Author/s:
Su, Z;Ho, M;Hao, Z;Lall, U;Sun, X;Chen, X;Yan, L

Title:
The impact of the Three Gorges Dam on summer streamflow in the Yangtze River Basin

Date:
2020-01-30

Citation:
Su, Z., Ho, M., Hao, Z., Lall, U., Sun, X., Chen, X. & Yan, L. (2020). The impact of the Three Gorges Dam on summer streamflow in the Yangtze River Basin. Hydrological Processes, 34 (3), pp.705-717. https://doi.org/10.1002/hyp.13619.

# The impact of the Three Gorges Dam on summer streamflow in the Yangtze River Basin

**Running title:** Impact of Three Gorges Dam on summer streamflow

Zhenkuan Su [1], Michelle Ho[2, 3], Zhenchun Hao [1, 4], Upmanu Lall [2, 5], Xun Sun [2, 6], Xi Chen [7], Longzeng Yan[8]

[1] State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, China

[2] Columbia Water Center, Columbia University, New York, NY 10027, USA

[3] CSIRO Land and Water, Canberra, ACT, 2601, Australia

[4] National Cooperative Innovation Center for Water Safety & Hydro-Science, Hohai University, Nanjing 210098, China

[5] Department of Earth and Environmental Engineering, Columbia University, New York, NY 10027, USA

[6] Key Laboratory of Geographic Information Science (Ministry of Education), East China Normal University, Shanghai 200241, China

[7] Bureau of Hydrology, Changjiang Water Resources Commission, Wuhan 430010, China

[8] Upper Changjiang River Bureau of Hydrology and Water Resources Survey, Bureau of Hydrology, Changjiang Water Resources Commission, Chongqing 400020, China



**Acknowledgements**

We would like to sincerely thank our colleagues at Hohai University and the Columbia Water Center, Columbia University for their constructive feedback and critique. This work was supported by the National Key Research Projects (Grant No.2018YFC1508001 and No.2016YFC0402704).

**Correspondence**

Zhenchun Hao, State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, China. E-mail: hzchun-hhu@163.com

# The impact of the Three Gorges Dam on summer streamflow in the Yangtze River Basin


## Abstract

The Three Gorges Dam is the world's largest capacity hydropower station located in the Hubei province along the Yangtze River in China, which began operations in 2003. The dam also functions to store and regulate the downstream releases of water in order to provide flood control and navigational support in addition to hydropower generation. Flow regulation is particularly important for alleviating the impacts of low and high flow events during the summer rainy season (June, July, and August). The impact of dam operations on summer flows is the focus of this work. Naturalized flows are modelled using a canonical correlation analysis and covariates of subbasin-scale precipitation resulting in good model skill with an average correlation of 0.92. The model is then used to estimate natural flows in the period after dam operation. A comparison between modelled and gauged streamflow post 2003 is made and the impact of the dam on downstream flow is assessed. Streamflow variability is found to be strongly related to rainfall variability. An analysis of regional streamflow variability across the Yangtze River Basin showed a mode of spatially negatively correlated variability between the upper and lower basin areas. The Three Gorges Dam likely mitigated the occurrence of high flow events at Yichang station located near the dam. However, the high flow at the remaining stations in the lower reach is not noticeably alleviated due to the diminishing influence of the dam on distant downstream flows and the impact of the lakes

downstream of the dam that act to attenuate flows. Three types of flow regime changes between naturalized and observed flows were defined and used to assess the changes in the occurrence of high and low flow events resulting from dam operations.

**Keywords:** Three Gorges Dam; Canonical correlation analysis; High flow; Low flow

## 1. Introduction

The Yangtze River Basin in China is home to one-third of the country's population and the largest capacity hydropower project in the world to date – the Three Gorges Dam (TGD) (Figure 1). The river is also known as "Changjiang" in China - its literal translation being "long river", which reflects its status as the longest river in Asia and is a significant region in terms of China's history, culture, and economy (Bermudez, 2015).

The basin's climate is characterized by complex regional subbasin-scale patterns of precipitation and temperature in terrain ranging from plateaus to plains. The combination of complexity in both climate and terrain results in widely varying rainfall-runoff processes across the basin that spans an area of 1.8 million $km^2$. Periodic floods and droughts have been experienced in the basin throughout history and a number of extreme events have captured the public's attention due to the loss of life and impacts on the regional economy (Bing et al., 2012; Wang et al., 2011; Xu et al., 2008; Yin et al., 2001). For example, in 1998 floods destroyed 4 970 000 houses (Zong et al., 2000) while a drought in 1994 was estimated to have resulted in economic losses of CNY 20 billion (approximately $2.3 billion USD) (Zhang et al., 2015).

The TGD was originally conceived almost a century ago but progress was interrupted by political and social upheavals. Construction of the TGD began in 1993 with flow regulation and water impoundment commencing in 2003. Chinese laws and regulations stipulate that the primary function of any water conservancy project must be flood storage and control. That is, flood control operations will take precedence over other functions such as hydropower, transport and water security. The dam is therefore operated to reduce the impacts of floods in spring and summer, while water is released over winter and autumn to maintain downstream flows for water supply and to facilitate navigation (Zhao et al., 2012). The Yangtze River has experienced a number of severe low flow events (e.g. in 2006, 2008, 2011) resulting in suspension of navigation in reaches downstream of the TGD, loss of residential water supply for hundreds of thousands in the Hubei Province (Qiu, 2011), and water shortages that impacted 18 million people during the 2006 autumn drought over the Sichuan and Chongqing provinces (Zhang et al., 2015). Such events have raised concerns regarding the effectiveness of the dam in providing seasonal flow regulation. Past studies have focused on the TGD's impacts on downstream lakes, river levels and flows (Chen et al., 2016; Jiang et al., 2014; Lai et al., 2014; Li et al., 2016; Mei et al., 2015; Wang et al., 2017; Wang et al., 2013), however, the changes in flow regimes between naturalized flows and observed flows impacted by the dam and the underlying mechanism have been largely overlooked. Consequently, an analysis that compares estimated naturalized flows against flows regulated by the TGD in the period post impoundment is important for determining whether dam operations likely reduced or exacerbated

low flow or high flow occurrences, and how the occurrence of these flow types have changed.

Numerous approaches for estimating naturalized flows have been used including distributed hydrological models and stochastic models. Hydrological models such as the physically based Soil and Water Assessment Tool (SWAT) model, MIKE SHE, and Variable Infiltration Capacity (VIC) model have been widely used in the analyses of streamflow modelling and their predictions (Kim et al., 2016; Liang et al., 1994; Vansteenkiste et al., 2013; Yuan et al., 2004). However, distributed hydrological models are limited by large data requirements, unclear parameter uncertainty, and inaccurate modelling of physical process (Gayathri et al., 2015; Pechlivanidis et al., 2011). Data constraints, such as limited data on catchment characteristics and land use limit the ability to accurately model hydrological processes using physically based models.

Alternatively, statistically-based models, such as regression and correlation models, are observation-oriented methods that seek to find functional relationships between explanatory and response variables (Elsanabary et al., 2015; Kwon et al., 2009; Steinschneider et al., 2016). These models are data orientated and do not explicitly consider the process driving the hydrological system. In addition, stochastic models are typically comprised of a relatively small number of model parameters. These models therefore allow for a parsimonious characterization of streamflow variability. Yangtze River Basin streamflow has been shown to be strongly related to rainfall particularly at coarser (seasonal to annual) timescales. For example, the correlation coefficient between annual precipitation and streamflow at Datong station in the Lower Yangtze River Basin is

103 0.85 (Zhang et al., 2011) suggesting that statistical rainfall-runoff models can be useful in the
104 region.

105 Here, we assess the impacts of the TGD on summer flows using a statistical approach to model
106 naturalized flows and comparing these modelled flows with observed flows in the period following
107 dam operation. The rainfall and streamflow data used in the analysis are described in Section 2
108 along with key hydrological features along the Yangtze River. The canonical correlation analysis
109 used to model streamflow across the basin is described in Section 3 and we address the issue of high
110 spatial correlation among both rainfall covariates and streamflow. In Section 4 we quantify the
111 relationship between summer (the wettest season) rainfall and streamflow along the main stem of
112 the Yangtze River downstream of the TGD. The spatial and temporal patterns of rainfall anomalies
113 over the whole basin are used to qualitatively assess whether changes in rainfall variability can
114 likely explain changes in streamflow in the post dam period compared with modelled streamflow.
115 We also assess the influence of the dam on summer flow by identifying changes in low, normal and
116 high flow regimes. We find that the TGD likely attenuates both high and low flows at downstream
117 locations close to the dam. However, the dam's impacts are largely eliminated on flows further
118 downstream.

## 119 2. Study area and data description

120 The TGD is located approximately 1880 km upstream of the estuary where the Yangtze River flows
121 into the East China Sea, and significant landmarks of the Dongting Lake (covering an area of ~

2625 km$^2$) and Poyang Lake (~4125 km$^2$) are located in this reach. To assess the likely impacts of dam operations on regulating summer high and low flows along the Yangtze River, we use streamflow data from four hydrological stations downstream of the dam located on the main stem at Yichang, Luoshan, Hankou, and Datong (Figure 1).

The Yichang station is known as the "Gateway to the Three Gorges" and is located approximately 40 km downstream of the dam. The Luoshan station is located close to the outlet of the Dongting Lake that aggregates the flows from two major subbasins (i.e. Yuanshui and Xiangjiang). As a result, streamflow at Luoshan station is impacted by natural water detention in the lake. Hankou station was selected due to its importance for flood monitoring given its location downstream of where the Han River, which is the largest tributary, joins the main stem of the Yangtze River. Similar to the Luoshan station, streamflows at Datong station are impacted by lake effects due to its location downstream of Poyang Lake. The station is also critical for monitoring environmental flows and outflows to the sea.

