# Short Rainfall Duration Evaluation in Design Storms Development

## Evaluasi Durasi Hujan Pendek Dalam Pengembangan Perancangan Badai

**Masimin** [1] **and Sobri Harun** [2]

[1] Department of Civil Engineering, Syiah Kuala University, Banda Aceh, Indonesia 23111, masimin@plasa.com
[2] Faculty of Civil Engineering, Universiti Teknologi Malaysia, 81300 Skudai, Johor Bahru, Malaysia.

## ABSTRACT

A 5-minute incremental rainfall data taken from 5 stations distributed in Johor southwest region was analyzed to study an inten-sity-duration-frequency (IDF) related to design storm development. The study includes a regional rainfall analysis, determination of probability distribution, estimation parameter and quantiles of the distribution, development of IDF, and the comparison study for IDF estimates based on daily rainfall. In the study, the duration pattern is set to 15-, 30-minute, 1-, 2-, 4-, 8-, 16- hour, and 1-day. An X10- test is used for regional analysis to define homogeneous region and probability weighted moments (PWM) is used to do a frequency analysis. To see the presence of trends, the Kendall's method is applied. It is found that the region is passed the tests to be the homogeneous region. Based on frequency analysis, the data tends to follow the generalized extreme value (GEV) and generalized logistic (GLOG) distribution. In regression analysis, the IDF-curve for return period of 2-, 5-, 10-, 20-, 50-, 100-, 200-, and 500-year tends to follow a power relationship. The IDF family curves calculated based on daily data gives over estimates about 10% compared to those of calculated based on the 5-minute incremental rainfall data.

**Keywords:** regional homogeneity, design storms, frequency analysis

## ABSTRAKSI

Data hujan tiap 5 menit yang diambil dari 5 stasiun yang tersebar di wilayah barat daya Johor dianalisis untuk mem-pelajari intensitas-dirasi-frekuensi (IDF) yang berhubungan dengan pengembangan perancangan badai. Penelitian ini mencakup analisis hujan regional, penentuan distribusi probabilitas, estimasi parameter dan quantil distribusi, pe-ngembangan IDF, dan studi perbandingan untuk estimasi IDF berdasarkan hujan harian. Dalam penenlitian ini pola durasi diatur dalam 15-, 30-menit, 1-, 2-, 4-, 8-, 16-jam, dan 1-hari. Pengujian X10 digunakan untuk analsisi regio-nal untuk menentukan keseragaman regional dan probabilitas berat momen (PWM) digunakan untuk nanlisis fre-kuensi. Untuk melihat munculnya tren, metoda Kendall digunakan. Hasil analisis menunjukkan bahwa wilayah yang diteliti adalah wilayahyang seragam. Berdasarkan analisis frekuensi, data cenderung mengikuti nilai akstrim yang u-mum (GEV) dan distribusi logistik umum (GLOG). Pada analisis regresi, kurva IDF pada periode 2-, 5-, 10-, 20-, 50-,100-, 200-, dan 500-tahun cenderung mengukuti persamaan power. Kurva famili IDF berdasarkan data harian memberikan over estimasi sekitar 10% dibandingkan dengan data yang dihitung berdasarkan data hujan tiap 5 menit.

**Kata-kata Kunci :** Keseragaman wilayah, perancangan badai, analisis frekuensi

## INTRODUCTION

Estimation of design storms is an unrest topic in statistical hydrology that needed in a broad spectrum in civil works especially for an urban drainage design. One type of design storms development is classified as frequency based storm that the method is based on statistical frequency analysis (Viessman and Lewis, 2003). Small size project with low hazard level is commonly using this method with return period is less than 100 years. This study covers a topic of design storms with a case study area of southwest of Johor, Malaysia. The study includes a regional rainfall analysis, determination of probability distributi-on, estimation parameter and quantiles of the distribution, deve-lopment of IDF, and the comparison study for IDF estimates ba-sed on daily rainfall data proposed by Mononobe (Sosrodarsono and Takeda, 1980).

A rational method is a simple formula and widely used in design storms that the flow rate (Q) is a function of runoff coe-fficient (C), intensity (I), and area (A). Two variables C and A are the physical measures and variable I is a hydrological condition at the region. The value of intensity (I) is based on rainfall duration that equals or greater than time of concentration (tc). Design storm is mostly applied to short rainfall durations that the availability of short incremental rainfall data becomes problem because most of rainfall stations only give the data based on daily data bases. Rain-fall intensity prediction based on daily rainfall data was proposed and called as Mononobe method with the relation as in equation (1).

$$I = \frac{R_T}{24}\left(\frac{24}{t_c}\right)^{2/3} \tag{1}$$

where, $I$ = rainfall intensity (mm/hour); $R_T$ = daily rainfall for the return period of T years; and $tc$ = rainfall duration equal to its time of concentration (minutes). Although the rational method is very popular and widely used in defining the peak flow rate for urban drainage project, it has limitation for its coverage area. The method is suitable for the coverage area of less than 600 acres and for the coverage areas of urban drainage are greater than 600 acres, the hydrograph analysis is warranted (Viessman and Lewis, 2003).

Daud et al. (2002) studied of extreme rainfall process in Peninsular Malaysia and they found that the GEV was the most suitable distribution to be applied for annual maxima rainfall data in Peninsula Malaysia. The regional study for the state of Selangor using long duration rainfall data, it was found that the region of was determined as a homogeneous region and the sui-table statistical distributions were GEV and GLOG distributions (Masimin and Harun, 2006). There is no information yet for the study of design storms using short duration rainfall data in the region including for Johor. Adamowski and Bougadis (2003) stu-died about IDF using rainfall data in Canada and they found that the presence trends in short rainfall duration was significant in design storms.

The objectives on the study consist of the development of IDF curves based on 5-minute incremental rainfall data and the

comparison of IDF family curves defined based on daily rainfall data records to the those of first objective.

