Author comments on RC2- egusphere-2025-1161

Comment 1 (About runoff data):

The runoff data product is actually simulated with VIC, a process based model, which uses one of the two meteorological datasets (the CN05.1) as input. In general, the uncertainties in the VIC data product that is being used as a proxy for "observed data" in terms of model training and evaluation should be brought up more clearly. This seems relevant in several places in the study. For example, the finding that the CN05 product leads to better performance is mainly because it is embedded in the flow data. This is explicitly mentioned in Line 470, and I would say that is extremely likely that it is what is happening. Meanwhile, there are other places in the paper where it is posed that CN05 must be the superior forcing product because it leads to better model results. For example, Line 32 in the abstract cites that "high-quality precipitation data better enables the model to simulate runoff processes". While this is surely true in general, I think the results of this study reflect the fact that one of the products was used to generate the original flow data, making it a more unfair comparison. With this, I'm not sure what the significance of comparing between the two meteorological products is here – since whichever product is used to generate the flow data product is likely to be the most useful input to another model that is trained to emulate the flow data. It would be better to bring in a third met product (or more) to really compare this, or drop this aspect of the study and focus on differences between models based on a single forcing product.

Response:

We thank you for this insightful and important comment, which raise a fundamental and valid concern regarding the experimental setup. The use of VIC-CN05.1 simulated streamflow as the evaluation target, despite being a practical choice under data constraints, does introduce potential biases, particularly when comparing models forced by CN05.1 versus ERA5-Land. The original rationale behind our experimental design was twofold:

First, the objective was to evaluate the performance and robustness of different hydrological modeling approaches (process-based, deep learning, and hybrid) over a large number of catchments in China. To do this consistently, we required a streamflow dataset that covers a broad spatial extent (544 basins) with unified meteorological forcing and standardized catchment boundaries. Due to limitations in the availability and accessibility of observed streamflow records across China, especially with regard to consistency in spatial delineation and temporal overlap, observed data could not be used to construct a large-sample dataset of sufficient quality. As a result, we adopted the VIC-CN05.1 runoff product, widely used in the Chinese hydrological community (Miao et al., 2020; Ma et al., 2024; Wang et al., 2024; Yu et al., 2025), as a surrogate 'reference' dataset. Second, the study aimed to examine whether different models would exhibit consistent relative performance rankings under different meteorological forcings. This comparison was designed not to determine which meteorological product is superior in absolute terms, but to assess the robustness of model performance across varying inputs. Nevertheless, we fully agree with the reviewer that the use of CN05.1 in both the VIC simulation and some of the model inputs can lead to biased performance comparisons that favor CN05.1-driven models due to internal consistency between input and evaluation target.

We acknowledge that this issue you raise was not sufficiently emphasized in the original manuscript. In the revised version, we have addressed this concern by clarifying the nature of the runoff product in the data section, explicitly stating that it is a simulated product driven by CN05.1 rather than observational data. Additionally, we have revised the abstract and conclusion to remove overly strong claims regarding the superiority of CN05.1, instead emphasizing the potential bias introduced by using CN05.1-driven runoff as the target. Furthermore, we have included a dedicated paragraph in the Discussion section to highlight this structural bias and discuss its implications for interpreting the model performance results, cautioning readers against over-interpreting the advantages of CN05.1. Lastly, we clarified the limited scope of the meteorological forcing comparison, noting that the findings do not necessarily indicate the real-world superiority of one forcing product over another, but rather reflect the internal consistency within the data generation process.

While we fully acknowledge that incorporating additional independent meteorological forcing products would enhance the comprehensiveness and generalizability of the results, such an extension would significantly increase the scope and complexity of the current study. Given our primary objective to establish a consistent baseline framework using the two most widely applied precipitation products in China, we chose to limit the comparison to CN05.1 and ERA5-Land in this initial analysis. Nevertheless, we consider the inclusion of more diverse meteorological forcings to be a valuable direction for future research, particularly for evaluating model robustness under broader climatic variability.

