WATER RESOURCES
AND THE REGIME OF WATER BODIES

Assessing the Influence of Upstream Basin and Climatic Characteristics on Post-Dam Downstream Streamflow Changes: Empirical Insights from Peninsular India

J. Sanyal*

Department of Geography, Presidency University, Kolkata, 700073 India
e-mail: joy.geog@presiuniv.ac.in
Received April 4, 2023; revised May 25, 2023; accepted May 30, 2023

Abstract—Size of the reservoir and dam operation are cited as the primary causes of changes in the post-dam downstream hydrograph in scientific literature. This study investigates whether upstream basin and climatological characteristics also contribute to such changes. Drawing upon the case study of 18 sites across Peninsular India, we employed Multiple Linear Regression (MLR) with two dam-related and five geophysical as well as climatological factors of upstream basins as predictors to determine key influencing causes. Pre- and post-dam flow duration curves were used to characterise changes in flow post dam construction. We considered periods with the least rainfall difference pre- and post-dam and eliminated rainfall's effect on downstream hydrograph changes. Two additional machine learning feature extraction algorithms, such as SelectK-Best with f_regression and mutual information score, were also deployed to identify the most influential driver of changes in the post-dam downstream hydrograph. The outcome of three separate methods provided a robust and unbiased result. The three methods agreed that while dams are the primary determinant of changes in the typical peak monsoon flow and overall flow variability, potential evapotranspiration and land use-soil-slope, represented by US Soil Conservation Service Curve Number (CN) of the upstream catchment, could also be key factors affecting changes in the post-dam flood peak. These findings highlight the importance of considering the geophysical and climatological characteristics of upstream catchments when assessing the impact of dam construction on downstream hydrology, and potentially aid in the placement, size, and operation of new and existing reservoirs.

Keywords: Basin characteristics, Post-dam downstream flow, Flow Duration Curve, Flow Variability, Peninsular India

DOI: 10.1134/S0097807823600729

INTRODUCTION

Dams have significant impacts on hydrology downstream due to their disruption of natural river regimes and the rerouting of river flows for human purposes [20]. The past century has seen the construction of nearly 45000 large dams worldwide, resulting in artificial flow regimes [37]. Developing nations are proposing thousands of new dams for hydropower production and water supply security, following in the footsteps of industrialized countries that built tens of thousands of dams in the twentieth century. This increased number of reservoirs has greatly expanded the amount of land available for irrigation, driving up demand for water [17].

Dams have significant effects on river hydrology, primarily through changes in the timing, magnitude, and frequency of low and high flows, resulting in a hydrologic regime that differs significantly from the natural flow regime that existed prior to impoundment [23]. A vast volume of literature is devoted to this issue. For example, changes in hydrologic regimes before and after dam construction throughout the United States is well documented [2]. Dams have the most significant impact on rivers with the highest yearly variations in flow, as they can exert immense control over downstream hydrology [12]. Spatially variable alterations in hydrological regimes due to reservoirs were demonstrated in China's middle and lower Yellow River [41]. More recently, researchers conducted a comprehensive analysis [31] of the hydrological regime change across all of China resulting from the construction of large dams. The study revealed typical mitigation of high pulses, low pulses, and rise and fall rates. In another study, the literature on the impact of dams on the hydrology and geomorphology of tropical rivers was reviewed [6], emphasizing the importance of incorporating before-after comparisons in such investigations. Various methods were utilized to evaluate the impact of reservoirs on downstream streamflow. A model was created to quantify the downstream attenuation of flood peaks [36], while

a continuous rainfall-runoff model was also employed to demonstrate the influence of reservoir storage capacity, release structures, operation rules, and inflows on regulated flood frequency [1]. Moreover, the Indicators of Hydrological Alteration (IHA) statistical measures and the Range Variability Approach (RVA) [28] were utilized in several studies [16, 21, 31] to assess dam-induced hydrological changes. Finally, a simpler approach was also taken by analyzing pre- and post-dam flow-duration curves to evaluate overall river regime changes [24].

We observed that most research has focused on individual reservoirs or a single river basin, except for regional studies conducted in the USA and more recently in China. In the past, research on the broader topic of downstream flow alteration caused by dams primarily aimed to identify changes in the hydrographs below dams and measure their extent. Less effort has been directed towards identifying the primary drivers of hydrological changes downstream of dams. Existing research has predominantly focused on determining the specific physical or operational attributes of the reservoir or dam that have the greatest impact on downstream hydrology. For example, the storage and yield ratio of reservoirs were found to influence the effect of reservoirs on downstream hydrologic conditions; reservoir systems with small storage ratios lead to much greater values of average annual instream flow than systems with large storage ratios [34]. The impact of the operation a large reservoir (China's Three Gorges Dam) on downstream water level and flow regime was also elaborately discussed [42]. The rainfall-reservoir composite index was devised to quantify the impact of reservoirs on downstream flood frequency [39]. Other studies found the capacity of the reservoir to be the most important factor determining the impact on downstream streamflow regime [27, 44]. Hydropower generation was also noted to exert primary control over changing the flow regime below dams [10]. Relevant scientific work commented that few site-specific studies constituted the majority of the literature evaluating the effects of reservoir operating rules on instream flow properties, and as a result, few general conclusions can be drawn [35]. To the knowledge of this author, no recent study has addressed this issue since this observation.

It may seem evident from the sparse literature available that the size and functioning of a reservoir are the sole determinants of alterations in the flow patterns downstream. No specific consideration was given to the role of the catchment characteristics of the upstream areas of the reservoirs in such regard. When reservoir capacity and dam operation mechanisms are considered at different sites in a region in conjunction with the physiographic and climatic characteristics of the upstream catchments, it is unclear how the flow below the location of the dam can be affected from pre- to post-dam conditions. This is due to the complex interplay between dam's functioning and various factors such as the slope, land use, moisture availability, evaporation rate, and other hydrological characteristics of the diverse upper catchments in different sites. The prevalent view is that only the capacity of the reservoir and its operation are responsible for affecting downstream hydrology. However, number of physiographic and climatological attributes of the upstream catchment have profound influence on the runoff generation process. These basin characteristics vary from one catchment to another and the recommended before-after research framework [6] can be the most effective means to capture their changing degree of influence in runoff generation.

Regional climate conditions, such as water supply (precipitation) and energy availability (PET), as well as corresponding underlying characterizations, such as vegetation, soil, and geography, regulate the regional hydrological cycle and energy balance [43]. The USDA SCS Curve Number (CN) can be used to account for the role of land use (vegetation), soil and geography (slope). PET is widely used for water balance in hydrological models [7]. Other basin related parameters, like Strahler's stream order is useful to describe the catchment area and average annual discharge [14]. The variation of these non-dam factors affects streamflow in any unregulated basin, and it is possible that these variations from one upper catchment to another provide a better explanation for the pre- to post-dam changes in downstream flow than the dam-related parameters.

The purpose of our study is to determine if size and operation of dams control all aspects of changes in downstream flow from before to after construction. We will be using empirical evidence from selected sites at a regional level to answer this scientific question. To do so, a before-after dam construction approach has been adopted, focusing on two aspects of changes in downstream hydrograph: 1) magnitude of high flow events and 2) the overall variability of streamflow discharge. We will further evaluate high flow events by categorizing them as a) rare very high flow events occurring approximately once a year (flood condition), and b) moderate high flow events occurring approximately four times a year (Typical monsoon peak flow). We have chosen Peninsular India as our study area due to its four large river basins and availability of information on numerous medium to large reservoirs online, making it an ideal region for selecting our sample sites.

There are several benefits of examining our current research question. Firstly, dams can alter downstream flow in dry seasons by diverting water for agricultural purposes via canals, curtailing the ecological flow. It is no longer feasible to exploit water resources for human needs without taking into account the ecological flow requirements. Thus, accurately pinpointing the cause(s) of variations in water availability throughout

the year is crucial for developing effective solutions. Additionally, reducing the magnitude of rare, very high flow events can disrupt the natural sediment transport regime of a river. As most sediment transport occurs during these infrequent events, this can have consequences for channel morphology and subsequent flooding patterns [3, 11]. Sediment transport from upstream naturally replenishes sediment stored in river beds, but the absence of sand flux due to dam building inhibits the sustainability of sand mining [13]. Finally, this overall research has the potential to guide decisions regarding the placement, size, and operation of new or existing reservoirs.

