Extraction of threshold for partial duration series (Case Study: Mashhad, Iran)

Hamid Reza Asgari, Department of Civil Engineering, Shirvan Branch, Islamic Azad University, Shirvan, Iran
He_asgari@yahoo.com
Ali Ghiami Bajgirani, Assistant Professor, Department of Civil Engineering, The Institute of Higher Education of Eqbal Lahoori, Mashhad, Iran
Shokoufeh Pourshahabi, Lecturer, Department of Civil Engineering, Eghbal Lahoori Institute of Higher Education, Mashhad, Iran
Bahare Asgari, Student of Civil Engineering Department, Khavaran Institute of Higher Education, Mashhad, Iran

Paper Reference Number: 6-11-16-0140
Name of the Presenter: Hamid Reza Asgari

Abstract

The Design of water structures requires an in-depth understanding of extreme events. Analysis of the extreme events includes considering the minimum or maximum values that the maximum amount of rainfall expected during a specific time period is an example of it. The partial duration series (PDS) deals with the selection of over-threshold values. In this paper, for extreme values of rainfall, a threshold level is selected which leads to the best fitness between the PDS extracted and generalized Pareto distribution (GPD). The daily rainfall data for 44 years of Mashhad synoptic station is used in this paper. The comparison between the right tail of the GPD curve and probability distribution for observed events shows that the threshold of 17 millimeters is a better option for extracting of PDS.

Key words: Partial duration Series, Generalized Pareto distribution, Rainfall threshold.

1. Introduction

The Design of water structures requires an in-depth understanding of probabilistic behavior of extreme events. Analysis of the extreme events includes considering the minimum or maximum values that the maximum amount of rainfall expected during a specific time period is an example of it. The data of extreme events can be selected as the annual maximum series (AM) or PDS. A series of the maximum event in each year (i.e. one event per year) leads to the AM series which is equal to the number of years. In the AM series, maximum values in the second or third position of one year may be significant quantities over the whole period, but do not enter the calculations. On the other side, the PDS series deals with the selection of peaks-over-threshold values (POT). A PDS series includes all peak values (X_s) that exceeds a specific threshold (S). In the PDS approach, the major problem is the choice of threshold. Although several methods have been proposed for threshold selection, but selecting it still depends on the expert judgments.

POT modeling is a compromise between AM and time series modeling. Van Montfort and Witter (1985) investigated the use of the GPD instead of the exponential distribution for the magnitude of exceedances. Lang et al. (1999) concluded that there is no unique specific value for threshold selection, but it is possible to specify a range of threshold values for which the POT approach leads to similar results. Onoz and Bayazit (2001) investigated the effect of the selected threshold in different frequency estimations. They showed that in some peak-over-threshold series, variance of the annual number of exceedances is significantly smaller or larger than the mean. Begueria (2004) considered the effect of the selected threshold in the estimation of GPD parameters. He determined a range of threshold values for which the data of 12 rain-gauge stations proved to fit adequately the GPD model. Michele and Salvadori (2004) showed that estimation of parameters in the GPD model does not require all data and only a proper subset of the sample is sufficient, containing the most relevant information for estimating a given parameter. In this paper, different methods of threshold selection and PDS modeling based on the GP distribution are discussed. The results obtained for 44 years of daily rainfall in Mashhad synoptic station are illustrated.

2. Data and Material

In this paper, the daily rainfall data for 44 years of Mashhad synoptic station in the period 1961-2004 has been used. The maximum daily rainfall observed was 52 mm among 16073 records. The average daily rainfall during the period was 0.7 mm and the standard deviation was 2.8.

3. POT Modeling

3-1-Threshold selection

With respect to the sampling variance of quantiles, Cunnane (1973) showed that mean number of peaks per year (μ) should be more than 1.65. Lang et al. (1999) recommended selecting the threshold with $\mu > 2$ or 3. Davison and Smith (1990) and Naden and Bayliss (1993) proposed to select the threshold value so as to be in the domain where the mean exceedance above threshold $E(X_s) - S$ is a linear function of the threshold level (S), where $E(X_s)$ is the mean value of exceedances. They showed that this method leads to good results when the POT distribution is fitted with GPD or an exponential distribution. Ashkar and Rousselle (1987) used the assumption of Poisson process as a basis for threshold selection. The threshold (S) is then selected so that the dispersion index ($I = Var(m_i)/E(m_i)$) is located in the interval $[I(0.05), I(0.95)]$, where m_i is the number of peaks above threshold in the year number i. Rosbjerg and Madsen (1992) proposed a method as $s = \hat{\mu} + k\hat{\sigma}$ based on the mean ($\hat{\mu}$) and standard deviation ($\hat{\sigma}$) of the original series. Begueria (2004) recommended $k = 3$.