In short, we agree with your comments. The details of our revised content are as follows:

"…

About Data: used to clarify the sources and characteristics of runoff data

The streamflow data used for model training and evaluation in this study are derived from the VIC-CN05.1 runoff product, which was generated by the VIC hydrological model driven by CN05.1 precipitation. It should be noted that this is not observational streamflow data, but a simulated product that serves as a consistent and spatially complete surrogate in the absence of publicly available observed daily streamflow records for a large number of catchments in China. While VIC-CN05.1 has been evaluated in previous studies, its use as a reference introduces potential structural bias, particularly when comparing models forced by the same precipitation product.

About Abstract: modify the conclusion in line 32

The results suggest that precipitation data quality plays a critical role in runoff simulation. However, since the runoff data used as the evaluation target were generated using CN05.1, the apparent performance advantages of models driven by CN05.1 may partly reflect internal consistency rather than inherent superiority of the forcing product.

About Conclusion: correspond to the abstract

The CN05.1-driven models showed relatively better performance, which may be influenced by the source of runoff data. Therefore, this finding should be interpreted with caution, as it may reflect structural consistency rather than intrinsic superiority.

About Discussion: add discussion corresponding to runoff data

One limitation of this study lies in the use of a simulated runoff product (VIC-CN05.1), generated using the CN05.1 precipitation data, as the reference for model evaluation. This introduces a structural bias that may favor models driven by CN05.1 due to internal consistency between inputs and targets. As a result, the observed performance advantage of CN05.1-driven models may not necessarily indicate the intrinsic quality of the precipitation product. The meteorological forcing comparison in this study should therefore be interpreted as an exploration of input-output consistency effects rather than a definitive evaluation of forcing product accuracy. Future studies incorporating additional meteorological datasets and observed streamflow records will be important for validating and extending these findings.

…”

Comment 2: *It is easy to get confused between model names. For example, authors could use subscripts on "EXP" and "XAJ" to indicate the process based, alternative hybrid, and differentiable hybrid model versions.*

Response:

Thank you for your helpful reminder. We agree that the original naming conventions for the models, particularly for EXP and XAJ in their different versions (process-based, alternative hybrid, and differentiable hybrid), may cause confusion for the reader. In the revised manuscript, we have adopted a consistent notation using subscripts to distinguish the different versions of each model. For example, we now denote the process-based version as EXP_{p}, the hybrid version as $EXP_{IN\text{-}LSTM}$, and the differentiable hybrid version as EXP_{dPL} (similarly for XAJ). These changes have been applied throughout the manuscript, including in figure legends, tables, and the methods section, to enhance clarity and facilitate comparison across models.

Comment 3: *Figure 4: Figure 4a makes it look like only VP is an input to the LSTM but I think it should be all the forcing variables? As with other figures more description here would be useful.*

Response:

Thank you for pointing this out. We agree that the current visual layout of Figure 4a may give the misleading impression that only vapor pressure (VP) is used as input to the LSTM, while in fact all meteorological forcing variables (precipitation, temperature, radiation, and vapor pressure) are jointly used as inputs to the LSTM. To address this, we have updated the figure to adjust the arrow connections and improve visual clarity. In addition, we have revised the figure caption and added a brief explanatory paragraph in the main text to provide a more detailed description of the model structure illustrated in both panels (a) and (b). These changes aim to help readers better understand the input structure and model components.

The modified Figure 4 is as follows:

"…

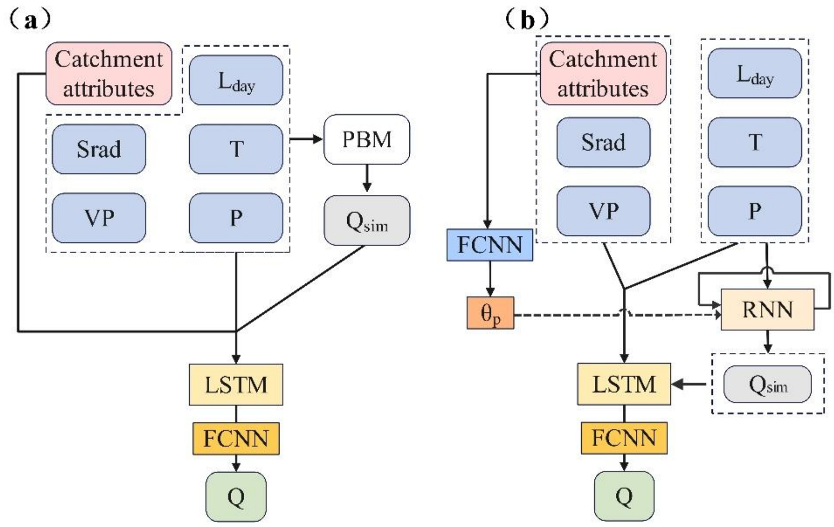

Figure 4. Structure of the hybrid hydrological models. (a) A standard hybrid model in which meteorological forcings (precipitation, temperature, radiation, and vapor pressure) and catchment attributes are used to train a data-driven model (LSTM-FCNN). The PBM component provides simulated streamflow (Q_sim) based on the meteorological forcings. (b) A differentiable hybrid model where catchment attributes are used to generate dynamic hydrological parameters via a fully connected network (FCNN), which are then passed to a differentiable process-based model. The process-based outputs (Q_sim) are fused with LSTM predictions for the final discharge estimate (Q)

…"

Comment 4: *Figure 5: the legends show the colors as continuous but the dots in the maps make it seem discrete (that there are 5 colors). It would be better if the legend reflected the ranges for these 5 colors.*

Response:

Thank you for the helpful suggestion. We agree that the current figure design, which uses continuous color bars alongside discretely colored basin points, could cause confusion. To address this, we have revised Figure 5 by adjusting the legends to reflect discrete color bins that correspond to the categories used in the maps. The updated legend now clearly shows the value ranges associated with each color, which improves the consistency and interpretability of the figure. The modified Figure 5 and its caption are as follows:

"…

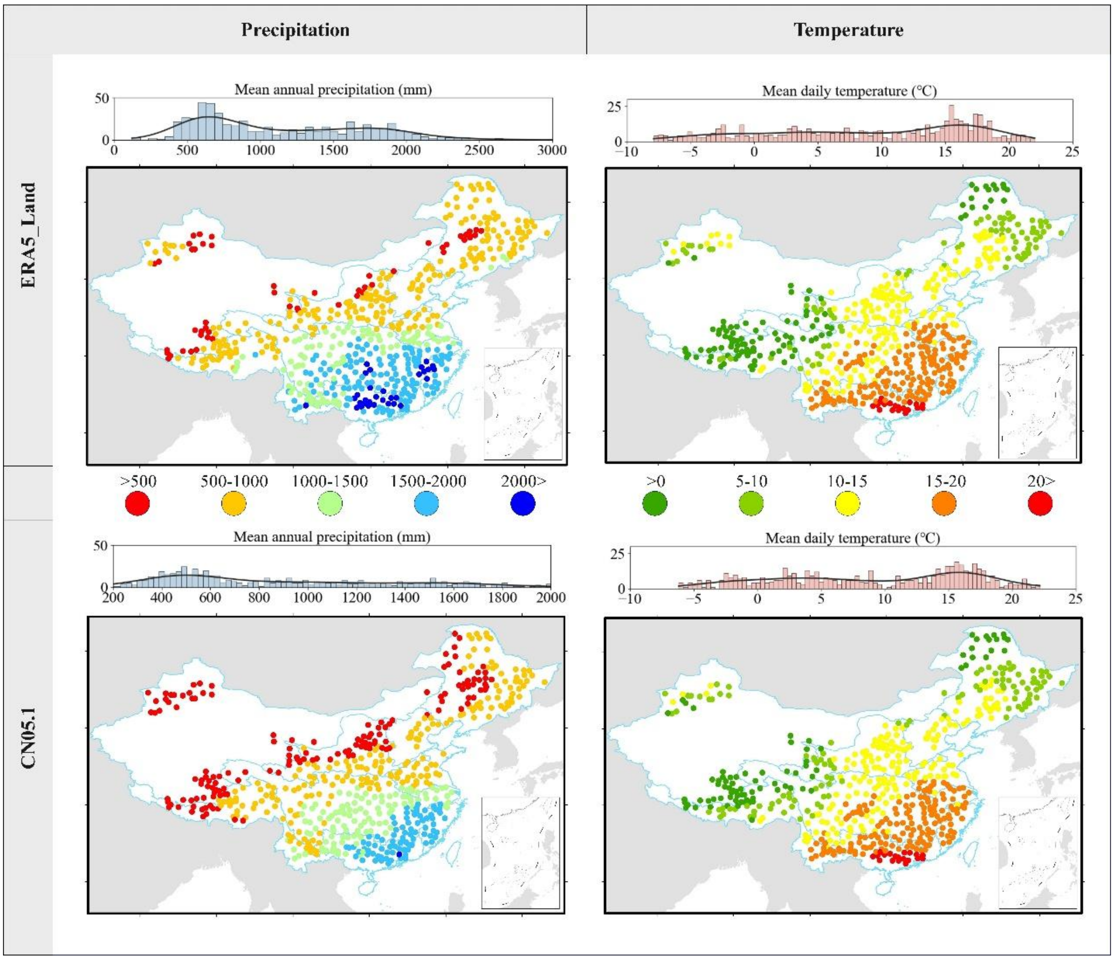

Figure 5. Spatial distribution of five hydro-climatic categories of basin-averaged precipitation and temperature (1 Oct 1975–30 Sep 2015). (a) Mean annual precipitation (mm): Top histograms display the continuous distribution of basin means. Maps use five discrete classes: <500, 500 to 1000, 1000 to 1500, 1500 to 2000, and >2000 mm, applied uniformly to both ERA5-Land (top row) and CN05.1 (bottom row). (b) Mean daily temperature (°C): Top histograms display the continuous distribution of basin means. Maps use five discrete classes: <5, 5 to 10, 10 to 15, 15 to 20, and >20 °C, applied identically to both datasets.

…"

Comment 5: *Figure 9: The color scales are all different, so it makes it hard to compare between the four panels. Same in Figure 10 – why are the color scales different for NSE for the two different forcing cases?*

Response:

Thank you for the reminder. We acknowledge that using consistent color scales across all panels would improve direct visual comparability. However, due to the differences in the value ranges and distributions of NSE under ERA5-Land and CN05.1 forcings, we chose to use separate color scales to better highlight the spatial variability and performance differences within each specific scenario. Unifying the color scales would compress the visual contrast in certain panels and potentially obscure important spatial patterns. We have added clarifying notes in the revised figure captions to explicitly state that the color scales are not unified, and we provide the respective NSE ranges for each panel. This should help readers interpret the visualizations more accurately and avoid misinterpretation. The revised Figures 9 and 10 and their captions are as follows:

"…

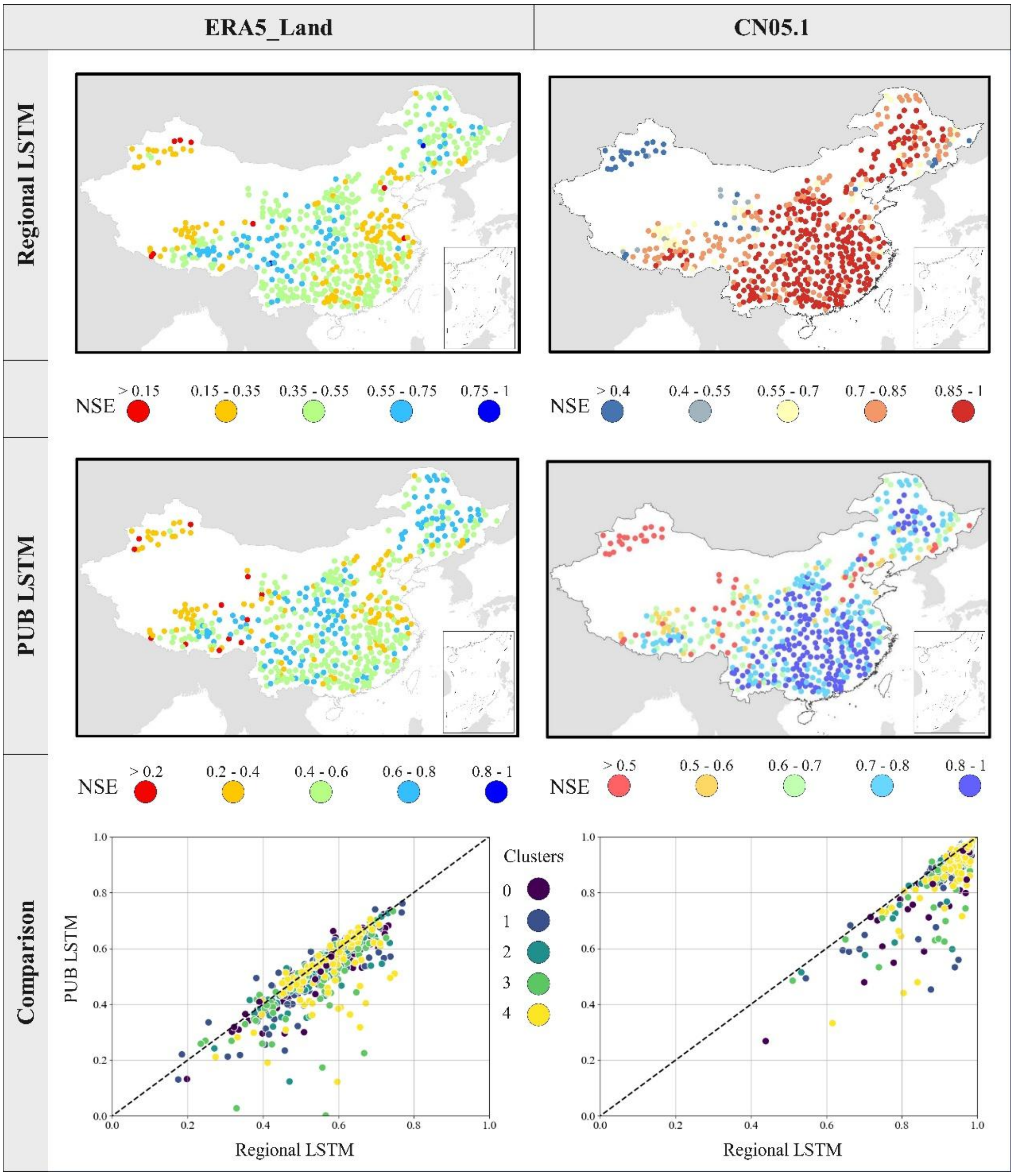

Figure 9. Performance of LSTM models using ERA5 (left column) and CN05.1 (right column) precipitation during regional modeling and PUB testing. Top row (maps): Spatial distribution of NSE during regional modeling. Middle row (maps): Spatial distribution of NSE during PUB testing. Bottom row (scatter): Basin-by-basin comparison of PUB NSE (vertical axis) vs. regional NSE (horizontal axis). Points are colored by clusters. The axes both span [0, 1].

