



# 1 The effect of Three Gorges Dam and rainfall on summer

## 2 flow risk over Yangtze River Basin

3 Zhenkuan Su<sup>1</sup>, Zhenchun Hao<sup>1</sup>, Michelle Ho<sup>2</sup>, Upmanu Lall<sup>2,3</sup>, Xun Sun<sup>2,4</sup>, Xi Chen<sup>5</sup>,  
4 Longzeng Yan<sup>6</sup>

<sup>1</sup> State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, China

<sup>7</sup> <sup>2</sup> Columbia Water Center, Columbia University, New York, NY 10027, USA

<sup>8</sup> <sup>3</sup> Department of Earth and Environmental Engineering, Columbia University, New York, NY 10027, USA

<sup>4</sup> Key Laboratory of Geographic Information Science (Ministry of Education), East China Normal University, Shanghai 200241, China

<sup>11</sup> <sup>5</sup> Bureau of Hydrology, Changjiang Water Resources Commission, Wuhan 430010, China

<sup>12</sup> <sup>6</sup>Upper Changjiang River Bureau of Hydrology and Water Resources Survey, Chongqing 400020, China

13 Correspondence to: Zhenkuan Su (zhenkuan.su@gmail.com)

14 **Abstract.** As the largest water conservancy project, Three Gorges Dam starts its impoundment in 2003 and  
15 henceforth the efficient operation of a multi-purpose dam has aroused a great concern on the effectiveness  
16 on flood control and water management over Yangtze River Basin. In this paper, we consider the relationship  
17 between rainfall from 136 weather stations and streamflow from 5 hydrological stations including Cuntan,  
18 Yichang, Luoshan, Hankou and Datong. Meanwhile, the spatial average rainfall over 21 subbasins was  
19 computed. The analysis of the correlation demonstrated that the correlation of spatial average rainfall and  
20 streamflow for each station is consistent with that between rainfall by stations and streamflow. Then, two  
21 options were selected to develop the linear models, including option a) using rainfall by stations to forecast  
22 streamflow and option b) using spatial average rainfall to forecast streamflow. The canonical correlation  
23 analysis enabled a large degree of spatially coherent information of rainfall by linear transforms to maximize  
24 the correlation of rainfall and streamflow for developing linear models. The model resulting from option b)



25 best fits the observations. Coefficient of determination for each model and statistics such as reduction of error,  
26 coefficient of efficiency under cross validation serve as a test of good performance of the model. An analysis  
27 of streamflow characteristics across the stations illustrated that the streamflow of Cuntan and Yichang have  
28 different modes of variability from that of the rest stations below the dam. The comparison of risk type  
29 changes between observations and predictions during the post-dam period showed the dam effectively  
30 relieved the risk of upstream while not always played a positive role in risk mitigation of downstream. The  
31 spatial and temporal patterns of rainfall anomalies over the whole basin helps to explain how the dam impact  
32 the streamflow. Some applications as to possible operational strategies are discussed.

33 **Keywords:** Three Gorges Dam, canonical correlation analysis, flood, drought, risk

34 **1. Introduction**

35 The Yangtze River, also known as Changjiang in China, is the longest river in Asia and plays a large role in  
36 history, culture and economy of China (Figure 1). Its river basin is home to one-third of the country's people.  
37 The whole drainage basin covers an area of 1.8 million km<sup>2</sup> and is comprised of a range of complex terrains  
38 from plateaus to plains. The basin's climate is characterized by complex regional subbasin scale patterns of  
39 precipitation and temperature. The combination of complexity in both climate and terrain results in widely  
40 varying rainfall-runoff process across the basin (Xu et al., 2008). Periodic floods and droughts in rainy  
41 seasons have been experienced in the basin throughout history and have significantly impacted those who  
42 live there and local economy (Heng and Xu, 1999; Wang et al., 2011; Bing et al., 2012; Zhang et al.,  
43 2007; Zhang and Zhou, 2015; Zong and Chen, 2000). For example, 4 970 000 houses were flattened in the  
44 flood of year 1998 and agricultural disaster area was 33 900 km<sup>2</sup> over upstream in the drought of year 2006.  
45 In response to these events, the Three Gorges Dam (TGD) was planned and began construction in 1993 with  
46 water storage commencing in 2003. The primary function of what is now the largest project in the world is  
47 to provide flood control and regulate seasonal variability in streamflow to facilitate water allocation (Gleick,  
48 2009). With the dam in operation since 2003, streamflow from flood seasons are stored and released for  
49 hydropower and water allocations during the dry seasons. In this way, the dam modulates the streamflow



50 distribution over the year (Wang et al., 2013a). Past studies have focused on how the dam has performed in  
51 relation to flood control in addition to other impacts such as water pollution and ecological problem (Wang  
52 et al., 2013a;Gao et al., 2013;Ma et al., 2016;Li and Qiu, 2016). Both the basin-wide drought in 2006 (Dai et  
53 al., 2011;Dai et al., 2012) and the 2016 flood have raised concerns regarding the effectiveness of the dam.  
54 Consequently, analyses of the dam impact on the streamflows are important.

55 Hydrological models such as the physically based Soil and Water Assessment Tool (SWAT) model and  
56 Variable Infiltration Capacity (VIC) model have been widely used in the analyses of streamflow modelling  
57 and their predictions (Zhang et al., 2012a;Kim et al., 2016;Liang et al., 1994;Yuan et al., 2004;Gayathri et  
58 al., 2015). The models are used for modelling both gauged and ungauged catchments, and help inform flood  
59 control operations, water resources management and climate induced runoff forecasting. However,  
60 distributed hydrological models such as MIKE SHE model and SWAT model are limited by aspects such as  
61 large data requirements, unclear parameter uncertainty and inaccurate physical process (Gayathri et al.,  
62 2015;Pechlivanidis et al., 2011). Moreover, the uneven distribution of rainfall stations and the complex  
63 climate and terrain conditions in Yangtze River Basin make it difficult for a single model to accurately  
64 characterize streamflow across the whole basin.

65 In contrast, statistically-based models, such as regression and correlation models, use observation-oriented  
66 methods and seek to find the functional relationship between explanatory and response variables(Elsanabary  
67 and Gan, 2015;Kwon et al., 2009;Lima and Lall, 2010;Chen et al., 2014;Steinschneider et al., 2016). These  
68 models are data orientated and do not take into the consideration process driving the hydrological system and  
69 typically comprised of a small amount of model parameters. The simplicity of these models thus allows one  
70 to apply relatively less variables to the streamflow forecasting.

71 Ultimately, the success of both distributed hydrological models and statistical models depends critically on  
72 the availability of data. In Yangtze River Basin, the streamflow is strongly related to rainfall. For example,  
73 the correlation coefficient between annual precipitation and streamflow at Datong station is 0.85 and the  
74 runoff coefficient is 45% (Zhang et al., 2011). But the use of highly correlated rainfall records among the  
75 adjacent gauges in a single model is problematic. From this point, we will develop a model considering the



76 inter-site correlations for forecasting seasonal streamflow at multiple sites, which is important for river  
77 system applications.

78 Flood and drought events of streamflow are typically characterized by above and below normal streamflow.  
79 The economic and environmental consequences of these events are among the most serious of all natural  
80 disasters. As the dam impounded water since 2003, the streamflow shows a different tendency compared to  
81 the records during the pre-dam period.

82 The objectives of the analyses presented in this paper are:

- 83 1) Investigate the relationship between rainfall and streamflow in flood seasons, and analyze the  
84 characteristics of streamflow over the whole basin and their relationships;
- 85 2) Develop a plausible statistical approach to model the streamflow across the Yangtze River Basin,  
86 considering the high correlations among the intensive rainfall stations as well as the correlations among the  
87 streamflow stations.
- 88 3) Evaluate the current dam operation for risk mitigation and guide future dam operation strategy.

