



Causal effects of dams and land cover changes on flood changes in mainland China

Wencong Yang^{1,2}, Hanbo Yang^{1,2}, Dawen Yang^{1,2}, and Aizhong Hou³

¹ Department of Hydraulic Engineering, Tsinghua University, Beijing 100084, China.

5 ² State Key Laboratory of Hydro-Science and Engineering, Tsinghua University, Beijing, 100084, China.

³ Hydrological Forecast Center, Ministry of Water Resources of the People's Republic of China, Beijing, China

Correspondence to: Hanbo Yang (yanghanbo@tsinghua.edu.cn)

Abstract. Quantifying the effects of human activities on floods is challenging because the knowledge and observations toward the effects are limited. Many previous methods fail to isolate different effects and reduce the uncertainty caused by small samples. We use panel regressions to derive the sensitivity of annual maximum discharges (Q) to the changing values of three human factors: urban areas, cropland areas, and reservoir indexes for large and middle dams. We also test whether the effects increase or decrease with increasing initial values of human factors. This method is applied in 2739 hydrological stations in China. Results show that a 1% increase of urban areas causes around a 2.9% increase of Q . Cropland areas have no significant effect on Q . Reservoir index has a decreasing effect: a 1 unit increase of reservoir index causes a decrease in Q from 23.1% to 5.4% for catchments with initial reservoir indexes from 0 to 5. From 1992 to 2017, increasing urban areas cause more than 10% increases in Q in 10.4% of the 2739 catchments, most of which are located in the North Plain of China. From 1960 to 2017, increasing reservoir indexes cause a more than 10% decrease of Q in 53.4% of 777 catchments with at least one dam. Among 1074 catchments with limited impacts from urban areas and reservoir indexes, 210 (19.6%) catchments have more than 10% unexplained decreasing rates in Q per decade during 1960-2017, and 62.4% of the 210 catchments are located in the middle and down streams of the Yellow River Basin and the upper streams of the Haihe River Basin. This study extends the panel regression method in hydrology and sheds light on the attribution of flood changes on a national scale.

1 Introduction

River flooding is one of the most severe disasters in the world. China experiences tremendous damages from floods in the past decades with expanding urban areas, a booming economy, and increasing populations (Du et al., 2019; Kundzewicz et al., 2019). Sharply changing flood characteristics make flood risk management more difficult. According to a national investigation of flood peak changes in China conducted by Yang et al. (2019), abrupt changes due to human activities are the predominant mode of flood changes. Understanding how floods change in a changing environment helps flood risk management in the future. Therefore, a quantitative attribution of flood changes is urgent on a national scale for policy decisions.



To detect flood changes and pinpoint the underlying reasons, scientists need to answer the following questions: 1. whether a factor affects floods? 2. If the effect presents, how strong is the effect? The drivers of flood changes can be classified into three categories: atmospheric factors, catchment factors, and river factors (Merz et al., 2012; Blöschl et al., 2015). Atmospheric factors refer to the meteorological forcing of water fluxes such as natural climate variability and anthropogenic climate change; catchment factors refer to the alternating physiographic characteristics of catchments, such as land cover changes; river factors refer to hydraulic infrastructures that change river morphology and flood routing, such as dams and channelization (Merz et al., 2012; Blöschl et al., 2015). Catchment and river factors are mainly attributed to human activities, which attract increasing attention in hydrological systems in the era of “socio-hydrology” (Di Baldassarre et al., 2019; Müller and Levy, 2019). However, quantifying human impacts on floods is challenging for the following reasons. Firstly, due to the highly unpredictable human behaviors, we have limited knowledge to reproduce the process of how human activities affect floods (Pande and Sivapalan, 2017). For example, the expansion of cropland and urban areas not only casts deterministic effects on floods through changing soil physics and surface roughness but also brings uncertain effects through irrigations and water diversions. We consider these effects “uncertain” because they are related to unknown human decisions. Secondly, the observations of human activities are limited (Pande and Sivapalan, 2017). In the example above, many regions lack long-term and large-scale data of soil physics, roughness, irrigations, and water diversions, which are highly dependent on a high-cost network of in-site measurements.

Previous studies have used three methods to quantify human impacts on floods. A first method is a model-based approach. This method regards the impacts of human activities as either the difference between actual observations and the model simulations of floods (Viglione et al., 2016; Lu et al., 2018) or flood changes with time-varying model parameters (Peña et al., 2016; Umer et al., 2019). However, this method suffers from limited model accuracy. The second method is a paired-catchment experiment. This method either compares the floods before and after human impacts in one catchment or compares floods in two groups of catchments with and without human impacts (Prosdocimi et al., 2015; Hodgkins et al., 2019). However, the comparisons above cannot rigorously isolate multiple impacts on floods since we cannot actually control everything except one targeted human factor (Runge et al., 2019). The third method is empirical variable dependence, i.e., using regressions or non-stationary probability distributions to link human factors to flood characteristics (FitzHugh et al., 2011; Prosdocimi et al., 2015; Bertola et al., 2019; De Niel and Willems, 2019). The third method is cost-efficient for large-scale studies, but it has two problems. Firstly, to derive the causal effects of human factors, all confounders — which correlates with human factors and floods at the same time — should be explicitly accounted for in the empirical relationships. However, defining numerous variables to represent confounders may be an endless task. For example, climatic confounders are ambiguous because floods are caused by different climatic factors (e.g., long rainfall, short rainfall, snowmelt, and rain on snow) in different regions (Stein et al., 2020; Yang et al., 2020a; Merz et al., 2020). Therefore, in a large-sample study, we do not have a unified regression form to control all possible variables for all catchments. Secondly, empirical methods require sufficient data for robust statistical inference, while flood samples are rare for each catchment.



Panel regression (Steinschneider et al., 2013; Wooldridge, 2016) solves the problems of the empirical method in two ways. Firstly, panel regression adds virtual variables to the regression to represent a fixed individual or regional effect (Steinschneider et al., 2013). In such a way, the regression can account for the effects of ambiguous confounders that are constant in time or region. Secondly, panel regression pools all samples into one model and trades space for time (Steinschneider et al., 2013). Therefore, the regression result is more reliable even with short flood records for each catchment. Although panel regression is a tool in economics, it has been introduced in hydrology to estimate the effects of forests on floods (Ferreira and Ghimire, 2012), urbanization on runoff coefficients (Steinschneider et al., 2013), dams on streamflow (McManamay et al., 2014), rainfall on streamflow (Bassiouni et al., 2016), deforestation on streamflow (Levy et al., 2018), urbanization on floods (Blum et al., 2020), and rain/snow fraction on floods (Davenport et al., 2020). However, these studies only focused on one factor at one time. Considering more human factors can provide a more comprehensive picture of the human impacts on floods. In addition, only Blum et al. (2020) and Davenport et al. (2020) tested the nonlinear effects of factors. Previous studies rarely examined whether the effects increase or decrease with increasing initial values of human factors.

In this study, supported by an unprecedentedly large dataset of Chinese floods from 2739 streamflow gauge stations, we quantify the national average sensitivities of annual flood peaks to changing urban areas, cropland areas, and reservoir indexes for large and middle dams using panel regression. We also design a workflow to test whether effects increase or decrease for catchments with increasing initial values of the targeted factor. The causal effects of factors distinguish the flood changes explained and unexplained by the three human factors in recent decades. This study is organized as follows. Section 2 introduces methods. Section 3 describes the data. Section 4 presents the results. Section 5 discusses the methods and the insights gained by this study. Section 6 gives conclusions.

2 Methods

2.1 Causal map of flooding

Causal maps depict the dependency relationship between variables, and they help discover confounders and focus on the causal effects of different factors when fitting a regression model (Pearl and Mackenzie, 2020). A causal effect is defined as the sensitivity of floods to a factor when all possible confounding variables are controlled. Similar to Blum et al. (2020), we draw a causal map of flooding in Fig. 1. This study estimates the causal effects of the changes in dams, urban areas, and cropland areas on floods, as the three dashed lines show in the figure. Variables lying above the dashed lines are unknown or unobserved mediators. Urban areas and cropland areas are interrelated because they may change into each other during the process of land cover change. We consider two major confounders. The first confounder is the regional time-varying confounder, which means that it changes over time but keeps spatially constant in a region. Climate is such a variable since it varies temporally and has spatial homogeneity. The trends of climate may correlate with human activities. For example, decreasing annual precipitation exacerbates water shortage and may therefore promote the reservoir constructions or the



100

implementation of the Grain for Green Project. The second confounder is the individual time-invariant confounder, which means that it keeps constant in time but varies by different catchments. This confounder is mainly represented by the characteristics of catchment landscapes, e.g., topography, soil types, geology. They may also correlate with human activities. For example, urban areas are likely built on flat and plain catchments.

