SEASONALITY MEASURES TO FIND CATCHMENT SIMILARITY FOR REGIONAL FLOOD FREQUENCY ANALYSIS OF CATCHMENTS IN SRI LANKA

by

Nimal P.D. Gamage

Abstract

Regional flood frequency analysis is used for the estimation of floods at sites where little or no data are available. It involves the identification of groups (or regions) of hydrologically homogeneous catchments and the application of a regional estimation method in the identified homogeneous region. An understanding of the hydrologic characteristics of a catchment is essential in order to obtain a reliable estimate of the relationship between extreme flood quantiles and the associated return periods.

Catchment similarity is expressed using seasonality measures derived from the mean date of occurrence of the annual maximum flood and its associated dispersion. A regionalization approach is applied within the region of influence (ROI) framework and has the advantage of reserving the use of information derived from flood magnitudes for the examination of the homogeneity of flood regions as opposed to first using this information to form regions. Regionalization technique is applied to a catchment in Sri Lanka and is demonstrated result in the identification of region that is effective for the estimation of extreme quantiles.

1.0 Introduction

When data at a given location are insufficient for a reliable estimation of flood quantiles, recourse must be made to regional frequency analysis. Regional flood frequency analysis makes use of similarities in characteristics of flood at different gauging stations within a region. Records from nearby stations are used to substitute spatial to temporal characteristics. A difficulty associated with flood frequency analysis is the lack of sufficient data. Consequently, it is difficult to identify the parent distribution from which a sample is drawn. Large standard errors of estimate result because of small sample sizes. Estimation of the magnitude of floods greater than the length of the available record represents an extrapolation in which the standard error of estimate increases rapidly with increasing return periods.

The first step in regional frequency analysis is identification of homogeneous regions in order for the results of frequency analysis to be valid. A region, in this context, has come to mean a collection of catchments, not necessarily geographically contiguous, that can be considered to be similar in terms of catchment hydrologic response. The goal of the regionalization process can thus be viewed as the identification of grouping of catchments (regions) that are sufficiently similar to warrant the combination and transfer of extreme flow information for sites within the region. The regions defined should thus be able to be considered homogeneous with respect to extreme flow characteristics.

This Paper describes the regionalization process, in the context of regional flood frequency analysis. This is followed by a description of seasonality measures that have been used to characterize the timing and regularity of flooding events for catchments. Seasonality measures are subsequently suggested as an appropriate basis for characterizing the similarity of the flooding response of catchments. A regionalization strategy is then developed and demonstrated through application to a collection of catchments in Sri Lanka.

2.0 Regionalization

Traditionally, the delineation of regions relied on geographic, political, administrative, or physiographic boundaries. The resulting regions were assumed to be homogeneous in terms of hydrologic response, but this cannot in general be guaranteed, particularly if the spatial variability of the physiographic and hydrologic characteristics is large. The importance of regional homogeneity has been demonstrated by, among others, Hosking et al. (1985), and Lettenmaier et al. (1987).

Eng. (Dr.) Nimal P D Gamage, BSc (Eng.) (Hons), MSc (Hons), PhD, AMIE(SL), Postdoctoral Research Fellow, Department of Environmental Science and Human Engineering, Saitama University, Japan.

Hydrologists have often used a regional regression model to estimate flow characteristics at ungauged sites based on catchment characteristics. Many studies use residuals from an overall regression relationship for an entire area, to create geographically contiguous regions based on the sign and magnitude of the residuals (Wandle, 1977). However, deriving regions using this method requires a large amount of subjective input.

A novel regionalization approach involves dispensing with fixed regions. This approach allows each site to have a potentially unique set of stations which constitutes the region for that site. Thus, it is possible for two neighbouring sites to have completely different sets of stations that represent the region for each site. This methodology, which involves the transfer of extreme flow information from similar stations to the site of interest, was first suggested by Acreman and Wiltshire (1987) and Acreman (1987). This approach requires the choice of a threshold value that functions as a cut-off point for the dissimilarity measure. Regionalization approaches, such as cluster analysis or the ROI approach, require the selection of variables that are used to define the pair-wise similarity (or dissimilarity) for the catchments. There are two general types of variables used to define similarity; physiographic catchment characteristics (such as the drainage area, catchment slope, etc.) and flood statistics (such as the L-moment ratios, or other statistical measures, estimated from the available flood series). Difficulties can occur when using physiographic catchment characteristics as similarity variables since similarity in physiographic catchment characteristic space does not necessarily imply similarity in hydrologic response. This discrepancy may arise from the complex interaction of physiographic parameters that are responsible for generating a hydrologic response from a given set of meteorological inputs (Zrinji and Burn, 1994). The same difficulty does not generally occur when flood statistics are used as the similarity variables. However, the problem with this latter option is that the flood statistics are then often used both to form the regions and to subsequently evaluate the homogeneity of the collection of catchments in the region. This often results in regions that are homogeneous but not necessarily effective for regional flood frequency analysis. In addition, the use of flood statistics as the basis for a similarity measure precludes direct consideration of ungauged catchments.

3.0 Seasonality Measures

The timing and regularity of flood events can be used as a measure of similarity in catchment hydrologic response. Similarities in the timing and the seasonality of the flood response for catchments might be expected to result from similarities in influential physiographic and meteorological characteristics. Such catchments can thus be considered as potential candidates for membership in the same region for regional frequency analysis. Although it would be possible to augment the seasonality measures with physiographic and meteorological characteristics, this is not done herein since it is often difficult to attain reliable estimates for such attributes. The use of additional measures of similarity could be expected to enhance the results from the catchment grouping process.

