Regional frequency analysis of rainfall extremes in Southern Malawi using the index rainfall and L-moments approaches

Cosmo S. Ngongondo · Chong-Yu Xu · Lena M. Tallaksen ·
Berhanu Alemaw · Tobias Chirwa

Published online: 26 May 2011

© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract Rainfall extremes often result in the occurrence of flood events with associated loss of life and infrastructure in Malawi. However, an understanding of the frequency of occurrence of such extreme events either for design or disaster planning purposes is often limited by data availability at the desired temporal and spatial scales. Regionalisation, which involves “trading time for space” by pooling together observations for stations with similar behavior, is an alternative approach for more accurate determination of extreme events even at ungauged areas or sites with short records. In this study, regional frequency analysis of rainfall extremes in Southern Malawi, large

parts of which are flood prone, was undertaken. Observed 1-, 3-, 5- and 7-day annual maximum rainfall series for the period 1978–2007 at 23 selected rainfall stations in Southern Malawi were analysed. Cluster analysis using scaled at-site characteristics was used to determine homogeneous rainfall regions. L-moments were applied to derive regional index rainfall quantiles. The procedure also validated the three rainfall regions identified through homogeneity and heterogeneity tests based on Monte Carlo simulations with regional average L-moment ratios fitted to the Kappa distribution. Based on assessments of the accuracy of the derived index rainfall quantiles, it was concluded that the performance of this regional approach was satisfactory when validated for sites not included in the sample data. The study provides an estimate of the regional characteristics of rainfall extremes that can be useful in among others flood mitigation and engineering design.

C. S. Ngongondo (✉) · C.-Y. Xu · L. M. Tallaksen
Department of Geosciences, University of Oslo, P.O. Box 1047,
Blindern, 0316 Oslo, Norway
e-mail: cosmon@student.matnat.uio.no;
cngongondo@chanco.unima.mw

C. S. Ngongondo
Department of Geography and Earth Sciences, Chancellor
College, University of Malawi, P.O. Box 280, Zomba, Malawi

B. Alemaw
Department of Geology, University of Botswana, Pr. Bag 0022,
Gaborone, Botswana

C.-Y. Xu
Department of Earth Sciences, Uppsala University, Uppsala,
Sweden

T. Chirwa
Department of Mathematical Sciences, Chancellor College,
University of Malawi, P.O. Box 280, Zomba, Malawi

T. Chirwa
Epidemiology and Biostatistics, School of Public Health,
Faculty of Health Sciences, University of the Witwatersrand,
Johannesburg, South Africa

Keywords Cluster analysis · L-moments · Regional frequency analysis · Index rainfall · Malawi

1 Introduction

The estimation of magnitudes and frequencies of extreme hydrometeorological events such as daily maximum rainfall is central in the design of hydraulic structures, flood plain zoning and economic estimation of flood protection projects (Noto and La Loggia 2009; Sarkar et al. 2009). Often the interest is in the very rare events with return periods (T) of above 50 or 100 years. This mainly owes to their destructive nature to life and infrastructures. However, reliable estimation of such extreme events requires very long station records if single station data are to be used. Availability and quality of such data are often a

challenge in many parts of the world, especially in the data scarce regions of Africa. The Southern Africa region is one such region as it is considered especially vulnerable to and ill-equipped (in terms of adaptation) for extreme events such as rainfall, droughts and flooding. This is due to a number of factors including extensive poverty, famine, disease and political instability (Williams et al. 2009). Regional frequency analysis (RFA) is a commonly used and practical means of providing information at sites with little or no data available (Zhang and Hall 2004).

Various regionalisation techniques have been developed and can be broadly classified into those used for prediction in data scarce areas and those used for RFA (Chen et al. 2006; Durrans and Tomic 1996; Mazvimavi et al. 2004; Sivapalan et al. 2003; Hosking and Wallis 1997). Techniques which have been widely applied in rainfall regionalisation include: linkage analysis (e.g. Jackson 1972); spatial correlation analysis (Gadgil et al. 1993); common factor analysis (e.g. Barring 1988); empirical orthogonal function analysis (e.g. Kulkarni et al. 1992); principal component analysis (PCA) (e.g. Baeriswyl and Rebetez 1996; Singh and Singh 1996); cluster analysis (e.g. Easterling 1989; Venkatesh and Jose 2007); combination of PCA and cluster analysis (e.g. Din-pashoh et al. 2004); L-moments in association with cluster analysis (e.g. Schaefer 1990; Guttman 1993; Wallis et al. 2007; Satyanarayana and Srinivas 2008) and a combination of L-moments and generalised least squares regression (Haddad et al. 2010).

Past rainfall regionalisation studies in Southern Africa include that of Jackson (1972), who used the spatial correlation based simple linkage analysis on rainfall data from 30 stations in Tanzania between 1930 and 1960. In that study, six rainfall homogenous regions were identified. PCA based studies include: Barring (1988) who identified five main rainfall regions in Kenya based on daily rainfall; Van Regenmortel (1995) in which Botswana was categorised into a single rainfall region based on a correlation matrix of 49 daily rainfall stations; and Unganai and Mason (2001) who used summer rainfall characteristics in Zimbabwe to identify two major rainfall regions. Parida and Moalafhi (2008) applied the L-moments approach to analyse annual rainfall series for 11 stations in Botswana for the period 1960–2003. The study established that the whole of Botswana behaved homogeneously and the Generalized Extreme Value (GEV) distribution was accepted as the best model for the data. In South Africa, Smithers and Schulze (2000a, b, 2001, 2003) also applied the L-moments approach in various regional rainfall frequency analysis.

Very few studies on extreme rainfall events for Malawi and specifically our study domain in Southern Malawi, an economically very important region, are documented in the literature. The most notable work on rainfall extremes in Malawi is probably that of Drayton (1980) who analysed

1-day annual maximum rainfall from 38 stations across Malawi using the Gumbel (Extreme Value I) distribution to derive estimates of point daily rainfall with return periods $T = 20$ and 50 years. New et al. (2006), however, reported that there is some evidence of increasing trends in regionally averaged rainfall on extreme precipitation days and in annual 1- and 5-day maximum rainfalls in Southern Africa. This apparent need to provide updated knowledge of rainfall extremes using well recognised regionalisation tools in a data scarce region partly motivated this study. Drayton (1980) also recommended the application of more robust computer based frequency analysis procedures.

The main aims of this study therefore, are (1) to improve the understanding of the regional rainfall characteristics in Southern Malawi as a key factor for regional flood estimation to support flood risk management, and (2) to provide a well designed and verified procedure for operational use for RFA of rainfall extremes in data scarce region of African countries. The aims will be achieved through the following specific objectives (1) to perform a RFA of 1-, 3-, 5- and 7-day annual maximum rainfall series (AMS1, AMS3, AMS5, AMS7 hereafter) in Southern Malawi using the well known L-moments approach; (2) to develop regional index rainfalls for the various annual maximum rainfall series; and (3) to evaluate the accuracy of the regional L-moments approach in estimating design rainfall at sites in the region through uncertainty assessments. To our best knowledge, the RFA approach used in this study, with the most up to date data available, is the first of its kind for rainfall extremes in Malawi.

2 Study area and data set

2.1 Study area

Southern Malawi has a total population of 5.3 million making it the most densely populated region in Malawi (Sajidu et al. 2008). The northern part of the area borders the southern shore of Lake Malawi in Mangochi district (Fig. 1). The Shire river, a tributary of the Zambezi river, is the major river in the area and drains Lake Malawi southward to the Zambezi (Jury and Gwazantini 2002; Jury and Mwafulirwa 2002). The Shire Highlands dominate the topography of the eastern part of the region with major relief features like Zomba and Mulanje mountain massifs. These two mountains rise to above 2,000 m above mean sea level (m a.s.l.) in the Lake Chilwa catchment area of the Chilwa-Phalombe plain. To the south of the region is the lower Shire river area, a low lying flood plain.

