Analysis of precipitation extremes in the Taihu Basin of China based on the regional L-moment method

Zhengzheng Zhou, Shuguang Liu, Yan Hu, Yuyin Liang, Hejuan Lin and Yiping Guo

ABSTRACT

China has suffered from increasingly severe flood events in recent years, most of which are caused by heavy rains. The substantial casualties and damage caused by flooding necessitates a better understanding of precipitation extremes, especially in heavily populated urban areas. Based on L-moments from a regional perspective, this paper analyzes precipitation extremes in the Taihu Basin, utilizing annual maximum daily precipitation and partial duration series at 96 rain gages. The comparison of regional and at-site analysis results shows that the former provides more robust estimates, especially in the upper tail of a distribution (higher quantiles). Also, the use of partial duration series, which captures more information about extreme events, was found to be preferable to describe the extreme precipitation events in the Taihu Basin. Given the recently observed more frequent occurrence and greater magnitude of precipitation extremes, it is suggested that the food design standard used in the basin should be updated, especially for the urbanizing zones.

Key words | extreme precipitation, L-moments, regional precipitation frequency analysis, Taihu Basin

Zhengzheng Zhou
Shuguang Liu (corresponding author)
Yuyin Liang
Department of Hydraulic Engineering,
Tongji University,
Shanghai,
China
E-mail: *liusgliu@tongji.edu.cn*

Zhengzheng Zhou
Shuguang Liu
UNEP-Tongji Institute of Environment for Sustainable Development,
Shanghai,
China

Yan Hu
Hejuan Lin
Bureau of Hydrology,
Taihu Basin Authority, Ministry of Water Resources,
Shanghai,
China

Yiping Guo
Department of Civil Engineering,
McMaster University,
Hamilton, ON,
Canada L8S 4L7

INTRODUCTION

Since the late 20th century, the frequency of extreme precipitation events has increased substantially in many countries worldwide, including China (IPCC 2013). Extreme precipitation events very often cause severe floods. In most of the southeast and northwest parts of China, the frequency of precipitation-induced flood events has increased significantly (Li *et al.* 2012). Meanwhile, the rapid urbanization in recent years in China greatly aggravates flood damage. It is imperative to investigate the extreme rainfall-induced floods in urban areas, and the analysis of extreme precipitation events is a critical component of engineering design to protect against such events. To date, the quantification of rainfall frequency in China is mainly based on at-site frequency analysis via the conventional moment method (CMM) using Pearson type III distributions (PE3) (Ministry of Water Resources China 2006). This method has been applied since the 1950s without any modifications or updates. Previous studies have demonstrated that this method leads to a lower estimation of extreme precipitation amounts and correspondingly lower flood design standards (Wang *et al.* 2008). Therefore, to update the current design precipitation atlas and to enhance flood protection capabilities, it is necessary to examine the precipitation extremes using other approaches.

Regional frequency analysis, which trades space for time, provides another framework for characterization of the frequency distribution of hydrological events

doi: 10.2166/nh.2016.071

(Norbiato *et al.* 2007) and improves at-site statistical characterization by incorporating spatial data from several sites in a homogeneous region. This method assumes that in a homogeneous region, the distributions of extremes are identical apart from a site-specific scaling factor. In this study, the index flood/precipitation, which is one of the commonly employed models, is used as a representative quantity of regional frequency analysis.

Similar to conventional moments, L-moments provide measures of distributional location (mean), scale (variance), skewness (shape), and kurtosis (peakedness) as well. Distribution parameters' estimation based on L-moments is robust to outliers and virtually unbiased for small samples (Hosking & Wallis 1997). It has been accepted as an accurate and robust method for selecting and parameterizing representative probability distribution functions (Lin & John 1993) and is used in many countries and regions worldwide (Parida *et al.* 1998; Sankarasubramanian & Srinivasan 1999; Yue & Wang 2004; Atiem & Harmancioğlu 2006; Liu *et al.* 2007; Kjeldsen & Jones 2009; Um *et al.* 2010; Chen & Hong 2012). In the United States, the L-moment method (LMM) for distribution parameter estimation has been applied as a design standard for hydrological works by the OHD of NOAA (Lin & John 1993; Lin *et al.* 2012). However, in China, there is still a lack of relevant research in urban areas. Thus it is worthwhile to test this approach and assess its applicability in China.

The Taihu Basin is one of the most important basins in China with substantial recent increases in urban land. It also suffers from massive and severe flood damage induced by precipitation events every year (Wu 2000a, 2000b; Wu & Guan 2000; Ou & Wu 2001). A better understanding of frequency of occurrence of extreme precipitation is urgently needed. Several studies about the extreme rainfall frequency in the Taihu Basin have already been conducted. Liang *et al.* (2013) compared LMM and CMM utilizing rainfall data in the Taihu Basin, and the results demonstrated a better performance of LMM. Zhou *et al.* (2014b) investigated the correlation between rainfall amounts observed in several small subbasins of the Taihu Basin. Zhou *et al.* (2014a) conducted at-site analysis at the rain gage with the longest rainfall records in the Taihu Basin using the parameter estimation methods of maximum likelihood and LMM. This study extends from those previous studies and provides a framework of precipitation frequency analysis using regional L-moments analysis. In this study, estimation of precipitation extremes is carried out based on L-moments from a regional perspective using two datasets, i.e., annual maximum daily precipitation (AMDP) and partial duration series (PDS). The estimations from the regional approach and the at-site approach are compared, and the results obtained using AMDP and PDS are also compared.

