Review of Manuscript

'A graph neural network approach to basin-scale river network learning: The role of physics-based connectivity and data fusion'

by Sun et al.

Dear Editor, dear Authors,

I have reviewed the aforementioned work. My conclusions and comments are as follows:

1. Scope

The article is well within the scope of HESS.

2. Summary

The authors propose a machine-learning (ML) approach to catchment hydrological modeling in daily resolution, consisting of graph neural networks (GNNs) pretrained on a distributed hydrological model (thus absorbing information about physical processes and subbasin connectivity as encoded in the model), fine-tuned on observed streamflow at available gauges (thus absorbing real-world information not captured by the hydrological model), and finally, if forecasting rather than simulation is the goal, a data-fusion approach (thus absorbing information of observations from the immediate past not captured by GNN modeling based on observed of forcing, but not streamflow). The authors demonstrate the approach at the example of the high alpine, snow-influenced East Taylor watershed (ETW) in Colorado, consisting of two adjacent watersheds, one with, one without human influence by a reservoir. The existing National Water Model (NWM) in the watershed is used as hydrological model. The authors demonstrate that the GNN can almost fully emulate the NWM (Kling-Gupta efficiency in the order of 0.9). As the NWM itself performs relatively poorly when compared to observations (KGE in the order of 0.5), especially for high flows, the GNN is shown to further profit from fine-tuning against streamflow observations, and from data fusion, propagating error corrections along the river network connectivity, in case of forecasting. On this basis, the authors explore GNN setup and training variants depending on the physics-based model resolution and parameter structure.

3. Evaluation

Overall, this is a very well-written manuscript on a relevant topic. The introduction provides a very good overview on the topic, the goals of the study, the experimental design, the methods and results are very clearly explained, and all conclusions by the authors are supported by the results. Congratulations. I have only very comments:

- NWM model quality: Overall, the NWM does not perform very well in the ETW basin, and so does the pretrained and fine-tuned GNN (see Table 2). Can you explain in more detail why this is so? Also, I have seen many studies where LSTMs were trained on watershed data in the US, in daily time steps, for basins with snow influence, with higher performance. While I see the benefits of the author's approach to emulate high-resolution physics-based watershed models, I wonder what KGE in testing could be achieved by LSTMs simply trained on the available gauges. I suspect they would outperform both the NWM and the GNN. Again, this will not invalidate this study, as the goal is somewhat different. Nevertheless, training an LSTM at least on some or one of the gauges (e.g. at the basin outlet) and presenting the results would help to put the results of the study into perspective.
- Fig 6: In the figure caption, subplot c) is not explained

Yours sincerely,

Uwe Ehret