Evaluation of the best fit distribution for partial duration series of daily rainfall in Madinah, western Saudi Arabia

FAHAD ALAHMADI[1,2], NORHAN ABD RAHMAN[1,3] & MOHAMED ABDULRAZZAK[3]

1 Department of Hydraulic and Hydrology, Faculty of Civil Engineering, Universiti Teknologi Malaysia, Skudai Campus, 81310 Johor Bahru Johor, Malaysia
f3@dr.com
2 Madinah Water Directorate, Madinah, Kingdom of Saudi Arabia
3 Department of Civil Engineering, College of Engineering, Taibah University, Madinah, Kingdom of Saudi Arabia

Abstract Rainfall frequency analysis is an essential tool for the design of water related infrastructure. It can be used to predict future flood magnitudes for a given magnitude and frequency of extreme rainfall events. This study analyses the application of rainfall partial duration series (PDS) in the vast growing urban Madinah city located in the western part of Saudi Arabia. Different statistical distributions were applied (i.e. Normal, Log Normal, Extreme Value type I, Generalized Extreme Value, Pearson Type III, Log Pearson Type III) and their distribution parameters were estimated using L-moments methods. Also, different selection criteria models are applied, e.g. Akaike Information Criterion (AIC), Corrected Akaike Information Criterion (AICc), Bayesian Information Criterion (BIC) and Anderson-Darling Criterion (ADC). The analysis indicated the advantage of Generalized Extreme Value as the best fit statistical distribution for Madinah partial duration daily rainfall series. The outcome of such an evaluation can contribute toward better design criteria for flood management, especially flood protection measures.

Key words rainfall frequency analysis; statistical distributions; L-moments; goodness-of-fit tests; arid region; Madinah, Saudi Arabia

INTRODUCTION

Extreme rainfall events in arid and extreme arid regions have resulted in loss of life and extensive infrastructures damage, especially in urban areas. Saudi Arabia, located in an arid region, has been experiencing major extreme rainfall events, generating flash floods with high frequencies and magnitudes in many of its major urban centres. The high urbanization rates in major cities, including Madinah, have resulted in extensive flood damage.

Modelling of daily rainfall using various statistical models has been carried out in different parts of the world to provide better understanding of its contribution in the generation of major flooding event and flood protection infrastructures. Rainfall frequency analyses have been used to develop the design storm needed as an input for hydraulic structure design and water resources planning and management. The analysis relates the magnitude of storm events to their occurrence frequency through an application of suitable probability distribution functions.

Estimation of extreme rainfall frequencies presents a challenge due to the extreme random nature of events and lack of long records. To overcome these constraints, some advanced statistical distributions are usually applied, e.g. the L-moment method in estimating the parameters. Different criteria were used to determine the best fit statistical distribution.

The Partial Duration Series (PDS) was recommended for modelling extreme data by several researchers (Hosking and Wallis 1987, Begueria 2005). For an arid region, PDS provide an alternative to overcome the disadvantages of the annual maximum series (AMS) as AMS uses the largest event in each year. Also, PDS avoids such problems by considering all independent peaks that exceed a specified threshold.

Selection of the best fit distribution and associated parameter estimation is an important step in frequency analysis. Laio *et al.* (2009), applied Akaike Information Criterion (AIC), Corrected Akaike Information Criterion (AICc), Bayesian Information Criterion (BIC) and Anderson-Darling Criterion (ADC) to verify and identify the probability distribution of hydrological extremes in selected catchments around the UK. They found the four techniques produced comparable quality results. Another study by Khaled and Rahman (2010) used the same criteria for the application of flood frequency analysis in Tasmania, Australia. However, their analyses indicated that ADC is more successful in recognizing the three-parameter distribution.

This study applied suitable techniques to select the best distribution function for rainfall analyses using partial duration series in order to evaluate the impact of extreme events in Madinah, one of the major urbanized cities in Saudi Arabia.

THE STUDY AREA

Madinah, the fourth largest city in the Kingdom of Saudi Arabia (KSA), is located at 24°28′N latitude and 39°36′E longitude (Fig. 1). The city lies approximately 160 km inland from the Red Sea coast, at an elevation of about 600 m above mean sea level. Since 1960 the Ministry of Agriculture and Water, recently re-named as the Ministry of Water and Electricity (MoWE), has established a hydrological network over different parts of the country, including the Madinah region. Madinah area is characterized by an arid climate with high temporal and spatial variability, rainfall takes place primarily during winter and spring. The winter and spring rainfall events are caused by a combination of disturbances from the winter Mediterranean and the Sudan trough (Subyani *et al.* 2010); usually they generate extreme convective rainfall events over Madinah and surrounding areas. The average annual rainfall depth is estimated at 55 mm and the highest daily rainfall occurred on 20 March 1993 with 89.6 mm depth in 18 hours duration.