METHODOLOGY

Theory of L-moments

L-moments are modifications of probability weighted moments (Greenwood *et al.* 1979) and are expectations of certain linear combinations of order statistics (Hosking 1990). Letting $X_{1:n} \le X_{2:n} \le \ldots \le X_{n:n}$ be the order statistics of a random sample of size n drawn from the distribution of X, define the r^{th} L-moment λ^r of variable X to be the quantities:

$$\lambda_r \equiv r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} EX_{r-k:r}, r = 1, 2, \ldots \quad (1)$$

Based on the above definition, the first four L-moments of X can be written as follows:

$$\lambda_1 = EX \quad (2)$$

$$\lambda_2 = \frac{1}{2} E(X_{2:2} - X_{1:2}) \quad (3)$$

$$\lambda_3 = \frac{1}{3} E(X_{3:3} - 2X_{2:3} + X_{1:3}) \quad (4)$$

$$\lambda_4 = \frac{1}{4} E(X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4}) \quad (5)$$

Similar to conventional comment ratios, Hosking (1990) defined L-moment ratios as:

$$t = \frac{\lambda_2}{\lambda_1},\ \mathrm{L{-}C_V} \tag{6}$$

$$t_3 = \frac{\lambda_3}{\lambda_2},\ \mathrm{L{-}C_S} \tag{7}$$

$$t_4 = \frac{\lambda_4}{\lambda_2},\ \mathrm{L{-}C_K} \tag{8}$$

L-moment ratios measure the shape of a distribution independently of its scale of measurement. L-C_v, L-C_s, and L-C_k are the most useful quantities for summarizing probability distributions. For example, the values of these ratios of some hydrologic quantities can be used to show the impacts of urbanization (Kjeldsen 2010).

The L-moment approach of regional frequency analysis

This approach starts by identifying the regional homogeneity and evaluating its validity through statistical tests. It investigates the data properties by L-statistics, groups the sites into homogeneous regions, and then estimates the distribution of extremes at a given site. The merit of regional analysis is that sampling variations in the parameter estimates are more unbiased, and high quantiles (the upper tail of a distribution) are more robust. Hence, it provides more accurate estimates as it combines data at different sites in one homogeneous region.

First, to screen for erroneous data and check the data independence, the statistic D_i is calculated based on the difference between L-moment ratios of a site and the average L-moment ratios of a group of similar sites. For a group containing N sites, D_i is:

$$D_i = \frac{1}{3}N(u_i - \bar{u})^T A^{-1}(u_i - \bar{u}),\ i = 1, 2, \ldots, N \tag{9}$$

where $u_i = [\mathrm{t}^{(i)}, \mathrm{t}_3^{(i)}, \mathrm{t}_4^{(i)}]$; $A = \sum_{i=1}^{N}(u_i - \bar{u})(u_i - \bar{u})^T$; $\bar{u} = \sum u(i)/N$. The critical values of D_i, based on number of site, are set to examine the site's discordancy (Hosking & Wallis 1997).

The statistic H is selected as the heterogeneity measure to identify the homogeneity. Random samples of size 500 is generated via Monte Carlo simulations and the variability of the L-statistics of the real region is compared with that of the simulated region. Assuming that the proposed region has N sites, with site i having a record length of n_i:

$$t^R = \frac{\sum_{i=1}^{N} n_i t(i)}{\sum_{i=1}^{N} n_i} \tag{10}$$

$$t_t^R = \frac{\sum_{i=1}^{N} n_i t_r^{(i)}}{\sum_{i=1}^{N} n_i} \tag{11}$$

where $t^{(i)}$, $t_3^{(i)}$ and $t_4^{(i)}$ are sample L-moment ratios of site i, t^R, t_3^R, and t_4^R are the regional average L-C_V, L-C_S, and L-C_K, respectively, weighted proportionally to the site's record length.

The weighted standard deviation of an L-moment ratio in a homogeneous region is:

$$V = \left\{ \frac{\sum_{i=1}^{N} n_i \left(t_r^{(i)} - t^R\right)^2}{\sum_{i=1}^{N} n_i} \right\}^{1/2} \tag{12}$$

The heterogeneity measure H is computed as:

$$H = \frac{(V - \mu_V)}{\sigma_V} \tag{13}$$

The criteria for assessing the homogeneous regions are (Hosking & Wallis 1990): if $H < 1$, the region is acceptably homogeneous; if $1 < H < 2$, the region is possibly homogeneous; and if $H > 2$, the region is definitely homogeneous.

Although there is a belief among hydrologists that the extreme hydrological events are best described by heavy-tailed distributions (Strupczewski *et al.* 2011), the behavior of the upper tail of a distribution needs to be examined with different datasets (El Adlouni *et al.* 2008). In this basin, to choose the best distributions for homogeneous

regions, five distributions covering a range of different tail weights are considered as candidates: generalized logistic (GLO), generalized extreme-value (GEV), generalized normal (GNO), generalized pareto (GPA), and Pearson type III (PE3) distributions. The goodness-of-fit statistic Z^{DIST} is computed to evaluate the performances of distributions. For a certain distribution, the goodness-of-fit test is carried out by first calculating the following quantities:

$$B_4 = N_{sim}^{-1} \sum_{m=1}^{N_{sim}} (t_4^{(m)} - t_4^R) \tag{14}$$

$$\sigma_4 = \left[(N_{sim} - 1)^{-1} \left\{ \sum_{m=1}^{N_{sim}} (t_4^{(m)} - t_4^R) - N_{sim} B_4^2 \right\} \right]^{1/2} \tag{15}$$

$$Z^{DIST} = \frac{(\tau_4^{DIST} - t_4^R + B_4)}{\sigma_4} \tag{16}$$

Then the calculated Z^{DIST} value is compared with the critical value. The critical value is set as $|Z^{DIST}| \leq 1.64$, corresponding to a level of significance of 10%. The selection of the final distribution is based on the minimum error achieved between the generated data and the real data.

The index-rainfall/index-flood, as suggested by Dalrymple (1960), is used for quantile estimation in this study. It is assumed that apart from a site-specific scale parameter or an index scale, frequency distributions at the sites in a homogeneous region are approximately identical. In a homogeneous region with N sites, the quantile function of the frequency distribution at site i is:

$$Q(F) = l_i q(F);\ i = 1, 2, \ldots, N, 0 < F < 1 \tag{17}$$

where l_i is the site-dependent scale factor, the index rainfall; $q(F)$ is the regional quantile function referred to as the regional growth factor (RGF) which represents the regional average T-year ($T = 1/(1-F)$) quantile of the chosen distribution.

