ELSEVIER

Journal of Hydrology 244 (2001) 86–96

Journal of Hydrology

www.elsevier.com/locate/jhydrol

Effect of the occurrence process of the peaks over threshold on the flood estimates

B. Önöz*, M. Bayazit

Civil Engineering Department, Hydraulics Division, Istanbul Technical University, Maslak, 80626 Istanbul, Turkey

Received 10 April 2000; revised 28 November 2000; accepted 26 January 2001

Abstract

Partial duration (peaks over threshold) series have been used as an alternative to annual maxima series in flood frequency analysis. Poisson process is usually assumed for the occurrence of peaks. In some peaks over threshold series, variance of the annual number of exceedances is significantly smaller (or larger) than the mean. In such cases binomial (or negative binomial) distribution has a better fit. Expressions are obtained for the estimation of the T-year flood and its sampling variance when binomial (or negative binomial) model is combined with the exponential distribution of peak magnitudes. It is shown that the results are almost identical to those obtained using the Poisson model, for which much simpler expressions are available.
© 2001 Elsevier Science B.V. All rights reserved.

Keywords: Flood; Partial duration series; Peaks over threshold; Poisson process; Binomial distribution; Negative binomial distribution

1. Introduction

1.1. Literature review

Partial duration flood series (peaks over threshold) have been proposed as an alternative to annual maximum flood series in flood frequency analysis. A partial duration series consists of all the peaks above a certain threshold level, whereas an annual maximum flood series contains only the maximum flood of each year. It has been argued that partial duration series method uses more information about the floods because it works with more elements than the annual maxima series method (Langbein, 1949). Furthermore, it has more physical relevance because it is based on models for the distribution of the annual number of exceedances of flood events and of their magnitudes. In practice, it is usually more meaningful to consider not just the annual maximum floods but also the floods that exceed the safe capacity. The idea is to derive the distribution of the magnitudes of annual floods from the assumed distributions of the annual number of occurrences and the magnitudes of the peaks over threshold.

Shane and Lynn (1964) assumed that the occurrence times of flood peaks followed a Poisson process and the flood magnitudes had the exponential distribution (since the upper tails of many commonly used distribution functions are approximately exponential). Kirby (1969) showed that the Poisson process could be justified as a limiting form of the randomly spaced Bernoulli trials at sufficiently small exceedance probabilities. The probability distribution and moments of the number of exceedances in a specified time interval approached those implied by the occurrence of trials in a Poisson process.

---

* Corresponding author. Fax: +90-212-285-30-02.
E-mail address: onoz@itu.edu.tr (B. Önöz).

0022-1694/01/$ - see front matter © 2001 Elsevier Science B.V. All rights reserved.
PII: S0022-1694(01)00330-4

Zelenhasic (1970) considered the two-parameter gamma distribution for the magnitude of exceedances, of which exponential distribution is a special case. He investigated the distribution of annual maximum floods assuming Poisson distribution for the annual number of events and exponential distribution for their magnitudes, in which case annual maxima is found to follow the Gumbel distribution. Analysing the data from two streams, he found that the Gumbel distribution has a better fit to the annual maximum floods when its parameters are estimated on the basis of the partial duration series.

Taesombut and Yevjevich (1978) compared a number of distributions for the annual number and magnitude of the exceedances. They did not derive expressions for the probability distribution of annual maxima except for the combination of Poisson and exponential models. They empirically investigated the efficiency of estimates of flood peaks of given return periods by using annual and partial flood series on two study cases. It was found that the estimates of annual floods of a certain return period from partial series had a smaller sampling variance than the estimates from annual series, if partial series had at least 1.65 exceedances per year on an average, as obtained theoretically by Cunnane (1973). Using 17 sets of partial flood series in the U.S. they showed that the Poisson or mixed Poisson process had the best fit to the observed frequency distribution of annual number of exceedances, closely followed by the negative binomial. Exponential or mixed exponential distribution had the best fit to the observed frequency distributions of the magnitude of exceedances.

NERC (1975); and Cunnane (1979) found that the data from 26 streams in Great Britain indicated that the number of peaks occurring each year was not a Poisson variate, its variance being significantly greater than its mean. They argued that the negative binomial distribution did not seem to offer a vast improvement although no statistical test was carried out.

Van Montfort and Witter (1985) investigated the use of the generalised Pareto distribution instead of the exponential distribution for the magnitude of exceedances and proposed some statistics to compare the goodness-of-fit of two distributions.

Ben-Zvi (1991) investigated the goodness-of-fit of the Poisson and negative binomial distributions to eight flood data series in Israel at a number of threshold discharges using the χ^2 test. He found that the acceptance rate of the fit of the negative binomial distribution to the samples was better than that of the Poisson distribution.

Rosbjerg et al. (1992) and Madsen et al. (1997) showed that the generalised extreme value (GEV) distribution is obtained for annual floods when Poisson process assumption for the occurrence of exceedances is combined with the generalised Pareto distribution for the magnitude of exceedances. They derived approximate expressions for the variance of the *T*-year event estimator.

Vukmirovic and Petrovic (1997) and Lang et al. (1997) considered binomial and negative binomial distributions for the annual number of peaks, and gave expressions for the distribution of annual maxima for these cases.

Recently, Lang et al. (1999) presented a state-of-the-art review of the peaks over threshold modeling. They discussed issues such as the threshold selection, models for the occurrence processes of the peaks and for the distribution of their magnitudes, and the correspondence between the peaks over threshold and annual maximum flood distributions. They concluded that the main difficulties of the peaks over threshold approach concern the selection of the threshold level and of the occurrence process.