The climate of Southern Malawi is tropical wet and dry, commonly known as Savanna. The main rain bearing system is the Inter-Tropical Convergence Zone (ITCZ), where the north easterly monsoon and south easterly trade winds

Fig. 1 Map of Central Africa (*left*) and Malawi (*middle*) showing the study area and stations in Southern Malawi (*right*). The arrow shows the location of Southern Malawi on the map of Malawi

converge. A distinct rainy season is experienced between November and April when over 80% of the annual rainfall occurs. Tropical cyclones originating from the Indian Ocean, frequently occur during the rainy season bringing very intense rainfall over few days. Annual rainfall varies from 700 mm in the low lying areas to 2,500 mm in highlands of Mulanje and Zomba. There is considerable sporadic winter rainfall locally called *chipirone* in the highlands during the period from May to August. The winter rains originate from an influx of cool moist south-eastern winds. Monthly average temperatures are around 10–16°C in the highlands and 21–30°C along the lower Shire valley (British Geological Survey 2004). Figure 1 shows the location of the study area and the geographic distribution of the rainfall stations used in this study.

Southern Malawi is economically very important. Over 95% of Malawi's hydropower is generated along the river Shire towards its confluence with the Zambezi River. Further, the Malawi National Contingency Plan (MNCP 2009) identified flooding triggered by heavy rainfall as the cause of more than 40% of disasters in Malawi since 1940. Recent flooding events as a result of heavy rainfall occurred in 1996/1997, 1998/1999, 2000/2001, and 2002/2003. The flooding often results in the loss of life, crops, property and vital infrastructure. Up to date information about rainfall extremes of the area is therefore crucial for ensuring the security of human lives and properties (Yang et al. 2010a).

2.2 Data availability

The study analysed AMS1, AMS3, AMS5 and AMS7 of rainfall derived from daily readings at 23 rainfall stations in Southern Malawi covering the period 1978–2007. These were sourced from the Malawi Department of Climate Change and Meteorological Services. Table 1 lists the stations including their location and elevation. Although some stations had gaps, none had a record less than 13 years.

3 Methodology

This section presents the regionalisation procedure applied. Initial data screening procedures for serial independence of the at-site data and stationarity of the at-site means are firstly presented. This is followed by a spatial independence check among the stations using Moran's I coefficient. The K-means clustering procedure, for the determination of the possible number of homogeneous regions, is then presented. Finally, a summary of the four steps in RFA using the L-moments approach is presented.

According to Fowler and Kilsby (2003), extreme value theory recommends the peak-over-threshold (POT) approach, which includes all events above some chosen threshold, for rainfall frequency estimation. In this study, however, AMS were chosen mainly to preserve the number of stations and

Table 1 Location and elevation of the rainfall stations in Southern Malawi

Serial	Station	Lat (°S)	Long (°E)	Elevation (m a.s.l.)	Serial	Station	Lat (°S)	Long (°E)	Elevation (m a.s.l.)
1	Balaka	14.92	34.87	625	13	Mimosa	16.08	35.58	652
2	Bvumbwe	15.92	35.07	1,146	14	Monkeybay	14.08	34.92	482
3	Chanco	15.38	35.35	886	15	Mwanza	15.62	34.52	1,260
4	Chichiri	15.80	35.05	1,132	16	Naminjiwa	15.77	35.67	773
5	Chikwawa	16.03	34.78	107	17	Nchalo	16.27	34.92	52
6	Chikweo	14.75	35.67	717	18	Neno	15.4	34.65	899
7	Chileka	15.68	34.97	767	19	Ngabu	16.5	34.95	102
8	Chingale	15.37	35.25	610	20	Ntaja	14.87	35.53	731
9	Liwonde	15.07	35.22	507	21	Satemwa	16.07	35.1	975
10	Makhanga	16.52	35.15	76	22	Thyolo	16.15	35.22	820
11	Makoka	15.52	35.22	1,029	23	ZombaRTC	15.5	35.32	915
12	Mangochi	14.43	35.25	482					

station years available. Some sites had missing values for considerable time periods. POT analysis requires a common period of analysis which would imply a reduction in the available number of stations and station years. Alternatively, missing values would have to be filled in. These conditions were considered challenging to meet due to low spatial correlations among stations in this sparsely gauged region (Ngongondo et al. 2011). The AMS procedures used in this study do not have to satisfy any of the above conditions.

3.1 Data screening

Frequency analysis requires that the at-site data are independent (without serial correlation and trends) and identically distributed (from the same population) (i.i.d.). Serial correlation within a time series reduces the effective sample size of the series compared with independent data (Matalas and Langbein 1962; Tallaksen et al. 2004). Independence was tested using lag-1 to lag-5 autocorrelation coefficients. Trend was determined using the non-parametric Mann–Kendall (MK) test (Mann 1945; Kendall 1975) as recommended by the World Meteorological Organisation (WMO 1988). For stations with data that is serially correlated and having trends, von Storch (1995) recommends pre-whitening of the series.

Another requirement is that the AMS1 at different stations in a homogeneous region should be spatially independent. High spatial cross-correlation between stations gives a lower degree of additional regional information to the site being studied than uncorrelated sites. The magnitude of cross-correlations therefore provides a measure of the amount of independent information contained in the regional data relative to the amount of station years. If not taken into consideration, cross-correlation leads to estimates that are less accurate than they would be if the

samples were spatially independent (Stedinger 1983; Schaefer 1990; Yue and Wang 2002). Moran's I test (Moran 1950) was used to test for spatial independence. The test is based on cross products of deviations from the mean and is computed from:

$$\beta = \frac{N}{S} \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2} \quad (1)$$

where x_i is the observed value at station i , x_j is the observed value at station j , N is the number of stations, and w_{ij} are the elements of the weight and $S = \sum_{i=1}^N \sum_{j=1}^N w_{ij} (i \neq j)$. Two neighbouring stations will have $w_{ij} = 1$ and 0 otherwise. Randomly arranged and uncorrelated values over space have Moran's I equal to its expected value $[-1/(n-1)]$ where n is the total number of locations (Khalili et al. 2007). A significance level of $\alpha = 0.05$ was used in these data screening tests.

3.2 Regional rainfall frequency analysis

The K-means cluster algorithm, a supervised approach, and Ward's hierarchical unsupervised technique were applied to identify homogeneous regions. The two clustering techniques were chosen owing to their successful application in other rainfall regionalisation studies (e.g. McQueen 1967; Hosking and Wallis 1997; Ward 1963; Ramos 2001). The L-moments algorithm (Hosking and Wallis 1997) and the index-flood procedure (Dalrymple 1960) were used for the RFA. L-moments, a linear combination of PWMs (Hosking 1986, 1990), are considered robust over the use of conventional moments in certain aspects. Firstly, they are relatively insensitive to outliers and do not have sample size related bounds as do conventional moments. Secondly, parameter estimations are more reliable and less biased than the conventional method of moment estimates, particularly

for small samples. Thirdly, they are usually computationally more tractable than maximum likelihood estimates and fourthly, estimators of L-moments are virtually unbiased (Hosking 1990). Vogel and Fennessey (1993) further highlight that the use of product moment ratio estimators of the coefficient of variation (C_v), skewness (γ) and kurtosis (k) exhibit substantial bias for samples less than 100, which is very common in hydrology. L-moments, however, are more robust for both large and small samples.

3.2.1 Homogeneous clusters

The number of homogeneous regions and their stations composition were initially defined using the K-means and Ward's clustering methods. As presented by Satyanarayana and Srinivas (2008), a data set is classified by K-means clustering through a certain number of clusters fixed a priori, assumed K clusters. If $Y = \{y_i/i = 1 \dots N\}$ is a set of N feature vectors (rain gauges in this study) in n -dimensional attribute space:

$$\{y_i = [y_{i1}, \dots, y_{ij}, \dots, y_{in}] < n \in \mathbb{R}^n\} \quad (2)$$

where y_{ij} is the value of attribute j in i th feature vector y_i . Each feature vector represents one of the N sites (rain gauges) in the study region. Variables influencing precipitation at a site or their principal components and its geographical location attributes are used. To avoid dominance of feature vectors with large absolute values (e.g. altitude), each feature vector is rescaled as:

$$x_{ij} = \frac{(y_{ij} - \bar{y}_j)}{\sigma_j} \quad \text{for } 1 \leq j \leq n, \quad (3)$$

where x_{ij} are the rescaled values of y_{ij} , σ_j is the standard deviation of attribute j , and $\bar{y}_j$ is the mean value of attribute j over all the N feature vectors. Through an iterative procedure, the K-means algorithm move the feature vectors from one cluster to another to minimize the objective function, F defined as:

$$F = \sum_{k=1}^K \sum_{j=1}^n \sum_{i=1}^{N_k} d^2(x_{ij}^k - x_j^k) \quad (4)$$

where the number of clusters K is set a priori; N_k is the number of feature vectors in cluster k ; x_{ij}^k denotes the rescaled value of attribute in the feature vector i assigned to cluster k ; x_j^k is the mean value of attribute j for cluster k , computed as:

$$x_j^k = \frac{\sum_{i=1}^{N_k} x_{ij}^k}{N_k} \quad (5)$$

Each feature vector i.e. rain gauge, is allocated to a cluster by minimising F in (4) whereby the distance of each feature vector from the center of the cluster (centroid) to

which it belongs is minimized. The number of clusters was determined using the PCA based scree plot. Total loadings across the abscissa show separation (also called a break) in fraction of total variance where the 'most important' components cease and the 'least important' components begin. On the other hand, Ward's clustering approach is unsupervised requiring no a priori setting of the number of clusters. The same attributes used in the K-means approach were used. The method has been widely applied in rainfall cluster analysis (e.g. Muñoz-Díaz and Rodrigo 2004).