Extension to PDS methods

For a more reliable estimation of precipitation extremes, it is very important to involve as many large events (peaks) as possible from the observed series. The AMDP series contains one annual peak per year and is accepted as an efficient and common series for frequency analysis in hydrology. However, regardless of whether the second largest event in a year exceeds the largest events of other years, the second largest event in any year will not be included in the AMDP series; thus the AMDP series may employ insufficient information (Zhou *et al.* 2014a). When dealing with values exceeding a certain threshold, PDS is preferred for hydrological frequency analysis, and is capable of capturing more information about extreme events than the AMDP series (Hosking & Wallis 1997; Norbiato *et al.* 2007, Pham *et al.* 2014a; Zhou *et al.* 2014a). This approach is effective for accurately estimating precipitation extremes, reducing the uncertainty and maximizing the utilization of available data (Pham *et al.* 2014b). Although there still exist some difficulties in defining PDS (Madsen *et al.* 1997), the method has been studied by more researchers in recent years.

PDS is a series of selected data whose magnitude is greater than a threshold value (u) or over a threshold characterized by the average number of peaks per year (λ). For a series of observations $x = \{x_1, x_2, \ldots, x_n\}$ over n year, PDS is a set containing m values, $y = \{y_1, y_2, \ldots, y_m\}$, which exceed a chosen threshold u, thus, $y_i > u$, $i = 1, 2, \ldots, m$; and $\lambda = m/n$. Here, the total number of peaks (m) is a random quantity chosen from the average number of peaks per year. The value of m depends on either the average number of peaks per year or the threshold value. Thus, compared with the AMDP series, the PDS is able to extract more information for the prediction of extreme rainfalls.

In this study, the main procedure of PDS data selection at site i is as follows:

1. Produce the AMDP series, $\{x_{i,\ \mathrm{AMDP}}\}$, with a data length of $N_{i,\ \mathrm{AMDP}}$.
2. Select the average number of peaks per year λ and determine the required data length $N'_{i,\ \mathrm{PDS}} = \lambda N_{i,\mathrm{AMDP}}$. Note that all the peaks are independent random variables.
3. Based on the required data length $N_{i,\ \mathrm{PDS}}$, where, e.g., $\lambda = 3$ and $N_{i,\ \mathrm{PDS}} = 3N_{i,\ \mathrm{AMDP}}$, determine the required threshold $x_{\min}$.
4. Based on $x_{\min}$, select observations from individual years of record and form the PDS series, $\{x_{i,\ \mathrm{PDS}}\}$, where $x_{i,\ \mathrm{PDS}} > x_{\min}$.

STUDY AREA AND DATASET

The Taihu Basin

The Taihu Basin (30.14 °–32.26 °N, 119.0 °–121.89 °E) is located in the Yangtze Delta plain at the mouth of the Yangtze River (Wang 2006); with an area of 36,900 km^2, it is one of the tributary regions of the Yangtze River Basin. Although the Taihu Basin only contains 0.4% of the total area of China, it is one of the most economically advanced areas of China (Yang & Wang 2003).

With a climate of subtropical humid monsoon, the Taihu Basin receives about 1,181 mm of rainfall per year (Gou *et al.* 2010). Precipitation has a considerable temporal and spatial variation in this region, generally decreasing from south to north. During the flooding season from May to September, plum rains with the characteristics of long duration, and heavy rains with the characteristics of great intensity and short duration both occur frequently (Wu 2000a, 2000b). Both types of rains contribute to serious flood risks. Considering that the region is suffering from frequent flood events with the recent process of fast urbanization, the Taihu Basin Authority (TBA) has updated its flood control plan in the Taihu Basin and has raised the longest return period for flood control up to 200 years in 2015 (Huang 2000; TBA 2008).

Dataset

The precipitation dataset used in this study is from 155 rain gages in the Taihu Basin maintained by the TBA. Considering that there is a large amount of randomness in the AMDP series, only the gages with a record length of at least 20 years were chosen first. Thus, the number of sites valid for further analysis is reduced from 155 to 96 (Figure 1). The data length of the 96 rain gages varies from 21 to 70 years (Figure 2); the maximum of the AMDP is 563.90 mm and the minimum is 25.00 mm, which shows a significant variation of annual maximum precipitation. For water resources management, the TBA has divided the basin into seven sub-regions based on its geographic and hydrometeorological characteristics. Region 6, a mountainous area in the southwest part of the Taihu Basin, has the highest mean value of the AMDP; while Regions 3 and 4, located in the north part of the basin with a typical plain area, have relatively smaller mean AMDP. The general information about these sub-regions is provided in Table 1. The at-site sample L-moment statistics L-C_v and L-C_k, both as functions of

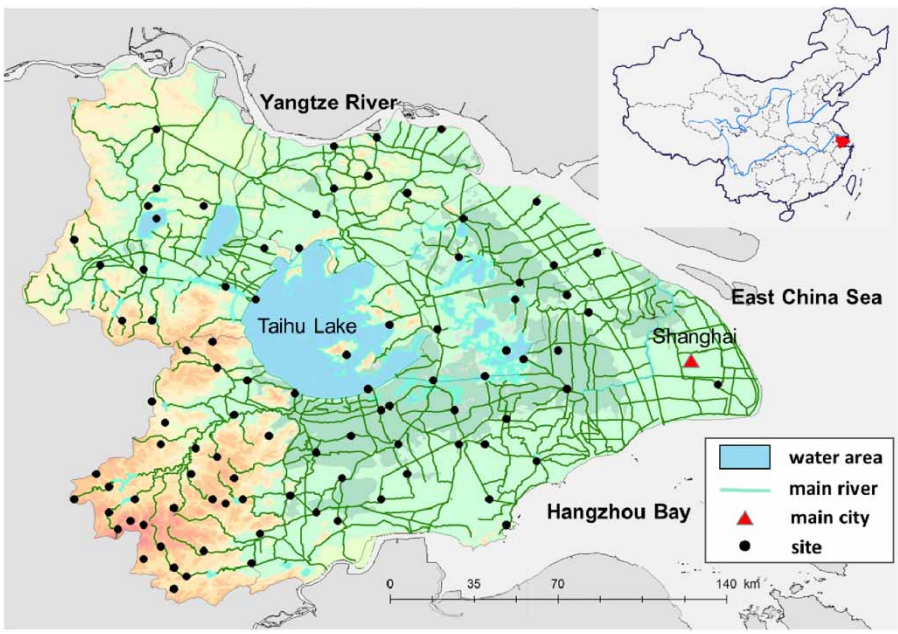

Figure 1 | Geographical locations of rain gages in the Taihu Basin.

