Dear Editor,

We sincerely thank you for your thorough review and insightful comments on our manuscript. In response, we have carefully revised the manuscript and addressed all comments in detail. A point-by-point response is provided below, indicating the changes made in the revised manuscript.

1.Comment:

Section Abstract: Backflow should be well addressed when it first occurs, for example, from the lake to river? The river–lake connection should be briefly introduced.

Response:

Thank you for this helpful suggestion. We agree that the definition of backflow and the river–lake connection should be clarified when first mentioned. In the revised abstract, we now explicitly define backflow as the intrusion of Yangtze River discharge into Poyang Lake, thereby clarifying its direction and physical meaning. In addition, the hydraulic connection between the Yangtze River and Poyang Lake at Hukou has been briefly emphasized to provide essential background for understanding backflow processes. The revised sentence in the Abstract now reads as follows:

Revised part is now included on line 18-21: Results indicate that backflow events, characterized by the intrusion of Yangtze River discharge into Poyang Lake, declined significantly in frequency, duration, and discharge, with dam regulation accounting for about 70% of the variation in backflow-event discharge.

2.Comment:

Section Abstract: The model seems to be quantitatively to examine the contribution of the upstream dam on the lake hydrology, however, the author's main results come off as rather trivial. Is presenting model contribution rates to two decimal places really necessary?

Response:

Thank you for this insightful comment. We agree that presenting contribution rates with excessive numerical precision may give an impression of triviality without adding substantive scientific value. In response, we have revised the Abstract to reduce

unnecessary numerical precision by reporting contribution rates using approximate values (e.g., "approximately 70%" and "about 30%"), and place greater emphasis on the hydrological mechanisms and system-level transitions induced by dam regulation, rather than on numerical magnitudes alone.

Specifically, the revised Abstract now highlights the cascading hydrological responses associated with changes in backflow, normalflow, flood, and drought events, as well as the shift in hydrological dominance at Hukou from mainstream water-level control to lake-outflow control during drought conditions. These revisions aim to clarify that the quantitative analysis is used to support process understanding, rather than to present isolated contribution percentages. The Abstract has been revised accordingly.

Revised part is now included on line 13-32: Taking the Three Gorges Project–Poyang Lake system, where the lake is hydraulically connected to the Yangtze River at Hukou, as a representative case, this study integrates the Light Gradient Boosting Machine (LightGBM) model with the SWAT model to analyze lake hydrological responses to dam regulation. Based on long-term runoff, water level, and meteorological records, simulated and observed hydrological events were compared to quantify the influence of the Three Gorges Project on Poyang Lake. Results indicate that backflow events, characterized by the intrusion of Yangtze River discharge into Poyang Lake, declined significantly in frequency, duration, and discharge, with dam regulation accounting for approximately 70% of the variation in backflow event discharge. The backwater effect at Hukou weakened substantially, inducing cascading hydrological responses across the lake. Normalflow events were notably prolonged, whereas flood events decreased in both frequency and magnitude, indicating a significant flood peak attenuation effect. In contrast, water levels and discharge during normalflow conditions declined, accompanied by an increased occurrence of drought events, for which dam regulation explained about 30% of the runoff variation. During drought events, discharge deviated substantially from model expectations, suggesting enhanced hydraulic resistance at Hukou and a shift from mainstream water-level dominance to lake-outflow dominance. Overall, the Three Gorges Dam has stabilized hydrological variability in the Poyang

Lake region while simultaneously increasing drought risk, reflecting a transition from a naturally regulated lake system toward a semi-natural, dam-regulated system. Future management should optimize dam operation strategies and strengthen river–lake connectivity to support sustainable water-resource and ecosystem management.

3.Comment:

Section Introduction: The authors stated that "These processes are difficult to accurately capture using traditional one-dimensional or quasi-two-dimensional models(Wu et al., 2022)." I am profoundly skeptical of this viewpoint. Moreover, the cited literature fails to substantiate the claim.

Response:

Thank you for this thoughtful comment. We agree that the original wording was too general and may have led to ambiguity regarding the applicability of different modeling approaches. One-dimensional and quasi-two-dimensional hydrodynamic models are well-established and effective tools for simulating specific hydrodynamic processes in river–lake systems, such as flood propagation or backwater effects, under prescribed boundary conditions(Changxin et al., 2015; Li et al., 2017).