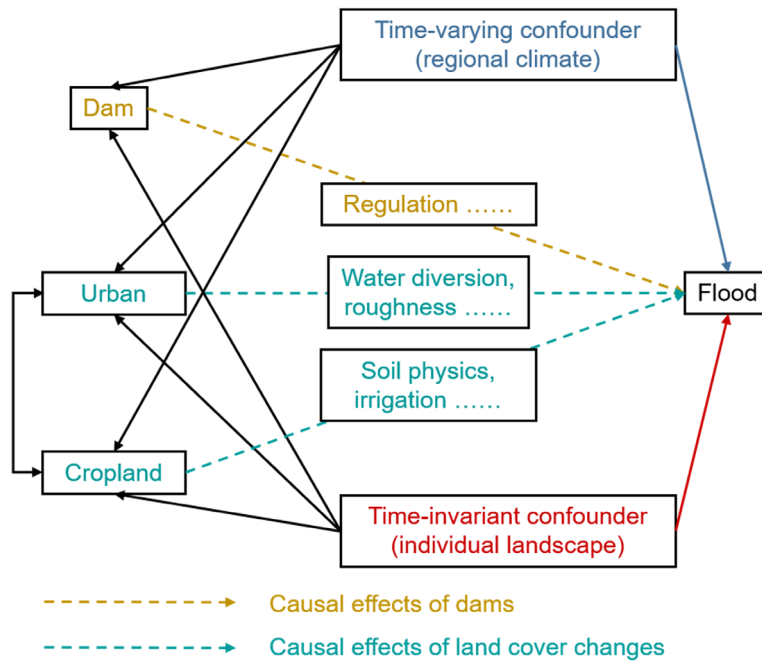


Figure 1. Causal map illustrating the relationships between human factors and floods.

2.2 Design of panel regression

105 Panel regression is a statistical technique for panel data (Steinschneider et al., 2013; Wooldridge, 2016). Panel data are observations on several subjects in different periods. Panel regression controls the constant effects of each subject or each period to mitigate regression bias due to omitted variables. Panel regressions in this study are extended from the equation in Blum et al. (2020) and are presented in Eq. (1) and (2) as follows.

$$\log(Q_{i,t}) = \alpha_i^{(1)} + g_1(Urban_{i,t}) + g_2(Crop_{i,t}) + \pi_{r,t}^{(1)} D_r D_t + \varepsilon_{i,t}^{(1)} \quad (1)$$

110 $\log(Q_{i,t}) = \alpha_i^{(2)} + g_3(RI_{i,t}) + \pi_{r,t}^{(2)} D_r D_t + \varepsilon_{i,t}^{(2)} \quad (2)$

$Q_{i,t}$ is the annual flood peak of catchment i in year t ($\text{m}^3 \cdot \text{s}^{-1}$). $Urban_{i,t}$ is the urban percentage of catchment i in year t (%). $Crop_{i,t}$ is the cropland percentage of catchment i in year t (%). $RI_{i,t} = \sum_j (A_i^{(j)} / A_i) \cdot DOR_i^{(j)}$, is the reservoir index of



catchment i in year t ; $DOR_i^{(j)}$ is the degree of regulation of reservoir j in catchment i , which is the ratio between the storage capacity and total annual flow of the reservoir; $A_i^{(j)}$ is the upstream area of reservoir j ; A_i is the area of catchment i . D_r is a region dummy which equals 1 or 0. D_t is a year dummy which equals 1 or 0. $\alpha_i^{(1)}$ and $\alpha_i^{(2)}$ are the time-invariant constant effects of catchment i in Eq. (1) and (2) respectively. $\pi_{r,t}^{(1)}$ and $\pi_{r,t}^{(2)}$ are the time-varying constant effects of region r in year t in Eq. (1) and (2) respectively. $\varepsilon_{r,t}^{(1)}$ and $\varepsilon_{r,t}^{(2)}$ are the model residuals in Eq. (1) and (2) respectively. The response functions $g_1(\cdot)$, $g_2(\cdot)$, and $g_3(\cdot)$ represent various response types of Q to different factors. We assigned *Urban*, *Crop*, and *RI* into two regressions because their data sources had different temporal lengths. Although *RI* may correlate with *Urban* and *Crop*, the effects of dams and land cover on floods can be derived independently since we have controlled their common drivers (Pearl and Mackenzie, 2020) in each equation, i.e., the regional time-varying term and the individual time-invariant term. A region consists of a group of spatially coherent catchments with a similar climate. Unlike Blum et al. (2020) who used predefined physiographic regions, we delineated regions by using the partitioning around medoids (PAM) algorithm (Reynolds et al., 2006) based on the distance matrix of all catchments defined as follow:

$$125 \quad dist(i, j) = dist_{KG}(i, j) + dist_{cen}(i, j) \quad (3)$$

$$dist_{KG}(i, j) = \frac{1}{2} \sum_{l=1}^{30} |k_i^{(l)} - k_j^{(l)}| \quad (4)$$

$$dist_{cen}(i, j) = \left(\frac{earth.dist(i, j)}{\max\{earth.dist(i, j) | \forall i, j\}} \right)^{1/2} \quad (5)$$

where $dist_{KG}(i, j)$ is the distance of Köppen–Geiger class (Beck et al., 2018) ratios between catchment i and j ; $k_i^{(l)}$ is the area percentage of Köppen–Geiger class l in catchment i ; $dist_{cen}(i, j)$ is the standardized distance between the geometric centers of catchment i and j ; $earth.dist(i, j)$ is the spherical distance on the earth between the geometric centers of catchment i and j .

The effect of a factor X on Q , i.e., the percentage change in Q given a fixed change in X , is expressed as:

$$\Delta Q(\%) = \Delta Q_{i,t} / Q_{i,t} = e^{g(X_{i,t} + \Delta X) - g(X_{i,t})} - 1 \quad (6)$$

We considered three types of response functions $g(\cdot)$. $g(X_{i,t}) = \beta X_{i,t}$ indicated a stable effect where the percentage change in Q only depended on ΔX ; $g(X_{i,t}) = \gamma X_{i,t}^2$ indicated an increasing effect where the percentage change in Q increased with increasing $X_{i,t}$; $g(X_{i,t}) = \theta X_{i,t}^{1/2}$ indicated a decreasing effect where the percentage change in Q decreased with increasing $X_{i,t}$. To determine the specific effect type, we tested the p values of the coefficients related to different $g(\cdot)$ functions in regressions, as shown in Fig. 2. Since the pooling samples were sufficient for statistical inference, we regarded p values smaller than 0.01 as significant.

140

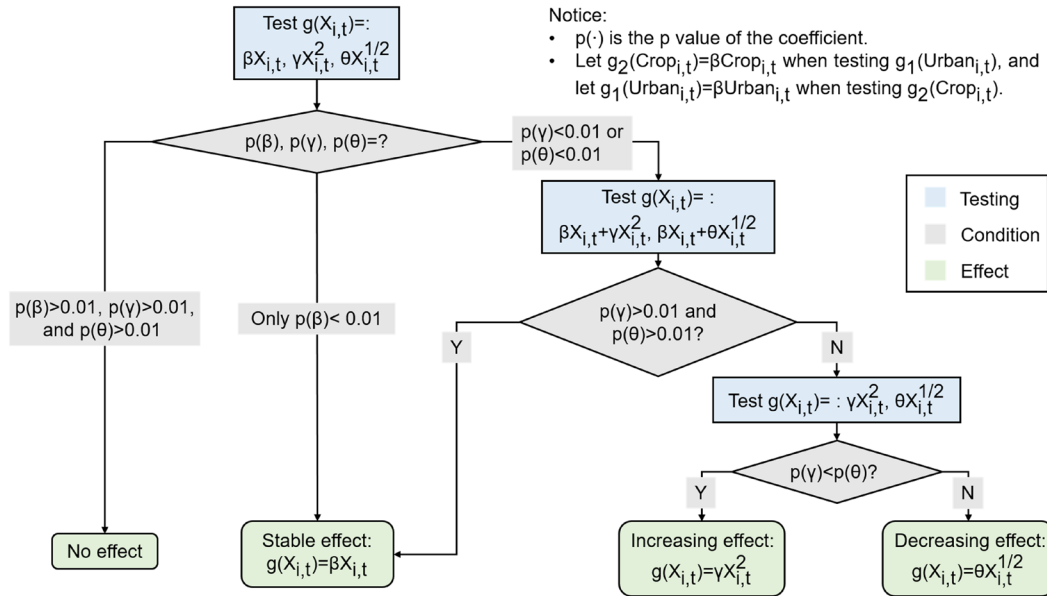


Figure 2. Workflow for testing the formulation of response functions $g(\cdot)$ for factors *Urban*, *Crop* and *RI*. All tests are conducted using Eq. (1) and (2).

