**Temporal clustering of floods and impacts of climate indices in the Tarim
River basin, China**

Xihui Gu, Qiang Zhang, Vijay P Singh, Yongqin David Chen, Peijun Shi

Corresponding author:

Qiang Zhang, Ph.D. Professor

State Key Laboratory of Earth Surface Processes and Resource Ecology,

Beijing Normal University,

Beijing 100875,

China

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

Temporal clustering of floods and impacts of climate indices in the Tarim River basin, China

Xihui Gu¹, Qiang Zhang^{2,3,4}, Vijay P Singh⁵, Yongqin David Chen⁶, Peijun Shi^{2,3,4}

1. Department of Water Resources and Environment, Sun Yat-sen University, Guangzhou 510275, China;

2. Academy of Disaster Reduction and Emergency Management, Beijing Normal University, Beijing 100875, China;

3. Key Laboratory of Environmental Change and Natural Disaster, Beijing Normal University, Beijing 100875, China;

4. State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing 100875, China;

5. Department of Biological and Agricultural Engineering and Zachry Department of Civil Engineering, Texas A&M University, College Station, Texas, USA;

6. Department of Geography and Resource Management, and Institute of Environment, Energy and Sustainability, The Chinese University of Hong Kong, Shatin, Hong Kong, China.

Abstract: The occurrence rates of floods in Tarim River basin, the largest arid basin in China, were estimated using the Peak-over-Threshold (POT) technique. The intra-annual, seasonal and inter-annual clustering of floods was then analyzed using the Cox regression model, month frequency method and dispersion index, respectively. Possible impacts of climate indices on the occurrence rates were also investigated. Both NAO and AO are selected as significant covariates to occurrence rates of floods in Tarim River basin by Cox regression model, suggesting occurrence of flood events is not

independent, but exhibits temporal clustering in intra-annual scale. On the basis of the results of the station and region-wide modeling by Cox regression model, we suggest using a model in which the rate of occurrence depends on monthly averaged NAO or AO. The Cox regression model not only can be used to assess the time-varying rate of flood occurrence, but also has the capability to forecast the predictors. Flood occurrence time and probability of exceedance are changing with climate index from negative to positive on both station and region scale. In addition, seasonal clustering of station-based floods and regional observed floods are also identified with mainly concentrating from June to August. Meanwhile, dispersion index is used to evaluate the inter-annual clustering of annual number of flood occurrences both on station and region. We found that inter-annual clustering of regional floods is more evident than that of station-based floods, indicating that regional observed flood records are generally over-dispersed with a tendency for flood events to cluster in time

Keywords: Cox regression model; dispersion index; clustering of floods; POT-based sampling technique; Tarim River basin

1. Introduction

Some studies indicated that climate warming may have the potential to accelerate the hydrologic cycle at global and continental scales (e.g., Alan et al., 2003; Bates et al., 2008). There is correlation between climate warming and heavy precipitation, which suggests that increasing temperature can greatly enhance precipitation intensity (Zhang et al., 2013a). Extreme weather and climatic events, such as extreme precipitation, longer dry spans, higher peak flows, and increased intensity of tropical cyclones have been observed more frequently during the past two or three decades (Zhang et al., 2013b; IPCC, 2013; World Bank, 2013; Apurv et al., 2015). Increasing occurrences of floods

have been arousing considerable concern in recent years.

Further, recent studies have examined clustering of occurrence of extratropical and tropical storms, i.e., whether the counts of extreme events vary in time or occur randomly with constant rate in time (e.g., Mailier et al., 2006; Vitolo et al., 2009, Villarini et al., 2010). These studies concluded that clustering was a significant feature of storm occurrences with periods of elevated activity followed by relative quiescence. More generally, the question of whether hydrometeorological events (e.g., extreme precipitation, floods) cluster aroused great human concerns due to the large economic repercussions. After analyzing the precipitation structure, Zhang et al. (2012) pointed out that severe droughts or floods were apt to occur in groups, which indicated a clustering behavior in extreme precipitation episodes. Due to flood occurrence process driven by both large-scale climate indices and antecedent rainfall in Iowa, Villarini et al. (2013) found that clustering was also as an important element of the flood occurrence process. Although models describing clustering of floods have been developed, such as Cox regression model and dispersion index (Villarini et al., 2013; Mediero et al., 2015), the researches about temporal clustering of flood events are still scarce. The previous clustering studies mainly focused on the point process modeling framework of flood occurrence, which can be represented as Poisson processes with randomly varying rate of occurrence (Smith and Karr, 1986; Silva et al, 2012; Villarini et al, 2013; Mediero et al, 2015). Poisson processes with constant rate in time represent flood occurrences which are not clustered, and vice versa. However, Silva et al. (2012) showed that the occurrence of flood events constitutes an inhomogeneous Poisson process in Portuguese watersheds, hence the occurrence rates are nonstationary, i.e. varying with time. Cox regression model built by Smith and Karr (1986), was used to model flood

occurrences in which the rate of occurrence depends on time-varying covariate processes. In the special case that the rate of occurrence is deterministic, the Cox process simplifies to a Poisson process. The work of Villarini et al. (2013) is among the few studies using Cox model with climate indices as time-varying covariates to analyze temporal clustering of flood events in Iowa. Moreover, the degree of flood clustering and departure from a Poisson process can be characterized by the index of dispersion that relates the variability in an observed series of annual number of flood occurrences to its expectation value (Mediero et al., 2015).

In this study, we examine the temporal clustering of flood events in Xinjiang, China. Floods in Xinjiang are one of the major natural hazards, inflicting massive losses on economic, agriculture and so on. Moreover, another important reason for focusing on Xinjiang area is that it is dominated by a typical arid climate and the Tarim River basin, the largest arid inland river basin in China. Water resources are the key to economic development, conservation of ecological environment, and agricultural irrigation in the basin. 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). Generally, the basin is characterized by low precipitation and high evaporation (Zhang et al., 2010; Sun et al., 2012). Due to the exceptional and unique role that water resources play in economic development, the Chinese government launched a five-year emergency water diversion program in 2000 with US\$1.3 billion earmarked for reclamation of the river and Taitema Lake (Tao et al., 2008). However, studies have indicated increasing heavy rainfall events and amplification of flood disasters since the 1980s (Zhang et al., 2016), 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). Further, it seems that glacier melt-induced floods in the basin have been increasing in recent years (e.g. Zhang et al., 1992).

