Water Resources Management (2005) 19: 363–374
DOI: 10.1007/s11269-005-2073-2

Analysis of Extreme Flood Events for the Pachang River, Taiwan

S. NADARAJAH[1,*] and J. T. SHIAU[2]
[1] *Department of Mathematics, University of South Florida, Tampa, Florida 33620, USA;*
[2] *Department of Water Resources and Environmental Engineering, Tamkang University, Tamsui, 251, Taiwan, ROC; e-mail: jtshiau@mail.tku.edu.tw*
*(*author for correspondence, e-mail: snadaraj@chuma1.cas.usf.edu)*

(Received: 14 September 2004; in final form: 23 December 2004)

Abstract. Flood events of the Pachang River, one of the major rivers in Taiwan, are modeled by extreme value distributions. Flood events are characterized by its peak, volume, duration and the time of peak. Flood volume and peak are fitted to a generalized extreme value distribution. Flood duration and the time of flood peak are incorporated into the model to detect possible trends. The results show that flood volume exhibits an upward trend with respect to flood duration, but flood peak exhibits a downward trend with respect to flood duration. There appears to be no significant trends with respect to time. Among other results, we provide estimates of return period for flood peak and flood volume, which could be used as measures of flood protection. This paper provides the first application of extreme value distributions to flood data from Taiwan.

Key words: generalized extreme value distribution, Gumbel distribution, partial duration series

1. Introduction

Severe damage caused by extreme hydrologic events, such as floods and droughts, had been recorded throughout the human civilization. In order to mitigate the negative impacts of extreme hydrologic events, the characteristics of such events (for example, rainfall intensity, flood peak, etc.) were used as a design criterion of hydraulic structures. However, available data of extreme events were insufficient to precisely assess the risk of extreme events. Therefore, probabilistic assessment was developed and extreme value theory is one of the common methods to investigate the extreme phenomenon of interest (see, for example, Stedinger *et al.*, 1993).

Extreme value theory deals with the stochastic behavior of the extreme values of a process. For a single process the behavior of the maxima can be described by the three extreme value distributions – Gumbel, Frechet and negative Weibull – as suggested by Fisher and Tippett (1928). Kotz and Nadarajah (2000) indicated that the extreme value distributions can be traced back to the work done by Bernoulli in 1709. Their first application to flood flow was probably made by Fuller in 1914. In the 1940s, Gumbel applied extreme value distributions to analyze flood flow data. In subsequent work he continued his discussion on estimation and forecasting of floods. Thereafter, several researchers have provided useful applications of extreme value distributions to flood data from the different regions of the world: see Pitlick (1994) and Provaznik and Hotchkiss (1998) for applications in the U.S.A.; El-Jabi *et al.* (1998) and Yue *et al.* (1999) for applications in Canada; Haktanir and Horlacher (1993) for applications in Germany; Enjo *et al.* (2001) for applications in Spain; Cannarozzo *et al.* (1995) and De Michele and Rosso (2001) for applications in Italy; Madesen *et al.* (1997) for applications in New Zealand; Vogel *et al.* (1993) for applications in Australia; Sankarasubramanian and Srinivasan (1999) for applications in India; Karim and Chowdhury (1995) for applications in Bangladesh; Phien and Laungwattanapong (1991) for applications in Thailand; Hoybye and Iritz (1997) for applications in mainland China; Sene *et al.* (2001) for applications in Lebanon; and, Mkhandi *et al.* (2000) for applications in southern Africa. For a review of applications of extreme value distributions to climate data see Farago and Katz (1990).

Flooding in Taiwan is a common phenomenon and often it causes considerable damage. For example, on August of 1997 heavy floods and destructive landslides in northern Taiwan caused millions of dollars in damage and economic losses, and killed over 40 people. Agriculture has been the hardest hit sector. In spite of this, there has been no work concerning extreme values of flood data from Taiwan. As a matter of fact, there has been little concerned with extreme values of any kind of climate data from Taiwan – the only piece of work that we are aware of is Yim *et al.* (1999) where extreme value distributions are applied to study the wind climate at Taichung harbour. Our paper provides the first application of extreme value distributions to flood data from Taiwan. Among other things, we provide estimates of return period for flood peak and flood volume, which could be used as measures of flood protection.

We use streamflow data of the Pachang River located in southern Taiwan. The Pachang river is one of the major rivers in Taiwan. It is 81 km long and has a drainage area of 475 square km. It passes through Chiayi and Tainan counties of Taiwan. In year 2000 the Pachang River drew a great deal of international attention after 4 workers were swept to their deaths by its raging flood waters. During a flood event in 2001 waters from the river reached up to one-storey high.