145 The mathematical assumptions of the panel regressions in this study are as follows: 1. No other important time-varying sub-regional variables that significantly affect both human factors and floods; 2. No interactions between human factors and regional or individual characteristics that produce significant spatially heterogeneous effects. The regressions and statistical tests were performed in R (R Core Team, 2019) using packages lfe (Gaure, 2019), lmtest (Hothorn et al., 2020), and sandwich (Zeileis et al., 2020).

150 2.3 Flood change quantification

After obtaining the coefficients in the response function $g(\cdot)$, we can derive actual flood changes attributed to factor changes in a long period. The regressions in this study calculate a common percentage change in all flood peaks rather than heterogeneous changes of different flood events, according to Eq. (6). Therefore, the percentage change in annual maximum discharges for catchment i from year t_1 to t_2 can be written as:

$$155 \Delta Q(\%) = \exp(g(X_{i,t_2}) - g(X_{i,t_1})) - 1 \quad (7)$$

To examine the changes in flood peaks unexplained by changing urban areas, cropland areas, and dams, we selected catchments with less than 10% changes in flood peaks due to those factors respectively. Specifically, for a factor X , we selected catchments with $|\exp(g(X_{i,t_2}) - 0) - 1| < 10\%$ where t_2 was the most recent year of the data. For the annual maximum discharges of each selected catchment, we used Theil–Sen slope estimator (Theil, 1992) to derive the trend and

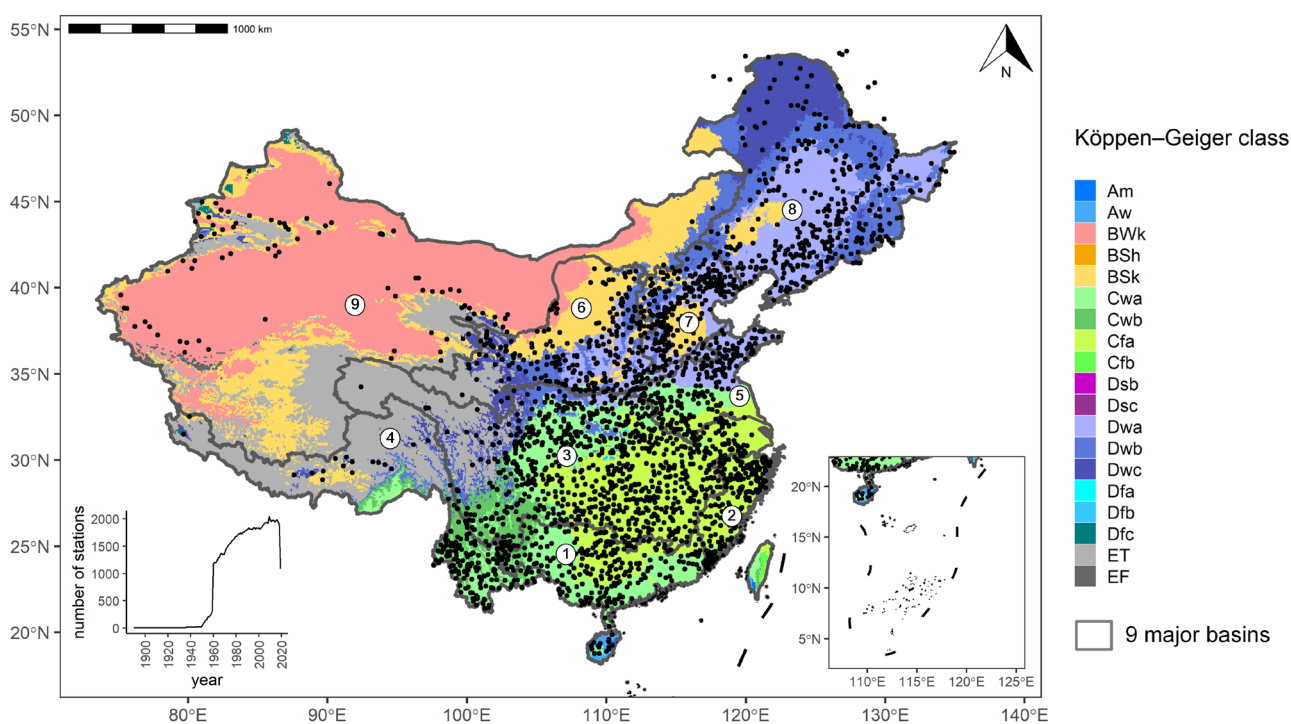
160 Mann-Kendall test (Mann, 1945) to derive the statistical significance.



3 Data

3.1 Streamflow data

Annual maximum instantaneous discharge data in 2739 streamflow gauge stations were obtained from the Ministry of Water Resources in China (<http://www.mwr.gov.cn/english/>). Figure 3 shows the outlet locations of all stations. The catchment areas are from 1 km² to 1,705,383 km² with a median of 1,660 km². Catchment boundaries were extracted using MERIT-Hydro hydrography data (Yamazaki et al., 2019). Differences between extracted catchment areas and reported areas were less than 20% for all catchments. We only used data from 1960 to 2019 because less than 1,000 stations had available data before 1960. Notice that a few stations in the northeast lie outside mainland China. They were not excluded from this study because all other data were globally available. The 1-km resolution data of Köppen-Geiger climate classes were obtained from Beck et al. (2018).



175 **Figure 3.** National 2739 streamflow stations and the number of stations with available annual maximum discharges each year. The Köppen-Geiger climate classes are obtained from Beck et al. (2018). The boundary lines delineate nine major river basins of China: 1. the Pearl River Basin, 2. the Southeast Basin, 3. the Yangtze River Basin, 4. the Southwest Basin, 5. the Huaihe River Basin, 6. the Yellow River Basin, 7. the Haihe River Basin, 8. the Songliao River Basin, and 9. the Continental Basin.



3.2 Land cover and dams data

Land cover maps were obtained from the CCI-LC product produced by the European Space Agency (ESA) Climate Change
180 Initiative (CCI). This product provides global yearly 300m-resolution land cover data in 1992-2015 in version 2.0.7 and
2016-2018 in version 2.1.1 (<http://maps.elie.ucl.ac.be/CCI/viewer/download.php>). Urban areas of catchments can be
extracted from the maps directly. Cropland areas consist of rain-fed cropland, irrigated or post-flooding cropland, and
mosaic cropland.

Dam data were available in the Global Reservoir and Dam (GRanD) v1.3 database (Lehner et al., 2011). GRanD collected
185 information about 7,320 global dams in 1948-2017 and recorded 923 dams with storage capacities larger than 10 million m³
in China. These 923 dams were categorized as large and middle dams, according to the Bulletin of the first National Water
Conservancy Survey (<http://www.chinawater.com.cn/ztgz/xwzt/2013slpczt/1/>). The total storage capacity of all 4,694 large
and middle dams is 861,961 million m³ in China, according to the bulletin. The total storage capacity of all the 923 GRanD
dams in China is 670,158 million m³, approximately 78% of that recorded in the bulletin. It suggests that the GRanD
190 database is reliable for quantifying the effects of large and middle dams on floods while it is unsuitable for considering small
dams. For simplicity, we use “dams” to represent large and middle dams in the rest of the paper. We obtained the locations,
upstream areas, storage capacities, and total annual flows of each dam from the database. The reservoir index can be
calculated with the information above.

3.3 Catchment selection for regression setup

We selected catchments with sufficient numbers of annual maximum discharges (Q) to fit Eq. (1) and (2), as shown in Table
195 1. For Eq. (1), the CCI-LC data were available in 1992-2018, and thus, we selected 1644 catchments with at least 20 years of
 Q data in this period. For Eq. (2), the discharge data were limited before 1960 and the GRanD data were available until 2017,
and thus, we selected 1744 catchments with at least 30 years of Q data in 1960-2017. The number of catchment groups k in
Section 2.2 had no optimal value. In light of the number of selected catchments in the regression models, we set k to be 10,
200 20, 30, 40, 50, and 60 to test the robustness of the models. We premised that groups were too small when $k \geq 70$, since the
average number of catchments in a group was smaller than 25 for Eq. (1) and (2).

Table 1. Setup of panel regression models for different factors.

