hydrology

MDPI

Article

A Novel Data-Driven Attribution Analysis of Long-Term Streamflow Changes in the Heavily Regulated, Data-Scarce Middle Reach of the Minjiang River

Minghao Chen [1,2,*], Cong Li [3] and Taihua Wang [3]

[1] PowerChina Huadong Engineering Corporation Limited, Hangzhou 311122, China
[2] PowerChina Chongqing Engineering Corporation Limited, Chongqing 401120, China
[3] State Key Laboratory of Hydroscience and Engineering, Department of Hydraulic Engineering, Tsinghua University, Beijing 100084, China
* Correspondence: chen_mh4@hdec.com

Abstract

Streamflow variations in the Middle Minjiang River Basin (MMR) are vital for the flood mitigation and water resources management of the Chengdu metropolitan area which is important for the development of Southwest China. However, how climate change, Chengdu metropolitan area and Zipingpu Reservoir influence streamflow in the MMR remains unclear. Hence, we coupled the Geomorphology-Based Ecohydrological Model (GBEHM), the Physic-aware Hybrid Learning (PaHL) model and the Extreme Gradient Boosting (XGBoost) model to reproduce streamflow variations at Pengshan station—the outlet cross section of MMR—from 1980 to 2019, subsequently performing attribution analysis. Annual streamflow at Pengshan station exhibits a decreasing trend from 1980 to 2019. Coupled simulations effectively reproduce daily streamflow at Pengshan station during 35 years, with values of *NSE*, R^2 and *KGE* exceeding 0.96. The dominant influence of anthropogenic disturbance on daily streamflow decrease is generally steady at Pengshan station, explaining 62.3% and 430.8% of it before and after the impoundment of Zipingpu Reservoir (in 2006), respectively. Majority of the climate change's influence is notably concentrated from June to September, suggesting a potential temporal imbalance in water resources and a threat of extreme hydrological events. Our study contributes to flood mitigation and water resources management in the MMR.

Keywords: Minjiang River Basin; distributed hydrological model; data-driven model; attribution analysis

Check for updates

Academic Editor: Mohammed Bari

Received: 29 May 2026
Revised: 22 June 2026
Accepted: 23 June 2026
Published: 25 June 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Due to climate change, extreme hydrological events including flash floods, flash droughts and drought-flood abrupt alternation events have notably increasing hazards and frequency in recent years [1–5]. Meanwhile, the greatest surge in global dam construction and the irreversible urbanization trend notably modify channel and surface runoff [6–8]. The complex coupling impact of climate change and anthropogenic disturbance on runoff poses a challenge to flood mitigation and water resources management.

The Chengdu metropolitan area is important for the development of Southwest China [9]. Situated in the middle reaches of the Minjiang River (MMR), the Chengdu metropolitan area is not only exposed to increasing flood risk from extreme rainfall, but also has notable impacts on downstream runoff through urbanization and other

human activities [10–12]. The streamflow variations in the MMR are also influenced by operation of the Zipingpu Reservoir. The Zipingpu Reservoir, one of the largest reservoirs in the Minjiang River, controls a drainage area of 22,662 km^2 above the dam site which accounts for 98% of the upper Minjiang River Basin [13–15]. This reservoir also regulates approximately 90% of the flood flows generated by rainstorms in the upper Minjiang River [14–16]. The Zipingpu Reservoir reached its full impoundment in 2006. On an annual cycle, the reservoir operates in three distinct periods: the flood period (June–September), the impoundment period (October–November), and the water supply period (December–May of the following year) [13–15]. During the flood period, the water level in front of the dam is lowered to and maintained at the flood-control limit of 850 masl. In the impoundment period, the water level in front of the dam is gradually raised to the normal pool elevation of 877 masl, In the water supply period, the Zipingpu Reservoir supplies water to the downstream basin while the water level in front of the dam is higher than 817 masl.

The streamflow variation in the Minjiang River is vital for the flood management of the Chengdu metropolitan area. Climate change and anthropogenic disturbance dominates the streamflow variation. Existing studies reported that beyond climatic drivers, human activities appear to dominate the observed runoff decline in the Minjiang River Basin. In the upper reaches of the Minjiang River Basin, water resources decreased from 1961 to 2012, while the contribution of climate change to changes in annual mean discharge was 23–41.4% [17–19]. From 1990 to 2019, in the middle and lower Minjiang River Basin, economic and landscape factors had significant negative impacts on runoff changes, while natural factors had positive impacts on streamflow changes at Chengdu cross section in the MMR [20,21]. Anthropogenic disturbance, vegetation, and climatic conditions accounted for 76.24, 13.62, and 10.14% of streamflow variations at the Gaochang station, respectively [22]. Based on studies of anthropogenic disturbance, reservoir construction highly influences the annual and monthly rainfall-runoff relationship in the Minjiang River Basin [23]. Existing studies analyzed flood events in the middle reach of the Minjiang River, pointing out that the development of infrastructure contributes to effective flood mitigation [10,12,24].

Although existing studies have comprehensively disentangled multiple influencing factors of the runoff variations in the upper and lower reaches of the Minjiang River Basin, the mechanisms by which climate change and anthropogenic disturbances influence streamflow in the MMR remain unclear [25]. This knowledge gap hinders a comprehensive understanding of streamflow variability across the Minjiang River Basin. Furthermore, an effective numerical model capable of simulating the influences Zipingpu Reservoir and Chengdu metropolitan area on streamflow variations within the MMR is to be developed. Although hydrodynamic simulation provides an effective mechanistic means of reproducing the effects of anthropogenic disturbances, the complex river network of Minjiang River near the Chengdu metropolitan area which comprises 12 major tributaries, poses a considerable challenge for hydrodynamic simulation in the MMR [10,25]. Consequently, most existing hydrodynamic simulations have been confined to limited urban areas within the MMR, primarily focusing on urban flood response patterns [9,12,26,27]. Large-scale hydrodynamic simulations in the main channel of the MMR have been mostly performed using the kinematic wave model [21,25,26,28]. While these simulations reveal how river hydrodynamic processes respond to reservoir operation and municipal engineering, they generally yield lower accuracy than those based on the dynamic wave model.

In this study, streamflow variations at the outlet cross section of the MMR was utilized to represent the influence of climate change and anthropogenic disturbance on down-

stream streamflow. The anthropogenic disturbance includes the impacts of Chengdu metropolitan area and the Zipingpu Reservoir. The Geomorphology-Based Ecohydrological Model (GBEHM) was coupled with a cascade of Physics-aware Hybrid Learning (PaHL) model and Extreme Gradient Boosting (XGBoost) model. This coupling was used as the hydrological-hydrodynamic coupling simulation to conduct scenario experiments, which help to disentangle the combined effects of multiple factors on streamflow variations. In this paper, we (1) conducted trend and mutation test of the streamflow variations within the MMR; (2) built data-driven models to reproduce hydrodynamic influence of anthropogenic disturbance; (3) conducted hydrological-hydrodynamic coupling simulations; and (4) evaluated the contribution of each influencing factor to the runoff variations in the Minjiang River. Our study is devoted to comprehensively evaluate the mechanism of streamflow variations in the MMR.

2. Study Area and Data

2.1. Study Area

The Mingjiang River Basin (27°49∼33°09′ N, 102°32′∼105°30′ E) is located in the upper reaches of the Yangtze River, as shown in Figure 1. The Minjiang River is a first-order tributary of the upper Yangtze River. The MMR, as long as 216 km, extends from Dujiangyan City to Leshan City, which serve as its upstream and downstream boundaries, respectively. The MMR is basically plain and belongs to the subtropical monsoon climate zone with abundant rainfall. The mean annual precipitation of the MMR is about 800–1200 mm. The Chengdu metropolitan area is located in the MMR, while the mainstream of the Minjiang River has a complex river network near the Chengdu City, consisting of multiple parallel tributaries including the Jinmahe River, the Jianganhe River and the Qingshuihe River [20,29].

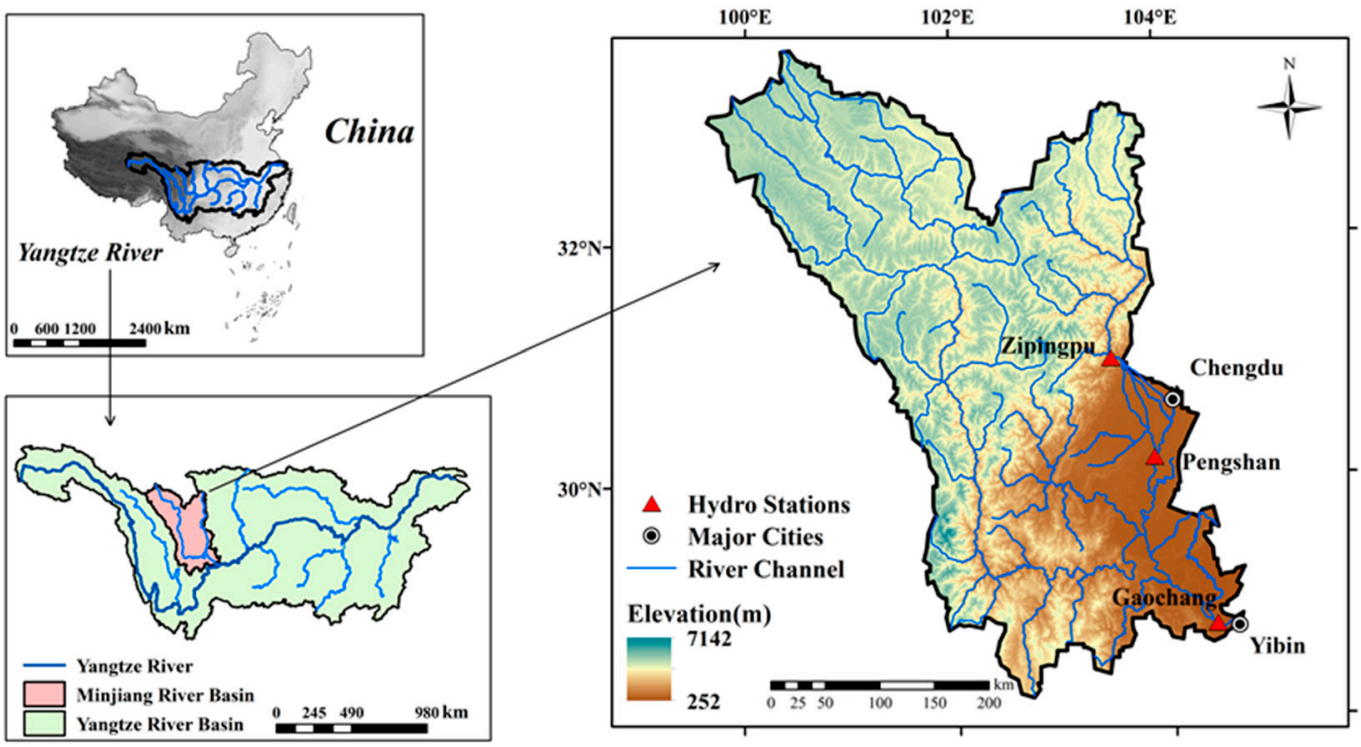


Figure 1. Location, river network and topography of the Minjiang River Basin.

The Zipingpu Reservoir which is located in the Dujiangyan City modifies upstream runoff of the Minjiang River within the MMR. The Zipingpu hydrological station monitors the runoff in the Minjiang River before and after the official impoundment of Zipingpu Reservoir, while the Pengshan and Gaochang hydrological stations monitor the streamflow at the outlet of the MMR and Minjiang River Basin, respectively. Thus, the streamflow variations in the Minjiang River within the MMR were represented by the streamflow variations at Pengshan station.

2.2. Data

The datasets for the GBEHM employed in this study consist of two groups: meteorological forcing data which are used to drive the GBEHM simulations, and underlying-surface datasets which are required for model parameterization.

In meteorological forcing data, the daily observations of near-surface meteorological variables which were obtained from the China Meteorological Administration (CMA) Surface Climate Daily Dataset v3.0 (http://data.cma.cn), while Daily precipitation was derived from the China Gauge-based Daily Precipitation Analysis (CGDPA) dataset [30]. All meteorological variables were prepared at a daily temporal resolution.