89 **2. Study area and data description**

90 To perform the basin-scale risk analysis, we selected streamflow data from five hydrological stations  
91 including Cuntan, Yichang, Luoshan, Hankou and Datong, along the main stream (Figure 1). Cuntan station  
92 is above the TGD is considered to be the starting point of the reservoir formation, while all other gauges are  
93 located downstream of the dam. The Yichang station is known as the “Gateway to the Three Gorges” and is  
94 located approximately 40 km downstream of the dam and 1 840 km from the estuary. Luoshan is located  
95 close to the outlet of the Dongting Lake, the majority of which is located in the Yangtze floodplain, resulting  
96 in the lake’s inundation level impacts with the gauge’s level. Hankou station provides critical streamflow  
97 monitoring for managing flood control operations given its location downstream of where the Hanjiang River  
98 join the main stem of the Yangtze River. For Datong station, the water level is impacted by the inflow from  
99 the main stream and the upper Poyang Lake. It is also the gauging station used to assess environmental flows  
100 and monitor outflow to the sea. The whole basin is divided into 21 subbasins based on the flow direction and



101 elevation, and the boundaries are obtained from the water authorities of Changing Water Resources  
102 Commission. Their locations have been shown in the Figure 1.  
103 The daily streamflow records cover the period from 1960 to 2014 for each station. This data was also obtained  
104 from the water authorities of Changjiang Water Resources Commission and have been checked and quality  
105 controlled. Daily streamflow was aggregated into monthly streamflow data using the arithmetic mean. The  
106 annual hydrograph shows the seasonality of streamflow (Figure 2), which is monsoon summer (June-July-  
107 August) flow dominating.  
108 The daily rainfall dataset is accessed from the National Meteorological Information Center  
109 (<http://data.cma.cn>). Data at 136 weather stations spans from 1960 to 2014 with less than 5% missing data,  
110 which are linearly interpolated using data from the closest two stations. And, the subbasin rainfall is also used  
111 in this paper. The rainfall over a given subbasin is the spatially averaged value at all rainfall stations within  
112 the subbasin and computed by the Thiessen method. This method assigns an area called Thiessen polygon to  
113 each rainfall station. The Thiessen polygon of a rainfall station is the region for which if we randomly choose  
114 any point in the polygon, that point is closer to this specific station than to any other stations. Thus, the rainfall  
115 recorded by each station should be weighted according to the polygon, it represents. Finally, we consider the  
116 summer JJA average streamflow and the sum of rainfall during the same period as variables for detecting  
117 summer risk changes over Yangtze River Basin.

118 **3. Predictive models for the summer flow**

119 In this section, summer (JJA) streamflow and rainfall were used to estimate the rainfall-runoff relationship  
120 through the use of canonical variables of canonical correlation analysis, and predictive models were designed  
121 using this relationship. The date of June 2003 is used to delineate the change from a free flowing river to an  
122 impounded river as this was the time at which storage in the reservoir behind the TGD began to accumulate.  
123 The analysis is therefore conducted using two time periods of pre and post June 2003 to assess the impact of  
124 the dam on runoff. The models are fit to the data prior to 2003 to quantify the rainfall-runoff relationship  
125 prior to major streamflow regulations. The basics of canonical correlation analysis are introduced in section



126 3.2 and the performance metrics are summarized in section 3.3.

127 **3.1 Model design**

128 We try to use both rainfall data over stations or spatial average rainfall over subbasins to model the average  
129 streamflow in summer by developing a log-linear model. The summer JJA flow was considered as dependent  
130 variable  $y_{t,s}$  in year  $t$  at station  $s$ , and logarithmically transformed as summer streamflow at five stations  
131 were approximately log-normally distributed. Rainfall from 136 weather stations and spatial average rainfall  
132 over 21 subbasins were considered as the independent variables  $x_{t,k}$  in year  $t$  over station or subbasin  $k$ .

133 Initial diagnostic analyses were conducted using the Pearson correlation analysis between the rainfall  $x_{t,k}$   
134 and the log transformed streamflow  $\ln(y_{t,s})$  (shown in Figure 3). Correlations between streamflow  $\ln(y_{t,s})$   
135 and the rainfall  $x_{t,k}$  across the whole basin show that the regions with a strong significantly correlation  
136 using spatial average rainfall are consistent with those using data by stations. It reflects the ability of both the  
137 station by station and the subbasin by subbasin regression methods used for rainfall to preserve local climate  
138 details in the streamflow modelling.

139 In terms of spatial distribution of the correlation map, the values vary over different parts of the whole basin.  
140 For the streamflow of Cuntan station, the most significantly correlated subbasins are Jinshajiangupper,  
141 Yalongjiang and Jialingjiang. These subbasins are the headwater catchments of the Yangtze River and feed  
142 into streamflow at the Cuntan gauge. However, the rainfall in Subbasin Xiangjiang located at downstream of  
143 the dam has negative correlation with the streamflow, indicating that different climate events occurred over  
144 northwestern and southeastern part of Yangtze River Basin. For other streamflow stations, details can be  
145 observed in Figure 3 with blue denoting high positive correlation and red denoting high negative correlation.

146 The preliminary results indicated that streamflow is highly correlated with both subbasin-scale rainfall and  
147 rainfall at specific locations. As a result, two predictor data sets were used to model streamflow as follows:

148 **a)** streamflow at each gauge (Cuntan, Yichang, Luoshan, Hankou and Datong) is modelled using gauged  
149 rainfall data from 136 weather stations as predictors in a regularized canonical correlation analysis;

150 **b)** streamflow at each gauge is modelled using the spatially averaged subbasin rainfall as predictors to model  
151 the streamflow in a canonical correlation analysis (without regularization).



152 The application of the canonical correlation analysis will be discussed in next section 3.2.

153 **3.2 Canonical correlation analysis**

154 In this paper, due to the adjacent locations for the rainfall stations and subbasins, the observed rainfall  $x_{t,k}$   
155 from the weather stations or subbasins do interact with each other, and actually the mutual correlation is very  
156 high, even greater than 0.9 for many stations or subbasins. To solve this problem, we implement canonical  
157 correlation analysis (CCA) (Hotelling, 1936) to process the dimension reduction since the rainfall records  
158 over stations have high spatial correlation. Unlike the methods such as principle component analysis (PCA)  
159 (Jolliffe, 2002) and archetype analysis (AA) (Cutler and Breiman, 1994; Steinschneider and Lall, 2015; Stone  
160 and Cutler, 1996) that are usually used for dimension reduction, CCA takes into account the relationship  
161 between the explanatory and target variables. Consequently, we employed CCA to maximize the correlation  
162 between  $\ln(y_{t,s})$  and  $x_{t,k}$ .

163 In CCA, we consider two data sets (i.e., rainfall  $x_{t,k}$  and streamflow  $\ln(y_{t,s})$ ), where  $x_{t,k}$  is the  
164 explanatory variable  $X$  and the  $\ln(y_{t,s})$  is the target variable  $Y$ . CCA rotates the two sets of variables to  
165 achieve two vectors:  $a$  and  $b$ , such that the random variables  $U = a'X$  and  $V = b'Y$  maximize the  
166 correlation between the rainfall and streamflow.  $U$  and  $V$  are then regarded as the first pair of canonical  
167 variables. Next, one seeks vectors maximizing the same correlation subject to the constraint that they are to  
168 be uncorrelated with the first pair of canonical variables. This gives the second pair of canonical variables.  
169 This procedure will be continued up to the minimum value between the number of explanatory variables and  
170 the number of target variables. We can understand in one way that the CCA transfers the multiple variables  
171 to less variables, so that we can use the correlation between the less variables to represent the relationship of  
172 the explanatory variables and target variables.

173 However, we employ regularized CCA to address the high dimensionality of rainfall variables in the case  
174 where station data are used to prevent overfitting. Regularization is a smoothing process in order to prevent  
175 overfitting and is akin to ridge regression (De Bie and De Moor, 2003). It is typically represented by the  
176 regularization parameter (lambda), which ranges from 0 to 1 with larger lambda indicating a higher degree  
177 of smoothing. The degree of freedom (as defined by Dijkstra (2014)) is limited to a maximum value of  $n-10$ ,



178 where  $n$  is the total number of observations. If the maximum value could not be achieved, lambda was set to  
179 one, otherwise lambda was evaluated to two significant figures. Both regularized CCA and non-regularized  
180 CCA are executed in R using the package “CCA” by Gonzalez et al. (2008).

