Application of Linear Moments and Uncertainty Analysis to Extreme Rainfall Events in Sabah

Shu Lim Chuah^{1,*}, Janice Lynn Ayog¹, and Nurmin Bolong¹

¹Universiti Malaysia Sabah, Faculty of Engineering, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia

Abstract. Linear moments (LM) has been applied in extreme rainfall study for several countries, including China, United States of America, and Peninsular Malaysia. In this study, the LM procedures were applied to extreme rainfall data corresponding to locations provided in Malaysia Urban Stormwater Manual (MSMA) to derive new design rainfalls. Different record lengths were considered to assess the changes in design rainfall, and Monte Carlo simulations were carried out to compute confidence interval of the derived design rainfalls. Based on the Goodness-of-Fit (GoF) test results, the Generalized Extreme Value (GEV) probability distribution was chosen to derive the design rainfalls. The updated design rainfalls at all four locations showed significant reduction at design rainfalls of 50-year ARI and above. The difference of the design rainfalls from shorter record lengths with respect to the full record length and the confidence intervals do not necessarily reduce with a longer record. In hypothetical cases where 100-yr ARI rainfall was added, the increase in design rainfalls did not exceed the upper bound of the confidence intervals. The derived confidence intervals hence allow for better risk assessment, and should be considered in the design of critical structures, i.e. dams.

1 Introduction

Design rainfalls are often applied in the design of hydraulic structures, such as monsoon drains and detention ponds. The design rainfalls for Sabah in both MSMA [1] and HP No. 26 [2] are currently based on rainfall data prior to 1980. The main aim of this study is to update the design rainfalls in Sabah using LM procedures. Secondary aims in this study are assessing the differences in the design rainfall and confidence intervals derived from different record lengths, and whether the derived confidence interval was sufficient to accommodate for the presence of outliers.

2 Literature Review

The first stage in extreme rainfall studies is the compilation of extreme rainfall events, using either the Annual Maxima (AM) or Peak-Over-Threshold (POT) approach. The POT approach ensures that other extreme rainfall events apart from the annual maxima are

* Corresponding author: chuahshulim@gmail.com

included [3, 4], but suffers from the subjective selection of a suitable threshold [5, 6] and potential correlation between extreme events [7]. In contrast, the AM approach is simple, hence remaining popular in extreme rainfall studies [7].

Common probability distributions considered in extreme rainfall studies include Gumbel, Generalized Extreme Value (GEV), Generalized Pareto (GPA), and Pearson III (P3). For the AM series, the GEV is the limiting distribution [8], with Gumbel being a special case of GEV. Similarly, the PoT series converge to GPA [9]. P3 has been considered as the most suitable distribution in several extreme rainfall studies [10, 11]. The distributions are often fitted to the rainfall data by estimating the parameters via statistical procedures, such as the Maximum Likelihood Estimation and Linear Moments (LM).

The LM procedures were first published by Hosking [12], and are generally viewed as robust to outliers as there is no squaring and cubing of the variables involved [13, 14]. The LM procedures produce only small bias in small samples [15]. The LM procedures have since been applied in several countries, including USA, China, and Malaysia [3, 16, 17, 18, 19]. The design rainfalls in Peninsular Malaysia were updated using the LM procedures [20].

GoF tests are used to assess the validity of the fitted distributions in representing the extreme rainfall events. The k-s test and AD tests perform better than the Chi-Squared [21], with the AD deemed as more powerful as it places more emphasis on tail end of the distribution [22]. The critical values for the k-s and AD tests change based on the theoretical probability distribution considered and number of parameters being estimated, with various works [23-26] published on this matter. The GoF has been by used to determine the best-fitting probability distribution in studies where multiple distributions are considered [27, 28]. GoF test results should however be treated with caution, as the results depends not only on the PD choice, but also the fitting procedures [4].