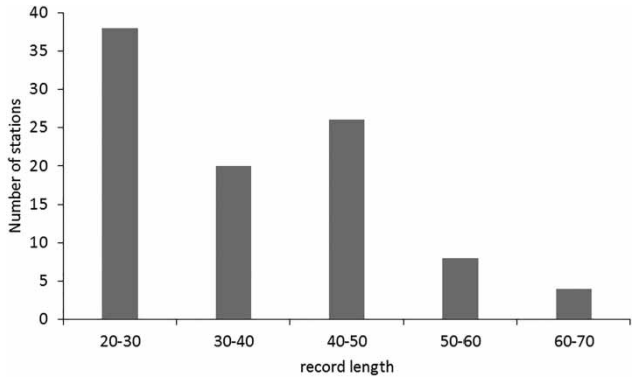

Figure 2 | Distribution of record lengths at 96 rain gages in the Taihu Basin.

Table 1 | General information about the seven sub-regions in the Taihu Basin

No.	Sub-region	Number of sites	Mean annual precipitation (mm)
Region 1	Taihu	7	1,156
Region 2	Hangjiahu	21	1,225
Region 3	Wuchengxiyu	7	1,074
Region 4	Yangchengdianmao	9	1,098
Region 5	Huxi	12	1,132
Region 6	Zhexi	34	1,441
Region 7	Pudongpuxi	6	1,105
Taihu Basin		96	1,181

L-C_s, are shown in Figure 3. The Pearson correlation coefficient between L-C_s and L-C_k is 0.73, which is higher than that between L-C_s and L-C_v.

RESULTS AND DISCUSSION

Analysis of precipitation extremes (AMDP/L-moment)

The results of discordancy measures D_i for the 96 rain gages are shown in Figure 4. Two sites in Region 6, site 6032 and site 6033, are considered as 'probably' discordant with $D_i = 4.74$ and $D_i = 3.79$, respectively. The large D_i indicates that the data series at the two sites might have some irregularities or inconsistencies, and are discordant with other sites in their group. To decide whether to include them or not in subsequent analysis, further investigation is required. Data series at site 6032 and site 6033 are plotted in Figures 5 and 6. At site 6032, an outlier (538.80 mm) in 1971 is observed in Figure 5, which is the second largest extreme in Region 6. Considering that the data length of site 6032 is only 30 years and the mean value of AMDP is 100.85 mm, the site discordancy is associated with this outlier. At site 6033, the data from 1985 to 2001 are missing (Figure 6). As site 6033 has the smallest L-C_s in Region 6 with a relatively short data length, the discordancy of site 6033 is probably caused by insufficient data. Furthermore, no gross errors or incorrect recording of data values were identified after the recheck of data validity. Also, there is no physical ground (such as topography, meteorology) to remove the two sites from Region 6 in subsequent analysis. Hence, data at all rain gages are tentatively retained for carrying out further analysis so as to use as much data as possible. In the following section, the issue of discordancy will be discussed again.

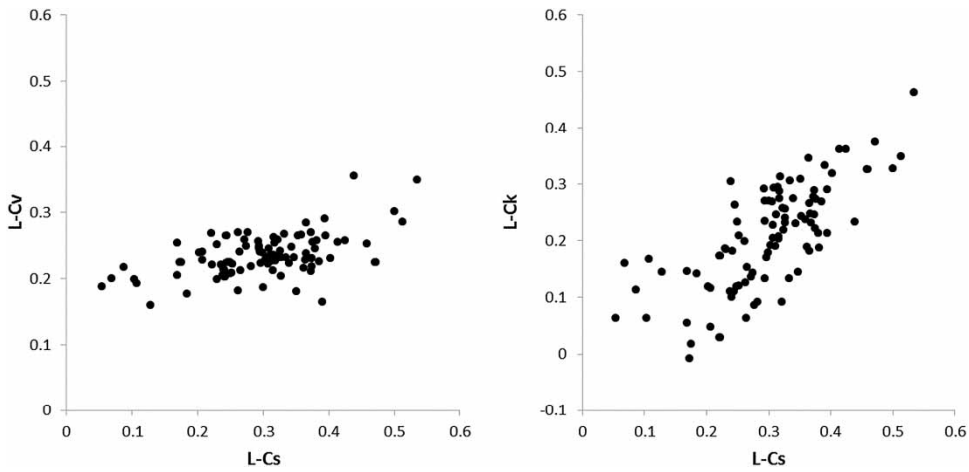

Figure 3 | L-moment statistics calculated from precipitation data of the Taihu Basin.

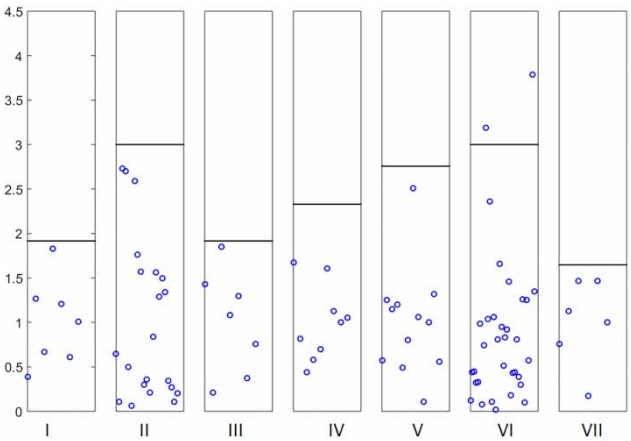

Figure 4 | D_i for 96 rain gages in seven sub-regions.

