USE OF PARTIAL FLOOD SERIES FOR ESTIMATING DISTRIBUTION OF MAXIMUM ANNUAL FLOOD PEAK

by
Viraphol Taesombut and Vujica Yevjevich

October 1978

HYDROLOGY PAPERS
COLORADO STATE UNIVERSITY
Fort Collins, Colorado

97

USE OF PARTIAL FLOOD SERIES FOR ESTIMATING DISTRIBUTION OF MAXIMUM ANNUAL FLOOD PEAK

by

Viraphol Taesombut and Vujica Yevjevich

HYDROLOGY PAPERS
COLORADO STATE UNIVERSITY
FORT COLLINS, COLORADO 80523

October 1978 No. 97

TABLE OF CONTENTS

TABLE OF CONTENTS, Cont.

ACKNOWLEDGMENT

The paper is the Ph.D. thesis of Viraphol Taesombut, with V. Yevjevich, Professor of Civil Engineering, as his adviser, major professor, and chairman of the graduate committee. The other committee members were: D. C. Boes, Associate Professor of Statistics, D. A. Woolhiser of the Agricultural Research Service, M. M. Siddiqui, Professor of Statistics, and T. G. Sanders, Assistant Professor of Civil Engineering.

The financial support of Mr. Viraphol Taesombut during his studies by the Kasetsart University, Bangkok, Thailand, is acknowledged. The acknowledgment goes also to the Department of Civil Engineering for financial support for numerical computations. The supports by National Science Foundation Grant ENG-74-17396, and Colorado State University Experiment Station Research Project on Floods (ES 1114), are duly acknowledged.

ABSTRACT

The estimation of probability distributions of maximum annual flood peak by using a combination of probability distributions of the number and the magnitude of flood peaks that exceed a selected truncation level, is the subject matter of this paper. This method of estimation is tested on the 17 daily streamflow series of gaging stations in the United States. Five discrete and six continuous probability distribution functions were used to fit their frequency distributions of the number and the magnitude of exceedances above the selected truncation level of partial flood series, respectively. From them the best fit functions are selected. For these functions inferred, the goodness-of-fit statistics are related to the truncation level of partial series. The probability distribution of the largest annual exceedance of the instantaneous flood peak (represented by highest daily discharge), with assumptions postulated and tested on these 17 time series. The sequential dependence in partial and annual flood peak series is also investigated, with the dependence of partial flood peak series increasing with a decrease of the truncation level. For the range of truncation levels studied in this paper, the average number of exceedances per year varied from one to four, with the sequential dependence relatively small for the rivers used as examples.

The mathematical model for generating samples of daily flow series is selected and refined. The daily flow series of two gaging stations, with different runoff regimes were used in testing the model. With the models based on statistics of the samples of the daily flow series of these two river examples, as well as of their derived annual flood peak series, the generated samples of daily flows showed the parameters to be close to the inferred parameters of historic daily flow series. The generated samples reproduced well the flood extremes. These samples then were used to investigate the properties of flood peaks.

By using the generated samples of daily flows, the efficiency of estimated annual flood peaks of given return periods was investigated by using both the annual and the partial flood peak series. The sampling variances of annual flood peaks of given return periods, obtained from each of these two flood peak series, were compared both analytically and experimentally from generated samples. The estimates of annual flood peaks of given return periods from the partial flood peak series showed a smaller sampling variance than the corresponding estimates from the annual flood peak series, when the average number of exceedances per year in partial flood series was at least 1.65 for an exact analytical comparison, and at least 1.50 for an approximate analytical comparison. The ratios of sampling variances of estimated annual flood peaks for these two approaches did not show a dependence on the sample size.

In case of the use of the empirical approach, the sampling variance of estimated annual flood peaks from the partial flood series showed to be smaller than the corresponding sample variance of estimated annual flood peaks from the annual flood series, for the range of investigated return periods and for the average number of exceedances in partial flood series of at least 1.95 for the sample sizes 10-25, and somewhat larger than 1.95 for the larger sample sizes.

Chapter I
INTRODUCTION

1.1 Introductory Remarks

Floods and droughts as runoff extremes represent some of the most damaging natural disasters, with which humanity has to live and struggle through all the recorded history. Floods represent a rapidly evolving disaster. Basic flood risks are functions of climatic factors, conditions of river basins, and the state and occupancy of river floodplains. These risks can be changed only by changing these factors, conditions, state, and occupancy.

The analysis of flood frequency distributions, as the inference making about flood probability distributions, plays a major role in hydrologic and economic evaluations of water resources projects and in establishing project design criteria. The large highway programs that include bridges and drainage design belong to the major undertakings which depend on flood frequency analysis, because they require large expenditures of public funds. In the construction of dams, the spillways account for a sizable portion of the total cost. The capacity of a spillway is governed by flood characteristics of given frequency or recurrence interval. Besides the usual needs for information on floods for most water resources projects, this information has become of paramount importance for flood insurance. The most reliable determination of flood levels for given return periods is needed in floodplain delineation and for establishing the appropriate flood insurance rates.

The definition of a flood often is not precise. In general, a flood is a relatively high flow that may or may not overtop the banks of a stream, and which may or may not cause damage. The general public usually refers to floods as high flows that cause damage. The water resource specialists frequently define floods as flows of the magnitude close to or higher than the one-year return flood. Usually, the maximum instantaneous annual flood peak discharges of each year are used in flood analysis. Often the maximum instantaneous flood peak discharge. To avoid the division of the water year by the arbitrariness in selecting 365 daily values of the calendar year, many countries use the water year beginning with usually lowest flows (dry season) rather than using January 1 as the year beginning.

Two types of flood peak series, the annual flood series and the partial flood series, are considered in this study. The annual series consists of the largest flood in each year, as defined above. The partial flood series consists of all well-defined flood peaks above a specified magnitude, often called the flood truncation level. Partial flood series is approximately derived from the mean daily flows, since instantaneous peak flows for events smaller than the annual maximum peak are not readily available. While the time series of mean daily peak flows are close to the instantaneous peak flows of large catchments, the partial series of mean daily flows is only an approximation to instantaneous peak partial series for small flashy catchments, since the instantaneous peaks are smoothed in daily averaging of flows.

1.2 Major Problems Needing Studies

A classical dilemma in flood frequency analysis is whether to use either the annual flood peak series, or the partial flood series of all the peaks above a given truncation level. The most frequent objection encountered with respect to the use of annual flood series is that it uses only one flood for each year. In certain cases the second largest flood in a year, which the annual flood series neglects, may outrank many annual floods of other years. The largest annual discharges in dry years of some rivers in arid or semiarid regions may be so small that calling them floods may be misleading. Another increasingly important shortcoming of annual flood series is that only a small number of floods is considered. On the other hand, the partial flood series appears to be more useful for theoretical analysis than the annual flood series, since the objections raised on annual flood series do not apply. The major drawback of partial flood series is that the sequence of flood events might not be independent time series since some flood peaks may occur on the recession limbs of preceding floods. However, the dependence of partial flood series is a function of the selected truncation level which defines a particular partial flood series. If the truncation level is selected in such a way that the average number of floods per year is greater than one, and the assumption of independence of these floods still valid, the partial flood series may become more useful for theoretical analyses than the annual flood series. Consequently, in order to ascertain whether the partial flood series is more efficient for estimating the flood values of given return periods than the annual flood series, the comparison of sampling variances of flood peaks of given return periods obtained from each of the two flood peak series needs to be investigated.

Another problem of continuing interest in flood frequency analysis in case of annual flood series with small sample size is the reliability of estimates of skewness coefficients of historical flood series. Regional estimates of these coefficients may be the only solution in case of short historical records.

To overcome the problems of short annual flood series, the consideration of all the flood peaks above the given truncation level, as partial flood series should be used. This approach provides an alternative approach of estimating the probabilities of annual flood peaks by a combination of distributions of the number and the magnitude of flood peaks above a suitable truncation level. This approach has two important advantages over the empirical distribution approach in using the annual flood series. First, the partial flood series would contain more floods than the annual flood series. Hence, in general, the estimate of parameters of annual flood distribution from the partial flood series would be subject to lesser uncertainty. Second, the theoretical expressions on annual flood distributions obtained through characteristics of partial floods have physical relevance and often are exact distributions rather than asymptotic.

Further studies are necessary in order to answer many questions arising in the use of partial flood series and achieve the dual goal of consistency and accuracy in estimating flood values for given flood return periods. For example, the range of suitable truncation levels to be used for defining partial flood series should be well investigated, as well as the probability distribution functions of the best fit for the frequency distributions of the number of floods exceeding the truncation level for a given time interval. Investigations are needed for probability distributions of the flood magnitudes of partial series, as well as the probability distributions of the largest flood for the same time interval. A question remains whether the dependence is significant in partial flood series for each truncation level.

1.3 Objectives of the Study

The major objectives of this study are:

(1) To estimate parameters of probability distributions of annual flood peaks by using the partial flood series, instead of estimating parameters directly from annual flood series.

(2) To compare efficiency in using annual and partial flood series for estimating annual flood peaks of given return periods by using the sampling variance of such estimates of annual flood peaks, estimated from each of the two series for various sample sizes and assumed probability models.

(3) To develop mathematical models for generating the long records of daily flow data for the use of records in the comparison of efficiency of estimates of annual flood peaks by using annual and partial flood series.

1.4 Procedures Used

The theory of probability distributions of partial series of flood peaks is outlined in Chapter III. It includes the outline of selected discrete distribution functions for the number of floods and the continuous distribution functions for flood magnitudes used for fitting the frequency distributions of the number and the magnitude of flood peaks above a given truncation level, respectively. The assumptions used in deriving the probability distribution of the largest flood peak in the year from the combination of distributions of number and magnitude of flood peaks of partial flood series are given. This chapter ends with a procedure used in comparing the sampling variances of annual flood peaks of given return periods obtained from annual and partial flood series.

The application of the theory of probability distribution of partial series of flood peaks to 17 stream flow gaging stations throughout the United States is presented in Chapter V. Probability distributions are selected from distribution functions outlined in Chapter III, by using the goodness-of-fit parameters in fitting the 17 frequency distribution functions for both the number and the magnitude of floods for the year as the time interval. The results of investigation on how the parameters of selected distribution functions change with the truncation level are then presented. The derivation of probability distribution of the largest flood peak in a year is given with the necessary assumptions postulated and tested on observed data. Chapter V further includes the study of dependence of annual flood series and partial flood series, as well as how the degree of dependence of partial flood series changes with the change of the truncation level.

Procedures used in developing the mathematical model of daily flow series are presented in Chapter IV. Results of generation of long records of daily flows, by using the parameters of the Boise River, Idaho and the Powell River, Tennessee, are shown in Chapter VI.

The comparison of efficiency in using annual and partial flood series for estimating annual flood peaks of given return periods, by using generated daily flow series, is presented in Chapter VII. The long records of generated daily flow series are used for verifying properties and assumptions, as required in the development of the partial flood series model.

Chapter II
REVIEW OF LITERATURE

2.1 Definitions of Annual and Partial Flood Series

Flood data is usually listed either in the form of an annual flood series or a partial flood series. Annual flood series is a sequence of annual floods, with annual flood defined as the largest instantaneous peak discharge of each year of record. Sometimes the maximum mean daily discharge of each year is used as this flood. The partial flood series are not as precisely defined as the annual flood series. The definition of partial flood series depends on the application of the flood frequency curves as well as the hydrologic river basin characteristics. Water Resources Council (1976) defines the partial flood series as a sequence of separate flood events. These separate floods are arbitrarily defined as events separated by at least as many days as five plus the natural logarithm of square miles of drainage area, with the requirement that the intermediate flows must drop below 75 percent of the lower of the two separate maximum daily flows. Zelenhasic (1970) and Rousselle (1972) defined the partial flood series as all flood peaks which are called exceedances above a given truncation level. In the case of a multiple peak flood hydrograph, only the largest discharge is considered to be the flood peak. This latter treatment of partial flood is an approximation, with an expectation that the independence of flood peaks would be closely preserved. It is feasible to separate a complex hydrograph in such a way as to obtain independent flood peaks; however, that approach would complicate the estimate of partial flood series with no significant advantage in modeling floods by the partial series approach.

2.2 Theories of Probability Distributions of Annual Flood Peaks

Large numbers of references are available on flood studies by using the statistical approaches. It would be beyond the scope of this study to cover all the methods proposed to date. For purposes of showing the versatility of approaches to the problem, some analytical methods that have been used by individuals and agencies in recent years will be reviewed.

Flood probability distribution functions have been tested empirically, and when found unsatisfactory they have been replaced by the new functions. It was found out relatively early that logarithms of annual flood peaks are often well fitted by the Gaussian normal function. Because of high skewness coefficient in flood frequency distributions, functions with such characteristics are looked for. If annual flood peaks could be considered as products of effects of a large number of random causal factors, it should be lognormally distributed, since logarithms of the variable could be considered as sums of effects of a large number of random causal factors, therefore normally distributed by the central limit theorem (Chow, 1954).

Based on the annual flood series of 1959 long-record river gaging stations in the United States, Beard (1954) concluded that with rare exceptions the logarithms of annual flood of mean daily flows are normally distributed.

Foster (1924) preferred to work with untransformed data and hence sought to fit the skewed distribution functions. He introduced the use of the Pearson Type III density function, with the empirical support for it from data, although some hydrologists consider the application to be somewhat difficult.

The Gumbel extreme value distribution is one of the three limiting forms of distributions of the largest member of a sample of N independent random variables from a distribution which satisfies certain conditions in the asymptotic behavior of its tails.

Extreme value theory indicates that if the random variable X_N is the maximum in a sample of size N from some population of x values, provided N is sufficiently large, the distribution of X_N is one of the three limiting forms, the choice depending on the parent distribution of x. Since the maximum daily flow in a year is the maximum of N = 365 values, Gumbel (1941) postulated that it should be distributed as the extreme value variable. However, the N values of daily flows are highly dependent and they are not identically distributed, since it is well known that daily flows are highly autocorrelated with periodic parameters. If, therefore, the annual flood peaks follow an extreme value distribution, it is for some other reason than those stated by Gumbel. If daily flows are not independent in the annual collection of 365 values, one may find a group of independent values to replace dependent values. Unfortunately, this group if determined would be so small that the assumption of a large sample would be violated. Furthermore, the critical assumption is the assumption that the parental population is made up of identically distributed random variables. It is not feasible to assume that the daily flows of the first of May have the same distribution as those of the first of December, as shown by Quimpo (1967) because the mean and the standard deviation of daily flows are periodic. Hence, the theoretical arguments that flood peaks follow an extreme value distribution are weak, not supported by time series properties. In addition, the problem is the selection of the type of extreme value distribution; according to extreme value theory, this distribution depends on the type of parental distribution, which is not known a priori.

At present, the latest word on frequency analysis of annual flood series in the United States may be the method adopted by Water Resources Council (1976). It is condensed in Bulletin No. 17, prepared by the Hydrology Committee of the Council. This bulletin is an extension of Bulletin No. 15, "A Uniform Technique for Determining Flood Flow Frequency" (U.S. Water Resources Council, 1967, and also Benson, 1968). At its inception the method raised controversies among water resources agencies. The suggested and adopted probability distribution function for flood peaks in the bulletin is the Pearson Type III distribution function applied to logarithms of the annual flood peaks, briefly called the Log-Pearson Type III function. Parameters of that function are expressed in terms of the mean, the standard deviation, and the skewness coefficient, computed for logarithms of annual flood peaks.

The unadjusted frequency curve is obtained by computing the logarithms of annual flood peaks which correspond to selected points on the frequency scale. Since the samples used in hydrologic studies are of finite sizes, an adjustment of the exceedance frequency is necessary. The magnitude of flows which correspond to each of the selected points is computed by:

$$\log Q = \bar{x} + KS ,$$

in which log Q = the logarithm of flow which corresponds to a specified value of the unadjusted exceedance probability, $\bar{x}$ = the mean of logarithms of sample values, S = the standard deviation of logarithms, K = the deviation from the mean $(x-\bar{x})/S$ (in the standard deviation units) of variable values with the exceedance probability P (unadjusted).

The Guidelines by Water Resources Council suggest a series of analytical and statistical refinements to improve the accuracy of frequency curves obtained by that procedure. Such a refinement is the elimination of the bias in relation to the average future expectation, by adjusting the exceedance probability P to an expected exceedance probability which accounts for the actual sample size. Another refinement relates to the skewness coefficient. Since hydrologic records are usually shorter than 100 years, the sample estimates of this coefficient are unreliable. Specifically, if records available are of 100 years or more, the station skewness coefficient should be used exclusively. For records of 25 to 100 years, a weighted skewness coefficient should be calculated in which the station skewness is given the weight of (N-25)/75, where N = the length of record, and the generalized skewness is given a weight of [1.0 - (N-25)/75]. Guidelines also provide adjustments for zero flow, incomplete records and the treatment of outliers.

A problem of increasing interest in flood frequency analysis is the reliable estimation of skewness coefficients of historical flood records. The result of experiments made by Matalas, Slack, and Wallis (1975), further commented by Klemes (1976), has shown that in applying the concept of regionalizing (and, even more, contouring; Hardison, 1974) the skewness coefficient of annual flood peaks has a serious fault. In their concluding remarks, Matalas, Slack, and Wallis (1975) caution that the regional estimates of the skewness coefficients should be conditioned on the record length N, because of the bias and boundness of the small sample extimates. More cautions on hydrologic grounds were advanced by Klemes (1976). First, that the regional estimates of skewness coefficients should be conditioned also on basin area and physiographic features. Second, that the skewness coefficients of annual flood peaks are likely to vary along the course of a river, with reversals in the direction of change. Similarly, the skewness coefficient of a tributary may be very different from that of the main river. Third, it follows that the regional estimates of skewness coefficient, even though they depend on sample size, basin area and main physiographic features, reflects only an overall average tendency of the regional skewness coefficient, so they cannot be expected to be good estimates for individual basins or gaging stations. They should not be used as design standards for assessing flood frequencies at individual sites.

Natural Environment Research Council (1975) of the United Kingdom adopted the general extreme value distribution (of which the Gumbel distribution is a special case) to achieve standardization of flood-frequency procedures used in the United Kingdom. As reported by the Council, seven distribution functions were tested by calculating the goodness of fit indices for 28 stations with 30 years or more of records in Great Britain and for seven stations of between 23 and 44 years in Ireland. The result of the test showed that the Pearson Type III and the log-Pearson Type III functions were sensitive to the formulation of tests, and their goodness-of-fit changed places in the order of merit when the type of test was changed. The general extreme value distribution was more stable, and for this and other reasons it was recommended as the first choice among distributions of annual flood peaks by the Council. However, when only a small sample is available, say N less than 25, the Council recommended that the Gumbel distribution be fitted if an estimate based on the sample data alone is required. However, it should not be used for gross extrapolations, because on the average this leads to an underestimation of peaks for high return periods.

Another distribution, which consists of a mixture of two distributions, was suggested for annual flood peaks by Singh and Sinclair (1972). If the annual flood peaks could be classified in some objective manner into two groups, between which there is a noticeable difference in the distribution of variate values, then the concept of a mixture of distributions may be useful. For example, the annual flood peaks might be classified according to whether they arise from thunderstorm rainfall or from other types of precipitation, or from snowmelt. This application by Singh and Sinclair is a device for introducing a five parameter distribution, while previously only two and three parameter distributions have been used. A mixture of two normal distributions, applied to logarithms of annual flood peaks, was proposed. However, this did not require a classification of flood records into two types, and the estimation of parameters of each component distribution and of the mixture parameter separately. The proposed method of estimation of parameters was by using only means and variances. They concluded that for the medium to high floods (of greatest interest to engineers and hydrologists) the prediction was satisfactory by this method.

2.3 Theories of Probability Distributions of Partial Series of Flood Peaks

The standard approach to the analysis of flood peaks consists roughly either by applying the limiting distributions of the maximum value in a sequence of independent, identically distributed random variables, such as Gumbel's approach, or simply by testing which theoretical distribution best fits the observed frequency distributions of annual flood peaks. A different approach to the problem of flood peak analysis is to use a stochastic model for the description and analysis of excessive stream flows, or the partial flood series.

Borgman (1963) discussed the meaning and implication of the return period. He proposed the risk criteria such as the encounter probability, distribution of the waiting time, distribution of the total damage, probability of zero damage, and the mean total damage. Each criterion was derived from three mathematical simplifications of the actual physical and engineering situation.

Shane and Lynn (1964) developed a probability model based on the time independent Poisson process and the theory of sums of a random number of random variables for using in the analysis of base-flow flood data. From the model, design equations were derived relating several commonly used measures of risk to the design discharge: recurrence interval distribution, encounter probability and expected recurrence interval. Furthermore, Shane and Lynn (1969) developed confidence limits along with a lower bound for the corresponding level of confidence for evaluating the effect of sampling errors on flood risk evaluation from base-flow flood data.

Kirby (1969) defined flood peaks as successes or exceedances in a sequence of randomly spaced Bernoulli trials, each representing the occurrence of a hydrograph peak. An arbitrary criterion for distinguishing between floods and ordinary hydrograph peaks was used. His model showed that, at sufficiently small exceedance probabilities, the probability distributions of times between exceedances and the number of exceedances approach those implied by trials from a Poisson process.

Although the theory of extreme values has been extended beyond Gumbel's distribution function, its applications to flood frequency analysis have been limited to that distribution, except for the applications made by Todorovic and his co-workers (Todorovic, 1970; Todorovic and Zelenhasic, 1970; Todorovic and Rousselle, 1971; Todorovic and Woolhiser, 1972), and Gupta, Duckstein, and Peebles (1976). Gumbel's distribution stems from applying the classical extreme value theory to a complete series (such as daily flows). As mentioned before, the mathematical assumptions underlying the classical extreme value theory are not applicable to most flood problems. However, the theory developed by Todorovic and his co-workers may be more meaningful for flood frequency analysis than the classical extreme value theory.

The first attempt to develop a theory by Todorovic (1970), Todorovic and Zelenhasic (1970), was based on stream flow partial duration series. The series of flows in a partial duration series within an arbitrary but fixed time interval is represented by a random number of random variables. The time dependent Poisson process was used to describe the distribution of the random number of exceedances. It is applied to stream flow by further assuming that the individual exceedances form a sequence of identically independent random variables which are represented by an exponential distribution. However, the theory is sufficiently general as to treat also the non-identically distributed exceedances. In addition, it is applicable over any arbitrary time interval of interest, such as season or a year.

From a physical point of view, this method appears more feasible for flood peaks than the classical extreme value theory for two reasons. First, when the truncation level which defines a partial flood peak series is taken adequately high, the assumption of stochastic independence among individual exceedances becomes reasonable. Second, the assumption that the number of exceedances in a fixed time interval is a random variable allows this approach to be applied to an arbitrary time interval, which is not true for the classical extreme value theory.

The extension of the above approach to flood frequency analysis by Todorovic and Rousselle (1971) was by realizing that for a time interval equal to a year the assumption for exceedances being identically distributed is unrealistic, since different storm types can produce different flood characteristics from one season to another. Accordingly, they derived a distribution function for the largest flood peak for the case where two or more different exceedance distribution functions occur within a time interval.

By considering the application of this approach for deriving the distribution function of the largest exceedance in a time interval, the question requiring attention is the independence of the event that exactly k exceedances occur in a given time interval and the event that all those k exceedances are less than or equal to the specified value. In other words, the question is whether the magnitude of those exceedances are independent of the number of exceedances in a time interval. The case pertaining to previous works by many authors is that the magnitude of exceedances are independent of the number of exceedances (Todorovic and Zelenhasic, 1970).

Todorovic (1971) used the above method, together with the mathematical assumptions of Todorovic and Zelenhasic (1970), to derive another important property of the extreme flood, namely, its time of occurrence within a selected time interval. The expression for the time of occurrence of the extreme flood obtained by Todorovic (1971) is exact. It was tested on two rivers in the United States by Todorovic and Woolhiser (1972). Gupta, Duckstein, and Peebles (1976) extended the work by Todorovic and Woolhiser (1972) and developed the expression for the joint distribution function of the largest flood peak and its time of occurrence. They also modified this expression, valid for the case of identically independent exceedances, to the case of independent but non-identically distributed exceedances.

2.4 Relationship between Annual and Partial Flood Series

The empirical relationship between the probability of annual flood series and the expectancy of partial flood series was investigated by Langbein (1949) and the corresponding relationship was derived by Chow (1950). Let P_p be the expectancy of a variate in the partial flood series being equal to or greater than x, and let m be the average number of events per year, or mN be the total number of events in N years of record. Then P_p/m is the probability of an event being equal to x or greater, and $1 - P_p/m$ is the probability of an event being less than x. Thus the probability of an event of magnitude x becoming a maximum of the m events in a year is $(1 - P_p/m)^m$. The probability approaches $\exp(-P_p)$ when P_p is small compared with m. Hence, the probability P_a of an annual flood series of magnitude x being equal to or exceeded is

$$P_a = 1 - \exp(-P_p)$$

or

$$P_p = -\ln (1 - P_a)$$

in which P_p approaches P_a as both P_p and P_a become small. The recurrence intervals in partial flood series are smaller than in annual flood series, but the differences become negligible for floods greater than about a five-year recurrence interval (Langbein, 1949).

In more mathematical terms, if partial flood models and annual flood models are derived under specified assumptions, and if these models are accepted, the theoretical relationship between annual and partial flood series may be derived. When the Poisson distribution for the number of exceedances in a year and the exponential distribution for the magnitude of these exceedances are assumed, combined they give a double exponential or Gumbel distribution of annual flood peak series (Zelenhasic, 1970). The double exponential distribution of annual flood series is an exact distribution, derived from the model of partial flood series under commonly used assumptions.

Cunnane (1973) used the above relationship for comparing the statistical efficiency of estimates, Q(T), of the T-year flood by using the annual and partial flood series. On the basis of commonly used assumptions, he concluded that the estimate of Q(T) of annual exceedance series (i.e., the exceedance series that has the average number of exceedances per year equal to one) has a larger sampling variance than the annual flood series estimate for the return periods greater than 10 years. For the same range of return periods the estimates of Q(T) of partial flood series have a smaller sampling variance than the estimates from annual flood series only if the partial flood series contained at least 1.65 N items, with N = the number of years of record. These results are based on a theoretical approach as well as on general assumption that the distribution of annual flood series is exactly a Gumbel distribution and the partial flood series is represented by the combination of the Poisson distribution for the number of exceedances and the exponential distribution for the magnitude of those exceedances.

2.5 Modeling Daily Flow Series

In the analysis of time series, their structure can be considered to be a combination of three components: trend component, periodic or cyclic component, and stochastic component. The trend component may occur as a result of either man-made changes within the watershed or by natural causes. The presence of a periodic component is attributed to astronomical cycles. The dependence among the successive values of the stochastic component are usually described by a deterministic model plus the independend stochastic component. If the trend does not exist or is not significant, the general structural model reduces to a combination of periodic parameters and a stochastic component. The description and separation of these deterministic and stochastic components of hydrologic time series are described by Yevjevich (1972c). If the time series can be separated into components, the generation of their new samples can be carried out by the reversed procedures.

Roesner and Yevjevich (1966) used a seasonal model for generation of monthly flows. The annual periodicities were in the mean and standard deviation of the series, while the dependence of stochastic components was fitted by the Markov models. Harmonics of periodic parameters were inferred by spectral analysis, and described by Fourier series. Quimpo (1967) followed a similar representation for daily flows. He applied this approach to daily runoff records from 17 gaging stations in the United States, and found that all the series of stochastic components satisfied approximately the second order autoregressive model.

Tao, Yevjevich and Kottegoda (1976), and also Tao (1973), using the same data as Quimpo (1967), made an extensive study of fitting the distribution functions to independent stochastic components for different time intervals of series. The important conclusions are: (1) In case of independent stochastic components of daily flow series, none of the probability distribution functions currently used for fitting the frequency distributions could pass the chi-square and Smirnov-Kolmogorov tests; (2) The double-branch gamma function gave the best goodness-of-fit among the distribution functions tested; (3) The logarithmic transformation provides some improvements in the analysis by assigning different weights to values of the original series and by reducing flow fluctuations in comparison with the original series; and (4) Errors in determining the number of significant harmonics and errors in estimating their Fourier coefficients greatly affect the accuracy of inferred periodic functions.

The use of a similar model for generating daily flows on British data was studied by Hall and O'Connell (1972). They transformed the original series by taking natural logarithms of daily flows and performing an analysis on the transformed series. Six harmonics were required to describe the periodic daily means and standard deviations. First-order Markov models, with a lognormal random component, were found to fit the stochastic component. They faced the same problems as Quimpo (1967), and Tao, Yevjevich and Kottegoda (1967), namely that it is difficult to find a good distribution function of independent stochastic component, as tested by commonly used test statistics. By generating new sequences of daily flows, equal in length to the historical record, they found that the daily means and daily standard deviations, as well as the flow duration curves of the generated flows, were remarkably similar to those of the historical data. However, during the summer half-year, a lesser fluctuation was apparent in the daily standard deviation of generated data.

So far, the classical hydrologic analysis by using Markovian or other linear models, with periodic parameters, has been successful in generating stream flow series with a long time interval, such as for weekly and monthly flows. Extensions of these models to daily flows have met a limited success. This is mainly due to high variation of flows, unconventional probability distributions of independent stochastic components, and a failure to simulate processes to transfer hydrograph characteristics into the historical flows (Kottegoda, 1972). Investigations on the origin of these failures in case of daily flow series were undertaken by Vargas (1977). By using the data generation method, he systematically checked each of the stages of modeling and estimation of model parameters, with the purpose of assessing whether failures originated by biases in estimation procedures, or by inappropriate models. He concluded that the inference on the number of significant harmonics in periodic parameters affected all stages of estimation. The underestimation of the number of harmonics in periodic daily means and standard deviations led to a rejection of the hypothesis of independence of stochastic components in the dependence models, while the overestimation seemed to have no effect. The estimation procedures are sensitive to the type of distribution used for the stochastic component. Procedures initially developed for the normal distribution are not sufficiently robust to be applied to non-normal dependent variables, especially those of highly skewed distribution functions.

The more or less similar approaches to the above described method have been proposed for generating daily stream flows. Green (1973) proposed the method based on the linear interpolation for the logarithms of 5-day average flows. The 5-day average flows were produced by using Kottegoda's model (1972). Beard (1967) used the procedure based on generation of monthly stream flows and subsequent allocation of the monthly total amount to each day. The daily flows were generated for those months, when flow fluctuations within a month were important. The daily flows generator consisted of a 2-pass generation by the use of a second-order Markov chain applied to standardized variates derived from a log-Pearson Type III distribution. He used a linear regression of the standard deviation of daily flow logarithms, within each month of record, and the logarithm of total flow for the month.

Natural Environmental Research Council (1975) of the United Kingdom studied the application of the shot noise model for generating daily flows. The flow was considered as the sum of a series of random impulses. Each impulse consisted of a sudden random rise of height Y which decayed exponentially. These impulses occurred as a Poisson process. The impulse height Y is a random variable which may be represented by an exponential, gamma, or a special form of the Pareto distribution.

Kelman (1977) developed a model which takes into consideration the diversity of physical factors that produce the stream flow. He divided the daily stream flow record Q_t into two sequences according to the increments $(Q_t - Q_{t-1})$. The positive increments, which assumed to be produced by bursts of surface and subsurface flow were characterized by a weak persistence. The negative increments were the consequence of watershed emptying process, and hence had a strong persistence. He represented the sequence of positive increments by a power transformed, truncated normal distribution with the first-order autoregressive model. The sequence of negative increments was obtained by assuming that recession discharges were a stochastic output of two linear reservoirs.

The literature is full of other approaches to generation of new samples of daily flow series. Only those have been reviewed herein, which have an influence on the content of this study.

Chapter III
THEORY OF PROBABILITY DISTRIBUTIONS OF PARTIAL SERIES OF FLOOD PEAKS

The objective of this chapter is to develop probability distributions of largest exceedances above selected truncation levels for the time interval of a year. These distributions estimated from partial flood series can then be used to estimate flood exceedances for given return periods. The approach by Todorovic and Zelenhasic (1970) is used as the basis in this study. Discrete and continuous distribution functions to be used in fitting the frequency distributions of the number and the magnitude of exceedances, respectively, are described. In addition a method used for comparing sampling variances of annual flood peaks of given return periods obtained from annual and partial flood series is presented.

The application of the theory of probability distributions of partial series of flood peaks to observed daily flows of 17 gaging stations in the United States is presented in Chapter V.

3.1 Phenomenological Considerations

According to Kirby (1969), any stream flow hydrograph can be interpreted as a sequence of nearly instantaneous hydrograph peaks separated by relatively longer periods of low flows. Because of the nature of the phenomenon, the number of these peaks in a given interval of time (0,t) and their magnitudes are random variables.

For a given truncation level Q_b, consider only those separate flood peaks Q_i in the time interval (0,t) that exceed Q_b (Fig. 3-1). It is necessary to define the separate flood peaks for the partial flood series. The definition normally depends on frequency analysis and the stream characteristics. As suggested by U.S. Water Resources Council (1976), the separate

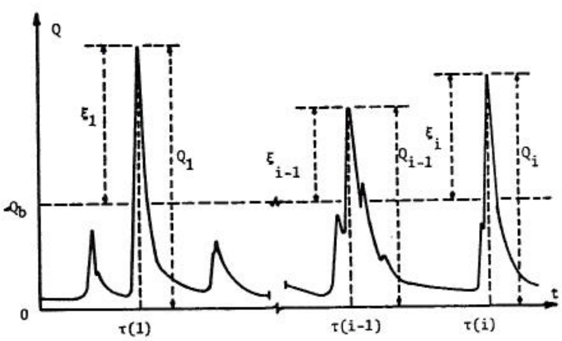

Fig. 3-1 Schematic Representation of a Stream Flow Hydrograph

flood peaks are arbitrarily defined as flood peaks separated by at least as many days as five plus the natural logarithm of the square miles of drainage area, with the requirements that the intermediate flows must drop below 75 percent of the lower of the two separate flood peaks. This criteria is used as the guideline in this study. However, in case of a river with highly fluctuating daily flow hydrograph, the time between the two successive flood peaks is taken to be somewhat less than that suggested by the Water Resources Council, since the intermediate flows drop much below the 75 percent of the lower of the two separate flood peaks. In any case, flood peaks can and are assumed to be precisely defined. By such definition, the separate flood peaks associated with a given truncation level, Q_b , are still separate flood peaks for the truncation level $Q_b' < Q_b$. In other words, the number of separate flood peaks above a given truncation level Q_b is a non-increasing function of Q_b. The number of separate flood peaks are the same for the various truncation levels that are smaller than the minimum flow of the considered time interval. For example, the number of separate flood peaks for the truncation level Q_b shown in Fig. 3-2 is 4, while for the truncation level Q_b' is 10.

Let define

$$\xi_i = Q_i - Q_b \tag{3-1}$$

in which $\xi_i > 0$ is a random variable for all i = 1,2,... With each ξ_i the time $\tau(i)$, when the corresponding separate flood peak has occurred (Fig. 3-1), is associated. The separate flood peak exceedance flows, ξ_i, from now on will be called the exceedances.

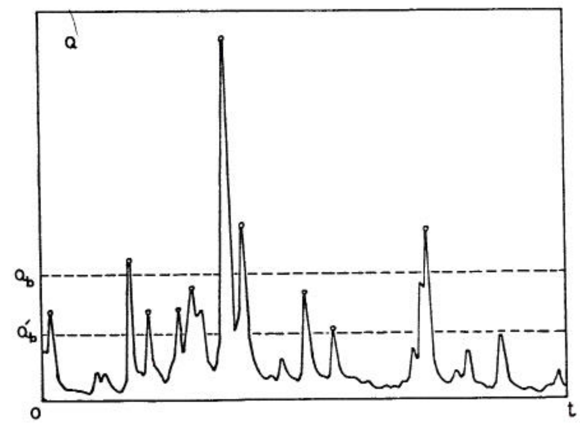

Fig. 3-2 Example of Extracting Partial flood Series from Daily Flow Hydrograph

Consider an interval of time (0,t) and denote by $\chi(t)$ the largest ξ_ν in this time interval. Since the number of ξ_ν in (0,t) is a random variable that depends on time t, $\chi(t)$ is defined as:

$$\chi(t) = \max_{\tau(\nu) \leq t} \xi_\nu \tag{3-2}$$

By virtue of definition it follows that for every $t \geq 0$ and $\Delta t > 0$

$$\chi(t) \leq \chi(t + \Delta t) \tag{3-3}$$

This implies that $\chi(t)$ is a stochastic process of non-decreasing sample functions.

In the following an attempt is made to determine a distribution function $F_t(x)$ of the stochastic process $\chi(t)$,

$$F_t(x) = P(\chi(t) \leq x) \tag{3-4}$$

However, before going into the derivation of this distribution, distributions of the number and the magnitude of exceedances must be developed.

It is important to note that a year is considered in this study as the time interval. The purpose of using the partial flood series is to include more data into analysis, especially in case of small sample sizes. If the year is divided into different time intervals such as seasons, it may not be feasible to analyze the distribution of the magnitude of exceedances which occur during a season such as summer, because of the small number of exceedances that may occur during that season. Since the number of exceedances is small, the estimated distribution parameters for the magnitude of exceedances may be unreliable. Therefore, the time interval of a year is considered and ξ_ν are assumed to be identically distributed random variables throughout the year.

3.2 Distributions of the Number of Exceedances

Let η be denoted as the number of exceedances in the time interval of a year. By definition η is a non-increasing function of the truncation level Q_b.

Denote $E_\nu = (\eta=\nu)$ then it follows that

$$E_i \cap E_j = \phi \quad \text{for all} \quad i \neq j \quad \text{and} \quad \bigcup_{\nu=0}^{\infty} E_\nu = \Omega,$$

where ϕ stands for the impossible event and Ω stands for the certain event. Hence E_ν, for $\nu = 0,1,2,\ldots$, is a discrete event representing a countable partition of Ω, and

$$P(E_\nu) = P(\eta=\nu) \tag{3-5}$$

is the probability that exactly ν exceedances occur in a year.

Following the previous works by several authors, the Poisson distribution has been widely used to fit frequency distributions of η. The mean is equal to variance in the Poisson distribution. Inspection of partial flood series obtained from the mean daily flow data for 17 gaging stations used in this study indicated that many series have the ratios of mean to variance far from unity. The reason is that the Poisson distribution has only one parameter and may not be sufficiently flexible to fit frequency distributions of η for all cases of the study. Furthermore, since the partial flood series is obtained from the daily flow series instead of from series of instantaneous discharges, the distribution of η may depart more or less from the Poisson distribution. Hence, the selection of the best discrete probability distribution to be used in fitting frequency distribution of η is needed and is studied by using the records of 17 daily flow series in the United States. The selected distributions for study, which are more or less similar to the Poisson distribution, their important properties and the method of estimating their parameters, are:

Poisson Distribution. The probability density function is

$$f(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!} \quad , \quad x = 0,1,2,\ldots, \tag{3-6}$$

with $\lambda>0$ as a parameter. The mean and variance are equal, or $E(x) = \text{var } x = \lambda$. The maximum likelihood estimate of λ is $\bar{x}$, or the mean of all the numbers of exceedances.

Mixed Poisson Distribution. Let $\lambda_1>0$ and $\lambda_2>0$ be parameters of two Poisson distributions, that are mixed in proportions p and 1-p, respectively. The probability density function of the mixed distribution is

$$f(x;p,\lambda_1,\lambda_2) = p\,\frac{e^{-\lambda_1}\lambda_1^x}{x!} + (1-p)\,\frac{e^{-\lambda_2}\lambda_2^x}{x!}\ ,\quad x = 0,1,2,\ldots, \tag{3-7}$$

where, without any loss of generality, $\lambda_1>\lambda_2$. The k-th factorial moment of x is

$$\nu_{(k)} = p\lambda_1^k + (1-p)\lambda_2^k \tag{3-8}$$

The mean and the variance are $E(x) = p\lambda_1 + (1-p)\lambda_2$, and $\text{var } x = p\lambda_1 + (1-p)\lambda_2 + p(1-p)(\lambda_1-\lambda_2)^2$. Hence, the mean is always smaller than the variance.

For this particular distribution, the maximum likelihood estimation is complicated. Two methods of estimation of parameters p,λ_1 and λ_2 are considered (Cohen, 1963). One is based on the first three sample moments, and the other on the first two sample moments and the sample frequency of zero.

For the first method, the estimates of p, λ_1, λ_2 are obtained by using the equations

$$\left.\begin{aligned} \hat{\lambda}_1 &= \tfrac{1}{2}\left(\theta + \sqrt{\theta^2 - 4\gamma}\right) \\ \hat{\lambda}_2 &= \tfrac{1}{2}\left(\theta - \sqrt{\theta^2 - 4\gamma}\right) \\ \text{and} \quad \hat{p} &= \frac{\bar{x} - \hat{\lambda}_2}{\hat{\lambda}_1 - \hat{\lambda}_2} \end{aligned}\right\} \tag{3-9}$$

with $\bar{x}$ the mean, and θ and γ defined as

$$\left.\begin{aligned} \theta &= \frac{\nu_{(3)} - \bar{x}\nu_{(2)}}{\nu_{(2)} - \bar{x}^2} \\ \text{and} \\ \gamma &= \bar{x}\theta - \nu_{(2)} \end{aligned}\right\} \tag{3-10}$$

The k-th factorial moment of x can be determined from the data by

$$\nu_{(k)} = \frac{1}{n}\sum_{x=0}^{m} x(x-1)\ldots(x-k+1)\,n_x \quad , \tag{3-11}$$

in which m = the largest observed value of x, n_x = the sample frequency of x, and n = the total sample size, i.e., n is the sum of m+1 values of n_x.

Estimates of p, λ_1, and λ_2 based on the first two sample moments and the sample frequency of zero are obtained by solving first the following equation for $\hat{\lambda}_2$ by an iterative procedure,

$$\frac{\bar{x} - \hat{\lambda}_2}{G(\hat{\lambda}_2) - \hat{\lambda}_2} = \frac{\frac{n_0}{n} - \exp(-\hat{\lambda}_2)}{\exp[-G(\hat{\lambda}_2)] - \exp(-\hat{\lambda}_2)} , \quad (3\text{-}12)$$

in which n_0 = the number of zero observations in the sample, $G(\hat{\lambda}_2)$ = a function of $\hat{\lambda}_2$, expressed by

$$G(\hat{\lambda}_2) = \frac{\nu_{(2)} - \bar{x}\hat{\lambda}_2}{\bar{x} - \hat{\lambda}_2} = \hat{\lambda}_1 \quad . \quad (3\text{-}13)$$

With $\hat{\lambda}_2$ determined from Eq. 3-12, $\hat{\lambda}_1$ follows from Eq. 3-13 as

$$\hat{\lambda}_1 = \frac{\nu_{(2)} - \bar{x}\hat{\lambda}_2}{\bar{x} - \hat{\lambda}_2} , \quad (3\text{-}14)$$

and $\hat{p}$ follows from the third expression of Eq. 3-9 as

$$\hat{p} = \frac{\bar{x} - \hat{\lambda}_2}{\hat{\lambda}_1 - \hat{\lambda}_2} \quad . \quad (3\text{-}15)$$

Hyper-Poisson Distribution. The probability density function is

$$f(x;\lambda,\theta) = \frac{\Gamma(\lambda)\ \theta^x}{F_1(1;\lambda;\theta)\Gamma(\lambda+x)} , \quad x = 0,1,2,\ldots, \quad (3\text{-}16)$$

where

$$F_1(1;\lambda;\theta) = 1 + \frac{\theta}{\lambda} + \frac{\theta^2}{\lambda(\lambda+1)} + \frac{\theta^3}{\lambda(\lambda+1)(\lambda+2)} + \cdots \quad (3\text{-}17)$$

is the confluent hypergeometric function with first argument equal to 1, and λ and θ parameters.

The distribution of Eq. 3-16 may be classified according to $\lambda=1$, $\lambda>1$, or $0<\lambda<1$. If $\lambda=1$, it reduces to the Poisson distribution. If $\lambda>1$, the variance exceeds the mean, and the distribution has been called "Super Poisson." If $0<\lambda<1$, the variance is exceeded by the mean, and the distribution has been called "Sub Poisson."

The mean of Eq. 3-16 is given by

$$\mu = \alpha_1 = \theta + (1-\lambda)\ (1-f_0) , \quad (3\text{-}18)$$

in which $f_0 = 1/F_1(1;\lambda;\theta)$. The higher moments about the origin are given recursively by

$$\alpha_{j+1} = (\theta-\lambda+1)\ \alpha_j + \theta[j\alpha_{j-1} + \binom{j}{2}\ \alpha_{j-2} + \cdots + \binom{j}{k}\ \alpha_{j-k} + \cdots + j\alpha_1 + 1],\ j = 1,2,\ldots \quad (3\text{-}19)$$

The variance is given by

$$\mu_2 = \theta(1+\mu) + \mu(1-\mu-\lambda) \quad (3\text{-}20)$$

The methods of estimation of parameters λ and θ are summarized as follows (Crow and Bardwell, 1963; Bardwell and Crow, 1964). The maximum likelihood estimates can be obtained by solving the equations

$$\left.\begin{aligned} &\frac{\hat{\theta}}{F_1}\frac{\partial F_1}{\partial\hat{\theta}} = \bar{x} \\ &\frac{1}{F_1}\frac{\partial F_1}{\partial\hat{\lambda}} - \Psi(\hat{\lambda}) + \frac{1}{n}\sum_{i=1}^{n}\Psi(\hat{\lambda}+x_i) = 0 \quad , \end{aligned}\right\} \quad (3\text{-}21)$$

where $\Psi(\hat{\lambda}) = \frac{\partial}{\partial\hat{\lambda}}\ln\Gamma(\hat{\lambda}) = \frac{\Gamma(\hat{\lambda})'}{\Gamma(\hat{\lambda})}$ is the digamma function.