Daily streamflow records were obtained from the Changjiang Water Resources Commission. Data have been checked and quality controlled and cover the period from 1960 to 2014 for each of the four stations. The monthly averaged streamflow data was obtained by averaging the daily flow rates within each month. The annual hydrograph shows the seasonality of streamflow (Figure 2), which is monsoon summer (June-July-August) flow dominated. As a result, the summer season (JJA) was selected to enable an analysis of changes in the occurrence of high and low flows. The summer

streamflow was logarithmically transformed as the streamflow at all four stations were approximately log-normally distributed.

Daily rainfall data was accessed from the China Meteorological Data Service Center (2016). Rainfall records that span from 1960 to 2014 with less than 5% missing values were selected resulting in data from 136 weather stations. For each station, missing values were infilled using the average values from the two nearest stations with rainfall recorded on that day. The intra-annual pattern of rainfall is similar to that of streamflow with maximum monthly precipitation occurring during summer. The gauged rainfall data was spatially aggregated using the 21 subbasin boundaries (Figure 1) obtained from the Changjiang Water Resources Commission. The rainfall over a given subbasin is the spatially averaged value across all rainfall stations within the subbasin using the Thiessen Polygon Method (Zhong et al., 2018). The spatial distribution of rainfall stations is highly heterogeneous. The Thiessen polygon method was applied as this approach accounts for the uneven distribution of rainfall sites, allowing for a more reasonable representation of aggregated rainfall. This method assigns an area called the Thiessen polygon (or the Voronoi diagram) to each rainfall station thereby providing an area-weighted average, which was applied to all stations within a subbasin to attain the aggregate subbasin rainfall.

## 3. Methodology

Here we present an approach for modelling unregulated summer streamflow using rainfall as a covariate. Canonical correlation analysis was selected to model the relationship between rainfall

and streamflow to account for rainfall induced changes of the streamflow across the four hydrological stations. The model performance metrics and method of analysis of the dam impact on summer streamflow were then described. The conveyance period for streamflow to travel from the upper catchment to the estuary is approximately 14 days (Chu et al., 2006), which is less than our 3 month time scale. Given the relatively coarse temporal resolution considered, we did not consider a temporal lag in our analysis.

### 3.1 Model design

The date of June 2003, hereon used to define the pre- and post-dam periods, delineates the change from a free-flowing river to an impounded river as this was the time at which storage in the reservoir behind the TGD began to accumulate. Initial diagnostic analyses were conducted using the rank correlation analysis between the rainfall $x_{t,k}$ in year $t$ at subbasin $k$, and the log transformed streamflow $\ln(y_{t,s})$ in year $t$ at station $s$ in the pre-dam period (shown in Figure 3). The summer flow data for all stations have the strongest positive correlations with spatially averaged rainfall over adjacent subbasins. However, rainfall over some subbasins far away from the streamflow stations has a negative correlation with the flow (e.g. the correlation of flow at Luoshan and rainfall over Mintuojiang in upper basin).

Physically, rainfall can only contribute to downstream streamflow, however, rainfall in a downstream subbasin may be indicative of regional climate that influences streamflow across the basin (e.g. the precipitation over a downstream catchment may be negatively correlated with

upstream streamflow). As a result, the model was designed using the spatially averaged rainfall over each subbasin.

The spatial and temporal relationship between rainfall and streamflow can be expressed by a formula $y = f(x)$, where $y$ is streamflow and $x$ is rainfall. The model was fit using data prior to 2003 in order to quantify the rainfall-runoff relationship prior to regulations by the TGD. The model was then used to estimate the streamflows post 2003 as informed by rainfall. The estimated flows are referred to as modelled naturalized flows here and are compared with the gauged flows post 2003 to assess the likely impacts of streamflow regulation by the TGD. A canonical correlation analysis (CCA) was applied to maximize the correlation between the covariates and the target variables.

### 3.2 Canonical correlation analysis

The proximity of the 21 subbasins results in covariates that are significantly correlated. Preliminary analysis (not presented here) showed that the correlations between many subbasins were greater than 0.9. Regression models require the explanatory variables to be significantly independent in order to avoid multicollinearity. Thus, canonical correlation analysis (CCA) (Hotelling, 1936) was implemented to allow for dimension reduction among the highly correlated rainfall data. CCA differs from other methods of dimension reduction such as principle component analysis (PCA) (Jolliffe, 2002) and archetype analysis (AA) (Cutler et al., 1994; Steinschneider et al., 2015; Stone et al., 1996) as CCA is constructed to maximize the correlation between the explanatory and target

variable through a linear transform. It has been used in applications of reconstructing streamflow using paleoclimate records (Ho et al., 2016), refining climate projections (Chen et al., 2017) and assessing regional flood frequency (Ouarda et al., 2001).

A brief description of CCA is presented here and we refer readers to Hotelling (1936) for more details. CCA consists of the linear transform of two variables to canonical form in order to maximize the correlation between them. Given two random variables $X$ and $Y$, where $X$ and $Y$ consist of $p$ and $q$ variables respectively, there exist two vectors $a$ (of length $p$) and $b$ (of length $q$) such that $U = a' X$ and $V = b' Y$ where the correlation between $U$ and $V$ is maximized. $U$ and $V$ yield the first pair of canonical variables. Subsequent canonical pairs may be derived but with the additional constraint that they are orthogonal to previously identified canonical variables. Dimension reduction is achieved by selecting fewer than min($p$, $q$) pairs of canonical variables. In this study, the explanatory variable $X$ is comprised of rainfall $x_{t,k}$ for $k \in$ 1:21 subbasins and the target variable $Y$ is the log-transformed streamflow $\ln(y_{t,s})$ for $s \in$ 1:4 and $t \in$ 1960:2014 for both $X$ and $Y$. In our study, the algorithm of CCA is implemented using the CCA package in R language by Gonzalez et al. (2008).

The first 4 pairs of canonical variables were retained to develop models relating rainfall and streamflow. For each pair, a linear regression equation was built using Equation (1).

$$V_i = \beta_i U_i + \alpha_i + \varepsilon_i \quad (1)$$

where, $i$ is a value ranging from 1 to 4, denoting the order of the 4 pairs of canonical variables. $\beta_i$

is the regression coefficient and $\alpha_i$ is the intercept in *i*-th linear equation. $\varepsilon_i$ is the error term and involves various uncertainties stemming from variability in streamflow unaccounted for by rainfall. Rainfall in the post dam period is used as covariates to estimate the naturalized streamflow using the canonical covariates *a* and Equation (1). Log-transformed streamflow is obtained by calculating *V* and back transforming *V* using canonical weights *b*. Streamflows were then back transformed from the log space.

### 3.3 Model performance metrics

A leave-*m*-out cross validation method was applied to verify the skill of the model. In this process, *m* randomly selected observations of streamflow data are retained for validation, while the remaining *n*-*m* data are used for model calibration, where *n* is the total number of observations. In this study, we first used $m = 5$ (approximately 10% of the total number of observations) and randomly selected *m* data points without replacement from the time series. The remaining 38 years of data were used for model calibration. The procedure was repeated 100 times to produce a set of validation metrics computed from all the cross validated models. Two metrics, the reduction of error (RE) and the coefficient of efficiency (CE) (Cook et al., 1994; Wilson et al., 2010) for those original variables rather than the canonical variables, were chosen to verify the performance and predictability of the model. These are defined as follows in Eq. (2) and Eq. (3):

$$\mathrm{RE} = 1 - \frac{\sum(x_i - \hat{x}_i)^2}{\sum(x_i - \bar{x}_c)^2} \quad (2)$$

$$\mathrm{CE} = 1 - \frac{\sum(x_i - \hat{x}_i)^2}{\sum(x_i - \bar{x}_v)^2} \quad (3)$$

Where, $x_i$ and $\hat{x}_i$ are the observed and predicted streamflow in year $i$ of the validation period, respectively. $\bar{x}_c$ is the mean of observations used for calibration and $\bar{x}_v$ is the mean of observations used for validation. Theoretically, the values of both RE and CE range from the negative infinity to a maximum of 1, with higher values indicating better prediction skill. A positive value indicates that the model has better predictive skill than using the climatology of the verification period. The RE and CE statistics are similar to the Nash-Sutcliffe efficiency score (McCuen et al., 2006; Nash et al., 1970) and have previously been used to assess the performance of streamflow models (Chen et al., 2014; Devineni et al., 2013; Ho et al., 2016).

### 3.4 Analysis of the impact of the TGD on summer streamflow

Based on the modelled streamflow across the four hydrological stations, 5-year moving averages of the modelled naturalized streamflow at each gauge were calculated. The observed and modelled streamflows are first qualitatively assessed. To analyze the variability of the non-regulated flow in both space and time principal component analysis (PCA) was applied.

Changes in streamflow during the post-dam period are induced by both rainfall variations and dam regulation. We therefore analyze and interpret variations in streamflow in the post dam period by considering the probable impacts of both variable rainfall and dam regulation. A qualitatively assessment of the influence of rainfall variability on summer flow was first investigated to show whether the changes in flow were possibly due to changes in rainfall. Then changes in flows at the four streamflow locations due to dam regulations were quantitatively assessed using estimates of

naturalized flows informed by rainfalls.

