Nonstationarity in the occurrence rate of floods in the Tarim River basin, China, and related impacts of climate indices

Xihui Gu, Qiang Zhang, Vijay P. Singh, Lin Liu

Corresponding author:

Qiang Zhang, Ph.D. Professor

Department of Water Resources and Environment (Department Chairman)

SunYat-sen University

Guangzhou 510275,

China

Tel: +86-20-84113730

Fax: +86-20-84113730

E-mail: zhangq68@mail.sysu.edu.cn (preferred contact address)

Nonstationarity in the occurrence rate of floods in the Tarim River basin, China, and related impacts of climate indices

Xihui Gu^{1,2}, Qiang Zhang^{1,2,3}, Vijay P. Singh⁴, Lin Liu^{3,5}

1. Department of Water Resources and Environment, Sun Yat-sen University, Guangzhou 510275, China;
2. Key Laboratory of Water Cycle and Water Security in Southern China of Guangdong High Education Institute, Sun Yat-sen University, Guangzhou 510275, China;
3. Guangdong Provincial Key Laboratory of Urbanization and Geo-simulation, School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China;
4. Department of Biological and Agricultural Engineering and Zachry Department of Civil Engineering, Texas A&M University, College Station, Texas, USA;
5. Department of Geography, University of Cincinnati, Cincinnati, OH 45221-0131.

Abstract: Amplification of floods in the Xinjiang, China, has been observed, but reports on their changing properties and underlying mechanisms are not available. In this study, occurrence rates of floods in the Tarim River basin, the largest inland arid river basin in China, were analyzed using the Kernel density estimation technique and bootstrap resampling method. Also analyzed were the occurrence rates of precipitation extremes using the POT (Peak over Threshold)-based sampling method. Both stationary and non-stationary models were developed using GAMLSS (Generalized Additive Models for Location, Scale and Shape) to model flood frequency with time, climate index, precipitation and temperature as major predictors. Results indicated: (1) two periods with increasing occurrence of floods, i.e., the late 1960s and the late 1990s with

considerable fluctuations around 2-3 flood events during time intervals between the late 1960s and the late 1990s; (2) changes in the occurrence rates of floods were subject to nonstationarity. A persistent increase of flood frequency and magnitude was observed during the 1990s and reached a peak value; (3) AMO (Atlantic Multidecadal Oscillation) and AO (Atlantic Oscillation) in winter were the key influencing climate indices impacting the occurrence rates of floods. However, NAO (North Atlantic Oscillation) and SOI (South Oscillation Index) are two principle factors that influence the occurrence rates of regional floods. The AIC (Akaike Information Criterion) values indicated that compared to the influence of climate indices, occurrence rates of floods seemed to be more sensitive to temperature and precipitation changes. Results of this study are important for flood management and development of mitigation measures.

Key words: Kernel estimation; GAMLSS; Occurrence rates of floods; Non-stationarity; flooding forecasting; Tarim River basin

1. Introduction

Streamflow is one of the most important components of the hydrological cycle. Recent years have witnessed increasing concern for global warming and impacts of warming temperature on environment and society. The warming climate has the potential to accelerate the hydrological cycle at global and continental scales (Alan et al., 2003; Bates et al., 2008; Zhang et al., 2013a). Observed the changes of extreme weather and climatic events, such as more extreme precipitation, lengthening dry span, higher peak flows, and increased intensity of most extreme tropical cyclones since the 1950s (Zhang et al., 2013b; IPCC, 2013; World Bank, 2013; Apurv et al., 2015) have increased risks of floods and droughts. Increasing occurrences of floods have been

arousing considerable concern in recent years. Caspary and Bárdossy (1995) analyzed yearly maximum runoff time series from the Enz River (SW Germany) at two stations from 1930 to 1994 by making a polynomial regression and fitting the Gumbel extreme value distribution to various time intervals. They showed an increase in flood risk and attributed it to an increase in the occurrence of zonal westerly circulation systems. Jacobeit et al. (2003) analyzed the same flood data and their relation to sea level pressure (SLP) fields and confirmed the relevance of westerly zonal circulation type for the occurrence of winter floods. Mudelsee et al. (2004) hypothesized that anthropogenically induced climate change would add to the risk of extreme river floods because a warmer atmosphere carries more water. Thus, more attentions should be paid to understand the flood processes and occurrences of floods, which will be conducive to reduce the losses caused by floods.

The aim of this study is the investigation of the validity of the stationarity assumption in the occurrence rate of floods. Stationary of a hydrologic time series is often defined as “free of trends, shifts, or periodicity (cyclicality)” (Salas, 1993). Therefore, most studies have check whether this assumption is valid or not based on analyzing slowly varying changes (trend analysis) or change point due to climate change and intensifying human activities, such as construction of water reservoirs and urbanization (Milly et al., 2008; Gilroy and McCuen, 2012; Silva et al., 2012; Zhang et al., 2014a, 2015a). Villarini et al (2009b) performed the flood frequency analysis for nonstationary annual peak records in an urban drainage basin from the point of trend analysis. The stationarity of annual flood peaks was also checked based on trend and change point analysis in the continental United States during the 20th century (Villarini et al., 2009a). However, an element that is often overlooked in analyses of stationarity of flood records is the presence of long-term persistence in the flood series

(Koutsoyiannis, 2002). Koutsoyiannis (2006) pointed out that in particular, some of the patterns observed in hydrologic series could be better explained by accounting for long-term persistence. Because the occurrence rates of floods are not the original flood series but estimated by kernel function (Mudelsee et al., 2003; Mudelsee et al., 2004), the stationarity in occurrence rates of floods may be not suitable to analyze presence of long-term persistence, although the concepts of stationarity and persistence of long-term should be analyzed jointly. In addition, Mudelsee et al (2003, 2004) analyzed the stationarity of occurrence rates of floods in central Europe over the past 500 years without considering the long-term persistence, too. Therefore, we investigate the stationarity in occurrence rates of floods mainly from the way of slowly varying changes.

The accuracy of hydrologic analysis and design may be influenced with violation of the stationarity assumption (Douglas et al., 2000). With this point in mind, Galloway (2011) asked: what do we do now if stationarity is dead? Therefore, he called for not only supporting additional study of the death of stationarity, its implications, and new approaches, but also providing more information to plan, design, and operate today's projects. Clarke (2007) indicated that the next few decades should witness a substantial increase in our understanding of the processes causing climate change, not only for the purpose of forecasting the development of such changes, but also for predicting the frequency of an event of a certain magnitude. Thus, a number of investigators have attempted to incorporate nonstationarity in flood frequency analysis (Silva et al., 2012; Zhang et al., 2014b, 2015b). Reports on occurrence rates of floods in arid regions with the consideration of nonstationarity are lacking. Therefore, our understanding of flooding processes in arid regions, particularly nonstationary flood behavior, under the influence of changes in climate is limited. Studies addressing this issue will help

understand regional hydrological responses of floods to global climate indices and management of water resources and agricultural irrigation, in arid areas such as the Tarim River basin (e.g. Jiang et al., 2005).