The objective of this study is not event-scale process simulation, but long-term attribution analysis based on multi-decadal hydrological observations. The key methodological challenge lies in disentangling the relative contributions of different driving factors—particularly large-scale engineering regulation—through the construction of a counterfactual baseline. This represents a fundamentally data-driven problem. Traditional hydrodynamic models are primarily designed for state simulation rather than for learning patterns from large observational datasets or directly supporting quantitative attribution over long time scales, and their application in this context is therefore limited.

Accordingly, we employ a hybrid framework combining SWAT and LightGBM. SWAT provides a physically based representation of natural runoff conditions, while LightGBM captures the observed river–lake response relationships from data. This complementary framework is specifically tailored to long-term attribution analysis and

enables a more direct and efficient separation of driving factors. The relevant statements in the manuscript have been revised to more clearly reflect these distinctions.

Revised part is now included on line 77-80: Traditional one-dimensional and quasi-two-dimensional hydrodynamic models have been widely and successfully applied to simulate hydrological processes under prescribed boundary conditions, particularly for event-scale analyses such as flood propagation and backwater effects (Changxin et al., 2015; Li et al., 2017)

4.Comment:

Section Introduction: paragraph 70-75, this part aims to highlight the importance of combined SWAT model and LightGBM model for the present work, however, this part should be moved elsewhere.

Response:

Thank you for this helpful suggestion. We agree that the original paragraph placed excessive emphasis on the modeling framework at an early stage of the Introduction, which could interrupt the logical development from problem background to research motivation.

In the revised manuscript, we have substantially restructured this part. The Introduction has been refocused on the scientific challenges associated with river–lake interaction analysis and the limitations of existing modeling approaches for long-term, attribution-oriented studies. Detailed explanations of the hybrid SWAT–LightGBM framework and its complementary roles have been streamlined and repositioned to later sections where methodological details are more appropriate. The revised paragraph now introduces the hybrid framework only briefly as a necessary methodological response to this challenge, without elaborating on implementation details.

Revised part is now included on line 76-89:

In this context, river–lake interaction mechanisms in connected lakes have become a major focus of recent research. Traditional one-dimensional and quasi-two-dimensional hydrodynamic models have been widely and successfully applied to simulate hydrological processes under prescribed boundary conditions, particularly for event-

scale analyses such as flood propagation and backwater effects (Changxin et al., 2015; Li et al., 2017). However, these models are primarily designed for reproducing individual events and are less suited for quantitatively disentangling the contributions of multiple driving factors from long-term observational records. This limitation is especially pronounced under conditions of persistent anthropogenic regulation, where system behavior evolves gradually and cumulative effects become increasingly important. To systematically characterize river–lake responses over long time scales and quantitatively distinguish natural variability from human regulation (e.g., the Three Gorges Project), this study develops a hybrid modeling framework that integrates physical mechanisms with data-driven approaches by coupling a watershed hydrological model (SWAT) with a Light Gradient Boosting Machine (LightGBM) model. The SWAT model is employed to simulate inflow processes under a natural (no-dam) scenario, while LightGBM learns river–lake response relationships from multi-decadal observations, enabling quantitative attribution of mainstream regulation and other anthropogenic influences.

5.Comment:

Section Introduction: paragraph 80-85, looking into the literature (e.g., backflow, Three Gorges Dam, Poyang Lake), it seems the author has missed a lot of earlier work, for example,

Zhang, Q., Li, L., Wang, Y. G., Werner, A. D., Xin, P., Jiang, T., & Barry, D. A. (2012). Has the Three-Gorges Dam made the Poyang Lake wetlands wetter and drier?. Geophysical research letters, 39(20).

Li, Y., Zhang, Q., Werner, A. D., Yao, J., & Ye, X. (2017). The influence of river-to-lake backflow on the hydrodynamics of a large floodplain lake system (Poyang Lake, China). Hydrological Processes, 31(1), 117-132.

Response:

We thank the reviewer for pointing out this important issue. We agree that the initial version of the manuscript did not sufficiently acknowledge several key early studies on the hydrological impacts of the Three Gorges Project and river–lake

interactions in the Poyang Lake–Yangtze River system.