Factor to be tested	Data period	Required Q length (years)	Number of stations	Preset number k of regions
<i>Urban & Crop</i>	1992-2018	≥ 20	1644	10, 20, 30, 40, 50, 60
<i>RI</i>	1960-2017	≥ 30	1744	10, 20, 30, 40, 50, 60



205 **4 Results**

4.1 The sensitivity of floods to human factors

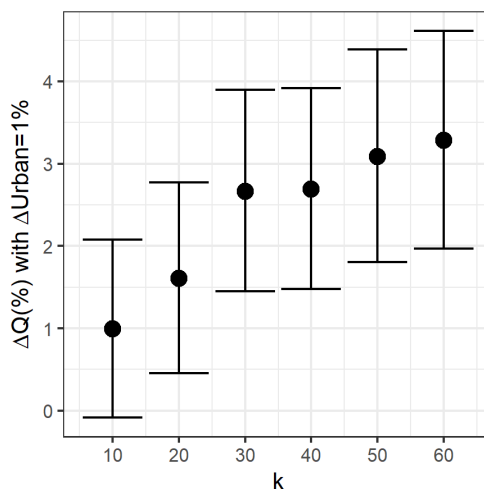
Table 2 shows the effect types of factors *Urban*, *Crop* and *RI* on Q according to Eq. (1) and (2). A positive or negative effect means Q increases or decreases with the increasing values of the factor respectively. Effects are consistent for varying values of k , with only one exceptional type of effect for *Urban* and *Crop*. *Urban* exhibits a positive and stable effect
 210 ($g(X_{i,t}) = \beta X_{i,t}$ and $\beta > 0$), which means a fixed percentage increase of *Urban* brings a fixed percentage increase of Q no matter how large the initial *Urban* is. *Crop* exhibits no effect. *RI* exhibits a negative and decreasing effect ($g(X_{i,t}) = \theta X_{i,t}^{1/2}$ and $\theta < 0$), which means a fixed increase of *RI* brings a lower percentage decrease of Q with a larger value of initial *RI*. The details about the regression coefficients (e.g., values, standard deviations, and p values) in the effect testing process (Fig. 2) can be seen in Table S1 and Table S2. The maps of catchment groups for all k values can also be seen in Fig. S1 and Fig. S2.

215

Table 2. Effect types on annual maximum discharges (Q) of factor urban percentage (*Urban*), cropland percentage (*Crop*) and reservoir index (*RI*). k is the preset number of catchment groups.

k	<i>Urban</i> effect	<i>Crop</i> effect	<i>RI</i> effect
10	No	No	Negative, decreasing
20	Positive, stable	No	Negative, decreasing
30	Positive, stable	No	Negative, decreasing
40	Positive, stable	Negative, stable	Negative, decreasing
50	Positive, stable	No	Negative, decreasing
60	Positive, stable	No	Negative, decreasing

Figure 4 shows the percentage change in Q caused by 1% increase of urban area according to Eq. (6). The values of $\Delta Q(\%)$
 220 are relatively consistent when the number of catchment groups $k \geq 30$. We regard the average value $\Delta Q = 2.9\%$ for $k \in \{30, 40, 50, 60\}$ as the sensitivity of Q to *Urban*. *Crop* has no significant effect on Q , therefore we do not calculate the corresponding sensitivity. Figure 5 shows the percentage change in Q caused by 1 unit increase of *RI* according to Eq. (6). The relationship between ΔQ and *RI* has little change with varying values of k . Therefore, the average values of coefficient θ for all six k values are considered to calculate the sensitivity of Q to *RI*. In addition, the absolute values of ΔQ decrease
 225 with increasing initial values of *RI*. For initial $RI = 0$, $\Delta Q = -23.1\%$ when *RI* increases by 1; for initial $RI = 5$, $\Delta Q = -5.4\%$ when *RI* increases by 1.



230 **Figure 4.** Percentage change in annual maximum discharge (Q) caused by 1% increase of urban area ($Urban$) based on different numbers of catchment groups (k). The error bars are 95% confidence intervals.

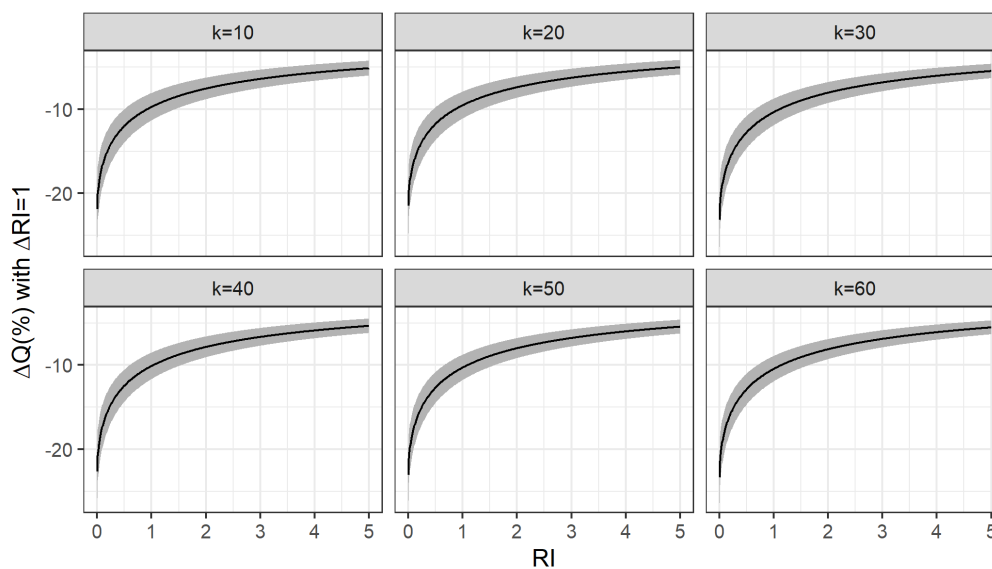


Figure 5. Percentage change in annual maximum discharge (Q) caused by 1 unit increase of reservoir index (RI) from different initial RI values based on different numbers of catchment groups (k). The shaded areas are 95% confidence intervals.

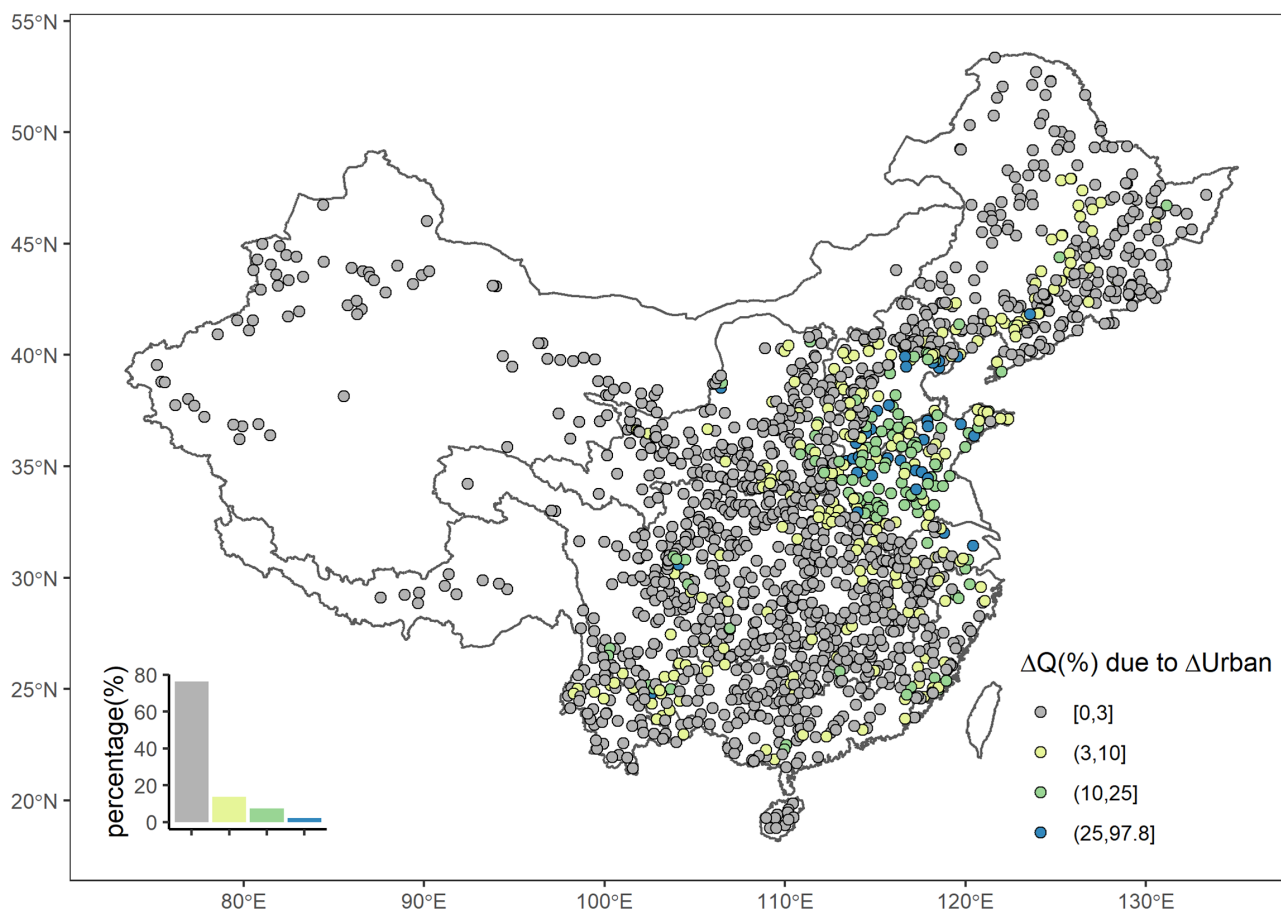
235

4.2 National flood changes and attributions

Figure 6 shows the changes in Q due to the changes in $Urban$ from 1992 to 2017 in 2739 catchments, according to Eq. (7). More than 10% of increases in Q attributed to increasing $Urban$ occur in 284 (10.4% of 2739) catchments, which are mainly



located in the North Plain of China, especially in the Huaihe River Basin and the middle and down streams of the Haihe
240 River Basin. Among these 284 catchments, 65 (2.4% of 2739) have more than 25% increases of Q attributed to increasing
Urban. The changes in Q due to the changes in *Crop* cannot be effectively quantified because *Crop* has no statistically
significant effects on Q according to the results in Section 4.1.