## DATA AVAILABILITY AND PRELIMINARY ANALYSIS

Five rainfall stations in Johor provide a 5-minute incrementtal rainfall for 25 years duration of record (1980-2004). Those five stations are Sta.1437116, 1534002, 15411139, 1636001, and 1732004. The first step in the study is to define an annual maxima rainfall depth based on the duration pattern. The 'block method' is used for data extraction that the block is defined based on duration pattern, so that it is not an event based but time based analysis. The reason of using the time base analysis is that a short duration rainfall is a 'cloud burst' and long duration rainfall is a multiple storms (Sveinsson et al., 2002). The event based analysis will give a small magnitude compared to those of time base data.

A preliminary data analysis is conducted by doing some sta-tistical tests included the independence and stationarity test and spatial correlation samples. The test for independence and station-nary is executed using equation (2) to (5) called the Wald-Wolfowitz testing method (WW-test) to test for the independence of a data set and to test for the existence of trend on it (Rao and Hamed, 2000). The WW-test showed that all the data set were passed to be independent as the statistical results were shown in Table 1.

$$u = (R - \overline{R}) / (\text{var}(R))^{1/2} \qquad (2)$$

$$R = \sum_{i=1}^{N} x_i x_{i+1} + x_i x_N \qquad (3)$$

$$\overline{R} = (s_1^2 - s_2) / (N-1) \qquad (4)$$

$$\text{var}(R) = (s_2^2 - s_4)/(N-1) - \overline{R}^2 + \\ (s_1^4 - 4s_1^2 s_2 + 4s_1 s_3 + s_2^2 - 2s_4)/((N-1)(N-2)) \qquad (5)$$

where $u$, $R$, $\overline{R}$ = the statistical test parameters; $s_r = Nm_r'$ and $m_r'$ is the $r^{th}$ moment of the sample about the origin.

The spatial sample correlation was used to examine the correlation of every pairs observed point data sets that was separated by some distances (Kottegoda & Rosso, 1997; Berndtsson & Niemczynowics, 1986). The test was executed using equation (6) and it is about 80 tests are executed applied to the eight durations pattern for data sets. The results are plotted as in Fig.1.

$$r(h) = \frac{\sum_{i=1}^{N} [z_i(u) - \hat{m}(u)][z_i(u+h) - \hat{m}(u+h))}{\{\sum_{i=1}^{N} [z_i - \hat{m}(u)]^2 \sum_{i=1}^{N} [z_i(u+h) - \hat{m}(u+h)]^2\}^{1/2}} \qquad (6)$$

where $r(h)$ = sample correlation; $\hat{m}(u)$ and $\hat{m}(u+h)$ are the sample means of the observations at the two points.

Fig.1 shows some information that some pairs of data points for a shorter distance (20-50 Km) give correlation values about zero and some data points for a long distance (70-100 Km) give correlation with greater values. Examining the location of the stations, the distance from the sea would be the significance parameter in partitioning the region to homogeneous sub regions. It is agree to the study of regional homogeneity using Selangor's rainfall data (Masimin & Harun, 2006).

## SPATIAL RAINFALL VARIABILITY

Rainfall depth varies in space or location when it examined for the same time of occurrence since storm has a center or storm eye which the highest magnitude occurs and it also has a travel path. The rainfall depth is decreasing upon the location is increasing in distance from the storm center. Since this study is intended to develop of design storms, the analysis is only using the daily annual maxima rainfall data, so it is only one rainfall data represented for a year of record. There were 5 data sets from 5 stations in southwest Johor that each station provided 25 numbers of data for 25 years length of record since 1980 to 2004. In some cases, the point of annual daily maxima rainfall data is not coincidence with an areal data.

Station data correlation is set for the correlation among 5 available stations to define their slopes. The analysis is done by plotting their rainfall data and formulated linear correlation by setting the intercept is equal to zero. The two sets of rainfall data are correlated when its formulation is Y $\approx$ X where Y is an ordinate of $1^{st}$ data set and X is an axis of $2^{nd}$ data set. When Y $\approx$ X or the slope is about unity applied, the two data sets can be treated as at-site data and the two locations are assumed to have no spatial rainfall variability and otherwise shall be treated individuallly. The slopes of linear regression are summaries in Table 2 and an example of the data plotting can be seen on Fig. 3.
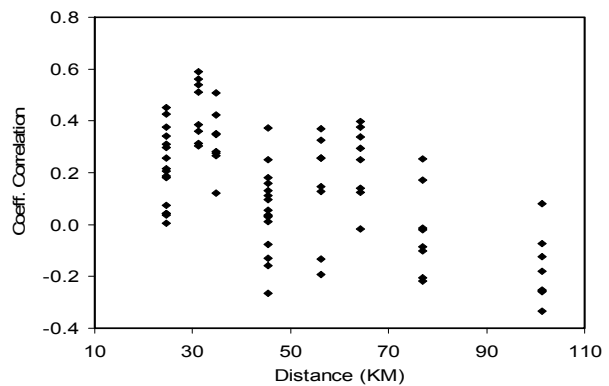


Fig. 1. Spatial correlation and distance plot for overall data among the stations

Fig. 2 shows that the data set from Sta1451139 had no correlation at all to the neighboring stations since its values are about 0.50 and the data from this station will be treated individually. The rest of four stations have a good correlation by showing the correlation values are about one, so that at-site data will applied to the sub-region. The conclusion of the preliminary analysis is that the region is divided into two sub-regions; sub-region I for northea-stern coastal area is represented by Sta1541139, and sub-region II for southwestern coastal area is represented by Sta1437116, 1636001, 1534002, and 1732004.