METHODS

The frequency analyses in this study focused on selecting the best fitting distribution for daily rainfall amount using several criteria of goodness-of-fit (GOF) test. The study is based on observed daily rainfall data and a 20 mm threshold is selected for partial duration series (PDS) analysis. The best distribution was chosen based on the minimum values specified by those criteria. Three steps were applied as follows; (1) applying model selection criteria, (2) estimating parameters using L-moments method, and (3) fitting probability distribution.

Applying the model selection criteria

The selection of the best fitting distribution has always been a key interest in the studies of rainfall amount. Thus, in this study the best fitting distribution for daily rainfall amount based on several criteria of GOF test was evaluated and also whether they came from the same probability distributions was checked. The rainfall analysis based on AIC, BIC, and ADC criteria, as adopted by Laio *et al.* (2009), was used for rainfall analysis.

(a) Akaike Information Criterion (AIC) – The Akaike information criterion (Akaike, 1973) is based on the use of Kullback-Leible's information as the discrepancy measure between the true model f(x) and the approximating model, $M_j = g_j(x,\vartheta)$. The AIC for the *j*th operational model can be computed as:

$$AIC = -2\prod(\theta) + 2p \tag{1}$$

When the sample size *n* is small, with respect to the number of estimated parameters *p*, the AIC may perform inadequately (Sugiura 1978). Therefore Sugiura derived a second-order variant of AIC, called AICc:

$$AICc = -2\prod(\theta) + 2p\,\frac{n}{(n-p-1)} \tag{2}$$

Burnham and Anderson (2002) recommend to use AICc when $n/p < 40$.

(b) Bayesian Information Criterion (BIC): The Bayesian Information Criterion was proposed by Schwarz (1978):

$$BIC = -2\prod(\theta) + \ln(n)p \tag{3}$$

(c) Anderson-Darling Criterion (ADC): The AIC and BIC are standard selection techniques, commonly used in many different fields; however, none of these methods was explicitly designed to deal with the small sample sizes and highly asymmetric distributions that are commonly encountered in hydrological applications. However, the Anderson-Darling test has

demonstrated good fit when applied to hydrological samples (e.g. Onoz and Bayazit 1995, Laio 2004, Viglione *et al*. 2007). The ADC applied in this study to verify positive results when the statistic is used for selection purposes. The Anderson-Darling criterion has the form;

$$\Delta AD[gj(x,\theta), fn(x)] = -n - \frac{1}{n}\sum_{i=1}^{i=n}[(2i-1)\ln[Gj(x_i,\theta)] + (2n+1-2i)\ln[1-Gj(x_i,\theta)] \quad (4)$$

Estimating parameters using L-moments computation

Estimating parameters of the statistical distribution using L-moments approach (Hosking 1990) was applied. L-moments are a sequence of statistics that delineate the shape of a probability distribution and also ease application and provide a greater degree of accuracy. L-moments have been widely used and are unbiased. Kochanek *et al.* (2010) indicated that the L-moments method provides better accuracy of the parameters than the method of moments (MoM) technique for smaller sample sizes. L-statistics for linear combinations of order statistics was used to derive estimation of the parameters of probability distributions.
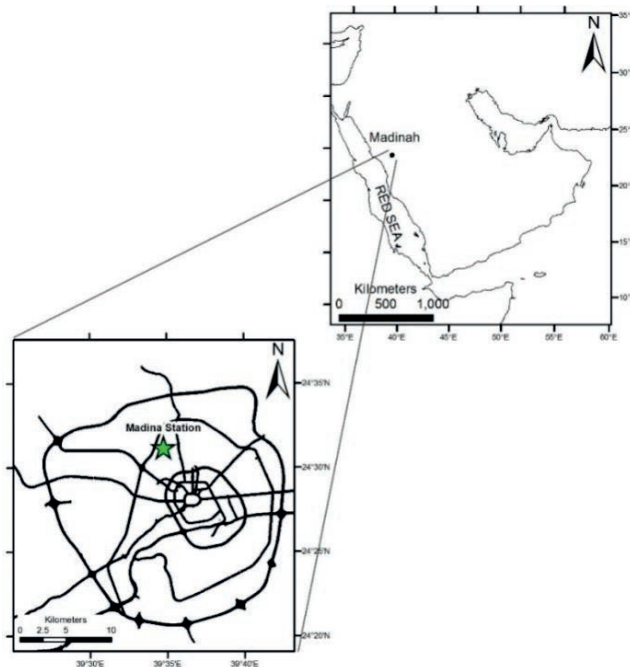

Fig. 1 The study area.