In recent years, several extreme event deemed highly improbable based on standard design rainfall procedures had occurred, such as the 2003 storm at the Panama Canal [8], and the 1999 storm at Venezuela [29]. This indicates on the importance of allowing for uncertainty inherent in the derivation of design rainfalls to be incorporated via the construction of confidence intervals. The construction of confidence intervals would allow for the design risks to be properly assessed [12].

3 Method and Material

Rainfall data from four (4) stations highlighted in Fig. 1 were considered in this study due to close proximity to locations of design rainfall in MSMA [1], with record lengths listed in Table 1 considered. The ‘Base’ scenario represents the full data record available.

Fig. 1. Rainfall Stations Considered in this Study

Table 1. Rainfall Record and Sample Size Considered for each Scenario

Scenarios	Inanam, Kuamut, Bongawan		Kalabalakan	
	Record Period	Sample Size	Record Period	Sample Size
1	1985 – 1995	11	1987 – 1995	9
2	1985 – 2000	16	1987 – 2000	14
3	1985 – 2005	21	1987 – 2005	19
4	1985 – 2010	26	1987 – 2010	24
Base	1985 – 2015	31	1987 – 2015	29

The maximum annual 1-hour rainfall was extracted from the existing rainfall records, with L-moments and ratios computed using equations by Hosking [30]. L-moment estimators were applied for probability distribution fitting, considering the Gumbel, GEV, P3, GPA, Normal, and Exponential (Exp) probability distributions. The k-s and AD tests were applied to assess the GoF of the fitted probability distributions for each scenario. The k-stest statistic, D^* , and AD test statistic, A^* , were computed using equations published by Stephens and others [23-26], as listed below:

$$D^* = n^{1/2} \max \{ \{ (i/n - F(x_i)) \}, \{ F(x_i) - (i - 1)/n \} \}, i = 1, 2, 3 \dots n, \tag{1}$$

$$A^2 = -[\sum_{i=1}^n (2i - 1) \{ \ln F(x_i) + \ln (1 - F(x_{n+1-i})) \}] / n - n, i = 1, 2, 3 \dots n \tag{2}$$

Normal Distribution: $A^* = (1 + 4/n + 25/n^2)$ (3)

Exponential Distribution: $A^* = (1 + 0.6/n)$ (4)

Gumbel, GEV, P3 Distributions: $A^* = (1 + 0.2/n^{1/2})$ (5)

where n is the sample size, $F(x)$ is the cumulative distribution function of the probability distribution, and x_i is the ordered data.

For the purposes of GoF, the P3 distribution was treated an extreme value distribution. Critical values published by Stephens and others [23-26] were referred to in this study, and listed in Tables 2 and 3. The distribution with the lowest test statistic value was deemed as the best fit. The design rainfall at 10, 50, and 100-years ARI were then extracted from the best-fit probability distribution, with the 95% confidence intervals constructed using Monte Carlo procedures outlined by Silva et al. [31]. The relative difference of the design rainfalls in various scenarios relative to the ‘Base’ scenario and the root mean square error (RMSE) of the Monte Carlo simulations were subsequently calculated and assessed.

Table 2. Critical Values for D^* and A^*

Probability Distribution	Critical Value for D^*	Critical Value for A^*	Critical Value for A^*
Normal	0.895	0.787	0.787
Exp	1.094	1.341	1.341
Gumbel, GEV, P3, GPA (D^* only)	0.819 @ $n = 10$; 0.843 @ $n = 20$; 0.860 @ $n = 50$	0.757	0.757

Note: n = sample size

Table 3. Critical Values for A^2 (GPA Only)

Shape Parameter, k	-0.9	-0.5	-0.2	-0.1	0	0.1	0.2	0.3	0.4	0.5
Critical Value	0.771	0.830	0.903	0.935	0.974	1.020	1.074	1.140	1.221	1.321

To further assess the confidence intervals, a 100-year ARI event derived from the fitted probability distribution was added to the original data series, and alternative ‘outlier’ design rainfalls derived at 10, 50, and 100-year ARI derived via re-fitting the same probability distribution based on the modified data series. The ‘outlier’ design rainfalls were then compared against the upper bound of the confidence intervals. All the procedures above were carried out using Microsoft Excel, a software capable of carrying out statistical analysis.

