

# High Risk Flash Flood Rainstorm Area Mapping And Its Application in Jiangxi Province, China

Hui Ding<sup>a</sup>, Yifan Liao<sup>b</sup>, Bingzhang Lin<sup>c</sup>

School of Hydrology and Water Resources Nanjing University of Information Science & Technology Nanjing, China

**Abstract**—The leading hydrologists around the world have been working hard to develop some kind of preventive measures to reduce the disastrous consequences of a flash flood in advance. For this purpose, a flash flood early-warning and forecasting system that can accurately and timely forecast an coming flash flood has being the research focus in this field, despite its difficulties and complexities. An ideal to specify those areas that are subject at high risk to flash flood in terms of precipitation intensity in a relatively large region is proposed in this paper. It is accomplished through the design of the High Risk Flash Flood Rainstorm Area (HRFFRA) with a certain return period for a given duration based on the application of the end-to-end Regional L-moments Approach to precipitation frequency analysis. A HRFFRA is defined as the area potentially under hitting by higher intense-precipitation for a given duration with certain return period that may cause a flash flood disaster in the area.

An example to develop the HRFFRA has been demonstrated in detail in this paper through the application of the Regional L-Moments Approach to precipitation frequency analysis in Jiangxi Province, South China Mainland. The high risk areas that will be hit by an forthcoming flash flood can be visually showed by the HRFFRA, with its help, hydrologists and governments can substantially reduce the disastrous outcome of a flash flood beforehand.

## 1 INTRODUCTION

A flash flood is a flood that rises and falls quite rapidly with little or no advance warning, usually caused by intense rainfall over a relatively small area. Flash flood and herein induced muddy flow is one of the major rainstorm-related disasters in China. The worst flash flood with muddy flow, caused by extremely heavy rainfall with soil erosion, occurred on 8/8/2010 in Zhouqu, Gansu Province, took 1467 lives with 198 missing. The outstanding flash flood with muddy flow caused by record heavy rainfall during the Molakot Typhoon on 8/8/2009, suffered in xiao-lin-zhuang, Taiwan, took more than 460 lives. The flash flood may also occur after the collapse of a human structure, for instance, the failure of the South Fork Dam located on the Little Conemaugh River 14 miles upstream of the town Johnstown, Pennsylvania, burst through on 5/31/1889, killed more than 2,200 people and caused US\$17 million of damage after several days of extremely heavy rainfall. The most recent catastrophe caused by flash flood across the world occurred in the town Kedarnath, India. With 5000 people approximately dead, it suffered extensive destruction from flash floods triggered by torrential rains during June 2013.

Changes in meteorological conditions and the utilization pattern of land have increased floods frequency across the globe [6]. Moreover, a research of Mazzarana et al. has found that extreme rainfall events which causes extreme flood events will become more and more

frequent across the world in the future [7]. That means we would face a more frequent threats of flash flood. The prevention of flash flood is and will be a tough challenge across the world. Especially for China, since hilly areas and mountains covers as much as two thirds of its land surface [11]. Since the 1980s, researchers in China began to study Flood hazard mapping, but it was not until recently that flash flood hazard mapping was paid attention. It is of significant importance to map and manage the natural hazard causing by flash floods.

The development of the HRFFRA based on statistical analysis of historical rainfall data, synoptic analysis of prevailing storm rainfalls as well as the field survey of historical flash flood events is presented in the paper. A HRFFRA is defined as the area potentially under hitting by higher intense-precipitation for a given duration with certain return period that may cause a flash flood disaster in the area. A Regional L-moments Method (RLMM) is utilized to statistical analysis of rainfall series. Furthermore, the spatial interpolation scheme is used in combination of the contouring technique to form the HRFFRA.