To quantitatively assess the differences between the modelled and observed streamflow during the post-dam period the frequency of high flow and low flow events and quantiles of summer flow were calculated for modelled and observed streamflow pre and post 2003. The comparison between streamflow before and after dam operation provides an approach to assessing the impact of the dam on either exacerbating or mitigating extreme events. We define a high flow occurrence as a streamflow event greater than the $75^{th}$ percentile summer flow of the pre-dam period. Similarly, a low flow occurrence is streamflow less than the $25^{th}$ percentile summer flow of the pre-dam period.

# 4. Results

## 4.1 Model results

The summer streamflows for all gauges along the main stem were modelled using the spatially averaged rainfall data and resulted in mean adjusted $R^2$ value of 0.80. The mean modelled streamflows along with the $5^{th}$-$95^{th}$ prediction interval are shown in Figure 4. The modelled streamflows closely match the observations, and the average correlation across all stations was 0.92. Some rainfall-runoff models have shown improved results when spatially averaged rainfall is used, compared with using gauged data (Tramblay et al., 2011).

To verify the performance and predictive skill of the model, cross validation was performed. The RE and CE values were calculated, and the results are presented in boxplots (Figure 5). There

resulted in a model with median RE and CE values of around 0.8 for all streamflow sites, indicating that the associated model predictions are more informative than climatology of the verification period.

### 4.2 Main modes of natural streamflow variability

Summer averaged streamflows at the four selected gauges within the Yangtze River Basin modelled using subbasin-scale rainfall were used to further analyze changes in streamflow response pre and post dam regulation. Both modelled and observed summer streamflow for the period 1960-2014 are shown in log space in Figure 6.

Prior to dam regulation the temporal variability of modelled streamflows closely matches that of the observations for all stations. The low flows in the early 1970s and high flows in the late 1990s (Gemmer et al., 2008) are visible in both the observed and modelled flows at all stations. The 5-year moving averages show the flows for the four stations fluctuate periodically. In the early 1970s all stations experienced negative summer flow anomalies. Negative streamflow anomalies are also observed and modelled in three of the stations (i.e. Luoshan, Hankou and Datong) in the late 1980s. In contrast, the early 1980s showed a persistent flood regime with positive anomalies seen in the 5-year moving averages for all stations.

PCA was used to extract the main spatiotemporal modes of logarithmically transformed modelled streamflows across all four stations to investigate their relationships both in space and time. A scree test was used to inform the number of PCs to retain (scree plot was shown in Figure S1 of the

supporting information). The first two PCs were retained and together explain over 97% of the streamflow variance (PCs were shown in Figure 7).

For PC1, the loading values are of the same sign across all stations representing a regionally cohesive signal. Both the flood in 1998 and drought in 1972 (Heng et al., 1999; Jiang et al., 2017; Yang et al., 1982; Zhang et al., 2015) are evident in the PC scores. During the 1998 floods, water transport from three regions, the Indian Ocean, the Bay of Bengal-South China Sea, and the Pacific Ocean, resulted in persistent heavy rainfall over the southwestern, southern and southeastern subbasins, respectively, culminating in a widespread severe flood (Jiang et al., 2017). Conversely, a large negative score for PC1 can be seen in 1978, when the middle and downstream Yangtze River Basin were struck by a severe drought event resulting from reduced summer precipitation (Zhang et al., 2015).

Unlike PC1, PC2 represents a non-homogeneous pattern of streamflow variability in the Yangtze River Basin (see loading values for PC2 in Figure 7). The loading values are of opposite sign for Yichang station close to the dam versus those located further downstream. This mode of variability distinguishes between streamflow in the mid to upper basin and the lower basin and is representative of the differences in regional climate and streamflow response. For example, Figure 6 shows anomalously low flows at Yichang station in 1969, however the remaining three downstream stations experience average to above average flows. A similar signal is seen in 1996 where flow at Yichang is average or slightly above average, while streamflows at the remaining three stations are

311 anomalously high. This signal is reversed in 1981 when flow at Yichang is above average and flows
312 at the downstream stations show negative anomalies.

313 **4.3 Influence of rainfall variability on summer streamflow**

314 High correlations between streamflow and precipitation suggest that streamflow variability is
315 strongly influenced by precipitation variability. The spatial and temporal patterns of rainfall
316 anomalies over the whole basin (Figure S2 of supporting information) are therefore first used to
317 qualitatively assess whether changes in rainfall variability can likely explain changes in streamflow
318 in the post dam period.

319 In the years of 2003 and 2010, anomalously high rainfalls were experienced in the subbasins located
320 the northern section of upstream of Yichang station (e.g., Jialingjiang basin), which should have
321 resulted in a high flow event. However, the dam operation likely reduced the magnitude of the flow
322 as evident by the reduced streamflow observations compared to the modelled flows at Yichang
323 station, which was likely due to decreased water releases thereby retaining the floodwaters and
324 reducing the downstream flows. Streamflow at the remaining three gauges further downstream were
325 not notably impacted by dam operations.

326 In 2008, streamflow observations at all four gauges recorded below average flows despite rainfall in
327 the northern section of the lower basin being slightly above average. It is likely that decreased
328 outflows from the dam in combination with decreased rainfall in the lower basin region resulted in
329 low flow conditions at Hankou and Datong stations downstream of the Dongting Lake.

In 2012, average rainfall was observed over the central region of the Yangtze River Basin with the exception of the Qingjiang subbasin, a relatively small subbasin located in the center of the basin (Figure 1). No notable change was observed at Yichang station when comparing modelled and observed streamflows. However, water supply from the lakes downstream (e.g. Dongting Lake and Poyang Lake), where above average rainfalls were observed, may have contributed to increased flows at Luoshan, Hankou and Datong stations compared with naturalized flows.

### 4.4 Dam impacts on summer streamflow

The modelled streamflows represent naturalized flows unimpacted by dam regulation and were compared with the observed flows for all stations in the post dam period to assess potential changes in the occurrence of high and low flows. The difference between the modelled and the gauged values therefore represents the difference in streamflow as a result of dam operations in addition to model errors. Figure 8 shows the distribution of high flow and low flow events for the non-regulated flows from 1960 to 2014 with results for the period of 2003-2014 obtained using modelled flows. Figure 8 shows that low flows and high flows may be wide spread across the basin (e.g. basin-wide high flows occurred in 1962, and basin-wide low flows occurred in 1971). These high and low flow also can co-occur in different regions of the basin (e.g. 1981 with low flow conditions in the lower basin and high flow conditions in the middle basin). Streamflows across the basin are neutral (i.e. in middle two quartiles) in approximately 16% of years.

The frequency of high flow and low flow events in the pre and post dam periods is shown in Table 1

for both modelled naturalized and observed flows. The frequency of high flow events for Yichang increased by 0.12 when comparing naturalized flows. The remaining stations show a decreased frequency in high flow events with a decrease in high flow frequency of 0.09 for Luoshan and Hankou and 0.03 for Datong. The frequency of low flow events for all stations increased by 0.04~0.12. Compared with the modelled naturalized flows, the frequency of high flow events at Yichang and Datong are reduced in the post-dam period, which indicates that the operation of the dam has likely contributed to the greatly reduced occurrence of high flow events. While the TGD appears to have attenuated the magnitude of high flow events, the occurrence of low flow events appear to have been magnified as seen when comparing modelled and observed low flow occurrences in the period post 2003.

The observed and modelled flows in the post dam period (2003-2014) are plotted as quantiles of the flows in the pre-dam period in Figure 9. The threshold lines of high flow and low flow events were plotted in Figure 9 shown by dashed and solid line respectively. These threshold values are used to identify three types of changes in occurrence of high flows or low flows at the four stations resulting from dam operation. The first type of flow regime change is defined as a year in which a high flow or low flow is identified in the naturalized flows, but has been alleviated by dam operations as evident in the observed flows. The second case is when no high flow or low flow is evident in the naturalized flows, but occurs in the observed record. The third case is when a high flow or low flow is observed in the naturalized flows, but the observed flows show the opposite

368 flow event of a low or high flow event respectively. The years in which these three different types
369 of flow regime changes occur are tabulated in Table 2.
370 From Table 2 and Figure 9, it can be seen that in the 12 years between 2003 and 2014 that the dam
371 likely mitigated high flow events at Yichang in three of these years (i.e. 2003, 2010, 2014) and at
372 Luoshan in 2014 exemplifying the first type of flow regime change. In contrast, these high flow
373 events were not alleviated by dam operations for the lower reach. These middle and lower basin
374 regions are influenced by lake effects upstream of the stations (Wang et al., 2017) and consequently
375 dam operations have limited noticeable impacts further downstream. An example of the second case
376 of flow regime change occurred in 2012, where either a high or low flow event is observed, but
377 does not appear in the modelled naturalized flows. Here, the naturalized flows are within the
378 $25^{th}$-$75^{th}$ percentile range at Luoshan, Hankou and Datong. However, the observed high flows
379 indicate that the combination of dam operations and lake effects resulted in increased flows at these
380 locations. Similarly, in 2013, a low flow event for Luoshan and Hankou was observed, while was
381 not evident in the modelled naturalized flows. A similar occurrence is seen in 2008 at Yichang and
382 Luoshan. In the same year of 2008, the third type of flow regime change occurred in the two
383 stations further downstream of the dam. That is, the naturalized flow for Hankou and Datong show
384 a high flow event while the observation indicates the occurrence of a low flow event. The flow for
385 Datong in 2013 shows a similar change in events.
386 There are a number of potential reasons for differences in the dam's impact on high flow and low

flow occurrence in the middle and lower portions of the basin. These reasons include: 1) The topography of the lower basin is dominated by relatively flat plains containing large areas of lakes (a total lake area of 15 000 $km^2$). These lakes naturally attenuate the tributary flows, which impact flows to the main stem. These lake effects on streamflow occur in addition to the impacts imposed by the TGD (Wang et al., 2017). 2) The Yangtze River Basin consists of 21 subbasins (see Figure 1), whose runoff form the main components of inflows at the four downstream hydrological stations. However, spatial heterogeneity in rainfall and catchment responses mean that high flow and low flow events do not necessarily co-occur across the different sub-basins. Perhaps more importantly, the geographical distribution of the precipitation in the Yangtze basin generally increases from northwest to southeast. The dam's locations in the drier part of the upper catchment limits the ability to attenuate floods resulting from storm events in the wetter eastern catchments downstream of the dam.