The Tarim River basin is the largest arid inland river basin in China. The total area of the basin is $1.02 \times 10^6 \text{ km}^2$ (996000 km^2 within Xinjiang), in which the land is divided into mountains (47%), plains (22%), deserts (31%), and an arable land area of $1.36 \times 10^6 \text{ hm}^2$ (Fan et al., 2011). The source of water in the basin is mainly glacial/snow melt which accounts for 48.2% of the total water volume (Chen et al., 2007). About 10 million people, including ethnic minorities-Uyghurs and Mongolians, live in the basin. More than 8 million people live in oases along the river banks and in alluvial plains downstream (Zhang et al., 2010). The basin is characterized by low precipitation and high evaporation (Zhang et al., 2010; Sun et al., 2012). In lowlands, the river is the most important source of water. Due to its exceptional and unique role in the economic development, the Chinese government launched a five-year emergency water diversion program in 2000 with 10.7 billion yuan (US\$1.3 billion) earmarked for reclamation of the river and Taitema Lake (Tao et al., 2008). However, the number of heavy rainfall events and the frequency of rainstorm flood disasters have been on the rise since the 1980s, and the increasing trend in flood disasters in Xinjiang since the mid-1980s can be attributed to the increasing trend in annual precipitation (Jiang et al., 2005). Observations further indicate that glacier melt-induced floods in the basin have been increasing in recent years (e.g. Zhang, 1992).

Only limited attention was paid to the relations between the occurrence rates of floods and climate indices. This study investigates the spatiotemporal relations between occurrence rates of floods and climate indices in the Tarim River basin with considering nonstationarity. The objectives of this study therefore are: (1) to analyze occurrence

rates of floods with consideration of nonlinearity; (2) to probe into causes behind occurrence rates of floods by building relations between occurrence rates of floods and climate indices; and (3) to investigate at-site probabilistic forecasting of flood frequency with climate indices as major influencing factors by the development of GAMLSS-based model. Results of this study are important for management of water resources and agricultural irrigation under the influence of climate indices.

2. Data

Data collected were streamflow, flood occurrences, precipitation, and temperature. Daily streamflow data of at least 37 years with covering different periods from 8 hydrological stations, shown in Table 1, were obtained from the Hydrological Bureau of the Tarim River basin which firmly controls the quality of data before release. Locations of the hydrological stations are shown in Fig. 1. The flood series in hydrological stations were obtained from Peak-over-Threshold (POT) sampling of daily streamflow data, which were used to analyze the nonstationarity in occurrence rates of floods in stations. Data on occurrences of floods and their timing for a period of 1950-2000 from 5 states, simply called as subregions, i.e., Aksu, Kezhou, Kashi, Hotan and Bazhou, were also collected from *Encyclopedia of Meteorological Disasters in China* (Wen, 2005). Each recorded flood with corresponding occurrence date in each state was collected, which was used to analyze the nonstationarity in occurrence rates of floods in regions. Also, daily precipitation and temperature data covering a period of 1961-2010 were collected from the meteorological stations closest to the hydrological stations (Fig. 1, Table 1).

Climate indices have a significant influence on climate change and hydrological cycle. Ran et al. (2015) indicated that Southern Oscillation Index (SOI) and North

Atlantic Oscillation (NAO) impact climate change and hydrological processes in Xinjiang. Kalra et al. (2013) developed a data-driven model incorporating oceanic atmospheric oscillations to increase the streamflow lead time in the Kaidu River, a major tributary of the Tarim River basin. The interannual and decadal scale oceanic-atmospheric oscillations, i.e., Pacific decadal oscillation (PDO), North Atlantic oscillation (NAO), Atlantic multidecadal oscillation (AMO), and El Niño-southern oscillation (ENSO), were used to generate streamflow volumes for the peak season (April-October) and the water year. In this case, NAO, AO (Atlantic Oscillation), SOI and AMO during winter season (December to March of the subsequent year) were taken as the major climate indices influencing hydrological processes in the basin. Monthly data of NAO, AO, SOI and AMO were extracted from <http://www.esrl.noaa.gov/psd/data/climateindices/list/>, and the period of aforementioned climate indices data was 1950-2010.

3. Methodology

3.1 Peak-over-Threshold (POT) based sampling

Flood series were obtained by two methods, i.e., annual peak flow and POT-based sampling technique. POT sampling helps obtain information on flood magnitude, occurrence rate and timing of flood events. The independence of flood peak was evaluated as (Lang et al., 1999):

$$\begin{cases} D > 5 + \log(A) \\ Q_{\min} < \frac{3}{4} \min(Q_1, Q_2) \end{cases} \quad (1)$$

where D denotes the waiting time between two flood peaks in days; A denotes the drainage area in km^2 ; and Q_1 and Q_2 denote the magnitudes of two flood peaks, respectively, in m^3/s ; $Q_{\min}$ is the minimum discharge between two successive peaks

Q_1 and Q_2 .

The Mean Exceedance above Threshold (MET) technique and Dispersion Index (DI) method were used to determine the range of threshold values. These two methods, together with the mean number of over-threshold flood events (MnT), were used to define the threshold value which satisfied the condition that the mean number of flood events ranged between 2.4-3 per year (Lang et al., 1999). The choice of threshold should satisfy three conditions: (1) MET has a linear relationship with thresholds; (2) In addition, DI with corresponding of thresholds should be at a confidence interval (for example 95% confidence interval in this paper) making POT series following the Poisson distribution; (3) MnT has a number between 2.4 and 3 with corresponding of thresholds. Finally, the biggest value of the thresholds in which satisfy the above three conditions was chose as the terminal threshold. These sampling techniques were also used for precipitation extremes considered in the study.

3.2 Nonparametric estimation of occurrence rates of floods

The kernel density estimation method is used to smooth POT-based flood series. The estimation of occurrence rates of the time-dependent extreme events can be computed as (Mudelsee et al., 2003; Mudelsee et al., 2004):

$$\lambda(t) = \frac{1}{h} \sum_{i=1}^m K\left(\frac{t - T_i}{h}\right) \quad (2)$$

where T_i is the timing of the i th flood event with unit in day; m is the number of floods based on the POT sampling technique; $K(\cdot)$ is the kernel function; and h is the bandwidth of the kernel function. The Gaussian kernel function is the widely-used kernel function, which can use the Fourier space and produce a smoothed estimation of the occurrence rates of extreme events (Mudelsee et al., 2003; Mudelsee et al., 2004):

$$K(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) \quad (3)$$

where $y = (t - T_i)/h$. Occurrence rates of an extreme event, $\lambda(t)$, denote the daily number of an extreme event exceeding threshold values given a certain time interval, t , with a unit of d^{-1} . The occurrence frequency of floods is usually at the annual time scale and hence the occurrence rates of flood, $\lambda(t)$, times 365.25, indicates the frequency of floods exceeding threshold values given time, t . The time interval of POT sampling-based time series is $[t_1, t_m]$. Since no data are available outside of the time interval, i.e., $[t_1, t_m]$, $\lambda(t)$ near the boundaries of the time interval is usually underestimated. In this case, a kind of pseudodata was used to reduce the error as a result of underestimated $\lambda(t)$. A mapping technique was used to produce the pseudodata (Mudelsee et al., 2004). pT is the pseudodata outside of the time interval of $[t_1, t_m]$ for the POT sampling-based series. For $t < t_1$, $pT[i] = t_1 - [T_i - t_1]$; and the same procedure was done for $t > t_m$. The extended series was 1.5 times longer than the original one. The computation of $\lambda(t)$ based on the extended data series was based on:

$$\lambda(t) = \frac{1}{h} \sum_{i=1}^{m^*} K\left(\frac{t - T_i^*}{h}\right) \quad (4)$$

where T_i^* is the timing of the i th flood event based on the extended data series with a unit in day; m^* is the sample size of the extended data series. Also, the selection of bandwidth, h , is important for the estimation of $\lambda(t)$. Too small bandwidth, h , selected for computation of $\lambda(t)$ will substantially influence the randomness on $\lambda(t)$; too large bandwidth, h , may cause over smoothing of the data series and hence details in information may be excluded. The cross validation method was used to determine the width of window (Mudelsee et al., 2003). For analysis of the annual change in $\lambda(t)$ without considering seasonal changes within the year, the width of window should be

longer than 365.25 days.

3.3 Confidence interval by Bootstrap technique

Bootstrap technique and Equation (4) can be combined to enable uncertainty analysis of the occurrence rate of floods, $\lambda(t)$, using the following procedure (Mudelsee et al., 2004):

- (1) Based on the extended data series, T_i^* , the simulated T^+ of the same series length can be obtained using the bootstrap technique;
- (2) The occurrence rates, $\lambda^+(t)$, of sample extreme events, T^+ , can be computed using Equation (4);
- (3) Steps (1) and (2) above will be repeated for 2000 times, and $\lambda^+(t)$ of 2000 samples can be obtained;
- (4) The 90% confidence interval for $\lambda^+(t)$ will be obtained using the quantile method.

3.4 Poisson regressive method

Based on T_{ij} , i.e., the timing of the j th flood event during the i th year, the annual frequency of an extreme event can be computed by:

$$N_i(t) = \sum_{j=1}^{n_i} 1(T_{ij} \leq t) \quad (5)$$

where t ranges between $[0, 365]$ or $[0, 366]$. The occurrence rates of extreme events usually follow the Poisson distribution:

$$P(N_i = k | \Lambda_i) = \frac{e^{-\Lambda_i} \Lambda_i^k}{k!} \quad (k = 0, 1, 2, \dots) \quad (6)$$

where N_i is a nonnegative random variable. The response variable of the Poisson regression is the occurrence rates of floods. In Poisson regression, the relation between Λ_i and covariate is:

$$\Lambda_i = \exp\left[\beta_0 + \beta_1 h_1(x_{1,i}) + \beta_2 h_2(x_{2,i}) + \dots + \beta_n h_n(x_{n,i})\right] \quad (7)$$

where $(x_{1,i}, x_{2,i}, \dots, x_{n,i})$ is the vector of covariates, for example NAO, AO, SOI and AMO in this study, and $(\beta_0, \beta_1, \dots, \beta_n)$ is the coefficient vector of covariates.

GAMLSS (Generalized Additive Models for Location, Scale and Shape) has been widely used for the development of regressive relations between response variable and covariate and specifically the identification of the significant covariate factor (Rigby and Stasinopoulos, 2005; Zhang et al., 2014a, 2014b, 2015a; Villarini et al., 2010, 2012; López and Francés, 2013). Selection of the covariate variable is critical for the analysis of physical causality between covariate variable and response variable. The hydrological stations considered in this study, except Alar station, are also located at the outlet of the mountainous drainage area and hydrological processes are influenced by climate change. In this case, NAO, AO, SOI and AMO in winter are accepted as the main climate indices which potentially influence runoff changes in the Tarim River basin.

Researches have corroborated that the aforementioned climate indices have definitive impacts on climate change and hydrological variations across Xinjiang, and also in the Tarim River basin (Kalra et al., 2013; Ran et al., 2015). Generally, the flood season of the Tarim River basin is mainly during May-September and taking winter climate indices as the predictor for short-term forecasting of occurrence rates of floods will be of considerable merit for management and mitigation of flood events. Besides, floods in Xinjiang are mainly the result of precipitation, temperature and melt snow/ice. Rainstorm-induced floods, temperature-induced floods and mixed type floods (floods due to combination of precipitation and temperature) account for 24%, 39% and 34% of total floods occurring in Xinjiang (Wen et al., 2005). Annual total precipitation, frequency of precipitation extremes and seasonal average temperature in winter are

taken as the main influencing factors impacting water resources of Tarim River basin. Seasonal average temperature in winter is taken as one of the major influencing factors because the increase of winter temperature can significantly increase the annual mean temperature (Zhang et al., 2009).

Therefore, four models are assessed in this study: Λ_i is constant (stationary model, Model 0); Λ_i with t as the covariate and changes with t (nonstationary model, Model 1); Λ_i with climate indices as covariates and the climate indices in this study are NAO, AO, SOI and AMO in winter) (nonstationary model, Model 2); Λ_i with annual total precipitation, frequency of precipitation extremes (POT-based sampled precipitation extreme series), and winter mean temperature as the covariate variable (Model 3). The cubic spline function was used as a smoothing function $h_n(x_{n,i})$, and AIC (Akaike information criterion) was accepted in this study as the criterion for the selection of optimal model and the significant covariate. Because of lacking a statistic way to evaluate the goodness of fit of the selected optimal models as a whole, analyzing the normality and independence of the residuals of each model is an effective verification way. Therefore, according to the previous studies (Rigby and Stasinopoulos, 2005; Villarini et al., 2010, 2012; López and Francés, 2013; Zhang et al., 2014a, 2014b, 2015a;), the Filliben coefficient were examined, and a visual inspection of diagnostic plots of the residuals (worm plots) was made.

4. Results and discussion

4.1 Changing properties of POT-based sampled flood and precipitation extremes

Based on the MnT and related curves, MET and related threshold curves, and DI and related curves (Zhang et al., 2016), the thresholds for floods and extreme precipitation events at each station are calculated, respectively. Hence, POT-based

sampling was done for daily streamflow and precipitation data to obtain the timing and frequency of extreme precipitation and streamflow events. Besides, the 5-year moving average and loess function were used to smooth and fit the POT-based sampled flood and extreme precipitation series for investigating the changing properties of these extreme events. Meanwhile, the modified Mann-Kendall test (Daufresne et al., 2009) was applied to detect possible trends in the 5-year moving average series of flood and extreme precipitation. Fig. 2 shows changes in flooding frequency. It can be depicted from Fig. 2 that floods at almost all hydrological stations in the Tarim River basin experienced two increasing periods, i.e., about the mid-1960s and about the late 1990s. Further, floods were fluctuating with a magnitude of 2-3 flood events during the 1970s and the 1980s. Furthermore, the loess curve shows that after 21st century flood events witnessed a downward tendency. However, the Dashankou station is an exception which is dominated by the increasing frequency of floods after 2000. Trends by the Modified Mann-Kendall trend test indicate that flood frequency at the Tongguziluohe, Xiehela and Alar stations is decreasing and the decreasing trends at the Xiehela and Alar stations are statistically significant at the 95% confidence level. A significant increasing flooding frequency can be observed at other stations except Kaqun station. Flood frequency in 5 subregions across the Tarim River basin was also analyzed (Fig. 3). Similar changing properties of flood frequency can be identified in the Aksu, Kezhou and Kashi regions, i.e., the occurrences of floods concentrated during the 1960s and the 1990s. An abrupt increase of flood frequency was observed after 1980 and then a slow decrease after 1995. Flood events were scarce before 1980 in the Hotan region but became frequent after 1980. Flood frequency in the Bazhou region boomed during the mid-1950s, and during 1980-2000. Generally, the period of 1950-2000 witnessed an amplification of floods in five subregions considered in the study (Fig. 3) and the

increasing trend was statistically significant at 95% confidence level. The nonstationarity of flood frequency in the Tarim River basin was statistically significant and posed new challenges for management of floods and mitigation of flood hazards.

