Author comments on CC1- egusphere-2025-1161

Comment 1: *Is there any reference or rationale for the determination of the watershed boundaries?*

Response:

Thank you for your comments. The determination of watershed boundaries is very important. Our previous manuscript did not provide a detailed description of our method for determining watershed boundaries. In fact, when we established the dataset, we used the watershed boundaries provided by the HydroSHEDS dataset. We then screened the watersheds based on their area and deleted the watersheds with an area of less than one thousand square kilometers. We have added a description of the method for determining watershed vector boundaries in the data section. The modified details are as follows:

"….

To delineate watershed boundaries consistently across China, we utilized the HydroSHEDS dataset (Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales; Lehner et al., 2008), which provides high-resolution hydrologic information derived from SRTM elevation data. Based on the D8 flow direction scheme, the outlet of each basin was determined and used to extract the upstream contributing area. To ensure data consistency and model applicability, we performed area-based filtering and excluded small basins with an area less than 1,000 km^2. This threshold was chosen to reduce the influence of spatial resolution mismatch and potential errors in meteorological data aggregation. As a result, a total of 544 basins were retained, representing a wide range of hydrological and climatic conditions across China. The final basin boundaries are shown in Figure 1(a), and outlet locations are provided in Supplementary Figure S1.

…."

Comment 2: *The abbreviations of the models in the article are very confusing. Please explain them uniformly in the appropriate place.*

Response:

Thank you for your reminder. Our study did use 7 different models. To make the article easier for readers to understand, we have declared an abbreviation for each model after it is mentioned for the first time. At the same time, we have summarized the detailed information of all the models used in this study into a table (as shown in Table 3). This table will be added to the new manuscript. The specific table content is as follows:

"….

Table 3 The details for all the models.

Categories	Model	Model inputs	Training set	Testing set
process-based model	EXP-HYDRO	meteorological forcings	October 1, 1975, to September 30, 1995	October 1, 1995, to September 30, 2015
	Xin'anjiang	meteorological forcings		
Deep learning model	LSTM	meteorological forcings, static basin attributes		
alternative hybrid model	EXP-IN-LSTM	EXP-HYDRO predicted runoff, meteorological forcings, static basin attributes		
	XAJ-IN-LSTM	Xin'anjiang predicted runoff, meteorological forcings, static basin attributes		
differentiable hybrid model	EXP-dPL	meteorological forcings, static basin attributes		
	XAJ-dPL	meteorological forcings, static basin attributes		

…."

Comment 3: *Line 284 describes the PUB test method. Why are the remaining 9 clusters used for training?*

Response:

Thank you for your careful review. We used 5-fold cross validation to test the generalization ability of the model when extrapolating ungauged basins. Specifically, we first randomly divided 544 basins into five clusters. Four of the five clusters were used as training sets, and the model was trained during the training period of these four clusters. Then the prediction performance of the model was tested using the remaining test period data. The purpose of this is to ensure that the model has neither been exposed to any data from the test basin during training nor learned any data from the test basin during the test period. Ensure that the test of the model's generalization ability spans both space and time. There is a problem with our statement here, and we will describe the pub test scheme in more detail in the new manuscript. The modified details are as follows:

"….

The validation process is as follows: the model is trained using the training period data from the basins in four of the clusters, and its performance is validated on the test period data from the remaining a cluster.

…."

Comment 4: *The authors claim that the differentiable mixed hydrological model can output unobserved intermediate hydrological variables, but there is no data to support this.*

Response:

Thank you very much for your valuable comment. One of the potential advantages of the differentiable hybrid hydrological model is that it can output relatively reliable intermediate hydrological variables under the premise of runoff as the target variable. In our original manuscript, we hoped to illustrate that the model can output reliable intermediate hydrological variables by calculating the closure of the water budget. Specifically, when the evapotranspiration output by the differentiable hybrid hydrological model is used to measure the water budget, the number of basins that achieve closed balance increases significantly. Therefore, we believe that the model can output relatively reliable intermediate hydrological variables on the basis of having ideal prediction performance, and its output evapotranspiration can also better conform to physical consistency, so the model can output relatively reliable intermediate hydrological variables.

However, as you mentioned, the reliability of the hydrological variables output by the model can be most directly illustrated by directly comparing the intermediate hydrological variables output by the differentiable hybrid model with the observed values. Therefore, we add a comparative analysis of the hydrological variables output by the model and the corresponding observed values (ERA5-Land). The specific results and analysis will be added to the new manuscript, as follows:

"….