181 Here, we got the first 5 pairs of canonical variables, which were used to fit the model of streamflow. For each  
182 pair, we can develop one linear regression equation using the linear model as following Eq. (1). Hence, a  
183 system of equations can be developed and this enables one to use rainfall to predict streamflow by solving it.

184 
$$V_i = \beta_i U_i + \alpha_i + \varepsilon_i \quad (1)$$

185 where,  $i$  is order of the 5 pairs of canonical variables ranging from 1 to 5.  $\beta_i$  is the regression coefficient  
186 and  $\alpha_i$  is the intercept in  $i$ -th linear model.  $\varepsilon_i$  is the error term.

187 **3.3 Model performance metrics**

188 Cross validation is used to verify the predictability of a model against overfitting. Here, we applied a leave-  
189  $m$ -out cross validation method, which is processed by leaving out  $m$  randomly selected data from the  
190 observations for validation and the remaining  $n-m$  data are used for model calibration, where  $n$  is the total  
191 number of observation. This procedure was be repeated 100 times, resulting in a set of validation metrics  
192 from all the models. In this study, we calculate two metrics to verify the performance and predictability of a  
193 model. These are the reduction of error (RE) and coefficient of efficiency (CE) (Cook et al., 1994; Wilson et  
194 al., 2010) and are similar to the Nash-Sutcliffe efficiency test. RE and CE are defined as following Eq. (2)  
195 and Eq. (3):

196 
$$RE = 1 - \frac{\sum(x_i - \hat{x}_i)^2}{\sum(x_i - \bar{x}_c)^2} \quad (2)$$

197 
$$CE = 1 - \frac{\sum(x_i - \hat{x}_i)^2}{\sum(x_i - \bar{x}_v)^2} \quad (3)$$

198 Where,  $x_i$  is the observation in year  $i$  in the validation period and  $\hat{x}_i$  is the associated predicted streamflow  
199 data by the model.  $\bar{x}_c$  and  $\bar{x}_v$  are the mean of observations used for calibration and validation, respectively.  
200 The theoretical limits for the values of both statistics range from a maximum of 1 to minus infinity, with  
201 larger value indicating better prediction skill. A positive value indicates that the model skill is better than  
202 using the climatology of the verification period. Some application in the streamflow modelling context could



203 be found in (Chen et al., 2014;Devineni et al., 2013;Ho et al., 2016).

204 **4. Results**

205 Two log-linear models were developed for predicting the summer streamflow at each of the 5 hydrological  
206 stations using the canonical variates derived from the predictors stipulated by option a) gauged rainfall and  
207 b) subbasin scale rainfall. A cursory comparison of the developed models was firstly performed by comparing  
208 the coefficient of determination for each model and then the cross-validation statistics computed over  
209 different blocks of data sets identify the best fitting model of the streamflow. Finally, risk change in summer  
210 streamflow was analyzed and the spatial variability of the rainfall over the whole basin during the post-dam  
211 period was explored to help explain the risk change.

212 **4.1 Model results**

213 Averaged flow in summer at each station along the main stream of Yangtze River was modelled using the  
214 canonical variables from option a) and b), respectively. The modelled results with 5<sup>th</sup>-95<sup>th</sup> prediction interval  
215 have been shown in Figure 4. The development of CCA linear models for each hydrological station either  
216 using option a) or option b) resulted in acceptable models of streamflow, and the mean adjusted R<sup>2</sup> for the  
217 two options are 1.0 and 0.80, respectively. Furthermore, the modelled streamflows and observations have a  
218 strong correlation with average values of 1.0 and 0.92 for option a) and option b), respectively, which indicate  
219 that the models are powerful to model and predict the streamflow.

220 While for the model from option a), the R<sup>2</sup> is close to 1, which seems that the modelled streamflows may  
221 overfit the observations. In fact, the spatial average rainfall over a given subbasin can represent the regional  
222 effect from a climate event and the corresponding model was superior to the model from option a), while a  
223 single rainfall station may misrepresent the amount of rainfall received by the watershed. Several researches  
224 have shown an improvement of rainfall-runoff models contributed by spatially distribution comes from  
225 accounting for rainfall variability (Andreassian et al., 2004;Tramblay et al., 2011). To verify this point, a  
226 detailed description of the cross validation for the model results will be further discussed and evaluate the



227 models in next section 4.2.

228 **4.2 Cross validation**

229 For the cross validation, we randomly draw 5-yr data (about 10% of the total number of observations) without  
230 replacement for validation from the time series and the remaining 38-yr data were used to develop a new  
231 model for estimating the regression parameters. This procedure was repeated 100 times to calculate the  
232 validation metrics of RE and CE. The RE and CE values were calculated for option a) and option b),  
233 respectively, and the results are shown in boxplots (Figure 5).

234 For option a), the median CE value for Yichang is slightly less than zero, indicating that the model predictions  
235 are less skillful than the climatology of the verification period. The median values of RE for Yichang and CE  
236 for Hankou and Datong are close to zero. For option b), we can observe that both the median RE and CE  
237 values for all streamflow sites are greater than zero, indicating that the model from option b) has better  
238 predictive skill than using climatology of the verification period. It reflects that the log-linear model using  
239 spatial average rainfall over subbasins without overfitting is better at predicting the summer flow with greater  
240 accuracy than that using rainfall over stations through an overall assessment. We therefore select the model  
241 from option b) (CCA using spatial average rainfall over subbasins) for further analysis. In next section, we  
242 use predictions of the streamflow during the post-dam period to represent the natural flow without the dam  
243 impact for risk change analysis.

244 **4.3 Risk analysis**

245 Summer average streamflows at the five selected gauges within the Yangtze River Basin were modelled using  
246 subbasin scale rainfall. The modelled streamflow represents naturalized flows and was separated into the pre  
247 and post dam periods to assess potential changes in flood and drought risk. Quantiles of summer flow after  
248 2003 relative to the streamflow during the pre-dam period was then used to help estimate the risk with which  
249 floods or droughts of certain magnitude may occur. Furthermore, the spatial patterns of rainfall over the  
250 whole basin help to interpret and analyze how the dam impact the streamflow for the several stations.

251 **4.3.1 Streamflow characteristics over the whole basin**



252 Both modelled and observed summer streamflow for the period 1960-2014 are shown in log space in Figure  
253 6. The modelled streamflows closely match the observed streamflows in the period prior to dam operation.  
254 In contrast, the modelled streamflows are noticeable different from the regulated streamflows after 2003,  
255 indicating the dam regulates summer flows.  
256 PCA was used to extract the main modes of variability of modelled natural log streamflows across all 5  
257 stations. A scree test was used to inform the number of PCs to retain (the plot is not shown). The first two  
258 PCs were retained that together explain over 95% of the variation (PCs shown in Figure 7).  
259 For PC 1, the same sign in the loading values across all inputs is basically showing a regionally cohesive  
260 signal. Both the flood in 1998 and drought in 2006 can be explained through the temporal variability of PC  
261 scores. The Yangtze River Basin experienced approximately twice the seasonal average rainfall during the  
262 summer of 1998 resulting in widespread flooding. This event is evident in PC 1. While the severe drought in  
263 more than a century struck southwest China and Sichuan Province in the summer of 2006 and water level of  
264 Yangtze was the lowest in the past century, we can see a large negative PC score in 2006.  
265 PC 2 mainly represents the difference in the upper and lower reach of Yangtze River Basin, and largely  
266 impacts the streamflow across the 5 stations by the regional climate events. Unlike in PC 1, there shows  
267 different loadings for the 5 stations, with positive values for Cuntan, Yichang and negative values for the rest  
268 stations locating in the middle and lower reach of Yangtze River. In 1969, the streamflow anomalies from  
269 upper to lower reach increased while the inverse trend was found in 1981. During the mid-1990s, from Cuntan  
270 to Datong station, the streamflows increased and this was detected in PC 2 with negative values.