Analysis of changing properties of precipitation extremes helps understand changing features of floods and was also done. Fig. 4 shows that precipitation extremes occurred mainly in two periods, i.e., the late 1960s and the late 1990s. Meanwhile, the frequency of precipitation extremes reached a peak value during the late 1990s. Higher fluctuations of precipitation were observed during 1970-1990, and frequency of precipitation extremes decreased after 2000. Visual comparison between Figs. 3, 4 and 5 indicates a strong influence of precipitation extremes on the occurrence of floods.

POT-based sampled series were analyzed in terms of the parameters of occurrence rates of floods that followed the Poisson process. The events will follow a stationary Poisson process when they occur independently with fixed augmentation, i.e., occurrence rate of extreme events during a certain period depends on the length of the time interval only. Hence, parameter, λ , of the occurrence rate was constant. When parameter, λ , of occurrence rate is a function of time, t , then the Poisson process is nonstationary. When the Poisson process is nonstationary, the variable $w_i = (T_i - t_1) / (T_m - t_1)$ (standardized POT-based sampled series) follows a uniform distribution in interval $[0, 1]$. The Kolmogorov-Smirnov (KS) goodness-of-fit test method was used to test whether the empirical distribution function, $\hat{F}(w_i)$, of w_i followed the uniform distribution $F(w) = w$, wherein $0 \leq w \leq 1$, which was used to evaluate whether the time series considered was nonstationary or not. It can be seen from Table 2 that the KS_D statistics for Yuzimenleke, Shaliguilanke and Huangshuigou stations are significant at 0.1 significance level and the POT-based sampled series is nonstationary; the KS_D statistics for Kaqun, Xiehela and Alar

stations are close to 0.1 significance level and the occurrence rate is nonstationary. The KS_D statistics of the POT-based sampled extreme precipitation series for Tongguziluoke, Kaqun, Shaliguilanke, Xiehela and Dashankou stations were larger than the threshold value of 0.1, implying that the occurrence rate of extreme precipitation events was also nonstationary. It should be noted here that the KS_D statistics of occurrence rates of flood events in all the subregions considered in this study were larger than the threshold value of 0.1, showing strong nonstationarity in the regional occurrence rates of floods. These results justify the application of nonstationary models in subsequent analysis.

4.2 Occurrence rates and nonstationarity of flood events

The Kernel density estimation technique and bootstrap resampling method were used to estimate the occurrence rates of extreme events, $\lambda(t)$, and the uncertainty based on the POT-based sampled timing of the flood and extreme precipitation events. Fig. 5 shows temporal variations of occurrence rates of flood events (FOR), $\lambda(t)$, at the hydrological stations considered in the Tarim River basin. Besides, Fig. 5 also shows the estimated boundary error of $\lambda(t)$ based on pseudodata. It can be seen from Fig. 5 that the occurrence rates of floods at different hydrological stations show slightly different changing properties. The occurrence rates of floods at the Xiehela, Alar and Tongguziluoke stations generally tended to decrease. However, the increasing tendency of occurrence rates of floods was observed at other hydrological stations, i.e., Yuzimenleke, Kaqun, Shaliguilanke, Huangshuigou, Dashankou. Nevertheless the occurrence rates of floods at all the hydrological stations shared similar changing properties. Two periods were identified with increasing occurrence rates of floods, i.e., the 1960s and the 1990s or the 2000s, with slight shifts of periods. The occurrence rates

tended to decrease after 2000. Moreover, the occurrence rates of floods at the Tongguziluoke station exhibited moderate changes but those at other stations showed a strong nonstationary Poisson process.

A greater nonstationarity was observed in the regional variations in the occurrence rates of floods. Fig. 6 shows similar changing properties of occurrence rates of floods in the five subregions across the Tarim River basin. The period of 1950-1970 witnessed decreasing occurrence rates of floods and the trough values of occurrence rates of floods were found in the early 1970s. An abrupt increase of occurrence rates of floods was observed after the early 1970s and this increase continued till the early 1990s. The peak values of occurrence rates of floods were detected during the early to mid-1990s. The occurrence rates of floods decreased after about the mid-1990s. From both the viewpoints of flood processes at hydrological stations or regional flood changes, similar changing properties of occurrence rates of floods were evident in the Tarim River basin. Two periods, the early 1960s and the early and the mid-1990s, had increasing occurrences of floods and decreasing occurrence rates between these two periods. Amplification of floods has also been observed in recent years, though the magnitudes of decrease in the occurrence rates of floods have been different in recent years, especially after about 2000. Floods in the Tarim River basin are caused mainly by precipitation changes (precipitation-induced floods) or melting of snowpack and glacier caused by increasing temperature (shorthand for temperature-induced floods) or both (shorthand for precipitation-temperature mixed floods). Precipitation-induced floods, temperature-induced floods, and precipitation-temperature mixed floods account for 24%, 39% and 34% of total floods occurring in the basin (Wen, 2005). The occurrence rates of precipitation extremes were also analyzed (Fig. 7). Comparison between Figs. 6 and 8 shows similar changing features of floods and precipitation extremes at the

Yuzimenleke, Shaliguilanke and Huangshuigou stations, indicating correlations between precipitation extremes and floods in the Tarim River basin. However, relatively different changing properties were found for precipitation extremes and floods at the Tongguziluoke, Kaqun, Xiehela and Dashankou stations, and floods at these stations were also influenced by other factors than precipitation, such as temperature and bursts of dams.

4.3 Relation between occurrence rates of floods and climate indices

The most hydrological stations considering in the study are located at the outlets of the mountainous drainage basins, which seldom affected by human activities, such as farming activities, construction of the reservoirs, and so on (Fig. 1). In addition, the mountainous drainage basins are mainly covered by snowpack and glacier with high elevations (Fig. 1). Moreover, Sun et al (2013) pointed out that the incidence of extreme hydrological events was greatly affected by weather systems and terrain features in Xinjiang. Therefore, the occurrence of floods may be mainly induced by climate change (Wang et al., 2015; Zhang et al., 2016). Climate indices in a specific tributary of the Tarim River basin, e.g. Kaidu River basin (Kalra et al., 2013), were taken as the major predictor of hydrological variations. Recently, Wang et al (2015) found that the winter AO and NAO showed significant influence on streamflow means and extremes in the Tarim River, and indicated that the teleconnection between climate indices and streamflow and extremes can be utilized in a model to predict the probability of hydrological indices in the Tarim River. Therefore, we further investigate the relation between occurrence rates of floods and climate indices in winter, such as AO, NAO, SOI and AMO (Kalra et al., 2013; Ran et al., 2015). Moreover, we try to conduct at-site probabilistic forecasts of occurrence rates of floods with climate indices as

covariates.