In the revised Introduction, we have expanded and strengthened the literature review by incorporating seminal and representative studies, including Zhang et al. (2012) and Li et al. (2017), as well as other relevant work addressing backflow processes, hydrodynamic responses, and long-term hydrological changes in Poyang Lake following the impoundment of the Three Gorges Project (e.g., Ye et al., 2013; Yuan et al., 2021). These studies are now explicitly cited when describing the alteration of outlet hydrodynamics, lake water-level regimes, and seasonal hydrological patterns under mainstream regulation.

Revised part is now included on line 60-70:

Against this backdrop, the Poyang Lake–Yangtze River system serves as a representative case for investigating the hydrological impacts of human regulation on connected lakes, owing to its unique river–lake connectivity and high sensitivity to mainstream water-level variations. Since impoundment began in 2003, the Three Gorges Project (TGP), the world's largest hydraulic infrastructure, has substantially altered the hydrological regime of the middle and lower Yangtze River through its "peak-shaving and valley-filling" operation strategy, exerting profound influences on outlet hydrodynamics and lake outflow processes (Ye et al., 2013; Yuan et al., 2021; Zhang et al., 2012). Previous studies have shown that following TGP impoundment, Poyang Lake has experienced persistently lower water levels, an earlier onset of the dry season, and a contraction of lake surface area. These changes reflect a systematic alteration of the lake's hydrological rhythm and water availability, posing new challenges to ecosystem stability(Guo et al., 2012; Tan et al., 2024; Tian et al., 2023, 2025).

6.Comment: Section 2.1, Poyang Lake, located on the southern bank of the middle–lower Yangtze River. Perhaps the location is wrong. It locates in the middle steam of the River? Please check it.

Response:

Thank you for this helpful comment. We agree that the expression "southern bank"

may lead to ambiguity in an English hydrological context. To avoid any potential misunderstanding regarding the geographical location of Poyang Lake, we have revised the sentence by removing the term "bank" and using a clearer regional description.

Revised part is now included on line 109-110:

Poyang Lake, located in the middle–lower reaches of the Yangtze River, is the largest freshwater lake in China (Feng et al., 2012; Wang et al., 2009).

7.Comment: Section 2.1, the authors used two hydrological stations Jiujiang and Hukou (Fig.1), however, these two stations are so close to each other. What are their respective purposes, then? I'm not sure if the author has taken into account the possible joint impact from the lake and the upper Yangtze River on the station hydrology.

Response:

Thank you for this valuable comment. We agree that the original description of the hydrological stations lacked sufficient detail, which may have caused some misunderstanding regarding the rationale for selecting the Hukou and Jiujiang stations.

In the revised manuscript, we have clarified that Hukou Station was deliberately chosen because it is situated at the sole outlet of Poyang Lake and is subject to the combined influence of lake outflow and the Yangtze River mainstream. Rather than being a limitation, this dual control is a key advantage: it enables Hukou to directly capture critical river–lake interactions, including variations in outflow and potential backflow events.

We have also expanded our justification for selecting Jiujiang Station, which is located approximately 19 km upstream of the lake–river confluence. Given the wide channel and complex bed morphology of the Yangtze River in this reach, water levels over longer distances exhibit weak hydrodynamic connectivity. Consequently, Jiujiang provides the most representative and hydrologically relevant upstream reference point while remaining closely tied to the dynamics of the confluence zone.

Revised part is now included on line 118-136:

As the hydrological junction of the Yangtze–Poyang Lake system, the Hukou Hydrological Station, situated at the outlet of Poyang Lake, directly captures the

dynamic exchange between the lake and the Yangtze River. At this location, hydrological conditions are jointly regulated by outflow from the lake and backwater effects from the river, enabling the station to record both lake-derived discharge and river-induced intrusion or flow reversal. For this reason, Hukou Station was selected as the primary site for characterizing river–lake interaction processes in this study.

To represent upstream hydrological conditions of the Yangtze River, water level data from a Yangtze River hydrological station were additionally employed Given the wide channel, complex bed morphology, and spatially variable bed elevations of the Yangtze River, water levels measured at stations located far from the lake–river junction are strongly influenced by local channel geometry and therefore cannot reliably reflect the backwater effects exerted by the Yangtze River at Hukou. For this reason, the Jiujiang Hydrological Station was selected. Located approximately 19 km upstream of the lake–river confluence, it remains unaffected by direct lake outflow while capturing hydrodynamic conditions spatially comparable to those at Hukou, thereby providing an appropriate reference for upstream forcing from the Yangtze River.

