50	34.74 24.48	.2143	205.38 175.24	146.61 142.89	.1590	.3536

*The null hypothesis is rejected at the five percent significance level.

Table 4-14. Comparisons for Sample Distributions for Functionals (ix), (x), and (xi), of Historic and Generated Series

	MAX RUN-SUM			ANNUAL TOTAL			DAILY MAXIMUM			
	$\bar{x}$ (HIST) (GEN)	s (HIST) (GEN)	T.S.7	$\bar{x}$ (HIST) (GEN)	s (HIST) (GEN)	T.S.7	$\bar{x}$ (HIST) (GEN)	s (HIST) (GEN)	T.S.7	d_{cr}
Columbia	3.70 3.71	1.04 1.19	.1800	33.66 33.25	6.43 5.90	.2422	2.50 2.15	.58 .61	*.3822	.3738
Kansas City	3.77 4.50	1.14 1.65	.3322	36.04 36.13	9.06 7.67	.1583	2.66 2.50	.83 .83	.2548	.3427
Springfield	4.61 4.49	1.61 1.50	.0843	38.46 37.26	7.64 6.20	.2087	3.03 2.36	.87 .65	*.4157	.3427
Raleigh-Durham	3.78 4.27	.95 1.39	.2867	41.35 41.44	4.80 6.62	.1524	2.89 2.41	.87 .64	.2886	.3536
Austin	5.25 5.02	3.25 1.84	.1800	33.02 32.45	10.09 8.25	.0800	3.75 3.03	2.10 .96	.1971	.2518
Rapid City	2.50 2.69	.92 1.35	.2422	16.39 16.88	3.55 3.51	.1622	1.73 1.29	.72 .42	.3089	.3738
Flagstaff	3.35 3.46	1.40 1.57	.1511	19.82 21.61	5.50 5.00	.2600	1.87 1.49	.51 .50	*.3822	.3738
Seattle-Tacoma	5.65 4.88	2.43 1.46	.1857	39.95 41.07	6.66 5.74	.1838	1.84 1.43	.56 .31	*.4495	.3536

*The null hypothesis is rejected at the five percent significance level

A Dual Model For Daily Streamflows

In this chapter a new approach for the stochastic modeling of daily streamflow is introduced. It should be pointed out at the outset that no *universality* is claimed for the model to be described. In fact, the attempt to develop a general model may have been the reason for failures of previous efforts to model daily flows. It is hardly conceivable that a simple scheme could model equally well the streams fed by snowmelt and the streams draining a tropical catchment, to give only an example. The model to be described here refers to catchments for which the direct runoff plays an important part in the composition of the total flow. Nevertheless, each catchment that qualifies for such a description must be studied on a case-by-case basis.

A dual approach is used, in the sense that the positive and the negative first-derivatives of the streamflow process can be modeled by two alternating intermittent stochastic processes.

In this chapter first the conceptual framework is set up, and then the technique described with the help of the case study of the Powell River, near Arthur, Tennessee. This river is described by Quimpo (1967) as having an accurate record from 1921 to 1960. The outlet drains an area of 683 square miles and is located at 36°32'N latitude and 83°38'W longitude. The mean daily flow is 1116 cfs. For a better insight into the type of streamflow studied, Figure 5-1 shows the hydrograph for the year of 1921, which is a fairly typical hydrograph.

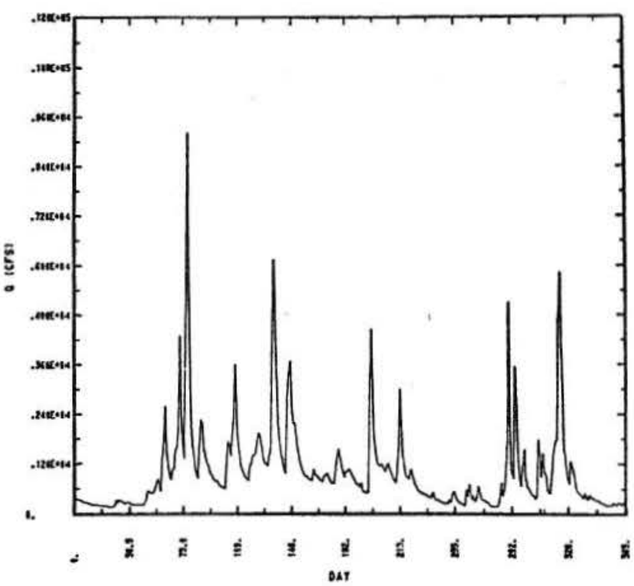


Fig. 5-1. Daily Flow Hydrograph of the Powell River for the Year of 1921.

5-1 The Conceptual Framework

The runoff at the outlet of a watershed is considered to be the sum of three components, namely,

$$q(t) = q_1(t) + q_2(t) + q_3(t) \tag{5-1}$$

Conceptually, these components have different physical characteristics, as in the case of underground flow and surface flow. Therefore, it is expected that these components will exhibit also different stochastic characteristics. Figure 5-2 gives an illustration of how the runoff formation is conceived in this study.

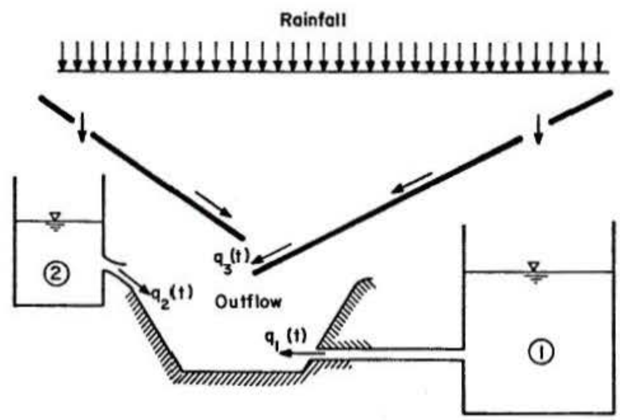


Fig. 5-2. Schematic Representation of Components in Streamflow.

$q_1(t)$ is the outflow from Reservoir No. 1, which simulates the groundwater storage; $q_2(t)$ is the outflow from Reservoir No. 2, which simulates the lumped storages of: (i) surface detention storage, (ii) bank storage and (iii) channel storage; and $q_3(t)$ is the direct runoff, which is composed mainly of the surface runoff and the precipitation over the stream surfaces. Like daily precipitation, daily direct runoff is an intermittent process.

There is no doubt that representing the retention capacity of a watershed by only two reservoirs is an oversimplification of the real situation. However, it is better than assuming the homogeneity of the whole process, as is usually done.

Ideally $q_3(t)$ depends mostly on factors external to a watershed. It can be thought as the *immediate* response of a catchment to the precipitation events. Therefore, it is modeled reasonably well by the methods as developed in Chapter III. However, a serious obstacle must still be removed, namely how to estimate the parameters of the process $q_3(t)$ if no realization of the process is available.

The fact is that only the time series of the total discharge, $q(t)$, is available. There is no way of splitting $q(t)$ into exactly its three components, $q_1(t)$, $q_2(t)$, and $q_3(t)$. A somehow arbitrary assumption is then necessary. It is possible that some modification would lead to a more realistic representation of the phenomena. The assumption is,

$$q_3(t) = \max(0, q(t) - q(t-1)) \tag{5-2}$$

Equation (5-2) says that the direct runoff is either zero or it is equal to the positive increment of the total discharge. In fact, if $q_3(t)>0$ one might expect that the reservoirs are partially replenished on the day t, and therefore it is likely that

$$q_1(t+1) + q_2(t+1) > q_1(t) + q_2(t)$$

or

$$(q_1(t+1) - q_1(t)) + (q_2(t+1) - q_2(t)) > 0 \tag{5-3}$$

Equation (5-2) simply says that the above positive quantity is equal to $q_3(t)$, or that

$$q_3(t) = (q_1(t+1) - q_1(t)) + (q_2(t+1) - q_2(t)),$$

$$\text{for } q_3(t) > 0 \tag{5-4}$$

From Eqs. (5-1) and (5-4) one can see that, whenever $q_3(t) > 0$

$$q_1(t) + q_2(t) = q(t-1) \tag{5-5}$$

Hence any rising limb of the hydrograph, say from day t_o, day t_f can be obtained if the value of $q(t_o)$ as well as of the succession $q_3(t_o),\ldots,q_3(t_f)$ are known. In order to have a rising limb all the values in the succession $q_3(t_o),\ldots,q_3(t_f)$ should be positive.

How to cope with the falling limbs of the hydrographs is the subject of Section 5-3. Next the process $q_3(t)$ is studied in more details.

5-2 Direct Runoff

As already mentioned, the process $q_3(t)$ is modeled according to the technique explained in Chapter III. Therefore, all the tests there explained, as well as the asymptotic covariance matrix, could conceivably be employed to the positive increments of streamflow. However this was not done in this investigation. Rather a simple procedure, demonstrated on data provided by the case study, was used to give a first insight on the potential benefits of this approach.

First, the observed streamflow data of the Powell River was processed following Eq. (5-2) to produce the time series $q_3(t)$. (The same symbol is used for convenience, either for the stochastic process or for the corresponding time series.) The data was divided into 26 seasons, each 14 days long, adding up to 364 days. For each season the parameters μ, σ, ρ, and α were estimated. This is essentially the same as done in Section 4-4, with the difference that there the year started on January 1 and here on October 1. A second difference is that no goodness of fit was tested in Section 4-4 because it was already done in Section 4-2 while focusing on the stationary case. For the $q_3(t)$-process the chi-square goodness-of-fit statistic was computed for each of the 26 marginal distributions (one for each season). The results are shown in Table 5-1. As usual the seasons marked with an asterisk are those which have the goodness-of-fit of the marginal distribution rejected at the 5 percent significance level. Those marked with a triangle are the cases with the rejection also at 1 percent significance level. The number of rejections was high: 7 cases at the 5 percent level and 2 at the 1 percent level. August through November, roughly the Autumn, seems to be the time of the year for which the positive increments were badly fitted by the model. Section 5-4 will reveal that this problem is serious enough to impede a reliable working of the model for this specific season. However it is likely that one will be more concerned on studying the Spring and Summer, rather than the Autumn, due to the timing of the floods. In Section 5-4 it will be shown that for this particular set of data the model can be applied for the whole year except for the Autumn.

Data from Table 5-1 was used to produce the periodic functions that represent the time variation of each one of the parameters, expressed in the general form by Eq. (4-6). The criterion for deciding which harmonics are relevant was described in Section 4-4. The results are summarized in Table 5-2. Plots of periodic functions μ_τ, σ_τ, ρ_τ, and α_τ for the daily flows of the Powell River are shown in Figures 5-3 and 5-4.

5-3 Outflow from the Watershed Storage

It was seen in Section 5-1 that, according to the proposed model, any falling limb of a hydrograph is the result of emptying the two reservoirs. The hydrograph values decrease only when $q_3(t) = 0$. Hence, the hydrograph recession curve is nearly independent of the characteristics of storm which causes the hydrograph rise. Only the states of the reservoirs, as well as their operating rules are relevant for this analysis. The description of reservoirs is then needed. It is assumed that both reservoirs are linear, meaning that the output $q_i(t)$, i = 1 and 2, is proportional to the storage $S_i(t)$. Or

$$q_i(t) = K_i S_i(t) \ , \ (i = 1 \text{ and } 2) \tag{5-6}$$

During the hydrograph recession part the input to reservoirs is zero with the continuity equation expressed in the simple form as

$$q_i(t) = \frac{-dS_i(t)}{dt} \ , \ (i = 1 \text{ and } 2) \tag{5-7}$$

If Eq. (5-6) is differentiated with respect to time t and then Eq. (5-6) used,

$$\frac{dq_i(t)}{dt} = -K_i q_i(t) \ , \ (i = 1 \text{ and } 2)$$

or

$$\frac{dq_i(t)}{q_i(t)} = -K_i dt \ , \ (i = 1 \text{ and } 2) \tag{5-8}$$

Table 5-1. Results for Goodness-of-Fit Statistics for the 26 Seasons of Daily Flows of the Powell River

Period	From-To	$\bar{\mu}$	σ	ρ	α	X^2(d.f.)
1	1 Oct-14 Oct	3.6296	6.9689	.2413	.4312	8.01(1)Δ
2	15 Oct-28 Oct	3.2556	6.8980	.7113	.4098	10.57(4)*
3	29 Oct-11 Nov	3.3940	7.1673	.6098	.3883	8.76(3)*
4	12 Nov-25 Nov	4.3443	9.8158	.5737	.3848	5.33(6)
5	26 Nov- 9 Dec	4.2393	13.5182	.6620	.4123	14.74(9)
6	10 Dec-23 Dec	11.2918	22.7644	.6352	.4500	6.19(11)
7	24 Dec- 6 Jan	13.5328	37.9861	.6001	.5197	20.54(13)
8	7 Jan-20 Jan	15.2014	37.4007	.6905	.5067	18.57(14)
9	21 Jan- 3 Feb	19.3112	43.0868	.5325	.5038	12.78(14)
10	4 Feb-17 Feb	28.4365	67.6299	.5148	.5626	17.57(16)
11	18 Feb- 3 Mar	36.3807	68.4997	.6526	.5623	13.55(14)
12	4 Mar-17 Mar	43.4370	96.6109	.6247	.6158	14.82(15)
13	18 Mar-31 Mar	29.4476	57.0287	.5679	.5421	10.52(14)
14	1 Apr-14 Apr	39.4548	60.2830	.5625	.5800	12.72(9)
15	15 Apr-28 Apr	35.1393	52.9360	.5672	.5483	14.30(10)
16	29 Apr-12 May	21.0269	35.0319	.6524	.5284	12.63(9)
17	13 May-26 May	19.7310	27.8554	.5145	.4901	7.00(7)
18	27 May- 9 Jun	14.4111	20.8414	.5221	.4903	14.14(8)
19	10 Jun-23 Jun	12.2671	26.4863	.4879	.5550	14.32(9)
20	24 Jun- 7 Jul	7.0326	14.4152	.3834	.4371	18.79(10)*
21	8 Jul-21 Jul	8.0993	18.3048	.4296	.4800	14.91(10)
22	22 Jul- 4 Aug	8.9538	17.0723	.2473	.5132	9.86(6)
23	5 Aug-18 Aug	6.7849	15.0933	.4047	.4694	24.88(9)Δ
24	19 Aug- 1 Sep	10.3664	15.5428	.5306	.4953	8.70(5)
25	2 Sep-15 Sep	7.2767	11.7339	.3901	.4904	9.17(3)*
26	16 Sep-29 Sep	7.6774	10.2326	.3397	.4761	7.30(2)*

*The test is rejected at the 5% significance level.
ΔThe test is rejected at the 1% significance level.

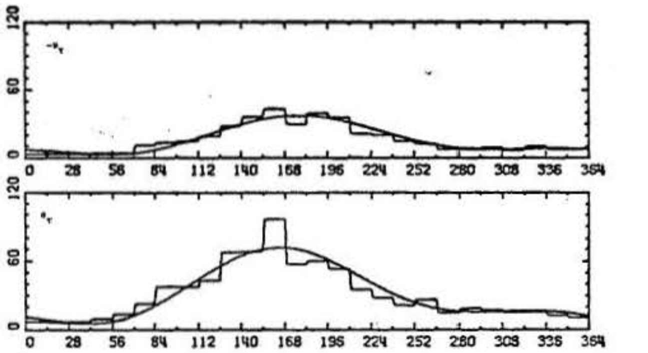


Fig. 5-3. The Periodic μ_τ and σ_τ for Daily Values of the Powell River.