#### 271 **4.3.2 Risk changes of summer flow during the post-dam period**

272 To quantitatively assess the difference in temporal modes of variability implied by the modelled and observed  
273 streamflow during the post-dam period, the probabilistic analysis is typically used. However, probabilistic  
274 analyses require suitably long records to fit distributions. As an alternative, quantiles of the streamflows were  
275 used to compare the modelled and observed streamflows in the pre and post dam periods at each streamflow  
276 location. Furthermore, we also define a flood occurrence as an event during which the quantile of streamflow  
277 is greater than the 75<sup>th</sup> percentile JJA flow of the pre-dam period, while a drought occurrence is less than the



278 25<sup>th</sup> percentile JJA flow of the pre-dam period. The threshold lines were plotted in Figure 8 with dashed line  
279 denoting the flood and black line denoting the drought.

280 The occurrence of flood and drought in the post dam period were assessed using the modelled streamflows,  
281 to represent theoretical flood and drought occurrences without the dam, and observed streamflows, to  
282 quantify the flood and drought mitigation capacities enabled by dam operations. These results are  
283 summarized in Figure 8 and Table 1. We can see that the dam mitigates the flood risk for Cuntan and Yichang  
284 over the upper reach purely considering the number of years (e.g., 2003, 2010, 2013). In contrast, the  
285 occurrence of floods and droughts are not noticeably alleviated by dam operations in the middle and lower  
286 reaches as these streamflows are impacted by Dongting Lake and Poyang Lake.

287 The dam influence may be qualitatively assessed by considering the spatial pattern of subbasin rainfall  
288 anomalies and observed streamflows in the period after 2003 (rainfall anomalies map shown in Figure 9). In  
289 both 2003 and 2010, anomalously high rainfalls were experienced in the subbasins upstream (e.g.,  
290 Jialingjiang) of the Cuntan and Yichang gauges resulting in flood conditions. However, flood impacts were  
291 alleviated through dam operations with lower observed flows at the Yichang gauge located downstream of  
292 the dam relative to the modelled flows. Decreased flows were also observed at the Cuntan gauge upstream  
293 of the dam in 2003, likely a result of reservoir backwater as the reservoir filled. However, streamflow at the  
294 remaining three gauges further downstream were not notably impacted by dam operations.

295 In 2008, the rainfall over the upstream portion of the basin were close to the average while lower basin  
296 experienced above average rainfall. What is worse is that a spurious “drought” is observed at Cuntan as the  
297 gauge was impounded by the reservoir and this resulted in that the backwater blocked the flow and reduced  
298 the flow velocity. Despite the above average rainfalls in the lower basin, reduced outflows from the dam as  
299 a result of average inflows into the dam meant that the region downstream of the dam was in an effective  
300 drought.

301 In 2012 and 2013, the rainfall over the upstream is approximately average, and the dam operations do not  
302 cause large changes in hydrological risk for Cuntan and Yichang. Conversely, negative rain anomalies over  
303 some subbasins surrounding the two lakes largely impacted the outflow from the lakes into the main Yangtze



304 stream in 2013. The abnormal rainfall and river-lake interaction resulted in smaller values of observations.

305 This was also detected in 2012 but not obvious in the spatial pattern.

306 **5. Conclusions and discussions**

307 Human activities and climate change irrefutably have a profound influence on the hydrologic cycle and  
308 intertwine with each influencing and affecting the behavior of the other (Wang et al., 2013b). This study  
309 examined the potential to use the basin-scale rainfall to directly model the streamflow and evaluate the effect  
310 of dam operations on summer flow risk. We considered two options in the model designing, which included  
311 the model using rainfall by stations and the model using spatially averaged rainfall by subbasins. Instead of  
312 the distributed physically-based models, a single statistical model was developed and it used relatively less  
313 variables to model streamflows but considered the spatial correlation of streamflow over the whole basin.

314 The CCA approach enabled a large degree of spatially coherent information of rainfall by linear transforms,  
315 with which the correlated rainfall and streamflow could be expressed using linear equations. The best fitting  
316 model was selected to be the model using spatial average rainfall over subbasins and cross-validation method  
317 served as a test of better performance with the result consistent with the expectations from fitting. This  
318 method can be extended to apply in the area where less available data was obtained and each independent  
319 variable has a high correlation with others.

320 The result of our study presented that, since 2003, there has been a certain degree of difference between the  
321 observations and the natural flows without dam running. Floods and droughts experienced in the post dam  
322 period were amplified, driven, or alleviated. In summer, the rainfall is the headwater of Yangtze's streamflow  
323 and the dam's behaviors are typically taken forcedly by the operators. Therefore, the dam regulation has the  
324 most significantly impact on the streamflow, compared to other driven factors such as land use change and  
325 suspended sediment. The analysis of variance of streamflow explained by each predictor could be the next  
326 research topic.

327 Changes in discharge and water level of the river change the blocking force of the river on the outflows from  
328 many large hydrological entities along the Yangtze River such as Dongting Lake, Poyang Lake and the key



329 tributaries (Guo et al., 2012; Zhang et al., 2012b; Dai et al., 2016; Lai et al., 2014). The TGD changes the  
330 Yangtze River discharge and water level and subsequently disturb the interrelations of the river and lakes,  
331 and affects the river flow and hydrologic cycle of the lakes. Similarly, in Zhan et al. (2015)'s study, they took  
332 triplicate samples of the river and Dongting Lake to determine the relationship between river-lake interaction,  
333 and found that during 3 different dispatching periods including water-supply, flood-storage and water-storage  
334 dispatching period, the lake is recharged through different rivers. Through the dam operation, it should be  
335 ensured that the water level and discharge over both upstream and downstream are not too low in dry seasons  
336 and not too high in wet seasons according to the rainfall distribution over the whole basin, and thus prevent  
337 the floods and droughts.

338 In this paper, we used the rainfall to model the streamflow and this method could also be applied to any  
339 regional areas where are not easy to directly relate the pre-season climate and streamflows. The approach  
340 demonstrated here enabled one to develop and test both the rainfall induced variations and changes due to  
341 the human activities on a river. Much research is also directed towards an analysis of the river-lake interaction  
342 in a changing spatial pattern of climate.

### 343 **Acknowledgements**

344 We would like to sincerely thank the editors and reviewers who provided valuable comments, Changjiang  
345 Water Resources Commission for providing the streamflow data, National Meteorological Information  
346 Center (<http://data.cma.cn>) for providing the rainfall dataset, our colleagues working at Hohai University and  
347 Columbia Water Center of Columbia University for going over the result. This work was partly supported by  
348 China Scholarship Council. It was also funded by the National Key Research Projects (Grant NO.  
349 2016YFC0402704) and National Natural Science Foundation of China (Grant NO. 41371047).

### 350 **References**

351 Andreassian, V., Oddos, A., Michel, C., Anctil, F., Perrin, C., and Loumagne, C.: Impact of spatial  
352 aggregation of inputs and parameters on the efficiency of rainfall-runoff models: A theoretical study using  
353 chimera watersheds, *Water Resour. Res.*, 40, doi:10.1029/2003wr002854, 2004.



354 Bing, L. F., Shao, Q. Q., and Liu, J. Y.: Runoff characteristics in flood and dry seasons based on wavelet  
355 analysis in the source regions of the Yangtze and Yellow rivers, *J. Geogr. Sci.*, 22, 261-272, 2012.

356 Chen, X., Hao, Z., Devineni, N., and Lall, U.: Climate information based streamflow and rainfall forecasts  
357 for Huai River basin using hierarchical Bayesian modeling, *Hydrol. Earth. Syst. Sc.*, 18, 1539-1548,  
358 doi:10.5194/hess-18-1539-2014, 2014.

359 Cook, E. R., Briffa, K. R., and Jones, P. D.: Spatial regression methods in dendroclimatology - a review and  
360 comparison of 2 techniques, *Int. J. Climatol.*, 14, 379-402, doi:10.1002/joc.3370140404, 1994.

361 Cutler, A., and Breiman, L.: Archetypal analysis, *Technometrics*, 36, 338-347, doi:10.2307/1269949, 1994.

362 Dai, L. Q., Mao, J. Q., Wang, Y., Dai, H. C., Zhang, P. P., and Guo, J. L.: Optimal operation of the Three  
363 Gorges Reservoir subject to the ecological water level of Dongting Lake, *Environ. Earth. Sci.*, 75, 2016.