Above analysis is based on the annual daily maxima rainfall data because of the purpose of the study is for design storm analysis. The same analysis could be done for the different types of rainfall duration pattern. When comparing the magnitude of rainfall depth between northeastern and southwestern coastal region, it is found that the rainfall depth in northeastern coastal region is about 60% higher compared to those of southwestern coastal region although it is only one station represented for northeastern coastal region.

Table 2. The slope magnitudes for station data distribution

| Stations | 1541 139 | 1437 116 | 153 002 | 1636 001 | 1732 004 |
|---|---|---|---|---|---|
| 1541139 | 1.0 | - | - | - | - |
| 1437116 | 0.593 | 1.0 | - | - | - |
| 1534002 | 0.594 | 1.012 | 1.0 | - | - |
| 1636001 | 0.592 | 0.967 | 0.927 | 1.000 | - |
| 1732004 | 0.582 | 1.044 | 1.003 | 0.800 | 1.0 |

## Spatial data variability

The spatial rainfall variability is analyzed based annual daily maxima rainfall magnitude for 5 stations taken from the same day of record. It is assumed that the values are to be used in rainfall frequency analysis. The results of data extraction for the annual daily maxima were plotted and presented in Fig.3 and Fig.4.
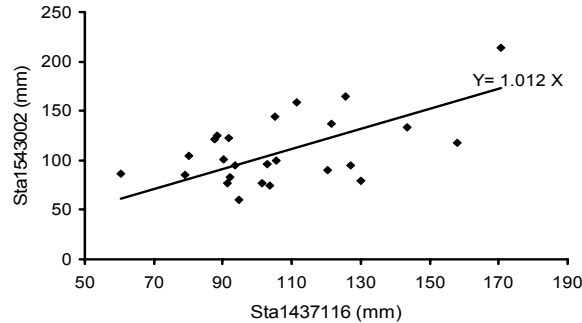


Fig. 2. Position of the stations and data correlation

Table 1. The statistical test results ($u$) of independence and stationary data set

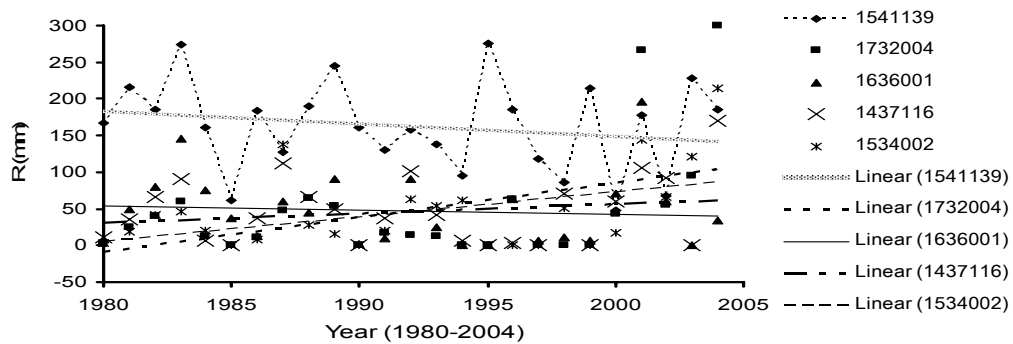| Sta | $u_{\alpha/2}$ | Duration pattern (min) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 15 | 30 | 60 | 120 | 240 | 480 | 960 | 1440 |
| 1437116 | | -0.25 | -0.28 | -0.32 | -0.33 | -0.33 | -0.32 | -0.38 | -0.43 |
| 1534002 | | -0.27 | -0.43 | -0.40 | -0.47 | -0.41 | -0.41 | -0.36 | -0.38 |
| 1541139 | 1.96 | 0.72 | -0.11 | -0.40 | -0.48 | -0.40 | -0.35 | -0.40 | -0.44 |
| 1636001 | | -0.16 | -0.27 | -0.38 | -0.40 | -0.35 | -0.40 | -0.45 | -0.57 |
| 1732004 | | -0.32 | -0.31 | -0.37 | -0.35 | -0.32 | -0.46 | -0.61 | -0.78 |

Resource: Station data correlation.



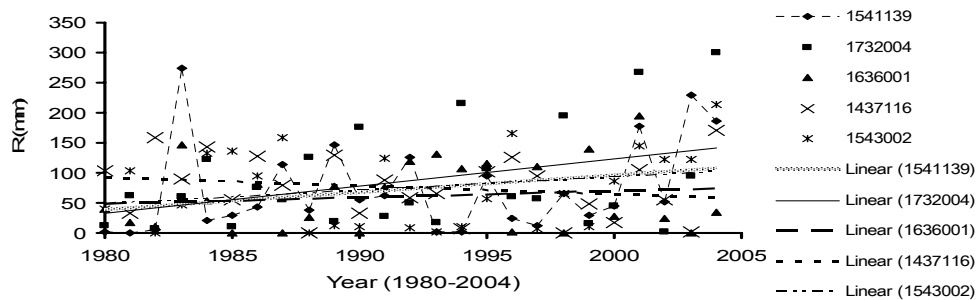Fig.3. Plot of data for the annual daily maxima values for Sta.1541139 and others



Fig.4. Plot of the annual daily maxima values of west coastal stations and Sta1541139

## THEORY AND METHODS

### Probability Weighted Moments (PWM)

Three of the more commonly can be used for parameter estimation were the method of moments (MOM), the maximum likelihood method (MLM), and the probability weighted moments (PWM). Among them, the PWM method gave parameter estimates comparable to MLM estimates, yet in some cases the estimation procedure was much less complicated and the computation is simpler (Rao & Hamed, 2000). The PWM's method was firstly introduced by Greenwood et al. (1979) based on the concept that the distributions whose inverse forms were explicitly defined, such as Turkey's lambda, may present problems in deriving their parameters by more conventional means. It has received considerable attention to the researchers and used it in their research works (Hosking et al., 1985; Hosking and Wallis, 1987; Rao & Hamed., 1997; Sveinsson et al., 2002; Kumar et al., 2005).