4 Results and Discussion

4.1 Goodness-of-Fit

A summary of the best fit for each scenario and station based on GoF tests is listed in Table 4, ranked in an ascending order. The GEV distribution provides the overall best performance in both k-s and AD tests, and was subsequently used in the derivation of the design rainfalls and confidence intervals. The Exp distribution was the worst performing, followed by the GPA distribution, with over half deemed as bad fits in the AD test. Although the Normal distribution shows good performance in the k-s test, its overall ranking falls behind both Gumbel and P3 in the AD test, indicating a lower fit for the Normal distribution at the tail ends. The P3 distribution slightly outperforms the Gumbel distribution, but has more bad fits in the AD test.

4.2 Design Rainfalls and Upper Bounds

The design rainfalls and the upper bounds based on the existing rainfall records are listed in columns ‘Ori’, and ‘UB in Table 5. The ‘outlier’ design rainfalls based on the addition of the 100-yr ARI event into the data series is listed in column ‘Out’ in the same table. The derived design rainfalls are similar to the MSMA design rainfalls only at 10-yr ARI, becoming lower at higher ARI’s of 50-year and above. The ‘outlier’ design rainfalls are higher due to the addition of the 100-year ARI event into the data series. The upper bounds were sufficient to cover for the ‘outlier’ design rainfalls at all rainfall stations.

4.3 Changes in Design Rainfall and Uncertainty

Changes in the design rainfall and the inherent uncertainty of the design rainfalls were quantified using the relative difference and RMSE, as listed in Table 6. The relative differences compared to the base scenario are generally not significant, with only several cases exceeding 10% at Inanam (scenario 3), and Kalabalakan (scenarios 1 & 2). There is no clear trend of reduction in both the relative difference and RMSE in relation to an increase in record length, with different trends occurring at each station. It may be surmised that longer record lengths do not necessarily lead to a reduction in design rainfall changes and its inherent uncertainty.

Table 4. Gof Ranking based on k-s and AD tests

Scenario s	Ranking based on Probability Distribution											
	k-s						AD					
	G	GEV	P3	GPA	Nor	Exp	G	GEV	P3	GPA	Nor	Exp
Inanam												
1	2	1	3	4	5	6	4	3	2	1	5	6
2	1	2	4	6	3	5	2	1	3	Fail	4	6
3	2	1	3	5	4	6	1	2	4	Fail	3	6
4	3	2	4	6	5	1	2	1	3	5	Fail	4
Base	2	1	3	5	6	4	2	1	3	4	Fail	6
Kuamut												
1	2	3	5	4	1	6	1	2	3	Fail	4	6
2	1	2	4	5	3	6	2	1	3	4	5	Fail
3	1	2	3	4	5	6	2	1	3	Fail	Fail	6
4	1	2	3	5	4	6	1	2	Fail	Fail	Fail	Fail
Base	2	3	4	5	1	Fail	1	2	3	Fail	Fail	Fail
Kalabalakan												
1	4	3	2	5	1	6	5	2	1	4	3	Fail
2	5	1	2	3	4	6	4	2	1	Fail	3	Fail
3	5	1	2	3	4	6	Fail	2	1	Fail	3	Fail
4	5	2	1	4	3	6	3	1	Fail	Fail	2	Fail
Base	Fail	2	1	4	3	6	3	1	Fail	Fail	2	Fail
Bongawan												
Base	5	3	1	4	2	6	Fail	2	Fail	Fail	1	Fail
2	5	3	4	1	2	6	5	2	3	1	4	6
3	5	4	3	2	1	6	5	2	3	1	4	Fail
4	5	4	3	1	2	6	4	3	2	1	5	6
Base	5	3	1	4	2	6	Fail	2	Fail	Fail	1	Fail