## 2 METHODOLOGY

The frequency analysis of extreme events is often limited by data availability at the desired temporal and spatial scales, while Regional Frequency Analysis (RFA), which involves “trading time for space” by pooling together data for stations with similar statistic characteristics, is an alternative approach for more accurate estimation of

\* Corresponding author: <sup>a</sup>dinghui\_09@nuist.edu.cn, <sup>b</sup>lyifan322@163.com, <sup>c</sup>lbz@nuist.edu.cn

extreme events even in ungauged area or area with short records. RFA does not intend to provide regional estimates [5], it only helps us to get more precise site-specific frequency estimations.

## 2.1 Regional L-Moments Analysis

- Probability distributions and parameter estimation.

Frequency analysis employs a limited data sample to estimate its underlying population by selecting and parameterizing a probability distribution, which is uniquely characterized by a set of parameters. In hydrologic frequency analysis, the number of parameters of a plausible distribution ranges from 2 to 5. Three-parameter distributions, such as Generalized Logistic (GLO), Generalized Extreme Value (GEV), Generalized Normal (GNO), Generalized Pareto (GPA), and Pearson Type III (PE3), behave both relatively reliably and flexibly, are often selected to represent the data. Sometimes, the four-parameter Kappa and the five-parameter Wakeby distributions are used. The parameters of a probability distribution have traditionally been estimated by the Conventional Moments Method (CMM) which can be expressed as:

$$\lambda_r = r^{-1} \sum_{k=1}^{r-1} (-1)^k \binom{r-1}{k} E[X_{r-k:r}], r=1,2, \quad (1)$$

Here,  $\mu$  is the mean of the sample data. The sample estimates based on CMM have some undesirable properties. The higher sample moments associated with skewness and kurtosis can be severely biased, and very sensitive or unstable to the presence of outliers in the data. If the data have a skewed distribution, the selection of the most fitted distribution could be unreliable and the quantile estimates are likely unreliable. Due to the linearity of its moment statistics, the L-moments Method (LMM) has become accepted as a more robust method for selecting and parameterizing fitted probability distribution functions. L-moments are expectations of certain linear combinations of order statistics [1]. Since it's a linear function of the order statistics, L-moments are more robust than conventional moments to the presence of outliers in the data, and provide more robust parameter estimates than the CMM.

Letting  $X_{1:n}, X_{2:n}, \dots, X_{n:n}$  be the order statistics of a random sample drawn from the population of  $X$ , the  $r$ th L-moment can be expressed as:

$$\lambda_r = r^{-1} \sum_{k=1}^{r-1} (-1)^k \binom{r-1}{k} E[X_{r-k:r}], r=1,2, \quad (2)$$

The following example taken from the 743 rain gauges in Jiangxi Province, China, demonstrates how differently the sample coefficient of skewness,  $C_s$ , and the sample coefficient of L-skewness,  $L-C_s$ , behaves in terms of bias when they were used separately to estimate the population parameters. The at-site  $C_s$  and  $L-C_s$  were computed and used as population parameters together with other parameter estimates to generate a plenty of synthetic data via M-C Simulation.  $C_s$  and  $L-C_s$  were calculated for each repeat of the simulation, then averaged over the repeats for each site. These averaged simulated  $C_s$  and  $L-C_s$  were separately compared with

their original values obtained from the real data at each station. As shown in Figs. 1-1~1-2  $L-C_s$  behaves much better than  $C_s$  in terms of bias. Sample  $C_s$  presents severe bias to its population  $C_s$ . The figures shown here used the GEV distribution. The same properties have been observed for other distributions.