## 5. Discussion and conclusions

A novel analysis of the impact of the Three Gorges Dam on the summer streamflow in the Yangtze River Basin was presented using canonical correlation analysis to quantify the relationship between the summer streamflow and rainfall across the Yangtze River Basin. The development of the CCA model was motivated by the need to investigate the potential impact of the TGD on summer streamflows by modelling naturalized streamflow and comparing these modelled flows with observed streamflows that are potentially influenced by dam operations.

Another model using rainfalls over each station as covariates was trialed, but the results were not presented in this paper and was hindered by the collinearity and the high dimension of the data. While using a weighted average of the rainfall results in a reduction in variance, this reduction is akin to the natural spatial aggregation represented in the streamflow. The smaller number of sub-basins results in the improved ability to resolve relationships between rainfall and streamflow using CCA compared with using a large number of rainfall stations.

The complexity of the model was constrained by the dearth of readily available data that would have improved the streamflow model such as data on catchment characteristics. The approach of CCA addressed the issues of high dimensionality and large spatial correlations of rainfall records over the whole basin.

Our study was focused on changes in occurrence of high flow and low flow events as informed by streamflow before and after the commencement of operations of the TGD in 2003 and evaluating the impact of dam operations on mitigating the magnitude of these events. The modelled naturalized streamflows generally closely match the observed streamflows for the four stations. Our analysis shows that the streamflows at Yichang station, located at a relative close proximity to the dam, are more noticeably impacted by the TGD as evident by the reduced occurrence of high flow events after 2003. Flows at the other three stations appear to be less impacted by the dam operations due to the increased distance from the dam, the impacts of the lakes located downstream of the dam (Gao et al., 2013; Guo et al., 2012; Lai et al., 2014) and the inability to attenuate flows generated by

runoff in the higher precipitation regions of the lower basin.

Three types of flow regime change were identified and summarized for the post dam period. The dam did not consistently mitigate the occurrence of either high or low flow events at the four locations considered. These changes in flow regimes, sometimes resulting in high or low flows, suggest that the dam operations are unable, in isolation, to alleviate high flow or low flow events in the lower basin as other factors likely influence these events at distant locations. Although many previous studies demonstrate that the flow regulation in impounded rivers are solely a result of dams (Nilsson et al., 2005), Dai et al. (2012)'s analysis support our result, where they concluded that the flows from the upstream basin of the TGD can only account for about 50% of changes in multiyear mean flow over the period 1950-2005 at Datong station, and the discharge from the downstream basin accounted for the remaining difference, where flows were influenced by other factors including the effects of Dongting Lake and Poyang Lake and potentially other changes in water use and management. Further, impacts of the TGD on flow so far downstream would largely be masked by significant inflow to the river from the wetter catchments in the eastern Yangtze River basin.

Several limitations could be addressed in future analyses. First, the model was developed using only summer rainfall excluding other factors, such as the land use change and changes in large scale atmospheric circulation (Xu et al., 2007; Xu et al., 2017). Theoretically, our analysis may result in a conservative estimate of the streamflow but it provides an original approach to assess the likely

444 influence of the TGD on streamflows. Second, the approach of CCA demonstrated here considered
445 both the correlations of the streamflow stations and spatial correlations of rainfall stations over the
446 entire Yangtze Basin. An improved understanding of the complex physical processes that influence
447 precipitation occurrence and runoff characteristics would aid in the understanding of changes in
448 runoff and streamflow in the Yangtze River catchment. Such future endeavors are needed to build
449 on past studies such as Chen et al. (2016) and Lai et al. (2014) that have contributed to this
450 understanding and to further our understanding of the influence of the TGD on streamflow. Third,
451 only seasonal high flow regimes were considered here rather than distinct flood events that likely
452 pose the greatest risk to downstream regions.

453 **Data Availability Statement**

454 Daily streamflow records were obtained from the Changjiang Water Resources Commission. While
455 this data are not publicly available, they may be requested from the Changjiang Water Resources
456 Commission. Daily rainfall data was accessed from the China Meteorological Data Service Center
457 (http://data.cma.cn/en).

458 **Supporting information**

459 Figure S1. Scree plot showing the variance of each PC for the modelled naturalized streamflows.
460 Figure S2. The spatial and temporal distributions of rainfall anomalies over subbasins of the
461 Yangtze River Basin.

462 **References**

463 Bermudez, C.: Culture and demography in the Yangtze River Basin, China,
464 https://hubpages.com/education/Culture-and-Demography-in-the-Yangtze-River-Basin, Accessed on:
465 10 Oct., 2015.
466 Bing, L. F., Shao, Q. Q., & Liu, J. Y. (2012). Runoff characteristics in flood and dry seasons based on wavelet
467 analysis in the source regions of the Yangtze and Yellow rivers. *Journal of Geographical Sciences*,
468 *22*(2), 261-272. https://doi.org/10.1109/RSETE.2011.5964374.
469 Chen, J., Finlayson, B. L., Wei, T., Sun, Q., Webber, M., Li, M., & Chen, Z. (2016). Changes in monthly
470 flows in the Yangtze River, China - With special reference to the Three Gorges Dam. *Journal of*
471 *Hydrology*, *536*, 293-301. https://doi.org/10.1016/j.jhydrol.2016.03.008.
472 Chen, X., Hao, Z., Devineni, N., & Lall, U. (2014). Climate information based streamflow and rainfall
473 forecasts for Huai River basin using hierarchical Bayesian modeling. *Hydrology and Earth System*
474 *Sciences*, *18*(4), 1539-1548. https://doi.org/10.5194/hess-18-1539-2014.
475 Chen, X., You, Q., Sielmann, F., & Ruan, N. (2017). Climate change scenarios for Tibetan Plateau summer
476 precipitation based on canonical correlation analysis. *International Journal of Climatology*, *37*,
477 1310-1321. https://doi.org/10.1002/joc.4778.
478 China Meteorological Data Service Center: CMDC, http://data.cma.cn/en, Accessed on: 1 Mar., 2016.
479 Chu, Z., Zhai, S., & Chen, X. (2006). Changjiang River sediment delivering into the sea in response to water
480 storage of Sanxia Rervoir in 2003. *Acta Oceanologica Sinica*, *25*(2), 71-79
481 Cook, E. R., Briffa, K. R., & Jones, P. D. (1994). Spatial regression methods in dendroclimatology - a review
482 and comparison of 2 techniques. *International Journal of Climatology*, *14*(4), 379-402.
483 https://doi.org/10.1002/joc.3370140404.
484 Cutler, A., & Breiman, L. (1994). Archetypal analysis. *Technometrics*, *36*(4), 338-347.
485 https://doi.org/10.2307/1269949.
486 Dai, Z. J., Chu, A., Stive, M. J. F., & Yao, H. Y. (2012). Impact of the Three Gorges Dam overruled by an
487 extreme climate hazard. *Natural Hazards Review*, *13*(4), 310-316
488 Devineni, N., Lall, U., Pederson, N., & Cook, E. (2013). A tree-ring-based reconstruction of Delaware River
489 Basin streamflow using hierarchical Bayesian regression. *Journal of Climate*, *26*(12), 4357-4374.
490 https://doi.org/10.1175/Jcli-D-11-00675.1.
491 Elsanabary, M. H., & Gan, T. Y. (2015). Weekly streamflow forecasting using a statistical disaggregation
492 model for the Upper Blue Nile Basin, Ethiopia. *Journal of Hydrologic Engineering*, *20*(5), 04014064.
493 https://doi.org/10.1061/(ASCE)HE.1943-5584.0001072.
494 Gao, B., Yang, D. W., & Yang, H. B. (2013). Impact of the Three Gorges Dam on flow regime in the middle
495 and lower Yangtze River. *Quaternary International*, *304*, 43-50.
496 https://doi.org/10.1016/j.quaint.2012.11.023.
497 Gayathri, K. D., Ganasri, B. P., & Dwarakish, G. S. (2015). A review on hydrological models. *Aquatic*
498 *Procedia*, *4*, 1007-1007. https://doi.org/10.1016/j.aqpro.2015.02.126.
499 Gemmer, M., Jiang, T., Su, B., & Kundzewicz, Z. W. (2008). Seasonal precipitation changes in the wet seaon
500 and their influence on flood/drought hazards in the Yangtze River Basin, China. *Quaternary*

*International*, *186*, 12-21. https://doi.org/10.1016/j.quaint.2007.10.001.

Gonzalez, I., Dejean, S., Martin, P. G. P., & Baccini, A. (2008). CCA: An R package to extend canonical correlation analysis. *Journal of Statistical Software*, *23*(12), 1-14. https://doi.org/10.18637/jss.v023.i12.

Guo, H., Hu, Q., Zhang, Q., & Feng, S. (2012). Effects of the Three Gorges Dam on Yangtze River flow and river interaction with Poyang Lake, China: 2003-2008. *Journal of Hydrology*, *416*, 19-27. https://doi.org/10.1016/j.jhydrol.2011.11.027.