Agriculture in the Tarim River basin is dependent on irrigation and is vulnerable to hydrologic extremes in the headwaters, such as floods and droughts (e.g. Zhang et al., 1992; Jiang et al., 2005). Possible clustering of floods and related causes are critical to water resources systems management and development of irrigation. Nevertheless, the following issues have not been addressed for the basin: (1) what are the clustering effects of flood events evaluated by both Cox regression model and index of dispersion, if any, in an arid river basin; (2) how the time-varying climate indices can be used to model the occurrence rates of floods both in hydrological station and subregions in Tarim River Basin, Xinjiang; and (3) can climatic indices be taken as the predictors for probabilistic forecasting of flood occurrence? Answers of these issues can help with management of water resources and agricultural irrigation in Xinjiang.

2. Data

Daily streamflow data of at least 37 years 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. The missing data were processed based on the method by Zhang et al. (2011). Specifically, the missing values of 1–2 days were replaced by the average of neighboring days. Adequate consecutive days with missing data were filled by the long-term average of

the same days of other years. 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 clustering 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 clustering in occurrence rates of floods in regions.

Most of 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, Wang et al., 2015; Zhang et al., 2016). 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.

Covariates used in this study are the North Atlantic Oscillation (NAO) and Arctic Oscillation (AO). The NAO and AO climate indices were chosen not only because of availability of daily data covering a long time period but also because of the strong relationship between NAO and AO and hydrometeorological variables over Xinjiang. Yang et al. (2010) found that winter NAO had significant positive correlation relationship with annual average temperature and winter temperature, while AO was also a major factor with strong relationship with temperature during almost all seasons. Compared with AO, spring NAO was the major factor influencing precipitation in Xinjiang (Yang et al., 2010). Ran et al. (2015) pointed out that Standardized Moisture

Index (SMI) had a general rising trend in Xinjiang which can be explained by invoking its climatic linkage with NAO. Kalra et al. (2013) developed a data-driven model incorporating oceanic atmospheric oscillations, which also included NAO, as predictors to increase the streamflow lead time in the Kaidu River, a major tributary of Tarim River basin. Recently, Wang et al. (2015) highlighted the importance of NAO and AO on the significant influence of stream means and extremes in Tarim River Basin.

The above studies have concentrated on the inter-decadal and inter-annual impacts of NAO and AO on hydrometeorological variables in Xinjiang. However, there is still a knowledge gap about what is the role of large-scale climate indices in driving the occurrence of hydrometeorological variables at intra-annual scale. It is fortunate that the Cox regression model provides a powerful statistical framework to check whether the rate of occurrence of a counting process depends on covariate processes at intra-annual scale (Smith and Karr, 1986). In recent years, the values of Pacific/North American Teleconnection (PNA) and North Atlantic Oscillation (NAO) averaged over the previous 14d and 28d before occurrence date of floods were composed as covariates and were used to analyze their roles in driving the occurrence of flooding by Cox regression model (Villarini et al., 2013). In addition, we used POT sampling to obtain occurrence time of flood events, and the waiting time between two flood peaks is usually more than 15d to satisfy the independence of flood peaks. Therefore, as similar as Villarini et al. (2013), we focus on the values of these climate indices averaged over the previous 15d (we refer to them as “NAO15” and “AO15”) and 30d (we refer to them as “NAO30” and ‘AO30’), to capture longer time scales of influence of these indices.

In this study, NAO and AO covering a period of 1950-2010 were specifically accepted as the climate indices having impacts on the timing of floods in the Tarim

River basin. Daily NAO and daily AO data were extracted from
<http://www.esrl.noaa.gov/psd/data/climateindices/list/>.

3. Methodology

3.1. Peak-over-Threshold (POT) based sampling

Flood series can be obtained using two techniques, i.e. annual peak flow and POT-based sampling technique. POT sampling can help gain information on the magnitude and timing of flood events. The independence of flood peaks can be evaluated by (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; A denotes the drainage area in km^2 ; and Q_1 and Q_2 denote the magnitudes of two flood peaks, respectively, with units 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 and 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 (Zhang et al., 2016).

Then, the time data of flood events and the number of flood occurrence in each year (count data, i.e. annual number of flood occurrences) were obtained. In addition, the time data is within intra-annual scale with a cycle of 365 or 366 (the occurrence location (date) of each flood event on corresponding year, e.g., 99, 191, 240, 114, 130, 194, 214, 202, ...). When using POT sampling method to generate the time data and count data, the independence of flood peaks has also been satisfied. In addition, unlike magnitude of floods is described by continuous distribution, both time data and count data are described by discrete distribution (i.e. point process). Therefore, the auto-correlated issue in flood events doesn't have to be considered as well as Silva et al. (2012) and Villarini et al. (2013).

3.2. Cox regression model

The Cox regression model was developed for counting processes by Smith and Karr (1986) for incorporating time-varying exogenous information into flood frequency analysis. The model is used to describe the temporal variations of flood occurrence rates or nonstationary Poisson process. If the occurrence rates of floods are stationary, the model is simplified to stationary Poisson process. The stationary Poisson flood process indicates no clustering and it is opposite for the nonstationary Poisson process. The Cox model was improved for flood frequency analysis by Smith and Karr (1986). Cox regression model has been widely used in survival analysis (e.g. Zivanovic et al., 2015), its application to the analysis of flood events has been scarce. Our study also adopted the Cox regression model for flood occurrences in which the rate of occurrence depends

on time-varying covariate processes.

The Cox regression model we used in this study has been described with considerable details in Smith and Karr (1986) and Villarini et al. (2013). Here, we only give a brief introduction. For time data of flood series, we can describe the point process using a counting process representation:

$$N_i(t) = \sum_{j=1}^{M_i} 1(T_{i,j} \leq t) \quad (2)$$

where $t \in [0, T]$, 0 is the starting time of each year, and T is the ending time of each year, 365 or 366; $T_{i,j}$ is the time of the j^{th} flood during year i ; M_i is the number of flood peaks during year i .

