

DESIGN FLOOD FORMULA DEVELOPMENT IN UNGAUGED CATCHMENTS, WEST JAVA INDONESIA: INDEX FLOOD AND L-MOMENT APPROACH

F. Mulyantari¹⁾, S.J. Sutanto^{1&2)}, T. Asyantina¹⁾

¹⁾Research Center for Water Resources, Ministry of Public Work Indonesia, Jl. Ir. H. Djuanda No.193, Phone: +62-22-2504053, 2501554, Fax: +62-22-2503357, Bandung 40135, Indonesia.

²⁾Now at Institute for Marine and Atmosphere Research Utrecht, University of Utrecht, Princetonplein 5, 3584CC, Utrecht, the Netherlands

Corresponding email: frantari@yahoo.com

Diterima: 13 Juli 2011; Disetujui: 27 September 2011

ABSTRAK

Data hidrologi yang baik dan berkualitas sangat dibutuhkan dalam analisis banjir di Indonesia, akan tetapi banyak terdapat daerah aliran sungai (DAS) yang mempunyai data pengamatan debit yang tidak cukup atau bahkan tidak terdapat data sama sekali. Studi ini mempunyai tujuan utama yaitu membuat suatu formula untuk menghitung besarnya debit banjir rencana pada DAS yang tidak mempunyai stasiun pengamatan muka air di Jawa Barat dengan menggunakan pendekatan indeks banjir dan momen-L. Beberapa metode telah digunakan dalam studi ini diantaranya: metode momen untuk analisis frekuensi, metode L-momen untuk analisis frekuensi wilayah dan uji keheterogenan dengan 50 iterasi, metode Montecarlo untuk pembangkit data sebanyak 1000 data, metode Ward Linkage untuk analisis pengelompokan. Hasil analisis menunjukkan bahwa debit banjir yang terjadi mengikuti distribusi GEV (Generalized Extreme Value). Perumusan Growth factor dibagi menjadi dua kelompok berdasarkan formula $100(L/A)*S$, sementara formula MAF (banjir tahunan rata-rata) dibagi menjadi tiga kelompok dibuat berdasarkan luas DAS dan hujan harian maksimum. Verifikasi rumus yang telah dibentuk menunjukkan hasil yang memuaskan dengan nilai BIASr 9%. Dengan perumusan yang dibentuk ini, maka besarnya banjir rencana pada DAS tidak terukur di Jawa Barat dapat dilakukan hanya dengan menggunakan data hujan dan karakteristik DAS.*

Katakunci: DAS tidak terukur, indeks banjir, L-momen, Ward Linkage, GEV dan perumusan banjir rencana

ABSTRACT

Robust hydrological data are needed to analyze the flood problem in Indonesia but many catchments in Indonesia show few flood data or even no data. Main objective of this research is to formulate the regional design flood for ungauged catchments in West Java using the index flood and L-moment approach. Some methods used in this research include: moment method for frequency analysis, L-moment method for regional frequency analysis and heterogeneity test with 50 iterations, Montecarlo method to generated 1000 years data, and Ward Linkage method for clustering. Result shows that the discharge stations follow the GEV (Generalized Extreme Value) distribution. The growth factor formula has been divided into two groups based on $100(L/A)*S$ formula while MAF (Mean Annual Flood) formula is divided into three groups based on catchment area and maximum daily precipitation. Verification shows acceptable result with BIASr 9%. Therefore, using the developed formula, design flood can be estimated for ungauged catchments in West Java by using precipitation and catchments characteristic data only.*

Keywords: Ungauged catchment, index flood, L-moment, Ward Linkage, GEV, design flood formula

INTRODUCTION

Flood is frequent disaster in Indonesia especially in the big cities; therefore the effective and efficient action is needed to overcome this problem. Hydrological data is the important key

for flood analysis but the data for analysis is not always available, unreliable and the length is too short for analysis. There are so many areas which do not have hydrological data especially in the remote area. The other reason is that the stations to record the data are broken or even lost due to

vandalism. To analyze the flood in ungauged catchment, a study needs to be performed to develop the formula. One of the methods to develop flood formula in an ungauged catchment is regional flood frequency analysis using index flood and L-moment method since a regional flood frequency analysis can be applied when no local data are available or the data are insufficient (Cunderlik and Burn, 2002; Leclerc and Ouarda, 2007). In addition, the regional flood frequency analysis can enhance the data from single site using the data from the other sites as long as they have similar frequency distribution (Saf, 2009).

Numerous regional analysis methods for hydrology have been carried out in the world both for high flow and low flow. Each method has assumptions and calibration based on the local condition. The use of recent method without any adaptation to the local condition will form the overestimate or underestimate of flood analysis (Stedinger *et al.*, 1993; Burn, 1990; Ferrarier *et al.*, 1993; Cannarozzo *et al.*, 1995; Hosking and Wallis, 1997). Therefore, if regional analysis is applied into precipitation data to calculate design flood, then the local adaptation is absolutely required.

Hosking and Walling (1997) described some regional flood approaches. The approach differences are the estimation of frequency distribution parameters such as mean and dispersion which is commonly written as variation coefficient and skewness coefficient. The combination of regional and individual methods is chosen if the data have specific condition such as: observation cluster that is considered as a homogenous region was homogeneity doubtful, individual skewness coefficient is more accurate than regional average, used for higher return period estimation and average variance coefficient is low (Hosking and Wallis, 1997). On the other hand, if the region is homogeneous and individual skewness coefficient is less accurate, index floods method should be chosen.

The index flood method uses average parameter from location and measured station, while variance coefficient and skewness coefficient are calculated from regional average. Dalrymple (1960) stated that this method can be used for flood analysis and simple. Furthermore, this method can be used for another data. Hosking and Wallis (1997) also suggest using index flood method in the regional frequency analysis with consideration from the previous research (Hosking *et al.*, 1985; Lettenmaier and Potter, 1985; Wallis and Wood, 1985; Lettenmaier *et al.*, 1987; Hosking and Wallis, 1988) which has accurate and trustworthy design flood estimation results.