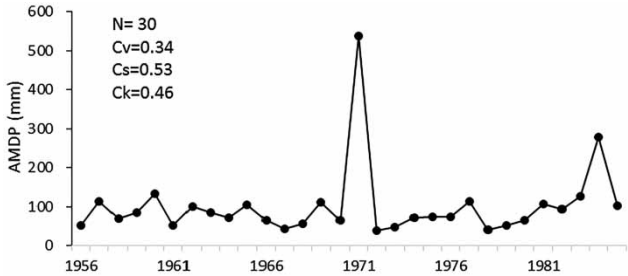

Figure 5 | Time series plot of the AMDP series at site 6032.

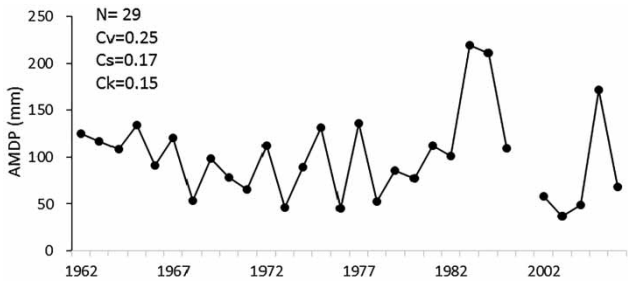

Figure 6 | Time series plot of the AMDP series at site 6033.

The results of the heterogeneity measures (Table 2) demonstrate that all the sub-regions are considered to be definitely heterogeneous except for Region 5, whose $H(3)$ is slightly over 1. $H(1)$ and $H(2)$, calculated from L-C_v/L-C_k and L-C_s/L-C_k, are more representative indexes of heterogeneity compared with $H(3)$. Thus, Region 5 is considered to be approximately heterogeneous.

The best distributions for the seven homogeneous regions are summarized according to Z^{DIST} (Table 3). GLO is chosen for Regions 2, 3, and 4. GEV are fitted for Regions 1 and 6. PE3 and GNO are fitted for Region 5 and Region 7, respectively. Formation of the seven sub-regions and their spatial distributions are provided in Figure 7.

Table 2 | Results of heterogeneity tests in the Taihu Basin

	$H(1)$	$H(2)$	$H(3)$
Region 1	−1.69	−1.11	−0.91
Region 2	−1.91	−2.92	−2.75
Region 3	−1.28	−1.26	0.12
Region 4	−0.31	0.44	0.36
Region 5	0.14	0.39	1.01
Region 6	−0.18	−1.95	−1.49
Region 7	−0.42	−1.32	−0.66

Following the spatial distribution of the homogeneous regions, RGFs are computed (Figure 8). It is shown that the RGF increases with the increase of return period in all the regions. For a return period of 200 years, the RGF is the largest in Region 2, and the smallest in Region 4. The precipitation at each gage for various return periods is estimated next. The regional precipitation estimations find that precipitation extremes decrease from the southwest to the northeast (figures are not shown here). The largest precipitation extremes occur in the mountainous area in the southwest of the Taihu Basin and the lowest extremes happen at the northeast plain zone. The second largest precipitation extremes occur in Shanghai. With return periods >100 years, the scale of rainfall area decreases and the centroid of rainfall moves towards north.

Figure 9 compares the results of regional analysis and at-site analysis at two gages. At site 6032, estimates via at-site analysis are significantly higher than those via regional analysis for larger return periods, largely due to the existence of an outlier in the data series. This demonstrates the robustness of the regional analysis. At site 6033, due to the lack of data from 1985 to 2001, estimates via regional analysis are higher than those via at-site analysis for larger return periods, which are closer to the observed data. It shows that regional analysis, utilizing the regional index, has an unavoidable influence on the upper tail of a distribution (higher quantiles).

Note that in regional frequency analysis, the aim is not to identify a 'true' distribution but to find a distribution that will yield accurate quantile estimates for each site using a single frequency distribution to fit data from several sites (Hosking

Table 3 | Summary of the best distributions for individual sub-regions based on Monte Carlo simulation tests

		Region 1	Region 2	Region 3	Region 4	Region 5	Region 6	Region 7
GLO	L-C_K	0.239	0.262	0.212	0.218	0.203	0.260	0.240
	Z	0.42	−0.16	−0.40	−0.23	4.55	1.54	1.19
GEV	L-C_K	0.212	0.239	0.178	0.186	0.167	0.237	0.212
	Z	−0.37	−1.25	−1.42	−0.78	2.75	0.12	0.43
GNO	L-C_K	0.191	0.213	0.165	0.171	0.157	0.211	0.192
	Z	−0.97	−2.48	−1.81	−1.25	2.27	−1.44	−0.15
PE3	L-C_K	0.156	0.169	0.142	0.144	0.137	0.167	0.156
	Z	−2.00	−4.58	−2.54	−2.09	1.30	−4.10	−1.15
GPA	L-C_K	0.139	0.171	0.097	0.106	0.082	0.168	0.139
	Z	−2.49	−4.58	−3.89	−3.30	−1.42	−4.05	−1.62
BEST		**GEV**	**GLO**	**GLO**	**GLO**	**PE3**	**GEV**	**GNO**

Figure 7 | Spatial distribution of the homogeneous regions in the Taihu Basin.

& Wallis 1997). In general, a homogeneous region is slightly heterogeneous, and there will be no single 'true' distribution that applies to each site. Thus, the results of frequency analysis for sites 6032 and 6033 are reasonable here.

Further analysis of heterogeneity

Considering that the bias in the estimation of RGF may be increased by heterogeneity if the 'probably' discordant sites are retained, the impacts of sites 6032 and 6033 on regional estimates are evaluated via sensitivity tests.