MATERIALS AND METHODS

Data

We considered the entire Peninsular India for this study. Peninsular India is drained by the east flowing major rivers, such as the Mahanadi, the Godavari, the Krishna and the Cauvery. The important west flowing rivers are the Narmada and the Tapi. There are numerous small rivers that originate in the Western Ghats Mountain, situated parallel to the west coast of India, and drains in the Arabian Sea traversing the narrow coastal strip.

The most crucial aspect of this study is to gather suitable data for pre- and post-dam comparison of flow characteristics. An acceptable study site for this purpose is required to meet two criteria: 1) it must have a dam/reservoir and a streamflow discharge monitoring station in a downstream location of the dam 2) at least ~5 years of flow data pertaining to the pre- and post-dam conditions with similar average mean annual rainfall before and after dam building (equal number of years before and after dam is also required).

It was difficult to find pairs of a dam and its downstream stream gauge. Although the Central Water Commission (CWC), Government of India, has compiled details of 2300 dams and hydrological data for 260 streamflow monitoring stations for Peninsular India [4], this study found only 18 study sites (pair of dam and stream gauge) that meet the above-mentioned criteria (Fig. 1). One of the primary reasons for the dearth of suitable study sites is the scarcity of pre-dam gauging records downstream of a dam. Furthermore, a large proportion of river monitoring sites in India began operations only in the mid 1970s. As a result, all dams that began operations prior to that date could not be considered due to the lack of pre-dam flow records. Only one dam qualified in the Cauvery River Basin, for example, because the majority of reservoirs in this basin were built in the late nineteenth and early twentieth centuries, and no pre-dam streamflow records exist. The requirement for very similar average annual rainfall in the pre- and post-dam periods narrowed the selection of sites even further. Con-

sequently, despite extensive efforts, this study only identified 18 pairs of dam and downstream river monitoring stations. Table 1 shows some key information about the dams/reservoirs that were chosen.

We used the daily gridded rainfall data provided by the Indian Meteorological Department, of $0.25° \times 0.25°$ cell size and supplied in netCDF format [25]. We created the drainage network for the entire Peninsular India using a coarse resolution GTOP30 DEM and carved out the watershed for each gauging station in shapefile format using ArcGIS. Later, a Matlab code was developed by incorporating the Maskregion routine [18] to extract daily rainfall from the IMD gridded rainfall product and the watershed boundary for each basin (zonal mean of grid-cells falling under the watershed boundary). The daily figures were then used to calculate the yearly figures. Obviously, the variation in precipitation in the pre- and post-dam period is a key factor that obscures the effect of dams in downstream streamflow alteration [38]. We tackled this challenge by considering the periods for which the difference between post- and pre dam rainfall is minimal (Fig. 2).

Measuring the Downstream Impact of Dams

One goal of this research is to quantify the impact of dams on high flow conditions in their respective downstream reaches. We used the flow duration curve (FDC) to quantify the downstream impact of impoundments following similar previous studies [38]. Specifically, the percentage chance that a flow magnitude would be equal or exceeded (P_p) was calculated from FDCs. The monitoring stations' daily streamflow data were arranged in descending order, and the P_p was calculated using the following formula:

$$P_p = m/(N+1) \times 100, \qquad (1)$$

where m is the rank of the discharge, and N is the number of observations used.

The figure for very low frequency high magnitude river discharge was obtained for $P_p < 0.5\%$ to represent the flood conditions. Relatively more frequent high flow magnitudes were derived from $P_p < 1\%$, which depict moderate high flow condition in peak monsoon season. Both P_p values were calculated for the pre- and post-dam states for the 18 study locations. In the next step, we determined the percentage change in the P_p values for both $P_p < 0.5\%$ and $P_p < 1\%$ from pre- to post-dam conditions. As dams are generally found to dampen flood peaks the percentage change values were expected to be positive from pre- to post-dam conditions with a few odd exceptions. P_p value of 0.5 indicates the highest two days of flow in a year and P_p of 1 approximately corresponds with the highest 3—4 days of annual flow magnitude.

Flow duration distributions derived from daily mean discharges are useful indicators of a streamflow record's temporal variability. When plotted on log
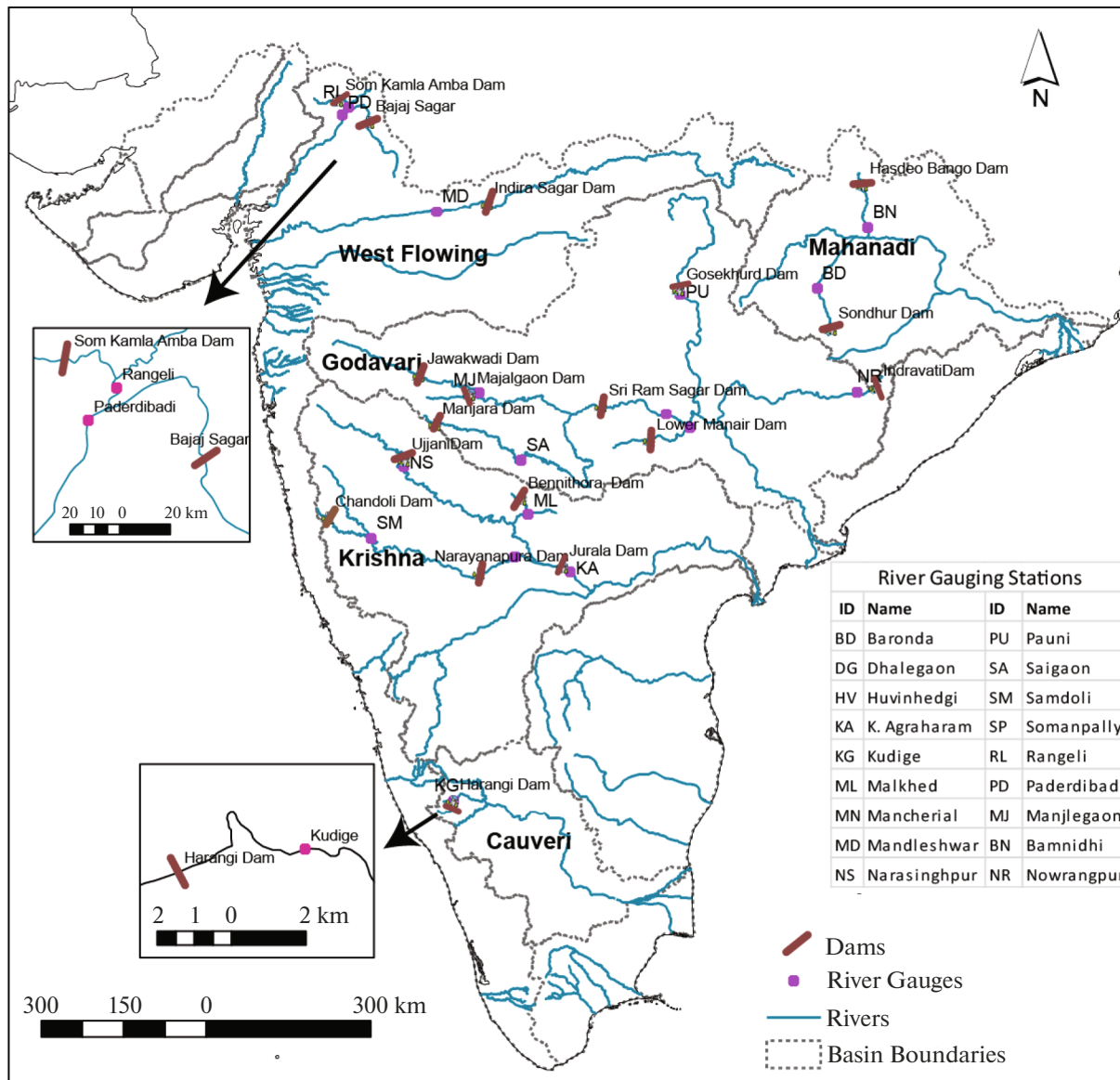
Fig. 1. Location of the 18 study sites of pairs of dam and its downstream river flow monitoring stations.

scale on both axes, a flow duration curve appears as a straight line. The slope of these lines indicates how fluctuating or stable the discharge record has been over a given time period [33]. The slope of the straight line was calculated using the flow duration curves for each study site's pre- and post-dam conditions. Figure 3 shows an example for the monitoring station Malkhed, which is located downstream of the Bennithora reservoir. Because of the longer residence time of water in the basin, reduced flow variability is expected after dam construction. Hence, we calculated the percentage decrease in the slope of the flow duration curve from pre-dam to post-dam conditions in order to obtain mostly positive numbers. To assess the impact of dams on downstream streamflow variability, the percentage decrease in slope was used as an

index. Discharge records less than 30 m³/s were ignored because the precision of such low flow observations is generally insufficient in India.