The two-moment estimates can be obtained by solving the following equation for $\hat{\theta}$ by an iterative procedure,

$$(1+\bar{x})\hat{\theta} = \alpha_2 + (\hat{\theta}-\alpha_2+\bar{x}^2)\ F_1[1;\frac{1}{\bar{x}}\{(1+\bar{x})\hat{\theta}+\bar{x}-\alpha_2\};\hat{\theta}]. \quad (3\text{-}22)$$

and then $\hat{\lambda}$ is found by using the equation

$$\hat{\lambda} = \frac{1}{\bar{x}}\ [(1+\bar{x})\hat{\theta} + \bar{x} - \alpha_2] \quad . \quad (3\text{-}23)$$

The explicit and simple modified moment estimates of λ and θ may be obtained in using the first three moments about the origin, α_j, $j = 1,2,3$, by

$$\left.\begin{aligned} &\hat{\theta} = \frac{\alpha_1\alpha_3 - \alpha_2^2}{2\alpha_1^2 + \alpha_1 - \alpha_2} \\ &\text{and} \\ &\hat{\lambda} = \hat{\theta} - \hat{\theta}' + 1 \end{aligned}\right\} \quad (3\text{-}24)$$

in which $\hat{\theta}'$ = a three-moment estimate of a convenient parameter

$$\hat{\theta}' = \frac{2\alpha_1\alpha_2 + \alpha_2 - \alpha_3}{2\alpha_1^2 + \alpha_1 - \alpha_2} \quad . \quad (3\text{-}25)$$

Modified moment estimates using the frequency for x=0 are obtained immediately by

$$\left.\begin{aligned} &\hat{\theta} = \frac{(1-f_0)\ \alpha_2 - \alpha_1^2}{1 - f_0\ (\alpha_1+1)} , \\ &\hat{\theta}' = \frac{\alpha_1 - f_0\ \alpha_2}{1 - f_0\ (\alpha_1+1)} , \\ &\text{and} \\ &\hat{\lambda} = \hat{\theta} - \hat{\theta}' + 1 \end{aligned}\right\} \quad (3\text{-}26)$$

Negative Binomial Distribution. The probability density function is

$$f(x;r,p) = \binom{r+x-1}{x} p^r q^x = \frac{\Gamma(x+r)\ p^r q^x}{x!\ \Gamma(r)}\ ,\ x = 0,1,\ldots, \qquad (3\text{-}27)$$

in which $r>0$, $0\leq p\leq 1$, and $q=1-p$.

The mean and variance are

$$\left.\begin{aligned} E(x) &= \frac{rq}{p} \\ &\text{and} \\ \operatorname{var} x &= \frac{rq}{p^2} \end{aligned}\right\} \qquad (3\text{-}28)$$

Hence, its mean is smaller than its variance.

The maximum likelihood estimates of parameters r and p can be obtained by solving the equations

$$\hat{p} = \frac{1}{1 + \bar{x}/\hat{r}}$$

and

$$\ln\hat{p} = \Psi(\hat{r}) - \frac{1}{n}\sum_{i=1}^{n}\Psi(x_i+\hat{r}) = -\frac{1}{n}\sum_{i=1}^{n} S_{x_i}\ , \qquad (3\text{-}29)$$

in which

$$S_{x_i} = \sum_{j=1}^{x_i} \frac{1}{r+j-1}\ ,\ x_i = 1,2,3,\ldots,$$

with n = the sample size, $\bar{x}$ = the mean of data values. The parameter r can be estimated by using an iterative procedure of the equation

$$\ln\left[\frac{1}{1 + \bar{x}/\hat{r}}\right] = -\frac{1}{n}\sum_{i=1}^{n} S_{x_i} \qquad (3\text{-}30)$$

Mixture of Two Geometric Distributions. A mixture distribution of two geometric distributions has the left side with a truncated geometric distribution and the right side with a standard geometric distribution. Guerrero-Salazar and Yevjevich (1975) used this distribution to fit frequency distributions of the longest run-length in case of samples of given sizes. The probability density function of this mixed distribution is

$$f(x;\alpha,\gamma,\theta_1,\theta_2) = \frac{\alpha(1-\theta_1)\theta_1^{\gamma-x}}{1-\theta_1^{\gamma+1}} I_{\{0,1,2,\ldots,\gamma\}} + \frac{(1-\alpha)\theta_2(1-\theta_2)^x}{(1-\theta_2)^{\gamma+1}} I_{\{\gamma+1,\ldots\ .\}} \qquad (3\text{-}31)$$

with θ_1 and θ_2 = the parameters of each part, respectively, γ = a location parameter and α = a partition parameter. The location γ is estimated either by the mode $\hat{\gamma}$ = m or by $\hat{\gamma}$ = m-1, α by $\hat{\alpha} = \sum_{i=1}^{\gamma} p_i$, θ_2 by $(\bar{x}_2 - \hat{\gamma})^{-1}$ with $\bar{x}_2$ the mean of sample values greater than or equal to $\hat{\gamma}+1$, and θ_1 by an iterative solution of the equation

$$\frac{1}{1-\hat{\theta}_1} = \frac{\hat{\gamma} - \bar{x}_1}{\hat{\theta}_1} + \frac{(\hat{\gamma}+1)\hat{\theta}_1^{\gamma}}{1-\hat{\theta}_1^{\gamma+1}}\ , \qquad (3\text{-}32)$$

with $\bar{x}_1$ = the mean of sample values which are smaller than or equal to $\hat{\gamma}$.

3.3 Distributions of the Magnitude of Exceedances

The other distributions that require investigations are the common distribution functions, H(x), i.e., $H(x) = P(\xi_\nu \leq x)$, of all exceedances ξ_ν, $\nu = 1,2,\ldots$, in a year. Presently, few theoretical grounds indicate the forms of distributions of all exceedances. Two probability functions have played an important role for the magnitude of flood peaks: gamma and exponential (Zelenhasic, 1970). In previous works of several authors the exponential distribution has been widely used in fitting frequency distributions of ξ_ν. Sufficient evidence does not exist to indicate the exponential distribution to be universally applicable. For that reason, and similar reasons in the case of the use of Poisson distribution for the number of exceedances, several distribution functions are selected to study their fits to frequency distributions of ξ_ν. The results are then compared with the goodness-of-fit statistics with that of the exponential distribution, in order to find out the best probability distribution functions for ξ_ν.

The selected continuous probability distribution functions for the study, their properties and the parameter estimation are:

Exponential Distribution. The Pearson Type III distribution function has three parameters, denoted by x_0, β and γ. The special case occurs when the lower bound $x_0 = 0$, giving the two-parameter gamma distribution. Another case arises when $\gamma = 1$ and $x_0 = 0$, giving the one-parameter exponential distribution. The probability density function of exponentially distributed random variables is

$$f(x;\beta) = \frac{1}{\beta} e^{-x/\beta},\quad x > 0 \qquad (3\text{-}33)$$

The mean and variance are

$$\left.\begin{aligned} \mu &= \beta \\ \sigma^2 &= \beta^2 \end{aligned}\right\} \qquad (3\text{-}34)$$

The moment and maximum likelihood estimates take the same form as

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x} \quad , \qquad (3\text{-}35)$$

with n = the sample size.

Gamma Distribution. The probability density function is

$$f(x;\beta,\gamma) = \frac{x^{\gamma-1} e^{-x/\beta}}{\beta^{\gamma} \Gamma(\gamma)} , \quad x > 0 \quad , \qquad (3\text{-}36)$$

in which $\Gamma(\gamma)$ = the complete gamma function, β and γ are scale and shape parameters, respectively. The mean and variance are $\mu = \beta\gamma$ and $\sigma^2 = \beta^2\gamma$.

The moment estimates of β and γ are obtained by

$$\left.\begin{aligned} \hat{\gamma} &= \frac{\bar{x}^2}{\hat{\sigma}^2} \\ \text{and} \quad \hat{\beta} &= \frac{\hat{\sigma}^2}{\bar{x}} \quad , \end{aligned}\right\} \qquad (3\text{-}37)$$

in which $\bar{x}$ and $\hat{\sigma}$ = the sample mean and sample standard deviation, respectively.

The maximum likelihood estimates of β and γ are obtained by solving the two equations

$$\left.\begin{aligned} \hat{\beta} &= \frac{\bar{x}}{\hat{\gamma}} \\ \text{and} \quad \ln\bar{x} - \ln\hat{\gamma} + \Psi(\hat{\gamma}) - \frac{1}{n}\Sigma\ln x_i &= 0 \quad . \end{aligned}\right\} \qquad (3\text{-}38)$$

The method of solving the second expression of Eq. 3-38 for $\hat{\gamma}$ is approximated by

$$\hat{\gamma} = \frac{1 + (1+4A/3)^{1/2}}{4A} - \Delta\hat{\gamma} \quad , \qquad (3\text{-}39)$$

in which A is defined by $A = \ln\bar{x} - \frac{1}{n}\Sigma\ln x_i$, and $\Delta\hat{\gamma}$ is approximated by $\Delta\hat{\gamma} = 0.04475(0.26)^{\hat{\gamma}}$.

Pearson Type III Distribution. The probability density function is

$$\left.\begin{aligned} f(x;x_0,\beta,\gamma) &= \frac{(x-x_0)^{\gamma-1} e^{-(x-x_0)/\beta}}{\beta^{\gamma} \Gamma(\gamma)} , \\ x_0 &\le x \le \infty , \end{aligned}\right. \qquad (3\text{-}40)$$

with γ = the shape parameter, β = the scale parameter, and x_0 = the location parameter.

The mean, variance and skewness are

$$\left.\begin{aligned} \mu &= x_0 + \beta\gamma \quad , \\ \sigma^2 &= \beta^2\gamma \quad , \\ g &= \frac{2}{\sqrt{\gamma}} \quad . \end{aligned}\right\} \qquad (3\text{-}41)$$

If $\bar{x}$, $\hat{\sigma}$ and $\hat{g}$ are the sample estimates of mean, standard deviation and skewness coefficient, the moment estimates of γ, β, and x_0 are obtained by

$$\left.\begin{aligned} \hat{\gamma} &= \frac{4}{\hat{g}^2} \\ \hat{\beta} &= \frac{1}{2}\hat{g}\hat{\sigma} \\ \hat{x}_0 &= \bar{x} - \hat{\beta}\hat{\gamma} = \bar{x} - \frac{2\hat{\sigma}}{\hat{g}} \end{aligned}\right\} \qquad (3\text{-}42)$$

The approximate maximum likelihood estimate of the lower bound x_0 is obtained by solving the following equation by an iterative procedure (Tao, Yevjevich, and Kottegoda, 1976)

$$\frac{1 + (1+4A/3)^{1/2}}{1 + (1+4A/3)^{1/2} - 4A} - (\bar{x}-\hat{x}_0)\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i - \hat{x}_0} = 0 \quad , \qquad (3\text{-}43)$$

in which

$$A = \ln(\bar{x}-\hat{x}_0) - \frac{1}{n}\sum_{i=1}^{n}\ln(x_i-\hat{x}_0) \quad . \qquad (3\text{-}44)$$

Once $\hat{x}_0$ is determined, the parameter γ is estimated by

$$\hat{\gamma} = \frac{1 + (1+4A/3)^{1/2}}{4A} - \Delta\hat{\gamma} \quad , \qquad (3\text{-}45)$$

with A given by Eq. 3-44 and $\Delta\hat{\gamma}$ approximated by $\hat{\gamma} = 0.04475(0.26)^{\hat{\gamma}}$.

The parameter β is then estimated by

$$\hat{\beta} = \frac{1}{\hat{\gamma}}(\bar{x} - \hat{x}_0) \qquad (3\text{-}46)$$

Weibull Distribution. The two-parameter Weibull distribution has the probability density function

$$f(x;a,b) = abx^{b-1}\exp(-ax^b), \quad x > 0 \qquad (3\text{-}47)$$

with $a > 0$ and $b > 0$ parameters. If $b = 1$, this distribution becomes exponential, with the parameter a.

The mean and variance are

$$\left.\begin{aligned} \mu &= a^{-1/b}\,\Gamma(1+b^{-1}) \quad , \\ \sigma^2 &= a^{-2/b}\,[\Gamma(1+2b^{-1}) - \Gamma^2(1+b^{-1})] \end{aligned}\right\} \qquad (3\text{-}48)$$

The maximum likelihood estimate of b is obtained by solving the following equation by an iterative procedure

$$\frac{n}{\hat{b}} + \sum_{i=1}^{n}\ln x_i - \frac{n}{\sum_{i=1}^{n} x_i^{\hat{b}}}\sum_{i=1}^{n}(x_i^{\hat{b}}\ln x_i) = 0 \quad . \qquad (3\text{-}49)$$

Once $\hat{b}$ is determined, the parameter a is estimated by

$$\hat{a} = \frac{n}{\sum_{i=1}^{n} x_i^{\hat{b}}} \qquad (3\text{-}50)$$

Three-Parameter Lognormal Distribution. The probability density function is

$$f(x;x_0,\mu_y,\sigma_y) =$$

$$\frac{1}{(x-x_0)\sigma_y\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left\{\frac{\ln(x-x_0)-\mu_y}{\sigma_y}\right\}^2\right], \qquad (3\text{-}51)$$

where x_0, μ_y and σ_y are parameters.

If a variate x follows the lognormal distribution, $y = \ln(x-x_0)$ has a normal distribution with mean μ_y and variance σ_y^2. The probability density function of the transformed variate y is then

$$f(y;\mu_y,\sigma_y) = \frac{1}{\sigma_y\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left\{\frac{y-\mu_y}{\sigma_y}\right\}^2\right] \qquad (3\text{-}52)$$

The mean and variance of x are

$$\left.\begin{aligned} \mu_x &= x_0 + \exp\left[\mu_y + \frac{1}{2}\sigma_y^2\right], \\ \sigma_x^2 &= e^{(2\mu_y+\sigma_y^2)} \cdot (e^{\sigma_y^2} - 1) \end{aligned}\right\} \qquad (3\text{-}53)$$

Equation 3-51 is the three-parameter lognormal distribution function. It becomes the two-parameter lognormal distribution function for $x_0 = 0$. The maximum likelihood estimate $\hat{x}_0$ is obtained by solving the following equation by an iterative procedure (Tao, Yevjevich and Kottegoda, 1976)

$$\left[\sum_{i=1}^{n} \frac{1}{x_i-\hat{x}_0}\right]\left[\frac{1}{n}\sum_{i=1}^{n} \ln^2(x_i-\hat{x}_0) - \left\{\frac{1}{n}\sum_{i=1}^{n} \ln(x_i-\hat{x}_0)\right\}^2\right.$$

$$\left. - \frac{1}{n}\sum_{i=1}^{n} \ln(x_i-\hat{x}_0)\right] + \sum_{i=1}^{n} \frac{\ln(x_i-\hat{x}_0)}{x_i-\hat{x}_0} = 0 \qquad (3\text{-}54)$$

with $\hat{x}_0$ determined from Eq. 3-54, and the maximum likelihood estimates of μ_y and σ_y are obtained by

$$\mu_y = \frac{1}{n}\sum_{i=1}^{n} \ln(x_i-\hat{x}_0) \qquad (3\text{-}55)$$

$$\sigma_y^2 = \frac{1}{n}\sum_{i=1}^{n} [\ln(x_i-\hat{x}_0) - \mu_y]^2 \qquad (3\text{-}56)$$

Mixed Exponential Distribution. A mixture of two exponential distributions is composed of two populations of type of Eq. 3-33, with parameters β_1 and β_2, respectively, and mixed in unknown proportions p and 1-p. The resulting probability density function is

$$f(x;p,\beta_1,\beta_2) = p\,\beta_1^{-1} e^{-x/\beta_1} +$$

$$(1-p)\,\beta_2^{-1} e^{-x/\beta_2}, \quad x > 0 \qquad (3\text{-}57)$$

in which $0 \leq p \leq 1$, $\beta_1 > 0$, and $\beta_2 > 0$ as parameters. The r-th moment about the origin is expressed by

$$\mu_r' = E(X^r) = p\Gamma(r+1)\beta_1^r + (1-p)\Gamma(r+1)\beta_2^r \quad . \qquad (3\text{-}58)$$

Hence, the mean and variance are

$$\mu = p\beta_1 + (1-p)\beta_2 \quad ,$$

$$\sigma^2 = p\beta_1(2\beta_1 - p\beta_1) + (1-p)\beta_2(\beta_2 + p\beta_2) -$$

$$2p(1-p)\beta_1\beta_2 \qquad (3\text{-}59)$$

The estimation of parameters p, β_1, β_2 by using the maximum likelihood method is complicated. A simple method of estimation is by using the first three sample moments.

Let m_1', m_2' and m_3' denote the first three sample moments about the origin for a sample of Eq. 3-57. Estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ are obtained by solving the following equation (Rider, 1961)

$$6(2m_1'^2-m_2')\hat{\beta}_j^2 + 2(m_3'-3m_1'm_2')\hat{\beta}_j + 3m_2'^2 - 2m_1'm_3' = 0 \qquad (3\text{-}60)$$

with j=1 or 2.

The estimate of p is then obtained from

$$\hat{p} = \frac{m_1' - \hat{\beta}_2}{\hat{\beta}_1 - \hat{\beta}_2} \quad . \qquad (3\text{-}61)$$

The two roots of Eq. 3-60 are $\hat{\beta}_1$ and $\hat{\beta}_2$, being immaterial of which root is designated $\hat{\beta}_1$ or $\hat{\beta}_2$. The estimate $\hat{p}$ of the proportion p, obtained by substituting $\hat{\beta}_1$ and $\hat{\beta}_2$, respectively, in Eq. 3-61, refers to the component having β_1 as parameter, and 1-p refers to the other component.

3.4 Distribution of the Largest Exceedance in a Year

The important probability distribution function in flood analysis, obtained by the use of partial flood series, is the distribution of the largest exceedance in a year. It enables the computation of flood peak values for given return periods. Denote by χ the largest exceedance of ξ_ν in a year. The distribution χ is denoted by

$$F(x) = P(\chi \leq x) , \quad \text{for} \quad x > 0 \quad . \qquad (3\text{-}62)$$

The distribution of the largest exceedance can be derived by using the combination of distributions of the number and the magnitude of exceedances (Zelenhasic, 1970)

$$F(x) = P(E_0) + \sum_{k=1}^{\infty} P(\max_{1 \leq \nu \leq k} \xi_\nu \leq x \cap E_k) \ , \qquad (3\text{-}63)$$

with $\chi = \max_{1 \leq \nu \leq k} \xi_\nu = \max(\xi_1, \xi_2, \ldots, \xi_k)$, $E_k = (\eta = k)$ being the event that exactly k exceedances occur in a year.

Under the assumptions:

(i) $\xi_1, \xi_2 \ldots$ are independent of η, and

(ii) $\xi_1, \xi_2, \ldots$ are mutually independent random variables with the common distribution function $H(x)$, i.e.,

$$H(x) = P[\xi_\nu \leq x] \ , \qquad (3\text{-}64)$$

Eq. 3-63 is simplified to read

$$F(x) = P(E_0) + \sum_{k=1}^{\infty} [\{H(x)\}^k \cdot P(E_k)] \qquad (3\text{-}65)$$

The validity of the above two assumptions will be investigated by using the observed data in Chapter V.

3.5 Evaluation of the Return Period

The rarity of a flood peak may be conveyed in a number of ways, each expressing the probability of its exceedance or nonexceedance during a time interval, or alternatively each flood value may be considered as a function of its associated value of return period. The flood value for a given return period has played a major role in hydrologic and economic evaluations of water resources projects. It is important to derive the relationship between flood magnitude and its return period by using the probability distribution of the largest exceedance in a year.

The time elapsing between successive peak flows to exceed a specified value x is a random variable. Its mean value is defined as the return period T of X. Following Rousselle (1972), let $\chi_1, \chi_2, \chi_3, \cdots$ be a sequence of maximal annual values or of the largest exceedance in a year and let

$$N_x = \min(\nu; \chi_\nu > x) \ , \quad \text{for} \quad x > 0 \quad . \qquad (3\text{-}66)$$

Hence

$$P(N_x = n) = P(\chi_1 \leq x, \chi_2 \leq x, \ldots, \chi_{n-1} \leq x, \chi_n > x) \qquad (3\text{-}67)$$

for n = 1,2,3,... .

Because χ_ν is assumed to be a sequence of independent random variables with the distribution

$$F(x) = P(\chi_\nu \leq x) \quad , \qquad (3\text{-}68)$$

then

$$P(N_x = n) = [F(x)]^{n-1} \ [1 - F(x)] \quad , \qquad (3\text{-}69)$$

with $E(N_x)$ = the average number of years for the first exceedance of x to occur obtained by

$$E(N_x) = \sum_{n=1}^{\infty} n[F(x)]^{n-1} \ [1 - F(x)] = \frac{1}{1 - F(x)} \quad . \qquad (3\text{-}70)$$

The x values of Eq. 3-70 represent the largest annual flood peak exceedance flows for specified return periods, $E(N_x)$, and the distribution $F(x)$ is given by Eq. 3-65.

The value of $E(N_x)$ obtained by Eq. 3-70 is the return period for annual flood series. This return period is different from the return period for partial flood series which can be alternately defined as follows:

With each selected truncation level Q_b, the series of exceedances ξ_ν are well defined (see Fig. 3-1). For a given $Q_d > Q_b$, consider only those ξ_ν that exceed Q_d and denoted by ξ_i^*. Let T^* represent the inter-event time between two successive ξ_i^*'s. The expected value of T^*, $E(T^*)$, is the return period for partial flood series corresponding to the flood peak Q_d and truncation level Q_b.

The purpose of this study is to use the partial flood series to estimate the distribution of annual flood series. Hence, the return period $E(N_x)$ obtained by Eq. 3-70 is used for this study.

3.6 Comparison of Efficiency of Estimates of Flood Peaks of Given Return Periods by Using Annual and Partial Flood Series

To answer whether partial flood series, obtained from the mean daily flow hydrographs, is more efficient in estimating flood peaks $\hat{Q}(T)$ of given annual return periods than annual flood series, the approach used is that of comparing the sampling variances, var $\hat{Q}(T)$, of $\hat{Q}(T)$ obtained from both flood series. Let $\hat{Q}(T)_a$ (see Eq. 3-76) and $\hat{Q}(T)_p$ (see Eq. 3-86) be the estimates of annual flood peaks of given return periods obtained directly from annual flood series and indirectly from partial flood series, respectively. If var $\hat{Q}(T)_p$ of $\hat{Q}(T)_p$, estimated from the partial flood series, is smaller than var $\hat{Q}(T)_a$ of $\hat{Q}(T)_a$ estimated from the annual flood series, then the partial flood series is said to be more efficient or more useful in estimating annual flood peaks than the annual flood series. For convenience, the annual flood peaks will be called flood peaks $Q(T)$ in this study.

A statistical model is chosen which gives the population partial flood series from which the model of annual flood series can be derived. Each sample of partial flood series gives an estimate of Q(T). The corresponding estimate of Q(T) can be obtained directly from the sample of annual flood series. Hence, each sample series provides two estimates of Q(T), one from partial and the other from annual flood series. From many samples, the sampling variances of these estimators can be obtained, and compared.

Use of Generated Samples of Daily Flows. To study the sampling variances of estimated flood values for given return periods, long records of mean daily flows are needed. Such long records would be considered to represent the known population, from which various small samples are drawn. For each small sample of mean daily flow, the partial and annual flood series are derived. It follows that the flood $\hat{Q}(T)$ of a given return period T can be estimated from each sample of both series. Sampling variances of $\hat{Q}(T)$ for each sample size of the two series are then computed by

$$\text{var } \hat{Q}(T) = \frac{1}{m-1} \sum_{i=1}^{m} [\hat{Q}_i(T) - \overline{Q(T)}]^2 \qquad (3\text{-}71)$$

where m = the total number of samples for given sample size N, $\hat{Q}_i(T)$ = the flood value from the i-th sample, i = 1,2,...,m, $\overline{Q(T)}$ = the mean of all $\hat{Q}_i(T)$ values.

A method for generation of long records of daily flows is needed, since such long records of historical daily flows are not available. A model for generation of daily flows is developed and used in order to generate a long record (such as 1000 or 2000 years) of daily flows. Procedures used in developing the daily flow model are presented in Chapter IV, with the application to the Boise River and the Powell River given in Chapter VI.

Selection of Models for Annual and Partial Flood Series. It is necessary to emphasize that the main purpose is to compare var $\hat{Q}(T)$ of $\hat{Q}(T)$, which results from the estimation of Q(T) from the annual flood series, with the corresponding var $\hat{Q}(T)$ from the estimation of $\hat{Q}(T)$ from partial flood series. The statistical model should be chosen in such a way as to represent the population partial flood series, from which the model of annual flood series can be derived.

The empirical relationship between expectancies of partial flood series and probabilities of annual flood series, which was suggested by Langbein (1949), is first considered since it is not dependent on the assumed flood model. Two main objections can be raised in using this empirical relationship, that can make the comparison of sampling variances of $\hat{Q}(T)$ either inappropriate or unfeasible: (1) For a given sample of N years, it is not feasible to estimate Q(T) for T greater than N years, by using the plotting position, because the extrapolation is needed to estimate $\hat{Q}(T)$, subject to errors; and (2) Though the number of floods in a partial flood series is greater than for an annual flood series, this greater number tend to include the lower floods, or floods of the low return period, with flood at the high return periods being generally close to or identical with those of the annual flood series. Hence, for a given sample of size N, the flood values for the return periods close to N years are generally of the same magnitude for both series. It follows that in the range of return periods close to N years the ratio of sampling variance of $\hat{Q}(T)$ would be close to unity.

It was shown by Zelenhasic (1970) for the partial flood series, when a Poisson distribution for the number of exceedances in a year and an exponential distribution for the magnitude of exceedances are good fits, and under some commonly used assumptions, that combined they give a double exponential or Gumbel distribution function for the annual floods. This theoretical finding is used in the comparison of sampling variances of $\hat{Q}(T)$ obtained from the two flood series. Only two parameters must be estimated from the available sample. This is an advantage in using this finding since the sampling variance of $\hat{Q}(T)$ depends on the sampling variances of estimates of distribution parameters as well as on the number of parameters used.

Derivation of Flood Magnitudes and their Sampling Variances, for Given Return Periods, from Annual Flood Series by Using Gumbel Distribution.

(a) Gumbel Distribution and Estimation of its Parameters. The probability density function is

$$f(x;u,\alpha) = \frac{1}{\alpha} \exp[-(\frac{x-u}{\alpha}) - e^{-(x-u)/\alpha}] \qquad (3\text{-}72)$$

and the distribution function is

$$F(x) = \exp[-e^{-(x-u)/\alpha}] \quad , \qquad (3\text{-}73)$$

in which u = the location parameter, and α = the scale parameter.

The mean, variance and skewness are

$$\left.\begin{aligned} \mu &= u + 0.5772\alpha \\ \sigma^2 &= \frac{1}{6}\pi^2\alpha^2 \\ g &\simeq 1.14 \end{aligned}\right\} \qquad (3\text{-}74)$$

The maximum likelihood estimates of u and α are obtained by solving the equations

$$\left.\begin{aligned} &\frac{-n + \Sigma e^{-y_i}}{\hat{\alpha}} = 0 \\ \text{and} \quad &\frac{n - \Sigma y_i + \Sigma y_i e^{-y_i}}{\hat{\alpha}} = 0 \end{aligned}\right\} \qquad (3\text{-}75)$$

in which n = the number of observations, $y_i = (x_i-u)/\alpha$ = the Gumbel standardized or reduced variate. The solution of Eq. 3-75 for $\hat{\alpha}$ and $\hat{u}$ are obtained by the Fisher method which uses the information matrix as an iterative procedure. A demonstration of the method is given by Jenkinson (1969), and also Natural Environment Research Council (1975) of the United Kingdom.

(b) Estimate $\hat{Q}(T)_a$ and its Sampling Variance. Let $\hat{Q}(T)_a$ denote the flood magnitude for a given

return period T obtained from the annual flood series by using Gumbel distribution. Hence, the estimate of $Q(T)_a$ can be obtained from

$$\hat{Q}(T)_a = \hat{u} + \hat{\alpha} y(T) \tag{3-76}$$

in which $y(T) = -\ln[-\ln(1 - \frac{1}{T})]$, the Gumbel reduced variate, and T = the return period.

The sampling variance of $Q(T)_a$ in Eq. 3-76 is

$$\text{var}\,[\hat{Q}(T)_a] = \text{var}\,\hat{u} + 2\,\text{cov}\,[\hat{u},\hat{\alpha}y(T)] + \text{var}\,[\hat{\alpha}y(T)] \tag{3-77}$$

The variance-covariance matrix of the maximum likelihood estimates of $\hat{u}$ and $\hat{\alpha}$ (Kimball, 1946) is

$$\begin{bmatrix} \text{var}\,(\hat{u}) & \text{cov}\,(\hat{u},\hat{\alpha}) \\ \text{cov}\,(\hat{u},\hat{\alpha}) & \text{var}\,(\hat{\alpha}) \end{bmatrix} = \frac{\alpha^2}{n}\begin{bmatrix} 1 + \frac{6}{\pi^2}(1-\gamma)^2 & \frac{6}{\pi^2}(1-\gamma) \\ \frac{6}{\pi^2}(1-\gamma) & \frac{6}{\pi^2} \end{bmatrix} = \frac{\alpha^2}{n}\begin{bmatrix} 1.11 & 0.26 \\ 0.26 & 0.61 \end{bmatrix} \tag{3-78}$$

Therefore, var $\hat{u}$ = 1.11 α^2/n, var $\hat{\alpha}$ = 0.61 α^2/n; and cov $(\hat{u},\hat{\alpha})$ = 0.26 α^2/n. By substituting these values into Eq. 3-77, then

$$\text{var}\,[\hat{Q}(T)_a] = \frac{\alpha^2}{n}\,[1.11 + 0.52\,y(T) + 0.61\,y^2(T)] \tag{3-79}$$

The var $\hat{Q}(T)_a$ obtained from Eq. 3-79 is the theoretical sampling variance, based on the assumption that the distribution of annual floods is exactly the Gumbel distribution. It will be used for comparison with the sampling variance $\hat{Q}(T)_a$ obtained by the empirical method of Eq. 3-71.

Derivation of Flood Magnitudes and their Sampling Variances, for Given Return Periods, from Partial Floods Series by Using Combination of Poisson and Exponential Distributions.

(a) Estimate $\hat{Q}(T)_p$. The distribution of the number of exceedances in a year is assumed Poissonian, and the distribution of the magnitude of exceedances exponential. The distribution of the largest exceedance in a year is then given by Eq. 3-63. In addition, two more assumptions described in Section 3.4 are used in this approach.

Hence, Eq. 3-65 can be applied as

$$F(x) = \sum_{k=0}^{\infty} [\{H(x)\}^k \cdot P(E_k)] \quad . \tag{3-80}$$

For the case under discussion, the common distribution function of all exceedances ξ_ν is

$$H(x) = 1 - \exp\,(-\frac{x}{\beta}) \;, \quad x \geq 0 \tag{3-81}$$

and the distribution of the number of exceedances is

$$P(E_k) = \frac{e^{-\lambda}\lambda^k}{k!} \tag{3-82}$$

By substituting Eq. 3-81 and Eq. 3-82 into Eq. 3-80, the distribution of the largest exceedance in a year becomes

$$F(x) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} (1-e^{-x/\beta})^k$$

which in the limit becomes

$$F(x) = e^{-\lambda e^{-x/\beta}} \tag{3-83}$$

The relationship of Eq. 3-70 between the distribution function of the largest exceedance and the return period is

$$T = \frac{1}{1-F(x)} \tag{3-84}$$

By eliminating F(x) from Eq. 3-83 and Eq. 3-84, the flood exceedance for a given return period is expressed by

$$x = \hat{\beta}\ln\hat{\lambda} + \hat{\beta}\,y(T) \tag{3-85}$$

with $y(T) = -\ln[-\ln(1-1/T)]$. Because $x = Q-Q_b$, in which Q_b = the truncation level which defines the partial flood series and Q = the annual floods above Q_b, the annual flood magnitude for a given return period T, denoted by $\hat{Q}(T)_p$, obtained indirectly from partial flood series, becomes

$$\hat{Q}(T)_p = Q_b + \hat{\beta}\ln\hat{\lambda} + \hat{\beta}\,y(T) \quad , \tag{3-86}$$

with $\hat{\beta}$ and $\hat{\lambda}$ the parameters estimated from the partial flood series.

Suppose that m peaks in excess of Q_b, with m the random variable, have occurred in n years. Let ξ_ν, $\nu = 1,2,\ldots,m$ denote these exceedances above Q_b. The maximum likelihood estimates of λ and β are

$$\hat{\lambda} = \frac{m}{n}$$

and

$$\hat{\beta} = \overline{\xi}_\nu \tag{3-87}$$

with $\overline{\xi}_\nu = \frac{1}{m}\sum_{\nu=1}^{m} \xi_\nu$, the mean of all exceedances.

(b) Sampling Variance of $\hat{Q}(T)_p$. The derivation of sampling variance of $\hat{Q}(T)_p$ is mainly based on work by Cunnane (1972). For more details the reader is referred to this work.

Since $\hat{\lambda}$ and $\hat{\beta}$ are the maximum likelihood estimates of λ and β, their variances and covariance are

$$\text{var } \hat{\beta} = \frac{\beta^2}{m}$$

$$\text{var } \hat{\lambda} = \frac{\lambda}{n} \tag{3-88}$$

and

$$\text{cov } (\hat{\beta},\hat{\lambda}) = 0$$

The sampling variance of $\hat{Q}(T)_p$ of Eq. 3-86 is, for Q_b a constant

$$\begin{aligned}\text{var } [\hat{Q}(T)_p] &= \text{var } (\hat{\beta}\ln\hat{\lambda}) + 2 \text{ cov } [\hat{\beta}\ln\hat{\lambda},\hat{\beta}y(T)] \\ &\quad + \text{var } [\hat{\beta}y(T)] \\ &= \text{var } (\hat{\beta}\ln\hat{\lambda}) + 2y(T) \text{ cov } (\hat{\beta}\ln\hat{\lambda},\hat{\beta}) \\ &\quad + y^2(T) \text{ var } (\hat{\beta})\end{aligned} \tag{3-89}$$

By using

$$\text{var } [f(x,y)] \approx (\frac{\partial f}{\partial x})^2 \text{ var } (x) + 2 (\frac{\partial f}{\partial x})(\frac{\partial f}{\partial y}) \text{ cov } (x,y) + (\frac{\partial f}{\partial y})^2 \text{ var } (y),$$

with derivatives with respect to x and y evaluated at the expected values of x and y, respectively, then var $(\hat{\beta}\ln\hat{\lambda})$ becomes

$$\begin{aligned}\text{var } (\hat{\beta}\ln\hat{\lambda}) &= (\ln\lambda)^2 \text{ var } (\hat{\beta}) + 2\beta \frac{\ln\lambda}{\lambda} \text{cov } (\hat{\beta},\hat{\lambda}) \\ &\quad + \frac{\beta^2}{\lambda^2} \text{var } (\hat{\lambda}) \\ &= (\ln\lambda)^2 \text{ var } (\hat{\beta}) + \frac{\beta^2}{\lambda^2} \text{ var } (\hat{\lambda})\end{aligned} \tag{3-90}$$

since cov $(\hat{\beta},\hat{\lambda}) = 0$. By using

$$\begin{aligned}\text{cov}[f(x,y),g(x,y)] &\approx (\frac{\partial f}{\partial x})(\frac{\partial g}{\partial x})\cdot\text{var}(x) + [(\frac{\partial f}{\partial x})(\frac{\partial g}{\partial y}) \\ &\quad + (\frac{\partial f}{\partial y})(\frac{\partial g}{\partial x})] \cdot \text{cov}(x,y) \\ &\quad + (\frac{\partial f}{\partial y})(\frac{\partial g}{\partial y})\cdot\text{var}(y)\end{aligned}$$

it follows

$$\begin{aligned}\text{cov}(\hat{\beta}\ln\hat{\lambda},\hat{\beta}) &= \ln\lambda\cdot\text{var}(\hat{\beta}) + \{(\ln\lambda)(0) + (\frac{\beta}{\lambda})(1)\} \text{ cov}(\hat{\beta},\hat{\lambda}) \\ &\quad + (\frac{\beta}{\lambda})(0)\cdot\text{var}(\hat{\lambda}) \\ &= \ln\lambda\cdot\text{var}(\hat{\beta})\end{aligned}$$

Substituting var$(\hat{\beta}\ln\hat{\lambda})$ and cov$(\hat{\beta}\ln\hat{\lambda},\hat{\beta})$ into Eq. 3-89

$$\begin{aligned}\text{var}[\hat{Q}(T)_p] &= (\ln\lambda)^2\cdot\text{var}(\hat{\beta}) + \frac{\beta^2}{\lambda^2}\text{var}(\hat{\lambda}) \\ &\quad + 2y(T)\cdot\ln\lambda\cdot\text{var}(\hat{\beta}) + y^2(T)\cdot\text{var}(\hat{\beta}) \\ &= (\ln\lambda)^2 \frac{\beta^2}{m} + \frac{\beta^2}{\lambda^2}\frac{\lambda}{n} + 2y(T)\cdot\ln\lambda\cdot \frac{\beta^2}{m} \\ &\quad + y^2(T) \cdot \frac{\beta^2}{m} \\ &= \frac{\beta^2}{m} [(\ln\lambda)^2 + \frac{m}{n\lambda} + 2y(T)\cdot\ln\lambda + y^2(T)] \\ &= \frac{\beta^2}{\lambda n} \{1 + [\ln\lambda + y(T)]^2\} ,\end{aligned} \tag{3-92}$$

since $m = n\lambda$.

Comparison of Sampling Variances of Flood Value for a Given Return Period Obtained from Annual and Partial Flood Series. Under the approaches used, the the sampling variance of $\hat{Q}(T)$ for annual and partial flood peak series can be obtained both theoretically and empirically. The following are procedures used in comparison of sampling variances $\hat{Q}(T)_a$ and $\hat{Q}(T)_p$ for a given return period T.

(a) Exact Theoretical Approach. Let $R_{v,1}$ be the ratio of the sampling variances $\hat{Q}(T)_a$ and $\hat{Q}(T)_p$ obtained theoretically from annual and partial flood series, respectively. Hence, for a given T, $R_{v,1}$ is obtained from Eqs. 3-79 and 3-92 as

$$R_{v,1} = \frac{\lambda\alpha^2[1.11+0.52\ y(T)+0.61\ y^2(T)]}{\beta^2\ \{1 + [\ln\lambda + y(T)]^2\}} \tag{3-93}$$

The relationships between parameters u, α for annual flood series and parameters λ, β for partial flood series can be derived analytically by comparing Eq. 3-76 and Eq. 3-86, under the assumption $\hat{Q}(T)_a = \hat{Q}(T)_p$. Hence for

$$\left.\begin{aligned}&\alpha = \beta \\ \text{and} \\ &u = Q_b + \beta\ln\lambda\end{aligned}\right\} \tag{3-94}$$

it follows

$$R_{v,1} = \frac{\lambda[1.11 + 0.52\ y(T) + 0.61\ y^2(T)]}{\{1 + [\ln\lambda + y(T)]^2\}} \tag{3-95}$$

Equation 3-95 shows how the ratio of sampling variances, obtained by the exact theoretical approach, varies with the return period T. For a given value of λ, the relationship between the ratio $R_{v,1}$ and the return period T expressed as the Gumbel reduced variate, y(T), can be derived. The results of these relationships, for the range of λ from 0.8 to 5.0, are shown in Fig. 3-3. It can be concluded from

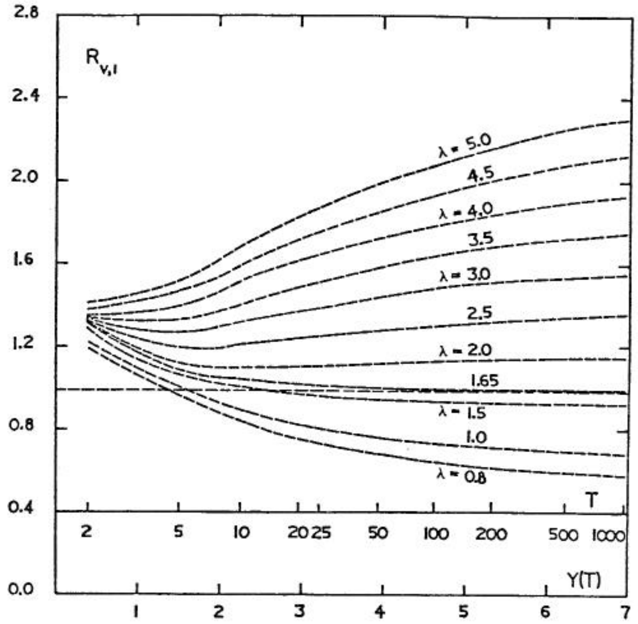

Fig. 3-3 Relationship between the Ratio $R_{v,1}$ of Sampling Variances $\hat{Q}(T)_a$ and $\hat{Q}(T)_p$, Based on Exact Theoretical Approach, and the Return Period T, for Given Values of λ.

Fig. 3-3 that, based on exact theoretical approach, the partial flood series estimate of Q(T) always has a smaller sampling variance than that of the annual flood series for the return period T less than 5 years. For the whole range of return periods, the partial flood series estimate of Q(T) has a smaller sampling variance than that of the annual flood series if the partial flood series value of λ is at least 1.65. The larger λ, the smaller is the sampling variance of the estimate of Q(T) by means of partial flood series. This result is later used for the comparison with the results obtained by using the approximate theoretical and the empirical approach, respectively.

(b) *Approximate Theoretical Approach.* In this particular approach, instead of using the relationships of Eq. 3-94, parameters u and α are estimated from the annual flood series of generated long record of daily flow series, and parameters λ and β from the corresponding partial flood series. Let $R_{v,2}$ denote the ratio of sampling variances of $\hat{Q}(T)_a$ and $\hat{Q}(T)_p$ by this approach, then

$$R_{v,2} = \frac{\lambda\alpha^2[1.11 + 0.52\ y(T) + 0.61\ y^2(T)]}{\beta^2\ \{1 + [\ln\lambda + y(T)]^2\}} \quad . \qquad (3\text{-}96)$$

The difference between $R_{v,1}$ and $R_{v,2}$ is that the difference between α and β is taken into consideration in computing $R_{v,2}$.

(c) *Empirical Approach.* In this approach, the sampling variance of $\hat{Q}(T)$ for each flood series is obtained empirically by using Eq. 3-71. Let $R_{v,3}$ denote the ratio of sampling variances of $\hat{Q}(T)_a$ and $\hat{Q}(T)_p$, then

$$R_{v,3} = \frac{\sum_{i=1}^{n} [\hat{Q}_i(T)_a - \overline{Q(T)_a}]^2}{\sum_{i=1}^{n} [\hat{Q}_i(T)_p - \overline{Q(T)_p}]^2} \quad , \qquad (3\text{-}97)$$

with subscripts a and p indicating the estimates obtained from annual flood series and partial flood series, respectively.

Chapter IV
MATHEMATICAL MODEL OF DAILY FLOWS

4.1 Purpose of Generation of Daily Flows

The main purpose of generating the long record of daily flows is to compare the efficiency of estimates of flood peaks of given return periods by using annual and partial flood series. This carried out by comparing the sampling variances of flood values for given return periods obtained for each of the two series, which in turn are derived from the generated daily flows. In addition, properties of partial flood series and approach are studied on this long record of generated daily flows.

4.2 Selection of Length of Generation of Daily Flows

It takes considerable computer time to generate daily flows for a long period of years. In this study, the main purpose of generating a long record of daily flows is to investigate properties of extreme large values. When most major floods occur in the wet seasons, it is feasible and sufficient to generate daily flows only within that wet season. Important reasons justify the generation of daily flows only within the wet seasons, namely this approach requires less computer time, less computer core storage, while a sufficiently accurate approach in the estimation of model parameters. The number of harmonics in description of periodic parameters in the model then is smaller than if daily flows are generated for the whole year, for the same accuracy, for the simple reason that each periodic parameter has more variation in the whole year than that in the selected wet season. The main disadvantage of this approach is the problem of some distortion of partial flood series, because some small flood peaks, greater than the truncation level Q_b, may occur in dry seasons. This distortion can be minimized by expanding somewhat the period of generation from wet season into the dry seasons. The lowest required truncation level defines partial flood series. The season of generation is then selected in such a way as to have most of flood peaks greater than this truncation level. The lowest truncation level is selected in this study so that the average number of flood exceedances per year is about 4 or 5.

4.3 Mathematical Model for Daily Flows

General Concepts. The mathematical model of daily flows, studied by Tao, Yevjevich, and Kottegoda (1976), was carefully reviewed. It is learned that it is relatively difficult to fit a probability distribution function to independent stochastic component of daily flows, because of its high skewness and kurtosis coefficients. Furthermore, it is difficult to remove completely the dependence from the dependent stochastic component after periodicities in the mean and standard deviation are removed. Vargas (1977) used the generation method to systematically check each stage of estimation procedure, to assess whether failures originate from estimation procedures or from inappropriate models. The number of selected harmonics for periodic parameters affect all the subsequent stages of estimation. Estimation procedures are sensitive to distribution of independent stochastic component.

By removing periodicities in the mean and standard deviation only, the remaining series is usually considered as stationary. In generating daily flows for the study of extreme large values, it is important to consider not only the eventual periodicities in autocorrelation coefficients, but also to preserve the skewness coefficient properties in generated data. The use of transformations may not only remove periodicities in the mean and standard deviation, but also periodicities in autocorrelation and skewness coefficients.

The three-parameter lognormal probability distribution has important advantages not to be overlooked. It provides a relatively simple method for preserving the first three moments of observed data, with its logarithmic transforms normally distributed by definition. It is attractive to transform the original data into normally distributed values as the first step of analysis, in order to use the two important properties of normal variables, namely that dependence structure does not affect the distribution, because the distribution of the sum of normal variables is normal, and that the second-order stationarity of normal variables implies the stationarity of high order also.

The flows of each individual day of the year are assumed to follow the lognormal distribution with three parameters: lower bound of original data and the mean and standard deviation of transformed data. They are estimated from historic data of each individual day. Logarithmic transformations are applied to historic data by using the lower bounds in order to transform the original values to normal variables. Periodicities in the mean and standard deviation of transformed values are then removed. By using the postulated dependence model, independent standard normal variable is then obtained.