Heng, L., & Xu, Z. K. (1999). The 1998 floods of the Yangtze river, China. *Nature & Resources*, *35*(3), 14-21

Ho, M., Lall, U., & Cook, E. R. (2016). Can a paleodrought record be used to reconstruct streamflow?: A case study for the Missouri River Basin. *Water Resources Research*, *52*. https://doi.org/10.1002/2015WR018444.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, *28*, 321-377. https://doi.org/10.1093/biomet/28.3-4.321.

Jiang, L., Ban, X., Wang, X., & Cai, X. (2014). Assessment of hydrologic alterations caused by the Three Gorges Dam in the middle and lower reaches of Yangtze River, China. *Water*, *6*, 1419-1434. https://doi.org/10.3390/w6051419.

Jiang, Z., Pu, J., Yang, H., & Ren, W. (2017). Diagnostic analysis of water vapor transport process during the catastropical flood period over Yangtze River Basin in 1998. *Transactions of Atmospheric Sciences*, *40*(3), 289-298. https://doi.org/10.13878/j.cnki.dqkxxb.201503250113.

Jolliffe, I. T. (2002). *Principle component analysis*. 2nd Ed., Springer.

Kim, S., Noh, H., Jung, J., Jun, H., & Kim, H. S. (2016). Assessment of the impacts of global climate change and regional water projects on streamflow characteristics in the Geum River Basin in Korea. *Water*, *8*(3). https://doi.org/10.3390/w8030091.

Kwon, H. H., Brown, C., Xu, K. Q., & Lall, U. (2009). Seasonal and annual maximum streamflow forecasting using climate information: Application to the Three Gorges Dam in the Yangtze River basin, China. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, *54*(3), 582-595. https://doi.org/10.1623/hysj.54.3.582.

Lai, X. J., Liang, Q. H., Jiang, J. H., & Huang, Q. (2014). Impoundment effects of the Three-Gorges-Dam on flow regimes in two China's largest freshwater lakes. *Water Resources Management*, *28*(14), 5111-5124. https://doi.org/10.1007/s11269-014-0797-6.

Li, D. Y., Lai, X. J., Dong, Z. C., & Luo, X. L. (2016). Effects of the Three Gorges Project on the environment of Poyang Lake. *Polish Journal of Environmental Studies*, *25*(6), 2477-2490

Liang, X., Lettenmaier, D. P., & Wood, E. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research*, *99*(D7), 14415-14428. https://doi.org/10.1029/94JD00483.

McCuen, R. H., Knight, Z., & Cutter, A. G. (2006). Evaluation of the Nash-Sutcliffe efficiency index. *Journal of Hydrologic Engineering*, *11*(6), 597-602. https://doi.org/10.1061/(ASCE)1084-0699(2006)11:6(597).

539 Mei, X., Dai, Z., van Gelder, P. H. A. J. M., & Gao, J. (2015). Linking Three Gorges Dam and downstream
540 hydrological regimes along the Yangtze River, China. *Earth and Space Science*, *2*(4).
541 https://doi.org/10.1002/2014EA000052.
542 Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion
543 of principles. *Journal of Hydrology*, *10*(3), 282-290. https://doi.org/10.1016/0022-1694(70)90255-6.
544 Nilsson, C., Reidy, C. A., Dynesius, M., & Revenga, C. (2005). Fragmentation and flow regulation of the
545 world's large river systems. *Science*, *308*(5720), 405-408. https://doi.org/10.1126/science.1107887.
546 Ouarda, T. B. M. J., Girard, C., Cavadias, G. S., & Bobée, B. (2001). Regional flood frequency estimation
547 with canonical correlation analysis. *Journal of Hydrology*, *254*(1-4), 157-173.
548 https://doi.org/10.1016/S0022-1694(01)00488-7.
549 Pechlivanidis, I. G., Jackson, B. M., Mcintyre, N. R., & Wheater, H. S. (2011). Catchment scale hydrological
550 modelling: A review of model types, calibration approaches and uncertainty analysis methods in the
551 context of recent developments in technology and applications. *Global Nest Journal*, *13*(3), 193-214
552 Qiu, J. (2011). China admits problems with Three Gorges Dam. *Nature*.
553 https://doi.org/10.1038/news.2011.315.
554 Steinschneider, S., Ho, M., Cook, E. R., & Lall, U. (2016). Can PDSI inform extreme precipitation?: An
555 exploration with a 500 year long paleoclimate reconstruction over the US. *Water Resources Research*,
556 *52*(5), 3866-3880. https://doi.org/10.1002/2016WR018712.
557 Steinschneider, S., & Lall, U. (2015). Daily precipitation and tropical moisture exports across the eastern
558 United States: An application of archetypal analysis to identify spatiotemporal structure. *Journal of*
559 *Climate*, *28*(21), 8585-8602. https://doi.org/10.1175/Jcli-D-15-0340.1.
560 Stone, E., & Cutler, A. (1996). Introduction to archetypal analysis of spatio-temporal dynamics. *Physica D*,
561 *96*(1-4), 110-131. https://doi.org/10.1016/0167-2789(96)00016-4.
562 Tramblay, Y., Bouvier, C., Ayral, P. A., & Marchandise, A. (2011). Impact of rainfall spatial distribution on
563 rainfall-runoff modelling efficiency and initial soil moisture conditions estimation. *Natural Hazards*
564 *and Earth System Sciences*, *11*(1), 157-170. https://doi.org/10.5194/nhess-11-157-2011.
565 Vansteenkiste, T., Tavakoli, M., Ntegeka, V., Willems, P., De Smedt, F., & Batelaan, O. (2013). Climate
566 change impact on river flows and catchment hydrology: a comparison of two spatially distributed
567 models. *Hydrological Processes*, *27*(25), 3649-3662. https://doi.org/10.1002/hyp.9480.
568 Wang, J., Sheng, Y., & Wada, Y. (2017). Little impact of the Three Gorges Dam on recent decadal lake decline
569 across China's Yangtze Plain. *Water Resources Research*, *53*(5), 3854-3877.
570 https://doi.org/10.1002/2016WR019817.
571 Wang, J., Sheng, Y. W., Gleason, C. J., & Wada, Y. (2013). Downstream Yangtze River levels impacted by
572 Three Gorges Dam. *Environmental Research Letters*, *8*(4).
573 https://doi.org/10.1088/1748-9326/8/4/044012.
574 Wang, M. J., Zheng, H. B., Xie, X., Fan, D. D., Yang, S. Y., Zhao, Q. H., & Wang, K. (2011). A 600-year
575 flood history in the Yangtze River drainage: Comparison between a subaqueous delta and historical
576 records. *Chinese Science Bulletin*, *56*(2), 188-195. https://doi.org/10.1007/s11434-010-4212-2.

Wilson, R., Cook, E., D'Arrigo, R., Riedwyl, N., Evans, M. N., Tudhope, A., & Allan, R. (2010). Reconstructing ENSO: the influence of method, proxy data, climate forcing and teleconnections. *Journal of Quaternary Science*, *25*(1), 62-78. https://doi.org/10.1002/jqs.1297.

Xu, J. J., Yang, D. W., Yi, Y. H., Lei, Z. D., Chen, J., & Yang, W. J. (2008). Spatial and temporal variation of runoff in the Yangtze River basin during the past 40 years. *Quaternary International*, *186*, 32-42. https://doi.org/10.1016/j.quaint.2007.10.014.

Xu, K., Brown, C., Kwon, H.-H., Lall, U., Zhang, J., Hayashi, S., & Chen, Z. D. (2007). Climate teleconnections to Yangtze river seasonal streamflow at the Three Gorges Dam, China. *International Journal of Climatology*, *27*(6), 771-780. https://doi.org/10.1002/joc.1437.

Xu, S., Zhang, Y., Dou, M., Hua, R., & Zhou, Y. (2017). Spatial distribution of land use change in the Yangtze River Basin and the impact on runoff. *Progress in Geography*, *36*(4), 426-436. https://doi.org/10.18306/dlkxjz.2017.04.004.

Yang, G., & Liang, P. (1982). The relationship between the flow pattern variation over lower troposphere in low latitudes and the persistent drought and flood in the middle and lower Yangtze valleys. *Plateau Meteorology*, *1*(3), 43-51

Yin, H., & Li, C. (2001). Human impact on floods and flood disasters on the Yangtze River. *Geomorphology*, *41*(2), 105-109. https://doi.org/10.1016/S0169-555X(01)00108-8.

Yuan, F., Xie, Z. H., Liu, Q., Yang, H. W., Su, F. G., Liang, X., & Ren, L. L. (2004). An application of the VIC-3L land surface model and remote sensing data in simulating streamflow for the Hanjiang River basin. *Canadian Journal of Remote Sensing*, *30*(5), 680-690. https://doi.org/10.5589/m04-032.

Zhang, L. X., & Zhou, T. J. (2015). Drought over East Asia: A review. *Journal of Climate*, *28*(8), 3375-3399. https://doi.org/10.1175/JCLI-D-14-00259.1.

Zhang, Z. X., Chen, X., Xu, C. Y., Yuan, L. F., Yong, B., & Yan, S. F. (2011). Evaluating the non-stationary relationship between precipitation and streamflow in Nine Major Basins of China during the past 50 years. *Journal of Hydrology*, *409*(1-2), 81-93. https://doi.org/10.1016/j.jhydrol.2011.07.041.

Zhao, J., Li, J., Dai, Z., Wang, Y., & Zhang, A. (2012). Analysis the runoff variation of Yangtze River in Yichang. *Resources Science*, *34*(12), 2306-2315

Zhong, W., Li, R., Liu, Y. Q., & Xu, J. (2018). Effect of different areal precipitation estimation methods on the accuracy of a reservoir runoff inflow forecast model. *IOP Conference Series: Earth and Environmental Science*, *208*, 012043. https://doi.org/10.1088/1755-1315/208/1/012043.