1.2. Objective of the study and outline of the paper

Poisson process is the most commonly used model for the occurrences of exceedances. Resulting Poisson distribution for the annual number of peaks over threshold has a variance equal to the mean. However, it has been observed that in certain cases the variance of observed number of exceedances is significantly different, usually larger, than the mean. In this study, two other models are considered for the process. These lead to binomial and negative binomial probability distributions, two-parameter distributions that are of the same family as the Poisson distribution. One of them has a variance smaller than the mean, and the other has a variance larger than the mean. These distributions are combined with the exponential distribution for the magnitude of the exceedances, and expressions for the resulting probability distributions of the annual maximum floods are theoretically derived. Sampling variances of the T-year return period floods are also estimated in each case.

This study aims at clarifying the role of the model chosen for the occurrence times of flood peaks over a threshold in flood frequency analysis. In a recent review Lang et al. (1999) discussed the selection of the occurrence process, which points out the significance of the subject. In practice it would be important to know if a model other than the Poisson process should be chosen when the variance and mean of the observed number of exceedances are significantly different. Thus it is required to find out how the expressions for the T-year flood and its sampling variance are changed when different models are adopted for the occurrence process.

The paper starts with the introduction of the models for the probability distribution of the annual number of exceedances. Binomial and negative binomial distributions are then combined with the exponential distribution, and formulas for the probability distribution function of the annual maxima are derived. Expressions for the T-year return period annual flood, x_T, are obtained in each case, and formulas for its asympotical sampling variance are derived in the Appendix. Finally, some applications are made to illustrate the use of the derived expressions.

2. Models for the distribution of the annual number of exceedances

In this study three models are considered for the probability distribution of m, the annual number of exceedances.

2.1. Poisson distribution

This is the model most commonly used for the distribution of the annual number of peaks above a certain threshold level x_0. A useful quality of this distribution is that if a process is Poissonian for a certain level x_0, it is also Poissonian for any other level higher than x_0 (Cunnane, 1979).

Poisson distribution has the probability mass function:

$$P(m=k) = e^{-\mu}\mu^k/k!, \qquad k=0,1,2,\ldots \tag{1}$$

Moments of variable m are

$$E(m) = E = \mu$$
$$\mathrm{Var}(m) = V = \mu \tag{2}$$
$$\mu_3(m) = \mu$$

where $E(m)$, $\mathrm{Var}(m)$ and $\mu_3(m)$ are the mean, variance, and third order central moment of m, respectively.

This distribution has a variance equal to the mean. It has been used by several researchers (e.g. Shane and Lynn, 1964; Kirby, 1969; Zelenhasic, 1970; Rosbjerg et al., 1992; Madsen et al., 1997) for modelling the probability distribution of the annual number of exceedances although in many cases the assumption of the equality of the variance to the mean does not hold (Taesombut and Yevjevich, 1978; Cunnane, 1979). Records of the peaks over threshold data of some streams showed that peaks sometimes occurred in bunches, interevent times not being identically distributed in conflict with the assumptions of Poisson process (NERC, 1975).

2.2. Binomial distribution

This two-parameter distribution has a mean higher than the variance. Its probability mass function is

$$P(m=k) = \binom{\gamma}{k}\alpha^k(1-\alpha)^{\gamma-k}, \qquad k=0,1,2,\ldots \tag{3}$$

where $\binom{k}{\gamma}$ is the number of combinations of γ different things taken k at a time:

$$\binom{\gamma}{k} = \gamma!/k!(\gamma-k)! \tag{4}$$

Moments of variable m are

$$E(m) = E = \alpha\gamma$$
$$\mathrm{Var}(m) = V = \alpha(1-\alpha)\gamma \tag{5}$$
$$\mu_3(m) = \alpha(1-\alpha)(1-2\alpha)\gamma.$$

2.3. Negative binomial distribution

In most cases annual number of occurrences of the peaks above threshold has a variance larger than its mean. Negative binomial distribution has this property, and has been used by Taesombut and Yevjevich (1978), Cunnane (1979), Ben-Zvi (1991), Vukmirovic and Petrovic (1997) and Lang et al. (1999) in the modelling of peaks over threshold. It is argued that this distribution accounts for the clustering of the peaks in some years (NERC, 1975; Lang et al., 1997). Lang (1999) showed that the waiting time of a process described by a negative binomial distribution is also exponentially distributed as is the case for the Poisson process, and explained the apparent paradox by the fact that the negative binomial distribution allows the clustering of events.

Formulas for this distribution are as follows:

$$P(m = k) = \begin{pmatrix} \gamma + k - 1 \\ k \end{pmatrix} \alpha^k (1 - \alpha)^\gamma, \tag{6}$$

$k = 0, 1, 2, \ldots$

$$E(m) = E = \alpha\gamma/(1 - \alpha) \tag{7}$$

$$\mathrm{Var}(m) = V = \alpha\gamma/(1 - \alpha)^2$$

$$\mu_3(m) = (1 + \alpha)\alpha\gamma/(1 - \alpha)^3$$

Williams and Bretherton (1963) provided the tables of this distribution.