Traditionally, the three extreme value distributions are applied to annual maximum floods. The major drawback with this is that annual maximum series uses only the largest event in each year, regardless of whether the second largest event in a year exceeds the largest event of the other year. Partial duration series makes more use of the data available and yield more accurate estimates of extreme events.

This study investigates the extreme value distributions of partial duration floods exceeding a given threshold. Two important flood characteristics, flood peak and flood volume, are modeled separately by the generalized extreme value (GEV) distribution. The GEV distribution has all the flexibility of the three extreme value distributions – Gumbel, Frechet and negative Weibull. It was developed by Jenkinson (1955); see also Hosking *et al.* (1985) and Galambos (1987). Its cumulative distribution function is:

$$F(x) = \exp\left\{-\left(1 + \xi\frac{x - \mu}{\sigma}\right)^{-1/\xi}\right\}, \tag{1}$$

where μ, σ, and ξ are referred to the location, scale, and shape parameters, respectively; and $1 + \xi(x - \mu)/\sigma$ should be greater than 0 in (1). The particular case of (1) for $\xi = 0$,

$$F(x) = \exp\left\{-\exp\left(-\frac{x - \mu}{\sigma}\right)\right\} \tag{2}$$

is the Gumbel distribution, while the cases $\xi > 0$ and $\xi < 0$ are known as the Frechet and the negative Weibull distributions, respectively. For flood flows, usually $\xi > 0$, although sometimes the Gumbel distribution is adequate.

Goel *et al.* (1998) and Yue *et al.* (1999) suggested that flood flows are not only determined by their peak and volume, but also by other characteristics such as flood duration and the time of flood peak. So it is natural to ask how the extreme values of flood volume and flood peak vary with respect to the duration and time. Because of factors such as global climate change and river basin change it is possible that the extreme values of flood volume and flood peak could exhibit an upward or a downward trend. An important part of our paper is to determine this variation and for this reason we shall consider a number of variants of (1) or (2) with the location parameter μ expressed as functions of flood duration and the time of flood peak.

2. The Data

The daily streamflow data from the Pachang River located in southern Taiwan is used. Thirty-nine yearly daily streamflow records, from 1961 to 1999, are employed to investigate the extreme floods. Flood events are defined as daily streamflow exceeding a given threshold, which is selected as 100 cms in this study, see Figure 1. The following four properties are used to characterize flood events. Flood peak is defined as the maximum daily flow during the flood period, flood volume is defined as the cumulative flow volume during the flood period, flood duration is defined as the time period when the flow exceeds the threshold, and the time of peak is defined as the number of days counting 1/1/1961 to the day of flood peak. Fifty flood events were abstracted from the daily streamflow data of the Pachang River.

For the fifty events abstracted, flood duration ranged from 2 days to 11 days while the time of flood peak ranged from 8/8/1961 (220 days from 1/1/1961) to 8/12/1999 (14103 days from 1/1/1961). Figure 2 shows the variation of both the flood volume and flood peak with respect to flood duration. Figure 3 shows the corresponding variation with respect to the time of peak.

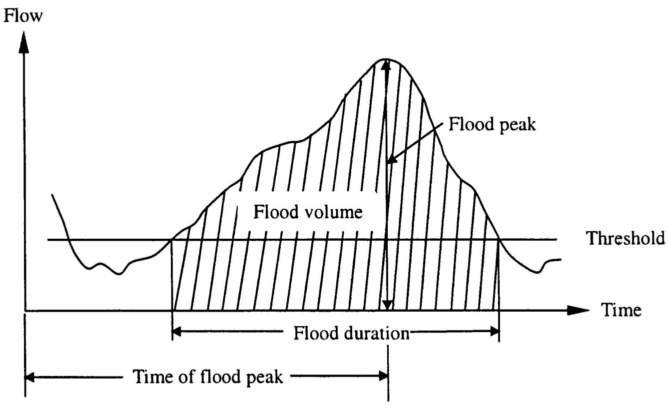

Figure 1. Characteristics of a flood event.