4.6 Validation of Intermediate Hydrological Variables

To further assess the physical consistency and interpretability of the differentiable hybrid hydrological model, we evaluated its ability to reproduce intermediate hydrological variables that were not used as training targets. Specifically, we focused on three key variables: evapotranspiration, soil water, and snowpack. These variables were output by the differentiable hybrid model during the testing period (1995–2015), and compared with corresponding reanalysis estimates from the ERA5-Land dataset. Figure x illustrates the time series of the spatial average (across 544 basins) for the three variables. All time series were normalized prior to plotting to facilitate shape comparison. As shown, the outputs from the hybrid model (blue lines) closely follow the seasonal and interannual variations of the ERA5-Land estimates (orange dashed lines) across all three variables. This suggests that the model not only provides accurate runoff predictions, but also captures physically plausible hydrological states and fluxes. To quantify this consistency at the basin scale, the Pearson correlation coefficient between the output time series and ERA5-Land reference time series was calculated for each basin and each variable. The median correlation across all basins reached 0.71 for evapotranspiration, 0.64 for soil water, and 0.52 for snowpack.

These results demonstrate the hybrid model's ability to internally simulate key components of the hydrological cycle, even though these variables were never directly used in model training or supervision. The snowpack correlation, although slightly lower, can be attributed to the fact that some basins are located in warm regions with negligible snow accumulation. Moreover, uncertainties in the reanalysis-based "reference" values may also affect the observed correlations. Overall, these findings confirm that the differentiable hybrid modeling framework can generate physically meaningful intermediate states, further supporting its potential for interpretable hydrological modeling and process-informed regional analysis.