2.4. Tests for the choice of the distribution

Taesombut and Yevjevich (1978) used the χ^2 test to determine the best-fit distribution. Cunnane (1979) proposed to use the dispersion index defined as:

$$I = \mathrm{Var}(m)/E(m) \tag{8}$$

The test statistic corresponding to the dispersion index is:

$$d = \sum_{i=1}^{N} \frac{(m_i - \bar{m})^2}{\bar{m}} = \frac{(N - 1)V}{E} \tag{9}$$

where m_i is the annual number of exceedances observed in the year i ($i = 1,2,\ldots,N$), and $\bar{m}$ is the mean of m_i. d follows a χ^2 distribution with $N - 1$ degrees of freedom. As the Poisson distribution has a dispersion index $I = 1$, the poissonian hypothesis is not rejected if the computed d value is in the range $(\chi^2_{\alpha/2}, \chi^2_{1-\alpha/2})$ where α is the significance level. If $d < \chi^2_{\alpha/2}$ binomial distribution can be preferred, if $d > \chi^2_{1-\alpha/2}$ negative binomial distribution can be preferred.

3. Probability distribution of annual maximum floods

The probability distribution function $F_x(x)$ of the annual maxima can be derived using the total probability theorem. If $G_y(y)$ is the probability distribution of the magnitude of exceedances $y = x - x_0$, we can write (Shane and Lynn, 1964):

$$F_x(x) = \sum_{k=0}^{\infty} P(m = k)[G_y(y)]^k \tag{10}$$

It will be assumed here that $G_y(y)$ is exponential with the parameter $\beta = E(y) = \mathrm{Var}^{0.5}(y)$:

$$G_y(y) = 1 - \exp(-y/\beta), \qquad y \geq 0 \tag{11}$$

Exponential distribution has been widely used to model the magnitude of the exceedances. If y is exponentially distributed for a threshold level x_0, it will also be exponential for any level higher than x_0 (Ashkar and Rousselle, 1983). Zelenhasic (1970) and Taesombut and Yevjevich (1978) used this distribution. It was shown that the exponential assumption did not fit the data of some streams for low threshold levels (NERC, 1975).

Another distribution for the magnitude of exceedances is the generalised Pareto, of which the exponential distribution is a special case. Van Montfort and Witter (1985) proposed this distribution. Rosbjerg et al. (1992) showed that exponential distribution should be preferred when the shape parameter k is less than 0.1 even if the parent distribution is generalised Pareto. Vukmirovic and Petrovic (1997) considered the Weibull distribution.

Exponential distribution of y will now be combined with the Poisson, binomial and negative binomial distributions, respectively, to derive the probability distribution $F_x(x)$ of annual maxima.

3.1. Poisson distribution

Zelenhasic (1970) obtained an expression for $F_x(x)$ when the number of exceedances are poissonian and their magnitude is exponential:

$$F_x(x) = \sum_{k=0}^{\infty} \frac{e^{-\mu}\mu^k}{k!}[1 - \exp(-y/\beta)]^k$$

$$= e^{-\mu} \exp\{\mu[1 - \exp(-y/\beta)]\}$$

$$= \exp\{ - \mu \exp[-(x - x_0)/\beta]\} \quad (12)$$

$F_x(x)$ is the Type I extremal (Gumbel) distribution, widely used in flood frequency analysis. It is known, however, that annual maximum floods of many streams do not follow this distribution.

3.2. Binomial distribution

Inserting Eq. (3) for $P(m = k)$ and Eq. (11) for $G_y(y)$ into Eq. (10):

$$F_x(x) = \sum_{k=0}^{\infty} \binom{\gamma}{k} \alpha^k (1 - \alpha)^{\gamma-k} [1 - \exp(-y/\beta)]^k$$

$$= (1 - \alpha)^{\gamma} \sum_{k=0}^{\infty} \binom{\gamma}{k} \left(\frac{\alpha}{1 - \alpha}\right)^k [1 - \exp(-y/\beta)]^k$$

$$= (1 - \alpha)^{\gamma} \left\{1 + \frac{\alpha}{1 - \alpha}\left[1 - \exp\left(-\frac{x - x_0}{\beta}\right)\right]\right\}^{\gamma} \quad (13)$$

using the binomial series expansion of $(1 + a)^{\gamma}$ where $1 + a$ is the quantity in brackets in Eq. (13). $(1 - \alpha)^{\gamma}$ is the probability that the annual flood does not exceed the threshold level x_0.

3.3. Negative binomial distribution

Inserting Eq. (6) for $P(m = k)$ and Eq. (11) for $G_y(y)$ into Eq. (10):

$$F_x(x)$$

$$= \sum_{k=0}^{\infty} \binom{k + \gamma - 1}{k} \alpha^k (1 - \alpha)^{\gamma} [1 - \exp(-y/\beta)]^k$$

$$= (1 - \alpha)^{\gamma} \left\{1 - \alpha\left[1 - \exp\left(-\frac{x - x_0}{\beta}\right)\right]\right\}^{-\gamma} \quad (14)$$

using the binomial series expansion of $(1 - b)^{-\gamma}$ where $1 - b$ is the quantity in brackets in Eq. (14).

Lang et al. (1997) obtained expressions equivalent to Eqs. (13) and (14) using an approach based on random censoring. Similar expressions were given by Vukmirovic and Petrovic (1995).

4. *T*-year flood and its sampling distribution

Annual flood of T-year return period has the probability of non-exceedance

$$F_x(x) = 1 - \frac{1}{T} = T_1 \quad (15)$$

Expressions for the T-year flood x_T corresponding to this probability and its asymptotic sampling variance are derived below.

4.1. Poisson distribution

Combining Eqs. (12) and (15) an expression for the estimate of the T-year flood is obtained:

$$\hat{x}_T = x_0 + \beta \ln \mu - \beta \ln(-\ln T_1) \quad (16)$$

Cunnane (1973) obtained the following expression for the asymptotic sampling variance of the estimate $\hat{x}_T$ computed from N-year long observations:

$$\mathrm{Var}(x_T) = (\beta^2/\mu N)\{1 + [\ln \mu - \ln(-\ln T_1)]^2\}. \quad (17)$$

Rosbjerg (1985) introduced a small sample correction factor to the above formula, which is of minor importance for $\mu N > 10$.