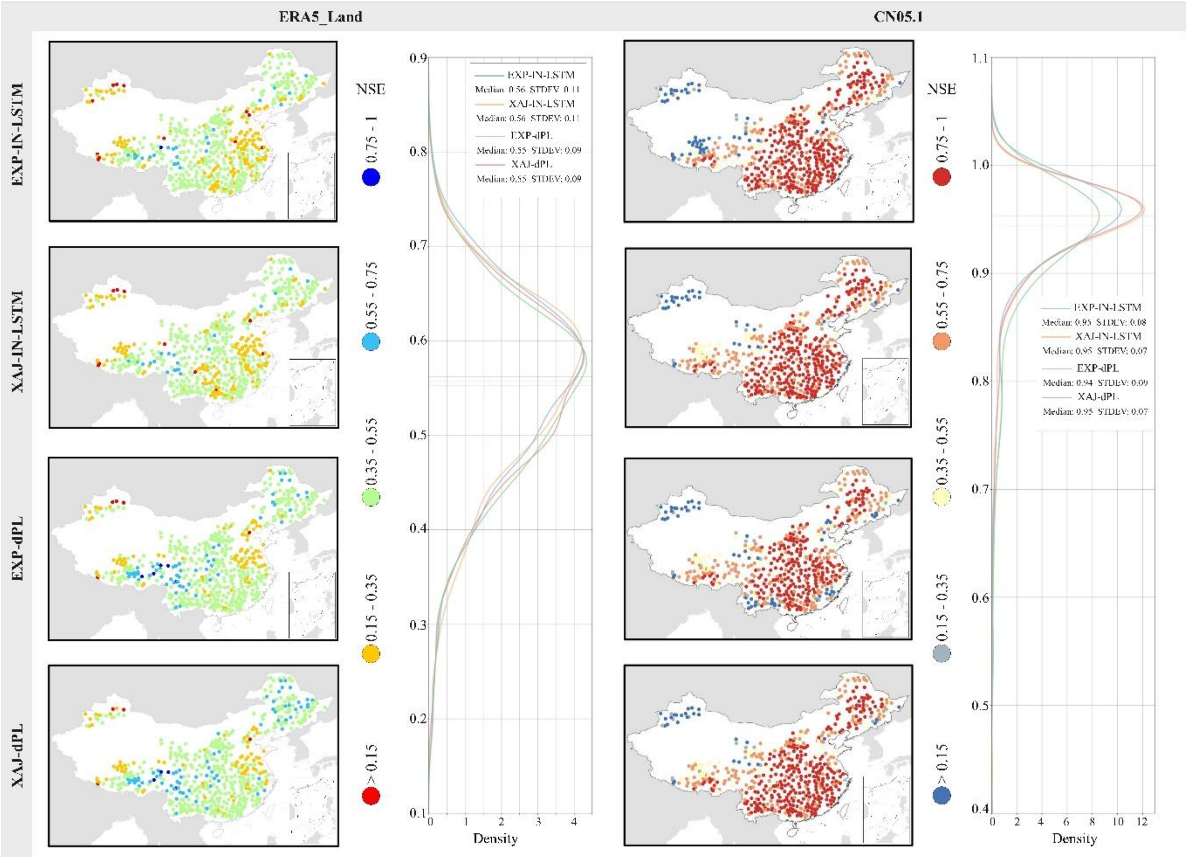

Figure 10. Four hybrid models prediction performances using ERA5-Land (left) and CN05.1 (right) precipitation data. Spatial maps: Colorbars cover ranges of NSE values for different models. Density plots: x-axes display the NSE values; y-axes are density.

…."

Comment 6: *Line 230: think there is a word missing "provided by the originates"*

Response:

Thank you for your correction. We have corrected the sentence to read:

"…

The runoff data used in this study originates from the VIC-CN05.1 dataset (Miao and Wang, 2020), which is consistent with the total runoff simulated by the Global Runoff Data Center (UNH/GRDC).

…"

We have carefully reviewed the manuscript for similar wording issues and revised them to improve clarity and readability.

Comment 7: *Line 421: cross to across*

Response:

Thank you for pointing out this typo. We have corrected “cross different regions” to “across different regions” in the revised manuscript.

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:

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

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

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

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

Patil, S., Stieglitz, M., 2014. Modelling daily streamflow at ungauged catchments: what information is necessary?: MODELLING DAILY STREAMFLOW AT UNGAUGED CATCHMENTS. Hydrol. Process. 28, 1159–1169. https://doi.org/10.1002/hyp.9660

Ren-Jun, Z., 1992. The Xinanjiang model applied in China. J. Hydrol. 135, 371–381. https://doi.org/10.1016/0022-1694(92)90096-E

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

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