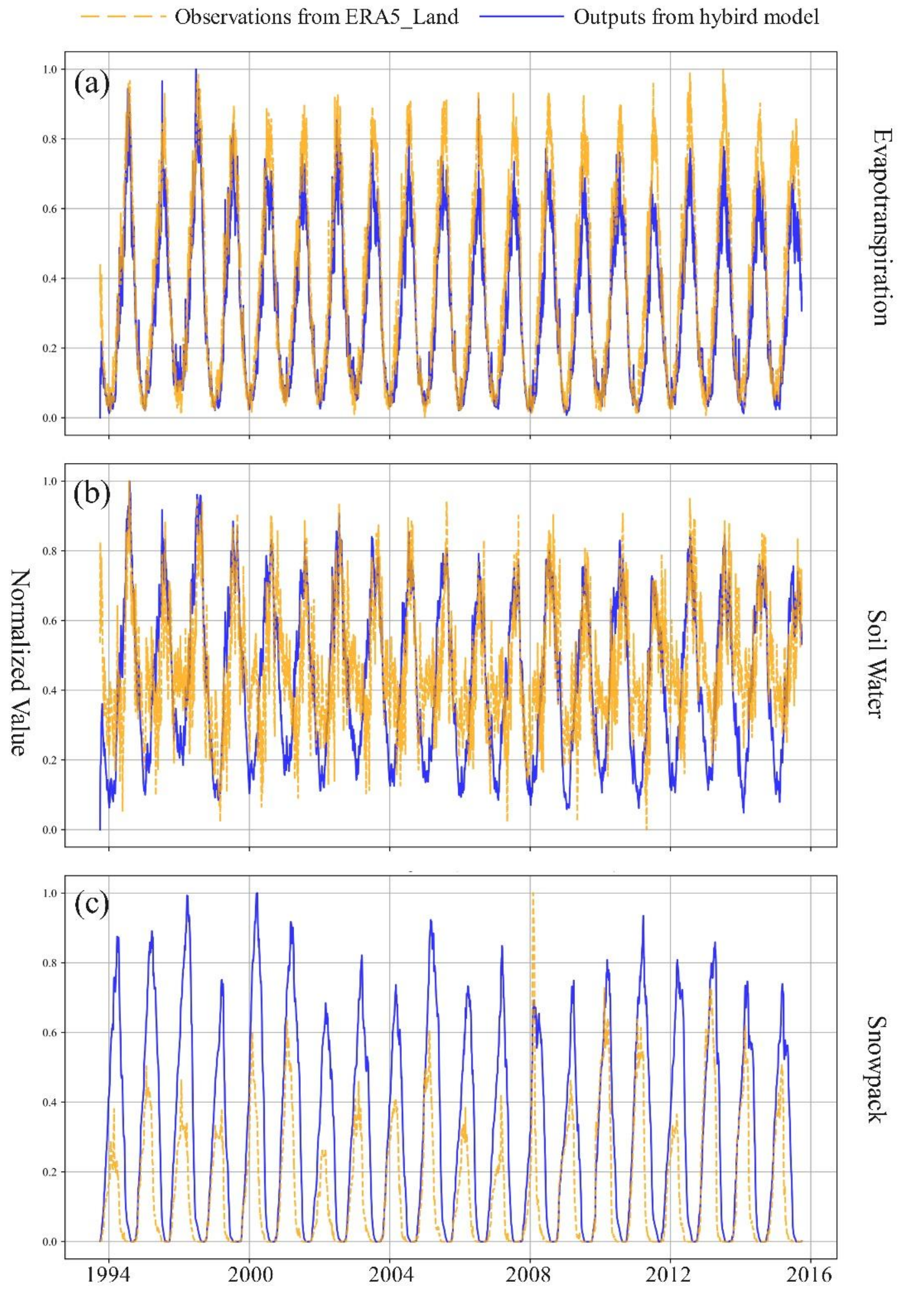

Figure x. Comparison between the outputs of the differentiable hybrid model and ERA5-Land reanalysis data for three intermediate hydrological variables (evapotranspiration, soil water, and snowpack) during the testing period (1995–2015). All values are normalized, and each curve represents the spatial average across 544 basins.

…."

Comment 5: *What does the spatial distribution map in Figure 12 mean? A detailed explanation* *should be given in the image caption.*

Response:

Thank you for your careful review and helpful suggestion. We agree that the caption of Figure 12 lacked sufficient detail to help readers interpret the spatial distribution maps. In the revised manuscript, we have clarified that the maps on the left side of each subpanel display the spatial distribution of NSE values under PUB test for each hybrid model, while the scatter plots on the right compare PUB and regional performances at the basin level. The colors of the dots represent the basin clusters. We have now updated the figure caption as follows:

"….

Figure 12. Comparison of PUB performances of four hybrid models (EXP-IN-LSTM, XAJ-IN-LSTM, EXP-dPL, and XAJ-dPL) under two meteorological datasets (ERA5-Land and CN05.1). For each model and forcing case, the left-side map shows the spatial distribution of the NSE values for each basin under PUB test, with warmer colors indicating lower NSE performance and cooler colors indicating higher performance. The right-side scatter plots compare PUB model performance (y-axis) to regional model performance (x-axis) for each basin. Each point represents a basin, and point colors correspond to different clusters. The dashed 1:1 line serves as a reference for comparing PUB and regional performances.

…."

Comment 6: *Why do you use the runoff predictions for water balance assessment? What is the purpose of calculating the water imbalance ratio? Please add an explanation.*

Response:

Thank you for your insightful reminder. In Section 4.5, we used the runoff predictions and evapotranspiration output by the differentiable hybrid model to calculate the water imbalance ratio. Because for basins with different dry and wet conditions, the imbalance degree of water conservation under the same error term is not exactly the same. For example, because the predicted runoff is 10 mm higher than the difference between precipitation and evapotranspiration, the reliability of the prediction results for basins with daily average precipitation of 50 mm and 1 mm is obviously far apart. Therefore, we chose to use the ratio of the error term to precipitation to measure the physical consistency of the runoff prediction results used in different basins and the reliability of the model output evapotranspiration. The resulting enhanced water balance closure ability shows the potential of differentiable hybrid models in improving the understanding of regional water resources availability. We have followed your suggestion and added relevant explanations. The added explanations are as follows:

"….

The water balance closure is assessed based on the outputs of the differentiable hybrid models, which are able to output predicted runoff and evapotranspiration (ET). To evaluate the physical consistency of these predictions, the sum of model-predicted streamflow and ET is compared against total precipitation during the testing period. To ensure comparability across catchments with varying climatic conditions, the water imbalance ratio (ε/P) is adopted. This ratio is defined as the absolute difference between precipitation and the sum of streamflow and ET, normalized by total precipitation. This ratio reflects the extent to which model outputs satisfy the fundamental water balance constraint. Using this ratio enables a consistent and interpretable measure of physical plausibility across diverse hydrological regimes. A lower imbalance ratio indicates better water budget closure, suggesting that the model not only produces accurate runoff predictions but also maintains hydrological consistency in terms of mass conservation. This evaluation provides insights into the model's ability to simulate key hydrological variables in a physically meaningful way.