$N_i(t)$ is Poisson process with an independent discrete interval and the distribution of counts is Poisson (Villarini et al., 2013), i.e.:

$$Pr\{N_i(t) = k\} = \frac{\exp\left\{-\int_0^t \lambda(u) du\right\} \left[\int_0^t \lambda(u) du\right]^k}{k!} \quad (3)$$

where $\lambda(u)$ ($u \in [0, T]$) is a non-negative function, representing occurrence rates of flood events. If $\lambda(u)$ is a constant when $u \in [0, T]$, then floods are a stationary Poisson process. However, clustering denotes a kind of nonstationary flood process, i.e. occurrence rate, $\lambda(u)$, is changing with time or covariate. The Cox process is a kind of double Poisson process: randomness of Poisson process and randomly changing occurrence rates of floods in a Poisson process.

In the Cox process, the occurrence rate of one event is in close relation with that of another event, and the increase or decrease of occurrence probability hinges on the properties of Poisson distribution. Given $\lambda(u)$, the occurrence rate of floods follows the Poisson distribution but the non-conditional distribution does not satisfy the Poisson

process. The Cox model is used in modeling random bursting and/or dormant condition of an event. In the Cox-based flood frequency analysis model developed by Smith and Karr (1986), the occurrence rate of floods, $\lambda(u)$, depends on covariate variables (Villarini et al., 2013), i.e.:

$$\lambda_i(t) = \lambda_0(t) \exp \left[\sum_{j=1}^m \beta_j Z_{ij}(t) \right] \quad (4)$$

where $\lambda_0(t)$ ($t \in [0, T]$) is a nonnegative time function and is also recognized as a baseline hazard function; $Z_{ij}(t)$ is the j^{th} covariate function for the i^{th} year; β_j is the coefficient of the j^{th} covariate function; and $\lambda_i(t)$ is the occurrence rate process of the i^{th} year, and also called as the conditional density function or hazard function.

The selection of optimal covariate variable, Z_j ($j = 1, \dots, m$), was based on the Akaike Information Criterion (AIC). Three methods were used in this study to determine the optimal covariate that can be used in modelling the occurrence rates of floods, i.e.: Chi-Square Goodness-of-Fit Test, Schoenfeld residual plot and Dfbeta plot. The Cox regression model assumes proportional hazards (PH), i.e. the absence of interaction between predictor and predictand (Villarini et al., 2013). The assumption behind Chi-Square test and Schoenfeld residual plot in testing whether the developed model can satisfy the Proportion-hazards model (PHM) is that the PHM assumption is satisfied when $P > 0.05$ of the Chi-Square test and no linear trends in the Schoenfeld residual. Dfbeta plot was used to test whether the occurrence rates of floods had significant impacts on the estimation of covariates. If the Dfbeta residuals were mostly near 0 implies then the observed occurrence rates of floods had no impacts on the estimated coefficients of covariates.

Wang et al. (2015) found that the AO and NAO showed significant influence on streamflow means and extremes, and pointed out 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 try to estimate the probabilistic forecasts of primary flood events by Cox regression model based on the study of Maia and Meinke (2010). Maia and Meinke (2010) have presented two examples for El Nino/Southern Oscillation (ENSO)-based forecasts by Cox regression model: the onset date of the wet season (Cairns, Australia) and total wet season rainfall (Quixeramobim, Brazil).

Cox regression model is semi-parametric approaches originally designed for modelling risks arising from time-to-event data. For flood risk management, cumulative distribution functions (CDFs) are an important source of information, on which Cox regression model provides a statistical framework for probabilistic forecasting with covariate. Hence, after referencing the way from Maia and Meinke (2010), we firstly try to evaluate the probabilities of floods to provide a way of risk assessments and more informative than central tendency measures. The quantification of optimal covariates and the developed Cox regression model helped compute the probability of exceedance of flood events by probability of exceedance functions or CDFs, $P(T > t)$, and the prediction of floods was done using the above-decided covariate (Maia and Meinke, 2010):

$$P(T > t; Z_j) = [P_0(t)]^{\exp[\beta_j Z_j(t)]} \quad (5)$$

where $P_0(t)$ acts as the base function for $P(T > t; Z_j)$.

3.3. Temporal clustering in flood occurrence rates

The Cox regression model was used to analyze the clustering of occurrence rates of floods with time data in intra-annual scale. In addition, inter-annual variations of clustering in occurrence rates of floods with count data can be evaluated by the

dispersion index, D , which allows for the quantification of deviation of annual occurrence rates from the expected occurrence rates. Mumby et al. (2011) analyzed temporal clustering of tropical cyclones and its ecosystem impacts using the dispersion index at yearly scale:

$$D = \frac{Var(N_i)}{E(N_i)} - 1 \quad (6)$$

where N_i denotes the number of floods during the i th year; $E(N_i)$ is the average number of floods; $Var(N_i)$ denotes the variance of number of floods; and D is the dispersion index. $D > 0$ implies a larger dispersion degree of the POT-based sampled flood series and shows clustering of floods with active and dormant periods. The bootstrap technique was used to quantify the confidence interval for D , and the Lagrange multiple statistic, LM, was used to evaluate significance level of D (Greene, 2003).

3.4 Monthly frequencies of flood events

Flood events with occurrence time were obtained by POT-based sampling and then the annual distribution of flood events was analyzed by monthly frequencies method (Mediero et al., 2015). Also, periods with abundant and/or deficient floods were identified. Monthly frequency analysis was done to investigate the annual distribution of flood events:

$$FF_m = \frac{F_m}{N} \frac{30}{n_m} \quad (7)$$

where F_m is the annual accumulative occurrence rates of floods during the m th month; N denotes the total occurrence rates of POT-based sampled floods; n_m is the number of days of the m th month; FF_m denotes the occurrence rates of floods during the m th month. If the occurrence of floods is not seasonal, FF_m should be 1/12; and for

seasonal occurrence rates of floods, the 95% confidence interval of FF_m can be quantified as (Mediero, et al., 2015):

$$\begin{cases} L_U^N = \frac{N + 11.491}{0.048N^{1.131}} \\ L_L^N = \frac{N - 27.832}{0.199N^{0.964}} \end{cases} \quad (8)$$

4. Results and discussion

4.1 Cox regression model analysis

Based on the MnT and related threshold curves, MET and related threshold curves, and DI and related threshold curves (Zhang et al., 2016), the thresholds for flood events at each station are calculated. Hence, POT-based sampling was done for daily streamflow data to obtain the time data and count data.