Directional statistics (Mardia, 1972) form the basis for defining similarity measures using the timing of flood events. Following Bayliss and Jones, (1993), we can define the date of occurrence of a flood as directional statistics by converting the Julian date of the flood occurrence for flood event 'i' to an angular value using:

$$\theta_i = (JulianDate)_i \left(\frac{2\pi}{365} \right) \tag{1}$$

where θ_i = angular value (in radian) for the flood event 'i'. Each flood date can now be interpreted as a vector with a unit magnitude and a direction given by θ_i. If we have a sample of 'n' flood events (representing 'n' years of extrean flow data), then the 'x' and 'y' coordinates of the mean flood date can be determined from:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} \cos(\theta_i) \tag{2}$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} \sin(\theta_i) \tag{3}$$

where $\bar{x}$ and $\bar{y}$ represent the x and y coordinates of the mean flood date and lie within, or on, the unit circle. The mean direction of the flood dates is then obtained from:

$$MD = \bar{\theta} \frac{365}{2\pi} \tag{4}$$

The mean direction, θ can be converted back to a day of the year using:

$$MD = \bar{\theta} \frac{365}{2\pi} \tag{5}$$

The variable MD thus represents a measure of the average time of occurrence of flood events for a given catchment. In addition to the mean date of flood occurrence, it is also possible to determine a measure of the variability of the *'n'* flood occurrence about this mean date. This can be determined by defining the mean resultant as:

$$\bar{r} = \sqrt{\bar{x}^2 + \bar{y}^2} \qquad (6)$$

where ' $\bar{r}$ ' provides a measure of the spread of the data. Values of close to unity indicate a catchment with a strongly seasonal flood response. Values close to zero indicate that there is a great deal of variability in the date of occurrence of flood events for the catchment. It is to be expected that catchments with similar values for *MD* and $\bar{r}$ might also exhibit similarities in other important hydrologic characteristics.

The seasonilty measure presented above can now be employed in the definition of the dissimilarity between catchments. This requires a single numerical value that will be used to define the seperation of two catchments in seasonility space. An appropriate measure can be obtained using the Euclidean distance between catchment in the space of the '*x*'- and '*y*-' coordinates of the mean flood date for the catchment. The distance measure is thus defined as:

$$d_s^{ij} = \left[(\bar{x}_i - \bar{x}_j)^2 + (\bar{y}_i - \bar{y}_j)^2 \right]^{1/2} \qquad (7)$$

where d^{ij}_s is the dissimilarity between catchments *i* and *j*, and '$\bar{x}$' and '$\bar{y}$' are the '*x*-' and '*y*-' coordinates of the mean flood date for catchment *i*. The value of 'd^{ij}_s' is invarient to the choice of the origin for the seasionality measure (i.e. the definition of the water year). This is one of the main benefits of using circular statistics to define the seasonality of the flood events (Mardia, 1972). The measure defined in Equation 7 gives the dissimilarity between two catchments and thus small values for 'd^{ij}_s' indicate that the corrosponding catchments exhibit similar seasonality in flood response.

4.0 Catchment Grouping Process

4.1 Regionalization

The region of influence (ROI) approach to regionalization is employed as the basis for the proposed catchment grouping process. In the region of influence approach, a potentially unique collection of catchments is defined as forming the ROI for each catchment. A ROI for a catchment can be viewed as a site-specific region consisting of a collection of gauged catchments that are useful for the transfer of extreme flow information for the estimation of extreme flow quantiles at the site of interest. There are several desired characteristics for a collection of catchments forming the ROI for a given site. The collection of catchments should be hydrologically homogeneous so that the transfer of extreme flow information is from sites that are hydrologically similar to the site of interest. There should be comparatively few 'outlier' sites within the homogeneous ROI as these could unduly influence the estimation of extreme flow quantiles. Finally, the site of interest should, ideally, be close to the 'centre' of the collection of catchments as measured in an appropriate similarity space. This latter requirement is to ensure that the site of interest is hydrologically representative of the sites in its region.

In defining an ROI, a threshold value is generally selected to determine a cut-off point for including stations in the ROI. A number of strategies are available for selecting the threshold value (Burn, 1990, Zrinji and Burn, 1994). The approach taken herein was to establish a target membership of 25 stations for each ROI. A target of 25 stations was selected based on results from Hosking and Wallis, (1996), demonstrating diminishing returns associated with increasing numbers of stations in a region. A larger target value might have been selected if the primary interest was in the extreme tails of the distribution. The resulting collection of catchments is then evaluated using a homogeneity test, described below, and if the ROI cannot be considered to be homogeneous, the membership in the ROI is revised. Revisions to the ROI membership are accomplished by sequentially deleting the catchment that is furthest from the catchment of interest and then re-evaluating the homogeneity of the revised ROI. This process is repeated until the ROI can be considered to be homogeneous.