…."

Comment 7: *There is an error in the labeling of Figure 11. Two sub-figures (b) appear. Please modify them.*

Response:

Thank you for your valuable correction. We have corrected the subfigure labels to avoid duplication. The Sankey diagram previously labeled as Figure 11(b) has now been relabeled as Figure 11(c), and the subsequent subplots have been updated accordingly. The modified Figure 11 and its caption are as follows:

"….

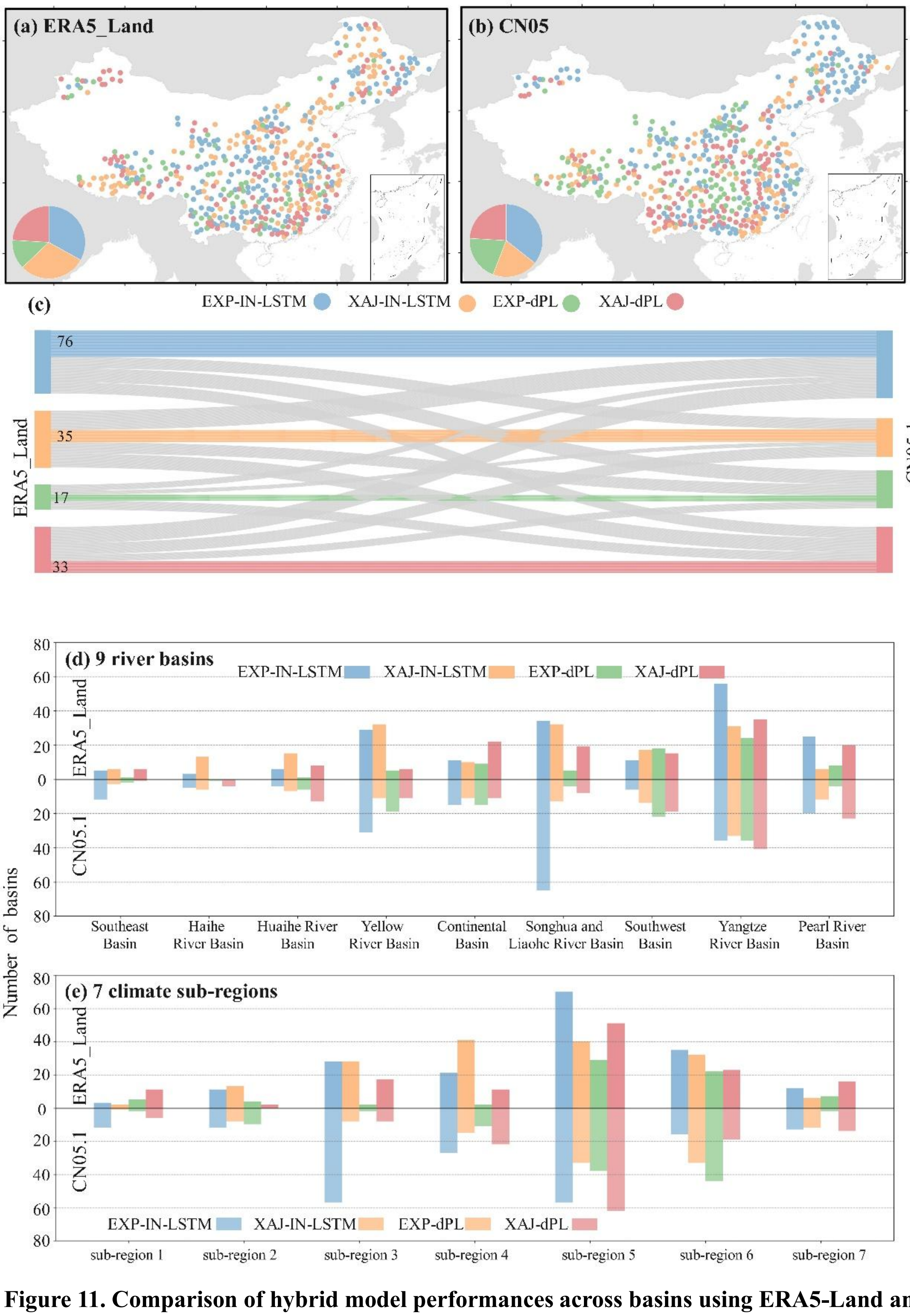

Figure 11. Comparison of hybrid model performances across basins using ERA5-Land and CN05.1 precipitation data. (a) Spatial distribution of the best-performing hybrid models under ERA5-Land forcing. (b) Spatial distribution under CN05.1 forcing. (c) Sankey diagram showing the consistency of best-performing models across datasets. (d) Number of best-performing basins for each model in the 9 major river basins. (e) Number of best-performing basins in the 7 climate sub- regions.