The criteria for choosing clustering variables largely depend on the major factors that influence rainfall in an area. Satyanarayana and Srinivas (2008) in India used 15 large-scale atmospheric variables of air temperature, geopotential height, specific humidity, zonal and meridional wind velocities, precipitable water and surface pressure in addition to latitude/longitude location, elevation and mean annual rainfall (MAR). Schaefer (1990) used at-site Mean Annual Precipitation (MAP) to define homogeneous areas in the regionalisation of precipitation annual maxima in Washington State, USA. Schaffer's assumption was that the MAP is numerically descriptive of arid versus semi-arid environments. In this study, the approach by Yang et al. (2010b), which used station latitude/longitude location, elevation and MAP in defining rainfall regions in the Pearl River Basin, China, was adopted. These four variables present a fair balance of commonly used attributes and were readily available for Southern Malawi. The variables were scaled to values between -1 and 1 to avoid bias of those variables with large absolute values. Using the K-means clustering algorithm, five simulations were performed with the number of clusters set between two and six.

Two cluster validity indices, i.e. Hubert's gamma coefficient and the Dunn index (Satyanarayana and Srinivas 2008) were applied to determine optimal partition provided by the K-means clustering algorithm. This was done by pair-wise comparisons of the compositions of the different clusters between two and six.

3.2.2 Discordancy measure D_i

The discordancy measure (D_i) is used to screen data within the identified homogenous region(s) (Hosking and Wallis 1997). Stations in a homogenous region having either gross errors in their data or not belonging to the region, will have $D_i \geq 3$. The case of gross errors in the data requires careful scrutinisation of the data. If gross errors are not detected in the data, but $D_i \geq 3$, it is recommended to relocate such stations to other regions (Hosking and Wallis 1997; Adamowski 2000). Further details of the discordancy measure are described by Hosking and Wallis (1997).

3.2.3 Heterogeneity measure

The heterogeneity measures H_n ($n = 1, 2, 3$) are used to verify whether the proposed sites make up a spatially homogeneous region with the same underlying distribution apart from a site-specific scale factor (Hosking and Wallis 1997). The check is based on observed and simulated dispersion of L-moments for a group of sites under consideration. A region is considered ‘acceptably homogenous’ if $H_n < 1$; ‘possibly heterogeneous’ if $1 < H_n < 2$; and ‘definitely heterogeneous’ if $H_n \geq 2$. A large positive value of H_1 indicates that the observed L-moments are more dispersed than what is consistent with the hypothesis of homogeneity. H_2 indicates whether the at-site and regional estimates are close to each other. A large value of H_2 indicates a large deviation between regional and at-site estimates. H_3 indicates whether the at-site and the regional estimates agree. Large values of H_3 suggest a large deviation between at-site estimates and observed data. However, H_1 is the primary measure for the heterogeneity test as both H_2 and H_3 rarely yield values larger than 2 even for grossly heterogeneous regions (Hosking and Wallis 1997; Yang et al. 2010b). Computation details for the calculation of H_n are given in Hosking and Wallis (1997).

3.2.4 Distribution selection

The best fitted distributions for each homogeneous regions can be identified by several means: visually on the L-moment ratio diagram, a plot of sample L-skewness (τ_3) versus L-kurtosis (τ_4) where the best distribution fits evenly through the cloud of points; the goodness of fit measure $|Z^{\text{Dist}}| \leq 1.64$; and the difference between sample kurtosis and distribution kurtosis ($|\tau_{4\text{sample}} - \tau_{4(\text{DIST})}|$), which should be the minimum. Hosking and Wallis (1997) presented details for the calculation the $|Z^{\text{Dist}}|$ statistics.

Six commonly applied distributions namely the Generalized Logistic (GLO), Generalized Extreme Value (GEV), Generalized Normal (GNO), Generalized Pareto (GPA), Pearson type III (PE3) and Wakeby (WAK) were tested. The five-parameter Wakeby distribution is included in case the choice of the candidate distributions is inconclusive. This normally occurs when the region is mis-specified as being homogeneous. The Wakeby therefore offers the best option for frequency analysis in such cases as it is more robust (Hosking and Wallis 1997).

3.2.5 Derivation of regional rainfall quantiles

Rainfall quantiles of the best frequency distribution were derived for each region using the index rainfall method, first introduced for floods (Dalrymple 1960). The procedure assumes that the frequency distributions of all sites in a

homogenous region are identical, except for a site-specific scale factor. The quantile estimates $\hat{Q}(F)$, with non-exceedance probability F , at a site in a region with N sites is then computed by: $Q_i(F) = l_1 q(F)$; where q is a common dimensionless quantile function (growth curve) and l_1 is a site-specific scaling factor, also called the index rainfall value, representing the T -year quantile of the normalized regional distribution. In this study, the mean (μ_i) was used as the site specific scale factor at a given location. At-site AMS1, AMS3, AMS5 and AMS7 values with return periods $T = 2, 10, 100$ and 1,000 years were estimated from the regional quantiles.

The accuracy of the estimated rainfall quantiles were assessed using Monte Carlo simulations (Hosking and Wallis 1997). According to the 5T guide (Robinson and Reed 1999), the T -year rainfall estimate in each of the regions will only be accurate up to 5T station-years of data available. To further assess the accuracy of the regional rainfall quantiles, regionally derived at-site rainfall estimates were firstly compared with those derived from fitting the best distribution to the at-site data. Secondly, at-site and regional based estimates for three validation stations were compared. Alumenda station has a 10 year record from 1998 to 2007, Toleza Farm has a 49 year record from 1941 to 1989, whereas Zomba Plateau has a 10 year record from 1983 to 1993. Alumenda and Zomba Plateau were selected for this assessment because of their short, but up to date record. On the other hand, the Toleza Farm record was long but not up to date because of a 10-year gap in the 1990s. The at-site and regional based RMSE values from Monte Carlo simulations were then compared.

Finally, at-site and regional based estimates for the AMS1 were compared to those derived for Malawi by Drayton (1980) (see Sect. 1). In this comparison, both the mean and the median were used as the site specific scale factor since Drayton (1980) used the median rainfall. However, differences can in this case be attributed to a range of factors such as different data periods and thus extreme events, different regionalisation approaches and possibly measurement practices.

For all the procedures presented above, various packages and macros of the free statistical software R were used (R Development Core Team 2008). The L-moments approach used the package *lmomRFA* (<http://cran.r-project.org/web/packages/lmomRFA/index.html>) by Hosking (2009) in R Software (R Development Core Team 2008).