Hosking (1985, 1990) has defined the *L*-moments (Rao & Hamed, 1997). The *L*-moments were analogous to conventional moments and were estimated by linear combination of order statistics. L-moment can also be expressed by linear combination of PWM. Thus, the procedures based on PWM and *L*-moments were equivalent. However, *L*-moments are more convenient because they were directly interpretable as measure of the scale and of the shape of probability distribution (Hosking & Wallis, 1987). Hosking (1990) has used *L*-moments ratio diagram to identify underlying parent distributions and *L*-moment ratios for testing hypotheses about different probability distributions. Hosking and Wallis (1993) extended the use of L-moments and developed statistics that can be used for regional frequency analysis to measure discordance, regional homogeneity, and goodness-of-fit.

Probability weighted moments $M_{prs}$ was proposed by Green-wood et al. (1979) with the following relationship.

$$M_{p,r,s} = E\{[x(F)]^p F^r [1-F]^s\}$$
$$= \int_0^1 [x(F)]^p F^r (1-F)^s dF \tag{7}$$

where *F* is the cumulative distribution and there were two moments to be considered:

$$M_{1,0,s} = \alpha_s = \int_0^1 x(F)(1-F)^s dF \tag{8}$$

$$M_{1,r,0} = \beta_r = \int_0^1 x(F)F^r dF \tag{9}$$

Both $\alpha_s$ and $\beta_r$ are linear in *x* and are of sufficient generality for parameter estimation. There are also relationship between $\alpha_s$ and $\beta_r$ as in (10) and (11)

$$\alpha_s = \sum_{k=0}^s \binom{s}{k}(-1)^k \beta_k \tag{10}$$

$$\beta_r = \sum_{k=0}^r \binom{r}{k}(-1)^k \alpha_k \tag{11}$$

The *L*-moments $\lambda_{r+1}$ were introduced by Hosking (1985, 1990) which were linear function of PWM's. It was more convenient than PWM's because of the direct interpretation of the measures of scale and shape probability distribution. In terms of PWM's, $\alpha_s$ and $\beta_r$ are defined as in (12).

$$\lambda_{r+1} = (-1^r)\sum_{k=0}^r p_{r,k}^* \alpha_k$$
$$= \sum_{k=0}^r p_{r,k}^* \beta_k \tag{12}$$

where

$$p_{r,k}^* = (-1)^{r-k}\binom{r}{k}\binom{r+k}{k} \tag{13}$$

For given ordered sample $x_i \leq \ldots \leq x_n, n > r,$ and $n > s$ the unbiased sample PWM's are calculated (Hosking 1986) using (14) and (15)

$$a_s = \tfrac{1}{n}\sum_{i=1}^n \binom{n-i}{s}x_i \bigg/ \binom{n-1}{s} \tag{14}$$

$$b_r = \tfrac{1}{n}\sum_{i=1}^n \binom{i-1}{r}x_i \bigg/ \binom{n-1}{r} \tag{15}$$

Alternatively, consistent, but not unbiased, plotting position estimators are obtained by using (16) and (17)

$$a_s = \hat{\alpha}_s = \tfrac{1}{n}\sum_{i=1}^n (1-P_{i:n})^s x_i \tag{16}$$

$$b_r = \beta_r = \tfrac{1}{n}\sum_{i=1}^n P_{i:n}^r x_i \tag{17}$$

where $P_{i:n}$ is the plotting position. The use of $P_{i:n} = (i - 0.35)/n$ usually gives good estimates of parameters and quantiles for generalized extreme value (GEV) (Hosking et al. 1985), the generalized Pareto (Hosking and Wallis, 1987), and Wakeby distributions (Lettenmaier et al. 1987). However Hosking and Wallis (1993) recommended using plotting position estimators when using a Wakebay distribution to estimate extreme upper tail quantiles in a regional frequency analysis and to use unbiased estimators under all other circumstances. Sample L-moments ($l_r$) are calculated by using (12) and (13), replacing $\alpha_s$ or $\beta_r$ with their sample estima-tes $a_s$ and $b_r$ from (16) and (17). The L-moments ratio, which are analogous to the conventional moment ratio, are defined by Hosking (1985, 1990) in (18) and 19)

$$\tau = \lambda_2 / \lambda_1 \tag{18}$$

$$\tau_r = \lambda_r / \lambda_2; \quad r \geq 3 \tag{19}$$

where $\lambda_1$ = measure of location; $\tau$ = measure of scale and dispersion ($LC_v$); $\tau_3$ = measure of skewness ($L$-$C_s$); and $\tau_4$ = measure of kurtosis ($L$-$C_k$). Sample L-moment ratio (denoted *t* and $t_r$) are calculated by using (18) and (19), substituting sample estimates $l_r$ for population value $\lambda_r$. The L-moment ratio offer an easy way to identify underlying, particularly skewed, distributions (Hosking, 1990).

### Generalized Extreme Value (GEV) Distribution

The GEV distribution was introduced by Jenkinson (1955 and 1969) to identify the frequency distribution of the largest values of meteorological data when the limiting form of the extreme value distribution was unknown. The cumulative density function of the GEV distribution was as in (20) (Kottegoda and Rosso, 1997; Burn (1990)) for $0 \leq x$ with $\lambda > 0$ and $r > 0$.

$$F(x_{max}) = \exp\{-[1-(k(x-\varepsilon)/\alpha)]^{1/k}\} \tag{20}$$

where $\alpha$ denotes a scale parameter, $\varepsilon$ denotes a location parameter, and $k$ is the shape parameter. The PWM's parameter estimates of the GEV are of the form in equation (21) (Hosking, et al., 1985).

$$\beta_r = (r+1)^{-1}[u + (\alpha/\hat{k})\{1 - (r+1)^{-\hat{k}}\Gamma(1+\hat{k})\}] \quad (21)$$

