The authors thank the reviewers for their positive evaluation of our work and for their helpful suggestions.

In the following, we provide answers to the reviewers‘ comments and how we addressed them. Original comments are provided in blue, and our responses are shown in black.

Review 1

I appreciate the authors' efforts in fully addressing my comments. A couple of suggestions as follows.

1. Line5: LSTM is commonly short for "Long Short-Term Memory" rather than "Long-Short-Term-Memory".

   We changed all cases of "Long-Short-Term-Memory" to "Long Short-Term Memory".

2. In Figure 2, it is not clear whether the dots are mean or median values, and the color of the vertical lines tends to be grey rather than black.

   Thank you for the comment. Indeed, the lines appear more grey because they are rather thin. We changed the caption of figure 2 as follows:

   Mean NSE (a) and KGE (b) performance of the LSTM ensembles consisting of 10 individually trained models for the test segments of the time series for streamflow simulation for each river gauge depending on the catchment area. The thin dark grey lines represent the best and worst members of the LSTM ensemble. The color shows the standard deviation (std) of the 10 individual LSTMs.

Review 2

Congratulations to the authors on this interesting paper. All the points I raised in the initial review have been sufficiently addressed. I have only noticed a few minor issues that can be corrected in the final version. Line referencing refers to the manuscript which includes track changes.

1. L303: delete “and”.

   We deleted "and".

2. L389-393: There are bias-correction methods that take into account changes in the temporal structure of the data. Readers may wonder why such a method was not used, given that the study is on flood events which are potentially affected by such changes. Thus, I suggest commenting this very briefly.

   Thank you for your comment. Indeed, there are more advanced bias-correction methods. We prefer here the quantile-delta mapping, because it adjusts the value range of the model data to match that of the training data while maintaining other model characteristics, particularly in the context of potential future climate changes. Additionally, the widespread floods are most likely not affected by small temporal or spatial shifts. We added a short comment at the end of the subsection:

   Although there are more advanced bias-correction methods that address these issues, we prefer quantile-delta mapping because it adjusts the value range of the model data to match that of the training data while maintaining other model characteristics, particularly in the context of potential future climate changes. Furthermore, the potential impact of minor spatial and temporal shifts in weather patterns is likely to be rather small in the context of widespread winter floods.