4 Results and discussions

4.1 Stationarity and serial independence check

The basic statistics for the stations, including the MK statistic at $\alpha = 0.05\%$ significance level, are presented in Table 2. The overall average for AMS1, AMS3, AMS5 and

Table 2 Basic statistics of the annual maximum 1-day rainfall series for Southern Malawi

Station	Mean (mm)				Mann–Kendall (MK)			
	AMS1	AMS3	AMS5	AMS7	AMS1	AMS3	AMS5	AMS7
Balaka	83.5	126.8	156.6	177.2	−0.14	0.57	0.75	0.98
Bvumbwe	78.1	120.0	151.3	175.4	1.82	1.12	2.14	1.30
Chanco	106.8	151.6	194.0	219.8	−0.02	−1.64	−0.84	−1.14
Chichiri	79.2	118.5	149.1	175.4	0.52	−0.39	0.00	0.46
Chikwawa	85.7	119.2	151.7	171.9	0.00	0.00	0.68	0.08
Chikweo	89.1	128.4	166.5	188.2	0.18	0.15	0.35	0.00
Chileka	79.0	119.9	139.6	154.5	−0.86	−0.70	−0.73	−0.59
Chingale	87.6	131.3	155.7	173.6	−0.82	−0.26	0.31	0.40
Liwonde	90.0	123.5	146.5	166.0	0.40	−1.00	−0.44	−0.58
Makhanga	77.7	119.0	137.6	156.5	1.96	2.55	2.01	2.08
Makoka	79.1	126.1	157.0	177.8	0.79	0.50	0.61	0.48
Mangochi	69.3	106.8	129.2	146.2	−0.43	0.43	0.32	0.07
Mimosa	108.4	169.2	201.9	230.7	−0.32	0.18	0.39	0.75
Monkeybay	94.1	149.2	181.8	215.9	−0.30	0.58	−0.02	0.32
Mwanza	83.4	136.1	164.0	188.4	0.49	0.10	0.49	0.71
Naminjiwa	92.5	136.7	163.3	186.7	1.61	1.69	1.52	0.95
Nchalo	74.6	110.3	136.4	154.1	1.20	1.80	1.93	2.21
Neno	102.6	151.4	183.2	211.7	0.32	−0.12	0.17	−0.20
Ngabu	80.1	116.0	147.5	171.3	1.61	2.00	2.19	1.94
Ntaja	84.7	124.7	149.1	166.8	1.06	1.97	1.44	0.76
Satemwa	82.8	130.5	165.7	190.7	0.44	0.71	2.25	1.81
Thyolo	89.0	130.8	163.5	189.4	−0.04	0.75	1.89	2.28
ZombaRTC	100.1	144.0	185.9	204.5	0.67	0.79	0.92	0.55

Bold values mean statistically significant at 0.05 level (exceed a critical value of 1.96)

AMS7 were 86.8, 130, 160 and 182 mm, respectively. The highest mean AMS1-AMS7 rainfall during the period were 108.4, 169, 202 and 231 mm all, recorded at Mimosa located in the eastern highlands which has the highest average annual rainfall in Malawi (Ngongondo et al. 2011).

None of the AMS1 MK trend statistics were significant at the $\alpha = 0.05$ level, and thus it can be assumed that all stations had stationary series in terms of mean values. However, statistically significant positive trends were exhibited by three stations for the AMS3 (Makhanga, Ngabu and Ntaja), four stations for the AMS5 (Bvumbwe, Makhanga, Ngabu and Satemwa), and three stations for the AMS7 series (Makhanga, Nchalo, and Thyolo). However, as most of the stations in the region did not have statistically significant trends, we could not reject the null hypothesis that the station trends in the region were homogenous. Hence, it was concluded that the significant trends were not significant at the regional level. We therefore accept that the region has statistically stationary trends. Annual, seasonal and monthly series analysed by Ngongondo et al. (2011) for the whole of Malawi also revealed statistically non-significant trends at $\alpha = 0.05$ level.

Further, the absolute values of the autocorrelation coefficients for lags 1 and 5, were not significant at the $\alpha = 0.05$ level. For a time series with n observations, the critical values at $\alpha = 0.05$ can be calculated from $1.96/\sqrt{n}$ (Douglas et al. 2000). We therefore accepted that the series were independently, identically distributed. Moran's I coefficient for all series suggested that cross-correlation among the stations were not statistically significant at $\alpha = 0.05$ for all series. The stations were therefore considered spatially independent in further computations.

4.2 Identification of homogenous regions

Prior to cluster analysis, the region was treated as one large homogeneous group and tested using the discordancy measure D_i and the heterogeneity measures H_1 , H_2 and H_3 . The region did not pass this homogeneity test because Liwonde station had a value $D_i \geq 3$ (the critical value for the 23 station grouping). A check of the data at the Liwonde station did not reveal any obvious inconsistencies. The heterogeneity values for the whole of Southern Malawi were: $H_1 = 1.64$, $H_2 = -0.66$, $H_3 = -0.57$. The region

should therefore be considered as possibly heterogeneous since $1 < H_1 < 2$, although the H_2 and H_3 values classified the area as acceptably homogeneous. The region did not pass the tests even after the removal of the Liwonde station and cluster analysis was therefore applied.

In the cluster analysis, the number of possible clusters was initially determined from a scree plot using a 15 cluster solution (figure not shown). Total loadings of the scree plot suggested between two and six possible clusters. The Hubert's gamma coefficients and the Dunn index cluster validity tests both suggested a three region partitioning with four, nine and ten stations. The results of Ward's clustering (dendrogram not shown) also agreed with the three clusters solution and station membership. However, in the suggested region 2, Chileka and Naminjiwa stations had discordancy values $D_i \geq 3$. These were subsequently relocated to region 3 based on similarities in their MAR and altitudes with stations in region 3. Drayton (1980) suggested three rainfall regions in Malawi as follows: Group 1 consisting of all areas where high rainfalls can be caused by convection over land surfaces; Group 2 areas where high rainfall events can be caused by both convection and orographic rainfall processes; and finally Group 3 where high rainfalls can be caused by either moist air convection from the lake surface or strong orographic barriers or a combination of both factors. The suggested regions in this study do not entirely agree with those suggested by Drayton (1980) in terms of their spatial distribution. Figure 2 shows stations in each group, whereas the D_i values and H_n measures for the three groups are presented in Table 3.

Fig. 2 Location of sites in the three rainfall clusters in southern Malawi

The first region (G1 hereafter) was unique in all cluster simulations. The G1 region had four stations namely Chikwawa, Makhanga, Nchalo and Ngabu. These stations are located in the predominantly semi-arid low lying Lower Shire valley. The region is in the southern arm of the Malawi Rift Valley with an average altitude of 84 m a.s.l. MAR for 1978–2007 was around 782 mm with average AMS1, AMS3, AMS5 and AMS7 series respectively of 80, 116, 143 and 163 mm. Rainfall in the region is strongly affected by rain shadow effects due to its low altitude. Intense rainfall in this region is most likely caused by convection over land surfaces (Drayton 1980).

The second region (G2 hereafter) was comprised of seven stations located along Lake Malawi, the upper Shire river basin and the surrounding medium altitude and plain areas with average altitude of 632 m a.s.l. The region had a MAR of around 901 mm with average AMS1, AMS3, AMS5 and AMS7 respectively of 85, 127, 155 and 176 mm. A combination of convective processes over land and adjacent water bodies (e.g. Lake Malawi and rivers) and rain shadow effects are major influences of intense rainfall in the region.

The third region (G3 hereafter) had 12 stations mostly located in the Southern Highlands with average altitude above 1,000 m a.s.l. MAR was 1,193 mm and average AMS1, AMS3, AMS5 and AMS7 of 90, 136, 168 and 192 mm respectively. G3 stations are all located in the high rainfall areas as discussed by Ngongondo et al. (2011). Convective, cyclonic and orographic rainfall processes influence extreme rainfall activities in this region. The G3 region face windward of south easterlies originating from the Indian Ocean. In addition, the region lies along paths tracked by tropical cyclones from the Indian Ocean. The highest recorded rainfall was observed at Zomba town (location of Zomba RTC and Chanco stations) when tropical cyclone “Edith” brought 509 mm of rainfall on 14 December 1946 (Drayton 1980).

4.3 Testing for heterogeneity (H)

The results for the heterogeneity test (H_n , $n = 1, 2, 3$) based on 1,000 simulations for each of the three homogeneous regions are shown in Table 5. For the AMS1 series, all three regions passed the heterogeneity test which suggest that for the AMS1 series the three regions can be considered as homogeneous since all $H_n < 1$. For the AMS3, AMS5 and AMS7, certain stations with shorter record lengths (marked asterisk in Table 3) had to be removed first for the groups to pass the heterogeneity test. However, we did not observe obvious inconsistencies in those stations and therefore subsequent analyses did not exclude these.