364 Dai, Z. J., Chu, A., Stive, M., Du, J. Z., and Li, J. F.: Is the Three Gorges Dam the cause behind the extremely  
365 low suspended sediment discharge into the Yangtze (Changjiang) Estuary of 2006?, *Hydrolog. Sci. J.*, 56,  
366 1280-1288, 2011.

367 Dai, Z. J., Chu, A., Stive, M. J. F., and Yao, H. Y.: Impact of the Three Gorges Dam overruled by an extreme  
368 climate hazard, *Nat. Hazards Rev.*, 13, 310-316, 2012.

369 De Bie, T., and De Moor, B.: On the regularization of canonical correlation analysis, *International  
370 Symposium on ICA and BSS*, Nara, Japan, 2003.

371 Devineni, N., Lall, U., Pederson, N., and Cook, E.: A tree-ring-based reconstruction of Delaware River Basin  
372 streamflow using hierarchical bayesian regression, *J. Climate*, 26, 4357-4374, doi:10.1175/Jcli-D-11-00675.1,  
373 2013.

374 Dijkstra, T. K.: Ridge regression and its degrees of freedom, *Qual. Quant.*, 48, 3185-3193,  
375 doi:10.1007/s11135-013-9949-7, 2014.

376 Elsanabary, M. H., and Gan, T. Y.: Weekly streamflow forecasting using a statistical disaggregation model  
377 for the Upper Blue Nile Basin, Ethiopia, *J. Hydrol. Eng.*, 20, 2015.

378 Gao, B., Yang, D. W., and Yang, H. B.: Impact of the Three Gorges Dam on flow regime in the middle and  
379 lower Yangtze River, *Quatern. Int.*, 304, 43-50, 2013.



380 Gayathri, K. D., Ganasri, B. P., and Dwarakish, G. S.: A review on hydrological models, *Aquatic Procedia*,  
381 4, 1007-1007, doi:10.1016/j.aqpro.2015.02.126, 2015.

382 Gleick, P. H.: Three Gorges Dam Project, Yangtze River, China, *Water Brief* 3, 2009.

383 Gonzalez, I., Dejean, S., Martin, P. G. P., and Baccini, A.: CCA: An R package to extend canonical correlation  
384 analysis, *J. Stat. Softw.*, 23, 1-14, 2008.

385 Guo, H., Hu, Q., Zhang, Q., and Feng, S.: Effects of the Three Gorges Dam on Yangtze River flow and river  
386 interaction with Poyang Lake, China: 2003-2008, *J. Hydrol.*, 416, 19-27, 2012.

387 Heng, L., and Xu, Z. K.: The 1998 floods of the Yangtze river, China, *Nature Resour.*, 35, 14-21, 1999.

388 Ho, M., Lall, U., and Cook, E. R.: Can a paleodrought record be used to reconstruct streamflow?: A case  
389 study for the Missouri River Basin, *Water Resour. Res.*, 52, doi:10.1002/2015WR018444, 2016.

390 Hotelling, H.: Relations between two sets of variates, *Biometrika*, 28, 321-377, doi:10.1093/biomet/28.3-  
391 4.321, 1936.

392 Jolliffe, I. T.: Principle component analysis, 2nd ed., Springer, N.Y., 2002.

393 Kim, S., Noh, H., Jung, J., Jun, H., and Kim, H. S.: Assessment of the impacts of global climate change and  
394 regional water projects on streamflow characteristics in the Geum River Basin in Korea, *Water-Sui*, 8, 2016.

395 Kwon, H. H., Brown, C., Xu, K. Q., and Lall, U.: Seasonal and annual maximum streamflow forecasting  
396 using climate information: Application to the Three Gorges Dam in the Yangtze River basin, China, *Hydrolog.  
397 Sci. J.*, 54, 582-595, 2009.

398 Lai, X. J., Liang, Q. H., Jiang, J. H., and Huang, Q.: Impoundment effects of the Three-Gorges-Dam on flow  
399 regimes in two China's largest freshwater lakes, *Water Resour. Manag.*, 28, 5111-5124, 2014.

400 Li, F. F., and Qiu, J.: Incorporating ecological adaptation in a multi-objective optimization for the Three  
401 Gorges Reservoir, *J. Hydroinform.*, 18, 564-578, 2016.

402 Liang, X., Lettenmaier, D. P., and Wood, E.: A simple hydrologically based model of land surface water and  
403 energy fluxes for general circulation models, *Journal of Geophysical Research*, 99, 14415-14428, 1994.

404 Lima, C. H. R., and Lall, U.: Climate informed monthly streamflow forecasts for the Brazilian hydropower  
405 network using a periodic ridge regression model, *J. Hydrol.*, 380, 438-449, 2010.



406 Ma, Y. Q., Qin, Y. W., Zheng, B. H., Zhao, Y. M., Zhang, L., Yang, C. C., Shi, Y., and Wen, Q.: Three Gorges  
407 Reservoir: Metal pollution in surface water and suspended particulate matter on different reservoir operation  
408 periods, *Environ. Earth. Sci.*, 75, 2016.

409 Pechlivanidis, I. G., Jackson, B. M., McIntyre, N. R., and Wheater, H. S.: Catchment scale hydrological  
410 modelling: A review of model types, calibration approaches and uncertainty analysis methods in the context  
411 of recent developments in technology and applications, *Global Nest J.*, 13, 193-214, 2011.

412 Steinschneider, S., and Lall, U.: Daily precipitation and tropical moisture exports across the eastern United  
413 States: An application of archetypal analysis to identify spatiotemporal structure, *J. Climate*, 28, 8585-8602,  
414 doi:10.1175/Jcli-D-15-0340.1, 2015.

415 Steinschneider, S., Ho, M., Cook, E. R., and Lall, U.: Can PDSI inform extreme precipitation?: An  
416 exploration with a 500 year long paleoclimate reconstruction over the US, *Water Resour. Res.*, 52, 3866-  
417 3880, 2016.

418 Stone, E., and Cutler, A.: Introduction to archetypal analysis of spatio-temporal dynamics, *Physica D*, 96,  
419 110-131, doi:10.1016/0167-2789(96)00016-4, 1996.

420 Tramblay, Y., Bouvier, C., Ayral, P. A., and Marchandise, A.: Impact of rainfall spatial distribution on rainfall-  
421 runoff modelling efficiency and initial soil moisture conditions estimation, *Nat. Hazard. Earth Sys.*, 11, 157-  
422 170, doi:10.5194/nhess-11-157-2011, 2011.

423 Wang, J., Sheng, Y. W., Gleason, C. J., and Wada, Y.: Downstream Yangtze River levels impacted by Three  
424 Gorges Dam, *Environ. Res. Lett.*, 8, 2013a.

425 Wang, M. J., Zheng, H. B., Xie, X., Fan, D. D., Yang, S. Y., Zhao, Q. H., and Wang, K.: A 600-year flood  
426 history in the Yangtze River drainage: Comparison between a subaqueous delta and historical records,  
427 *Chinese Sci. Bull.*, 56, 188-195, 2011.

428 Wang, Y., Ding, Y. J., Ye, B. S., Liu, F. J., Wang, J., and Wang, J.: Contributions of climate and human  
429 activities to changes in runoff of the Yellow and Yangtze rivers from 1950 to 2008, *Sci China Earth Sci*, 56,  
430 1398-1412, 2013b.

431 Wilson, R., Cook, E., D'Arrigo, R., Riedwyl, N., Evans, M. N., Tudhope, A., and Allan, R.: Reconstructing



432 ENSO: the influence of method, proxy data, climate forcing and teleconnections, *J. Quaternary Sci.*, 25, 62-  
433 78, doi:10.1002/jqs.1297, 2010.

434 Xu, J. J., Yang, D. W., Yi, Y. H., Lei, Z. D., Chen, J., and Yang, W. J.: Spatial and temporal variation of runoff  
435 in the Yangtze River basin during the past 40 years, *Quatern. Int.*, 186, 32-42, 2008.

436 Yuan, F., Xie, Z. H., Liu, Q., Yang, H. W., Su, F. G., Liang, X., and Ren, L. L.: An application of the VIC-3L  
437 land surface model and remote sensing data in simulating streamflow for the Hanjiang River basin, *Can. J.*  
438 *Remote Sens.*, 30, 680-690, 2004.

