Response to Reviewer 2's Comments:

This manuscript presents a hybrid modelling framework coupling WRF-Hydro with an LSTM-Attention residual corrector to attribute runoff decline and hydrological drought intensification in the Xilin River Basin, Inner Mongolia, over 1980 to 2020. The topic is timely; the intra-annual seasonal attribution is a genuine contribution over most comparable studies and using deep learning strictly as a residual corrector rather than a process replacement is a thoughtful design choice. However, several major methodological concerns must be resolved before the findings can be accepted. I detail these below.

Special comments:

(1) LSTM hyperparameters are not reported. Number of layers, hidden units, dropout rate, learning rate, optimizer, batch size, sequence length, and early stopping criteria are nowhere in the paper. For an AI-integrated study, this is an unacceptable omission. It is impossible to evaluate whether the network is appropriately sized, whether it is overfitting the training period, or whether the residual learning is physically meaningful.

Response:

We thank the reviewer for raising this important point. For studies that integrate deep learning, comprehensive disclosure of hyperparameters is essential for evaluating model capacity, overfitting risk, and the physical interpretability of residual learning. This aspect was insufficiently addressed in the original manuscript. In the revised manuscript, we have added a detailed description in Section 3.3 and included Table S1 summarizing the key hyperparameters of the LSTM-Attention residual correction model. After applying the sliding window (N_IN=6, N_OUT=3, stride=1), the training period (1980–1996, 204 months) yields 196 supervised samples, and the validation period (1997–2000) yields 40 samples. This configuration is well matched to the sample size and task complexity of our study. Importantly, because the model is designed to correct systematic residuals of WRF-Hydro rather than to simulate the full runoff series, the learning objective is more focused and can be effectively captured without resorting to a overly deep or wide network architecture.

To mitigate overfitting, we adopted a multi-layered strategy (detailed in Table S1). The training, validation, and simulation periods are strictly separated in chronological order to prevent information leakage. The model also uses dropout (0.3) and early stopping based on validation loss (patience = 20, with the best weights restored). As shown in Fig. 5 of the original manuscript, the gap between training and validation metrics is small (R^2 decreases by only about 0.08), showing no obvious signs of overfitting. Furthermore, during the 2001–2020 simulation period, WH-DL consistently outperforms WRF-Hydro across all metrics, indicating that the residual correction generalizes well beyond the training period rather than mere memorizing the training observations.

Regarding the physical meaning of residual learning, the inputs to LSTM-Attention are climate drivers (e.g., precipitation and temperature) that are directly linked to runoff generation and are consistent with the forcing data of WRF-Hydro. Under the same climate forcing, the model learns how the systematic bias of the physical model varies with different climate conditions, rather than performing purely data-driven curve fitting. Fig. 5 of the original manuscript shows that the improvements of WH-DL are mainly concentrated in low-flow periods and peak flow simulations, which aligns with the known limitations of

WRF-Hydro in capturing low-flow processes and snowmelt peaks in semi-arid basins. This indicates that the residual correction is precisely targeting the weak components of the physical model. The deep learning module serves only as a residual correction layer added on top of the physical simulation and does not replace physical processes (Cho and Kim, 2022; Kraft et al., 2022); the attribution analysis therefore remains grounded in physical mechanisms.

Table S1. Key hyperparameters of the LSTM-Attention residual correction model.

Hyperparameter	Value
Input window length	6 months
Forecast horizon	3 months
LSTM layers / units	1 layer, 64 units
Attention mechanism	Dense (1) scoring + Softmax temporal weighting
Dropout rate	0.3
Learning rate	1e-3
Optimizer	Adam
Batch size	16
Max epochs	150
Early stopping patience	20 epochs

We thank the reviewer again for this constructive comment, which has improved the transparency and assessability of the methodology.

(2) The baseline period "no human influence" assumption is contradicted by the authors' own supplemental data. Fig. S4 clearly shows water withdrawal, sheep population, and grazing intensity all increasing monotonically from 1980 onward. The authors acknowledge this as a limitation but never estimate the resulting bias on the 61%/39% attribution split. Presenting this split as a point estimate without uncertainty bounds is scientifically overconfident. Even a simple sensitivity test varying the change-point year by ±2 years, or assuming a linearly increasing baseline human influence, would substantially strengthen the quantitative credibility of the central finding.