The study about regional flood analysis in Indonesia is very scarce. One study in the past was done by Puslitbang Air (Research Center for Water Resources) and Institute of Hydrology, UK in 1983 using a different method (see also Wharton and Tomlinson, 1999 for application example). This paper will discuss about regional flood analysis in ungauged catchments using index flood and L-moment method to replace the previous study in Indonesia especially in West Java Province. The L-moment method is used since this method is superior to other methods and has been used worldwide (Chen *et al.*, 2006). Thus, the objective of this research is to formulate the regional design flood for ungauged catchment in West Java using index flood and L-moment approach.

METHODOLOGY

The first analysis step such as: data screening, individual frequency analysis in every station, regional design rainfall, catchment characteristic, etc have been carried out in this study, but the results will not be discussed and shown in this paper. This paper will discuss starting from regional frequency flood analysis with L-moment method, discordancy test, clustering analysis, regional distribution determination, regional homogeneity test, growth factor analysis, index flood, and the last is model validation.

Regional flood frequency analysis has been carried out using L-moment method which is started with the estimation of L-moment parameters ($L-C_v$, $L-C_s$ and $L-C_k$). L-moment method is a linier combination from Probability Weighted Moments (PWM) that can eliminate outliers and be accurate for short data recording. This L-moment describe as:

$$\beta_r = E\{X[F(X)]^r\} \quad (1)$$

$F(X)$ is cumulative distribution function X for $r = 0$, $\beta_0 = \text{mean (average)}$. Generally estimated by:

$$\beta_0 = \frac{1}{n} \sum_{j=1}^{n-r} \binom{n-j}{\dots r} X(f) \quad (2)$$

With:

$$r: 1, \dots, n-1$$

For every distribution in PWMs function, L-moment is expressed as:

$$\lambda_1 \quad , \quad \beta_0 \quad (3)$$

$$\lambda_2 \quad , \quad 2\beta_1 - \beta_0 \quad (4)$$

$$\lambda_3 \quad , \quad 6\beta_2 - 6\beta_1 + \beta_0 \quad (5)$$

$$\lambda_4 \quad , \quad 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \quad (6)$$

$$\tau = \lambda_2/\lambda_1 = L-C_V = t, \text{ sample} \quad (7)$$

$$\tau_3 = \lambda_3/\lambda_2 = L-C_S = t_3, \text{ sample} \quad (8)$$

$$\tau_4 = \lambda_4/\lambda_2 = L-C_k = t_4, \text{ sample} \quad (9)$$

L-moment diagram is a correlation between τ_3 and τ_4 for various distributions. The fitted distribution selection based on the average regional τ_3 and τ_4 , is closest to the distribution lines and shown in the diagram. This diagram shows also in the data distribution, forming a group or scattered, as an indication of deviation of location probably not included in the cluster. L-moment diagram can show if a gauging station is not a member of a cluster when parameters deviate from the data cluster. Discordancy test can show the deviation value. The deviated gauging station is considerable not to use. The regional deviation (Hosking and Wallis, 1997) is described as:

$$D_i = \frac{1}{3} N(u_i - \bar{u})^T A^{-1} (u_i - \bar{u}) \quad (10)$$

$$u_i = (t^i t_3^i t_4^i)^T \quad (11)$$

$$\bar{u} = \frac{1}{N} \sum_{i=1}^N u_i \quad (12)$$

$$A = \sum_{i=1}^N (u_i - \bar{u})(u_i - \bar{u})^T \quad (13)$$

N is gauged number from one cluster, u_i is vector or metric from τ : $L-C_V$, τ_3 : L-skewness, τ_4 : L-Kurtosis. T is transpose from metric and D_i is value of deviation. Critical value for discordancy test (D_i) is shown in Table 1 below.

Table 1 Discordancy value for D_i

N (station number)	Critical Value (D_i)	N (station number)	Critical Value (D_i)
5	1.333	10	2.491
6	1.648	11	2.632
7	1.917	12	2.757
8	2.140	13	2.869
9	2.329	14	2.971
		≥ 15	3

The regional homogeneity development has been done using Ward Linkage clustering analysis method with homogeneity test from L-moment. After a homogeneous region is formed, the regional frequency distribution determination can be performed. To test the homogeneity of region or group and the selection of group distribution, the data series in each group needs to be extended to 1000 years (Hosking and Wallis, 1997) with the Montecarlo method and it is made of 50 pieces of artificial groups for each group. Regional estimation for L-moment parameter is t^i , t_3^i and t_4^i for location i . Differences of data recorded (n) deviation are described as:

$$t^R = \frac{\sum_{i=1}^N n_i t^{(i)}}{\sum_{i=1}^N n_i} \quad (14)$$

$$V = \left\{ \frac{\sum_{i=1}^N n_i (t^{(i)} - t^R)^2}{\sum_{i=1}^N n_i} \right\}^{0.5} \quad (15)$$

Deviation (σ_V) and average (μ_V) from V can be estimated from the simulation. The heterogeneity value can be estimated by:

$$H = \frac{(V - \mu_V)}{\sigma_V} \quad (16)$$

A cluster is homogeneous if $H < 2$ and definitely heterogeneous if $H \geq 2$ (source: Hosking and Wallis 2007). In contrast, Robson and Reed (1999) as cited in Lim and Lye (2003), extended this criteria by suggesting that if $2 < H \leq 4$ a region can be considered as heterogeneous.