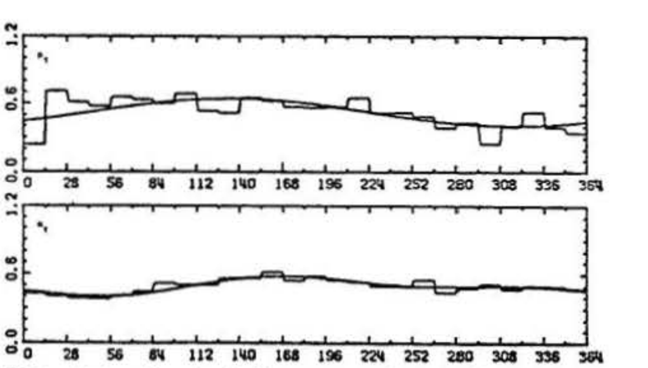


Fig. 5-4. The Periodic ρ_τ and α_τ for Daily Values of the Powell River.

Integrating Eq. (5-8) between 0 and t yields

$$\ln \frac{q_i(t)}{q_i(0)} = -K_i t \ , \ (i = 1 \text{ and } 2)$$

or

$$q_i(t) = q_i(0)e^{-K_i t} \tag{5-9}$$

Equation (5-9) is the well known exponential recession curve. It is obvious that the outflow discharge from the ith linear reservoir, during a recession period, depends only on the initial discharge $q_i(0)$ and on the reservoir characteristic K_i. Therefore any recession curve can be expressed by

$$q(t) = q_1(0)e^{-K_1 t} + q_2(0)e^{-K_2 t} \ , \ t \leq \ell \tag{5-10}$$

where for convenience t=0 indicates the beginning of the recession curve, and ℓ is the length of the recession considered.

For

$$W = \frac{q_1(0)}{q(0)} \tag{5-11}$$

Equation (5-10) may be rewritten as

$$q(t) = q(0)[W e^{-K_1 t} + (1-W)\ e^{-K_2 t}] \tag{5-12}$$

or for $\gamma_1 = e^{-K_1}$ and $\gamma_2 = e^{-K_2}$

$$q(t) = q(0)[W\gamma_1{}^t + (1-W)\ \gamma_2{}^t] \tag{5-13}$$

K_1 and K_2 are constants that must be estimated. On the other hand, W indicates how the maintenance of the hydrograph is split between the two reservoirs, after a storm has occurred. Since the initial states of reservoirs are expected to vary from one recession curve to another, W cannot be conceived as a constant; rather its visualization as a random variable seems feasible. Therefore, in order to use Eq. (5-12) in the generation of new samples, not only the values of K_1 and K_2 must be known but also the probability distribution of W, with q(0) always known.

It is reasonable to estimate K_1 and K_2 in such a way that the *theoretical recession curves* will resemble the observed recession curves. In the more specific terms, the estimation of K_1 and K_2 should be taken in the framework of the following optimization problem:

$$\min_{K_1, K_2} \sum_{r=1}^{n} \sum_{t=1}^{\ell(r)} \{q'(t,r) - q'(0,r)[w(r)e^{-K_1 t} + (1-w(r))e^{-K_2 t}]\}^2 \tag{5-14}$$

where $\ell(r)$ is the length of the rth recession curve; n is the number of recession curves in the historic data; $q'(t,r)$ is the observed discharge on the t-th day of the rth recession curve; and $w(r)$ is the outcome of the random variable W, associated with the rth recession curve.

For any pair (K_1, K_2) the objective function of Eq. (5-14) can only be evaluated if the outcomes $w(r)$, r = 1, 2, ..., are known. Again, it is reasonable to assume that each $w(r)$ is such that the differences between the rth theoretical and the observed recession curve values are minimized. By this reasoning, each $w(r)$ can be found by solving the equation

$$\frac{\partial}{\partial\, w(r)} \left[\sum_{t=1}^{l(r)} \{q'(t,r)-q'(0,r)[w(r)e^{-K_1 t} + (1-w(r))e^{-K_2 t}]\}\right] = 0 \qquad (5\text{-}15)$$

or

$$w(r) = \frac{\sum_{t=1}^{l(r)} q'(t,r) - q(0,r)\sum_{t=1}^{l(r)} e^{-K_2 t}}{q(0,r)\sum_{t=1}^{l(r)} (e^{-K_1 t} - e^{-K_2 t})} \qquad (5\text{-}16)$$

Several numerical algorithms are available for solving the optimization problem defined by Eq. (5-14). Among them is the Rosen Algorithm, as a quite convenient one. It is a *mountain climbing* type of technique, based on the gradient projection method. A detailed description of the algorithm is given by Kuester et al. (1973). Here it is sufficient to say that the only requirements for the algorithm are: (i) the objective function, which is given by Eq. (5-14); (ii) the first-derivatives of the objective function, which can be obtained by a proper use of Eq. (5-14); and (iii) the linear constraints, given as,

$$0 < \gamma_1 < 1 \quad \text{or} \quad 0 < K_1 < \infty$$

and

$$0 < \gamma_2 < 1 \quad \text{or} \quad 0 < K_2 < \infty \qquad (5\text{-}17)$$

Attention is called to the fact that each time the value of the pair (K_1, K_2) is changed, the observations $w(r)$ are reassessed by using Eq. (5-16). Also, one should expect from the way the conceptual model was set that $\gamma_1 > \gamma_2$ (or $K_1 < K_2$), although this does not constitute a constraint.

For the Powell River daily flow data, the application of the algorithm yields

$$\left.\begin{aligned} \gamma_1 &= 0.8971 \rightarrow K_1 = 0.1086/\text{day} \rightarrow \frac{1}{K_1} = 9.2091 \text{ days} \\ \gamma_2 &= 0.5029 \rightarrow K_2 = 0.6874/\text{day} \rightarrow \frac{1}{K_2} = 1.4548 \text{ days} \end{aligned}\right\} \qquad (5\text{-}18)$$

It is of interest to check how the theoretical recession functions obtained by the above procedure, fit their observed counterparts. Figure 5-5 gives this visual comparison for the recession curves of the daily flow series of the Powell River during the year 1921 for recessions which were longer than four days. This choice is an arbitrary selection, imposed by the practical difficulty of plotting all the recessions registered in 40 years. Attention is called to the fact that in general the curves would not be well fitted by straight lines. This means that the representation of the watershed storage by a *single* linear reservoir would not be appropriate.

Once the values K_1 and K_2 are estimated the next problem is how to statistically describe the random variable W. The set of outcomes of this variable is simultaneously obtained with K_1 and K_2. In principle, one might expect any outcome w to lie between 0 and 1. A value of $w > 1$ would indicate a reversion of the direction of flow related to the second reservoir. Analogously $w < 0$ would indicate a reversion of the direction of flow coming from the first reservoir. These flow reversions are anticipated to be rare, but when one of them does occur, it is necessary to assert the rules which govern the inflow hydrographs, rather than the outflow hydrograpns. This leads to the assumption that the characteristics of flow either from the reservoir to outlet o: from the outlet to the reservoir are identical.

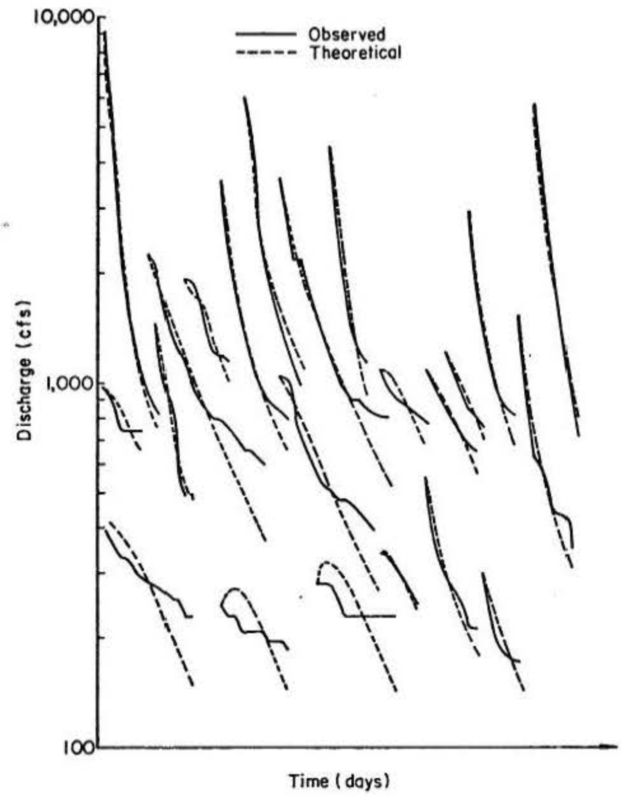


Fig. 5-5. Comparison Between Theoretical and Observed Recession Curves, of Daily Flow Series of the Powell River, for the Year 1921.

Qualitatively, one might expect $E[W|q(0)]$ to be small whenever the initial discharge $q(0)$ is large. Indeed high flows are associated with high retention in the storages that the second reservoir is supposed to represent. Consequently, its share of the flow supply should be higher initially than the flow supply which corresponds to the first reservoir. The first reservoir is characterized by a high storage capacity, which makes its contribution, $q_1(t)$, reasonable stable. Whenever the initial discharge is small, it is likely that the total flow will be sustained entirely by the outflow from the first reservoir, i.e.,

$$\lim_{q(0)\to 0} E[W|q(0)] = 1 \ .$$

A mathematical representation that fits the above qualitative descriptions is given by

$$E[W|q(0)] = e^{-\Psi q(0)} \ , \ \Psi > 0 \qquad (5\text{-}19)$$

For each historical recession curve one pair of values $[q(0,r), w(r)]$ is available, where r stands for the rth recession. These pairs can then be used to estimate the value of Ψ, by the least squares method. For the daily flow sequences of the Powell River, the value of Ψ is 0.000160. The coefficient of correlation between $q(0)$ and $\log w$ is -0.6737.

In general the random variable W might be expressed by

$$W = e^{-\Psi q(0)} + Z \tag{5-20}$$

where Z is another random variable.

For each recession the corresponding outcomes of Z can be obtained by solving Eq. (5-20) for Z. In case of the daily flow series of the Powell River,

$$z(r) = w(r) - \exp(-0.000160\ q(0,r)),\ r = 1, 2, \ldots, n \tag{5-21}$$

The next thing to do is to test whether the sample of Z may be considered as drawn from a normal probability distribution. This was tested for the daily flows of the Powell River. The chi-square goodness-of-fit test statistic is 42.70, with 36 degrees of freedom. Therefore, the hypothesis of normality could not be rejected at the 5 percent significance level. The sample mean and standard deviation of Z are 0.07334 and 0.25604, respectively. With these last estimates and test, one can then generate the new series.

5-4 Testing the Model

The utility of the model depends on its capacity to generate new series, to be considered the outcomes of the same stochastic process from which the historic series is observed. In Section 4-5, a way of comparing the properties of historic series with the properties of generated series was presented. Here a similar comparison is given for the model of daily flow series, as applied to the Powell River series.

The generation procedure is performed in the following steps.

Step I: Generate the intermittent process $q_3(t)$. This is accomplished by following the procedure explained in Section 4-4. The parameters used are given in Table 5-2.

Step II: Select a value of the discharge for the beginning of new samples. The mean discharge is a good choice for this value.

Step III: Generate for each day, according to:

(a) If $q_3(t) > 0$, take Step III(b); otherwise, go to Step III(c);

(b) Make $q(t) = q(t-1) + q_3(t)$, and go back to Step III(a);

(c) If $q_3(t-1) > 0$, go to Step III(d); otherwise, go to Step III(e);

(d) Find $E[W|q(t-1)] = e^{-0.00016q(t-1)}$. Sample from the normal distribution a value for z. For the Powell River Z comes out as (N(0.07334, 0.06556). Then find w by Eq.(5-20), and define $\delta_1 = wq(t-1)$ and $\delta_2 = (1-w)q(t-1)$;

(e) Make $\xi_1 = \gamma_1\delta_1$ and $\xi_2 = \gamma_2\delta_2$, so that $q(t) = \xi_1+\xi_2$. Go to Step III(f).

(f) Make $\delta_1 = \xi_1$ and $\delta_2 = \xi_2$, and go back to Step III(a).

The above step-by-step procedure was used to generate 40 years of data for daily flows of the Powell River. The hydrograph for the first year of the generated series is plotted in Figure 5-6. This particular hydrograph year is given because the first year of the historic record had previously been used. For the model being good, Figures 5-1 and 5-6 show two different realizations of the same hydrologic process. These samples are different, but the *pattern* of the series is expected to be similar. This approach is based on a subjective inference, with the individual assessment whether the hydrograph of Figure 5-6 *looks like* the historic sample of Figure 5-1, in a general hydrologic sense.

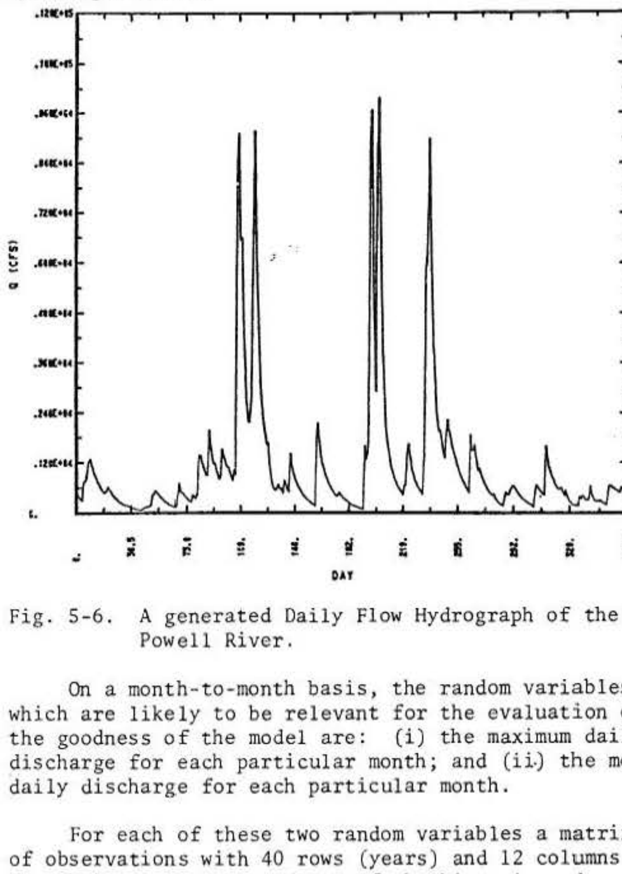


Fig. 5-6. A generated Daily Flow Hydrograph of the Powell River.

On a month-to-month basis, the random variables which are likely to be relevant for the evaluation of the goodness of the model are: (i) the maximum daily discharge for each particular month; and (ii) the mean daily discharge for each particular month.