439 Zhan, L. C., Chen, J. S., Zhang, S. Y., Huang, D. W., and Li, L.: Relationship between Dongting Lake and  
440 surrounding rivers under the operation of the Three Gorges Reservoir, China, *Isot. Environ. Health. S.*, 51,  
441 255-270, 2015.

442 Zhang, L. X., and Zhou, T. J.: Drought over East Asia: A review, *J. Climate*, 28, 3375-3399, 2015.

443 Zhang, N., He, H. M., Zhang, S. F., Jiang, X. H., Xia, Z. Q., and Huang, F.: Influence of reservoir operation  
444 in the upper reaches of the Yangtze River (China) on the inflow and outflow regime of the TGR-based on the  
445 improved SWAT model, *Water Resour. Manag.*, 26, 691-705, 2012a.

446 Zhang, Q., Chen, J. Q., and Becker, S.: Flood/drought change of last millennium in the Yangtze Delta and its  
447 possible connections with Tibetan climatic changes, *Global Planet. Change*, 57, 213-221, 2007.

448 Zhang, Q., Li, L., Wang, Y. G., Werner, A. D., Xin, P., Jiang, T., and Barry, D. A.: Has the Three-Gorges Dam  
449 made the Poyang Lake wetlands wetter and drier?, *Geophys. Res. Lett.*, 39, 2012b.

450 Zhang, Z. X., Chen, X., Xu, C. Y., Yuan, L. F., Yong, B., and Yan, S. F.: Evaluating the non-stationary  
451 relationship between precipitation and streamflow in nine major basins of China during the past 50 years, *J.*  
452 *Hydrol.*, 409, 81-93, 2011.

453 Zong, Y. Q., and Chen, X. Q.: The 1998 flood on the Yangtze, China, *Nat. Hazards*, 22, 165-184, 2000.

454

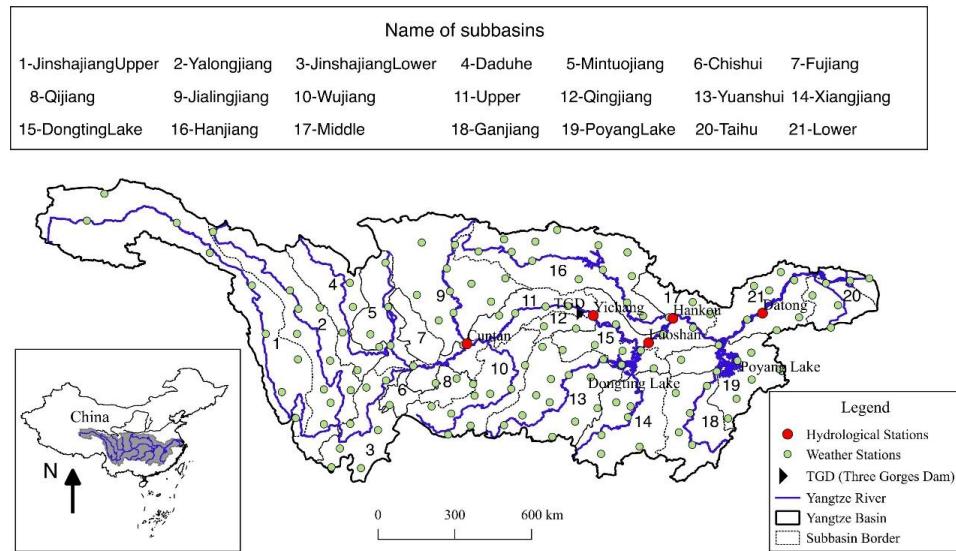
455



456 **Table 1: The years when the risk type changed during the post-dam period. Three cases are considered, which**  
457 **include Case 1: “None→Risk” denotes hydrological extremes from none to having; Case 2: “Risk→None” denotes**  
458 **hydrological extremes from having to none; Case 3: “Risk→Opposite Risk” denotes hydrological extremes**  
459 **changed from flood to drought or from drought to flood.**

Stations	None→Risk	Risk→None	Risk→Opposite Risk
Cuntan	08	03,10,13	14
Yichang	08	03,10	14
Luoshan	08,12,13	04,14	
Hankou	12,13		08
Datong	12	07	08,13

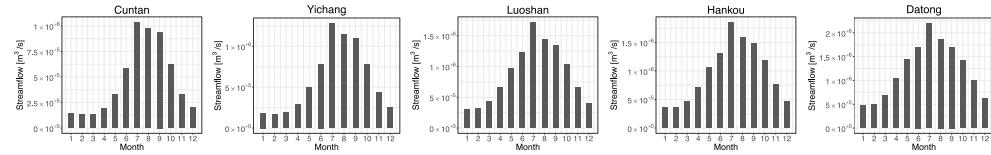
460  
461



462

463 **Figure 1: The locations of the rainfall stations, hydrological stations and subbasins used in the analyses, the**  
464 **Yangtze River Basin shown in grey in the bottom left. The subbasins are annotated with number and names are**  
465 **shown in the top of this figure.**

466

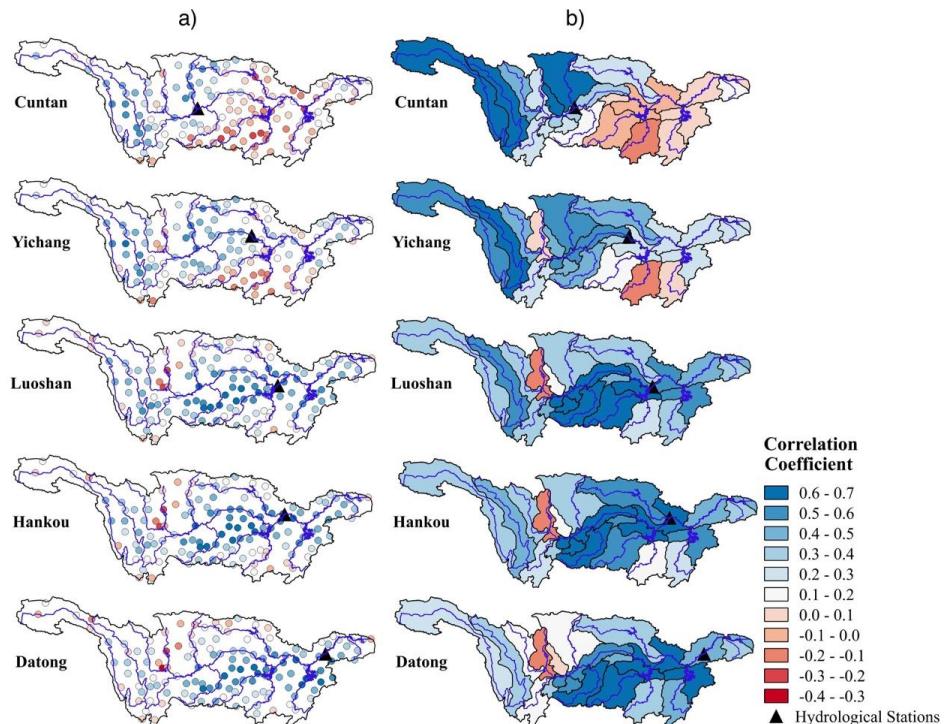


467

468 **Figure 2: Monthly average streamflow for each station over Yangtze River Basin.**