3-2-Choice of the model for exceedances magnitudes

After threshold selection, statistical distributions with positive skew can be used. Both GP and exponential distribution has been widely used by researchers for POT modeling (Van Montfort and Witter, 1986; Madsen and Rosbjerg, 1997; Onoz and Bayazit, 2001; Begueria, 2004). The GP distribution is described by a shape parameter k, a scale parameter α, and a location parameter s, and has the following cumulative distribution function:

$$G_s(X) = P(X \le x \mid s,\alpha,k) = 1-(1-k\frac{(X-s)}{\alpha})^{\frac{1}{k}} \qquad k \neq 0 \tag{1}$$

$$G_s(X) = P(X \le x \mid s,\alpha,k) = 1-\exp(-\frac{X-s}{\alpha}) \qquad k = 0 \tag{2}$$

The GP distribution converts to the exponential distribution, when $k=0$. Rosbjerg et al. (1992) showed that exponential distribution should be preferred, when $0<k<0.1$. For $k<0$, the distribution has a long tail at the right and the GP distribution can be used. For $k>0$, the right tail of the distribution is limited, $s \le X \le s+\alpha/k$, and GPD should be used with caution. The lower bound is the threshold s and there is no need to estimate it in POT modeling, but the estimates of GP parameters, α and k, can be obtained by different procedures.

4. Results and Analysis

For modeling the PDS, mean number of peaks per year (μ) and the mean exceedance above threshold $E(X_s)-S$ are plotted versus a range of threshold values (Fig. 1). As illustrated in this figure, the condition $\mu>1.65$ will be met when the threshold is less than 20 mm and the mean exceedance above threshold $E(X_s)-S$ is a linear function of the threshold level (S), when the threshold is less than 19 mm. This implies that a threshold equal to or less than 19 mm is suitable for the GP distribution. The method proposed by Rosbjerg and Madsen (1992) leads to the threshold of 9.1 mm, when $k=3$.

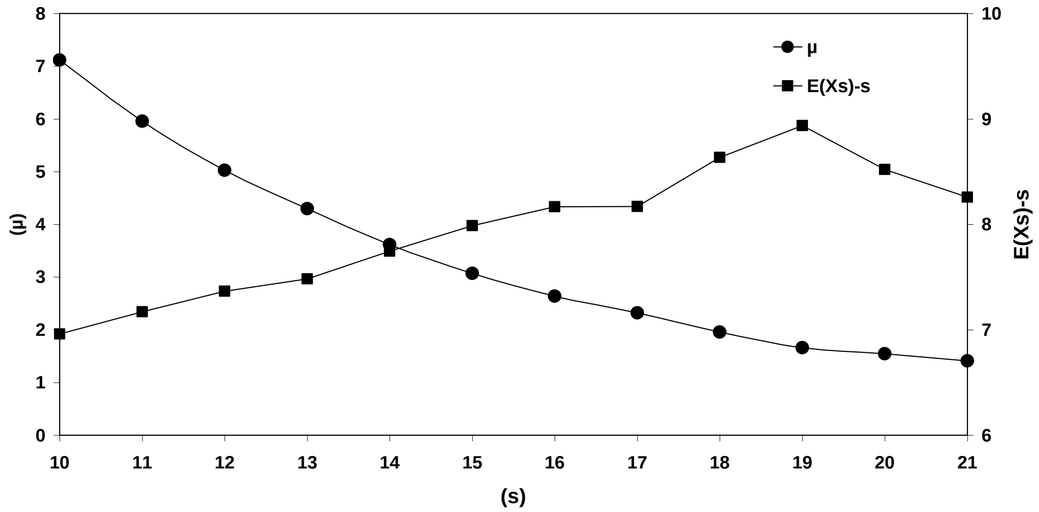

Fig 1: Mean number of peaks, μ, and the mean exceedance, $E(X_s)-S$, versus threshold

In order to determine the magnitudes of the events, the GP model is used that is valid for the threshold up to 19 mm. The parameters of the model are estimated using maximum likelihood method. Fig. 2 shows that for threshold less than 15 mm, where $k>0$, the GP distribution has not a long tail at right and analysis of the extreme events can not be used. On the other hand, for some thresholds (the range of 6 to 15), shape parameter is less than 0.1, so it is better to use the exponential distribution instead of GPD. In fact, for thresholds more than 15 mm that $k<0$, GPD can be used for extreme values. Thus, the GP analysis is limited to the thresholds 16 to 19. The comparison between the right tail of the GPD curve and probability distribution for observed events shows that the threshold of 17 mm leads to better fitness than the other thresholds (Fig. 3). The horizontal axis shows rainfall and the vertical axis shows log (P (X> x)).