245 **Figure 6.** Changes in annual maximum discharges (Q) due to the change in urban area (*Urban*) from 1992 to 2017 in 2739
catchments. The boundary lines delineate nine major river basins of China.

Figure 7 shows the changes in Q due to the changes in *RI* from 1960 to 2017 in 777 out of 2739 catchments with at least one
dam, according to Eq. (7). In 415 (53.4% of 777) catchments, increasing *RI* leads to more than 10% decreases in Q . It
250 indicates that flood peaks are likely to decrease severely if dams are built in the catchment. Among these 415 catchments, 68
(8.8% of 777) have more than 25% decreases of Q attributed to increasing *RI*. Spatially, the impacts of dams on floods are



larger in northern basins (the Huaihe River Basin, the Haihe River Basin, the Yellow River Basin, and the Songhua and Liaohe River Basin) than that in southern basins (the Yangtze River Basin, the Southeast River Basin, the Southwest River Basin, and the Pearl River Basin). In the northern basins, increasing RI leads to more than 10% and 25% decreases of Q in 66.1% and 20.0% catchments, respectively. By comparison, in southern basins, increasing RI leads to more than 10% and 25% decreases of Q in only 45.4% and 2.0% catchments, respectively.

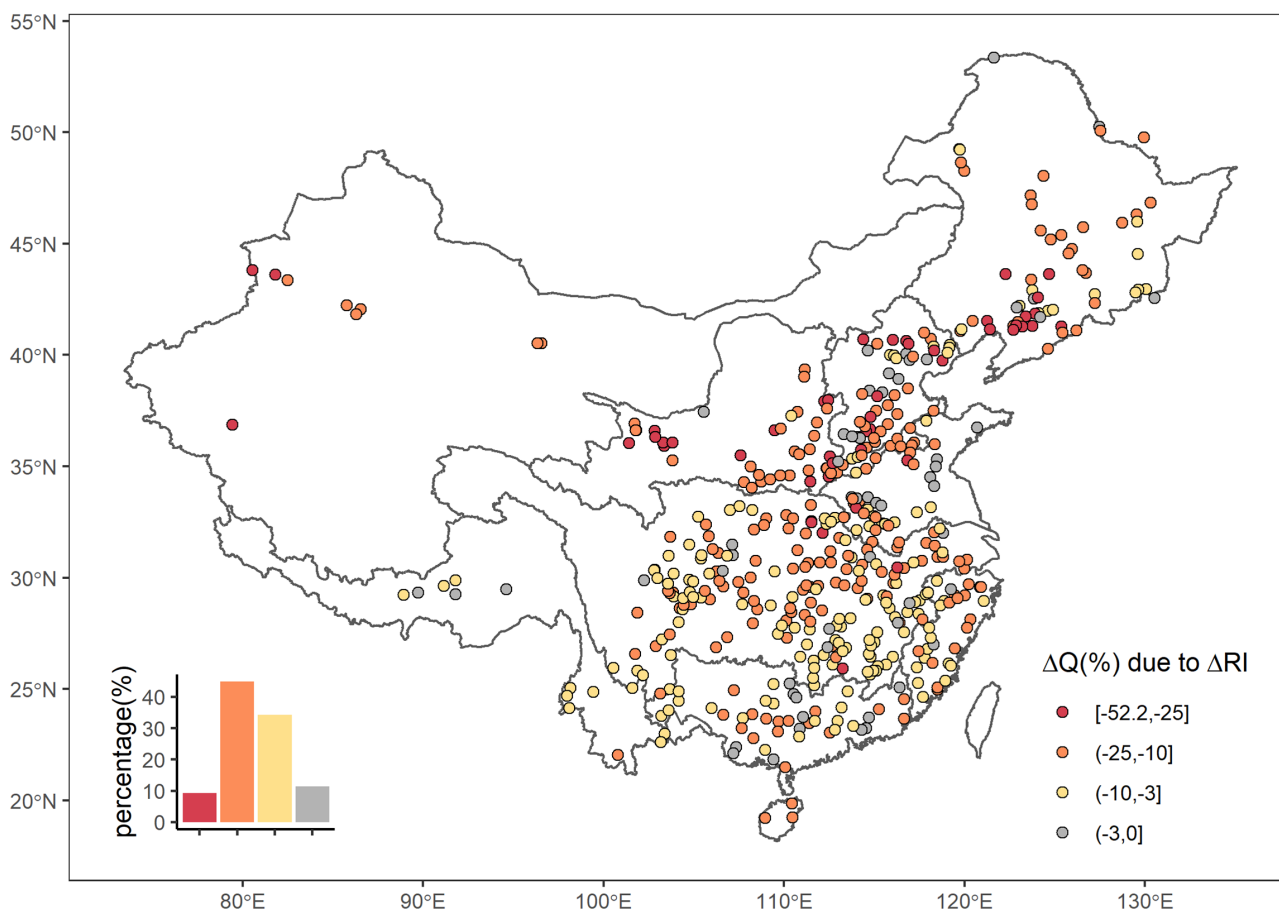


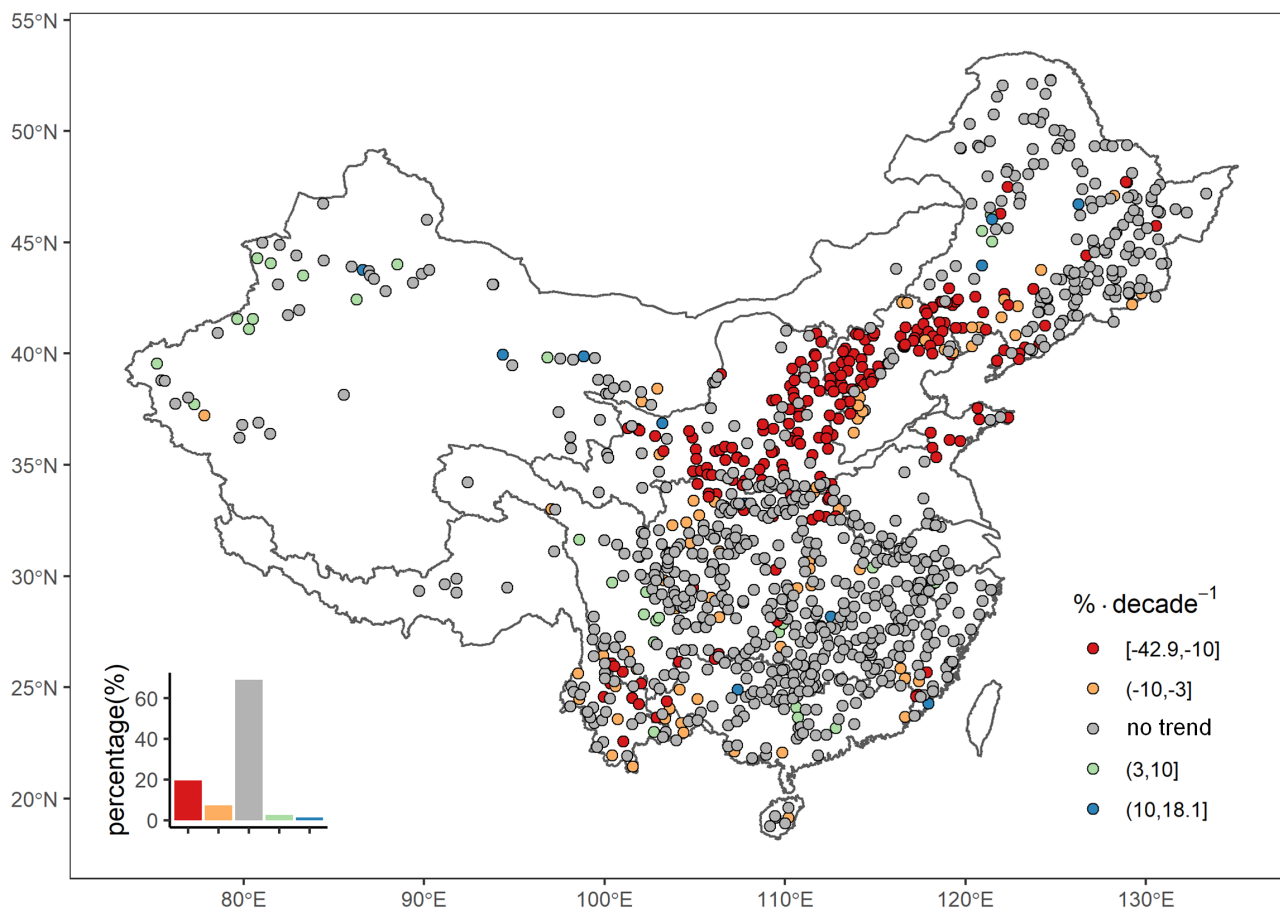
Figure 7. Changes in annual maximum discharges (Q) due to the change in reservoir index (RI) from 1960 to 2017 in 777 out of 2739 catchments with at least one dam. The boundary lines delineate nine major river basins of China.