4.2 Homogeneity Test

The homogeneity test used in this study was proposed by Hosking and Wallis, (1993), and is based on various orders of sample L-moments ratios calculated from the magnitudes of the peak flows. The homogeneity test is based on the idea that in a homogeneous collection of catchments, all catchments should have the same population L-moments. This is, however, not true for the sample L-moments due to the sampling variation. The question arises as to whether differences in sample L-moments indicate different populations or arise from sampling error. Therefore, the level of sampling, or random. error has to be determined. Simulations can be used to establish an acceptable variability level. The homogeneity test has three different levels that are evaluated by the variation of L-moments of different order. In this study, the homogeneity test based on the LC_V is used. More information about L-moments are given in Gamage, (2001). The test is based on a weighted variance of LC_V, *V*, such that the statistics calculated is:

$$V = \frac{\sum_{i=1}^{N} n_i \left(\tau_{2(i)} - \bar{\tau}_2 \right)^2}{\sum_{i=1}^{N} n_i} \qquad (08)$$

where 'n_i' = record length at site 'i;' N = number of sites in the region. It is necessary to determine the mean and the standard deviation of V to calculate the homogeneity measure. This can be done through simulation where a realization of the statistic, labled V_k, is calculated from the k^{th} simulation. The mean and standard deviation of the simulated population of the V_k values are defined as μv and σv respectively. The homogeneity measure is then defined as:

$$H = \frac{V - \mu_v}{\sigma_v} \qquad (09)$$

Acording to Hosking and Wallis, (1993), a region can be declared homogeneous if the value of H is less than 1, possibly homogeneous if the value of H is greater than or equal to 1 but less than 2, and definitively heterogeneous if the value of H is greater than or equal to 2.

5.0 Application

5.1 Description of Data Set

The use of seasonality measures for regional flood frequency analysis is illustrated using a collection of 68 catchments located in Sri Lanka. The data of the catchments were obtained from Master Plan for the Electricity Supply of Sri Lanka, (1987). A total of 127 catchments of Sri Lanka are listed in the Master Plan for the Electricity Supply of Sri Lanka, (1987). Out of this set of catchments, 68 passed the screening criteria of having at least 10 years of observation. It should be noted that the record length threshold selected (10 years) is not intended to signify that reliable at-site flood frequency analysis could be conducted for catchments having a record length as short as the threshold value.

It is decided to use the annual maximum flood series for the analysis of this study as the extraction of the partial duration series of floods is less straightforward because of occasional bunching of flood peaks. The possibility that such peak magnitudes are not statistically independent has led to a certain amount of unease about the validity of statistical methods used with this model.

Locations of the 68 sites selected after screening are shown in Figure 1. The selected data set contains a total of 1564 stream flow observations of the period 1940 to 1985. The range of record lengths at the gauging stations are from 10 to 42 years of annual maximum daily flow values with a mean record length of 23 years and a median length of 22.5 years. The catchment drainage areas range from 53 km^2 to 9606 km^2 with an average value of 1005 km^2 and a median value of 415 km^2. Other information of the selected sites is given in Table 1.

Figure 2 displays the results from calculating the seasonality statistics for the catchments wherein each catchment is plotted in the space defined by $\bar{x}$ and $\bar{y}$. For calculating the seasonality statistics, the time of occurrence of a flood event is taken as the day on which the peak flow occurs during the event with the largest peak flow magnitude in each year. Figure 2 reveals that the flooding regime for the collection of catchments exhibits a high degree of seasonality in that there are two comparatively small ranges in the values for mean date of occurrence of floods. The mean date of flood occurrence of the two ranges fall between Day 316 (November 11) to Day 22 (January 22) and Day 165 (June 14) to Day 278 (October 05). The value of '$\bar{r}$' range from 0.193 to 0.94 with an average value of 0.578 and a median value of 0.525. The magnitude of $\bar{r}$ for a catchment can be determined in Figure 2 from the closeness of the plotted point to the edge of the unit circle with points on the unit circle corresponding to $\bar{r}$ =1.

It should be noted that during the period of the available data set, discharge of some of the catchments were regulated by upstream reservoirs or by a large number of small tanks (Master Plan for the Electricity Supply of Sri Lanka, 1987). In the selected data set only 46 catchments were not regulated by upstream reservoirs or by small tanks and indicated in Table 1. The discordancy statistics of these catchments were analyzed by Gamage, (2001).

