A NEW APPROACH OF RAINFALL FREQUENCY ANALYSIS USING EVENT-MAXIMUM RAINFALLS

Ke-Sheng Cheng
Dept. of Bioenvironmental Systems Engineering
National Taiwan University
Taipei, Taiwan
rslab@ntu.edu.tw

Bo-Yu Chen
Dept. of Bioenvironmental Systems Engineering
National Taiwan University
Taipei, Taiwan
r06h41003@g.ntu.edu.tw

***Abstract*—Frequency analysis is a crucial step in hydrological engineering design. By estimating the probability distribution of annual maximum rainfalls of a specific duration, the exceedance probability and average recurrence interval, i.e., return period, of certain rainfall depths can be calculated. Annual maximum series (AMS) is most widely used for rainfall frequency analysis. By choosing the maximum rainfall in each year and regarding each year as a "block", the distribution of annual maximum rainfalls could be approximated by the Generalized Extreme Value (GEV) distribution, according to the Extremal Type Theorem. However, two concerns may arise when using AMS. AMS of specific durations were selected without considering rainfall events and thus AMS may be composed of rainfalls of two or more separated storm events, particularly for AMS of longer durations. As the result, frequency analysis using AMS tends to overestimate rainfall depth. However, in reality, such AMS rainfall amounts might not result in severe flooding since flood flow of the first storm could have receded before the beginning of the second storm. The second concern is the assumption of annual maximum rainfall distribution. According to the Extremal Type Theorem, the block size (number of events in each year) should be the same for AMS to be approximated be the GEV distribution. This is obviously not the case in the real world. The annual number of storm occurrences is not a constant, so the GEV approximation of AMS rainfalls is not theoretically granted. Therefore, we propose an event-maximum-rainfalls (EMR) based approach for frequency analysis. For any given design duration, we firstly extracted event-maximum-rainfalls of individual storm events by considering the minimum inter-event time of different storm categories including Meiyu, typhoons, convective storms, and winter frontal rainfalls. Annual counts of storm events and event-maximum rainfalls were modeled by the Poisson distribution and Pearson Type III distribution, respectively. For each storm category, a storm-type-specific mixture distribution of annual maximum rainfalls was then derived. Finally, the cumulative distribution function of the over-all (considering all storm types) annual maximum rainfalls was obtained from the storm-type-specific mixture distributions. Stochastic simulation was conducted to demonstrate the proposed EMR-based approach. The EMR-based approach is superior to the traditional AMS-based approach in terms of the biasedness and mean squared error. The EMR-based approach is particularly useful in situations of short record length and outlier presence.**

Keywords—Event-maximum series, annual-maximum series, frequency analysis

I. INTRODUCTION

Annual maximum series (AMS) combined with the generalized extreme value(GEV) distribution is widely used in frequency analysis.[1][2][4] The AMS method samples maximum rainfall in each year and estimate the underlying distribution of annual maximum rainfalls (AMRs). Then design rainfalls of specific exceedance probabilities (or return periods) are determined.

The AMS method is straightforward and easy to implement, however, it selects the largest total rainfall depths of individual years without considering rainfall events. As a result, AMRs may consist of rainfalls from two or more rainfall events. Moreover, the AMS method selects only the highest rainfall in each year, resulting in waste of information. In Taiwan, there are more than 300 automated rainfall stations with 20 – 30 year hourly rainfall records. Such rainfall data have not been used for AMRs rainfall frequency analysis due to short record lengths.

Another concern about the AMS method arises when using the GEV distribution for AMRs. According to extremal type theorem, if there exists constant sequence $\{a_n>0\}$ and $\{b_n\}$ such that the distribution of n^{th} order statistic(among n samples) would converge to a non-degenerate distribution G as n approcimate infinity, then G is a member of the GEV family[3]:

$$G(z) = \exp\{-[1+\xi(\frac{z-\mu}{\sigma})]^{-1/\xi}\} \quad (1)$$

with

$$1+\xi\left(\frac{z-\mu}{\sigma}\right)>0, -\infty<\mu<\infty, \sigma>0, -\infty<\xi<\infty$$

AMS method uses GEV approximation by regarding each year as a "block" with n rainfall samples and annual maximum rainfall is the n^{th} order statistic. However, the approximation is only valid when the sample size n is large. Also, the annual count of events varies each year, making the AMRs not suitable for GEV distribution. In addition to the AMS method, partial duration series (PDS) method is another frequently used method in frequency analysis.[2][7][9][10] The method includes the well-known peak over threshold (POT) method which uses the generalized Pareto distribution(GPD) for design rainfall estimation. The method separate rainfall record into independent events by setting a threshold value and ignore values below the threshold. PDS method could preserve more samples than AMS method. The challenge for PDS method is the choice of the threshold value. To maintain the independence between rainfall events, the threshold value should not be too low while the higher the threshold is, the more data may loss. There's no absolute criterion to select the threshold, so the PDS method is not as popular as AMS method.