469

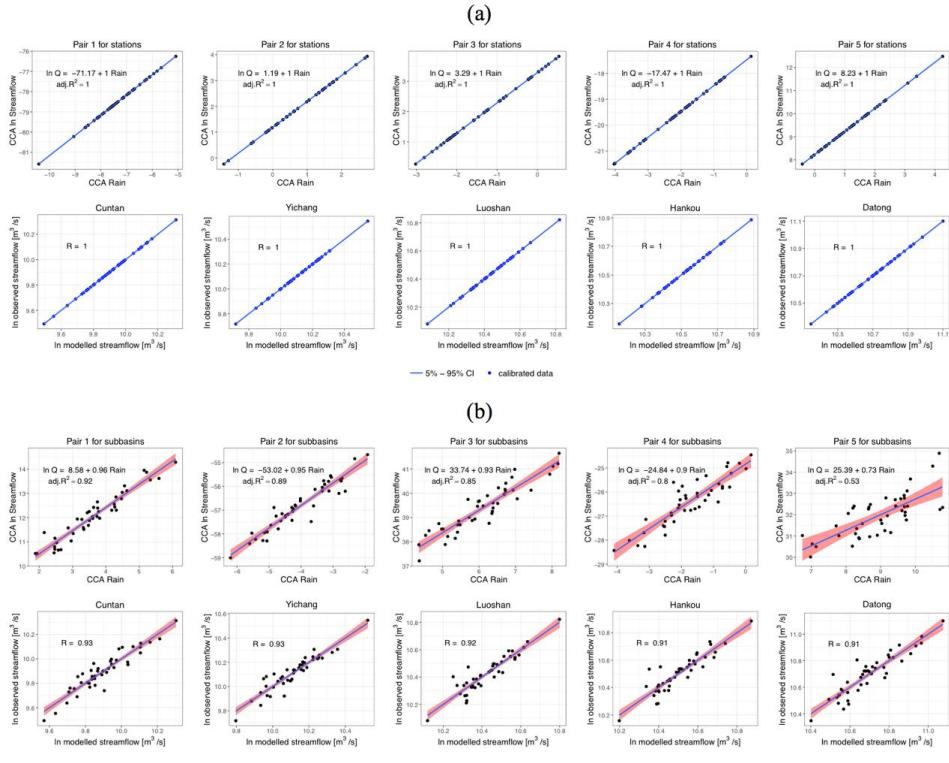
470



471

472 **Figure 3: Pearson correlation map between the average streamflow in summer (JJA) for each hydrological station**  
473 **and a) rainfall over stations, b) spatial average rainfall over subbasins in summer (JJA).**

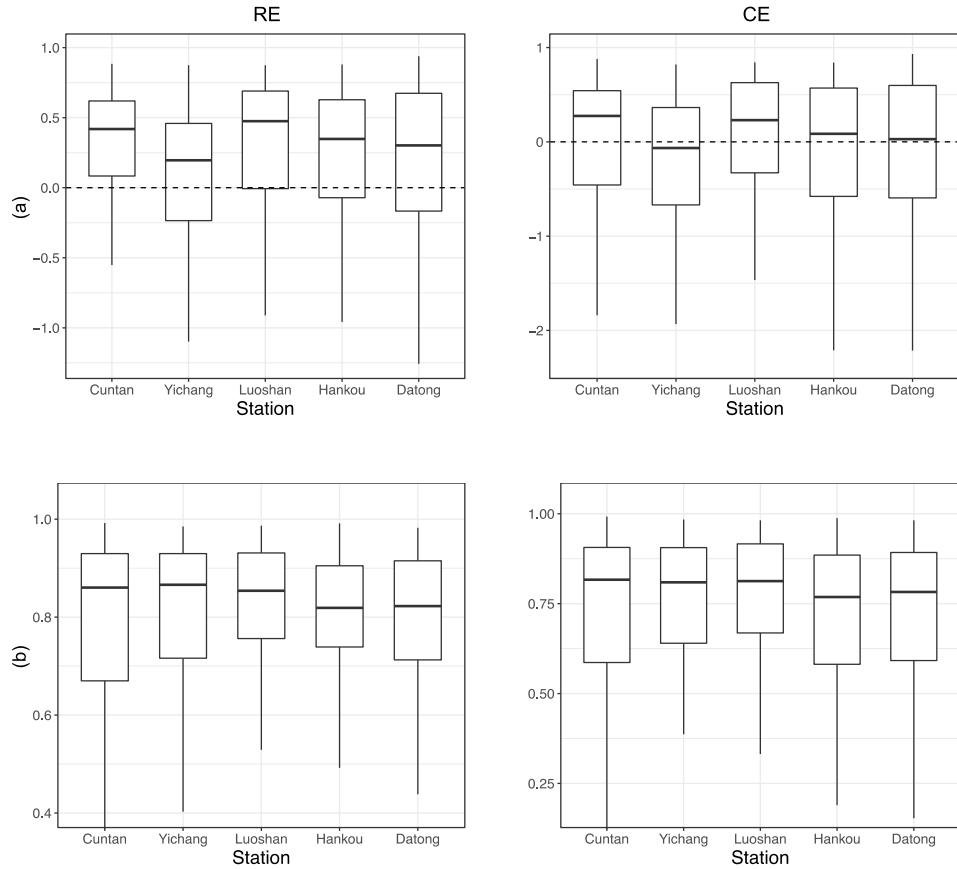
474



475

476 **Figure 4: Model results with 5<sup>th</sup>-95<sup>th</sup> prediction interval for the streamflows over the Yangtze River Basin**  
 477 calibrated using available data (1960-2002) and the canonical variate of rainfall by stations (a) and by subbasins  
 478 (b), respectively. The first row in both (a) and (b) is the result of developed models using 5 pairs of canonical  
 479 variables, and second row is the result of observed and modelled streamflow for each hydrological station.

480  
 481

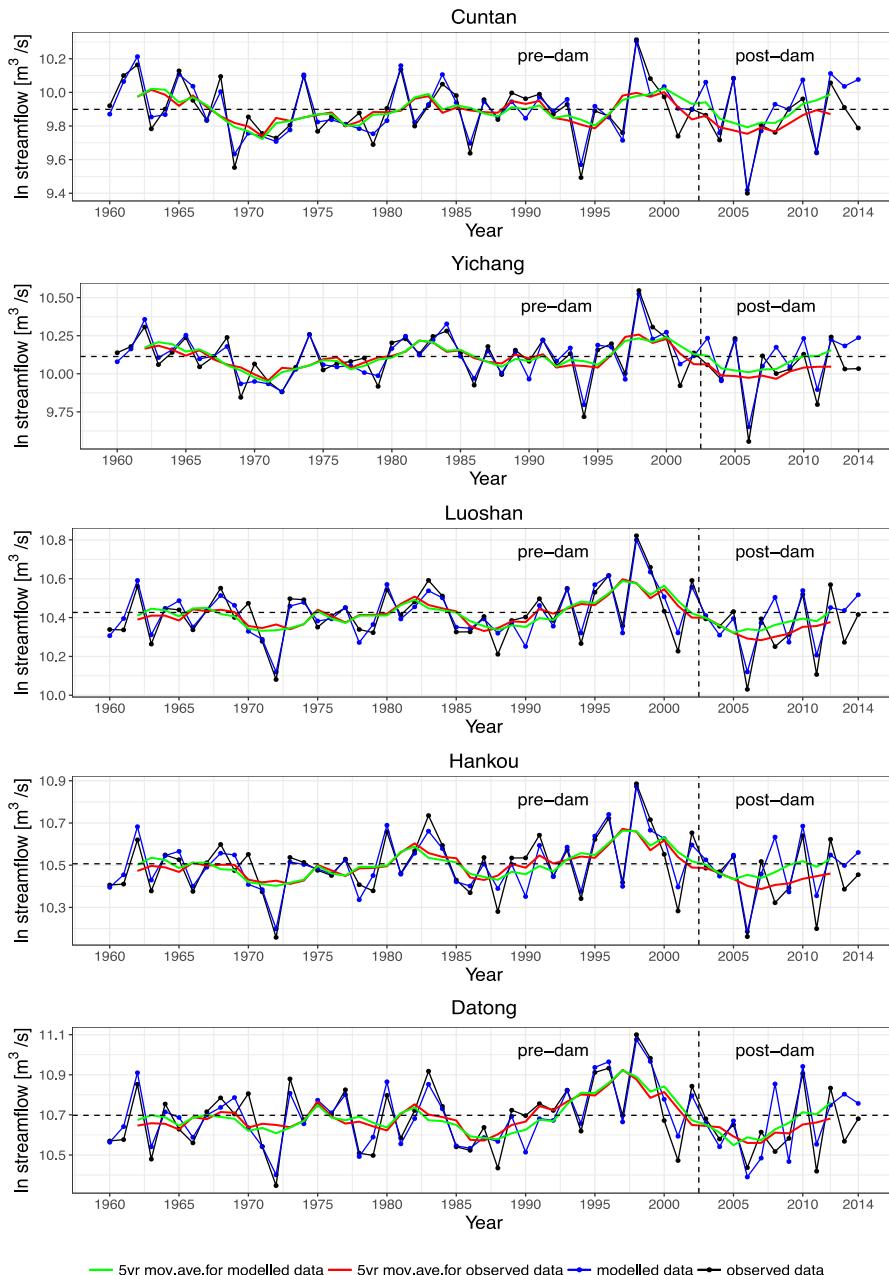


482

483 **Figure 5: Boxplots of RE and CE under cross validation. (a) and (b) on the left side indicate the two different  
484 predictor data sets from option a) and b), respectively.**