Potential Drivers of Post-Dam Changes in Downstream Streamflow

There are several potential watershed-related drivers of changes in streamflow characteristics before and after the construction of a dam, in addition to dam-linked factors. The causal relationships between these drivers and changes in streamflow can be complex. In this study, we only considered factors that can be easily quantified using national and global data sources. While there may be other important factors that could better explain the observed phenomenon, obtaining

Table 1. Important information regarding the selected 18 pair of dams and their respective downstream stream gauge

Sl. no.	Major River Basin	Gauge	Upstream Dam	Year of Completion*	Pre-dam Period Considered	Post-dam Period Considered
1	Mahanadi	Bamnidhi	Manimata Hasdeo Bango	1990	1979–1988	1995–2004
2		Baronda	Sondur	1989	1978–1987	1995–2004
3	Krishna	Malkhed	Bennithora	2001	1991–2000	2002–2011
4		Samdoli	Warna	2000	1988–1997	2006–2016
5		Huvinhedgi	Narayanpura	1982	1976–1981	2009–2014
6		Narsingpur	Ujjini	1980	1970–1979	1986–1995
7		K. Agraharam	Jurala	1996	1982–1989	1997–2004
8	Narmada	Mandlshwar	Indira Sagar	2005	1994–2004	2008–2018
9	Cauvery	Kudige	Harangi	1982	1974–1981	2006–2013
10	Mahi	Rangeli	Som Kamla Amba	2000	1978–1999	1995–2016
11		Paderdibadi	Bajaj Sagar	1985	1978–1984	1992–1998
12	Godavari	Somanpally	Lower Manair	1996	1982–1989	1997–2004
13		Dhalegaon	Jayakwadi	1976	1966–1975	1985–1994
14		Saigaon	Manjara	1980	1968–1979	2004–2015
15		Mancherial	Sriram Sagar	1977	1967–1976	2002–2012
16		Nowrangpur	Upper Indravati	2001	1987–1998	2002–2013
17		Manjlegaon	Majalgaon	1986	1977–1983	1989–1995
18		Pauni	Goshikhurd	2010	1980–1984	2011–2015

* Source of the data-National Register of Large Dams (Central Water Commission, India, 2018).

quantifiable data for these factors on a regional scale may be impossible. Additionally, due to the small sample size of 18, we kept the number of independent variables relatively small to ensure a meaningful regression analysis. Our focus was on primary factors related to dams, such as their size and operation. To determine size, we considered reservoir capacity, while for dam operation, we used the demand for irrigation water as a substitute variable since no information was available to describe dam operation in the wider context of India. We also considered five watershed and climatic attributes of the upstream catchments and one locational factor. Figure 4 illustrates the variations in these dam and non-dam characteristics across the candidate study sites, and each of these factors is discussed in detail in the following section.

(1) Dam-related factors

(a) Capacity of the reservoir behind the dam.

The reservoir's large capacity allows it to store more water, significantly increasing the residence time of water in the basin. Higher reservoir capacity can also help to moderate flood waves in high flow conditions, which is one of the main topics of this study. Larger reservoirs also result in more evaporation, which is likely to have an impact on the flow variability downstream, which we are also investigating as part of this research. In general, an increase in reservoir size is

expected to have a greater impact on the downstream high flow frequency pattern and overall flow variability. The gross capacity of each reservoir studied was obtained from the National Register of Large Dams [4].

(b) Cultivable command area (CCA) of the dams as a surrogate for dam operation.

