This manuscript introduces an explainable deep learning (EDL) model that merges the Xin'anjiang (XAJ) hydrological model's principles with recurrent neural network (RNN) units for flood simulation. The EDL model, tested in two Chinese river basins, outperforms benchmark models with high accuracy, demonstrating how incorporating physical constraints into deep learning architectures can enhance both simulation accuracy and model interpretability for hydrological modeling. These findings contribute to improved flood forecasting capabilities, offering a promising approach for combining traditional hydrological knowledge with advanced machine learning techniques. But several critical issues require attention.

Points for the Authors to Consider

1. Novelty and Positioning within Existing Literature

The integration of physical models with deep learning has evolved rapidly, with existing approaches broadly categorized into loose coupling (e.g., using physical constraints in loss functions, post-processing physical model outputs with DL) and tight coupling (e.g., embedding physical equations into neural network architectures) mentioned in this manuscript. While the manuscript briefly mentions some loose coupling methods, it overlooks key advancements in both paradigms. For instance, prior studies have initialized DL model weights using physical simulations (Read et al., 2019; Yang et al., 2020), combined offline-trained DL modules with physics-based models (Li et al., 2014). For tight coupling frameworks, differentiable models (DM) (Shen et al., 2023) demonstrate how end-to-end training enables deep integration of physical and neural components, going beyond parameter learning to include module replacement and hybrid architectures. This spectrum of tight coupling approaches is further illustrated by physics-embedded models like the Mass-Conserving LSTM (Hoedt et al., 2021) and RNNs with Mass-Conserving-Perceptron (MCP) (Wang and Gupta, 2024), which directly incorporate conservation laws into network architectures.

Another innovative approach involves using physics-based equations as the fundamental time units within RNN structures (He et al., 2024). Notably, many contemporary studies employ hybrid coupling strategies that combine both loose and tight coupling elements. For instance, DM frameworks frequently utilize parameter learning while simultaneously replacing specific physics-based modules with neural network subcomponents (Feng et al., 2022).

For this manuscript, the proposed EDL model appears functionally similar to existing hybrid approaches, particularly the physics-embedded models with some neural network modules to postprocess the results (Jiang et al., 2020). Without a clearer differentiation from these methods—such as explicating how the XAJRNN layer uniquely preserves hydrological parameters or interacts with LSTM modules—the claimed novelty remains ambiguous. A more thorough literature review and direct comparisons with recent physics-guided DL frameworks would better contextualize the work's contributions.

2. Methodological Transparency and Reproducibility

The manuscript lacks critical details about the XAJRNN implementation. For example, the mathematical formulation linking Nash unit hydrographs to neural network operations (Appendix A, Eqs. A30–A32) is insufficient to reconstruct the model. Key questions remain unresolved: How are XAJ parameters (e.g., tension water capacity) represented in the hybrid architecture? What specific components of the XAJ model are preserved versus replaced by DL modules? Without open-source code or a complete algorithmic description, the reproducibility and reliability of the results are compromised.

The post-processing modules (normalization and LSTM layers in Figure 4) also warrant deeper analysis. Their impact on model performance—particularly in extreme events where the EDL model shows better performance (Tables 1–2)—is unclear. Ablation studies comparing the full EDL model to standalone XAJRNN and post-processing components would clarify their relative contributions.

3. Validation and Generalizability

The evaluation limited to two basins raises concerns about generalizability. Deep learning models typically require diverse datasets to demonstrate robustness, yet the study does not test the EDL model in data-scarce regions or basins with contrasting hydrological regimes. Expanding the validation to additional basins and explicitly analyzing performance across flood magnitudes (e.g., low, medium, and extreme flows) would strengthen the claims.

Specific Comments

1. Line 43 contains a typographical error where "deep." appears. This should be corrected to "deep".

2. In lines 78-81, the final example in this paragraph would benefit from improved readability. Adding a transitional conjunction like "but" before the final point would enhance the flow of the argument.

3. At line 90, where the abbreviation "EDL" first appears in the manuscript, it should be properly introduced by providing its full name, "explainable deep learning", before using the abbreviated form.

4. Line 301 contains a subject-verb agreement error in the phrase "The evaluation metrics includes". The verb should be corrected to "include".

5. The reference to "high flow range" at line 322 lacks quantitative specificity. This should be made more precise by providing a numerical threshold, such as "flows larger than [specific value] m^3/s".