To preserve as many data as possible and to avoid the dilemma of deciding threshold, [6] used minimium inter-event time as criterion to separate independent rainfall event and use the generalized Poisson and GPD for frequency analysis. The result showed that the estimation from the event-based method is lower than the traditional approach and is more appropriate for hydraulics design. In this paper, we proposed a new event-maximum-rainfalls (EMR) based approach for frequency analysis. The method separate rainfall record into different types of independent events and estimate the underlying AMR distribution.

II. METHOD

Let n be the number of typhoon events in each year and $\{X_1 \ldots X_n\}$ be the rainfall depth of typhoon that are independent and identically distributed with distribution F_X. Then, the AMR of typhoon and its distribution can be obtained by order statistic:

$$Y = \max\{X_1, X_2, \ldots\ldots, X_{n-1}\ X_n\}$$

$$F_Y(y) = F_X(y)^n$$

Since the number of typhoon events in each year is not constant, it may be assumed to follow poisson distribution with parameter λ [5], then the AMR distribution of typhoon can be written as:

$$F_Y(y) == \sum_{n=0}^{\infty} P(n;\lambda) * F_X(y)^n$$

$$= \sum_{n=0}^{\infty} \frac{e^{-\lambda}\lambda^n}{n!} * F_X(y)^n = e^{-\lambda[1-F_X(y)]} \qquad (2)$$

In Taiwan, rainfall events could be classified into four types - typhoon, convective storm, mei-yu and frontal rain, so the true AMR is the maximum value among these different types of AMR. Let Z be the True AMR, Y_1, Y_2, Y_3, Y_4 are the AMR correspond to typhoon, convective storm, mei-yu and frontal rain, then since all rainfall events are independent, the distribution of Z is:

$$F_Z(z) = F_{Y_1}(y_1) * F_{Y_2}(y_2) * F_{Y_3}(y_3) * F_{Y_4}(y_4)$$

$$= e^{-\lambda_1[1-F_{x_1}(z)]-\lambda_2[1-F_{x_2}(z)]-\lambda_3[1-F_{x_3}(z)]-\lambda_4[1-F_{x_4}(z)]}$$

Which is the true distribution of AMR used in EMR-based frequency analysis.

III. SIMULATION STUDY

To compare the EMR based approach with traditional AMS based approach, we use Monte Carlo simulation to simulate 44 years rainfall records of typhoon and mei-yu from poisson and gamma distribution. Then the L-moment method is used for parameter estimation in both methods to estimate return level with corresponding return period. Finally, bias and mean square error(MSE) are calculated to evaluate these methods.

After deciding record length N=44 years and given parameters of poisson and gamma distribution(Table I), the steps of each simulation run are as follow:

In each year i,

(1) Generate event number n_{i1} and n_{i2} of typhoon and mei-yu events from corresponding poisson distribution.
(2) Generate n_{i1} typhoon rainfall $\{X_{i,1}, \ldots, X_{i,n_{i1}}\}$ of typhoon and n_{i2} mei-yu rainfall depth $\{Y_{i,1}, \ldots, Y_{i,n_{i2}}\}$ from corresponding gamma distribution.
(3) Let $a_i = \max\{X_{i,1}, \ldots, X_{i,n_{i1}}, Y_{i,1}, \ldots, Y_{i,n_{i2}}\}$ be the i^{th} element of annual maximum rainfall series.
(4) Repeat (1)-(3) N times to obtain EMR series and AMS series.
(5) Conduct event-based frequency analysis and AMS frequency analysis. Gamma distribution is used to estimate rainfall distribution in EMR method while GEV distribution and Pearson type III distribution(PTIII) are used for AMS estimation.