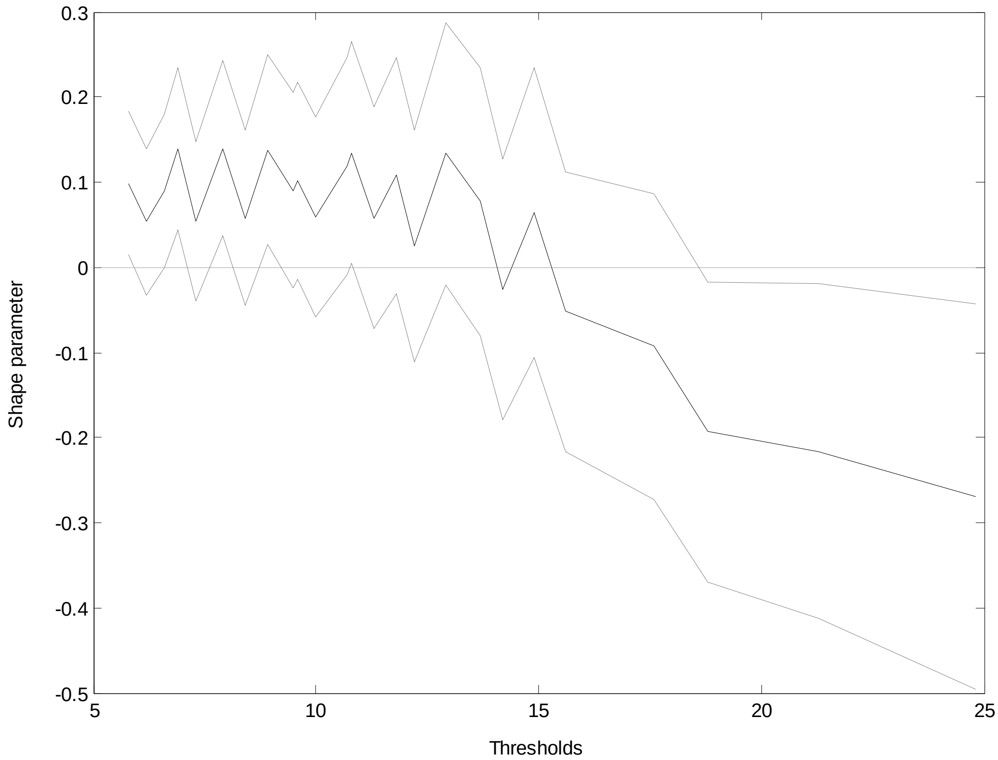

Fig 2: Changes in the shape parameter in GPD versus threshold (confidence level of 95%)

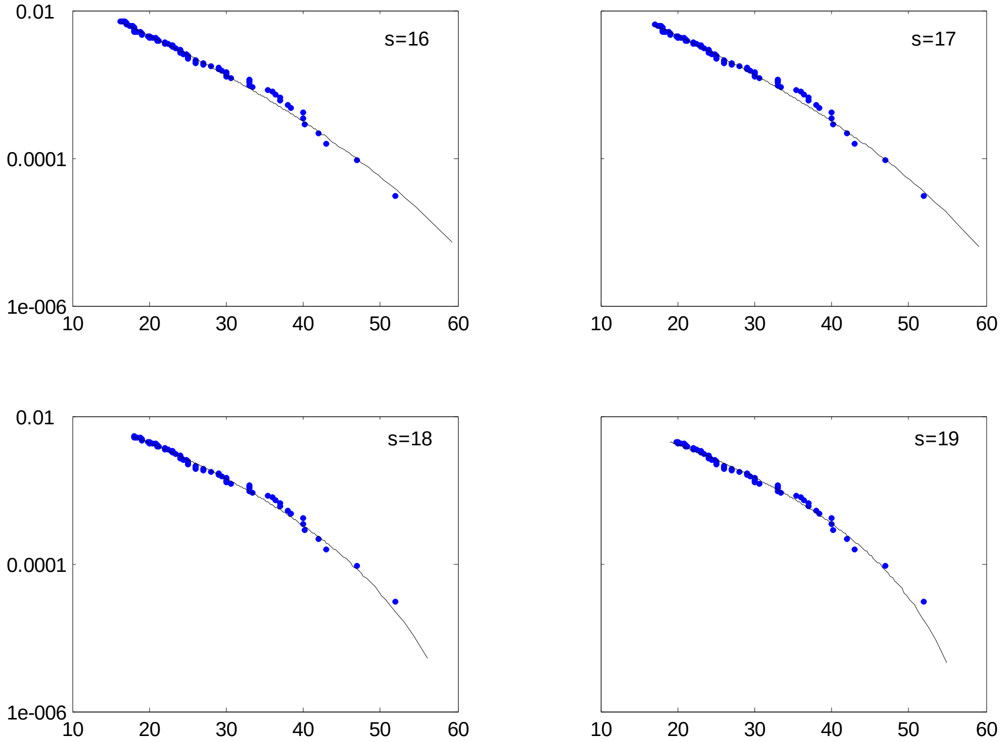

Fig 3: The right side of the GP distribution versus threshold.