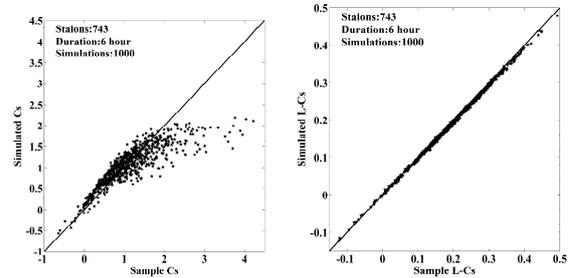


Fig. 1. Real data  $C_s$  vs. simulated  $C_s$  (2) Real data  $L-C_s$  vs. simulated  $L-C_s$

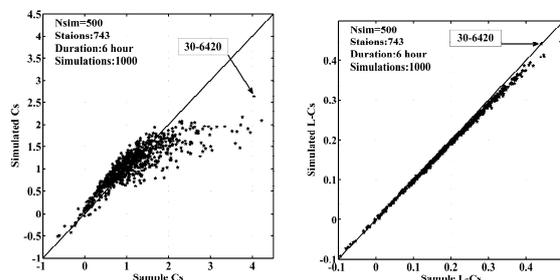


Fig. 2. (1) Sample data  $C_s$  vs. simulated  $C_s$  (2) Sample data  $L-C_s$  vs. simulated  $L-C_s$  at #30-6420

In the original data set, station #30-6420 in Jiangxi Province contains an unusual outlier of 346 mm (7/03/2009) in 6-hour AMS compared with its average 86 mm. The CMM could not handle such an outlier to make a credible estimate of  $C_s$  even with a synthetic data length of 500 years, while the LMM handles it very well. Figs. 2-1~2-2 provide a comparison when the GEV is fitted to the synthetic data. Similar results are obtained for other distributions. Clearly,  $L-C_s$  estimated by L-moments is more robust to the outliers than  $C_s$  estimated through CMM. In fact, the study showed that, when data size reached 100 years at all sites, the  $L-C_s$  estimates are quite stable regardless of outliers present in the data.

- The index-flood procedure

In the NWS updates of precipitation frequency, a Regional L-moments Analysis (RLMA) with a so-called index-flood procedure [2][3] was used. The basis of the index-flood procedure is the assumption that the frequency distributions at the stations in a homogenous region are identical apart from a site-specific scaling factor. The index-flood is a location estimator, usually the at-site sample mean  $\bar{x}_i$ . The stations are assumed to have similar climatology with respect to extreme precipitation. The stations in the predefined homogenous region share a common dimensionless frequency distribution. The frequency values  $q_{Tj}$  at several desired return periods of the dimensionless regional distribution are called Regional Growth Factors. Then the site-specific quantiles  $Q_{Tji}$  can be written as:

$$Q_{Tji} = \bar{x}_i * q_{Tj}, j = 2, 5, \dots, 100, 1000y \quad (3)$$

Table2 illustrates that the RFA is better than the at-site analysis in terms of uncertainties of the quantiles. The comparison was done as follows: the stations with the longest record were selected from the 84 daily regions in the Ohio River Basin that were previously developed and tested to be homogenous for NOAA Atlas 14 Volume 2. Each selected site underwent frequency analysis through at-site analysis and RLMA. Then, a plenty of synthetic data were generated to investigate the variation of the quantiles under the two scenarios. The coefficient of variation,  $C_v$ , was used as an index to describe the uncertainties. Table 2 demonstrates that the  $C_v$  of quantiles in the RFA is much smaller than in the at-site analysis, proved that quantiles estimated through RFA are much more robust than through at-site analysis.

**TABLE I.** UNCERTAINTIES OF AT-SITE ANALYSIS VS. REGIONAL ANALYSIS

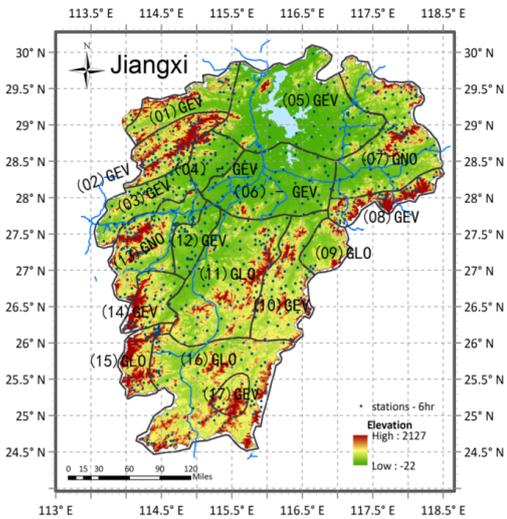
	2y	5y	10y	25y	50y	100y	200y	500y	1000y
At-site	0.031	0.049	0.050	0.073	0.095	0.121	0.015	0.193	0.228
RF A	0.035	0.036	0.037	0.041	0.045	0.050	0.055	0.064	0.072