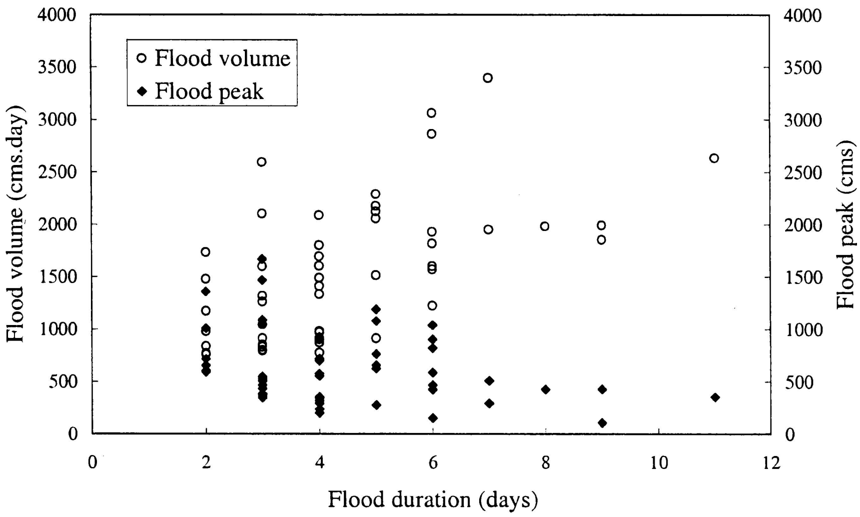

Figure 2. Flood volume and flood peak with respect to flood duration.

3. Model Fitting

The basic model fitted was (1) with μ, σ, and ξ constant (to be referred to as Model 1), using method of maximum likelihood to accomplish the fitting (Prescott and Walden, 1980). As mentioned in Section 1, sometimes the Gumbel distribution gives as good a fit as (1) for flood flow data, so we also fitted (2) with μ and σ constant (to be referred to as Model 2).

Model 2 is a submodel of model 1, so a standard way of determining the best fit model is the likelihood ratio test (Wald, 1943). If L_1 is the maximum likelihood for the three parameter model 1 and L_2 is the maximum likelihood for the two parameter model 2, then under the simpler model the test statistic $\lambda = -2 \log (L_2/L_1)$ would be assumed to be distributed as a chi-square variable with 1 degree of freedom

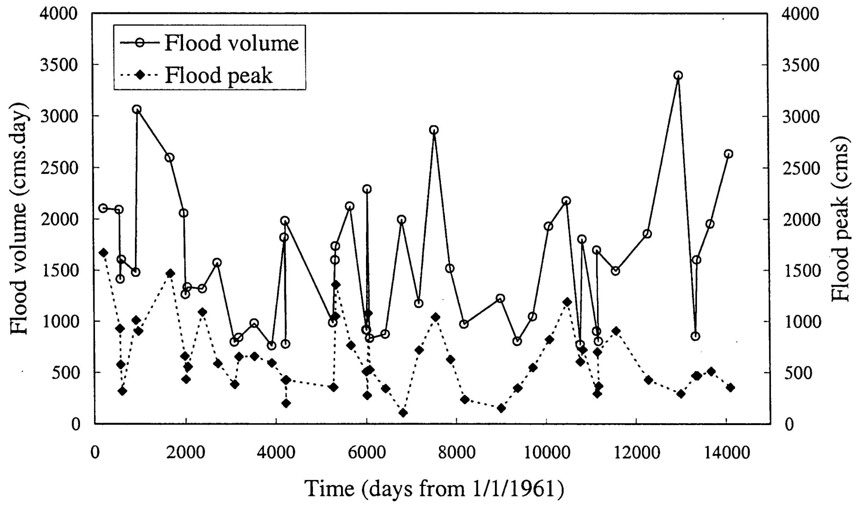

Figure 3. Flood volume and flood peak with respect to the time of flood peak.

(since the number of parameters differ by 1). In hypothesis testing problems this would be asymptotically true as the number of data tends to infinity. Thus, at the 5% significance level, the simpler two parameter model would be preferred if $-2\log(L_2/L_1) < \chi^2_{1,0.95} = 3.841$. In practice, because of the lack of complete independence of the partial duration floods, this would probably have to be interpreted conservatively.

Figures 2 and 3 suggest that flood volume and flood peak could possibly exhibit significant trends with respect to flood duration and time. To investigate this, we tried a number of variations of (2). These included a three parameter model with μ allowed to vary linearly with flood duration (D):

$$\text{Model 3: } \mu = a + bD, \quad \sigma \text{ constant;}$$

a three parameter model with μ allowed to vary linearly with the time of peak (T):

$$\text{Model 4: } \mu = c + dT, \quad \sigma \text{ constant;}$$

and a four parameter model with μ allowed to vary linearly with both flood duration (D) and the time of peak (T):

$$\text{Model 5: } \mu = e + fD + gT, \quad \sigma \text{ constant.}$$

We also fitted variants of (2) with σ allowed to vary with flood duration and the time of peak, but these models did not appear to improve significantly on the ones we have mentioned above. The standard likelihood ratio test was used to determine whether the trends described by the models 3–5 are significant or not.