After the groups distribution was selected, estimates of the parameters of the selected distribution for each group can be done and proceed with the estimation of Growth Factor which is a dimensionless. The general formula for index flood describe as:

$$y = -\ln \left(\ln \frac{T}{T-1} \right) \quad (18)$$

$$X_T = 1 + \frac{\alpha}{\xi * k} (1 - e^{-ky}) \quad (19)$$

With:

X_T , regional growth factor at return period T year

α , scale

ξ , location

k , shapes

The GEV (Generalized Extreme Value) parameter in this study is derived and developed from the basic formula described below:

$$y = -k^{-1} \log \{ 1 - k(x - \xi) / \alpha \}, k \neq 0 \quad (20)$$

$$y = (x - \xi) / \alpha, k = 0 \quad (21)$$

$$F(x) = e^{-e^{-y}} \quad (22)$$

$$x(F) = \xi + \alpha \{ 1 - (-\log F)^k \} / k, k \neq 0 \quad (23)$$

$$x(F) = \xi - \alpha \log(-\log F), k = 0 \quad (24)$$

$$\lambda_1 = \xi + \alpha \{ 1 - \Gamma(1+k) \} / k \quad (25)$$

$$\lambda_2 = \alpha (1 - 2^{-k}) \Gamma(1+k) / k \quad (26)$$

$$k = 7.8590c + 2.9554c^2 \quad (27)$$

$$c = \frac{2}{3 + \tau_3} - \frac{\log 2}{\log 3} \quad (28)$$

$$\alpha = \frac{\lambda_2 k}{(1 - 2^{-k})\Gamma(1 + k)} \quad (29)$$

$$\xi = \lambda_1 - \alpha \{1 - \Gamma(1 + k)\} / k \quad (30)$$

Where $\Gamma(x)$ is the gamma function:

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt \quad (31)$$

Gamma function can be seen on the gamma table. For simplification, the gamma function can be expressed as:

For $0.01 \leq n < 1$, then

$$\Gamma(n) = 0.889n^{-1.0243} \quad (32)$$

With relative mean bias (BIASr) 0.0291, and determination coefficient $R^2 = 0.9983$.

For $1 \leq n < 2$, then

$$\Gamma(n) = 0.4382784n^2 - 1.2964092n + 1.8445612 \quad (33)$$

With BIASr 0.00217, and determination coefficient $R^2 = 0.9934$.

For $2 \leq n < 3$, then

$$\Gamma(n) = 0.55769n^2 - 1.76852n + 2.286302 \quad (34)$$

With BIASr 0.03195 and determination coefficient $R^2 = 0.9775$.

The next step is the regional analysis to obtain an index flood from every gauged station. The common index flood is the mean annual flood (MAF) (Wharton and Tomlinson, 1999; saf, 2009; Malekinezhad, 2010). Design flood at any station can be estimated from the index flood at that location multiplied by the Growth Factor from the group in which the post is located. For the ungauged catchments, regression analysis between the MAF with catchment characteristics is conducted in a measurable location, afterward the MAF values in the ungauged catchment can be obtained based on the formula developed in gauged catchments.

Model verification is performed by removal some gauged stations from the group which has a long record period and then the estimation of design flood with various return periods is using the new developed formula. Verification is done by using bias (BIAS) and relative mean bias (BIASr) method, which is a correlation between design flood from simulation and actual design flood. Bias and relative mean bias can be described as:

$$BIAS = \left| \frac{Q_{comp} - Q_{obs}}{Q_{obs}} \right| \quad (35)$$

$$BIASr = \frac{1}{n} \sum_{n=1}^n Abs \left[\frac{Q_{comp} - Q_{obs}}{Q_{obs}} \right] \quad (36)$$

With:

BIAS ,bias error

BIASr ,relative mean bias

Q_{comp} , regional design flood from simulation

Q_{obs} , actual point design flood

n , number of data

RESULTS AND DISCUSSION

1 Discordancy Test

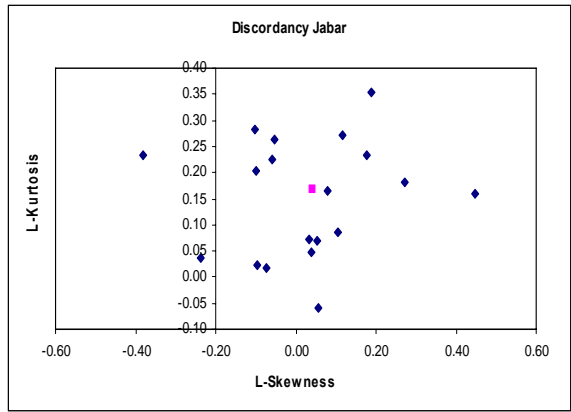
The discordancy test is used to determine the deviation in the gauging station group shown by L-moment diagram, that is gauging station parameter deviated farthest from group. If deviation occurs, the deviated gauging station needs to be separated, and simulation will be re-simulate with the remaining data. The result shows that 23 gauging stations passed the test with discordancy score below 3 except Cikapundung-Maribaya station. Regional L-moment parameters for 23 gauging stations are $L-C_v$ (t^R) 0.218, $L-Skew$ (t_3^R) 0.048 and $L-Kurt$ (t_4^R) 0.184. Discordancy test results are shown in Figure 1.

The black arrows in Figure 1a and 1b show the location of Cikapundung-Maribaya stations in the graphs which have discordancy. From the graph we can conclude that the discordancy is triggered by the highest $L-kurt$ and $L-C_v$, not by the highest of $L-skewness$. Regional frequency distribution of all data is indicated by the closeness regional parameters of L-moment which is the individual average to L-moment diagram. Visually, it seems that the regional average near to GLO (Generalized Logistic) distribution.