Reservoir yield is a function of reservoir storage and its operation. Reservoir yield accounts for canal diversion, requirement for hydropower generation as well as spillway discharge and other in-stream releases. Diversion of water to meet the irrigation demand is the most important part of dam operation that determines a reservoir's storage condition and in-stream release of water. The frequency of spillway discharge is lower in reservoirs with higher irrigation water demand. Such reservoirs rarely reach full reservoir capacity and usually have enough cushion to absorb flood waves from upstream areas during storm events. The demand for irrigation from all dams in Peninsular India is not readily available. It also varies from year to year, depending on the weather and the crops grown. As a result, we used CCA as a proxy for irrigation water demand. CCA is defined as land that is suitable for cultivation and can be irrigated by a dam. For major reservoirs in Peninsular India, this information is
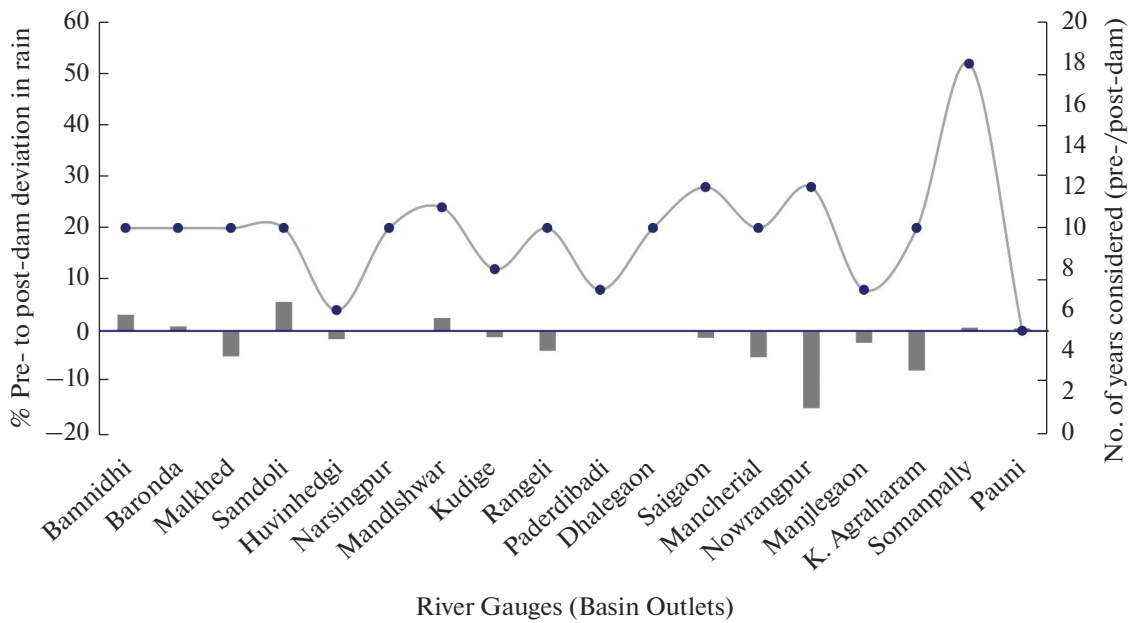
Fig. 2. Difference between pre- and post-dam average rainfall in the reservoirs' catchments (shown in columns) and the number of years considered for assessing the situation before and after dam construction (shown in line graph); total number of years (both pre- and post-dam) under consideration is $2x$ of the numbers plotted in the line graph as equal time span considered for pre- and post-dam condition.
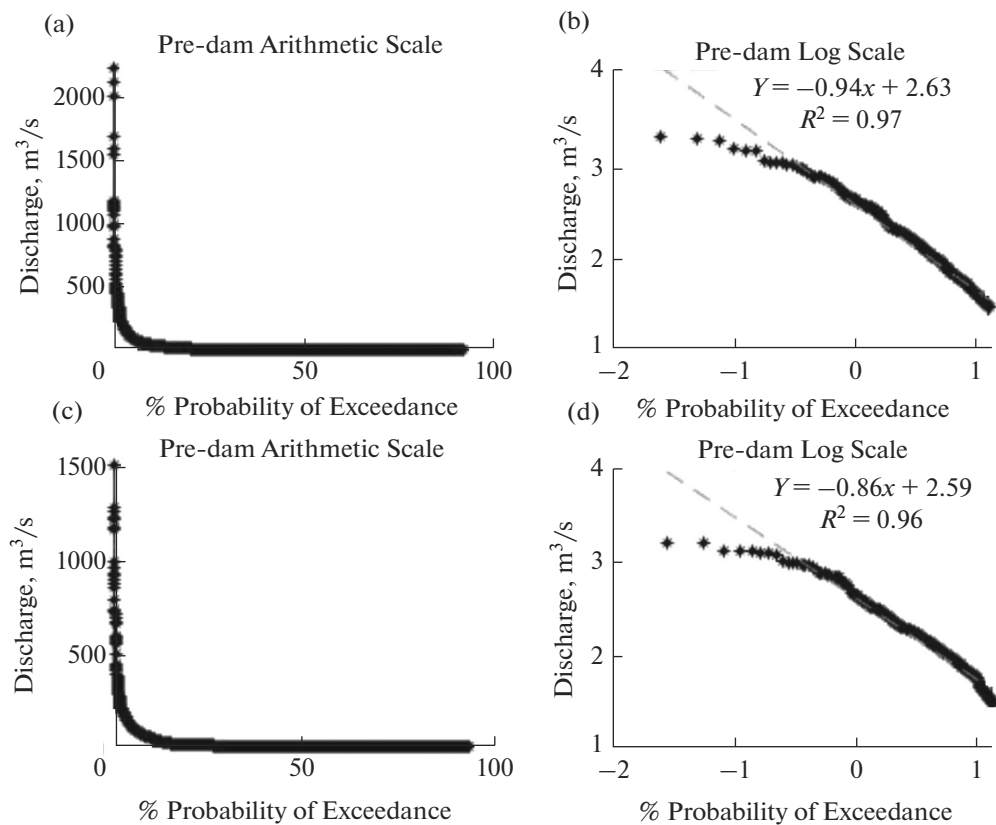
Fig. 3. Assessment of flow variability by plotting the pre- and post-dam flow duration curve in log-log scale (Base 10). The coefficient of x presents the slope the regression line, which is taken as an index of flow variability at the basin outlet (Malkhed gauging station in this example).
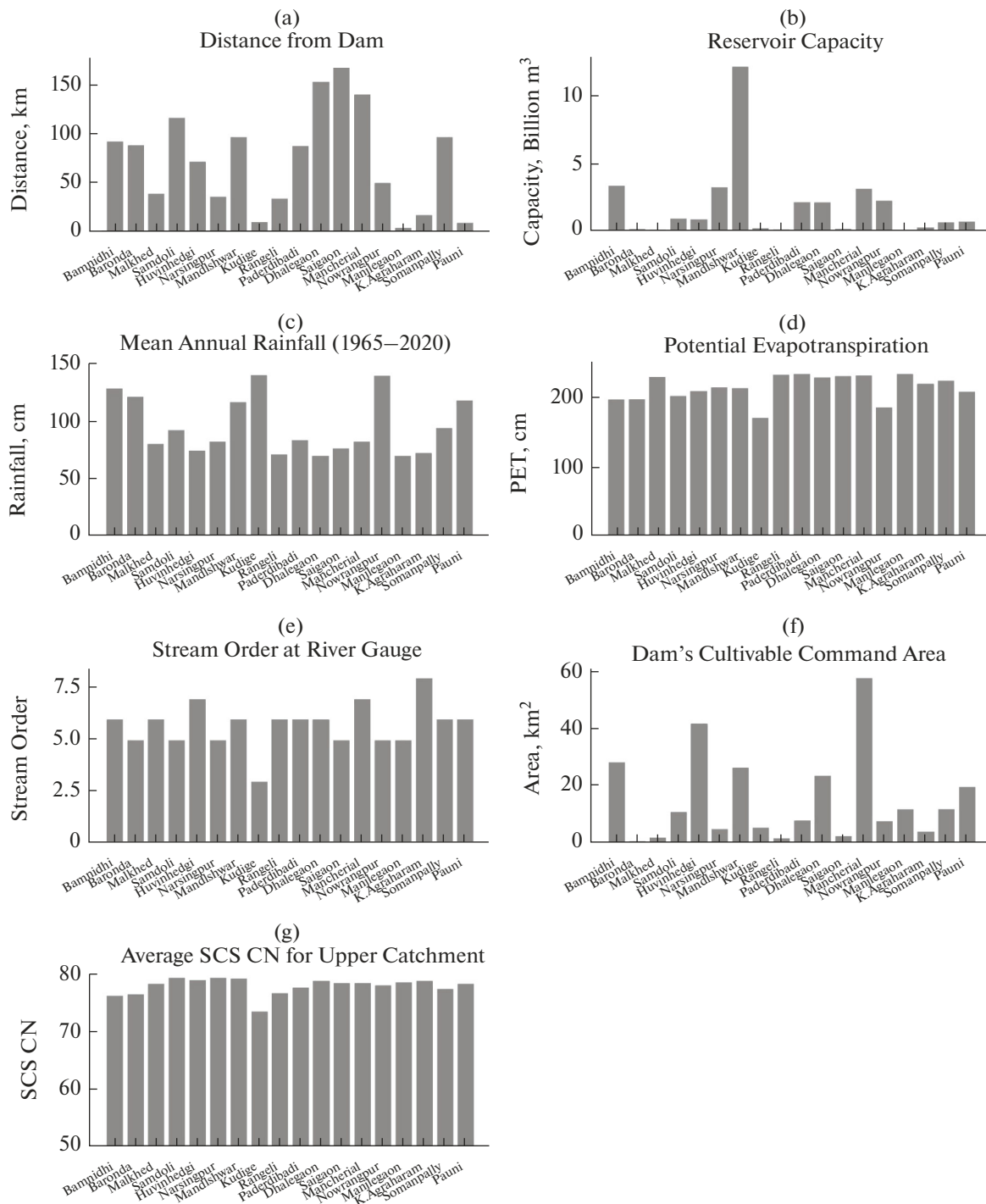
Fig. 4. The quantitative characteristics of probable drivers of varied streamflow response to dam construction at 18 basin outlets situated downstream of dams.

readily available online (see Table ST1 in Supplementary Materials).

(2) Non-dam related factors

(a) SCS CN of the basin.

The US Soil Conservation Service Curve Number (CN) is a widely used measure to account for the com-bined effect of land use/land cover, drainage charac-teristics of soil and the slope of the basin. It is an empirically derived dimensionless entity where a higher number represents less abstraction of the rain-fall with more potential for surface runoff and lower numbers characterise significant percolation and less

surface runoff. We derived the basin averaged CN value for each of the 18 candidate basins from the global curve number raster database (GCN250) of 250 m grid size [15]. The runoff coefficient of an uncontrolled basin was found to be very sensitive to the CN value. We envisaged that construction of a dam is likely to affect two basins with different runoff coefficients in a different manner in terms of the altered downstream flow regime. Hence SCS CN was considered as one of the important factors that may explain the varied response of streamflow to impoundments.

(b) Potential Evapotranspiration (PET).

Under adequate water supply, potential evapotranspiration (PET) is the highest mode of water loss. Changes in PET affect surface water yield by altering hydrological factors and water balance processes [22]. PET has crucial bearings on the agriculture induced water consumption as well as the evaporation from the surface of the reservoir. It was regarded as one of the factors with high potential for explaining the causes of different dams' varied effect on the downstream flow. We used the global PET raster data which is provided at 30 arc-second resolution [45] and derived basinwise lumped PET values using GIS.

(c) Mean annual rainfall.