Response:

We thank the reviewer for highlighting the issue of attribution uncertainty quantification. We have conducted a sensitivity analysis on the choice of change-point year and revised the relevant statements accordingly. We acknowledge that the strict assumption of "no human influence during the baseline period" does not fully hold. A more accurate description is that human influence was relatively limited during the baseline period (1980–2000), whereas the change period (2001–2020) reflects the combined effects of climate change and human activities. The revised manuscript has adjusted the wording accordingly and avoids absolute terms such as "negligible". Specifically:

(1) In Section 3.5 of the original manuscript (Line 324), "During the baseline period, runoff changes were minimally influenced by human activities and can be neglected." has been revised to "During the baseline period, runoff variability was predominantly driven by climate variability, with limited anthropogenic influence."

(2) In Section 4.1.1 (Line 366), "Accordingly, 1980–2000 was designated as the baseline period. During this stage, runoff changes were primarily driven by climatic factors, with negligible human activity impacts." has been revised to "Accordingly, 1980–2000 was designated as the baseline period, during which runoff variability was predominantly driven by climate variability, with limited anthropogenic influence."

Following the reviewer's suggestion, we redefined the baseline and change periods using 1999, 2000, 2002, and 2003 as alternative change-points years centered around 2001, and repeated the attribution analysis (Figure S1). Within this ±2-year perturbation range, the human contribution remained between 57.45% and 67.99%, while the climate contribution between 32.01% and 42.55%. Human activities remained the dominant driver (>50%) in all perturbation scenarios, and the original estimate of 61.04% (based on 2001) falls within this range, indicating that the attribution result is robust.

Further analysis shows that the human contribution increases as the change-point year is set later, while the climate contribution decreases correspondingly. A later change-point excludes earlier years of moderate human activity from the change period and concentrates the change period on the more recent stage of stronger human activity, thereby increasing the contrast in human activity intensity between the change period and the baseline period and raising the estimated human contribution. This systematic pattern also supports the plausibility of the sensitivity analysis.

Based on these results, we have expanded the key quantitative statements in the original manuscript from point estimates to expressions that include uncertainty ranges. We have also systematically expanded the discussion of the baseline-assumption limitation in Section 5 (Line 515). The original text:

"Consequently, it strengthens the robustness and applicability of attribution analysis. Nevertheless, the attribution framework in this study assumes that runoff variations before the change-point year were not influenced by human activities. This assumption introduces uncertainty into the quantitative analysis and may bias the calculation of contribution rates. This represents a limitation of the study."

has been revised to:

"Consequently, it strengthens the robustness and applicability of attribution analysis. Nevertheless, the attribution framework assumes that runoff variations during the baseline period were predominantly driven by climate variability, with limited anthropogenic influence. To test the robustness of the attribution to this baseline assumption, we repeated the attribution using change-point years of 1999, 2000, 2002, and 2003 (Figure S1). Across the ±2-year perturbation, the human contribution ranged from 57.45% to 67.99% and the climate contribution from 32.01% to 42.55%, with human activities remaining the dominant driver (>50%) in all scenarios. Across all tested partitions, human activities remain the dominant driver, indicating that the central attribution finding is not sensitive to the specific choice of change-point year."

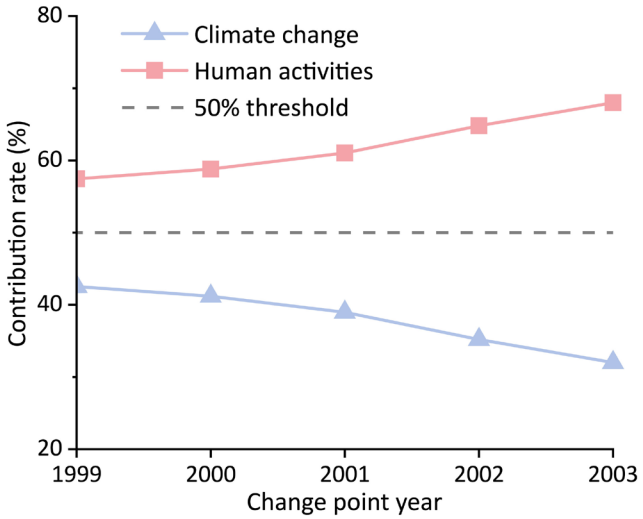

Figure S1. Sensitivity of attribution results to change-point year selection.