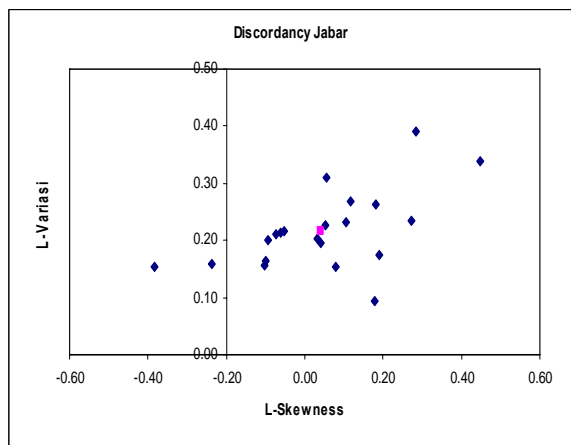
2 Clustering

Clustering analysis has an objective to classify the observations data so that observations in the same cluster are similar in some sense. The general parameters used for clustering are physically parameters such as location coordinate, catchment area, land slope, and statistically parameters such as coefficient variance, mean, and coefficient skewness. Clustering analysis can be hierarchical and non hierarchical. Hierarchical method is the process of group formation through agglomerative while non hierarchical method is done by splitting up the group (Gong and Richman, 1995).

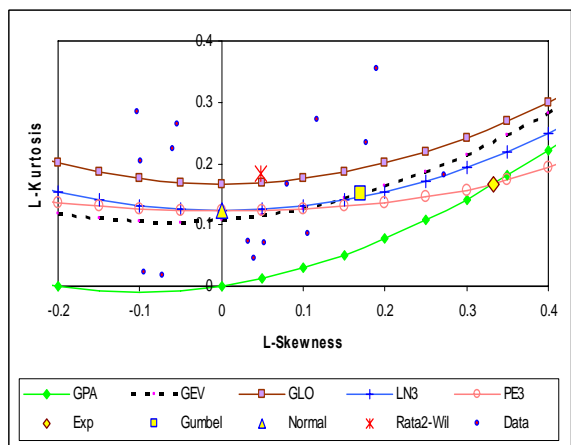
The method used in this study is hierarchy clustering using Ward's Linkage method (Malekinezhad, 2010) and this study is using physically parameters such as catchment area, river length and slope. This clustering has been carried out with the objective to have a better BIASr value compare with one group analysis (see Table 2 for the result).



(a)



(b)



(c)

Figure 1 (a) L-moment diagram of $L-skew$ and $L-kurt$, (b) L-moment diagram of $L-skew$ and $L-C_v$, (c) regional L-moment diagram

Table 2 Clustering result for West Java stations

No	Name	A (km ²)	L (km)	S	L/A	100* (L/A)*S	Cluster
1	Ciliman Munjul	327.7	45.72	0.014764	0.14	0.206	1
2	Cisadea Cikarang	240.7	14.23	0.038651	0.06	0.2285	1
3	Ciberang Sebagi	304.9	62.46	0.014209	0.21	0.2911	1
4	Citarik Pajangan	229	24.23	0.042303	0.11	0.4476	1
5	Ciasem Curuggagung	93.92	18.23	0.032227	0.19	0.6255	1
6	Cikapundung Gandok	91.23	19	0.036184	0.21	0.7536	1
7	Ciresea Cengkong	69.62	11.28	0.56516	0.16	0.9157	1
8	Cidurian Parigi	674.2	65.63	0.000381	0.1	0.0336	2
9	Ciujung Kragilan	1836	98.76	0.005569	0.05	0.03	2
10	Cimanuk Tomo	1971	113.1	0.005858	0.06	0.0336	2
11	Cijolang Cikadu	383.2	40.29	0.004964	0.11	0.0522	2
12	Cipunegara Kiarapayung	879.8	66.31	0.006975	0.08	0.0526	2
13	Citarum Nanjung	1733	80.82	0.0116	0.05	0.0541	2
14	Cibeureum Neglasari	124	12.86	0.005832	0.1	0.0605	2
15	Citarum Dayeuh Kolot	1332	69.11	0.013565	0.05	0.0704	2
16	Ciseel Binangun	325.1	27.2	0.009191	0.08	0.0769	2
17	Ciujung Rangkasbitung	591.8	55.13	0.009976	0.09	0.0929	2
18	Cilutung Damkamun	614.5	52.44	0.013587	0.09	0.1159	2
19	Cimanuk Leuwigoong	472.8	37.87	0.015844	0.08	0.1269	2
20	Cimanuk Leuwidaun	350.2	29.63	0.016875	0.09	0.1428	2
21	Cimanuk Bojongloa	286.3	25.29	0.017794	0.09	0.1572	2
22	Cilangla Leuwinekteuk	188.2	15.47	0.021816	0.08	0.1793	2
23	Cikapundung Maribaya	73.53	29.69	0.009262	0.4	0.374	3