4.2. Binomial distribution

Combining Eqs. (13) and (15):

$$(1 - \alpha)^{\gamma} \left\{1 + \frac{\alpha}{1 - \alpha}\left[1 - \exp\left(-\frac{\hat{x}_T - x_0}{\beta}\right)\right]\right\}^{\gamma} = T_1 \quad (18)$$

Solving for x_T:

$$\hat{x}_T = x_0 + \beta \ln \alpha - \beta \ln(1 - T_1^{1/\gamma}). \quad (19)$$

An expression for the asymptotically sampling variance of $\hat{x}_T$ is derived in Appendix A, with the

Table 1
Characteristics of the peaks over threshold series

Series No	Station No.	Station name	Basin area (10^6 m^2)	Record length N (yr)	Threshold x_0 (m^3/s)	Number of peaks
1	28804	Trent at Trent Bridge, UK	7490	86	300 (150)	216
2	3183500	Greenbrier, at Alderson, W.Vir., USA	3500	101	650	180

following result:

$$\begin{aligned}\mathrm{Var}(\hat{x}_T) = \beta^2/N\Bigg\{&\left(2+\frac{3\alpha-1}{\alpha\gamma}\right)\frac{(1-\alpha)^2}{\alpha^2}\\ &+\frac{1}{\alpha\gamma}\ln^2\left(\frac{\alpha}{1-T_1^{1/\gamma}}\right)+\left[2(1-\alpha)\gamma-\frac{(1-2\alpha)^2}{\alpha}\right]\\ &\times\frac{1-\alpha}{\alpha^2\gamma^3}\frac{T_1^{2/\gamma}\ln^2 T_1}{(1-T_1^{1/\gamma})^2}-\frac{2(1-\alpha)}{\alpha^2\gamma^2}\\ &\times\left[\frac{(1-2\alpha)^2}{\alpha}-2\gamma(1-\alpha)\right]\frac{T_1^{1/\gamma}\ln T_1}{1-T_1^{1/\gamma}}\Bigg\} \qquad (20)\end{aligned}$$

4.3. Negative binomial distribution

Combining Eqs. (14) and (15):

$$(1-\alpha)^\gamma\left\{1-\alpha\left[1-\exp\left(-\frac{\hat{x}_T-x_0}{\beta}\right)\right]\right\}^{-\gamma}=T_1 \quad (21)$$

Solving for $\hat{x}_T$:

$$\hat{x}_T = x_0-\beta\ln\left(\frac{1-\alpha}{\alpha}\right)-\beta\ln(T_1^{-1/\gamma}-1). \qquad (22)$$

Asympotical sampling variance of the estimate of $\hat{x}_T$ is (see Appendix A):

$$\begin{aligned}\mathrm{Var}(\hat{x}_T) = \beta^2/N\Bigg\{&\left(2-\frac{2}{\gamma}-\frac{1}{\alpha\gamma}\right)\frac{1}{\alpha^2}+\frac{1-\alpha}{\alpha\gamma}\\ &\times\ln^2\left[\frac{(1-\alpha)(T_1^{-1/\gamma}-1)}{\alpha}\right]\\ &+\frac{2\gamma-((1+\alpha)^2/\alpha)}{\alpha^2\gamma^3}\frac{T_1^{-2/\gamma}\ln^2 T_1}{(T_1^{-1/\gamma}-1)^2}\\ &+\frac{2}{\alpha^2\gamma^2}\left[\frac{(1+\alpha)^2}{\alpha}-2\gamma\right]\frac{T_1^{-1/\gamma}\ln T_1}{T_1^{-1/\gamma}-1}\Bigg\}. \qquad (23)\end{aligned}$$

4.4. Comparison of T-year flood magnitudes of the three distributions

For the negative binomial distribution, combining Eqs. (16) and (22), with $(1-\alpha)/\alpha = 1/(I-1)$ and $\gamma=\mu/(I-1)$, the ratio R_{NP} of the T-year flood estimates of the negative binomial and Poisson distributions is obtained as:

$$R_{NP}=\frac{(\hat{x}_{TN}-x_0)/\beta}{(\hat{x}_{TP}-x_0)/\beta}=\frac{\ln(I-1)-\ln(T_1^{-(I-1/\mu)}-1)}{\ln\mu-\ln(-\ln T_1)}. \qquad (24)$$

Similarly, combining Eqs. (16) and (19), with $\alpha = 1-I$ and $\gamma=\mu/(1-I)$, the ratio R_{BP} of the T-year flood estimates of the binomial and Poisson distributions is:

$$R_{BP}=\frac{(\hat{x}_{TB}-x_0)/\beta}{(\hat{x}_{TP}-x_0)/\beta}=\frac{\ln(1-I)-\ln(1-T_1^{(1-I)/\mu})}{\ln\mu-\ln(-\ln T_1)}. \qquad (25)$$

Both ratios tend toward unity when the return period T is high ($T_1 \cong 1$).

For $\mu=2$ and $I=1.4$ for the negative binomial ($I=1/1.4$ for the binomial) models, the relative difference between the estimates of the Poisson and negative binomial (or binomial) models is less than 2% even when T is as small as 5 years. It is seen that the T-year flood magnitudes estimated using the Poisson, binomial and negative binomial distribution assumptions for the annual number of exceedances are practically identical.

5. Examples

Models proposed for the times of occurrence of the exceedances are applied to two data sets. Characteristics of the data used in the study are given in Table 1.

