RN-5

PARTIAL DURATION SERIES MODELS

SATISH CHANDRA
DIRECTOR

STUDY GROUP

S M SETH

N K GOEL

NATIONAL INSTITUTE OF HYDROLOGY
JAL VIGYAN BHAVAN
ROORKEE - 247667

1984-85

CONTENTS

ABSTRACT

The basic requirement of flood frequency analysis is to know the probability with which a flow Q is exceeded during a stated design life of a particular project. The partial duration series models provide an alternative method of estimating floods from the number and magnitude of exceedances over a suitable threshold. The other models are annual peak series models and time series models.

Various partial duration series models have been proposed in literature. Most of the models assume Poisson distribution for the distribution of number of exceedances in a year and exponential distribution for the magnitude of exceedances. These models mainly differ from each other in the way the number of peaks every year over a threshold are treated. In the simplest model for partial duration series the variation between years and between seasons, in the number of peaks exceeding the threshold is ignored and a constant number of exceedances is assumed to occur each year. There are models which are capable of accomodating varying number of occurrences of events within the year and different distribution of peak magnitudes during different seasons of the year.

In the view note a theoretical comparison of annual maximum series models and partial duration series models

have been presented. The literature review reveals that the assumptions regarding distribution of flood magnitudes are important as they influence the estimates of the floods of high return periods very significantly.

1.0 INTRODUCTION

One of the major problems faced in hydrology is the estimation of design flood from fairly short data. If the length of data is more, say 1000 years or more than the same data can be used to realistically estimate the design flood but generally the length of data available is very less. As such the available sample data is used to fit frequency distribution, which in turn is used to extrapolate from recorded events to design events.

Flood frequency analysis is needed to relate the rarity of flood to its magnitude. The three approaches for flood frequency analysis are based on the analysis of the (i) time series, (ii) the peaks over a threshold or partial duration series and (iii) the annual maximum series. In most of the cases annual maximum series is used for flood frequency analysis. The first uses the whole flow hydrograph, the second only those peaks which exceed an arbitrary threshold and the third the highest peak in each year of record.

The Time Series Model

In this model the flow hydrograph is considered as a time series of daily flows. The time series of mean daily flows closely represents instantaneous peak flows on large catchments but on small flashy catchments ,this would not be necessarily so as flood peaks could be smoothed out by daily averaging.

If Q(t) is a flow on day t, a time series model may be written as the sum of deterministic and stochastic components.

Flow on day t = deterministic component
+ stochastic component
= (Trend component + periodic component)
+ stochastic component ...(1)

In such a situation this type of model allows both estimation of parameters and model formulation to proceed together through the three components, beginning with trend and finishing with the stochastic effect. Examples of the use of such a model for mean daily flows have been given by Quimpo (1967) on United States data and by Hall and O'Connell (1972) on Brithish data. A specific example of a time series model is the ' shot noise model'. In shot noise model flows are treated as a series of impulses and decays. The theory of shot noise model was developed by G.Weiss of Imperial College, London, working under the direction of Professors Cox and O'Donnel (NERC 1975).

Partial Duration Series Model

The partial duration series model concerns the distribution of the number and magnitude of peak flows that exceed a threshold such as q_o in figure 1 which shows part of a continuous record of flow in a river. Such peak flows are said to consitute a partial duration series. The threshold level q_o may be raised or lowered so as to involve a desirable number of peaks per year (λ). The value of λ is generally in the range from 2 to 5.

In order to make the analysis tractable, it is assumed that the individual peaks $q_1, q_2, q_3, \ldots\ldots$, etc., represent independent hydrometeorological events and these are not serially correlated. The peaks such as q_2' and q_4' which do not have definite ascensions and recessions and which seem to be associated with q_2 and q_4 respectively are not considered.

The distribution of interevent time ζ_i, i = 1,2,3.... between successive exceedances is also important. The joint distribution of the ζ_i values specify a stochastic process which is found by the times of peak flows exceeding q_o. Most of the models proposed in literature assume Poisson distribution for number of exceedances and exponential distribution for magnitude of exceedances.

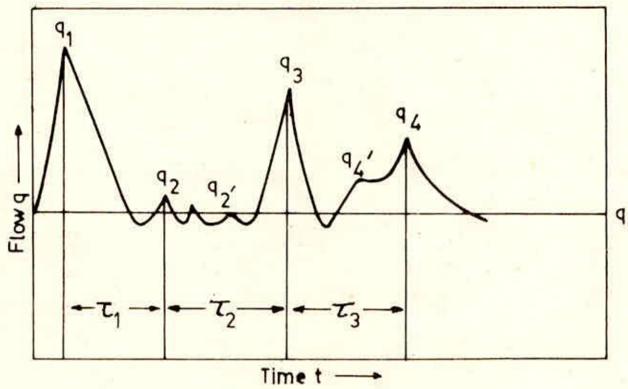

FIGURE 1 - PARTIAL DURATION SERIES MODEL

Annual Maximum Model

This is a special case of a time series model in which the unit of time is one year and the flow representing that time is the highest flow during the year. In practice this time series is statistical rather than stochastic since there is no dependence between successive peaks which may be considered as identically and inaependentallv distributed. Annual maximum approach has gained popularity because of the theory of extremes by Gumbel (1941-1945).

