Reviewer #1

The manuscript presents a well-structured study that integrates LSTM-based runoff generation with GNN-based routing and shows clear performance improvements across a large European basin. The approach is timely and relevant, and the results are generally convincing. However, several important issues should be addressed to strengthen the manuscript before publication.

We sincerely thank Reviewer #1 for the comments. Below we address each comment in detail.

1. Influence of subbasin partitioning not discussed

I think the performance of spatiotemporal LSTM–GNN models is highly sensitive to how the watershed is discretized, including the number, size, and topology of subbasins. The manuscript relies entirely on the predefined 530 subbasins from LamaH-CE, without examining how alternative partitioning choices might affect routing behavior or GNN performance. A short sensitivity analysis or, at minimum, a more explicit discussion of this limitation would enhance the rigor of the study.

We agree that subbasin discretization can influence the behavior and performance of spatiotemporal LSTM–GNN models. While our study relies on the 530 subbasins, we explicitly investigated the influence of topological and static attributes associated with this partitioning on model performance. In particular, Figures 6 and 7 analyze how performance improvements (ΔNSE and ΔKGE) relate to network topology and subbasin characteristics, including upstream contributing node counts, total degree, betweenness centrality, and catchment area. These attributes are direct consequences of the chosen discretization and river network structure. Our results show that improvements from the GNN-based routing are associated with network connectivity and upstream aggregation properties, indicating that the model is indeed sensitive to the underlying subbasin topology. Larger contributing areas and more connected nodes benefit most from explicit routing, while headwater basins show smaller gains, which is hydrologically consistent.

2. Need for simple process-based baseline models

To better isolate the contribution of the GNN routing module, it would be helpful to compare the proposed architecture with simple process-based hydrological baselines, not only deep learning models. For instance, the LSTM component could be replaced with a conceptual model such as HBV, and the GNN routing could be benchmarked against a minimal routing scheme (e.g., a basic topography-driven kinematic routing approximation). Such comparisons would clarify whether the observed improvements truly arise from explicit graph-based routing.

We appreciate this suggestion. We agree that comparing against process-based hydrological models would provide additional valuable insights into the contribution of the GNN routing module. However, the primary objective of our study is to evaluate and compare different *AI-based approaches* for rainfall-runoff modeling, specifically, to demonstrate the added value of incorporating explicit spatial routing through GNNs compared to spatially-lumped LSTM approaches that dominate current deep learning applications in hydrology. Our experimental design deliberately focuses on isolating the contribution of the GNN routing component within a consistent deep learning framework (LSTM-GNN vs. LSTM alone). This approach allows us to demonstrate that the performance improvements observed in our study can be directly attributed to the addition of the GNN-based routing module, rather than differences in runoff generation mechanisms or model complexity.

Also, it should be consider that, incorporating conceptual hydrological models (e.g., HBV) or process-based routing schemes would introduce additional assumptions, parameter calibration requirements, and sources of uncertainty, which could obscure the specific contribution of the GNN routing component.

We have added following suggestions as a suggestion to revised manuscript that such hybrid comparisons for future research, particularly for assessing complementarities between physically based and graph-based routing approaches.

"In addition, hybrid comparisons that combine graph-based routing with simple process-based runoff or routing schemes could help further clarify the complementary roles of physical and data-driven approaches."

3. Title is overly broad

The phrase "Is All You Need" feels too strong relative to the actual scope of the study and may overstate the generality of the conclusions.

We thank the reviewer for this comment. The intent of the title is not to claim universal sufficiency of the proposed approach, but rather to emphasize the critical importance of explicit runoff routing in AI-based rainfall–runoff modeling—a component that is often overlooked or treated implicitly in existing deep learning studies. The phrasing "*Is All You Need*" is intentionally used as a nod to a well-established convention in the artificial intelligence literature (Vaswani et al., 2017), where it highlights a key conceptual contribution rather than asserting literal completeness. In this context, the title underscores our central finding: that explicitly modeling routing through a graph-based framework is a crucial ingredient for improving spatial consistency and predictive performance in large, connected river basins.

Vaswani et al., (2017). *Attention is all you need*. Advances in neural information processing systems, 30.

4. Figure readability

Several figures use font sizes that are difficult to read (e.g., legends of Fig. 5 and 7). Improving resolution and enlarging axis labels and legends would help improve overall presentation quality.

We have thoroughly revised Figures in the manuscript to improve readability and presentation quality.