Reviewer #1

We would like to sincerely thank the reviewer for their constructive and thoughtful comments.

Below, we provide our responses to each of the reviewer's comments, along with line references to indicate where the changes have been made in the revised manuscript. The added text is shown in italics here and highlighted in yellow in the revised manuscript.

In the abstract, the authors can add some quantitative results to show how much this data assimilation approach significantly improves hydrologic model's ability to forecast extreme river flows.

To address this comment, we have added a paragraph to the abstract. Please refer to lines 45-52 in the revised version. The added text is also included below for your reference:

"The study found that data assimilation improved streamflow forecasting during Hurricane Harvey, enhancing the SAC-SMA model's accuracy across most USGS stations on the peak flow day. However, data assimilation had little effect on streamflow forecasting for Hurricane Rita. In Rita, the streamflow surged dramatically in a single day (from 28 m^3/s to 566 m^3/s), causing the model to miss the high flow event despite accurate initialization the day before. For Hurricanes Ivan and Matthew, data assimilation improved peak flow forecasts by 21% to 46% in Mobile and 5% to 46% in Savannah, with improvements varying by station location."

How the different sources of uncertainties are considered in HEAVEN should be mentioned.

The data assimilation method employs the weak-constraint 4DVAR cost function (Eq. 1), which incorporates three covariance matrices: B, R, and Q. These matrices represent errors in the initial condition, observations, and model structure, respectively. Additionally, the uncertainty associated with the forcing data is considered within the sequential assimilation process. In EPFM, we assume that there is an error in the forcing data. Based on this assumption, we add white noise to the forcing variables, creating an ensemble of forcing data that is then used to drive the hydrological model. As a result, this data assimilation method accounts for all sources of uncertainty.

To address this comment, we have incorporated the above text with some modifications to align it with the revised version. Please refer to lines 350-363 in the revised version. The added text is also included below for your reference:

"The DA method utilizes the weak-constraint 4DVAR cost function (Eq. 1), which accounts for multiple sources of uncertainty by incorporating three key covariance matrices: B, R, and Q. These matrices represent different types of errors: B accounts for errors in the initial condition, R represents observational errors, and Q captures model structural errors. By explicitly modeling these errors, the method provides a more comprehensive and realistic

representation of the uncertainty in the system. In addition to these sources of uncertainty, the method also considers the uncertainty associated with the forcing data. In the context of the EPFM approach, it is assumed that errors exist in the forcing data, which can significantly affect model predictions. To account for this, we introduce white noise to the forcing variables, effectively perturbing the forcing data. This process generates an ensemble of forcing data, which is then used to drive the hydrological model. Thus, the DA method is designed to account for all major sources of uncertainty—initial condition errors, observational errors, model structural errors, and errors in the forcing data. By incorporating these uncertainties into the assimilation process, the method enhances the accuracy and reliability of the model predictions."

It would be better to introduce figure 2 in the beginning of section 2.4.

Thank you for your comment. It would have been ideal to include this figure at the beginning of Section 2. However, since the figure is designed to summarize the methodology with references to equations and detailed explanations of the method, we felt it was more appropriate to place it toward the end of Section 2.4. This placement allows the figure to effectively summarize the approach discussed in this section.

In figure 4, the validation accuracy in terms of R is higher than that during calibration in Savanah, but the validated RMSE is also higher than the latter. Why?

Thank you for your comment. We used two distinct time periods for model calibration and validation, and we have double-checked the results to confirm their accuracy. While the correlation coefficient (R) during the validation period is slightly higher than during calibration, the simulated values in the validation period exhibit a higher bias. Therefore, we cannot conclude that the model validation outperforms the calibration results.

In figures 6-7, Streamflow forecast improved by data assimilation during peak flow conditions across the USGS stations are demonstrated. The POI ranges from 0-100%. Is there any negative value?

No – there was not any negative values.

It would be better to draw a scatter plot to demonstrate the streamflow forecast skills during peak flow conditions with/without data assimilation.

Thank you for the suggestion. We have already presented this result in Figure 7, as well as in supplementary Figures S2 and S3. Therefore, generating an additional figure would not have any added value.