Table 1 Calculated Seasonality Measures of the selected sites

Site Number	Name	River	Elev. (m)	Area (km^2)	Num. of Records	Lat (N) (Deg.)	Lon (E) (Deg.)	r - bar	MD (Deg.)
1	Glencourse	Kelani Ganga	18	1463	36	6.9750	80.1808	0.425	190.5
2	Metiyadola	Kelani Ganga	20	606	34	7.0261	80.2739	0.405	215.1
3*	Algoda Bridge	Sitawaka Ganga	26	345	15	6.9486	80.2611	0.490	180.4
4*	Deraniyagala	Sitawaka Ganga	82	152	28	6.9250	80.3375	0.514	185.4
5*	Kitulgala	Kelani Ganga	56	388	38	6.9917	80.4125	0.609	205.7
6*	Mousakelle	Maskeli Oya	1158	122	18	6.8375	80.5500	0.439	181.6
7*	Imbulana	Gurugoda Oya	26	329	26	7.0631	80.2611	0.391	203.4
8	Laxapana	Maskeli Oya	854	168	26	6.8875	80.5111	0.480	238.9
9*	Holambuwa	Gurugoda Oya	53	155	15	7.1931	80.2625	0.363	189.4
10	Hanwella	Kelani Ganga	16	1782	10	6.9097	80.0792	0.411	270.8
11*	Putupaula	Kalu Ganga	2	2598	37	6.6111	80.0653	0.401	211.3
12*	Millakanda	Kuda Ganga	17	769	28	6.6236	80.1736	0.262	228.2
13*	Malwala	Kalu Ganga	18	329	24	6.6875	80.4233	0.511	187.6
14*	Nambapana	Kalu Ganga	13	629	22	6.6864	80.4181	0.497	207.5
15*	Ellagawa	Kalu Ganga	4	1393	26	6.7311	80.2167	0.456	195.4
16*	Dela	Wey Ganga	29	220	27	6.6222	80.4528	0.265	249.3
17*	Kekulegama	Kukule Ganga	200	334	10	6.5556	80.3417	0.460	202.6
18*	Agaliya	Gin Ganga	10	696	38	6.1875	80.1958	0.350	232.7
19*	Jesmin Dam	Gin Ganga	27	361	10	6.3444	80.3333	0.318	208.2
20*	Bopagoda	Nilwala Ganga	18	442	42	6.1556	80.4847	0.273	215.9
21*	Bingimahara	Nilwala Ganga	26	333	12	6.2111	80.4778	0.350	184.0
22*	Julampitiya	Urubokka Oya	62	141	14	6.1944	80.7444	0.636	335.6
23*	Embilipitiya	Walawe Ganga	31	1580	22	6.3444	80.8986	0.358	355.2
24*	Samanalawewa	Walawe Ganga	364	353	16	6.6750	80.8014	0.193	242.8
25	Lunugamwehera	Kirindi Oya	35	913	25	6.3528	81.2194	0.525	356.0
26*	Wellawaya	Kirindi Oya	154	160	11	6.7319	81.1069	0.372	22.1
27*	Kuda Oya	Kuda Oya	81	291	17	6.5250	81.1236	0.561	352.9
28	Kataragama	Menik Ganga	34	787	34	6.4194	81.3292	0.652	343.8
29*	Nakkala	Kumbukkan Oya	160	216	10	6.8917	81.2958	0.677	354.3
30*	Siyambalanduwa	Heda Oya	55	295	20	6.9056	81.5444	0.891	6.5
31	Thottama	Pannela Oya	34	95	14	7.1083	81.6903	0.924	13.2
32*	Periya Aru	Magalwatuwan	36	119	32	7.5000	81.4889	0.899	6.5
33*	Nilobe	Rumbukkan Oya	44	159	24	7.5111	81.3778	0.908	7.6
34*	Maha Oya	Maha Oya	39	300	11	7.5333	81.3583	0.940	12.1
35	Welikanda	Maduru Oya	29	1062	29	7.9361	81.2500	0.914	4.8
36	Manampitiya	Mahaweli Ganga	32	7418	33	7.9111	81.0861	0.790	361.2
37	Galoya Junction	Gal Oya	84	199	12	8.1403	80.8528	0.526	347.1
38	Angamedilla	Amban Ganga	67	1363	20	7.8500	80.9028	0.913	362.3
39*	Peradeniya	Mahaweli Ganga	463	1167	39	7.2583	80.5903	0.558	233.7
40*	Gurudeniya	Mahaweli Ganga	425	1418	33	7.2750	80.6750	0.488	223.6
41*	Nalanda	Nalanda Oya	243	126	10	7.6708	80.6375	0.802	342.4
42	Weragantota	Mahaweli Ganga	76	4092	39	7.3167	80.9861	0.623	360.4
43*	Elehera	Amban Ganga	133	774	38	7.6792	80.7569	0.861	361.3
44*	Morape	Kotmale Oya	640	555	32	7.5111	80.6222	0.389	246.0
45*	Teldeniya	Hulu Ganga	411	160	23	7.2944	80.7667	0.597	362.8
46*	Randenigala	Mahaweli Ganga	141	2365	25	7.2028	80.9361	0.279	329.9
47*	Kandaketiya	Badulu Oya	126	387	17	7.1750	81.0056	0.743	362.3
48*	Watawala	Mahaweli Ganga	829	65	18	6.9472	80.5361	0.649	224.9
49*	Talawakelle	Kotmale Oya	1200	297	28	6.9403	80.6625	0.474	219.2
50*	Talawakanda	uma Oya	575	520	12	7.0083	80.9736	0.562	347.9
51	Allai Kantali	Mahaweli Ganga	8	9606	16	8.3167	81.1667	0.886	363.8
52*	Holbrook	Agra Oya	1346	121	18	6.8806	80.6944	0.625	234.1
53*	Welimada	Uma Oya	1070	179	17	6.9042	80.9083	0.717	358.3
54*	Wellewela	Kalu Ganga	140	194	12	7.6069	80.8333	0.916	14.3
55*	Moragahamulla	Galmal Oya	427	73	16	7.2825	80.8072	0.716	347.3
56	Pangurugaswewa	Yan Oya	21	1311	33	8.7486	80.8792	0.864	356.8
57	Horowpotana	Yan Oya	44	942	31	8.5767	80.8783	0.826	352.5
58	Wahalkada	Hamillawa Oya	31	91	18	8.7267	80.8514	0.865	347.4
59	Kapachchi	Malwathu Oya	36	2117	35	8.5986	80.2722	0.773	351.9
60	Tekkam	Aruvi Aru	18	3071	11	8.7486	80.1833	0.712	359.6
61	Nochchiyagama	kala Oya	24	1948	24	8.1972	80.0972	0.758	350.1
62	Mahauswewa	Mi oya	35	595	29	7.9639	80.0689	0.502	345.3
63	Tabbowa	Mi oya	10	1078	17	8.0472	79.9181	0.429	332.5
64	Chilaw	Deduru Oya	1	2611	19	7.6000	79.8161	0.747	316.4
65*	Alawwa	Maha Oya	49	804	23	7.2917	80.2403	0.332	251.9
66*	Badalgama	Maha Oya	12	1360	30	7.3028	79.9806	0.375	277.7
67*	Giriulla	Maha Oya	27	1191	21	7.3250	80.1153	0.438	274.5
68*	Karasnagala	Attanagalu Oya	28	53	14	7.1125	80.1708	0.466	164.9