Fitting probability distribution

The model selection tests were applied based on selected the statistical analysis approach from Hosking and Wallis (2005) using the generalized extreme-value (GEV), Normal (NOR), Extreme Value Type 1 (EV1), lognormal (LN3), Pearson type III (PE3) and Log Pearson type III (LP3). The six distributions are widely used in hydrologic frequency analysis. In this study nsRFA (Viglione 2013) and Lmomco (Asquith 2013) packages under R software are used.

RESULTS AND DISCUSSION

Several categories of statistical distributions and fitting tests provided the following observations. Comparing to AMS, PDS produced less range, higher average, small variance and standard deviation. Also the PDS has higher minimum value median. The goodness-of-fit criteria tests (AIC, AICc, BIC and ADC) were calculated for different statistical distributions (NOR, LN3, EV1, GEV, PE3 and LP3), as shown in Table 1. The best distribution was chosen based on the minimum values of GOF criteria. It is found that the GEV is the best fit distribution, and the second best fit distribution is Log Pearson type III.

Table 1 Summary of goodness-of-fit tests.

	AIC	AICc	BIC	ADC
NOR	196.5	197.2	198.7	4.2525
LN3	179.0	179.7	181.2	2.1986
EV1	180.3	180.9	182.5	2.6082
GEV	*162.9*	*164.3*	*166.2*	*0.1385*
PE3	167.8	169.2	171.1	0.8867
LP3	164.2	165.5	167.5	0.3728

Table 2 L-moment parameters estimation.

Parameter	Value	Parameter	Value
L1 (Mean)	34.29	L5	1.17
L2 (L-Scale)	8.92	L-CV (t)	0.26
L3	5.08	L-Skew (t3)	0. 57
L4	3.12	L-Kurt (t4)	0.35
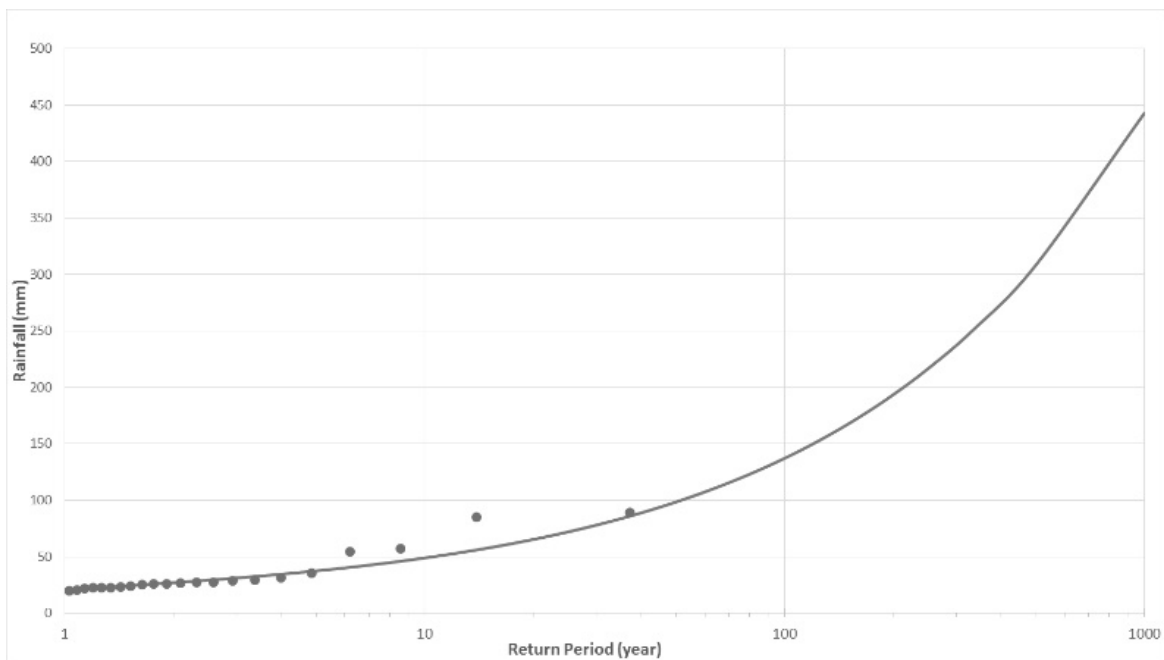
Fig. 2 Rainfall frequency curve.

The values of L-moment are then estimated and shown in Table 2, L-CV value indicates that the PDS of Madinah daily rainfall has minor variability, and L-Skew shows very large skewness, which is typical in the arid region of Saudi Arabia. Such a characteristic implied a convective-type storm occurring during a short period.