Table 5. Design Rainfalls

Scenarios	Design Rainfall								
	10-yr ARI			50-yr ARI			100-yr ARI		
	Ori	Out	UB	Ori	Out	UB	Ori	Out	UB
Inanam (Compared against Kota Kinabalu in MSMA)									
MSMA	91.9			124.1					
1	86.8	97.6	107.9	110.0	140.7	167.3	120.7	164.7	212.0
2	91.9	101.2	113.5	120.5	147.5	179.9	134.2	173.1	223.6
3	97.4	105.1	116.5	127.5	149.5	181.7	141.7	173.0	225.1
4	94.3	99.7	111.0	119.6	134.6	161.5	131.2	152.2	194.1
Base	93.5	97.7	104.7	115.5	126.2	147.0	124.8	139.4	172.1
Kuamut (Compared against Kuamut in MSMA)									
MSMA	88			106			116		
1	87.8	91.0	93.3	92.4	97.2	103.6	93.5	98.8	109.5
2	85.9	89.8	94.3	95.7	104.2	113.0	99.1	109.9	122.5
3	89.1	92.5	96.7	100.0	107.4	115.8	103.9	113.2	125.8
4	89.8	93.3	99.4	105.1	114.0	128.8	111.5	123.4	146.0
Base	89.4	91.6	95.3	98.6	102.5	109.2	101.3	105.9	115.0
Kalabalakan (Compared against Tawau in MSMA)									
MSMA	78								
1	78.9	89.6	95.8	97.2	126.0	146.8	105.0	145.0	181.9
2	73.9	80.4	88.0	89.3	105.1	122.7	95.4	116.5	142.6
3	81.3	87.9	98.1	103.7	121.3	149.0	113.5	138.0	178.0
4	85.1	91.2	101.8	111.3	128.0	153.1	123.2	146.6	187.4
Base	82.4	87.4	96.7	108.3	122.1	147.7	120.3	139.7	178.7
Bongawan (Compared against Sandakan in MSMA)									
MSMA	103			130					
1	94.7	102.9	110.1	108.6	125.4	138.4	113.3	134.1	158.5
2	93.5	96.7	100.1	100.3	105.3	112.3	102.0	107.6	118.2
3	91.5	94.0	97.2	98.3	102.3	108.5	100.1	104.7	114.0
4	94.9	97.4	101.2	103.9	108.4	115.3	106.5	111.8	121.9
Base	95.8	98.2	103.4	105.8	110.0	118.8	108.8	113.8	126.0

Table 6.Relative Difference and RMSE in Design Rainfalls

Scenarios	Relative Difference (RD) and RMSE for Various ARI					
	10-yr		50-yr		100-yr	
	RD	RMSE	RD	RMSE	RD	RMSE
Inanam						
1	7.1	9.72	4.7	23.94	3.3	35.75
2	1.7	9.73	4.4	24.78	7.5	36.81
3	4.2	8.84	11	23.03	14	34.47
4	0.9	7.23	3.6	17.85	5.1	25.51
Base		5.89		13.67		19.13
Kuamut						
1	1.8	3.39	6.3	5.27	7.7	6.64
2	4	4.42	3	8.05	2.2	10.41
3	0.3	4.11	1.5	7.65	2.6	10.00
4	0.4	4.71	6.7	10.21	10	14.04
Base		3.10		5.14		6.36
Kalabalakan						
1	4.3	8.89	10	20.63	13	30.06
2	10	6.56	18	14.30	21	19.90
3	1.3	7.58	4.3	17.80	5.6	25.35
4	3.2	7.28	2.8	11.46	2.4	13.27
Base		6.99		17.05		24.54
Bongawan						
1	1.2	7.68	2.7	13.89	4.1	18.63
2	2.5	3.65	5.2	5.76	6.3	7.14
3	4.5	3.10	7.1	4.86	8	5.99
4	1	3.38	1.8	5.51	2.1	6.87
Base		3.48		5.79		7.19