For the quantification of relation between climate indices in winter and occurrence rates of floods, GAMLSS model was developed to identify the principal climate index that had the most significant impact on the occurrence rates of floods in the basin. Four categories of GAMLSS models were developed to explore the relations between occurrence rates of floods and covariate variables: stationary model (Model 0) and nonstationary models (Models 1, 2 and 3). Table 3 lists the AIC values, Filliben coefficients and function relations between the occurrence rates of floods and climate indices. In Model 3 (Table 3), x_1 , x_2 and x_3 denote, respectively, the occurrence rates of POT-based sampled precipitation extremes, annual mean precipitation amount, and daily mean temperature in winter season. It can be seen from Table 3 that in Model 2, AMO was the most optimal covariate variable at 4 hydrological stations, followed by AO. In Model 3, however, annual mean precipitation amount and daily mean temperature in winter season were the optimal covariate variables at 4 hydrological stations. The AIC value for Model 1 with time, t , as the covariate variable was the largest, showing that the performance of Model 1 was not statistically satisfactory. However, the AIC values for Models 2 and 3 with precipitation and temperature as covariate variables were smaller than the AIC value for Model 0, the stationary model, indicating that the performance of models can be improved if the factors influencing the occurrence rates of floods were included in the model, which can help understand physical mechanisms underlying the occurrence rates of floods.

The Filliben coefficients (Table 3) of all the models were higher than 0.95, showing acceptable performance of these models. Worm plots also show that these models are acceptable (Fig. 8). Because there are no strong distinctions among models, we only displayed the worm plots of YZ (Yuzimenleke) station as an example (Fig. 8).

Fig. 9 shows results of Model 1 and Model 2 (only displaying the stations with bigger differences between different models). It can be seen that Model 1 can roughly explain temporal variations of occurrence rates of floods at the hydrological stations. Results of Model 1 show that the occurrence rates of floods increased after 1990. However, the 5% and 95% quantile curves of the occurrence rates of floods before 1990 were almost straight lines and did not reflect the random variation of the occurrence rates. Fig. 9 also shows that Model 2 captured the fluctuations of occurrence rates of floods and dispersion (only displaying the stations with better fluctuations in model 3). Considering the changes in the occurrence rates of floods at the Tongguziluoke station as an example, the 5% and 95% quantile curves of the occurrence rates of floods were increasing generally with fluctuations. Besides, the median curves of Model 2 also describe the fluctuations of occurrence rates, showing a better performance than Model 1. Fig. 10 shows that Model 3 performed better in describing fluctuations and dispersion of occurrence rates than Model 2. The AIC values for Model 3 were the smallest amongst these four categories of models which further corroborated the satisfactory performance of Model 3. Particularly, Model 3 better modeled the occurrence rates of floods at the Yuzimenleke, Shaliguilanke and Huangshuigou stations. Increasing occurrence of floods were identified during the 1970s and the 1990s, as described earlier.

In a similar vein, 3 categories of GAMLSS models, except Model 3, were developed for exploring relations between occurrence rates of regional floods and covariate variables. Table 4 lists the AIC values, Filliben coefficients and function relations between occurrence rates of regional floods and covariate variables. It can be seen from Table 4 that NAO and SOI were the factors most significantly influencing the occurrence rates of regional floods in the five subregions considered in the study. The

influences of AO and AMO on the occurrence rates of regional floods were not statistically significant. The AIC values for Model 1 and Model 2 were greater than those for stationary Model 0, implying unequivocal nonstationarity in the occurrence rates of regional floods, and these results can be used in the forecasting of occurrence rates of regional floods in the Tarim River basin. Similarly, worm plots shown also indicate satisfactory modelling performance of the three categories of models, which are not displayed here (the similar as Fig. 8). Table 4 shows that the Filliben coefficients of these models were larger than 0.95, implying satisfactory performance. Fig. 3 shows the results of Model 1 and Model 2. The quantile curves and the median curves of Models 2 and 3 show that the quantile and median curves of Model 3 exhibit frequent fluctuations, describing changes in the occurrence rates of regional floods. However, the quantile curves and the median curves of Models 2 only approximately describe the general changing tendency, but not details of changes in the occurrence rates of regional floods. Atmospheric circulation patterns have a direct impact on regional and global precipitation and temperature changes, and hence indirectly impact regional flooding. However, precipitation and temperature changes can directly influence flooding, and it is particularly true in the Tarim River basin. Therefore, flooding in the Tarim River basin is more sensitive to precipitation and temperature changes. Results of this study show that NAO and SOI are two factors that significantly influence the occurrence rates of floods at the hydrological stations as well as the occurrence rates of regional floods. Thus, NAO and SOI can be taken as predictors for short-term forecasting of occurrence rates of floods in the Tarim River basin. Besides, precipitation and temperature can also be taken as predictors for forecasting of occurrence rates of floods. The outputs of GCMs based on various climate scenarios can be used for long-term forecasting of occurrence rates of floods, providing a reference for management of water resources

and agricultural irrigation. Of course, forecasting results of occurrence rates of floods based on GCMs output may be subject to large uncertainty. However, it is more important to be roughly right, than precisely wrong (Maia and Meinke, 2010).

4.4 Probabilistic forecasting occurrence rates of floods

Probabilistic forecasting of flood behavior is of great importance for flood management and mitigation. Taking climate indices as covariate variables, forecasting of the occurrence rates of floods was done at the hydrological stations and of the occurrence rates of regional floods for the subsequent 15 years. With the identified significant climate indices as the covariate variables, the nonstationary model captured the changing properties of occurrence rates of floods at the hydrological stations, as shown by quantile curves and median curves in Fig. 11. It can be seen that the quantile curves reflected the amplification of floods after the 1990s and the fluctuations and dispersion of floods were captured satisfactorily. Fig. 11 also shows that the 5% and 95% quantile curves described all data points of the occurrence rates of floods, implying that nonstationary models with climate indices as the covariate variables satisfactorily modeled the occurrence rates of floods. We also model the the occurrence rates of regional floods during 1986-2000 by taking the period of 1950-1985 as the model validation period using the nonstationary model developed with climate indices as the covariate variables, which are not displayed here. Like Fig. 11, the nonstationary model satisfactorily captured the fluctuation, dispersion and even extreme values of occurrence rates of regional floods. However, compared to the results shown in Fig. 11, more evident amplification of regional floods can be identified after the 1980s. Moderate changes of the occurrence rates of regional floods were observed before the 1980s. Therefore, climate indices, such as those in winter, can be taken as appropriate

predictors for forecasting of flood behavior.