Table 3 Discordance and heterogeneity fit results for southern Malawi

Series	Region	Sites (Di)	Heterogeneity		
			H1	H2	H3
AMS1	G1	Chikwawa (1), Makhanga (1), Nchalo (1), Ngabu (1)	−1.93	0.39	0.15
	G2	Balaka (1.81), Chikweo (1.74), Chingale (0.86), Liwonde (1.9), Mangochi (0.32), Monkeybay (0.33), Ntaja (0.05)	0.89	−0.28	−0.25
	G3	Bvumbwe (1.06), Chanco (2.07), Chichiri (0.25), Chileka (1.63), Makoka (0.29), Mimosa (0.46), Mwanza (0.49), Naminjiwa (1.59), Neno (1.46), Satemwa (0.57), Thyolo (0.33), ZombaRTC (1.8)	0.82	−1.11	−0.22
AMS3	G1	Chikwawa (1), Makhanga (1), Nchalo (1), Ngabu (1)	0.14	−0.72	−0.22
	G2	Balaka (0.47), Chikweo (1.41), Chingale (1.11), Liwonde (1.70), Mangochi (1), Monkeybay (0.25), Ntaja (1.05)	0.67	−1.04	1.09
	G3	Bvumbwe (0.84), Chanco (0.46), Chichiri (1.41), Chileka (1.97), Makoka (1.28), Mimosa (0.13), Mwanza (0.67), Naminjiwa (0.83), Neno (0.60), Satemwa (0.51), Thyolo (0.37), ZombaRTC (2.93)	−0.22	−0.81	−1.15
AMS5	G1	Chikwawa (1), Makhanga (1), Nchalo (1), Ngabu (1)	−0.43	−1.2	−1.06
	G2	Balaka (0.43), Chikweo (0.90), Chingale (1.56), Liwonde (1.48), Mangochi (1.28), Monkeybay (0.33), Ntaja (1.01)	1.78	1	1.04
	G3	Bvumbwe (0.71), Chanco (1.46), Chichiri (0.40), Chileka (0.10), Makoka (1.87), Mimosa (0.61), Mwanza (2.24), Naminjiwa (0.96), Neno (2.02), Satemwa (0.55), Thyolo (0.09), ZombaRTC*	−0.83	−0.02	−0.7
AMS7	G1	Chikwawa (1), Makhanga (1), Nchalo (1), Ngabu (1)	−0.7	−0.97	−1.34
	G2	Balaka (0.97), Chikweo (1.31), Chingale (1.33), Liwonde*, Mangochi (0.39), Monkeybay (1.0), Ntaja*	0.01	0.92	0.68
	G3	Bvumbwe (0.38), Chanco (0.92), Chichiri (0.77), Chileka (1.36), Makoka (2.10), Mimosa (0.73), Mwanza (1.04), Naminjiwa (0.95), Neno*, Satemwa (0.96), Thyolo (0.79), ZombaRTC*	−0.68	−1.25	−1.83

Stations with shorter record lengths are marked with asterisk (*)

Table 4 Goodness of fit test results for candidate distributions

Series	Region	Acceptable distributions	Best distribution ($\min Z_{\text{crit}}^{\text{Dist}} \leq 1.64 $)	$ Z $ (best fit)	$\text{Min} \tau_{4(\text{sample})} - \tau_{4(\text{Dist})} $	DIST
AMS1	G1	GLO, GEV, GNO, PE3	GLO	0.41	0.02	GEV
	G2	GLO, GEV	GLO	1.04	0.014	GLO
	G3	GLO, GEV, GNO, PE3	GEV	0.44	0.007	GEV
AMS3	G1	GEV, GNO, PE3, GPA	PE3	0.71	0.025	PE3
	G2	GLO, GEV, GNO	GLO	0.28	0.0015	GLO
	G3	GEV, GNO, PE3	PE3	0.26	0.0057	PE3
AMS5	G1	GEV, GNO, PE3	PE3	0.37	0.0124	PE3
	G2	GEV, GNO, PE3	PE3	0.01	0.0004	PE3
	G3	GLO, GEV, GNO, PE3	GEV	0.25	0.004	GEV
AMS7	G1	GEV, GNO, PE3, GPA	PE3	0.19	0.0093	PE3
	G2	GEV, GNO, PE3	PE3	0.28	0.0085	PE3
	G3	GLO, GEV, GNO	GLO	0.56	0.0163	GEV

4.4 Goodness-of-fit measure (Z) and derivation of the regional growth curves

Goodness of fit (Z) test results for candidate distributions in the three homogeneous regions are shown in Table 4.

Acceptable distributions are all those satisfying the criteria $Z \leq |1.64|$ whereas the best distribution is the one satisfying the criteria $\min Z_{\text{crit}}^{\text{Dist}} \leq |1.64|$ among the acceptable distributions. L-moment ratio diagrams showing the location of regional average L-Cs and L-Ck and their

Fig. 3 L-moment ratio diagrams for the candidate distributions with regional average L-skewness and L-kurtosis. Filled triangle G1; filled circle G2; filled square G3

theoretical relationships with the different candidate distributions, are shown in Fig. 3a, b, c and d for AMS1, AMS3, AMS5 and AMS7, respectively.

From Table 4, it can be seen that the GEV is acceptable in all regions, whereas the GPA is the least acceptable appearing in region G1 for AMS3 and AMS7 only. The best distributions, i.e. with $\min Z_{\text{Crit}}^{\text{Dist}} \leq |1.64|$ and $\min |\tau_{4(\text{sample})} - \tau_{4(\text{Dist})}|$, were the GEV (G1 and G3 for AMS1, G3 for AMS5 and AMS7), PE3 (G1 and G3 for AMS3, G1 and G2 for AMS5 and AMS7) and GLO (G2 for AMS1 and AMS3 and G3 for AMS7). The L-moment ratio plot in Fig. 3 further confirms that these distributions were indeed closest to the regional weighted L-moments means and that the GEV is the best for AMS7 in G3. Hosking and Wallis (1997) recommended the use of four or five parameter distributions e.g. Wakeby or Kappa if the regional L-moment average lies above the GLO line as found for AMS1 and AMS3 in G2. In these cases, the Kappa distribution approximates the GLO. Hence, the choice of the GLO could be justified.

Table 5 shows the location, scale and shape parameters respectively (ξ , a and κ), of the acceptable distributions as well as the five-parameter WAK distribution in each region. Table 6 shows the T -year regional quantile estimates, 90% error bounds and the RMSE values from Monte Carlo simulations.

From Table 6, all series show that the G3 region's quantile estimates were not the highest, despite the G3 region being located mostly in the high rainfall areas of southern Malawi highlands. G2 is composed of stations in the upper Shire river valley and along the Lake Malawi region. The G2 region also had the highest AMS1 although its regional average maximum rainfall AMS1 was lower than that of the G3 region.

RMSE values for the regional quantiles increased with return period (Table 6). For the AMS1, regions G1 and G3 had relatively lower RMSE values for the accepted GEV distribution. This is reasonable as the GEV distribution is deemed suitable for estimation of extremes with $T \leq 500$ years (Norbiato et al. 2007). Further, the 5T rule suggests

Table 5 Regional parameter estimates for the selected distributions

Series	Group	Distribution	Parameters				
			ξ	α	κ	γ	δ
AMS-1	G1	GEV	0.849	0.267	0.013		
		WAK	0.480	1.131	2.485	0.151	0.230
	G2	GEV	0.994	0.195	−0.057		
		GLO	0.911	0.187	−0.266		
	G3	WAK	0.101	15.271	27.845	0.340	0.082
		GEV	0.854	0.231	−0.054		
AMS-3	G1	WAK	0.506	1.598	7.305	0.311	−0.033
		PE3	1.000	0.378	0.959		
	G2	WAK	0.415	1.279	7.462	0.552	−0.272
		GLO	0.944	0.195	−0.170		
	G3	WAK	0.328	2.811	7.298	0.328	0.015
		PE3	1.000	0.323	1.181		
AMS-5	G1	WAK	0.507	1.366	7.944	0.387	−0.137
		PE3	1.000	0.382	0.997		
	G2	WAK	0.000	31.348	56.796	0.592	−0.293
		PE3	1.000	0.354	0.737		
	G3	WAK	0.413	1.316	3.592	0.321	−0.071
		GEV	0.869	0.210	−0.045		
AMS-7	G1	WAK	0.586	0.845	3.647	0.218	0.059
		PE3	1.000	0.389	1.238		
	G2	WAK	0.000	45.018	80.802	0.548	−0.219
		PE3	1.000	0.339	0.876		
	G3	WAK	0.448	1.351	6.008	0.433	−0.207
		GLO	0.950	0.138	−0.207		
		GEV	0.871	0.203	−0.058		
		WAK	0.570	1.279	6.078	0.244	0.023

that rainfall estimates in G1, which had 102 station years, is reliable up to $T = 500$. In G3, which had 310 station years, reliability of rainfall estimates goes beyond $T = 1,000$ years. Therefore, all rainfall estimates with return periods larger than $T = 100$ years should be treated with caution. In region G2, the supposedly best fitted GLO distribution had higher RMSE values in the upper tail (results not included in Table 6) as compared to the GEV distribution. The 95% error bounds for the GLO were also higher than those of the GEV. On this basis, the GEV distribution is recommended for application in all the three regions for AMS1. The Wakeby or Kappa distribution, considered more robust to frequency distribution misspecification (Wallis et al. 2007), can be used as an additional source of information, especially in G2, whenever the GEV suggest unrealistic or doubtful estimates.