For each of these two random variables a matrix of observations with 40 rows (years) and 12 columns (months) was constructed out of the historic and generated samples. Let us designate these matrices by $\{F_{ij}\}$; $i = 1, 2, \ldots, 40$; $j = 1, 2, \ldots, 12$. For a month j the sample marginal distributions are available for the historic and generated series. The Smirnov two-sample can then be applied. The critical values at the 5 percent significance level are given by Eq. (4-5); for $g_1 = 40$ and $g_2 = 40$ it is 0.304. The test statistics T.S.7 given by Eq. (4-4) are displayed in the last columns of Tables 5-3 and 5-4. In these tables the values of

$$\bar{F}_j = \frac{1}{40} \sum_{i=1}^{40} F_{ij} \tag{5-22}$$

and

$$\mathrm{std}(F_j) = \sqrt{\frac{\sum_{i=1}^{40} (F_{ij}-\bar{F}_j)^2}{39}} \tag{5-23}$$

are also shown for the historic and generated series, respectively.

Table 5-3. Maximum Daily Flows for Each Month

MONTH	MEAN		STD.DEV.		T.S.7
	HIST.	GEN.	HIST.	GEN.	
October	843.2	758.0	1649.6	568.2	.350*
November	2574.3	833.6	3522.9	990.8	.275
December	5328.4	4615.7	4706.5	5235.4	.200
January	7890.0	12347.2	6190.5	11286.5	.275
February	8615.6	11248.0	5377.4	9995.9	.125
March	7204.2	7888.8	4565.6	7156.4	.150
April	5000.1	6350.9	3270.7	5755.6	.100
May	3772.6	4638.5	3944.9	3164.6	.250
June	2320.4	2487.5	2910.2	2153.4	.225
July	2289.5	1408.7	2324.7	1002.4	.225
August	1501.4	1680.9	1745.8	970.1	.300
September	747.8	1346.8	904.2	838.0	.550*

*The test is rejected at the 5% significance level.

As usual the deviations marked by an asterisk are those of the rejection of the hypothesis of statistical equality of samples, at the 5 percent significance level. Using jointly the results given in the two tables, one can see that the period of time between August and November is characterized by a rejection of the model. Fortunately, the remainder of the year

Table 5-4. Mean Daily Flows for Each Month

MONTH	MEAN		STD.DEV.		T.S.7
	HIST.	GEN.	HIST.	GEN.	
October	235.2	293.8	192.2	233.7	.200
November	584.0	196.4	586.1	160.4	.425*
December	1288.4	1274.9	1053.9	1200.0	.125
January	1981.2	3394.2	1224.6	2390.9	.300
February	2396.9	3186.6	1217.7	2648.6	.175
March	2310.3	2450.3	1136.8	1943.8	.225
April	1612.4	1698.6	753.6	1290.9	.150
May	1122.7	1563.0	786.5	1065.7	.300
June	652.9	880.0	492.1	736.0	.200
July	610.3	550.3	440.2	374.3	.125
August	437.0	707.1	384.7	400.7	.400*
September	241.7	578.8	173.2	352.2	.550*

*The test is rejected at the 5% significance level.

shows the model to be accepted. In Section 5-2, while studying the process $q_3(t)$, it was found that a reasonable fit could not be obtained for the Autumn data. This is likely also the reason for a poor performance of the overall model during this specific season.

Chapter VI

CONCLUSIONS AND RECOMMENDATIONS

Several further research possibilities of the dual streamflow model look promising, such as:

(i) As the direct-runoff $q_3(t)$ is supposed to represent the portion of the input to the watershed which is not retained by any river basin storage, it is likely that the parameters of $q_3(t)$ are strongly related to those that define the precipitation for the area. A joint study of the two processes could yield results valid for regional applications.

(ii) The constants K_1 and K_2, associated with the linear reservoirs, are estimated by an iterative algorithm. They define the operation rules of the two reservoirs. As these two reservoirs conceptually represent the watershed retention capacity, K_1 and K_2 must be related to physiographic characteristics of the catchment. An estimation procedure that could employ this additional information would represent a new dimension in the stochastic hydrologic modeling.

The performance of the developed precipitation model was tested for its goodness. The model is capable of producing such generated samples that resemble the historic series, in a stochastic sense.

The examined data show that the daily precipitation series cannot be assumed to be a sequence of independent events. Therefore, the capability of the model to reproduce the serial dependence of the processes is an essential feature to its good performance.

It is also shown that the model can be used for the generation of simultaneous precipitation series for several dependent-station processes.

The proposed streamflow model has a physically justified basis. It is possible to generate new samples with the complex characteristics of daily streamflow. The intermittent model fairly fits the positive first-differences of daily streamflow. Representation of the recession parts of hydrographs as a stochastic output from the two linear reservoirs was successful.

REFERENCES

Adamowski, K. and Smith, A. F., "Stochastic Generation of Rainfall," Journal of the Hydraulics Division, Proc. Am. Soc. Civil Eng., HY11, pp. 1935-1944, 1972.

Beard, L. R., "Simulation of Daily Streamflow," Technical Paper No. 6, Hydrology Engineering Center, U. S. Army, 1968.

Bradley, J., "Distribution Free Statistical Tests," Prentice-Hall Inc., 1968.

Bras, R. L. and Rodriguez-Iturbe, I., "Rainfall-Runoff as Spatial Stochastic Processes: Data Collection and Synthesis," Ralph M. Parsons Laboratory Report No. 196, Massachusetts Institute of Technology, 1975.

Brown, J. W., Walsh, M. R., McCarley, R. W., Green, A. J., and West, H. W., "Models and Methods Applicable to Corps of Engineers Urban Studies," U. S. Army Engineer Waterways Experiment Station, Hydraulics Laboratory, 1974.

Bryson, M. C., "Heavy-Tailed Distributions: Properties and Tests," Technometrics, Vol. 16, No. 1, pp. 61-68, 1974.

Clarke, R. T., "Mathematical Models in Hydrology," Food and Agriculture Organization (FAO) of the United Nations, Drainage Paper No. 19, 1973.

Cohen, A. C., Jr., "Simplified Estimators for the Normal Distribution When Samples are Singly Censored or Truncated," Technometrics, Vol. 1, No. 3, pp. 217-237, 1959.

Das, S. C., "The Fitting of Truncated Type III Curves to Daily Rainfall Data," Australian Journal of Physics, 7, pp. 298-304, 1955.

Fleming, G., "Computer Simulation Techniques in Hydrology," American Elsevier Publishing Co., 1975.

Franz, D. D., "Hourly Rainfall Generation for a Network," Proc. Symposium on Statistical Hydrology, Tucson, Arizona (1971), USDA Misc. Publ. 1275, pp. 147-153, 1974.

Gabriel, K. R. and Neumann, J., "A Markov Chain Model for Daily Rainfall Occurrence at Tel Aviv," Quarterly Journal Royal Meteorological Society, No. 88, pp. 90-95, 1962.

Grace, R. A. and Eagleson, P. S., "The Synthesis of Short Time Increment Rainfall Sequences," Ralph M. Parsons Laboratory Report No. 91, Massachusetts Institute of Technology, 1966.

Green, J. R., "A Model for Rainfall Occurrence," Journal of Royal Statistical Society, Series B, 26, pp. 345-353, 1964.

Green, N. M. D., "A Synthetic Model for Daily Streamflow," Journal of Hydrology, 20, pp. 351-367, 1973.

Gupta, V. K. and Duckstein, L., "A Stochastic Analysis of Extreme Droughts," Water Resources Research, Vol. 11, No. 2, pp. 221-228, 1975.

Hollander, M. and Proschan, F., "Test for the Mean Residual Life," Biometrika, 62, 3, pp. 585-593, 1975.

Ison, N. T., Feyerherm, A. M., and Bark, L. D., "Wet Period Precipitation and the Gamma Distribution," Journal of Applied Meteorology, Vol. 10, pp. 658-665, 1971.

Kendall, M. G. and Stuart, A., "The Advanced Theory of Statistics," Hafner Publishing Co., 1963.

Khanal, N. N. and Hamrick, R. L., "A Stochastic Model for Daily Rainfall Data Synthesis," Proc. Symposium on Statistical Hydrology, Tucson, Arizona (1971), USDA Misc. Publ. 1275, pp. 197-210, 1974.

Kottegoda, N. T., "Stochastic Five Day Stream Flow Model," Journal of the Hydraulics Division, Proc. Am. Soc. Civil Eng., HY9, pp. 1469-1485, 1972.

Kuester, J. L. and Mize, J. H., "Optimization Techniques with Fortran," McGraw-Hill, 1973.

Lawrance, A. J. and Kottegoda, N. T., "Stochastic Modelling of Hydrologic Series," to appear in the Journal of the Royal Statistical Society, 1976.

Matalas, N. C., "Mathematical Assessment of Synthetic Hydrology," Water Resources Research, Vol. 3, No. 4, pp. 973-945, 1967.

Mielke, P. W., Jr., "Another Family of Distributions for Describing and Analysing Precipitation Data," Journal of Applied Meteorology, Vol. 12, pp. 275-280, 1973.

Mood, A. M., Graybill, F. A., and Boes, D. C., Introduction to the Theory of Statistics," McGraw-Hill, 1974.

Nicks, A. D., "Stochastic Generation of the Occurrence, Pattern, and Location of Maximum Amount of Daily Rainfall," Proc. Symposium on Statistical Hydrology, Tucson, Arizona (1971), USDA Misc. Publ. 1275, pp. 154-171, 1974.

Pattison, A., "Synthesis of Hourly Rainfall Data," Water Resources Research, Vol. 1, No. 4, pp. 489-498, 1965.

Pfaehler, R., discussion on "Duration Curves," by H. A. Foster, Paper No. 1885, Trans. Am. Soc. Civil Eng., pp. 1236-1237, 1933.

Quimpo, R. G., "Stochastic Model of Daily River Flow Sequences," Hydrology Papers, No. 18, Colorado State University, Fort Collins, Colorado, 1967.

Regier, M. H. and Hamdan, M. A., "Correlation in a Bivariate Normal Distribution with Truncation in Both Variables," Australian Journal of Statistics, 13(2), pp. 77-82, 1972.

Rhenals-Figueredo, A. E., Rodriguez-Iturbe, I., and Schaake, J. C., "Bidimensional Spectral Analysis of Rainfall Events," Ralph M. Parsons Laboratory Report No. 193, Massachusetts Institute of Technology, 1974.

Rosenbaum, S., "Moments of a Truncated Bivariate Normal Distribution," Journal of the Royal Statistical Society, Series B, 23, pp. 405-408, 1961.

Saldarriaga, J. and Yevjevich, V., "Application of Run-Lengths to Hydrologic Series," Hydrology Papers, No. 40, Colorado State University, Fort Collins, Colorado, 1970.

Tallis, G. M., "The Moment Generating Function of the Truncated Multi-Normal Distribution," Journal of the Royal Statistical Society, Series B, 23, pp. 233-239, 1961.

Tao, P., "Distribution of Hydrologic Independent Stochastic Components," Ph.D. Dissertation, Colorado State University, Fort Collins, Colorado, 1973, (see also Hydrology Papers, No. 82, Colorado State University, 1976).

Todorovic, P., "A Mathematical Study of Precipitation Phenomena," Technical Report CET67-68T65, Colorado State University, Fort Collins, Colorado, 1968.

Todorovic, P. and Yevjevich, V., "Stochastic Process of Precipitation," Hydrology Papers, No. 35, Colorado State University, Fort Collins, Colorado, 1969.

Todorovic, P. and Woolhiser, D. A., "Stochastic Model of Daily Rainfall," Proc. Symposium on Statistical Hydrology, Tucson, Arizona (1971), USDA Misc. Publ. 1275, pp. 232-246, 1974.

Todorovic, P. and Woolhiser, D. A., "Stochastic Structure of the Local Pattern of Precipitation," to appear in the book Stochastic Approaches to Water Resources, to be published by Water Resources Publications, Fort Collins, Colorado, 1976.

Verschuren, J. P., "A Stochastic Analysis of Precipitation," Ph.D. Dissertation, Colorado State University, Fort Collins, Colorado, 1968.

WIC, "Climates of the States," Water Information Center Inc., Port Washington, New York, 1974.

Wiser, E. H., "Persistence Models in Precipitation Analysis," Ph.D. Dissertation, North Carolina State University at Raleigh, 1964.

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

Yevjevich, V., "Structural Analysis of Hydrologic Time Series," Hydrology Papers, No. 56, Colorado State University, Fort Collins, Colorado, 1972(b).

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

First and Second Derivatives of the Log-Likelihood Function

Define

$$T(\nu,i,j,k) = \sum^{\ell} \left[\frac{\phi(\xi)}{\Phi(\xi)}\right]^{i} \nu^{j^{\alpha}} (\log \nu)^{k} ,$$

where ν is a dummy variable that can represent "x_i", "y_i", x_iy_i", and "z_i"

$$\ell = \begin{cases} n_3, & \text{if } \nu \equiv z \\ n_2, & \text{otherwise} \end{cases}$$

$$\xi = \frac{-\rho\nu^{\alpha}-\mu(1-\rho)}{\sigma\sqrt{1-\rho^2}}$$

and

$$I(i,j,k) = \int_{-\infty}^{\infty} \phi(t)\, t^{i} [\phi(\varepsilon)]^{j} [\Phi(\varepsilon)]^{k}\, dt$$

where

$$\varepsilon = \frac{-\mu-\sqrt{\rho}\,\sigma t}{\sigma\sqrt{1-\rho}}$$

and

$$\beta(i) = I(i,1,1)\left[\frac{2I(0,1,1)}{I(0,0,2)} - \frac{\mu}{\sigma\sqrt{1-\rho}}\right] - \sqrt{\frac{\rho}{1-\rho}}\, I(i+1,1,1) - I(i,2,0) \qquad i = 0,1.$$

$$\frac{\partial LL}{\partial \mu} = \frac{-2n_1}{\sigma\sqrt{1-\rho}} \frac{I(0,1,1)}{I(0,0,2)} + \frac{T(x;0,1,0) + T(y;0,1,0) - 2n_2\mu}{\sigma^2(1+\rho)} - \frac{\sqrt{1-\rho}\, T(z;1,0,0)}{\sigma\sqrt{1+\rho}} + \frac{T(z;0,1,0) - n_3\mu}{\sigma^2}$$

$$\frac{\partial LL}{\partial \sigma} = \frac{2n_1\mu\, I(0,1,1)}{\sigma^2\sqrt{1-\rho}\, I(0,0,2)} - \frac{2n_2 + n_3}{\sigma} + \frac{T(x;0,2,0) + T(y;0,2,0) - 2\rho T(xy;0,1,0)}{\sigma^3(1-\rho^2)} + \frac{2\mu\left[n_2\mu - T(x;0,1,0) - T(y;0,1,0)\right]}{\sigma^3(1+\rho)} + \frac{\rho T(z;1,1,0) + \mu(1-\rho)T(z;1,0,0)}{\sigma^2\sqrt{1-\rho^2}} + \frac{T(z;0,2,0) - 2\mu T(z;0,1,0) + n_3\mu^2}{\sigma^3}$$