Let Q_0 be the regional precipitation estimates (with return periods = 25, 50, 100, 200 years) without site A, and Q_1 be the regional precipitation estimates with site A included; the sensitivity index is defined as $p = |Q_1 - Q_0|/Q_0$. The threshold is set as $p = 5\%$, if $p < 5\%$ site A can be retained; otherwise, the site may be deleted or grouped to another region.

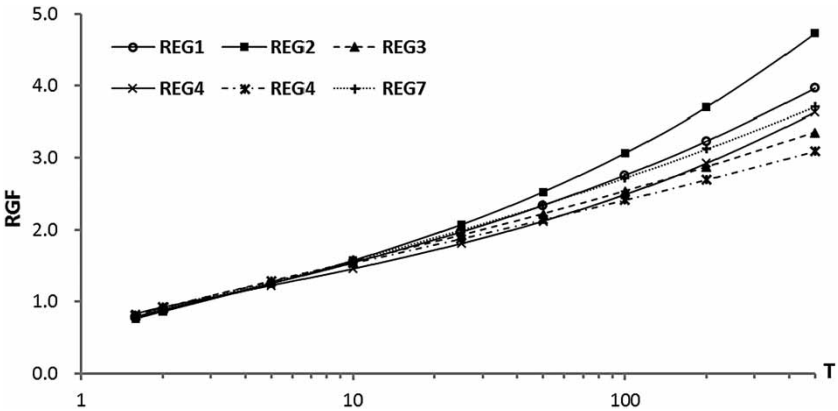

Figure 8 | RGF versus return period (T) for seven sub-regions.

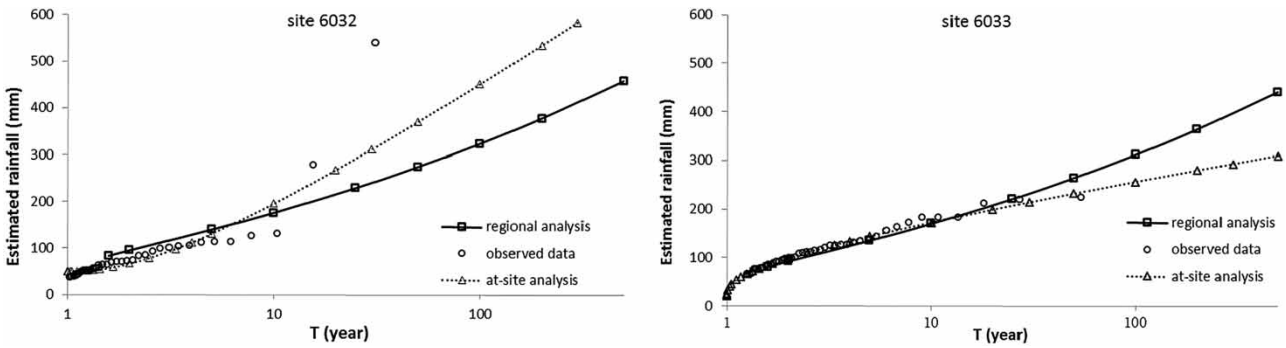

Figure 9 | AMDP frequency analysis results for sites 6032 and 6033 (Region 6, GEV).

The sensitivity test result is summarized in Table 4. It shows that the p value for both sites are <5% when the return period is smaller than 200 years. This indicates that Region 6 can be treated as a homogenous region and the estimations obtained in the previous sections are acceptable.

PDS data analysis

The same procedure of analysis is carried out to test the use of the PDS series instead of the AMDP series. The L-statistics based on the AMDP series and the PDS series for the seven sub-regions are provided in Table 5. Compared

Table 4 | Result of sensitivity tests (p value expressed as percentage) in Region 6

	Return period (year)				
Removed sites (Region 6)	**10**	**25**	**50**	**100**	**200**
6033	0.13	0.05	0.20	0.44	0.67
6032	0.13	0.05	0.20	0.44	0.67
6032 and 6033	0.45	0.69	0.85	0.97	1.11

Table 5 | L-statistics of different regions

	L-C_v	**L-C_s**	**L-C_k**
Region 1	0.23	0.30	0.22
	0.16	0.41	0.25
Region 2	0.25	0.34	0.25
	0.23	0.34	0.26
Region 3	0.23	0.23	0.21
	0.23	0.27	0.24
Region 4	0.20	0.25	0.20
	0.21	0.24	0.18
Region 5	0.22	0.21	0.11
	0.23	0.31	0.19
Region 6	0.24	0.33	0.22
	0.22	0.34	0.26
Region 7	0.23	0.30	0.19
	0.17	0.40	0.22

with the results based on AMDP, the estimated L-C_v values based on PDS are smaller, and the L-C_s and L-C_k values are larger. It demonstrates that the PDS series has a smaller sampling variance and a larger skewness and kurtosis, which results in the increase of estimates for each site in the upper tail of their distributions. Furthermore, the distributions for the seven sub-regions based on the PDS series are different from those based on the AMDP series. It is well-recognized that the best distribution is influenced by property of the data (Wang *et al.* 2011) and the model of GPA distributions using the PDS series is the most acceptable method (Begueria 2005; Trefry *et al.* 2005; Pham *et al.* 2014b). In this study, the GNO distribution is identified as the best distribution for Regions 1 and 2. The GLO distribution is best fitted for Region 4. The GPA distribution is best fitted for four regions: Regions 3, 5, 6, and 7. Therefore, the results obtained in this study verify that the PDS data are most closely fitted by GPA distributions in regional analysis.

The difference in estimates from the use of the AMDP series and the PDS series is also calculated. Treating estimates based on PDS as the true values, the difference in estimates, i.e., estimates based on AMDP minus estimates based on PDS for the same return periods, can be viewed as the bias of the AMDP estimates, which are plotted in Figure 10. It can be seen from Figure 10 that, for return periods <2 years, AMDP estimates are clearly lower than PDS estimates; for return periods >5 years, both positive and negative biases increase with the increase of return periods. For example, for a return period of 500 years, the largest bias reaches 44.61 mm.