260

Figure 8 shows the trends of Q during 1960-2017 in 1074 catchments with at least 40 years of data, $Urban < 3.4\%$ and $RI < 0.16$ in 2017. These catchments were selected based on the thresholds that ensure each factor leads to no more than 10% changes in Q , as stated in Section 2.3. In these catchments, the changes in Q are unexplained by the changes in $Urban$, $Crop$, and RI . More than 10% decreasing rates per decade are detected in 210 (19.6% of 1074) catchments, among which 131 (62.4%



265 of 210) are located in the middle and down streams of the Yellow River Basin and the upper streams of the Haihe River Basin. Such regional coherence of similar trends cannot be found in other regions.



270 **Figure 8.** Trends of annual maximum discharges (Q) during 1960–2017 in the 1074 catchments with at least 40 years of data, $Urban < 3.4\%$ and $RI < 0.16$ in 2017. These catchments are considered to have less than 10% changes in Q attributed to $Urban$ or RI . Catchments with statistically insignificant trends are labeled as “no trend”. The boundary lines delineate nine major river basins of China.



5 Discussion

275 5.1 Strengths and limitations of panel regressions

We use panel regressions to derive the causal effects of urban areas, cropland areas, and dams on annual maximum discharges across mainland China. In this study, the panel regressions exhibit the following strengths. 1. We obtain a nationally generalizable sensitivity of floods to each human factor. This sensitivity helps understand the overall added risks of specific human activity on floods on a national scale. In addition, with quantitative sensitivity, scientists are able to select catchments with limited impacts of dams and land cover changes for studying the effects of climate change, e.g., Blöschl et al. (2019). 2. Compared with previous studies using panel regressions in hydrology (Ferreira and Ghimire, 2012; Steinschneider et al., 2013; McManamay et al., 2014; Bassiouni et al., 2016; Levy et al., 2018; Blum et al., 2020; Davenport et al., 2020), we take a further step by considering multiple types of human impacts simultaneously and distinguishing their increasing or decreasing effects. Blum et al. (2020) and Davenport et al. (2020) considered non-linear forms of response functions for the targeted factors, but they did not distinguish increasing and decreasing effects. These improvements provide a more comprehensive understanding of human impacts on floods.

The limitations are as follows. 1. The assumptions in the regressions are difficult to test. As stated in Section 2, we assume no more important time-varying sub-regional confounders and no interaction terms between human factors and regional or individual characteristics. Testing these assumptions requires detailed information about catchment characteristics such as topography and geology. Moreover, adding too many variables into the regressions will decrease model interpretability. 2. The method cannot distinguish the heterogeneous effects of human factors on different floods. As stated in Section 2.3, the method derives a common percentage change in all flood peaks given changing human factors, which means no changes in coefficients of variation. However, practically, the variability of floods may change by human activities. For example, reservoirs tend to regulate extreme floods but omit small floods. 3. This study does not comprehensively assess the effects of total human impacts on floods. We omit many other human factors due to the lack of data. For example, the data about water diversion, irrigation, channelization, and afforestation on a national scale are currently not available to the public.

5.2 Consistency with knowledge and other large-sample studies

We detect a stable positive effect of urban areas on floods. In theory, expanding urban areas magnify floods in two major ways. Firstly, natural soil grounds are replaced by impervious surfaces, which lead to more rainfall water appearing on the surface rather than infiltrating into the soil (Villarini and Slater, 2017). Second, urban areas have smooth surfaces, where floods propagate faster and become more flashy (Mogollón et al., 2016). This study finds a 2.9% increase in annual maximum discharges given a 1% increase in urban areas. This finding accords with the result from a national investigation in the US (Blum et al., 2020), which reported a 3.3% increase in annual maximum discharges based on panel regressions.

The cropland areas impose no significant impacts on floods according to our results. Theoretically, expanding cropland areas affect floods in many ways. For example, during agricultural practices, soil depths may decrease due to erosion while



increase due to soil compaction (Rogger et al., 2017). Cropland may also bring artificial drainages that lower groundwater tables (Rogger et al., 2017). Some effects may be offset by others, which masks the relationship between cropland areas and floods. Similar to our result, Bertola et al. (2019) found that agricultural land-use intensification rarely caused flood changes in 95 catchments of Austria using covariate-based non-stationary flood probability distributions. To our knowledge, large-
310 sample studies are limited on the relationship between cropland and floods. Therefore, more detailed in-site investigations are required to uncover the causal chain from cropland changes to flood changes.

This study suggests that dams have a negative decreasing effect on floods. Generally, dams buffer water during floods and thus decrease flood peaks. More dams may not necessarily decrease floods at a constant rate because existing dams with sufficient storage capacities are already capable to control floods. This effect was confirmed by Wang et al. (2017), who
315 used detailed conceptual models of reservoir regulation and found that the mean annual floods had a slowing decrease with increasing reservoirs. In a large-sample study on 4859 catchments in the US (FitzHugh et al., 2011), median annual 1-day maximum flows were estimated to decrease by more than 20% when the storage ratios, i.e., the total storage capacity of upstream dams divided by average annual runoff, were larger than 1. If a dam with storage capacity equaling the annual runoff is established at the outlet of the catchment without any dam before, both the reservoir index defined in this study and
320 the storage ratio defined by FitzHugh et al. (2011) increase from 0 to 1. In this special case, the annual maximum discharges decrease by 23.1% in this study, comparable to the 20% decrease from FitzHugh et al. (2011). It is noteworthy that this study only focuses on the effects of human factors on annual maximum discharges. Generally, the effects are larger for less frequent floods. Zhao et al. (2020) investigated floods in 1403 catchments in the US and found a decrease of 100-year floods by more than 60% in 47% of catchments with a dam upstream.

325 **5.3 Insights toward a national investigation of flood changes**

This study takes the first step to explain flood changes quantitatively on a national scale in China. In this study, urbanization and dam constructions significantly change annual maximum discharges in the middle and down streams of the Yellow River Basin and the Haihe River Basin, where step changes were detected by Yang et al. (2019). As a major human residence with a high density of population, the North Plain of China experiences fast urbanization in recent years (Du et al.,
330 2018), which brings larger flood risks on lives and properties. In addition, the degree of dam regulation is larger in northern China because the annual runoff is much smaller than that in wet southern China. In this study, after removing the catchments with nonnegligible impacts of urbanization and dams, the unexplained decreasing trends occur in the middle and down streams of the Yellow River Basin and the upper streams of the Haihe River Basin, where decreasing trends were also derived by Yang et al. (2019). Yang et al. (2019) interpreted these trends as the results of soil conservation practices (Bai et al., 2016) and decreasing extreme rainfall (Yang et al., 2013; Wu et al., 2016). Besides, other reasons include decreasing soil moisture (Cheng et al., 2015; Yang et al., 2020a) and the impacts of cascade small soil-retaining dams (Yang et al., 2020b). It indicates that the impact factors of floods are complex in this region and further studies are required.



6 Conclusions

We conducted a data-based analysis on the causal effects of human impacts on floods using a panel regression on a national
340 scale, based on annual maximum discharges (Q) from 2739 stations in China, CCI-LC land cover data, and GRanD dam data.
Furthermore, we derived nationally generalizable information about the sensitivity of Q to human factors, namely the
changes in urban areas, cropland areas, and reservoir indexes for large and middle dams, and then determined the explained
and unexplained changes by the human factors based on the sensitivity. The major findings are as follows.