Figure 2 provides a comparison between the observed data and the GEV distribution, indicating a reasonable fitting. The rainfall depths at different average recurrence intervals (ARI) are then computed using the GEV distribution, as shown in Table 3.

The occurrence of an extreme event in Madinah in the past was in the range of 50–70 mm depth with a 25–50 year return period. However, the recent frequency of extreme events in the city with high intensity rainfall, in combination with the increased size of impervious areas resulting from accelerated urbanization, has generated a flood hydrograph with high peak rises occurring during a short period. In addition, the encroachment on main streams through the city has resulted in inundation of large flood plains, with high impact vulnerability from the current application for 25 to 50 year return periods. Thus future design criteria to reduce flood risk needs to apply 100, even 200 year return periods.

Table 3 Rainfall depths at different return periods using GEV method.

Probability of non-exceedence	Return period (years)	Rainfall depth (mm)
0.9	10	49
0.96	25	72
0.98	50	98.6
0.99	100	137.1
0.995	200	193.3

CONCLUSIONS

The paper evaluated the application of rainfall partial duration series at Madinah city, a vast urbanized area in Saudi Arabia. The analyses examined four model selection criteria: Akaike Information Criterion (AIC), Corrected Akaike Information Criterion (AICc), Bayesian Information Criterion (BIC) and Anderson-Darling Criterion (ADC). The analysis indicates that the generalized extreme-value distribution appears to provide the best fit to the Madinah rainfall data. The findings from this study can contribute to the application of flood frequency analysis in the other areas of the Saudi Arabia region with similar geomorphologic and hydrological characteristics. Further evaluation is warranted in the future if longer records and real rainfall coverage is available.

REFERENCES

Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. In: *Second International Symposium on Information Theory* (pp. 267–281). AkademinaiKiado.

Asquith, W. (2013) lmomco–L-moments, trimmed L-moments, L-comoments, censored L-moments, and many distributions. R package version, 1.7.8.

Begueria, S. (2005) Uncertainties in partial duration series modelling of extreme related to the choice of the threshold value, *Journal of Hydrology* 303, 215–230.

Haddad, K. and Rahman, A. (2011). Selection of the best fit flood frequency distribution and parameter estimation procedure: a case study for Tasmania in Australia. *Stochastic Environmental Research and Risk Assessment 25*(3), 415–428.

Hershfield, D. M.(1973) On the probability of extreme rainfall events. *Bull. American Meteorological Society* 54, 1013–1018.

Hosking, J. R. M. (1990) L-moments: analysis and estimation of distributions using linear combinations of order statistics. *J. R. Stat. Soc. Ser. B* 52(1):105–124.

Hosking, J. R. M. and Wallis, J. R. (1987) Parameter and quintile estimation for the generalized Pareto distribution, *Technometrics* 29, 339–349.

Hosking, J. R. M. and Wallis, J. R. (1997) *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press, NY, USA. Johnson.

Kochanek, K., Strupczewski, W. G. and Markiewicz, W. (2010) On feasibility of L-moments method for distributions with cumulative distribution function. In: *International Workshop Advances in Statistical Hydrology*, Taormina, Italy.

Laio, F. (2004) Cramer–von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters. *Water Resources Research* 40(9).

Laio, F., Di Baldassarre, G., and Montanari, A. (2009) Model selection techniques for the frequency analysis of hydrological extremes. *Water Resources Research* 45(7).

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

Önöz, B. and Bayazit, M. (1995) Best-fit distributions of largest available flood samples. *Journal of Hydrology* 167(1), 195–208.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* 6(2), 461–464.

Stedinger, J. R., Vogel, R. M. and Foufoula-Georgiou, E. (1993) Frequency analysis of extreme events, In: *Handbook of Hydrology* (ed. by D. R. Maidment), McGraw Hill, New York, NY, 18.1–18.66.

Subyani, A. M., Al-Modayan, A. A. and Al-Ahmadi, F. S. (2010) Topographic, seasonal and aridity influences on rainfall variability in western Saudi Arabia. *Journal of Environmental Hydrology* 18(2).

Sugiura, N. (1978) Further analysts of the data by akaike's information criterion and the finite corrections: Further analysts of the data by akaike's. *Communications in Statistics-Theory and Methods* 7(1), 13–26.

Viglione, A., Laio, F. and Claps, P. (2007). A comparison of homogeneity tests for regional frequency analysis. *Water Resources Research* 43(3), W03428.

Viglione, A. (2011) nsRFA: non-supervised regional frequency analysis. R package version 0.7–10.