5 Conclusion

The updated design rainfalls allow for engineering structures to be designed accordingly with more recent data, extending the range to beyond 50-year ARI, which was previously not available in the existing design rainfalls. Derivation of design rainfalls using the L-moments procedure may be applied to even record lengths of 10 years, with generally no significant difference to the design rainfalls derived at longer record lengths. Although the inherent uncertainty in design rainfall does not necessarily reduce with an increasing record length, the full record length should still be used as it is the best representation of the rainfall characteristics.

The design rainfalls become much higher when outliers are added to the existing data series, highlighting the dangers of carrying out engineering design without considering the inherent uncertainty of the design rainfalls. It is recommended for confidence level bands to be constructed in the development of design rainfall curves. The upper bound of these design rainfalls may then be applied for critical engineering structures, such as dams.

Deep gratitude is given to the Malaysian Government, with this study being financially supported under the Malaysian Ministry of Education Research Acculturation Collaborative Effort (RACE) Grant No. RACE0020-TK-2015.

References

- [1] Manual Saliran Mesra Alam, Chapter 13: Design Rainfall, 14, (2001)
- [2] K.S. Low, M.F. Haji Mahmood and S. Salleh, *Hydrological Procedure No. 26*, Estimation of Design Rainstorm in Sabah and Sarawak, (1983)
- [3] Z. Li, C. Li, Z. Xu and X. Zhou, Frequency analysis of precipitation extremes in Heihe River basin based on generalized Pareto distribution, *Stoch. Environ. Res. Risk Assess.*, **28**, 1709-1721, (2014)
- [4] D. Huard, A. Mailhot and S. Duchesne, Bayesian estimation of intensity-duration-frequency curves and of the return period associated to a given rainfall event, *Stoch. Environ. Res. Risk Assess.*, **24**, 337, (2010)
- [5] F.J. Acero, J.A. Garcia and M.C. Gallego, Peaks-over-threshold study of trends in extreme rainfall over the Iberian Peninsula, *Journal of Climate*, **24**, 1090, (2011)
- [6] A. Mailhot, S. Lachance-Cloutier, G. Talbot and A-C. Favre, Regional estimates of intense rainfall based on the peak-over-threshold (POT) approach, *Journal of Hydrology*, **476**, 188, (2013)
- [7] A. Ben-Zvi, Rainfall intensity-duration-frequency curves from large samples, *11th International Conference on Urban Drainage, Edinburgh, Scotland, UK*, (2008)
- [8] E. Shamir, K.P. Georgakakos and M.J. Murphy Jr., Frequency analysis of the 7-8 December 2010 extreme precipitation in the Panama Canal watershed, *Journal of Hydrology*, **480**, 140, (2013)
- [9] H. Van de Vyver and G.R. Demaree, Construction of intensity-duration-frequency (IDF) curves for precipitation at Lubumbashi, Congo, under the hypothesis of adequate data, *Hydrological Sciences Journal*, **55**(4), 556-557, (2010)
- [10] S. Sarkar, N.K. Goel and B.S. Mathur, Development of isopluvial map using L-moment approach for Tehri-Gahrwal Himalaya, *Stoch. Environ. Res. Risk Assess.*, **24**, 414, (2010)
- [11] M.M. Rashid, S.B. Faruque and J.B. Alam, Modelling of short duration rainfall intensity duration frequency (SDR-IDF) equation for Sylhet City in Bangladesh, *ARPN Journal Science of Technology*, **2**(2), 92-95, (2012)
- [12] J.R.M. Hosking, L-moments analysis and estimation of distributions using linear combination of order statistics, *J. of the Royal Statistical Society Series B*, **52**(1), (1990)
- [13] D. Koutsoyiannis and G. Baloutos, Analysis of a long record of annual maximum rainfall in Athens, Greece, and design rainfall inferences, *Natural Hazards*, **29**, 37, (2000)
- [14] S.S. Eslamian and H. Feizi, Maximum monthly rainfall analysis using L-moments for an arid region in Isfahan province, Iran, *J. of Applied Meteorology and Climatology*, **46**, 495, (2007)
- [15] R.V. Ramana, B. Chakaravorty, N.R. Samal, N.G. Pandey and P. Mani, Development of intensity duration frequency curves using L-moment and GIS technique, *J. of Applied Hydrology*, **XXI No.1 & 2**, 88-100, (2008)
- [16] J.P. Raiford, N.M. Aziz, A.A. Khan and D.N. Powell, Rainfall depth-duration-frequency relationships for South Carolina, North Carolina, and Georgia, *American Journal of Environmental Science*, **3**(2), 78 – 84, (2007)