5. Conclusions

Occurrence rates of floods and precipitation extremes were obtained using the POT-based sampling technique which maintains the useful information of occurrence rates and timing of extreme events. Nonstationarity and uncertainty were evaluated using the kernel density estimation method and bootstrap resampling technique. GAMLSS was developed to determine relations between timing and occurrence rates of floods, climate indices, precipitation and temperature, and major influencing factors were identified. The following conclusions are drawn from this study:

(1) Two periods with increasing occurrences of floods have been identified, i.e., the late 1960s and the late 1990s, though these two periods shift little from one hydrological station to another. The time interval between these two periods witnessed moderate and lower occurrence rates of floods. Changing properties of regional floods are similar those of floods at the hydrological stations. However, increasing magnitudes of occurrence rates and increased number of floods are far higher than those observed at the hydrological stations. Analysis of occurrence rates of precipitation extremes partly match with the occurrence rates of floods, implying an influence of changes in precipitation extremes on the occurrence of floods in both space and time across the Tarim River basin, and other influencing factors, such as snow melt due to temperature increase, cannot be ignored.

(2) Occurrence rates of floods in the Tarim River basin seem to follow nonstationary Poisson process the Tarim River basin, e.g. occurrence rate changes of floods at the Yuzimenleke, Shaliguilanke and Huangshuigou stations. However, more nonstationarity is observed in the occurrence rates of regional floods, though similar changing properties of floods are identified at the hydrological stations and for regional

floods. A persistent increase of occurrence rates of regional floods has been detected after the 1990s, implying even higher occurrence rates of floods with stronger intensity of regional floods than station-based observed changes of floods.

(3) GAMLSS models considering climate indices or precipitation and temperature capture critical changing properties of occurrence rates of floods. Results indicate that AMO and AO in winter are the key climate indices influencing the occurrence rates of floods. However, NAO and SOI are two principle factors influencing the occurrence rates of regional floods. The AIC values indicate that compared to the influence of climate indices, the occurrence rates of floods seem to be more sensitive to temperature and precipitation changes.

(4) Nonstationary models including influencing factors such as climate indices, temperature, and precipitation as predictors capture the major changing properties of occurrence rates of floods with acceptable forecast accuracy. Besides, dispersion characteristics and fluctuations of the occurrence rates of floods can also be well explained by the nonstationary models, showing their feasibility and applicability in modelling and forecasting of flooding behavior.

Acknowledgments: This work is financially supported by the National Science Foundation for Distinguished Young Scholars of China (Grant No.: 51425903), the Natural Foundation of Anhui Province (Grant No.: 1508085MD65), Project supported by the Funds for International Cooperation and Exchange of the National Natural Science Foundation of China (Grant No. 51210013), and is fully supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CUHK441313). Detailed information such as data can be obtained by writing to the corresponding author at zhangq68@mail.sysu.edu.cn. Our cordial gratitude should be extended to the editor, Prof. Dr.

S.A.P. L. Cloetingh, and reviewers for their pertinent and constructive suggestions and comments which are greatly helpful for further improvement of the quality of this manuscript.

References:

- Alan, D.Z., Justin, S., Edwin, P.M., Bart, N., Eric F.W., and Dennis P.L., 2003. Detection of intensification in global- and continental-scale hydrological cycles: temporal scale of evaluation. *Journal of Climate* 16, 535-547.
- Apurv, T., R. Mehrotra, A. Sharma, M.K. Goyal, S. Dutta, 2015. Impact of climate change on floods in the Brahmaputra basin using CMIP5 decadal predictions. *Journal of Hydrology* 527, 281-291.
- Bates, B., Kundzewicz, Z., Wu, S., and Palutikof, J.: Climate change and water, Tech. rep., Intergovernmental Panel on Climate Change, 2008.
- Caspary, H.J., and A. Bárdossy, 1995. Markieren die Winterhochwasser 1990 und 1993 das Ende der Stationarität in der Hochwasserhydrologie infolge von Klimaänderungen? *Wasser Boden*, 47(3), 18-24.
- Chen, Y.N., Li, W.H., Xu, C.C., Hao, X.M., 2007. Effects of climate change on water resources in Tarim River Basin, Northwest China. *Journal of Environmental Sciences* 19, 488–493.
- Clarke, R., 2007. Hydrological prediction in a non-stationary world. *Hydrology and Earth System Sciences* 11, 408-414.
- Daufresne, M., Lengfellner, K., Sommer, U., 2009. Global warming benefits the small in aquatic ecosystems. *Proc. Natl. Acad. Sci. U.S.A.* 106(31), 12788-12793.
- Fan, Y., Y. Chen, W. Li, H. Wang, X. Li, 2011. Impacts of temperature and precipitation on runoff in the Tarim River during the past 50 years. *Journal of Arid*

Land 3(3), 220-230.

Douglas, E.M., R.M. Vogel, C.N. Kroll., 2000. Trends in floods and low flows in the United States: impact of spatial correlation. *Journal of Hydrology* 240, 90-105.

López, J. and Francés, F., 2013. Non-stationary flood frequency analysis in continental Spanish rivers, using climate and reservoir indices as external covariates. *Hydrol. Earth Syst. Sci.* 17, 3189–3203.

Galloway, E.G., 2011. If stationarity is dead, what do we do now? *Journal of The American Water Resources Association* 47(3), 563-570.

Gilroy, K.L. and McCuen, H.R., 2012. A nonstationary flood frequency analysis method to adjust for future climate change and urbanization. *Journal of Hydrology* 414-415, 40-48.

IPCC, 2013. Summary for policymakers. In: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (Eds.), *Climate Change 2013: The Physical Science Basis, Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Jacobeit, J., R. Glaser, J. Luterbacher, and H. Wanner, 2003. Links between flood events in central Europe since AD 1500 and large-scale atmospheric circulation modes, *Geophysical Research Letters* 30(4), 1172, doi:10.1029/2002GL016433.

Jiang, F., C. Zhu, G. Mu, R. Hu, Q. Meng, 2005. Magnification of flood disasters and its relation to regional precipitation and local human activities since the 1980s in Xinjiang, northwestern China. *Natural Hazards* 36, 307-330.

Kalra, A., Li, L., Li, X., Ahmad, S., 2013. Improving streamflow forecast lead time using oceanic-atmospheric oscillations for Kaidu River basin, Xinjiang, China.