This parameter represents a river basin's overall moisture condition and hydrological regime. We considered relatively short periods (<10 years for some basins) for which records of river discharges were available and the rainfall amount were similar during the pre- and post-dam condition (Fig. 2). However, these short periods are generally insufficient to fully understand the overall moisture availability in a watershed. Various aspects of basin hydrology, such as the groundwater component, may operate on longer cycles. Additionally, the cropping pattern in an area, which depends largely on the overall water availability, also significantly influences streamflow. To capture these less dynamic characteristics of basin hydrology, it was necessary to consider a long-term measure of average annual rainfall. The long-term annual average rainfall also indicates the likelihood of a few very high flow events, which are more common in moist climates than in dry climates. The impact of dam construction on downstream flow pattern is likely to vary across river basins with varying annual average rainfall. Hence, this parameter was included in the set of potential drivers of the variation in dam response from basin to basin. We calculated the annual average for each basin using IMD gridded rainfall data [25] for 46 years (1965−2020). Matlab was used to extract data from netCDF format, as discussed in Material and Methods Data section.

(d) Strahler's stream order at basin outlet (river gauge).

Strahler's stream ordering method [32] gives all streams with no tributaries an order of $k = 1$. When two (or more) streams of the same order k meet downstream, they form a stream with an order of $k + 1$. When a stream of order k meets one of a lower order, the downstream order remains unchanged. Strahler's stream order at the basin outlet has an indirect control over the manner in which surface runoff turns into stream discharge and flows to the basin outlet. Higher order stream generally indicates higher connectivity in the basin. Provided the size, shape and other physiographic factors are similar between two basins, the surface runoff is likely to reach early at the basin outlet where the stream order is higher. Hence, the effect of dam construction could affect the frequency of high magnitude flow events differently at basin outlets where the stream order is different. For this reason, we selected Strahler's stream order at basin outlet as one of our explanatory factors.

(e) Distance from the dam to the basin outlet (river gauge).

This is significant because the flow hydrograph downstream of a dam experiences attenuation and translation, with the degree of attenuation increasing with distance. It is caused primarily by friction between the channel bed and the vegetation in the riparian zone. A significant amount of flowing water percolates underground through the streambed in dry conditions with loose bed material. This aspect is considered as an important factor that decides the intensity of downstream flow change post impoundment [40].

Determining the Significant Factors Affecting Downstream Flow Response to Dam Construction Across Peninsular India

Due to the interconnection and complexity of the potential explanatory factors we relied on statistical methods to determine whether watershed and climatological characteristics of the upstream contributing areas were significant drivers of changes in different aspects of the downstream streamflow following the construction of the dam. We primarily adopted a linear multiple regression (backward) method to this end. However, we also employed univariate feature selection methods commonly used in machine learning, such as SelectKBest technique [26] for the same purpose. SelectKBest is a feature selection technique in machine learning, which is used to select a specified number of the most informative features from a dataset. It was recently used for assessing the impact of climate and soil properties on the yields of crops [30]. In this study, the univariate statistical test based on ANOVA F-test and Mutual Information score were used as part of SelectKBest technique. To minimize the uncertainty in the findings, particularly due to the limited sample size of 18 dams in this study, we utilized three distinct techniques to identify key factors that contributed to the varying downstream response among the candidate dams.

In a backward linear multiple regression model, all explanatory variables (X_1, X_2...) (the factors mentioned in Potential drivers of post-dam changes in downstream streamflow section) are entered together in the model. The variable showing a level of significance(p) of ≤0.05 are retained in the model while the variables reporting $p > 0.1$ are dropped from the model as they are not deemed statistically significant. The model runs in steps and the final model contains the set of variables (if any) that can explain the variation in the dependant variable (Y) in a statistically significant manner. The R^2 value denotes the percentage variation of the dependant variable that could be explained by the combination of explanatory variables. The ANOVA table reveals the overall p value of the final model and the standardised coefficient (also known as BETA coefficient) is useful in understanding the relative importance of explanatory factors if more than one is found to be statistically significant. Finally, the VIF value denotes the degree of collinearity or mutual correlation among the explanatory variables. Ideally VIF should be near 1 but values up to 5 are considered acceptable [9].

SelectKBest with f_regression is a feature selection method in scikit-learn, a popular machine learning library in Python. f_regression works in four steps:

(a) Given a set of features X and a target variable y, f_regression calculates the correlation between each feature and the target variable.

(b) Then, it converts these correlations into F-scores, which measure the ratio of explained variance between the feature and the target variable relative to the unexplained variance.

(c) Finally, f_regression returns the F-scores for each feature, which can be used as a score to select the k best features. The features with the highest F-scores are considered to have the strongest linear relationship with the target variable.

The mutual information score is a measure of the dependency between two variables and is used as the criterion for selecting the best features. In the context of SelectKBest with mutual information score, the algorithm calculates the mutual information between each feature and the target variable. The features with the highest mutual information scores are selected as the best features, while the others are discarded. When examining the distributions of two discrete variables, mutual information is straightforward. However, it can be modified for use with numerical input and output data [29]. The mutual info regression function of the scikit-learn machine learning library implements mutual information for feature selection with numeric input and output variables.

We considered the seven potential explanatory causes as independent variables, which comprises both dam-related factors and upstream watershed as well as climatic characteristics and computed three

aforementioned models using the following response (Y) variables:

(a) the percentage change in the P_p values for $P_p < 0.5\%$, characterising flood conditions, from pre- to post-dam conditions (Henceforth, $P_p < 0.5\%$)

(b) the percentage change in the P_p values for $P_p < 1\%$, typical of peak monsoon flow, from pre- to post-dam conditions (Henceforth, $P_p < 1\%$)

(c) the percentage decrease in the slope of the flow duration curve from the pre- to post-dam condition (Henceforth, Variability Change). It represented the impact on the overall river regime, including the low flow conditions, crucial for determining the environmental flow in the downstream reach.

RESULTS

We adopted multiple linear regression (MLR) method as the primary technique and the identified statistically significant influencing factors were shown in Table 2. SelectKBest with f_regression and with Mutual Information Score also identified the important causes of change in the same three response variables as mentioned in the previous section. We did not use more complicated techniques, such as partial linear regression (PLR), which was used in similar context [5], because partial regression model may provide a good fit to the data and accurate predictions, but the linear terms are more difficult to interpret than in the standard linear model [8].

The only statistically significant cause of changes in the moderately high flow ($P_p < 1\%$) post-dam construction was related to cultivable command area (CCA), which is directly related to diversion of water through irrigation canals, a matter within the purview of dam operation. The overall model shows a statistically significant (+ve BETA value of 0.628) positive relationship (Model p value 0.05). According to the R^2 value, the model can explain 40% of the variability in the dependent variable.

When flood conditions ($P_p < 0.5\%$) are considered, four potential drivers were found to be statistically significant with a statistically significant overall model ($p = 0.020$). This model explains 57% of variability in Y. PET was the most influential factor, followed by Mean Annual Rainfall, CCA, and stream order at the basin outlet (river monitoring station), as indicated by the BETA coefficients. The first three factors are positively correlated with the dependent variable, while the result (−ve BETA value of −0.449) indicates that a higher stream order at basin outlets increases rather than decreases the magnitude of the $P_p < 0.5\%$ flow in the post-dam scenarios. As the VIF value was less than 5, the possibility of collinearity among the independent variables was deemed insignificant. It is noted that all factors other than CCA are not related to the

Table 2. Salient results of linear multiple regression (Backward) models established to find the causes behind pre- to post-dam downstream change in high flow conditions and the variability in the overall flow regime using 18 pairs of dam and their downstream river gauges in Peninsular India

Y	Model R^2	Model p Value (ANOVA)	Significant Influencers	p	t	Standardized Coefficients (BETA)	Collinearity Statistics (VIF)
Pre- to Post-dam Percentage Reduction in $P_p < 1\%$	0.40	0.005	Cultivable Command Area, ha	0.005	3.231	0.628	N/A
Pre- to Post-dam Percentage Reduction in $P_p < 0.5\%$	0.57	0.020	Mean Annual Rainfall, mm	0.025	2.523	0.902	3.841
			PET, mm	0.011	2.957	1.075	3.973
			Stream Order	0.091	−1.825	−0.449	1.820
			Cultivable Command Area, ha	0.014	2.820	0.590	1.318
Pre- to Post-dam Percentage Reduction in Flow Variability	0.37	0.032	Mean Annual Rainfall, mm	0.077	1.899	0.391	1.003
			Cultivable Command Area, ha	0.033	2.352	0.484	1.003

upstream dams per se but to the nature of the upstream catchments.