For AMS3, the accepted PE3 distribution had relatively lower RMSE values for G1 and G3 regions. This indicates more reliability of the quantiles even for return periods $T \geq 100$ years. The acceptable GLO distribution in region

G2, however, had high RMSE values in the upper tail, suggesting the unreliability of quantiles with return period $T \geq 100$ years.

Similarly, in both AMS5 and AMS7, lower RMSE values at large return periods for the PE3 (regions G1 and G2) and GEV (G3 region) distributions suggest higher reliability of quantile estimates.

4.4.1 Comparison between regional and at-site based design values

Design rainfall estimates derived from regional quantiles were compared with those derived from fitting the acceptable distribution based on at-site L-moment ratios. Simulation results of the RMSE values were used as the comparison basis. Lower RMSE values give an indication of better accuracy.

RMSE values for regional based rainfall estimates were mostly lower than at-site based RMSE for the GEV distribution in all regions, in particular in the extreme upper tail where $F \geq 0.99$ or $T \geq 100$, as shown in selected AMS1 plots in Fig. 4a–f. In the lower tail, RMSE values of site based estimates and regional based estimates are similar. Despite the GLO distribution being accepted as the best for the G2 region, its RMSE values were higher than those of the GEV at all sites. This could be an indication that the GLO distribution was either misspecified as the best for region G2 or that some stations had erroneous data.

Finally, three stations not used neither in the cluster analysis nor in the derivation of the regional quantiles, were used to test the accuracy of the design rainfall estimates. All three stations passed the independence and stationarity tests with serial correlation coefficients (lag-1 and lag-5) and Mann–Kendall test statistic not significant at $\alpha = 0.05$.

Alumenda station has an observed maximum AMS1 of 130 mm (mean 87.26 mm), whereas Toleza has a maximum AMS1 of 149.4 mm (mean 71.3 mm). The observed maximum AMS1 for Zomba plateau is 162 mm (mean 119 mm). Figure 5 shows the regional and at-site estimates of design rainfall at these three stations based on the GEV distribution. It is seen that on the one hand, the design values of AMS1 rainfall for different return periods are in approximate agreement between the two approaches, except for Zomba plateau where larger deviations are found for increasing return periods. On the other hand, the estimation uncertainty as measured by the 95% error bounds is much smaller for the regional based approach than that for the at-site estimates. From Fig. 4, regional based RMSE values for Alumenda ranged from 1.2% ($T = 2$) to a maximum of 28.0% ($T = 1,000$). These were lower than the at-site based RMSE, which ranged between 8.9% ($T = 2$) and 177% ($T = 1,000$). For $T = 100$, RMSE

Table 6 Regional quantiles, 90% error bounds and RMSE (%) values for the three regions

Series	Region	Distribution	T (years): F :	2 0.5	10 0.9	100 0.99	1,000 0.999
AMS1	G1	GEV	$\hat{q}(F)$	0.95	1.44	2.04	2.61
			RMSE (%)	0.01	0.03	0.14	0.33
			Error bounds				
			Lower	0.92	1.39	1.78	2.00
			Upper	0.98	1.50	2.34	3.33
			$\hat{q}(F)$	0.91	1.47	2.59	4.62
	G2	GEV	RMSE (%)	0.02	0.02	0.20	0.69
			Error bounds				
			Lower	0.88	1.42	2.30	3.65
			Upper	0.94	1.52	3.06	6.46
			$\hat{q}(F)$	0.94	1.41	2.06	2.79
			RMSE (%)	0.01	0.01	0.09	0.24
	G3	GEV	Error bounds				
			Lower	0.92	1.38	1.91	2.40
			Upper	0.95	1.44	2.27	3.36
AMS3	G1	PE3	$\hat{q}(F)$	0.94	1.51	2.13	2.69
			RMSE (%)	0.02	0.03	0.12	0.21
			Error bounds				
			Lower	0.91	1.44	1.92	2.32
			Upper	0.97	1.58	2.40	3.18
			$\hat{q}(F)$	0.94	1.46	2.30	3.50
	G2	GLO	RMSE (%)	0.01	0.02	0.13	0.38
			Error bounds				
			Lower	0.91	1.42	2.08	2.87
			Upper	0.97	1.51	2.62	4.51
			$\hat{q}(F)$	0.94	1.43	2.01	2.55
			RMSE (%)	0.01	0.02	0.06	0.11
	G3	PE3	Error bounds				
			Lower	0.92	1.40	1.91	2.36
			Upper	0.95	1.47	2.16	2.83
AMS5	G1	PE3	$\hat{q}(F)$	0.94	1.51	2.15	2.73
			RMSE (%)	0.02	0.03	0.11	0.20
			Error bounds				
			Lower	0.90	1.45	1.96	2.38
			Upper	0.97	1.58	2.43	3.25
			$\hat{q}(F)$	0.96	1.47	2.01	2.47
	G2	PE3	RMSE (%)	0.01	0.02	0.07	0.13
			Error bounds				
			Lower	0.94	1.42	1.86	2.21
			Upper	0.98	1.52	2.17	2.75
			$\hat{q}(F)$	0.95	1.38	1.97	2.60
			RMSE (%)	0.01	0.01	0.08	0.21
	G3	GEV	Error bounds				
			Lower	0.93	1.35	1.84	2.26
			Upper	0.96	1.41	2.16	3.11

Table 6 continued

Series	Region	Distribution	T (years): F :	2 0.5	10 0.9	100 0.99	1,000 0.999
AMS7	G1	PE3	$\hat{q}(F)$	0.92	1.52	2.23	2.89
			RMSE (%)	0.02	0.03	0.12	0.23
			Error bounds				
			Lower	0.89	1.46	2.01	2.51
			Upper	0.96	1.60	2.53	3.45
			$\hat{q}(F)$	0.95	1.47	2.03	2.51
	G2	PE3	RMSE (%)	0.01	0.02	0.07	0.13
			Error bounds				
			Lower	0.93	1.43	1.89	2.27
			Upper	0.97	1.52	2.20	2.83
			$\hat{q}(F)$	0.95	1.38	1.94	2.52
			RMSE (%)	0.01	0.01	0.08	0.19
	G3	GEV	Error bounds				
			Lower	0.93	1.35	1.80	2.19
			Upper	0.96	1.41	2.12	2.99

values were 11.8 and 49.3% for the regional and at-site based estimates, respectively. For Toleza Farm, regional RMSE values ranged between 1.1% to a maximum of 38.0% while at-site based RMSE values ranged between 3.6–76.6%. The RMSE values for the 100-year estimates were 12.4% for the regional based and 24.3% for the site based estimates. Zomba plateau had regional RMSE values from 0.89% ($T = 2$ years) to 23% ($T = 1,000$ years). The at-site RMSE's were 17.0% ($T = 2$ years) to 69.0% ($T = 1,000$ years).

Thus, it can be concluded that the regional based estimates have smaller uncertainty than the at-site based estimates, and that, based on the RMSE values, regional based rainfall estimates also have better accuracy than the at-site based estimates. However, these estimates are only reliable up to $T = 100$ after which very high RMSE values were noted.

4.4.2 Comparison with previous studies

Drayton (1980) analysed AMS1 using an at-site approach with the EV1 distribution. Table 7 provides the ranked AMS1 rainfall extremes in Malawi from that study based on data from 1895 to 1978.