Zong, Y. Q., & Chen, X. Q. (2000). The 1998 flood on the Yangtze, China. *Natural Hazards*, *22*(2), 165-184. https://doi.org/10.1023/A:1008119805106.

611 **Table 1** The frequency of high flow and low flow events pre and post 2003. The modelled
612 frequency for all stations from 2003 to 2014 is only influenced by rainfall, and the observed
613 frequency is influenced by rainfall and TGD.

| Stations | High flow events | | | Low flow events | | |
|---|---|---|---|---|---|---|
| | Observed (1960-2002) | Modelled (2003-2014) | Observed (2003-2014) | Observed (1960-2002) | Modelled (2003-2014) | Observed (2003-2014) |
| Yichang | 0.30 | 0.42 | 0.17 | 0.21 | 0.25 | 0.33 |
| Luoshan | 0.26 | 0.17 | 0.17 | 0.21 | 0.33 | 0.42 |
| Hankou | 0.26 | 0.17 | 0.17 | 0.21 | 0.25 | 0.42 |
| Datong | 0.28 | 0.25 | 0.17 | 0.30 | 0.42 | 0.50 |
| Influence factors | Observations | Rainfall | Rainfall + TGD | Observations | Rainfall | Rainfall + TGD |

614
615

**Table 2** The years when the flow regime changed during the post-dam period. Three cases are considered, which include Case 1: "Risk to none" denotes hydrological extremes from occurrence to none; Case 2: "No risk to risk" denotes hydrological extremes from none to occurrence; Case 3: "High/low flow event to low/high flow event" denotes hydrological extremes changed from high flow event to low flow event or from low flow event to high flow event. The signs "▲" and "▼"with different colors represent the high flow event and low flow event respectively, red color is Case 1, blue is Case 2, green is Case 3, black means no risk changed.

| Year | Yichang | | Luoshan | | Hankou | | Datong | |
|---|---|---|---|---|---|---|---|---|
| | modelled | observed | modelled | observed | modelled | observed | modelled | observed |
| 2003 | ▲ | | | | | | | |
| 2004 | ▼ | ▼ | ▼ | | | | ▼ | ▼ |
| 2005 | ▲ | ▲ | | | | | | |
| 2006 | ▼ | ▼ | ▼ | ▼ | ▼ | ▼ | ▼ | ▼ |
| 2007 | | | | | | | ▼ | |
| 2008 | | ▼ | | ▼ | ▲ | ▼ | ▲ | ▼ |
| 2009 | | | ▼ | ▼ | ▼ | ▼ | ▼ | ▼ |
| 2010 | ▲ | | ▲ | ▲ | ▲ | ▲ | ▲ | ▲ |
| 2011 | ▼ | ▼ | ▼ | ▼ | ▼ | ▼ | ▼ | ▼ |
| 2012 | ▲ | ▲ | | ▲ | | ▲ | | ▲ |
| 2013 | | | | ▼ | | ▼ | ▲ | ▼ |
| 2014 | ▲ | | ▲ | | | | | |

625 **Figure 1** The Yangtze River Basin with the main river stem shown in bold and subbasins delineated
626 and annotated with names shown above the figure. The location of the rainfall stations used in the
627 analysis are shown in green. The four hydrological stations that recorded streamflow are shown in
628 red.

629 **Figure 2** Monthly average streamflow (bar plot) for each streamflow station and rainfall amount
630 averaged over the entire Yangtze River Basin.

631 **Figure 3** Rank correlation map between the spatial average rainfall over subbasins in summer (JJA)
632 and the log-transformed average streamflow for each hydrological station in summer (JJA) in
633 pre-dam period.

634 **Figure 4** Model results with $5^{th}$-$95^{th}$ prediction interval for the streamflows over the Yangtze River
635 Basin calibrated using available data (1960-2002) and the canonical variate of subbasin-scale
636 rainfall. The subplots in first row are the result of developed models using four pairs of canonical
637 variables, and second row is the result of observed and modelled streamflow for each hydrological
638 station.

639 **Figure 5** Boxplots of RE (left plot) and CE (right plot) from cross validation with results from the
640 model using sub-basin scale rainfall data.

641 **Figure 6** Model results and observations with 5-yr moving average of the streamflow over Yangtze
642 River Basin. The horizontal dashed line is the average pre-dam period streamflow for each station
643 and the vertical dashed line represents the time before and after the dam starting to retain water (i.e.

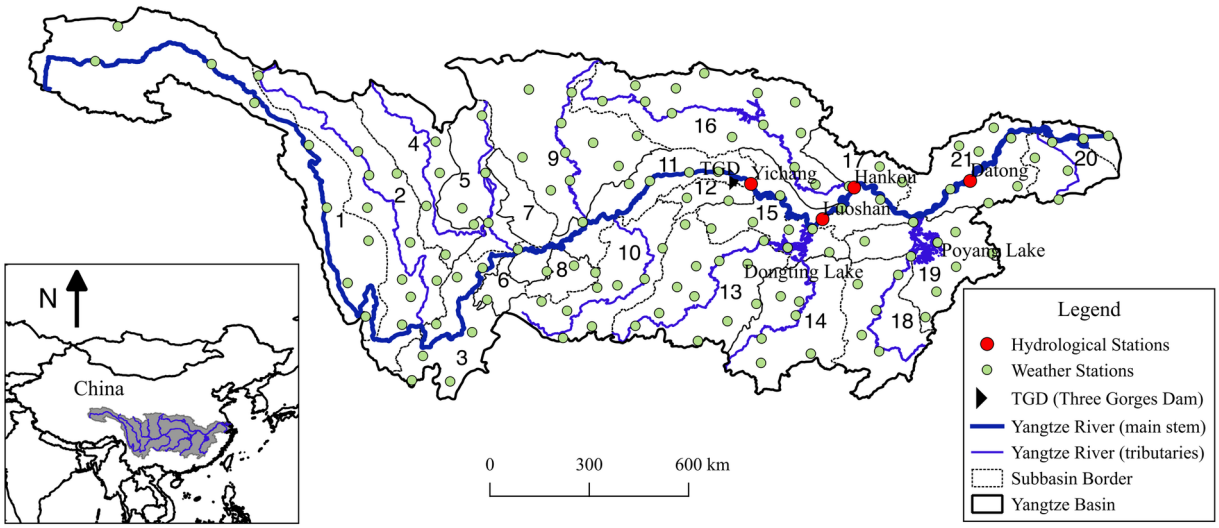
644 Jun. 2003). The blue line with a maker denotes the modelled naturalized streamflow, and the black
645 line with a marker denotes the observed streamflow. The five-year moving averages were also
646 shown with green line and red line representing the modelled naturalized streamflow and observed
647 streamflow, respectively.

648 **Figure 7** The first two PCs of summer average streamflows modelled using CCA across the four
649 hydrological stations and their 5-yr moving average over Yangtze River Basin (left), and the
650 corresponding loadings for each station (right). The percentage variance explained is shown in the
651 parenthesis.

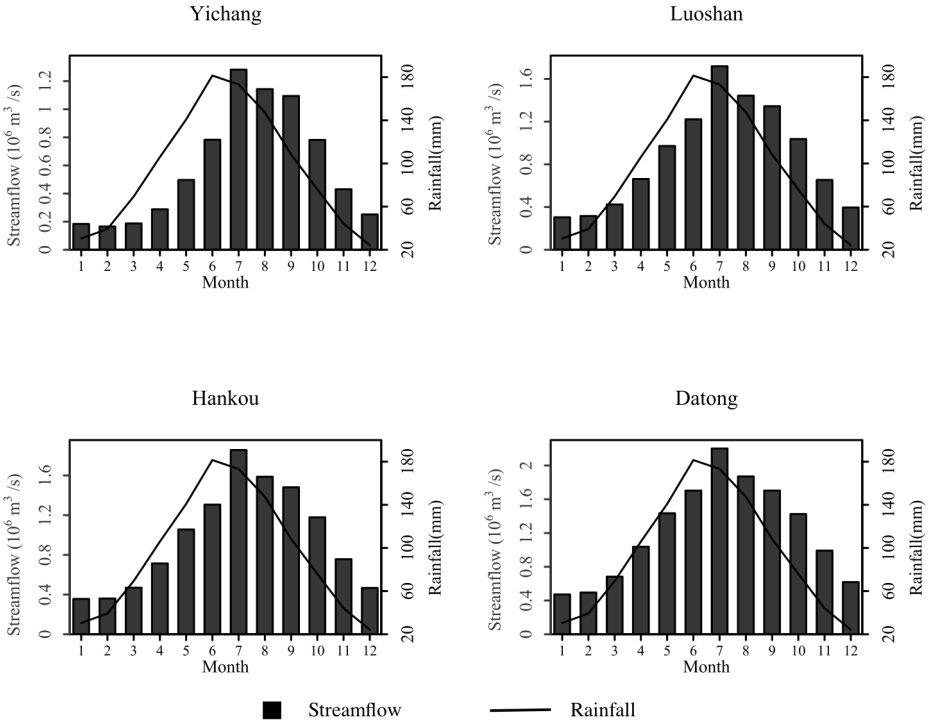
652 **Figure 8** Distribution of high flow and low flow events for the non-regulated flow from 1960 to
653 2014. Observed streamflows are used in the 1960-2002 period, while naturalized flows modelling
654 using canonical correlation analysis are used in the 2003-2014 period.