In annual maximum models the distribution of annual maximum peaks is required. As an example Gumbel found empirical support for the hypothesis that annual maximum peaks are distributed according to the extreme value type 1 distribution (known as the Gumbel distribution). Log Pearson type III distribution has been recommended by WRC (1967) and accepted by United States Federal Agencies. In India flood frequency analysis is generally carried out using annual maximum peak series. For small return periods (less than 5 to 10 years) the partial duration series models would give better estimate as the information contained in the sample is more. To study suitability and capability of partial duration series models, relevant literature has been reviewed.

Relation Between Time Series Model, Partial Duration Series Model and Annual Maximum Model

The time series model, partial duration series model and annual maximum model can be related on the basis of value of Q for given T in the parent process and the value of the expressionn for Q(T) evaluated in each of the three models which approximate it. The models can be related other way also i.e. by relating T given by various models for a particular Q. Each of these values T_{TS}, T_{PDS}, T_{AM} can be deduced from knowledge of the parent stochastic process. In T_{AM} each year is condensed into a single time unit. In peaks over a threshold model time is measured on a continuous scale and a value of T can be interpreted in the same way as in the parent process. In the time series model time is also measured in discrete steps but because these are small, their use has no practical effect on the interpretation of return period.

For practical purposes it can be assumed that T_{PDS} and T_{TS} are equal to T. Langbein (1949) showed that when T is small T_{AM} differs appreciably from T_{PDS} and hence from T but differs by only one half year at large values of T. The difference has been illustrated in Figure 2. It is clear from the figure that in practical situations there is no need to distinguish between the values of $Q(T)$, $Q(T)_{TS}$, $Q(T)_{PDS}$ and $Q(T)_{AM}$ when T is large.

Langbein's theory which relates T_{AM} and T_{PDS} is presented in brief in the forth coming section.

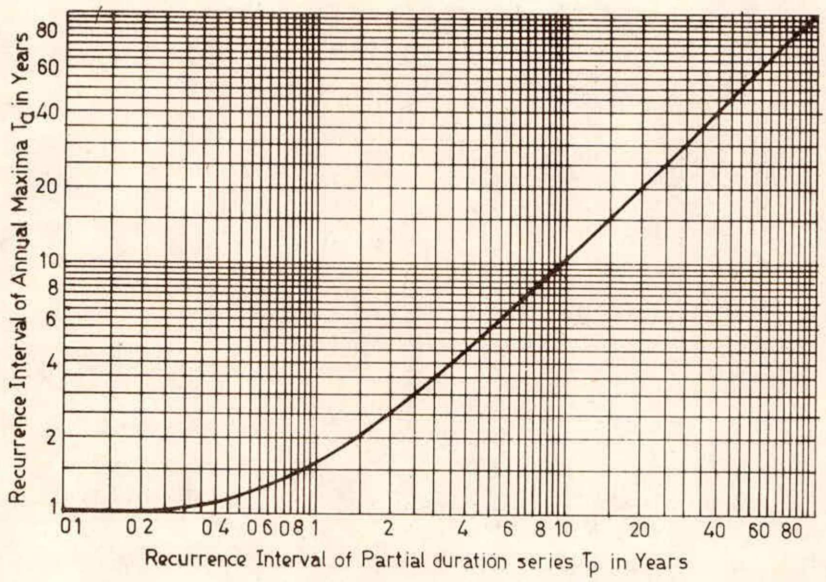

FIGURE 2 - RELATION BETWEEN RETURN PERIOD OF A GIVEN MAGNITUDE IN ANNUAL MAXIMUM AND PDS MODELS. (after Langbein 1949)

Langbein Theory

The recurrence intervals calculated from two approaches (annual maxima series and partial duration series) are not directly mutually comparable. The practitioners often face the difficulty that a design criterion given in terms of a recurrence interval of one kind cannot be interpreted as a hydrological value corresponding to that of other kind. For instance, a 5 year flood of a partial duration series and 5-year flood of annual maximum series don't coincide in magnitude. Langbein (1949) gave a solution to this problem. He derived a relation between the two recurrence intervals $T_{AM}(T_a)$ and $T_{PDS}(T_p)$ corresponding to the same event, say a flood greater than q_m as follows:

$$T_a = 1/(1-\exp(-1/T_p)) \qquad \ldots(2)$$

The following is the derivation given by Langbein.

The probability that an arbitrary sample in a partial duration series of M exceedances (over q_M in N years series) is greater than or equal to q_m(the mth largest in M exceedances) is m/M or ε/n, where n is the average number of exceedances over q_m in a year i.e.

$$n = \frac{M}{N} \qquad \ldots(3)$$

and ε is a fraction of m to N i.e.

$$\varepsilon = \frac{m}{N} \text{ or } \frac{1}{T_p} \qquad \ldots(4)$$

i.e. the inverse of the recurrence interval T_p. It follows that the probability of n events in a year being consecutively less than q_m is

$$p = (1 - \frac{\varepsilon}{n})^n \qquad \ldots(5)$$

if all M events are perfectly independent.

If $\varepsilon/N << 1$, then the probability p of above equation approaches:

$$(1 - \varepsilon/n)^n \simeq e^{-\varepsilon}$$

Accordingly, since the probability that at least one event greater than or equal to q_m occurs in a year is :

$$P_a = 1 - p \simeq 1-e^{-\varepsilon} \qquad \ldots(6)$$

The expected recurrence interval or a year having at least one such event is

$$T_a = \frac{1}{P_a} = \frac{1}{(1 - e^{-\varepsilon})}$$

$$\text{or} \quad T_a = 1/(1 - \exp(-\frac{1}{T_p})) \qquad \ldots(7)$$

which corresponds to Langbein's formula.