• Homogenous Regions

The key to RLMA is to construct “homogenous” regions, which assumes frequency distributions at different sites in the region are identical apart from a scale factor. This assumption is rarely valid in the real world, because it requires the same statistics at different sites. However, an acceptable threshold of variation for these statistics can be established to measure the homogeneity. Variation of the  $L-C_v$  over stations in a region,  $V_{L-C_v}$ , is investigated via M-C simulation. The standardized  $V_{L-C_v}$ ,  $H_1$ , serves as an index to measure the homogeneity.

$$H_1 = (V_{L-C_v} - \bar{V}_{L-C_v}) / \sigma_{L-C_v} \quad (4)$$

Hosking suggests the region is “acceptably homogenous” if  $H_1 < 2$ [3]. In our studies this criteria has been found to be appropriate. However, in precipitation frequency studies, we found that sole use of  $H_1$  was not enough. The effect of  $L-C_s$  on the formation of homogenous regions was also considered. Attention was paid to those sites that have too high or too low values of  $L-C_s$  in comparison with the regional averaged  $L-C_s$ . Fig.3 shows the regionalization of 6-hour AMS in Jiangxi Province, with 17 homogeneous regions grouped. Note that the regionalization of the 6-hour data was done based on the 743 rain gauges of 20-years plus inside Jiangxi Province only without creation of a buffer zone surrounding it as data limitation for this study.

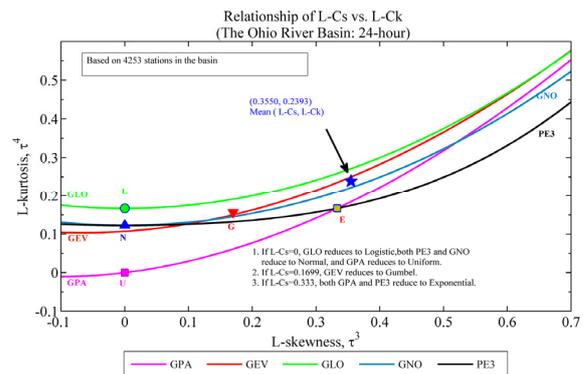


**Fig. 3.** Regionalization of 6-hr data for Jiangxi

• Goodness-of-fit

Three tests were applied to select the best distribution in each region.

The M-C Simulation test. In Fig. 4, the relation curves of the  $L-C_s$  vs.  $L-C_k$  for common theoretical distributions were plotted. The averaged  $L-C_s$  and  $L-C_k$  generated through 1,000 M-C simulations, plotted as a point in the diagram is an example from the Ohio River Basin 24-hour data over 4,253 daily stations. The goodness-of-fit is then judged by the deviation from the mean point to the distributions in the  $L-C_k$  scale. To account for sampling variability, the deviation is standardized, denoted as  $Z_{DIST}$ . For a confidence level, say 90%, a distribution is acceptable if  $|Z_{DIST}| \leq 1.64$ . Among accepted distributions, the distribution with the smallest  $|Z_{DIST}|$  is the most appropriate distribution [1]. Fig. 5 indicated GEV is the most appropriate distribution for the 4,253 stations as a whole in the Ohio River Basin.