We also considered the reviewer's second alternative—assuming a linearly increasing human influence during the baseline period. However, this approach would require additional parameters that are difficult to constrain independently, thereby introducing new sources of uncertainty. In contrast, the change-point perturbation approach requires fewer assumptions and is more straightforward to interpret, therefore we adopted it in this study.

We thank the reviewer again for this suggestion. This additional analysis has substantially strengthened the quantitative robustness of the attribution results and provides a more solid scientific basis for the central conclusions.

(3) WRF dynamical downscaling configuration is entirely absent. The paper states ERA5 was downscaled to 12.5 km via WRF, but no physics schemes, domain configuration, boundary condition settings, or validation of the downscaled meteorology against station observations are provided anywhere. Since this downscaled forcing drives the entire WRF-Hydro simulation, errors introduced here propagate into all subsequent results. This is a critical gap in methodological transparency.

Response:

We thank the reviewer for the suggestion regarding the structural completeness of the manuscript. We fully agree that a detailed description of the dynamical downscaling configuration and its validation is essential, accordingly, we have added the corresponding description in Section 3.2 of the revised manuscript as follows.

The WRF model (version 4.5.2) was used to dynamically downscale the ERA5 reanalysis. The simulations employs the Lambert Conformal projection centered at 35.0°N, 105.0°E, with standard parallels at 10.0°N and 60.0°N. The domain covers the study area at a horizontal resolution of 12.5 km, and uses 35 vertical levels with the model top at 50 hPa. The simulation period is 1979–2020, with the first year serving as the spin-up year. ERA5 provides the initial and lateral boundary conditions, which are updated every 3 hours, and model outputs are saved at hourly intervals. The physical parameterization schemes used are listed in Table S2.

Table S2. WRF physics parameterization schemes.

Process	Scheme
Microphysics	Thompson
Longwave / Shortwave radiation	RRTMG
Planetary boundary layer / Surface layer	MYJ
Cumulus convection	Kain-Fritsch
Land surface	Noah-MP
Land use	MODIS
Sea surface temperature	Time-varying SST update

In addition to the above details on the WRF configuration and physics schemes, we have added a validation of the downscaled temperature and precipitation fields in the revised manuscript. As precipitation and temperature are the dominant meteorological drivers controlling runoff generation and the land-surface water balance in WRF-Hydro, the validation focuses on these two variables and uses the correlation coefficient (CC), bias, and root-mean-square error (RMSE) at the daily scale. The validation results are presented below:

Due to the sparse observational network within the basin, with only one meteorological station (Xilinhot) available, station observations alone are insufficient to fully represent the basin-wide climatic variability. Therefore, we used the CN05.1 dataset as the reference data, which is a high-resolution gridded observational dataset developed by interpolating daily observations from more than 2,400 national meteorological stations across China (Wu and Gao, 2013). Considering that precipitation and temperature are the dominant meteorological drivers of runoff generation and land-surface water balance in WRF-Hydro, the validation of the forcing data was focused on these two variables. Figure S2 presents the comparison between the daily meteorological forcing data and the reference observations over the Xilin River Basin. The results show that WRF-simulated temperature agrees well with the reference data(CC=0.99, bia=0.69 °C, RMSE=2.2 °C). For precipitation, WRF also reproduces the daily variability reasonably well(CC=0.65, =bias=0.12 mm d^{-1}, RMSE=mm d^{-1}). Overall, these validation results indicate that the WRF downscaled data used in this study have acceptable accuracy for the study area and provide reliable support for the subsequent WRF-Hydro simulations and attribution analysis.