We created a partial regression plot (Fig. 5) to visualize the individual relationship between the response variable $P_p < 0.5\%$ and each of the six potential drivers, given that the rest of the predictor variables are also in the model. This figure further attests the relationship revealed by the BETA coefficients in Table 2. Accordingly, PET (Panel a) depicted a strong linear positive correlation with $0.5\% < P_p < 0.5\%$, conditioned upon others. Another statistically significant estimator, namely Mean Annual Rainfall also revealed a negative relationship with the response variable (Fig. 5c) but the relationship appears to be less linear than the case for PET.

Although the dam-related CCA emerged as the statistically significant controlling factor for all three dependent variables (as shown in Table 2), for flood conditions (where $P_p < 0.5\%$), the linear relationship with the dam-related factor (CCA) was found to be weak, when conditioned on all other independent variables (as shown in Fig. 5b). This weak linear relationship is also reflected in the rather low beta coefficient of 0.59 (as shown in Table 2).

MLR revealed that a statistically significant relationship existed between reduction in flow variability and Mean Annual Rainfall and the CCA (Model p value of 0.032). The influencing factors are a combination of dam-related and upstream catchment related attributes. The post-dam streamflow variability decreases more when the upstream dam diverts more water to irrigation canals (higher CCA) and is associated with generally moist climate (greater annual average precipitation). Both independent variables lack evidence of multicollinearity, as evidenced by their VIP values of less than 1.

By comparing the MLR results (Table 2) with the most influential factors identified by SelectKBest using f_regression and Mutual Information Score (Table 3), it was observed that all three methods converged on the same conclusion regarding the pre- to post-dam changes in the typical peak monsoon flow and the overall variability of the downstream river regime. The role of dam-related factors was found to be paramount. However, it cannot be unequivocally stated that dam-related factors are the most crucial in determining the pre- to post-dam changes in the magnitude of rare high flow events, which are typically associated with flood conditions. Instead, through the use of the MLR and Mutual Information Score method, it has been identified that PET and SCS CN are the most influential factors that contribute to post-dam changes in flow magnitude. These factors are independent characteristics of the upstream catchments that are not associated with the capacity or operation of the dams.

DISCUSSION

In this experiment, we considered periods for which the difference in rainfall between the pre- and post-dam situation was minimum. By doing so, we eliminated the influence of rainfall in affecting the
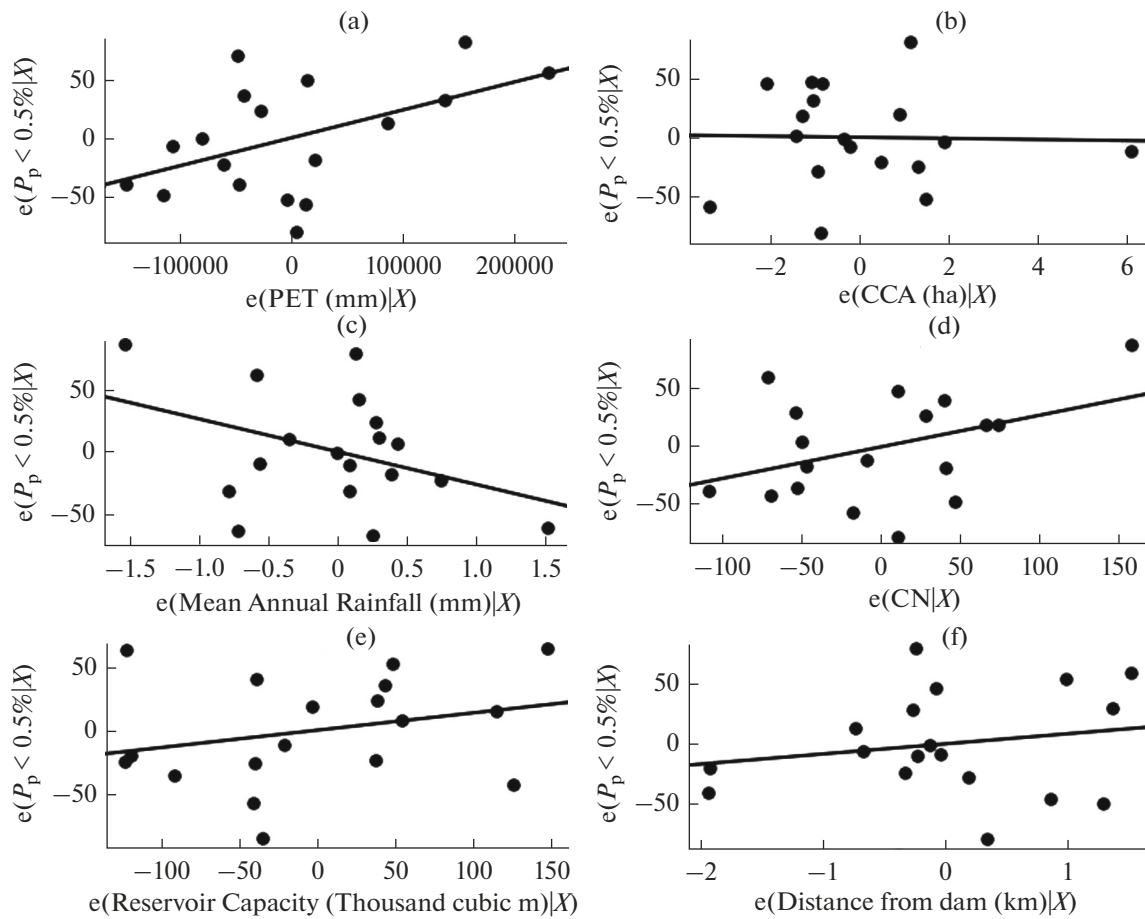
Fig. 5. Partial regression plot showing the linear relationship between change in very high magnitude streamflow ($P_p < 0.5\%$) before and after dam construction and each of the six predictor variables while considering the influence of other six predictors in the multiple regression.

pre- to post-dam changes in the downstream hydrograph. This allowed us to make a comparative assessment of the impact of dam-related and upstream catchment-related factors on altering post-dam downstream streamflow. The consideration of mean annual rainfall as an explanatory variable was also not likely to interfere with this process as it was calculated based on 46 years of annual rainfall data (1965–2020). This lengthy period covers both the pre- and post-dam durations for all 18 sites by a significant margin. As a result, this estimator is unlikely to introduce bias in the pre- or post-dam streamflow. Our findings suggest that as far as the changes in the post-dam magnitude of very high flow events are concerned, physical characteristics and climatic conditions of the upstream basin, such as PET and SCS CN, can play decisive roles.

When PET is high, it indicates that conditions are favourable for evapotranspiration and that soil water is likely to evaporate and transpire rapidly. This reduces the amount of available water for runoff, which can result in decreased flows. In contrast, when PET is low, less water is lost to evapotranspiration and more water is available for runoff, which can result in increased flows. In the pre-dam condition, PET can play a crucial role in determining the severity of very high flow events, such as floods. If PET levels are high, they can reduce the amount of water available for runoff, making flooding less likely. When a dam is constructed in such basins, the post-dam reduction in the magnitude of high flow events is pronounced because a) higher PET will evaporate more water, leaving more cushion or empty capacity for the reservoir to absorb flood waves, and b) due to the higher PET, the peak inflow itself will be of lower magnitude, resulting in less frequent spillway discharge. Given that the capacity and operation schedule of the reservoirs at two sites are comparable, the reservoir at the location with the higher PET is likely to experience a more pronounced decrease in the frequency and magnitude of rare extremely high flow events.