The time data is used to analyze the relations between occurrence rates of floods and large-scale climate indices by Cox regression model in Tarim River basin (Tables 2 and 3). It can be depicted from Table 2 and Table 3 that NAO30 and AO30 are two principal climatic indices that influence the flood occurrence rates in Tarim River basin with log link function. We also fitted a stratified Cox model to examine which, if any, of these covariates were important when we pooled individual stations and subregions together, respectively. The stratified Cox model is a kind of extension of the Cox regression model in which the data are divided into strata and each stratum shares the same coefficients of the covariate with other strata, but each has an individual baseline hazard function (e.g. Villarini et al., 2013). In this study, each stratum is represented by a station or subregion. NAO30 is the important covariates according to AIC for both stations and subregions. In addition, the values of coefficient β are positive (negative), implying that large (small) values of NAO and AO would result in an increased

(decreased) rate of occurrence of flooding over Tarim River basin. This result provides supporting evidence of the importance of NAO and AO in describing the occurrence of flooding over Tarim River basin.

Of all the predictors, those averaged over a 30d period are the ones that the most frequently selected in this study. These results suggest that the average state a month prior to the occurrence of a flood event is more important than the state of the atmosphere 15d prior to the flood. Some studies have indicated the delayed effect between streamflow and climate indices (e.g., Lu et al., 2011; Zaroug et al., 2014). According to previous brief introduction, NAO and AO have evidential correlation with precipitation and temperature in Tarim River basin, in which floods are mainly caused by heavy precipitation and glacier melt related to temperature. Therefore, due to the fact that NAO and AO affect runoff by influencing precipitation and temperature, we assume that there may be more waiting time about NAO and AO influencing on flood events.

Before further analysis of the optimized Cox regression model, the validity of assumption of the Cox-based proportion-hazards regression was evaluated using Chi-Square test and Schoenfeld residuals. Tables 2 and 3 indicate that the P values of Cox regression models for station-based and regional floods were larger than 0.05, satisfying the assumption of the proportion-hazard regression. Moreover, we selected two stations (subregions) with largest and smallest P values, respectively, for further analysis of Schoenfeld residuals and Dfbeta residuals to test the fitting performance of Cox regression model (Fig. 2)

It can be observed from Fig. 2 that for station-based and regional floods, Schoenfeld residuals did not have a linear tendency and all residuals scattered randomly within the time interval considered in this study (left column of Fig. 2). This result further

corroborates the validity of assumption of the Cox proportion-hazard regression. The temporal distribution of Dfbeta residuals indicated that the residuals between modelled and observed station-based and regional floods were mostly near 0, showing a minor influence of the observed flood series on the coefficients of climate indices. Thus, it can be tentatively concluded that the occurrence rates of station-based and regional floods are influenced by large-scale climate indices, and NAO and AO are the significant covariates impacting station-based and regional floods. These results imply that occurrence rates of floods are not independent, as generally assumed, showing temporal clustering.

4.2 Temporal variation of occurrence rates of floods

Based on the developed Cox regression model and the selected key influencing climate indices, the annual occurrence rates of floods ($\lambda_i(t)$) at each individual station were obtained (Fig. 3). For reference, we also include the solid curves showing long-term annual average occurrence rates of floods and representing the behavior of the corresponding inhomogeneous Poisson process. It can be depicted from Fig. 3 that flood occurrence rates at all the hydrological stations had bimodal and/or unimodal distributions and the highest flood occurrence rate was about 0.15. The annual occurrence rates of floods at Tongguziluoke, Shaliguilanke, Xiehela, Huangshuigou, Dashankou and Alaer stations had unimodal distributions with higher flood occurrence rates during July and August, implying high flood activity during these two months with occurrence rates > 0.05 . Diverse distribution patterns of flood occurrence rates were identified at the Yuzimenleke and Kaqun stations, implying diverse factors influencing flood occurrence rates at these two stations. These periods with highest occurrence rates of floods were corresponding to June, July or early August in each

station and these periods were dominated by higher temperature-induced melting of snow/glacier and frequent rainstorms.

The magnitudes of POT-based sampled flood series and Julian time were plotted to show the degree to which the estimated occurrence rates of floods matched the observed ones (Fig. 4). It can be seen from Fig. 4 that flood occurrences at all the hydrological stations clustered during a certain time interval. Specifically, floods at the Tongguziluoke, Yuzimenleke, Kaqun, Xiehela, Huangshuigou and Alaer stations occurred mainly during June to September with flood peak values, and long-term annual average daily streamflow has a unimodal distribution pattern. High flow regimes at the Shaliguilanke and Dashankou stations occurred consecutively within a longer time interval without remarkable flood peak values. The flood peaks are relatively flattened when compared to those at other stations. The temporal variations of long-term annual average daily streamflow and flood peaks (Fig. 4) were in agreement with estimation of occurrence rates of floods (Fig. 3). Specifically, the higher values occurring during the months for long-term annual average daily streamflow and flood peaks are in line with estimation of occurrence rates of floods. Also, the temporal distribution patterns of the long-term annual average daily streamflow and flood peaks as shown in Fig. 4 are consistent with those of estimation of occurrence rates of floods as shown in Fig. 3. Besides, both occurrence rates of floods and long-term annual average daily streamflow had a unimodal distribution pattern, implying high temporal clustering of flood events, which may pose a serious challenge for mitigation of flood hazards.

In comparison to POT-based sampled occurrence rates of floods at hydrological stations, the estimated occurrence rates of regional floods can mirror the actual temporal properties of the occurrences of regional floods (Fig. 5). It can be seen from Fig. 5 that there are remarkable differences in the temporal distribution of occurrence rates of regional floods among these five subregions. The occurrence rates of regional floods in the Aksu and Kashi had a unimodal pattern with peak values during July and the highest occurrence rate of about 0.1. The occurrence rates of regional floods in Kezhou scattered sporadically and randomly during March and August without discernible peak values. The occurrence of regional floods in Kezhou reached a peak value during May and then decayed gradually. This result shows that temperature changes, which impact the melting of snow and/or glacier, significantly influence the occurrence rates of regional floods in the Kezhou. The occurrence rates of regional floods in Bazhou fluctuated at a higher level during May and August with the occurrence rate about 0.05 and even higher than 0.1, being far higher than the occurrence rates of regional floods in other four subregions. The temperature kept increasing during May, July and August, which enhanced the melting of snow or glacier. Meanwhile, most of annual precipitation concentrated on May, July and August. The two factors jointly caused higher occurrence rates of floods during these three months.