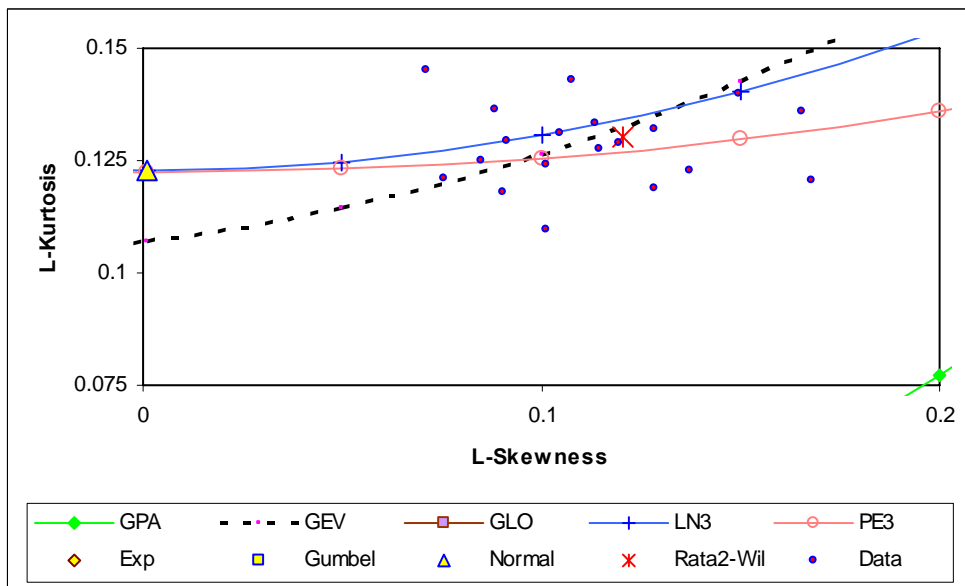


Figure 2 Visual distributions test using 1000 years data generated

Table 2 shows that discharge stations are divided into three groups. Cikapundung-Maribaya is not used for the following analysis because only one station is in group three. Group determination is based on how much is the river length divided by catchment area and multiplied by the slope of the channel and hundred ($100 \cdot L/A \cdot S$). Group one is the group which data has $L/A \cdot S \cdot 100$ value more than 0.2, group two is the group which data has $L/A \cdot S \cdot 100$ value in between 0-0.2.

3 Homogeneity Test and Distribution Determination

Homogeneity test is conducted to determine whether data originate from the same population because variability of one group of data is quite significant. This test used assumption that all of the stations are grouped together into one group. Thus all of the stations in that group follow one frequency distribution which has the same physically and characteristic. This means that the group has one growth factor value.

Regional frequency distribution test is performed after data have been generated 1000 years with 50 iteration following gamma distribution. The regional frequency distribution used in this test are Log Normal type 3, Pearson type 3, GEV and GPA. Graphical test result shows that the distribution is close to GEV and Pearson type 3 (see Figure 2).

Analytical test was done using Z-distribution test with formula 14-16 for each group. The result for group one is similar (0.269) between Log Normal type 3 and GEV distribution. The result for group two is almost similar between Log Normal type 3 and GEV, 0.206 and 0.233 respectively. This study is using GEV distribution instead of Log Normal type 3 because L-moment analytical solution for Log Normal type 3 is not explicitly solved (Hosking and Wallis, 2005). Chen et al, 2003

also noted that Log Normal type 3 has no explicit form of quantile function even though they used Log Normal type 3 in their research.

Heterogeneity test has been performed by using generated data (1000) and 50 iterations for each group. The average magnitude (μV and standard deviation (σV) of V is calculated to obtain the value of H that reflects the region homogeneity. $L-C_V$ parameter used in the homogeneity test should be divided with average $L-C_V$ but this was done on the iterations data only, not the original data. From the heterogeneity analysis, H value is 2.79 for group one and -0.022 for group two, therefore it can be said that stations in group one might be heterogenic and stations in group two is homogeny. Although group one has H value more than 2, it can be accepted if H value is less than 4 (Wang, 2000). H value shows the homogeneity rate. The higher the H value, the lesser the homogeneity rate.

4 Growth Factor Analysis

Growth factor is a correction coefficient and it is estimated by GEV distribution parameter in this study. The formulation to estimate the growth factor parameter with GEV distribution has more than one undefined constant, while the defined constant for L-moment analysis are t ($L-C_V$), t_3 (L-Skewness) and t_4 (L-Kurtosis). Therefore, to estimate α parameter, the derivation of basic formula (formulas 18-34) needs to be done and optimized in order to obtain lowest absolute error. The new formulas can be described as:

$$\alpha = 0,7384\tau_3 + 0,1318 \quad (37)$$

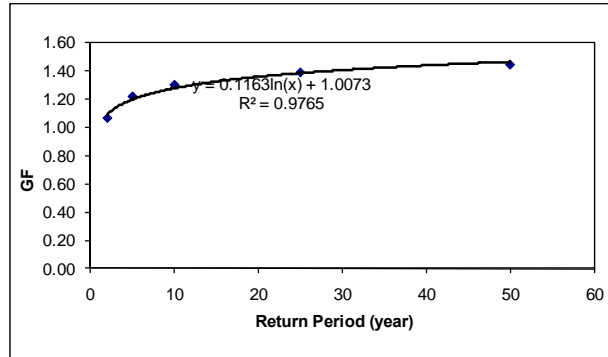
$$\xi = -1,1108\tau_3 + 1,0505 \quad (38)$$

The growth factor can be estimated by using the above equations (formula 37 and 38). Growth factor results are shown on Table 3.

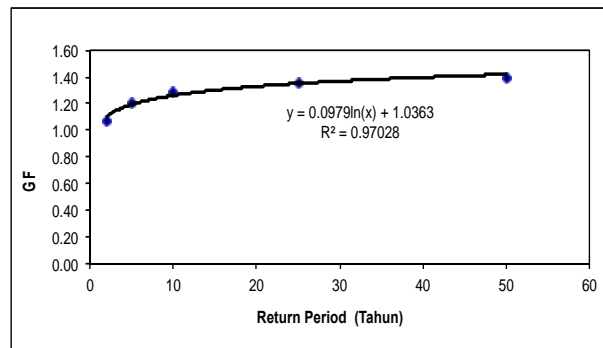
Table 3 (a) Parameter GEV, (b) growth factor

(a)	No	Cluster	N	t	t3	t4	c	K	$\Gamma(k)$	$\Gamma(k+1)$	α	ξ	λ_1	λ_2
	1	Cluster 1	7	0.21	0.04	0.17	0.026	0.209	4.412	0.923	0.164	1.003	7.757	1.662
	2	Cluster 2	15	0.189	0.022	0.140	0.031	0.246	3.736	0.920	0.149	1.026	7.223	1.365