To minimize the effect of the selected number of harmonics in periodic parameters on all subsequent stages of estimation in this model, the numbers of harmonics of all fitted periodic parameters are estimated from the original data and not from transformed data.

Relationship between Moments of Normal and Lognormal Variables. If historic data follow a three-parameter lognormal distribution, the generated data should resemble historic data in terms of mean, standard deviation and skewness coefficient, by using the relations of moments of the two processes (Matalas, 1967).

Let "a" be the lower bound of variable X, with (X-a) lognormally distributed; then Y = ln(X-a) is normally distributed. The mean μ_x, variance σ_x^2, and skewness γ_x are related to the lower bound a, mean μ_y and variance σ_y^2 of Y by

$$\mu_x = a + \exp(\frac{\sigma_y^2}{2} + \mu_y) \tag{4-1}$$

$$\sigma_x^2 = \exp[2(\sigma_y^2 + \mu_y)] - \exp(\sigma_y^2 + 2\mu_y) \tag{4-2}$$

and

$$\gamma_x = \frac{\exp(3\sigma_y^2) - 3\exp(\sigma_y^2) + 2}{[\exp(\sigma_y^2) - 1]^{1.5}} \quad . \tag{4-3}$$

Autocorrelation between Normal and Lognormal Processes. For X lognormal, with Y = ln(X-a) normal, the first-order autocorrelation $\rho_x(1)$ of X is expressed in terms of the first-order autocorrelation $\rho_y(1)$ of Y by

$$\rho_x(1) = \frac{\exp[\sigma_y^2 \rho_y(1)] - 1}{\exp(\sigma_y^2) - 1} \qquad (4\text{-}4)$$

So that

$$\rho_y(1) = \frac{1}{\sigma_y^2} \ln \{1 + \rho_x(1) [\exp(\sigma_y^2) - 1]\} . \qquad (4\text{-}5)$$

It can be proved that Eqs. 4-4 and 4-5 are valid for any time lag k (Mejia and Rodriguez-Iturbe, 1974). Hence,

$$\rho_x(k) = \frac{\exp[\sigma_y^2 \rho_y(k)] - 1}{\exp(\sigma_y^2) - 1} \qquad (4\text{-}6)$$

and

$$\rho_y(k) = \frac{1}{\sigma_y^2} \ln \{1 + \rho_x(k) [\exp(\sigma_y^2) - 1]\} . \qquad (4\text{-}7)$$

Cross Correlation between Normal and Lognormal Processes. Let X_1 and X_2 be the original variables (two different populations), with means μ_{x_1} and μ_{x_2} and standard deviations σ_{x_1} and σ_{x_2}, and the cross correlation coefficient ρ_x. Variables X_1 and X_2 are three-parameter lognormal with

$$\left.\begin{aligned} Y_1 &= \ln(X_1 - a_1) \\ &\text{and} \\ Y_2 &= \ln(X_2 - a_2) \end{aligned}\right\} \qquad (4\text{-}8)$$

normal. Let μ_{y_1}, μ_{y_2}, σ_{y_1}, σ_{y_2}, and ρ_y represent means, standard deviations and cross correlation coefficient for Y_1 and Y_2. The relation of ρ_x and ρ_y is (Mejia, Rodriguez-Iturbe, and Cordova, 1974)

$$\rho_y = \frac{1}{\sigma_{y_1}\sigma_{y_2}} \ln \{1 + \rho_x [[\exp(\sigma_{y_1}^2) - 1][\exp(\sigma_{y_2}^2) - 1]]^{1/2}\} \qquad (4\text{-}9)$$

If $\sigma_{y_1} = \sigma_{y_2}$, Eq. 4-9 is reduced to Eq. 4-7.

Application of Modeling Concepts to Daily Flows. Consider the matrix of daily flows:

$$X_{p,\tau} = \begin{bmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,\tau} & \cdots & X_{1,\omega} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,\tau} & \cdots & X_{2,\omega} \\ \vdots & \vdots & & \vdots & & \vdots \\ X_{p,1} & X_{p,2} & \cdots & X_{p,\tau} & \cdots & X_{p,\omega} \\ \vdots & \vdots & & \vdots & & \vdots \\ X_{n,1} & X_{n,2} & \cdots & X_{n,\tau} & \cdots & X_{n,\omega} \end{bmatrix} \qquad (4\text{-}10)$$

with p = 1,2,...,n, representing the year number in the record, τ, $\tau = 1,2,\ldots,\omega$, the day number within the year, running cyclically from 1 to ω, n = the total number of years, and ω = the total number of days in the wet season.

By considering daily flows to be from different populations for different days, the modeling concepts outlined above can be applied, provided that the marginal distribution of daily flows for each individual day, or each column of matrix of Eq. 4-10, is lognormal. For example, if $X_{1,\tau}, X_{2,\tau}, \ldots, X_{n,\tau}$, for the day τ, is three-parameter lognormal, with mean $\mu_{x,\tau}$, standard deviation $\sigma_{x,\tau}$, skewness coefficient $\gamma_{x,\tau}$, and lag-one serial correlation coefficient $\rho_{x,\tau}(1)$, the procedures explained above can be used to generate daily flows for the day τ for as many years as required. Equation 4-9 is used to preserve the serial correlation between the successive days, valid for any lag k (k days apart). It is then expressed by

$$\rho_y(k,\tau) = \frac{1}{\sigma_{y,\tau-k}\sigma_{y,\tau}} \ln \{1 + \rho_x(k,\tau) \cdot [[\exp(\sigma_{y,\tau-k}^2) - 1][\exp(\sigma_{y,\tau}^2) - 1]]^{1/2}\} \qquad (4\text{-}11)$$

with $\rho_x(k,\tau)$ and $\rho_y(k,\tau)$ = the lag-zero cross correlation coefficients between the day τ-k and the day τ of $X_{p,\tau}$ and $Y_{p,\tau}$, respectively. For convenience and understanding, they are called the k-th order serial correlation coefficients of daily flows.

Removal of Periodic Parameters. The nonparametric methods may be used to remove periodic parameters from a time series (Tao, Yevjevich and Kottegoda, 1976). In case of daily flows, the total number of statistics in the nonparametric method is very large in comparison with the total number of statistics in the parametric method. Since it is impossible to estimate so many parameters accurately from a limited size of sample series, these estimates must be subject to large sampling errors in the nonparametric method. The general objective of mathematical modeling of deterministic-stochastic processes is to condense information by developing models which use the number of parameters parsimoniously. Since the nonparametric method does not satisfy this objective, it is not used in this study.

Let the periodic parameters be symbolized by ν_τ. The mathematical description of periodic variation of ν_τ is represented by the Fourier series analysis as

$$\nu_\tau = \bar{\nu} + \sum_{j=1}^{m} C_j \cos\left(\frac{2\pi j \tau}{\omega} + \theta_j\right) , \qquad (4\text{-}12)$$

in which $\bar{\nu}$ = the average value of ν_τ, C_j = the amplitude, θ_j = the angular phase, j = the index sequence of harmonics, m = the total number of significant harmonics, and ω = the period in days.

The alternative form to Eq. 4-12 is

$$\nu_\tau = \bar{\nu} + \sum_{j=1}^{m} (A_j \cos \frac{2\pi j\tau}{\omega} + B_j \sin \frac{2\pi j\tau}{\omega}) \quad , \qquad (4\text{-}13)$$

with A_j and B_j the Fourier coefficients, estimated from the ω values of $\hat{\nu}_\tau$ (where $\hat{\nu}_\tau$ are sample values), by

$$A_j = \frac{2}{\omega} \sum_{\tau=1}^{\omega} \hat{\nu}_\tau \cos \frac{2\pi j\tau}{\omega}$$

and

$$B_j = \frac{2}{\omega} \sum_{\tau=1}^{\omega} \hat{\nu}_\tau \sin \frac{2\pi j\tau}{\omega} \qquad (4\text{-}14)$$

with the amplitude and angular phase expressed as

$$C_j = \sqrt{A_j^2 + B_j^2}$$

and

$$\theta_j = \tan^{-1}(-\frac{B_j}{A_j}) \qquad (4\text{-}15)$$

Let $s^2(\nu_\tau)$ be the variance of computed ν_τ. For a harmonic j, var $h_j = (A_j^2 + B_j^2)/2$. The ratio

$$\Delta P_j = \frac{\text{var } h_j}{s^2(\nu_\tau)} \qquad (4\text{-}16)$$

represents the part of the variance of ν_τ explained by the j- th harmonic. Hence, the explained variance of k harmonics is

$$P_k = \sum_{i=1}^{k} \Delta P_i \quad . \qquad (4\text{-}17)$$

This explained variance is used as the criterion for selecting the number of significant harmonics.

The first periodicity to be removed is in the skewness coefficient. It is accomplished by using the logarithmic transformation, with the lower bound a_τ periodic. The symbol $X_{p,\tau}$ stands for values of an observed daily flow series with p and τ previously defined. Let $Y_{p,\tau}$ denote the transformed variables, then

$$Y_{p,\tau} = \ln(X_{p,\tau} - a_\tau) \quad . \qquad (4\text{-}18)$$

Since $X_{p,\tau}$ is assumed as lognormally distributed with the lower bound a_τ, $Y_{p,\tau}$ must be normally distributed with mean $\mu_{y,\tau}$ and standard deviation $\sigma_{y,\tau}$. Therefore, the periodicity in the skewness coefficient has been removed.

The removing of periodicities in the mean and standard deviation of the transformed variable, $Y_{p,\tau}$, is made by

$$\varepsilon_{p,\tau} = \frac{Y_{p,\tau} - \mu_{y,\tau}}{\sigma_{y,\tau}} \quad , \qquad (4\text{-}19)$$

in which $\varepsilon_{p,\tau}$ = the standardized stochastic component of $Y_{p,\tau}$, a dependent, normally distributed variable with mean zero and variance unity.

Dependence Models for Stationary Stochastic Components. The $\varepsilon_{p,\tau}$ variable, obtained by removing the periodicities in the mean, standard deviation, and skewness coefficient of $X_{p,\tau}$, is stationary time series provided the autocorrelation coefficients are not periodic. The models for dependence of $\varepsilon_{p,\tau}$ may be: moving average, linear autoregressive, a combination of the two, and other schemes. Since the autoregressive linear models have been found in practice to be very useful in hydrology, they are applied in this study.

The dependence of a stochastic hydrologic series can be approximated by various orders of linear autoregressive models. The first-, second-, and third-order autoregressive linear models are most commonly used rather than the higher-order models.

The general m-th order autoregressive linear model is

$$\varepsilon_{p,\tau} = \sum_{k=1}^{m} \alpha_{k,\tau}\, \varepsilon_{p,\tau-k} + \sigma_{\xi,\tau}\, \xi_{p,\tau} \qquad (4\text{-}20)$$

with $\alpha_{k,\tau}$ = the autoregressive coefficients, which are functions of serial correlation coefficients $\rho_{k,\tau}$, which are either periodic or nonperiodic, $\sigma_{\xi,\tau}$ = the standard deviation of $\xi_{p,\tau}$ which is periodic if $\alpha_{k,\tau}$ are periodic, and $\xi_{p,\tau}$ = a standardized variable independent of $\varepsilon_{p,\tau-k}$. Since $\varepsilon_{p,\tau}$ is normally distributed, $\xi_{p,\tau}$ should be independent, normally distributed variable, with mean zero and variance unity. The serial correlation coefficient $\rho_{k,\tau}$ of the lag k is

$$\rho_{k,\tau} = \frac{\text{cov}(\varepsilon_{p,\tau}, \varepsilon_{p,\tau-k})}{\sigma_{\varepsilon_\tau}\, \sigma_{\varepsilon_{\tau-k}}} \qquad (4\text{-}21)$$

Equations used for determining the coefficients $\alpha_{k,\tau}$, and $\sigma_{\xi,\tau}$, with k = 1,2, and 3, are given below.

(a) *First-Order Model*. The first-order autoregressive linear model is

$$\varepsilon_{p,\tau} = \alpha_{1,\tau}\, \varepsilon_{p,\tau-1} + \sigma_{\xi,\tau}\, \xi_{p,\tau} \quad . \qquad (4\text{-}22)$$

The parameters $\alpha_{1,\tau}$ and $\sigma_{\xi,\tau}$ are expressed as

$$\alpha_{1,\tau} = \rho_{1,\tau} \qquad (4\text{-}23)$$

and

$$\sigma^2_{\xi,\tau} = 1 - \alpha^2_{1,\tau} = 1 - \rho^2_{1,\tau} \qquad (4\text{-}24)$$

(b) *Second-Order Model*. The second-order autoregressive linear model is

$$\varepsilon_{p,\tau} = \alpha_{1,\tau}\, \varepsilon_{p,\tau-1} + \alpha_{2,\tau}\, \varepsilon_{p,\tau-2} + \sigma_{\xi,\tau}\, \xi_{p,\tau} \quad . \qquad (4\text{-}25)$$

The parameters $\alpha_{k,\tau}$, k = 1,2, can be obtained from the following linear equations.

$$\begin{bmatrix} 1 & \rho_{1,\tau-1} \\ \rho_{1,\tau-1} & 1 \end{bmatrix} \begin{bmatrix} \alpha_{1,\tau} \\ \alpha_{2,\tau} \end{bmatrix} = \begin{bmatrix} \rho_{1,\tau} \\ \rho_{2,\tau} \end{bmatrix} \quad (4\text{-}26)$$

with the solution

$$\left.\begin{aligned} \alpha_{1,\tau} &= \frac{\rho_{1,\tau} - \rho_{2,\tau}\,\rho_{1,\tau-1}}{1 - \rho^2_{1,\tau-1}} \\ \alpha_{2,\tau} &= \frac{\rho_{2,\tau} - \rho_{1,\tau}\,\rho_{1,\tau-1}}{1 - \rho^2_{1,\tau-1}} \end{aligned}\right\} \quad (4\text{-}27)$$

The variance $\sigma^2_{\xi,\tau}$ is

$$\sigma^2_{\xi,\tau} = 1 - \alpha^2_{1,\tau} - \alpha^2_{2,\tau} - 2\alpha_{1,\tau}\,\alpha_{2,\tau}\,\rho_{1,\tau-1} \quad (4\text{-}28)$$

(c) *Third-Order Model.* The third-order autoregressive linear model is

$$\varepsilon_{p,\tau} = \sum_{k=1}^{3} \alpha_{k,\tau}\,\varepsilon_{p,\tau-k} + \sigma_{\xi,\tau}\,\xi_{p,\tau} \quad (4\text{-}29)$$

The parameters $\alpha_{k,\tau}$, for k = 1,2, and 3, can be obtained from the following linear equations

$$\begin{bmatrix} 1 & \rho_{1,\tau-1} & \rho_{2,\tau-1} \\ \rho_{1,\tau-1} & 1 & \rho_{1,\tau-2} \\ \rho_{2,\tau-1} & \rho_{1,\tau-2} & 1 \end{bmatrix} \begin{bmatrix} \alpha_{1,\tau} \\ \alpha_{2,\tau} \\ \alpha_{3,\tau} \end{bmatrix} = \begin{bmatrix} \rho_{1,\tau} \\ \rho_{2,\tau} \\ \rho_{3,\tau} \end{bmatrix} \quad (4\text{-}30)$$

The variance $\sigma^2_{\xi,\tau}$ is

$$\sigma^2_{\xi,\tau} = 1 - \alpha^2_{1,\tau} - \alpha^2_{2,\tau} - \alpha^2_{3,\tau} - 2\alpha_{1,\tau}\,\alpha_{2,\tau}\,\rho_{1,\tau-1} - 2\alpha_{1,\tau}\,\alpha_{3,\tau}\,\rho_{2,\tau-1} - 2\alpha_{2,\tau}\,\alpha_{3,\tau}\,\rho_{1,\tau-2} \quad (4\text{-}31)$$

If order of the linear autoregressive model is selected, parameters $\alpha_{k,\tau}$ and $\sigma_{\xi,\tau}$ can be estimated from the sample autocorrelation coefficients. Finally, the independent standardized normal random variable $\xi_{p,\tau}$ is computed from the $\varepsilon_{p,\tau}$ series.

4.4 Estimation of Parameters of Daily Flow Series

For parameters defined in terms of moments and time lags, the standard errors of their estimates increase with an increase of the moment order and time lag. The larger the standard error, the greater the bias is likely to be. As suggested by Matalas (1967), bias may be minimized but not completely eliminated. One technique for minimizing bias is regionalization, which takes the form of relating the parameters, estimated from the historic sequences at a number of sites in a basin, to certain meteorologic and physiographic characteristics of the basin. Another technique for minimizing bias is the use of maximum likelihood estimation of parameters, since the standard errors of these estimates are smaller than those for estimates based on moments. However, maximum likelihood estimators are not always statistically unbiased, and they cannot be determined without making an assumption about the underlying probability distributions.

The stages of parameter estimation are important in the study of daily flow series. The numbers of significant harmonics of periodic parameters are decisive since the parametric method is used. The approach in this study is to estimate parameters directly from historic data, not from sequences from which periodicities in other parameters have been removed, to avoid the effect of the selected number of significant harmonics for those parameters in the further stages of estimation.

The parameters used in generation are those of the three-parameter lognormal distribution: a_τ, $\mu_{y,\tau}$ and $\sigma_{y,\tau}$, the autoregressive coefficients, $\alpha_{k,\tau}$, with k = 1,2,...,m, and the standard deviation of residuals, $\sigma_{\xi,\tau}$. Fourier series were not applied directly in fitting periodic functions to estimates of these parameters. Fourier series are used for fitting parameters which are estimated directly from observed data because parameters a_τ, $\mu_{y,\tau}$, $\sigma_{y,\tau}$, $\alpha_{k,\tau}$ and $\sigma_{\xi,\tau}$ can be derived from them.

(a) *Estimation of Lower Bound, a_τ.* For each particular day τ, the mean $\mu_{x,\tau}$, standard deviation $\sigma_{x,\tau}$, and skewness coefficient $\gamma_{x,\tau}$ of $X_{p,\tau}$ are estimated by the method of moments. Parameters a_τ, $\mu_{y,\tau}$ and $\sigma_{y,\tau}$ of the lognormal variable $Y_{p,\tau}$ are then computed simultaneously by Eqs. 4-1 through 4-3. The main problem in this estimation is to obtain a reliable estimate of the skewness coefficient from the historic record, especially in case of small sample sizes. Since the skewness coefficient is a function of the first three moments, the standard error of its estimate is high and it is also biased.

Because the distribution of daily flow series for each day is assumed to be a three-parameter lognormal distribution, the maximum likelihood method can be applied in order to minimize biases in parameter estimation. The lower bound a_τ for each day τ is obtained by maximum likelihood method in solving Eq. 3-54 by an iterative procedure. Since it is a nonlinear equation, it has more than one solution. In applying iteration, the initial or starting value is important, to guarantee solution convergence. The starting value of a_τ is first assumed to be close to the observed $X_{min}(\tau)$, such as 0.975 $X_{min}(\tau)$, where $X_{min}(\tau) = \min[X_{p,\tau}$, p = 1,2,..., n, for fixed τ]. If the iteration diverges to values greater than $X_{min}(\tau)$, the new starting value less than the first one is assumed and so on. The purpose is to obtain $\hat{a}_\tau$ that has the value nearest to, but less than $X_{min}(\tau)$.

The fourier series is applied for fitting the periodicity in a_τ. The explained variance of Eq. 4-17 is used as the criterion for determining the number of significant harmonics. For days of the year for which the skewness coefficients are high, the maximum likelihood estimates $\hat{a}_\tau$ tend to be positive and close to observed $X_{min}(\tau)$. By using the Fourier series, some days have the fitted values of a_τ greater than the observed $X_{min}(\tau)$. Hence, some other consideration in selecting the number of significant harmonics for a_τ is that the number of days that have the fitted $\hat{a}_\tau$ greater than $X_{min}(\tau)$ should be very small.

Two alternative methods may be used to estimate parameters $\mu_{y,\tau}$ and $\sigma_{y,\tau}$ after the periodic a_τ has been computed: (1) By using the estimated values of a_τ (not the fitted periodic function values) in Eqs. 3-55 and 3-56 for estimating $\mu_{y,\tau}$ and $\sigma_{y,\tau}$, respectively, with the Fourier series applied in fitting the periodic $\hat{\mu}_{y,\tau}$ and $\hat{\sigma}_{y,\tau}$; and (2) By using the fitted periodic function values of a_τ in Eqs. 3-55 and 3-56 for estimating $\mu_{y,\tau}$ and $\sigma_{y,\tau}$, respectively, with the Fourier series then applied for fitting the periodic $\hat{\mu}_{y,\tau}$ and $\hat{\sigma}_{y,\tau}$. The experience of this study is that neither of these two approaches should be used. In the first approach, the Fourier series analysis is applied for fitting estimates of each periodic parameter, a_τ, $\mu_{y,\tau}$, and $\sigma_{y,\tau}$ independently. The problem arises because the fitted a_τ, $\mu_{y,\tau}$ and $\sigma_{y,\tau}$ for any day τ are not matched among themselves, giving rise to distortions in patterns of periodic functions $\mu_{x,\tau}$ and $\sigma_{x,\tau}$, with a large number of negative daily flows produced in generation procedure. By using the second approach the distortions are decreased. By fitting a periodic function to $\hat{a}_\tau$ by using a certain number of harmonics, some days have the fitted a_τ greater than $X_{min}(\tau)$, affecting then the estimations of $\mu_{y,\tau}$ and $\sigma_{y,\tau}$. The consequence is in distortions in patterns of $\mu_{x,\tau}$ versus τ, and $\sigma_{x,\tau}$ versus τ, in generated daily flow series in comparison with those of historic data.

To overcome this difficulty, Fourier series are used to fit the estimates of periodic parameters $\mu_{x,\tau}$ and $\sigma_{x,\tau}$, which are estimated directly from observed data. The periodic $\mu_{y,\tau}$ and $\sigma_{y,\tau}$, are then derived from fitted periodic functions of $\hat{a}_\tau$, $\hat{\mu}_{x,\tau}$ and $\hat{\sigma}_{x,\tau}$ by using their relationships.

(b) *Estimation of Mean* $\mu_{x,\tau}$ *and Standard Deviation* $\sigma_{x,\tau}$. The mean $\mu_{x,\tau}$ and the standard deviation $\sigma_{x,\tau}$ for each day τ are estimated from the observed data by

$$\hat{\mu}_{x,\tau} = \frac{1}{n}\sum_{p=1}^{n} X_{p,\tau} \qquad (4\text{-}32)$$

and

$$\hat{\sigma}_{x,\tau} = \left[\frac{1}{n-1}\sum_{p=1}^{n} (X_{p,\tau} - \mu_{x,\tau})^2\right]^{1/2} \quad , \qquad (4\text{-}33)$$

in which n = the number of years of observations.

The periodic $\hat{\mu}_{x,\tau}$ and $\hat{\sigma}_{x,\tau}$ are then fitted by Fourier series. First, the explained variance is used as the criterion for selecting the number of significant harmonics. Then, the final consideration in selecting the number of significant harmonics for $\hat{a}_\tau$, $\hat{\mu}_{x,\tau}$ and $\hat{\sigma}_{x,\tau}$ was for each day τ to have distributions of $\hat{\varepsilon}_{p,\tau}$, by Eq. 4-19, close to a normal distribution, on the average, and as much so as the chi-square test permitted.

The mean $\hat{\mu}_{y,\tau}$ and the standard deviation $\hat{\sigma}_{y,\tau}$ of $Y_{p,\tau}$ are then obtained from the fitted periodic functions to $\hat{a}_\tau$, $\hat{\mu}_{x,\tau}$, and $\hat{\sigma}_{x,\tau}$, by solving Eqs. 4-1 and 4-2, namely

$$\mu_{y,\tau} = \ln(\hat{\mu}_{x,\tau} - \hat{a}_\tau) - \frac{1}{2}\sigma^2_{y,\tau} \quad , \qquad (4\text{-}34)$$

and

$$\sigma^2_{y,\tau} = \ln\left[1 + \left(\frac{\hat{\sigma}_{x,\tau}}{\hat{\mu}_{x,\tau} - \hat{a}_\tau}\right)^2\right] \quad . \qquad (4\text{-}35)$$

(c) *Estimation of Serial Correlation Coefficients*, $\rho_x(k,\tau)$. The serial correlation coefficients $\rho_x(k,\tau)$ of $X_{p,\tau}$ are estimated from sample series by (see Tao, Yevjevich and Kottegoda, 1976)

$$r_x(k,\tau) = \frac{\sum_{p=1}^{n}\left[X_{p,\tau} - \frac{1}{n}\sum_{p=1}^{n}X_{p,\tau}\right]\left[X_{p,\tau-k} - \frac{1}{n}\sum_{p=1}^{n}X_{p,\tau-k}\right]}{\left[\sum_{p=1}^{n}(X_{p,\tau} - \frac{1}{n}\sum_{p=1}^{n}X_{p,\tau})^2\right]^{1/2}\left[\sum_{p=1}^{n}(X_{p,\tau-k} - \frac{1}{n}\sum_{p=1}^{n}X_{p,\tau-k})^2\right]^{1/2}} \qquad (4\text{-}36)$$

The Fourier series are used to fit the periodic values $r_x(k,\tau)$, k = 1,2,..., m, with m = the order of the autoregressive linear model.

Estimates of serial correlation coefficients, $r_y(k,\tau)$, of $Y_{p,\tau}$ are obtained from the fitted periodic functions to $r_x(k,\tau)$ by using Eq. 4-11 in the form

$$r_y(k,\tau) = \frac{1}{\sigma_{y,\tau-k}\,\sigma_{y,\tau}} \ln\left[1 + \rho_x(k,\tau)\{[\exp(\sigma^2_{y,\tau-k})-1][\exp(\sigma^2_{y,\tau})-1]\}^{1/2}\right] \qquad (4\text{-}37)$$

Finally, the autoregressive coefficients $\alpha_{k,\tau}$ and the standard deviation $\sigma_{\xi,\tau}$ are derived from periodic functions of $r_y(k,\tau)$ by replacing $\rho_{k,\tau}$ by $r_y(k,\tau)$ in Eqs. 4-23 and 4-24 for the first-order model, in Eqs. 4-27 and 4-28 for the second-order model, and Eqs. 4-30 and 4-31 for the third-order model, respectively.

4.5 Problem of Generated Negative Flows

By definition, the parameter a_τ is a lower bound for observed values of $X_{p,\tau}$. By nature, $X_{p,\tau}$ should be positive or zero. This implies that a_τ should always be positive in hydrologic applications. However, this interpretation may not be necessary, because a_τ can be positive or negative, and in fact is usually negative (Burges, Lettenmaier and Bates, 1975). According to experience of this study, a_τ depends on the skewness coefficient of $X_{p,\tau}$. The smaller the skewness coefficient, the more opportunity for a_τ to be negative. Because

$$X_{p,\tau} = a_\tau + \exp(Y_{p,\tau}) \tag{4-38}$$

the second term at the right side of Eq. 4-38 is always positive. If a_τ is positive, no problem arises with negative values of $X_{p,\tau}$. If a_τ is negative, on occasion $X_{p,\tau}$ may be negative since the normal distribution assigns non-zero probability to negative values.

In this study, the possible minimum value of $X_{p,\tau}$ is assumed zero, not allowing for negative values. The following procedure is used to minimize the effect of generated negative values. By considering a_τ negative, for $X_{p,\tau} = 0$ then

$$Y^*_{p,\tau} = \ln(-a_\tau) \quad , \tag{4-39}$$

following with

$$\varepsilon^*_\tau = \frac{\ln(-a_\tau)-\mu_{y,\tau}}{\sigma_{y,\tau}} \quad . \tag{4-40}$$

The value ε^*_τ is used as the lower limit of $\varepsilon_{p,\tau}$. If a generated value of $\varepsilon_{p,\tau}$ is smaller than ε^*_τ, it is set equal to ε^*_τ, with the process of generation continuing. The negative aspect may be in decreasing slightly the variance of generated series.

4.6 Generation Procedure

The procedure, parameters, and equations used to generate new daily flow samples are summarized as:

STEP 1 Obtain the Fourier parameters such as the number of significant harmonics, mean and Fourier coefficients, A and B of periodic parameters a_τ, $\mu_{x,\tau}$, $\sigma_{x,\tau}$ and $r_x(k,\tau)$, k = 1,2,..., m, with m = the order of the autoregressive linear model used.

STEP 2 Compute periodic parameters a_τ, $\mu_{x,\tau}$, $\sigma_{x,\tau}$, and $r_x(k,\tau)$ by using Eq. 4-13.

STEP 3 Derive periodic parameters $\mu_{y,\tau}$ and $\sigma_{y,\tau}$ from a_τ, $\mu_{x,\tau}$, $\sigma_{x,\tau}$ by using Eqs. 4-34 and 4-35.

STEP 4 Compute $r_y(k,\tau)$, k = 1,2,..., m, from $r_x(k,\tau)$ and $\sigma_{y,\tau}$ by using Eq. 4-37.

STEP 5 Compute autoregressive coefficients $\alpha_{k,\tau}$, and the standard deviation of residuals $\sigma_{\xi,\tau}$ from $r_y(k,\tau)$ by using Eqs. 4-23 and 4-24 for the first order model, Eqs. 4-27 and 4-28 for the second-order model, Eqs. 4-30 and 4-31 for the third order model, respectively.

STEP 6 Compute lower limits ε^*_τ for days that have negative values of a_τ by using Eq. 4-40.

STEP 7 Generate standard normal random variables $\xi_{p,\tau}$, for p = 1,2,...,n, τ = -10,-9,-8,...,-1,1,2,...,ω, where n = the total number of years of generated samples, and ω the within-the-year period of generation (note that $\xi_{p,\tau}$, for τ = -10,-9,...-1, are used as to avoid biases at generated τ = 1,2,...).

STEP 8 Introduce dependence to $\xi_{p,\tau}$ series, as the $\varepsilon_{p,\tau}$ series, by Eq. 4-20; check for every day τ whether a_τ is negative, and if $\varepsilon_{p,\tau}$ is smaller than the lower limit ε^*_τ, set $\varepsilon_{p,\tau} = \varepsilon^*_\tau$, and continue generation.

STEP 9 Discard the $\varepsilon_{p,\tau}$-values for τ = -10,-9,...-1, with only $\varepsilon_{p,\tau}$, τ = 1,2,..., , used in the next step.

STEP 10 Compute the transformed series $Y_{p,\tau}$ by

$$Y_{p,\tau} = \varepsilon_{p,\tau}\sigma_{y,\tau} + \mu_{y,\tau} \quad .$$

STEP 11 Produce generated daily flow series of $X_{p,\tau}$ from $Y_{p,\tau}$ by the inversed transformation

$$X_{p,\tau} = a_\tau + \exp(Y_{p,\tau}).$$

Chapter V
APPLICATION OF THEORY OF PROBABILITY DISTRIBUTIONS OF PARTIAL SERIES OF FLOOD PEAKS

The theory of probability distributions of partial series of flood peaks is applied to the observed data. The partial flood series are obtained from 17 sets of daily flow series for gaging stations located throughout the United States. Discrete and continuous probability distribution functions, described in Chapter III, are applied to frequency distributions of the number and the magnitude of exceedances above the selected truncation level of partial flood peak series, respectively, in order to find out the best fitting functions. After the best distribution functions are inferred, the change of the goodness of fit indices of selected functions with the truncation level is investigated. Also changes of parameters of these distributions with the truncation level are studied. In addition, derivation of probability distribution function of the largest exceedance in the year is presented.

The statistical dependence of partial and annual flood series is investigated at the end of this chapter. Also, for partial flood series, the study of the change in series dependence with the change in truncation level is included.

5.1 Research Data Used

The data used in this study (Quimpo, 1967; Tao, 1973) contain 17 series of daily flows from which 17 sets of partial flood series are derived. These 17 daily flow series are from runoff records published by the U.S. Geological Survey under the condition that the flows are sufficiently virgin, or have not been altered by significant man-made diversions or flow regulations.

The names of gaging stations, their locations, drainage areas, mean flow, and other pertinent information are given in Table 5-1, with the approximate geographic location of these stations shown in Fig. 5-1.

5.2 Chi-Square Goodness-of-Fit Test Statistic

Generally, a goodness-of-fit statistic is useful to discriminate between fits of different probability distribution functions to the same frequency distribution. If a single sample is available, the goodness-of-fit statistic of each fitted distribution function is computed, and the distribution selected with the smallest statistic. Several test statistics may be used in testing goodness-of-fit of probability distribution functions. The chi-square and the Smirnov-Kolmogorov goodness-of-fit test statistics are well known and frequently applied in statistics and hydrology. Test by Smirnov-Kolmogorov statistic is nonparametric or distribution-free. However, in case of goodness of fit, parameters of hypothetical distribution functions, which are fitted to frequency distribution, are estimated from the sample data, the Smirnov-Kolmogorov test is not appropriate and not used in this study.

For chi-square statistic, the range of variable values is divided into k mutually exclusive and exhaustive class intervals, each with a class frequency O_j and expected class probability E_j $(j=1,2,...,k)$. The quantity $(O_j - E_j)^2$ is used as a measure of departure from E_j, but they cannot be compared from one class to another without scaling each class interval

Table 5-1. Stations Selected for Investigation

Station Number	USGS Station Number	River	Location		Area (Sq. Mi.)	Records Available	Mean Daily Flow	Standard Deviation	Remarks on Accuracy of Record*
			Latitude	Longitude					
1	1B.6265	Tioga near Erwins, N. Y.	42°07'	77°08'	1370.0	1921-1960	1378.6	2777.8	Excellent. Fair during periods of ice effect.
2	4.0710	Oconto near Gillett, Wisconsin	44°52'	88°18'	678.0	1921-1960	543.5	441.0	Good. Fair during periods of ice effect.
3	7.0670	Current at Van Buren, Mo.	37°00'	91°01'	1667.0	1922-1960	1921.0	2694.3	Good. Poor during periods of ice effect.
4	14.1590	Mckenzie at Mckenzie Br., Ore.	44°11'	122°08'	345.0	1924-1960	1638.2	744.4	Excellent
5	8.3335	Neches near Rockland, Tex.	31°02'	94°24'	3539.0	1924-1960	2385.2	3813.0	Good
6	13.1850	Boise near Twin Springs, Idaho	43°40'	115°44'	830.0	1921-1960	1172.7	1458.6	Excellent. Good during periods of ice effect.
7	11.2750	Falls Creek near Hetch-hetchy, Cal.	37°58'	119°46'	45.2	1923-1960	141.2	234.2	Good. Fair during periods of ice effect.
8	3A.1835	Greenbrier near Alderson, W. Va.	37°44'	80°38'	1357.0	1921-1960	1885.5	3053.4	Good. Poor during periods of ice effect.
9	6B.8905	Delaware at Valley Falls, Kansas	39°21'	95°27'	922.0	1923-1960	375.9	1617.7	Good. Fair during periods of ice effect.
10	6A.0375	Madison near W. Yellowstone, Mont.	44°39'	111°04'	419.0	1924-1960	458.6	190.7	Excellent. Good during periods of ice effect.
11	3B.5320	Powell near Arthur, Tenn.	36°32'	83°38'	683.0	1921-1960	1116.1	1739.0	Good
12	12.1150	St. Maries near Lotus, Idaho	47°15'	116°38'	437.0	1923-1960	515.0	762.3	Good. Poor during periods of ice effect.
13	2A.0160	Cowpasture near Clifton Forge, Va.	37°48'	79°46'	456.0	1926-1960	515.6	762.3	Good
14	3A.2695	Mad near Springfield, Ohio	39°55'	83°52'	1474.0	1921-1960	487.2	686.7	Good
15	11.2665	Merced at Pohono Br., Yosemite, Cal.	37°43'	119°40'	321.0	1921-1960	595.7	979.4	Good
16	1B.3295	Batten kill at Battenville, N. Y.	43°06'	73°25'	394.0	1923-1960	722.9	722.9	Good. Fair during periods of ice effect.
17	5.3620	Jump near Sheldon, Wisconsin	45°18'	90°57'	574.0	1921-1960	505.0	1162.0	Good. Fair during periods of ice effect.

*According to USGS, the classification of the records are excellent, good, fair, or poor depending on whether errors in them are less than 5, 10, or 15 percent or greater than 15 percent, respectively.

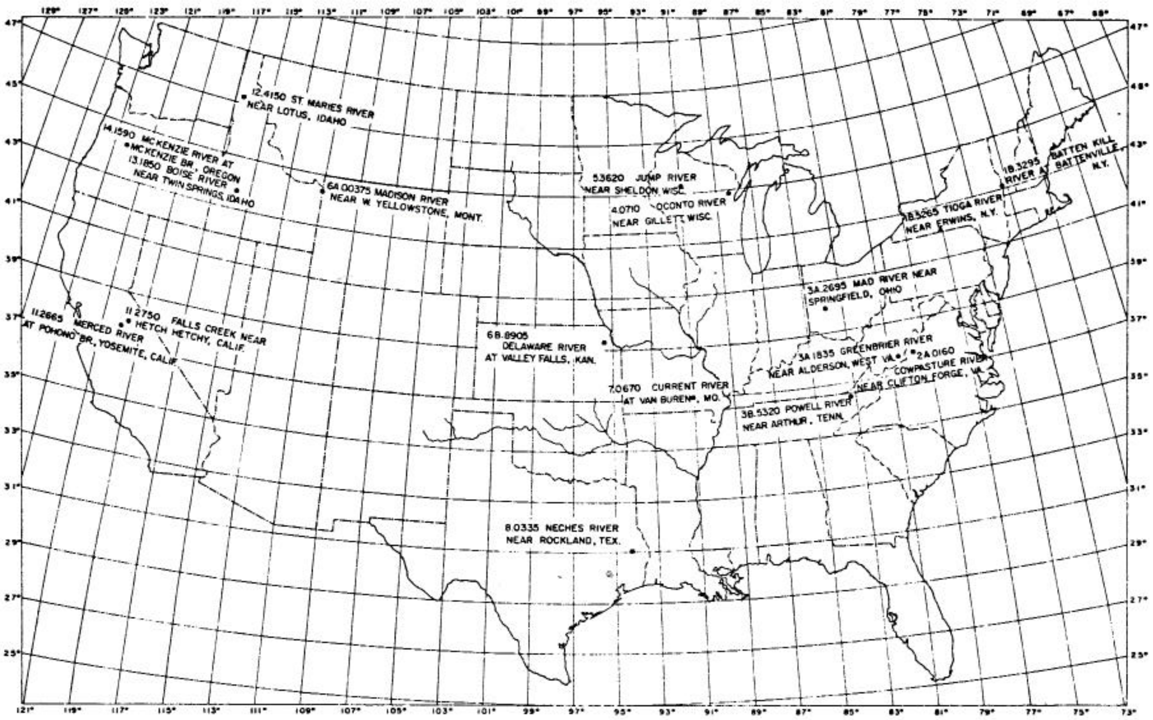

Fig. 5-1 Geographic Distribution of Selected Stations

proportionally to E_j. The measure used is $(O_j - E_j)^2/E_j$ and the test statistic of a fit is

$$\chi^2 = \sum_{j=1}^{k} \frac{(O_j - E_j)^2}{E_j} \quad . \tag{5-1}$$

This statistic is asymptotically chi-square distributed, with k-1 degree of freedom. When population parameters are estimated from sample data, the number of degree of freedom is then decreased by the number of estimated parameters. For m parameters, the total number of degree of freedom is

$$f = k - 1 - m \tag{5-2}$$

5.3 Distribution of the Number of Exceedances

The 17 sets of partial flood series are obtained from the 17 sets of daily flow series as described above. The series are exceedances above the selected truncation level Q_b and are functions of it. For each station, 9-11 truncation levels were selected so that the average number of exceedances per year, $\bar{\eta}$, would vary from 1 to 4.5.

Comparison of Discrete Probability Distribution Functions for the Number of Exceedances. Five discrete probability distributions outlined in Chapter III were fitted to frequency distributions of η for the 17 stations. The chi-square test statistics were calculated for all five distributions: Poisson, mixed Poisson, Hyper-Poisson, negative binomial, and the mixture of two geometrics, as well as for all stations.

The comparison of best fits of selected distribution functions is made in two steps: (1) Compare the goodness-of-fit statistics for each station series and various truncation levels; and (2) Compare the goodness-of-fit statistics for the 17 stations. The selected distribution is that one which has, on the average, the smallest goodness-of-fit statistic for all the selected truncation levels and for all the stations.

The chi-square statistic is affected by the degree of freedom, which in turn depends on the number of parameters of distribution functions used. Instead of comparing the computed chi-square statistics directly, the exceedance probability of these chi-square values is used, in order to remove in comparison the effect of degrees of freedom. The function with the largest exceedance probability of the computed chi-square is conceived as the best fit distribution. If 95 percent significance level is used in testing the goodness of fit, the fitted function that has the exceedance probability of computed chi-square less than 5 percent is rejected. Let χ^2 denote the computed chi-square, and $P(\chi^2)$ its non-exceedance probability, or the exceedance probability is $1 - P(\chi^2)$. The computed $1 - P(\chi^2)$ for five distributions, for various Q_b, and for station 1, are given as an example in Table 5-2.

For each Q_b four distributions (except Poisson) are ranked by their statistic values, from the largest to the smallest. The distribution that gives the largest value of $1 - P(\chi^2)$ is ranked No. 1 for a given Q_b. At the bottom of each column the sum of the ranks attributed to each distribution, and the total number of times (or Q_b) that the distribution was rejected by the chi-square test at the 95 percent significance level are shown.

Table 5-2. Comparison of Goodness-of-Fit Statistic Based on $1 - P(\chi^2)$ for Distribution Functions of the Number of Exceedances, for Various Truncation Levels and Station No. 1

Truncation Level	Number of Class Intervals	Statistic $1 - P(\chi^2)$				
		Poisson	Mixed Poisson or Poisson	Hyper-Poisson	Negative Binomial	Mixture of Geometrics
10800	10	0.005	0.068^{2*}	0.054^{3*}	0.081^{1*}	0.028^{4*}
11500	10	0.426	0.426^{2}	0.384^{3}	0.428^{1}	0.218^{4}
12000	10	0.297	0.423^{2}	0.402^{4}	0.511^{1}	0.405^{3}
12500	10	0.730	0.730^{1}	0.728^{2}	0.725^{3}	0.488^{4}
13500	10	0.591	0.626^{3}	0.705^{2}	0.723^{1}	0.601^{4}
14000	9	0.648	0.648^{1}	0.617^{2}	0.575^{3}	0.529^{4}
14500	9	0.833	0.833^{1}	0.618^{3}	0.730^{2}	0.580^{4}
15000	9	0.298	0.298^{1}	0.105^{3}	0.176^{2}	0.024^{4}
16000	9	0.398	0.398^{1}	0.214^{3}	0.267^{2}	0.056^{4}
17000	9	0.435	0.435^{1}	0.098^{3}	0.298^{2}	0.009^{4}
18000	8	0.955	0.955^{1}	0.807^{3}	0.863^{2}	0.582^{4}
Sum of Ranks		--	16	31	20	43
Number of Times Distribution is Rejected by Chi-Square Test		1	0	0	0	3

*Rank of $1 - P(\chi^2)$ attributes to each distribution function for a given truncation level Q_b. Distribution function which has the smallest number of rank is considered as the best fit function for a given Q_b.

Distributions are further ranked on the basis of sums of ranks for each station. Entries in Table 5-3 give sums of ranks, for different truncation levels of all the stations. For each station four distributions are ranked by sums of ranks for all the truncation levels, from the smallest to the largest. Sums of the new ranks attributed to each distribution are shown at the bottom of each column.

The number of times (or Q_b) that distribution functions are rejected by the chi-square test at the 95 percent significance level for each station are given in Table 5-4. The total number of times and the percentage of times that a distribution function is rejected, for all Q_b and for all stations, are shown at the bottom of each column.

These data show that:

(i) The one-parameter Poisson distribution cannot pass the chi-square test at the 95 percent significance level for all the stations studied;

(ii) Based on the results of Tables 5-3 and 5-4, the mixed Poisson or only the Poisson distribution, as the case may be, give the best fit among all the considered discrete distribution functions;

(iii) The mixed Poisson or the Poisson distribution, as the case may be, pass the chi-square test for all stations and for the most interesting range of truncation levels; and

(iv) The four-parameter distribution does not give any improvement in the goodness-or-fit for the test criterion selected.

In conclusion, the Poisson distribution with one parameter is not always sufficient to fit frequency distributions of η for all stations. It does not pass the chi-square test at the 95 percent significance level for all stations. However, for all stations and all selected Q_b, the percentage of times that it is rejected by chi-square test is 26.55 percent.

The other four distributions also were applied with the goodness-of-fit tested by chi-square statistic. On the average for 17 stations, the mixed Poisson or the Poisson as the case may be, gives the best fit among the considered distribution functions. This distribution can pass chi-square tests with 95 percent significance level for the range of interesting Q_b and for all stations. The percentage of times that it is rejected by the chi-square test is 5.08 percent.

Table 5-3. Goodness-of-Fit for Distribution Function of Number of Exceedances Based on Sums of Ranks of All Truncation Levels, for $1 - P(\chi^2)$ Statistic

Station Number	Sums of Ranks of All Truncation Levels			
	Mixed Poisson or Poisson	Hyper-Poisson	Negative Binomial	Mixture of Geometrics
1	16^{1*}	31^{3*}	20^{2*}	43^{4*}
2	15^{1}	25^{3}	15^{2}	35^{4}
3	29^{3}	13^{1}	24^{2}	44^{4}
4	16^{1}	21^{2}	28^{3}	35^{4}
5	25^{3}	20^{1}	25^{2}	30^{4}
6	26^{3}	15^{1}	22^{2}	27^{4}
7	28^{2}	23^{1}	29^{3}	30^{4}
8	16^{1}	31^{3}	22^{2}	41^{4}
9	26^{2}	16^{1}	26^{3}	32^{4}
10	17^{1}	35^{4}	31^{3}	27^{2}
11	20^{1}	22^{2}	24^{3}	44^{4}
12	20^{1}	24^{2}	28^{3}	38^{4}
13	19^{1}	21^{2}	27^{3}	43^{4}
14	27^{2}	14^{1}	29^{3}	40^{4}
15	26^{4}	18^{1}	26^{3}	20^{2}
16	17^{1}	22^{2}	24^{3}	37^{4}
17	21^{1}	38^{4}	22^{2}	29^{3}
Sums of Ranks	29**	34	44	63

*Rank based on sums of ranks of all Q_b, attributes to each distribution function for a given station. Distribution function which has the smallest number of rank is considered as the best fit function for a given station.