Fig. 6 shows the estimated occurrence rates of regional floods and occurrence of real-world regional floods (Fig. 6). It can be seen that most floods occurred during the periods with higher occurrence rates of floods. The floods that occurred in Aksu and Kashi mainly concentrated during July and August. The flood occurrences in Kezhou

were however distributed temporally in an even way. Floods in Bazhou occurred mainly during May to August. Fig. 6 also shows an amplification of flooding after the 1980s and it is particularly the case in the subregions of Aksu, Kashi and Bazhou. Scarce occurrences of floods can be observed in Hotan.

4.3 Probability of exceedance for timing of flood

Management of flood risk is one of the critical tasks of governmental decision making agencies. The cumulative probability distribution or probability of exceedance provides critical information for the management of flood risk. Also, forecasting of occurrence probability is a key step for the effective management and decision making. The Cox regression model with optimized climate indices as predictors was used to forecast the probability of exceedance of regional or station-based floods (Figs. 7 and 8).

It can be seen from Fig. 7 that probability of exceedance at Tongguziluohe, Xiehela and Alaer stations trend to increase when climate indices tend to be positive, and vice versa at the other stations. Meanwhile, the flood occurrence time with the same probability of exceedance tends to be delayed under climate indices tending to be positive at Tongguziluohe, Xiehela and Alaer stations, and vice versa at the other stations. Fig. 7 also shows larger changing magnitudes of probability of exceedance at Yuzimenleke, Kaqun, Shaliguilanke and Xiehela stations as a result of changes of climate indices, implying higher sensitivity of floods to changes of climate indices at these four stations. That is benefit for forecasting of occurrence probability for floods

with climate index as predictor.

Forecasting of the probability of exceedance of regional floods (Fig. 8) is of great practical value for the regional mitigation of floods, management of agricultural irrigation and for security of life and property. Fig. 8 also shows that regional floods in the northwest regions of the Tarim River basin, i.e. Aksu, Kezhou and Kashi, are subject to decreasing probability of exceedance when climate indices tend to be positive, and vice versa on the other regions. In another words, a higher climate index can drive regional floods in the Aksu, Kezhou and Kashi to a lower probability of exceedance. In addition, under climate index tending to be positive, the occurrence time of regional floods in Aksu, Kezhou and Kashi tended to be earlier with a similar probability of exceedance. However, different responses of regional floods to climate index changes were identified in southeastern parts of the Tarim River basin, i.e. Hotan and Bazhou, when compared to those in Aksu, Kezhou and Kashi. Meanwhile, responses of regional floods to climate index changes were subject to higher sensitivity with larger changing magnitudes of probability of exceedance when compared to station-based floods.

The forecasting performances of Cox regression model should be validated. However, the results of the Cox regression model forecasting is not specific values but probability of exceedance curve corresponding to specific climate index value, while traditional methods such as root mean squared error (RMSE) are not suitable. Because there is a range among climate index values (in this paper, the climate index varies from -2.5 to 2.5), the probability of exceedance curve changes with climate index values. Therefore, we can't evaluate the performances of the probability of exceedance curve

with observations. Despite Maia and Meinke (2010) pointed out probability curves should be drawn with a thick felt pen, they still used the probability of exceedance curve of Cox regression model forecasting for the time to wet season onset without any validation. In addition, they agreed with the words said by the economist, John Maynard Keynes, that it is more important to be roughly right, than precisely wrong (Corazza, 1999). At present, we used Chi-Square test, Schoenfeld residual test and Dfbeta plot to evaluate the goodness of fit of the selected optimal Cox regression model with final climate index (Fig. 2) as well as Villiarini et al. (2013) and Maia and Meinke (2010). Due to the good performances of the selected optimal Cox regression model, it can be used to calculate the probability of exceedance curve corresponding to each specific final-selected climate index. The further work about validating the performances of the forecasts should be supplemented when the Cox regression model is developed with suitable validation method.

Though our study showed the relationships between probability of exceedance and climate indices, some studies indicated that climate indices such as ENSO events can be accurately predicted one to two years in advance using a physical model of the coupled ocean-atmosphere system (e.g., Whitaker et al., 2011; Gutierrez and Dracup, 2001). Therefore, identifying such relationships holds potential as a powerful forecasting tool for water resources and floods. However, we should know that the reliability of water resources and flood predictions is heavily dependent on the accuracy of the predictions of climate indices.

4.4 Seasonal and inter-annual clustering of flood occurrence

The Cox regression model was used to analyze the changes in occurrence rates of floods with time data at intra-annual scale. Further analysis was done on seasonal and inter-annual clustering of floods. Seasonal variations of flood occurrence are critical for the identification of active and dormant flood periods and management of floods, which were evaluated by monthly frequencies of flood occurrences method (Mediero et al., 2015). Significant flood-poor and flood-rich months can be identified by comparing the expected sampling variability of monthly flood occurrences when no seasonality is assumed with the sampling variability of the observed monthly flood occurrences in a given record. In a non-seasonal model floods have the same probability of occurrence in all months, while in a seasonal model they are more likely to occur in given months.

Fig. 9 shows that the highest month frequency at all the hydrological stations was observed during July and August and the frequency ranged between 0.4-0.5. Therefore, July and August were the periods dominated by flood activities. Besides, the long durations dominated by dormant flood activities was identified during January to April, and September to December. The month frequency of flood occurrence during dormant periods was almost zero and even zero. Dormant periods of floods were characterized by little precipitation and lower air temperature which limited the flood occurrence.

The month frequency of regional floods (Fig. 9b) indicated similar changing patterns of flood occurrence as those at the hydrological stations. The period of June to August was dominated by active flooding with month frequency between 0.2-0.3. Two peak values of flood occurrence were found in Kezhou, i.e. May and August, implying that increased temperature and increased precipitation both triggered frequent floods during different months. When compared to the length of time interval with dormant flood activities, the shorter length of periods with dormant regional flood activities was

shorter.

