We thank the review for the positive and constructive feedback. Please find below the answers to your questions. All answers will be implemented in the revised version of the manuscript. Our reply to each of your questions/suggestions can be found below.

Answers to Anonymous Referee #1' s comments:

Comment 1: *Are precipitation, temperature, and humidity enough as input variables for your neural networks?*

Authors' answer: In this study, precipitation, temperature, and humidity are the only variables available to develop the models, hence they were selected as primary input variables based on their direct impact on flood generation processes. The sensitivity analysis that we have performed and presented in Section 3.4 of the discussion manuscript, showed that N-HiTS and N-BEATS models maintained high performance even when individual meteorological variables were excluded, indicating the models' robustness to input variations. This robustness likely stems from the models' ability to learn essential patterns from historical discharge data alone, particularly the rainfall-runoff relationships in headwater streams. We will add more clarity in that section to reflect that this is the main conclusion and this is the reason why the selected variables are enough.

Comment 2: *The forcing station is a single point in the watershed while the runoff generation should be attributed to the water convergence involving a large area of the watershed, do you think a single station can represent these complex processes at large areas?*

Authors' answer: Thank you for this important remark. Indeed, a single station may not fully represent the spatial heterogeneity of larger watersheds. We acknowledge that using a single station can provide localized information and data for small watershed, such as headwater streams where runoff generation is more responsive to immediate meteorological conditions. However, for the particular case we have applied the methods to, capture significant runoff and flood dynamics in these streams, with a single station. Therefore, we will mention these important aspects in the conclusion section of the paper, especially that for broader areas, more spatially distributed data could improve model accuracy.

Comment 3: *You mentioned, your models predicted one hour ahead? Is this meaningful for flood prediction? In other words, is this enough time to escape once people know the flood will arrive one hour later.*

Authors' answer: You raised a valid point. Although a one-hour lead time may seem brief, it is meaningful in the context of flash flood prediction for headwater streams, where rapid response is essential. Although, we tested the model to forecast up to 72 hours in advance based on National Weather Service near real time river forecast data. We found that one-hour lead time provides the best near real time performance. We will include this remark in the conclusion part of the revised manuscript, as this is an important remark for the readers, thank you for your comment.

Comment 4: *Did you train each NN model for each watershed? Trained based on one watershed and then transferred to the other one? Or trained both watersheds together?*

Authors' answer: Thank you for your comment. Each neural network model was trained separately for each watershed to account for local hydrological characteristics. The independent training approach enabled the models to capture watershed-specific runoff generation and flood dynamics, optimizing prediction accuracy within each unique environment. Transfer learning was not implemented. We will include this remark in the new version of the manuscript to make it clear for the readers.