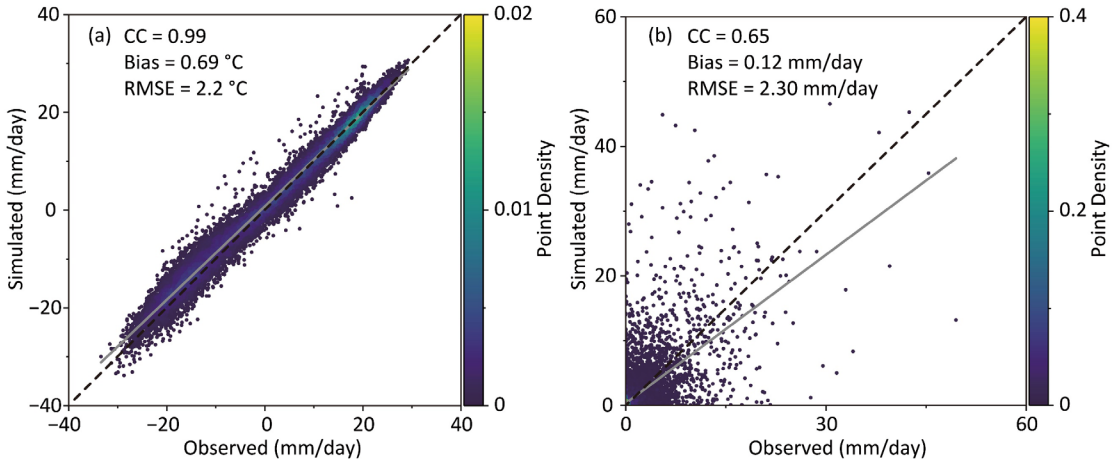

Figure S2. Comparison of daily WRF downscaled temperature (a) and precipitation (b) against reference observations.
Note: CC, correlation coefficient; Bias, mean bias; RMSE, root mean square error.

(4) Physical consistency of the DL-corrected runoff is never verified. The LSTM corrector adds a learned residual to WRF-Hydro output. The paper never checks whether the corrected runoff series satisfies approximate water balance closure, whether corrections are physically bounded (e.g., no negative runoff), or whether the magnitude of corrections is physically interpretable. A neural network trained to minimize residuals can produce corrections that are statistically optimal but physically spurious, particularly in the simulation period where conditions may differ from training.

Response:

We thank the reviewer for this valuable suggestion, which helps strengthen the rigor of the manuscript. In the revised manuscript, we have added a physical-plausibility diagnosis of the deep-learning–corrected runoff series (Q_sim). Across all 492 months from 1980 to 2020, the simulated runoff values are non-negative, with no violations of basic physical constraints. To assess water-balance consistency, given that groundwater observations in the basin are difficult to obtain and that a full ET-based closure involves multiple components that are hard to constrain independently, we adopted the runoff coefficient (Q/P) as a simplified constraint metric (Yilmaz et al., 2008; Sawicz et al., 2011). During the baseline period, Q_sim/P (1.58%) is close to Q_obs/P (1.61%), indicating that LSTM-Attention reasonably captures the actual hydrological response within the training domain. During the change period, Q_sim/P (1.44%) exceeds Q_obs/P (0.96%). This pattern aligns with our attribution logic: Q_sim reflects baseline-extrapolated runoff, whereas Q_obs additionally captures human-induced reductions. The overall runoff coefficient of the basin is consistent with values of about 0.02 reported for the Xilin River Basin in previous studies (Tang et al., 2014), and falls within the typical range for semi-arid basins.

In terms of correction magnitude, we computed the long-term mean of the correction term $r_f(t) = Q_{sim}(t) - Q_{WH}(t)$ for each period. The annual mean $|r_f|$ accounts for 6.0% of Q_sim during the baseline period and increases to 27.7% during the change period. LSTM-Attention learns the bias of WRF-Hydro relative to observations during the baseline period and extrapolates this bias to the change period. The smaller correction during the baseline period indicates that WRF-Hydro already captures the physical processes of the basin reasonably well during the calibration period, whereas the larger

correction during the change period reflects the amplification of systematic biases in WRF-Hydro under extrapolated climate conditions. This change is an inherent extrapolation limitation of the physical model itself.