Underlying-surface datasets include topographic data, vegetation conditions data, soil texture information, soil hydraulic parameters, and land use data. Topographic data were provided by the Shuttle Radar Topography Mission (SRTM) digital elevation model with a spatial resolution of 90 m [31]. Vegetation conditions were obtained from the GIMMS LAI4g and FPAR4g remote-sensing products [32]. Soil texture information was extracted from the Harmonized World Soil Database (HWSD) [33]. The soil hydraulic parameters required by the van Genuchten formulation were adopted from the China Dataset of Soil Hydraulic Parameters [34]. Land use data at a spatial resolution of 100 m were provided by the Data Center for Resources and Environmental Sciences, Chinese Academy of Sciences (http://www.resdc.cn). More detail information of hydrological database can be found in [17].

The data-driven model and hydrological model adopted the daily water level and discharge data from the Pengshan, Zipingpu and Gaochang stations (Table 1) to reproduce daily hydrodynamic processes. These data were obtained from the Changjiang Water Resources Commission of the Ministry of Water Resources and the Hydrological Statistical Year book. The daily discharge of small tributaries and interval confluence of the Minjiang River were also simulated by the GBEHM. The data were used for simulation in the mainstream of the Minjiang River from 1980 to 2019. Meanwhile, the GBEHM was used to reproduce daily natural discharge at three stations for the following periods: Zipingpu station (1986–2019), Pengshan station (1988–2006), and Gaochang station (2000–2019). As shown in Figure 2, both precipitation and streamflow at Zipingpu and Pengshan stations exhibited decreasing trends over the period of 1980–2019.

Table 1. Basic information of the gauges in the study area.

ID	Station Name	Type	Time Scale	Period
1	Zipingpu	Hydrologic station	Daily	1980–1985
2	Pengshan	Hydrologic station	Daily	1980–1987, 2007–2019
3	Gaochang	Hydrologic station	Daily	1980–1999

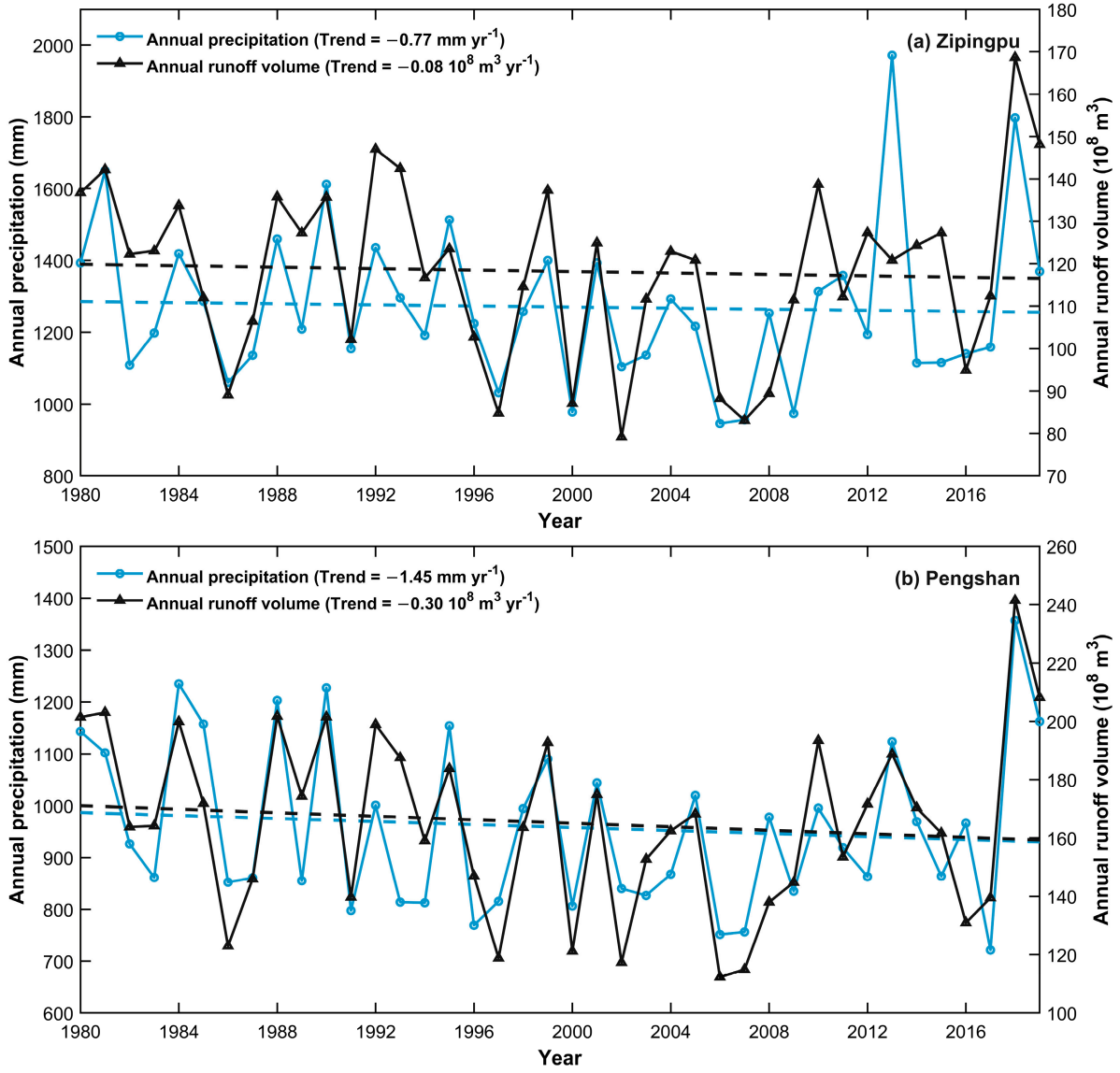


Figure 2. Long-term variations in precipitation and streamflow at Zipingpu and Pengshan station from 1980 to 2019.

3. Methodology

3.1. Trend and Mutation Test of Streamflow

Non-parametric trend tests can be used to test for significant trends in time series. The Mann–Kendall test (M-K test) was adopted to evaluate the long-term trend of mean annual discharge at the Pengshan station from 1980 to 2019. The M-K test is known as a nonparametric test used for trend and mutation detection in time series [35,36]. For a time-series of n observations, $X = x_1, x_2, \ldots, x_n$, we adopted Equations (1) and (2) to compute the M-K trend statistic (S).

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sign(x_j - x_i) \quad (1)$$

$$sign(x_j - x_i) = \begin{cases} 1, x_j > x_i \\ 0, x_j = x_i \\ -1, x_j < x_i \end{cases} \quad (2)$$

If the length of observations n exceeds 10, the standard normal variable Z can be used to evaluate the trend of observations series, as illustrated in Equations (3) and (4). If all values of observation series are distinct, the variance of $Var(S)$ can be computed by Equation (4). If the value of Z exceeds 0, this time-series has increasing trends; otherwise, the time-series has decreasing trends. If $|Z| \geq Z_{1-(\alpha/2)}$, the null hypothesis of no trend is rejected at the significance level α. The values of $Z_{1-(\alpha/2)}$ at the 5% and 10% significance levels are 1.96 and 1.28, respectively.

$$Z = \begin{cases} \frac{S-1}{\sqrt{Var(S)}}, S > 0 \\ 0, S = 0 \\ \frac{S+1}{\sqrt{Var(S)}}, S < 0 \end{cases} \quad (3)$$

$$Var(S) = \frac{n(n-1)(2n+5)}{18} \quad (4)$$

Given the impoundment of Zipingpu Reservoir, the mean annual discharge at Pengshan station was derived from GBEHM simulations for the pre-impoundment period and from observations for the post-impoundment period, while the trend and mutation test was conducted for both the pre-impoundment period and the post-impoundment period.

An autocorrelation function (ACF) test was performed on the simulated series, because significant autocorrelation in the simulations could inflate the significance of subsequent trend analyses on the mean annual discharge. The autocorrelation coefficient at lag k was computed by Equation (5):

$$r_k = \frac{\sum_{t=1}^{N-k}(x_t - \overline{x})(x_{t+k} - \overline{x})}{\sum_{t=1}^{N-k}(x_t - \overline{x})^2} \quad (5)$$

where $\overline{x}$ indicates the arithmetic mean of the time series; N indicates the length of time series, respectively. The confidence interval at significance level α was computed by Equation (6):

$$CI = \pm \frac{Z_{1-\alpha/2}}{\sqrt{N}} \quad (6)$$

For lags $k > 0$, significant autocorrelation is identified when $|r_k|$ exceeds the 95% confidence limit.

3.2. Coupling Simulation and Scenario Experiments

Before the impoundment of Zipingu Reservoir, streamflow in the Minjiang River within the MMR was influenced by the Chengdu metropolitan area and climate change, whereas for the after-impoundment period, it was also influenced by operation of the reservoir. A hydrological-hydrodynamic coupling simulation was conducted to present the contributions of anthropogenic disturbance and climate change on the streamflow variations at the outlet of the MMR. The GBEHM reproduced the non-reservoir runoff process in the upper and middle Minjiang River basins, simulating streamflow at Zipingpu station over the entire study period and at Pengshan station before the impoundment of Zipingpu Reservoir. The cascade of the PaHL model and XGBoost model quantified the anthropogenic streamflow regulation in the Minjiang River at the daily scale. In the MMR, streamflow at Zipingpu station, tributaries of the Minjiang River and lateral flow to the Minjiang River, simulated by the GBEHM, were used as boundary conditions for data-driven models, while the output of these models was the streamflow at Pengshan station. For the coupling simulation, streamflow at Zipingpu station, tributaries of the Minjiang River and lateral flow to the Minjiang River represent the influence of climate

change on the runoff generation within the MMR, while the key parameters of data-driven models reflect the effects of anthropogenic disturbance.

Given that Zipingpu Reservoir poses an influence on downstream streamflow from 2007 to 2019, the study period was initially divided into three equal parts: the initial period (1980–1993) without major human disturbances, the pre-reservoir period (1994–2006) with urbanization but before reservoir impoundment, and the post-reservoir period (2007–2019) with both urbanization and reservoir operation. Hence, the influences of climate change and the Chengdu metropolitan area dominate the streamflow variation between the initial period and the pre-reservoir period, whereas the change between the pre-reservoir period and the post-reservoir period is attributable to all three factors. If the M-K test detected the mutation in the pre-impoundment period, the length of the initial period and the pre-reservoir period will be accordingly adjusted. The states of three influencing factors vary across three periods. According to the distinct combinations of these states, seven scenarios were constructed as illustrated in Table 2. Since the simulated streamflow at Pengshan station reflects the response of streamflow to the influencing factors, the differences between simulations across periods quantify the temporal variation in their influence.

Table 2. The state of each influencing factor in scenario experiments.

Scenario	Baseline_1	Climate_1	Human_1	Baseline_2	Climate_2	Human_2	Baseline_3
State of climate change	1980–1992	1994–2006	1980–1992	1994–2006	2007–2019	1994–2006	2007–2019
State of Chengdu metropolitan area	1980–1992	1980–1992	1994–2006	1994–2006	1994–2006	2007–2019	2007–2019
State of Zipingpu Reservoir						2007–2019	2007–2019

In all six scenarios, Baseline_1~Baseline_3 were selected as the baseline scenarios in the initial period, the pre-reservoir period and the post-reservoir period, respectively. From the initial period to the pre-reservoir period, the contribution of each influencing factor on runoff variations were computed as follows:

$$CC = ((Q_{cliamte_1} - Q_{baseline_1}) + (Q_{cliamte_2} - Q_{baseline_2}))/2 \quad (7)$$

$$CM = ((Q_{human_1} - Q_{baseline_1}) + (Q_{baseline_2} - Q_{climate_1}))/2 \quad (8)$$

where *CC* and *CM* indicate the contributions of climate change and Chengdu metropolitan area respectively. $Q_{baseline_1}$, $Q_{cliamte_1}$, Q_{human_1}, and $Q_{baseline_2}$ indicate simulated discharge at multiple scenarios, respectively.