The value of parameter $k$ is given as (22) with $C = (2/(3+t_3)-0.6309$ (Sveinsson et al., 2002).

$$\hat{k} = 7.8590C + 2.9954C^2 \quad (22)$$

Once the value of $k$ is obtained $\hat{\alpha}$ and $\hat{\mu}$ are estimated by equation (23) and (24) where $b_0, b_1, b_2$ are the sample estimates of $\beta_0, \beta_1, \beta_2$.

$$\hat{\alpha} = l_2\hat{k}/\{\Gamma(1+\hat{k})(1 - 2^{-\hat{k}})\} \quad (23)$$

$$\hat{u} = l_1 + (\hat{\alpha}/\hat{k})\{\Gamma(1+\hat{k}) - 1\} \quad (24)$$

The inverse form of the distribution function is written as (25) and by substituting $F = 1 – 1/T$ where $T$ is the return period, the T-year quantile estimate is obtained by using (26),

$$x = u + (\alpha/k)\{1 - (-\log F)^k\} \quad (25)$$

$$\hat{x}_T = \hat{u} + (\hat{\alpha}/\hat{k})[1 - \{-\log(1 - (1/T))\}^{\hat{k}}] \quad (26)$$

## Generalized Logistic (GLOG) Distribution

The GLOG distribution function is defined as in (27), that

$$F(x) = [1 + \{1 - k((x-\varepsilon)/\alpha)\}^{1/k}]^{-1} \quad (27)$$

where ,
$$\begin{cases} \varepsilon + (\alpha/k) \le x < \alpha & k < 0 \\ -\infty < x \le \varepsilon + (\alpha/k) & k > 0 \end{cases} \cdots \quad (28)$$

The parameter estimates and quantile are calculated using (29) to (32) with $\hat{k} = -t_3$ and $F = 1 – (1/T)$ where $T$ is the return period in years.

$$\hat{\alpha} = l_2/[\Gamma(1+\hat{k})\Gamma(1-\hat{k})] \quad (29)$$

$$\hat{\varepsilon} = l_1 + (l_2 - \hat{\alpha})/\hat{k} \quad (30)$$

$$x = \varepsilon + (\alpha/k)[1 - \{1 - F)/F\}^k] \quad (31)$$

$$\hat{x} = \hat{\varepsilon} + (\hat{\alpha}/\hat{k})[1 - (T-1)^{-\hat{k}}] \quad (32)$$

## Regional Homogeneity

Current point rainfall frequency analysis techniques used in engineering design are outdated and should be developed based on regional analysis (Adamowski et. al., 1996). Regional frequency analysis uses data from several sites to estimate quantiles of underlying variable at each site in the region of consideration (Cunane, 1988). The analysis involves identification of the region, i.e., the site that belong to the region, testing whether the proposed region is homogeneous, choice of the distribution to fit the region data, and estimation of parameters and quantiles (Sveinsson et al., 2002). The identification of the region is something subjective and based on site characteristics; delineation of the region can be approached. The site characteristics includes: the latitude (°), longitude (°), alti-tude (°), concentration of precipitation (%), mean annual precipi-tation (mm), rainfall seasonability (category), and distance from the sea (m) (Smithers and Schulze, 2001).

The second step is testing for regional homogeneity of the regional data that is very important for the regional frequency analysis (Dinpashoh, 2004). In practice, homogeneity is judged by the variability, among sites, of coefficient of variance $Cv$ and

/or the skew coefficient $Cs$, of their $L$-moment equivalent, or dimensionless quantiles (Hosking, 1990; Fill and Stedinger, 1995). Fill and Stedinger (1995) studied some homogeneity tests included an index flood or Dalrymple's test, a normalized quantile test based upon $L$-moment parameter estimation ($X$-$10$ test), and the method of moment $Cv$ test ($MoM$-$Cv$ test). Other $L$-moment based tests was $H$-$W$ test proposed by Hosking and Wallis (1993) that the test would be equivalent to the $X$-$10$ test if the at-site quantiles were estimated using a fix value of shape parameter $k$.

Based on Lu and Stedinger (1992), the $X$-$10$ test was always more powerful than the other two tests and also the $X$-$10$ test for the GEV distribution has similar power to the $L$-$Cv$ regional homo-geneity test used by Hosking and Wallis (1993, 1997). The $X$-$10$ test was chosen to be used in the study since the GEV distribution was suspected to be applied to the region. Given the GEV distri-bution with $L$-moments parameter estimation and unit mean, so the quantiles are as in (33) that $\kappa$ is defined using (22). The regional estimate based on at-site data and statistical test are calculated by using (34) and (35). In (33), (34), and (35), $\hat{\tau}_2$ and $\hat{\tau}_3$ are the at-site $L$-$Cv$ and $L$-$Cs$ estimates, $\Gamma$ (.) is the gamma function, $m$ is the number of sites, $n$ is the length of data, and $R$ denoted as regional.

$$\hat{x}_{10} = \begin{cases} 1 + \dfrac{\hat{\tau}_2}{1 - 2^{-\hat{\kappa}}}[1 - \dfrac{0.10536^{\hat{\kappa}}}{\Gamma(1-\hat{\kappa})}] & \kappa \ne 0 \\ 1 + 2.4139\hat{\tau}_2 & \kappa = 0 \end{cases} \quad (33)$$

$$\hat{x}_{10}^R = \sum_{i=1}^{m} n_i\hat{x}_{10}^i / \sum n_i \quad (34)$$

$$\chi_{L-M^2} = \sum_{i=1}^{m} (\hat{x}_{10}^i - \hat{x}_{10}^R)^2 / \text{var}(\hat{x}_{10}^i) \quad (35)$$

The denominator, var($\hat{x}_{10}^i$) can be estimated by its asymptotic value as in (33) (Lu and Stedinger, 1992) and $\alpha$ is calculating using (23). The test statistics (35) has approximately a $Chi$-$square$ test with ($m$-$1$) degree of freedom.

$$\text{var}(\hat{x}_{10}^i) = (\alpha^2/n)\{\exp[A]\} \text{ , where} \quad (36)$$

$$A = [a_0 + a_1\exp(-\kappa) + a_2\kappa^2 + a_3\kappa^3] \quad (37)$$

There are some methods can be used in parameter and quantile estimates in regional frequency analysis and one of them is a at-site method (Durrans and Kirby 2004)). This method uses all data in the homogeneous region and the data is to be treated as a single sample.

