

Dear Authors,

Thank you for your thoughtful and informative responses to my review of your paper: “The need for uncertainty: why probabilistic LSTMs are key to improving flood predictions and enabling learned warning rules”. I think that this paper is highly impactful, and support its publication. I’m sorry for my delay in this second review. There are just a few outstanding issues that you might consider during the proofing phase:

- Your tracked changes document in Table 2 shows the DRN epochs changed from 10 to 15, but your clean document shows 10 still. Although the text on the tracked changes on line 953 also shows 10, so I assume 10 is the final number of epochs for DRN. No need to change anything if 10 is the final number, just pointing this out.
- It is not clear why you changed table 5 from DRN to CMAL. Was this a typo in the original draft, or intentionally changed? I looked through all your responses and didn’t see an explanation of this change, but maybe I missed it.
- Many of your changes were not described sufficiently in the response letters, which led to a delay in my review. For instance, Reviewer 2 asked for an explanation of how hits were calculated, and you simply said “Noted and will be added in the revised manuscript.” I believe I found it on Line 282, but would have been nice to include that in your response.
- On line 344 of the revised manuscript, you state “all within the 99th percentile” in reference to observed flood peak. In your response letter you state that “By definition, a predicted 90th (or 99th) percentile event is expected to be exceeded in 10% (or 1%) of cases, respectively.” Which I understand and agree with, and I appreciate your Figure F2 showing the distributions and probability plot. On line 382 in the revised manuscript it states, “about 83% of the observed values are within the 99th quantile of their respective predictive distributions”. Is this in conflict with the quote above? Then in your response letter “consequently, particularly rare or unprecedented events, such as those beyond the empirical range of the training data, are not expected to be reliably enclosed by lower quantiles.” Am I correct on interpreting that the missing 16% of the observations are the ‘particularly rare or unprecedented?’ I’m sure I am simply misunderstanding, but might be worth sharpening up that language.

Thanks