From the pre-reservoir period to the post-reservoir period, the contribution of each influencing factor on runoff variations were calculated as follows:

$$CC = ((Q_{climate_2} - Q_{baseline_2}) + (Q_{baseline_3} - Q_{human_2}))/2 \quad (9)$$

$$CM + ZR = ((Q_{human_2} - Q_{baseline_2}) + (Q_{baseline_3} - Q_{climate_2}))/2 \quad (10)$$

where *CC* and *CM* + *ZR* indicate the influence of climate change and the coupling influence of Chengdu metropolitan area and Zipingpu Reservoir, respectively. $Q_{bseline_2}$, $Q_{climate_2}$, Q_{human_2}, $Q_{baseline\text{-}3}$ indicate simulated discharge at multiple scenarios, respectively. The framework of coupling simulation and scenario experiments is shown in Figure 3.

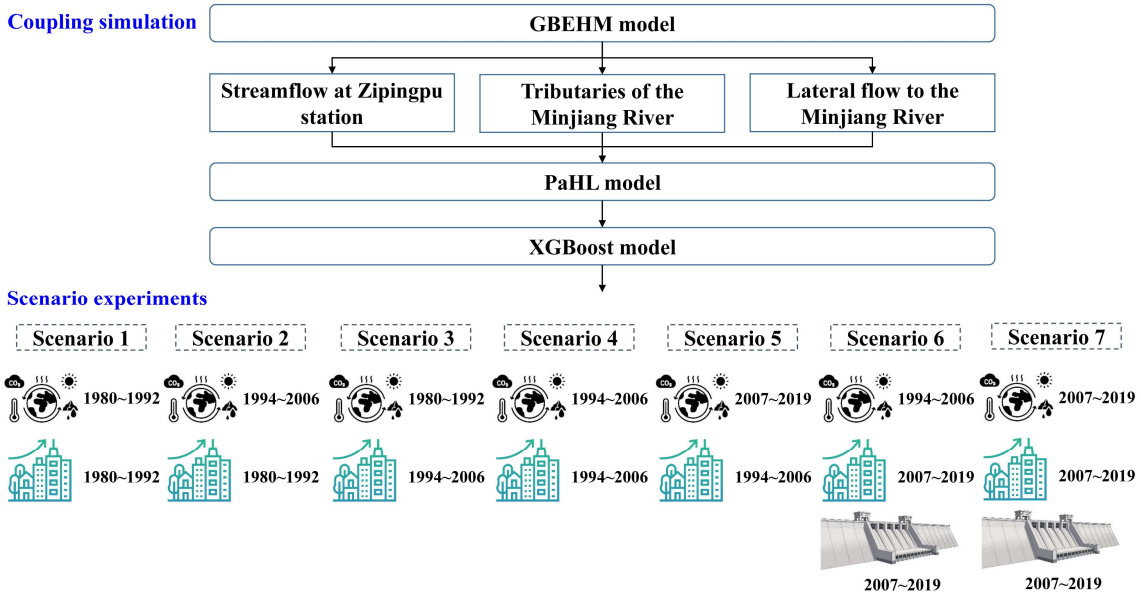


Figure 3. The framework and data flow of coupling simulation and scenario experiments.

3.3. Construction of the PaHL Model

The Physic-aware Hybrid Learning (PaHL) model constructs an idealized river channel strengthening the inherent linear relationships within the nature channel. The governing equations of the PaHL model were derived from the Saint-Venant equations, and a genetic algorithm (GA) is employed to optimize the key model parameters. Detail introduction of the GA module in the PaHL model could be found in Appendix B.

The main channel of Minjiang River within the MMR could be regarded as a 1D channel, thus the hydrodynamic process in this channel follows the Saint-Venant equations:

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = q_L \tag{11}$$

$$\frac{\partial Q}{\partial t} + \frac{\partial}{\partial x}(\alpha Q u) + gA\frac{\partial Z}{\partial x} + gA\frac{Q^2 n^2}{AR^{\frac{4}{3}}} = q_L V_x \tag{12}$$

where A indicates the area of cross section, m^2; Q indicates the discharge, m^3/s; q_L indicates the single-width interval confluence, m^2/s; a indicates the momentum adjustment coefficient, dimensionless; Z indicates water level, m; g indicates gravitational acceleration, m/s^2; u indicates velocity, m/s; n indicates roughness, dimensionless; R indicates hydraulic radius, m; V_x indicates the component of interval confluence flow velocity, m/s; t indicates time, s; x indicates length of the channel, m.

The typical solution of the Saint-Venant equations adopted finite difference method and Preissmann four-point eccentric scheme to discretize Equations (11) and (12). The linear implicit differential form of Equation (11) is as follows:

$$(Z_j^n, Z_{j+1}^n, Z_j^{n+1}, Q_j^n, Q_{j+1}^n, Q_j^{n+1}, Q_{j+1}^{n+1}, E)b_1 = Z_{j+1}^{n+1} \tag{13}$$

$$(Z_j^n, Z_{j+1}^n, Z_j^{n+1}, Q_j^n, Q_{j+1}^n, Q_j^{n+1}, Q_{j+1}^{n+1}, E)b_1 = Z_{j+1}^{n+1} \tag{14}$$

where the superscript and subscript of variables indicate time step and cross section number, respectively; C_j relates to the width of cross section B, m; D_j is a linear function of Q_{j+1}, Q_j, Z_{j+1}, Z_j at the nth time step, b_1, and b_2 are column vectors composed of coefficients which can be obtained using the least squares method. Assuming that changes in channel

width B are neglected, coefficients in b_1, and b_2 are constant.According to Equation (11), the temporal changes in A in is influenced by the spatial changes in Q, thus:

$$\frac{\partial A}{\partial Z}\frac{\partial Z}{\partial t}+\frac{\partial Q}{\partial x}=q_L \tag{15}$$

In Equation (15), we made a central difference for Q and a forward difference for Z, ignoring the single-width interval confluence q_L was ignoring. We constructed an iterative algorithm to calculated character water level Z_c at the tth time step, as illustrated in Appendix A:

$$Z_c=a(Z_{up}-Z_{down})+b(Z_{up}+Z_{down}) \tag{16}$$

$$Z_c^{t+1}=Z_c^t+({q_L}^t-\frac{({Q_{down}}^t-{Q_{up}}^t)}{\Delta x})\Delta t/t_1 \tag{17}$$

where Z_{up}, Z_{down}, Q_{up} and Q_{down} are discharge and water level at the upstream and downstream cross sections; a, b and t_1 are parameters to be determined. The initial parameters, namely Z_c, q_L, Q_{up} and Q_{down} were calculated by observed data, thus we can obtain Z_c at any time step on the basis of appropriate estimation of coefficients a, b and t_1. Subsequently, unknown Z_{down} was calculated by Z_c and observed Z_{up}, according to Equation (16). We adopted a linear algorithm and a nonlinear algorithm to calculate unknown Q_{down}: (1) yielding the column vectors b_1, and b_2 by solving Equations (13) and (14) simultaneously, then Z_{down} and Q_{down} can be obtained by recursions; (2) getting Z_c and Z_{down} at any time step, yielding the column vectors b_2 by solving Equation (14), then calculating Q_{down} according to Equation (14).

For the purpose of strengthening inherent linear relationship between the upstream and downstream water level and discharge, the upstream water level and discharge were iterated by data assimilation:

$$Variable_i=(Variable_o+\sum_{i=1}^{n}Variable_sIndicator_s)/(1+\sum_{i=1}^{n}Indicator_s) \tag{18}$$

where $Variable_o$, $Variable_i$ and $Variable_s$ indicate the observed, the iterated and the simulated water level or discharge at upstream cross section, respectively. $Indicator_s$ indicates the simulation error indicators, including the Nash–Sutcliffe efficiency (NSE), Coefficient of Determination (R^2), Kling–Gupta Efficiency (KGE). n indicate the numbers of algorithms for calculating unknown Z_{down} and Q_{down}.

$$NSE=1-\sum_{i=1}^{n}(M_i-S_i)^2/\sum_{i=1}^{n}(M_i-\overline{M})^2 \tag{19}$$

$$R^2=(\sum_{i=1}^{n}(M_i-\overline{M})(S_i-\overline{S})/\sqrt{\sum_{i=1}^{n}(M_i-\overline{M})^2\sum_{i=1}^{n}(S_i-\overline{S})^2})^2 \tag{20}$$

$$KGE=1-\sqrt{(r-1)^2+(\alpha-1)^2+(\beta-1)^2} \tag{21}$$

where S and M indicate simulated values and measured values, respectively; $\overline{S}$ and $\overline{M}$ indicate the mean simulated and measured values, respectively; r, α and β indicate Pearson correlation coefficient, variability ratio and bias ratio, respectively. Within the PaHL iterative scheme, the simulated downstream discharge is progressively updated at each iteration based on the discrepancy between the assimilated discharge and the observed downstream discharge. As the iteration proceeds, the residual will gradually decreases, and the assimilated discharge converges toward the observed downstream discharge. The convergence behavior is governed by Equation (15), which ensures that the iterative update reduces the residual to an acceptable magnitude.

Above all, the PaHL model has three modules, namely the data assimilation module, the Genetic Algorithm (GA) module, and the Physics module. The structure of the PaHL model and data flow within two adjacent iterations were shown in Figure 4. In the $(N + 1)^{th}$ iteration, observed upstream discharge and water level were assimilated with simulated downstream discharge and water level which were output by the former iteration, according to Equation (18), while the error indicators adopted by Equation (18) were also calculated in the former iteration. The GA module and the Physics module were tightly coupled in a bidirectional manner, where the key parameters including a, b and t_1 as well as the error indicators were calculated and exchanged. Specifically, the discharge and water level during the simulation period were solved iteratively by the Physics module, according to Equations (16) and (17). The simulation outputs, together with corresponding key parameters and error indicators, were subsequently fed into the GA module. The GA module was employed to achieve the best estimation of key parameters. Finally, the Physics module outputted the best simulation results and corresponding error indicators to the next iteration.

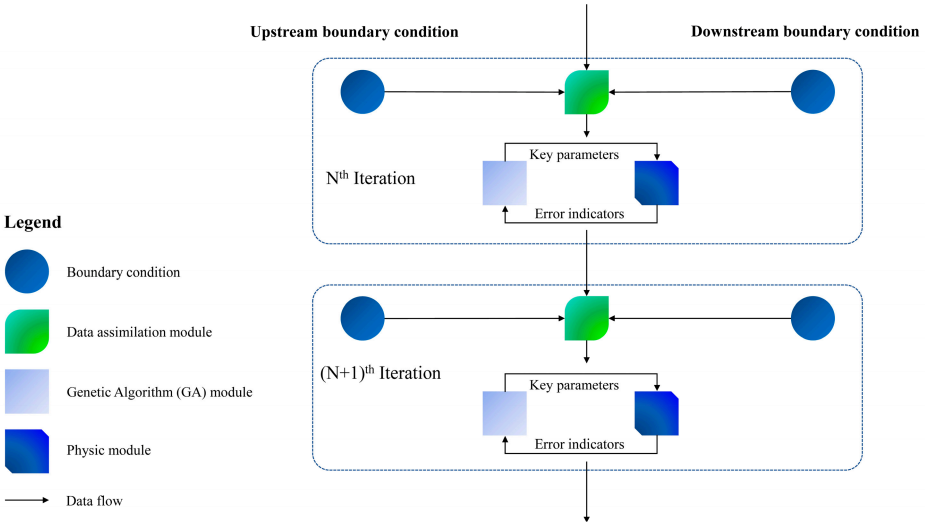


Figure 4. The structure and data flow of the PaHL model with 2 iterations.

3.4. Construction of the XGBoost Model

XGBoost (eXtreme Gradient Boosting) is an ensemble learning framework based on gradient-boosted decision trees [37]. It enhances traditional Gradient Boosting Decision Tree model with regularization, second-order approximation of the loss, column subsampling, and tree complexity control, effectively preventing overfitting and accelerating training. XGBoost supports custom loss functions, handles missing values automatically, and is robust to feature collinearity, making it ideal for structured tabular data and nonlinear regression tasks. Its core objective is as follows:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \tag{22}$$

where l and Ω are the loss and a regularization term, respectively. By iteratively adding weak learners, XGBoost minimizes residuals and achieves strong predictive performance.