The dispersion index, D , is often used to evaluate the inter-annual clustering of flood occurrence with count data, i.e. annual number of flood occurrences. Inter-annual clustering properties of station-based and regional floods were determined as listed in Tables 4 and 5. It can be seen from Table 4 that the D values were smaller than 0 at most hydrological stations, i.e. Tongguziluoke, Kaqun, Xiehela, Huangshuigou, Dashankou and Alaer, implying no inter-annual clustering of floods. In this case, it was called under-dispersion and the occurrence of flood events was on a regular basis. The D values were larger than 0 at the Yuzimenleke and Shaliguilanke stations, showing evident but not significant inter-annual clustering of floods at these two stations.

Results of dispersion index for regional floods are listed in Table 5. It can be seen from Fig. 5 that the D values for regional floods were all larger than 0, showing the clustering of floods at the inter-annual time scale. Moreover, the D values for regional floods in the Aksu, Kashi and Hotan subregions were statistically significant at 0.1 significance level, and the D values for regional floods in the Kashi and Hotan subregions were even significant at 0.001 significance level. These results showed strong inter-annual clustering of regional floods. The inter-annual clustering of flood events indicated flood occurrence trended to cluster during a certain time interval when compared to other time intervals.

5. Conclusions

Intra-annual clustering of floods in Tarim River basin, a typical arid basin in China, is analyzed based on POT-based sampled flood series and also real-world observed regional flood occurrence records using Cox regression model. Seasonal and inter-annual variations of temporal clustering of floods are also investigated using monthly

frequency analysis and dispersion index. The following conclusions are drawn from this study:

(1) Based on results by Cox regression model, NAO and AO are accepted as the key factors impacting the flood occurrence rates in Tarim River basin. Moreover, NAO30 and AO30 are taken as the principal influencing climate indices at 6 out of 8 hydrological stations. Further, AO30 is recognized as the critical climate index impacting the occurrence rates of regional floods in 3 out of 5 subregions considered in the study. On the basis of the results of the station and region-wide modeling by Cox regression model, we suggest using a model in which the rate of occurrence depends on monthly averaged NAO or AO.

(2) Dependence of the rate of flood occurrence on NAO and AO points to intra-annual clustering as an important element of the flood occurrence process on both station and region scales. The higher values of occurrence rates of floods in stations almost concentrate during June to August with above 0.1. Temporal distribution of occurrence rates of regional floods is remarkably different from one region to another across the Tarim River basin, indicating that changes of flood occurrence due to different causes, such as melting of snow or precipitation changes.

(3) The probability of exceedance of floods under the influence of climate indices provides valuable information for flood risk management. The Cox regression model not only can be used to assess the time-varying rate of flood occurrence, but also has the capability to forecast the predictors. Flood occurrence time and probability of exceedance are changing with climate index from negative to positive on both station and region scale. However, larger changing magnitudes of probability of exceedance at regions than stations as a result of changes of climate indices, implies higher sensitivity of floods to changes of climate indices.

(4) Similar seasonal distribution features of month frequency of flood occurrence can be identified for station-based POT sampling floods and regional observed floods. Active flood periods are mostly during June to August. Compared to the seasonal distribution of flood occurrence of station-based floods, more even seasonal distribution of flood occurrence can be identified for regional floods with shorter dormant flood duration.

(5) With respect to inter-annual clustering of floods, stronger clustering of annual number of flood occurrences can be observed for regional floods than for station-based floods. Specifically, a significant inter-annual clustering of regional floods can be identified in the Aksu, Kashi and Hotan subregions, leading to a larger variability in the annual number of flood occurrences. These results indicate that regional observed flood records are generally over-dispersed, indicating a tendency for flood events to cluster in time.

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), 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. Sierd Cloetingh, and two anonymous reviewers for their professional and pertinent comments and revision suggestions which are greatly helpful for further improvement of the quality of this manuscript.

References

599 Alan, D.Z., Justin, S., Edwin, P.M., Bart, N., Eric, F.W., Dennis, P.L., 2003. Detection
 600 of intensification in global- and continental-scale hydrological cycles: temporal
 601 scale of evaluation. *J. Clim.* 16, 535–547.
 602 Apurv, T., Mehrotra, R., Sharma, A., Goyal, M.K., Dutta S., 2015. Impact of climate
 603 change on floods in the Brahmaputra basin using CMIP5 decadal predictions.
 604 *Journal of Hydrology* 527, 281-291.
 605 Bates, B., Kundzewicz, Z., Wu, S., Palutikof J., 2008. Climate change and water, Tech.
 606 rep., Intergovernmental Panel on Climate Change.
 607 Corazza, M., 1999. Merton-like theoretical frame for fractional Brownian motion in
 608 finance. In *Current Topics in Quantitative Finance*, Canestrelli E (ed.). Physica
 609 Verlag: Heidelberg; 37.
 610 Greene, W.H., 2003. *Econometric Analysis*, Ed. Prentice Hall, New Jersey.
 611 Gutierrez, F., Dracup, J.A., 2001. An analysis of the feasibility of long-range
 612 streamflow forecasting for Colombia using El Nino-Southern Oscillation
 613 indicators. *J. Hydrol.* 246, 181–196.
 614 IPCC. 2013. Summary for policymakers. In: Stocker TF, Qin D, Plattner G-K, Tignor
 615 M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (Eds.), *Climate*
 616 *Change 2013: The Physical Science Basis*, Contribution of Working Group I to the
 617 *Fifth Assessment Report of the Intergovernmental Panel on Climate Change*.
 618 Cambridge University Press, Cambridge, United Kingdom and New York, NY,
 619 USA.
 620 Jiang, F., Zhu, C., Mu, G., Hu, R., Meng, Q., 2005. Magnification of flood disasters and
 621 its relation to regional precipitation and local human activities since the 1980s in
 622 Xinjiang, northwestern China. *Natural Hazards* 36, 307-330.
 623 Kalra, A., Li, L., Li, X., Ahmad, S., 2013. Improving streamflow forecast lead time

624 using oceanic-atmospheric oscillations for Kaidu River basin, Xinjiang, China.
 625 Journal of Hydrologic Engineering 18(8), 1031-1040.