This study employed daily discharge and water-level records from Hukou Station (spanning the pre-TGP period 1970–2002 and the post-TGP period 2003–2020) along with water-level data from Jiujiang Station to analyze the long-term evolution of four key hydrological event types in Poyang Lake: floods, normal-flow periods, droughts, and backflow events.

8.Comment: Section 2.2.2, the authors only used 1970 as the reference year, the reason should be well explained and a one-year timeframe lacks representativeness. In addition, the MK test method is very simple, but it still requires citation of the original literature. Response:

We sincerely appreciate the reviewer's valuable comment. We agree that adopting 1970 as a single reference year may lack representativeness. Therefore, the cumulative anomaly analysis in Section 2.2.2 has been revised using the long-term mean (1970–2020) as the baseline, allowing for a more robust characterization of long-term variations. Corresponding text, figures, and calculations have been updated.

Furthermore, citations to the original Mann–Kendall test methodology have been added, as suggested, to strengthen the methodological description..

Revised part is now included on line 156-159:

To investigate the long-term evolution of water level and discharge, the cumulative difference method was employed to derive interannual variation trends. Using the multi-year mean as the reference baseline, annual mean water level (H) and discharge (Q) were converted into interannual anomalies (ΔH and ΔQ), which were then cumulatively summed to obtain long-term variation curves.

line 163-166:

Mutation years were identified at the intersection of the UF and UB statistic curves within the confidence interval, thereby revealing the structural changes in hydrological processes before and after the TGP operation (Kendall, 1948; Mann, 1945).

line 233-237:

From 1970 to 2020, the cumulative monthly anomalies of runoff (ΔQ) and water level (ΔH) displayed evident fluctuations and a distinct transition in trends around 2003, corresponding to the initial impoundment of the Three Gorges Reservoir. Before 2003, the cumulative water level anomaly increased gradually, whereas the cumulative runoff anomaly showed weak upward fluctuations. After 2003, both cumulative anomalies exhibited a pronounced downward trend (Fig.3).
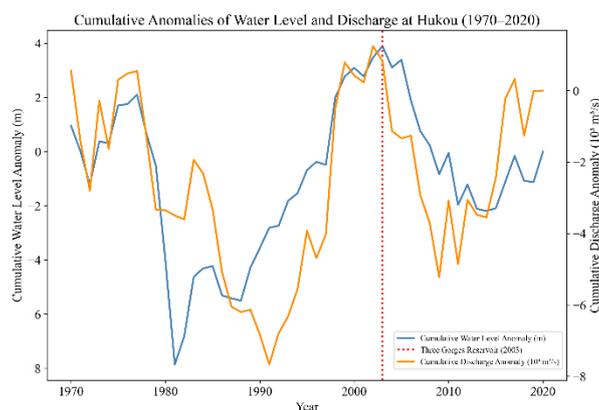

Figure 3 Trends in cumulative monthly variations of runoff and water level

9.Comment: Section 2.2.3, the authors stated that the LightGBM has the ability to

reproduce the lake–river confluences exhibiting backflow phenomena. I am uncertain how the authors' understanding of the model's nonlinear capture capability leads to the conclusion that it can accurately represent a specific hydrological phenomenon like backflow.

Response:

We thank the reviewer for the valuable comment regarding our description of the role of LightGBM in the analysis of backflow phenomena. We acknowledge that the original wording was not sufficiently precise and may have been misleading, as it could be interpreted as suggesting that LightGBM directly reproduces or simulates the physical dynamics of lake–river backflow. We would like to clarify that this is not the case.

In this study, the primary purpose of LightGBM is not to simulate or reconstruct the backflow process itself, but rather to establish a robust, data-driven relationship between the inflows from the five tributary rivers and the corresponding lake outflow under non-backflow conditions. Specifically, the LightGBM model is trained exclusively on historical periods without backflow influence, allowing it to learn the typical nonlinear response of lake outflow to tributary inflows under normal hydraulic conditions. Once trained, the model is used as a baseline reference to estimate the lake outflow that would be expected for a given inflow during backflow events, in the absence of backflow effects.