In this study, the PaHL model adopted 21 iterations to achieve satisfactory simulation accuracy. The simulated downstream discharge and water level from all but the first iteration were then fed into the XGBoost model, which learned a mapping from the simulation results of the PaHL model to the observed downstream discharge. The outputs of the

XGBoost model were adopted as the final simulation results of the data-driven models, as shown in Figure 5.

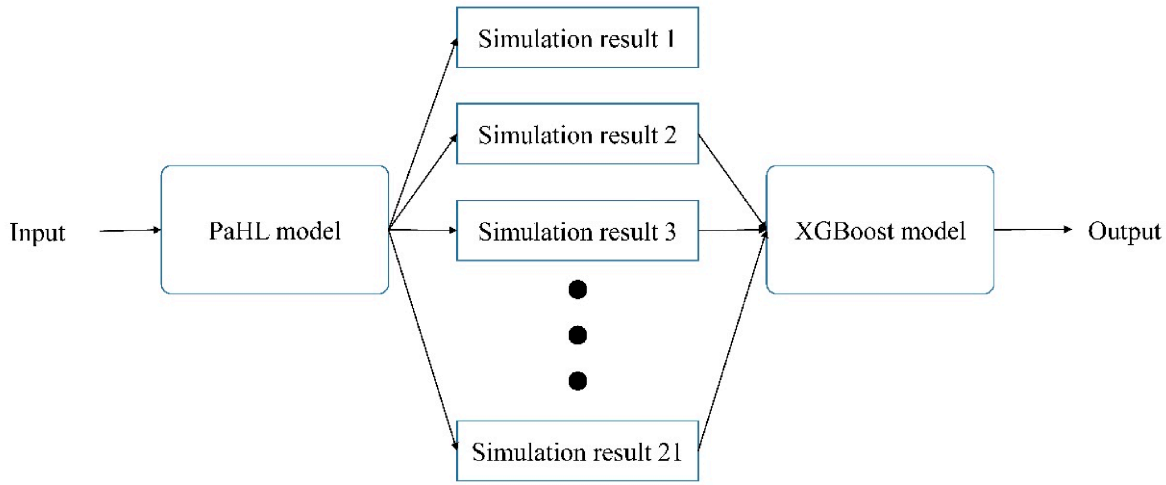


Figure 5. The data flow between the PaHL model and the XGBoost model.

3.5. Geomorphology-Based Ecohydrological Model

The Geomorphology-Based Ecohydrological Model (GBEHM) is a physically based, distributed ecohydrological model developed to simulate coupled water, energy, and vegetation processes in river basins with complex terrain [38,39]. In this research, we adopted a previously established GBEHM to reproduce the hydrological process in the Minjiang River Basin [17].

The GBEHM adopts a geomorphology-informed spatial framework where snow, soil, and vegetation are treated as active components of the land surface system, allowing key ecohydrological feedbacks to be captured while maintaining computational efficiency at the basin scale [24,34,40].

The GBEHM adopts a multi-layer snow scheme that explicitly resolves mass and energy balance within the snowpack to simulate snow accumulation and melt [41,42]:

$$C_{snow}\frac{\partial T_{snow}}{\partial t} - L_f\frac{\partial \rho_{ice}\theta_{ice}}{\partial t} = \frac{\partial}{\partial z}\left(k_{snow}\frac{\partial T_{snow}}{\partial z}\right) + \frac{\partial I_r}{\partial z} + Q_r \quad (23)$$

where C_{snow} is the volumetric heat capacity of snow; T_{snow} is the snow temperature; t denotes time; L_f is the latent heat of fusion of ice; ρ_{ice} is the density of ice; θ_{ice} is the volumetric ice content; z is the vertical coordinate measured downward from the snow surface; k_{snow} is the thermal conductivity of snow; I_r represents radiative flux penetrating into the snowpack; Q_r denotes the heat input associated with precipitation.

A fully coupled framework that accounts for conductive heat transfer, advective heat transport associated with water movement, and latent heat exchange during freezing and thawing is used to solve the soil thermal dynamics [41,43]:

$$\rho_{ice}\lambda_f\frac{\partial \theta_{ice}}{\partial t} - C_s\frac{\partial T}{\partial t} + \frac{\partial}{\partial z}(k_s\frac{\partial T}{\partial z}) + \rho_{ice}C_{liq}\frac{\partial q_{liq}T}{\partial z} = 0 \quad (24)$$

where T denotes soil temperature; q_{liq} is vertical liquid-water flux between soil layers; ρ_{liq} is the densities of liquid water; λ_f is the latent heat of fusion for the ice–water phase change; C_s is the bulk volumetric heat capacity of the soil matrix; k_s is the soil thermal conductivity; C_{liq} is the specific heat capacity of liquid water.

The GBEHM employs the Richards equation to describe vertical water movement in soil. Unsaturated hydraulic conductivity is formulated as a function of unfrozen

liquid water content and an ice-impedance factor following a modified Mualem–van Genuchten formulation:

$$K = f_{ice} K_s \left(\frac{\theta_{liq} - \theta_r}{\theta_s - \theta_r}\right)^{\frac{1}{2}} \left[1 - \left(1 - \left(\frac{\theta_{liq} - \theta_r}{\theta_s - \theta_r}\right)^{\frac{1}{m}}\right)^m\right]^2 \quad (25)$$

where K denotes the vertical hydraulic conductivity; K_s is saturated hydraulic conductivity; θ_{liq} represents the volumetric content of unfrozen liquid water; θ_s and θ_r are saturated and residual water contents, and m is a dimensionless shape parameter of the soil water retention curve [44]; The factor f_{ice} is an ice-impedance term that decreases with increasing volumetric ice content and is commonly specified using empirical relationships [45].

The photosynthesis is simulated using the biochemically based Farquhar model, in which canopy assimilation is limited by either Rubisco activity or electron transport [46]:

$$A = \min(w_c, w_j) \quad (26)$$

where A is the net photosynthetic rate; w_c is the Rubisco-limited carboxylation rate, and w_j is the electron-transport-limited carboxylation rate.

Besides the *NSE* and R^2, the percent bias (*PBIAS*) was also adopted to evaluate the accuracy of simulation results:

$$PBIAS = \frac{\sum_{t=1}^{T}(S - M)}{\sum_{t=1}^{T} M} \times 100\% \quad (27)$$

where S and M indicate simulated values and measured values, respectively.

4. Results

4.1. Performance Evaluation of the GBEHM

The error indicators of simulation results obtained by the GBEHM were summarized in Table 3. The observed and simulated daily runoff show good overall agreement during the calibration (1983–1992) and validation (1993–1999) periods at Gaochang station, with *NSE* and R^2 generally above 0.8 and the *PBIAS* of −6.75%. The GBEHM also simulated daily runoff at Zipingpu and Pengshan stations. The simulation accuracy at Zipingpu station from 1980 to 1985 is slightly worse than that at Gaochang station, with *NSE* and R^2 below 0.80, and the *PBIAS* of −6.38%. From 1980 to 1985, *NRE* and R^2 of simulated discharge at Pengshan station range from 0.77 to 0.88, while *PBIAS* of simulation results is 7.13%. Overall, the GBEHM adequately represents dominant runoff generation and routing processes in the MMR.

Table 3. Performance statistics of daily runoff simulations at Gaochang hydrological station during calibration and validation periods.

Region	Station	Time Period	*NSE*	R^2	*PBIAS* (%)
Minjiang River	Gaochang	Calibration (1983–1992)	0.86	0.88	−8.23
		Validation (1993–1999)	0.81	0.84	−6.75
	Pengshan	Validation (1980–1985)	0.77	0.88	7.13
	Zipingpu	Validation (1980–1985)	0.74	0.76	−6.38

4.2. Long-Term Variations in Streamflow

For the pre-impoundment period, mean annual discharge at Pengshan station which was simulated by the GBEHM was used for the trend and mutation test. The result of

autocorrelation test for the pre-impoundment mean annual discharge at Pengshan station is shown in Figure 6. None of the autocorrelation coefficients at lags $k \geq 1$ exceeded the 95% confidence limit of ±0.38, indicating the absence of autocorrelation.

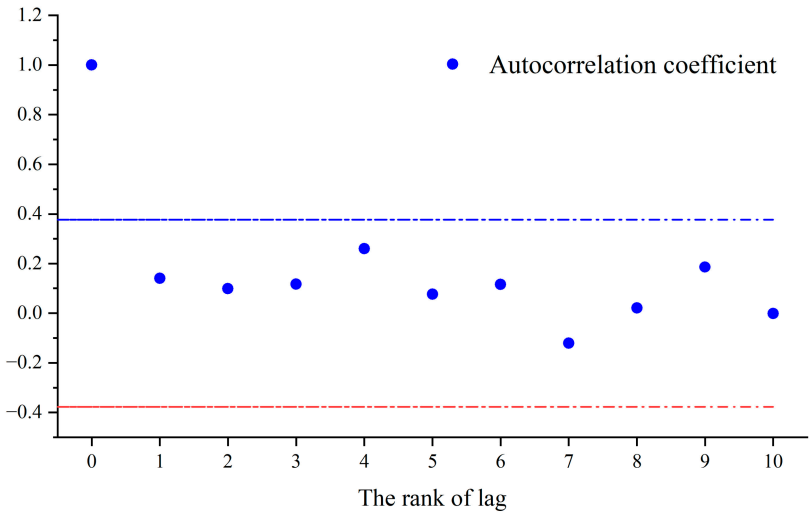


Figure 6. Autocorrelation coefficients for the pre-impoundment mean annual discharge at Pengshan station.

The mean annual discharge at Pengshan station exhibited a decreasing trend during the pre-impoundment period and an increasing trend during the post-impoundment period, as shown in Figure 7. The trend and mutation of mean annual discharge discharge at Pengshan station was tested by M-K test, as shown in Figure 8. During the pre-impoundment period and the post-impoundment period, the M-K test yielded Z statistics of −2.877 and 1.769, respectively, with a critical value of ±1.96 at the 5% significance level. Accordingly, a significant decreasing trend in mean annual discharge was detected at Pengshan station before the impoundment of the Zipingpu Reservoir, while no significant trend was found after the impoundment. Meanwhile, no mutation was detected in the mean annual discharge at Pengshan station, as shown in Figure 8.

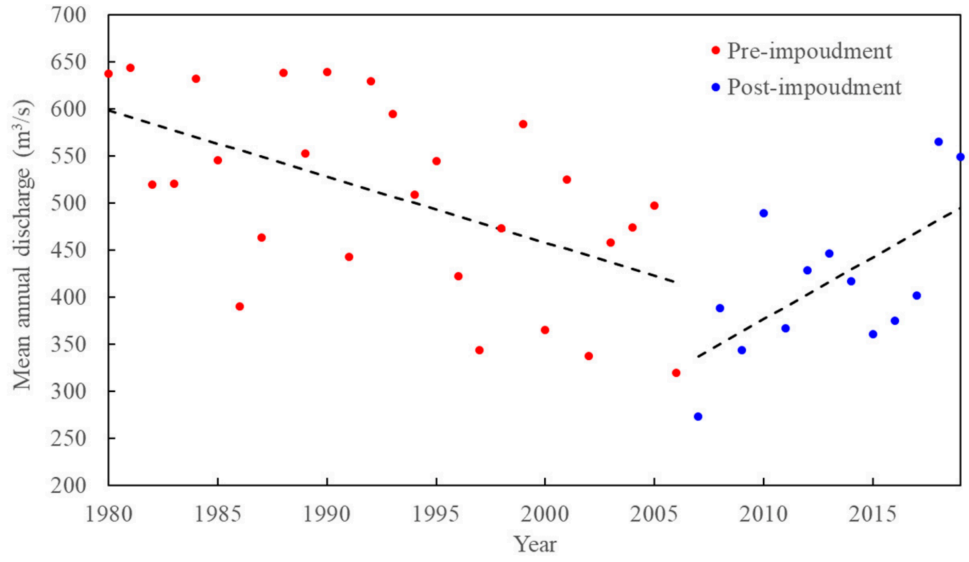


Figure 7. The trend of mean annual discharge at Pengshan station.