Regarding the physical consistency between the training and application periods, the small difference between Q_sim and Q_obs during the baseline period is itself a direct outcome of the training convergence. The real test of the physical consistency of the correction comes from the application period. If the correction stems solely from numerical optimization without a physical basis, the deviation patterns during the application period would lack systematic structure (Coron et al., 2012). However, the monthly deviations of Q_sim from Q_obs during the change period exhibit a structural pattern that aligns with the seasonal cycle of human activities in the basin. The deviations are mainly concentrated during the snowmelt peak (April), the irrigation season (July–August), and early autumn (September–October), which are precisely the stages most strongly affected by human activities in the basin (anthropogenic water diversion during the spring snowmelt period, agricultural irrigation in summer, the lagged effects of agricultural water consumption, and the sustained influence of grazing). These periods are consistent with the human-activity-dominated stages identified in Section 4.3 of the original manuscript. The deviation in March is close to zero, which is consistent with the conclusion in the original manuscript that the runoff increase during the early snowmelt period is dominated by climate change with a weaker contribution from human activities. This structural deviation, which corresponds to the seasonal pattern of human activities, provides strong evidence that the correction is physically interpretable (Frame et al., 2021).

(5) Noah-MP parameterization options are undisclosed. Noah-MP has over 40 configurable physics options governing snow, ET, soil hydrology, and runoff generation. The specific combination used is never stated. Given that the paper's main climatic drivers are PET and snowmelt, both directly controlled by Noah-MP options, this is not a minor omission. Different Noah-MP configurations can yield substantially different attribution results.

Response:

We thank the reviewer for raising this important point. The original manuscript only made a general reference to "the Noah-MP land surface scheme" without listing specific option numbers, which was an oversight in our description. In the revised manuscript, we have added Table S3 in the Appendix, which lists the key physical parameterization options used in the offline Noah-MP module of WRF-Hydro, covering important processes including dynamic vegetation, stomatal conductance, runoff, rainfall–snowfall partitioning, snow albedo, surface resistance, and radiative transfer.

Among these, the option most directly relevant to the present study is the runoff scheme. We adopted RUNOFF_OPTION = 3, the Original Noah scheme with free drainage at the bottom of the soil column. On the one hand, TOPMODEL-based schemes (Options 1 and 2) require parameterization based on high-quality topographic indices and long-term groundwater-table observations, which are not available for the Xilin River Basin and introduce greater parameter uncertainty. On the other hand, this scheme has been successfully applied and validated in runoff simulations over arid and semi-arid inland river basins in northern China (Yu et al., 2023).

For the options most directly related to evapotranspiration and snowmelt processes in this study, we adopted the Ball-Berry stomatal conductance scheme, the Noah soil moisture stress scheme, and the Sakaguchi-Zeng surface resistance scheme. This combination is a common configuration of Noah-MP for semi-arid regions and can simultaneously represent the suppression of transpiration under drought stress and the differentiated processes of evaporation from snow surfaces and bare soil. For rainfall–snowfall partitioning, we used the Jordan scheme. The Xilin River Basin has no glacier coverage and limited snowfall, so the snowmelt process is relatively simple. The Jordan scheme is widely applied and provides a stable and reliable representation of phase partitioning under the low-elevation, semi-arid conditions of this basin.

Table S3. Summary of default physical options in Noah-MP model

Noah-MP options	Options	Description
Dynamic vegetation option	4	Table LAI with maximum FVEG
Canopy stomatal resistance option	1	Ball-Berry scheme
BTR option	1	Noah soil moisture stress
Runoff option	3	Original Noah runoff with free drainage
Surface drag option	1	Monin-Obukhov similarity
Frozen soil option	1	Linear effects, more permeable
Supercooled water option	1	No iteration
Radiative transfer option	3	Two-stream applied to vegetated fraction
Snow albedo option	2	CLASS scheme
Partitioning precipitation into rainfall and snowfall	1	Jordan
TBOT option	2	Prescribed deep-soil temperature
Temperature time scheme	3	Semi-implicit
Glacier option	2	Simple glacier
Surface resistance option	4	Sakaguchi-Zeng resistance

(6) The model abbreviation switches between "WH-LA" (Fig. 5 caption) and "WH-DL" (throughout the text) with no explanation. Pick one and apply it consistently.

Response:

We thank the reviewer for pointing out this inconsistency. The alternation between "WH-LA" and "WH-DL" in the original manuscript indeed caused confusion. In the revised manuscript, we have unified the abbreviation as "WH-DL" (WRF-Hydro coupled with deep learning) and have double checked and revised the main text, figure captions, and all related locations to ensure consistent usage throughout.