By comparing this baseline prediction with the actually observed lake outflow, we are able to quantitatively assess the deviation induced by backflow events. In addition, by combining these results with inflows simulated by the SWAT model under a no-human-intervention scenario, we further derive the corresponding natural lake outflow and compare it with observations, thereby quantifying the impact of the Three Gorges Dam regulation on lake hydrological processes.

To avoid misunderstanding and to improve clarity, we have revised Section 2.2.3 in the manuscript to more accurately describe the functional role of LightGBM and to clearly distinguish its data-driven baseline estimation from any physical representation of backflow dynamics.

Revised part is now included on line 171-216:

The Light Gradient Boosting Machine (LightGBM) is an ensemble learning framework based on Gradient Boosting Decision Trees (GBDT), which constructs a robust predictive model by iteratively training multiple weak learners and combining them through weighted aggregation. Compared with conventional GBDT implementations (e.g., XGBoost), LightGBM is specifically optimized for large-scale datasets and high-dimensional feature spaces, offering advantages in computational efficiency, memory usage, and predictive accuracy (Mienye and Sun, 2022). Runoff-related processes are often highly nonlinear. Traditionally, such relationships have been modeled using physics-based hydrological models, such as SWAT, VIC, and HEC-HMS, which typically require extensive parameter calibration and involve substantial computational costs. In contrast, the tree-based structure of LightGBM enables flexible learning of complex nonlinear relationships directly from observations, making it suitable for characterizing the empirical response of lake outflow to tributary inflows under normal hydrodynamic conditions (Bian et al., 2023; Kumar et al., 2023).

It should be emphasized that LightGBM is a purely data-driven model and does not explicitly incorporate physical constraints such as mass conservation or energy balance. To avoid interference from extreme conditions and potential overfitting, this study exclusively employed normal-flow (non-backflow) events to construct separate LightGBM models for the pre- and post-Three Gorges Project (TGP) periods.

(1) Model inputs include discharge from the five major tributaries flowing into Poyang Lake—the Ganjiang, Fuhe, Xinjiang, Raohe, and Xiushui Rivers—together with Hukou discharge and seasonal indicators such as month.

(2) During training, the model automatically learns the nonlinear relationships between tributary inflows and lake outflow, including the relative contributions of different rivers and lagged effects reflecting delayed downstream responses to upstream inflows. Monthly information is transformed using sine and cosine functions to represent intra-annual hydrological periodicity and seasonal hydrological conditions.

(3) The dataset is split chronologically into a training set (80%) and a testing set (20%) to preserve temporal integrity and prevent information leakage. A small learning rate

combined with a relatively large number of boosting iterations is adopted, along with subsampling and feature sampling strategies to enhance generalization. Early stopping based on the validation root mean square error (RMSE) is applied to mitigate overfitting. (4) Model performance is evaluated using the coefficient of determination (R^2), calculated separately for the training and testing datasets.

(5) Based on historical normal-flow events, baseline models of Hukou outflow are established for both the pre- and post-TGP periods. These trained models are then used to estimate the theoretically expected Hukou outflow under different hydrological scenarios in the absence of backflow effects, providing a reference framework for quantitatively assessing changes in lake hydrological regimes associated with TGP operation. (Figure 2).

10.Comment: Section 2.2.3, the model seems to use discharge data from the basin river inflows, but these key stations were not shown in Fig. 1.

Fig. 3 should be moved into the Section "3 Results". In addition, the unit of the metric RMSE is m3/s. Apparently, the authors may have overlooked this unit issue.

Response:

We thank the reviewer for the valuable comment, We have added hydrological station markers to Fig. 1 to explicitly indicate the locations of key stations used for discharge data collection. The revised Fig. 1 now clearly displays all critical monitoring points referenced in Section 2.2.3.

Fig. 3 has been relocated to Section 3 Results (now Fig. 6 in Section 3.3), where it directly supports the analysis of hydrological trend shifts. The associated text has been updated to reference the new location. The unit for RMSE has been corrected to m³/s in relevant Fig. 6. This correction aligns with the physical interpretation of RMSE for discharge data.