* No upstream intervention

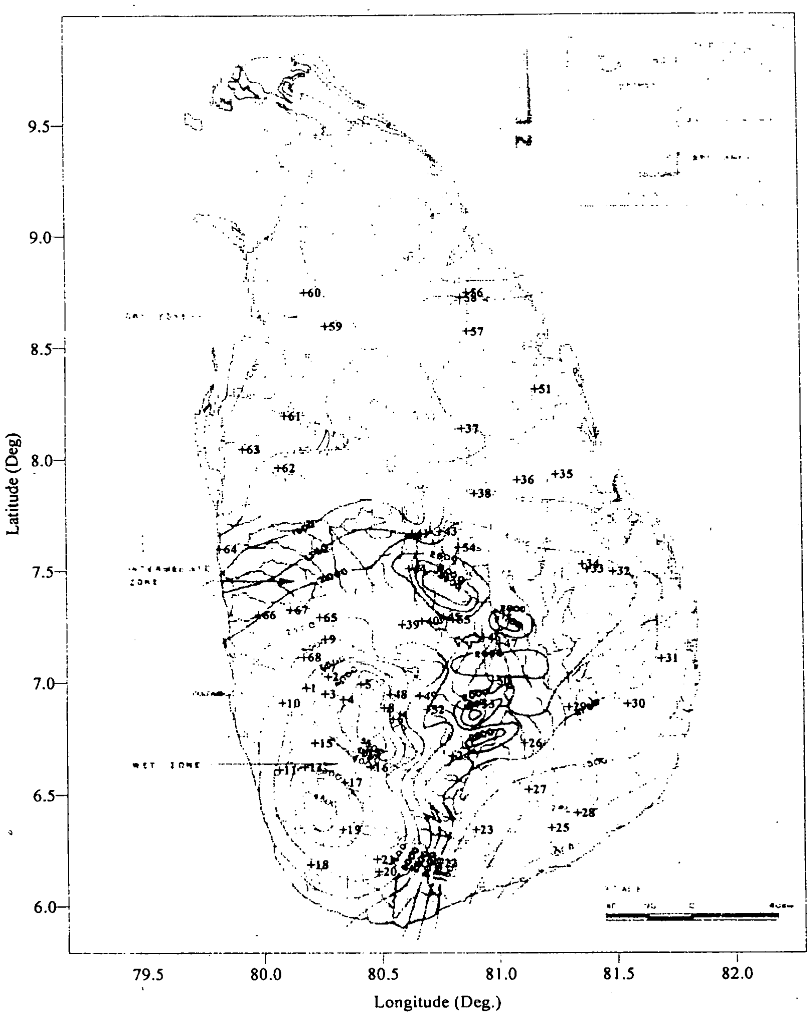

Figure 1 Location of streamflow gauging stations (Base Map Source: Master Plan for the Electricity Supply of Sri Lanka, (1987)).

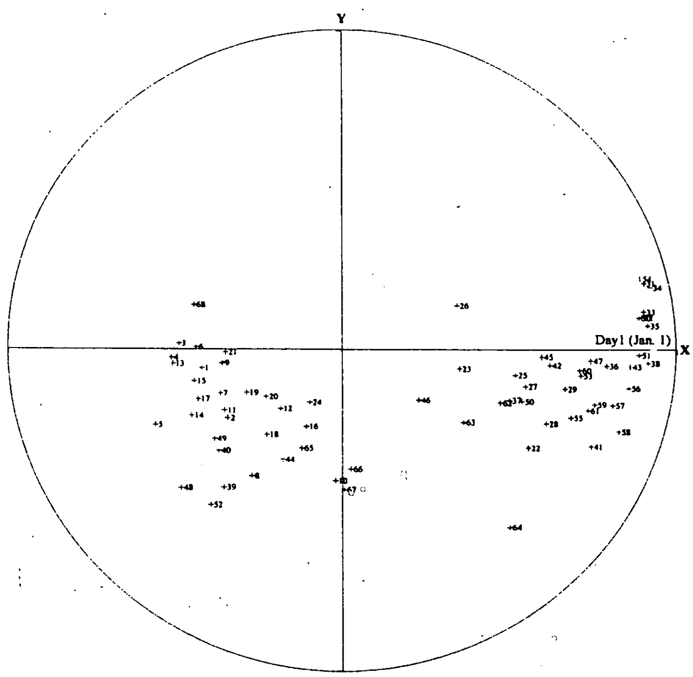

Figure 2 Representation of catchments in terms of Seasonality Measures

5.2 Results

The regionalization approach based on the seasonality measures was demonstrated using the catchment Kitulgala (Site 5). There was no specific reason for selecting this catchment. Variations of the moving averages of the flood record at Kitulgala are shown in Figure 3. Table 2 shows the results for the catchments for which the regionalization approach based on the seasonality measure. The identified 25 catchments in the ROI shows a very high heterogeneity with H=8.2. The membership in the ROI was revised by sequentially deleting the catchments that is furthest from the Kitulgala. This method was able to identify a homogeneous ROI with 6 sites for Kitulgala. The heterogeneity measure of the identified ROI was 0.6 and the sites in the region are listed in Table 2. The location of the sites are shown in Figure 4.