CONCLUSIONS

In this study, regional analysis based on L-moment in the Taihu Basin are carried out with the AMDP series and the PDS series, and the performance of AMDP and PDS are investigated. The regional analysis procedure was found to be more effective for extreme precipitation estimation in this basin as compared with at-site analysis. The procedure is therefore recommended for other cases.

It was noted that the discordancy measures D_i of site 6032 and site 6033 do show a slight discordancy. However, after more detailed investigation of the data series and examination of the sensitivity test results, the two sites are still accepted. This verification procedure provides a practical way for engineers to use in the flood design in this area. The L-moment-based heterogeneity measure identifies that the seven sub-regions in the Taihu Basin are acceptably homogeneous even though Region 5 shows a slight heterogeneity in $H(3)$. As the basin comprises different climatic regions, the seven hydrological sub-regions are appropriate for water resources management purposes. The different types of distributions suitable for the seven different sub-regions partly reflect the different climatic conditions.

Robustness and accuracy of regional L-moment analysis is demonstrated through the comparison of at-site analysis and regional analysis results for sites 6032 and 6033. It shows that outliers have a significant influence on the upper tail of a distribution, and regional analysis helps to

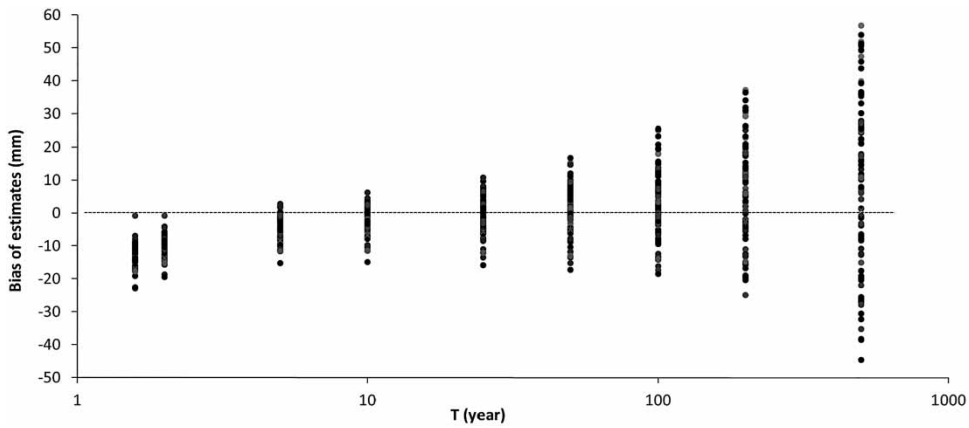

Figure 10 | The bias of AMDP estimates at 96 rain gauges (treating corresponding PDS estimates as true values).

reduce the influence of insufficient data at individual sites. Based on the results of this study, the analysis method may even be used in ungauged areas.

PDS considers all of the extreme values and the use of the PDS series enables a more complete analysis of extreme events than the use of the AMDP series, even with some outliers. For small return periods, estimates based on the AMDP series are lower than those based on PDS. It was found that PDS provides higher quantile estimates for practical purposes in the upper tail of a distribution; and the GPA distribution is suitable for regions in the Taihu Basin. Although the PDS series is still under-used relative to the AMDP series because of its complexity, it is recommended that the PDS should be used for the study of extreme hydrological events. More research based on the PDS series is preferable to describe the extreme hydrological events in the Taihu Basin.

In the mountainous areas of the west part of the Taihu Basin, extreme precipitation is estimated to be greater. Thus, it is suggested that the flood design standard should be updated in this region considering the more frequent occurrence and greater magnitude of precipitation extremes. In addition, this work can be extended to other regions of China.

ACKNOWLEDGEMENTS

This paper is the result of part of a research initiative titled 'Key Analysis for Flood Control in Rural Areas' sponsored by National Science & Technology Pillar Program (Grant No. 2014BAL05B02). This paper also addresses the research topic under 'The Application of Hydrometeorological L-moments in Flood Control Plan', a project which is supported by the Ministry of Water Resources P.R. China (No. 201001047-02). Our greatest gratitude goes first and foremost to Professor Bingzhang Lin, Dean of Applied Hydrometeorological Research Institute of Nanjing University of Information Science & Technology, China. Professor Lin provided us with many valuable suggestions. The authors would like to thank all the reviewers and editors for their helpful comments and suggestions on improving the quality of this paper.

REFERENCES

Atiem, I. A. & Harmancioğlu, N. B. 2006 Assessment of regional floods using L-moments approach: the case of the River Nile. *Water Resour. Manage.* **20** (5), 723–747.

Begueria, S. 2005 Uncertainties in partial duration series modelling of extremes related to the choice of the threshold value. *J. Hydrol.* **303** (1–4), 215–230.

Chen, L. H. & Hong, Y. T. 2012 Regional Taiwan rainfall frequency analysis using principal component analysis, self-organizing maps and L-moments. *Hydrol. Res.* **43** (5), 275–285.

Dalrymple, T. 1960 Flood frequency analyses. *U.S. Geol. Sttrv. Water Supply Paper* 1543–A.

El Adlouni, S., Bobée, B. & Ouarda, T. 2008 On the tails of extreme event distributions in hydrology. *J. Hydrol.* **355**, 16–33.

Gou, H. L., Wu, H. Y. & Liu, S. G. 2010 Primary analysis of natural disasters in Taihu Basin. *China Flood Drought Manage.* **1**, 55–58.

Greenwood, J. A., Landwehr, J. M., Matalas, N. C. & Wallis, J. R. 1979 Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water Resour. Manage.* **15**, 1049–1054.

Hosking, J. R. M. 1990 L-moments: analysis and estimation of distributions using linear combinations of order statistics. *J. Roy. Stat. Soc. Ser B* **52**, 105–124.

Hosking, J. R. M. & Wallis, J. R. 1997 *Regional Frequency Analysis: An Approach Based on L-moments*. Cambridge University Press, Cambridge, UK.

Huang, X. W. 2000 *Basin Planning and Comprehensive Treatment in Taihu Basin*. China Water Power Press, Beijing, China. (in Chinese)

IPCC 2013 *Climate Change 2013: The Physical Science Basis*. Cambridge University Press, Cambridge, UK.