Table 2
T-year floods and their variances for series no. 1

Return period T (yr)	$\hat{x}_T$ (m^3/s) Gumbel	Poisson and Negative binomial	Var $\hat{x}_T$ (m^6/s^2) Gumbel	Poisson	Negative binomial
$x_0 = 300$ m^3/s					
25	996	902	3463	1780	1817
50	1124	1005	4774	2403	2441
100	1251	1107	6305	3119	3157
250	1418	1242	8669	4208	4246
500	1544	1343	10716	5142	5179
$x_0 = 150$ m^3/s					
25	996	1017	3463	1727	1754
50	1124	1143	4774	2241	2268
100	1251	1268	6305	2820	2847
250	1418	1432	8669	3686	3714
500	1544	1555	10716	4420	4447

5.1. Series no. 1

Most of the peaks over threshold series have a larger variance of the annual number of peaks than its mean. This is the case for series no. 1, of which the mean is $E = 2.51$ and the variance is $V = 3.52$ at the threshold level $x_0 = 300$ m^3/s. Parameters of the negative binomial distribution are estimated from Eq. (7) as $\alpha = 0.287$, $\gamma = 6.24$. Dispersion index test has the test statistic

$$d = \frac{(N-1)V}{E} = 119.2 > \chi^2_{85,\,0.975} = 112.$$

Parameter of the Poisson distribution is $\mu = E = 2.51$.

The hypothesis of Poisson distribution is rejected at the level of significance $\alpha = 0.05$. The χ^2 test is also applied to test the goodness-of-fit of the Poisson and negative binomial probability distributions. Using 9 class intervals ($m = 0,1,\ldots,8$) the Poisson distribution gives

$$\chi^2 = 19.1 > \chi^2_{7,0.95} = 14.1$$

and the negative binomial distribution gives

$$\chi^2 = 6.2 < \chi^2_{6,0.95} = 12.6.$$

Thus the Poisson distribution is rejected whereas the negative binomial distribution is not rejected at the level $\alpha = 0.05$.

Exponential distribution fitted to the magnitudes of the peaks has the parameter $\beta = E(y) = 146.3$. Tests given by Van Montfort and Witter (1985) are applied to check the adequacy of the exponential distribution against the generalised Pareto distribution. The statistic G (maximum/median) is equal to

$$G = 807/85 = 9.5.$$

For the sample size of peaks $n = EN = 216$, 0.05 and 0.95 quantiles of G in exponential samples are 5.9 and 12.2, respectively. The exponential distribution hypothesis is not rejected. The statistic $r\sqrt{n}$ is equal to

$$r\sqrt{n} = -0.075\sqrt{216} = -1.1.$$

For $n = 216$, 0.05 and 0.95 quantiles of $r\sqrt{n}$ are -1.78 and 1.59, respectively. Again, exponentiality is not rejected.

The study is repeated at the lower threshold level $x_0 = 150$ m^3/s in order to see the effect of lowering the threshold. There are 454 peaks in 86 years. The parameters are $E = 5.28$ and $V = 7.38$. Parameter of the Poisson model is $\mu = E = 5.28$. Parameters of the negative binomial distribution are $\alpha = 0.285$, $\gamma = 13.2$. Dispersion index test statistic is

$$d = \frac{(N-1)V}{E} = 118.8 > \chi^2_{85,0.975} = 112.$$

Again, Poisson model is rejected at 0.05 level of significance. Applying the χ^2 test with 13 class intervals ($m = 0,1,\ldots,12$)

$$\chi^2 = 33.3 > \chi^2_{11,0.95} = 19.7$$

for the Poisson distribution, whereas the negative

Table 3
T-year floods and their variances for series no. 2

Return period T(yr)	$\hat{x}_T$(m^3/s)		Var $\hat{x}_T$(m^6/s^2)		
	Gumbel	Poisson and binomial	Gumbel	Poisson	Binomial
25	1881	1835	9343	8336	8289
50	2109	2055	12879	11509	11460
100	2335	2274	17009	15194	15144
250	2632	2562	23886	20858	20807
500	2857	2779	28910	25747	25696

binomial distribution gives

$$\chi^2 = 15.7 < \chi^2_{10,0.95} = 18.3.$$

The test rejects the Poisson model but not the negative binomial model at the level $\alpha = 0.05$. Exponential distribution with the parameter $\beta = E(y) = 178.4$ has a good fit to the magnitudes of the peaks above 150 m^3/s.

Results are similar to those at the higher threshold level.

5.2. Series no. 2

Very few partial duration series has a variance of the annual number of exceedances smaller than its mean (Taesombut and Yevjevich, 1978). Series no. 2 has the mean $E = 1.78$ and variance $V = 1.61$. Although the variance is smaller than the mean, the dispersion index test does not reject the hypothesis of Poisson distribution:

$$d = 90.4 > \chi^2_{100,0.025} = 74.2.$$

The χ^2 test with six class intervals ($m = 0,1,\ldots,5$) rejects neither the hypothesis of Poisson distribution nor that of the binomial distribution. For Poisson distribution:

$$\chi^2 = 1.07 < \chi^2_{4,0.95} = 9.5$$

for binomial distribution with parameters $\alpha = 0.095$, $\gamma = 18.6$:

$$\chi^2 = 0.91 < \chi^2_{3,0.95} = 7.8.$$

Both distributions seem to have a very good fit to the annual number of peaks of the observed series.