Revised part is now included on

line 138:
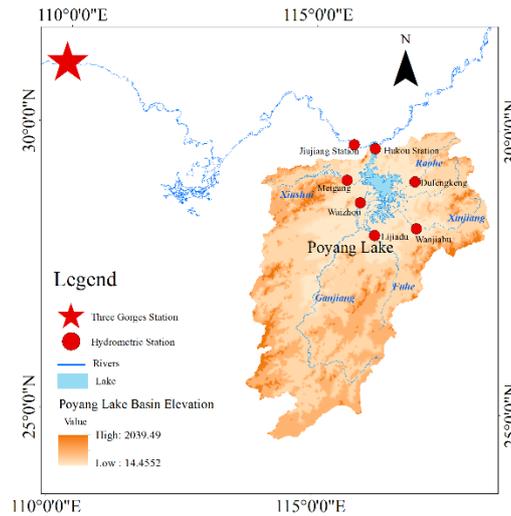
line 279-285:

Model results indicate high predictive accuracy for normal-flow events (Figure 6). The pre-TGP model achieves an $R^2 = 0.95$ with an RMSE $= 828.56$ m³/s, demonstrating its ability to effectively capture observed variations in Hukou outflow. Similarly, the post-TGP model yields an R^2 of 0.94 and an RMSE of 861.14 m³/s, comparable to pre-TGP performance. Overall, the LightGBM models exhibit stable predictive capability across different periods and serve as a reliable baseline tool for quantifying the hydrological impacts of the Three Gorges Project on Poyang Lake.
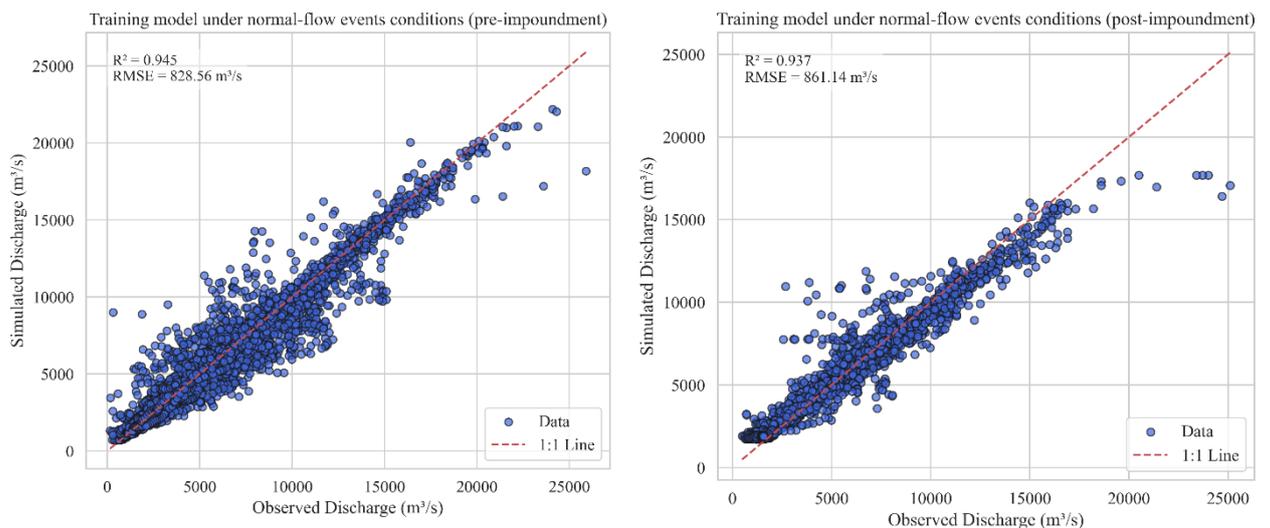

Figure 6 Validation of Simulated vs. Observed Values for the Normal-Flow Training Model

11. Comment: Section 2.2.4, The construction for the SWAT model of the Poyang Lake

basin should be very complex, involving the data collection, many hydrological parameters, and basin reservoirs. The model construction process is described inadequately, and a complete presentation of the calibration and validation procedures is lacking. This is an issue that cannot be overlooked. The absence of these critical components significantly undermines my confidence in the subsequent results.