(7) Figure 7's caption is excessively long and explains the three drought classification types that should have been defined in the Methods section. Move the classification definitions to Section 3.4.

Response:

We thank the reviewer for this constructive suggestion, we have moved the classification of the three types of human activity impacts on drought to the end of Section 3.4, and we have simplified the caption of Figure 7 and added "definitions provided in Section 3.4" at the end to point readers to the Methods. The specific revisions are as follows.

A new paragraph has been added at the end of Section 3.4:

Furthermore, based on the comparison between observed and simulated runoff, this study classifies the impact of human activities on hydrological drought events into three categories (Figure 7). Specifically, when the simulated runoff indicates drought (SRI-12 < −0.5) and the observed runoff is higher than the simulated runoff, human activities are considered to have alleviated the drought. When both the observed and simulated runoff indicate drought but the simulated runoff is higher than the observed runoff, human activities are considered to have intensified the existing drought. When the observed runoff indicates drought but the simulated runoff does not, human activities are considered to have triggered a drought event that would not otherwise have occurred.

Figure 7. Three types of human activity impact patterns on hydrological drought evolution. Note: The black solid line represents observed SRI-12 values; the red solid line represents simulated values. The black dashed line (SRI-12 = −0.5) indicates the drought threshold. The blue, red, and yellow shaded areas correspond to alleviation, intensification, and triggering effects, respectively; definitions are provided in Section 3.4.

(8) The Discussion repeats several findings verbatim from the Results section (e.g., the 58.87% Snow-M contribution, the 61.04% human attribution). Discussion sections should interpret and contextualize findings, not restate them.

Response:

We thank the reviewer for helping us strengthen the Discussion. We have made the following revisions in Section 5.

(1) Section 5.1, second paragraph. Original: "Despite rainy season precipitation accounting for 85.33% of annual totals, its conversion efficiency to runoff remains limited, contributing only 50.89%." Revised to: "Although the rainy season dominates the annual precipitation budget, its conversion efficiency to runoff is disproportionately low, indicating that a substantial fraction of incoming water is consumed by evapotranspiration before reaching the channel network."

(2) Section 5.1, third paragraph. Original: "This corroborates the finding that runoff reduction during this period accounts for 58.87% of the total annual reduction." Revised to: "This mechanism explains why the snowmelt season emerges as the dominant temporal contributor to annual runoff reduction, reflecting the cryosphere's outsized leverage on hydrological response in cold semi-arid basins where snow functions as the principal seasonal storage buffer."

We have also checked the manuscript and confirmed that "61.04%" does not appear in the Discussion. This value is only used in the Abstract, Section 4.3 (Results), and the Conclusion, where it serves as a quantitative summary of the paper's central finding consistent with the function of each of these sections, and we have therefore retained it.

(9) The paper does not state the number of WRF-Hydro parameters that were calibrated, the calibration algorithm used, or the objective function. This is standard information for any distributed hydrological modelling paper.

Response:

We thank the reviewer for pointing out the inadequate description of the calibration methodology in the original manuscript. We have added a complete description of the calibration procedure in the revised manuscript.

Following previous studies in arid and semi-arid basins (Guo et al., 2024; Yu et al., 2023) and considering the sensitivity characteristics of WRF-Hydro parameters, we calibrated seven key parameters that strongly influence runoff simulation, covering soil hydraulic properties, surface runoff generation, overland flow routing, and channel routing. The physical meaning, default value, candidate range, and final calibrated value of each parameter are summarized in Table S4. The calibration was carried out using the one-at-a-time (OAT) method: while keeping the other parameters at their current optimal values, each target parameter was adjusted individually by scanning its candidate range given in Table S4, with the calibration period set to 1981–1985, and the optimal value was selected based on the simulation performance during the calibration period. This procedure was applied sequentially to all seven parameters. The calibration used R^2, NSE, and KGE jointly as evaluation criteria, and the parameter combination that achieved the best overall performance across these three metrics was selected to enhance the robustness of the calibration.