## RESULTS AND DISCUSSION

### Identification of Homogeneous Region

In regional rainfall frequency analysis, it is difficult step to identify the homogenous region. Based on physical data, the study case, Johore Bahru, is in one degree both for latitude and longitude. The annual mean of rainfall is not quite different in their magnitudes that the higher values are found in the inland region. Two physical variables, the altitude and the distance from the sea, are to be considered in partitioning the region. The region is influenced by body of water, the Strait of Malacca on the South-West (SW) direction and the South China Sea on the East. The rainfall study for Klang Valley as part of the State of Selangor was also found that the high magnitude of rainfall occurred in the inland region (Mohammad Kasim and Way, 1995). Similar study, it was found that the whole region of the

State Selangor was also can be treated as one homogeneous region (Masimin, 2006)

Based on above information, the area of Johore Bahru was expected to be a one homogeneous region and directed to do a sta-tistical test. There were 5 stations distributed in Johore Bahru that their samples were used to test regional homogeneity as a whole region (*WR*). The *X-10* test using (33) to (37) was used in the study and the results are presented in Table 3 showing the values of statistical *X-10* test were less than Chi-square value. It meant than the data sets for each duration pattern were passed the statistical test and the region was concluded to be a homogeneous region. The selection of the distributions will defined in the next section.

## Selection of Distribution

The GEV and GLOG distributions were an appropriate selection of probability distribution based on previous study by Daud et. al. (2002) and Masimin & Harun (2006). The plotting diagram for statistical regional parameters based on *L*-moment for eight data sets and two types of distribution can be examined on Fig.5. Based on Fig.5, three data set for shorter duration pattern (15-, 30-, and 60-minute) follow the GLOG distribution, while the others five follow the GEV distribution as presented in Table 3.

Table 3: The results for *X-10* test

| Location | Duration (min) | *X-10* Test | $\chi^2_{0.9;df=4}$ |
|---|---|---|---|
| | 15 | 2.66 | |
| | 30 | 4.78 | |
| | 60 | 0.63 | |
| Whole | 120 | 1.08 | 7.78 |
| Region | 240 | 0.53 | |
| (*WR*) | 480 | 1.54 | |
| | 960 | 1.07 | |
| | 1440 | 0.70 | |

## Parameter and Quantile Estimates

Upon the known of the probability distributions, the estimation of parameters and quantiles can be determined. In the study, frequency analysis was using at-site data that all data of the region was to be treated as in one sample. The parameter estimation and quantile estimation were only applied for GEV distribution and they were determined by using (20) to (24). The values of parameter estimates for the GEV distribution were as in Table 4 and these parameters were used to calculate their quantile estimates as in Fig.6 of the plotting data.
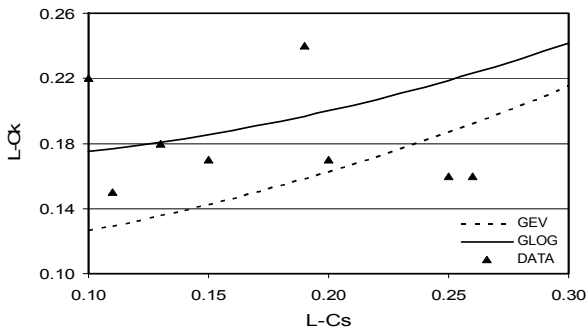
Fig.5. Plotting diagram for *L*-moments parameters with GLOG and GEV distribution

Table 4: Statistical parameters based on *L*-moments

| Duration (min) | Whole Region (*WR*) | | | |
|---|---|---|---|---|
| | *L*-Cv | *L*-Cs | *L*-Ck | Distribution |
| 15 | 0.21 | 0.19 | 0.24 | GLOG |
| 30 | 0.16 | 0.10 | 0.22 | GLOG |
| 60 | 0.16 | 0.13 | 0.18 | GLOG |
| 120 | 0.17 | 0.11 | 0.15 | GEV |
| 240 | 0.18 | 0.15 | 0.17 | GEV |
| 480 | 0.20 | 0.20 | 0.17 | GEV |
| 960 | 0.23 | 0.25 | 0.16 | GEV |
| 1440 | 0.23 | 0.26 | 0.16 | GEV |

To generalize the results in constructing IDF's family, the regretssion analysis in performing the relationship between duration and intensity was done and the power relationship was the appropriate one as in (38) where *a* and *b* were parameters to be defined.

$$I_T = at^b \qquad (38)$$

In (38), $I_T$ = rainfall intensity for the T-year return period (mm/hour); *t* = duration equal to the time concentration (minutes); *a* and *b* = parameters. The results of regression analysis were presented in Table 5 showing the values of *a*, *b*, and $R^2$ for every proposed return period. Looking at the value of $R^2$ that mostly about 0.99, it meant that the power relationship for IDF's family was very rational. The final result of IDF's family was presented in Fig.7.