When the annual maximum floods are distributed according to a specified frequency distribution with *cdf F*, the *T*-year event can be calculated as:

$$X_T = F^{-1}\left(1 - \frac{1}{T}\right) \quad (11)$$

Consider a homogeneous region with *N* sites, each site *i* having sample size n_i and observed annual maximum series x_{ij}, $j = 1, ..., n_i$. The annual maximum series from a homogeneous region are identically distributed except for a site specific scaling factor, viz., the index-flood. At each site the annual maximum series is normalised using the index-flood as:

$$z_{ij} = \frac{x_{ij}}{\mu_i} \quad (12)$$

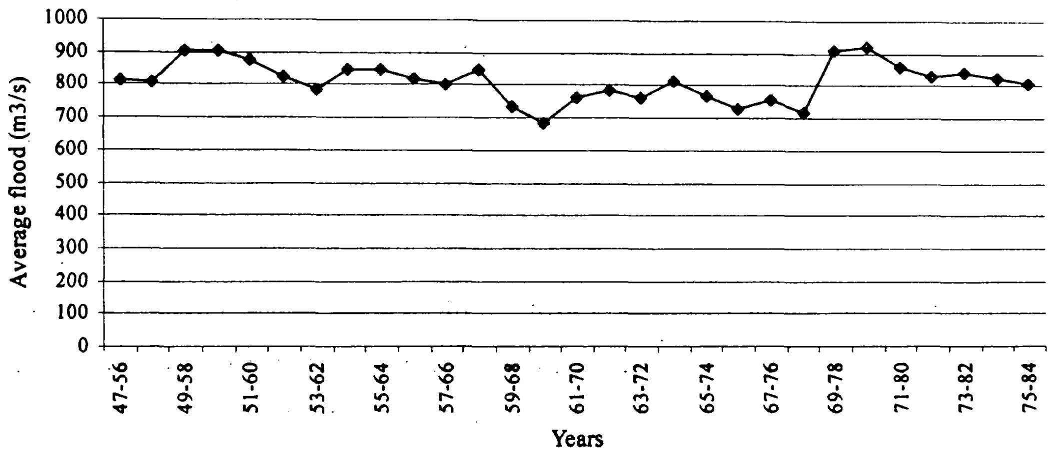

Figure 3 Variation of 10 years moving averages of the annual maximum floods at Kitulgala

Table 2 Heterogeneity measure for the selected regions

Sites in the region of influence of Kitulgala	H
3, 4, 5, 6, 7, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 39, 40, 44, 48, 49, 52, 65, 68	8.2
4, 5, 6, 7, 48, 68	0.6

5.2.1 Estimation of quantiles for the selected homogenous ROI

The *T*-year event X_T is defined as the event exceeded on average once every *T* years and is given as

$$P[X > x_T] = \frac{1}{T} \quad (10)$$

where μ_i is the mean annual flood at site *i*, which is often used as the index-flood. The sample L-moment ratios are estimated at each site and the regional record length weighted average L-moment ratios are calculated as:

$$\hat{\lambda}_r^R = \frac{\sum_{i=1}^{N} n_i \hat{\lambda}_r^{(i)}}{\sum_{i=1}^{N} n_i}, \quad r = 1,2 \quad (13)$$

where $\hat{\lambda}_r^{(i)}$ is the r^{th} order sample L-moment ratio at site *i*, and $\hat{\lambda}_r^R$ is the r^{th} order regional average sample L-moment ratio.

The parameters of a regional frequency distribution can be estimated using the method of L-moments together with the regional average sample L-moment ratios, as shown, by Stedinger et al, (1993) and Hosking and Wallis, (1996). The calculated quantile values of the regional distributions are given in Table 3. Finally, the T-year event at site i can be estimated as

$$\hat{X}_{T,i} = \hat{\mu}_i \hat{z}_T \quad (14)$$

where $\hat{\mu}_i$ is the mean annual flood at site *i*, and $\hat{z}_T$ is the regional growth curve. The regional growth curve is the (1 - 1/T)-quantile of the regional distribution of the normalised annual maximum series as defined through Equation (12).

The test described above has been applied to the selected homogeneous region. For the region the data was tested against the GLO, GEV, GNO, P3 and the GPA distributions, and the results are shown in Table 4. This method identifies the GLO, GEV, GNO, and P3 distributions as valid regional distributions and rejects the GPA distribution. However, the statistic Z^{DIST}

Table 3 Quantiles of the regional distribution selected for Kitulgala

(1-1/T)	Quantiles of the Regional Distribution					
	General Logistic	General Extreme Value	General Normal	Pearson Type III	General Pareto	Wakeby
0.01	0.263	0.302	0.335	0.395	0.428	0.314
0.02	0.314	0.345	0.366	0.408	0.435	0.338
0.05	0.401	0.416	0.425	0.441	0.455	0.402
0.1	0.488	0.49	0.489	0.488	0.49	0.487
0.2	0.606	0.596	0.589	0.574	0.565	0.605
0.5	0.88	0.871	0.867	0.86	0.854	0.862
0.9	1.603	1.643	1.664	1.697	1.723	1.673
0.95	1.969	2.007	2.022	2.036	2.047	2.042
0.99	3.109	3.008	2.936	2.806	2.699	2.947
0.999	5.913	4.972	4.502	3.883	3.429	4.371

5.2.2 Search for a suitable distribution for the selected homogeneous ROI using Goodness-of-fit test

The goodness-of-fit test described by Hosking and Wallis, (1996) is based on a comparison between sample L-kurtosis and population L-kurtosis for different distributions. The test statistic is termed Z^{DIST} and given as

$$Z^{DIST} = \frac{\left(\tau_4^{DIST} - t_4^R + B_4\right)}{\sigma_4} \quad (15)$$

where DIST refers to a candidate distribution. τ_4^{DIST} is the population L-kurtosis of selected distribution, t_4^R is the regional average sample L-kurtosis, B_4 is the bias of regional average sample L-kurtosis, and σ_4 is the standard deviation of regional average sample L-kurtosis.