**The smallest number of sums of ranks indicates that the mixed Poisson or Poisson distribution gives the best fit.

Table 5-4. Number of Times (or Truncation Levels) that Distribution Functions are Rejected by Chi-Square Test at the 95 percent Significance Level, for Fitting the Frequency Distributions of Number of Exceedances

Station Number	Total Times or Truncation Levels	Number of Times that Distribution is Rejected				
		Poisson	Mixed Poisson or Poisson	Hyper-Poisson	Negative Binomial	Mixture of Geometrics
1	11	1	0	0	0	3
2	9	1	0	0	0	1
3	11	9	1	1	1	1
4	10	1	0	0	1	1
5	10	3	0	0	1	0
6	9	1	1	2	1	1
7	11	2	1	2	1	1
8	11	0	0	0	0	0
9	10	7	2	1	2	2
10	11	0	0	7	1	0
11	11	0	0	0	0	3
12	11	10	1	3	0	0
13	11	0	0	0	0	0
14	11	4	1	0	1	2
15	9	2	2	3	4	1
16	10	0	0	0	0	0
17	11	6	0	0	0	2
Total	177	47	9	19	13	18
Percent	100	26.55	5.08	10.73	7.34	10.17

Change of Goodness-of-Fit Statistic of Selected Distribution with a Change of Truncation Level. The chi-square statistic, expressed as the exceedance probability $1 - P(\chi^2)$, of the selected distribution, is investigated to determine its change with the change of truncation level. For each station, 9-11 truncation levels were selected in such a way that $\bar{\eta}$ varies from 1 to 4.5. For convenience, the truncation level is expressed in terms of $\bar{\eta}$. The relationships between $1 - P(\chi^2)$ and $\bar{\eta}$ for all stations are plotted and studied. By using the method of interpolation, the average $1 - P(\chi^2)$ at the particular $\bar{\eta}$ in the range of 1 to 4.5 are obtained.

The change of the average $1 - P(\chi^2)$ for all stations with the truncation level for the Poisson distribution and the mixed Poisson distribution are shown by a dotted line and a full line in Fig. 5-2, respectively.

Figure 5-2 shows that:

(i) The goodness-of-fit by chi-square statistic and for both distributions tend to be better as the truncation level increases;

(ii) The mixed Poisson distribution, when it can be applied, is an improvement, especially for the smaller truncation levels; and

(iii) There is a tendency for the Poisson distribution to be rejected on the average for truncation levels that have $\bar{\eta}$ greater than 5.5, approximately. For some stations the Poisson distribution is rejected for most truncation levels of the practical range.

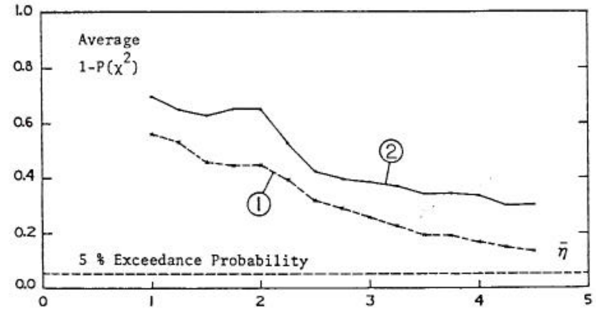

Fig. 5-2 Relationship of Average $1 - P(\chi^2)$ to the Truncation Level, Expressed as the Average Number of Exceedances per year, $\bar{\eta}$ for: (1) Poisson Distribution, and (2) Mixed Poisson Distribution, as Averages for All 17 Stations

Changes in Parameters of Selected Distribution with a change of Truncation Level. In order to study how the Poisson or mixed Poisson distribution can be applied, ratios $R_{m,v}$ of mean to variance of frequency distributions of η are plotted against the truncation levels for all the stations. Results are shown in Fig. 5-3 for stations nos. 1-9, and Fig. 5-4 for stations nos. 10-17. Departure from the Poisson distribution depends on the departure of the ratio $R_{m,v}$ from unity. When the ratio is greater than unity the Poisson distribution may be still applicable, such as for cases of stations nos. 6, 10 and 15. For ratios less than one, such as for stations nos. 3, 9, 12 and 17, the application of Poisson distribution is rejected by the chi-square test. In these particular cases the mixed Poisson with three parameters can be applied, since one of its properties is for the variance to be greater than the mean.

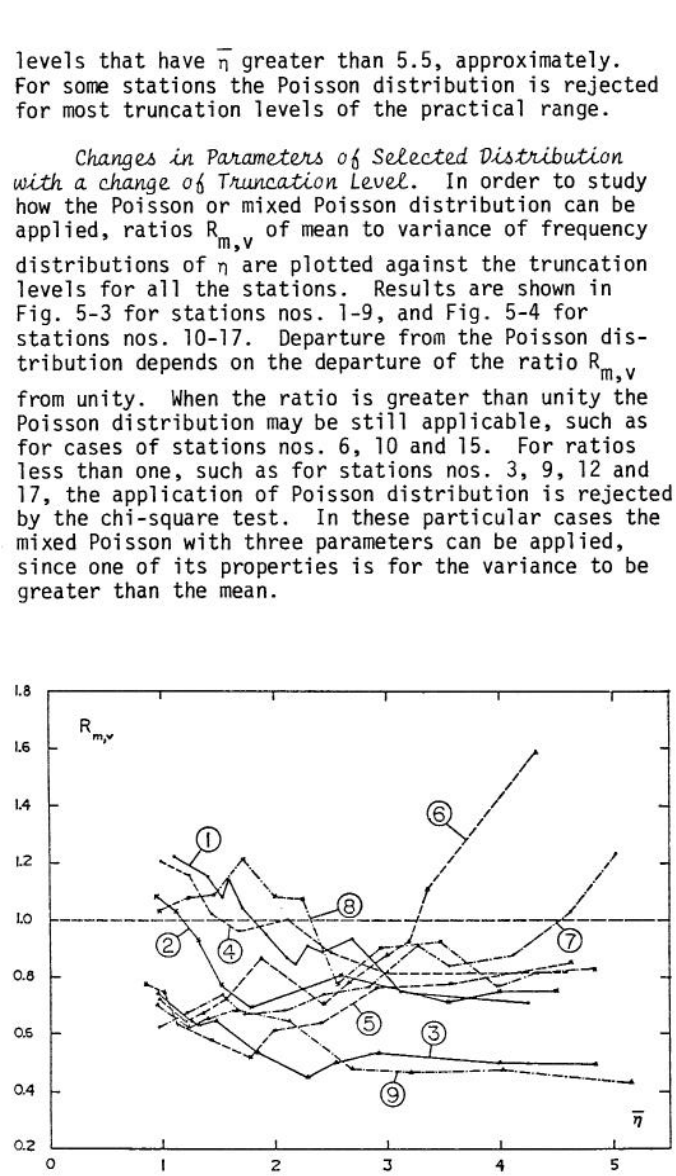

Fig. 5-3 Relationship between Ratio $R_{m,v}$ of Mean to Variance and Truncation Level (Expressed as the Average Number of Exceedances per Year, $\bar{\eta}$) for Distributions of the Number of Exceedances, Station Nos. 1-9

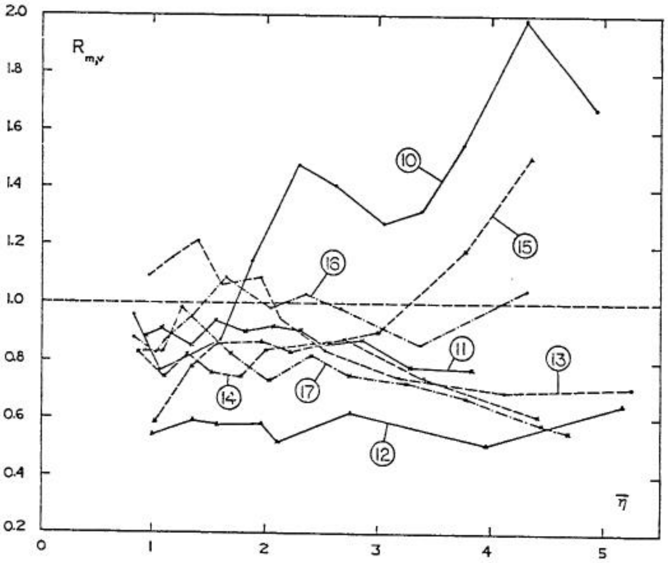

Fig. 5-4 Relationship between Ratio $R_{m,v}$ of Mean to Variance and Truncation Level (Expressed as the Average Number of Exceedances per Year, $\bar{\eta}$) for Distributions of the Number of Exceedances, Station Nos. 10-17

For stations for which the daily flow series fluctuates highly, the ratio $R_{m,v}$ of the number of exceedances tends to be smaller than unity. In this case the mixed Poisson distribution is applicable. In the opposite case, the Poisson distribution is acceptable. For stations such as nos. 1, 2 and 13, both distributions should be applied in passing the chi-square test for the whole range of Q_b.

Since the Poisson distribution has only one parameter, λ, and the maximum likelihood estimate of λ is the average number of exceedances, λ decreases with an increase of Q_b.

The mixed Poisson distribution has three parameters: λ_1, λ_2 and p, with $\lambda_1 > \lambda_2$, and p the proportion for Poisson with parameter λ_1. By definition, $\lambda_1 > \lambda > \lambda_2$, and $0 > p > 1$, where λ is a parameter if the Poisson distribution is applied. The larger p in the mixed Poisson distribution, the closer the values of parameters λ_1 and λ. Table 5-5 shows how p, λ_1, and λ_2 for the frequency distributions of η for station no. 17 change with the change of Q_b. For this case, p decreases with an increase of Q_b and of $R_{m,v}$. This table also shows when $R_{m,v}$ is close to unity, as Q_b increases, the Poisson distribution becomes a good fit. For example, for Q_b = 5500 cfs, $R_{m,v}$ approaches to unity and $\lambda_2 \approx \lambda$, hence, p approaches to zero.

5.4 Distribution of the Magnitude of Exceedance

The frequency distributions of the magnitude of exceedance of partial flood series for 17 daily flow series are used in the study for selecting the probability distribution function of best fit. The chi-square statistic is used for testing the goodness-of-fit. The procedure used for the number of exceedances is also applied for this case.

Comparison of Continuous Probability Distribution Functions for Magnitude of Exceedances. The investigation is divided into two steps. In the first, preliminary step, the five continuous distribution functions: exponential, gamma, Pearson Type III, Weibull and lognormal are used to fit the frequency distributions of the magnitude of exceedance ξ_v. Detailed computations are not presented. Results show that exponential, gamma, and Weibull distributions have a more close fit than the other distributions. It is difficult to distinguish which one of these three distributions fits best. The percentage of times that each distribution is rejected by the chi-square test at the 95 percent significant level for all stations did not come to be 5 percent or less. These investigations passed to the second step.

A mixed exponential distribution with three parameters was applied, or an exponential distribution if a mixed exponential cannot be applied. The goodness-

Table 5-5. The Change of Parameters λ_1, λ_2, and p with the Change of Truncation Level for Distributions of Number of Exceedances, for Station No. 17

	Truncation Level, Q_b										
	1670	1850	2000	2500	3000	3500	4000	4500	5000	5500	6000
Mean or λ	5.154	4.667	4.462	3.769	3.256	2.744	2.410	2.026	1.692	1.256	1.077
Ratio $R_{m,v}$	0.548	0.550	0.577	0.671	0.741	0.746	0.814	0.725	0.822	0.991	0.827
p	0.512	0.554	0.559	0.479	0.318	0.385	0.486	0.199	0.181	0.00002	0.179
λ_1	7.168	6.419	6.070	5.186	4.821	3.966	3.174	3.787	2.980	26.509	2.092
λ_2	3.039	2.491	2.426	2.468	2.528	1.979	1.690	1.589	1.407	1.256	0.855
χ^2 by Mixed Poisson	9.251	6.530	9.755	17.362	14.560	14.444	7.197	3.076	2.461	5.102	3.709
χ^2 by Poisson	36.588	26.658	33.084	35.119	27.798	18.980	8.455	8.842	4.468	5.098	6.747

of-fit is compared for the three distributions. Sums of ranks for all the truncation levels for the statistic $1 - P(\chi^2)$ for each station are given in Table 5-6. The sums of new ranks for all stations and for the three distributions are determined and given at the bottom line of each column. The number of times (or Q_b) that a distribution is rejected by the chi-square test at the 95 percent significance level for each station is given in Table 5-7, with the total number of times of rejection shown at the bottom line of each column.

The results show that the mixed exponential or the exponential distribution give the best goodness-of-fit by chi-square test statistic. The percent of times that this distribution is rejected by the test is 13.57. The percent of times that it is rejected by the test is greater than five. Only station no. 5 was rejected by the chi-square test at the 95 percent significance level for all selected Q_b. This number affects the total percent of times. However, for this station, only five out of eight truncation levels were rejected by the chi-square test at the 97.5 percent significance level, and four out of eight truncation levels at the 99 percent significance level.

Table 5-6. Goodness-of-Fit for Distribution Function of Magnitude of Exceedances Based on Sums of Ranks of All Truncation Levels for $1 - P(\chi^2)$ Statistic

Station Number	Sums of Ranks of All Truncation Levels: Mixed Exponential or Exponential	Gamma	Weibull
1	17^{1*}	18^{2*}	19^{3*}
2	14^1	17^2	17^3
3	17^1	18^2	19^3
4	16^2	17^3	15^1
5	23^3	16^2	9^1
6	22^3	17^2	9^1
7	8^1	24^3	16^2
8	10^1	24^3	14^2
9	15^1	20^3	19^2
10	17^2	19^3	12^1
11	14^1	18^3	16^2
12	10^1	19^3	19^2
13	17^3	15^1	16^2
14	14^1	17^2	17^3
15	16^2	9^1	17^3
16	10^1	23^3	15^2
17	15^1	27^3	18^2
Sums of Ranks	26**	41	35

*Rank based on sums of ranks of all Q_b, attributes to each distribution function for a given station. Distribution function which has the smallest number of rank is considered as the best fit function for a given station.

**The smallest number of sums of ranks indicates that the mixed Exponential or Exponential distribution gives the best fit.

Table 5-7. Number of Times (or Truncation Levels) that Distribution Functions are Rejected by Chi-Square Test at the 95 Percent Significance Level, for Fitting the Frequency Distributions of Magnitude of Exceedances

Station Number	Number of Times that Distribution is Rejected: Exponential	Mixed Exponential or Exponential	Gamma	Weibull
1	4	3	6	6
2	0	0	0	0
3	1	1	0	0
4	1	1	0	0
5	8	8	8	8
6	1	0	0	0
7	8	1	8	8
8	0	0	0	0
9	2	0	0	0
10	0	0	1	1
11	0	0	0	0
12	0	0	0	0
13	2	2	1	1
14	0	0	0	0
15	1	1	0	1
16	5	0	6	5
17	10	2	7	4
Total	43	19	37	34
Percent	30.71	13.57	26.43	24.29

Two reasons may be responsible for the chi-square test to reject this particular river: (1) The partial flood series is approximately derived from the mean daily flow series and not from the instantaneous flow peak series; and (2) This effect may be reinforced by the large outliers, since all the considered distributions were rejected. For this particular river, the catchment area of about 3539 square miles is the largest among all the considered 17 rivers.

Change of Goodness-of-Fit Statistic of Selected Distribution with a Change of Truncation Level. The same procedure outlined is used here as for the distribution of the number of exceedances, for the change of goodness- of-fit statistic with a change of the truncation level.

The changes of the statistic of the average $1 - P(\chi^2)$, for all stations with the change of the truncation level (in this case expressed by $\bar{\eta}$) for the exponential and the mixed exponential distributions are shown by the dotted line and full line in Fig. 5-5.

Figure 5-5 shows, for the averages of all stations, that:

(i) The goodness-of-fit of chi-square statistics for both distributions tend to increase for high truncation levels;

(ii) The mixed exponential distribution, when it can be applied, gives a goodness-of-fit improvement, especially for the low truncation levels; and

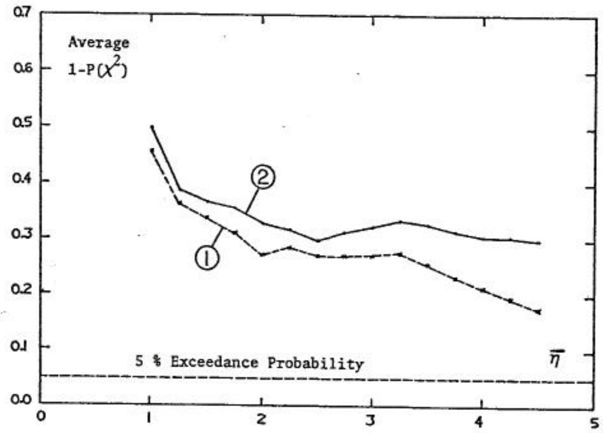

Fig. 5-5 Relationship of Average $1 - P(\chi^2)$ to the Truncation Level, Expressed as the Average Number of Exceedances per Year, $\bar{\eta}$ for: (1) Exponential Distribution, and (2) Mixed Exponential Distribution, as Averages for All 17 Stations.

(iii) The goodness-of-fit for the exponential distribution decreases rapidly with a decrease of the truncation level in the range of lower truncation levels (or for $\bar{\eta}$ larger than 3).

Changes in Parameters of Selected Distribution with a Change of Truncation Level. Exponential distribution parameter, β, estimated by the sample mean by the maximum likelihood implies that the change of the mean with the truncation level is the change of parameter β with truncation level. By studying the 17 frequency distributions of ξ_ν, it is not conclusive how β changes with Q_b. For stations nos. 2,5,9 and 16, β clearly increases with an increase of Q_b. For stations nos. 6 and 10, β clearly decreases with an increase of Q_b. For other stations, the change in β with Q_b is not clear.

Three statistics, the coefficients of variation, skewness and kurtosis of frequency distributions of ξ_ν are investigated to find the ranges in which the mixed exponential distribution should be applied. In case of coefficients of variation, the results are not quite conclusive, except that the mixed exponential distribution can be applied only in the range of high coefficients of variation, and not applied if it is less than unity. The population coefficient of variation of exponential distribution is unity, so that it is applicable in a range of values close to unity.

The mixed exponential distribution can be applied if the skewness coefficient γ is greater than two, except for station no. 13. For it the distribution is not applicable in the range of coefficients 2.4 to 3.1, but is applicable for values greater than 3.1. Figure 5-6 shows how the skewness coefficients of distributions of ξ_ν for stations nos. 1-9 change with Q_b. The variation in skewness with Q_b for stations now. 10-17 are shown in Fig. 5-7. The higher the skewness coefficient, the more opportunity is there for the mixed exponential distribution to be applicable, with a better goodness-of-fit than for the exponential distribution.

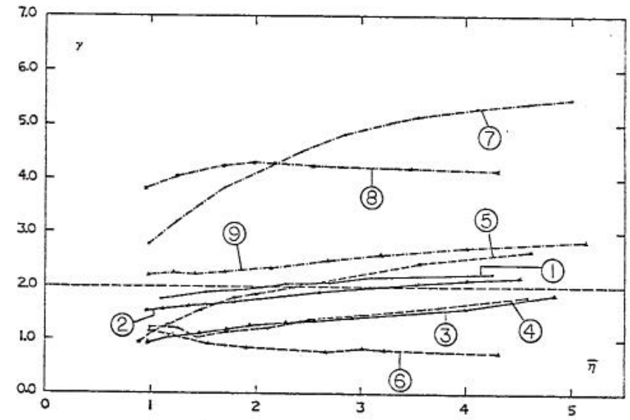

Fig. 5-6 Variation of Skewness Coefficients γ of Distributions of the Magnitude of Exceedance with the Truncation Level (Expressed as the Average Number of Exceedances per Year, $\bar{\eta}$) for Station Nos. 1-9

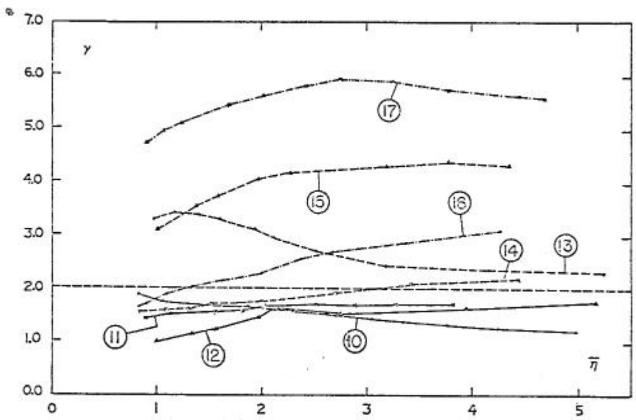

Fig. 5-7 Variation of Skewness Coefficients γ of Distributions of the Magnitude of Exceedance with the Truncation Level (Expressed as the Average Number of Exceedances per Year, $\bar{\eta}$) for Station Nos. 10-17

The mixed exponential distribution can be applied if the kurtosis coefficient is approximately greater than 7.5, except for station no. 13. For it the mixed exponential distribution is not applicable in the range of coefficients 13.5 to 16.8, but is applicable for values above that range. The larger a value of the kurtosis coefficient, the better the application of the mixed exponential distribution.

For the mixed exponential distribution, with three parameters β_1, β_2, and p, and without any loss of generality, let $\beta_1 > \beta_2$, and let p indicate the proportion for the exponential with the parameter $_1$. By definition, $\beta_1 > \beta > \beta_2$, $0 < p < 1$, where β is a parameter if the exponential distribution is applied. For stations that the mixed exponential distribution can be applied for the whole range of Q_b, such as stations nos. 7, 8, 9, 15 and 17, the proportion parameter p tends to increase with an increase of Q_b. Table 5-8 shows how parameters β_1, β_2, and p for station no. 17 change with Q_b. The change of moments and chi-square statistic for both distributions, are also included.

Table 5-8. The Change of Moments, Parameters p, β_1 and β_2, and Chi-Square Statistic with the Truncation Level for the Distribution of the Magnitude of Exceedance, for Station No. 17

	Truncation Level, Q_b									
	1850	2000	2500	3000	3500	4000	4500	5000	5500	6000
Mean or β	3060.4	3047.5	3054.7	2992.8	2998.7	2867.2	2864.6	2884.9	3296.5	3307.8
Coefficient of Variation	1.19	1.21	1.24	1.30	1.35	1.46	1.54	1.61	1.54	1.61
Skewness	5.558	5.606	5.739	5.813	5.828	5.764	5.632	5.449	5.109	4.940
Kurtosis	52.02	52.06	51.50	50.61	48.26	45.87	42.33	38.45	32.26	29.38
p	0.014	0.016	0.020	0.031	0.041	0.064	0.082	0.102	0.111	0.135
β_1	14969.2	14580.2	14077.9	12837.7	12370.5	11149.6	10772.7	10509.8	11015.5	10746.8
β_2	2892.6	2862.8	2825.9	2674.3	2601.1	2301.1	2157.6	2017.2	2329.2	2143.2
χ^2 by Mixed Exponential	3.275	3.387	4.003	3.202	3.860	5.933	8.152	9.684	3.472	2.569
χ^2 by Exponential	26.861	28.521	30.550	36.879	38.491	56.742	43.652	43.589	27.572	28.670

5.5 Probability Distribution of the Largest Exceedance

The main purpose of the study of partial flood series is to develop the probability distribution of the largest exceedance in a year. This distribution can be then used to estimate flood exceedances for given annual return periods. It can be derived by using the combination of distributions of the number and the magnitude of exceedances above the selected truncation level.

Let η represent the number of exceedances in a year and $\{\xi_\nu\}_1^\infty$, represent a sequence of the magnitude of those exceedances. It is shown in Section 3.4 that the distribution of the largest exceedance in a year is expressed by

$$F(x) = P(\eta=0) + \sum_{k=1}^{\infty} P[\max_{1\le\nu\le k} \xi_\nu \le x \cap \eta=k], \qquad (5\text{-}3)$$

with $\max_{1\le\nu\le k} \xi_\nu = \max(\xi_1,\xi_2,\ldots,\xi_k)$ = the random variable which represents the largest exceedance in a year.

Two assumptions are used in order to simplify the application of Eq. 5-3. The first assumption is that $\{\xi_\nu\}_1^\infty$ are independent of η. The second assumption is that $\{\xi_\nu\}_1^\infty$ are mutually independent random variables with the common distribution function H(x).

The test whether $\{\xi_\nu\}_1^\infty$ are independent of η, the exceedances ξ_ν are divided into groups which have the same number of exceedances per year, η. Because of short sample data, the $\{\xi_\nu\}_1^\infty$ are divided into only two groups. The exceedances $\{\xi_\nu\}_1^\infty$ with small η are combined into one group, while other exceedances with larger η are considered as the other group. For example, the first group may consist of all exceedances with η = 1, 2, and 3. All other exceedances are then considered as the second group. The idea is that the total number of exceedances in each group should be close together. The two-sample-Smirnov-Kolmogorov test is then used to test the hypothesis that the distributions of the ξ_ν's corresponding to the two groups are identical. Under the null hypothesis of equality of the two distributions, the statistic

$$\Delta = \max_x |H_1(x) - H_2(x)| \qquad (5\text{-}4)$$

has some distribution whose 95 percent quantile is approximated by

$$\Delta_c = 1.358 \sqrt{\frac{n_1 + n_2}{n_1 n_2}}, \qquad (5\text{-}5)$$

where $H_1(x)$ is the sample distribution function of ξ_ν's corresponding to the first group with sample size n_1 and $H_2(x)$ is the sample distribution function of the ξ_ν's of the second group with sample size n_2.

The results of the test for 24 exceedance series of 12 stations, each with two selected truncation levels, are shown in Table 5-9. This table includes station number, truncation level, average number of exceedances per year, sample size of each group, the computed Δ, and the critical value Δ_c. All of the computed Δ are less than Δ_c. This implies that the null hypothesis cannot be rejected at the 95 percent significance level. That is, the hypothesis that

Table 5-9. Two-Sample-Smirnov-Kolmogorov Test for the Hypothesis that the Distributions of the ξ_ν's Corresponding to the Two Selected Groups are Identical

Station Number	Truncation Level Q_b	Average Number of ξ_ν per Year, $\bar{\eta}$	Sample Sizes n_1	n_2	Smirnov-Kolmogorov Statistics Computed Values Δ	Critical Values Δ_c
1	11500	2.69	56	49	0.102	0.266
1	12500	2.28	52	37	0.096	0.292
2	1000	3.54	64	74	0.150	0.232
2	1250	2.59	53	48	0.130	0.271
3	6000	4.00	58	94	0.190	0.227
3	9000	2.55	53	44	0.090	0.277
4	2500	4.58	89	76	0.090	0.212
4	3250	2.50	44	46	0.200	0.286
5	5000	3.56	51	77	0.210	0.245
5	7000	2.42	43	44	0.185	0.291
6	4000	3.00	68	49	0.220	0.255
6	4250	2.67	45	59	0.209	0.269
7	600	4.14	92	61	0.157	0.224
7	700	3.27	70	51	0.235	0.250
9	4000	4.03	66	83	0.170	0.224
9	6000	2.67	44	55	0.100	0.275
11	7000	3.28	62	66	0.070	0.240
11	8000	2.51	55	43	0.105	0.276
12	2500	2.76	58	44	0.200	0.272
12	3250	1.97	34	39	0.290	0.319
16	2500	4.27	90	68	0.095	0.218
16	3000	2.65	47	51	0.130	0.275
17	3000	3.26	57	70	0.150	0.242
17	3500	2.74	51	56	0.050	0.263

the distribution of the magnitude of exceedances does not depend on the value of η cannot be rejected and hence it will be assumed that the $\{\xi_\nu\}_1^\infty$ and η are independent.

The study of dependence of successive exceedances for a selected Q_b is presented later in this chapter.

Under the conditions that $\{\xi_\nu\}_1^\infty$ are independent of η, and $\{\xi_\nu\}_1^\infty$ are mutually independent random variables with the common distribution function H(x), the distribution of the largest exceedance, Eq. 5-3, is

$$F(x) = P(\eta=0) + \sum_{k=1}^{\infty} [\{H(x)\}^k \cdot P(\eta=k)] \qquad (5\text{-}6)$$

The distribution of η, $P(\eta=k)$, used is either the Poisson distribution,

$$P(\eta=k) = \frac{e^{-\lambda}\lambda^k}{k!} \qquad (5\text{-}7)$$

or the three parameter mixed Poisson distribution,

$$P(\eta=k) = \frac{p\, e^{-\lambda_1}\lambda_1^k}{k!} + (1-p)\frac{e^{-\lambda_2}\lambda_2^k}{k!} \qquad (5\text{-}8)$$

with λ = parameter of Poisson, and p, λ_1, and λ_2 = parameters of mixed Poisson distribution, respectively.

The common distribution function of ξ_ν's is either the exponential distribution,

$$H(x) = 1 - \exp(-\frac{x}{\beta}) \qquad (5\text{-}9)$$

or the three-parameter mixed exponential distribution,

$$H(x) = p'[1 - \exp(-\frac{x}{\beta_1})] + (1-p')[1 - \exp(-\frac{x}{\beta_2})] \qquad (5\text{-}10)$$

with β = the parameter of exponential, and p', β_1, and β_2 = parameters of mixed exponential distribution, respectively.

5.6 Statistical Dependence in Partial Flood Peak Series

One of the often-stated drawbacks of the partial flood series is that successive series values are not independent. To test whether the series of the magnitude of exceedances is independent stochastic process, the correlogram of each series of 17 sets of ξ_ν is investigated for independence with 95 percent tolerance limits. Lags from one to one-third of total number of exceedances are checked, for various Q_b. The 95 percent tolerance limits, r_u and r_ℓ, for an independent series are given by

$$r_{u,\ell} = \frac{-1 \pm t\sqrt{N-k-2}}{N-k-1}, \qquad (5\text{-}11)$$

with k = the lag, t = 1.96 = the value of standard normal distribution for a two-tail test that $\rho_k = 0$ for $k > 0$ at the 95 percent level, and N = the sample size. The range of Q_b, the number of Q_b within the range, total number of computed r_k, the number of times and the percent of times that r_k is outside the 95 percent tolerance limits for each series are shown in Table 5-10. If Q_b is selected in such a way that $\bar{\eta}$ varies approximately from 1 to 4, the percent of the total number of r_k of all series outside the 95 percent tolerance limits is 4.37, or less than the expected value of 5.00. Only two out of 17 series have the percent of r_k outside the 95 percent tolerance limits, 7.04 and 8.66, which are more than the expected value of 5.00.

Since the first-order serial correlation coefficient, r_1, is most important in non-periodic series, some further information is provided in Table 5-11. Values r_1 in the range of $\bar{\eta}$ from 1 to 4, are approximately within the 95 percent tolerance limits. The change of $\bar{r}_1$ for all the 17 series, with Q_b are shown in Fig. 5-8 and the upper and lower tolerance limits included. This figure shows the average r_1 of all stations within the 95 percent tolerance limits for the range of $\bar{\eta}$ from 1 to 4.5. When Q_b decreases so that $\bar{\eta}$ is greater than 4.5, $\bar{r}_1$ tends to fall outside the 95 percent tolerance limit. For more details, the relationship between r_1 and Q_b, in such a range of Q_b that $\bar{\eta}$ varies from 1 to 4 or 5, and for each station, is given in the appendix. The relationship between

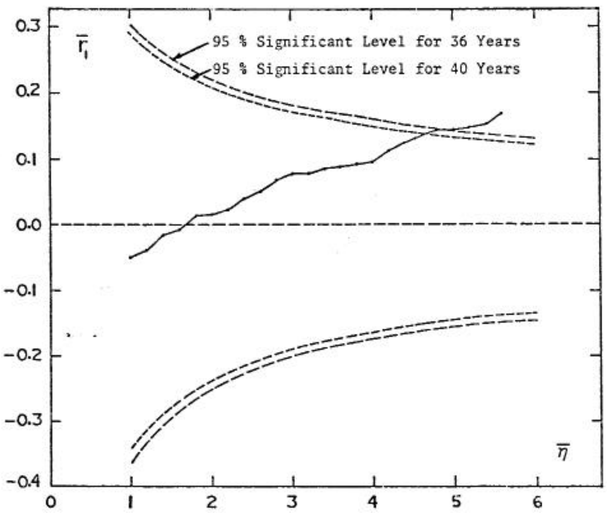

Fig. 5-8 Relationship between the Average First-Order Serial Correlation Coefficient, $\overline{r}_1$ and the Truncation Level, Expressed as the Average Number of Exceedances per Year, $\overline{\eta}$

Table 5-10. Summary of Study of Dependence of Successive Exceedances above the Truncation Levels

Station Number	Range of Q_b Expressed by $\overline{\eta}$	Number of Q_b in the Range	Total Number of Computed r_k	Number of Times r_k is outside 95% T.L.	Percent of Times r_k is outside 95% T.L.
1	1.13-4.10	11	283	14	4.95
2	0.97-4.03	9	255	3	1.18
3	0.97-5.61	11	373	17	4.56
4	0.97-4.58	9	222	11	4.95
5	0.89-4.64	10	258	7	2.71
6	1.00-4.31	11	336	14	4.17
7	1.24-5.49	11	402	11	2.74
8	1.00-4.82	11	355	25	7.04
9	0.97-6.86	10	358	31	8.66
10	0.83-4.94	11	352	17	4.83
11	0.92-3.82	11	303	13	4.29
12	0.78-5.89	10	271	12	4.43
13	0.97-5.97	11	348	16	4.60
14	0.87-6.05	10	326	16	4.91
15	0.64-5.10	10	323	8	2.48
16	0.86-4.27	9	237	12	5.06
17	1.08-5.15	11	420	9	2.10
		176	5402	236	4.37

r_1 and Q_b, for the whole range of Q_b, and for station no. 9 is also given as an example in the appendix. Table 5-12 shows the number and the percent of times that the first 15 values of r_k of all series, with such a range of Q_b that $\overline{\eta}$ varies from 1 to 4.5, are outside the 95 percent tolerance limits. The overall average percent is 4.01.

Table 5-11. Range of Truncation Levels with First Serial Correlation Coefficients either within or outside the 95 Percent Tolerance Limits

Station Number	Range of $\overline{\eta}$ with r_1 Within 95 Percent T.L.	$\overline{\eta}$ with r_1 Outside 95 Percent T.L.
1	1.13-4.10	
2	0.97-4.03	
3	0.89-5.61	
4	0.97-4.30	4.58, 5.47
5	0.89-4.20	4.64, 5.36, 6.14
6	1.00-3.50	4.31
7	1.24-5.89	
8	0.82-4.82	
9	0.97-4.80	5.16, 6.86
10	0.50-2.56;3.30-4.10	2.61, 3.06, 4.31, 4.94
11	0.85-3.82	
12	0.65-4.40	5.89
13	1.17-5.97	
14	0.77-3.80	4.41, 6.05, 9.26, 10.28
15	0.41-4.88	5.10
16	0.86-4.27	
17	0.92-5.15	

Table 5-12. Number and Percent of Times that the First 15 Values of r_k of All Series with such Truncation Levels for $\overline{\eta}$ to Vary Approximately from 1 to 4.5, that are outside 95 Percent Tolerance Limits

Lag Number	Number of Times that the First 15 Values of r_k of All Series are Outside 95 Percent T.L.	Percentage
1	9	5.36
2	1	0.60
3	5	2.98
4	4	2.38
5	5	2.98
6	7	4.17
7	9	5.36
8	3	1.79
9	5	2.98
10	11	6.55
11	10	5.95
12	15	8.93
13	4	2.38
14	10	5.95
15	3	1.79
Total	101	4.01

It can be concluded from the study of correlograms, in such a range of Q_b that $\overline{\eta}$ varies from 1 to 4, that the dependence in the partial series of exceedances is not significant from the point of view of practical applications. If Q_b is lower than this range, the dependence may not be neglected, and it tends to increase with a decrease of Q_b.

5.7 Statistical Dependence of Annual Flood Peak Series

Annual flood peaks are commonly assumed to be a series of independent events. If these flood peaks are not independent, an effect would be the underestimation of the sampling variance of the T-year flood.

The dependence of annual flood peak series is also studied by using their correlograms. Each annual flood series of the 17 stations is tested for significant departure from independence. The number of years of available records varies from 36 to 40 years. The first 15 r_k values of each series are checked for the number and percent of r_k values that are outside the 95 percent tolerance limits for each station, and results are shown in Table 5-13. The number and percent of stations with r_k outside the tolerance limits for each lag k are shown in Table 5-14. The percent of times that an r_k is outside the 95 percent tolerance limits is 4.31, which is less than the expected value of 5.00. The first and second serial correlation coefficients of all stations are within the 95 percent tolerance limits. The lag with the maximum number of stations outside the tolerance limits (3 out of 17) is k = 5. If the first ten lags, instead of the first 15 lags, were considered, the percent of times that r_k are outside the 95 percent tolerance limits is 3.53, which is also less than the expected value of 5.00. It can be concluded, from the data used that the annual flood peak series are approximately independent series.

Table 5-13. Number and Percent of Times that r_k of Annual Flood Series of Each Station are outside 95 Percent Tolerance Limits

Station Number	Number of Lags that r_k Outside 95 Percent T.L.	Lag Numbers for r_k Outside 95 Percent T.L.	Percent
1	0		0.00
2	1	7	6.67
3	1	12	6.67
4	0		0.00
5	2	3, 9	13.33
6	2	5, 12	13.33
7	2	5, 13	13.33
8	0		0.00
9	1	11	6.67
10	0		0.00
11	0		0.00
12	0		0.00
13	0		0.00
14	0		0.00
15	1	5	6.67
16	1	11	6.67
17	0		
	11		4.31

Table 5-14. Number and Percent of Stations for Any Specified Lag k, that r_k are outside 95 Percent Tolerance Limits

Lag Number	Number of Stations that r_k are Outside 95 Percent T.L.	Percent
1	0	0.00
2	0	0.00
3	1	5.88
4	0	0.00
5	3	17.65
6	0	0.00
7	1	5.88
8	0	0.00
9	1	5.88
10	0	0.00
11	2	11.76
12	2	11.76
13	1	5.88
14	0	0.00
15	0	0.00
Total	11	4.31

Chapter VI
GENERATION OF DAILY FLOWS OF TWO CASES FOR TESTING APPLICABILITY OF THE DEVELOPED MODELS

Two series of daily flows, one for the Boise River near Twin Springs, Idaho, and the other for the Powell River near Arthur, Tennessee, are used herein to test the methods developed, and to estimate parameters of daily flow model, as outlined in Chapter IV. Rivers with different characteristics of daily flow series are selected. Patterns of daily flow series vary, depending upon the geographic location and climatic conditions of their river basins. The Boise River has a smooth daily flow series, as well as smooth estimated daily means and daily standard deviations over 365 days, since most runoff comes from snowmelt. The Powell River has rather a highly fluctuating daily flow series, also resulting in highly fluctuating of estimated daily means and daily standard deviations, since most runoff comes from rainfall.

6.1 Generation of Long Daily Flow Samples in Case of Boise River

Selection of Season for Generation. The Boise River daily flow hydrograph indicates significant floods only within the wet season of 5 months, or 150 days, February 28 through July 27, as the season for generating of daily flows. For the truncation level of partial flood series, selected in such a way that the average number of exceedances per year is about 4, only two out of 168, or 1.19 percent of flood exceedances occur outside this season. These two floods are not significant in their magnitude. Hence, the distortion in partial flood series by generating daily flows only within the selected season is not significant.

Test of Lognormal Distribution for Daily Flows. For each day of the selected season, the daily flow sample is of size equal to the number of years of available records, or 40 years in this case. The maximum likelihood estimates of a_τ, $\mu_{y,\tau}$ and $\sigma_{y,\tau}$, of each day τ, are obtained by Eqs. 3-54, 3-55, and 3-56, respectively. By using the lower bound and logarithmic transformation, the transformed variable $Y_{p,\tau}$ is computed by $Y_{p,\tau} = \ln(X_{p,\tau} - \hat{a}_\tau)$, which is then standardized by using estimates $\hat{\mu}_{y,\tau}$ and $\hat{\sigma}_{y,\tau}$. If $X_{p,\tau}$ for each day τ is a three-parameter lognormally distributed variable, the standardized transformed variable will be normally distributed with mean zero and variance unity.

Fits of normal distributions to frequency distributions of standardized transformed variables for each day are tested by chi-square test. Eight class intervals which equal probability are used for this test. Results for days 3,6,9,..., 150, are shown in Table 6-1. The maximum and average chi-square values are 13.608 and 4.974, respectively. For four degrees of freedom, the 95 percent critical value of chi-square is 9.49; only five out of 50, or 10 percent of the computed chi-square values are outside the tolerance limit. Although this percent outside the tolerance limit is greater than the expected value of 5 percent, the chi-square values outside the limit are mostly close to it. Hence, distribution of daily flows of the Boise River are approximately three-parameter lognormal.

Test of Independence of Daily Flow Series. For each day the serial correlation coefficients of daily flow series are computed and checked for departures from independence. The time interval between two successive values of daily flow series in this case is a year. The first ten lags of correlograms are checked for dependence. Results of this test for series of days 3, 6, 9, ..., 150, are given in Table 6-2. The percent of times that r_k is outside the 95 percent tolerance limits is only 3, which is smaller than the expected value of 5 percent. All r_1, shown in Table 6-2, are within the 95 percent tolerance limits. Considering each individual day, the daily flow series is independent lognormal variable. However, the successive days of an entire series are highly serially correlated, as well known for the entire daily flow series.

Table 6-1. Results of Tests for Fits of Lognormal Distribution to Daily Flows of Individual Days

Day Number	Chi-Square	Day Number	Chi-Square	Day Number	Chi-Square
3	1.608	54	6.333	105	2.733
6	2.408	57	6.100	108	5.608
9	5.008	60	6.500	111	4.900
12	13.608*	63	3.133	114	8.008
15	11.933*	66	4.408	117	3.500
18	4.808	69	3.700	120	9.875*
21	4.008	72	2.100	123	1.700
24	3.700	75	9.133	126	4.808
27	7.208	78	3.208	129	2.008
30	10.608*	81	12.075*	132	3.133
33	9.133	84	6.333	135	2.733
36	5.608	87	2.008	138	2.008
39	4.333	90	5.533	141	2.008
42	3.933	93	0.408	144	4.008
45	5.133	96	5.208	147	6.275
48	2.500	99	2.808	150	5.133
51	1.608	102	4.333		

*Chi-square outside the 95% tolerance limit.

Maximum chi-square value = 13.608

Average chi-square value = 4.974

Estimation of Parameters. The parametric method was used to highly reduce the total number of parameters to be estimated. The maximum likelihood estimates of the lower bound a_τ for $\tau = 1,2,\ldots,150$, are shown as curve (1) in Fig. 6-1. The daily means $\mu_{x,\tau}$ and the daily standard deviations $\sigma_{x,\tau}$ of $X_{p,\tau}$ are shown as curves (1) in Figs. 6-2 and 6-3, respectively. Periodicities exist in all of these curves, even though only the wet season is considered. These periodic parameters are fitted by Fourier series harmonics. The number of significant harmonics for each periodic parameter is estimated by using procedure outlined in Section 4.4. Results of the selected number of significant harmonics as well as the Fourier coefficients of a_τ, $\mu_{x,\tau}$, and $\sigma_{x,\tau}$ are given in Table 6-3. The fitted functions for these numbers of significant harmonics of a_τ, $\mu_{x,\tau}$, and $\sigma_{x,\tau}$ are shown as curves (2) in Figs. 6-1, 6-2 and 6-3,

Table 6-2. Results of Tests of Independence of Daily Flow Series of Individual Days

Day Number	r_1	Number of Times r_k is Outside T.L. Limits	Day Number	r_1	Number of Times r_k is Outside T.L. Limits	Day Number	r_1	Number of Times r_k is Outside T.L. Limits
3	-0.1175	1	54	0.0018	0	105	-0.0801	1
6	-0.1760	0	57	-0.1249	0	108	-0.0076	1
9	-0.2065	0	60	0.1313	0	111	-0.1036	1
12	-0.0095	1	63	-0.0736	0	114	-0.0707	1
15	-0.0064	0	66	-0.0943	0	117	-0.0093	0
18	-0.0258	0	69	0.1062	0	120	-0.0505	0
21	-0.0450	0	72	-0.1193	0	123	-0.0411	0
24	-0.0252	0	75	-0.1735	1	126	0.0127	0
27	-0.0593	0	78	0.0515	0	129	0.1493	0
30	-0.2225	0	81	0.1784	0	132	0.0628	0
33	-0.1437	0	84	0.1971	0	135	0.1386	0
36	-0.0423	0	87	-0.0183	0	138	0.1559	0
39	0.0543	1	90	-0.0342	2	141	0.1365	0
42	0.0194	1	93	0.0366	1	144	0.0878	0
45	-0.0793	0	96	0.1274	1	147	0.0837	0
48	-0.1051	0	99	0.1885	1	150	0.0783	0
51	-0.1964	0	102	0.0126	1			

95% significant upper limit for r_1 = 0.2874;

95% significant lower limit for r_1 = -0.3401

Percent of times r_k is outside the T.L. limits =

$$\frac{15}{50 \times 10} \times 100 = 3\%$$

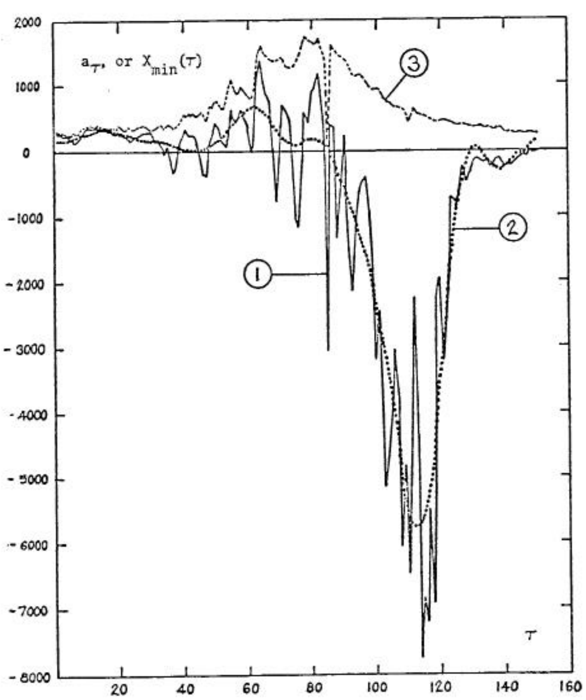

Fig. 6-1 Lower Bound a_τ with: (1) Maximum Likelihood Estimates, (2) Fitted Periodic Function, and (3) Observed Minimum Values $X_{min}(\tau)$ of $X_{p,\tau}$

respectively. The lower bounds in Fig. 6-1 have the minimum and negative values during days from 100 to 120. The value of a_τ depends on the skewness coefficient $\gamma_{x,\tau}$ of $X_{p,\tau}$. In days for which $\gamma_{x,\tau}$ are small, shown as curve (1) in Fig. 6-4, a_τ tends to be negative, and positive for days which have higher values of $\gamma_{x,\tau}$.