$$\frac{\partial LL}{\partial \rho} = \frac{-n_1\mu}{\sigma\sqrt{(1-\rho)^3}} \frac{I(0,1,1)}{I(0,0,2)} - \frac{n_1}{\sqrt{\rho(1-\rho)^3}} \frac{I(1,1,1)}{I(0,0,2)} + \frac{n_2}{(1-\rho^2)} + \frac{\mu\left[n_2\mu - T(x;0,1,0) - T(y;0,1,0)\right]}{\sigma^2(1+\rho)^2} - \frac{T(z;1,1,0)}{\sigma\sqrt{(1-\rho^2)^3}} + \frac{(1+\rho^2)T(xy;0,1,0) - \rho\left[T(x;0,2,0) + T(y;0,2,0)\right]}{\sigma^2(1-\rho^2)^2} + \frac{\mu T(z;1,0,0)}{\sigma\sqrt{1-\rho^2}\,(1+\rho)}$$

$$\frac{\partial LL}{\partial \alpha} = \frac{\mu(1-\rho)\left[T(x;0,1,1) + T(y;0,1,1)\right] - T(x;0,2,1) - T(y;0,2,1) + \rho T(xy;0,1,1)}{(1-\rho^2)\sigma^2} + \frac{\mu T(z,0,1,1) - T(z;0,2,1)}{\sigma^2} - \frac{\rho T(z;1,1,1)}{\sigma\sqrt{1-\rho^2}} + \frac{2n_2 + n_3}{\alpha} + T(xy;0,0,1) + T(z,0,0,1)$$

$$\frac{\partial^2 LL}{\partial \mu^2} = \frac{-2n_1}{\sigma^2(1-\rho)} \frac{\beta(0)}{I(0,0,2)} \quad \frac{-2n_2}{\sigma^2(1+\rho)} \quad \frac{-n_3}{\sigma^2} + \frac{\rho\sqrt{1-\rho}}{\sigma^3(1+\rho)^{3/2}} T(z;1,1,0) + \sqrt{\left(\frac{1-\rho}{1+\rho}\right)^3} \frac{\mu}{\sigma^3} T(z;1,0,0) - \frac{(1-\rho)}{\sigma^2(1+\rho)} T(z;2,0,0)$$

$$\frac{\partial^2 LL}{\partial \mu \partial \sigma} = \frac{2n_1}{\sigma^3(1-\rho)} \frac{\beta(0)}{I(0,0,2)} + \frac{2n_1}{\sigma^2\sqrt{1-\rho}} \frac{I(0,1,1)}{I(0,0,2)} + \frac{2\left[2n_2\mu - T(x;0,1,0) - T(y;0,1,0)\right]}{\sigma^3(1+\rho)} + \frac{2\left[n_3\mu - T(z;0,1,0)\right]}{\sigma^3} + \frac{\sqrt{1-\rho}\left[(1+\rho)\sigma^2 - (1-\rho)\mu^2\right]T(z;1,0,0) - 2\mu\rho\sqrt{1-\rho}\,T(z;1,1,0)}{\sigma^4(1+\rho)^{3/2}} + \frac{\mu(1-\rho)T(z;2,0,0)}{\sigma^3(1+\rho)} - \frac{\rho^2\, T(z;1,2,0)}{\sigma^4\sqrt{(1+\rho)^3(1-\rho)}} + \frac{\rho T(z;2,1,0)}{\sigma^3(1+\rho)}$$

$$\frac{\partial^2 LL}{\partial\mu\partial\rho} = \frac{-n_1}{\sigma(1-\rho)^{3/2}}\,\frac{I(0,1,1)}{I(0,0,2)}\left[\frac{\mu\beta(0)}{\sigma} + \frac{\beta(1)}{\sqrt{\rho}}\right]$$

$$- \frac{n_1}{\sigma(1-\rho)^{3/2}}\,\frac{I(0,1,1)}{I(0,0,2)}$$

$$+ \frac{2n_2\mu - T(x;0,1,0) - T(y;0,1,0)}{\sigma^2(1+\rho)^2}$$

$$+ \frac{[\sigma^2(1+\rho) - \mu^2(1+\rho)]T(z;1,0,0)}{\sigma^3\sqrt{(1+\rho)^5(1-\rho)}}$$

$$+ \sqrt{\frac{(1-\rho)}{(1+\rho)^5}}\,\frac{\mu}{\sigma^3}\,T(z;1,1,0) + \frac{\rho\, T(z;1,2,0)}{\sigma^3\sqrt{(1+\rho)^5(1-\rho)^3}}$$

$$+ \frac{\mu T(z;2,0,0)}{\sigma^2(1+\rho)^2} - \frac{T(z;2,1,0)}{\sigma^2(1+\rho)^2(1-\rho)}$$

$$\frac{\partial^2 LL}{\partial\mu\partial\alpha} = \frac{T(x;0,1,1)+T(y;0,1,1)}{(1+\rho)\sigma^2} + \frac{T(z;0,1,1)}{\sigma^2}$$

$$+ \frac{\rho[\rho T(z;1,2,1) + \mu(1-\rho)T(z;1,1,1)]}{\sigma^3\sqrt{(1+\rho)^3(1-\rho)}} - \frac{T(z;2,1,1)}{\sigma^2(1+\rho)}$$

$$\frac{\partial^2 LL}{\partial\sigma^2} = \frac{-2n_1\mu^2}{\sigma^4(1-\rho)}\,\frac{\beta(0)}{I(0,0,2)} - \frac{4n_1\mu}{\sigma^3\sqrt{1-\rho}}\,\frac{I(0,1,1)}{I(0,0,2)} + \frac{2n_2+n_3}{\sigma^2}$$

$$+ \frac{3[2\rho T(xy;0,1,0) - T(x;0,2,0) - T(y;0,2,0)]}{\sigma^4(1-\rho^2)}$$

$$+ \frac{3[2\mu T(z;0,1,0) - T(z;0,2,0) - \rho_3\mu^2]}{\sigma^4}$$

$$+ \frac{6\mu[T(x;0,1,0)+T(y;0,1,0)-n_2\mu]}{\sigma^4(1+\rho)}$$

$$+ \frac{\sqrt{1-\rho}\,[\mu^3(1-\rho) - 2\mu\sigma^2(1+\rho)\ \ T(z;1,0,0)]}{\sigma^5\sqrt{(1+\rho)^3}} + \frac{\rho^3 T(z;1,3,0)}{\sigma^5\sqrt{(1-\rho^2)^3}}$$

$$+ \frac{\rho[3\mu^2(1-\rho) - 2\sigma^2(1+\rho)]T(z;1,1,0) + 3\mu\rho^2 T(z;1,2,0)}{\sigma^5\sqrt{(1+\rho)^3(1-\rho)}}$$

$$- \frac{\mu[(1-\rho)\mu T(z;2,0,0) + 2\rho T(z;2,1,0)]}{\sigma^4(1+\rho)} - \frac{\rho^2 T(z;2,2,0)}{\sigma^4(1-\rho^2)}$$

$$\frac{\partial^2 LL}{\partial\sigma\partial\rho} =$$

$$\frac{n_1\mu}{\sigma^2(1-\rho)^2\ I(0,0,2)}\left[\frac{\mu\beta(0)}{\sigma} + \frac{\beta(1)}{\sqrt{\rho}}\right] + \frac{n_1\mu}{\sigma^2\sqrt{(1-\rho)^3}}\,\frac{I(0,1,1)}{I(0,0,2)}$$

$$+ \frac{2\{\rho[T(x;0,2,0)+T(y;0,2,0)] - (1+\rho^2)T(xy;0,1,0)\}}{\sigma^3(1-\rho^2)^2}$$

$$+ \frac{2\mu[T(x;0,1,0)+T(y;0,1,0)-n_2\mu]}{\sigma^3(1+\rho)^2}$$

$$+ \frac{\mu[\mu^2(1-\rho) - \sigma^2(1+\rho)]\,T(z;1,0,0)}{\sigma^4(1+\rho)^2\sqrt{1-\rho^2}}$$

$$+ \frac{[\sigma^2(1+\rho) - \mu^2(2\rho^2-3\rho+1)]\,T(z;1,1,0) - \mu\rho(2-\rho)T(z;1,2,0)}{\sigma^4\sqrt{(1+\rho)^5(1-\rho)^3}}$$

$$+ \frac{\mu[T(z;2,1,0) - \mu T(z;2,0,0)]}{\sigma^3(1+\rho)^2} - \frac{\rho^2 T(z;1,3,0)}{\sigma^4\sqrt{(1-\rho^2)^5}} + \frac{\rho T(z;2,2,0)}{\sigma^3\sqrt{(1-\rho^2)^2}}$$

$$\frac{\partial^2 LL}{\partial\sigma\partial\alpha} = \frac{-2\{\mu(1-\rho)[T(x;0,1,1)+T(y,0,1,1)] - T(x;0,2,1) - T(y;0,2,1) + \rho T(xy;0,1,1)\}}{(1-\rho^2)\sigma^3}$$

$$+ \frac{2[T(z;0,2,1) - \mu T(z;0,1,1)]}{\sigma^3}$$

$$+ \frac{\rho[\sigma^2(1+\rho) - \mu^2(1-\rho)]\,T(z;1,1,1)}{\sigma^4\sqrt{(1+\rho)^3(1-\rho)}}$$

$$+ \frac{\rho[\rho T(z;2,2,1) + \mu(1-\rho)T(z;2,1,1)]}{\sigma^3(1-\rho^2)}$$

$$- \frac{\rho^2[\rho T(z;1,3,1) + 2\mu(1-\rho)T(z;1,2,1)]}{\sigma^4\sqrt{(1-\rho^2)^3}}$$

$$\frac{\partial^2 LL}{\partial\rho^2} = \frac{-n_1\mu}{2\sigma(1-\rho)^3\ I(0,0,2)}\left[\frac{\mu\beta(0)}{\sigma} + \frac{\beta(1)}{\sqrt{\rho}}\right]$$

$$- \frac{-n_1}{2\sqrt{\rho}\,(1-\rho)^3\ I(0,0,2)}\left[\frac{\mu\beta(1)}{\sigma} + \frac{1}{\sqrt{\rho}}\left[\frac{2[I(1,1,1)]^2}{I(0,0,2)} - \frac{\mu I(2,1,1)}{\sigma\sqrt{1-\rho}} - \frac{\rho}{1-\rho}\,I(3,1,1) - I(2,2,0)\right]\right]$$

$$- \frac{3n_1\mu}{2\sigma\sqrt{(1-\rho)^5}}\,\frac{I(0,1,1)}{I(0,0,2)} + \frac{n_1\ (1-4\rho)}{2\sqrt{\rho^3}\,(1-\rho)^5}\,\frac{I(1,1,1)}{I(0,0,2)}$$

$$+ \frac{n_2(1+\rho^2)}{(1-\rho^2)^2} + \frac{2\mu[T(x;0,1,0)+T(y;0,1,0)-n_2\mu]}{\sigma^2(1+\rho)^3}$$

$$+ \frac{2\rho(3+\rho^2)T(xy;0,1,0) - (1+3\rho^2)[T(x;0,2,0)+T(y;0,2,0)]}{\sigma^2(1-\rho^2)^3}$$

$$+\frac{\mu[\sigma^2(2\rho-1)(1+\rho)^2+\mu^2(1-\rho)]T(z;1,0,0)}{\sigma^3\sqrt{(1+\rho)^7(1-\rho)^3}}$$

$$+\frac{\mu(1-2\rho)T(z;1,2,0)-[3\sigma^2\rho(1+\rho)+\mu^2(1-\rho)(2-\rho)]T(z;1,1,0)}{\sigma^3\sqrt{(1+\rho)^7(1-\rho)^5}}$$

$$-\frac{\mu^2T(z;2,0,0)}{\sigma^2(1+\rho)^3(1-\rho)}+\frac{2\mu T(z;2,1,0)}{\sigma^2(1+\rho)^3(1-\rho)^2}+\frac{\rho T(z;1,3,0)}{\sigma^3\sqrt{(1-\rho^2)^7}}$$

$$-\frac{T(z;2,2,0)}{\sigma^2\sqrt{(1-\rho^2)^3}}$$

$$\frac{\partial^2LL}{\partial\rho\partial\alpha}=\frac{-\mu[T(x;0,1,1)+T(y;0,1,1)]}{\sigma^2(1+\rho)^2}+\frac{(1+\rho^2)T(xy;0,1,1)}{\sigma^2(1-\rho^2)^2}$$

$$-\frac{2\rho[T(x;0,2,1)+T(y;0,2,1)]}{\sigma^2(1-\rho^2)^2}$$

$$-\frac{[\sigma^2(1+\rho)+\mu^2\rho(1-\rho)]T(z;1,1,1)}{\sigma^3\sqrt{(1+\rho)^5(1-\rho)^3}}+\frac{\rho^2T(z;1,3,1)}{\sigma^3\sqrt{(1-\rho^2)^5}}$$

$$+\frac{\rho[\mu(1-\rho)T(z;2,1,1)-T(z;2,2,1)]}{\sigma^2(1-\rho^2)^2}+\frac{\rho\mu T(z;1,2,1)}{\sigma^3\sqrt{(1+\rho)^5(1-\rho)}}$$

$$\frac{\partial^2LL}{\partial\alpha^2}=\frac{\mu[T(x;0,1,2)+T(y;0,1,2)]}{\sigma^2(1+\rho)}$$

$$+\frac{\rho T(xy;0,1,2)-2[T(x;0,2,2)+T(y;0,2,2)]}{\sigma^2(1-\rho^2)}$$

$$+\frac{\mu T(z;0,1,2)-2T(z;0,2,2)}{\sigma^2}-\frac{\rho^2T(z;2,2,2)}{\sigma^2(1-\rho^2)}$$

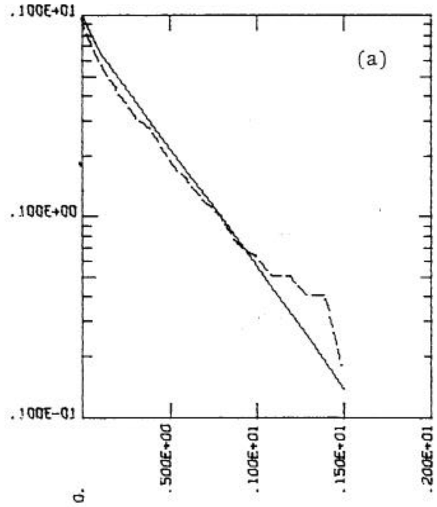
$$+\frac{\rho^2[\rho T(z;1,3,2)+\mu(1-\rho)T(z;1,2,2)]}{\sigma^3\sqrt{(1-\rho^2)^3}}-\frac{2n_2+n_3}{\alpha^2}$$

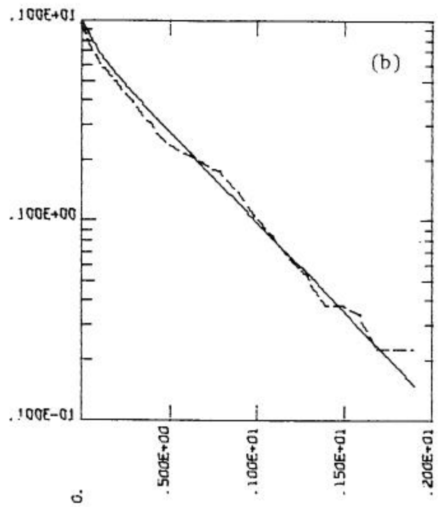
Appendix B-1

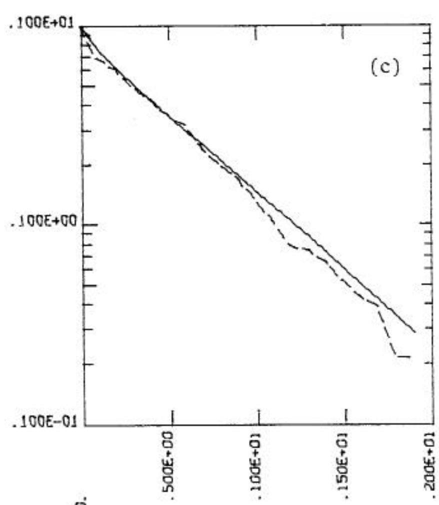
Results of the Application of the Model to the Kansas City Precipitation Series

