Response to Reviewer 1's Comments:

The manuscript investigates runoff decline in semi-arid regions using a hybrid WRF-Hydro and LSTM-Attention framework. While the integration of physical modeling with error correction is interesting, several critical issues regarding data validation and methodology transparency must be addressed.

Special comments:

1.Scalability concerns

In distributed hydrological modeling, it is a standard and reasonable paradigm to conduct in-depth physical mechanism analysis within small or medium sized catchments. However, the authors are encouraged to supplement the discussion by clarifying to what extent the core findings can be generalized to other typical basins globally.

Response:

We thank the reviewer for this valuable suggestion regarding the generalizability and applicability of our framework. Following this suggestion, we have added the following clarification to the revised manuscript:

The hybrid framework proposed in this study is methodologically transferable. It integrates a distributed physically based hydrological model with deep-learning-based residual correction, and further applies this combined framework to attribute the respective impacts of climate change and human activities. This approach has the potential to be extended to other basins with similar hydroclimatic settings, particularly arid and semi-arid regions where observational data are relatively scarce and hydrological processes are jointly influenced by climate variability and human activities. However, deep learning methods inevitably simplify the physical processes represented in hydrological models (Reichstein et al., 2019). To mitigate this limitation, our study only uses deep learning to predict and correct the systematic residuals of the physical model, thereby improving simulation accuracy while preserving physical consistency to the greatest extent possible. It should also be noted that the quantitative results of this study are basin-specific. For example, the identified change points in runoff and the optimized model parameter values are both influenced by local climatic conditions, underlying surface characteristics, and the intensity of human activities. Therefore, when applying this framework to other basins, it remains necessary to recalibrate the model and redefine attribution periods based on local data and regional characteristics. Overall, the hybrid modeling and attribution framework developed in this study provides a promising approach for hydrological studies in data-limited basins. In future work, we will further integrate multi-source datasets and test the applicability of this framework across larger spatial scales and a wider range of basin types.

2.Validation of Downscaled Data

Line 102 mentions that dynamical downscaling techniques were applied to generate high-resolution meteorological forcing data. While downscaling increases spatial resolution, it does not inherently guarantee improved accuracy, and regional climate models are prone to systematic biases. The manuscript currently lacks a comparative validation between the 12.5 km WRF outputs and actual ground

meteorological observations within the basin. Although an LSTM-Attention module is used for error correction, a rigorous physics based framework requires a quality assessment of the initial meteorological input sources.

Response:

We thank the reviewer for this important comment. We agree that although dynamical downscaling increases the spatial resolution of meteorological data, it does not necessarily guarantee improved simulation accuracy. Therefore, before using the WRF downscaled outputs as meteorological forcing for WRF-Hydro, we supplemented the manuscript with an evaluation of the forcing data quality. Given that precipitation and temperature are the key meteorological drivers controlling runoff generation and land-surface water balance in WRF-Hydro, our validation focused on these two variables. We evaluated the forcing data at the daily scale using the correlation coefficient (CC), bias, and root mean square error (RMSE). The validation results and corresponding explanation are as follows:

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, this study used the CN05.1 dataset as the reference data to evaluate the reliability of the forcing data. CN05.1 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 1 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, with a CC of 0.99, a bias of 0.69 °C, and an RMSE of 2.2 °C. For precipitation, WRF also reproduces the daily variability reasonably well, with a CC of 0.65, a bias of 0.12 mm d^{-1}, and an RMSE of 2.3 mm d^{-1}. Overall, these validation results indicate that the WRF downscaled forcing data used in this study have acceptable accuracy in the study area and can provide reliable support for the subsequent WRF-Hydro simulations and attribution analysis.