Kjeldsen, T. R. 2010 Modelling the impact of urbanization on flood frequency relationships in the UK. *Hydrol. Res.* **41** (5), 391–405.

Kjeldsen, T. R. & Jones, D. A. 2009 A formal statistical model for pooled analysis of extreme floods. *Hydrol. Res.* **40** (5), 465–480.

Li, J., Dong, W. J. & Yan, Z. W. 2012 Changes of climate extremes of temperature and precipitation in summer in eastern China associated with changes in atmospheric circulation in East Asia during 1960–2008. *Chinese Sci. Bull.* **57** (15), 1856–1861. (in Chinese)

Liang, Y. Y., Liu, S. G., Zhong, G. H., Zhou, Z. & Hu, Y. 2013 Comparison between conventional moments and L-moments in rainfall frequency analysis for Taihu Lake Basin. *J. China Hydrol.* **33** (4), 16–21. (in Chinese)

Lin, B. Z. & John, L. V. 1993 A comparison of L-moments with method of moments. Engineering hydrology. In: *ASCE Proceedings of the Symposium, San Francisco, California, USA*, pp. 443–448.

Lin, B. Z., Yan, G. X. & Zhang, Y. P. 2012 *Hydrometeorology Promotes the Development of Key Project Analysis of Engineering Hydrology Calculation*. New Development of Hydrological Science and Technology in China, Nanjing, China.

Liu, S. G., Zhu, H. L. & Ma, G. F. 2007 Flood frequency study in the lower reach of the Yellow River by regional L-moments analysis method. *Method. Hydrol.* **311**, 257–263.

Madsen, H., Rasmussen, P. F. & Rosbjerg, D. 1997 Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events 1. At-site modeling. *Water Resour. Res.* **33** (4), 747–757.

Ministry of Water Resources of the People's Republic of China. 2006 *Regulation for Calculating Design Flood of Water Resources and Hydropower Projects*. China Water Power Press, Beijing, China.

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** (3–4), 149–166.

Ou, Y. L. & Wu, H. Y. 2001 Great Flood in Taihu Basin in *1999*. China Waterpower Press Beijing, China. (in Chinese)

Parida, B. P., Kachroo, R. K. & Shrestha, D. B. 1998 Regional flood frequency analysis of Mahi-Sabarmati Basin (Subzone 3-a) using index flood procedure with L-moments. *Water Resour. Manage.* **12** (1), 1–12.

Pham, H. X., Shamseldin, A. Y. & Melville, B. 2014a Statistical properties of partial duration series: case study of North Island, New Zealand. *J. Hydrol. Eng.* **19** (4), 807–815.

Pham, H. X., Shamseldin, A. Y. & Melville, B. W. 2014b Statistical properties of partial duration series and its implication on regional frequency analysis. *J. Hydrol. Eng.* **19** (7), 1471–1480.

Sankarasubramanian, A. & Srinivasan, K. 1999 Investigation and comparison of sampling properties of L-moments and conventional moments. *J. Hydrol.* **218** (1–2), 13–34.

Strupczewski, W. G., Kochanek, K., Markiewicz, I., Bogdanowicz, E., Weglarczyk, S. & Singh, V. P. 2011 On the tails of distributions of annual peak flow. *Hydrol. Res.* **42** (2–3), 171–192.

TBA 2008 *Flood Control Planning in Taihu Basin*. Beijing Ministry of Water Resources, Beijing, China. (in Chinese)

Trefry, C. M., Watkins, D. W. & Johnson, D. 2005 Regional rainfall frequency analysis for the state of Michigan. *J. Hydrol. Eng.* **10** (6), 437–449.

Um, M. J., Yun, H., Cho, W. & Heo, J. H. 2010 Analysis of orographic precipitation on Jeju-Island using regional frequency analysis and regression. *Water Resour. Manage.* **24** (7), 1461–1487.

Wang, T. S. 2006 *Flood Control and Water Resource Management in Taihu Basin*. China Water Power Press, Beijing, China. (in Chinese)

Wang, J. Q., Yao, H. M. & Guan, T. S. 2008 Analysis of skewness coefficient of rainstorm and precipitation. *Adv. Water Sci.* **17** (3), 365–370. (in Chinese)

Wang, W., Wang, X. G. & Zhou, X. 2011 Impacts of Californian dams on flow regime and maximum/minimum flow probability distribution. *Hydrol. Res.* **42** (2), 275–289.

Wu, H. Y. 2000a Comparison analysis of plum rain disasters in typical years in Taihu Basin. *J. China Hydrol.* **20** (4), 54–57. (in Chinese)

Wu, T. L. 2000b 1999 Catastrophic flood in Taihu Basin and the consideration for Taihu flood control planning. *J. Lake Sci.* **21** (1), 6–11. (in Chinese)

Wu, H. Y. & Guan, W. Q. 2000 Great Flood in Taihu Basin in *1991*. China Water Power Press, Beijing, China. (in Chinese)

Yang, G. S. & Wang, D. J. 2003 *Economic Development, Water Environment, Water Disaster in Taihu Basin*. Science Press, Beijing, China. (in Chinese)

Yue, S. & Wang, C. Y. 2004 Possible regional probability distribution type of Canadian annual streamflow by L-moments. *Water Resour. Manage.* **18** (5), 425–438.

Zhou, Z. Z., Liu, S. G., Hua, H., Chen, C. S., Zhong, G. H., Lin, H. J. & Huang, C. W. 2014a Frequency analysis for predicting extreme precipitation in Changxing Station of Taihu Basin, China. *J. Coast. Res.* **68**, 144–151.

Zhou, Z. Z., Liu, S. G., Liang, Y. Y. & Lin, H. J. 2014b Correlation analysis and its processing method in Taihu Basin, China. In: *The Eleventh ISOPE Pacific/Asia Offshore Mechanics Symposium*, 12–14 October, Shanghai, China.

First received 31 March 2015; accepted in revised form 25 April 2016. Available online 6 June 2016