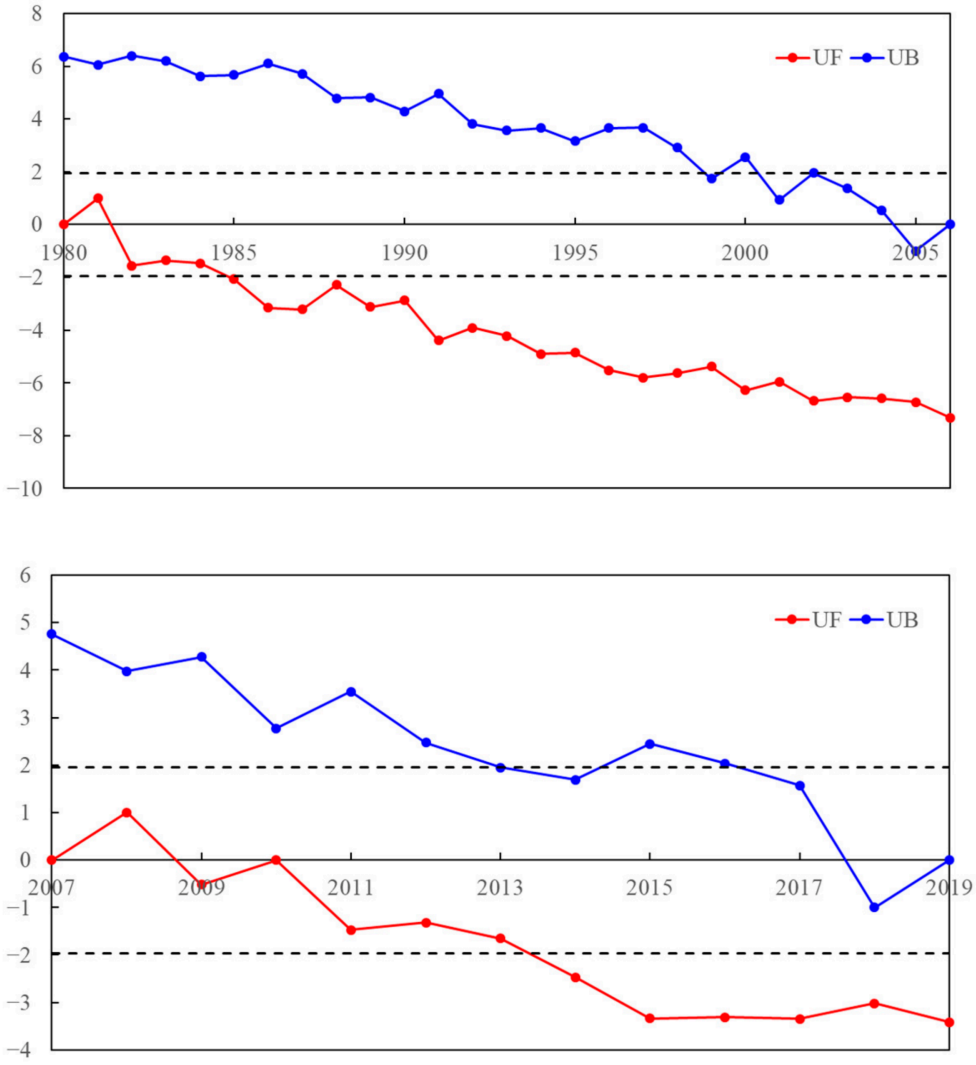


Figure 8. Results of the M-K test for the mean annual discharge at Pengshan station before (1980–2006) and after (2007–2019) the impoundment of Zipingpu Reservoir, respectively.

Therefore, the temporal division of the study period into three periods follows the initial scheme, that is, each simulation period lasts for 13 years. As illustrated in Table 4, the maximum, mean and minimum values of annual average discharge in each simulation period simultaneously exhibit a decreasing trend, indicating a decline in annual mean discharge at Pengshan station from 1980 to 2019.

Table 4. Maximum, mean and minimum values of annual average discharge in the initial period (1980–1992), pre-reservoir period (1994–2006) and post-reservoir period (2007–2019), respectively.

Simulation Period	Initial Period	Pre-Reservoir Period	Post-Reservoir Period
Maximum annual average discharge (m^3/s)	643.86	583.54	564.68
Mean annual average discharge (m^3/s)	557.98	450.11	415.60
Minimum annual average discharge (m^3/s)	389.97	319.20	273.93

4.3. Performance of the Cascade of Data-Driven Model

Given the absence of significant autocorrelation in the simulation results of GBEHM, the entire study period was evenly divided into three 13-year periods. For each period, the cascade of data-driven models used 9 years for calibration, 2 years for validation, and 2 years for testing.

Water level and discharge at Zipingpu and Pengshan stations, as well as the discharge of tributaries and lateral inflow were adopted to construct the PaHL and XGBoost models, while the water level was estimated by water level-discharge curve. For the PaHL model, the GA module automatically calibrates key parameters. Specifically, key parameter a ranges from -9.70 to 9.83, from 3.74 to 9.91 and from -2.94 to 9.54 in the initial period, pre-reservoir period and the post-reservoir period, respectively. Corresponding values of key parameter b ranges from 0.02 to 1.98, from 0.10 to 0.51 and from 0.54 to 1.89, respectively, while the value of t_1 was related to the values of a and b.

For the XGBoost model, given the limited size of the training dataset (~5000 records) relative to the 40 predictor features, we constrained the number of boosting rounds to 500. This setting acts as a capacity control that provides sufficient ensemble depth to capture the underlying interactions between features. The remaining hyper-parameters were calibrated through systematic trials aimed at minimizing error indicators in the calibration period. Throughout the calibration process, we monitored error trends to avoid overfitting and underfitting. Eventually, the XGBoost model was configured with the following hyper-parameters: the number of boosting trees was set to 500, the learning rate to 0.03, and the maximum tree depth to 8. To further control overfitting, the subsample ratio of training instances and the column subsample ratio of features were both set to 0.9, and L2 regularization was applied with a lambda of 1.0. The objective function adopted the standard squared error loss.

The *NSE*, R^2 and *KGE* were adopted to evaluate simulation errors at Pengshan station. As Table 5 illustrated, during the initial period and pre-reservoir period, the cascade of data-driven models achieved good accuracy, with values of three error indicators exceeding 0.97. However, while values of three error indicators approximate 0.99 in the calibration period during the post-reservoir period, the accuracy of data-driven models obviously decreases in the validation and test periods, with values of three error indicators ranging from 0.50 to 0.88. As shown in Figure 9, the simulations derived by data-driven models generally capture the peaks and overall variations in daily discharge at Pengshan station, demonstrating an effective reproduction of the streamflow dynamics.

Table 5. Performance statistics of daily streamflow simulated by the cascade of models at Pengshan station during calibration, validation and test periods in initial period (1980–1992), pre-reservoir period (1994–2006) and post-reservoir period (2007–2019), respectively.

Period	Simulation Period	*NSE*	R^2	*KGE*
Initial period	Calibration (1980–1988)	1.00	1.00	1.00
	Validation (1989–1990)	0.99	0.99	0.99
	Test (1991–1992)	0.99	0.99	0.99
Pre-reservoir period	Calibration (1994–2002)	0.99	0.99	0.99
	Validation (2003–2004)	0.96	0.96	0.97
	Test (2005–2006)	0.96	0.96	0.97
Post-reservoir period	Calibration (2007–2015)	0.99	0.99	0.99
	Validation (2016–2017)	0.50	0.50	0.72
	Test (2018–2019)	0.77	0.77	0.88

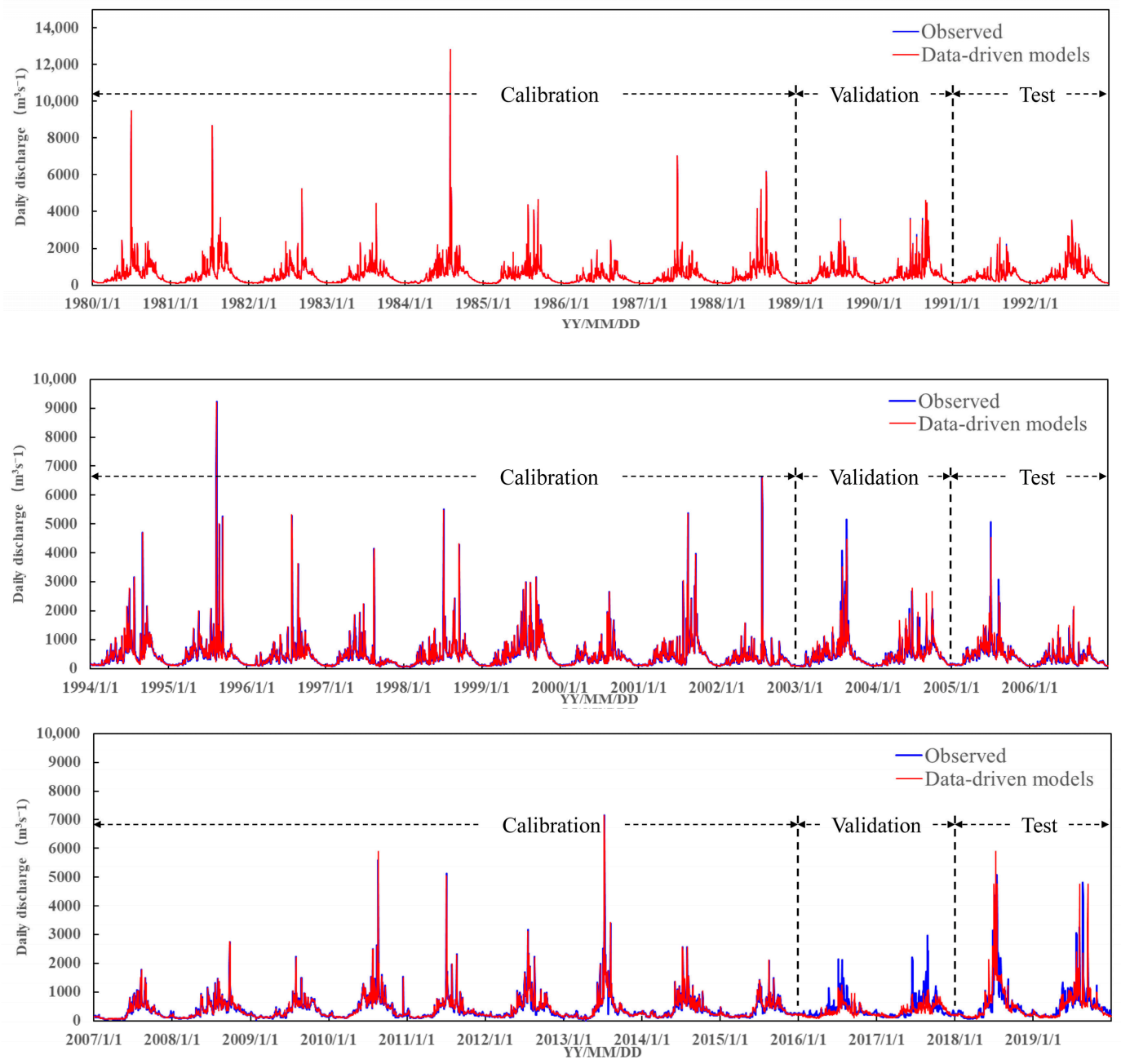


Figure 9. Comparison of observed and daily runoff by the data-driven models at Pengshan station in the initial period (1980–1992), the pre-reservoir period (1994–2006) and the post-reservoir period (2007–2019), respectively, with calibration, validation and test periods indicated.

4.4. Attribution Analysis of Variations in Streamflow Variations

Scenario experiments were on the basis of daily streamflow simulations at Pengshan station. Subsequently, the influences of Chengdu metropolitan area (CM), climate change (CC), and the Zipingpu Reservoir (ZR) on daily discharge variations at Pengshan Station were calculated between the baseline period and the recent period.

As illustrated in Table 6, the initial period and pre-reservoir period were chosen as the baseline period and the recent period, respectively. Compared with the initial period, mean daily discharge at Pengshan station decreases for 105.3 m^3/s, while climate change and Chengdu metropolitan area explained 37.7% (−39.7 m^3/s) and 62.3% (−65.6 m^3/s) of this decrease, respectively.