…."

We would like to thank the editors and reviewers once again for their valuable suggestions on our manuscript. We have incorporated these suggestions into the revised manuscript. Looking forward to hearing from you.

Chunxiao Zhang

Corresponding author

E-mail address: zcx@cugb.edu.cn

References:

Gray, L. C., Zhao, L., & Stillwell, A. S., 2023. Impacts of climate change on global total and urban runoff. Journal of Hydrology, 620, 129352. https://doi.org/10.1016/j.jhydrol.2023.129352

Jiang, S., Zheng, Y., Solomatine, D., 2020. Improving AI System Awareness of Geoscience Knowledge: Symbiotic Integration of Physical Approaches and Deep Learning. Geophysical Research Letters 47, e2020GL088229. https://doi.org/10.1029/2020GL088229

Koya, S. R. , & Roy, T. . (2024). Temporal fusion transformers for streamflow prediction: value of combining attention with recurrence. Journal of Hydrology, 637.https://doi.org/10.1016/j.jhydrol.2024.131301

Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A.K., Hochreiter, S., Nearing, G.S., 2019. Toward Improved Predictions in Ungauged Basins: Exploiting the Power of Machine Learning. Water Resources Research 55, 11344–11354. https://doi.org/10.1029/2019WR026065

Nai, C., Liu, X., Tang, Q., Liu, L., Sun, S., & Gaffney, P. P. J. (2024). A novel strategy for automatic selection of cross-basin data to improve local machine learning-based runoff models. Water Resources Research, 60, e2023WR035051. https://doi.org/10.1029/2023WR035051

Nearing, G., Cohen, D., Dube, V., Gauch, M., Gilon, O., Harrigan, S., Hassidim, A., Klotz, D., Kratzert, F., Metzger, A., Nevo, S., Pappenberger, F., Prudhomme, C., Shalev, G., Shenzis, S., Tekalign, T. Y., & Weitzner, D., 2024. Global prediction of extreme floods in ungauged watersheds. Nature, 627(8004), 559-563. https://doi.org/10.1038/s41586-024-07145-1

Zhan, Y. , Guo, Z. , Yan, B. , Chen, K. , Chang, Z. , & Babovic, V., 2024. Physics-informed identification of pdes with lasso regression, examples of groundwater-related equations. Journal of Hydrology, 638. https://doi.org/10.1016/j.jhydrol.2024.131504.

Lehner, B., Verdin, K., Jarvis, A., 2008. New Global Hydrography Derived From Spaceborne Elevation Data. Eos Trans. Am. Geophys. Union 89, 93–94. https://doi.org/10.1029/2008EO100001.

Tan, Xuejin, Liu, B., Tan, Xuezhi, Chen, X., 2022. Long-Term Water Imbalances of Watersheds Resulting From Biases in Hydroclimatic Data Sets for Water Budget Analyses. Water Resour. Res. 58, e2021WR031209. https://doi.org/10.1029/2021WR031209.

Wang, S., Huang, J., Li, J., Rivera, A., McKenney, D.W., Sheffield, J., 2014. Assessment of water budget for sixteen large drainage basins in Canada. J. Hydrol. 512, 1–15. https://doi.org/10.1016/j.jhydrol.2014.02.058.

Luo Y, Ma N, Zhang Y. Divergent vegetation greening's direct impacts on land-atmosphere water and carbon exchanges in the northeastern Tibetan Plateau. Global and Planetary Change. 2025;251: 104825. https://10.1016/j.gloplacha.2025.104825.

Liu H, Wang J. Analyzing the impact of different forms of precipitation (rain and snow) on drought. Journal of Water and Climate Change. 2025;16(3):1119-1135. https://10.2166/wcc.2025.691.

Niu, Z., Liu, Y., Feng, L. et al. Spatio-temporal distribution characteristics of different types of compound extreme climate events in the Yangtze River Basin. Nat Hazards. 2025;121, 9261–9280. https://doi.org/10.1007/s11069-025-07165-8.