The goodness of fit these models was examined by QQ plots, where the observed quantile is plotted against the quantile predicted by the fitted model. For example, to check the goodness of fit of model 1 for flood volume data, we would plot the sorted values (in the ascending order) of the observed flood volume data versus

$$\mu - \frac{\sigma}{\xi}\left[1 - \left\{-\log\left(\frac{i}{51}\right)\right\}^{-\xi}\right], \quad i = 1, \ldots, 50,$$

which are the quantiles expected under model 1. Similarly to check the goodness of fit of model 2 for flood volume data, we would plot the sorted values of the observed flood volume data versus

$$\mu - \sigma \ \log\left\{-\log\left(\frac{i}{51}\right)\right\}, \quad i = 1, \ldots, 50.$$

For models 3–5 we take a slightly different approach since $\mu = \mu(D, T)$ is varying: transform the observed data X to the new variable Z defined by

$$Z = \mu_0 + \sigma_0 \frac{X - \mu(D, T)}{\sigma},$$

for some constants μ_0 and σ_0 (for convenience chosen to be the fitted values from the two parameter model 2). Then, according to any of the models 3, 4 or 5, for all D and T, Z will follow the distribution (2) with parameters μ_0 and σ_0. QQ plots were derived for the transformed variable Z.

4. Results and Discussion

4.1. FLOOD VOLUME DATA

When fitted using the method of maximum likelihood, model 1 gave the estimates $\mu = 1217.9$, $\sigma = 473.8$ and $\xi = 0.126$ with the negative logarithm of the maximum likelihood $-\log L_1 = 390.7$. On the other hand, model 2 gave the estimates $\mu = 1251.6$ and $\sigma = 503.9$ with $-\log L_2 = 391.0$. Since $-2\log(L_2/L_1) = 2(391.0 - 390.7) < 3.841$ it follows by the standard likelihood ratio test that model 2 should be preferred to model 1 (in other words, the Gumbel distribution provides as good a fit as the more general GEV distribution). This finding is supported by Figure 4, where we have plotted the QQ plots for models 1 and 2 – there is little to choose between the two as the plots look very similar.

Model 3 yielded the estimates $a = 517.2$, $b = 180.3$ and $\sigma = 391.9$ with $-\log L_3 = 378.9$. Comparing with model 2, we see $-2\log(L_2/L_3) = 24.073 > 3.841$, so flood volume exhibits a highly significant upward trend with respect to flood duration. We can say that extreme values of flood volume increase by 180.3 cms per unit change in flood duration.

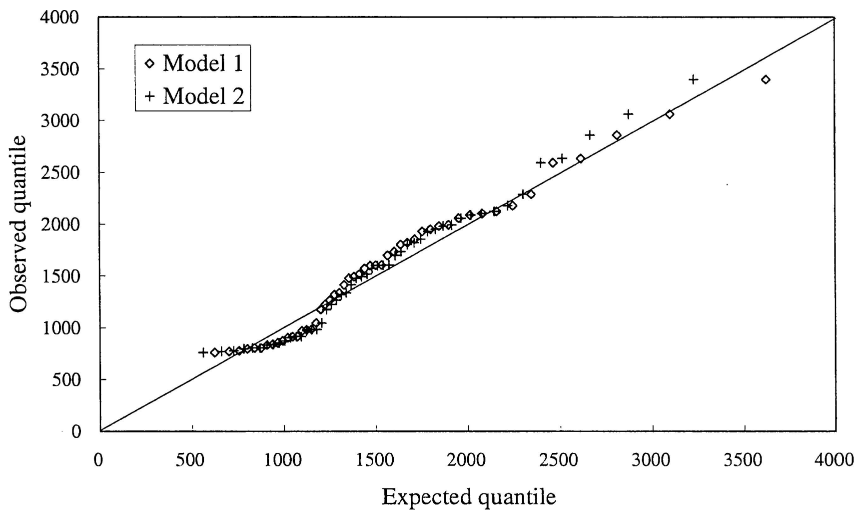

Figure 4. QQ plots for flood volume data, models 1 and 2.