The following description has been added in Section 3.2 of the revised manuscript:

Following previous studies in arid and semi-arid basins (Guo et al., 2024; Yu et al., 2023) and considering the sensitivity characteristics of WRF-Hydro parameters, seven parameters with strong influence on runoff simulation were selected for calibration (Table S4). The calibration was carried out using the one-at-a-time (OAT) method: while keeping the other parameters at their current optimal values, each target parameter was adjusted individually by scanning its candidate range, and its optimal value was selected based on the simulation performance during the calibration period (1981–1985). The calibration used R^2, NSE, and KGE jointly as evaluation criteria, and the parameter combination that achieved the best overall performance across these three metrics was selected to enhance the robustness of the calibration.

We sincerely thank the reviewer for the valuable suggestions on methodological description, model transparency, and attribution uncertainty.

Table S4 Model parameters considered in the calibration.

Name	Description	Default	Candidate values	Optimal value
bexp	Pore size distribution index	×1	0.1, 0.4, 0.7, 1, 3, 5, 7, 10	0.7
dksat	Saturated hydraulic conductivity	×1	0.1, 0.5, 0.7, 1, 2, 3, 5, 10	1
smcmax	Saturation soil moisture content	×1	0.1, 0.5, 0.7, 1, 1.5, 2, 3, 5	0.5
REFKDT	Surface runoff parameter	3	1, 2, 3, 3.5, 4, 4.5, 5, 10	4.5
slope	Openness of Bottom drainage	0.1	0.01, 0.03, 0.05, 0.07, 0.1,	0.05

	boundary		0.2, 0.3, 0.5	
OVROUGHRTAC	Overland Manning roughness multiplier	1	0.01, 0.1, 0.3, 0.5, 0.7,1, 3, 5	0.5
mann	Channel Manning roughness	×1	0.1, 0.2, 0.3, 0.5, 0.7, 1, 2, 5	0.5

Note: For parameters with a default value of ×1, the candidate and optimal values also represent multipliers applied to the spatially distributed defaults.

References

Cho, K. and Kim, Y.: Improving streamflow prediction in the WRF-Hydro model with LSTM networks, Journal of Hydrology, 605, 127297, https://doi.org/10.1016/j.jhydrol.2021.127297, 2022.

Coron, L., Andréassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M., and Hendrickx, F.: Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments, Water Resources Research, 48, W05552, https://doi.org/10.1029/2011WR011721, 2012.

Frame, J. M., Kratzert, F., Raney II, A., Rahman, M., Salas, F. R., and Nearing, G. S.: Post-processing the National Water Model with Long Short-Term Memory Networks for streamflow predictions and model diagnostics, JAWRA Journal of the American Water Resources Association, 57, 885–905, https://doi.org/10.1111/1752-1688.12964, 2021.

Guo, S., Tian, L., Chen, S., Liang, J., Tian, J., Cao, B., Wang, X., and He, C.: Analysis of effects of vegetation cover and elevation on water yield in an alpine basin of the Qilian Mountains in Northwest China by integrating the WRF-Hydro and Budyko framework, Journal of Hydrology, 629, 130580, https://doi.org/10.1016/j.jhydrol.2023.130580, 2024.

Kraft, B., Jung, M., Körner, M., Koirala, S., and Reichstein, M.: Towards hybrid modeling of the global hydrological cycle, Hydrology and Earth System Sciences, 26, 1579–1614, https://doi.org/10.5194/hess-26-1579-2022, 2022.

Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., and Carrillo, G.: Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA, Hydrology and Earth System Sciences, 15, 2895–2911, https://doi.org/10.5194/hess-15-2895-2011, 2011.

Tang, X.-W., Wu, J.-K., Xue, L.-Y., Zhang, M.-Q., Barthold, F., Breuer, L., and Frede, H.-G.: Major ion chemistry of surface water and its controlling factors in the Xilin River Basin, Environmental Science (China), 35, 131–142, 2014.

Yilmaz, K. K., Gupta, H. V., and Wagener, T.: A process-based diagnostic approach to model evaluation: application to the NWS distributed hydrologic model, Water Resources Research, 44, W09417, https://doi.org/10.1029/2007WR006716, 2008.

Yu, E., Liu, X., Li, J., and Tao, H.: Calibration and evaluation of the WRF-Hydro Model in simulating the streamflow over the arid regions of Northwest China: a case study in Kaidu River Basin, Sustainability, 15, 6175, https://doi.org/10.3390/su15076175, 2023.