# Author comments on CC2- egusphere-2025-1161

***Comment 1:*** *Why did the author only choose two sets of precipitation data, or did the temperature data also come from two sets of data products? A more detailed description of the data source is needed.*

**Response:**

Thank you for your careful review and comments. In our study, we selected two sets of precipitation data and corresponding temperature data, mainly based on the following considerations:

1. Data accuracy and reliability: The two datasets we selected have been widely used in previous studies (Xie et al., 2025; Yu et al., 2025; Ma et al., 2024). Therefore, we believe that these data can provide a solid foundation for our research and provide effective reference for future research.
2. Purpose of comparative analysis: Using more data products for model comparison can indeed improve the credibility of research results. However, due to the length of the article, we selected two data sets in this study. In the future, we will refer to more data sets for more in-depth research.
3. Availability and coverage: We also considered the spatial resolution and temporal coverage of the data when selecting data. At the same time, in order to ensure the fairness of the comparison of different data, we need to consider the consistency of the start and end time of the selected data sets. After comprehensive consideration, we finally selected the ERA5-Land and CN05.1 data sets and ensured that their training time and test time were exactly the same.

We have further described the sources of the data in detail in the manuscript so that readers can understand our data selection process more clearly. The specific contents after modification are as follows:

"…

The meteorological data used in this study are sourced from the ERA5-Land and CN05.1 (Gao et al., 2013) datasets. Both datasets provide multiple meteorological variables, including daily precipitation and 2-meter air temperature, which are used consistently throughout this study. The full list of meteorological forcing elements and their corresponding units is shown in Table 1. These two datasets were selected to extract basin-scale meteorological forcings based on the following considerations:

(1) ERA5-Land dataset: Although previous studies have noted that ERA5-Land data may exhibit certain deviations in East Asia, the dataset has several notable advantages. It offers a wide range of meteorological variables such as precipitation, temperature, radiation, humidity, and wind speed. Additionally, it spans a long historical period at a daily resolution, making it highly suitable for large-sample hydrological modeling across extended time scales.

(2) CN05.1 dataset: This gridded dataset is interpolated from over 2,400 national meteorological stations across China. In addition to precipitation, it provides high-resolution air temperature data, making it suitable for regional-scale climate analysis. Owing to its dense observational basis, CN05.1 is considered more accurate in reflecting local meteorological trends and spatial heterogeneity.

To ensure consistency, each model experiment used both precipitation and temperature from the same data source. ERA5-Land precipitation was used alongside ERA5-Land temperature, and similarly for CN05.1. These variables were aggregated to the catchment scale using area-weighted averaging. By incorporating both a global reanalysis product (ERA5-Land) and a regionally calibrated observational product (CN05.1), this study aims to evaluate the robustness of hydrological models under different meteorological forcing conditions, and to examine the sensitivity of model performance to data source selection. This dual-dataset strategy also provides useful insights for regions where observational data may be sparse or incomplete.

…"

***Comment 2:*** *The description of the hybrid model structure in Section 3.4 is confusing. Please try to describe the operating logic of the two hybrid models separately.*

**Response:**

Thank you for your constructive feedback. We agree that the original description of the hybrid model structures in Section 3.4 could be clearer. In the revised manuscript, we have reorganized and rewritten this section to separately and more explicitly describe the operating logic of the alternative hybrid model and the differentiable hybrid model. The modified content is as follows：

"…

Both alternative hybrid modeling and differentiable hybrid modeling schemes are designed to combine the advantages of process-based models (PBMs) and deep learning (DL) models. The architectures of these two types of models are illustrated separately in Figure 4, and their working mechanisms are described below.

(1) Alternative hybrid modeling scheme

In this approach (Figure 4a), the PBMs (EXP or XAJ) are used first to simulate runoff based on meteorological inputs. The LSTM model then serves as a post-processing tool, taking both the PBM-simulated runoff and additional inputs—including the original meteorological forcing and static basin attributes as inputs. The LSTM is trained to learn and correct the discrepancies between the PBM outputs and the target runoff data. This method leverages the LSTM's ability to capture residual nonlinear relationships, thereby compensating for limitations of the PBMs in representing complex processes across large-sample and diverse basins.

(2) Differentiable hybrid modeling scheme

In this approach (Figure 4b), the discrete ordinary differential equations that define the hydrological processes in the PBMs are encoded directly into recurrent neural network (RNN) units, allowing the model to be fully differentiable. Static basin attributes are incorporated into the framework through a neural network-based parameterization scheme, which dynamically generates hydrological parameters. This allows the model to adapt the physical parameter values based on basin characteristics, overcoming the limitations of traditional PBMs that rely on fixed parameters. The model is trained end-to-end using backpropagation through time, enabling joint optimization of both hydrological states and parameters. In addition to discharge simulation, this framework also enables the output of intermediate hydrological variables, such as soil moisture and evapotranspiration, facilitating process-level interpretation and diagnostics.