Urban areas have a positive and stable effect on floods, i.e., a 1% increase in urban areas causes a 2.9% increase in annual
345 maximum discharges. Cropland areas have no significant effect on Q . Reservoir index has a negative and decreasing effect
on Q , i.e., the decrease of Q caused by a 1 unit increase of reservoir indexes ranges from 23.1% to 5.4% corresponding to
initial reservoir indexes from 0 to 5.

From 1992 to 2017, more than 10% of increases in Q were caused by increasing urban areas in 10.4% of the 2739
catchments. These catchments are mainly located in the North Plain of China, especially in the Huaihe River Basin and the
350 middle and down streams of the Haihe River Basin. From 1960 to 2017, among the 777 catchments with at least one dam,
53.4% have more than 10% decreases in Q caused by increasing reservoir indexes. Spatially, the impacts of dams on floods
are larger in northern basins, including the Huaihe River Basin, the Haihe River Basin, the Yellow River Basin, and the
Songhua and Liaohe River Basin, where 66.1% catchments have more than 10% decreases of Q attributed to increasing
reservoir indexes. Among 1074 catchments with less than 10% changes in Q caused by urban areas or dams, 210 (19.6%)
355 catchments have a more than 10% decrease per decade during 1960-2017, and 62.4% of the 210 catchments are located in
the middle and down streams of the Yellow River Basin and the upper streams of the Haihe River Basin.

This study extends the panel regression method to quantify the effects of multiple human factors on floods, which helps
understand the causes of flood changes on a national scale in China. Future studies may collect more data to consider more
human factors and quantify the effects on different return periods of floods.

360 Data availability

Annual maximum discharge data are obtained from the Water Resources Information Center of the Ministry of Water
Resources in China (<http://www.mwr.gov.cn/english/>). CCI-LC data are obtained from the ESA Climate Change Initiative -
Land Cover project 2017 (<http://maps.elie.ucl.ac.be/CCI/viewer/download.php>). GRanD data are obtained from Global Dam
Watch (<http://globaldamwatch.org>).

365 Author contribution

Wencong Yang: Conceptualization, Methodology, Software, Visualization, Writing – original draft preparation.
Hanbo Yang: Conceptualization, Data curation, Funding acquisition, Supervision, Writing – review & editing.



Dawen Yang: Data curation, Funding acquisition.

Aizhong Hou: Data curation.

370 **Competing interests**

The authors declare that they have no conflict of interest.

Financial support

This research was partially supported by funding from the National Natural Science Foundation of China (Grant Nos. 51979140 and 41661144031), the Scientific Research Project from the China Three Gorges Corporation (Grant No. 202003098), and the National Program for Support of Top-notch Young Professionals.

References

- Bai, P., Liu, X. M., Liang, K., and Liu, C. M.: Investigation of changes in the annual maximum flood in the Yellow River basin, China, *Quatern Int*, 392, 168-177, 10.1016/j.quaint.2015.04.053, 2016.
- Bassiouni, M., Vogel, R. M., and Archfield, S. A.: Panel regressions to estimate low-flow response to rainfall variability in ungaged basins, *Water Resour Res*, 52, 9470-9494, 10.1002/2016wr018718, 2016.
- 380 Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., and Wood, E. F.: Present and future Köppen-Geiger climate classification maps at 1-km resolution, *Sci Data*, 5, ARTN 180214, 10.1038/sdata.2018.214, 2018.
- Bertola, M., Viglione, A., and Blöschl, G.: Informed attribution of flood changes to decadal variation of atmospheric, catchment and river drivers in Upper Austria, *J Hydrol*, 577, ARTN 12391910.1016/j.jhydrol.2019.123919, 2019.
- 385 Bloeschl, G., Hall, J., Viglione, A., Perdigão, R. A. P., Parajka, J., Merz, B., Lun, D., Arheimer, B., Aronica, G. T., Bilbashi, A., Bohac, M., Bonacci, O., Borga, M., Canjevac, I., Castellarin, A., Chirico, G. B., Claps, P., Frolova, N., Ganora, D., Gorbachova, L., Gul, A., Hannaford, J., Harrigan, S., Kireeva, M., Kiss, A., Kjeldsen, T. R., Kohnova, S., Koskela, J. J., Ledvinka, O., Macdonald, N., Mavrova-Guirguinova, M., Mediero, L., Merz, R., Molnar, P., Montanari, A., Murphy, C., Osuch, M., Ovcharuk, V., Radevski, I., Salinas, J. L., Sauquet, E., Sraj, M., Szolgay, J., Volpi, E., Wilson, D., Zaimi, K., and Zivkovic, N.: Changing climate both increases and decreases European river floods, *Nature*, 573, 108+, 10.1038/s41586-019-1495-6, 2019.
- 390 Blöschl, G., Gaal, L., Hall, J., Kiss, A., Komma, J., Nester, T., Parajka, J., Perdigão, R. A. P., Plavcova, L., Rogger, M., Salinas, J. L., and Viglione, A.: Increasing river floods: fiction or reality?, *Wires Water*, 2, 329-344, 10.1002/wat2.1079, 2015.



- 395 Blum, A. G., Ferraro, P. J., Archfield, S. A., and Ryberg, K. R.: Causal Effect of Impervious Cover on Annual Flood
Magnitude for the United States, *Geophys Res Lett*, 47, ARTN e2019GL086480, 10.1029/2019GL086480, 2020.
- Cheng, S. J., Guan, X. D., Huang, J. P., Ji, F., and Guo, R. X.: Long-term trend and variability of soil moisture over East
Asia, *J Geophys Res-Atmos*, 120, 8658-8670, 10.1002/2015jd023206, 2015.
- Davenport, F. V., Herrera-Estrada, J. E., Burke, M., and Diffenbaugh, N. S.: Flood Size Increases Nonlinearly Across the
400 Western United States in Response to Lower Snow-Precipitation Ratios, *Water Resour Res*, 56, ARTN
e2019WR025571, 10.1029/2019WR025571, 2020.
- De Niel, J., and Willems, P.: Climate or land cover variations: what is driving observed changes in river peak flows? A data-
based attribution study, *Hydrol Earth Syst Sc*, 23, 871-882, 10.5194/hess-23-871-2019, 2019.
- Di Baldassarre, G., Sivapalan, M., Rusca, M., Cudennec, C., Garcia, M., Kreibich, H., Konar, M., Mondino, E., Mard, J.,
405 Pande, S., Sanderson, M. R., Tian, F. Q., Viglione, A., Wei, J., Wei, Y. P., Yu, D. J., Srinivasan, V., and Blöschl, G.:
Sociohydrology: Scientific Challenges in Addressing the Sustainable Development Goals, *Water Resour Res*, 55, 6327-
6355, 10.1029/2018wr023901, 2019.
- Du, S. Q., He, C. Y., Huang, Q. X., and Shi, P. J.: How did the urban land in floodplains distribute and expand in China from
1992-2015?, *Environ Res Lett*, 13, ARTN 034018, 10.1088/1748-9326/aaac07, 2018.
- 410 Du, S. Q., Cheng, X. T., Huang, Q. X., Chen, R. S., Ward, P. J., and Aerts, J. C. J. H.: Brief communication: Rethinking the
1998 China floods to prepare for a nonstationary future, *Nat Hazard Earth Sys*, 19, 715-719, 10.5194/nhess-19-715-
2019, 2019.
- Ferreira, S., and Ghimire, R.: Forest cover, socioeconomics, and reported flood frequency in developing countries, *Water
Resour Res*, 48, ArtN W08529, 10.1029/2011wr011701, 2012.
- 415 FitzHugh, T. W., and Vogel, R. M.: The impact of dams on flood flows in the United States, *River Res Appl*, 27, 1192-1215,
10.1002/rra.1417, 2011.
- Gaure, S.: lfe: Linear group fixed effects, available at: <https://cran.r-project.org/web/packages/lfe/index.html>, R package
version 2.8-5.1, 2019.
- Hodgkins, G. A., Dudley, R. W., Archfield, S. A., and Renard, B.: Effects of climate, regulation, and urbanization on
420 historical flood trends in the United States, *J Hydrol*, 573, 697-709, 10.1016/j.jhydrol.2019.03.102, 2019.
- Hothorn, T., Zeileis, A., Farebrother, R., Cummins, C., Millo, G., and Mitchell, D.: lmtest: Testing Linear Regression
Models, available at: <https://cran.r-project.org/web/packages/lmtest/index.html>, R package version 0.9-38, 2020.
- Kundzewicz, Z. W., Su, B. D., Wang, Y. J., Xia, J., Huang, J. L., and Jiang, T.: Flood risk and its reduction in China, *Adv
Water Resour*, 130, 37-45, 10.1016/j.advwatres.2019.05.020, 2019.
- 425 Lehner, B., Liermann, C. R., Revenga, C., Vorosmarty, C., Fekete, B., Crouzet, P., Doll, P., Endejan, M., Frenken, K.,
Magome, J., Nilsson, C., Robertson, J. C., Rodel, R., Sindorf, N., and Wisser, D.: High-resolution mapping of the
world's reservoirs and dams for sustainable river-flow management, *Front Ecol Environ*, 9, 494-502, 10.1890/100125,
2011.