Exponential distribution with the parameter $\beta = E(y) = 313$ has a very good fit to the magnitudes of the peaks. Statistics have the values:

$$G = 8.35, \qquad r\sqrt{n} = -0.34.$$

For $n = 180$, 0.05 and 0.95 quantiles of G are 5.5 and 12.5, respectively. Corresponding quantiles of $r\sqrt{n}$ are -1.78 and 1.22. Neither test rejects the hypothesis of exponential distribution.

5.3. Estimates of T-year floods and their sampling variances

For both series, T-year floods and their sampling variances are estimated for various values of the return period ($T = 25$, 50, 100, 250 and 500 years) assuming Gumbel distribution for the annual maxima series, Poisson and negative binomial (for series no. 1) or binomial (series no. 2) models for the partial duration series. Parameters of the Gumbel distribution are estimated by the maximum likelihood method. For other models, formulas derived in this study are used. Results are given in Tables 2 and 3.

Estimates of x_T are the same for Poisson and binomial (or negative binomial) models. These estimates are smaller than those of the corresponding estimates of the Gumbel distribution of annual maxima. For series no. 1, the choice of the threshold level affects the estimates of x_T. For $x_0 = 150$ m^3/s, $\hat{x}_T$ values are larger than those of $x_0 = 300$ m^3/s, and close to the Gumbel estimates. Variances of $\hat{x}_T$ are slightly larger for negative binomial and slightly smaller for binomial models than those of the Poisson model in these examples. Var($\hat{x}_T$) of Gumbel model are much larger than those obtained using the (negative) binomial or Poisson models.

6. Conclusions

Partial duration series of some streams have annual number of exceedances that has significantly larger (smaller) variance than the mean. In such cases negative binomial (binomial) model provides a better fit to the process than the Poisson model. However, flood estimates based on the negative binomial or binomial models when combined with the exponential distribution for the magnitudes of peaks are almost identical to those obtained using the Poisson model, as is also the case for their sampling variances. This result is in agreement with the findings of Kirby (1969) and Cunnane (1979) and makes it unnecessary to prefer the binomial or negative binomial models even when the Poisson process hypothesis is rejected by the statistical tests. It is easier to use the Poisson model because it leads to much simpler expressions for the *T*-year flood and its sampling variance. The choice of the threshold value has an important effect on the flood estimates, and needs to be studied further.

Acknowledgements

Authors express their thanks to two anonymous reviewers for the comments that helped to improve the paper.

Appendix A. Asymptotical variances of *T*-year flood estimates

Probability distribution of annual floods based on binomial or negative binomial distributions of mean annual number of exceedances (Eqs. (13) and (14)) have three parameters: α and γ of binomial or negative binomial distribution, and β of exponential distribution. It has been shown that the magnitudes of exceedances are independent of the number of exceedances (Taesombut and Yevjevich, 1978). Therefore, it will be assumed here that α and β are independent, and so are γ and β.

A.1. Variances of parameters

α and γ are functions of E and V, the mean and variance of annual number of exceedances. Asymptotical sampling variances of their estimates can be computed as

$$\mathrm{Var}(\hat{\alpha}) = \left(\frac{\partial\alpha}{\partial E}\right)^2 \mathrm{Var}(\hat{E}) + \left(\frac{\partial\alpha}{\partial V}\right)^2 \mathrm{Var}(\hat{V}) + 2\frac{\partial\alpha}{\partial E}\frac{\partial\alpha}{\partial V}\mathrm{Cov}(\hat{E},\hat{V})$$

$$\mathrm{Var}(\hat{\gamma}) = \left(\frac{\partial\gamma}{\partial E}\right)^2 \mathrm{Var}(\hat{E}) + \left(\frac{\partial\gamma}{\partial V}\right)^2 \mathrm{Var}(\hat{V}) + 2\frac{\partial\gamma\partial\gamma}{\partial E\partial V}\mathrm{Cov}(\hat{E},\hat{V}), \tag{A1}$$

where

$$\mathrm{Var}(\hat{E}) = V/N, \qquad \mathrm{Var}(\hat{V}) = 2V^2/N,$$

$$\mathrm{Cov}(\hat{E},\hat{V}) = \mu_3/N.$$

For binomial distribution, with $\alpha = 1 - (V/E)$ and $\gamma = E^2/(E - V)$ from Eq. (5):

$$\frac{\partial\alpha}{\partial E} = \frac{V}{E^2} = \frac{1-\alpha}{\alpha\gamma}, \qquad \frac{\partial\alpha}{\partial V} = -\frac{1}{E} = -\frac{1}{\alpha\gamma},$$

$$\frac{\partial\gamma}{\partial E} = \frac{E(E-2V)}{(E-V)^2} = \frac{2\alpha-1}{\alpha^2},$$

$$\frac{\partial\gamma}{\partial V} = \frac{E^2}{(E-V)^2} = \frac{1}{\alpha^2}, \qquad \mathrm{Var}(\hat{E}) = \frac{\alpha(1-\alpha)\gamma}{N},$$

$$\mathrm{Var}(\hat{V}) = \frac{2\alpha^2(1-\alpha)^2\gamma^2}{N},$$

$$\mathrm{Cov}(\hat{E},\hat{V}) = \frac{\alpha(1-\alpha)(1-2\alpha)\gamma}{N}.$$

Substituting the above expressions into Eq. (A1) and rearranging, the asymptotical sampling variances of the estimates of α and γ are obtained:

$$\mathrm{Var}(\hat{\alpha}) = \left(2 + \frac{3\alpha-1}{\alpha\gamma}\right)\frac{(1-\alpha)^2}{N}, \tag{A2}$$

$$\mathrm{Var}(\hat{\gamma}) = \left[2(1-\alpha)\gamma - \frac{(1-2\alpha)^2}{\alpha}\right]\frac{(1-\alpha)\gamma}{\alpha^2 N}. \tag{A3}$$