As demonstrated in Table 7, most 1-day rainfall extremes in Malawi occur in the southern region. The earliest record of 1-day rainfall extreme was 315 mm recorded on 15 January 1895 at Lauderdale Estate in the Shire Highlands. The largest event was recorded at Nkhotakota in 1956 with 572.5 mm.

Fig. 4 Regional and at-site based AMS1 RMSE values for various return periods at selected stations: **a** and **b** in G1; **c** and **d** in G2; **e** and **f** in G3

The largest recorded 1-day event to date in the southern region was 509 mm reported on 14.12.1946 at Zomba Town as a result of the combined effects of Cyclone “Edith” and orographic processes (Drayton 1980). Studies on AMS3, AMS5 and AMS7 for Malawi have not been documented.

Our results suggest that there are both similar and different results from those by Drayton (1980). For the common stations used, Table 8 shows the 20-year rainfall estimates from the two studies.

The comparison in Table 8 suggests that most estimates are comparable. The differences can be attributed to the different periods used and also the methodologies. It appears that the maximum AMS1 for the stations reported in Table 7 did not occur between 1978 and 1980, the common period of the data used in these two studies.

5 Conclusions

RFA of AMS1, AMS3, AMS5 and AMS7 using 23 rainfall stations in Southern Malawi has been implemented based on the well known index (flood) and L-moments methods. All stations passed some minimum requirements for RFA and were considered i.i.d., i.e. they passed the independence and stationarity tests based on their autocorrelations, Mann–Kendall statistics and spatial correlation.

The 23 stations did not constitute one homogeneous region. The k-means cluster analysis and Ward’s classification suggested three homogeneous rainfall regions: G1 in the Lower Shire valley, G2 in the Lake Malawi and Upper Shire plains and G3 in the Southern Highlands. Although Monte Carlo simulations for AMS1 identified the GLO distribution as the best for region G2 and the GEV in G1

Fig. 5 Design storms from at-site (*shaded*) and regional analysis for Alumenda in G1, Toleza Farm in G2 and Zomba Plateau with 95% error bounds as *bars*

and G3 regions, further accuracy assessments suggested that the GEV distribution is the best model for AMS1 in all three regions. Accuracy for rainfall estimates for return period $T > 100$ was, however, low for AMS1. More station years, either from longer records or more stations in the regions, would be required for rainfall estimates above $T = 100$ years. In the AMS3, AMS5 and AMS7, most regions accepted the PE3 distribution, followed by the GEV and least the GLO distribution. Our results are in

general agreement with Sarkar et al. (2009) who reported that most extreme flood events find their quantiles in the GEV, GLO and PE3 distributions.

The performance of the derived regional quantiles at validation sites was satisfactory and had smaller uncertainty as compared to at-site estimates. However, there is a need to develop a procedure for using at-site characteristics for estimating the mean or median rainfall at ungauged areas in the regions. This would enable the testing of the

Table 7 One-day rainfall extremes in Malawi from 1895 to 1978

Rank	Rainfall (mm)	Date	Station	Region	Type
1	572.5	21.2.1957	Nkhotakota	Central	Convictional
2	509	14.12.1946	Zomba Town	South ^a	Cyclone/orographic
3	355.6	6.3.1914	Mulanje	South ^a	Cyclone
4	315.2	15.1.1895	Lauderdale	South ^a	Cyclone
5	307.8	15.2.1950	Chigweje	South ^a	Cyclone
6	298.5	1967	Chintheche	North	Convictional
7	279.4	18.12.1904	Thornwood	South ^a	Cyclone
8	269	14.3.1967	Blantyre	South ^a	Convictional
9	260.6	14.12.1959	Dowa	Central	Convictional
10	260.4	14.3.1966	Chintheche	North	Convictional
11	255.8	28.2.1966	Zomba Plateau	South ^a	Convictional
12	>254	4.4.1956	Widespread	Widespread ^a	Cyclone

Source: Drayton (1980)

^a Stations located in Southern Malawi

Table 8 Comparison of some regional and at-site AMS1 statistics calculated by this study with those from Drayton (1980)

Region	Station	T = 20-year rainfall estimates			Maximum	
		Drayton	Regional	Site	Post 1978	Pre 1980
G1	Chikwawa	152.5	138.2	132.7	144.7	172.5
	Makhanga	120.1	125.3	124.1	138.5	118.1
G2	Mangochi	200.7	121.1	117.8	153.0	229.1
	Monkeybay	142.2	164.2	175.4	210.7	Na
G3	Bvumbwe	117.0	124.8	111.0	127.7	Na
	Chileka	139.7	126.3	137.5	169.5	Na
	Mimosa	154.9	173.2	178.7	185.2	Na
	Zomba Plateau	198.9	190.2	184.3	162	255.2

regional index rainfall quantiles that have been developed at ungauged sites. Depending on data availability, it would also be interesting to compare these results with those from a POT analysis.

Acknowledgments This research study is part of the project on capacity building in water sciences for the better management of water resources in Southern Africa (NUFUPRO-2007) funded by the Norwegian Programme for Development, Research and Education (NUFU) and we acknowledge their support. We also thank the Malawi Department of Climate Change and Meteorological Services for providing the rainfall data.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

Adamowski K (2000) Regional analysis of annual maximum and partial duration flood data by nonparametric and L-moment methods. *J Hydrol* 229:219–239

- Baeriswyl PA, Rebetez M (1996) Regionalisation of precipitation in Switzerland by means of principal component analysis. *Theor Appl Climatol* 58:31–41
- Barring L (1988) Regionalisation of daily rainfall in Kenya by means of common factor analysis. *J Climatol* 8:371–389
- British Geological Survey (2004) Groundwater quality: Malawi. British National Environment Research Council
- Chen YD, Huang G, Shao QX, Xu C-Y (2006) Regional low flow frequency analysis using L-moments for Dongjiang Basin in China. *Hydrol Sci J* 51:1051–1064
- Dalrymple T (1960) Flood frequency methods. U.S. Geological Survey, Water supply paper, 1543A, 11–51
- Dinpashoh Y, Fakheri-Fard A, Moghaddam M, Jahanbakhsh S, Mirnia M (2004) Selection of variables for the purpose of regionalization of Iran's precipitation climate using multivariate methods. *J Hydrol* 297(1–4):109–123
- Douglas EM, Vogel RM, Kroll CN (2000) Trends in floods and low flows in the United States: impact of spatial correlation. *J Hydrol* 240:90–105
- Drayton RS (1980) An analysis of maximum point daily rainfall in Malawi. Malawi Water Resources Division, WRD no. TP 6, WRB (Series) no. TP 6:1–16
- Durrans SR, Tomic S (1996) Regionalization of low-flow frequency estimations: an Alabama case study. *Water Resour Bull* 32:23–37
- Easterling DA (1989) Regionalization of thunderstorm rainfall in the contiguous United States. *Int J Climatol* 9(6):567–579