…”

***Comment 3:*** *Line251: Which of the 6 categories the 15 attributes belong to needs additional explanation, or should be added to Table 2.*

**Response:**

Thank you for your suggestion. We neglected to explain the attribute categories of each watershed. Your suggestion is very valuable. It is more intuitive and clear to mark the attribute categories directly in Table 2. We have revised Table 2 in the new manuscript. The specific content is as follows "…

**Table 2** Static basin attributes data for 544 basins.

| Attribute | Categories | Description | Unit | Source |
|---|---|---|---|---|
| area | topography | Basin area | km2 | This study |
| srftopo | topography | Surface (rock + ice) elevation | m | Amante and Eakins (2009) |
| slope_avg | topography | Mean subgrid slope (inner slope) | m/m | Amante and Eakins (2009) |
| wcap | Soil | Maximum soil water capacity | $Kg/m^2$ | Hagemann and Stacke (2015) |
| wava | Soil | Plant available water | $Kg/m^2$ | Hagemann and Stacke (2015) |
| Fveg | vegetation | Fractional vegetation cover climatology relative to LSM | / | Hagemann (2002) |
| Lai | vegetation | Leaf area index | $m^2/m^2$ | Hagemann (2002) |
| p_mean | climate | Mean daily precipitation | m | This study |
| pet_mean | climate | Mean daily potential evapotranspiration | m | This study |
| aridity | climate | Ratio of Mean PET to Mean Precipitation | - | This study |
| frac_snow | climate | Fraction of precipitation falling on days with temp < 0 ◦C | - | This study |
| high_prec_freq | climate | Frequency of days with ≤ 5× mean daily precipitation | - | This study |
| high_prec_dur | climate | Average duration of high precipitation events | - | This study |
| low_prec_freq | climate | Frequency of dry days (< 1 mm/day) | - | This study |
| low_prec_dur | climate | Average duration of dry periods | - | This study |

…"

***Comment 4:*** *Line455: The author mentioned here the accuracy of climate characteristics and rainfall data. Among the 15 attributes, which meteorological data product is used to calculate "p_mean", "pet_mean", etc., or did the author use other methods?*

**Response:**

Thank you for your reminder. Regarding the watershed attribute data used in the study, we would like to make the following clarifications:

1. All watershed attribute data belonging to climate types are calculated from the meteorological time series of each watershed (1975.10.1-2015.9.30), including: "p_mean", "pet_mean", "aridity", " frac_snow", "high_prec_freq", "high_prec_dur", "low_prec_freq", "low_prec_dur". The calculation of these attributes is consistent with their descriptions.
2. There are actually two sets of climate-related watershed attributes, corresponding to the two datasets used in the study. When the data provided by ERA5-Land is used for model training

and testing, the watershed attributes are also calculated based on the ERA5-Land data; when the data provided by CN05.1 is used for model training and testing, the watershed attributes are also calculated based on the CN05.1 data (except "pet_mean" and "aridity").

3. Since the CN05.1 dataset does not provide potential evapotranspiration data, when the data provided by CN05.1 is used for model training and testing, "pet_mean" and "aridity" are calculated based on the ERA5-Land data.

To clarify this issue, we have added the following description in the 2.3 Static catchment attributes of the revised manuscript:

"…

In this study, catchment attributes related to climatic characteristics—specifically including p_mean, pet_mean, aridity, frac_snow, high_prec_freq, high_prec_dur, low_prec_freq, and low_prec_dur—were calculated from the corresponding meteorological forcing time series during the training and testing periods (1975/10/01–2015/09/30).To ensure consistency, we generated two sets of climate-related attributes, corresponding to the two meteorological datasets used in this study (ERA5-Land and CN05.1). When a model is trained and tested using ERA5-Land forcing, its associated catchment attributes are also calculated based on ERA5-Land data; similarly, when CN05.1 is used as the forcing, the climate-related attributes are calculated from the CN05.1 time series.

A specific exception is made for pet_mean and aridity, as the CN05.1 dataset does not directly provide potential evapotranspiration (PET). In these cases, PET-related attributes are consistently calculated based on the ERA5-Land dataset, even when CN05.1 is used for precipitation and temperature. This compromise ensures both the availability and consistency of these critical attributes while maintaining reasonable comparability between the two modeling configurations.