Asymptotical covariance of $\hat{\alpha}$ and $\hat{\gamma}$ can be computed as follows:

$$\mathrm{Cov}(\hat{\alpha},\hat{\gamma}) = \frac{\partial\alpha}{\partial E}\frac{\partial\gamma}{\partial E}\mathrm{Var}(\hat{E}) + \frac{\partial\alpha}{\partial V}\frac{\partial\gamma}{\partial V}\mathrm{Var}(\hat{V}) + \left(\frac{\partial\alpha}{\partial E}\frac{\partial\gamma}{\partial V} + \frac{\partial\alpha}{\partial V}\frac{\partial\gamma}{\partial E}\right)\mathrm{Cov}(\hat{E},\hat{V}). \tag{A4}$$

Substituting the expressions for the partial derivatives and sampling variances and covariance of the estimates of E and V into Eq. (A4) and rearranging:

$$\mathrm{Cov}(\hat{\alpha},\hat{\gamma}) = \left[\frac{(1-2\alpha)^2}{\alpha} - 2(1-\alpha)\gamma\right]\frac{1-\alpha}{\alpha N}. \tag{A5}$$

β is the mean of the exponentially distributed magnitude of exceedances whose variance is β^2. Therefore, the sampling variance of the estimate of β is

$$\mathrm{Var}(\hat{\beta}) = \frac{\beta^2}{E(m)N} = \frac{\beta^2}{\alpha\gamma N} \tag{A6}$$

because the mean number of exceedances in N years is $E(m)N$.

For negative binomial distribution, with $\alpha = 1 - (E/V)$ and $\gamma = E^2/(V-E)$ from Eq. (7):

$$\frac{\partial\alpha}{\partial E} = -\frac{1}{V} = -\frac{(1-\alpha)^2}{\alpha\gamma},$$

$$\frac{\partial\alpha}{\partial V} = \frac{E}{V^2} = \frac{(1-\alpha)^3}{\alpha\gamma},$$

$$\frac{\partial\gamma}{\partial E} = \frac{E(2V-E)}{(V-E)^2} = \frac{(1+\alpha)(1-\alpha)}{\alpha^2},$$

$$\frac{\partial\gamma}{\partial V} = -\frac{E^2}{(V-E)^2} = -\frac{(1-\alpha)^2}{\alpha^2},$$

$$\mathrm{Var}(\hat{E}) = \frac{\alpha\gamma}{(1-\alpha)^2 N}, \qquad \mathrm{Var}(\hat{V}) = \frac{2\alpha^2\gamma^2}{(1-\alpha)^4 N},$$

$$\mathrm{Cov}(\hat{E},\hat{V}) = \frac{(1+\alpha)\alpha\gamma}{(1-\alpha)^3 N}.$$

Substituting these into Eqs. (A1) and (A4) and rearranging:

$$\mathrm{Var}(\hat{\alpha}) = \left(2\gamma - \frac{1}{\alpha} - 2\right)\frac{(1-\alpha)^2}{\gamma N}, \tag{A7}$$

$$\mathrm{Var}(\hat{\gamma}) = \left[2\gamma - \frac{(1+\alpha)^2}{\alpha}\right]\frac{\gamma}{\alpha^2 N}, \tag{A8}$$

$$\mathrm{Cov}(\hat{\alpha},\hat{\gamma}) = \left[\frac{(1+\alpha)^2}{\alpha} - 2\gamma\right]\frac{1-\alpha}{\alpha N}, \tag{A9}$$

$$\mathrm{Var}(\hat{\beta}) = \frac{\beta^2}{E(m)N} = \frac{(1-\alpha)\beta^2}{\alpha\gamma N}. \tag{A10}$$

A.2. Variances of T-year flood estimates

T-year flood x_T based on the binomial and negative binomial distributions of the annual number of exceedances is given by Eqs. (19) and (21), respectively. $\hat{x}_T$ is a function of the parameters α, β and γ. Its asymptotical sampling variance can be computed as

$$\mathrm{Var}(\hat{x}_T) = \left(\frac{\partial x_T}{\partial\alpha}\right)^2\mathrm{Var}(\hat{\alpha}) + \left(\frac{\partial x_T}{\partial\beta}\right)^2\mathrm{Var}(\hat{\beta}) + \left(\frac{\partial x_T}{\partial\gamma}\right)^2\mathrm{Var}(\hat{\gamma}) + 2\frac{\partial x_T}{\partial\alpha}\frac{\partial x_T}{\partial\gamma}\mathrm{Cov}(\hat{\alpha},\hat{\gamma}) \tag{A11}$$

assuming that β is independent of α and γ.

For binomial distribution (Eq. (19))

$$\frac{\partial\hat{x}_T}{\partial\alpha} = \frac{\beta}{\alpha}, \qquad \frac{\partial\hat{x}_T}{\partial\beta} = \ln\frac{\alpha}{1-T_1^{1/\gamma}},$$
$$\frac{\partial\hat{x}_T}{\partial\gamma} = -\frac{T_1^{1/\gamma}\ln T_1}{1-T_1^{1/\gamma}}\frac{\beta}{\gamma^2}. \tag{A12}$$