(b)	No	Cluster	y							ξ	α	K	X_T						
			T2	T5	T10	T25	T50	T100	T200				T2	T5	T10	T25	T50	T100	T200
	1	Cluster 1	0.4	1.5	2.3	3.2	3.9	4.6	5.3	1.003	0.164	0.209	1.061	1.215	1.298	1.386	1.441	1.488	1.529
	2	Cluster 2	0.4	1.5	2.3	3.2	3.9	4.6	5.3	1.026	0.149	0.246	1.079	1.213	1.284	1.356	1.400	1.436	1.467



(a)



(b)

Figure 3 The derivation of growth factor, (a) group one, (b) group two

Note:

N , number of gauged stations
 t , variation coefficient
 t_3 , skewness coefficient
 t_4 , kurtosis coefficient
 α , shape
 ξ , scale
 k , location
 X_T , growth factor

After the growth factor value is identified, design flood can be estimated by using the growth factor and MAF observed. Design flood from simulation is compared with the actual design flood. The lower the absolute error, the better the calculation.

The relative mean bias analysis (BIASr) using formula 35 and 36 resulted that the high error seen at high return period, while the low error seen at low return period (0.01-0.66). BIASr value give unsatisfied result, thus the formula needs to be multiplied by reduction factor.

The formulation of reduction factor is done by optimization of minimum value of BIASr. In this process, reduction factor is divided into two parts, namely for the return period of 2 years and for periods of more than two years. The new formulas are described as a follow:

Group one:

For $T > 2$ years

$$FR_{>2} = 1.01925 * T^{0.04329} \quad (39)$$

For $T = 2$ year

$$FR_{=2} = 0.8472 * T^{0.04329} \quad (40)$$

Group two:

For $T > 2$ years

$$FR_{>2} = 1,0307 * T^{0.03109} \quad (41)$$

For $T = 2$ year

$$FR_{=2} = 0.8908 * T^{0.03109} \quad (42)$$

Note:

FR_T , reduction factor for T return period

T , return period (year)

In addition, the derivation of Growth Curve needs to be done to calculate the Growth Factor with different return periods. This derivation is made based on the previous growth factor. The result of the growth factor derivation can be seen on Figure 3.

The derivation of the new growth factor is feasible with R^2 0.9765 and 0.9703 for group one and two respectively. Optimization of this equation is made to have better BIASr values. The new growth factor and complete formula to calculate design flood is described below.

Group one:

$$X_T = 0.1067 * \ln(T) + 1.0427 \quad (43)$$

With KAR value 1.46%

For $T > 2$ years

$$Q_T = (1.01925 * T^{0.04329}) * (0.1067 * \ln(T) + 1.0427) * MAF \quad (44)$$

For $T = 2$ year

$$Q_T = (0.8472 * T^{0.04329}) * (0.1067 * \ln(T) + 1.0427) * MAF \quad (45)$$

$R^2 = 0.9438$ and KAR value = 1 %

Group two:

$$X_T = 0.0882 * \ln(T) + 1.0712 \quad (46)$$

With KAR value 1.39%

For $T > 2$ years

$$Q_T = (1.0307 * T^{0.03109}) * (0.0882 * \ln(T) + 1.0712) * MAF \quad (47)$$

For $T = 2$ year

$$Q_T = (0.8908 * T^{0.03109}) * (0.0882 * \ln(T) + 1.0712) * MAF \quad (48)$$

$R^2 = 0.9731$ and BIASr value = 0.5%

Note:

X_T , growth factor

T , return period

Q_T , design flood

MAF, mean annual flood

MAF values used in this calculation are observed MAF. The BIASr result is summarized on Table 4 for original calculation, without reduction factor and with reduction factor.

Table 4 shows that BIASr values with reduction factor are better than without reduction factor. Cluster one has higher BIASr value, more than 20% for return period 25 and 50 years, while cluster two has lower BIASr value with maximum value 10.3% for return period 50 years. This phenomena show that the homogeneity rate determines BIASr value, cluster two has H value - 0.0233 and cluster one is 2.792.

Table 4 Summarize of BIASr values

Cluster	Cases	Return period (years)				
		2	5	10	25	50
1	Original	13,0	10,2	18,6	25,8	29,6
	Without FR	19,0	10,2	19,0	25,8	29,5
	With FR	7,4	7,7	14,7	21,6	26,7
2	Original	7,2	7,6	11,2	13,4	14,5
	Without FR	12,1	7,6	11,7	13,5	13,6
	With FR	4,0	4,5	5,5	7,6	10,3
Combination 1 and 2	With FR	5,1	5,5	8,4	12,3	15,5