Correspondingly, for the post-impoundment period, the pre-reservoir period and post-reservoir period were chosen as the baseline period and the recent period, respectively. During the post-reservoir period, mean daily discharge at Pengshan station is 45.6 m^3/s lower than that during the pre-reservoir period. Climate change and anthropogenic disturbance including Chengdu metropolitan area and Zipingpu Reservoir account for −330.8% (150.8 m^3/s) and 430.8% (196.3 m^3/s) of this change, respectively.

Overall, anthropogenic disturbance is the dominant role of streamflow variations at the outlet of the MMR, which is obvious during the post-impoundment period.

Table 6. Contribution rates of the Chengdu metropolitan area (CM), climate change (CC), and the Zipingpu Reservoir (ZR) to variations in mean and median daily discharge difference at Pengshan Station for each period.

Baseline Period	**Initial Period**			**Pre-Reservoir Period**		
Recent period	Pre-reservoir period			Post-reservoir period		
Mean daily discharge in the baseline period (m^3/s)	557.9			452.6		
Influencing factor	CC	CM	CC and CM	CC	CM and ZR	CM, ZR and CC
Mean daily discharge changes (m^3/s)	−39.7	−65.6	−105.3	150.8	−196.3	−45.6
Contribution rate	37.7%	62.3%	100.0%	−330.8%	430.8%	100.0%

4.5. Mechanisms of Influencing Factors on Streamflow Variations at Multiple Timescales

At the annual scale, the influences of climate change and anthropogenic disturbance on discharge variations at Pengshan station were calculated by daily results of scenario experiments, as shown in Figure 10. During the pre-reservoir period and post-reservoir period, annual discharge at Pengshan station in each year was compared to its order-matched counterpart during the baseline period (the initial period and the pre-reservoir period). Before the impoundment of Zipingpu Reservoir, climate change altered mean annual discharge at Pengshan station by −232.19–132.17 m^3/s, resulting in an overall decrease. For the post-impoundment period, the influence of climate change generally increases the mean annual discharge at Pengshan station, altering this discharge by −54.8–402.90 m^3/s. The anthropogenic disturbance has a steady influence on the mean annual discharge at Pengshan station, reducing it by 39.94–98.61 m^3/s during the pre-impoundment period and by 167.00–285.15 m^3/s during the post-impoundment period, respectively.

At the monthly scale, the influence of climate change increases the mean monthly discharge at Pengshan station from February to May and decreases this discharge from June to January of the next year before the impoundment of Zipingpu Reservoir. During the post-impoundment period, the influence of climate change leads to a significant increase in the mean monthly discharge at Pengshan station from June to October by 148.07–668.68 m^3/s, while the effects in other months were much weaker. In contrast to climate change, whose influence mechanism may have varied, anthropogenic disturbance influences streamflow through a consistent mechanism before and after impoundment. The anthropogenic disturbance generally poses a generally decreasing effect on mean monthly discharge at Pengshan station from March to October, and an increasing effect from November to January of the next year. However, this decrease became more pronounced after the impoundment of Zipingpu Reservoir, with a reduction of −33.55–−49.57 m^3/s in the pre-impoundment period and −156.53–−172.29 m^3/s in the post-impoundment period.

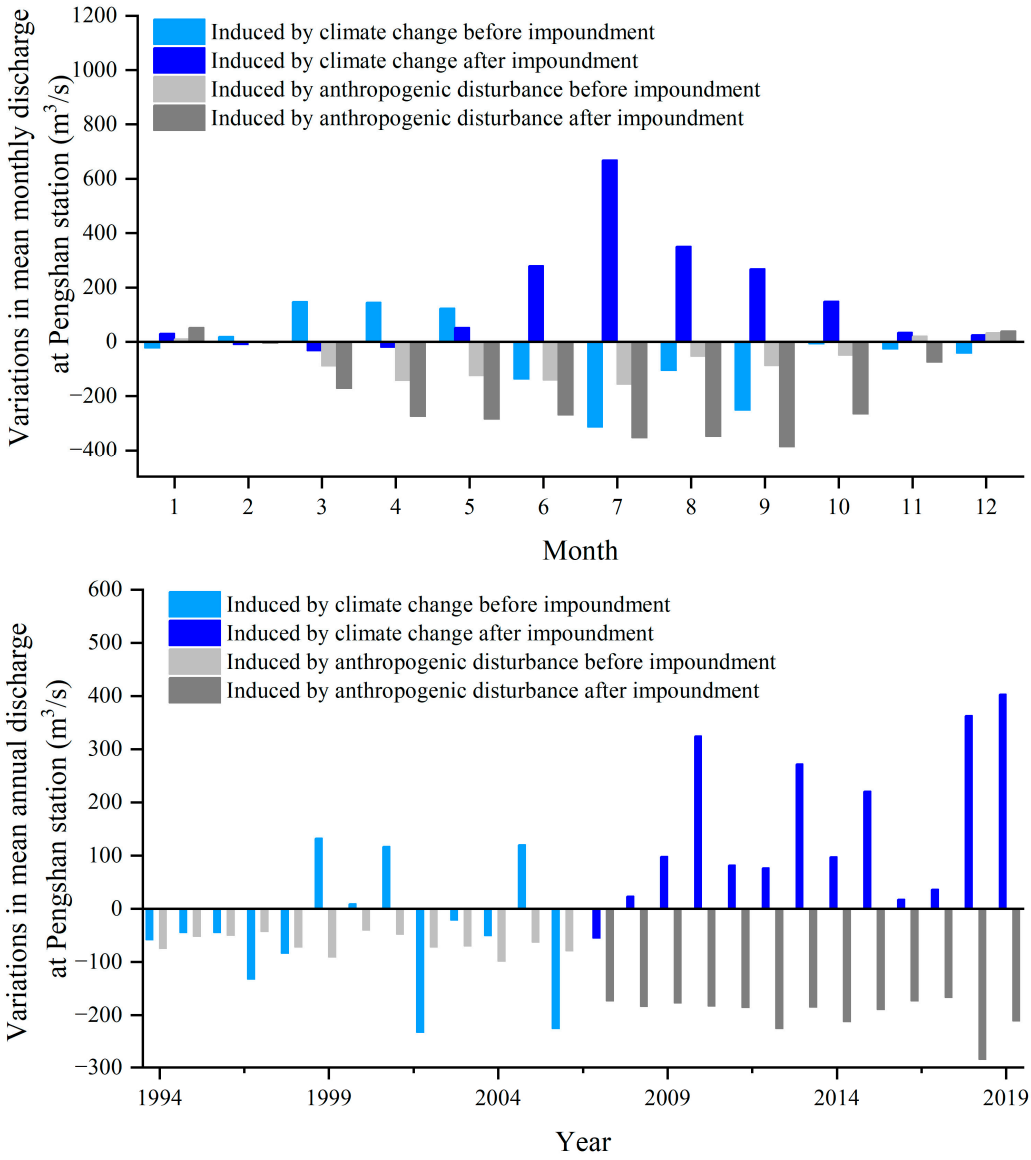


Figure 10. Variations in discharge at Pengshan station induced by climate change and anthropogenic disturbance at the annual and monthly scales, respectively, during the pre-impoundment and post-impoundment periods.

5. Discussion

5.1. Performance and Advantage of the Cascade of Data-Driven Models

Accurate hydrodynamic modeling requires detailed topography in cross sections, operation schedules of hydraulic engineering, and correct boundary conditions. In the MMR, however, the combined effects of a complex river network and dense hydraulic engineering pose considerable difficulties for model setup and reliable hydrodynamic simulation [12,22,23,26,47–49]. To date, hydrodynamic simulations have rarely quantified how the Zipingpu Reservoir and the Chengdu metropolitan area affect runoff at the outlet cross section of the MMR.

The PaHL model is governed by the Saint-Venant equations, which effectively capture complex hydrodynamics such as pronounced backwater effects and frequent flow reversals. The Physics module of the PaHL model linearizes the upstream–downstream boundary relationship, whereas the XGBoost model effectively maps the outputs of the PaHL model to observed downstream discharge. Therefore, the cascade of the PaHL model and XGBoost model successfully reproduced the combined effects of climate change and anthropogenic

disturbance on runoff at the outlet cross section of the MMR, as illustrated in Table 5 and Figure 9. From 1980 to 2015, the cascade of data-driven models achieved high accuracy, with values of *NSE*, R^2 and *KGE* ranging from 0.96 to 1.00. These metrics collectively confirm that the cascade of the PaHL model and XGBoost model accurately reproduce the timing, variability, and overall water balance of the observed daily streamflow during 35 years. The high NSE further demonstrates its reliability in capturing high-flow events, while the ideal KGE components (r, α, β) guarantee a faithful representation of the frequency distribution of streamflow. However, the pronounced decrease in accuracy during the validation and test periods of the post-reservoir period suggests that the data-driven cascade simulations may be over-fitted. Further work is therefore required to improve model performance.

Overall, this cascade of data-driven models overcame the difficulties arising from scarce topographic data and unknown scheduling rules of dense hydraulic engineering, capturing the hydrodynamic impacts of spatially distributed anthropogenic factors and climate change on the streamflow at the outlet of the MMR. Thereby, the cascade of data-driven models provides a useful numerical model for quantifying the influences of climate change and anthropogenic disturbance.

5.2. Influence Mechanisms of Climate Change and Anthropogenic Disturbance

Our study indicates that the influence of anthropogenic disturbance accounts for 62.3% of the variations in daily streamflow at the outlet of the MMR before the impoundment of Zipingpu Reservoir, and its influence becomes even more pronounced at the annual and monthly scales in the post-impoundment period. Since previous studies generally focus on attribution analysis of streamflow variations in the upper and lower reaches of the Minjiang River Basin, our study complements existing knowledge by revealing the influence mechanisms of climate change and anthropogenic disturbance on the streamflow variations within the MMR [18,19,22,29]. Overall, the influence of anthropogenic disturbance contributes to around 60% of the streamflow variations in the upper, middle and lower reaches of the Minjiang River Basin, before the impoundment of Zipingpu Reservoir. Meanwhile, our results suggest that the anthropogenic disturbance pronouncedly reduces downstream discharge during the flood period (June to September), which is consistent with the prevailing view that the development of infrastructure contributes to effective flood mitigation [12,23,26].

Our study further reveals detailed influence mechanisms of anthropogenic disturbance, based on effective reproducing of human-induced streamflow variations at a fine time scale. At the annual scale, the anthropogenic decrease the streamflow at Pengshan station by 167.00–285.15 m^3/s (33.58–64.41%) during post-impoundment period and by 39.94–98.61 m^3/s (8.09–16.65%) during pre-impoundment period, respectively. As shown in Figure 10, the magnitude of this influence has remained generally steady across 13 years. The operation of Zipingpu Reservoir could partly explain this steady decrease in mean annual discharge at Pengshan station, given the influence of urbanization and other human activities in the Chengdu metropolitan area [12,23,26].

As shown in Figure 10, according to the results of monthly attribution analysis, the human-induced decrease and climate-induced increase in streamflow at Pengshan station occur simultaneously during the flood period, reflecting anthropogenic modification of flood flows. Meanwhile, anthropogenic disturbance reduces streamflow at Pengshan station by 75.43–265.30 m^3/s (23.64–45.60%) from October to November, a reduction that is more significant after the impoundment of Zipingpu Reservoir. This intensification can be partly attributed to the post-flood impoundment of Zipingpu Reservoir. However, although Zipingpu Reservoir supplies water to downstream users, monthly streamflow at Pengshan station increases only slightly from December to the following January by

38.79–51.92 m^3/s (26.79–49.47%) and decreases by 172.29–285.27 m^3/s (53.30–60.25%) from March to May. This decrease may suggest that water demand in the Chengdu metropolitan area exceeds the supply capacity of Zipingpu Reservoir from March to May.

In summary, our study shows that the anthropogenic disturbance reduces downstream streamflow through similar mechanisms during the pre- and post-impoundment periods, whereas the impoundment and operation of Zipungpu Reservoir merely modulate the magnitude of streamflow variations. However, as shown in Figure 10 and illustrated in Table 6, climate change alters downstream streamflow through different mechanisms. While streamflow at Pengshan station exhibits decrease trend after the impoundment of Zipingpu Reservoir, climate change raises this streamflow at all three time scales. Notably, at the monthly scale, most of the climate-driven streamflow change at the Pengshan station is concentrated between June and September, increasing mean monthly discharge by 267.54–668.67 m^3/s (36.03–44.59%) in the post-impoundment period but reducing it by 105.93–314.55 m^3/s (38.21–94.79%) in the pre-impoundment period. Accordingly, our findings highlight a marked temporal unevenness in the influence of climate change on streamflow variations in the MMR, suggesting a potential temporal imbalance in water resources and an increasing threat of extreme hydrological events.