Chow (1950) discussed Langbein's formula and pointed out that the difference between T_a and T_p, evaluated by the relative difference $(T_a - T_p)/T_a$ is less than 5% for $T_p \geq 10$ years and greater than 10% for $T_p \leq 5$ years. Chow further writes that " in ordinary engineering practice a five percent difference is tolerable and that, the two methods give essentially, identical results for intervals greater than about ten years".

Relative Merits of Different Models

The classical dilemma in flood frequency analysis is whether to use annual maximum model or partial duration series model. Relative demerits and merits of annual maximum model and PDS model are discussed in the forth coming section.

Demerits of annual maximum model

The most frequent objection for the use of annual maximum model 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 out rank many annual floods of other years. The maximum annual discharges in dry years of some rivers in arid or semi arid regions may be so small that calling them floods may be misleading.

Another short coming of annual flood series approach is that only a small number of floods is considered. The estimate of skewness coefficient of historical flood series will not be reliable in case of annual flood series with small sample size.

Advantages of partial duration series model

(i) The partial duration series model contains more floods than annual maximum model, so the estimate of parameters of annual flood distribution from the partial flood series would be subjected to lesser uncertainty.

(ii) The theoretical expressions for annual flood distribution obtained through characteristics of partial floods have physical relevance and often are exact distributions rather than asymptotic (Viraphol et al 1978).

(iii) When the truncation level which defines a partial flood peak series is taken adequately high, the assumption of stochastic independence among individual exceedances become reasonable.

(iv) 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.

2.0 REVIEW OF LITERATURE

The earliest approaches to the estimation of future floods on a drainage basin were based upon simple empirical formulae involving the correlation of pask peak discharges with various parameters. The most popular of these parameters were the area, width and length of the basin.

Statistical methods were introduced in hydrology about sixty years ago. The main objective of these methods is to fit theoretical distributions to flood data. The mean, the standard deviation and the coefficient of skewness of the flood magnitudes are used to fit the parameters of the distribution function. The work on partial duration series model started with the theory of Langbein (1949).

Empirical formulae and statistical methods have been reviewed briefly in sections 2.1 and 2.2. The extensive review and suitability and applicability of partial duration series model is given in section 2.3 and 2.4.

2.1 Empirical Formulae

Formulae for the maximum expected flood involving drainage area only are of the general type

$$Q = C\,A^n \qquad \ldots(8)$$

where, C is a coefficient depending upon the characteristics of the drainage basin, A the drainage area in square miles, n a constant varying from 0.5 to 1.0 and Q is the flood flow rate in cusecs (Benson, 1962).

Some empirical formulae also use length of the drainage basin, the storm rainfall in inches and coefficients related to regional and drainage basin characteristics alongwith drainage area.

The rational method used in the design of storm water drains for small area is

$$Q = C\, i\, A \qquad \ldots (9)$$

in which C is the runoff coefficient varying from 0.1 to 0.8, i is the intensity of rainfall in inches per hour, A is the drainage area in acres and Q is the discharge in cusecs.

Richards (1955) described methods which were developed for certain streams or for certain areas. Varshney (1979) gives extensive list of empirical formulae developed for Indian catchments. Some of the important formulae are Dikens, Ryves, Inglis, Fanning, Charmier, Craig, Rhind etc.

These formulae are applied in cases where rainfall and catchment conditions are similar to those from which they were derived. The inconsistent results from the application of such empirical formulae have made their use very limited.

2.2 Statistical Methods

The first published papers which present the statistical analysis of flood data with the expressed intention of estimating the frequency of occurrence of floods of various magnitudes were by Robert E.Horton and Weston E. Fuller. Horton (1913) discussed briefly the applications of the Gaussion law of probability to frequency analysis.

Fuller (1914) introduced the fundamental idea that floods are unlimited statistical variables and that flood will always be exceeded in size some day. He gave the first comprehensive study of statistical methods applied to floods by plotting flood flows on semi-log paper. Fuller's analysis resulted in the following empirical formula:

$$Q = Q_{average} \left(1 + 0.8 \log_{10} T \right) \qquad \ldots(10)$$

where,

$Q_{average}$ is the average of the recorded annual maximum 24 hour average flow in cfs., T is the number of years in the period under consideration and Q is the average maximum flood flow for a time period of length T years. This formulae is based upon observed data without including any concept from the theory of probability.

In a systematic statistical study of floods Hazen (1914) supplemented the work of Fuller by constructing the normal and log normal probability papers for plotting floods. This replaced the estimation of parameters by linear fits on semi-log or log-log paper.

The plotting of data on probability paper requires the use of plotting positions and most of the methods for their determination are empirical. Chow (1964) gives a summary of the formulae presented by Hazen, Weibull, Beard etc.

Langbein (1949) gave the relationship between flood expectancies in the partial duration series and the probability of the corresponding flood as an annual flood provided they are derived from a time homogeneous process. The recurrence intervals in the partial duration series are smaller than in the annual flood series, but the difrerences become negligible for floods greater than about a five year recurrence interval.

Based on the annual maximum flows of 159 long-record river-measurements stations in the U.S.A. Beard (1954) concluded that with rare exceptions the logs of the annual maxima of mean daily flows are normally distributed.