In general, higher curve numbers (CN) lead to greater peak flows because they represent a greater proportion of the rainfall becoming direct runoff,

Table 3. Comparison of the influential predictor variables identified by Multiple Linear Regression, SelectKBest with f_regression and with Mutual Information Score

Response Variables	Methods	Most Influencing Factor	Dam-related
Pre- to post-dam change in P_p < 1% (Typical Monsoon High Flow)	Multiple Linear Regression	Cultivable Command Area	✓
	SelectKBest with f_regression	Cultivable Command Area	
	SelectKBest with Mutual Information Score	Cultivable Command Area	
Pre- to post-dam change in P_p < 0.5% (Flood Condition)	Multiple Linear Regression	PET	×
	SelectKBest with f_regression	Cultivable Command Area	✓
	SelectKBest with Mutual Information Score	SCS Curve Number (CN)	×
Pre- to post-dam change in downstream flow variability (River Regime)	Multiple Linear Regression	Cultivable Command Area	✓
	SelectKBest with f_regression	Reservoir Capacity	✓
	SelectKBest with Mutual Information Score	None*	–

* Mutual Information Score did not identify any influencing factor.

assuming other factors such as the size and shape of the watershed remains constant. In an uncontrolled basin, as the CN value increases, the amount of runoff increases, resulting in a flashy peak flow with smaller lag time. In the post-dam situation, the high magnitude peak is attenuated by the reservoir to a significant extent. Thus, the difference in the peak flow magnitude between the pre- and post-dam scenario is accentuated. Conversely, for the upstream catchments with lower CN values the natural flood peak is itself low and the reduction in the post-dam condition is limited.

The most influential drivers identified by SelectK-Best with f_regression and with Mutual Information Score differed with MLR on number of occasions. Notably, f_regression assumes that the relationships between the features (independent) and the target (dependant or response) variable are linear, making it similar to MLR in some ways. However, mutual information, which is calculated between two variables, measures the degree to which one variable's uncertainty is reduced when the value of the other variable is known. It differs from MLR and related theories and is suitable for nonlinear correlation detection [46]. SCS CN, which takes into account land use, soil, and slope, may have a nonlinear relationship with P_p < 0.5% that the other two methods missed.

The capacity of each reservoir appears intuitively to be a potentially powerful dam-related cause of post-dam changes in the river regime. While MLR did not find this factor statistically significant, f_regression detects it as the most influential covariate for explaining the Variability Change (Table 3). The reason for this difference in the results could be due to the fact that MLR builds the model sequentially, whereas f_regression constructs the model simultaneously. As per the official online documentation of SelectKBest with f_regression, there is no sequential application, and most importantly, unlike MLR, p-values are never used.

The MLR results (Table 2) indicate that Variability Change is expected to be greater in moist river basins, where high flow events are more frequent. Flow variability decreased in post-dam periods as a result of the increased canalization of irrigation water. During the dry season, irrigation consumes even the baseflow accumulating in the reservoir from its upstream catchment. As a result, after the construction of a dam, the controlled rivers' flow variability is reduced.

We acknowledge that the sample size of this study (18) is relatively small due to the difficulty involved in finding a pair of dam and downstream monitoring site that meets all of the experiment's requirements. Hence, the inferences drawn from this small sample size are subject to some degree of uncertainty. As it was beyond our control to increase the sample size, we tried to compensate for it by validating the MLR results with two additional modern tech-

niques of data science. It played a crucial role in reducing the uncertainty in determining the important environmental controls on the variation of dams' effect in Peninsular India.

The values for the coefficient of determination (R^2) listed in Table 2 are not particularly high. We attempted to explain dimensionless response variables using a variety of 1D ('mm' for PET & Rainfall, 'km' for distance from dam), 2D ('ha' for CCA), and 3D ('m^3' for reservoir capacity) explanatory factors. These quantities are measured for purposes wholly unrelated to the purpose of this study, but we were able to establish a statistically significant relationship between them and our response variables. Under such conditions high R^2 value for the models is not expected in pragmatic terms.

This study has produced valuable datasets that consist of pairs of reservoirs and their corresponding downstream stream gauges, where streamflow data is available for both pre- and post-dam periods with similar annual average rainfall. These datasets can be utilized in future research to investigate the correlation between climate, basin characteristics, and dams in South Asia.

CONCLUSIONS

The aim of our study was to challenge the widely accepted notion that the size and operation of reservoirs are the primary drivers of changes in downstream hydrographs following dam construction. We proposed a research hypothesis that the watershed and climatic characteristics of the upstream catchment also contribute significantly to this phenomenon. To test this hypothesis, we selected 18 sample sites from the major river basins of Peninsular India where streamflow records exist downstream of reservoirs for at least five years before and after dam construction. We eliminated the effect of varying rainfall amounts during the pre- and post-dam conditions by considering periods where the difference between post and pre-dam average rainfall was minimal. We evaluated the post-dam changes in the downstream streamflow records in three dimensions: (1) rare high flow events associated with flood conditions, (2) relatively more frequent high flow events typically associated with peak monsoon, and (3) the overall variability of the flow regime.

In this study, we employed multiple linear regression (MLR) and two machine learning feature extraction algorithms to determine the key drivers of change in the downstream hydrograph at 18 sample locations. We used various dam-related, geophysical, and climatic characteristics of the basin as independent variables to explain the post-dam changes. The findings from all three techniques agreed that dam-related factors, such as capacity and operation, were the primary cause of post-dam changes in moderately high flow and overall flow variability. However, the post-dam changes in rare peak flow may possibly be attributed to characteristics of upstream basins, such as energy availability (PET), and the combined effect of land use, drainage properties of soils, and slope (SCS CN).

This study offers a decision-making tool for water resource managers that can be applied on a macro scale. It helps identify the appropriate course of action for maintaining environmental flow below dams, while also minimizing disruption to flooding patterns after dam construction. Our research has empirically demonstrated that in large geographic regions like Peninsular India, the watershed and climatic components of upstream contributing areas are critical in understanding variations in downstream streamflow following dam construction. Therefore, these factors require greater attention in managing water resources.

CONFLICT OF INTEREST

The author declares to have no conflicts of interest.

REFERENCES

1. Ayalew, T.B., Krajewski, W.F., and Mantilla, R., Exploring the effect of reservoir storage on peak discharge frequency, *J. Hydrol. Eng.*, 2013, vol. 18, no. 12, pp. 1697−1708.

2. Benito, G. and Thorndycraft, V.R., Palaeoflood hydrology and its role in applied hydrological sciences, *J. Hydrol.*, 2005, vol. 313, nos. 1−2, pp. 3−15.

3. Bhattacharyya, K., *The Lower Damodar River, India: Understanding the Human Role in Changing Fluvial Environment*, New York: Springer, 2011.

4. CWC, Central Water Commission, Government of India, 2023. https://cwc.gov.in Accessed 30 January, 2023.

5. Das, S., Dynamics of streamflow and sediment load in Peninsular Indian rivers (1965−2015), *Sci. Total Environ.*, 2021, vol. 799, pp. 149372.

6. Chong, X.Y., Vericat, D., Batalla, R.J., Teo, F.Y., Lee, K.S.P., and Gibbins, C.N., A review of the impacts of dams on the hydromorphology of tropical rivers, *Sci. Total Environ.*, 2021 vol. 794, pp. 148686.

7. Earls, J. and Dixon, B., A comparison of SWAT model-predicted potential evapotranspiration using real and modelled meteorological data, *Vadose Zone J.*, 2008, vol. 7, no. 2, pp. 570−580.

8. Fox, J., *Applied Regression Analysis and Generalized Linear Models*, Sage Publications, 2015.

9. Gareth, J., Daniela, W., Trevor, H., and Robert, T., *An Introduction to Statistical Learning: with Applications in R.*, New York: Spinger, 2013.

10. Gierszewski, P.J., Habel, M., Szmańda, J., and Luc, M., Evaluating effects of dam operation on flow regimes and riverbed adaptation to those changes, *Sci. Total Environ.,* 2020, vol. 710, pp. 136202.

11. Ghosh, S. and Guchhait, S.K., Hydrogeomorphic variability due to dam constructions and emerging

problems: a case study of Damodar River, West Bengal, India, *Environ. Dev. Sustainability*, 2014, vol. 16, pp. 769−796.

12. Graf, W.L., Downstream hydrologic and geomorphic effects of large dams on American rivers, *Geomorphol.*, 2006, vol. 79, nos. 3−4, pp. 336−360.