5.3. Uncertainty of Attribution Analysis

Errors in coupled hydrodynamic–hydrological simulations, the confounded effects of the Zipingpu Reservoir and the Chengdu metropolitan area and the short length of the study period contribute to uncertainties in identifying the influencing mechanisms.

In this study, outputs of the GBEHM were used to reproduce the streamflow without the influence of Zipingpu Reservoir, resulting in error propagation from the GBEHM into the cascade of data-driven models. Furthermore, the PaHL model adopts static key parameters instead of dynamic parameters, lowering simulation accuracy. These shortcomings may lead to unsatisfactory model performance during the validation and test periods in the post-reservoir period.

While variations in impact of anthropogenic disturbance on downstream runoff could be partly explained by the operation of the Zipingpu Reservoir, the available observed downstream discharge data are scarce. As a result, the influences of the reservoir and the Chengdu metropolitan area are both reflected in discharge variations at the same station, leaving their combined effects confounded and in need of further disentanglement.

To investigate the mechanisms driving streamflow variations in recent decades, we focused our analysis on the period 1980–2019. Data availability prevented a contribution analysis for the period after 2020 where extreme hydrological events have become increasingly frequent. Addressing these recent changes thus requires further investigation [1–5].

While the climatic and anthropogenic impacts on streamflow vary across timescales, we simplified this complexity using intra-annual and monthly mean values. Given the simulation uncertainties and the scarcity of observational data, the estimated contributions of climatic and anthropogenic factors should be regarded as a preliminary attribution, rather than as definitive conclusions; nonetheless, our analysis provides an initial quantification of the relative importance of these confounded drivers. Further work is required to resolve the range of their relative contributions at finer timescales.

6. Conclusions

In this study, the GBEHM were coupled with a cascade of data-driven models to reproduce influence of climate change and anthropogenic disturbance on streamflow variations at the outlet cross section of the MMR. We conducted trend and mutation test of the streamflow variations within the MMR from 1980 to 2019, and attribution analysis of the streamflow variations was conducted by scenario experiments. The major conclusions were as follows.

(1) Before the impoundment of Zipingpu Reservoir, mean annual discharge at the outlet of the MMR exhibits a significant decreasing trend, whereas it showed an increasing trend during the post-impoundment period. However, no statistically significant mutation was detected during the study period.

(2) The coupled numerical models provide effective reproduction of daily streamflow at Pengshan station during 35 years, with values of *NSE*, R^2 and *KGE* ranging from 0.96 to 1.00.

(3) During the impoundment period, mean daily discharge at Pengshan station decreases by 105.3 m^3/s, while climate change and anthropogenic disturbance explain 37.7% (-39.7 m^3/s) and 62.3% (-65.6 m^3/s) of this decrease, respectively. After the impoundment of Zipingpu Reservoir, mean daily discharge at Pengshan station decreases by 45.6 m^3/s, while climate change and anthropogenic disturbance account for -330.8% (150.8 m^3/s) and 430.8% (196.3 m^3/s) of this change, respectively.

(4) Before and after the impoundment of Zipingpu Reservoir, anthropogenic disturbance generally reduces mean annual and monthly streamflow at Pengshan station. Climate change alters downstream streamflow through different mechanisms, while most of the climate-driven streamflow change at Pengshan station is concentrated from June to September. This finding suggests a potential temporal imbalance in water resources and an increasing threat of extreme hydrological events.

Author Contributions: M.C.: Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing—Original draft, Writing—Review and Editing, Project administration, Funding acquisition. C.L.: Writing—Original draft. T.W.: Methodology, Writing—Review and Editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the project of the coupled hydrological–hydrodynamic simulation platform for river basins in the Chengdu–Chongqing region (Grant No. KY2024-CQ-02-01) and the National Natural Science Foundation of China (Grant No. U2340208).

Data Availability Statement: The data supporting the reported results are legally protected and cannot be shared by the authors. Access to these data is restricted by applicable regulations.

Acknowledgments: We gratefully acknowledge the financial support provided by the project of the coupled hydrological–hydrodynamic simulation platform for river basins in the Chengdu–Chongqing region and the National Natural Science Foundation of China.

Conflicts of Interest: Author Minhao Chen was employed by the PowerChina Huadong Engineering Corporation Limited and PowerChina Chongqing Engineering Corporation Limited. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Appendix A

$$\frac{\partial A}{\partial Z}\frac{\partial Z}{\partial t} + \frac{\partial Q}{\partial x} = q_L \tag{A1}$$

In Equation (A1), we made a central difference for Q and a forward difference for Z:

$$\frac{\partial A}{\partial Z}\frac{\Delta Z}{\Delta t} + \frac{(Q_{down} - Q_{up})}{\Delta x} = q_L \tag{A2}$$

where Δx indicates the length of the river channel; Q_{down} and Q_{up} indicate discharge at the downstream and upstream cross sections of the river channel, respectively.

Further, we approximated water level of the center cross section of the river channel, which is called the characteristic cross section of the river channel, by a linear combination of water level at the downstream and upstream cross section of the river channel, thus:

$$\frac{\partial A_c}{\partial Z_c(Z_{up}, Z_{down})}\frac{\Delta Z_c(Z_{up}, Z_{down})}{\Delta t} + \frac{(Q_{down} - Q_{up})}{\Delta x} = q_L \tag{A3}$$

where Z_c, Z_{down} and Z_{up} indicate the water level at the characteristic cross section, downstream cross section and upstream cross section, respectively; A_c indicates the area of characteristic cross section; Z_c is approximated as a linear function of Z_{up} and Z_{down}.

Cross sections in channels may highly complex geometric configurations. Thus, we assumed that the area of any cross section can be approximated as a polynomial function of water depth:

$$A = C_1 + t_1 Z + t_2 Z^2 + t_3 Z^3 \cdots \tag{A4}$$

where C_1 (m^2), t_1 (m), t_2 (dimensionless), t_3 (m^{-1}) are undetermined coefficients.

According to Equation (A4), the partial derivative of A_c to Z_c in Equation (A3) can be approximated as a function of Z_c, thus:

$$\Delta Z_c(Z_{up}, Z_{down}) = (q_L - \frac{(Q_{down} - Q_{up})}{\Delta x})\Delta t / (t_1 + 2t_2 Z_c + 3t_3 Z_c^2 + \cdots) \tag{A5}$$

According to Equation (A5), temporal variations in water level at any time step can be approximated by a function of water level and spatial variables of discharge. We adopted the variational approach to obtain the best form of the partial derivative of A_c to Z_c, reducing the approximation error. The variational approach was utilized to evaluate the relationship between the form of the partial derivative of A_c to Z_c and estimation error of Z_c, then:

$$I = \int_{t'}^{t} f(Z_c, Z_c{}')dt = \int_{t'}^{t} (q_L - \frac{(Q_{down} - Q_{up})}{\Delta x}) / (t_1 + 2t_2 Z_c + 3t_3 Z_c^2 + \cdots) - \frac{\partial Z_{c-obs}}{\partial t} dt \tag{A6}$$

where t' indicates the initial time; $Z_{c\text{-}obs}$ indicates the observed Z_c. There is no term related to $Z_c{}'$, thus when we obtained the minimum of I in Equation (A6) at any time step, the Euler–Lagrange equation as following should hold at any time step:

$$\frac{\partial f}{\partial Z_c} - \frac{d}{dt}(\frac{\partial f}{\partial Z_c{}'}) = 0 \tag{A7}$$

There is no term related to $Z_c{}'$ in Equation (A6), accordingly:

$$-(q_L - \frac{(Q_{down} - Q_{up})}{\Delta x}) / (t_1 + 2t_2 Z_c + 3t_3 Z_c^2 + \cdots)^2 (2t_2 + 6t_3 Z_c + \cdots) = 0 \tag{A8}$$

It is obvious that Equation (A8) holds identically for any time step when coefficients t_2, $t_3 \cdots$ equal 0. Therefore, the best form of Equation (A7) is as follows:

$$\Delta Z_c(Z_{up}, Z_{down}) = (q_L - \frac{(Q_{down} - Q_{up})}{\Delta x})\Delta t / t_1 \tag{A9}$$

Here, we assumed a linear relationship between Z_c and Z_{up}, Z_{down} has a form as follows:

$$Z_c = a(Z_{up} - Z_{down}) + b(Z_{up} + Z_{down}) \tag{A10}$$

We conducted a recursive process to calculate Z_c as follows:

$$Z_c^{t+1} = Z_c^t + ({q_L}^t - \frac{({Q_{down}}^t - {Q_{up}}^t)}{\Delta x})\Delta t/t_1 \tag{A11}$$

Appendix B

In the PaHL model, the GA was mainly employed to obtain appropriate estimation of coefficients a, b and t_1. The GA is a heuristic optimization method inspired by natural selection, operating through iterative evolution of a population of candidate solutions. Each individual is encoded as a chromosome representing the decision variables, and evolves via selection, crossover, and mutation operators to maximize a predefined fitness function. The GA have three operators as following: the selection, the crossover, and the mutation.

The selection operator applies stochastic uniform selection where the probability of an individual being selected is proportional to its fitness rank:

$$P_i = \frac{r_i}{\sum_{j=1}^{N} r_j} \tag{A12}$$

where r_i is the rank of the ith individual, N is the population size, respectively.

The crossover operator used scattered crossover. For two parent chromosomes P_1 and P_2, an offspring o is generated as follows:

$$o_k = \begin{cases} p_{1k}, if\ r_k \le 0.5 \\ p_{2k}, otherwise \end{cases} \tag{A13}$$

where $r_k \sim$U(0, 1) is a random number drawn independently for each gene k. The crossover fraction was set to 0.8.

The mutation operator adopts the Gaussian mutation to add a normally distributed perturbation to each gene:

$${x'}_k = x_k + \sigma \cdot N(0,1) \tag{A14}$$

where σ is the mutation scale, set to 1.0 initially and shrunk linearly with generations: $\sigma_g = \sigma_0 \cdot \left(1 - \frac{g}{G_{\max}}\right)$, with g indicating the current generation and $G_{\max} = 100$.

Meanwhile, two individuals ($E = 2$) with the highest fitness were guaranteed to survive to the next generation unchanged. For the calibration length LL, the value was rounded to the nearest integer after crossover and mutation: $LL' = round(LL)$. The fitness function was specifically designed to balance calibration accuracy and generalization capability:

$$\begin{aligned} F = {} & \alpha \cdot (QNSE_{cal} + ZNSE_{cal} + {QR^2}_{cal} + {ZR^2}_{cal}) + \\ & \beta (QNSE_{val} + ZNSE_{val} + {QR^2}_{val} + {ZR^2}_{val}) - \lambda (a + b) \end{aligned} \tag{A15}$$

where $QNSE$ and $ZNSE$ denote the Nash–Sutcliffe efficiency for discharge and water level, respectively; QR^2 and ZR^2 are the corresponding coefficients of determination, while subscripts $_{cal}$ and $_{val}$ refer to the calibration and validation periods. Weights were set as $\alpha = 0.6$ and $\beta = 0.4$, and the regularization coefficient $\lambda = 0.1$ was applied to penalize extreme parameter values. The GA was configured with a population size of 50, scattered crossover, Gaussian mutation, and a maximum of 100 generations.