5. Conclusions

As it was observed, the major problem in the PDS modeling is the choice of threshold. Choosing a threshold that will produce results distinct than the adjacent values, is not possible. As illustrated in Figure 3, the results of the threshold of 16 to 19 are very similar.

Therefore, a set of techniques should be used to determine an acceptable range for threshold and to model peaks-over-threshold properly. Choosing a threshold level of 17 mm, leading to $\mu > 2$ that proves that the PDS series is twice as length as the AM series. Increasing data can help in better estimation of parameters of the model. The climate and the main goal for determination of rainfall threshold (such as floods, erosion, etc.) are also important factors that will be helpful in selecting the threshold apart from the statistical analysis. After choosing a threshold, all of the peaks-over-threshold can be analyzed statistically and the maximum daily rainfall for various return periods can be determined. In conclusion, determination of the initial data for such analysis is a process of trial and error and the threshold depends on the type of statistical distribution.

References

Ashkar, F., & Rousselle, J. (1987). Partial duration series modeling under the assumption of a poissonian flood count. *Journal of Hydrology,* 90, 135–144.

Begueria, S. (2004). Uncertainties in partial duration series modelling of extremes related to the choice of the threshold value. *Journal of Hydrology*, 1-16.

Castillo, E., Hadi, A.S., Balakrishnan, N., & Sarabia, J.M. (2005). *Extreme Value and Related Models with Applications in Engineering and Science*. John Wiley Publication, NJ.

Coles, S., Pericchi, L.R., & Sisson, S. (2003). A fully probabilistic approach to extreme rainfall modeling. *Journal of Hydrology*, 273, 35-50.

Cunnane, C. (1973). A particular comparison of annual maxima and partial duration series methods of flood frequency prediction. *Journal of Hydrology*, 18, 257–271.

Davison, A.C., & Smith, R.L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society B* 52 (3), 393–442.

Madsen, H., & Rosbjerg, D. (1997). The partial duration series method in regional index-flood modeling. *Water Resources Research,* 33 (4), 737–746.

Naden, P.S., & Bayliss, A.C. (1993). Flood estimation: peak-overthreshold techniques. *MAFF Conf. of River and Coastal Engineers*, University of Loughborough, UK, 5–7 July, pp. 9.1.1– 9.1.18.

Michele, D.C., & Salvadori, G. (2004). Some hydrological applications of small sample estimators of Generalized Pareto and Extreme Value distributions. *Journal of Hydrology,* 301, 37-53.

Katz, R.W. (1999). Extreme value theory for precipitation: sensitivity analysis for climate change. *Advances in water Resources,* 23, 133-139.

Lang, M., Quarda, T.B.M.J., & Bobee, B. (1999). Towards operational guidelines for over- threshold modeling. *Journal of Hydrology*, 225, 103-117.

Onoz, B., & Bayazit, M. (2001). Effect of the occurrence process of the peaks over threshold on the flood estimates. *Journal of Hydrology,* 244, 86-96.

Pawlowsky-Glahn, V., Tolosana-Delgado, R., & Egozcue, J.J. (2005). Scale Effect in Hazard Assessment- Application to Daily Rainfall. *Advance in Geosciences*, 2, 117-121.

Rosbjerg, D., & Madsen, H. (1992). On the choice of threshold level in partial duration series. In: Østrem, G. (Ed.), *Nordic Hydrological Conference*, Alta. NHP Rep. 30, Oslo, pp. 604–615.

Rosbjerg, D., Madsen, H., & Rasmussen, P.F. (1992). Prediction in partial duration series with generalized Pareto distributed exceedances. *Water Resources Research,* 28 (11), 3001±3010.

Van Montfort, M.A.J. (1985). Testing exponentiality against generalized Pareto distribution. *Journal of Hydrology*, 78, 305-315.

Van Montfort, M.A.J., & Witter, J.V. (1986). The Generalized Pareto distribution applied to rainfall depths. *Hydrological Sciences Journal*, 31 (2), 151–162.

Wilks, D.S. (1995). *Statistical Methods in the Atmospheric Sciences: An Introduction.* Academic Press.

Yoo, C., Lee, J.S., & Yoon, Y.N. (2001). Climatological Thresholds of Daily Rainfall. *Journal of Hydrologic Engineering*, Vol. 6, No. 5, 443-449.