Moran (1957) chose the log normal and log Pearson type III distributions to fit fifty annual values of extreme monthly flows and to show that the errors in the estimation of a flood corresponding to a given probability arises from two sources: (i) The uncertainty as to the mathematical form of the distribution (ii) The uncertainty arising from the statistical errors of estimation of the parameters of the distribution which occur because of the finiteness of the length of record.

Kendall (1959) discussed the relationship between the risk of occurrence of an event in a given period of time and its return period.

Riggs (1961) derived a relation between magnitude, design period in years and probability of not exceeding that magnitude in the design period from a cumulative frequency curve.

Nash and Amorocho (1966) showed that the extrapolation of magnitude frequency relationship obtained from finite samples is not too hazardous when the form of the frequency distribution is known for the population of floods. They made a plea for research to establish if possible the true form of the frequency distributions of floods. The magnitude corresponding to any given probability or return period can be estimated subject to error coming from two different causes: (i) failure of the universe of floods on the catchment to conform to the assumed probability distribution, (ii) sampling error due to non-representativeness of the record from which the numerical values of the parameters of the frequency distribution are estimated.

Bulletin No.13 (1966) describes the methods most commonly used by Federal Agencies for making frequency studies of runoff at individual streamflow station and provides an extensive list of applications of frequency analysis.

Benson (1968) as the Chairman of the work group on Flow Frequency Methods, Hydrology Committee, WRC, studied the most commonly used methods of flood frequency analysis and compared the results by applying these methods to a selected group of log-record representative sites in different parts of the country. He showed that there are large differences in the predicted floods when different distributions are assumed particularly for larger recurrence intervals.

In 1967 U.S.Water Resources Council (1967) adopted the

the log Pearson type III distribution (of which log-normal is a special case) to achieve standardization of flood-frequency procedures used by Federal Agencies.

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 application of Todorovic (1970).

Chander et al.(1978) applied Box-Cox transformation to flood frequency analysis and established its suitability. Hadgraft (1982) compared 12 candidate flood distributions for flood data from over 40 gauging stations located in Quanbland, Australia with records in excess of 40 years and proved that Box-Cox transformation is better than log normal and log Pearson type III.

Walter C.Boughton introduced log Boughton distribution and gave fitting procedures for the distribution (Boughton 1980,1983). Houghton (1978) introduced Wakeby distribution for modelling flood flows.

Rao D.V. (1980 a and 1980 b) evaluated log Pearson type III distribution and gave method of mixed moments for its parameters estimation.

Flood Studies Report, Vol.1, NERC(1975) is a major contribution to statistical methods.

2.3 Partial Duration Series Models

The limiting distribution of the maximum term in a sequence of independent identically distributed random indices variables ξ_1, ξ_2, ξ_3 .... ξ_{N_n} with random indices $(N_n)_1^\infty$ was first analysed by Berman (1962). Provided that sequences $(N_n)_1^\infty$ and $(\xi_n)_1^\infty$, are independent and $N_n \to \infty$ with probability 1 as $n \to \infty$ Berman has shown that the limiting distribution of the maximum term is a mixture of distributions obtained in the case of independent random variables. Barndorff Nielsen (1964) extended Berman's results to the case where $P \lim_{n \to \infty} (N_n/n) = \rho$ holds, where ρ is a positive random variable, i.e. $P(\rho > 0) = 1$.

Borgman (1961) gives a simplified technique for computing the probability that a near extreme occurrence of a physical phenomenon will exceed a selected value. Further Borgman (1963) discussed the return period concept with other risk criteria such as (i) encountered probability (ii) distribution of waiting time (iii) distribution of total damage (iv) probability of zero damage and (v) mean total damage. These are derived from three different sets of initial assumptions.

Shane and Lynn (1964) developed a probability model based on the time independent Poisson process and 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 design discharge : recurrence interval distribution, encounter probability and expected recurrence interval. Furthermore, Shane and Lynn (1969) developed confidence limits alongwith 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) considered flood peaks as the successes or exceedances in a sequence of randomly spaced Bernoulli trials representing the occurrence of hydograph peaks and he adopted an arbitrary criterion for classifying hydrograph peaks into floods and non floods.

Todorovic (1967 a, 1967 b, 1968) applied the theory of stochastic processes to a non decreasing sample function and showed its application in the application of precipitation and sediment transport. He indicated the conditions for which the Poisson, the Gamma and other distributions can be used.

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). 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 streamflow 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 streamflow by further assuming that the individual exceedances form a sequence of identically independent random variables which are represented by an exponential distribution. The theory is sufficiently general as to treat also the non-identically distributed exceedances. In addition, this theory is applicable over an arbitrary time interval of interest, such as season or year. Todorovic and Zelenhasic (1970) apply above model to 72 year record of the Susquehanna river at Wilkes-Barre, Pensylvania. They conclude that observed and theoretical results seem to agree fairly well. Application of the model on the Greenbrier river at Alderson, West Virginia has also shown good agreement between theoretical and observed results.

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.

Todorovic and Roussell (1971) extended the work of Todorovic and Zelenhasic (1970) 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. The results are applied to the 72- year record of the Greenbrier River at Alderson, West Virginia. The theoretical and observed results agree reasonably well.

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, it's time of occurrence with a selected time interval. The expression for the time of occurrence of the extreme flood obtained by Todorovic (1971) is exact.

Todorovic and Woolhiser (1972) applied the above theory to two rivers of United States. Good agreement between observed and theoretical distributions for two rivers where floods are caused by snowmelt and intense rainfall indicates that the assumptions involved are not unduly restrictive.