485

486



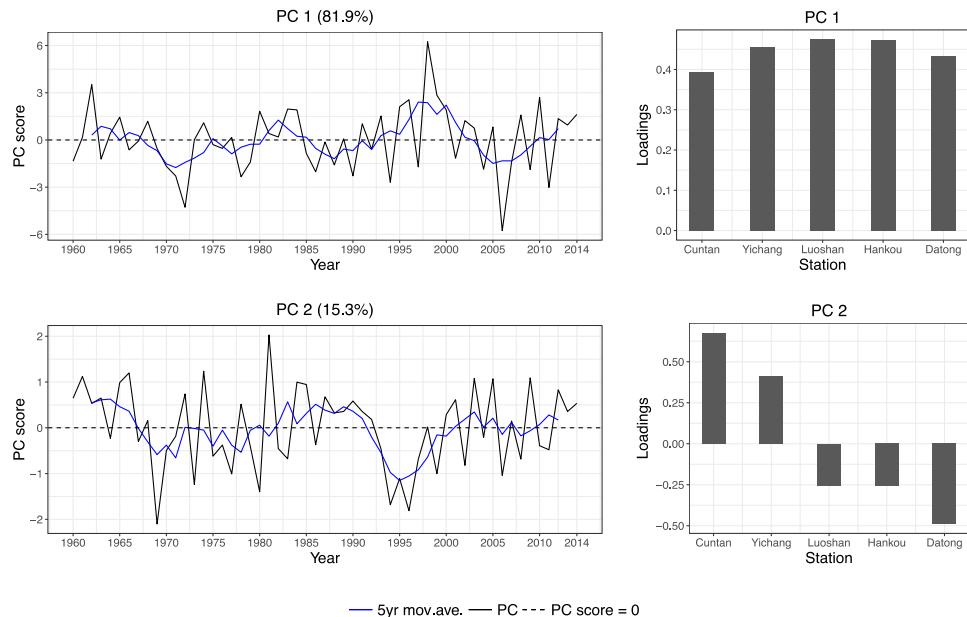
487

488 **Figure 6: Model results and observations with 5-yr moving average of the streamflow over Yangtze River Basin.**

489 **The dashed line is the average streamflow for each station during pre-dam period.**



490



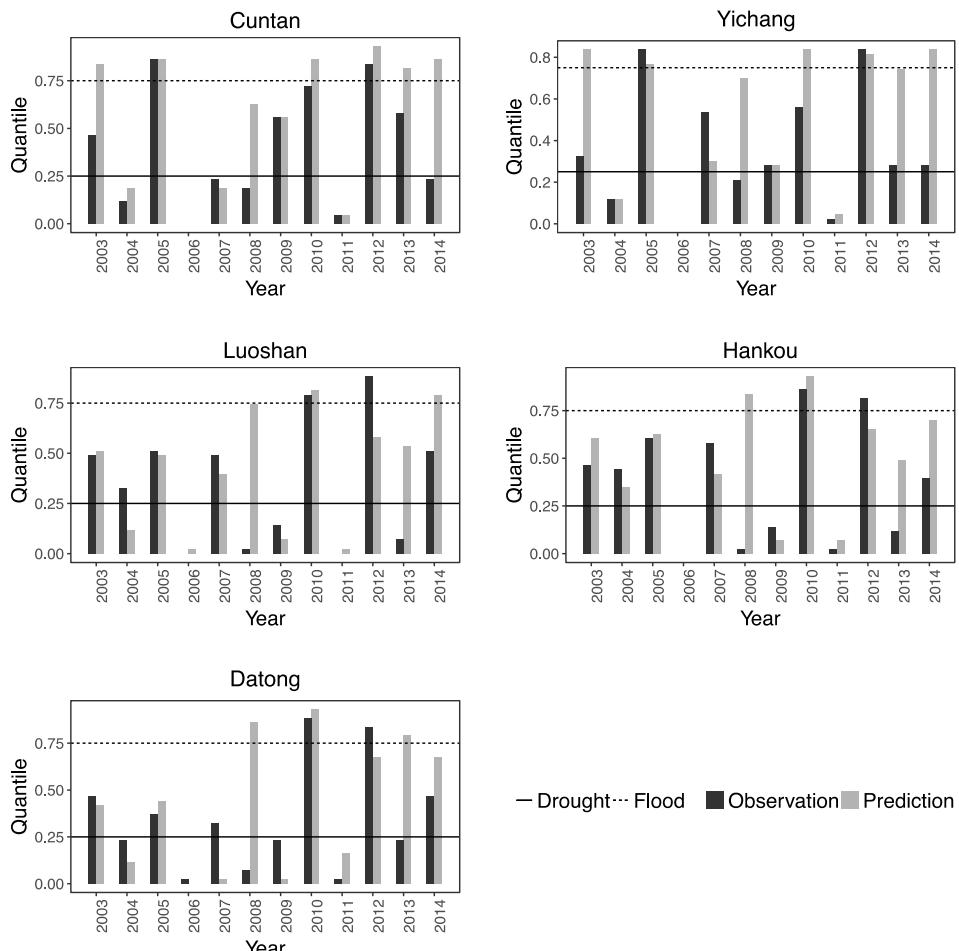
491

492 **Figure 7: The first two PCs of summer average streamflows across the 5 hydrological stations and their 5-yr**  
493 **moving average over Yangtze River Basin (left), and the corresponding loadings for each station (right). The**  
494 **percentage variance explained is shown in the parenthesis.**

495



496



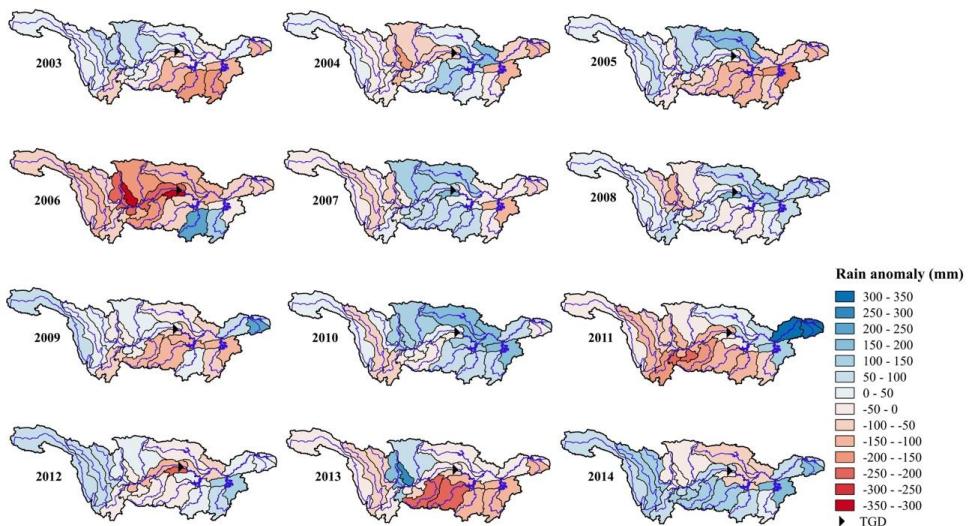
497

498 **Figure 8: Comparison of quantiles of streamflow between observations and predictions for each station. Based on**  
499 **observations during the pre-dam period, we obtained an ensemble of time series ranging from the largest to the**  
500 **smallest, then calculate the quantile of the streamflow relative to the time series. The dashed line denotes the flood**  
501 **and black line denotes the drought.**

502



503



504

505 **Figure 9: The spatial and temporal distribution of rainfall anomalies over subbasins of Yangtze River Basin.**