**Fig. 4.** M-C simulation test for Ohio daily data stations

Root Mean Square Error (RMSE) of the sample L-moments. RMSE is originated from the Four Criteria test [4] Unlike the M-C simulation test, which emphasizes the effect of mean on deviation, the sample  $L-C_s$  and  $L-C_k$  at all sites are used to assess the variability. A weighted RMSE calculated for each of the plausible distributions serves as an index in the test. The distribution with the smallest RMSE is the most appropriate distribution. In eq.5,  $S_i, L-C_k$  and  $D_i, L-C_k$  are the sample  $L-C_k$  and the distribution’s  $L-C_k$ , respectively.

$$RMSE = \left( \sum_{i=1}^N n_i (S_{i,L-C_k} - D_{i,L-C_k})^2 \right)^{1/2} / \left( \sum_{i=1}^N n_i \right)^{1/2}, i = 1, 2, 3, \dots, N \quad (5)$$

Real-data-check test [4] Comparing a quantile estimate from a fitted distribution at a given return period with the real data series, an empirical exceedance frequency,  $F_{i,T_j}$ , to the quantiles can be calculated, which is then compared with its corresponding theoretical probability  $P_{T_j}$ . A relative error, RE, can be calculated over several return periods to reflect the degree of the match between them. The smaller the RE is, the better the fitting will be. Due to sampling error, the RE for a single site or a few sites is meaningless. However, the regional averaged RE calculated over a number of sites is of statistical significance and can be used as an index for the goodness-of-fit.

$$RE = \sum_{j=1}^N \frac{(F_{i,T_j} - P_{T_j})}{P_{T_j}} \bigg/ \sum_{j=1}^N 1, i=1,2,3,\dots,N, j=2y,\dots,100y \quad (6)$$

A final decision can be made based on a summary of the above three tests. It should be done on a region-by-region basis. The final results provide a strong statistical basis for goodness-of-fit. Sensitivity tests are then used to ensure there is a smooth transition in terms of variation of the RGFs across regions. The best fitted distribution of 6-hour data for each region in Jiangxi is given in Fig. 3 above.

## 2.2 Internal Consistency Check and Adjustment

The precipitation frequency is performed separately at each duration; 1-hr, 3-hr, ..., 24-hr. Sometimes estimate curves for two adjacent durations may meet or cross at some frequency. This result, based on sound statistical analysis, is physically unreasonable. The causes of such an anomaly are primarily discontinuities in selection and parameterization of distributions between durations, data sampling variability, and the application of average conversion factors to convert 1-hr data to 60-min and to convert 1-day data to 24-hr for a specific site. Such anomalies, i.e. internal consistency violations, were removed by distributing the surplus of the ratio of the longer duration vs. the shorter duration of the previous frequency at a constant slope to the anomalous frequency and higher through 10,000-year plus, keeping the pattern of the ratios consistent until it converges at 1.0 after 10,000-year. An example the adjustment at site 30-9660 was shown in Table 2. And the frequency estimate curves for 1-hr, 3-hr and 6-hr before and after the adjustment for site 30-9660 was shown in fig. 5. It is clearly that after the adjustment the anomalies were missing.

TABLE II. INTERNAL CONSISTENCY ADJUSTMENT 6-HR TO 3-HR AT 30-9660

	2-y	5-y	10-y	25-y	50-y	100-y
before	1.12	1.1	1.08	1.06	1.05	1.03
after	1.12	1.1	1.08	1.06	1.05	1.03
	200-y	500-y	1000-y	5000-y	10000-y	
before	1.02	1	0.99	0.95	0.94	
after	1.02	1.01	1.01	1.01	1	