Gupta, Duckstein and Peebles (1976) extended the work of Todorovic and Woolhiser (1972) and developed the expression for the joint distribution function of the largest flood peak and its time of occurrence. They derived distribution function of the time of occurrence of the largest flood for the

Mississippi river, St. Paul, Minnisota and Licking river, Catawba, Kentucky. They also modified this expression, valid for the case of identically independent exceedances, to the case of independent but non-identically distributed exceedances and applied to above rivers. Guptaet al conclude that:

(i) A partial duration series is formally represented as a random sequence of iid (independently indentically distributed) random variables (flood exceedances) under certain assumptions which appear physically reasonable. This formalism leads to an exact expression for the joint distribution function of the largest exceedance and its time of occurrence within a fixed time interval which may be selected arbitrarily.

(ii) If the individual exceedances are assumed to be exponentially distributed and the number of exceedances is governed by a Poisson process, then the maximum exceedance is distributed like a double exponential distribution function. This functional form although similar to Gumbel's classical distribution function represents an exact expression where as Gumbel's distribution is an asymptotic result.

(iii) The expression for the joint distribution function is generalized after assuming that the exceedances are non-identically distributed from one season to another but are identically distributed within a season. The number and the length of each season can be selected arbitrarily. The expression by Todorovic and Roussell (1971) is shown to be the marginal of this joint distribution function.

(iv) Two rivers in the United States are analysed for the time occurrence of the maximum flood to test any possible improvement in the fit to the empirical distribution if one adopts the nonidentically distributed exceedance approach. These results don't show any conclusive improvement in the fit.

(v) The use of a flood frequency distribution derived from the partial duration series seems to have considerable operational appeal in comparison with 'curve fitting' a frequency function to the yearly peaks because (a) by estimating the parameters from the partial duration series, rather than from the yearly maximum flows, the parameters uncertainty can be reduced and (b) the flood frequency expressions admit some theoretical justification and are exact rather than asymptotic.

Todorovic (1978a) presents stochastic models of extreme flows and their application to design. He also gives various assumptions made in the formulation of models.

Todorovic (1978 b) discusses the two approaches i.e. annual flood series and partial duration series, for frequency analysis. Three stochastic models based on partial duration series are also presented. Each model depends on certain assumptions concerning properties of exceedances of base level q_o. The second and third models represent vis-a-vis the first one in the sense that they are based on less restrictive assumptions. Each exceedance is characterized by its duration and its volume. The distribution function of the largest volume in an interval of time (O, t) is also given.

This distribution function has been applied to Greenbrier river at Alderson,West Virginia. Good agreement between theoretical and observed distributions shows that the assumptions concerning the exceedances are not unduly restrictive.

Viraphol and Yevjevich (1978) estimates the probability distribution 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. This method of estimation is tested on the 17 daily streamflow series of gauging station in the United States. Five discrete and six continuous probability distribution functions were used to fit the frequency distributions of the number and magnitude of exceedances above the selected truncation level of partial flood series respectively. From the them the best fit functions are selected. The independence of partial flood series is checked.

By using the generated samples of daily flows the efficiency of estimated annual flood peaks of given return periods has been investigated by using both the annual and the partial flood peak series. The findings of Viraphol and Yevjevich (1978) are summarized below :

(i) Either the mixed Poisson or Poisson distribution have the best fit to frequency distributions of the number of exceedances per year.

(ii) Either the mixed exponential or exponential distribution have the best fit to frequency distributions of the magnitude of exceedances.

(iii) With one to four average number of exceedances per year 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.

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

(v) The generated samples of daily flows have properties close to the properties of corresponding historic daily flows.

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

(vii) 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.

(viii) 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.

(ix) When the model of partialflood 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.

(x) By using the observed and generated samples of daily flows, the partial flood series model given below

$$F\ (X) = P\ (\eta = 0) + \sum_{k=1}^{\infty} \left((H\ (X)\ )^{k} . P\ (\eta = K) \right) \quad \ldots (11)$$

gives a better fit of frequency distributions of the largest exceedance than the commonly assumed partial flood series model given by

$$F\ (x) = e^{-\lambda\ e^{-x/B}} \quad \ldots (12)$$

This is especially true for low truncation levels and for rivers with highly fluctuating daily flows.

Ashkar and Rousselle (1983 a) gives comments on the truncation used in partial flood series models. Water Resources Council (1976) defined the partial flood series as a sequence of flood events separated by at least as many days as five plus the natural logarithm of square miles of drainage area. This in addition to the arbitrary imposed requirement that the intermediate flows between two consecutive peaks must drop below 75% of the lower of two separate maximum daily flows. The purpose of these restrictions imposed on the interarrival time between two successive flood events is to minimize the stochastic dependence between flood exceedances.

Ashkar et al finally conclude that both the Poisson distribution as a model for flood frequency and exponential distribution as a model for flood magnitude once found applicable with a certain truncation level should remain so with any higher level of truncation level should remain so with any higher level of truncation also. A great degree of freedom is left to the engineer, therefore, to choose the truncation level that he finds adequate for the problem at hand without having to worry too much about the sensitivity of the obtained results: The truncation level should be sufficiently high so as to satisfy the Poisson model.

Guidelines such as those put forward by the Water Resources Council in relation to the partial flood series models should be followed with caution and the effect of any restrictions put on such models on their underlying assumptions and mathematical characteristics should always be examined.