Model 4 yielded the estimates $c = 1299.2$, $d = -0.007$ and $\sigma = 502.6$ with $-\log L_4 = 390.9$. Comparing with model 2, we see $-2\log(L_2/L_4) = 0.128 < 3.841$, so there is no evidence of significant trend with respect to time.

Model 5 yielded the estimates $e = 630.0$, $f = 193.1$, $g = -0.025$ and $\sigma = 502.6$ with $-\log L_5 = 377.8$. Comparing with model 3, we see $-2\log(L_3/L_5) = 2.254 < 3.841$, so model 5 does not improve significantly on model 3.

Hence we conclude that model 3 is the most reasonable one for flood volume data. Figure 5 shows the QQ plot for model 3. The fit is quite acceptable and

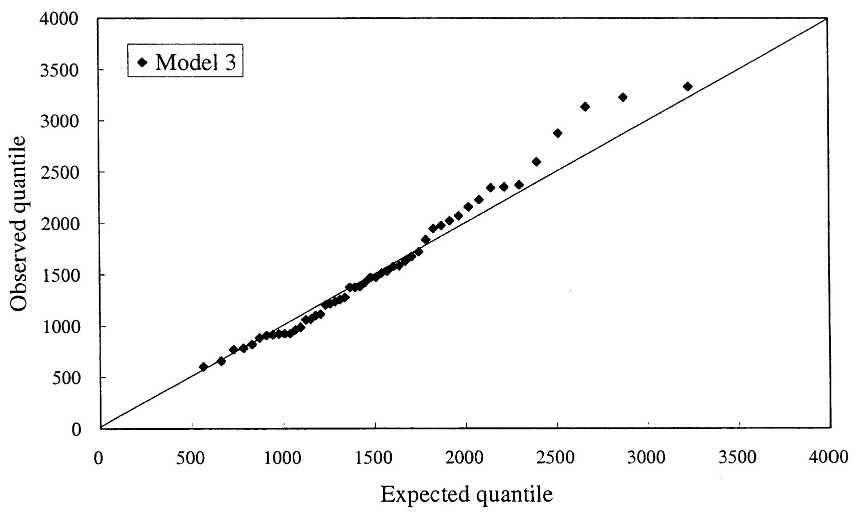

Figure 5. QQ plot for flood volume data, model 3.

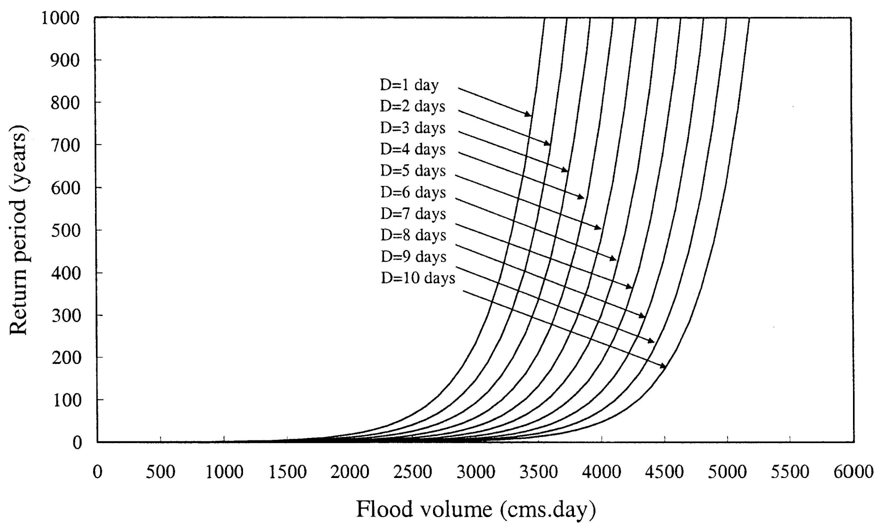

Figure 6. Return period for flood volume, model 3.

comparison with Figure 4 suggests that this fit is considerably better than the fit of the Gumbel or the GEV distribution with constant parameters. It is generally the case that there are few significant departures on the QQ plots.

If a random variable X is selected by the partial duration series method, then the return period of X exceeding a given value x is:

$$T = \frac{E(L)}{1 - F(x)}, \tag{3}$$

where $E(L)$ is the mean inter-arrival time between events and $F(x)$ is the cumulative distribution function of X. So, under the best fitting model 3, the return period of flood volume exceeding a given value x is:

$$T = \frac{E(L)}{1 - \exp\left\{- \exp\left(-\frac{x-517.2-180.3d}{391.9}\right)\right\}}.$$

The $E(L)$ is estimated from the observed data to be 283.2 days or 0.6526 years. Figure 6 shows how the return period varies with respect to the level of flood volume for a range of flood durations.