- Journal of Hydrologic Engineering 18(8), 1031-1040.
- Koutsoyiannis, D., 2002. The Hurst phenomenon and fraction Gaussian noise made easy. *Hydrological Sciences-Journal*, 47(4), 573-59.
- Koutsoyiannis, D., 2006. Nonstationarity versus scaling in hydrology. *J. Hydrol.*, 324, 239–254.
- Lang, M., Ouardab, T.B.M.J., Bobee, B., 1999. Towards operational guidelines for over-threshold modeling. *Journal of Hydrology* 255, 103-117.
- Maia, A.H.N. and Meinke, H., 2010. Probabilistic methods for seasonal forecasting in a changing climate: Cox-type regression models. *International Journal of Climatology* 30, 2277-2288.
- Milly, P.C.D., J. Betancourt, M. Falkenmark, R.M. Hirsch, Z.W. Kundzewicz, D.P. Lettenmaier, and R.J. Stouffer, 2008. Stationarity Is Dead: Whither Water Management? *Science* 319(1), 573-574.
- Mudelsee, M., Börngen, M., Tetzlaff, M., Grünewald, U., 2003. No upward trends in the occurrence of extreme floods in central Europe. *Nature* 425, 166-169.
- Mudelsee, M., M. Börngen, G. Tetzlaff, and U. Grünewald, 2004. Extreme floods in central Europe over the past 500 years: Role of cyclone pathway “Zugstrasse Vb”. *Journal of Geophysical Research* 109, D23101, doi:10.1029/2004JD005034.
- Rigby, R.A., Stasinopoulos, D.M., 2005. Generalized additive models for location, scale and shape. *Applied Statistics* 54(3), 507-554.
- Silva, A.T., M.M. Portela, and M. Naghettini, 2012. Nonstationarities in the occurrence rates of flood events in Portuguese watersheds. *Hydrology and Earth System Sciences* 16, 241-254.
- Sun, G., Chen, Y., Li, W., Pan, C., Li, J., Yang, Y., 2013. Spatial distribution of the extreme hydrological events in Xinjiang, north-west of China. *Nat Hazards* 67, 483–495.

- Sun, P., Zhang, Q., Lu, X.X., Bai, Y., 2012. Changing properties of low flow of the Tarim River basin: Possible causes and implications. *Quaternary International* 282, 78-86.
- Ran, M., Zhang, C., Zhao, F., 2015. Climatic and hydrological variations during the past 8000 years in northern Xinjiang of China and the associated mechanisms. *Quaternary International* 358, 21-34.
- Salas, J. D., 1993. *Analysis and modeling of hydrologic time series*, in *Handbook of Hydrology*. Edited by D. Maidment, pp. 19.1–19.72, McGraw-Hill, New York.
- Tao, H., Gemmer, M., Song, Y., Jiang, T., 2008. Eco-hydrological responses on water division in the lower reaches of the Tarim River, China. *Water Resources Research* 44, W08422. doi:10.1029/2007WR006186.
- Villarini, G., Serinaldi, F., Smith, J.A., Krajewski, W.F., 2009a. On the stationarity of annual flood peaks in the continental United States during the 20th century. *Water Resources Research*, 45, W08417. doi:10.1029/2008WR007645.
- Villarini, G., Smith, J.A., Serinaldi, F., Bales, J., Bates, P.D., Krajewski, W.F., 2009b. Flood frequency analysis for nonstationary annual peak records in an urban drainage basin. *Advances in Water Resources* 32, 1255–1266.
- Villarini, G., Vecchi, G.A., Smith, J.A., 2010. Modeling the Dependence of Tropical Storm Counts in the North Atlantic Basin on Climate Indices. *Monthly Weather Review* 138, 2681-2704.
- Villarini, G., Smith, J.A., Serinaldi, F., Ntelekos, A.A., Schwarz, U., 2012. Analyses of extreme flooding in Austria over the period 1951–2006. *Int. J. Climatol.* 32, 1178–1192.
- Wang, H., Chen, Y. and Li, W., 2015. Characteristics in streamflow and extremes in the Tarim River, China: trends, distribution and climate linkage. *Int. J. Climatol.* 35, 761–776.
- Wen, K.G., 2005. *Encyclopedia of Meteorological Disasters in China*. Beijing:

Meteorological Press (in Chinese).

World Bank, 2013. Turn Down the Heat: Climate Extremes, Regional Impacts, and the Case for Resilience. A report for the World Bank by the Potsdam Institute for Climate Impact Research and Climate Analytics. Washington, DC: World Bank. License: Creative Commons Attribution- NonCommercial- NoDerivatives3.0 Unported license (CCBY-NC-ND 3.0).

Zhang, Q., C.-Y. Xu, Z. Zhang, Y.D. Chen. 2009. Changes of temperature extremes for 1960-2004 in Far-West China. Stochastic Environmental Research and Risk Assessment 23(6), 721-735.

Zhang, Q., Xu, C.-Y., Tao, H., Jiang, T., Chen, Y.D., 2010. Climate changes and their impacts on water resources in the arid regions: a case study of the Tarim River basin, China. Stochastic Environmental Research and Risk Assessment 24(3), 349-358.

Zhang, Q., J. Li, V.P. Singh, M. Xiao, 2013a. Spatio-temporal relations between temperature and precipitation regimes: Implications for temperature-induced changes in the hydrological cycle. Global and Planetary Change 111, 57-76.

Zhang, Q., M. Xiao, V.P. Singh, X. Chen, 2013b. Copula-based risk evaluation of droughts across the Pearl River basin, China. Theoretical and Applied Climatology 111(1), 119-131.

Zhang, Q., X. Gu, V.P. Singh, M. Xiao, C.-Y. Xu, 2014a. Stationarity of annual flood peaks during 1951-2010 in the Pearl River basin, China. Journal of Hydrology 519, 3263-3274.

Zhang, Q., X. Gu, V.P. Singh, M. Xiao, 2014b. Flood frequency analysis with consideration of hydrological alterations: changing properties, causes and implications. Journal of Hydrology 519, 803-813.

- Zhang, Q., X. Gu, V.P. Singh, M. Xiao, X. Chen, 2015a. Evaluation of flood frequency under non-stationarity resulting from climate change and human activities in the East River basin, China. *Journal of Hydrology* 527, 565-575.
- Zhang, Q., X. Gu, V.P. Singh, M. Xiao, C.-Y. Xu, 2015b. Flood frequency under the influence of trends in the Pearl River basin, China: changing patterns, causes and implications. *Hydrological Processes* 29, 1406-1417.
- Zhang, Q., X. Gu, V.P. Singh, P. Sun, X. Chen, D. Kong., 2016. Magnitude, frequency and timing of floods in the Tarim River basin, China: changes, causes and implications. *Global and Planetary Change* 139, 44-55.
- Zhang, X., 1992. Investigation of glacier bursts of the Yarkant River in Xinjiang, China. *Annals of Glaciology* 16, 135-139.

Fig. 1 Locations of meteorological and hydrological stations in the Tarim River basin, China.

Fig. 2 Temporal variations of frequency of extreme flood events at the hydrological stations in the Tarim River basin, China. The thick solid black curves are the 5-year moving average; Z values are the MK trend test results of the 5-year moving average; $Z > 1.96$ indicates significant trends. The thick solid red curves are the Loess-based fit curves with degree of freedom of 0.2. The stations are: TG: Tongguziluoke; YZ: Yuzimenleke; KQ: Kaqun; SL: Shaliguilanke; XH: Xiehela; HS: Huangshuigou; DS: Dashankou; AL: Alaer.

Fig. 3 Temporal variations of frequency of extreme flood events in the subregions in the Tarim River basin, China. The thick solid black curves are the 5-year moving average; Z values are the MK trend test results of the 5-year moving average; $Z > 1.96$ indicates significant trends. The thick solid red curves are the Loess-based fit curves with degree of freedom of 0.2.