If the Poisson and exponential distributions are to be used, then the choice of the base level should be made primarily on mathematical grounds rather than on economic or engineering considerations. A truncation level suggested by a particular set of circumstances not based on the statistical characteristics of the streamflow data may not be high enough to give a good Poisson or exponential fit. In this case one should either refer to other models or try different base levels if this is feasible.

In the application of partial flood series models of flood analysis it is occasionally observed that successive exceedances are correlated. To reduce this correlation some investigators tend to impose certain restrictions on the interarrival times of flood events in order that these events will not occur close together in bunches. For example, Water Resources Council (1976) put this kind of restriction when it defined partial flood series as flood events separated by at least as many days as 5 plus the natural logarithm of drainage area taken in square miles with the requirement that the intermediate flows should drop below 75% of the lower of the two separate maximum daily flows. Cunnane (1979) used data from 26 gauging stations on 20 catchments in Great Britain and applied restrictions in the form of ' arbitrary but consistent' rules to deal with the problem of which peaks to include or exclude when peaks occurred close together. Two neighbouring peaks were included only if (a) the flow between them dropped to less than two thirds of the earlier two and (b) the time between the peaks exceeded 3 T_p, where T_p is the average time to peak of the first five 'clean' hydrographs on the record. Ashkar and Rousselle (1983 b) show analytically how such restrictions interfere with the underlying hypothesis of the Poisson process commonly used to model flood counts, and caution against imposing restrictions that may render this simple and appealing model inapplicable.

2.4 Suitability and Applicability of Partial Duration Series Models

Cunnane (1973) gives a method for comparing the statistical efficiency of the estimate of T year flood by two different approaches i.e. annual maximum model and partial duration series model. On the basis of commonly used assumptions Cunnane shows that for return periods greater than about 10 years the annual exceedance estimate of Q(T) has larger sampling variance than the annual maxima series. Annual exceedance series is nothing but partial duration series in which threshold has been chosen in such a way that series contains only N (where N is the number of years in a record) floods. It consists of the N largest floods in the record. He also shows that for same range of return periods the partial duration series estimate of Q (T) has smaller sampling variance than the annual maximum series estimate only if partial duration series contains at least 1.65 N items, where N is the number of years of record.

Calenda, Petacia and Togna (1977) give theoretical probability distribution of critical hydrologic events by the partial duration series method. They conclude that partial duration series method should be employed when high frequency events are involved. In this case the Poisson model may lead to a substantial underestimate of the mean recurrence intervals. They suggest the use of Polya stochastic process instead of Poisson model. Polya stochastic process implies that each occurrence probability depends on previous events. This process yields a negative binomial distribution of the number of exceedances in a time interval.

Valadares Tavares and Evaristo Da Silva (1983) evaluates the relative efficiency of annual maxima method and partial duration series method considering separately independent and autocorrelated exceedances and find

(i) The estimation variance of the flood magnitude associated with a given return period can be under estimated (or over estimated) by the approximation given by Cunnane (1973) if the average number of exceedances λ is higher (or smaller) than 2.

Cunnane (1973) gives the following expression for the variance of Q(T)

$$VAR\ (Q'_T)_{PD} = \alpha'^2\ N^{-1}\ (1 + (\ln \lambda + Z_T)^2/\lambda \quad ...(13)$$

where,

$VAR\ (Q'_T)_{PD}$: Variance of T year estimate for partial duration series

N : Number of years in the record

Z_T : Standard Gumbel variate for return period T.

(ii) The partial duration series method has a significantly lower estimation variance than the annual maximum method if $\lambda > 2$, assuming that the exceedances can be modelled by Poisson process and are an iid sequence of negative exponential variables. This reduction of estimation variance increases with the return period and with λ.

(iii) The truncation level q_o should be as low as possible providing that the model assumptions are still valid but when q_o is decreased the hypothesis which is first violated is usually the one regarding the serial independence of the exceedances. In this case another model for autocorrelated exceedances following a negative exponential law with a constant serial autocorrelation should be used. The estimation variance of flood magnitude increases with ρ_1, and partial duration series method no longer compares favourably with the annual maxima method. This means that the advantage of partial duration method is strongly conditioned by the validity of the assumption concerning the serial independence of exceedances.

Takeuchi (1984) evaluates Langbein's theory (1949) which relates hydrological recurrence intervals calculated from an annual maximum series and from a partial duration series. He suggests an alternative derivation procedure. The resultant formula is identical to Langbein's but the conditions to be satisfied for the formula to hold good is replaced by a new, more relaxed one. He reconfirms the validity of Langbein formula (1949) and Chow's discussion (1950).

Rosbjerg (1985) further extends the work of Tavares and Da Silva (1983). He introduces a correction factor in the variance formula given by Tavares and Da Silva. This is done in order to reduce the deviations between theoretical and Monte-Carlo generated samples. In the dependent case a formula for the variance of the T-year estimate has been developed and shown to be in fine agreement with Monte-Carlo based variance calculations. It has been finally concluded that T year estimates from annual maximum series and partial duration series become very close to each other for large return periods.