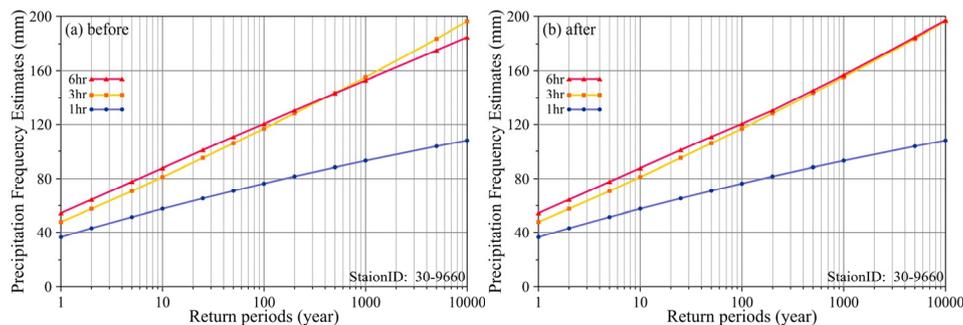


Fig. 5. The frequency estimate curves for 1-hr, 3-hr and 6-hr before (a) and after the internal consistency adjustment, station 30-9660, Jiangxi

## 2.3 Spatial Consistency Check and Adjustment

Due to lack of rainfall stations, there may be gradients appeared in estimated quantiles between regions. An appropriate spatial adjustment to reasonably smooth the quantiles over regions is required. A two-way adjustment called “Two-time back-forth interpolation” scheme is proposed: firstly to develop a computation grid net to cover the study area, with a resolution equivalent in density to the rainfall stations available in the area, as shown in Fig. 6, secondary to spatially interpolate the quantiles at stations into each grids as pseudo stations,

thirdly to interpolate the quantiles at grids back to the stations. A study showed that the RE between the empirical frequencies and the quantiles is reduced after the spatial adjustment on average over the entire study area. Figs. 7-1~7-2 gave a general view of the results for the spatial consistency check and adjustment of the 6-hour data. The quantiles in Fig. 7-2 looks more smooth spatially.

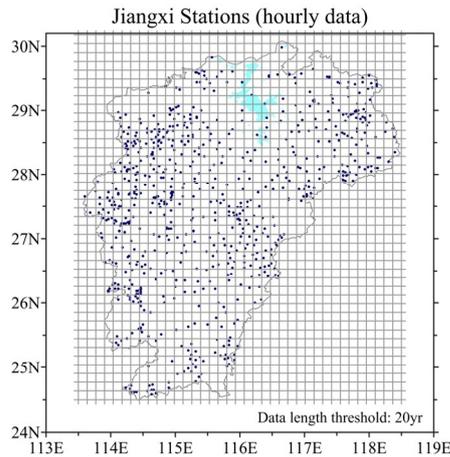


Fig. 6. The computation grid net for spatial adjustment

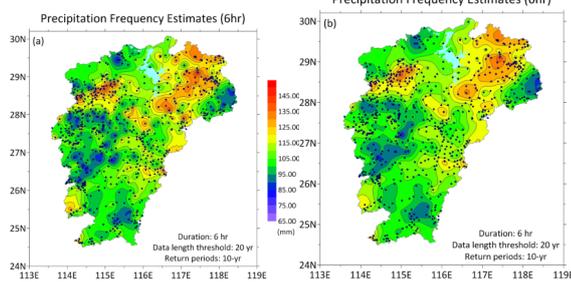


Fig. 7. (a) Before spatial adjustment (b) After spatial adjustment

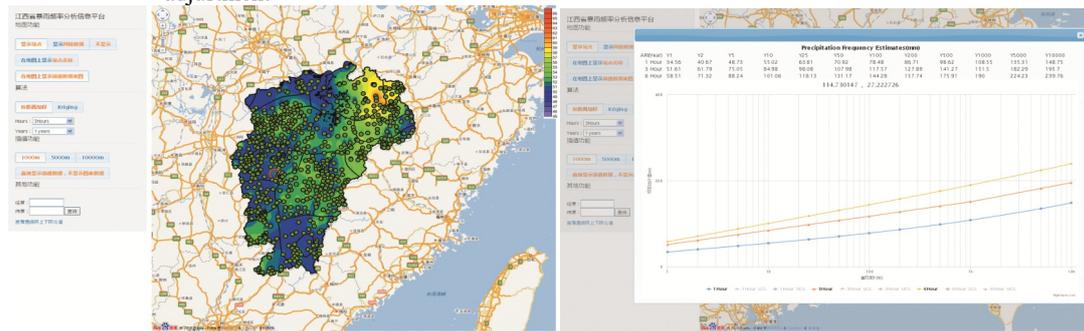


Fig. 8. (upper) The episode showing the study area with stations (lower) The episode showing the quantiles in graph and table