Response:

We sincerely thank the reviewer for this critical comment. We fully acknowledge that the original description of the SWAT model construction was insufficiently detailed, which compromised the transparency of our methodology. In response, we have significantly expanded Section 2.2.4 to comprehensively address all aspects of model construction, data sourcing, parameterization, and validation.

Revised part is now included on

line 209-230:

The SWAT model inputs include digital elevation data, land-use data, soil data, hydrological data, and meteorological data (Table 1). The hydrological dataset consists of monthly discharge observations from key control stations on the five major tributaries within the basin, namely the Lijiadu Station on the Fuhe River, Meigang Station on the Raohe River, the Xinjiang Station, Dufengkeng Station on the Xiushui River, and Waizhou Station on the Ganjiang River. Meteorological inputs comprise daily mean air temperature, daily maximum and minimum temperature, mean relative humidity, daily precipitation, small- and large-pan evaporation, sunshine duration, mean atmospheric pressure, and mean wind speed.

As this study aims to quantify the impacts of hydraulic engineering—particularly the Three Gorges Project—the SWAT model was constructed under naturalized conditions, excluding reservoir regulation and other anthropogenic interventions. Model simulations were driven by observed monthly runoff from the five tributaries for the period 2005–2017. The calibration period was set to 2005–2011, while the validation period covered 2011–2016. A total of 27 key hydrological parameters were calibrated, and satisfactory model performance was achieved at all five control stations (Table 2. 3). The simulation results meet the requirements for quantitative hydrological process

analysis, with Nash–Sutcliffe efficiency (NSE) values exceeding 0.5.

Based on the calibrated SWAT model, naturalized monthly runoff from the five tributaries was used to generate corresponding simulated outflow at the Hukou outlet. By comparing the simulated naturalized outflow with observed discharge at Hukou, the impacts of the Three Gorges Project on lake outflow dynamics were quantitatively assessed.

Table 1 Data sources

Raster Data	References	Data scale	Time
Digital Elevation Model	www.data.tpdc.ac.cn	annual	2010
Landuse	www.resdc.cn	annual	2010
Soil Hydrologic Group	www.data.tpdc.ac.cn	annual	2010
	www.data.cma.cn	daily	2005—2017
Runoff	Hydrological Information Yearbook	monthly	2005—2017

Table 2 SWAT Model Key Parameters

Parameter	Units	Type
Surface Temperature Adjustment Factor	-	Basin Variable
Soil Temperature Adjustment Factor	-	Basin Variable
Maximum Soil Moisture Capacity	mm	Basin Variable
Minimum Soil Moisture Capacity	mm	Basin Variable
Percolation Index	-	Basin Variable
Surface Runoff Lag Time	days	Basin Variable
Groundwater Recharge Threshold	mm	HRU Variable
Slope Factor	-	HRU Variable
Maximum Canopy Storage Capacity	mm	HRU Variable
Overland Flow Manning's Roughness	-	HRU Variable
Soil Evaporation Compensation Factor	-	HRU Variable
Plant Uptake Compensation Factor	-	HRU Variable

Curve Number Adjustment Factor	-	Management Parameter
Biomixing Coefficient	-	Management Parameter
Soil Depth Adjustment Factor	-	Management Parameter
Soil Bulk Density Adjustment	g/cm³	Soil Parameter
Available Water Capacity Adjustment	mm	Soil Parameter
Hydraulic Conductivity Adjustment	mm/d	Soil Parameter
Soil Albedo Adjustment Factor	-	Soil Parameter
Groundwater Delay Time	days	Groundwater Variable
Baseflow Alpha Factor	-	Groundwater Variable
Minimum Groundwater Flow Threshold	mm	Groundwater Variable
Groundwater Re-evaporation Factor	-	Groundwater Variable
Minimum Re-evaporation Threshold	mm	Groundwater Variable
Recharge Depth Adjustment Factor	-	Groundwater Variable
Channel Manning's Roughness	-	Channel Parameter
Channel Diffusion Coefficient	m/s	Channel Parameter

Table 3 Performance of the SWAT model during calibration and validation periods

River	R^2	NSE
Ganjiang	0.79	0.79
Raohe	0.64	0.62
Xiushui	0.67	0.65
Fuhe	0.66	0.66
Xinjiang	0.60	0.60