3.0 REMARKS

i. In practical situations there is no need to distinguish between $Q(T)$, $Q(T)_{TS}$, $Q(T)_{PDS}$ and $Q(T)_{AM}$ when T is large. NERC (1975)

ii. Langbein Formula (1949) which relates recurrence intervals calculated from an annual maximum series and partial duration series still holds good. Tekeuchi (1984)

iii. The dependence of partial flood series decreases with increase in truncation level.

iv. Either the mixed Poisson or Poisson distribution have the best fit to frequency distributions of the number of exceedances per year. Viraphol and Yevjevich (1978)

v. Either the mixed exponential or exponential distribution have the best fit to frequency distributions of the magnitude of exceedances. Viraphol and Yevjevich (1978)

vi. The partial duration series method has a significantly lower estimation variance than the annual maxima method if $\lambda > 2$

vii. If peaks over a threshold are correlated then model suggested by Tavares et al. (1983) should be used.

viii) Assumption regarding distribution to be adopted for flood magnitudes is important. Depending upon the distribution selected, estimate of the floods of high return periods have a wide range of variation.

In partial duration series models most of the progress has been made in the development of science. The applicability and suitability of partial duration series models proposed by Todorovic in various papers and also by Tavares et.al. (1983) needs to be studied for Indian rivers where flows are correlated. The results of partial duration series models should also be compared with the results of annual maximum models, particularly for short records/limited data.

REFERENCES

1. Ashkar,F. and J.Rousselle(1983 a),'Some Remarks on the Truncation used in Partial Flood Series Models', Water Resources Research,19/2 : pp.477-488.

2. Ashkar,F. and J.Rousselle (1983 b), ' The Effect of certain Restrictions Imposed on the Interarrival Times of Flood Events on the Poisson Distribution Used for Modelling Flood Counts', Water Resources Research,19/2 : pp.481-485.

3. Barndorff-Nielson,O.,(1964),'On the limit Distribution of the Maximum of a Random Number of Independent Random Variables', Acta.Math.Acad.Sci.Hungar,15 :pp. 399-403.

4. Beard, L.R.(1954),' Estimation of Flood Probabilities,' Proc. ASCE,80 : Sept 438,pp.21.

5. Beard, L.R.(1962),' Statistical Methods in Hydrology', Sacramento,Calif,1 Vol.(various pagings),Revised Edition (Published under Civil Works Investigations Project CW-151 by the U.S.Army Corps of Engineers).

6. Benson, M.A. (1962),' Evoluation of Methods for Evaluating the Occurrence of Floods', U.S.G.S. Water Supply Paper 1580 A ;pp.30.

7. Benson, M.A. (1968), ' Uniform Flood Frequency Estimating Methods for Federal Agencies', Water Resources Research, 4/5 : pp.891-908.

8. Berman,S.M.,(1962), ' Limiting Distribution of the Maximum Term in Sequence of Dependent Random Variables', Ann. Math.Statist.33 : 894-908.

9. Borgman,L.E.(1961),' The Frequency Distribution of Near Extremes', J.Geophys.Res.66/10 : pp.3295-3307.

10. Boughton,W.C.,(1980),' A Frquency Distribution of Annual Floods', Water Resources Research,Vol.16/2 : pp.347-354.

11. Boughton,W.C. and E.D.Shirly (1983),'Fitting Procedures for the Log Boughton Distribution', Journal of Hydraulic Division,ASCE,Vol.109/1 :pp. 579-589.

12. Calenda,G,A.,Petaccia,A.Togna,(1977),' Theoretical Probability Distribution of Critical Hydrologic Events by the Partial-Duration Series Method', Journal of Hydrology,33: pp.233-245.

13. Chander,S., S.K.Spolia and A.Kumar,(1978),' Flood Frequency Analysis by Power Transformation',Journal of Hydraulic Division,ASCE,Vol.104/2,pp.1495-1504.

14. Chow, V.T.,(1950),' Discussion of Annual Floods and the Partial-Duration Flood Series', Transactions of the American Geophysical Union,31(6) :pp.939-941.

15. Chow,V.T.,(1964), Editor-In-Chief, 'Hand Book of Applied Hydrology', Mc Graw-Hill Book Co., N.Y.

16. Cunnane,C.,(1973),' A Particular Comparison of Annual Maxima and Partial Duration Series Methods of Flood Frequency Prediction', Journal of Hydrology,20: pp.257-271.

17. Fuller,W.E.,(1914),' Flood Flows', Trans.ASCE, 77, Paper 1293 : pp.564-694.

18. Gumbel,E.J.,(1958),' Statistics of Extremes', Columbia University Press, N.Y.

19. Gupta,V.K., L.Dukstein and R.W.Peebles(1976),' On the Joint Distribution of the Largest Flood and its Time of Occurrence', Water Resources Research,12/2 : pp.295-304.

20. Hadgraft, R.G.,(1982),'An Empirical Study of Flood Frequency Distributions," Hydrol.Rep.000716.P.R., Surface Water Branch,Queensland Water Resources Comm. Brisbane, Australia.

21. Hall, M.J. and P.E. O'Connell (1972), ' Time Series Analysis of Mean Daily River Flows', Water and Water Engineering, 76 pp.125-133.

22. Hazen,A., (1914),' Storage to be Provided in Impounding Reservoirs for Muncipal Water Supply', Trans.ASCE, 77 : pp. 1539-1640.

23. Horton,R.E.,(1913), ' Frequency of Recurrence of Hudson River Floods', U.S.Weather Burea Bull.Z, pp.109-112.

24. Houghton,J.C.,(1978), 'Birth of a Parent: The Wakeby Distribution for Modelling Flood Flows', Water Resources Research,Vol.14 : pp.1105-1109.

25. Kendall,G.R.(1959),' Statistical Analysis of Extreme Events', Natural Res.Counc.Can.,Ass.Comm.Geod,Geophys. Proce.Hydrol.Symp.,Ottawa,Canada, 1: pp.54-78.