Tablę B-1-1. Results Obtained in Case the Year is Divided in Twelve Seasons, for the Kansas City Station

PERIOD	PARAMETERS				ASYMPTOTIC COVARIANCE MATRIX				T.S.1	T.S.2	T.S.3	T.S.4	T.S.5
	μ	σ	ρ	α	($X\ 10^{-6}$)				(d.f.)	(1d.f.)	(4d.f.)	(4d.f.)	N(0,1)
001-032	-.4441	.5470	.3792	.6747	2770	-1955 2110	- 562 803 6788	950 -1278 - 190 2448	10.670 (5)	19.423	10.568	1.136	-4.556
033-060	-.3786	.5431	.3085	.6695	2273	-1582 1884	- 367 650 7492	724 -1166 - 157 2398	10.465 (5)	12.583	4.533		-4.387
061-092	-.3330	.6636	.3864	.6133	1983	-1089 1502	- 294 638 4607	97 - 428 - 51 1408	21.586 (9)	29.020	8.553	1.062	-7.038
093-120	-.2658	.6082	.3218	.6838	1692	- 915 1369	- 219 543 5465	109 - 567 - 45 1876	10.649 (8)	17.755	1.979	/	-3.894
121-152	-.2726	.6982	.2045	.6478	1736	- 904 1322	- 111 329 5160	- 27 - 267 - 10 1385	11.040 (11)	7.991	7.534	4.707	-3.888
153-180	-.3091	.8008	.1938	.6870	2594	-1349 1902	- 143 409 5922	- 147 - 69 - 4 1761	16.252 (12)	6.536	.489		-1.213
181-212	-.5109	.9052	.2613	.6753	3951	-2271 2741	- 324 652 5885	- 138 - 35 - 14 1748	8.247 (12)	11.112	.784	5.433	-1.950
213-240	-.4765	.7886	.3720	.6664	3740	-2130 2579	- 437 902 5984	73 - 391 - 68 2078	6.026 (10)	20.713	4.547		-2.640
241-272	-.6562	.9435	.3162	.6131	5480	-3245 3528	- 474 883 6373	- 106 - 55 - 33 1669	17.592 (11)	15.034	1.988	7.135	-6.422
273-300	-.7926	.9314	.4489	.6934	8716	-5310 5204	-1135 1627 7001	176 - 423 - 125 2948	10.477 (8)	23.827	2.983		-2.368
301-332	-.6288	.7389	.2785	.5655	4815	-3117 3085	- 513 779 8215	570 - 782 - 98 1719	11.950 (6)	9.310	2.267	1.563	-8.177
333-360	-.4955	.5818	.4570	.6962	3895	-2685 2803	- 836 1165 6899	1224 -1577 - 285 3163	6.931 (5)	25.057	.734		-2.933







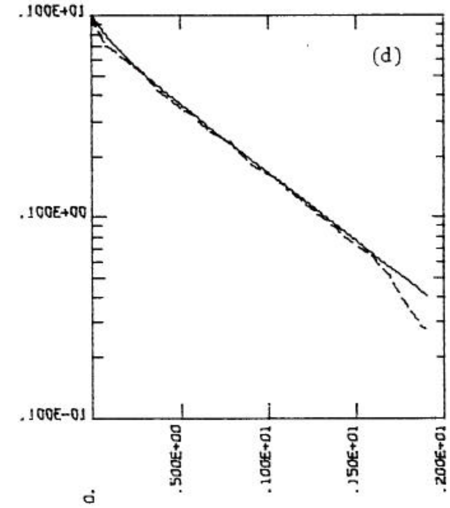


Fig. B-1-1. Plot of Log $\overline{F}(x)$ Versus x for Kansas City Data: (a) Period 93-120, (b) Period 121-152, (c) Period 153-180, and (d) Period 181-212.

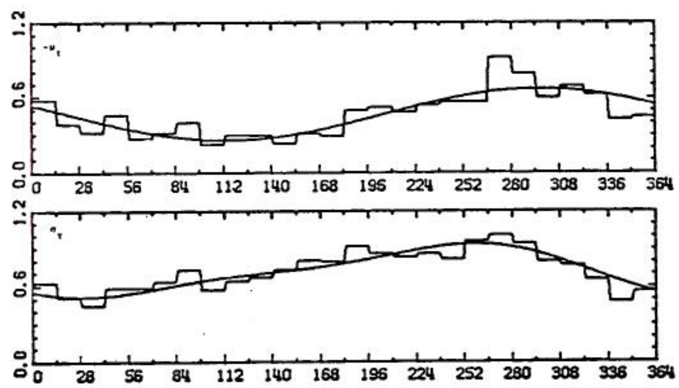


Fig. B-1-2.

The Periodic μ_τ and σ_τ for Daily Values of the Kansas City Precipitation Station.

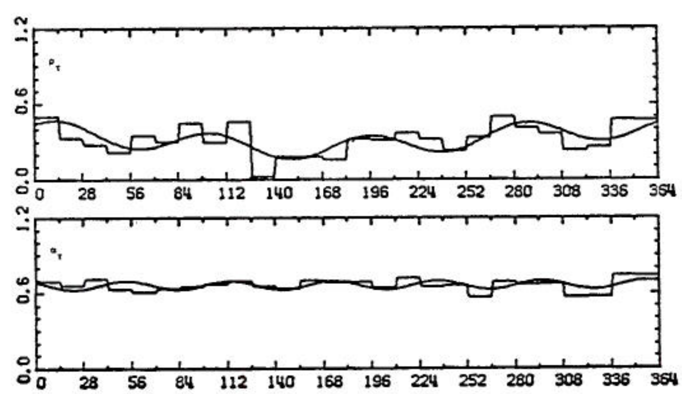


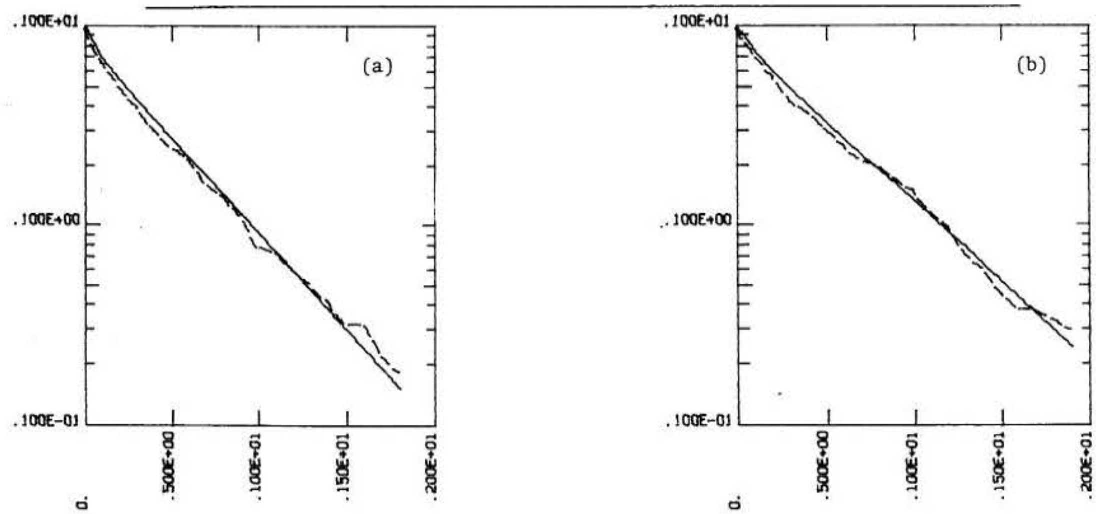
Fig. B-1-3. The Periodic ρ_τ and α_τ for Daily Values of the Kansas City Precipitation Station.

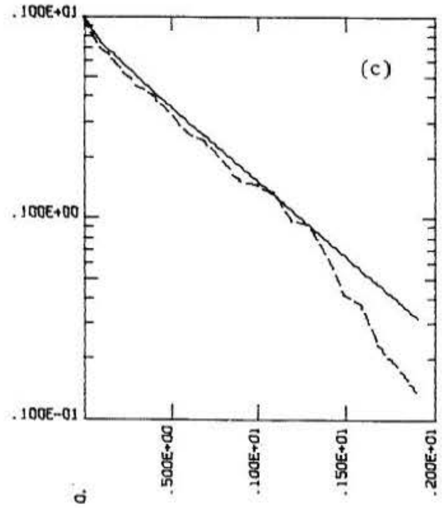
Appendix B-2

Results of the Application of the Model
to the Springfield Precipitation Series

Table B-2-1. Results Obtained in Case the Year is Divided in Twelve Seasons, for the Springfield Station

PERIOD	PARAMETERS μ	σ	ρ	α	ASYMPTOTIC COVARIANCE MATRIX (X 10^{-6})				T.S.1 (d.f.)	T.S.2 (1d.f.)	T.S.3 (4d.f.)	T.S.4 (4d.f.)	T.S.5 N(0,1)
001-032	-.4192	.6168	.3896	.5634	2436	-1561 1863	- 425 747 5590	391 - 706 - 98 1434	5.035 (6)	24.474	5.015	9.149	-8.527
032-060	-.3157	.5917	.3459	.5871	1906	-1133 1559	- 293 626 5831	233 - 640 - 64 1524	10.794 (7)	18.955	9.230		-7.294
061-092	-.2571	.6006	.2712	.6152	1395	- 761 1149	- 151 395 5038	85 - 449 - 27 1312	16.871 (8)	14.343	2.831	4.490	-5.982
093-120	-.3080	.6941	.2885	.6704	2182	-1155 1635	- 206 547 5708	19 - 390 - 31 1798	6.327 (9)	13.854	5.717		-3.237
121-152	-.2838	.7643	.2510	.6722	2071	-1042 1515	- 153 438 4871	- 103 - 131 - 12 1465	17.041 (12)	12.660	2.041	4.913	-2.976
153-180	-.3748	.8293	.2893	.6797	3180	-1692 2256	- 293 675 5799	- 136 - 99 - 13 1829	15.006 (11)	13.647	3.714		-2.397
181-212	-.4639	.8136	.2748	.6122	3197	-1821 2222	- 267 596 5816	- 7 - 217 - 26 1462	13.923 (10)	12.340	3.755	6.231	-6.464
213-240	-.4605	.7334	.2317	.6094	3289	-1991 2347	- 270 558 7427	198 - 518 - 42 1751	5.812 (8)	7.324	3.228		-5.787
241-272	-.5317	.8373	.3968	.5987	3938	-2252 2632	- 477 909 5201	18 - 242 - 52 1515	12.114 (10)	26.198	5.217	5.632	-5.714
273-300	-.7495	.9162	.2436	.6227	7525	-4685 4611	- 546 886 9345	156 - 371 - 54 2244	2.574 (8)	6.481	1.292		-4.766
301-332	-.4271	.7066	.3644	.5781	2645	-1543 1907	- 347 705 5322	162 - 453 - 56 1378	13.029 (9)	21.736	1.902	4.720	-7.979
333-360	-.4745	.6955	.4420	.5866	3502	-2124 2485	- 584 1023 5744	335 - 677 - 110 1778	15.226 (7)	28.613	2.668		-6.278





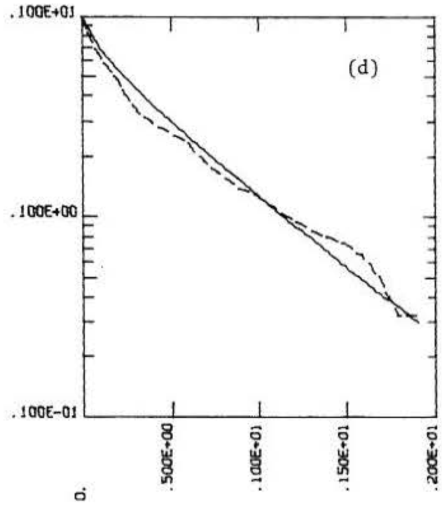


Fig. B-2-1. Plot of Log $\overline{F}(x)$ Versus x for Springfield Data: (a) Period 93-120, (b) Period 121-152, (c) Period 153-180, and (d) Period 181-212.

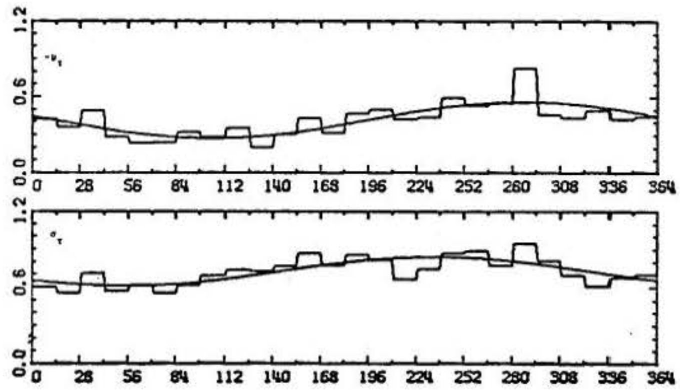


Fig. B-2-2.

The Periodic μ_τ and σ_τ for Daily Values of the Springfield Precipitation Station.

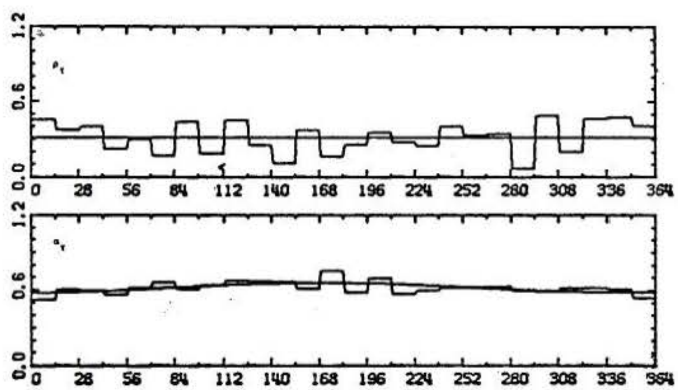
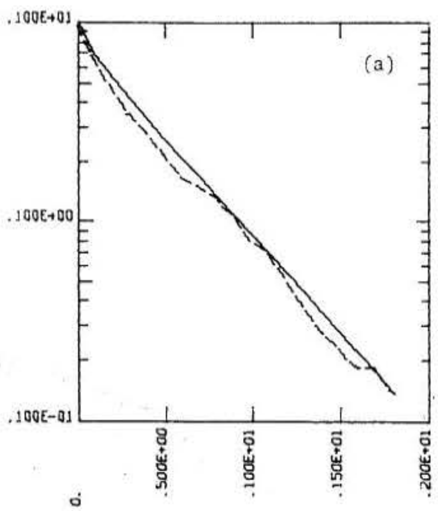


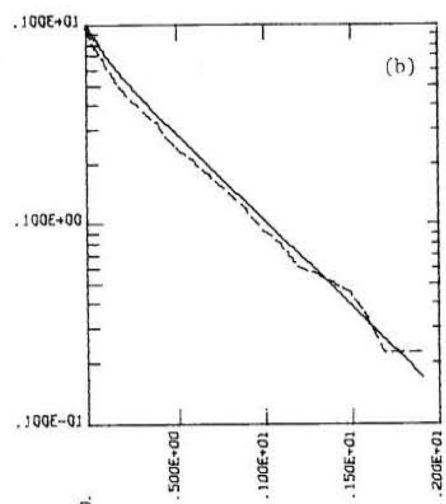
Fig. B-2-3.