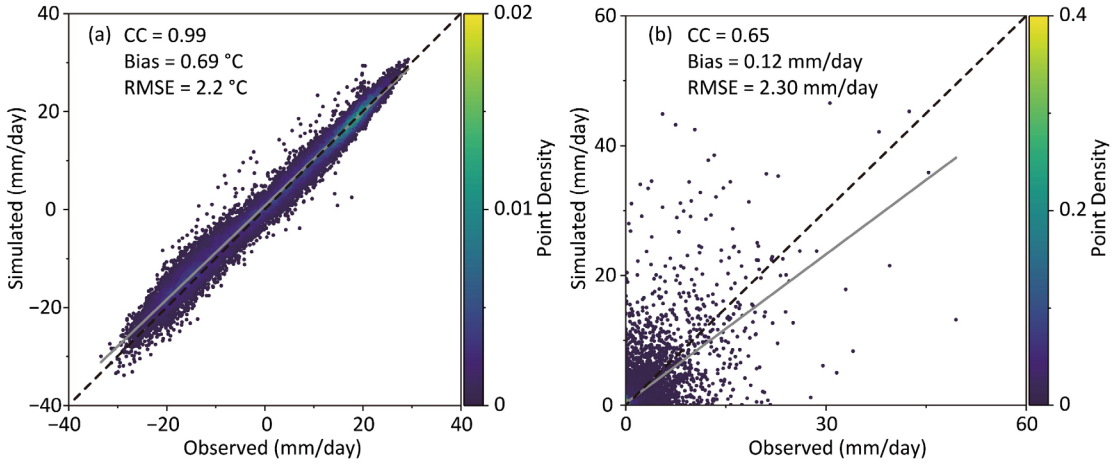

Figure 1. Comparison of daily WRF downscaled temperature (a) and precipitation (b) against reference observations.

Note: CC, correlation coefficient; Bias, mean bias; RMSE, root mean square error.

3.Transparency of WRF-Hydro Parameter Calibration

Line 140 states that "parameters sensitive to hydrological processes were selected for calibration," but the specific parameters are not identified. It is recommended to include a table (in Supplementary) listing the names, physical meanings, initial ranges, and final calibrated values of the key parameters to enhance the transparency and reproducibility.

Response:

We thank the reviewer for pointing out this issue in our manuscript. To improve the transparency and reproducibility of the WRF-Hydro parameter calibration process, we have added Table 1 to the Supplement. This table lists the key calibration parameters, including their names, physical meanings, default values, candidate values tested during calibration, and final calibrated values. For parameters with a default value denoted as "×1", the candidate and optimal values represent multiplicative scaling factors applied to the spatially distributed default parameter fields. The parameter calibration was conducted using a one-at-a-time (OAT) approach, and the corresponding description has been added to the revised manuscript.

Revised text in the manuscript:

Referring to previous studies conducted in arid and semi-arid basins (Guo et al., 2024; Yu et al., 2023), and considering the sensitivity characteristics of WRF-Hydro parameters, this study selected several parameters with strong influences on runoff simulation for calibration (Table 1). The calibration was performed using a one-at-a-time (OAT) approach, in which each target parameter was adjusted individually while keeping all other parameters unchanged, and the optimal value was determined based on the runoff simulation performance.

Table 1 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 boundary	0.1	0.01, 0.03, 0.05, 0.07, 0.1, 0.2, 0.3, 0.5	0.05
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.

4.Sample Size and Overfitting Prevention

According to Line 159, the training period covers 1980–1996. If monthly resolution data were used, the sample length would be 204? Could the authors clarify the exact sample size? Given such a limited

dataset, what specific measures were implemented to prevent the deep learning model from overfitting? Additionally, please provide the key hyperparameters to facilitate reproducibility and further academic exchange.

Response:

We thank the reviewer for this valuable comment. We agree that sample size is an issue that requires careful consideration in deep learning applications. Therefore, we adopted multiple strategies during model training to reduce the risk of overfitting. Specifically, the training period covered 1980–1996, corresponding to 204 monthly time steps. Because the model was trained using a sliding-window strategy to construct supervised samples, with an input window of N_IN = 6, an output horizon of N_OUT = 3, and a stride of 1, the actual number of supervised training samples was 196. The validation period (1997–2000) included 48 monthly time steps, corresponding to 40 supervised samples.

In addition, the LSTM-Attention model in this study was not used to directly predict the full runoff series, but rather served as a correction module for the systematic residuals of WRF-Hydro. Compared with direct runoff prediction, the residual series has a narrower dynamic range and lower nonlinearity, which may help reduce the risk of overfitting under limited-sample conditions.