Return levels correspond to return period 5, 10, 25, 50, 100, 200 years are calculated. The total simulation run is set to 1000 and the return levels from EMR and AMS are compared to theorecical values in Table II to calculate bias and MSE. Results are shown in Fig 1. .

Clearly, the AMS-GEV method has the largest bias and MSE while EMR approach has the smallest. As return period increases, the bias and MSE of three approach increase since the cumulative distribution function become flatter in the tail of the distribution and small bias in parameter estimation would cause larger bias in quantile estimation comparing to smaller return period. The poor performance of AMS-GEV approach may due to small and non-constant sample size in each year causing the distribution unable to converge, and thus unstable parameter estimation with larger bias in each simulation run. Changing the record length from 44 years to 20 years would result in similar bias and larger MSE with similar trend. Indicating that for short record length, the EMR approach is still more suitable for estimation.

IV. Real Data Analysis and Discussion

We choose three rain gauge stations in northern, central and southern part of Taiwan to compare frequency analysis result of two methods. L-moment method is used for parameter estimation and the L-momnet ration diagram with acceptance region is used for goodness of fit test.[8] Both EMR and AMS methods used PTIII distribution for rainfall distribution. Rainfall records are first separated into different types of independent storm events using minumum inter-event time and season (Table III). Since AMS is mostly consists of typhoon, mei-yu and convective storm, we would use these three types of storm in EMR frequency analysis.

The design durations are 1, 2, 3, 6, 12, 24, 48, 72 hours and the return periods are 5, 10, 25, 50, 100, 200 years. The estimated return levels are shown in table IV .

Table I. parameter settings for simulation study

	Gamma-scale	Gamma-shape	Poisson-λ
Typhoon	141.19,	1.34	2.09
Mei-yu	53.87	0.68	11.45

Table II. Theoretical Values of return level with corresponding return period

Return period	5	10	25	50	100	200
Theoretical value	352	459	596	698	799	899

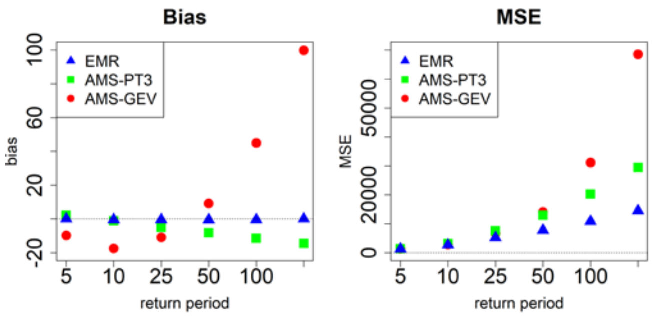

Fig.1 Bias and MSE of Return level of EMR and AMS frequency analysis

The estimation of AMS method is higher than the one from EMR method in all design durations, this difference is due to the way AMS method selects samples. AMS method does not consider rainfall events, thus when using moving window to search annual maximum rainfall, the selected sample may consists of two or more rainfall events. Most hydrologic and hydraulic design assume the rainfall comes from single event, so the estimation from AMS method would cause overestimation. The effect is more apparent in the difference between 48-hour and 72-hr return, in EMR method, the return level has little difference while in AMS method the difference is obvious. In Taiwan, there are only a few rainfall events lasting longer than 3 days, so when using 72-hour design duration, most selected rainfall events actually having durations close to 48-hour and the estimated return level is similar to the 48-hour result. On the other hand, AMS method could select more than one event, so the return level in higher duration would tempt to be overestimated

By removing the largest value in EMR/AMR series and calculate the return level again (Table V), it is possible to compare how sensitve both methods are to their sample (or outlier). Result shows that the AMS result changes much more than the EMR result. Due to the small sample size, AMS method is more unstable than EMR method. When confronting an extreme rainfall event in the rainfall record, AMS is easier to be influenced and to have larger estimation.