The Periodic ρ_τ and σ_τ for Daily Values of the Springfield Precipitation Station.

Appendix B-3

Results of the Application of the Model to the Raleigh-Durham Precipitation Series





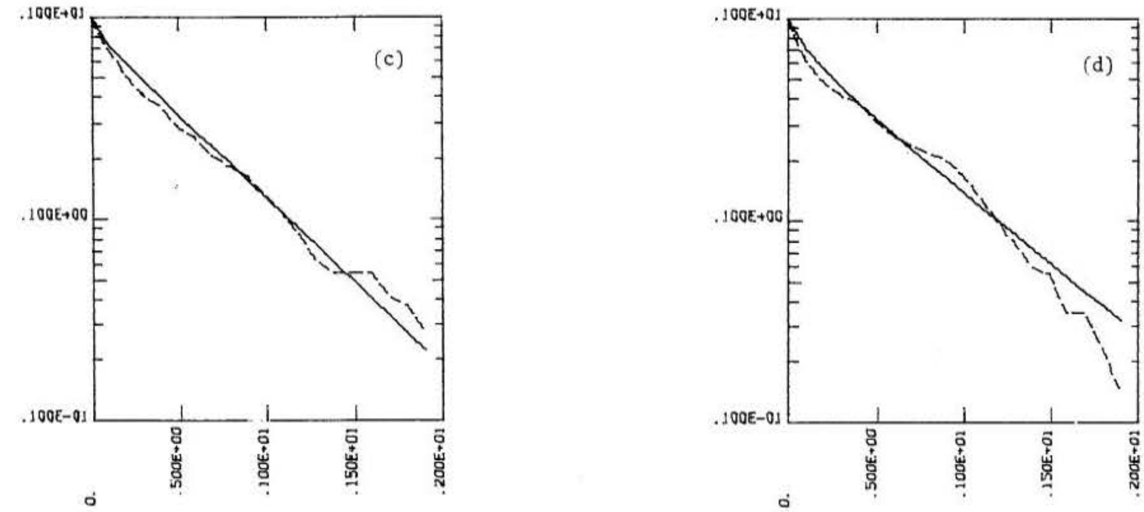


Fig. B-3-1. Plot of Log $\overline{F}(x)$ Versus x for Raleigh-Durham Data: (a) Period 121-152, (b) Period 153-180, (c) Period 181-212, and (d) Period 213-240.

Table B-3-1. Results Obtained in Case the Year is Divided in Twelve Seasons, for the Raleigh-Durham Station

PERIOD	PARAMETERS μ	σ	ρ	α	ASYMPTOTIC COVARIANCE MATRIX ($X\ 10^{-6}$)				T.S.1 (d.f.)	T.S.2 (1d.f.)	T.S.3 (4d.f.)	T.S.4 (4d.f.)	T.S.5 N(0,1)
001-032	-.3736	.6950	.2862	.6505	2436	-1393 1812	- 247 571 6050	126 - 484 - 43 1768	8.118 (9)	13.211	3.204	4.495	-5.214
032-060	-.2535	.6452	.2164	.7514	1872	- 985 1482	- 139 406 6401	29 - 524 - 21 2349	24.353 (9)	7.282			.571
061-092	-.2998	.6339	.3228	.7352	1850	-1014 1450	- 219 550 5384	126 - 581 - 50 2153	6.417 (8)	18.498	1.035	6.059	-1.990
093-120	-.3590	.6592	.2513	.7530	2550	-1526 1978	- 285 581 7271	271 - 785 - 54 2710	11.755 (7)	8.584	2.093		-2.160
121-152	-.3467	.6944	.2101	.6681	2221	-1260 1683	- 156 403 6284	82 - 440 - 24 1781	6.529 (9)	7.011	5.431	2.699	-4.397
153-180	-.4193	.7565	.3156	.6643	3413	-1924 2425	- 354 784 6788	69 - 432 - 49 2138	4.845 (9)	13.796	2.173		-4.313
181-212	-.2985	.7620	.3029	.6816	2357	-1186 1709	- 203 578 5071	- 97 - 180 - 34 1686	17.880 (12)	17.017	1.149	.628	-3.365
213-240	-.3537	.8015	.2735	.6255	3172	-1679 2278	- 261 664 6356	- 109 - 152 - 24 1685	17.650 (10)	11.246	4.383		-3.234
241-272	-.8665	1.0351	.4521	.6518	10116	-6125 6003	-1197 1709 6545	- 200 17 - 80 2440	17.382 (9)	26.576	3.922	12.059	-4.112
273-300	-.8053	.9473	.4502	.6897	9917	-6033 5893	-1290 1822 7651	134 - 397 - 130 3187	6.383 (8)	22.644	6.498		-2.932
301-332	-.4739	.7305	.3322	.6469	3350	-2010 2355	- 407 776 6459	240 - 573 - 70 1956	19.083 (8)	16.201	2.735	9.362	-3.706
333-360	-.4387	.6600	.3212	.7907	3257	-2025 2398	- 418 800 7639	530 -1055 - 121 3444	4.237 (7)	12.551	2.550		.002

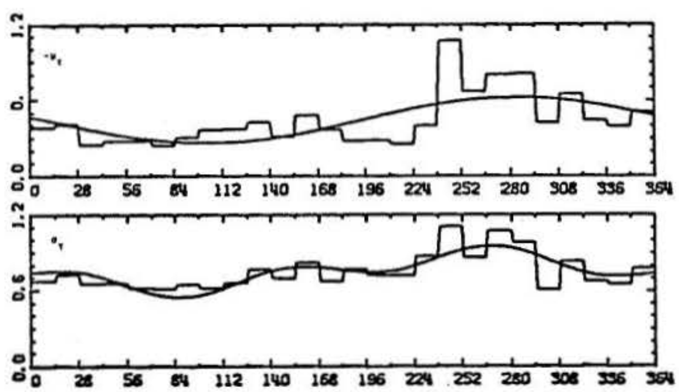


Fig. B-3-2.

The Periodic μ_τ and σ_τ for Daily Values of the Raleigh-Durham Precipitation Station.

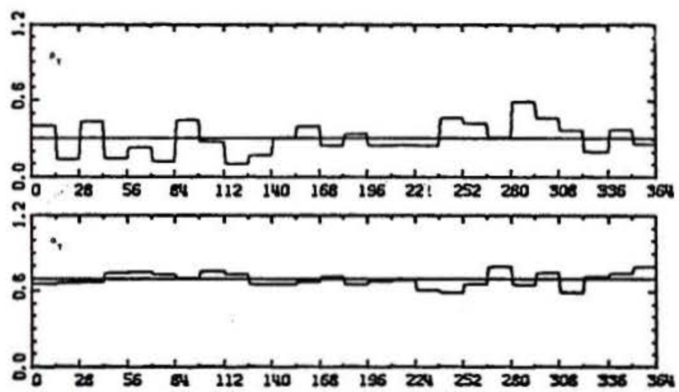


Fig. B-3-3.

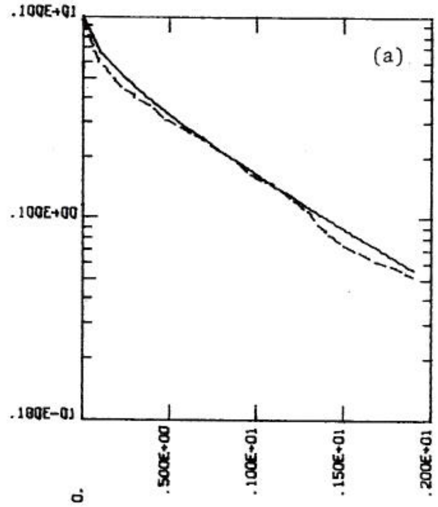
The Periodic ρ_τ and α_τ for Daily Values of the Raleigh-Durham Precipitation Station.

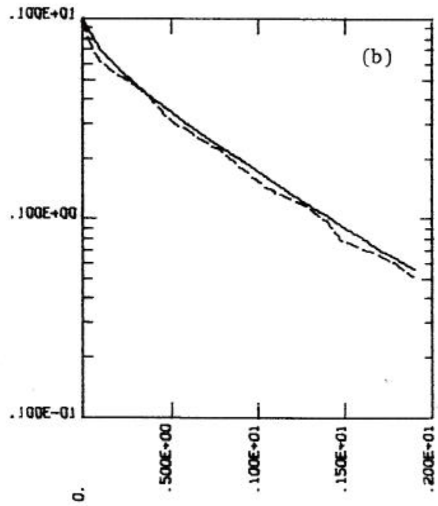
Appendix B-4

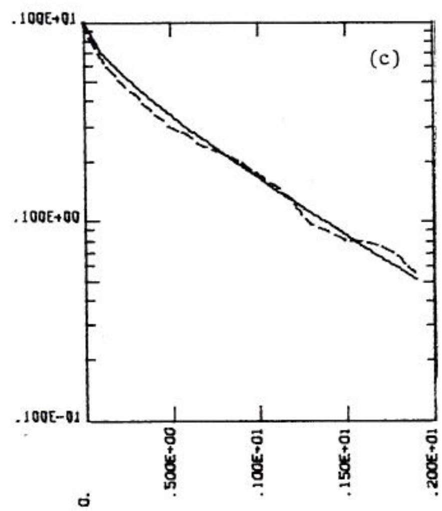
Results of the Application of the Model to the Austin Precipitation Series

Table B-4-1. Results Obtained in Case the Year is Divided in Twelve Seasons, for the Austin Station

PERIOD	PARAMETERS μ	σ	ρ	α		ASYMPTOTIC VARIANCE-COVARIANCE ($X\ 10^{-6}$)			T.S.1 (d.f.)	T.S.2 (1d.f.)	T.S.3 (4d.f.)	T.S.4 (4d.f.)	T.S.5 N(0,1)
001-032	-.4589	.6432	.4417	.5723	841	- 482 559	- 124 234 1693	99 -195 - 19 495	34.283 (11)	93.511	6.071	7.810	-14.807
033-060	-.4906	.7326	.3905	.5753	1131	- 628 731	- 133 270. 2045	59 -159 - 18 543	27.749 (13)	65.116	1.842		-12.879
061-092	-.6050	.7568	.3571	.5929	1345	- 791 819	- 146 259 2229	108 -187 - 23 580	23.074 (13)	52.826	7.211	7.991	-10.963
093-120	-.6060	.9293	.3835	.5601	1902	-1112 1265	- 207 384 2049	- 43 - 12 - 15 501	21.488 (18)	64.543	9.426		-11.141
121-152	-.6270	.9511	.3765	.5952	1792	-1065 1201	- 192 349 1829	- 52 5 - 12 499	26.883 (20)	70.383	.584	1.505	- 8.761
153-180	-.8966	1.0167	.4861	.6171	4410	-2933 2813	- 660 828 2236	- 17 - 48 - 40 802	26.069 (16)	84.268	7.783		- 7.695
181-212	-.9792	.9902	.4937	.5731	4545	-2961 2635	- 667 772 2208	108 -154 - 52 702	16.346 (15)	87.787	7.580	?	-11.752
213-240	-1.0460	.9695	.4263	.6171	5859	-3720 3083	- 735 813 3362	285 310 - 73 1048	8.120 (12)	48.091	7.908		- 8.147
241-272	-.7895	.9902	.4545	.5814	2804	-1795 1830	- 387 551 1872	- 34 - 18 - 24 559	25.744 (18)	90.777	5.555	8.703	-10.748
273-300	-.9755	1.0664	.4596	.5600	5980	-4174 3888	- 895 1042 2530	- 70 14 28 688	21.779 (17)	71.480	8.772		-10.504
301-332	-.6761	.8256	.4767	.5908	1770	-1045 1071	- 248 385 1787	89 -164 - 33 601	11.971 (14)	99.947	4.469	3.409	-10.580
333-360	-.5951	.7980	.4799	.5587	1620	- 929 1013	- 230 388 1850	67 -154 - 30 564	32.065 (14)	97.713	4.105		-13.726







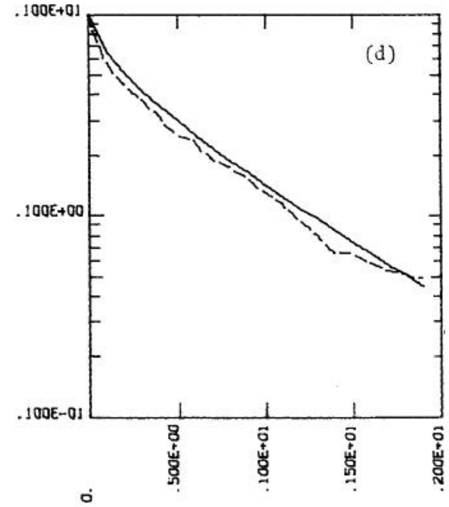


Fig. B-4-1. Plot of Log $\overline{F}(x)$ Versus x for Austin Data: (a) Period 93-120, (b) Period 121-152, (c) Period 153-180, and (d) Period 181-212.

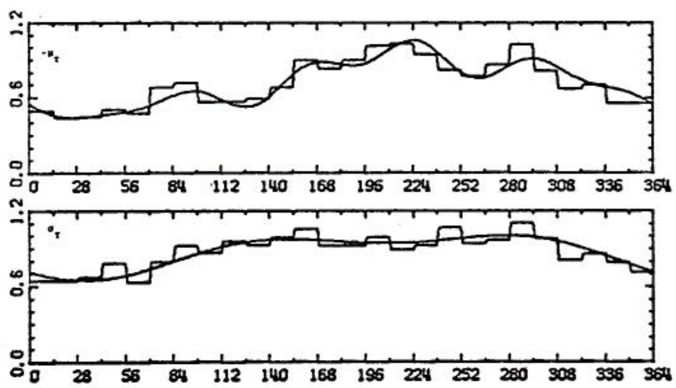


Fig. B-4-2.
The Periodic μ_τ and σ_τ for Daily Values of the Austin Precipitation Station.