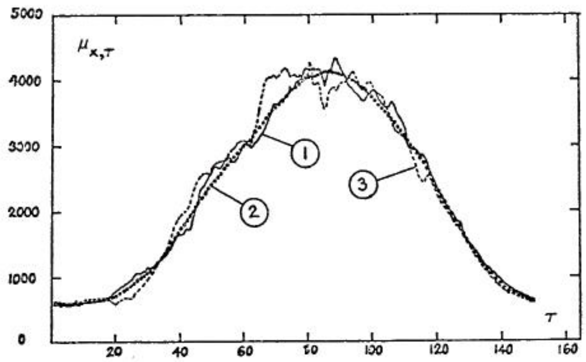

Fig. 6-2 Daily Means, $\mu_{x,\tau}$, with: (1) Estimates from Historic Data, (2) Fitted Periodic Function, and (3) Estimates from Generated Sample

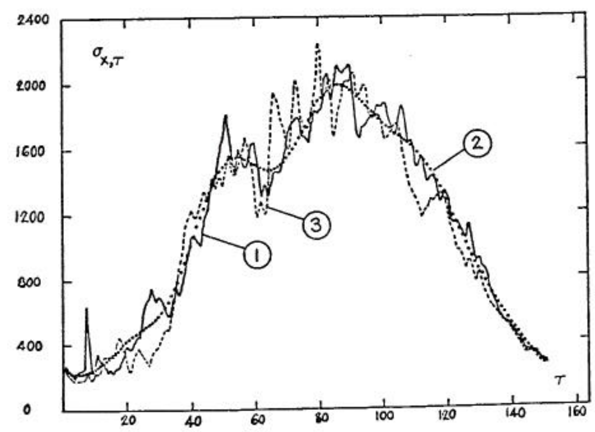

Fig. 6-3 Daily Standard Deviations, $\sigma_{x,\tau}$, with: (1) Estimates from Historic Data, (2) Fitted Periodic Function, and (3) Estimates from Generated Sample

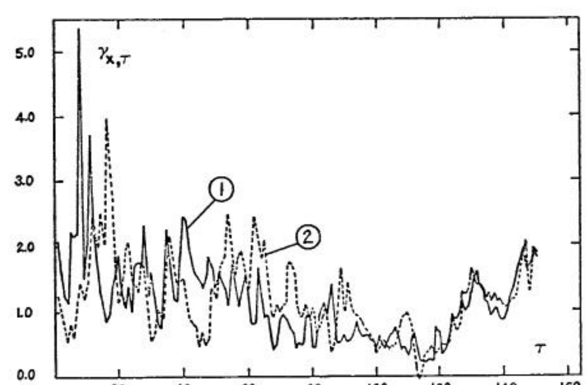

Fig. 6-4 Daily Skewness Coefficients, $\gamma_{x,\tau}$, with: (1) Estimates from Historic Data, and (2) Estimates from Generated Sample

Table 6-3. Number of Significant Harmonics, Explained Variances, and Fourier Coefficients of Periodic Parameters $\mu_{x,\tau}$, $\sigma_{x,\tau}$ and a_τ of $X_{p,\tau}$ Series

Periodic Parameter	Number of Harmonics Used	Explained Variance	Fourier Coefficients								
			Mean	A							
				1	2	3	4	5	6	7	8
$\mu_{x,\tau}$	4	0.988	2243.90	-1617.85	-14.27	12.92	10.42				
$\sigma_{x,\tau}$	5	0.968	1776.71	-756.40	-109.99	42.99	-33.30	-10.48			
a_τ	8	0.830	-671.86	61.66	1276.71	-209.24	-640.97	47.90	205.21	128.99	-52.21
				B							
				1	2	3	4	5	6	7	8
$\mu_{x,\tau}$	4	0.988	2243.90	-707.72	118.10	17.30	49.48				
$\sigma_{x,\tau}$	5	0.968	1776.71	-299.27	-12.99	-24.32	69.71	-56.11			
a_τ	8	0.830	-671.86	1629.77	-208.60	-738.59	86.58	401.10	116.56	-156.79	-48.72

For the estimation of parameters of dependent stochastic component is usually selecting of the order of the autoregressive model, and when necessary the number of significant harmonics of $r_x(k,\tau)$, and $k = 1,2,\ldots,m$, with m = the order of the model. The procedure is not used herein in the study of dependent component of transformed variable $Y_{p,\tau}$, in order to avoid errors resulting from removal of periodicities in a_τ, $\mu_{y,\tau}$ and $\sigma_{y,\tau}$. In this study the Fourier series analysis is used to fit periodicities in $r_x(k,\tau)$. Then values of $r_y(k,\tau)$ are obtained from $r_x(k,\tau)$. by Eq. 4-11. The order of the autoregressive model and the number of significant harmonics of $r_x(k,\tau)$ are selected by comparing the average values of $r_x(k,\tau)$, the shapes of the $r_x(k,\tau)$ curves, and the patterns of the daily flow hydrographs of generated samples with those of the corresponding historic data.

The first-order autoregressive model was not used since it produced generated daily flow samples with more fluctuation of daily flow hydrograph than the historic flows. The shape of the $r_x(k,\tau)$ curves of the generated samples did not sufficiently coincide with those of historic data. Improvements were significant by using the third-order autoregressive model, and it was selected for this study.

The computed $r_x(1,\tau)$, $r_x(2,\tau)$, and $r_x(3,\tau)$ are shown as curves (1) in Figs. 6-5, 6-6 and 6-7, respectively. The Fourier coefficients and the number of significant harmonics for each series of $r_x(k,\tau)$ are given in Table 6-4. The fitted $r_x(k,\tau)$ curves, k = 1, 2, and 3, with the number of significant harmonics of Table 6-4, are shown as curves (2) in Figs. 6-5, 6-6 and 6-7, respectively. However, by using these Fourier coefficients, the average values of $r_x(1,\tau)$, $r_x(2,\tau)$ and $r_x(3,\tau)$ of the generated

Table 6-4. Number of Significant Harmonics, Explained Variances, and Fourier Coefficients of Periodic Parameters $r_x(1,\tau)$, $r_x(2,\tau)$ and $r_x(3,\tau)$ of $X_{p,\tau}$ Series

Periodic Parameter	Number of Harmonics Used	Explained Variance	Fourier Coefficients								
			Mean	A				B			
				1	2	3	4	1	2	3	4
$r_x(1,\tau)$	4	0.252	0.9331 (0.9575)	-0.0058	-0.0084	0.0052	0.0039	-0.0180	-0.0190	-0.0170	-0.0097
$r_x(2,\tau)$	4	0.441	0.8755 (0.8951)	0.0058	-0.0140	0.0127	0.0043	-0.0463	-0.0403	-0.0380	-0.0113
$r_x(3,\tau)$	4	0.542	0.8168 (0.8520)	0.0226	-0.0236	0.0188	0.0026	-0.0704	-0.0594	-0.0565	-0.0141

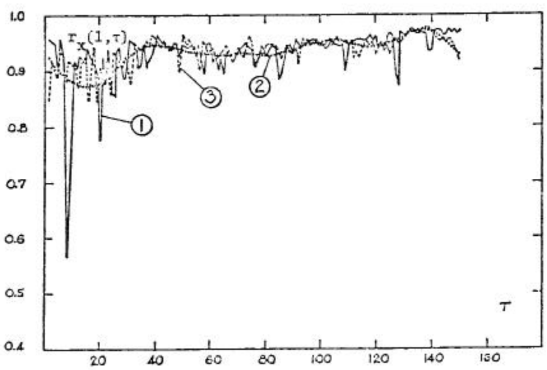

Fig. 6-5 First Serial Correlation Coefficients, $r_x(1,\tau)$, with: (1) Estimates from Historic Data, (2) Fitted Periodic Function, and (3) Estimates from Generated Sample

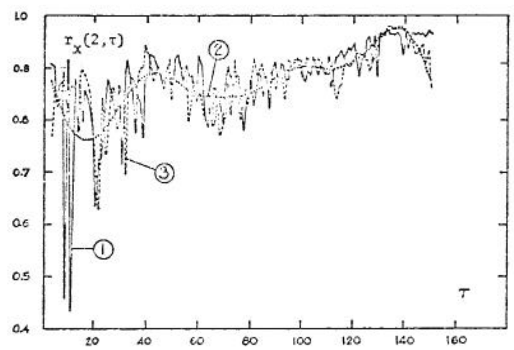

Fig. 6-6 Second Serial Correlation Coefficients, $r_x(2,\tau)$, with: (1) Estimates from Historic Data, (2) Fitted Periodic Function, and (3) Estimates from Generated Sample

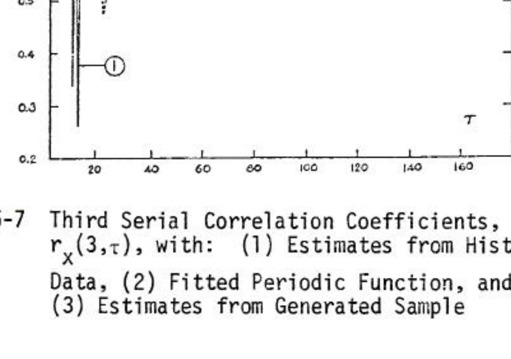

Fig. 6-7 Third Serial Correlation Coefficients, $r_x(3,\tau)$, with: (1) Estimates from Historic Data, (2) Fitted Periodic Function, and (3) Estimates from Generated Sample

sample tended to be smaller than the average values of $r_x(1,\tau)$, $r_x(2,\tau)$ and $r_x(3,\tau)$ of the historic data, respectively. These effects likely result from biases in estimates of $r_x(k,\tau)$, since the sample size is only 40 years. Estimates are adjusted for biases by increasing the average values of $r_x(1,\tau)$, $r_x(2,\tau)$ and $r_x(3,\tau)$ from 0.9331, 0.8755 and 0.8168 to 0.9575, 0.8951 and 0.8320, respectively.

In conclusion, the total number of parameters used in the generation of daily flow samples by the mathematical model of this study depends on the number of significant harmonics used for fitting the series of a_τ, $\mu_{x,\tau}$, $\sigma_{x,\tau}$ and $r_x(k,\tau)$, k = 1,2,...,m, by Fourier analysis. In case of the Boise River, and using the wet season of 150 days, 9 parameters are used for $\mu_{x,\tau}$, 11 parameters for $\sigma_{x,\tau}$, 17 parameters for a_τ, 9 parameters for each of $r_x(k,\tau)$, k = 1, 2, and 3. The total number is 64. In case of the non-parametric method, the total number of parameters would be 900.

Generation of Long Daily Flow Samples. Fifty samples of daily flows, each 40 years long, were generated for wet season of 150 days. The total number of generated years was 2000. The set of 40 year samples is selected for comparison of characteristics of generated daily flows with the corresponding characteristics of historic series of the same sample size. The model for generation should preserve the mean, standard deviation, skewness coefficient, and the first three serial correlation coefficients, or distributions of historic daily flows of each day. The degree of preservation of these properties depends on how well daily flows of each day are fitted by the three-parameter lognormal distribution, and the third-order autoregressive model of dependent stochastic component, as well as how well the model parameters are estimated.

6.2 Comparison of Characteristics of Generated Daily Flows with Corresponding Characteristics of Historic Daily Flows in Case of Boise River

The practical use of a model ultimately depends on its capacity to generate new samples that preserve characteristics of historic series. The main objective of generating new daily flow samples is to study properties of annual and partial flood peak series, but not to check how correctly the model preserves characteristics of historic series. Therefore, the purpose of comparison of characteristics of generated series with corresponding historic series is to ascertain whether generated series preserve in practical terms some characteristics of historic series, at least for purposes of this study.

Comparison Based on Daily Flow Series. The comparison of characteristics of generated series with those of historic series is made in four steps:

(1) Two typical daily flow hydrographs of historic data, considering only the selected wet season, as shown in Figs. 6-8 and 6-9, are visually compared with two typical daily flow hydrographs of generated sample, as shown in Figs. 6-10 and 6-11. Though the generated daily flow hydrographs have somewhat more fluctuating and sharper peaks than those of historic data, general patterns are similar.

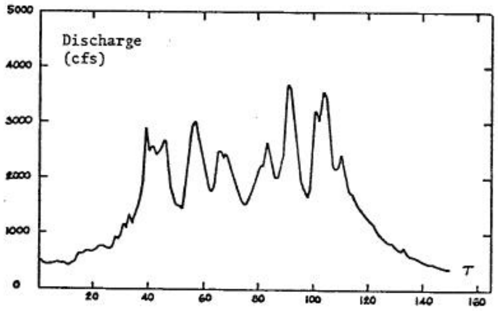

Fig. 6-8 Historic Daily Flow Hydrograph, Year 1930

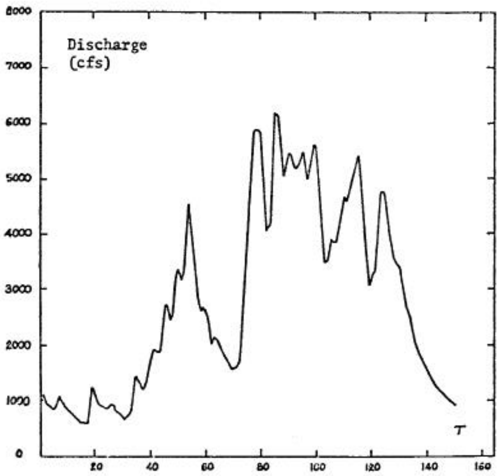

Fig. 6-9 Historic Daily Flow Hydrograph, Year 1950

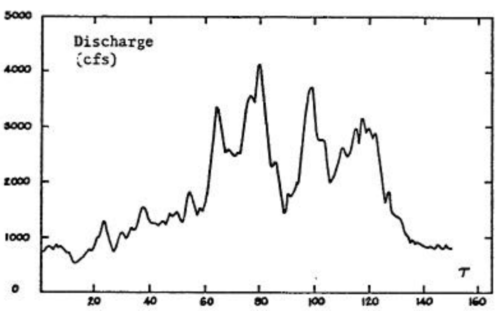

Fig. 6-10 Generated Daily Flow Hydrograph

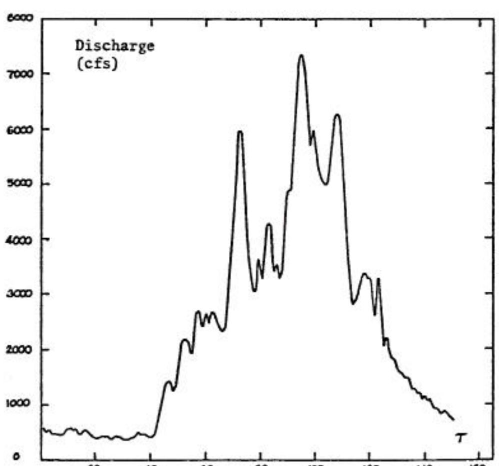

Fig. 6-11 Generated Daily Flow Hydrograph

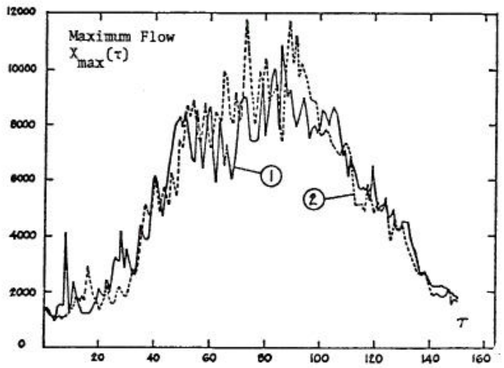

Fig. 6-12 Maximum Flow of 40 Years for Each Day τ, $X_{max}(\tau)$, with: (1) Estimates from Historic Data, and (2) Estimates from Generated Sample

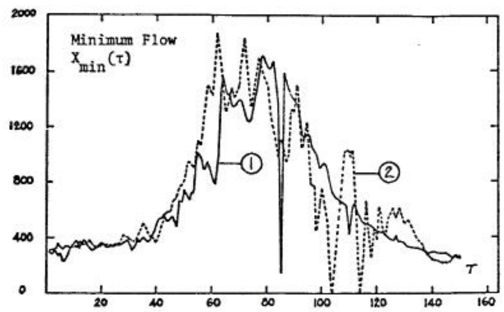

Fig. 6-13 Minimum Flow of 40 Years for Each Day τ, $X_{min}(\tau)$, with: (1) Estimates from Historic Data, and (2) Estimates from Generated Sample

(2) Sequences of $\mu_{x,\tau}$, $\sigma_{x,\tau}$ and $\gamma_{x,\tau}$, for sample no. 5 of generated daily flows are shown as curves (3) in Figs. 6-2 and 6-3, and curve (2) of Fig. 6-4, respectively. Figure 6-12 shows the visual comparison of maximum flow $X_{max}(\tau)$ for each day of historic data with the corresponding maximum flow of generated sample (from sample no. 5), for the same sample size of 40 years. In case of minimum flow $X_{min}(\tau)$ for each day, the visual comparison is shown in Fig. 6-13. These figures display how the model preserves the general patterns of daily flows via $\mu_{x,\tau}$, $\sigma_{x,\tau}$, $\gamma_{x,\tau}$, $X_{max}(\tau)$ and $X_{min}(\tau)$.

(3) Sequences of $r_x(1,\tau)$, $r_x(2,\tau)$ and $r_x(3,\tau)$, estimated from generated sample (sample no. 5), shown as curves (3) in Figs. 6-5, 6-6 and 6-7, respectively, are compared visually with the corresponding estimates of historic data, shown as curves (1) of those figures. By adjusting for biases of estimates for average values of $r_x(1,\tau)$, $r_x(2,\tau)$ and $r_x(3,\tau)$, the model seems to preserve well general characteristics of these periodic parameters of historic data.

(4) The general mean, standard deviation, skewness coefficient, kurtosis coefficient, and the first three serial correlation coefficients of all data for each sample of generated daily flows of sample size of 40 years are given in Table 6-5. Also, the number of adjustments for negative flows, as well as the maximum flow of each sample are included in this table. The average values for the set (of 50 samples), the standard deviation, the maximum value and the minimum value of each statistic are given at the bottom of each column. Estimates of historic data of corresponding

Table 6-5. Statistics of Generated Daily Flows for Each Sample of 40 Years of Records of the Set

Sample Number	Number of Adjustments for Neg. Flows	Moments of Daily Flow Series				Average Serial Correlation Coefficient			Maximum Flow of Sample
		Mean	Std.Dev.	Skewness	Kurtosis	$\bar{r}_x(1,\tau)$	$\bar{r}_x(2,\tau)$	$\bar{r}_x(3,\tau)$	
1	8	2178.52	1727.05	1.480	5.572	0.9339	0.8706	0.8043	11247
2	3	2287.31	1649.66	1.037**	3.781**	0.9237	0.8514	0.7762	11516
3	4	2179.79	1727.58	1.366	5.463	0.9341	0.8756	0.8156	13581
4	0	2367.37	1905.51	1.539	6.552	0.9348	0.8751	0.8131	16636
5	2	2251.80	1793.26	1.292	4.744	0.9342	0.8772	0.8168	11754
6	10	2115.60	1629.04	1.342	5.101	0.9339	0.8764	0.8151	11716
7	5	2307.31	1919.85	1.480	5.729	0.9347	0.8777	0.8214	13479
8	3	2362.96	1884.12	1.592	6.647	0.9333	0.8731	0.8127	15387
9	2	2138.79	1622.67	1.230	4.710	0.9336	0.8733	0.8130	12566
10	4	2258.60	1783.81	1.337	5.257	0.9323	0.8683	0.8022	14562
11	4	2224.57	1881.01	1.971*	9.883*	0.9312	0.8666	0.7995	18143
12	4	2336.44	1840.73	1.348	5.258	0.9356	0.8801	0.8260	14101
13	2	2257.66	1742.40	1.149	4.104	0.9298	0.8636	0.7975	10688
14	2	2279.79	1745.47	1.268	5.077	0.9281	0.8637	0.8000	12461
15	5	2338.21	1916.98	1.498	5.684	0.9387*	0.8874*	0.8347	12236
16	2	2137.89	1624.64	1.334	5.008	0.9312	0.8691	0.8065	10813
17	0	2269.52	1671.07	1.103	4.173	0.9314	0.8690	0.8032	11318
18	0	2193.49	1637.20	1.197	4.591	0.9257	0.8529	0.7786	11364
19	4	2207.11	1828.87	1.795	8.695	0.9301	0.8684	0.8025	19691*
20	1	2442.61	1894.68	1.338	5.215	0.9334	0.8756	0.8176	13333
21	2	2230.74	1662.27	1.133	4.332	0.9324	0.8689	0.8046	11507
22	0	2297.40	1722.52	1.126	4.010	0.9365	0.8806	0.8224	11586
23	2	2178.30	1759.41	1.489	5.987	0.9308	0.8672	0.8055	12792
24	2	2274.16	1825.92	1.307	4.726	0.9307	0.8673	0.8014	11676
25	2	2336.67	1826.29	1.246	4.879	0.9360	0.8804	0.8261	15272
26	2	2251.94	1816.49	1.533	6.733	0.9376	0.8874*	0.8374*	16072
27	0	2375.05	1910.94	1.438	5.905	0.9376	0.8820	0.8261	14675
28	4	2279.57	1842.84	1.326	4.889	0.9338	0.8748	0.8137	12572
29	5	2266.16	1898.62	1.863	8.610	0.9316	0.8647	0.7981	16592
30	6	2367.16	1934.81	1.464	6.499	0.9303	0.8655	0.8020	18477
31	6	2186.04	1753.79	1.414	5.353	0.9323	0.8711	0.8102	12352
32	0	2319.53	1821.18	1.396	5.453	0.9313	0.8694	0.8080	13777
33	0	2128.38	1548.12**	1.187	4.685	0.9255	0.8535	0.7833	10967
34	3	2222.85	1737.83	1.327	5.303	0.9357	0.8786	0.8239	13731
35	0	2247.05	1724.81	1.211	4.268	0.9314	0.8673	0.8002	10837
36	5	2308.03	1952.03*	1.702	7.453	0.9376	0.8851	0.8327	15474
37	1	2267.32	1787.38	1.381	5.158	0.9385	0.8862	0.8332	11160
38	10	2387.74	1817.61	1.196	4.697	0.9336	0.8720	0.8102	13549
39	0	2315.03	1793.03	1.243	4.549	0.9282	0.8647	0.7993	11857
40	4	2198.08	1838.89	1.479	5.458	0.9339	0.8754	0.8198	12072
41	3	2203.97	1823.04	1.764	8.873	0.9335	0.8751	0.8134	18307
42	4	2226.80	1689.09	1.195	4.414	0.9271	0.8572	0.7872	11792
43	5	2185.28	1713.33	1.464	6.017	0.9272	0.8592	0.7936	13098
44	2	2263.62	1702.60	1.147	3.988	0.9279	0.8591	0.7912	10027
45	3	2251.12	1715.06	1.258	4.854	0.9233**	0.8470**	0.7741**	12620
46	8	2111.53**	1618.45	1.154	4.088	0.9304	0.8700	0.8104	9368**
47	1	2245.22	1719.48	1.314	5.198	0.9378	0.8850	0.8296	12055
48	1	2162.97	1653.65	1.240	4.582	0.9343	0.8714	0.8082	10914
49	0	2233.23	1663.94	1.441	6.026	0.9306	0.8680	0.8021	13134
50	0	2462.84*	1896.03	1.317	5.538	0.9289	0.8625	0.7958	15534
Average	2.92	2258.38	1771.90	1.369	5.475	0.9321	0.8706	0.8084	13209
St.Dev.	2.53	81.08	100.08	0.214	1.316	0.135	0.125	0.011	2351.03
Maximum	10	2462.84	1952.03	1.971	9.883	0.9387	0.8874	0.8374	19691
Minimum	0	2111.53	1548.12	1.037	3.781	0.9233	0.8470	0.7741	9368
Historic Data		2243.90	1776.71	1.219	4.156	0.9331	0.8755	0.8168	10800

*maximum, **minimum

statistics are given at the last line of each column for comparison purposes.

The average values of all samples for the mean, the standard deviation and the first three serial correlation coefficients of generated data agree well with the corresponding estimates from historic data. The average values of all samples for skewness and kurtosis coefficients are somewhat greater than for historic data. This may result from large extreme values since samples of very high values of skewness and kurtosis coefficients are also samples with large extreme values generated. These large extreme values may also be responsible for the average value of maximum flows of 40 year samples to be greater than those of historic data.

The effects of negative flows on the above statistics of generated daily flows depend on the number of adjustments for negative flows. On the average for all samples of the set, only 2.92 values out of 6000 values, or 0.049 percent per sample of 40 years of daily flows, were adjusted for negative values, so effects of adjustment can be considered as small and neglected as such.

Comparison Based on Annual Flood Series. The most important test of applicability of the model for this study is to find out whether it preserves the extreme values, expecially the annual flood series. For each 40-year sample of generated set of daily flows, the annual flood series is computed. Moments and statistics for each annual flood series are estimated and results given in Table 6-6. For all samples the values of the mean, standard deviation and maximum and minimum value of each statistic are given at the bottom of each column. Corresponding statistics for historic data are also given at the last line of each column. The model seems to preserve some statistics of historic annual flood series. The average values of all samples for skewness and kurtosis coefficients are somewhat greater than those of historic series, likely as effects of large extremes generated in some samples. Some samples, however, have skewness and kurtosis coefficients smaller than those of historic annual flood series. By comparison, the differences between the average values of all generated samples and the historic values of both skewness and kurtosis in case of the daily flow series are smaller than those for the annual flood series, the effect may be due to the small sample sizes (40 years) of the annual flood series since the reliability of estimating skewness and kurtosis depends on the available sample sizes.

Comparison based on statistics of daily flow series and their annual flood peak series of generated samples of daily flows show similar characteristics to those of historic data. Therefore, generated samples produce extreme values which can be used for the study of flood peaks.

6.3 Generation of Long Daily Flow Samples in Case of Powell River

Selection of Season for Generation. The Powell River daily flow hydrograph indicates significant floods only within the wet season of 9 months, or 270 days, October 31 through July 27, as the season for generating daily flows. For the truncation level of partial flood series, selected in such a way that the average number of exceedances per year is about 4, only 0.67 percent of flood exceedances occur outside this season.

Table 6-6. Statistics of Annual Flood Peak Series of Generated Samples, Each Sample of 40 Years of Daily Flows

Sample Number	Statistics of Annual Flood Series: Mean	Std.Dev.	Skewness	Kurtosis	Minimum	Maximum
1	6848.18	2174.40	0.488	2.339	3595	11247
2	6556.85	1581.24**	0.906	5.184	3445	11516
3	6926.13	2229.61	1.315	5.324	3583	13581
4	7239.49	2669.48	1.709	6.926	4000	16636
5	6999.31	2133.42	0.326	3.041	3146	11754
6	6350.59	1860.02	0.508	4.111	2691	11716
7	7448.42	2454.61	0.609	3.042	3693	13479
8	7380.16	2912.26	1.013	3.824	3802	15387
9	6459.19	1863.06	1.109	4.937	3567	12566
10	6971.75	2233.47	1.328	5.673	3616	14562
11	7436.62	3288.75*	1.717	6.360	2807	18143
12	7038.59	2281.45	1.186	5.272	3529	14101
13	6769.10	1798.28	0.512	2.534	4049	10688
14	6781.57	2164.44	0.823	3.452	3518	12461
15	6749.82	2526.85	0.813	3.228	2532	12236
16	6547.02	1705.98	0.852	3.105	4191	10813
17	6498.82	1744.90	0.462	3.746	3057	11318
18	6382.24	1726.99	0.737	4.216	3160	11364
19	7525.66	3020.49	1.852	9.051	3449	19691*
20	7417.66	2334.94	0.630	2.852	4273*	13333
21	6750.09	1804.98	0.590	3.375	3577	11507
22	6683.13	1792.18	0.083	3.567	2784	11586
23	6418.03	2170.21	0.973	5.068	2874	12792
24	7003.62	2111.88	0.431	2.561	4198	11676
25	7000.94	2137.73	1.461	7.672	3667	15272
26	6938.35	2534.02	1.508	6.430	3585	16072
27	6997.03	2498.91	1.406	5.211	3921	14675
28	7081.05	2071.19	0.664	3.725	3664	12572
29	7061.46	2863.17	1.235	5.474	2931	16592
30	7573.07*	2557.22	1.902	10.927*	3841	18477
31	6872.05	1993.84	0.894	3.998	3429	12352
32	7072.93	2191.16	1.077	4.959	3727	13777
33	6308.03	1759.14	0.848	3.953	3765	10967
34	6559.45	2373.36	0.736	4.162	2490**	13731
35	7022.75	1894.02	0.158	2.618	3750	10837
36	7547.29	2882.07	1.175	4.305	3970	15474
37	6966.69	2218.90	0.308	2.374	3385	11160
38	7263.88	2092.31	0.680	4.276	3637	13549
39	6899.95	1940.68	0.467	3.235	3487	11857
40	6662.41	2312.09	0.557	3.049	2826	12072
41	6799.08	2818.13	2.097*	9.992	3085	18307
42	6815.37	1967.64	0.520	3.245	3359	11792
43	6720.46	2455.28	0.950	4.122	2931	13098
44	6554.35	1783.30	0.140	2.254	3703	10027
45	6774.47	1877.58	0.855	4.241	3400	12620
46	6213.67**	1755.69	0.015**	2.179**	3117	9368**
47	6788.39	1927.83	0.792	3.876	3728	12055
48	6536.07	2035.04	0.283	2.587	3019	10914
49	6632.15	2170.71	1.198	4.357	4018	13134
50	7570.54	2468.23	1.400	5.672	4113	15534
Average	6888.58	2203.26	0.886	4.433	3474	13209
Std.Dev.	349.15	390.32	0.498	1.918	446.59	2351.03
Maximum	7573.07	3288.75	2.097	10.927	4273	19691
Minimum	6213.67	1581.24	0.015	2.179	2490	9368
Historic Sample	6430.00	1989.01	0.0531	2.184	2870	10800

*maximum, **minimum

Test of Lognormal Distribution for Daily Flows. Fits of normal distributions to frequency distributions of standardized transformed variables for each day are tested by chi-square test with 8 class intervals of equal probability. Results for days 1, 6, 11, ..., 261, are shown in Table 6-7. The maximum and average chi-square values are 13.808 and 5.456, respectively. For 4 degrees of freedom, the 95 percent critical value of chi-square is 9.49; only 6 out of 53, or 11.32 percent of the computed chi-square values are outside the tolerance limit.

Estimation of Parameters. The maximum likelihood estimates of a_τ for $\tau = 1,2,\ldots,270$, are shown as curve (1) in Fig. 6-14. The daily means $\mu_{x,\tau}$ and daily standard deviations $\sigma_{x,\tau}$ of $X_{p,\tau}$ are shown as curves (1) in Figs. 6-15 and 6-16, respectively. Periodicities exist in all of these curves with more fluctuating around the periodicities than for the Boise River. Results of the selected number of significant harmonics and the Fourier coefficients of a_τ, $\mu_{x,\tau}$ and $\sigma_{x,\tau}$ are given in Table 6-8. The fitted

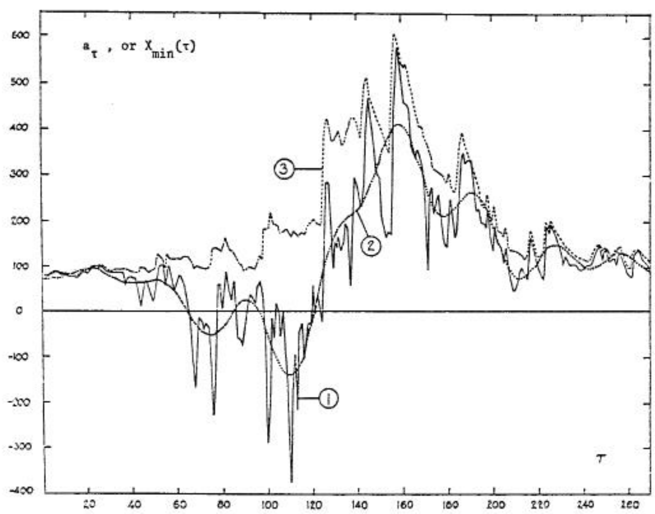

Fig. 6-14. Lower Bound a with: (1) Maximum Likelihood Estimates, (2) Fitted Periodic Function, and (3) Observed Minimum Values $X_{min}(\tau)$ of $X_{p,\tau}$

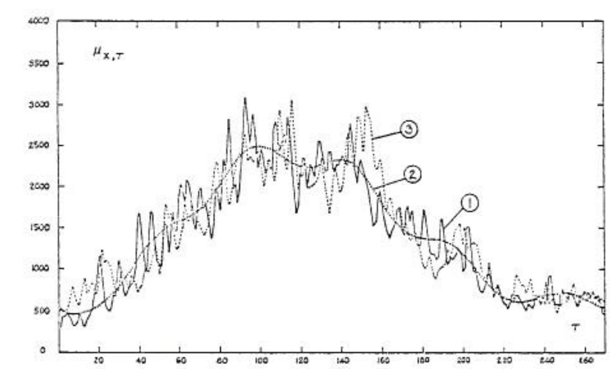

Fig. 6-15 Daily Means, $\mu_{x,\tau}$, with: (1) Estimates from Historic Data, (2) Fitted Periodic Function, and (3) Estimates from Generated Sample

Table 6-7. Results of Tests for Fits of Lognormal Distributions to Daily Flows of Individual Days

Day Number	Chi-Square	Day Number	Chi-Square	Day Number	Chi-Square
1	7.208	86	13.808*	176	7.208
6	2.733	91	6.808	181	4.100
11	6.275	96	11.408*	186	6.008
16	3.533	101	3.933	191	11.608*
21	5.875	106	4.100	196	6.500
26	5.933	111	2.275	201	4.208
31	2.900	116	11.408*	206	3.300
36	3.300	121	7.300	211	1.533
41	1.608	126	8.900	216	3.300
46	4.500	131	6.900	221	4.008
51	5.208	136	3.300	226	2.333
56	2.333	141	0.808	231	7.700
61	3.208	146	10.100*	236	3.133
66	5.533	151	3.408	241	5.300
71	7.075	156	2.333	246	4.333
76	8.275	161	6.100	251	8.733
81	4.808	166	1.608	256	13.533*
		171	5.133	261	2.408

functions for those numbers of selected harmonics of a_τ, $\mu_{x,\tau}$ and $\sigma_{x,\tau}$ are shown as curves (2) in Figs. 6-14, 6-15 and 6-16, respectively. Number of days having negative a_τ in Fig. 6-14 is less than that of the Boise River since, on the average, the values of skewness coefficients of daily flows of the Powell River, shown as curve (1) in Fig. 6-17, are greater, The fitted function of a_τ using the selected harmonics results in some days, the fitted a_τ are greater than the observed minimum values of $X_{p,\tau}$ for the same days. Improvements were not significant by increasing the number of significant harmonics.

The third-order autoregressive model was selected to represent the dependence of stochastic component of daily flow series. The estimates of $r_x(1,\tau)$, $r_x(2,\tau)$ and $r_x(3,\tau)$ are shown as curves (1) in Figs. 6-18, 6-19 and 6-20, respectively. The Fourier coefficients and the selected number of significant harmonics for each series of $r_x(k,\tau)$ are given in Table 6-9. The fitted functions of $r_x(k,\tau)$, for k = 1, 2 and 3 are shown as curves (2) in Figs. 6-18, 6-19 and 6-20, respectively. By using these Fourier coefficients, the average value of $r_x(1,\tau)$ of the generated sample was smaller than the corresponding average value of $r_x(1,\tau)$ of the historic data, while the average values of $r_x(2,\tau)$ and $r_x(3,\tau)$ of the generated sample were larger. Estimates are adjusted for biases by increasing the average value of $r_x(1,\tau)$ from 0.8305 to 0.8480, and by decreasing the average values of $r_x(2,\tau)$ and $r_x(3,\tau)$ from 0.6133 and 0.4723 to 0.6100 and 0.4700, respectively.

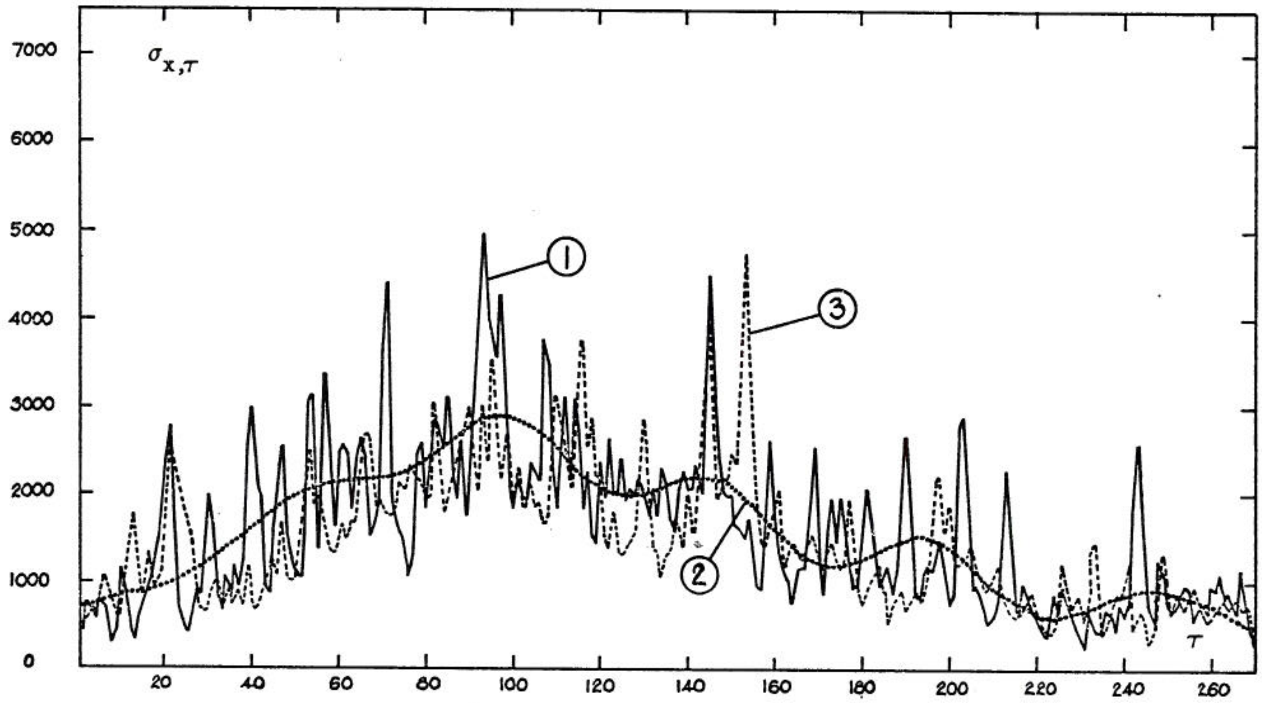

Fig. 6-16 Daily Standard Deviations, $\sigma_{x,\tau}$, with: (1) Estimates from Historic Data, (2) Fitted Periodic Function, and (3) Estimates from Generated Sample

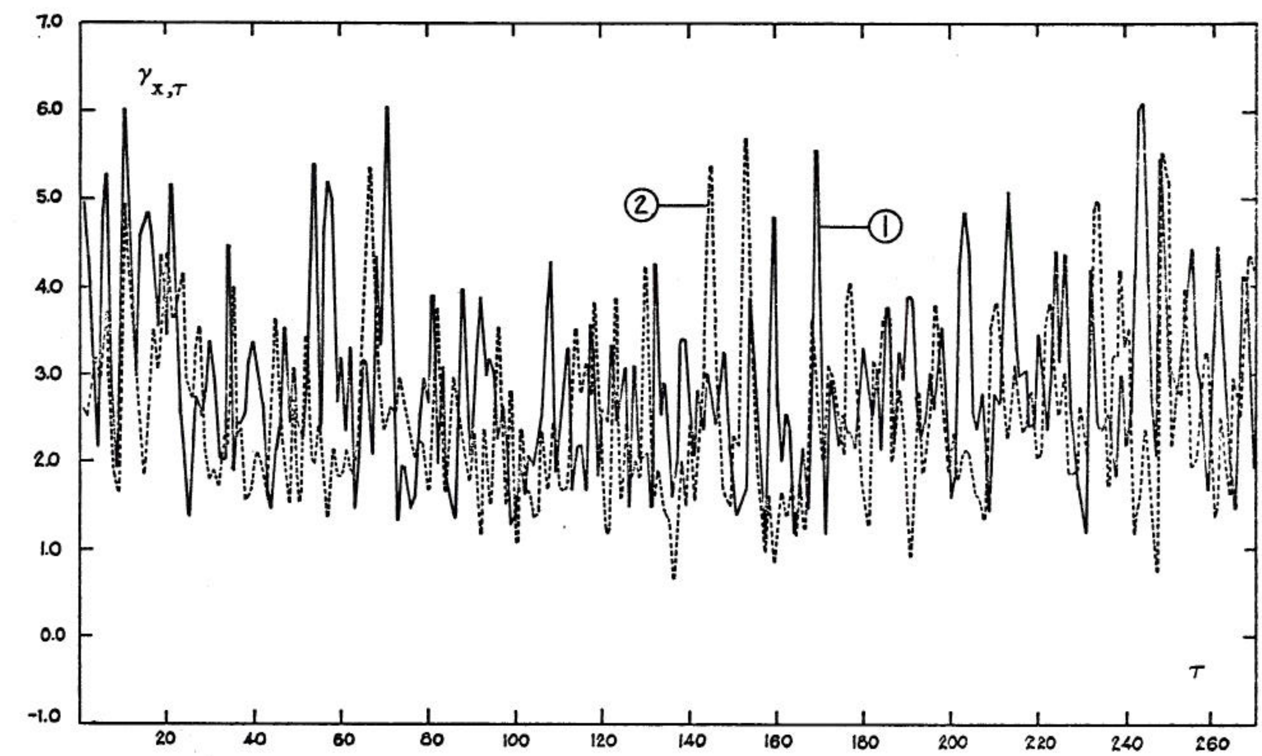

Fig. 6-17 Daily Skewness Coefficients, $\gamma_{x,\tau}$, with: (1) Estimates from Historic Data, and (2) Estimates from Generated Sample

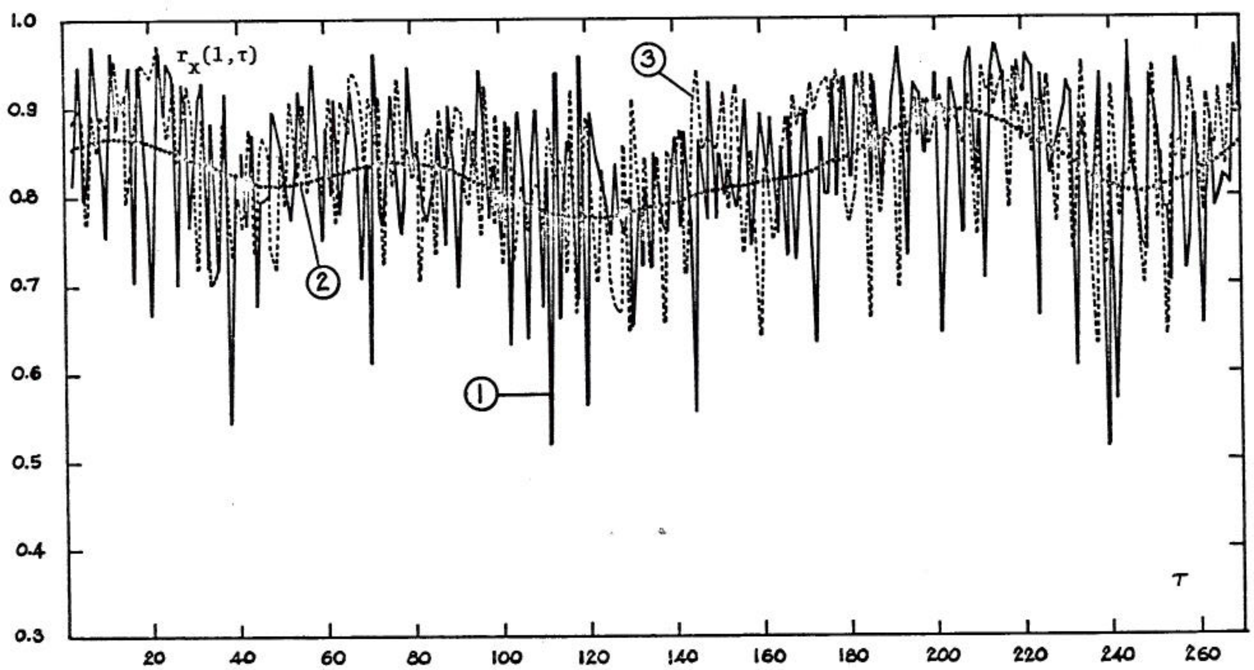

Fig. 6-18 First Serial Correlation Coefficients, $r_x(1,\tau)$, with: (1) Estimates from Historic Data, (2) Fitted Periodic Function, and (3) Estimates from Generated Sample