To further mitigate overfitting, we adopted a strict time-based data split (training: 1980–1996; validation: 1997–2000; application: 2001–2020) to avoid information leakage. During training, early stopping was applied based on the validation loss (patience = 20, with restoration of the best weights), and dropout (rate = 0.3) was introduced. Moreover, the network architecture was intentionally kept simple, consisting of a single-layer LSTM with 64 hidden units. In addition, the normalization parameters for all input and output variables were calculated exclusively from the training-period statistics and then consistently applied to both the validation period and the independent application period. The key hyperparameters have been summarized in Table 2.

Table 2. 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
Optimizer	Adam
Batch size	16
Max epochs	150
Early stopping patience	20 epochs

5.Link Human Activity Data with Attribution Results

The authors list data regarding human activities in lines 355-365. Please clarify how these multi-source

datasets were statistically linked to the "effect of human activities" derived from attribution analysis. While "Correlation analysis" is explicitly mentioned in the technical flowchart (Figure 2), there appears to be a lack of corresponding methodological description or results in the main text. Please confirm if this analysis was performed and provide detailed evidence.

Response:

We thank the reviewer for this helpful suggestion regarding the manuscript structure. We had conducted the correlation analysis during the initial stage of the study design, but it was not included in the original manuscript in order to maintain focus on the main conclusions. Following the reviewer's suggestion, we have now incorporated this analysis into the revised manuscript, making it consistent with the "Correlation analysis" component shown in the workflow of Figure 2. It should be noted that this correlation analysis was not intended to replace the attribution framework in quantitatively separating the effects of human activities. Rather, it was used as a supplementary interpretation of the attribution results. Specifically, it serves to examine the consistency between different human activity indicators and variations in hydrological drought, thereby improving the interpretability of the attribution results and the overall coherence of the manuscript. Revised text in the manuscript:

To enhance the interpretability of the attribution results, we further analyzed the relationships between hydrological drought (SRI-12) and representative climatic factors as well as human activity indicators (Figure 2). The results show that hydrological drought is significantly negatively correlated with PET, cropland, forestland, water withdrawal, and grazing intensity, while it is significantly positively correlated with snowmelt, precipitation, and grassland. These results suggest that hydrological drought is jointly regulated by water supply and atmospheric evaporative demand. Precipitation and snowmelt are key sources of runoff recharge in arid and semi-arid regions and are therefore closely linked to hydrological processes (Berghuijs et al., 2014; Chai et al., 2025; Li et al., 2024). In contrast, enhanced evaporative demand tends to accelerate soil water depletion and reduce basin runoff generation, thereby intensifying hydrological drought.

In addition, grassland and forestland exhibit a significant negative correlation (−0.90). Since 2000, a series of ecological restoration policies, such as afforestation and the Grain-for-Green Program, have been implemented by the government (Li et al., 2023). While these measures have improved the ecological environment, they may also alter basin water balance by enhancing vegetation transpiration (Liu et al., 2026). In particular, under a warming climate and conditions of insufficient water supply, dense vegetation cover may further aggravate hydrological drought by increasing evapotranspiration and consuming deeper soil water storage (Schwärzel et al., 2020; Zhou et al., 2024).