Substituting Eqs. (A2)–(A6) and (A12) into Eq. (A11) and rearranging, an expression for the asymptotic sampling variance of $\hat{x}_T$ is obtained:

$$\mathrm{Var}(\hat{x}_T) = \frac{\beta^2}{N}\left\{\left(2 + \frac{3\alpha-1}{\alpha\gamma}\right)\frac{(1-\alpha)^2}{\alpha^2} + \frac{1}{\alpha\gamma}\ln^2\left(\frac{\alpha}{1-T_1^{1/\gamma}}\right) + \left[2(1-\alpha)\gamma - \frac{(1-2\alpha)^2}{\alpha}\right] \times \frac{1-\alpha}{\alpha^2\gamma^3}\frac{T_1^{2/\gamma}\ln^2 T_1}{(1-T_1^{1/\gamma})^2} - \frac{2(1-\alpha)}{\alpha^2\gamma^2} \times \left[\frac{(1-2\alpha)^2}{\alpha} - 2\gamma(1-\alpha)\right]\frac{T_1^{1/\gamma}\ln T_1}{1-T_1^{1/\gamma}}\right\}. \tag{A13}$$

For negative binomial distribution: (Eq. (22))

$$\frac{\partial \hat{x}_T}{\partial \alpha} = \frac{\beta}{\alpha(1-\alpha)},$$

$$\frac{\partial \hat{x}_T}{\partial \beta} = -\ln\left[(T_1^{-1/\gamma}-1)\left(\frac{1-\alpha}{\alpha}\right)\right], \tag{A14}$$

$$\frac{\partial \hat{x}_T}{\partial \gamma} = \frac{T_1^{-1/\gamma}\ln T_1}{T_1^{-1/\gamma}-1}\,\frac{\beta}{\gamma^2}.$$

Substituting Eqs. (A7)–(A10) and (A14) into Eq. (A11) and rearranging

$$\begin{aligned}\mathrm{Var}(\hat{x}_T) = \frac{\beta^2}{N}\Bigg\{&\left(2-\frac{2}{\gamma}-\frac{1}{\alpha\gamma}\right)\frac{1}{\alpha^2} \\ &+\frac{1-\alpha}{\alpha\gamma}\ln^2\left[\frac{(1-\alpha)(T_1^{-1/\gamma}-1}{\alpha}\right] \\ &+\frac{2\gamma-\dfrac{(1+\alpha)^2}{\alpha}}{\alpha^2\gamma^3}\,\frac{T_1^{-2/\gamma}\ln^2 T_1}{(T_1^{-1/\gamma}-1)^2} \\ &-\frac{2}{\alpha^2\gamma^2}\left[\frac{(1+\alpha)^2}{\alpha}-2\gamma\right]\frac{T_1^{-1/\gamma}\ln T_1}{T_1^{-1/\gamma}-1}\Bigg\}.\end{aligned} \tag{A15}$$

References

Ashkar, F., Rousselle, J., 1983. Some remarks on the truncation used in partial flood series models. Water Resour. Res. 19 (2), 477–480.

Ben-Zvi, A., 1991. Observed advantage for negative binomial over Poisson distribution in partial duration series. Stochastic Hydrol. Hydraul. 5, 135–146.

Cunnane, C., 1973. A particular comparison of annual maxima and partial duration series methods of flood frequency prediction. J. Hydrol. 18, 257–271.

Cunnane, C., 1979. A note on the Poisson assumption in partial duration series models. Water Resour. Res. 15 (2), 489–494.

Kirby, W., 1969. On the random occurrence of major floods. Water Resour. Res. 5 (4), 778–789.

Lang, M., Rasmussen, P., Oberlin, G., Bobée, B., 1997. Over-threshold sampling: modeling of occurrences by renewal processes. Revue des Sciences de l'Eau 3, 279–320.

Lang, M., 1999. Theoretical discussion and Monte-Carlo simulations for a negative binomial process paradox. Stoch. Env. Res. Risk Assessment 13, 183–200.

Lang, M., Ouarda, T.B.M.J., Bobée, B., 1999. Towards operational guidelines for over-threshold modelling. J. Hydrol. 225, 103–117.

Langbein, W.B., 1949. Annual floods and the partial-duration flood series. Transactions, AGU 30 (6), 879–881.

Madsen, H., Rasmussen, P.F., Rosbjerg, D., 1997. Comparison of annual maximum series and partial duration series methods for modelling extreme hydrologic events. 1. At-site modelling. Water Resour. Res. 33 (4), 747–757.

NERC (National Environment Research Council), 1975. Flood Studies Report, vol. 1.

Shane, R.M., Lynn, W.R., 1964. Mathematical model for flood risk evaluation. J. Hydraul. Div. ASCE 90 (HY6), 1–20.

Taesombut, V., Yevjevich,V., 1978. Use of partial duration series for estimating the distribution of maximum annual flood peak. Hydrology Paper no. 97, Colorado State University, Fort Collins, CO, p. 71.

Rosbjerg, D., 1985. Estimation in partial duration series with independent and dependent peak values. J. Hydrol. 76, 183–195.

Rosbjerg, D., Madsen, H., Rasmussen, P.F., 1992. Prediction in partial duration series with generalized Pareto distributed exceedances. Water Resour. Res. 28 (11), 3001–3010.

Van Montfort, M.A.J., Witter, J.V., 1985. Testing exponentiality against generalized Pareto distribution. J. Hydrol. 78, 305–315.

Vukmirovic, V., Petrovic, J., 1997. Flood flow analysis using renewal processes. PHI-V, No.11, Unesco, Paris. 5. FRIEND-AMHY Annual Conference, Thessaloniki. 25–27 September 1995, pp. 159–169.

Williamson, E., Bretherton, M.H., 1963. Tables of the negative binomial probability distribution. Wiley, London, p. 275.

Zelenhasic, E., 1970. Theoretical probability distributions for flood peaks. Hydrology paper no. 42, Colorado State University, Fort Collins, CO, p. 35.