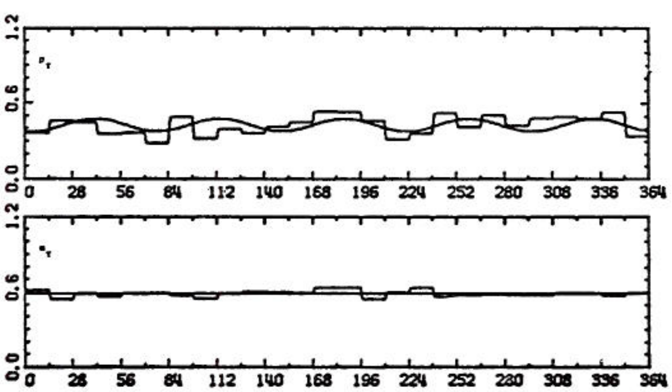


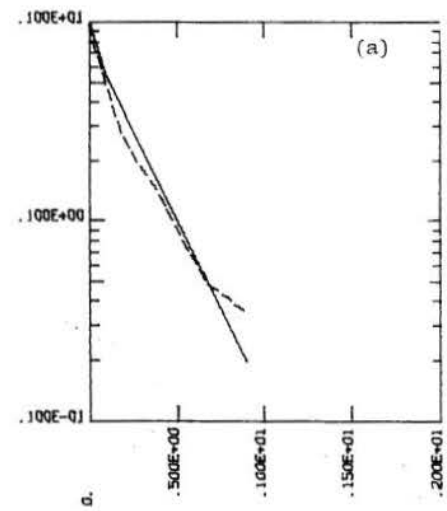
Fig. B-4-3.
The Periodic ρ_τ and α_τ for Daily Values of the Austin Precipitation Station.

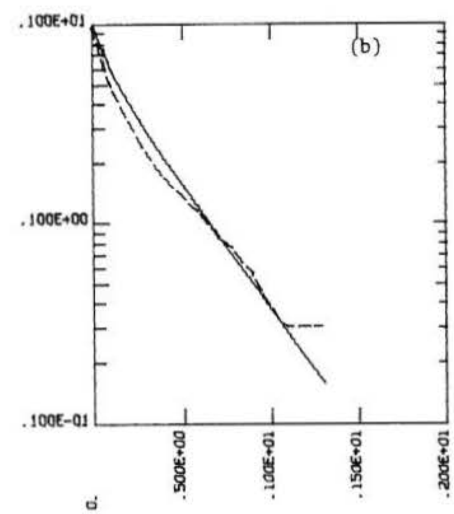
Appendix B-5

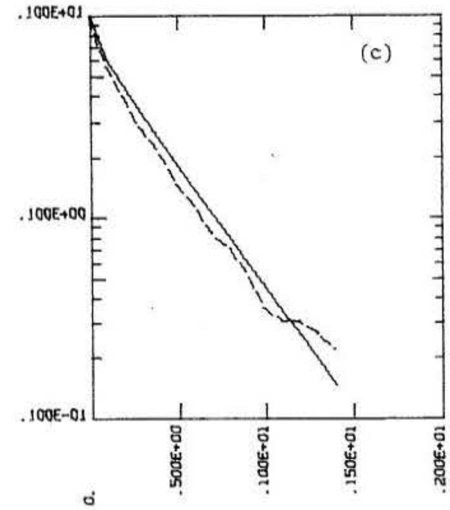
Results of the Application of the Model to the Rapid City Precipitation Series

Table B-5-1. Results Obtained in Case the Year is Divided in Twelve Seasons, for the Rapid City Station

PERIOD	PARAMETERS μ	σ	ρ	α	ASYMPTOTIC COVARIANCE MATRIX ($X\ 10^{-6}$)	T.S.1 (d.f.)	T.S.2 (1d.f.)	T.S.3 (4d.f.)	T.S.4 (4d.f.)	T.S.5 N(0,1)
001-032	-.1457	.1640	.3687	.8567	906 - 884 - 489 1979 994 638 -2319 9736 -1060 6610	1.033 (1)	13.310	7.286	3.798	-2.133
033-060	-.1402	.2739	.3707	.6928	642 - 586 - 225 669 1037 476 -1472 7041 - 222 3134	2.131 (1)	17.827	1.373		-5.179
061-092	-.2012	.3383	.4050	.6514	1022 - 884 - 355 790 1316 629 -1435 6357 - 269 2613	4.347 (2)	22.344	1.570	11.446	-6.466
093-120	-.2123	.4596	.3848	.6510	1358 - 844 - 282 357 1389 655 -1059 6457 - 108 2349	5.330 (4)	20.015	7.545		-5.248
121-152	-.1592	.5292	.3549	.5899	1211 - 569 - 178 24 1058 513 - 525 5118 - 31 1431	5.591 (6)	22.304	3.717	13.269	-7.355
153-180	-.0903	.5274	.2486	.5994	1136 - 454 - 103 - 85 974 382 - 459 5999 - 8 1501	4.235 (7)	9.858	4.585		-5.611
181-212	-.3232	.5551	.2561	.5915	2033 -1320 - 263 399 1745 533 - 868 7700 - 68 1800	7.339 (5)	8.339	3.574	4.545	-6.818
213-240	-.3269	.4754	.2105	.6825	2223 -1628 - 273 915 2012 496 1503 10444 - 129 3057	3.693 (3)	4.247	5.594		-2.881
241-272	-.4190	.5220	.3795	.6344	3228 -2325 - 686 1164 2552 994 -1574 8536 - 250 2764	3.249 (4)	15.584	1.903	3.506	-5.665
273-300	-.5795	.5066	.4446	.6793	9256 -6436 -2007 4068 5401 2121 -3706 13814 - 910 5867	2.984 (1)	12.595	2.251		-4.344
301-332	-.2467	.2848	.3264	.6964	1785 -1586 - 595 1845 1741 766 -2187 10178 - 600 3985	7.061 (2)	10.171	9.074	2.980	-6.396
333-360	-.1678	.2034	.2139	.8030	1049 -1016 - 319 1764 1193 459 -2216 12452 - 605 5406	5.039 (2)	3.808	9.341		-2.902







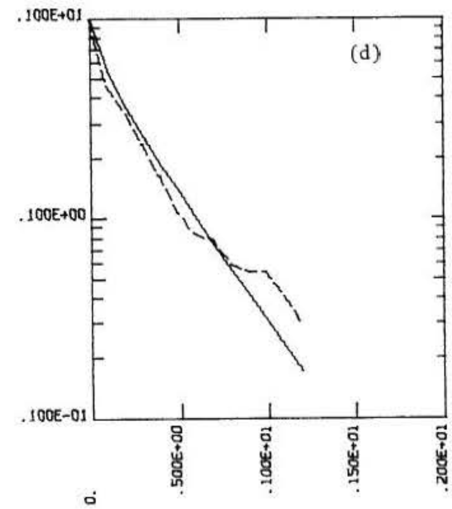


Fig. B-5-1. Plot of Log $\overline{F}(x)$ Versus x for Rapid City Data: (a) Period 93-120, (b) Period 121-152, (c) Period 153-180, and (d) 181-212.

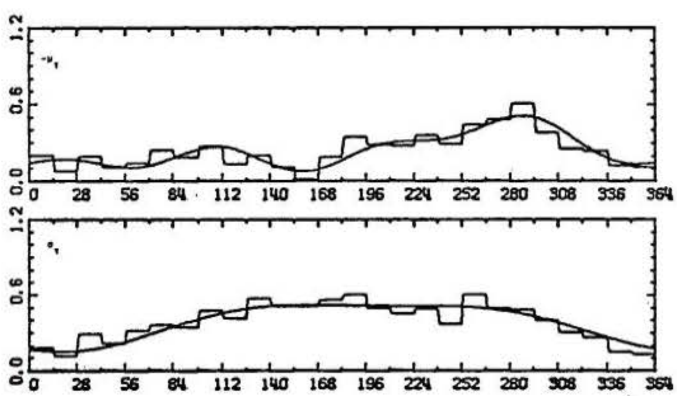


Fig. B-5-2.
The Periodic μ_τ and σ_τ for Daily Values of the Rapid City Precipitation Station.

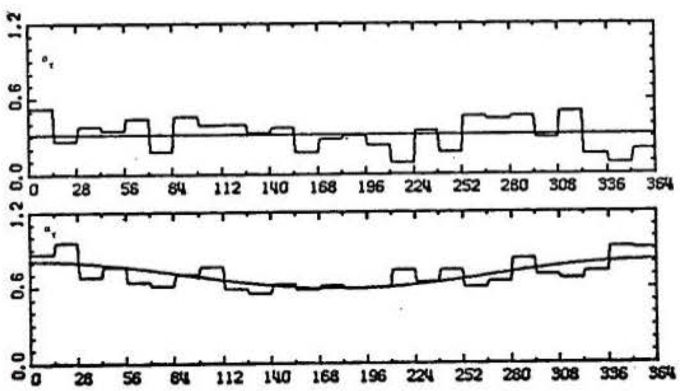
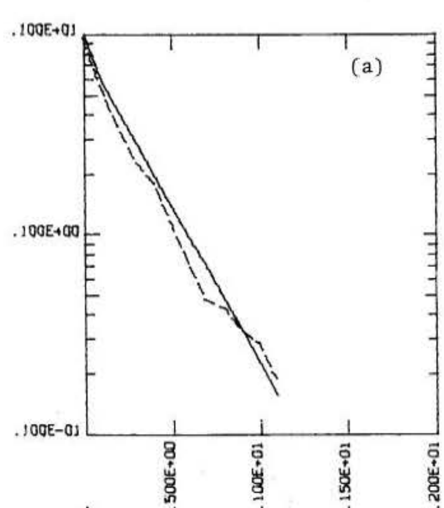
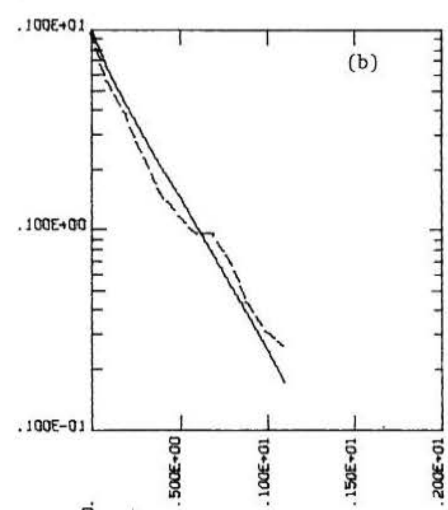


Fig. B-5-3.
The Periodic ρ_τ and α_τ for Daily Values of the Rapid City Precipitation Station.

Appendix B-6

Results of the Application of the Model to the Flagstaff Precipitation Series





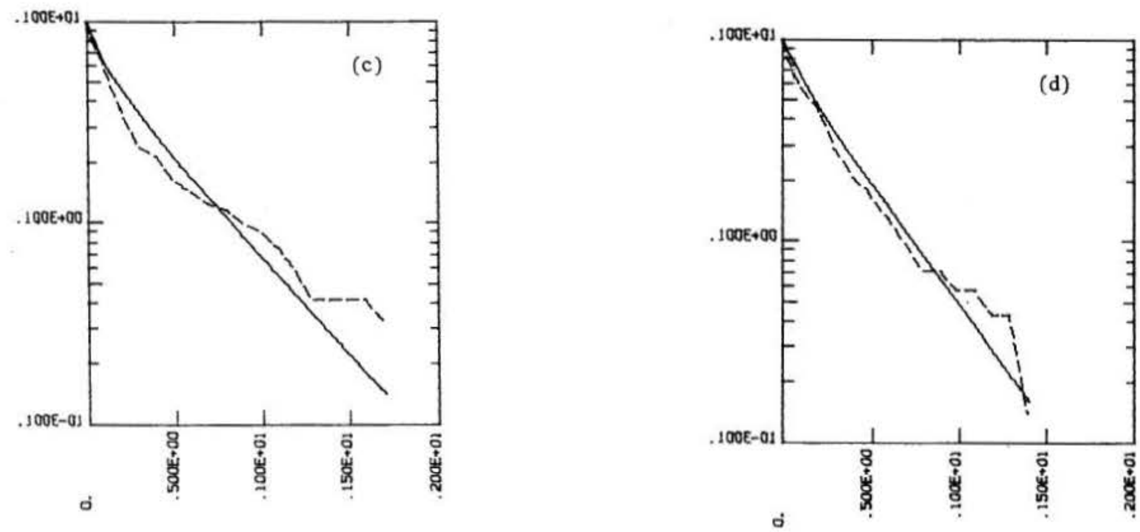


Fig. B-6-1. Plot of Log $\overline{F}(x)$ Versus x for Flagstaff Data: (a) Period 181-212, (b) Period 213-240, (c) Period 241-272, and (d) Period 273-300.

Table B-6-1. Results Obtained in Case the Year is Divided in Twelve Seasons, for the Flagstaff Station

PERIOD	PARAMETERS μ	σ	ρ	α	ASYMPTOTIC COVARIANCE MATRIX ($X\ 10^{-6}$)				T.S.1 (d.f.)	T.S.2 (1d.f.)	T.S.3 (4d.f.)	T.S.4 (4d.f.)	T.S.5 N(0,1)
001-032	- .5328	.6354	.6656	.7497	5036	-3295 3552	-1139 1544 3876	1242 -1734 - 335 4272	5.492 (5)	60.055	2.067	3.840	-1.321
033-060	- .5328	.6802	.5605	.7234	5556	-3513 3841	-1153 1685 6106	945 -1468 - 280 3987	3.735 (5)	35.279	3.187		-1.592
061-092	- .3965	.5364	.5364	.7739	2928	-2007 2385	- 736 1138 5459	1095 -1697 - 290 3971	6.226 (5)	37.810	7.361	5.261	-1.677
093-120	- .5836	.5730	.2416	.8133	7472	-5218 4663	- 925 1105 16024	3076 -3152 - 467 6542	3.546 (4)	3.791	1.118		- .697
121-152	- .6252	.4912	.6284	.8227	14284	9750 7504	-3451 2985 8610	8128 -6534 -1640 10788	.179 (1)	29.906	11.251	1.881	- .741
153-180	-1.0089	.7884	.6288	.5757	27036	-16681 12492	-4414 3987 9481	4133 -3410 - 690 5194	3.165 (2)	25.922	2.515		-6.504
181-212	- .1950	.4943	.3625	.6654	1205	- 665 1150	- 213 536 5522	175 - 779 - 54 1985	5.409 (6)	20.816	39.621	7.685	-4.714
213-240	- .1770	.4978	.2337	.6674	1255	- 678 1205	- 134 391 7112	134 - 799 - 26 2146	12.934 (6)	7.442	6.750		-4.730
241-272	- .5968	.7217	.5098	.6275	5779	-3632 3793	-1042 1526 6481	746 -1101 - 207 2736	14.631 (6)	30.172	14.230	1.702	-6.567
273-300	- .8045	.7253	.5145	.7108	14155	-9116 7614	-2546 2717 10953	3008 -2890 - 597 5852	3.551 (3)	19.096	5.374		-2.975
301-332	- .7056	.7354	.5235	.6798	8255	-5279 4960	-1475 1871 7421	1376 -1629 - 318 3834	5.143 (5)	27.546	.701		-3.578
333-360	- .6828	.8005	.5768	.6917	8820	-5493 5540	-1612 2154 6371	812 -1210 - 268 3875	6.445 (6)	35.404	4.035		-3.270

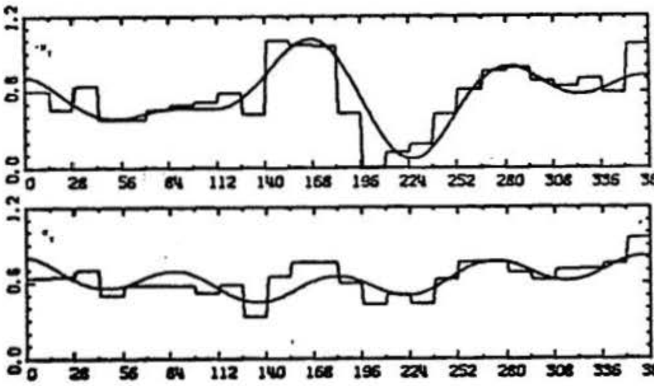


Fig. B-6-2.
The Periodic μ_τ and σ_τ for Daily Values of the Flagstaff Precipitation Station.

