I thank the authors for addressing my comments, including taking care of notational details in the equations.

The comments were extremely helpful and we are grateful that the reviewer put enough care into reading the paper especially to catch the notation issues.

I think it is a useful contribution that shows an attractive alternative to DA/inverse modeling for incorporating real-time observations of hydrological model outputs, especially for the type of models used here (LSTM).

A few remaining remarks:

1. The authors are now more nuanced in their conclusions, except for the last paragraph which makes assertions that are not supported by this study. Basically, the authors 'suspect' that their case study results are generally applicable, without giving convincing arguments or citing literature. Why not stay with the facts and simply state that more extensive and systematic benchmarking is needed to draw more general conclusions?

We agree completely, and after re-reading the last paragraph it is not good. I rewrote the last paragraph.

2. In their response, the authors mention they also did experiments where DA hyperparameters were tuned by basin. It would be interesting to mention this in the paper as a discussion item (or appendix), and give a brief quantitative indication of how much it improved things.

This is interesting, however it is a very large amount of work to make this happen. All of the experiments were re-run to account for the changes in the revisions. Incidentally, all DA and AR code was rewritten mostly from scratch during the revision because of major code refactors in the NeuralHydrology codebase between submission times. The results of the experiments were almost identical, so we are confident that there were no functional changes in the code. We would need to re-do all of these per-basin experiments in order to add them to the open source repository. This would be several days of work and weeks of run-time on GCP. Given that these results do not change the conclusions in any substantive way, we do not see this as a useful investment of time or resources.

3. Figures: some of the fonts are too small, e.g. Fig. 5

Thank you, all figures are updated with increased font size with the exception of one of the appendix figures (G1). This figure is very crowded and

Additionally, we updated the color scheme on Figure 5 to be colorblind compatible. The other figures already used colorblind-friendly color cycles.