Table 4. Parameter estimates based on the GEV distribution

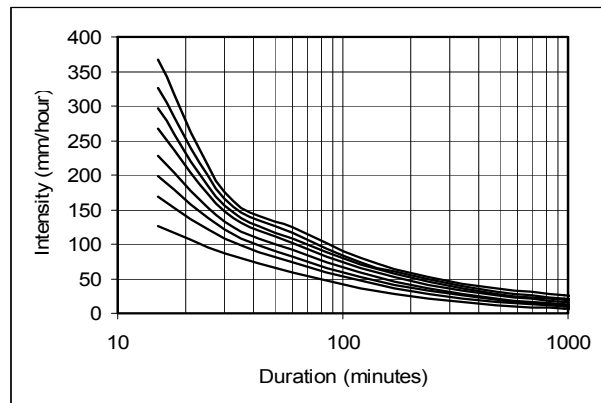| Duration (min) | Whole region (*WR*) | | | |
|---|---|---|---|---|
| | $\hat{\alpha}$ | $\hat{u}$ | $t_3$ | Distribution |
| 15 | 9.33 | 27.94 | 0.19 | GEV |
| 30 | 10.73 | 39.58 | 0.10 | GEV |
| 60 | 13.88 | 53.45 | 0.13 | GEV |
| 120 | 20.03 | 65.55 | 0.11 | GEV |
| 240 | 23.68 | 29.41 | 0.15 | GEV |
| 480 | 27.95 | 86.05 | 0.20 | GEV |
| 960 | 34.42 | 95.55 | 0.25 | GEV |
| 1440 | 37.47 | 105.47 | 0.26 | GEV |

Fig.6. Plotting data for the intensity quantile estimates

Fig.7 was the IDF's curves family that developed based on 5-minute incremental rainfall data; it was needed to compare the cur-ve with those of Mononobe method as in (1). By applying the daily (1440-minute) rainfall data for the return period T = 100 years, that was $R^{24}_{100} = 345.10 \ mm$, the IDF-curve based on Mononobe method was defined and the result was compared to those of this study as in Fig.4 especially for T = 100 years. The result of the com-parison was presented in Fig.8 showing the two IDF-curve, that both of them were identical in pattern. When analyzing the magnitude differences, the IDF-curve based on

Mononobe method gave an over estimate value about 10% for the common duration pattern applying in urban drainage projects that was 30 to 100 minutes. It is noted that the rational method is only applied to the area less than 600 acres or 248 ha and for the area greater than those, it is recommended to use hydrograph analysis with the input is a fractional rainfall data. It is a future research topic of the writer.

Table 5: Parameter estimates for power relationship

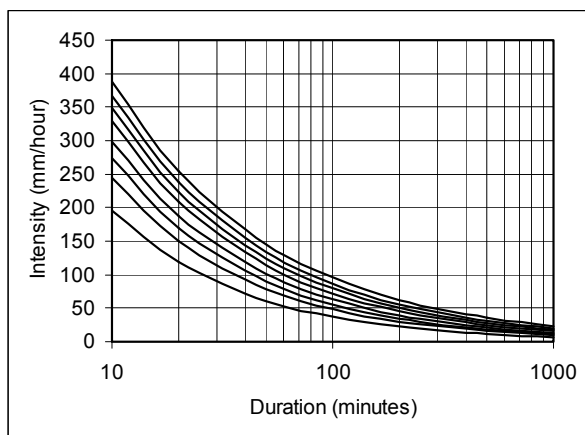| $T$ (year) | $a$ | $b$ | $R^2$ |
|---|---|---|---|
| 2 | 1024.6 | -0.7206 | 0.9932 |
| 5 | 1230.2 | -0.7019 | 0.9965 |
| 10 | 1334.5 | -0.6883 | 0.9980 |
| 20 | 1413.5 | -0.6745 | 0.9988 |
| 50 | 1489.5 | -0.6562 | 0.9986 |
| 100 | 1529.3 | -0.6422 | 0.9974 |
| 200 | 1556.4 | 0.6279 | 0.9950 |
| 500 | 1575.4 | -0.6089 | 0.9900 |



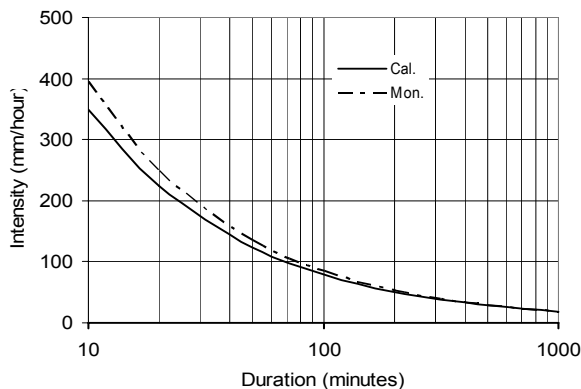Fig.7. The family of intensity duration frequency (IDF-curves)



Fig.8. The comparison for IDF-curve based on calculation and Mononobe method.

## CONCLUSIONS

As the conclusion, the study revealed some of the following information

1. The whole region of the area of Johore Bahru can be treated as one regional homogeneous area that matching to the study by Daud et. al. (2002).
2. Two types of probability distributions, GEV and GLOG distributions, were applied to the all duration pattern rainfall data in Johore Bahru.

3. Looking at the statistical $L$-moments analysis, the higher $L$-Ck, the sample tends to follow the GLOG distribution and it follows the GEV distribution for the lower $L$-Ck.
4. The terms of quantile estimates that it was intended to use in design storms development, GEV distribution was proposed to be used in analysis.
5. By comparing the IDF-curve proposed by Mononobe method that predicting the intensity based daily rainfall data, both IDF curves were identical with Mononobe method gave the result of 10% over estimate.

The general conclusion of the study of design storms development is that the homogeneous region and the type of probability distribution are identified including their quantile estimates. Since the study is using the rational method with its limitation, it is identified that the further study related to the topic of design storms is a fractional rainfall characteristics as an input for hydrograph analysis that to be applied for greater coverage areas.

## REFERENCES

Adamowski, K., Alila, J. & Pilon, P.J. (1996). "Regional rainfall distribution for Canada." *Atmospheric Research* 42, 75-88.