- [17] P.S. Chu, X. Zhao, Y. Ruan and M. Grubbs, Extreme rainfall events in the Hawaiian Islands, *J. of Applied Meteorology and Climatology*, **48**, 502-516 (2009)
- [18] M.D. Norlida, I. Abutan, R. Abdullah, A.S. Yahaya, O. Sazali, M.D. Mohd Nor and M.S. Lariyah, Intensity-duration-frequency estimation using generalized Pareto distribution for urban area in a tropical region, *12nd International Conference on Urban Drainage, Porto Alegre/Brazil*, (2011)
- [19] S.C. Liew, S.V. Raghavan and S-Y. Liong, Development of intensity-duration-frequency curves at ungauged sites: risk management under changing climate, *Geosciences Letter*, **1**(8), (2014)
- [20] Reviewed and updated the Hydrological Procedure No.1 (Estimation of Design Rainstorm in Peninsular Malaysia), Volume I – Main Report, 32, (2010)
- [21] N.Z. Abidin, M.B. Adam and H. Midi, The goodness-of-fit test for Gumbel distribution: a comparative study, *Matematika*, **28**(1), 35, (2012)
- [22] A. Ben-Zvi and B. Azmon, Joint use of L-moment diagram and goodness-of-fit test: a case study of diverse series, *Journal of Hydrology*, **198**, 246, (1997)
- [23] M.A. Stephens, EDF statistics for goodness of fit and some somparisons, *J. of the American Statistical Association*, **69**(347), 732, (1974)
- [24] M.A. Stephens, Goodness of fit for the Extreme Value Distribution, *Biometrika*, **64**(3), 587, (1977)
- [25] M. Chandra, N.D. Singpurwalla and M.A. Stephens, Kolmogorov Statistics for tests of fit for the Extreme Value and Weibull Distributions, *J. of the American Statistical Association*, **76**(375), 730, (1981)
- [26] V. Choulakian and M.A. Stephens, Goodness-of-Fit tests for the Generalized Pareto Distribution, *Technometrics*, **43**(4), 480, (2001)
- [27] B. Mohyment, G.R. Demaree and D.N. Faka, Establishment of IDF-curves for precipitation in the tropical area of Central Africa – comparison of techniques and results, *Natural Hazards and Earth System Sciences*, **4**, 375-387, (2004)
- [28] O.O. Olofintoye, B.F. Sule and A.W. Salami, Best-fit probability distribution model for peak daily rainfall of selected cities in Nigeria, *New York Science Journal*, **2**(3), (2009)
- [29] S. Coles, L.R. Pericchi and S. Sisson, A fully probabilistic approach to extreme rainfall modelling, *J. of Hydrology*, **273**, 35-50, (2003)
- [30] J.R.M Hosking and J.R. Wallis, *Regional frequency analysis: an approach based on L-moments*, Cambridge University Press, (1997)
- [31] A.T. Silva, M.M. Portela, J. Baez and M. Naghettini, Construction of confidence intervals for extreme rainfall quantiles, *Risk Analysis VIII*, 299, (2012)