655 **Figure 9** Comparison of observed and modelled naturalized flows for the four streamflow stations
656 plotted as quantiles of pre-dam flows. The dashed line denotes the $75^{th}$ percentile (high flows are
657 defined as flows in the top quartile) and the solid line denotes the $25^{th}$ percentile (low flows are
658 defined as flows in the bottom quartile).

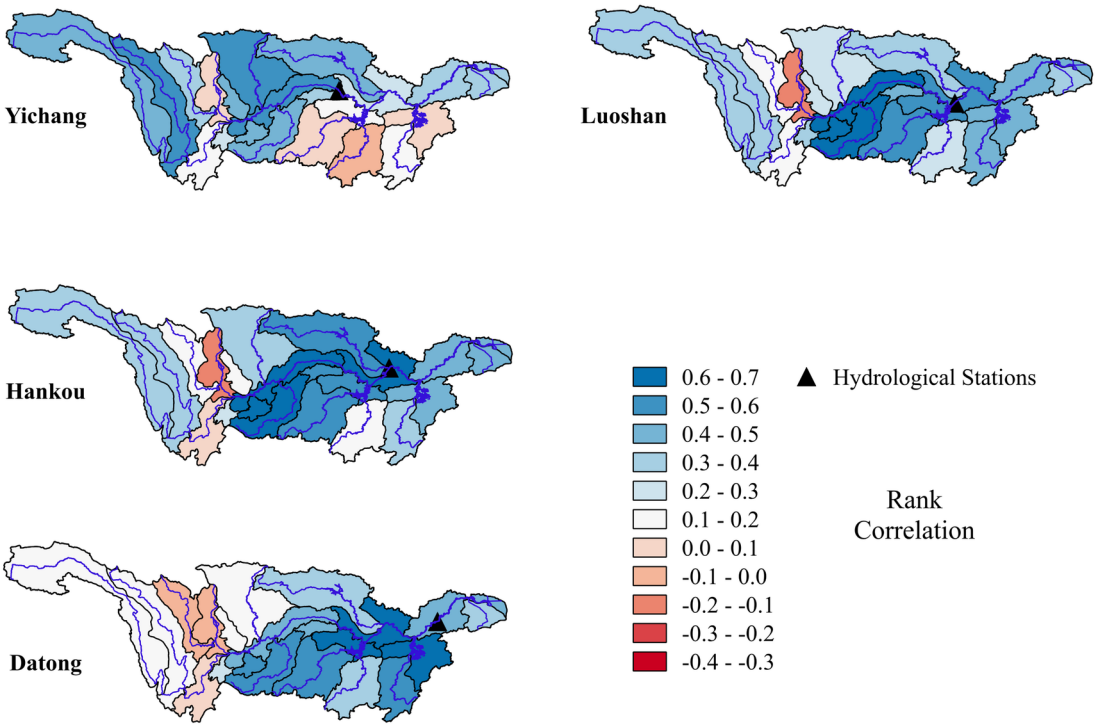
| Name of subbasins | | | | | | |
|---|---|---|---|---|---|---|
| 1-JinshajiangUpper | 2-Yalongjiang | 3-JinshajiangLower | 4-Daduhe | 5-Mintuojiang | 6-Chishui | 7-Fujiang |
| 8-Qijiang | 9-Jialingjiang | 10-Wujiang | 11-Upper | 12-Qingjiang | 13-Yuanshui | 14-Xiangjiang |
| 15-DongtingLake | 16-Hanjiang | 17-Middle | 18-Ganjiang | 19-PoyangLake | 20-Taihu | 21-Lower |