13. Hackney, C.R., Darby, S.E., Parsons, D.R., Leyland, J., Best, J.L., Aalto, R., et al. River bank instability from unsustainable sand mining in the lower Mekong River, *Nat. Sustainability*, 2020, vol. 3, no. 3, pp. 217−225.

14. Hughes, R.M., Kaufmann, P.R., and Weber, M.H., National and regional comparisons between Strahler order and stream size, *J. North Am. Benthol. Soc.,* 2011, vol. 30, no. 1, pp. 103−121.

15. Jaafar, H.H., Ahmad, F.A., and El Beyrouthy, N., GCN250, new global gridded curve numbers for hydrologic modeling and design, *Sci. Data*, 2019, vol. 6, no. 1, pp. 1−9.

16. Javaid, S., Bhat, W.A., Ahmed, R., Rather, A.F., Ahmad, S.T., and Ahmed, P., Assessing changing flow regime of upper and middle reaches of Narmada river using the indicators of hydrological alterations (IHA) metrics, *ISH J. Hydraul. Eng.*, 2023. https://doi.org/10.1080/09715010.2023.2173026

17. Kilic, S., Evrendilek, F., Berberoglu, S.Ü.H.A., and Demirkesen, A.C., Environmental monitoring of land-use and land-cover changes in a Mediterranean region of Turkey, *Environ. Monit. Assess.*, 2006, vol. 114, no. 1, pp. 157−168.

18. Kumar, A., Maskregion (https://www.mathworks.com/matlabcentral/fileexchange/71500-maskregion), MATLAB Central File Exchange. Accessed August 12, 2022.

19. Laarne, P., Zaidan, M.A., and Nieminen, T., ENNEMI: Non-linear correlation detection with mutual information, *SoftwareX*, 2021, vol. 14, pp. 100686.

20. López-Moreno, J.I., Beguería, S., and García-Ruiz, J.M., The management of a large Mediterranean reservoir: storage regimens of the Yesa Reservoir, Upper Aragon River Basin, central Spanish Pyrenees, *Environ. Manage.*, 2004, vol. 34, no. 4, pp. 508−515.

21. Lu, X.X. and Chua, S.D.X., River Discharge and Water Level Changes in the Mekong River: Droughts in an Era of Mega-Dams, *Hydrol. Processes*, 2021, vol. 35, no. 7, pp. e14265.

22. Luo, K., Tao, F., Deng, X., and Moiwo, J.P., Changes in potential evapotranspiration and surface runoff in 1981−2010 and the driving factors in Upper Heihe River Basin in Northwest China, *Hydrol. Processes*, 2017, vol. 31, no. 1, pp. 90−103.

23. Magilligan, F.J. and Nislow, K.H., Changes in hydrologic regime by dams. *Geomorphol.*, 2005, vol. 71, nos. 1−2, pp. 61−78.

24. Marcinkowski, P. and Grygoruk, M., Long-term downstream effects of a dam on a lowland river flow regime: Case study of the Upper Narew, *Water*, 2017, vol. 9, no. 10, pp. 783.

25. Pai, D.S., Rajeevan, M., Sreejith, O.P., Mukhopadhyay, B., and Satbha, N.S., Development of a new high spatial resolution (0.25 × 0.25) long period (1901−2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region, *Mausam*, 2014, vol. 65, no. 1, pp. 1−18.

26. Pedregosa, F., Varoquaux, G., Gramfort, A., et al., Scikit-learn: Machine learning in Python, *J Mach Learn Res.*, 2011, vol. 12, pp. 2825−2830.

27. Poff, N.L. and Zimmerman, J.K., Ecological responses to altered flow regimes: a literature review to inform the science and management of environmental flows, *Freshwater Biol.*, 2010, vol. 55, no. 1, pp. 194−205.

28. Richter, B., Baumgartner, J., Wigington, R., and Braun, D., How much water does a river need?, *Freshwater Biol.*, 1997, vol. 37, no. 1, pp. 231−249.

29. Ross, B.C., Mutual information between discrete and continuous data sets, *PloS one*, 2014, vol. 9, no. 2, pp. e87357.

30. Sihi, D., Dari, B., Kuruvila, A.P., Jha, G., and Basu, K., Explainable Machine Learning Approach Quantified the Long-Term (1981−2015) Impact of Climate and Soil Properties on Yields of Major Agricultural Crops Across CONUS, *Front. Sustain. Food Syst.*, 2022, vol. 6, pp. 847892. https://doi.org/10.3389/fsufs.2022.847892

31. Song, X., Zhuang, Y., Wang, X., et al., Analysis of hydrologic regime changes caused by dams in China. *J. Hydrol. Eng.,* 2020, vol. 25, no. 4, pp. 05020003.

32. Strahler, A.N., Quantitative analysis of watershed geomorphology, *Eos, Transactions Am. Geophys. Union*, 1957, vol. 38, no. 6, pp. 913−920.

33. Subramanya, K., *Engineering Hydrology*, Fourth Edition, New Dehli: McGraw-Hill Education, 2013.

34. Vogel, R.M., Lane, M., Ravindiran, R.S., and Kirshen, P., Storage reservoir behavior in the United States, *J. Water Resour. Plann. Manage.*, 1999, vol. 125, no. 5, pp. 245−254.

35. Vogel, R.M., Sieber, J., Archfield, S.A., Smith, M.P., Apse, C.D., and Huber-Lee A., Relations among storage, yield, and instream flow, *Water Resour. Res.*, 2007, vol. 43, W05403. https://doi.org/10.1029/2006WR005226

36. Volpi, E., Di Lazzaro, M., Bertola, M., Viglione, A., and Fiori, A., Reservoir effects on flood peak discharge at the catchment scale, *Water Resour. Res.*, 2018, vol. 54, no. 11, pp. 9623−9636.

37. WCD., *Dams and Development: A New Framework for Decision-Making: The Report of the World Commission on Dams*, Lomdon: Earthscan, 2000.

38. Wellmeyer, J.L., Slattery, M.C., and Phillips, J.D., Quantifying downstream impacts of impoundment on flow regime and channel planform, lower Trinity River, Texas, *Geomorphol.*, 2005, vol. 69, nos. 1−4, pp. 1−13.

39. Xiong, B., Xiong, L., Xia, J., Xu, C. Y., Jiang, C., and Du, T., Assessing the impacts of reservoirs on downstream flood frequency by coupling the effect of scheduling-related multivariate rainfall with an indicator of reservoir effects, *Hydrol. Earth Syst. Sci.*, 2019, vol. 23, no. 11, pp. 4453−4470.

40. Williams, G.P. and Wolman, M.G., Downstream effects of dams on alluvial rivers, *US Government Printing Office*, 1984, vol. 1286.

41. Yang, T., Zhang, Q., Chen, Y. D., Tao, X., Xu, C.Y., and Chen, X., A spatial assessment of hydrologic alteration caused by dam construction in the middle and

lower Yellow River, China, *Hydrol. Processes*, 2008, vol. 22, no. 18, pp. 3829−3843.

42. Yang, Y., Zhang, M., Zhu, L., Liu, W., Han, J., and Yang, Y., Influence of large reservoir operation on water-levels and flows in reaches below dam: Case study of the Three Gorges Reservoir, *Sci. Rep.*, 2017, vol. 7, no. 1, pp. 15640.

43. Zhang, D., Liu, X., Zhang, Q., Liang, K., and Liu, C., Investigation of factors affecting intra-annual variability of evapotranspiration and streamflow under different climate conditions, *J. Hydrol.*, 2016, vol. 543, pp. 759−769.

44. Zhao, C., Li, Y., Li, H., Wang, G., and Li, Z., Impacts of cascade reservoirs on downstream hydrology: A case study of the Lancang River Basin in China, *J. Hydrol.*, 2019, vol. 574, pp. 337−348.

45. Zomer, R.J., Xu, J., and Trabucco, A., (2022). Version 3 of the Global Aridity Index and Potential Evapotranspiration Database, *Sci. Data*, 2022, vol. 9, no. 1, pp. 1−15.

Publisher's Note.