Table III. Event Number of each types of storm and record length of three stations

station	Convective storm	Meiyu	Typhoon	Record length
01B030(WuDuh ,North)	260	230	147	32
01F680(TBK,ce ntral)	355	300	70	42
467480(Chiayi, South)	754	504	92	44

Table IV. Return Level estimated from EMR method and AMS-PTIII method

	Annual Maximum Rainfall					
Dur\yr	5	10	25	50	100	200
1	83.06	110.38	150.09	181.82	214.52	247.94
2	118.26	168.88	248.67	315.09	384.97	457.4
3	148.47	209.78	301.89	376.75	454.59	534.63
6	217.61	302.27	422.15	516.56	613.12	711.29
12	300.45	409.39	557.82	672.17	787.72	904.18
24	389.03	514.85	680.73	806.01	931.17	1056.24
48	460.45	599.99	781.77	918.03	1053.5	1188.53
72	495.21	639.63	827.06	967.21	1106.4	1244.89
	Event Maximum Rainfall					
1	81	93	108	119	130	141
2	121	139	163	180	197	214
3	146	170	199	221	243	264
6	204	244	295	332	370	407
12	288	355	441	504	566	628
24	378	476	600	692	783	872
48	430	544	690	797	903	1008
72	431	545	691	798	904	1009

V. Conclusion

We propose a new approach for rainfall frequency analysis using event-maximum-rainfalls. Simulation study shows that the EMR method has better performance over traditional AMS approach in terms of biasedness and mean square error. Having more samples than AMS method, estimation from EMR method is more stable for both short and long record length. In real data analysis, effect of sample size and outlier to both methods were discussed. The result shows that EMR method is preferred to avoid overestimation and to reduce bias.

Table V. Return Level estimated from EMR method and AMS-PTIII method after removing largest value

	Annual Maximum Rainfall					
Dur\yr	5	10	25	50	100	200
1	76.02	89.75	107.54	120.83	134.03	147.15
2	108.78	132.03	163.77	188.24	213	237.95
3	135.31	167.3	210.33	243.24	276.35	309.61
6	198.03	250.83	320.32	372.75	425.09	477.37
12	276.25	353.83	454.43	529.62	604.26	678.5
24	363.77	463.9	591.9	686.68	780.26	872.93
48	434.81	552.01	701.46	811.94	920.91	1028.73
72	470.4	595.59	755.55	873.96	990.85	1106.59
	Event Maximum Rainfall					
1	77	89	104	116	127	138
2	114	132	155	172	189	205
3	138	161	191	213	234	256
6	195	236	288	327	366	404
12	280	350	438	502	565	627
24	371	472	599	691	782	872
48	422	540	687	796	902	1008
72	423	541	689	797	903	1009

References

[1] Al Mamoon, A, Rahman, A, 2017, "Selection of the best fit probability distribution in rainfall frequency analysis for Qatar"", Natural hazards , 2017, Vol.86(1), p.281-296

[2] Chin, RJ, Lai, SH, Chang, KB, Jaafar, WZW & Othman, F, 2018, "Statistical Analysis Towards Improvement of Design Rainfall: Case Study of Peninsular Malaysia", Iranian journal of science and technology. Transactions of civil engineering. , 2018, Vol.42(2), p.121-131

[3] Coles, S. An introduction to statistical modeling of extreme values. Springer, 2001.

[4] Deka, S, Borah, M, Kakaty, SC, 2011, "Statistical analysis of annual maximum rainfall in North-East India: an application of LH-moments", Theoretical and applied climatology , 2011, Vol.104(1-2), p.111-122

[5] Haight, F. A., 1967, Handbook of the Poisson Distribution. John Wiley & Sons, New York.

[6] Jun, C.H., Qin, X., Tung, Y.K. & Michele, C.D., 2017, "Storm event-based frequency analysis method", 2017, Hydrology Research (2017) 49 (3): 700-710.

[7] S. Beguerı´a, 2005, "Uncertainties in partial duration series modelling of extremes related to the choice of the threshold value", Journal of Hydrology 303 (2005) 215–230

[8] Wu, Y.C., Liou, J.J., Su, Y.F. & Cheng, K.S., " Establishing acceptance regions for L-moments based goodness-of-fit tests for the Pearson type III distribution", 2012, Stoch Environ Res Risk Assess (2012) 26: 873

[9] W. Zucchini & P. T. Adamson, 1989, "Bootstrap confidence intervals for design storms from exceedance series", Hydrological Sciences Journal, 34:1, 41-48.

[10] Zhou, ZZ, Liu, SG, Hu, Y, Liang, YY, Lin, HJ, Guo, YP, 2017, "Analysis of precipitation extremes in the Taihu Basin of China based on the regional L-moment method", Hydrology Research, 2017, Vol.48(2), p.468-479