- Fowler HJ, Kilsby CG (2003) A regional frequency analysis of United Kingdom extreme rainfall from 1961 to 2000. *Int J Climatol* 23:1313–1334
- Gadgil S, Yadumani, Joshi NV (1993) Coherent rainfall zones of the Indian region. *Int J Climatol* 13:547–566
- Guttman NB (1993) The use of L-moments in the determination of regional precipitation climates. *J Climatol* 6:2309–2325
- Haddad K, Rahman A, Green J (2010) Design rainfall estimation in Australia: a case study using L moments and generalized least squares regression. *Stoch Environ Res Risk Assess*. doi:10.1007/s00477-010-0443-7
- Hosking JRM (1986) The theory of probability weighted moments. Research report RC 12210, IBM Research, Yorktown Heights
- Hosking JRM (1990) L-moments: analysis and estimation of distributions using linear combinations of order statistics. *J R Stat Soc Ser B* 52:105–124
- Hosking JRM (2009). Regional frequency analysis using L-moments, *lmomRFA* R package, version 2.2. <http://CRAN.R-project.org/package=lmomRFA>
- Hosking JRM, Wallis JR (1997) Regional frequency analysis: an approach based on L-moments. Cambridge University Press, Cambridge
- Jackson IJ (1972) The spatial correlation of fluctuations in rainfall over Tanzania: a preliminary analysis. *Arch Meteorol Geophys Bioklimatol Ser B* 20:167–178
- Jury MR, Gwazantini ME (2002) Climate variability in Malawi, part 2: sensitivity and prediction of lake levels. *Int J Climatol* 22:1303–1312
- Jury MR, Mwafulirwa ND (2002) Climate variability in Malawi, part 1: dry summers, statistical associations and predictability. *Int J Climatol* 22:1289–1302
- Kendall MG (1975) Rank correlation methods, 4th edn. Charles Griffin, London
- Khalili M, Leconte R, Brissette F (2007) Stochastic multisite generation of daily precipitation data using spatial autocorrelation. *J Hydrometeorol* 8:396–412
- Kulkarni A, Kripalani RH, Singh SV (1992) Classification of summer monsoon rainfall patterns over India. *Int J Climatol* 12:269–280
- Malawi National Contingency Plan 2009–2010 (2009) Government of Malawi
- Mann HB (1945) Nonparametric test against trend. *Econometrica* 13:245–259
- Matalas NC, Langbein W (1962) Information content of the mean. *J Geophys Res* 67(9):3441–3448
- Mazvimavi D, Meijerink AMJ, Stein A (2004) Prediction of base flows from basin characteristics: a case study from Zimbabwe. *Hydrol Sci J* 49:703–715
- McQueen JB (1967). Some methods for classification and analysis of multivariate observations. In: Proceedings of 5th Berkeley symposium on mathematical statistics and probability, Berkeley, University of California Press, vol 1, pp 281–297
- Moran PAP (1950) Notes on continuous stochastic phenomena. *Biometrika* 37:17–23
- Muñoz-Díaz D, Rodrigo FS (2004) Spatio-temporal patterns of seasonal rainfall in Spain (1912–2000) using cluster and principal component analysis: comparison. *Ann Geophys* 22(5):1435–1448
- New M, Hewitson B, Stephenson DB, Tsiga A, Kruger A, Manhique A, Gomez B, Coelho CAS, Masisi DN, Kululanga E, Mbambalala E, Adesina F, Saleh H, Kanyanga J, Adosi J, Bulane L, Fortunata L, Mdoka ML, Lajoie R (2006) Evidence of trends in daily climate extremes over southern and west Africa. *J Geophys Res* 111:D14102
- Ngongondo C, Xu C-Y, Gottschalk L, Alemaw B (2011) Evaluation of spatial and temporal characteristics of rainfall in Malawi: a case of data scarce region. *Theor Appl Climatol J*. doi:10.1007/s00704-011-0413-0
- Norbiato D, Borga M, Sangati M, Zanon F (2007) Regional frequency analysis of extreme precipitation in the eastern Italian Alps and the August 29, 2003 flash flood. *J Hydrol* 345:149–166. doi:10.1016/j.jhydrol.2007.07.009
- Noto LV, La Loggia G (2009) Use of L-moments approach for regional flood frequency analysis in Sicily, Italy. *Water Resour Manag* 23:2207–2229. doi:10.1007/s11269-008-9378-x
- Parida BP, Moalafhi DB (2008) Regional rainfall frequency analysis for Botswana using L-Moments and radial basis function network. *Phys Chem Earth* 33:614–620
- Ramos MC (2001) Divisive and hierarchical clustering techniques to analyse variability of rainfall distribution patterns in a Mediterranean region. *Atmos Res* 57:123–138
- Robinson A, Reed D (1999) Flood estimation hand book: statistical procedure for flood frequency estimation, vol 3. Institute of Hydrology, Wallingford
- Sajidu SMI, Masamba WRL, Thole B, Mwatseteza JF (2008) Groundwater fluoride levels in villages of Southern Malawi and removal studies using bauxite. *Int J Phys Sci* 1:1–11
- Sarkar S, Goel NK, Mathur BS (2009) Development of isoplethial map using L-moment approach for Tehri-Garhwal Himalaya. *Stoch Environ Res Risk Assess* 24:411–423
- Satyanarayana P, Srinivas VV (2008) Regional frequency analysis of precipitation using large-scale atmospheric variables. *J Geophys Res* 113:D24110. doi:10.1029/2008JD010412
- Schaefer MG (1990) Regional analyses of precipitation annual maxima in Washington State. *Water Resour Res* 26:119–131
- Singh KK, Singh SV (1996) Space–time variation and regionalization of seasonal and monthly summer monsoon rainfall on sub-Himalayan region and Gangetic plains of India. *Clim Res* 6:251–262
- Sivapalan M, Takeuchi K, Franks SW, Gupta VK, Karambiri H, Lakshmi V, Liang X, McDonnell JJ, Mendiando EM, O’Connell PE, Oki T, Pomeroy JW, Schertzer D, Uhlenbrook S, Zehe E (2003) IAHS decade on predictions in ungauged basins (PUB), 2003–2012: shaping an exciting future for the hydrological sciences. *Hydrol Sci J* 48:857–880
- Smithers JC, Schulze RE (2000a) Development and evaluation of techniques for estimating short duration design rainfall in South Africa. WRC report no. 681/1/00. Water Research Commission, Pretoria, 356 pp
- Smithers JC, Schulze RE (2000b) Long duration design rainfall estimates for South Africa. WRC report no. 811/1/00. Water Research Commission, Pretoria, 69 pp
- Smithers JC, Schulze RE (2001) A methodology for the estimation of short duration design storms in South Africa using a regional approach based on L-moments. *J Hydrol* 241:41–52
- Smithers JC, Schulze RE (2003) Design rainfall and flood estimation in South Africa. WRC report no. 1060/01/03. Water Research Commission, Pretoria, 155 pp
- Stedinger JR (1983) Estimating a regional flood frequency distribution. *Water Resour Res* 19(2):503–510
- Tallaksen LM, Madsen H, Hisdal H (2004) Frequency analysis, hydrological drought—processes and estimation methods for streamflow and groundwater, developments in water sciences, vol 48. Elsevier Science, The Netherlands
- R Development Core Team (2008) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. ISBN 3-900051-07-0. <http://www.R-project.org>
- Unganai LS, Mason SJ (2001) Spatial characterisation of Zimbabwe summer rainfall for the period 1920–1996. *S Afr J Sci* 97:425–431
- Van Regenmortel G (1995) Regionalization of Botswana rainfall during the 1980s using principal component analysis. *Int J Climatol* 15:313–323
- Venkatesh B, Jose M (2007) Identification of homogeneous rainfall regimes in parts of Western Ghats region of Karnataka. *J Earth Syst Sci* 116(4):321–329

- Vogel RM, Fennessey NM (1993) Should L-moment replace product moment ratios. *Water Resour Res* 29:1745–1752
- Von Storch H (1995) Misuses of statistical analysis in climate research. In: von Storch H, Navarra A (eds) *Analysis of climate variability: applications of statistical techniques*. Springer, Berlin
- Wallis JR, Schaefer MG, Barker BL, Taylor GH (2007) Regional precipitation frequency analysis and spatial mapping for 2-hour and 24-hour durations for Washington State. *Hydrol Earth Syst Sci* 11(5):415–442
- Ward JH (1963) Hierarchical grouping to optimize an objective function. *J Am Stat Assoc* 58:236–244
- Williams C, Kniveton D, Layberry R (2009) Rainfall variability and extremes over Southern Africa: assessment of a climate model to reproduce daily extremes. *Geophys Res Abst* 11 EGU 2009-12317, EGU General Assembly 2009
- WMO (1988) Analysing long time series of hydrological data with respect to climate variability and change, WCAP-3, WMO-TD 224
- Yang T, Xu C-Y, Shao Q-X, Chen X (2010a) Regional flood frequency and spatial patterns analysis in the Pearl River Delta region using L-moments approach. *Stoch Environ Res Risk Assess* 24:165–182
- Yang T, Shao QX, Hao Z-C, Chen X, Zhang Z, Xu C-Y, Sun L (2010b) Regional frequency analysis and spatiotemporal pattern characterization of rainfall extremes in Pearl River Basin, Southern China. *J Hydrol* 380(3–4):386–405
- Yue S, Wang CY (2002) Applicability of pre-whitening to eliminate the influence of serial correlation on the Mann–Kendall test. *Water Resour Res* 38(6):1068–1075
- Zhang JY, Hall MJ (2004) Regional flood frequency analysis for the Gan-Ming River basin in China. *J Hydrol* 296:98–117