4.2. FLOOD PEAK DATA

Model 1 gave the estimates $\mu = 472.1$, $\sigma = 251.2$ and $\xi = 0.066$ with the negative logarithm of the maximum likelihood $-\log L_1 = 357.3$. On the other hand, model 2 gave the estimates $\mu = 481.1$ and $\sigma = 257.9$ with $-\log L_2 = 357.5$. Since $-2\log(L_2/L_1) = 2(357.5 - 357.3) < 3.841$ it follows again that model 2

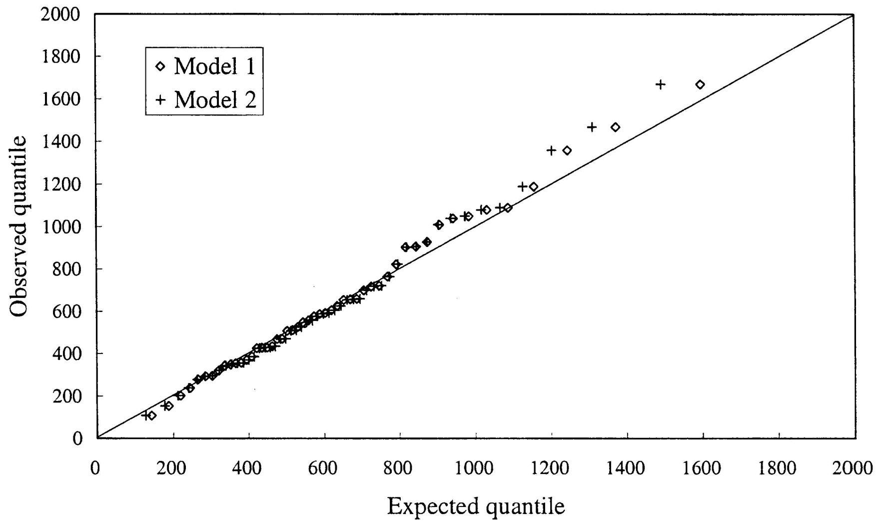

Figure 7. QQ plots for flood peak data, models 1 and 2.

should be preferred to model 1 (in other words, the Gumbel distribution provides as good a fit as the more general GEV distribution). This finding is supported by Figure 7, where we have plotted the QQ plots for models 1 and 2. There is little to choose between the two plots.

Model 3 yielded the estimates $a = 668.5$, $b = -40.7$ and $\sigma = 242.3$ with $-\log L_3 = 354.8$. Comparing with model 2, we see $-2\log(L_2/L_3) = 5.381 > 3.841$, so flood peak exhibits a significant downward trend with respect to flood duration. We can say that extreme values of flood peak decrease by 40.7 cms per unit change in flood duration.

Model 4 yielded the estimates $c = 580.5$, $d = -0.014$ and $\sigma = 254.2$ with $-\log L_4 = 356.3$. Comparing with model 2, we see $-2\log(L_2/L_4) = 2.282 < 3.841$, so again there is no evidence of significant trend with respect to time.

Model 5 yielded the estimates $e = 707.1$, $f = -36.6$, $g = -0.008$ and $\sigma = 241.6$ with $-\log L_5 = 354.4$. Comparing with model 3, we see $-2\log(L_3/L_5) = 0.780 < 3.841$, so model 5 again does not improve significantly on model 3.

Hence we conclude as before that model 3 is the most reasonable model for flood peak data. Figure 8 shows the QQ plot for model 3. The fit looks quite acceptable and comparison with Figure 7 suggests that model 3 provides at least as good a fit as the Gumbel or the GEV distribution with constant parameters.

Using Equation (3), the return period of flood peak exceeding a given value x is:

$$T = \frac{E(L)}{1 - \exp\left\{-\exp\left(-\frac{x - 668.5 + 40.7d}{242.3}\right)\right\}},$$