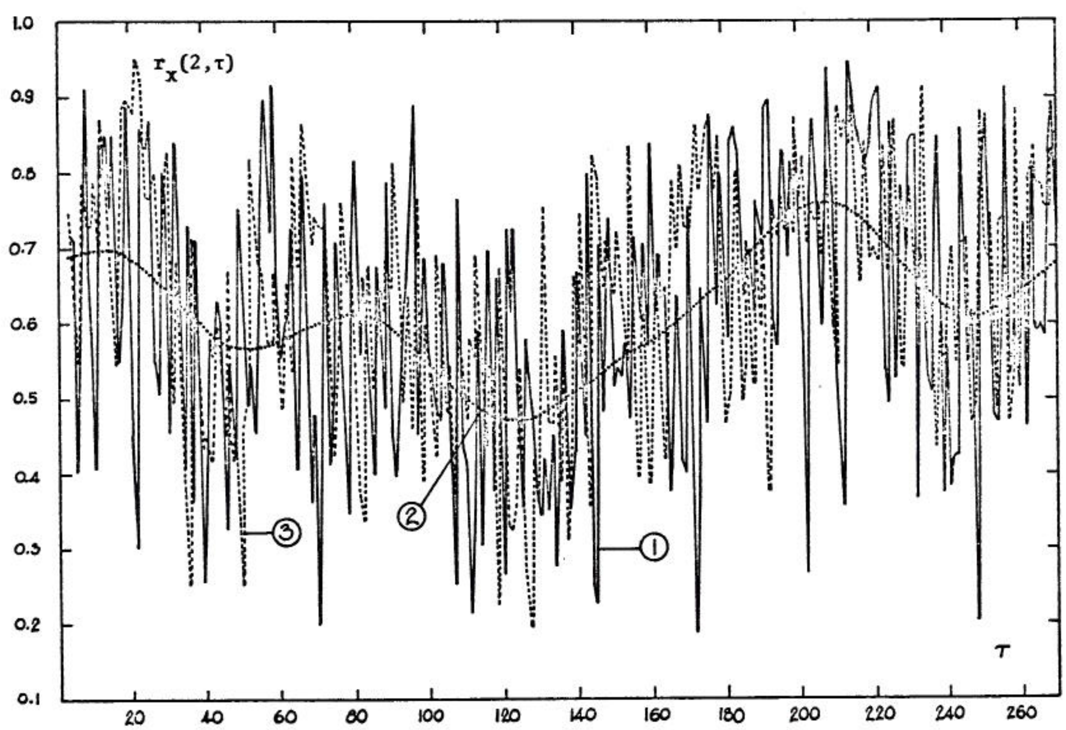

Fig. 6-19 Second Serial Correlation Coefficients, $r_x(2,\tau)$, with: (1) Estimates from Historic Data, (2) Fitted Periodic Function, and (3) Estimates from Generated Sample

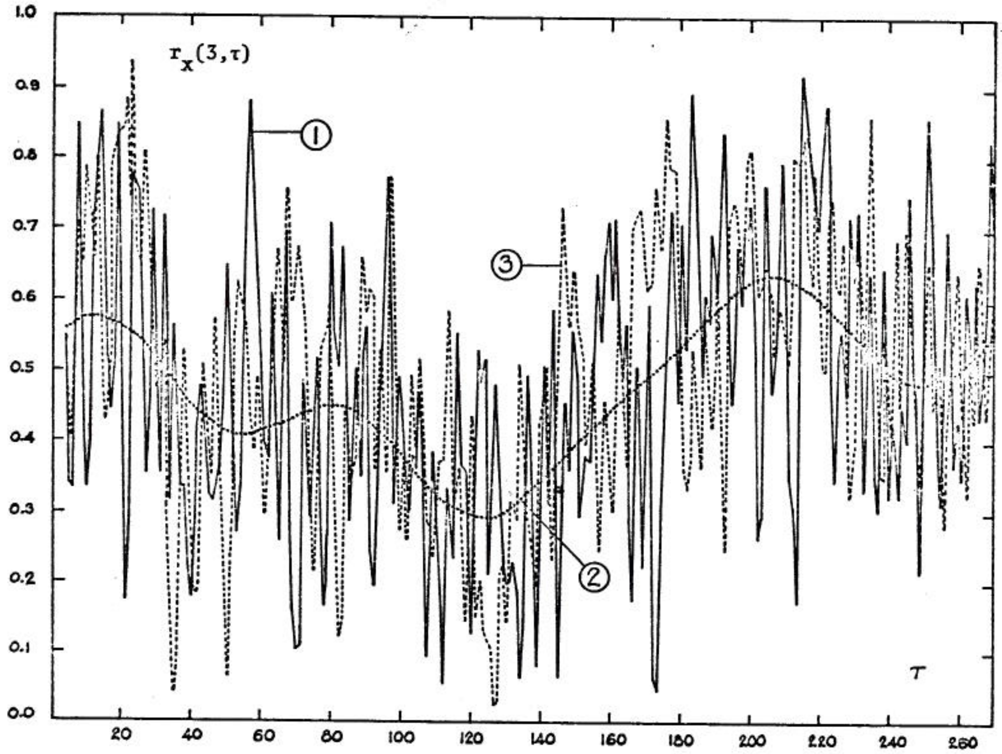

Fig. 6-20 Third Serial Correlation Coefficients, $r_x(3,\tau)$, with: (1) Estimates from Historic Data, (2) Fitted Periodic Function, and (3) Estimates from Generated Sample

Table 6-8. Number of Significant Harmonics, Explained Variances, and Fourier Coefficients of Periodic Parameters $\mu_{x,\tau}$, $\sigma_{x,\tau}$ and a_τ of $X_{p,\tau}$

Periodic Parameter	Number of Harmonics Used	Explained Variance	Fourier Coefficients									
			Mean	A								
				1	2	3	4	5	6	7	8	9
$\mu_{x,\tau}$	6	0.908	1400.14	-875.05	0.834	3.184	0.851	-27.36	15.003			
$\sigma_{x,\tau}$	6	0.568	1569.03	-694.28	-145.14	42.044	-1.867	-52.128	-16.619			
a_τ	9	0.792	113.16	-35.013	30.646	-9.268	-1.072	-12.652	11.199	1.929	-8.531	-6.325
				B								
				1	2	3	4	5	6	7	8	9
$\mu_{x,\tau}$	6	0.908	1400.14	373.52	-114.81	-38.262	-34.287	-113.38	58.944			
$\sigma_{x,\tau}$	6	0.568	1569.03	588.37	-135.26	-17.165	-32.354	-172.204	141.935			
a_τ	9	0.792	113.16	-111.228	90.905	-51.926	20.883	-20.098	10.309	13.744	-38.561	11.724

Table 6-9. Number of Significant Harmonics, Explained Variances, and Fourier Coefficients of Periodic Parameters $r_x(1,\tau)$, $r_x(2,\tau)$ and $r_x(3,\tau)$ of $X_{p,\tau}$ Series

Periodic Parameter	Number of Harmonics Used	Explained Variance	Fourier Coefficients								
			Mean	A				B			
				1	2	3	4	1	2	3	4
$r_x(1,\tau)$	4	0.1107	0.8305 (0.8480)	0.0184	-0.0218	0.0105	0.0113	-0.0181	0.0097	0.0128	0.0143
$r_x(2,\tau)$	4	0.1927	0.6133 (0.6100)	0.0633	-0.0444	0.0274	0.0172	-0.0571	0.0163	0.0214	0.0272
$r_x(3,\tau)$	4	0.2374	0.4723 (0.4700)	0.0794	-0.0462	0.0342	0.0149	-0.0783	0.0238	0.0210	0.0321

In conclusion, the total number of parameters for the daily flow model in case of the Powell River is 72, with 13 parameters used for $\mu_{x,\tau}$, 13 parameters for $\sigma_{x,\tau}$, 19 parameters for a_τ, and 9 parameters for each of $r_x(k,\tau)$, k = 1, 2, and 3.

Generation of Long Daily Flow Samples. Similar to the case of the Boise River, 50 samples of daily flows, each 40 years long, were generated for wet season of 270 days. The total number of generated years was 2000.

6.4 Comparison of Characteristics of Generated Daily Flows with Corresponding Characteristics of Historic Daily Flows in Case of Powell River

Comparison Based on Daily Flow Series.

(1) A typical daily flow hydrograph of historic data, considering only the selected wet season as shown in Fig. 6-21, is visually compared with a typical daily flow hydrograph of generated sample, as shown in Fig. 6-22.

(2) Sequences of $\mu_{x,\tau}$, $\sigma_{x,\tau}$ and $\gamma_{x,\tau}$, for sample no. 1 of generated daily flows are shown as curves (3) in Figs. 6-15 and 6-16, and as curve (2) in Fig. 6-17, respectively. Figures 6-23 and 6-24 show the visual comparison of maximum flow, $X_{max}(\tau)$, and minimum flow, $X_{min}(\tau)$, for each day of historic data with the corresponding maximum flow and minimum flow of generated sample (sample no. 1), for the same sample size of 40 years, respectively.

(3) Sequences of $r_x(1,\tau)$, $r_x(2,\tau)$ and $r_x(3,\tau)$ estimated from generated sample (sample no. 1), shown as curves (3) in Figs. 6-18, 6-19 and 6-20, respectively, are visually compared with the corresponding estimates from historic data, shown as curves (1) of those figures.

(4) The general mean, standard deviation, skewness, kurtosis, the first three serial correlation coefficients, the number of adjustments for negative flows, and the maximum flow, of all data for each sample of generated daily flows, 40 years long, are given in Table 6-10. The average values for the set (of 50 samples), the standard deviation, the maximum and minimum value as well as the historic value for each statistic are given at the bottom of each column.

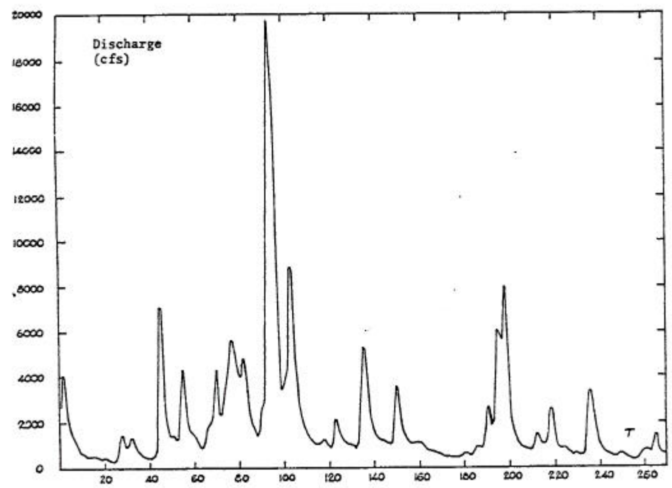

Fig. 6-21 Historic Daily Flow Hydrograph, Year 1950

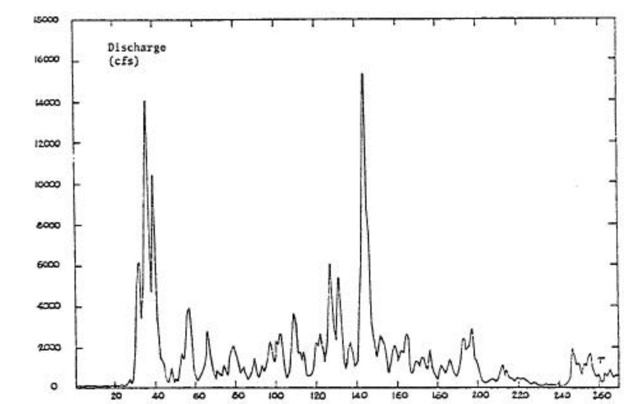

Fig. 6-22 Generated Daily Flow Hydrograph

Table 6-10. Statistics of Generated Daily Flows for Each Sample of 40 Years of Records of the Set

Sample Number	Number of Adjustments for Neg. Flows	Moments of Daily Flow Series				Average Serial Correlation Coefficient			Maximum Flow of Sample
		Mean	Std.Dev.	Skewness	Kurtosis	$\overline{r_x}(1,\tau)$	$\overline{r_x}(2,\tau)$	$\overline{r_x}(3,\tau)$	
1	0	1402.95	1703.35	3.773	29.159	0.8430	0.6324	0.5000	30213
2	0	1368.35	1705.23	5.781	98.436	0.8283	0.6009	0.4602	52902
3	2	1339.10	1597.78	3.707	25.067	0.8331	0.6089	0.4723	20813
4	0	1467.59	1960.17	4.964	49.147	0.8414	0.6296	0.4964	35041
5	1	1409.91	1692.12	3.966	32.248	0.8381	0.6244	0.4984	28414
6	0	1334.59	1599.77	5.095	60.767	0.8389	0.6286	0.5043	36237
7	0	1411.76	1800.29	5.289	56.255	0.8339	0.6142	0.4849	36352
8	0	1470.14	2042.24	5.078	50.306	0.8429	0.6346	0.5078	42139
9	0	1333.29	1621.36	3.849	27.456	0.8365	0.6182	0.4848	23199
10	1	1438.59	1731.86	3.676	24.776	0.8332	0.6111	0.4755	23736
11	0	1383.19	1921.81	6.706	88.806	0.8431	0.6291	0.4883	46005
12	0	1490.37	1882.01	4.056	32.565	0.8364	0.6169	0.4861	35075
13	0	1368.06	1741.50	5.507	61.308	0.8252	0.5903	0.4502	33254
14	0	1362.99	1748.98	4.128	30.776	0.8397	0.6293	0.5016	27186
15	1	1468.80	1910.88	5.139	54.974	0.8455	0.6414	0.5168	41596
16	0	1323.79	1650.28	4.269	36.908	0.8372	0.6248	0.4997	31551
17	0	1467.32	1774.66	3.891	30.849	0.8324	0.6050	0.4678	30300
18	0	1339.02	1808.08	7.372	122.990	0.8261**	0.5970**	0.4629**	49998
19	0	1535.36*	1969.89	4.387	37.156	0.8392	0.6236	0.4928	32127
20	1	1467.53	2000.51	4.830	45.733	0.8355	0.6226	0.4954	40418
21	0	1428.87	1853.75	4.491	35.198	0.8437	0.6395	0.5126	25273
22	1	1418.27	1725.86	3.466	22.442	0.8362	0.6154	0.4814	24412
23	0	1444.39	2003.10	5.557	57.345	0.8333	0.6163	0.4899	38498
24	0	1412.48	1678.14	3.218	19.161	0.8316	0.6054	0.4707	23397
25	0	1414.09	1983.64	6.999	117.991	0.8310	0.6245	0.4942	61041
26	0	1458.40	1846.95	4.192	31.041	0.8488*	0.6495*	0.5314*	27598
27	0	1343.19	1630.04	4.013	31.127	0.8355	0.6192	0.4915	27315
28	2	1397.54	1788.07	4.679	48.809	0.8451	0.6363	0.5060	40768
29	0	1443.79	1885.05	4.513	40.777	0.8299	0.6060	0.4781	32640
30	0	1401.62	•1733.18	3.602	22.761	0.8398	0.6275	0.4967	21590
31	1	1418.95	1831.66	4.210	32.928	0.8349	0.6149	0.4814	30847
32	0	1352.76	1695.75	5.472	61.127	0.8264	0.5976	0.4638	34386
33	1	1403.86	1768.50	3.739	24.062	0.8368	0.6172	0.4813	20396
34	2	1407.98	1750.91	4.362	39.170	0.8320	0.6079	0.4736	29143
35	0	1431.31	1907.13	4.574	35.616	0.8357	0.6147	0.4777	27244
36	0	1412.87	1884.84	4.681	41.634	0.8368	0.6239	0.5061	35227
37	2	1517.4	2045.33*	6.589	110.18	0.8382	0.6276	0.4976	61091
38	1	1382.90	1659.51	4.026	32.026	0.8311	0.6032	0.4663	28657
39	0	1381.68	1853.63	4.529	33.983	0.8414	0.6298	0.5021	24398
40	0	1394.23	1829.48	4.799	46.179	0.8422	0.6353	0.5072	33176
41	1	1354.83	1644.73	3.411	20.538	0.8360	0.6206	0.4957	19906
42	1	1349.73	1676.83	3.912	29.640	0.8343	0.6111	0.4718	28829
43	1	1412.75	1705.97	3.209**	18.136**	0.8334	0.6128	0.4806	19355**
44	1	1427.42	1857.82	6.886	129.480	0.8274	0.5975	0.4631	59205
45	0	1267.31**	1522.04**	3.804	27.662	0.8327	0.6112	0.4826	23075
46	0	1409.25	1851.80	4.464	34.931	0.8464	0.6474	0.5281	25713
47	0	1297.18	1630.28	4.466	39.244	0.8404	0.6291	0.4969	32285
48	0	1402.83	1675.34	3.635	24.501	0.8325	0.6083	0.4734	24243
49	0	1447.75	1904.92	7.446	147.020	0.8326	0.6084	0.4693	57776
50	0	1385.06	2041.88	10.075*	215.940*	0.8433	0.6340	0.5019	66092*
Average	0.400	1404.07	1795.30	4.770	51.334	0.8345	0.6195	0.4884	34002.54
Std.Dev.	0.656	54.10	122.94	1.319	39.007	0.061	0.046	0.017	11857.74
Maximum	2	1535.36	2045.33	10.075	215.940	0.8488	0.6495	0.5314	66092
Minimum	0	1267.31	1522.04	3.209	18.136	0.8261	0.5970	0.4629	19355
Historic Data	-	1400.14	1569.03	4.225	30.297	0.8305	0.6133	0.4723	28300

* maximum, **minimum

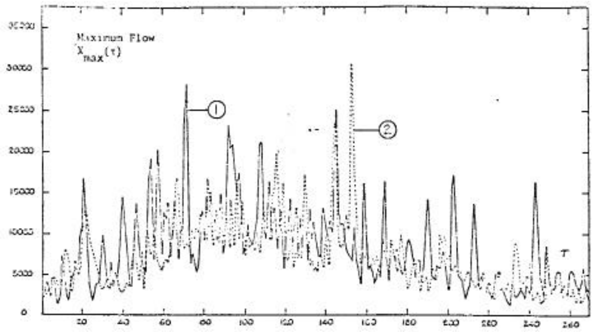

Fig. 6-23 Maximum Flow of 40 Years for Each Day τ, $X_{max}(\tau)$, with : (1) Estimates from Historic Data, and (2) Estimates from Generated Sample

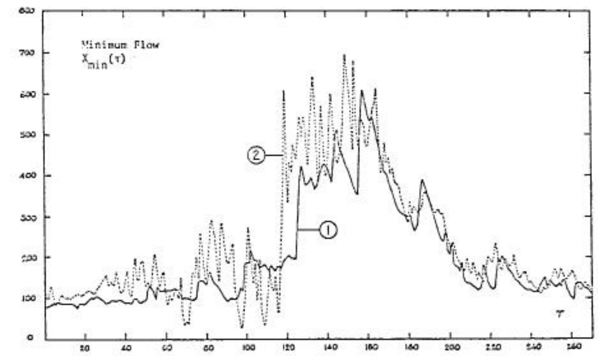

Fig. 6-24 Minimum Flow of 40 Years for Each Day τ, $X_{min}(\tau)$, with: (1) Estimates from Historic Data, and (2) Estimates from Generated Sample

The average value of all samples for the mean and the first three serial correlation coefficients of generated samples agree well with the corresponding estimates from historic data. The average values of all samples for the standard deviation, skewness and kurtosis as well as the maximum flow of each sample are somewhat greater than for the historic data.

The effects of adjustment for negative flows on the above statistics of generated daily flows are not significant since, on the average for all samples of the set, only 0.400 values out of 10,800 values or 0.0037 percent for sample of 40 years of daily flows, are adjusted for the negative values.

Comparison Based on Annual Flood Series. For each 40 year sample of generated set of daily flows, the annual flood series is computed. Moments and statistics for each annual flood series are estimated and results given in Table 6-11. For all samples, the values of the mean, standard deviation, maximum value and minimum value, as well as the historic value of each statistic are given at the bottom of each column for comparison purpose.

Similar to the case of the Boise River, comparisons based on statistics of daily flow series and their annual flood peak series of generated samples of daily flows show similar characteristics to those of historic data, therefore, generated samples produce extreme values which can be used for the study of flood peaks.

Table 6-11. Statistics of Annual Flood Peak Series of Generated Samples, Each Sample of 40 Years of Daily Flows

Sample Number	Statistics of Annual Flood Series					
	Mean	Std.Dev.	Skewness	Kurtosis	Minimum	Maximum
1	10927.95	5272.95	1.716	7.291	3869	30213
2	11392.78	8031.13	3.862	22.179	4546	52902
3	10854.23	4206.34	0.6712	2.916	4257	20813
4	12311.90	7120.14	2.077	7.357	4047	35041
5	11606.19	5360.20	1.453	5.908	3414	28414
6	10436.22	6989.33	2.099	7.821	4048	36237
7	12416.98	7705.30	1.519	4.960	4266	36352
8	13222.07	7390.74	1.835	8.373	3636	42139
9	10950.29	4733.73	0.787	3.466	3781	23199
10	10983.98	4753.60	0.862	3.988	4329	23736
11	12686.24	9161.01	2.091	7.831	3641	46005
12	12214.13	6340.93	1.614	6.691	3884	35075
13	12830.39	7793.28	1.303	4.484	4235	33254
14	12151.52	5223.63	0.867	4.142	4270	27186
15	11612.45	7635.04	2.207	9.193	2844**	41596
16	10787.36	5272.58	1.846	8.521	3654	31551
17	11649.36	5291.69	1.642	6.642	4518	30300
18	11558.75	8321.41	3.205	15.301	3868	49998
19	13177.97	6362.54	1.528	5.056	5808*	32127
20	13507.99	7082.30	1.717	7.705	3951	40418
21	11539.02	5657.89	1.032	3.504	4006	25273
22	11321.69	4046.33	1.110	4.998	5340	24412
23	12647.83	7866.62	1.685	6.019	4686	38498
24	10168.24	3969.32	1.340	5.918	4420	23397
25	13075.18	9915.20	3.525	18.061	4665	61041
26	10883.96	5886.33	1.329	4.427	3305	27598
27	10663.42	4683.75	1.449	6.141	4070	27315
28	11853.03	6288.06	2.672	14.206	3551	40768
29	12560.80	6928.61	1.564	5.165	4539	32640
30	11199.29	4410.31	0.268**	2.767	3196	21590
31	11728.24	5775.20	1.366	5.469	3889	30847
32	11333.71	7194.48	1.771	6.024	3649	34386
33	10838.98	4552.77	0.591	2.673	3780	20396
34	11538.36	6074.04	1.473	4.932	3941	29143
35	11774.56	6304.09	1.012	3.161	4511	27244
36	11674.17	6648.23	1.578	6.231	3546	35227
37	14603.34*	9459.35	3.362	18.261	5441	61091
38	11108.03	5012.69	1.476	5.966	4883	28657
39	11597.87	5973.67	0.861	2.666**	4408	24398
40	11836.02	6793.56	1.775	6.585	3668	33176
41	10944.54	3852.09	0.346	2.781	4110	19906
42	10832.05	5226.96	1.498	6.022	4583	28829
43	10776.18	3484.65**	0.610	3.610	4085	19355**
44	12554.12	8874.73	4.038*	23.665*	3883	59205
45	9836.30**	4308.80	1.143	4.575	3995	23075
46	11887.24	5723.28	0.917	3.424	4127	25713
47	10982.94	5596.46	1.872	7.816	4431	32285
48	10989.43	4499.21	1.147	4.886	4542	24243
49	12500.23	8971.60	3.745	20.528	3672	57776
50	13079.42	11261.23*	3.534	16.838	4258	66092*
Average	11712.14	6308.25	1.699	7.553	4121	34002
Std.Dev.	1337.90	1742.33	0.924	5.306	553.63	11857.74
Maximum	14603.34	11261.23	4.038	23.665	5808	66092
Minimum	9836.30	3484.65	0.268	2.666	2844	19355
Historic Data	13828.50	5144.76	0.938	3.908	6070	28300

*maximum, **minimum

Chapter VII
TWO STUDY CASES OF EFFICIENCY OF ANNUAL AND PARTIAL FLOOD SERIES

The long records of generated daily flow series for two study cases are used for comparing efficiency of using annual and partial flood series for estimating flood peaks of given return periods in this chapter. It includes the comparison of sampling variances of flood values for given return periods obtained from annual and partial flood series. Comparison of sampling mean square errors of estimates of flood values for given return periods in case of use of annual and partial flood series is also investigated. The long records of generated daily flow series are used for verifying properties and assumptions, as required in the development of the partial flood series model.

7.1 Annual and Partial Flood Series of Generated Daily Flows

The annual flood series of generated samples are largest flood peaks of daily flow for generated wet season flows. For each of two study cases, the Boise River and the Powell River, a total sample of generated 2000 years gives flood series of 2000 values of daily flows.

Similarly, partial flood series for given truncation levels are obtained from generated daily flows. In case of the Boise River, the lowest truncation level Q_b was selected in such a way that the average number of exceedances per year, estimate of λ, is 5.166. Table 7-1 gives the change of λ with the change of Q_b for all generated data of 2000 years. This table includes also the average value of the magnitude of all exceedances, estimate of β, and $Q_b + \beta \ln \lambda$ for each Q_b. In case of the Powell River, the changes of λ, β, and $Q_b + \beta \ln \lambda$ with the change of Q_b for all generated data are given in Table 7-2.

To ascertain whether annual and partial flood series, obtained from long sample of generated daily flows could be used for the study, relationships between frequencies of annual and expectancies of partial flood series are empirically determined from generated data and compared with the expected relationship resulting from the Langbein method (Langbein, 1949), independent of any assumption underlying probability distribution functions. The magnitude of floods corresponding to specified exceedance probabilities are computed by linear interpolation at the expected probability plotting positions, $m/(n+1)$, with m = rank in descending order, and n = number of years of records, for annual flood series. Corresponding expectancies of partial flood series are established by counting the total number of flood peak exceedances above each magnitude and dividing it by the total number of years of records. The computed expectancies of partial flood series are then compared with corresponding expectancies obtained from the Langbein method. The relationships between annual flood frequency and partial flood expectancy, as obtained by empirical method from generated data and by Langbein's method, for Q_b of 2870 cfs in case of the Boise River, and for Q_b of 4500 cfs in case of the Powell River, are given in Tables 7-3 and 7-4, respectively. For a given exceedance probability of annual flood series, the partial flood expectancy, obtained by the empirical method, agrees well with

Table 7-1. Variations of Values λ, β, and $Q_b + \beta \ln \lambda$ with Truncation Level Q_b, for Partial Flood Series Sample, 2000 Years Long, for the Boise River

	Truncation Level, Q_b									
	2870	3500	4000	4500	5000	5250	5500	5750	6000	6250
λ	5.166	4.156	3.427	2.645	2.091	1.819	1.559	1.352	1.182	1.009
β	2136.0	1947.0	1807.3	1770.7	1679.5	1660.7	1668.3	1655.6	1630.0	1635.9
$Q_b + \beta \ln \lambda$	6377.5	6273.6	6226.0	6222.3	6238.9	6243.6	6239.7	6249.3	6272.6	6264.7

Note: u = 5877.84; α = 1751.65

Table 7-2. Variations of Values λ, β, and $Q_b + \beta \ln \lambda$ with Truncation Level Q_b, for Partial Flood Series Sample, 2000 Years Long, for the Powell River

	Truncation Level, Q_b									
	4500	5000	5500	6000	6500	7000	7500	8000	8500	9500
λ	4.476	3.721	3.134	2.630	2.237	1.926	1.654	1.417	1.250	0.940
β	3335.5	3461.3	3568.1	3704.7	3813.1	3890.1	3992.7	4123.3	4139.9	4333.2
$Q_b + \beta \ln \lambda$	9494.6	9544.7	9565.9	9576.1	9567.0	9545.5	9500.9	9421.5	9165.2	9228.4

Note: u = 9114.29; α = 4114.69

Table 7-3. Relationships between Annual Flood Frequency and Partial Flood Expectancy for Truncation Level of 2870 cfs, for the Boise River

	Annual Flood Series		Partial Flood Series			
			Empirical Method		Langbein's Method	
Flood Magnitude	Exceedance Probability	Return Period	Expectancies	Return Period	Expectancies	Return Period
3525	0.980	1.02	3.406	0.29	3.912	0.26
3664	0.970	1.03	3.233	0.31	3.507	0.29
3805	0.960	1.04	3.077	0.32	3.219	0.31
3922	0.950	1.05	2.927	0.34	2.996	0.33
4404	0.900	1.11	2.302	0.43	2.303	0.43
5005	0.800	1.25	1.709	0.58	1.609	0.62
5539	0.700	1.43	1.248	0.80	1.204	0.83
6010	0.600	1.67	0.957	1.04	0.916	1.09
6516	0.500	2.00	0.690	1.45	0.693	1.44
7093	0.400	2.50	0.496	2.02	0.511	1.96
7782	0.300	3.33	0.320	3.12	0.357	2.80
8566	0.200	5.00	0.197	5.08	0.223	4.48
9650	0.100	10.00	0.090	11.11	0.105	9.49
10987	0.050	20.00	0.046	21.51	0.051	19.50
11364	0.040	25.00	0.037	27.03	0.041	24.50
12619	0.020	50.00	0.017	58.82	0.020	49.50
13770	0.010	100.00	0.008	117.65	0.010	99.50
15450	0.005	200.00	0.004	250.00	0.005	199.50
18140	0.002	500.00	0.002	500.00	0.002	499.50
18476	0.001	1000.00	0.001	1000.00	0.001	999.50

Table 7-4. Relationships between Annual Flood Frequency and Partial Flood Expectancy for Truncation Level of 4500 cfs, for the Powell River

	Annual Flood Series		Partial Flood Series			
			Empirical Method		Langbein's Method	
Flood Magnitude	Exceedance Probability	Return Period	Expectancies	Return Period	Expectancies	Return Period
4194	0.980	1.02	4.170	0.24	3.912	0.26
4518	0.970	1.03	4.145	0.24	3.507	0.29
4733	0.960	1.04	3.847	0.26	3.219	0.31
4919	0.950	1.05	3.598	0.28	2.996	0.33
5727	0.900	1.11	2.700	0.37	2.303	0.43
6926	0.800	1.25	1.843	0.54	1.609	0.62
7985	0.700	1.43	1.336	0.75	1.204	0.83
9047	0.600	1.67	1.001	1.00	0.916	1.09
10120	0.500	2.00	0.746	1.34	0.693	1.44
11415	0.400	2.50	0.544	1.84	0.511	1.96
13079	0.300	3.33	0.362	2.76	0.357	2.80
15240	0.200	5.00	0.224	4.46	0.223	4.48
19183	0.100	10.00	0.100	9.95	0.105	9.49
23736	0.050	20.00	0.048	20.83	0.051	19.50
25152	0.040	25.00	0.038	26.32	0.041	24.50
30213	0.020	50.00	0.018	55.56	0.020	49.50
35041	0.010	100.00	0.008	125.00	0.010	99.50
42139	0.005	200.00	0.004	250.00	0.005	199.50
59205	0.002	500.00	0.002	500.00	0.002	499.50
61090	0.001	1000.00	0.001	1000.00	0.001	999.50

that of the Langbein method, for both study cases. The expectancy of partial flood, obtained by the empirical method, may depend on Q_b especially in the range of low return periods. The effect of Q_b on expectancies of partial flood series for high return periods is not very high. Because of these agreements of expectancies of partial flood, for a given frequency of annual flood, the annual and partial flood series derived from generated daily flows of both the Boise River and the Powell River seem feasible for purposes of this study.

7.2 Comparison of Efficiency of Annual and Partial Flood Series by Using Ratios of Sampling Variances

Approaches used for comparison of sampling variances $\hat{Q}(T)$, outlined in Section 3.6, are applied here to generated data. The ratio of sampling variances of $\hat{Q}(T)$ based on the exact theoretical approach, $R_{v,1}$, given by Eq. 3-95 is shown in Fig. 3-3 as $R_{v,1}$ versus the return period T for a given Q_b, or the value of λ. The derivation of $R_{v,1}$ depends on the relations between parameters: α, u (of annual flood series model), and $\hat{\alpha}$, β (of partial flood series model), as shown in Eq. 3-94. To take into consideration differences between α and β, and u and $Q_b + \beta \ln\lambda$, ratios $R_{v,2}$ or var $\hat{Q}(T)$ based on the approximate theoretical approach, as shown in Eq. 3-96 are investigated.

In case of the empirical approach, the ratio of var $\hat{Q}(T)$, denoted by $R_{v,3}$, is obtained by Eq. 3-97. For each flood series, the long sample of 2000 years is divided into small samples, each of size N. For each small sample, the estimates $\hat{Q}(T)_a$ and $\hat{Q}(T)_p$ are obtained by Eqs. 3-76 and 3-86, respectively. Hence, for n samples each size N the ratio $R_{v,3}$ is obtained for a given return period by Eq. 3-97.

Boise River. The α and u values estimated from annual flood series of 2000 years, are 1751.65 and 5877.84, respectively. Theoretically $\alpha = \beta$, with β estimated from partial flood series. Table 7-1 shows how β varies with Q_b. For this case, α is in the range of computed β. As shown by Eq. 3-94, u = $Q_b + \beta \ln\lambda$. Table 7-1 gives $Q_b + \beta \ln\lambda$, estimated from partial flood series for various Q_b. It is seen that $Q_b + \beta \ln\lambda$ for the range of Q_b used is somewhat greater than u.

By substituting the estimates $\hat{\alpha}$, $\hat{\lambda}$, and $\hat{\beta}$ for each Q_b into Eq. 3-96, relationships between $R_{v,2}$ and T are obtained for various Q_b and are shown in Fig. 7-1. By comparing Fig. 3-3 with Fig. 7-1, relationships between ratio of var $\hat{Q}(T)$ and T, based on the exact (Fig. 3-3) and approximate (Fig. 7-1) theoretical approaches, are similar except for high values of $\hat{\lambda}$. For $1.00 \leq \hat{\lambda} \leq 2.25$, $R_{v,2}$ is greater than $R_{v,1}$ for the whole range of T considered. For $\hat{\lambda} \geq 3.4$, the relationship of $R_{v,2}$ and T tends to be unclear. In general, the larger λ, the greater the value is $R_{v,2}$. Figure 7-1 shows also that $R_{v,2}$ for $\hat{\lambda} = 5.166$ are smaller than $R_{v,2}$ for $\hat{\lambda} = 3.427$, for any T. This may come from an error in estimating β, since $\hat{\beta}$ for $\hat{\lambda} = 5.166$ is much larger than $\hat{\beta}$ for $\hat{\lambda} = 3.427$. It can also concluded from Fig. 7-1 that for the studied range of Q_b, the partial flood series estimates

$\hat{Q}(T)$ have a smaller sampling variance than the annual flood series estimates, if $\hat{\lambda}$ of partial flood series is at least 1.40. It should be stressed that ratios $R_{v,1}$ and $R_{v,2}$ do not depend on the sample size.

For empirical approach, results of relationships between $R_{v,3}$ and T for various Q_b, for N = 10,20,25, 40,50, and 100, are shown in Figs. 7-2 through 7-7, respectively. In general, curves of these figures are similar to those of Figs. 3-3 and 7-1, and this is especially the case for curves of Fig. 7-1. The conclusions from Figs. 7-2 through 7-7 are:

(i) The ratio $R_{v,3}$ seems to depend on the sample size. For small N, $R_{v,3}$ for a given T and for a given

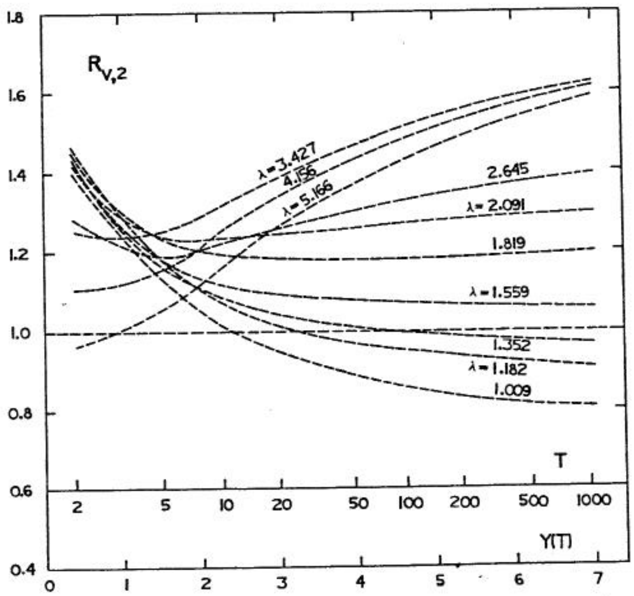

Fig. 7-1 Variation of $R_{v,2}$ with the Return Period T for the Range of λ from 5.166 to 1.009 (Q_b from 2870 to 6250 cfs), for the Boise River

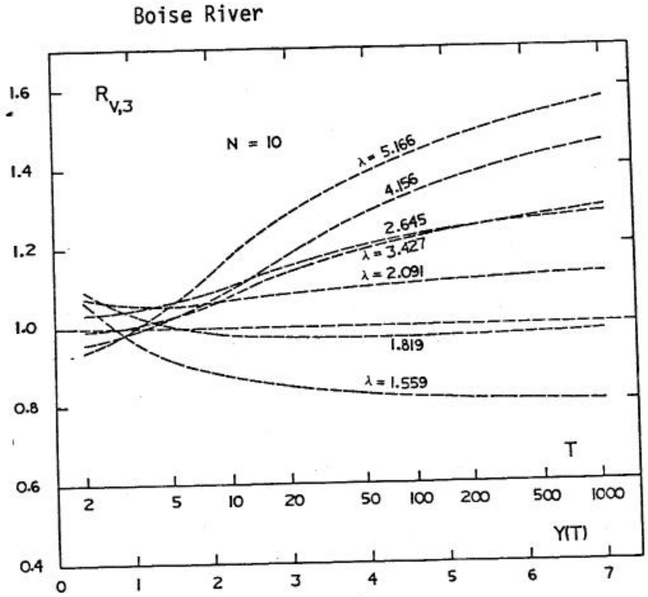

Fig. 7-2 Variation of $R_{v,3}$ with the Return Period T for the Range of λ from 5.166 to 1.559 (Q_b from 2870 to 5500 cfs), for N = 10, and for the Boise River Generated Samples

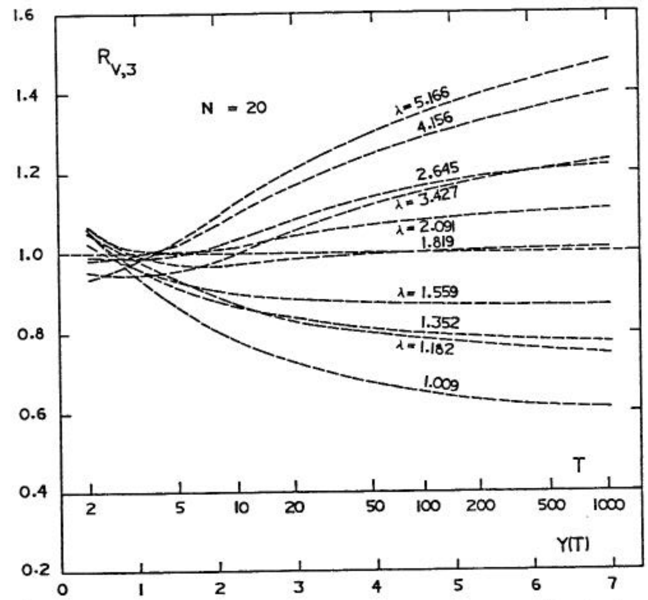

Fig. 7-3 Variations w$R_{v,3}$ with the Return Period T for the Range of λ from 5.166 to 1.009 (Q_b from 2870 to 6250 cfs), for N = 20, and for the Boise River Generated Samples

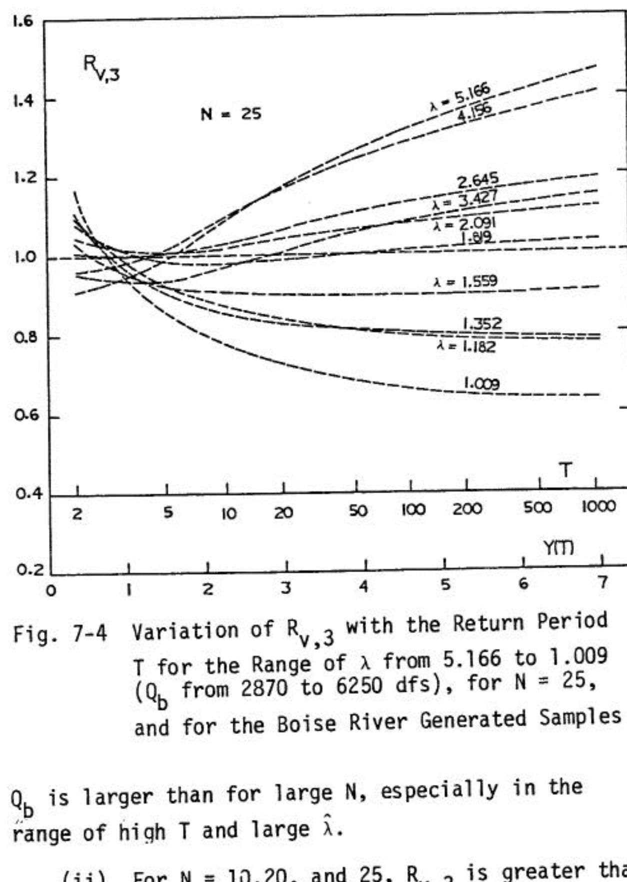

Fig. 7-4 Variation of $R_{v,3}$ with the Return Period T for the Range of λ from 5.166 to 1.009 (Q_b from 2870 to 6250 dfs), for N = 25, and for the Boise River Generated Samples

Q_b is larger than for large N, especially in the range of high T and large $\hat{\lambda}$.

(ii) For N = 10,20, and 25, $R_{v,3}$ is greater than unity for the studied range of return periods if $\hat{\lambda}$ approximately is at least 1.90; for larger N, $\hat{\lambda}$ should be somewhat greater than 1.90 for $R_{v,3}$ to be greater than unity.

(iii) For large N, the number of generated samples is small, with $R_{v,3}$-curve unreliable since some of them for small $\hat{\lambda}$ fall above the curves with larger

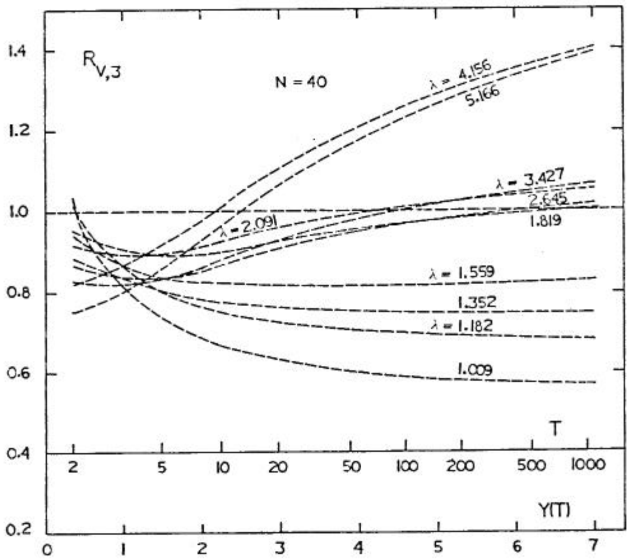

Fig. 7-5 Variation of $R_{v,3}$ with the Return Period T for the Range of λ from 5.166 to 1.009 (Q_b from 2870 to 6250 cfs), for N = 40, and for the Boise River Generated Samples

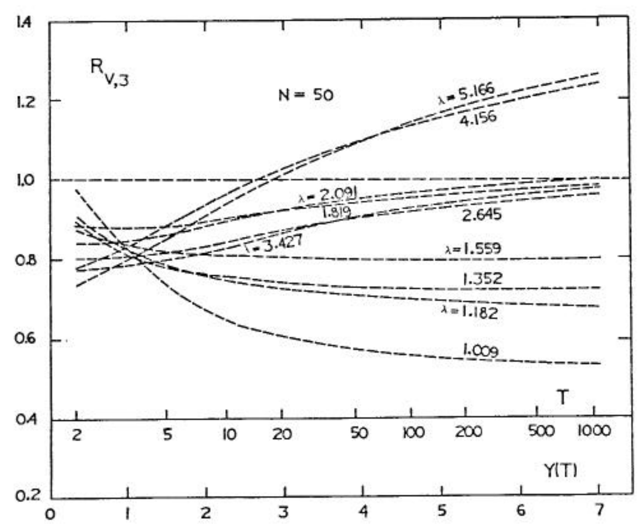

Fig. 7-6 Variation of $R_{v,3}$ with the Return Period T for the Range of λ from 5.166 to 1.009 (Q_b from 2870 to 6250 cfs), for N = 50, and for the Boise River Generated Samples

$\hat{\lambda}$, which by theory should not be true for the derived flood models.

Powell River. The α and u values estimated from annual flood series of 2000 years, are 4114.69 and 9114.29, respectively. Table 7-2 shows how β and $Q_b + \beta \ln\lambda$, estimated from partial flood series, vary with Q_b.