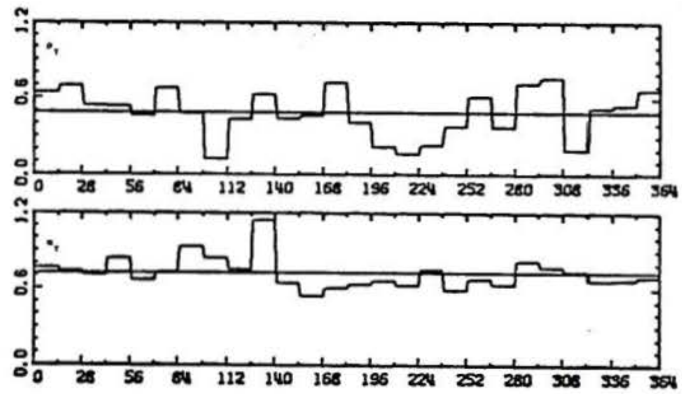


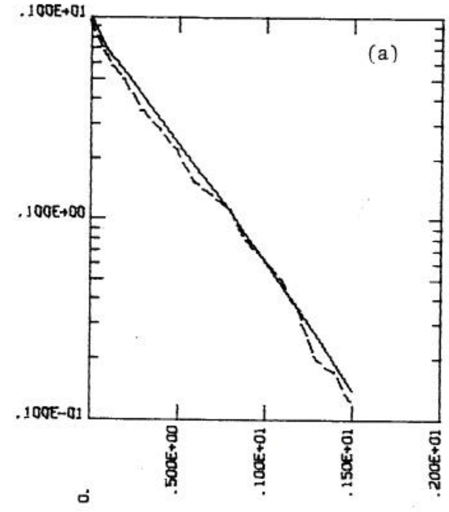
Fig. B-6-3.
The Periodic ρ_τ and α_τ for Daily Values of Flagstaff Precipitation Station.

Appendix B-7

Results of the Application of the Model to the Seattle-Tacoma Precipitation Series

Table B-7-1. Results Obtained in Case the Year is Divided in Twelve Seasons, for the Seattle-Tacoma Station

	PARAMETERS				ASYMPTOTIC COVARIANCE MATRIX ($\times 10^{-6}$)								
PERIOD	μ	σ	ρ	α					T.S.1 (d.f.)	T.S.2 (1d.f.)	T.S.3 (4d.f.)	T.S.4 (4d.f.)	T.S.5 N(0,1)
001-032	.1706	.4655	.4602	.6299	601	- 62 345	- 19 327 2400	- 265 - 126 - 89 1046	9.960 (10)	62.657	5.574	1.156	-1.627
033-060	.0453	.4593	.4172	.6089	689	- 133 517	- 77 391 3414	- 206 - 303 - 41 1207	11.589 (8)	40.589	13.064		-4.875
061-092	.0704	.3525	.4375	.6882	364	- 9 316	- 25 265 2668	- 239 - 385 - 77 1407	4.708 (6)	54.015	11.180	.538	-1.325
093-120	-.0423	.3720	.3893	.6778	493	- 159 576	- 82 340 4065	- 64 - 632 - 41 1755	8.122 (5)	31.782	1.341		-2.988
121-152	-.2146	.4192	.3273	.6688	942	- 663 1047	- 198 430 5603	393 - 934 - 99 1948	7.048 (4)	17.716	11.644	7.808	-4.923
153-180	-.1885	.3725	.3502	.6778	897	- 692 1132	- 250 502 6204	519 -1179 - 156 2370	4.547 (3)	18.280	2.848		-4.708
181-212	-.4345	.4316	.5197	.7257	3893	-2944 2813	-1101 1286 6910	2432 -2533 - 629 4430	2.147 (2)	30.143	8.387	3.789	-2.344
213-240	-.3873	.4975	.5622	.6450	2808	-2006 2339	- 751 1122 5108	1098 -1577 - 290 2912	2.505 (4)	41.227	6.219		-4.886
241-272	-.2306	.4611	.5423	.6770	1155	- 716 1161	- 303 643 3496	335 - 914 - 136 2102	1.397 (5)	59.112	19.301	4.139	-4.279
273-300	-.0384	.4677	.4438	.7187	785	- 234 679	- 109 460 3605	- 141 - 496 - 59 1867	9.338 (8)	42.042	1.532		-1.378
301-332	.0947	.4987	.3704	.6391	661	- 119 437	- 44 314 3025	- 216 - 174 - 43 1081	16.982 (10)	37.029	3.583	7.279	-3.064
333-360	.1816	.4326	.3430	.6504	566	- 36 320	1 270 3217	- 319 - 172 - 90 1246	25.928 (10)	30.622	1.962		-2.155



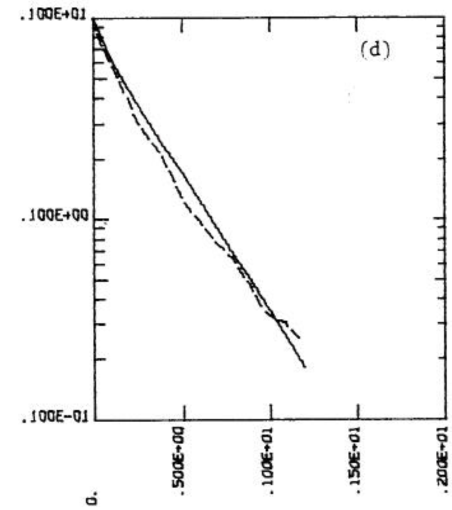
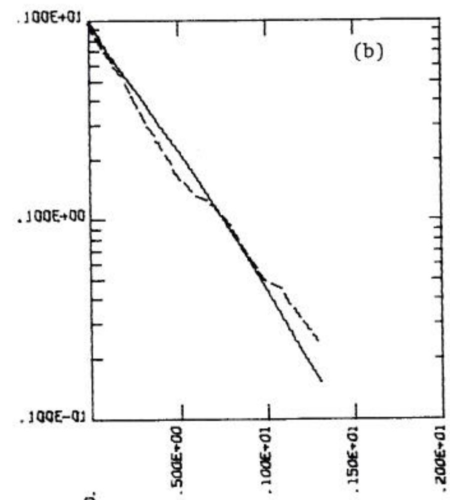


Fig. B-7-1. Plot of Log $\overline{F}(x)$ Versus x for Seattle-Tacoma Data: (a) Period 301-332, (b) Period 333-360, (c) Period 1-32, and (d) Period 33-60.



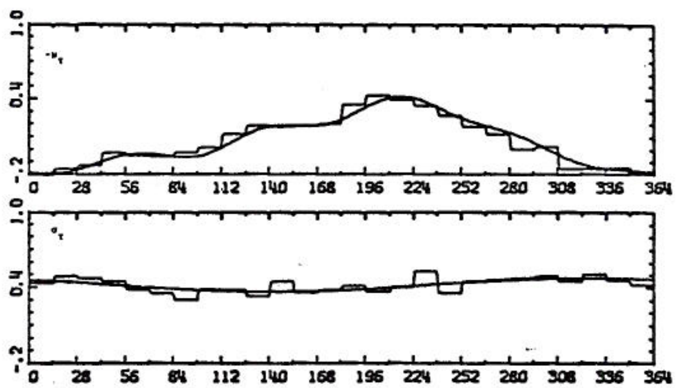
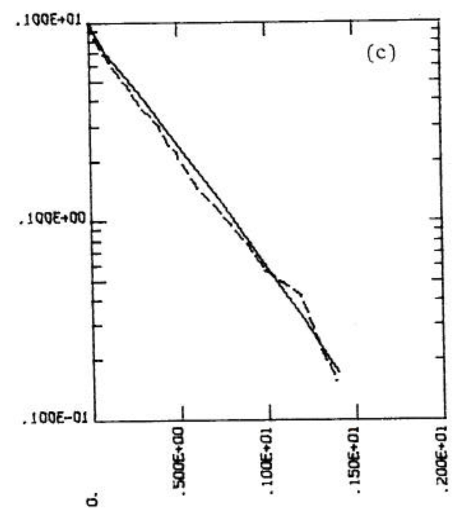


Fig. B-7-2.

The Periodic μ_τ and σ_τ for Daily Values of the Seattle-Tacoma Precipitation Station.



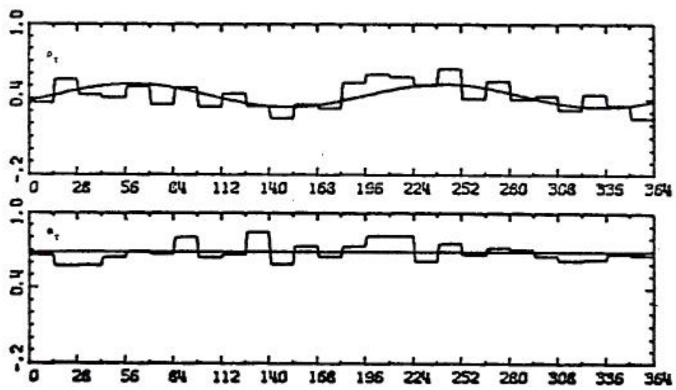


Fig. B-7-3.

The Periodic ρ_τ and α_τ for Daily Values of the Seattle-Tacoma Precipitation Station.

Appendix C

Critical Values for the Chi-Square Probability Distribution

Degrees of Freedom	5 Percent Significance Level	1 Percent Significance Level
1	3.84	6.63
2	5.99	9.21
3	7.81	11.3
4	9.49	13.3
5	11.1	15.1
6	12.6	16.8
7	14.1	18.5
8	15.5	20.1
9	16.9	21.7
10	18.3	23.2
11	19.7	24.7
12	21.0	26.2
13	22.4	27.7
14	23.7	29.1
15	25.0	30.6
16	26.3	32.0
17	27.6	33.4
18	28.9	34.8
19	30.1	36.2
20	31.4	37.6
21	32.7	38.9
22	33.9	40.3
23	35.2	41.6
24	36.4	43.0
25	37.7	44.3
26	38.9	45.6
27	40.1	47.0
28	41.3	48.3
29	42.6	49.6
30	43.8	50.9

Key Words: Time Series, Intermittent Processes, Daily Precipitation, Daily Runoff.

Abstract: A model for description and generation of new samples of intermittent daily precipitation series is developed. The basic assumption is that precipitation is a result of truncating a non-intermittent process. Classical methods for modeling the time dependence in this latter process can then be applied. The univariate non-intermittent process permits then an extension to multivariate case. Specific tests, related to stationarity and time independence of the process, are formulated. The model is tested on series of several precipitation stations in USA. Results have been found satisfactory.

Another model, in this case for the description and generation of new samples of daily streamflow, is also developed.

Key Words: Time Series, Intermittent Processes, Daily Precipitation, Daily Runoff.

Abstract: A model for description and generation of new samples of intermittent daily precipitation series is developed. The basic assumption is that precipitation is a result of truncating a non-intermittent process. Classical methods for modeling the time dependence in this latter process can then be applied. The univariate non-intermittent process permits then an extension to multivariate case. Specific tests, related to stationarity and time independence of the process, are formulated. The model is tested on series of several precipitation stations in USA. Results have been found satisfactory.

Another model, in this case for the description and generation of new samples of daily streamflow, is also developed.

Key Words: Time Series, Intermittent Processes, Daily Precipitation, Daily Runoff.

Abstract: A model for description and generation of new samples of intermittent daily precipitation series is developed. The basic assumption is that precipitation is a result of truncating a non-intermittent process. Classical methods for modeling the time dependence in this latter process can then be applied. The univariate non-intermittent process permits then an extension to multivariate case. Specific tests, related to stationarity and time independence of the process, are formulated. The model is tested on series of several precipitation stations in USA. Results have been found satisfactory.

Another model, in this case for the description and generation of new samples of daily streamflow, is also developed.

Key Words: Time Series, Intermittent Processes, Daily Precipitation, Daily Runoff.

Abstract: A model for description and generation of new samples of intermittent daily precipitation series is developed. The basic assumption is that precipitation is a result of truncating a non-intermittent process. Classical methods for modeling the time dependence in this latter process can then be applied. The univariate non-intermittent process permits then an extension to multivariate case. Specific tests, related to stationarity and time independence of the process, are formulated. The model is tested on series of several precipitation stations in USA. Results have been found satisfactory.

Another model, in this case for the description and generation of new samples of daily streamflow, is also developed.

The basic assumption is that the rising and falling limbs of discharge hydrographs can be modeled individually as two difference, intermittent processes, also physically different. The rising limb process is mainly due to factors external to watersheds. It is modeled similarly as the intermittent precipitation process. The falling limb is conceived as governed by regularities of water outflow from watersheds, with the watershed storage and outflow represented by two linear reservoirs. The model is tested for a case study. Results are satisfactory in reproducing the combined process.

<u>Reference</u>: Kelman, Jerson; Colorado State University, Hydrology Paper No. 89 (February 1977), Stochastic Modeling of Hydrologic, Intermittent Daily Processes.

The basic assumption is that the rising and falling limbs of discharge hydrographs can be modeled individually as two difference, intermittent processes, also physically different. The rising limb process is mainly due to factors external to watersheds. It is modeled similarly as the intermittent precipitation process. The falling limb is conceived as governed by regularities of water outflow from watersheds, with the watershed storage and outflow represented by two linear reservoirs. The model is tested for a case study. Results are satisfactory in reproducing the combined process.

<u>Reference</u>: Kelman, Jerson; Colorado State University, Hydrology Paper No. 89 (February 1977), Stochastic Modeling of Hydrologic, Intermittent Daily Processes.

The basic assumption is that the rising and falling limbs of discharge hydrographs can be modeled individually as two difference, intermittent processes, also physically different. The rising limb process is mainly due to factors external to watersheds. It is modeled similarly as the intermittent precipitation process. The falling limb is conceived as governed by regularities of water outflow from watersheds, with the watershed storage and outflow represented by two linear reservoirs. The model is tested for a case study. Results are satisfactory in reproducing the combined process.

<u>Reference</u>: Kelman, Jerson; Colorado State University, Hydrology Paper No. 89 (February 1977), Stochastic Modeling of Hydrologic, Intermittent Daily Processes.

The basic assumption is that the rising and falling limbs of discharge hydrographs can be modeled individually as two difference, intermittent processes, also physically different. The rising limb process is mainly due to factors external to watersheds. It is modeled similarly as the intermittent precipitation process. The falling limb is conceived as governed by regularities of water outflow from watersheds, with the watershed storage and outflow represented by two linear reservoirs. The model is tested for a case study. Results are satisfactory in reproducing the combined process.

<u>Reference</u>: Kelman, Jerson; Colorado State University, Hydrology Paper No. 89 (February 1977), Stochastic Modeling of Hydrologic, Intermittent Daily Processes.

LIST OF PREVIOUS 25 PAPERS

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 Western United States, by Albert G. Law, May 1974.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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