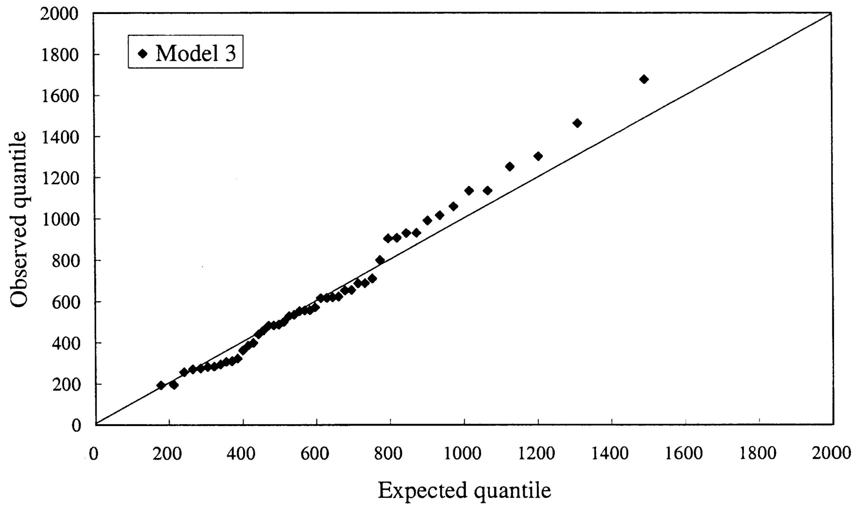

Figure 8. QQ plot for flood peak data, model 3.

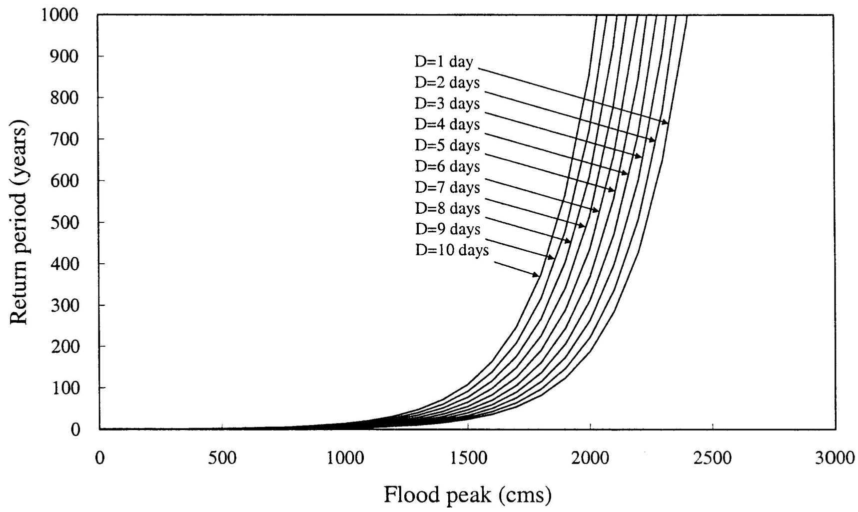

Figure 9. Return period for flood peak, model 3.

for the best fitting model 3. Figure 9 shows how the return period varies with respect to the level of flood peak for a range of flood durations.

5. Conclusions

This study provides the first application of extreme value distributions to flood data from Taiwan. The partial duration flood events of the Pachang River in southern Taiwan are used in the analysis. The following conclusions are drawn:

1. The Gumbel distribution (particular case of the GEV distribution when the shape parameter equals to zero) provides a reasonable model for both flood volume and flood peak.
2. Flood volume has a highly significant upward trend with respect to flood duration corresponding to an increase of 180.3 cms per unit change in flood duration.
3. Flood peak has a significant downward trend with respect to flood duration corresponding to a decrease of 40.7 cms per unit change in flood duration.
4. Both flood volume and flood peak exhibit no significant trends with respect to time.

Acknowledgments

The authors would like to thank the referees and the editor for carefully reading the paper and for their great help in improving the paper.

References

Cannarozzo, M., Dasaro, F., and Ferro, V., 1995, 'Regional rainfall and flood frequency-analysis for Sicily using the 2-component extreme-value distribution', *Hydrological Sciences Journal – Journal Des Sciences Hydrologiques* **40**, 19–42.

El-Jabi, N., Ashkar, F., and Hebabi, S., 1998, 'Regionalization of floods in New Brunswick (Canada)', *Stochastic Hydrology and Hydraulics* **12**, 65–82.

Enjo, M. A. A., Agudo, J. P., and Viqueira, F. D. F., 2001, 'Regional flood analysis at the Atlantic coast of Galicia northwest of Spain', *Ingenieria Hidraulica en Mexico* **16**, 67–76.

Farago, T. and Katz, R., 1990, 'Extremes and Design Values in Climatology', World Meteorological Organization, WCAP-14, WMO/TD-No. 386.

Fisher, R. A. and Tippett, L. H. C., 1928, 'Limiting forms of the frequency distribution of the largest or smallest member of a sample', *Proceedings of the Cambridge Philosophical Society* **24**, 180–290.