Fig. 4 Temporal variations of frequency of extreme precipitation events at the meteorological stations in the Tarim River basin, China. The thick solid black curves are the 5-year moving average; Z values are the MK trend test results of the 5-year moving average; $Z > 1.96$ indicates significant trends. The thick solid red curves are the Loess-based fit curves with degree of freedom of 0.2. The stations are: TG: Tongguziluo; YZ: Yuzimenleke; KQ: Kaqun; SL: Shaliguilanke; XH: Xiehela; HS: Huangshuigou; DS: Dashankou; AL: Alaer.

Fig. 5 Occurrence rates of floods (FOR) at the hydrological stations in the Tarim River basin. The stations are: TG: Tongguziluohe; YZ: Yuzimenleke; KQ: Kaqun; SL: Shaliguilanke; XH: Xiehela; HS: Huangshuigou; DS: Dashankou; AL: Alaer.

Fig. 6 Occurrence rates of floods in the subregions (RFOR) of the Tarim River basin. The subregions are: Aksu, Kezhou, Kashi, Hotan and Bazhou, and the locations of these subregions can be found in Fig. 1.

Fig. 7 Occurrence frequency of extreme precipitations at the meteorological stations in the Tarim River basin. The stations are: TG: Tongguziluoke; YZ: Yuzimenleke; KQ: Kaqun; SL: Shaliguilanke; XH: Xiehela; HS: Huangshuigou; DS: Dashankou; AL: Alaer.

Fig. 8 Worm plots showing the performance of GAMLSS models (Model0, Model1, Model2, and Model3) in modeling frequency of floods with YZ (Yuzimenleke) station as an example.

Fig. 9 GAMLSS-based modeling of flooding frequency at the hydrological stations with time (Model 1) and climate indices (Model 2) as covariate variables. The stations are: TG: Tongguziluoke; SL: Shaliguilanke; HS: Huangshuigou; AL: Alaer.

Fig. 10 GAMLSS-based modeling of flooding frequency at the hydrological stations with time and precipitation and temperature indices (Model 3) as covariate variables. The stations are: YZ: Yuzimenleke; SL: Shaliguilanke; HS: Huangshuigou; DS: Dashankou.

Fig. 11 Prediction of flooding frequency of the next 15 years using GAMLSS model with climate index as the covariate variable (Model 2) at the hydrological stations. The white dots indicate annual numbers of floods at hydrological stations. The dark gray region the area between the 25th and 75th percentiles, and the light gray region the area between the 5th and 95th percentiles, are the same as the predicted regions with biases.

Table 1 The statistical information of daily streamflow and daily precipitation data in hydrological and meteorological stations and data on flood occurrences in subregions in the Tarim River basin.

Stations	Streamflow series			Precipitation series		Subregions	Number of floods
	Period of streamflow series	Threshold value (m ³ /s)	bandwidth (days)	Threshold value (mm)	bandwidth (days)		bandwidth (days)
Tongguziluoke	1962-2008	292	2295	3.43	1917	Aksu	1570
Yuzimenleke	1957-2009	137	1441	4.77	1185	Kezhou	1325
Kanqun	1957-2009	659	1825	4.49	1852	Kashi	1355
Shaliguilanke	1956-2007	330	1175	14.46	2040	Hotan	2046
Xiehela	1956-2007	290	1939	5.53	1719	Bazhou	1681
Huangshuigou	1957-2008	40.5	1411	17.48	1709		
Dashankou	1972-2008	225	2349	13.82	1728		
Alaer	1958-2008	534	1938	3.89	2087		

Note: Threshold value indicates Peak-over-Threshold (POT) sampling threshold value.
Bandwidth indicates the bandwidth of kernel function.

Table 2 KS_D values showing whether the timing of extreme streamflow and precipitation events are from the uniform distribution.

Stations	Streamflow		Precipitation		Subregions	Number of regional floods	
	KS_D	Threshold value of 0.1	KS_D	Threshold value of 0.1		KS_D	Threshold value of 0.1
Tongguziluoke	0.047	0.092	0.116	0.091	Aksu	0.164	0.094
Yuzimenleke	0.092	0.088	0.079	0.091	Kezhou	0.222	0.125
Kanqun	0.071	0.089	0.107	0.091	Kashi	0.242	0.102
Shaliguilanke	0.112	0.087	0.107	0.090	Hotan	0.309	0.130
Xiehela	0.079	0.089	0.130	0.090	Bazhou	0.266	0.110
Huangshuigou	0.120	0.088	0.042	0.091			
Dashankou	0.070	0.102	0.098	0.090			
Alaer	0.061	0.087	0.044	0.091			

Note: KS_D indicates the statistical test value of Kolmogorov-Smirnov.

Table 3 Performance of four GAMLSS-based models (Model 0, Model 1, Model 2 and Model 3) in modeling station-based number of floods in hydrological stations. The AIC values are for the optimal models; the Filliben coefficients are used to evaluate the goodness-of-fit performance of optimal models (Filliben coefficient > 0.95 means high goodness-of-fit performance of the optimal models. cs() denotes cubic spline function. Variables without cs() denote linear function.

Stations	Model 0		Model 1		Model 2			Model 3		
	AIC	Filliben	AIC	Filliben	AIC	Λ_i	Filliben	AIC	Λ_i	Filliben
Tongguziluoke	168	0.99	174	0.99	165*	AO+AMO	0.98	169	x3	0.99
Yuzimenleke	203	0.98	205	0.99	203	AMO	0.99	186*	x1	1.00
Kanqun	186	0.98	188	0.99	182	AMO	0.98	174*	x3	0.98
Shaliguilanke	205	0.99	197	0.99	200	SOI	0.99	183*	cs(x2)	0.99
Xiehela	180	0.99	181	0.99	182	AO	0.98	162*	x2	0.99
Huangshuigou	197	0.98	197	0.99	195	AMO	0.99	169*	x2+x3	1.00
Dashankou	138	0.99	139	0.98	139	AO	0.99	138*	x3	0.99
Alaer	181	0.96	186	0.97	179	NAO+SOI	0.99	168*	x2	0.96

Note: bold numbers denoted that AIC values of Model 0 are larger than those of Model 1, Model 2 or Model 3. AIC values with * denote that this AIC value is the smallest.

Table 4 Performance of three GAMLSS-based models (Model 0, Model 1 and Model 2) in modeling station-based number of floods in subregions. The AIC values are for the optimal models; the Filliben coefficients are used to evaluate the goodness-of-fit performance of optimal models (Filliben coefficient > 0.95 means high goodness-of-fit performance of the optimal models. $cs()$ denotes cubic spline function. Variables without $cs()$ denote linear function.

Subregions	Model 0		Model 1		Model 2		
	AIC	Filliben	AIC	Filliben	AIC	Parameters	Filliben
Aksu	201	0.972	197*	0.989	199	NAO	0.988
Kezhou	141	0.997	131*	0.989	139	NAO+SOI	0.991
Kashi	196	0.977	179*	0.996	190	$cs(NAO)+cs(SOI)$	0.989
Hotan	132	0.985	123*	0.994	123	$cs(SOI)$	0.990
Bazhou	169	0.984	153*	0.977	157	NAO+SOI	0.986

Note: Black bold numbers mean that $AIC \leq \text{Model 0}$; Numbers with * denote the smallest AIC values amongst all the models considered in this study.