A four parameter kappa distribution is fitted to the regional average sample L-moment ratios. The kappa distribution is used to simulate 500 regions similar to the observed region, and from these simulations B_4 and σ_4 are estimated. Further details of the procedure are described by Hosking and Wallis, (1996). Hosking and Wallis, (1996), recommend that only distributions for which $|Z^{DIST}| \leq 1.64$ should be considered as suitable regional distributions for the particular region.

suggests that either GNO or GEV would be a better choice, as the statistic Z^{DIST} is very low.

5.2.3 Comparison of distributions using Probability plots

As pointed out by Hosking et al, (1985), comparison of different regional frequency distributions against observed data cannot be used to discriminate between different distributions, as the observed data represents only one of an infinite number of realisations of the 'true' underlying population. However, the probability plots may reveal tendencies such as systematic regional bias in the estimation of the extreme events.

Table 4 The test statistic Z^{DIST} of regional distribution based on L-kurtosis.

Distribution	Z-Value	Rank
General Logistic (GLO)	0.59*	3
General Extreme Value (GEV)	-0.07*	2
General Normal (GNO)	-0.05*	1
Pearson Type III (P3)	-1.23*	4
General Pareto (GPA)	-1.78	5

* Accepted as regional distribution

To assess how well the proposed regional frequency distributions fit to the observed annual maximum series calculated using L-Moments for the selected site are shown in Figures 5 to 9. The empirical exceedance probabilities for the ordered observations *X(i)* were calculated using the median probability plotting position as described by Stedinger et al. (1993), and shown in Equation 16.

$$P\left[X > x_{(i)}\right] = \frac{i - 0.3}{n + 0.4} \qquad (16)$$

where i is the rank of the ordered observation and n is the total number of observations.

6.0 Discussion and Conclusion

Analysis of the seasonality measure clearly indicated the two seasons prevailing in Sri Lanka. However, it should be noted that no unregulated flood data could be found in the season falls between Day 316 (November 11) to Day 22 (January 22). Even though it is possible to identify ROI's of the catchments fall in this season, it is not possible to check the homogeneity of the regions included with those catchments. However, the mean annual floods of the catchments in this seasonality space is highly seasonal, compared to the catchment in the other seasonality space, with $\bar{r}$values of most of the catchments close to unit circle (Figure 2).

Few catchments falling in to the ROI of Kitulgala indicate high degree of heterogeneity between the catchments in Sri Lanka. The identification of a suitable regional distribution of the selected homogeneous ROI with 6 sites was based on a Goodness-of-fit test and evaluated using the probability plots. The goodness-of-fit test statistic, however, suggest that region should preferably use GNO or GEV distributions. The probability plots of the selected site indicate marginal variations of the estimates of all the distributions for smaller return period floods of less than 10 years.

Calculation of quantiles using the seasonality-based ROI can involve the weighting of the information from the stations in the ROI in accordance with the similarity of each station with the site of interest. This was not done in the present work but formulating preferred approaches to this task remains an avenue for further investigation.

The regionalization approach presented herein is directly applicable to sites that are gauged. The analysis of ungauged sites might be achieved by estimating the values for the seasonality statistics at an ungauged site using available physiographic and meteorologic information. Alternatively, gauged sites that are similar to an ungauged site could be identified and seasonality statistics estimated for the ungauged site as a combination of the values at the gauged sites that are similar to the ungauged site. This too remains an avenue for further research.

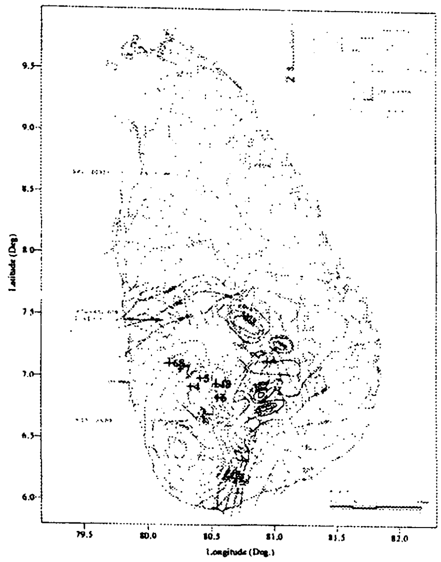

Figure 4 Geographic display of the region of influence for site 5, Kitulgala.