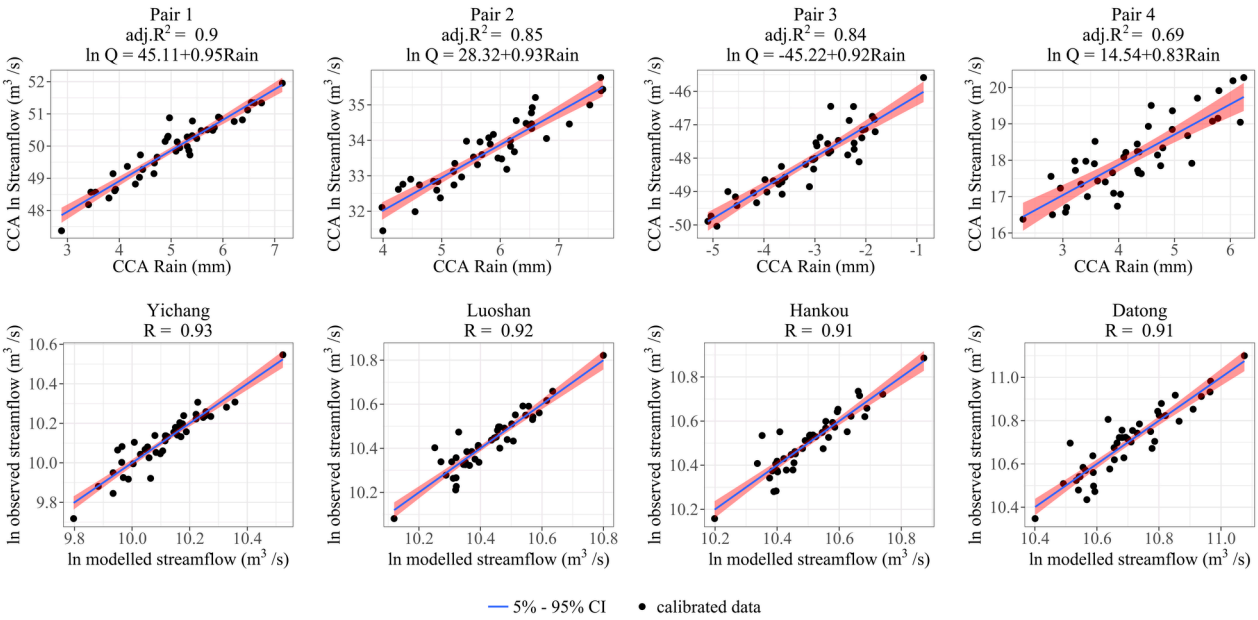

HYP_13619_Figure 1_final.tif



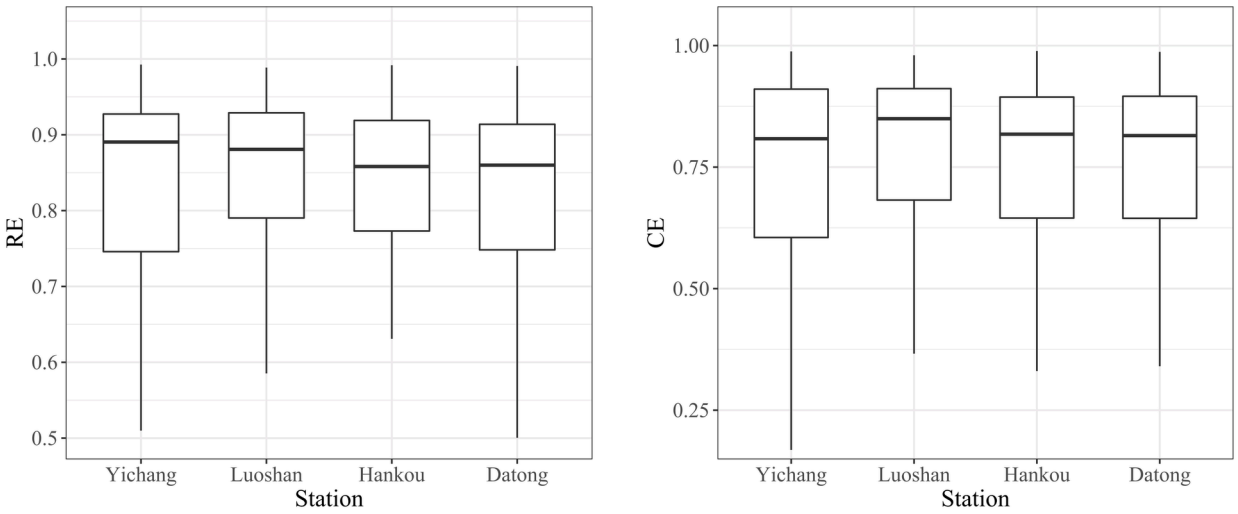

HYP_13619_Figure 2_final.tif



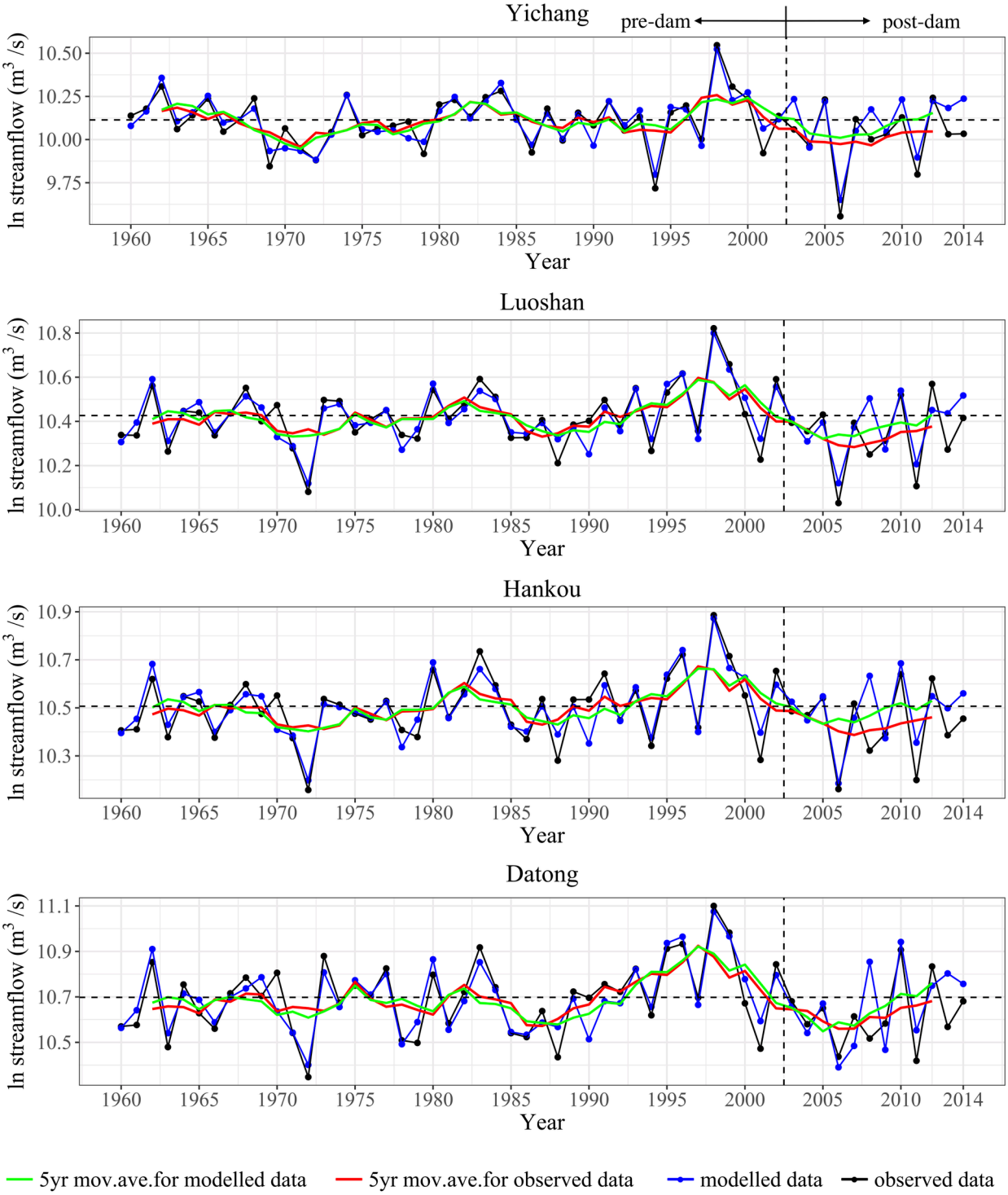

HYP_13619_Figure 3_final.tif



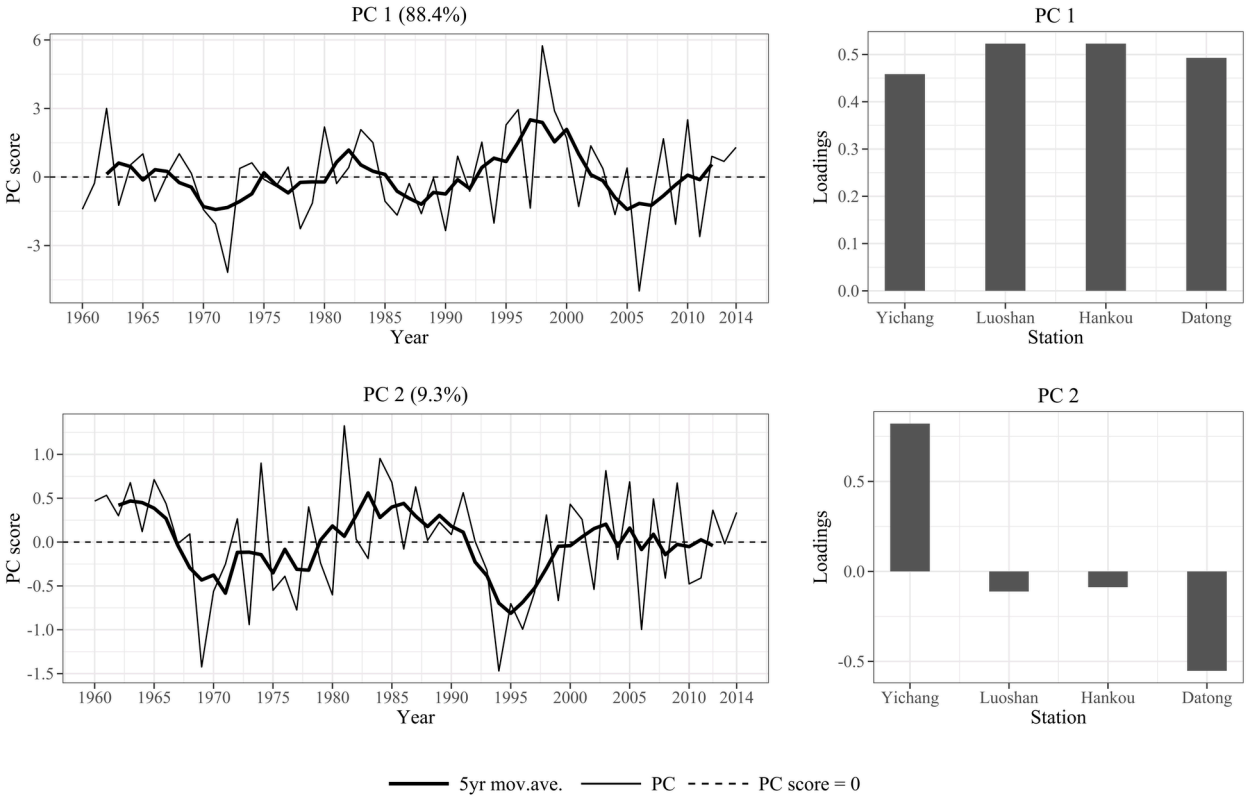

HYP_13619_Figure 4_final.tif



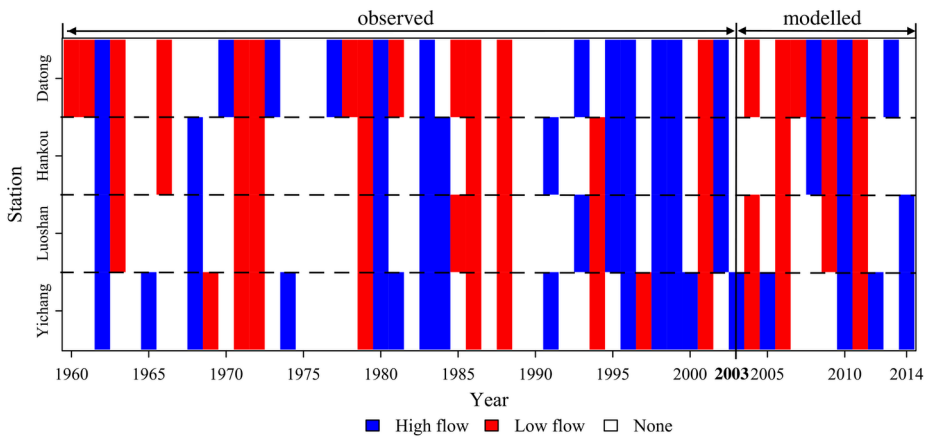

HYP_13619_Figure 5_final.tif



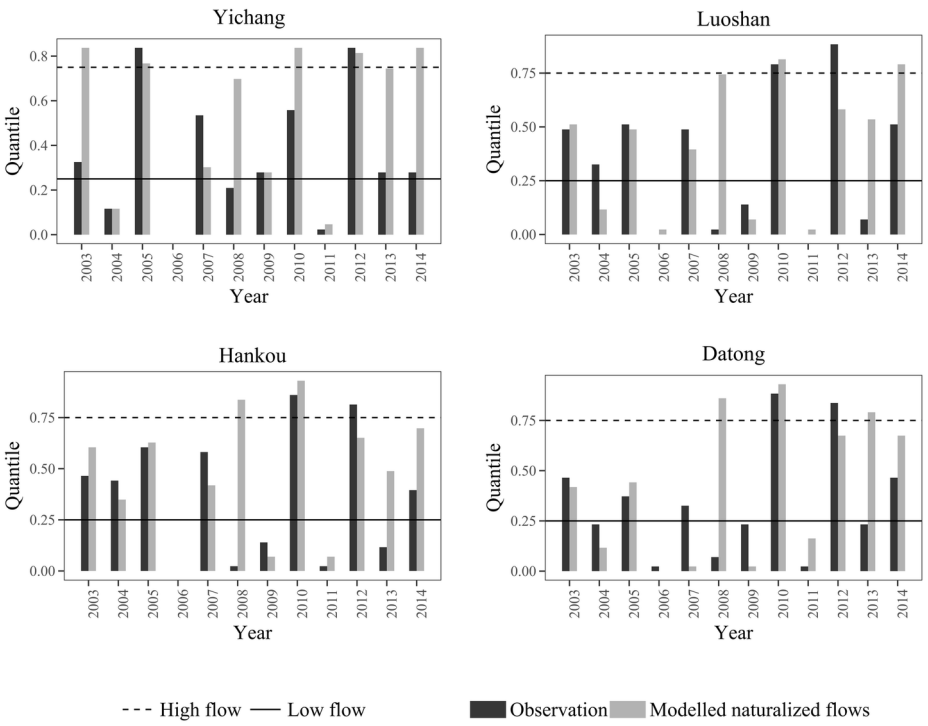

HYP_13619_Figure 6_final.tif




HYP_13619_Figure 7_final.tif




HYP_13619_Figure 8_final.tif




HYP_13619_Figure 9_final.tif