626 Lang, M., Ouadab, T.B.M.J., Bobee, B., 1999. Towards operational guidelines for
 627 over-threshold modeling. Journal of Hydrology 255, 103-117.

628 Lu, A., Jia, S., Zhu, W., Yan, H., Duan, S., Yao, Z., 2011. El Nino-Southern Oscillation
 629 and water resources in the headwaters region of the Yellow River: links and
 630 potential for forecasting. Hydrol. Earth Syst. Sci. 15, 1273–1281.

631 Maia, A.H.N., Meinke, H., 2010. Probabilistic methods for seasonal forecasting in a
 632 changing climate: Cox-type regression models. Int. J. Climatol. 30, 2277–2288.

633 Mailier, P.J., Stephenson, D.B., Ferro, C.A.T., 2006. Serial clustering of extratropical
 634 cyclones. Monthly Weather Review 134(8), 2224-2240.

635 Mediero, L., Kjeldsen, T.R., Macdonald, N., Kohnova, S., Merz, B., Vorogushyn, S.,
 636 Wilson, D., Albuquerque, T., Blöschl, G., Bogdanowicz, E., Castellarin, A., Hall,
 637 J., Kobold, M., Kriauciuniene, J., Lang, M., Madsen, H., Onusluel Gül, G.,
 638 Perdigão, R.A.P., Roald, L.A., Salinas, J.L., Toumazis, A.D., Veijalainen, N.,
 639 Óðinn Þórarinnsson, 2015. Identification of coherent flood regions across Europe
 640 by using the longest streamflow records. Journal of Hydrology 528, 341–360.

641 Mumby, P.J., Vitolo, R., Stephenson, D.B., 2011. Temporal clustering of tropical
 642 cyclones and its ecosystem impacts. Proceedings of the National Academy of
 643 Sciences of the United States of America 108(43), 17626-17630.

644 Ran, M., Zhang, C., Feng, Z., 2015. Climatic and hydrological variations during the
 645 past 8000 years in northern Xinjiang of China and the associated mechanisms.
 646 Quaternary International 358, 21-34.

647 Silva, A.T., Portela, M.M., Neghettini, M., 2012. Nonstationarities in the occurrence

648 rates of flood events in Portuguese watersheds. *Hydrol. Earth Syst. Sci.* 16, 241–
649 254.

650 Smith, J.A., Karr, A.F., 1986. Flood frequency analysis using the Cox regression model.
651 *Water Resources Research* 22(6), 890-896.

652 Sun, G., Chen, Y., Li, W., Pan, C., Li, J., Yang, Y., 2013. Spatial distribution of the
653 extreme hydrological events in Xinjiang, north-west of China. *Nat. Hazards* 67,
654 483–495.

655 Sun, P., Zhang, Q., Lu, X.X., Bai, Y., 2012. Changing properties of low flow of the
656 Tarim River basin: Possible causes and implications. *Quaternary International* 282,
657 78-86.

658 Tao, H., Gemmer, M., Song, Y., Jiang, T., 2008. Eco-hydrological responses on water division in
659 the lower reaches of the Tarim River, China. *Water Resources Research* 44, W08422.
660 doi:10.1029/2007WR006186.

661 Villarini, G., Vecchi, G.A., Smith, J.A., 2010. Modeling of the dependence of tropical
662 storm counts in the North Atlantic Basin on climate indices. *Monthly Weather*
663 *Review* 138(7), 2681-2705.

664 Villarini, G., Smith, A.J., Vitolo, R., Stephenson, D.B., 2013. On the temporal
665 clustering of US floods and its relationship to climate teleconnection patterns.
666 *International Journal of Climatology* 33, 629-640.

667 Vitolo, R., Stephenson, D.B., Cook, I.M., Mitchell-Wallace, K., 2009. Serial clustering
668 of intense European storms. *Meteorologische Zeitschrift* 18(4), 411-424.

669 Wang, H., Chen, Y., Li, W., 2015. Characteristics in streamflow and extremes in the
670 Tarim River, China: trends, distribution and climate linkage. *Int. J. Climatol.* 35,
671 761–776.

672 Wen, K.G., 2005. Encyclopedia of Meteorological Disasters in China. Beijing:
673 Meteorological Press (in Chinese).

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

679 Whitaker, D.W., Wasimi, S.A., Islam, S., 2011. The El Nino Southern Oscillation and
680 long-range forecasting of flows in the Ganges. *International J. Climatol.* 21, 77–87.

681 Yang, Y.H., Chen, Y.N., Li, W.H., Wang, M.Z., Sun, G., 2010. Impacts of climatic
682 change on river runoff in northern Xinjiang of China over last fifty years. *Chinese*
683 *Geography Sciences* 20(3), 193-201.

684 Zaroug, M.A.H., Eltahir, E.A.B., Giorgi, F., 2014. Droughts and floods over the upper
685 catchment of the Blue Nile and their connections to the timing of El Niño and La
686 Niña events. *Hydrol. Earth Syst. Sci.* 18, 1239–1249.

687 Zhang, Q., Xu, C.-Y., Tao, H., Jiang, T., Chen, Y.D., 2010. Climate changes and their
688 impacts on water resources in the arid regions: a case study of the Tarim River
689 basin, China. *Stochastic Environmental Research and Risk Assessment* 24(3), 349-
690 358.

691 Zhang, Q., Singh, V.P., Li, J., Chen, X., 2011. Analysis of the periods of maximum
692 consecutive wet days in China. *Journal of Geophysical Research* 116, D23106,
693 doi:10.1029/2011JD016088.

694 Zhang, Q., Zhou, Y., Singh, V.P., Li, J., 2012. Scaling and clustering effects of extreme
695 precipitation distributions. *Journal of Hydrology* 454-455, 187-194.

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

Zhang, Q., Gu, X., Singh, V.P., Sun, P., Chen, X., Kong, D., 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.

Zivanovic, R., Milloy, M.J., Hayashi, K., Dong, H., Sutherland, C., Kerr, T., Wood, E., 2015. Impact of unstable housing on all-cause mortality among persons who inject drugs. *BMC Public Health* 15(106), 1479-1479.