Galambos, J., 1987, *The Asymptotic Theory of Extreme Order Statistics* (second edition). Krieger, Melbourne, Florida.

Goel, N. K., Seth, S. M., and Chandra, S., 1998, 'Multivariate modeling of flood flows', *Journal of Hydraulic Engineering, ASCE* **124**, 146–155.

Haktanir, T. and Horlacher, H. B., 1993, 'Evaluation of various distributions for flood frequency-analysis', *Hydrological Sciences Journal – Journal Des Sciences Hydrologiques* **38**, 15–32.

Hosking, J. R. M., Wallis, J. R., and Wood, E. F., 1985, 'Estimation of the generalized extreme value distribution by the method of probability weighted moments', *Technometrics* **27**, 251–261.

Hoybye, J. and Iritz, L., 1997, 'Analysis of extreme hydrological events in a monsoon climate catchment: The Hongru River, China', *Hydrological Sciences Journal – Journal Des Sciences Hydrologiques* **42**, 343–356.

Jenkinson, A. F., 1955, 'The frequency distribution of the annual maxima (or minima) values of meteorological elements', *Quarterly Journal of the Royal Meteorological Society* **81**, 158–171.

Karim, M. A. and Chowdhury, J. U., 1995, 'A comparison of 4 distributions used in flood frequency-analysis in Bangladesh', *Hydrological Sciences Journal – Journal Des Sciences Hydrologiques* **40**, 55–66.

Kotz, S. and Nadarajah, S., 2000, *Extreme Value Distributions: Theory and Applications*. Imperial College Press, London.

Madsen, H., Pearson, C. P., and Rosbjerg, D., 1997, 'Comparison of annual maximum and partial duration series methods for modeling extreme hydrological events. 2. Regional modeling', *Water Resources Research* **33**, 759–769.

De Michele, C. and Rosso, R., 2001, 'Uncertainty assessment of regionalized flood frequency estimiates', *Journal of Hydrological Engineering* **6**, 453–459.

Mkhandi, S. H., Kachroo, R. K., and Gunasekara, T. A. G., 2000, 'Flood frequency analysis of southern Africa: II. Identification of regional distributions', *Hydrological Sciences Journal – Journal Des Sciences Hydrologiques* **45**, 449–464.

Phien, H. N. and Laungwattanapong, N., 1991, 'At-site flood frequency-analysis for Thailand', *Water SA* **17**, 147–154.

Pitlick, J., 1994, 'Relation between peak flows, precipitation, and physiography for 5 mountainous regions in the western USA', *Journal of Hydrology* **158**, 219–240.

Prescott, P. and Walden, A.T., 1980, 'Maximum likelihood estimation of the parameters of the generalized extreme-value distribution', *Biometrika* **67**, 723–724.

Provaznik, M. K. and Hotchkiss, R. H., 1998, 'Analysis of gauging station flood frequency estimates in Nebraska using L-moments and region of influence methods', *General Design and Roadside Safety Features Transportation Research Record* **1647**, 53–60.

Sankarasubramanian, A. and Srinivasan, K., 1999, 'Investigation and comparison of sampling properties of L-moments and conventional moments', *Journal of Hydrology* **218**, 13–34.

Sene, K. J., Houghton-Carr, H. A., and Hachache, A., 2001, 'Preliminary flood frequency estimates for Lebanon', *Hydrological Sciences Journal – Journal Des Sciences Hydrologiques* **46**, 659–676.

Stedinger, J. R., Vogel, R. M., and Foufoula-Georgiou, E., 1993, 'Frequency analysis of extreme events', in D. Maidment (ed), *Handbook of Hydrology*, chapter 18.

Vogel, R. M., McMahon, T. A., and Chiew, F. H. S., 1993, 'Floodflow frequency model selection in Australia', *Journal of Hydrology* **146**, 421–449.

Wald, A., 1943, 'Tests of statistical hypotheses concerning several parameters when the number of observations is large', *Transactions of the American Mathematical Society* **54**, 426–483.

Yim, J. Z., Lin, J. G., and Hwang, C. H., 1999, 'Statistical properties of the wind field at Taichung harbour, Taiwan', *Journal of Wind Engineering and Industrial Aerodynamics* **83**, 49–60.

Yue, S., Ouarda, T. B. M. J., Bobee, B., Legendre, P., and Bruneau, P., 1999, 'The Gumbel mixed model for flood frequency analysis', *Journal of Hydrology* **226**, 88–100.