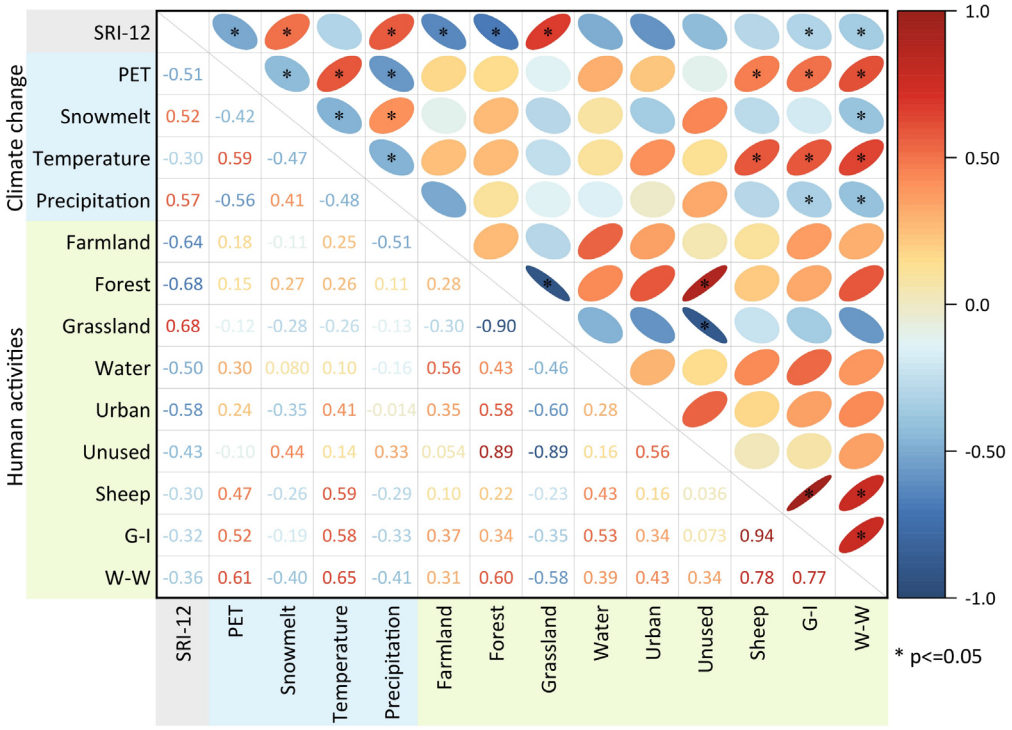

Figure 2. Correlations of hydrological drought with major climatic factors and human activity indicators.
Note: G-I, grazing intensity; W-W, water withdrawal.

We thank the reviewer again for the valuable comments. These comments will also be reflected in the revised manuscript.

References

Berghuijs, W. R., Woods, R. A., and Hrachowitz, M.: A precipitation shift from snow towards rain leads to a decrease in streamflow, Nature Climate Change, 4, 583-586, https://doi.org/10.1038/nclimate2246, 2014.

Chai, Y., Miao, C., Gentine, P., Mudryk, L., Thackeray, C. W., Berghuijs, W. R., Wu, Y., Fan, X. W., Slater, L., Sun, Q., and Zwiers, F.: Constrained Earth system models show a stronger reduction in future Northern Hemisphere snowmelt water, Nature Climate Change, 15, 514-520, https://doi.org/10.1038/s41558-025-02308-y, 2025.

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.

Li, B., Sun, T., Tian, F., Tudaji, M., Qin, L., and Ni, G.: Hybrid hydrological modeling for large alpine basins: a semi-distributed approach, Hydrology and Earth System Sciences, 28, 4521-4538, https://doi.org/10.5194/hess-28-4521-2024, 2024.

Li, J., Huang, L., Cao, W., Wang, J., Fan, J., Xu, X., and Tian, H.: Benefits, potential and risks of China's grassland ecosystem conservation and restoration, Science of The Total Environment, 905, 167413, https://doi.org/10.1016/j.scitotenv.2023.167413, 2023.

Liu, Y., Li, Z., Chen, Y., Gao, L., Zhang, L., Turnadge, C., He, B., Zhou, X., Duan, W., Li, B., Fang, G., and Huang, W.: Vegetation dynamics offset nearly 30% of precipitation's impact on runoff and are amplified by aridity, One Earth, 9, 101612, https://doi.org/10.1016/j.oneear.2026.101612, 2026.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195-204, https://doi.org/10.1038/s41586-019-0912-1, 2019.

Schwärzel, K., Zhang, L., Montanarella, L., Wang, Y., and Sun, G.: How afforestation affects the water cycle in drylands: A process-based comparative analysis, Global Change Biology, 26, 944-959, https://doi.org/10.1111/gcb.14875, 2020.

Wu, J. and Gao, X.: A gridded daily observation dataset over China region and comparison with the other datasets, Chinese Journal of Geophysics-Chinese Edition, 56, 1102-1111, https://doi.org/10.6038/cjg20130406, 2013.

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.

Zhou, J., Wang, Y., Li, R., He, H., Sun, H., Zhou, Z., Zhao, Y., Zhang, P., and Li, Z.: Response of deep soil water deficit to afforestation, soil depth, and precipitation gradient, Agricultural and Forest Meteorology, 352, 110024, https://doi.org/10.1016/j.agrformet.2024.110024, 2024.