…"

***Comment 5:*** *The author uses the Budyko curve to examine the watershed water balance in Section 4.1, while in Section 4.5, the water budget closure method is employed. Why are different methods used to verify the watershed's water balance situation?*

**Response:**

Thank you for your careful review. As you said, we used different methods to measure the water balance of the basin in Section 4.1 and Section 4.5. This is because we have different purposes in measuring the water balance at the basin scale in the two stages of the study, so we need to use appropriate methods. Specifically:

In Section 4.1, we further explored the differences and deviations between the two data sets after extracting the meteorological and runoff data of each basin in order to evaluate the number of basins that violate the water-heat balance when using different meteorological data. At this time, the meteorological and runoff time series used include the complete training period and test period (1975-2015).

The water balance evaluation work in Section 4.5 is mainly to verify the physical consistency of the prediction results of different models. Therefore, the runoff data used are the predicted values output by each model. At the same time, in order to avoid the unfairness of the absolute size of the water balance error terms of different dry and wet basins, we added the water imbalance ratio to measure the physical consistency of the prediction results of different models. At this time, the meteorological and runoff time series used only include the test period (1995-2015).

In order to better explain why we use different methods to measure water balance and avoid causing similar confusion to readers, we have added corresponding explanations in Sections 4.1 and 4.5. The modified details are as follows:

"…

4.1 Meteorological forcing assessment

…

The Budyko framework is employed to assess the overall water balance behavior of each basin over the full period (1975–2015), based on precipitation, potential evapotranspiration, and runoff data. This analysis aims to evaluate the consistency and deviation patterns between meteorological forcing and runoff datasets under different data sources. Specifically, it provides a diagnostic tool for detecting basins with potentially unbalanced water budgets, which may indicate issues in either meteorological forcing or runoff simulations. Therefore, the Budyko curve here serves as a reference framework for identifying data-driven inconsistencies across basins and forcing products.

…

4.5 Evaluation of water budget closure

The water balance assessment in this section is focused on evaluating the physical consistency of the model-predicted runoff during the testing period (1995–2015). The water budget closure analysis is used to compare precipitation, model-simulated runoff, and potential evapotranspiration for each model, aiming to quantify the degree of water balance closure in the model outputs. To account for differences in hydrological regimes, a water imbalance ratio (ε/P) is adopted as a metric to ensure comparability across basins. This approach provides insight into the process realism and hydrological plausibility of each model's predictions.

…"

***Comment 6:*** *There are still many available high-quality meteorological data products. I can understand the author's decision to limit the scope of the article to control its length. However, this needs to be clarified in the conclusion section of the article.*

**Response:**

Thank you very much for your understanding and reminder. As you pointed out, there are many high-quality meteorological data products available. Our research selected two data products for relevant experiments and analysis after considering the time span and resolution. In the future, there will be more other high-quality meteorological data products. Therefore, our research focuses more on providing references by playing the role of cases. Your suggestion is very pertinent, and we have added an explanation of this point in the conclusion. The specific content is as follows:

"…

It is worth noting that although this study employs only two widely-used meteorological datasets, there exist many other high-quality meteorological products that could also support large-sample hydrology modeling. The choice to limit our analysis to these two products was based on considerations of data accessibility, spatial and temporal resolution, and the desire to maintain clarity and focus in model comparison. Future work may incorporate a broader set of meteorological forcings to further evaluate the robustness and generality of hydrological model performance across varying data sources. This study thus serves as an initial benchmark for such efforts in China, providing a reference framework for subsequent research.

…."

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:


Addor, N., Newman, A.J., Mizukami, N., Clark, M.P., 2017. The CAMELS data set: catchment attributes and meteorology for large-sample studies. Hydrol. Earth Syst. Sci. 21, 5293–5313. https://doi.org/10.5194/hess-21-5293-2017

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

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

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

Xie Y, Kong D, Zhang Y, et al.Evaluation of the performance of multiple reanalysis forcing data in potential evapotranspiration estimation and its implication for actual evapotranspiration modeling.Journal of Hydrology[2025-07-07]. https://doi.org/10.1016/j.jhydrol.2025.133472.

Yu X, Zhang Q, Zeng X, et al.The distribution and driving climatic factors of agricultural drought in China: Past and future perspectives.Journal of Environmental Management, 2025, 377. https://doi.org/10.1016/j.jenvman.2025.124599.

Wang Z, Li M, Zhang X,et al.Prediction of long-term future runoff under multi-source data assessment in a typical basin of the Yangtze River.Journal of Hydrology: Regional Studies, 2024, 56(000). https://doi.org/10.1016/j.ejrh.2024.102053.

Ma X, Wang A. Evaluation and Uncertainty Analysis of the Land Surface Hydrology in LS3MIP Models Over China.Earth & Space Science, 2024, 11(7). https://doi.org/10.1029/2023EA003391.

Miao Y, Wang A. Evaluation of Routed-Runoff from Land Surface Models and Reanalyses Using Observed Streamflow in Chinese River Basins.Journal of Meteorological Research, 2020, 34(1):73-87. https://doi.org/10.1007/s13351-020-9120-z.