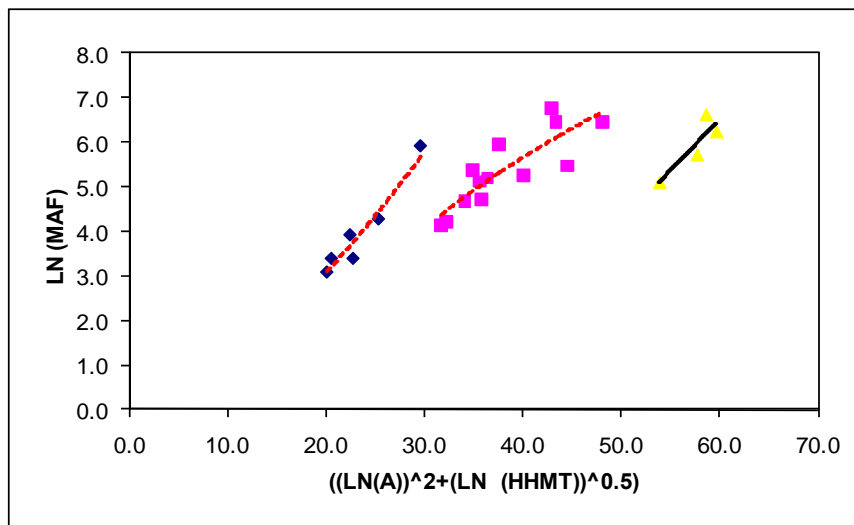


Figure 4 MAF, catchment area and maximum daily precipitation relationship graph

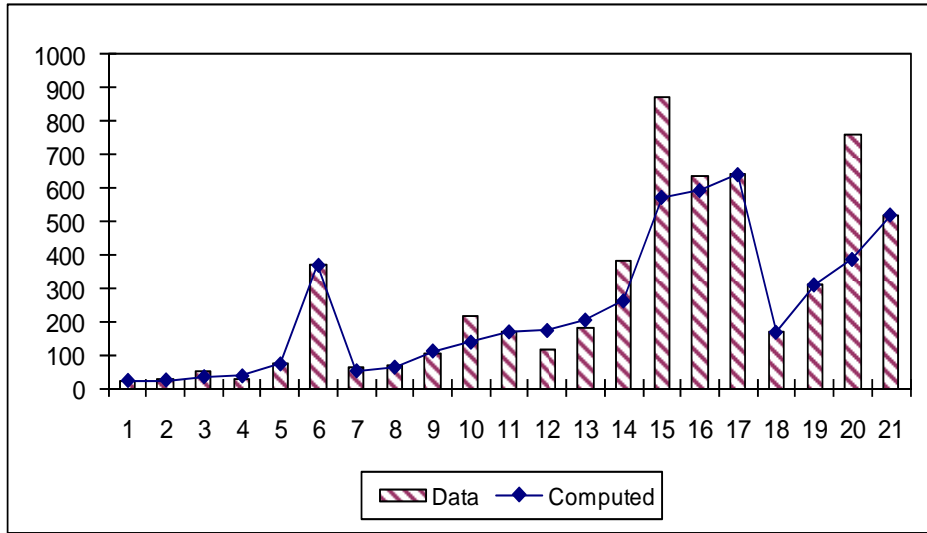


Figure 5 MAF calculated and MAF observation comparison

Table 5 Growth factor verification

No	Cluster	y							ξ	α	K	X _T						
		T2	T5	T10	T25	T50	T100	T200				T2	T5	T10	T25	T50	T100	T200
1	1 cluster	0.4	1.5	2.3	3.2	3.9	4.6	5.3	1.020	0.153	0.236	1.074	1.213	1.287	1.363	1.410	1.449	1.482

Table 6 Kiarapayung design flood estimation and BIAS

No	Name	Q _{comp} (m³/s)					Q _{obs} (m³/s)					BIAS				
		PU2	PU5	PU10	PU25	PU50	PU2	PU5	PU10	PU25	PU50	PU2	PU5	PU10	PU25	PU50
1	Kiarapayung	687.23	776.5	823.66	872.5	902.3	629.7	794.4	888.7	994.96	1066.35	0.09	0.02	0.07	0.12	0.15

5 MAF Analysis

Mean Annual Flood (MAF) is a common variable to determine the index flood values (Wharton and Tomlinson, 1999). MAF can be calculated using regression analysis based on the relationship between MAF and catchment characteristic. After some regression analyses using catchment characteristic and MAF, the dominant variable is catchment area with $R^2=0.674$. The other important parameter is precipitation. Thus the MAF analysis will be calculated using catchment area and maximum daily precipitation. Result of regression and optimalization between MAF, catchment area and maximum daily precipitation is shown on Figure 4.

Figure 4 shows that the data are scattered into three groups, group A has X value in between 20-30, group B has X value in between 30-50 and group C has X value in between 50-60. X value is described as follow.

$$X = (\ln(A))^2 + (\ln(HHMT))^{0.5} \quad (49)$$

Where A is catchment area and HHMT is average maximum daily precipitation in every year.

Therefore, the further MAF calculation is divided into three groups. Formula for each group is described as follow.

Group A:

$$\ln(MAF) = 0.0165 * X^2 - 0.5239 * X + 6.9568 \quad (50)$$

Group B

$$\ln(MAF) = -0.012 * X^2 + 1.093654 * X - 18.8815 \quad (51)$$

Group C

$$\ln(MAF) = 0.0178 * X^2 - 1.8219 * X + 51.7147 \quad (52)$$

Group A has R^2 value 0.9607 and BIASr 12.5%, group B has R^2 value 0.92 and BIASr 18.8%, and group C has R^2 value 0.78 and BIASr 12.4%. Comparison of MAF calculated and MAF observation can be seen on Figure 5.

6 Model Validation

Validation has been done using split sampling method. Cipunegara-Kiarapayung station has been taken out from the group and uses for validation as an ungauged area. Growth factor analysis was repeated without using that station

and was implemented to estimate design flood (see Table 5). Growth factor analysis has been done using one group calculation only, not divided into two groups. MAF is estimated using catchment characteristic as a follow:

Catchment area: 879.8 km²

River length: 66.31 km

Slope: 0.00697

HHMT: 116.8 mm

X is calculated using formula 49 with the result 48.15 and categorized into group B. MAF has been calculated using formula 51 with the result 639.92 m³/sec. This value is multiplied by the new growth factor value to have the design flood with various return periods (see Table 6).

Table 6 shows that BIAS values are 9%, 2%, 7%, 12% and 15% for return period 2, 5, 10, 25, and 50 years respectively. BIASr value is higher in the highest return period; this might be caused by the length of the data. The length of data used in this study is 10-25 years. It is advisable to calculate the maximum return period twice as the length of the data (max 50 years return period). Model verification shows that the developed model gives good result to calculate design flood in ungauged catchment with BIASr 9%.