6. Several instances throughout the text (lines 335, 339, 340, 342, 390, 391, 393, 404, 407, 409) incorrectly use "DEL" when referring to the model. These should all be

corrected to "EDL" for consistency. A thorough check of the entire manuscript for this error is recommended.

7. Tables 2 and 3 present negative values for the ΔT (timing error) metric without sufficient explanation. The authors should clarify the hydrological significance of these negative values in the context of flood peak timing predictions.

8. The manuscript currently lacks proper reference to Tables 2 and 3 in the text. When first discussing results from these tables, explicit references such as "as shown in Table 2" or "see Table 3" should be included to guide readers.

9. At line 388, there appears to be confusion between "XAJRNN" and "XAJ". These are distinct entities and should not be used interchangeably. The text should clarify their relationship and differences.

Conclusion

The EDL model represents a promising step toward integrating physical hydrology with deep learning. However, the manuscript currently understates its methodological limitations and overstates its novelty relative to existing hybrid approaches. Addressing these issues—through expanded validation, methodological transparency, and clearer positioning within the literature—would significantly enhance its contribution to the field.

Recommendation: Major revision required.

References

Feng, D., Liu, J., Lawson, K., Shen, C., 2022. Differentiable, Learnable, Regionalized Process-Based Models With Multiphysical Outputs can Approach State-Of-The-Art Hydrologic Prediction Accuracy. Water Resour. Res. 58, e2022WR032404. https://doi.org/10.1029/2022wr032404

He, L., Shi, L., Song, W., Shen, J., Wang, L., Hu, X., Zha, Y., 2024. Synergizing Intuitive Physics and Big Data in Deep Learning: Can We Obtain Process Insights While Maintaining State-Of-The-Art Hydrological Prediction Capability? Water Resour. Res. 60, e2024WR037582. https://doi.org/10.1029/2024WR037582

Hoedt, P.-J., Kratzert, F., Klotz, D., Halmich, C., Holzleitner, M., Nearing, G.S., Hochreiter, S., Klambauer, G., 2021. MC-LSTM: Mass-Conserving LSTM, in: Meila, M., Zhang, T. (Eds.), Proceedings of the 38th International Conference on Machine Learning, Proceedings of Machine Learning Research. PMLR, pp. 4275–4286.

Jiang, S., Zheng, Y., Solomatine, D., 2020. Improving AI System Awareness of Geoscience Knowledge: Symbiotic Integration of Physical Approaches and Deep Learning. Geophys. Res. Lett. 47, e2020GL088229. https://doi.org/10.1029/2020gl088229

Li, Z., Kan, G., Yao, C., Liu, Z., Li, Q., Yu, S., 2014. Improved Neural Network Model and Its Application in Hydrological Simulation. J. Hydrol. Eng. 19, 04014019. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000958

Read, J.S., Jia, X., Willard, J., Appling, A.P., Zwart, J.A., Oliver, S.K., Karpatne, A., Hansen, G.J.A., Hanson, P.C., Watkins, W., Steinbach, M., Kumar, V., 2019. Process-Guided Deep Learning Predictions of Lake Water Temperature. Water Resour. Res. 55, 9173–9190. https://doi.org/10.1029/2019wr024922

Shen, C., Appling, A.P., Gentine, P., Bandai, T., Gupta, H., Tartakovsky, A., Baity-Jesi, M., Fenicia, F., Kifer, D., Li, L., Liu, X., Ren, W., Zheng, Y., Harman, C.J., Clark, M., Farthing, M., Feng, D., Kumar, P., Aboelyazeed, D., Rahmani, F., Beck, H.E., Bindas, T., Dwivedi, D., Fang, K., Höge, M., Rackauckas, C., Roy, T., Xu, C., Lawson, K., 2023. Differentiable modeling to unify machine learning and physical models and advance Geosciences. Nat. Rev. Earth Environ. 4, 552–567. https://doi.org/10.1038/s43017-023-00450-9

Wang, Y.-H., Gupta, H.V., 2024. Towards Interpretable Physical-Conceptual Catchment-Scale Hydrological Modeling Using the Mass-Conserving-Perceptron. Water Resour. Res. 60, e2024WR037224. https://doi.org/10.1029/2024WR037224

Yang, S., Yang, D., Chen, J., Santisirisomboon, J., Lu, W., Zhao, B., 2020. A physical process and machine learning combined hydrological model for daily streamflow simulations of large watersheds with limited observation data. J. Hydrol. 590, 125206. https://doi.org/10.1016/j.jhydrol.2020.125206