

Reviewer 3 feedback

I am satisfied with the authors' response and revision. I only have two small points for them to address. Other than that, I think this work is ready for publication.

Thank you for the taking the time to review our manuscript.

General comments:

Line 104: Please justify why the AR model, rather than the LSTM without AR, is used in this work. Also, (if possible,) add some discussion about the potential impacts on the conclusions if the non-AR LSTM were used.

We've added the justification for an AR-model on L104 and a brief comment on the possible effects:

L105: AR LSTMs are used since they are more accurate than non-AR LSTMs (Nearing, et al., 2022), and because the Canadian hydrometric network is largely available in real-time.

L288: A limitation of this experiment, and the subsequent experiments, is that they were conducted using AR LSTMs. If AR inputs are not available, the model hyperparameters would need to be reconfigured and model performance would be expected to decrease. While the experimental results are expected to transfer to non-AR models, this would need to be empirically confirmed.

Line 255-257: Based on my understanding and the response from the authors, experiment 2b repeats 2a with cluster-based temporal undersampling. Here, it should be 'This experiment provides a comparison point between models trained in 32 basins without CUS_Q and models trained on 64 basins with CUS_Q', not CUS_B. Please confirm it.

Thank you for identifying this error; the section has been revised:

L252: This experiment provides a comparison point between models trained in 32 basins without CUS_Q and models trained on 64 basins with CUS_Q, as both configurations have the same number of samples.