References

1. Al-Rawas, G.; Nikoo, M.R.; Al-Wardy, M. A review on the prevention and control of flash flood hazards on a global scale: Early warning systems, vulnerability assessment, environmental, and public health burden. *Int. J. Disaster Risk Reduct.* **2024**, *115*, 105024. [CrossRef]
2. Bjerke, J.W.; López-Blanco, E.; Tømmervik, H.; Striberny, A.; Davids, C.; Ólafsdóttir, R.; Karlsen, S.R.; Sandström, P.; Turunen, M.; Rikkonen, T.; et al. Nordic boreo-arctic lands under rapid climatic change: A review of recent and future trends and extreme events. *Earth-Sci. Rev.* **2025**, *261*, 105012.
3. Chatzistratis, D.; Velegrakis, A.F.; Chalazas, T.; Alves, B.; Schiavon, E.; Monioudi, I.; Armaroli, C. European Early Warning Systems for coastal floods: User needs, system availability and pertinent policy and legislation. *Int. J. Disaster Risk Reduct.* **2025**, *126*, 105602. [CrossRef]
4. Yan, J.; Zheng, T.; Zhao, J.; Luo, J. Bridging rarity of drought-flood abrupt alternation events with reverse distributions or reverse copulas. *Water Res.* **2025**, *287*, 124433. [CrossRef] [PubMed]
5. New, M.; Hulme, M.; Jones, P. Representing twentieth-century space–time climate variability. Part II: Development of 1901–96 monthly grids of terrestrial surface climate. *J. Clim.* **2000**, *13*, 2217–2238.
6. Grill, G.; Lehner, B.; Thieme, M.; Geenen, B.; Tickner, D.; Antonelli, F.; Babu, S.; Borrelli, P.; Cheng, L.; Crochetiere, H.; et al. Mapping the world's free-flowing rivers. *Nature* **2019**, *569*, 215–221. [CrossRef] [PubMed]
7. Huang, S.; Gan, Y.; Chen, N.; Wang, C.; Zhang, X.; Li, C.; Horton, D.E. Urbanization enhances channel and surface runoff: A quantitative analysis using both physical and empirical models over the Yangtze River basin. *J. Hydrol.* **2024**, *635*, 131194. [CrossRef]
8. Maavara, T.; Chen, Q.; Van Meter, K.; Brown, L.E.; Zhang, J.; Ni, J.; Zarfl, C. River dam impacts on biogeochemical cycling. *Nat. Rev. Earth Environ.* **2020**, *1*, 103–116. [CrossRef]
9. Zhou, L.; Liu, L. Enhancing dynamic flood risk assessment and zoning using a coupled hydrological-hydrodynamic model and spatiotemporal information weighting method. *J. Environ. Manag.* **2024**, *366*, 121831. [CrossRef]
10. Feng, S.; Yamamoto, T. Preliminary research on sponge city concept for urban flood reduction: A case study on ten sponge city pilot projects in Shanghai, China. *Disaster Prev. Manag. Int. J.* **2020**, *29*, 961–985. [CrossRef]
11. Dui, H.; Zhang, H.; Wu, S. A IoT-based novel methodology to optimize multidimensional flood resilience of drainage systems for sponge city. *Sustain. Cities Soc.* **2025**, *130*, 106523.
12. Li, H.; Yang, D.; Zhu, Z.; Wei, Y.; Zhou, Y.; Ishidaira, H.; Commey, N.A.; Cheng, H. Flood risk analysis of urban agglomerations in the yangtze river basin under extreme precipitation based on remote sensing technology. *Remote Sens.* **2024**, *16*, 4289. [CrossRef]
13. Liao, N.; Zhang, L.; Chen, M.; Li, J.; Wang, H. The influence mechanism of water level operation on algal blooms in canyon reservoirs and bloom prevention. *Sci. Total Environ.* **2024**, *912*, 169377. [PubMed]
14. Tang, W. Research on Optimal Scheduling of Zipingpu Reservoir Based on Downstream Ecological Environment Demand. Master's Thesis, Sichuan University, Chengdu, China, 2023. (In Chinese) [CrossRef]
15. Yang, L.; You, L.H. Multi-objective dispatching operation practice of Zipingpu Reservoir during flood season. *Sichuan Hydro Power* **2018**, *37*, 24–26, 30. (In Chinese)
16. Shi, R.J.; Wang, T.H.; Yang, D.W.; Yang, Y.T. Streamflow decline threatens water security in the upper Yangtze river. *J. Hydrol.* **2022**, *606*, 127448. [CrossRef]
17. Chen, M.; Li, C.; Hu, Z.; Wang, T.; Tang, L.; Yang, D. Shifting water and energy controls on hydrological processes under climate change in the topographically complex Min–Tuo River Basin, southwest China. *J. Hydrol. Reg. Stud.* **2026**, *65*, 103352. [CrossRef]
18. Hou, J.; Ye, A.; You, J.; Ma, F.; Duan, Q. An estimate of human and natural contributions to changes in water resources in the upper reaches of the Minjiang River. *Sci. Total Environ.* **2018**, *635*, 901–912. [CrossRef] [PubMed]
19. Zhang, Y.; Ye, A.; You, J.; Jing, X. Quantification of human and climate contributions to multi-dimensional hydrological alterations: A case study in the Upper Minjiang River, China. *J. Geogr. Sci.* **2021**, *31*, 1102–1122. [CrossRef]
20. Liu, L.; Ao, T.; Zhou, L.; Takeuchi, K.; Gusyev, M.; Zhang, X.; Wang, W.; Ren, Y. Comprehensive evaluation of parameter importance and optimization based on the integrated sensitivity analysis system: A case study of the BTOP model in the upper Min River Basin, China. *J. Hydrol.* **2022**, *610*, 127819. [CrossRef]
21. Liu, Y.; Jiang, B. Research on the Development of Medium-sized Cities from the Perspective of a Metropolitan Area—A Case Study of Ziyang City in the Chengdu Metropolitan Area. *Int. J.* **2022**, *2*, 1–10.
22. Geng, K.; Chen, X.; Zheng, M.; Gao, Y.; Gu, Z.; Yao, H. The influence of human activities on rainfall-runoff relationships at different time scales in the Minjiang River Basin. *Theor. Appl. Climatol.* **2024**, *155*, 8435–8454. [CrossRef]
23. Wei, Y.; Li, H.; Zhou, Y.; Commey, N.A.; Zhu, Z.; Ishidaira, H.; Jiang, Z.; Yang, D. Flood vulnerability assessment of urban agglomeration under the background of urban expansion. *J. Hydrol.* **2025**, *662*, 133962. [CrossRef]
24. Yang, L.; Zeng, S.; Xia, J.; Wang, Y.; Huang, R.; Chen, M. Effects of the Three Gorges Dam on the downstream streamflow based on a large-scale hydrological and hydrodynamics coupled model. *J. Hydrol. Reg. Stud.* **2022**, *40*, 101039. [CrossRef]

25. Zhao, W.; Zhao, Y.; Zhao, Q.; Wang, X.; Su, T.; Guo, Y. Simulating Rainfall for Flood Forecasting in the Upper Minjiang River. *Water* **2025**, *18*, 4. [CrossRef]
26. Li, Y.; Ma, L.; Huang, J.; Disse, M.; Zhan, W.; Li, L.; Zhang, T.; Sun, H.; Tian, Y. Machine learning parallel system for integrated process-model calibration and accuracy enhancement in sewer-river system. *Environ. Sci. Ecotechnol.* **2024**, *18*, 100320. [PubMed]
27. Xu, Y.; Yang, Y.; Wang, Z.; Xiong, J.; Yong, Z.; Zhang, X.; Liu, J.; Zhu, A.; Li, D. Impact of different rainfall patterns on flood risk on the urban-rural gradient. *J. Environ. Manag.* **2025**, *394*, 126583. [CrossRef]
28. Fang, S.; Chen, M.; Li, J.; Zhang, L.; Zhang, Y.; Yang, C.; Wang, H. Evaluation and optimization of urban hydrological connectivity based on graph theory: A case study in Chengdu, China. *Ecol. Inform.* **2024**, *82*, 102749. [CrossRef]
29. Zhai, H.; Wang, M.; Shen, D.; Hu, B.; Li, Y. Analysis of runoff variation and driving factors in the Minjiang River Basin over the last 60 years. *J. Water Clim. Change* **2022**, *13*, 3675–3691. [CrossRef]
30. Shen, Y.; Xiong, A.Y. Validation and comparison of a new gauge-based precipitation analysis over mainland China. *Int. J. Climatol.* **2016**, *36*, 252–265.
31. Jarvis, A.; Reuter, H.I.; Nelson, A.; Guevara, E. Hole-Filled SRTM for the Globe Version 4, Available from the CGIAR-CSI SRTM 90m Database. 2008. Available online: https://srtm.csi.cgiar.org/ (accessed on 20 May 2026).
32. Cao, S.; Li, M.; Zhu, Z.; Wang, Z.; Zha, J.; Zhao, W.; Duanmu, Z.; Chen, J.; Zheng, Y.; Chen, Y.; et al. Spatiotemporally consistent global dataset of the GIMMS leaf area index (GIMMS LAI4g) from 1982 to 2020. *Earth Syst. Sci. Data* **2023**, *15*, 4877–4899. [CrossRef]
33. FAO/IIASA/ISRIC/ISS-CAS/JRC. *Harmonized World Soil Database*, version 1.2; FAO: Rome, Italy; IIASA: Laxenburg, Austria, 2012.
34. Dai, Y.J.; Shangguan, W.; Duan, Q.Y.; Liu, B.Y.; Fu, S.H.; Niu, G. Development of a China Dataset of Soil Hydraulic Parameters Using Pedotransfer Functions for Land Surface Modeling. *J. Hydrometeorol.* **2013**, *14*, 869–887. [CrossRef]
35. Mann, H.B. Nonparametric tests against trend. *Econom. J. Econom. Soc.* **1945**, *13*, 245–259. [CrossRef]
36. Kendall, M.G. *Rank Correlation Methods*, 4th ed.; Charles Griffin: London, UK, 1975.
37. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. *arXiv* **2016**. [CrossRef]
38. Yang, D.; Herath, S.; Musiake, K. Development of a geomorphology-based hydrological model for large catchments. *Proc. Hydraul. Eng.* **1998**, *42*, 169–174. [CrossRef]
39. Yang, D.W.; Gao, B.; Jiao, Y.; Lei, H.M.; Zhang, Y.L.; Yang, H.B.; Cong, Z.T. A distributed scheme developed for eco-hydrological modeling in the upper Heihe River. *Sci. China-Earth Sci.* **2015**, *58*, 36–45.
40. Gao, B.; Yang, D.W.; Qin, Y.; Wang, Y.H.; Li, H.Y.; Zhang, Y.L.; Zhang, T.J. Change in frozen soils and its effect on regional hydrology, upper Heihe basin, northeastern Qinghai-Tibetan Plateau. *Cryosphere* **2018**, *12*, 657–673.
41. Flerchinger, G.; Saxton, K. Simultaneous heat and water model of a freezing snow-residue-soil system I. Theory and development. *Trans. ASAE* **1989**, *32*, 565–571.
42. Qin, Y.; Yang, D.W.; Gao, B.; Wang, T.H.; Chen, J.S.; Chen, Y.; Wang, Y.H.; Zheng, G.H. Impacts of climate warming on the frozen ground and eco-hydrology in the Yellow River source region, China. *Sci. Total Environ.* **2017**, *605*, 830–841. [CrossRef] [PubMed]
43. Yang, D.; Herath, S.; Musiake, K. A hillslope-based hydrological model using catchment area and width functions. *Hydrol. Sci. J.* **2002**, *47*, 49–65. [CrossRef]
44. Van Genuchten, M.T. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci. Soc. Am. J.* **1980**, *44*, 892–898.
45. Hansson, K.; Simunek, J.; Mizoguchi, M.; Lundin, L.C.; van Genuchten, M.T. Water flow and heat transport in frozen soil: Numerical solution and freeze-thaw applications. *Vadose Zone J.* **2004**, *3*, 693–704. [CrossRef]
46. Farquhar, G.D.; Caemmerer, S.V.; Berry, J.A. A biochemical model of photosynthetic CO_2 assimilation in leaves of C3 species. *Planta* **1980**, *149*, 78–90. [CrossRef] [PubMed]
47. De la Fuente, L.A.; Ehsani, M.R.; Gupta, H.V.; Condon, L.E. Towards interpretable LSTM-based modelling of hydrological systems. *Hydrol. Earth Syst. Sci. Discuss.* **2023**, *2023*, 1–36. [CrossRef]
48. Feng, Z.-K.; Zhang, J.-S.; Niu, W.-J. A state-of-the-art review of long short-term memory models with applications in hydrology and water resources. *Appl. Soft Comput.* **2024**, *167*, 112352.
49. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.