Adamowski, K. and Bougadis, J. (2003). "Detection of trends in annual extreme rainfall." *Hydrol. Processs.* **17**, 3547-3560.

Berndtsson, R. & Niemczynowics, J. (1986). "Spatial and temporal characteristics of high-intensive rainfall in northern Tunisia." *Journal of Hydrology*, 87, 285-298.

Burn, D.H. (1990). "Evaluation of regional flood frequency analysis with a region of influence approach." *Water Resources Research*, 26(10), 2257-2265.

Chow, V.T. (1964). *Handbook of applied hydrology*, McGraw-Hill Book Co., New York.

Cunnane, C. (1988). "Methods and merits of regional flood frequency analysis." *J. Hydrology*, 100, 269-290.

Dinpashoh, Y., Fakheri-Fard, A., Moghaddam, M., Jahanbakhsh, S. & Mirnia, M. (2004). "Selection of variables for the purpose of regionalization climate using multivariate methods." *J. Hydrology*, 297, 109-123.

Daud, Z.M., Mohd Kassim, A.M., Mohd Desa, M.N., & Nguyen, V.T.V. (2002). "Statistical analysis of at-site extreme rainfall process in Peninsular Malaysia." *Proc. of 4th Int. Conference of FRIEND*, Cape Town, March 2002, IAHS Publication (274), 1-8.

Durrans, S.R. & Kirby, J.T. (2004). "Regionalization of extreme precipitation estimates for the Alabama rainfall atlas." *Journal of Hydrology*, 295,101-107.

Fill, H.D. & Stedinger, J.R. (1995). "Homogeneity tests based upon Gumbel distribution and a critical appraisal of Dalrymple's test." *J. Hydrology*, 166, 81-105.

Greenwood, J.A., Landwehr, J.M., Matalas, N.C., & Wallis, J.R. (1979). "Probability weighted moments: Definition and relation to parameters of several distributions expressable in inverse form." *Water Resources Research*, 15(5), 1049-1054.

Hosking, J.R.M., Wallis, J.R. & Wood, E.F. (1985). "Estimation of the generalized extreme-value distribution by the method of probability-weighted moments." *Technometrics*, 27(3), 251-261.

Hosking, J.R.M. & Wallis, J.R. (1987). "An 'index flood' procedure for regional rainfall frequency analysis." *Eos Trans. Am. Geophys. Un.*, 68, 312.

Hosking, J.R.M. & Wallis, J.R. (1987). "Parameter and quantile estimation for the generalized Pareto distribution." *Technometrics,* August 1987, 29 (3), 339-349.

Hosking, J.R.M, (1990). "L-moments: Analysis and estimation of distributions using liner combinations of order statistics." *J. R. Stat. Society*, 52 (1), 105-124.

Hosking, J.R.M. & Wallis, J.R. (1993). "Some statistics useful in regional frequency analysis." *Water Resources Research*, 29(2), 271-281.

Hosking, J.R.M. (1996). *Fortran routines for use with the method of L-moments*. RESEARCH REPORT, IBM Research Division, Yorktown Heights, NY 10598.

Kottegoga, N.T. & Rosso, R. (1997) *Statistics, probability, and reliability for civil and environmental engineers*, The Mc Graw-Hill Companies, Inc., New York, USA.

Kumar, R. & Chatterjee, C. (2005). "Regional flood frequency analysis using L-moments for North Brahmaputra region of India." *J. Hydrol. Engg.,* 10(1), 1- 7.

Lettenmaier, D.P., Matalas, N.C. &Wallis, J.R. (1987). "Estimation of parameters and quantiles of Wakeby distribution, part 1 and 2." *Water Resources Research,* 15(6), 1361-1379.

Lu, L.H. & Stedinger, J.R. (1992). "Sampling variance of normalized GEV/PWM quantile estimators and a regional homogeneity test." *J. Hydrology*, 138, 223-245.

Lu, L.H. & Stedinger, J.R (1992). "Variance of two and three-parameter GEV/PWM quantile estimators: formulae, confidence intervals, and a comparison." *J. Hydrology*, 138, 247-267.

Masimin & Harun, S. (2006). "Analysis for annual maximum pre-cipitation series for State of Selangor using probability weigh-ted moments." *Proc. of Hydrological Sciences for Managing Water Resources in the Asian Developing World Con-ference*, Guangzhou, P.R. China.

Mohd. Kasim, A.H., & Way, L.K. (1995). *Frequency analysis of annual maximum rainfall for the Klang river basin using GEV distribution.* Faculty of Civil Engineering, Universiti Teknolo-gi Malaysia, Johor Bahru, Malaysia.

Rao, A.R & Hamed, K.H. (1997). "Regional frequency analysis of Wabash river flood data by L-Moments." *J. Hydrol. Engg.,* 2(4), 160-179.

Rao, A.R & Hamed, K.H. (2000). *Flood frequency analysis.* CRC Press, New York, USA.

Rao, A.R. & Srinivas, V.V. (2006). "Regionalization of water-sheds by fuzzy cluster analysis." *J. Hydrology*, 318, 57-79.

Smithers, J.C. & Schulze, R.E. (2001). "A methodology for the estimation of short duration design storms in South Africa using regional approach based on L-moment." *J Hydrology*, 241, 42-52.

Sosrodarsono, S. & Takeda T. (1980). *Hidrologi untuk Pengairan*, Pradnya Paramita, Jakarta.

Sveinsson, O.G.B., Salas, J.D. & Boes, D.C. (2002). "Regional frequency analysis of extreme precipitation in Northeastern Colorado and Fort Collins flood of 1997." *J. Hydrol Engg.*, 7(1), 49-63.

Viessman, W. & Lewis, G.L. (2003). *Introduction to hydrology*, (5th ed) Prentice Hall, New York.