26. Kirby,W.,(1969),' On the Random Occurrence of Major Floods', Water Resources Research, 54 : pp 778-784.

27. Langbein,W.B.,(1949),' Annual Floods and partial Duration Flood Series', Transactions of the American Geophysical Union,30 : pp.879-881.

28. Moran, P.A.P.,(1957),' The Statistical Treatment of Flood Flows', Transactions American Geophysic. Union, 38/4 : pp.519-523.

29. Nash, J.E., and J.Amorocho,(1966),' The Accuracy of the Prediction of Floods of High Return period',Water Resources Research,2/2 : pp.191-198.

30. Natural Environment Research Council, (1975),' Flood Studies Report', Vol.1, U.K.

31. Quimpo,R.G.,(1967),' Stochastic Model of Daily River Flow Sequence', Hydrology Paper 18,pp.30,Colorado State University,Fort Collins.

32. Rao,D.V.,(1980),' Log Pearson Type III Distribution-A generalized Evaluation', Journal of Hydraulic Division, ASCE,Vol.106/5.

33. Rao,D.V.,(1980), ' Log Pearson Type III Distribution - Method of Mixed Moments', Journal of Hydraulic Division ASCE,Vol.106/6.

34. Richards, B.D., (1955), ' Flood Estimation and Control' Chapman and Hall, Ltd., London,3rd Rev.Ed : pp.187.

35. Riggs,H.C., (1961), ' Frequency of Natural Events', Proc. ASCE,Journal Hydraul.Div.,87/HY1,Paper 2706: pp.15-26.

36. Rosbjerg,D.,(1985),' Estimation in Partial Duration Series with Independent and Dependent Peak Values', Journal of Hydrology,76: pp.183-195.

37. Rousselle, J., (1972),' On Some Problems of Flood Analysis', Ph.D Dissertation,Colorado State University, Fort Collins.

38. Shane,R.M. and W.R.Lynn,(1964),' Mathematical Model for Flood Risk Evaluation', Proceedings of the ASCE, Journal of the Hydraulics Division,90/HY6 :pp.1-20.

39. Shane, R.M. and W.R.Lynn,(1969),' Peak Flood Discharge Confidence Limits', Proceedings of the ASCE,Journal of

the Hydraulics Division,95/HY1 : pp.161-173.

40. Takeuchi,K.,(1984),' Annual Maximum Series and Partial Duration Series-Evaluation of Langbein's Formula and Chow's Discussion' In : G.E.Stout and G.H.Davis (Editors),Global Water : Science and Engineering - The Ven Te Chow Memorial Volume. Journal of Hydrology,68 : pp.275-284.

41. Tavares, L.V. and J.E.Da Silva,(1983), ' Partial Duration Series Method Revisited,' Journal of Hydrology, 64 : pp.1-14.

42. Todorovic, P. and V.Yevjevich, (1967 a), ' A Particular Stochastic Process as Applied to Hydrology', Proc. Intern. Hydrol. Symp., Fort Collins,Colorado,1 : pp. 298-305.

43. Todorovic, P.,(1967 b),'A Stochastic Process of Monotonous Sample Functions', Math.Inst. of the Republic of Serbia, 4/19 : pp.149-158.

44. Todorovic,P. and E.Zelenhasic,(1970),' A Stochastic Model for Flood Analysis', Water Resources Research,6/6 : pp.1641-1648.

45. Todorovic,P., (1970),' On Some Problems Involving Random Number of Random Variables', The Annals of Mathematical Statistics, 41/3 : pp.1059-1063.

46. Todorovic,P. (1971), ' On Extreme Problems in Hydrology', Paper presented at Joint Statistics Meetings, American Statist. Assoc. and Inst. of Math.Statist., Colorado State University,Fort Collins, Co.

47. Todorovic, P. and J.Rousselle, (1971), ' Some Problems of Flood Analysis', Water Resources Research,7/5 : pp. 1140-1150.

48. Todorovic, P. and D.A.Woolhiser,(1972), ' On the time when the Extreme Flood Occurs', Water Resources Research, 8/6 : pp.1433-1438.

49. Todorovic, P., (1978), ' Stochastic Models of Extreme Flows and Their Application to Design ; International Symposium on Risk and Reliability in Water Resources, Proceedings Vol.II, University of Waterloo,Ontario, Canada.

50. U.S.Inter-Agency Committee on Water Resources,(1966), ' Methods of Flow Frequency Analysis', Sub Committee on Hydrology, Inter-Agency Committee on Water Resources, U.S.Government Printing Office, Washington,D.C.,Bull. 13 : pp.42.

51. Varshney, R.S., (1979), ' Engineering Hydrology'. Nem Chand & Bros., Roorkee : pp.917.

52. Viraphol Tarsombut and Yevjevich,(1978),' Use of Partial Flood Series for Estimating Distribution of Maximum Annual Flood Peak', Hydrology Papers 97, Colorado State University,Fort Collins,Co.

53. Water Resources Council, Hydrology Committee,(1967), ' A Uniform Technique for Determining Flood Flow Frequencies', Bulletin 15, Washington,D.C.

54. Water Resources Council, Hydrology Committee(1976), ' Guidelines for Determining Flood Flow Frequency', Bulletin 17, Washington,D.C.

55. Zelenhasic,E.,(1970), ' Theoretical Probability Distributions for Flood Peaks', Hydrology Paper No.42, Colorado State University,Fort Collins.