The variations of $R_{v,2}$ with T for the range of Q_b from 4500 to 9500 cfs are shown in Fig. 7-8. This figure shows that for the studied range of Q_b, the partial flood series estimates $\hat{Q}(T)$ have a smaller sampling variance than the annual flood series esti-

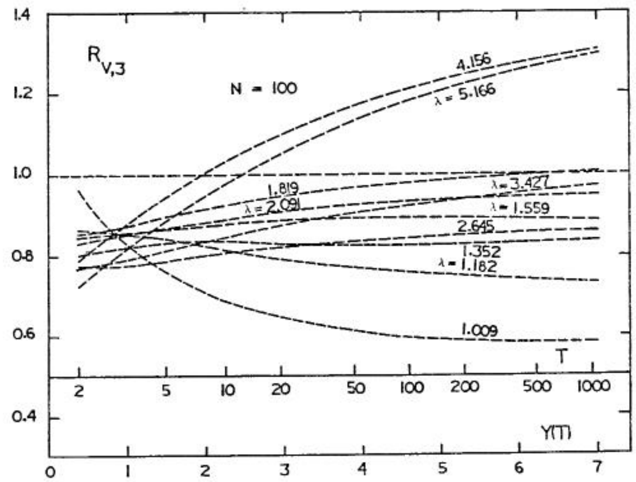

Fig. 7-7 Variation of $R_{v,3}$ with the Return Period T for the Range of λ from 5.166 to 1.009 (Q_b from 2870 to 6250 cfs), for N = 100, and for the Boise River Generated Samples

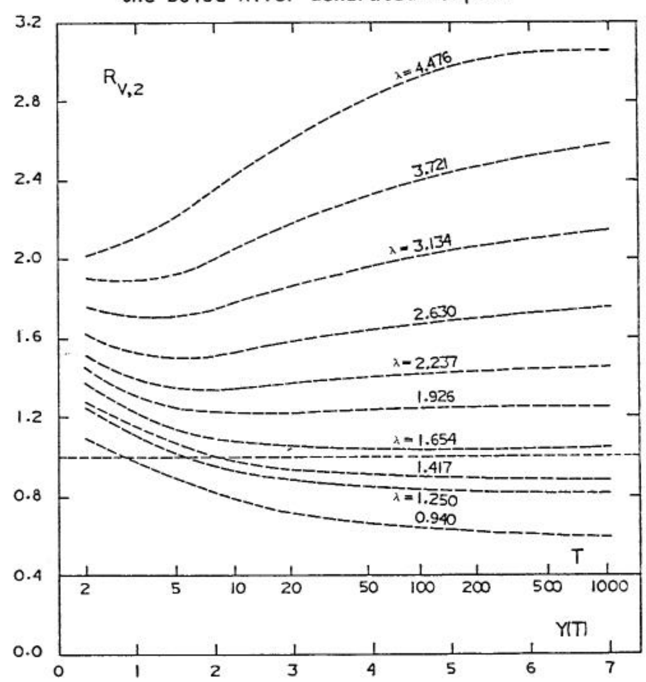

Fig. 7-8 Variation of $R_{v,2}$ with the Return Period T for the Range of λ from 4.476 to 0.940 (Q_b from 4500 to 9500 cfs), for the Powell River

mates, if $\hat{\lambda}$ of partial flood series is at least 1.60. For $\hat{\lambda} \geq 1.60$, $R_{v,2}$ is larger than $R_{v,1}$ for a given T, especially for large $\hat{\lambda}$.

In case of the empirical approach, relationships between $R_{v,3}$ and T for various Q_b, for N = 10,20,25, 40,50, and 100, are shown in Figs. 7-9 through 7-14, respectively. The conclusions from Figs. 7-9 through 7-14 are:

(i) The ratio $R_{v,3}$ seems to depend on the sample size. For small N, $R_{v,3}$ for a given T and for a given Q_b is larger than for large N, especially in the range of high T and large $\hat{\lambda}$.

(ii) For N = 10, $R_{v,3}$ is greater than unity for range of high T if $\hat{\lambda}$ approximately is at least 1.70; for larger N, $\hat{\lambda}$ should be greater than 1.70 for $R_{v,3}$ to be greater than unity.

(iii) For a given N, $R_{v,3}$ increases with decrease of Q_b or with an increase of $\hat{\lambda}$.

(iv) For a given N and for large $\hat{\lambda}$, $R_{v,3}$ in case of the Powell River are greater than $R_{v,3}$ in case of the Boise River, especially in the range of high return periods.

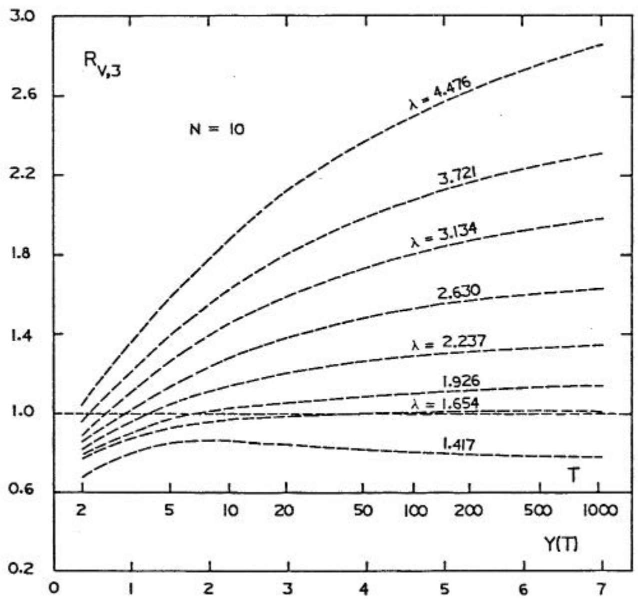

Fig. 7-9 Variation of $R_{v,3}$ with the Return Period T for the Range of λ from 4.476 to 1.417 (Q_b from 4500 to 8000 cfs), for N = 10, and for the Powell River Generated Samples

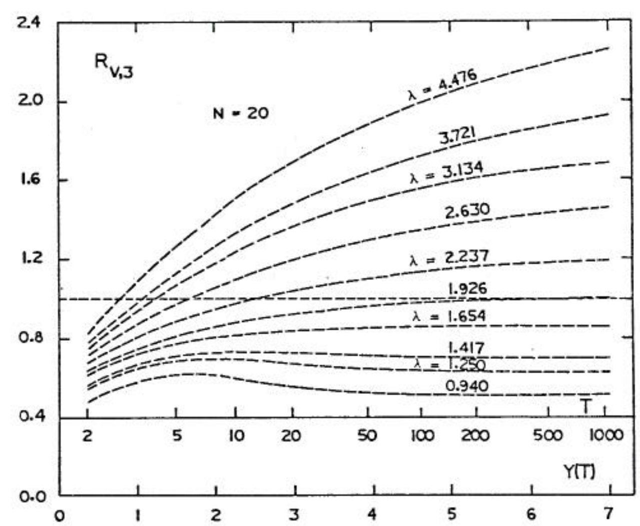

Fig. 7-10 Variation of $R_{v,3}$ with the Return Period T for the Range of λ from 4.476 to 0.940 (Q_b from 4500 to 9500 cfs), for N = 20, and for the Powell River Generated Samples

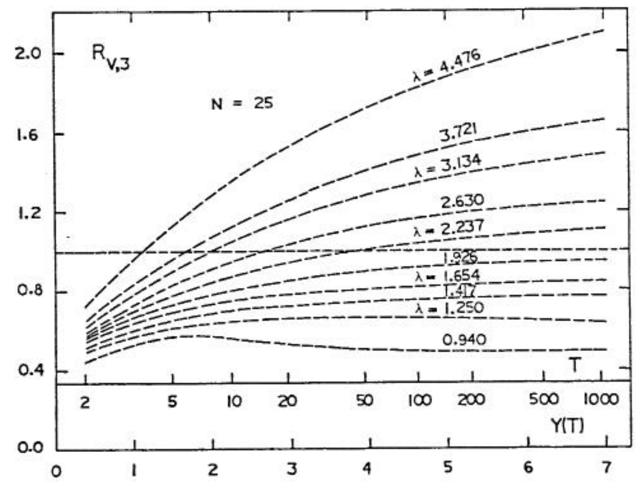

Fig. 7-11 Variation of $R_{v,3}$ with the Return Period T for the Range of λ from 4.476 to 0.940 (Q_b from 4500 to 9500 cfs), for N = 25, and for the Powell River Generated Samples

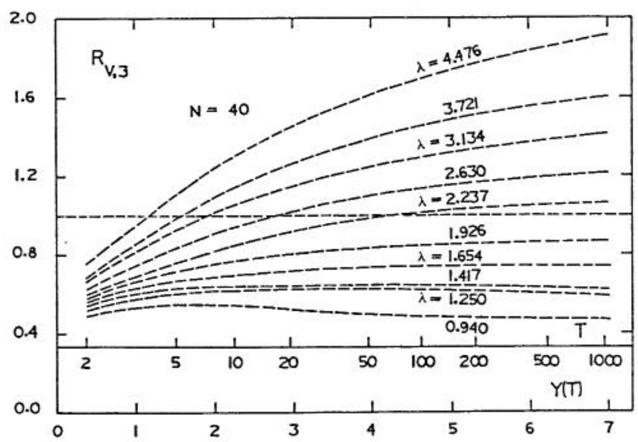

Fig. 7-12 Variation of $R_{v,3}$ with the Return Period T for the Range of λ from 4.476 to 0.940 (Q_b from 4500 to 9500 cfs), for N = 40, and for the Powell River Generated Samples

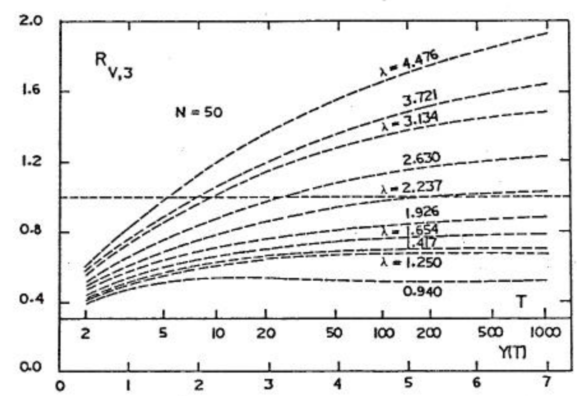

Fig. 7-13 Variation of $R_{v,3}$ with the Return Period T for the Range of λ from 4.476 to 0.940 (Q_b from 4500 to 9500 cfs), for N = 50, and for the Powell River Generated Samples

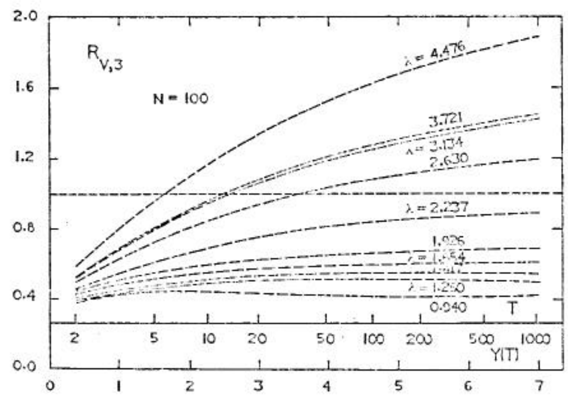

Fig. 7-14 Variation of $R_{v,3}$ with the Return Period T for the Range of λ from 4.476 to 0.940 (Q_b from 4500 to 9500 cfs), for N = 100, and for the Powell River Generated Samples

In conclusion, based on the average of the two study cases, the estimates of Q(T) from partial flood series have a smaller sampling variance than the estimates from annual flood series, if partial flood series have $\hat{\lambda}$ at least 1.65 based on the exact theoretical approach, and 1.50 based on the approximate theoretical approach. The conclusion in case of the exact theoretical approach is similar to that concluded by Cunnane (1973). Ratios of var $\hat{Q}(T)$ in case of exact and approximate theoretical approaches do not depend on the sample size. For the empirical approach, $\hat{\lambda}$ should be at least 1.95 for $R_{v,3}$ to be greater than unity for the range of N from 10 to 25, and for the range of high return periods.

The case of var $\hat{Q}(T)$ obtained from the partial flood series to be less than the corresponding var $\hat{Q}(T)$ obtained from the annual flood series implies than that the partial flood series is more efficient or more useful for estimating annual flood peaks of given return periods than the annual flood series.

Results obtained from the exact theoretical, approximate theoretical, and empirical approaches are somewhat different. Differences depend: (i) on how well the assumed flood models represent the true population models for both flood series, (ii) on the validity of assumptions used in deriving var $\hat{Q}(T)$ for each flood series, (iii) for the empirical approach, on the accuracy in estimating var $\hat{Q}(T)$ for each flood series, which estimates depend on the number of available samples, and (iv) eventually on the reproducibility of properties of the daily flow process by generating new samples.

In case of the exact theoretical approach, $R_{v,1}$ increases with an increase of λ, for a given T, and it approaches infinity as λ approaches infinity which can be seen from Eq. 3-95. However, there is a limit for λ which has the value much smaller than infinity. The value of λ may be close to infinity if the instantaneous flow hydrograph is used and every point of instantaneous flow above Q_b is considered as the partial flood series. In case of the use of mean daily flow hydrograph, the possible maximum value of λ is 365, if all of mean daily flows are above Q_b and they are considered as partial flood series. However, by the definition of partial flood series used in this study (Section 3.1), the value of λ must be much smaller than 365 since only separate flood peaks above Q_b are considered. By considering the assumptions used for deriving the assumed partial flood series model such as the indpendence of the successive exceedances, their validities tend to be supported by the observed data only within the range of Q_b such that λ is not greater than 4 or 5. Furthermore, in case of the empirical approach, the partial flood series is derived from the generated daily flow data which are generated only within the wet season of the year. Hence, for the range of Q_b such that λ is greater than 4 or 5, the distortion of partial flood series may not be neglected since some partial flood peaks may occur outside the selected wet season. For these reasons, the comparison of sampling variances of Q(T) of both flood peak series is studied only for the range of λ up to about 5.

7.3 Comparison of Theoretical and Empirical Sampling Variances of Estimates of Flood Values for Given Return Periods

Using Annual Flood Series. Let R_a denote the ratio of var $\hat{Q}(T)$ estimated from annual flood series by using empirical and theoretical approaches. Then, from Eqs. 3-71 and 3-79,

$$R_a = \frac{N \sum_{i=1}^{n} [\hat{Q}_i(T)_a - \overline{Q(T)_a}]^2}{\alpha^2 (n-1)[1.11 + 0.52\, y(T) + 0.61\, y^2(T)]} , \qquad (7\text{-}1)$$

with N = the sample size in years, n = the number of of generated samples of size N in the empirical approach, α = the model parameter estimated from 2000 values of annual flood series, $\hat{Q}_i(T)_a$ = the Q(T) estimate from the i-th sample.

Variations of R_a with the return period T, expressed in terms of y(T), for N = 10,25,50 and 100, in case of the Boise River and the Powell River, are shown in Figs. 7-15 and 7-16, respectively. In case of the Boise River, the average values of R_a for the return periods are 1.05, 1.19, 1.15 and 1.39 for N = 10,25,50 and 100, respectively. For the Powell River, the average values of R_a for the return periods are 1.57, 1.30, 1.35 and 1.45 for N = 10,25, 50 and 100, respectively. For both cases, R_a tends to be constant in the range of high T, for a given N. On the average, R_a tends to increase with N, and the estimated values of R_a for all cases are greater than one; indicating that var $\hat{Q}(T)_a$ based on the theoretical approach is less than the corresponding var $\hat{Q}(T)_a$ based on the empirical approach.

Using Partial Flood Series. Let R_p denote the ratio of var $\hat{Q}(T)$ obtained from partial flood series by using the empirical and theoretical approaches. Then, from Eqs. 3-71 and 3-92.

$$R_p = \frac{\lambda N \sum_{i=1}^{n} [\hat{Q}_i(T)_p - \overline{Q(T)_p}]^2}{\beta^2 (n-1)\ [1 + \{\ln\lambda + y(T)\}^2]} , \qquad (7\text{-}2)$$

with N = the sample size, n = the number of generated samples of size N in the empirical approach, λ and β = the model parameters estimated from 2000 years of

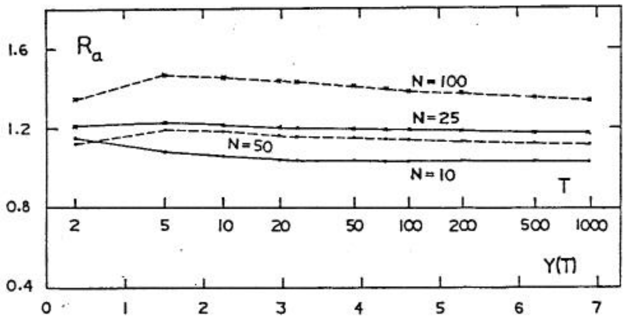

Fig. 7-15 Variation of R_a with the Return Period T for N = 10, 25, 50 and 100, for the Boise River

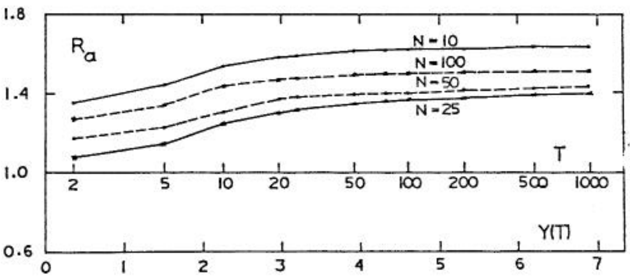

Fig. 7-16 Variation of R_a with the Return Period T for N = 10, 25, 50, and 100, for the Powell River

generated partial flood series, $\hat{Q}_i(T)_p$ = Q(T) estimate from the i-th sample.

For a given T, let $\overline{R_p}$ be the average of R_p for all selected Q_b, and for a given N. Variations of $\overline{R_p}$ with T for N = 10,25,50 and 100, in case of the Boise River and the Powell River, are shown in Figs. 7-17 and 7-18, respectively. $\overline{R_p}$ tends to be constant for a high return period, but to increase with its decrease. Considering the whole range of T, $\overline{R_p}$ increases with an increase of N. For the Boise River, the average values of $\overline{R_p}$ for all the return periods studied are 1.39, 1.44, 1.59 and 1.90 for N = 10,25, 50 and 100, respectively. In case of the Powell

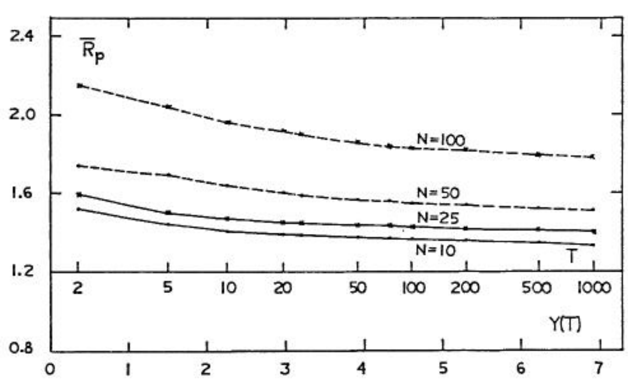

Fig. 7-17 Variation of $\overline{R_p}$ with the Return Period T for N = 10, 25, 50 and 100, for the Boise River

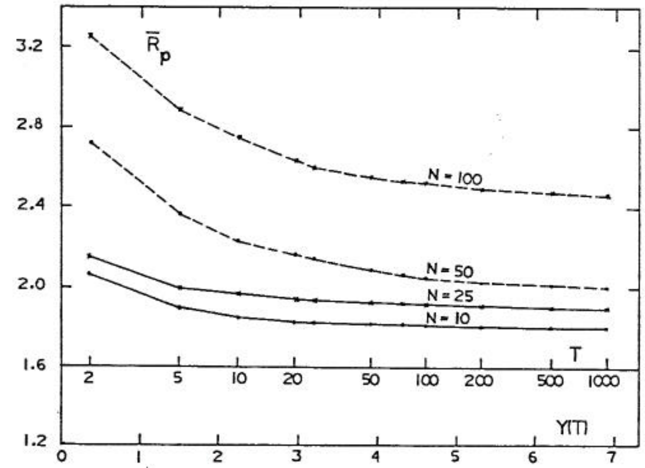

Fig. 7-18 Variation of $\overline{R_p}$ with the Return Period T for N = 10, 25, 50 and 100, for the Powell River

River, the average values of $\overline{R_p}$ are 1.85, 1.95, 2.21 and 2.69 for N = 10,25,50 and 100, respectively. Comparing for a given N and a given T, $\overline{R_p}$ is greater than R_a. In using both flood series, var $\hat{Q}(T)$ in case of the theoretical approach is smaller than the corresponding var $\hat{Q}(T)$ in case of the empirical approach.

7.4 Comparison of Sampling Mean Square Errors of Flood Values for Given Return Periods

Nash and Amorocho (1966) concluded that the flood magnitude of any given T can be estimated subject to error resulting from two different causes: (i) Failure of the model to conform to the universe of flood peaks of a catchment; and (ii) sampling errors resulting from non-representativeness of the record from which the model parameters are estimated. To test the accuracy of the assumed model to predict the flood peak Q(T) for a given T from each flood series, by considering the bias term, the comparison of sampling mean square errors in $\hat{Q}(T)$ is investigated on the long sample of generated data, considering it as an assumed population. For a given N, the sampling mean square error M of $\hat{Q}(T)$ is computed by

$$M = \frac{1}{n} \sum_{i=1}^{n} [\hat{Q}_i(T) - Q(T)]^2 , \qquad (7\text{-}3)$$

with Q(T) = the expected value of flood peak for a given T, $\hat{Q}_i(T)$ = the Q(T) estimate from the i-th sample, i = 1,2,...,n, and n = the number of samples of size N.

By expanding Eq. 7-3, then

$$M = \frac{1}{n} \sum_{i=1}^{n} [\hat{Q}_i(T) - \overline{Q(T)}]^2 + [\overline{Q(T)} - Q(T)]^2 , \qquad (7\text{-}4)$$

with $\overline{Q(T)}$ = the mean of all $\hat{Q}_i(T)$'s.

The term $[\overline{Q(T)} - Q(T)]^2$ results from the failure of the assumed model to conform with the population of flood peak properties and the bias in estimating model parameters.

For each flood series, the generated sample of 2000 years is split into two equal groups. The first group is assumed to be the population of flood series. The estimate of Q(T) for each T from the first group of long sample of annual flood series is assumed to be known population value. For the second group, the long sample of 1000 years is divided into small samples, each with size N. The value of M of each flood series is then computed by Eq. 7-3.

Let R_m denote the ratio of M of annual flood series sample to the corresponding M of partial flood series sample. Then,

$$R_m = \frac{\sum_{i=1}^{n} [\hat{Q}_i(T)_a - Q(T)]^2}{\sum_{i=1}^{n} [\hat{Q}_i(T)_p - Q(T)]^2} \qquad (7\text{-}5)$$

with a and p standing for annual partial flood series, respectively.

After computing R_m for various T by using Q(T) estimated from the first group, another set of R_m is obtained by interchanging groups in the same procedure. The average value $\overline{R_m}$ for these two steps is then obtained for each T.

Boise River. Variations of estimated $\overline{R_m}$ for each T and various Q_b are shown in Figs. 7-19 through 7-24 for N = 10,20,25,40,50 and 100, respectively. Except for the range of high T, these figures show that, on the average, $\overline{R_m}$ increases with an increase of Q_b up to Q_b such that $\hat{\lambda}$ is about two. For $\hat{\lambda} < 2$, $\overline{R_m}$ decreases with an increase of Q_b. By considering Eq. 7-4 and the range of low Q_b, the first right term in this equation is favorable to the use of partial flood series in estimating flood peaks, while the second right term is unfavorable for this purpose. In case of partial flood series, the first term decreases with an increase of λ, while the second term may increase or decrease, likely increasing with an increase of λ, especially in the range of large λ.

It is interesting to note that $\overline{R_m}$ for a given T is very sensitive to the population value Q(T). The values of Q(T) for various T and for the long samples of the first group and the second group of annual flood series are shown in Tables 7-5 and 7-6, respectively. Estimates of Q(T) of Gumbel distribution, denoted by $Q(T)_a$, and of assumed partial flood series model, denoted by $Q(T)_p$, from both long samples of the second and first groups are shown in Tables 7-5 and 7-6, respectively. For the range of T from 500 to 1000 years, the estimates of $Q(T)_p$ for Q_b = 3500, 4000, 4500 are closer to estimates of Q(T) obtained by using the plotting position than are the estimates of $Q(T)_a$. Hence, for these ranges of T and Q_b, $\overline{R_m}$ are very high as shown in Figs. 7-19 through 7-24.

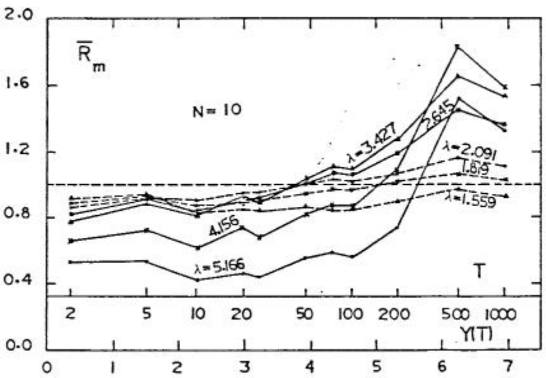

Fig. 7-19 Variation of $\overline{R_m}$ with the Return Period T for the Range of λ from 5.166 to 1.559 (Q_b from 2870 to 5500 cfs) for N = 10, for the Boise River Generated Samples

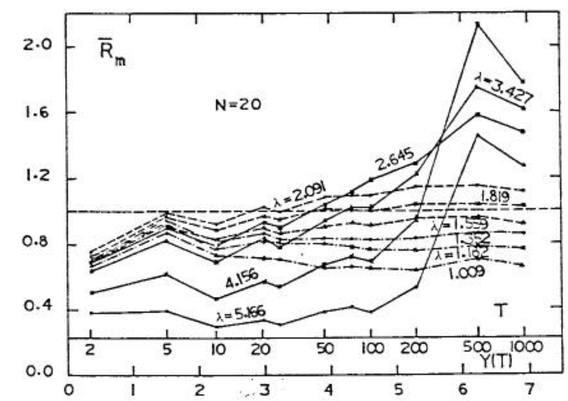

Fig. 7-20 Variation of $\overline{R_m}$ with the Return Period T for the Range of λ from 5.166 to 1.009 (Q_b from 2870 to 6250 cfs) for N = 20, for the Boise River Generated Samples

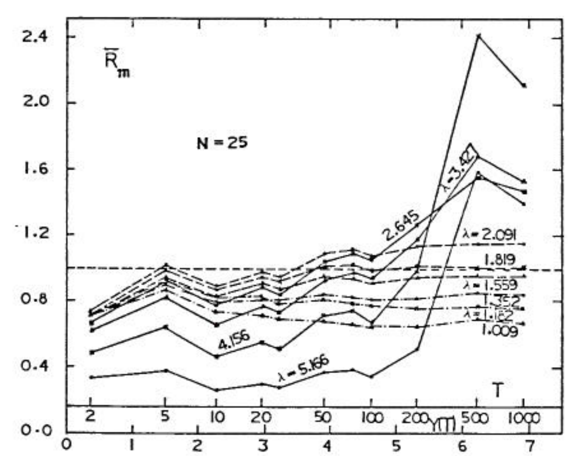

Fig. 7-21 Variation of $\overline{R_m}$ with the Return Period T for the Range of λ from 5.166 to 1.009 (Q_b from 2870 to 6250 cfs), for N = 25, for the Boise River Generated Samples

Table 7-5. Assumed Population of Flood Peaks for Various Return Periods, Obtained by Using: (1) Plotting Position Method, Q(T), for the First Generated Sample of 1000 Years, (2) Gumbel Distribution, $Q(T)_a$, (3) Assumed Partial Flood Series Model, $Q(T)_p$, for the Second Generated Sample of 1000 Years, for the Boise River

	Return Period, T										
	2	5	10	20	25	50	75	100	200	500	1000
Q(T) by Plotting Position (First 1000 Year Sample)	6482	8542	9665	11221	11430	12460	13428	14103	15386	18140	19690
$Q(T)_a$	6520	8533	9867	11146	11551	12801	13527	14042	15277	16908	18140
$Q(T)_p$ for Q_b = 2870	7267	9704	11317	12864	13355	14868	15747	16369	17864	19838	21329
$Q(T)_p$ for Q_b = 3500	7066	9267	10724	12121	12565	13930	14724	15286	16637	18418	19765
$Q(T)_p$ for Q_b = 4000	6968	9017	10375	11677	12090	13362	14102	14625	15883	17543	18798
$Q(T)_p$ for Q_b = 4500	6942	8935	10255	11522	11923	13160	13880	14388	15612	17227	18447
$Q(T)_p$ for Q_b = 5000	6914	8788	10028	11218	11596	12759	13434	13913	15063	16580	17727
$Q(T)_p$ for Q_b = 5500	6912	8785	10025	11214	11592	12755	13430	13908	15058	16575	17721
$Q(T)_p$ for Q_b = 6000	6930	8748	9951	11105	11472	12599	13255	13719	14835	16306	17419

Table 7-6. Assumed Population of Flood Peaks for Various Return Periods, Obtained by Using: (1) Plotting Position Method, Q(T), for the Second Generated Sample of 1000 Years, (2) Gumbel Distribution, $Q(T)_a$, (3) Assumed Partial Flood Series Model, $Q(T)_p$, for the First Generated Sample of 1000 Years, for the Boise River.

	Return Period, T										
	2	5	10	20	25	50	75	100	200	500	1000
Q(T) by Plotting Position (Second 1000 Year Sample)	6538	8638	9617	10755	11131	12851	13522	13752	15533	18304	18476
$Q(T)_a$	6520	8477	9773	11015	11410	12624	13330	13830	15031	16616	17814
$Q(T)_p$ for Q_b = 2870	7052	9457	11049	12577	13061	14554	15421	16035	17512	19459	20931
$Q(T)_p$ for Q_b = 3500	6903	9116	10581	11987	12433	13806	14604	15169	16527	18320	19674
$Q(T)_p$ for Q_b = 4000	6806	8853	10208	11508	11920	13190	13928	14451	15707	17365	18618
$Q(T)_p$ for Q_b = 4500	6794	8815	10152	11435	11842	13096	13825	14341	15581	17217	18452
$Q(T)_p$ for Q_b = 5000	6786	8719	9998	11226	11615	12815	13512	14005	15192	16757	17940
$Q(T)_p$ for Q_b = 5500	6785	8694	9958	11170	11554	12739	13427	13915	15086	16632	17799
$Q(T)_p$ for Q_b = 6000	6802	8679	9922	11115	11493	12659	13336	13815	14967	16488	17637

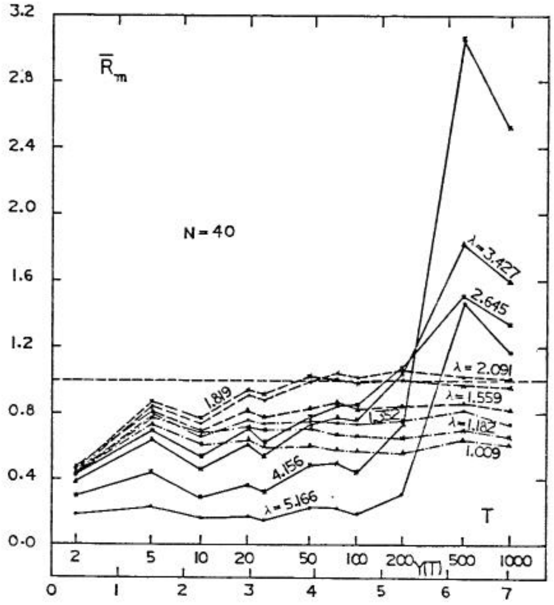

Fig. 7-22 Variation of $\overline{R_m}$ with the Return Period T for the Range of λ from 5.166 to 1.009 (Q_b from 2870 to 6250 cfs), for N = 40, for the Boise River Generated Samples

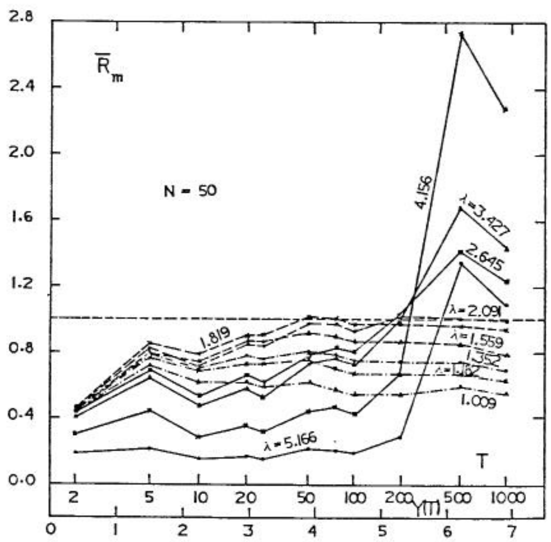

Fig. 7-23 Variation of $\overline{R_m}$ with the Return Period T for the Range of λ from 5.166 to 1.009 (Q_b from 2870 to 6250 cfs), for N = 50, for the Boise River Generated Samples

Powell River. Variations of estimated $\overline{R_m}$ for each T and various Q_b are shown in Figs. 7-25 through 7-30 for N = 10,20,25,40,50 and 100, respectively. These figures show that in the range of low T, $\overline{R_m}$

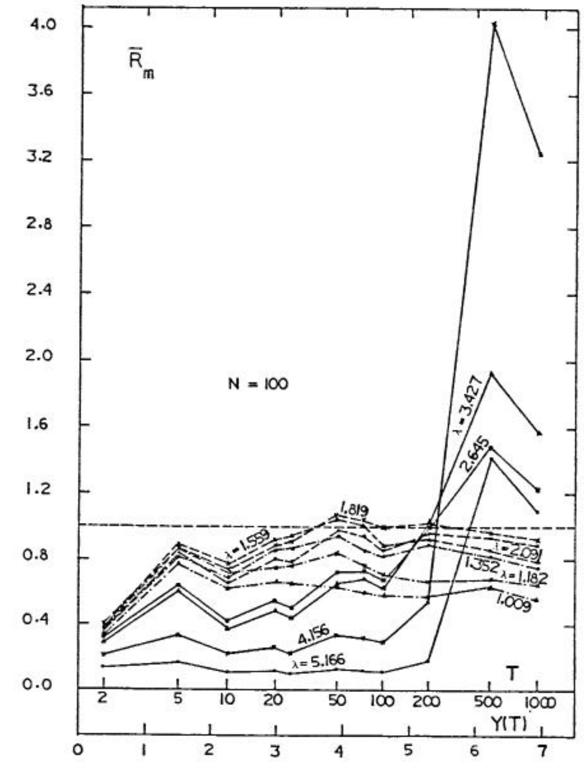

Fig. 7-24 Variation of $\overline{R_m}$ with the Return Period T for the Range of λ from 5.166 to 1.009 (Q_b from 2870 to 6250 cfs), for N = 100, for the Boise River Generated Sample

decreases with an increase of Q_b, while in the range of high T it increases with an increase of Q_b. By comparing with results of the study of the sampling variance of $\hat{Q}(T)$, the sampling mean square error of $\hat{Q}(T)$ for the partial flood series is influenced by the bias term in the range of low Q_b.

The population values Q(T) for various T and for the long samples of the first group and the second group of annual flood series are shown in Tables 7-7 and 7-8, respectively. Estimates of $Q(T)_a$ and $Q(T)_p$, from both long samples of the second group and first group are also shown in Tables 7-7 and 7-8, respectively. By comparing Tables 7-7 and 7-8 with Tables 7-5 and 7-6, the assumed flood models for both flood series do not predict well the population Q(T) in case of the Powell River, especially in the range of high T. Consequently, the bias terms for both flood series are larger for the Powell River than for the Boise River.

To determine the effect of the assumed population values Q(T) by using the plotting position method of generated data series, $Q(T)_a$ and $Q(T)_p$ are used as population values Q(T) in Eq. 7-5 for both the annual and partial flood series. The results indicate that the relationship of $\overline{R_m}$ to T is generally similar to the relationship between $R_{v,3}$ and T for each sample size N.

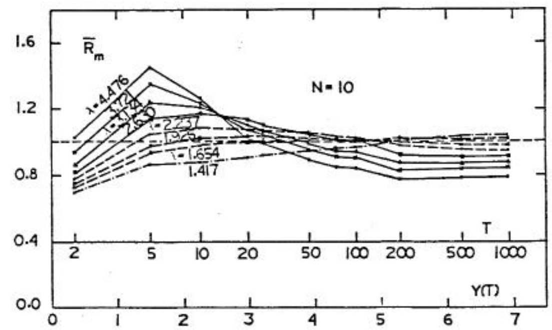

Fig. 7-25 Variation of $\overline{R_m}$ with the Return Period T for the Range of λ from 4.476 to 1.417 (Q_b from 4500 to 8000 cfs), for N = 10, for the Powell River Generated Samples

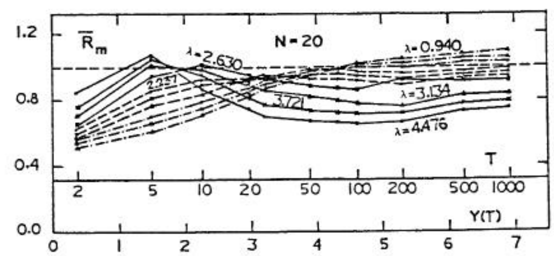

Fig. 7-26 Variation of $\overline{R_m}$ with the Return Period T for the Range of λ from 4.476 to 0.940 (Q_b from 4500 to 9500 cfs), for N = 20, for the Powell River Generated Samples

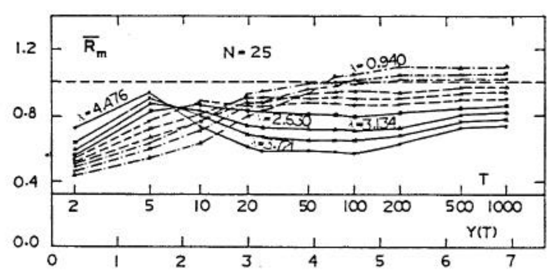

Fig. 7-27 Variation of $\overline{R_m}$ with the Return Period T for the Range of λ from 4.476 to 0.940 (Q_b from 4500 to 9500 cfs), for N = 25, for the Powell River Generated Samples

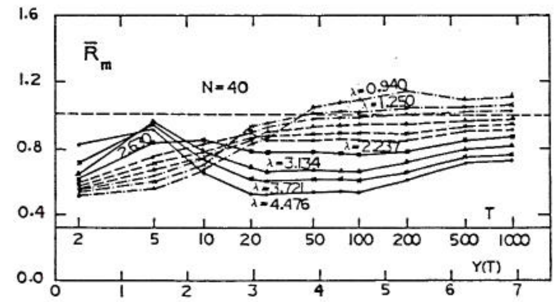

Fig. 7-28 Variation of $\overline{R_m}$ with the Return Period T for the Range of λ from 4.476 to 0.940 (Q_b from 4500 to 9500 cfs), for N = 40, for the Powell River Generated Samples

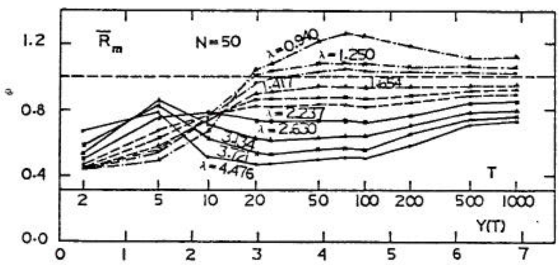

Fig. 7-29 Variation of $\overline{R_m}$ with the Return Period T for the Range of λ from 4.476 to 0.940 (Q_b from 4500 to 9500 cfs), for N = 50, for the Powell River Generated Samples

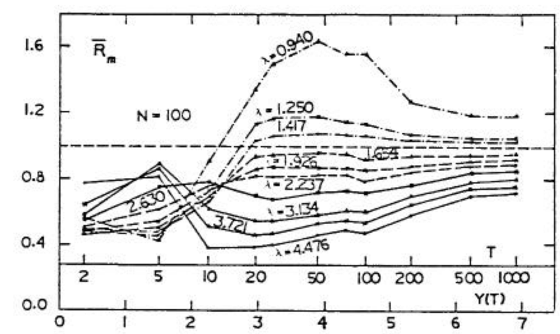

Fig. 7-30 Variation of $\overline{R_m}$ with the Return Period T for the Range of λ from 4.476 to 0.940 (Q_b from 4500 to 9500 cfs), for N = 100, for the Powell River Generated Samples

The comparison of sampling mean square errors of $\hat{Q}(T)$ for annual and partial flood series depends on the assumed population value Q(T), which is not known in practical cases. $\overline{R_m}$ is sensitive to the assumed population value Q(T). However, if Q(T) is assumed to be estimated from the long sample of annual flood series by the method of plotting position, it can be concluded from the use of partial flood series that the first right term in Eq. 7-4 decreases with a decrease of Q_b, while the second right term tends to increase with a decrease of Q_b, especially in the range of low Q_b. If the partial flood series model is developed in such a way that the assumptions required for its derivation are still valid or supported by observed data in the range of low Q_b, the partial flood series will be more efficient in estimating flood peaks of given return periods than the annual flood series, especially in cases of small sample sizes.

Table 7-7. Assumed Population of Flood Peaks for Various Return Periods, Obtained by Using: (1) Plotting Position Method, Q(T), for the First Generated Sample of 1000 Years, (2) Gumbel Distribution, $Q(T)_a$, (3) Assumed Partial Flood Series Model, $Q(T)_p$, for the Second Generated Sample of 1000 Years, for the Powell River

	Return Period, T										
	2	5	10	20	25	50	75	100	200	500	1000
Q(T) by Plotting Position (First 1000 Year Sample)	10241	15284	19477	24482	26053	32119	35016	36351	42136	52896	61033
$Q(T)_a$	10525	15069	18077	20962	21877	24697	26336	27496	30284	33963	36744
$Q(T)_p$ for Q_b = 4500	10347	14053	16506	18860	19606	21907	23243	24189	26464	29465	31733
$Q(T)_p$ for Q_b = 5000	10422	14254	16790	19224	19996	22374	23755	24734	27085	30188	32533
$Q(T)_p$ for Q_b = 6000	10528	14634	17352	19960	20787	23335	24816	25864	28385	31709	34222
$Q(T)_p$ for Q_b = 7000	10540	14816	17648	20364	21226	23880	25423	26514	29139	32602	35220
$Q(T)_p$ for Q_b = 7500	10524	14938	17860	20664	21553	24292	25884	27011	29720	33294	35995
$Q(T)_p$ for Q_b = 8000	10472	15081	18133	21061	21989	24850	26512	27689	30518	34251	37071
$Q(T)_p$ for Q_b = 9500	10404	15125	18252	21251	22202	25133	26837	28042	30940	34764	37654

Table 7-8. Assumed Population of Flood Peaks for Various Return Periods Obtained by Using: (1) Plotting Position Method, Q(T), for the Second Generated Sample of 1000 Years, (2) Gumbel Distribution, $Q(T)_a$, (3) Assumed Partial Flood Series Model, $Q(T)_p$, for the First Generated Sample of 1000 Years, for the Powell River

	Return Period, T										
	2	5	10	20	25	50	75	100	200	500	1000
Q(T) by Plotting Position (Second 1000 Year Sample)	9961	15222	18930	23070	24392	28653	31445	32893	42274	61087	66087
$Q(T)_a$	10719	15503	18671	21709	22673	25642	27368	28589	31525	35399	38327
$Q(T)_p$ for Q_b = 4500	11087	14942	17495	19943	20720	23112	24503	25487	27853	30975	33334
$Q(T)_p$ for Q_b = 5000	11204	15219	17877	20427	21236	23728	25176	26201	28665	31916	34373
$Q(T)_p$ for Q_b = 6000	11341	15633	18474	21200	22065	24729	26277	27373	30006	33482	36109
$Q(T)_p$ for Q_b = 7000	11403	15944	18951	21835	22750	25569	27207	28366	31154	34831	37610
$Q(T)_p$ for Q_b = 7500	11404	16041	19112	22056	22990	25868	27541	28725	31571	35326	38164
$Q(T)_p$ for Q_b = 8000	11393	16131	19267	22276	23231	26171	27880	29089	31997	35833	38732
$Q(T)_p$ for Q_b = 9500	11243	16382	19784	23048	24083	27272	29126	30438	33592	37753	40898

7.5 Distribution of the Number of Exceedances

One of the assumptions in deriving the commonly assumed partial flood model is the use of Poisson distribution for the number of exceedances η. In the developed model either the three-parameter mixed Poisson distribution or the simple Poisson distribution were found applicable. The following is the test on how the mixed Poisson distribution improves the goodness of fitting the frequency distributions of η for given Q_b and N, as derived from the long generated sample.

The first group of N = 1000 for the case of partial flood series of generated daily flows is used for investigation. For a given N, this long sample is divided into n small samples of equal size. For each small sample, the chi-square test statistic is used as criterion for fitting the frequency distributions of η both by the Poisson and by the mixed Poisson distributions, with the number of class intervals varying from 8 for the highest Q_b to 12 for the lowest Q_b, for N = 25, and from 10 to 13 for N = 50 and 100, and from 12 to 15 for N = 1000, respectively, for the Boise River. In case of the Powell River, the number of class intervals varying from 9 to 12 for N = 25, from 11 to 15 for N = 50 and 100, and from 12 to 20 for N = 1000, respectively. Average values for all n small samples of chi-square statistic are computed for both distributions, with the results for N = 25, 50, 100 and 1000 shown in Fig. 7-31 for the Boise River, and Fig. 7-32 for the Powell River, respectively.