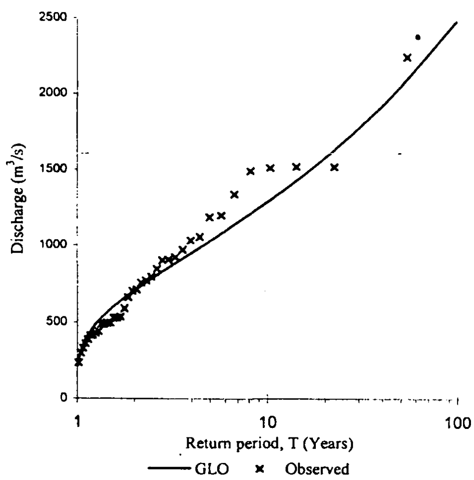

Figure 5 Probability plot for Kitulgala using General Logistic distribution

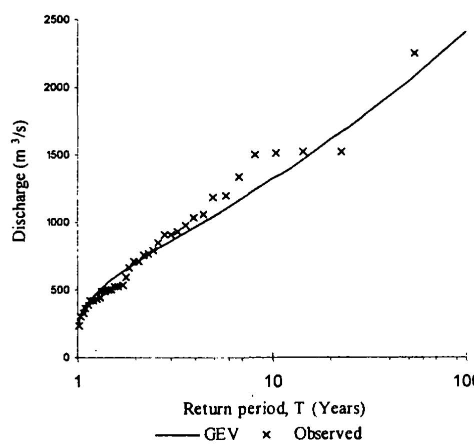

Figure 6 Probability plot for Kitulgala using General Extreme Value distribution

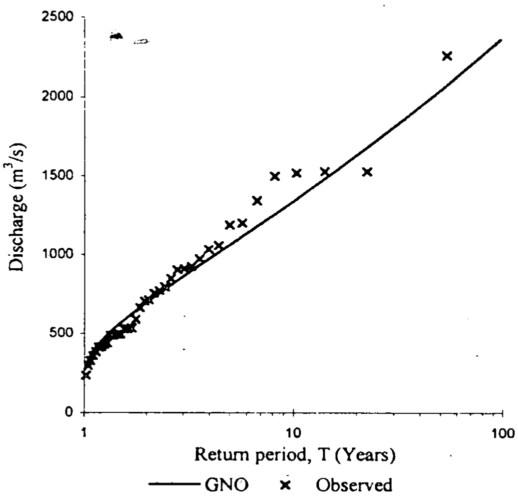

Figure 7 Probability plot for Kitulgala using General Extreme Value distribution

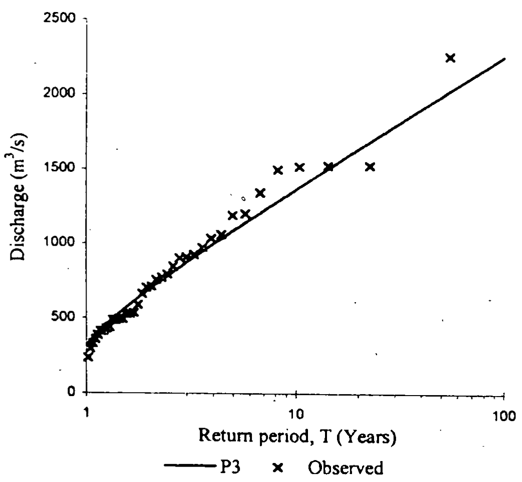

Figure 8 Probability plot for Kitulgala using Pearson Type III distribution

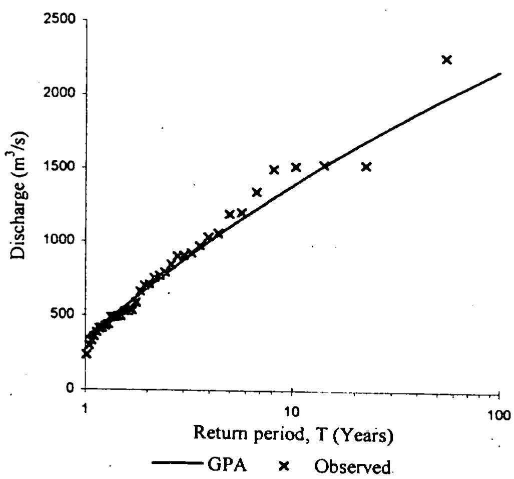

Figure 9 Probability plot for Kitulgala using General Pareto distribution

7.0 REFERENCES

Acreman, M.C.. 1987. Regional Flood Frequency Analysis in the UK: Recent Research-New Ideas. Institute of Hydrology, Wallingford. UK.

Acreman, M.C.. Wiltshire. S.E., 1987. Identification of regions for regional flood frequency analysis. EOS 68 (44), 1262 (abstract).

Baylis.s, A.C., Jones, R.C.. 1993. Peaks-Over-Threshold Flood Database: Summary Statistics and Seasonality. IH Report No, 121. Institute of Hydrology, Wallingford, UK.

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

Gamage, N.P.D., 2001. Analysis of regional homogeneity of catchments in SriLanka by L - Moments. "Engineer", Journal of the Institution of Engineers, Sri Lanka. Vol: xxxiv, No: 2, May 2001. pp: 78 - 87.

Hosking, J.R.M., Wallis, J.R., 1996. Regional Frequency Analysis: An Approach Based on L-moments. Cambridge University Press.

Hosking, J.R.M., Wallis, JR., 1993. Some statistics useful in regional frequency analysis. Water Resour. Res., 29, 271-281.

Hosking, J.R.M., Wallis, J.R., Wood, E.F., 1985. An appraisal of the regional flood frequency procedure in the UK "Flood Studies Report", Hydrol. Sd. J.. 30 (1), 85-109.

Lettenmaier, D.P., Wallis, J.R., Wood, E.F., 1987. Effect of regional heterogeneity on flood frequency estimation, Water Resour. Res., 23 (2), 3 13-323.

Mardia, K.V., 1972. Statistics of Directional Data. Academic Press, New York. NY.

Master Plan for the Electricity Supply of Sri Lanka, 1987: Report Published by Ceylon Electricity Board. Volume S-1, July 1987.

Reed, D.W., 1994. Plans for the Flood Estimation Handbook, Proceedings, MAFF Conference of River and Coastal Engineers. Loughborough, UK. pp. 8.3.1-8.3.8.

Stedinger, J.R., Vogel, R.M., Foufoula-Georgiou, E., 1993. Frequency analysis of extreme events. In: Maidment, DR. (Ed.). Handbook of Hydrology. McGraw-Hill, New York.

Wandle, S.W., 1977. Estimating the magnitude and frequency of flood on natural streams in Massachusetts. US Geological Survey Water Resources Investigations 77-39, 27 pp.

Zrinji. Z., Burn, D.H., 1994. Flood frequency analysis for ungauged sites using a region of influence approach, J. Hydrol,, 153. 1-21