CONCLUDING REMARKS

Catchment characteristic is priority in this analysis to analyze clustering and MAF. Hosking and Wallis (1997) recommend that H value should be less than 2. However, many further studies extent this limit into less than 4 (e.g. Wang, 2000). This extension is also supported by some studies carried out by Research Center for Water Resources (Puslitbang Air, 2006; Puslitbang Air, 2007; Puslitbang Air, 2008; Puslitbang Air 2009). This paper also proved that the results from the developed formula using H value less than 4 are acceptable. The important thing that should be taking into account is: the more homogeny the cluster, the lowest the BIASr value.

Index flood and L-moment method proved to be useful methods to calculate the design flood as long as no discordance in the data. Discordance station makes the group heterogenic and the requirement for L-moment analysis is the data should be homogeny. Moreover, the developed formula is stable and feasible. Although the verification process did not include the reduction factor, the BIASr value is 15% for 50 years return period and it will reduce if reduction factor is included in the verification process. Based on the verification results thus the developed formula can

be used in ungauged catchments, in West Java Province for flood designing and planning.

ACKNOWLEDGEMENTS

This study was financed by the Ministry of Public Work, Indonesia. We thank to Research Center for Water Resources staffs for their great job and effort to do this study. In addition, the authors extend their thanks to reviewers for their important suggestions and improving the text.

REFERENCES

- Burn D. 1990. Evaluation of regional flood frequency analysis with a region of influence approach. *Water Resources Research WREARQ 26*: 2257-2265.
- Cannarozzo M., D.Asaro F., Ferro V. 1995. Regional rainfall and flood frequency analysis for Sicily using the two component extreme value distribution. *Hydrological Sciences Journal 40*: 19-42.
- Cunderlik J.M., Burn D.H. 2002. Analysis of the linkage between rain and flood regime and its application to regional flood frequency estimation. *Journal of Hydrology 261*, p.115-131.
- Chen Y.D., Huang G., Shao Q., Xu C. 2006. Regional analysis of low flow using L-moment for Dongjiang basin, South China. *Hydrological Sciences Journal, 51:6, 1051-1064*.
- Dalrymle T. 1960. Flood frequency analyses. Water supply paper 1543-A. *Us Geological Survey, pp. 11-51*.
- Ferrari E., Gabriele S., Villani P. 1993. *Combined regional frequency analysis of extreme rainfalls and floods*. IAHS PUBLICATION: 333-333.
- Gong X., Richman M. 1995. On the Application of Cluster Analysis to Growing Season Precipitation Data in North America East of the Rockies. *Journal of Climate 8*: 897-931.
- Hosking J., Wallis J., Wood E. 1985. An appraisal of the regional flood frequency procedure in the UK Flood Studies Report. *Hydrological Sciences Journal 30*: 85-109.
- Hosking J and Wallis J. 1988. The effect of intersite dependence on regional flood frequency analysis. *Water Resources Research 24*: 588-600.
- Hosking J.R.M and Wallis J.R. 1997. Regional Frequency Anlysis. *An Approach Based on*

- L-Moments*. Cambridge University Press, Cambridge, United Kingdom.
- Hosking J.R.M and Wallis J.R. 2005. *Regional frequency analysis: an approach based on l-moments (PDO)*. Genie Civil, Batiment, Travaux Publics.
- Institute of Hydrology. 1983. *Flood Design Manual for Java and Sumatra*, Wallingford, Oxon, England.
- Lettenmaier D., Potter K. 1985. Testing flood frequency estimation methods using a regional flood generation model. *Water Resources Research* 21: 1903-1914.
- Lettenmaier D., Wallis J., Wood E. 1987. Effect of regional heterogeneity on flood frequency estimation. *Water Resources Research* 23: 313-323.
- Leclerc M., Ouarda T.B.M.J. 2007. Non-stationary regional flood frequency analysis at ungauged sites. *Journal of Hydrology* 343, p.254-265.
- Lim Y.H and Lye L.M. 2003. Regional flood estimation for ungauged basins in Sarawak Malaysia, *Hydrological Sciences* 48(1).
- Malekinezhad H., Nachtnebel H.P., Klik A. 2010. Comparing the index flood and multiple regression method using L-moments. *Physics and chemistry of the earth*, p.1-7, doi:10.1016/j.pce.2010.07.013.
- Puslitbang Air. 2006. Aplikasi teknologi dan metodologi hidrologi untuk mengetahui karakteristik hidrologi pada suatu DAS. IP.07.02/04/La-HITA/2006.
- Puslitbang Air. 2007. Penelitian parameter banjir regional untuk estimasi banjir rencana pada DAS tidak terukur. IP.07.02/07/La-HITA/2007.
- Puslitbang Air. 2008. Penelitian pengendalian day rusak air. Volume II, IP.0101/05/La-HITA/2008.
- Puslitbang Air. 2009. Penelitian parameter banjir regional untuk estimasi banjir rencana pada DAS tidak terukur. IP.0703/04/La-HITA/2009.
- Stedinger J.R., Richard M. Vogeldan Efi Foufoula-Georgiou. 1993. Frequency analysis of extreme events. *Handbook of Hydrology*, edited by D.R. Maidment, Chapter 18. McGraw-Hill, New York.
- Saf B. 2009. Regional flood frequency analysis using L-moment for the West Mediterranean Region of Turkey. *Water Resour Manage* 23, p. 531-551. doi: 10.1007/s11269-008-9287-z.
- Wallis J.R., Wood E.F. 1985. Relative accuracy of log Pearson III procedures. *J. Hydraul Div Am Soc Civ Eng* 111(7): 1043-1056.
- Wang Y. 2000. *Development of methods for regional flood estimates in the Province of British Columbia Canada*. The University of British Columbia, Canada.
- Wharton G and Tomlinson J.J. 1999. Flood discharge estimation from river channel dimensions: results of applications in Java, Burundi, Ghana and Tanzania. *Hydrological Sciences* 44(1).