- 430 Levy, M. C., Lopes, A. V., Cohn, A., Larsen, L. G., and Thompson, S. E.: Land Use Change Increases Streamflow Across
the Arc of Deforestation in Brazil, *Geophys Res Lett*, 45, 3520-3530, 10.1002/2017gl076526, 2018.
- Lu, W. W., Lei, H. M., Yang, D. W., Tang, L. H., and Miao, Q. H.: Quantifying the impacts of small dam construction on
hydrological alterations in the Jiulong River basin of Southeast China, *J Hydrol*, 567, 382-392,
10.1016/j.jhydrol.2018.10.034, 2018.
- Mann, H. B.: Nonparametric Tests Against Trend, *Econometrica*, 13, 10.2307/1907187, 1945.
- 435 McManamay, R. A.: Quantifying and generalizing hydrologic responses to dam regulation using a statistical modeling
approach, *J Hydrol*, 519, 1278-1296, 10.1016/j.jhydrol.2014.08.053, 2014.
- Merz, B., Vorogushyn, S., Uhlemann, S., Delgado, J., and Hundecha, Y.: HESS Opinions 'More efforts and scientific rigour
are needed to attribute trends in flood time series', *Hydrol Earth Syst Sc*, 16, 1379-1387, 10.5194/hess-16-1379-2012,
2012.
- 440 Merz, R., Tarasova, L., and Basso, S.: The flood cooking book: ingredients and regional flavors of floods across Germany,
Environ Res Lett, 15, ARTN 114024, 10.1088/1748-9326/abb9dd, 2020.
- Mogollon, B., Frimpong, E. A., Hoegh, A. B., and Angermeier, P. L.: Recent Changes in Stream Flashiness and Flooding,
and Effects of Flood Management in North Carolina and Virginia, *J Am Water Resour As*, 52, 561-577, 10.1111/1752-
1688.12408, 2016.
- 445 Muller, M. F., and Levy, M. C.: Complementary Vantage Points: Integrating Hydrology and Economics for Sociohydrologic
Knowledge Generation, *Water Resour Res*, 55, 2549-2571, 10.1029/2019wr024786, 2019.
- Pande, S., and Sivapalan, M.: Progress in socio-hydrology: a meta-analysis of challenges and opportunities, *Wires Water*, 4,
ARTN e1193, 10.1002/wat2.1193, 2017.
- Pearl, J., and Mackenzie, D.: *The book of why: the new science of cause and effect*, Basic Books, New York, 2020.
- 450 Pena, L. E., Barrios, M., and Frances, F.: Flood quantiles scaling with upper soil hydraulic properties for different land uses
at catchment scale, *J Hydrol*, 541, 1258-1272, 10.1016/j.jhydrol.2016.08.031, 2016.
- Prosdocimi, I., Kjeldsen, T. R., and Miller, J. D.: Detection and attribution of urbanization effect on flood extremes using
nonstationary flood-frequency models, *Water Resour Res*, 51, 4244-4262, 10.1002/2015wr017065, 2015.
- Reynolds, A. P., Richards, G., de la Iglesia, B., and Rayward-Smith, V. J.: Clustering Rules: A Comparison of Partitioning
455 and Hierarchical Clustering Algorithms, *Journal of Mathematical Modelling and Algorithms*, 5, 475-504,
10.1007/s10852-005-9022-1, 2006.
- R Core Team: *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna,
Austria, available at: <http://www.R-project.org>, 2019.
- Rogger, M., Agnoletti, M., Alaoui, A., Bathurst, J. C., Bodner, G., Borga, M., Chaplot, V., Gallart, F., Glatzel, G., Hall, J.,
460 Holden, J., Holko, L., Horn, R., Kiss, A., Kohnova, S., Leitinger, G., Lennartz, B., Parajka, J., Perdigao, R., Peth, S.,
Plavcova, L., Quinton, J. N., Robinson, M., Salinas, J. L., Santoro, A., Szolgay, J., Tron, S., van den Akker, J. J. H.,



- Viglione, A., and Blöschl, G.: Land use change impacts on floods at the catchment scale: Challenges and opportunities for future research, *Water Resour Res*, 53, 5209-5219, 10.1002/2017wr020723, 2017.
- 465 Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., Glymour, C., Kretschmer, M., Mahecha, M. D., Munoz-Mari, J., van Nes, E. H., Peters, J., Quax, R., Reichstein, M., Scheffer, M., Scholkopf, B., Spirtes, P., Sugihara, G., Sun, J., Zhang, K., and Zscheischler, J.: Inferring causation from time series in Earth system sciences, *Nat Commun*, 10, ARTN 2553, 10.1038/s41467-019-10105-3, 2019.
- Stein, L., Pianosi, F., and Woods, R.: Event-based classification for global study of river flood generating processes, *Hydrol Process*, 34, 1514-1529, 10.1002/hyp.13678, 2020.
- 470 Steinschneider, S., Yang, Y. C. E., and Brown, C.: Panel regression techniques for identifying impacts of anthropogenic landscape change on hydrologic response, *Water Resour Res*, 49, 7874-7886, 10.1002/2013wr013818, 2013.
- Theil, H.: A Rank-Invariant Method of Linear and Polynomial Regression Analysis, in: Henri Theil's Contributions to Economics and Econometrics, *Advanced Studies in Theoretical and Applied Econometrics*, 345-381, 1992.
- Umer, Y. M., Jetten, V. G., and Ettema, J.: Sensitivity of flood dynamics to different soil information sources in urbanized areas, *J Hydrol*, 577, ARTN 123945, 10.1016/j.jhydrol.2019.123945, 2019.
- 475 Viglione, A., Merz, B., Dung, N. V., Parajka, J., Nester, T., and Blöschl, G.: Attribution of regional flood changes based on scaling fingerprints, *Water Resour Res*, 52, 5322-5340, 10.1002/2016wr019036, 2016.
- Villarini, G., and Slater, L.: Climatology of Flooding in the United States, in: *Oxford Research Encyclopedia of Natural Hazard Science*, 2017.
- 480 Wang, W., Li, H. Y., Leung, L. R., Yigzaw, W., Zhao, J. S., Lu, H., Deng, Z. Q., Demisie, Y., and Blöschl, G.: Nonlinear Filtering Effects of Reservoirs on Flood Frequency Curves at the Regional Scale, *Water Resour Res*, 53, 8277-8292, 10.1002/2017wr020871, 2017.
- Wooldridge, J. M.: *Introductory econometrics a modern approach [Hauptband] [Hauptband]*, 2016.
- Wu, Y. J., Wu, S. Y., Wen, J. H., Xu, M., and Tan, J. G.: Changing characteristics of precipitation in China during 1960-2012, *Int J Climatol*, 36, 1387-1402, 10.1002/joc.4432, 2016.
- 485 Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., and Pavelsky, T. M.: MERIT Hydro: A High-Resolution Global Hydrography Map Based on Latest Topography Dataset, *Water Resour Res*, 55, 5053-5073, 10.1029/2019wr024873, 2019.
- Yang, L., Villarini, G., Smith, J. A., Tian, F. Q., and Hu, H. P.: Changes in seasonal maximum daily precipitation in China over the period 1961-2006, *Int J Climatol*, 33, 1646-1657, 10.1002/joc.3539, 2013.
- 490 Yang, L., Wang, L. C., Li, X., and Gao, J.: On the flood peak distributions over China, *Hydrol Earth Syst Sc*, 23, 5133-5149, 10.5194/hess-23-5133-2019, 2019.
- Yang, W. C., Yang, H. B., and Yang, D. W.: Classifying floods by quantifying driver contributions in the Eastern Monsoon Region of China, *J Hydrol*, 585, ARTN 124767, 10.1016/j.jhydrol.2020.124767, 2020a.



- 495 Yang, X. N., Sun, W. Y., Mu, X. M., Gao, P., and Zhao, G. J.: Run-off affected by climate and anthropogenic changes in a
large semi-arid river basin, *Hydrol Process*, 34, 1906-1919, 10.1002/hyp.13702, 2020b.
- Zeileis, A., Lumley, T., Graham, N., and Koell, S.: sandwich: Robust Covariance Matrix Estimators, available at:
<https://cran.r-project.org/web/packages/sandwich/index.html>, R package version 3.0-0, 2020.
- Zhao, G., Bates, P., and Neal, J.: The Impact of Dams on Design Floods in the Conterminous US, *Water Resour Res*, 56,
500 ARTN e2019WR025380, 10.1029/2019WR025380, 2020.