Table 1 Station-based streamflow and regional flood records of the Tarim River basin

No.	Stations	Abbreviation	Streamflow series	Subregions	Flood record
1	Tongguziluke	TG	1962-2008	Aksu	1950-2000
2	Yuzimenleke	YZ	1957-2009	Kezhou	1950-2000
3	Kaqun	KQ	1957-2009	Kashi	1950-2000
4	Shaliguilanke	SL	1956-2007	Hotan	1950-2000
5	Xiehela	XH	1956-2007	Bazhou	1950-2000
6	Huangshuigou	HS	1957-2008		
7	Dashankou	DS	1972-2008		
8	Alaer	AL	1958-2008		

Table 2 Statistical summary of the Cox regression model using climate indices as predictors for floods at hydrological stations

Stations	Selected climate indices	AIC	Coefficient β	P values*
Tongguziluo	NAO30	1052.42	-0.18	0.75
Yuzimenleke	NAO15	1178.91	0.24	0.6
Kaun	AO30	1159.61	0.26	0.66
Shaliguilan	AO30	1197.78	0.38	0.08
Xiehela	NAO30	1169.3	-0.3	0.31
Huangshuigou	NAO30	1191.21	0.17	0.34
Dashankou	NAO30	813.17	0.07	0.58
Alaer	AO15	1206.55	-0.29	0.8
Station stratum	NAO30	8972.89	0.08	0.06

*Note: P values are for chi-square test, which are higher than 0.05 to satisfy the PHM assumption.

Table 3 Summary statistics of the Cox regression model using climate indices as predictors for regional floods

Stations	Selected climate indices	AIC	Coefficient β	P values*
Aksu	AO30	964.0438	0.1907	0.1042
Kezhou	AO15	313.5139	0.2244	0.691
Kashi	NAO15	682.853	0.1455	0.0726
Hotan	AO30	213.515	-0.3914	0.5497
Bazhou	AO30	526.6694	-0.6184	0.8973
Station stratum	NAO30	2701	0.1497	0.1386

*Note: P values are for chi-square test, which are higher than 0.05 to satisfy the PHM assumption.

Table 4 Dispersion index of the occurrence rates of floods at the hydrological stations in the Tarim River basin.

Stations	<i>D</i>	Upper 95% CI	Lower 95% CI	<i>P</i> value
Tongguziluohe	-0.32933	-0.57043	-0.07323	0.095779
Yuzimenleke	0.021193	-0.29237	0.336809	0.992092
Kaun	-0.26684	-0.57444	0.089732	0.148494
Shaliguilanke	0.201854	-0.18124	0.580945	0.367954
Xiehela	-0.3011	-0.59652	-0.01325	0.111911
Huangshuigou	-0.0991	-0.40228	0.227451	0.552747
Dashankou	-0.19011	-0.52126	0.160143	0.361847
Alaer	-0.2872	-0.61722	0.008154	0.128293

Note: CI denotes confidence interval; bold number denotes that the dispersion index is significant at 0.1 significance level.

Table 5 Dispersion index of the occurrence rates of regional floods in the Tarim River basin.

Stations	<i>D</i>	Upper 95% CI	Lower 95% CI	<i>P</i> value
Aksu	0.35968	-0.03905	0.74366	0.092634
Kezhou	0.156923	-0.17143	0.499561	0.497854
Kashi	0.765957	0.188152	1.319566	0.000222
Hotan	0.704615	0.028	1.361739	0.000701
Bazhou	0.16	-0.17875	0.508923	0.488245

Note: CI denotes confidence interval; bold numbers denote that the dispersion index is significant at 0.1 significance level.

Fig. 1 Locations of hydrological 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. 2 Goodness-of-fit performance of Cox regressive model for occurrence timing of station-based and regional floods across the Tarim River basin. The day 1 is corresponding to January 1.

Fig. 3 Temporal changes in the occurrence rates of floods at hydrological stations across the Tarim River basin. The solid black curves denote the long-term annual average occurrence rate of floods. The stations are: TG: Tongguziluke; YZ: Yuzimenleke; KQ: Kaqun; SL: Shaliguilanke; XH: Xiehela; HS: Huangshuigou; DS: Dashankou; AL: Alaer. The day 1 is corresponding to January 1.

Fig. 4 Temporal changes of POT-based magnitude of flood peak and long-term daily streamflow at hydrological stations across the Tarim River basin. To enhance visibility of temporal changes, the long-term daily average streamflow is computed as 5 times of the actual long-term daily streamflow. The stations are: TG: Tongguziluoke; YZ: Yuzimenleke; KQ: Kaqun; SL: Shaliguilanke; XH: Xiehela; HS: Huangshuigou; DS: Dashankou; AL: Alaer. The day 1 is corresponding to January 1.

Fig. 5 Temporal changes in occurrence rates of floods within different subregions of the Tarim River basin. The way of calculation about occurrence rates of floods within subregions is the same as in hydrological stations (Figure 4). The solid black curves denote the long-term annual average occurrence rates of floods. The day 1 is corresponding to January 1.

Fig. 6 Annual variations in occurrence rates of regional floods across the Tarim River basin. The white circles indicate the occurrence of a flood event in subregions. The day 1 is corresponding to January 1. The shaded parts indicate the occurrence rate process of each year, $\lambda_i(t)$.

Fig. 7 Forecasted probabilities of flood exceedance using the selected climatic indices as the predictor at the hydrological stations across the Tarim River basin. From left to right three gray curves denote 5%, 50% and 95% percentiles respectively. CI is the short name of “climate index”. The stations are: TG: Tongguziluoke; YZ: Yuzimenleke; KQ: Kaqun; SL: Shaliguilanke; XH: Xiehela; HS: Huangshuigou; DS: Dashankou; AL: Alaer. The day 1 is corresponding to January 1.

Fig. 8 Forecasted probabilities of regional flood exceedance using the selected climatic indices as the predictor across the Tarim River basin. The way of calculation about probabilities of exceedance of floods within subregions is the same as in hydrological stations (Figure 8). From left to right three gray curves denote 5%, 50% and 95% percentiles respectively. CI is the short name of “climate index”. The day 1 is corresponding to January 1.

Fig. 9 Monthly flood frequency at the hydrological stations (a) and also monthly flood frequency of regional floods (b). The stations are: TG: Tongguziluoke; YZ: Yuzimenleke; KQ: Kaqun; SL: Shaliguilanke; XH: Xiehela; HS: Huangshuigou; DS: Dashankou; AL: Alaer.