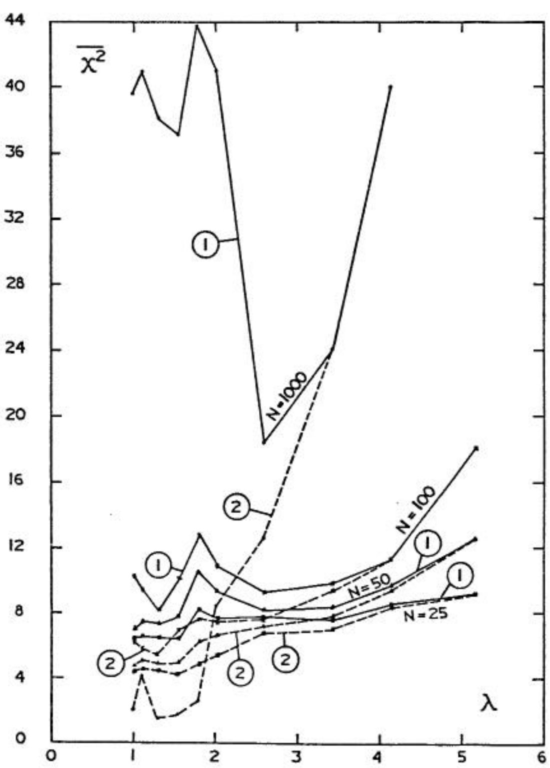

Fig. 7-31 Variation of Average Chi-Square $\overline{\chi}^2$ with Truncation Level (Expressed by λ) for Fitting Frequency Distributions of the Number of Exceedances by: (1) Poisson Distribution, and (2) Mixed Poisson Distribution, for N = 25, 50, 100 and 1000, for the Boise River

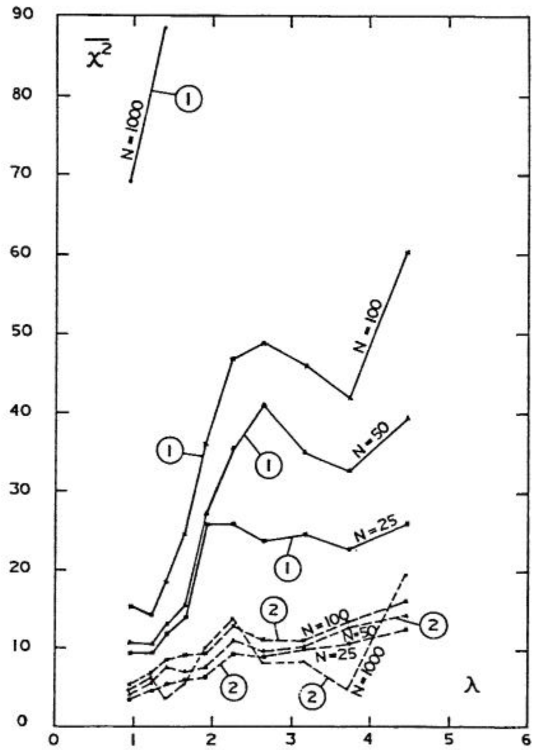

Fig. 7-32 Variation of Average Chi-Square $\overline{\chi}^2$ with Truncation Level (Expressed by λ) for Fitting Frequency Distributions of the Number of Exceedances by: (1) Poisson Distribution, and (2) Mixed Poisson Distribution, for N = 25, 50, 100 and 1000, for the Powell River

For both study cases, the mixed Poisson distribution gives significant improvements in goodness of fit, especially if applied for large N. In case of the Boise River, the ratio $R_{m,v}$ of mean to variance of frequency distributions of η decreases with an increase of Q_b. For example, for N = 1000 years, $R_{m,v}$ varies from 1.767 for the lowest Q_b to 0.784 for the highest Q_b. Hence, for the range of high Q_b which $R_{m,v}$ less than unity, the mixed Poisson distribution is well applicable.

In case of the Powell River, $R_{m,v}$ increases with an increase of Q_b. For N = 100 years, $R_{m,v}$ varies from 0.662 for the lowest Q_b to 0.749 for the highest Q_b. The mixed Poisson distribution is well applied throughout the range of Q_b considered.

7.6 Distribution of the Magnitude of Exceedances

Similar to the case of the distribution of η, the following is the test on how the mixed exponential distribution improves the goodness in fitting the frequency distributions of the magnitude of exceedances ξ_ν. The procedure and data used are the same as for the case of distribution of η. The numbers of class intervals are 9,12,15 and 20, for N = 25,50,100 and 1000, respectively.

The average values of the chi-square statistic for all small samples for exponential and mixed exponential distributions are shown in Fig. 7-33 for the Boise River and Fig. 7-34 for the Powell River, respectively. In case of the Boise River, skewness coefficients of the frequency distributions of ξ_ν for various Q_b are not very large. The mixed exponential distribution could be applied only in a few cases. For N = 1000, the skewness coefficient varies from 1.699 for the lowest Q_b to 2.109 for the highest Q_b. However, the results of the study daily flow of 17 stations show that the mixed exponential distribution could be well applied for cases of skewness coefficient greater than two. Therefore, for N = 1000, the mixed exponential distribution can be applied for the range of high Q_b. For N = 25,50 and 100, improvements of goodness of fit of this distribution are relatively small, since it is applied only for a few cases out of n small samples.

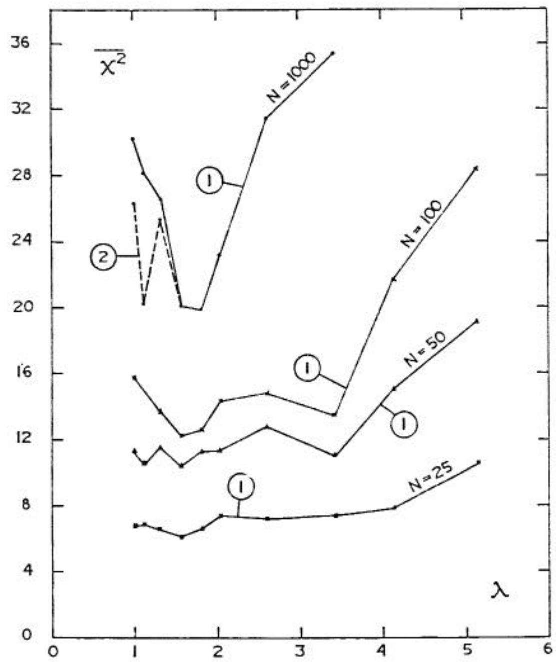

Fig. 7-33 Variation of Average Chi-Square $\overline{\chi}^2$ with Truncation Level (Expressed by λ) for Fitting Frequency Distributions of the Magnitude of Exceedances by: (1) Exponential Distribution, and (2) Mixed Exponential Distribution, for N = 25, 50, 100 and 1000, for the Boise River

In case of the Powell River, the mixed exponential distribution can be applied throughout the range of Q_b with significant improvements in goodness of fit, especially if applied for large N. The skewness coefficients of the frequency distributions of ξ_ν for various Q_b are greater than two. For N = 1000, the skewness coefficient varies from 3.504 for the lowest Q_b to 2.966 for the highest Q_b. Hence, the mixed exponential distribution can be well applied throughout the range of Q_b.

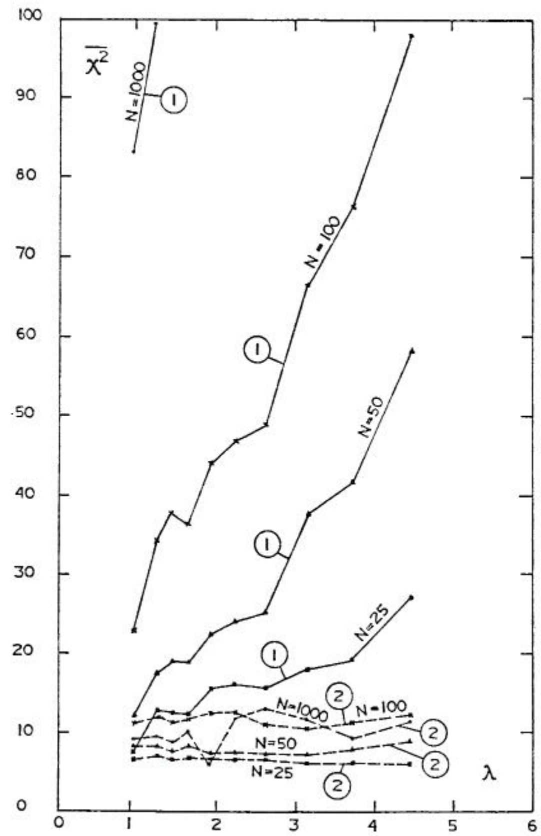

Fig. 7-34 Variation of Average Chi-Square $\overline{\chi}^2$ with Truncation Level (Expressed by λ) for Fitting Frequency Distributions of the Magnitude of Exceedances by: (1) Exponential Distribution, and (2) Mixed Exponential Distribution, for N = 25, 50, 100 and 1000, for the Powell River

7.7 Dependence between the Magnitude of the Largest Exceedance and the Number of Exceedances

One of the assumptions used in deriving the probability distribution of the largest exceedance in a year is that $\{\xi_\nu\}_1^\infty$ are independent of η, as given in Section 5.5. This does not mean that the magnitude of the largest exceedance χ is also independent of η. Instead, the magnitude of the largest exceedance in a year still depends on η of that year.

Assuming that $\{\xi_\nu\}_1^\infty$ are independently identically distributed with common distribution function H(x) represented by the exponential distribution with parameter β, and the distribution of η is represented by the Poisson distribution with parameter λ. Hence, the distribution of the largest exceedance, given that $\eta = k$, is identical to the distribution of $Z_1 + Z_2 + \ldots + Z_k$, where Z_1 is the minimum of the k exponential distributions with parameter β, which is also the exponential distribution with parameter β/k, Z_2 is the minimum of the k-1 exponential distributions with parameter β, which is also the exponential distribution with parameter $\beta/(k-1)$, and so on. Therefore, the expectation of the largest exceedance, χ, given that $\eta = k$, is

$$E[\chi|\eta=k] = \sum_{j=1}^{k} E[Z_j]$$

$$= \beta[\frac{1}{k} + \frac{1}{(k-1)} + \dots + \frac{1}{2} + 1] \qquad (7\text{-}6)$$

and the variance of χ is

$$\mathrm{var}[\chi|\eta=k] = \sum_{j=1}^{k} \mathrm{var}[Z_j]$$

$$= \beta^2[\frac{1}{k^2} + \frac{1}{(k-1)^2} + \dots + \frac{1}{2^2} + 1] \qquad (7\text{-}7)$$

since $Z_1, Z_2, \dots, Z_k$ are independent random variables according to the property of the lack-of-memory.

Equation 7-6 shows how the expected value of χ depends on η. The first group of N = 1000 for the case of partial flood series of generated daily flows is used for investigation the dependence between the expected value of χ and η. The estimate of β is obtained from the long sample of 1000 years for each Q_b. Let m denote the number of years out of 1000 years that $\eta = k$, for a given k. For each of m years that $\eta = k$, the largest exceedance is obtained. The average value of the largest exceedances for all m years divided by β is also obtained for a given $\eta = k$. This result can be compared with the assumed theoretical values as shown by Eq. 7-6.

For the Boise River, the relationships between the average value of the largest exceedance divided by β and η, for Q_b = 2870, 4000, 5000 and 6000 cfs, are shown in Fig. 7-35. In case of the Powell River, the results of Q_b = 5000, 6000, 6500, and 7500 cfs, are shown in Fig. 7-36. Both figures also include

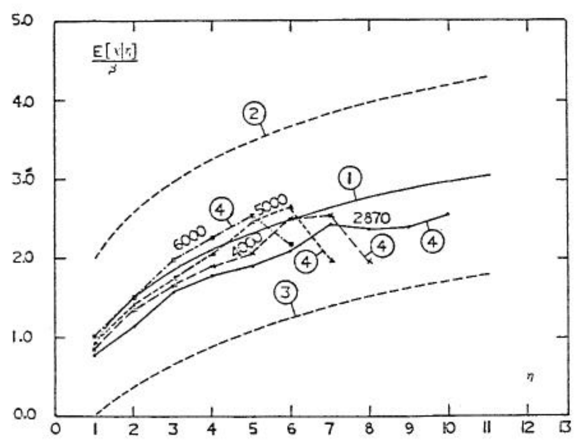

Fig. 7-35 Relationships between the Average Value of the Largest Exceedance Divided by β, $E[\chi|\eta]/\beta$, and the Number of Exceedances in a Year, η, for the Boise River, with: (1) Assumed Theoretical Values, Eq. 7-6, (2) and (3) Upper and Lower Limits, Eq. 7-8, and (4) Computed Values, for Q_b = 2870, 4000, 5000 and 6000 cfs

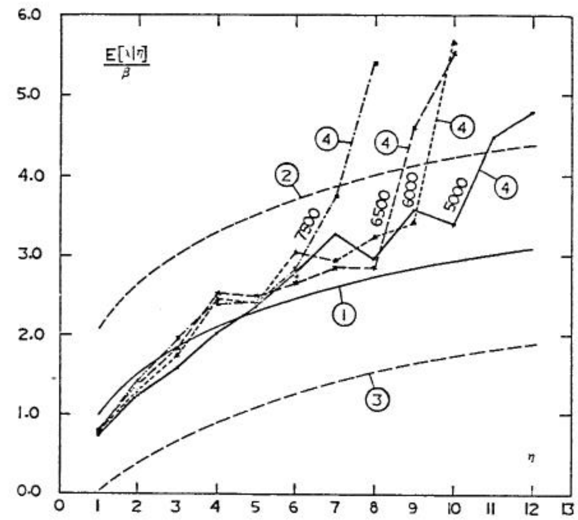

Fig. 7-36 Relationships between the Average Value of the Largest Exceedance Divided by β, $E[\chi|\eta]/\beta$, and the Number of Exceedances in a Year, η, for the Powell River, with: (1) Assumed Theoretical Values, Eq. 7-6, (2) and (3) Upper and Lower Limits, Eq. 7-8, and (4) Computed Values, for Q_b = 5000, 6000, 6500 and 7500 cfs

the theoretical relationships between $E[\chi|\eta=k]/\beta$ and $\eta = k$, as well as the upper and the lower limits (one standard deviation from the mean) which are expressed as

$$\frac{E[\chi|\eta=k]}{\beta} \pm \frac{\sqrt{\mathrm{var}[\chi|\eta=k]}}{\beta} \qquad (7\text{-}8)$$

with $E[\chi|\eta=k]$ and $\mathrm{var}[\chi|\eta=k]$ given by Eqs. 7-6 and 7-7, respectively.

Figures 7-35 and 7-36 show that the magnitude of the largest exceedance tends to depend on the number of exceedances in a year even though $\{\xi_\nu\}_1^\infty$ are independent of η, especially in the range of small η. For large η, the results obtained by using the generated data are not conclusive since the number of years that $\eta = k$, or the value of m, is very small.

7.8 Comparison of Goodness-of-Fit Statistics in Fitting Frequency Distributions of the Largest Exceedance in Using Both the Developed and Commonly Assumed Models

The developed probability distribution of the largest exceedance in a year is obtained from Eq. 5-6, with the distribution of the number of exceedance in a year either the mixed Poisson distribution or the Poisson distribution, the ξ_ν distribution either the mixed exponential distribution or the exponential distribution, as the case will be. For the commonly assumed model, this probability distribution is obtained from Eq. 3-83, with the distribution of η is assumed only Poissonian, the ξ_ν distribution is assumed exponential. In order to test how this developed model improves the goodness of fit of frequency distributions of the largest exceedance, the two models were fitted to frequency distributions of the largest exceedance obtained from the first group of

Table 7-9. Chi-Square Test Statistics for Frequency Distributions of the Largest Exceedance, for the Commonly Assumed and Developed Models, and for the Boise River

Truncation Level	λ	Commonly Assumed Model (Eq. 3-83)		Developed Model (Eq. 5-6)			
		Chi-Square Statistic		Chi-Square Statistic			
		Computed Value	Critical Value	Computed Value	Critical Value	Type of Distribution of η	Type of Distribution of ξ_ν
2870	4.973	87.96	27.6	87.96	27.6	P	E
3500	3.959	46.16	27.6	46.16	27.6	P	E
4000	3.279	30.87	27.6	30.87	27.6	P	E
4500	2.514	33.70	27.6	24.58	25.0	MP	E
5000	1.975	38.30	27.6	22.15	25.0	MP	E
5250	1.725	29.83	27.6	20.37	25.0	MP	E
5500	1.488	34.97	27.6	19.18	25.0	MP	E
5750	1.299	31.36	27.6	17.77	22.4	MP	ME
6000	1.120	32.14	27.6	14.95	22.4	MP	ME
6250	0.976	36.21	27.6	18.74	22.4	MP	ME

Note: P = Poisson; MP = Mixed Poisson; E = Exponential; ME = Mixed Exponential.

Table 7-10. Chi-Square Test Statistics for Frequency Distributions of the Largest Exceedance, for the Commonly Assumed and Developed Models, and for the Powell River

Truncation Level	λ	Commonly Assumed Model (Eq. 3-83)		Developed Model (Eq. 5-6)			
		Chi-Square Statistic		Chi-Square Statistic			
		Computed Value	Critical Value	Computed Value	Critical Value	Type of Distribution of η	Type of Distribution of ξ_ν
4500	4.808	774.681	27.6	22.171	22.4	MP	ME
5000	3.995	604.238	27.6	18.595	22.4	MP	ME
5500	3.403	511.105	27.6	17.317	22.4	MP	ME
6000	2.840	360.062	27.6	20.465	22.4	MP	ME
6500	2.406	260.850	27.6	15.843	22.4	MP	ME
7000	2.080	224.111	27.6	10.005	22.4	MP	ME
7500	1.800	188.086	27.6	11.814	22.4	MP	ME
8000	1.561	170.503	27.6	11.965	22.4	MP	ME
8500	1.363	152.042	27.6	11.642	22.4	MP	ME
9500	1.018	96.373	27.6	8.999	22.4	MP	ME

Note: P = Poisson; MP = Mixed Poisson; E = Exponential; ME = Mixed Exponential.

partial flood series of N = 1000 years of the long generated sample of 2000 years. The chi-square statistic is used to test the goodness of fit. Results of the computed and the 95 percent critical value of chi-square statistics with 20 class intervals for various Q_b and for the two models, are given in Table 7-9 in case of the Boise River, and in Table 7-10 for the Powell River. Each table also gives, for each Q_b, the type of the best distributions for η, and for ξ_ν, used in the developed model.

The developed model represents an improvement in the goodness of fit in comparison with the commonly assumed model, especially in case of the Powell River. The chi-square values for the commonly assumed model are very large in the range of low Q_b. However, the developed model seems to improve significantly the goodness of fit in this range of Q_b.

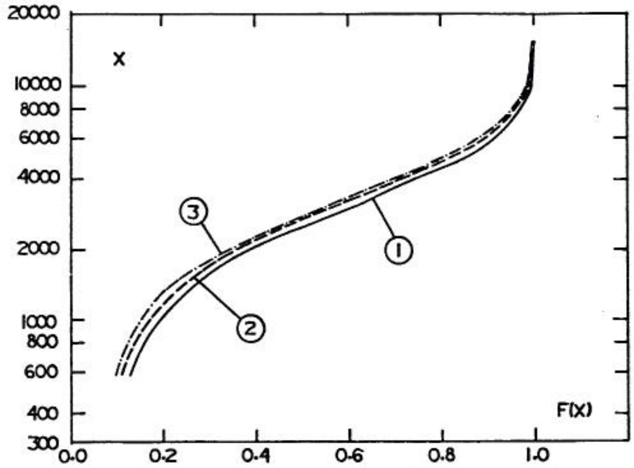

Fig. 7-37 Cumulative Distributions of the Largest Exceedances in a Year with: (1) Observed Frequency Distribution, (2) Fitted Distribution by the Developed Model, and (3) Fitted Distribution by the Commonly Assumed Model, Q_b = 5000 cfs, for the Boise River

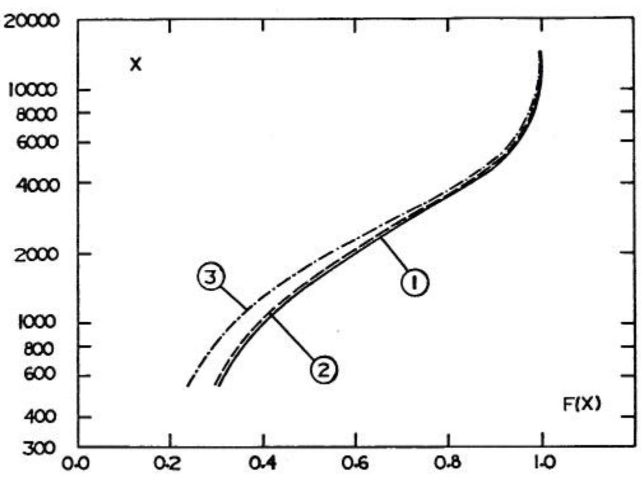

Fig. 7-38 Cumulative Distributions of the Largest Exceedance in a Year with: (1) Observed Frequency Distribution, (2) Fitted Distribution by the Developed Model, and (3) Fitted Distribution by the Commonly Assumed Model, Q_b = 5500 cfs, for the Boise River

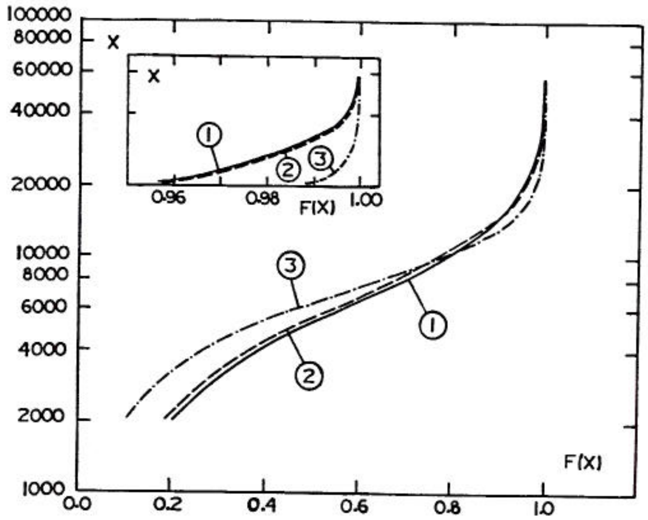

Fig. 7-39 Cumulative Distributions of the Largest Exceedance in a Year with: (1) Observed Frequency Distribution, (2) Fitted Distribution by the Developed Model, and (3) Fitted Distribution by the Commonly Assumed Model, Q_b = 5000 cfs, for the Powell River

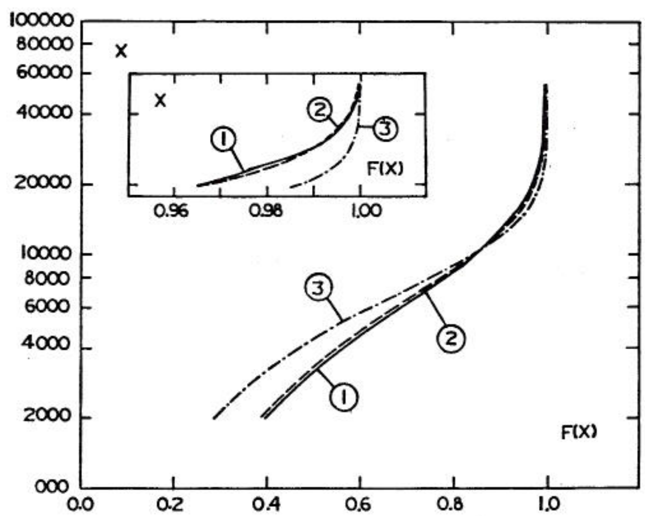

Fig. 7-40 Cumulative Distributions of the Largest Exceedances in a Year with: (1) Observed Frequency Distribution, (2) Fitted Distribution by the Developed Model, and (3) Fitted Distribution by the Commonly Assumed Model, Q = 7000 cfs, for the Powell River

The fitted distribution functions of the largest exceedance in a year, based on both models, as well as the corresponding frequency distributions of the largest exceedance are shown in Figs. 7-37 and 7-38 for Q_b = 5000 and 5500 cfs, respectively, in case of the Boise River. In case of the Powell River, the results are shown in Figs. 7-39 and 7-40 for Q_b = 5000 and 7000 cfs, respectively. These figures show that the distribution function of the largest exceedance obtained by means of the developed model has a better agreement with the observed frequency distribution of the largest exceedance than the corresponding distribution function obtained by means of the commonly assumed model. In the range of high return periods, the developed model predicts well the flood value Q(T) for a given T.

Chapter VIII
CONCLUSIONS

The topics investigated in this study belong basically into three areas: (i) development of the partial flood series model; (ii) development of the model for generation of daily flow series; and (iii) comparison of efficiency of estimates of annual flood peaks for given return periods by using annual and partial flood peak series.

8.1 Development of Partial Flood Series Model

Conclusions drawn from the development of the partial flood series model are:

(1) Either the mixed Poisson or Poisson distribution have the best fit, among all the considered discrete distributions, to frequency distributions of the number of exceedances per year;

(2) Either the mixed exponential or exponential distribution have the best fit, among all the considered continuous distributions, to frequency distributions of the magnitude of exceedances;

(3) In the range of truncation levels studied for partial flood series, with the average number of exceedances per year varying from one to four, the dependence of successive exceedances is not significant. When the truncation level is relatively low, the dependence may not be negligible, increasing with a decrease of the truncation level; and

(4) The series of annual flood peaks can be considered as approximately independent.

8.2 Development of Model for Generation of Daily Flow Series

The mathematical model, developed for generation of samples of daily flows, and based on statistics of both the daily flow series and annual flood peak series, leads to these conclusions:

(1) The generated samples of daily flows have properties close to corresponding properties of historic daily flow series; and

(2) The generated samples reproduce well the extremes, so that these samples can be used for the study of properties of flood peaks.

8.3 Comparison of Efficiency of Using Annual and Partial Flood Series

The study of generated long samples of daily flows, used to investigate the efficiency of using annual and partial flood series, leads to these conclusions:

(1) Estimates of annual flood peaks of given return periods from the partial flood series have smaller sampling variances than the corresponding estimates from the annual flood series, when the average number of exceedances per year in partial flood series is at least 1.65 for the exact theoretical approach, and at least 1.50 for the approximate theoretical approach. The conclusion in case of the exact theoretical approach is similar to that concluded by Cunnane (1973);

(2) Ratios of sampling variances of estimated annual flood peaks in case of exact theoretical and approximate theoretical approaches do not depend on the sample size;

(3) In case of the empirical approach, the sampling variance of annual flood peaks estimated from the partial flood series is smaller than the corresponding sampling variance of annual flood series for the range of investigated return periods, when the average number of exceedances in partial flood series is at least 1.95 for sample sizes 10-25, and somewhat larger than 1.95 for larger sample sizes.

(4) For each flood series and for each sample size, the sampling variance of estimated annual flood peaks for given return periods by using the theoretical approach is smaller than the corresponding sampling variance of estimates in the empirical approach, with differences of these sampling variances increasing with an increase of the sample size, and being greater for partial flood series for a given sample size than for annual flood series;

(5) Comparison of sampling mean square errors of estimates of annual flood peaks for given return periods in case of use of the annual and partial flood series depends on the corresponding population flood peaks, if assumed to be known;

(6) Assumed population flood peaks are sensitive to the ratio of sampling mean square errors, and in such a way that if flood peaks are assumed to be estimated from the generated samples of annual flood series, the sampling variances of estimates from partial flood series decrease with a decrease of the truncation level, while the bias in estimates tends to increase with a decrease of the truncation level, especially in the range of low values of truncation levels;

(7) When the model of partial flood series is developed with assumptions for its derivation supported by data for low truncation levels, the partial flood series is more efficient or more useful in estimating annual flood peaks than the annual flood series, especially in case of small sample sizes;

(8) By using the observed and generated samples of daily flows, the partial flood series model, developed in this study (Eq. 5-6), gives a better fit of frequency distributions of the largest exceedance than the commonly assumed partial flood series model (Eq. 3-83), especially for low truncation levels and for rivers with highly fluctuating daily flows.

REFERENCES

Bardwell, G. E., and Crow, E. L., 1964, A Two-Parameter Family of Hyper-Poisson Distributions, Journal of the American Statistical Association, Vol. 59, pp. 133-141.

Beard, L. R., 1954, Estimation of Flood Probabilities, Proc. ASCE, Vol. 80, pp. 438-459.

Beard, L. R., 1967, Simulation of Daily Streamflow, Proc. The International Hydrology Symposium, Fort Collins, Colorado, Vol. 1, Paper No. 78, pp. 624-632.

Benson, M. A., 1968, Uniform Flood Frequency Estimating Methods for Federal Agencies, Water Resources Research, Vol. 4, No. 5, pp. 891-908.

Borgman, L. E., 1963, Risk Criteria, Proceedings of the American Society of Civil Engineers, Journal of the Waterways and Harbors Division, Vol. 89, No. WW3, pp. 1-35.

Burges, S. J., Lettenmaier, D. P., and Bates, C. L., 1975, Properties of the Three-Parameter Log Normal Probability Distribution, Water Resources Research, Vol. 11, No. 2, pp. 229-235.

Chow. V. T., 1950, Discussion on Annual Floods and the Partial Duration Flood Series, by W. B. Langbein, Transactions, Am. Geophys. Union, Vol. 31, pp. 939-941.

Chow, V. T., 1954, The Log-Probability Law and Its Engineering Applications, Proceedings of the American Society of Civil Engineers, Vol. 80, pp. 1-25.

Cohen, A. C., 1963, Estimation in Mixtures of Discrete Distributions; Classical and Contagious Discrete Distributions, Edited by Patil, G. P., Statistical Publishing Society, Calcutta, India, pp. 373-378.

Crow, E. L., and Bardwell, G. E., 1963, Estimation of the Parameters of the Hyper-Poisson Distributions; Classical and Contagious Discrete Distributions, Edited by Patil, G. P., Statistical Publishing Society, Calcutta, India, pp. 127-140.

Cunnane, C., 1973, A Particular Comparison of Annual Maxima and Partial Duration Series Methods of Flood Frequency Prediction, Journal of Hydrology, Vol. 18, pp. 257-271.

Foster, H. A., 1924, Theoretical Frequency Curves and Their Applications to Engineering, Transactions of the American Society of Civil Engineers, Vol. 87, pp. 142-173.

Green, N. M. D., 1973, A Synthetic Model for Daily Streamflow, Journal of Hydrology, Vol. 20, pp. 351-364.

Guerrero-Salazar, P., and Yevjevich, V., 1975, Analysis of Drought Characteristics by the Theory of Runs, Hydrology Paper No. 80, Colorado State University, Fort Collins.

Gumbel, E. J., 1941, The Return Period of Flood Flows, Annals of Mathematical Statistics XII, pp. 163-190.

Gupta, V. K., Duckstein, L., and Peebles, R. W., 1976, On the Joint Distribution of the Largest Flood and Its Time of Occurrence, Water Resources Research, Vol. 12, No. 2, pp. 295-304.

Hall, M. J., and O'Connell, P. E., 1972, Time Series Analysis of Mean Daily River Flows, Water and Water Engineering, Vol. 76, pp. 125-133.

Hardison, C. H., 1974, Generalized Skew Coefficients of Annual Floods in the United States and Their Applications, Water Resources Research, Vol. 10, No. 4, pp. 745-752.

Jenkinson, A. F., 1969, Statistics of Extremes, Estimation of Maximum Floods, Chapter 5, WMO Technical Note 98, pp. 193-227.

Kelman, J., 1977, Stochastic Modeling of Intermittent Daily Hydrologic Series, Ph.D. Dissertation, Colorado State University, Fort Collins.

Kimball, B. F., 1946, Sufficient Statistical Estimation Functions for the Parameters of the Distribution of Maximum Values, Annals of Math. Stat., Vol. 17, pp. 299-309.

Kirby, W., 1969, On the Random Occurrence of Major Floods, Water Resources Research, Vol. 5, No. 4, pp. 778-784.

Klemes, V., 1976, Comment on "Regional Skew in Search of a Parent," by N. C. Matalas, J. R. Slack, and J. R. Wallis, Water Resources Research, Vol. 12, No. 6, p. 1325.

Kottegoda, N. T., 1972, Stochastic Five Daily Stream Flow Model, Proceedings of the ASCE, Journal of the Hydraulics Division, Vol. 98, HY9, pp. 1469-1485.

Langbein, W. B., 1949, Annual Floods and the Partial-Duration Flood Series, Transactions, American Geophysical Union, Vol. 30, No. 6, pp. 879-881.

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

Matalas, N. C., Slack, J. R., and Wallis, J. R., 1975, Regional Skew in Search of a Parent, Water Resources Research, Vol. 11, No. 6, pp. 815-826.

Mejia, J. M., and Rodriguez-Iturbe, I., 1974, Correlation Links between Normal and Log Normal Processes, Water Resources Research, Vol. 10, No. 4, pp. 689-690.

Mejia, J. M., Rodriguez-Iturbe, I., and Cordova, J. R., 1974, Multivariate Generation of Mixtures of Normal and Log Normal Variables, Water Resources Research, Vol. 10, No. 4, pp. 691-693.

Nash, J. E., and Amorocho, J., 1966, The Accuracy of the Prediction of Floods of High Return Period, Water Resources Research, Vol. 2, No. 2, pp. 191-198.

Natural Environment Research Council, The United Kingdom, 1975, Flood Studies Report, Vol. 1, Chapters 1 and 2.

Quimpo. R. G., 1967, Stochastic Models of Daily Rivers Flow Sequences, Hydrology Paper No. 18, Colorado State University, Fort Collins.

Rider, P. R., 1961, The Method of Moments Applied to a Mixture of Two Exponential Distributions, Annals of Math. Statist., Vol. 32, pp. 143-147.

Roesner, L. A., and Yevjevich, V., 1966, Mathematical Models for Time Series of Monthly Precipitation and Monthly Runoff, Hydrology Paper No. 15, Colorado State University, Fort Collins.

Rousselle, J., 1972, On Some Problems of Flood Analysis, Ph.D. Dissertation, Colorado State University, Fort Collins.

Shane, R. M., and Lynn, W. R., 1964, Mathematical Model for Flood Risk Evaluation, Proceedings of the ASCE, Journal of the Hydraulics Division, Vol. 90, No. HY6, pp. 1-20.

Shane, R. M., and Lynn, W. R., 1969, Peak Flood Discharge Confidence Limits, Proceedings of the ASCE, Journal of the Hydraulics Division, Vol. 95, No. HY1, pp. 161-173.

Singh, K. P., and Sinclair, R. A., 1972, Two-Distribution Method for Flood-Frequency Analysis, Proceedings of the ASCE, Journal of the Hydraulics Division, Vol. 98, No. HY1, pp. 29-44.

Tao, P. T., 1973, Distribution of Hydrologic Independent Stochastic Components, Ph.D. Dissertation, Colorado State University, Fort Collins.

Tao, P. T., Yevjevich, V., and Kottegoda, N. T., 1976, Distribution of Hydrologic Independent Stochastic Components, Hydrology Paper No. 82, Colorado State University, Fort Collins.

Todorovic, P., 1970, On Some Problems Involving Random Number of Random Variables, The Annals of Mathematical Statistics, Vol. 41, No. 3, pp. 1059-1063.

Todorovic, P., 1971, On Extreme Problems in Hydrology, Paper Presented at Joint Statistics Meetings, Amer. Statist. Assoc. and Inst. of Math. Statist., Colorado State University, Fort Collins.

Todorovic, P., and Zelenhasic, E., 1970, A Stochastic Model for Flood Analysis, Water Resources Research, Vol. 6, No. 6, pp. 1641-1648.

Todorovic, P., and Rousselle, J., 1971, Some Problems of Flood Analysis, Water Resources Research, Vol. 7, No. 5, pp. 1144-1150.

Todorovic, P., and Woolhiser, D. A., 1972, On the Time When the Extreme Flood Occurs, Water Resources Research, Vol. 8, No. 6, pp. 1433-1438.

Vargas-Semprun, D., 1977, On the Stochastic Modeling of Daily Streamflows, Ph.D. Dissertation, Colorado State University, Fort Collins.

Water Resources Council, Hydrology Committee, 1967, A Uniform Technique for Determining Flood Flow Frequencies, Bulletin 15, Washington, D.C.

Water Resources Council, Hydrology Committee, 1976, Guidelines for Determining Flood Flow Frequency, Bulletin 17, Washington, D.C.

Yevjevich, V., 1972a, Probability and Statistics in Hydrology, Water Resources Publications, Fort Collins, Colorado.

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

Yevjevich, V., 1972c, Structural Analysis of Hydrologic Time Series, Hydrology Paper No. 56, Colorado State University, Fort Collins.

Zelenhasic, E., 1970, Theoretical Probability Distributions for Flood Peaks, Hydrology Paper No. 42, Colorado State University, Fort Collins.

APPENDIX

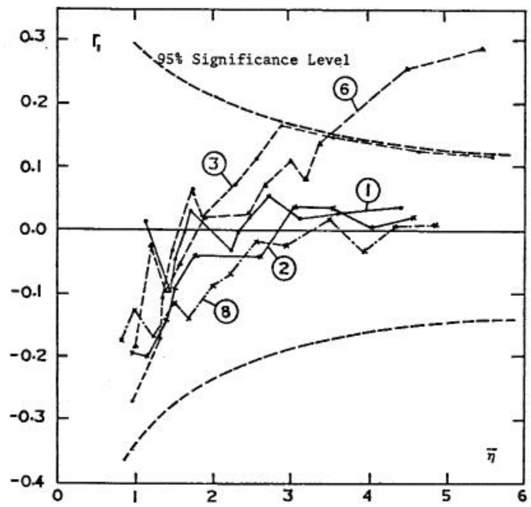

Fig. A-1. Relationship between the First-Order Serial Correlation Coefficient, r_1, of Series of Exceedances, and the Truncation Level, Q_b, Expressed as the Average Number of Exceedances per Year, $\bar{n}$, for Station Nos. 1, 2, 3, 6, and 8

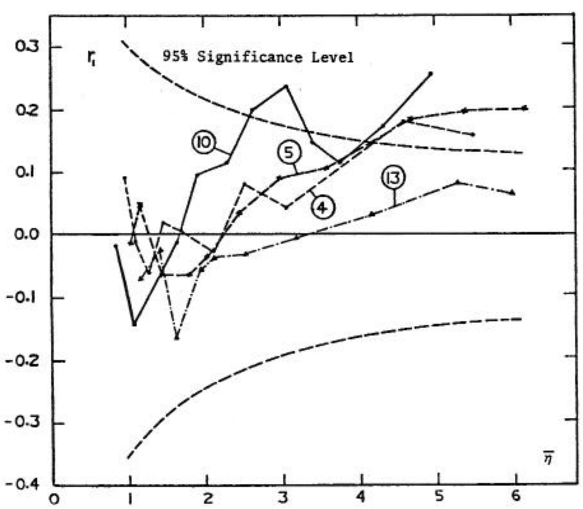

Fig. A-2. Relationship between the First-Order Serial Correlation Coefficient, r_1, of Series of Exceedances, and the Truncation Level, Q_b, Expressed as the Average Number of Exceedances per Year, $\bar{n}$, for Station Nos. 4, 5, 10, and 13

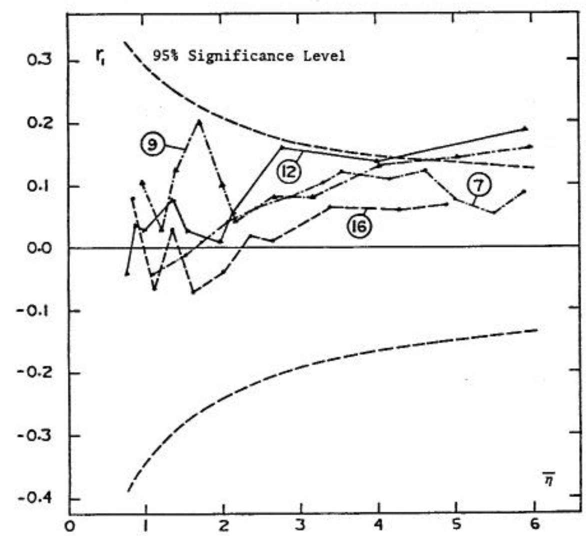

Fig. A-3. Relationship between the First-Order Serial Correlation Coefficient, r_1, of Series of Exceedances, and the Truncation Level, Q_b, Expressed as the Average Number of Exceedances per Year, $\bar{n}$, for Station Nos. 7, 9, 12, and 16

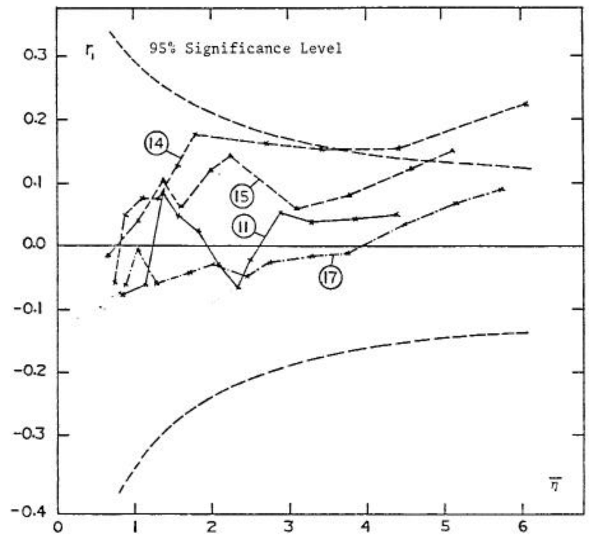

Fig. A-4. Relationship between the First-Order Serial Correlation Coefficient, r_1, of Series of Exceedances, and the Truncation Level, Q_b, Expressed as the Average Number of Exceedances per Year, $\bar{\eta}$, for Station Nos. 11, 14, 15, and 17

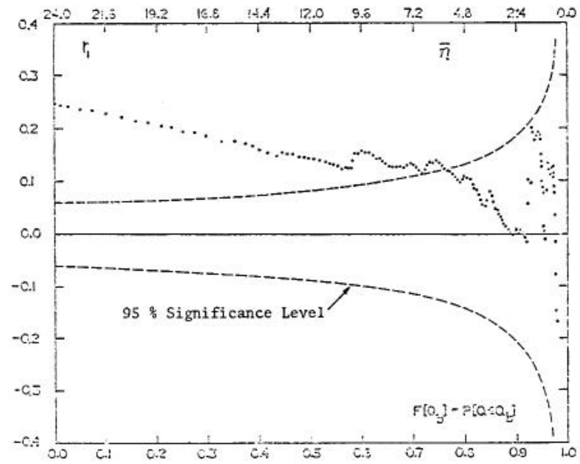

Fig. A-5. Relationship between the First-Order Serial Correlation Coefficient, r_1, of Series of Exceedances, and the Truncation Level, Q_b, (Expressed as the Average Number of Exceedances per Year, $\bar{\eta}$) as well as the Non-Exceedance Probability of Separate Flood Peaks, $F(Q_b)$, for the Whole Range of Q_b and for Station No. 9

Key Words: Floods, Flood Frequency, Flood Peaks, Annual Flood Series, Partial Flood Series, Information Content, Modeling Series, Generation of Samples.

Abstract: A method of estimating parameters of probability distributions of maximum annual flood peak by using the combination of distributions of the number and the magnitude of flood peaks for a selected truncation level is the subject matter of the paper. The 17 daily discharge series were used for testing of the method. The relationship of the goodness-of-fit statistics for selected distribution functions and the truncation level are investigated. The sequential dependence in partial and annual flood peak series were weak but showed the dependence of partial flood peak series to increase as the truncation level decreased.

Key Words: Floods, Flood Frequency, Flood Peaks, Annual Flood Series, Partial Flood Series, Information Content, Modeling Series, Generation of Samples.

Abstract: A method of estimating parameters of probability distributions of maximum annual flood peak by using the combination of distributions of the number and the magnitude of flood peaks for a selected truncation level is the subject matter of the paper. The 17 daily discharge series were used for testing of the method. The relationship of the goodness-of-fit statistics for selected distribution functions and the truncation level are investigated. The sequential dependence in partial and annual flood peak series were weak but showed the dependence of partial flood peak series to increase as the truncation level decreased.

Key Words: Floods, Flood Frequency, Flood Peaks, Annual Flood Series, Partial Flood Series, Information Content, Modeling Series, Generation of Samples.

Abstract: A method of estimating parameters of probability distributions of maximum annual flood peak by using the combination of distributions of the number and the magnitude of flood peaks for a selected truncation level is the subject matter of the paper. The 17 daily discharge series were used for testing of the method. The relationship of the goodness-of-fit statistics for selected distribution functions and the truncation level are investigated. The sequential dependence in partial and annual flood peak series were weak but showed the dependence of partial flood peak series to increase as the truncation level decreased.

Key Words: Floods, Flood Frequency, Flood Peaks, Annual Flood Series, Partial Flood Series, Information Content, Modeling Series, Generation of Samples.

Abstract: A method of estimating parameters of probability distributions of maximum annual flood peak by using the combination of distributions of the number and the magnitude of flood peaks for a selected truncation level is the subject matter of the paper. The 17 daily discharge series were used for testing of the method. The relationship of the goodness-of-fit statistics for selected distribution functions and the truncation level are investigated. The sequential dependence in partial and annual flood peak series were weak but showed the dependence of partial flood peak series to increase as the truncation level decreased.

The samples of daily flow series were generated by a refined model. Two station series with different water regimes are used to test the method developed.

The estimates of annual flood peaks of given return periods from the partial flood peak series showed a smaller sampling variance than the corresponding estimates from the annual flood peak series when the average number of exceedances per year in partial flood series was at least 1.65 for an exact analytical approach, and at least 1.50 for an approximate analytical approach.

The samples of daily flow series were generated by a refined model. Two station series with different water regimes are used to test the method developed.

The estimates of annual flood peaks of given return periods from the partial flood peak series showed a smaller sampling variance than the corresponding estimates from the annual flood peak series when the average number of exceedances per year in partial flood series was at least 1.65 for an exact analytical approach, and at least 1.50 for an approximate analytical approach.

The samples of daily flow series were generated by a refined model. Two station series with different water regimes are used to test the method developed.

The estimates of annual flood peaks of given return periods from the partial flood peak series showed a smaller sampling variance than the corresponding estimates from the annual flood peak series when the average number of exceedances per year in partial flood series was at least 1.65 for an exact analytical approach, and at least 1.50 for an approximate analytical approach.

The samples of daily flow series were generated by a refined model. Two station series with different water regimes are used to test the method developed.

The estimates of annual flood peaks of given return periods from the partial flood peak series showed a smaller sampling variance than the corresponding estimates from the annual flood peak series when the average number of exceedances per year in partial flood series was at least 1.65 for an exact analytical approach, and at least 1.50 for an approximate analytical approach.

LIST OF PREVIOUS 25 PAPERS

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

No. 94 Fluctuations of Wet and Dry Years, an Analysis by Variance Spectrum, by Vujica Yevjevich, August 1977.

No. 95 Mathematical Modeling of Sediment Deposition in Reservoirs, by Jose Luis Lopez S., July 1978.

No. 96 Effects of Surface Roughness and its Spatial Distribution on Runoff Hydrographs, by Yao-Huang Wu, Vujica Yevjevich and David A. Woolhiser, October 1978.