### 3 HIGH RISK FLASH FLOOD RAINSTORM MAPPING

Based on the results of quantiles for various durations such as 1-hour, 3-hour, 6-hour over a range of return periods from 1-yr, 2-yr to 100-yr, 200-yr up to 1,000-yr, an interactive visualization product, a frequency data server, has been developed to get spatial interpolation results for a given duration and for a desired return period. Figs. 8-1~8-2 are the two episodes as examples from the data server. Fig. 8-2 is a pop-up picture obtained instantaneously by moving the cursor showing in Fig. 8-1 as a red circle to a point inside the study area. This visual frequency data server is further employed to get the HRFRA for a given duration and a desired return period.

### 4 CONCLUSIONS

Other aspects such as intersite dependency, confidence limits, ratios of PDS vs AMS have been extensively studied and developed for this study and omitted here for compactness.

Due to the Difficulties and complexities exists in accurate and timely prediction of an incoming threatening flash flood, The HRFRA atlas with short durations, derived through spatial interpolation scheme and contouring technique, based on the quantiles calculated through RFA in combination with LMM, provides a scientific and practical foundation for early-warning and forecasting of flash hazard.

### REFERENCES

- Hosking, J. R. M. (1989). Some theoretical results concerning L-moments, IBM TJ Watson Research Center.

- Hosking, J. and J. Wallis (1993). "Some statistics useful in regional frequency analysis." *Water Resources Research* 29(2): 271-281.
- Hosking, J. R. M., Wallis, J. R. (1997). "Regional Frequency Analysis: An Approach based on L-Moments."
- Lin, B., Vogel J. L. (1993). "A comparison of L-Moments with method of Moments."
- Lin, B., Bonnin, G., M., Martin, D. L., Parzybok, T., Yekta, M., and Riley, D. (2006). *Regional Frequency Studies of Annual Extreme Precipitation in the United States based on Regional L-moments Analysis*. World Environmental and Water Resource Congress 2006@ sExamining the Confluence of Environmental and Water Concerns, ASCE.
- Mahmood, S., Khan A. ul. H., and Ullah, S. (2016). "Assessment of 2010 flash flood causes and associated damages in Dir Valley, Khyber

- Pakhtunkhwa Pakistan." *International Journal of Disaster Risk Reduction* 16: 215-223.
7. Mazzarana, B., Hubl, J., and Fuchs, S. (2009). "Improving risk assessment by defining consistent and reliable system scenarios." *Nat. Hazard Earth Syst* 9: 145-159.
  8. Wallis, J. R. M. (1980). "Risk and uncertainties in the evaluation of flood events for the design of hydraulic structures." *Fondazione Politecnica del Mediterrances, Catania, Italy*: 3-36.
  9. Wallis, J. R. M. (1982). "Hydrologic problems connected with oil-shale development. " *Environmental Systems and Management*: 85-102.
  10. Wallis, J., M., Schaefer, M. G., Barker, B. L., Taylor, G. M. (2007). "Regional precipitation-frequency analysis and spatial mapping for 24-hour and 2-hour durations for Washington State." *Hydrology and Earth System Sciences* 11(1): 415-442.
  11. Zhang, D.W., Quan, J., and Zhang, H. B., et al. (2015). "Flash flood hazard mapping: A pilot case study in Xiapu River Basin, China." *Water Science and Engineering* 8(3): 195-204.