Reviewer #2

We would like to sincerely thank the reviewer for their constructive and thoughtful comments. Their feedback greatly contributed to improving the clarity and quality of the manuscript. We appreciate the time and effort spent on providing valuable suggestions, which have helped enhance the overall presentation of the study.

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.

Abstract:

• Please add some quantitative results to the abstract to highlight the improvements achieved.

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."

Introduction:

• Line 55-56: Please move the sentence related to the current flood $ to line 50 before the mention of the climate change. It will help the flow of the paragraph to talk first about the current time, and then shift the focus to future and climate change impact.

We have revised the manuscript as per the reviewer's suggestion and relocated the sentence accordingly. Please refer to lines 57-59 in the revised version.

• Line 113: compared to what? If you mean open loop state that please, if you mean other methods, maybe you want to add a few words to indicate that.

We have incorporated the reviewer's suggestion and revised the sentence accordingly. Please refer to lines 118-123 in the revised version. The added text is also included below for your reference:

"The probability distributions of both model states and parameters are recursively and independently updated at each time step as new observations become available. These approaches yield more accurate state and parameter estimates compared to open-loop

simulation (without data assimilation), allowing the modeling system to evolve consistently over time. As a result, this leads to improved model predictions while accounting for uncertainties."

• Line 122: Add why focusing only on the peak flow. I assume this is since peak value matters the most in the inundation mapping and flooding extent, however, please explicitly mention as why this is the focus of the study.

We have added a paragraph to the revised manuscript to clarify this point and explain why the focus of this study is on forecasting peak flows. Please refer to lines 131-138 in the revised version. The added text is also included below for your reference:

"Although the capability of this method, in conjunction with hydrological models for streamflow forecasting, has been demonstrated in previous studies, its ability to capture peak flows induced by heavy rainfall from hurricanes—common in southeastern regions—has not been explored. This study aims to address this gap and contribute to enhancing the resiliency of the SEUS, a region particularly vulnerable to extreme flooding due to hurricanes and tropical cyclones. By improving streamflow forecasting during such events, the study seeks to better inform flood management strategies and mitigate the impacts of future extreme weather events in these high-risk areas."

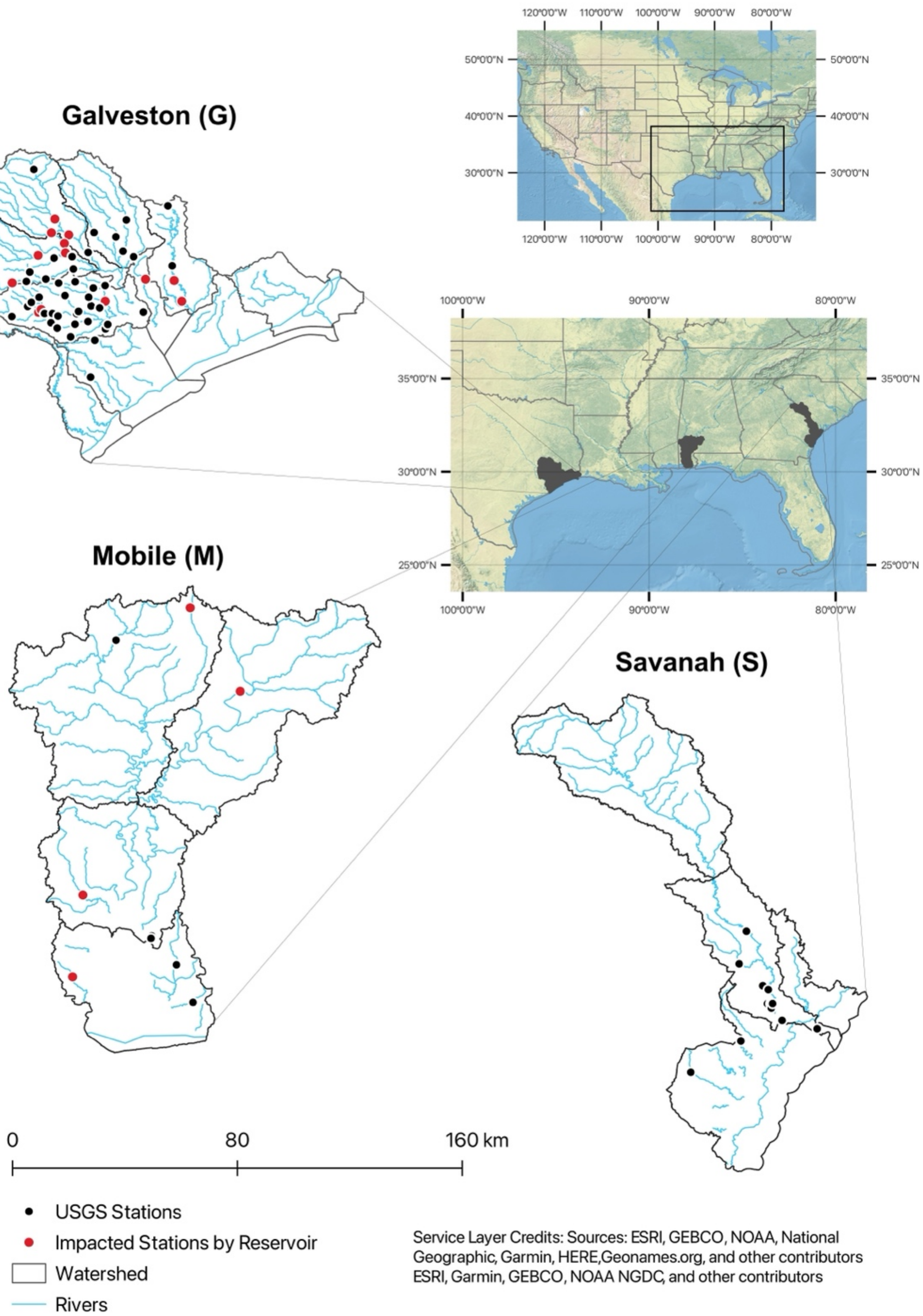
Study Area:

• Figure 1: Could you add the main rivers to the zoomed panels in the watersheds? It helps visualize the USGS gauge locations (on main step/small creek) better.

• Figure 1: Would it be possible to add the polygons used in the SAC-SMA? There is only mention of the HUC8s in the domain, is that the level of refined lumped basins? I did not find further information in the manuscript text.

• Figure 1: How about color coding the USGS gauges to those that are impacted by regulation/lake/reservoirs versus the natural gauges? Or maybe we could use different symbols. This would help with the calibration/validation verification and discussions.

To address the three comments above, we have revised and regenerated the figure and incorporated all of your suggestions. Please see the revised Figure below. We included this in the revised manuscript. Please refer to lines 155-156 in the revised version.



• Line 178: please cite the NOAA report.

We have included the reference to the NOAA report in the revised manuscript. Please refer to lines 181-182 in the revised version.

• Line 191-192: Please add a reference for the claim related to the climate change impact.

We have included two references for this statement in the revised manuscript. Please refer to line 208 in the revised version.

Datasets:

• Line 213: Please cite the package Climata.

We have included the GitHub page for Climata in the revised manuscript. Please refer to line 229 in the revised version.

• Line 230: It states NLDAS is a dataset reasonable for operational purposes. Could you cite? I think there are better options available for precipitation currently, for example MRMS which has a better spatial resolution, and is used in operational models. If you have an example of operational use of NLDAS it would be nice to mention here.

To address this comment, we have removed the word "operational" from the text and revised the sentence for improved clarity. Please refer to lines 246-247 in the revised version. The added text is also included below for your reference:

"… which has been widely used to derive hydrology and land surface models."

• Line 237: how about using semi-distributed instead of spatially lumped?

We have changed "spatially lumped" to "lumped-parameter model" (Gourley et al., 2013) in the revised text. Please refer to line 253 in the revised version.

References

Gourley, J. J., Flamig, Z. L., Hong, Y., & Howard, K. W. (2014). Evaluation of past, present and future tools for radar-based flash-flood prediction in the USA. *Hydrological Sciences Journal*, *59*(7), 1377–1389. https://doi.org/10.1080/02626667.2014.919391

• Line 240: I think it is more appropriate to refer the readers to the main SAC-SMA reference and use Abbazadeh et al. 2018 study as a second reference.

Done. Please refer to line 257 in the revised version.

• Line 242, is there a more recent reference, application that could be used?

We have added additional references here to address the reviewer's comment. Please refer to line 257 in the revised version.

• Line 245: Why 6 hourly? Is this used in operational? If yes, could you provide citation or reference? I have seen hourly studies before, but not sure why one do the 6 hourly unless due to limitation of the forcing. Later in the results it was mentioned the model outputs are hourly.

The SAC-SMA model is generally run at a 6-hour time step but can run at any time step. Inputs to the SAC-SMA model are 6-hour mean areal precipitation (MAP) and 6-hour mean areal potential evapotranspiration (MAPE)." this text is borrowed from National Weather Services website. We followed this description and use it in the manuscript. We also double checked the SAC-SMA model and its code configuration, and this was correct. The model output is at a daily time scale. Thank you for pointing this out; we have corrected it in the revised manuscript. Please refer to lines 423-424 in the revised version.

If the above hyperlinked page does not work, please use this link: https://www.weather.gov/owp/oh_hrl_papers_distmodel_sacbstg3#:~:text=The%20SAC%2DSMA%20model%20is%20generally%20run%20at%20a%206,areal%20potential%20evapotranspiration%20(MAPE).

Data Assimilation:

• This section was hard to follow. It would benefit greatly of rethinking the presentation. This paper is not focused on introducing any of the methodology used here, therefore, authors could drop most of mathematics and instead use telling schematics (expanding the figure 2) that could help narrate the workflow. Given the previous papers were filled with extensive discussions of the mathematics behind it and probably less space for graphics and schematics, maybe you could do that instead in this paper. I would highly recommend the author thinks differently about presentation of the information in this section. Adding information on what parameters and what states were updated, would be beneficial to the reader.

Thank you for your comment and suggestion. In the initial draft of the manuscript we submitted to the journal, we had indeed written the data assimilation section quite close to what you've suggested here. However, the editor recommended that this section should be more self-contained, ensuring that readers can understand the methodology without needing to refer back to the original paper. The editor emphasized that the information provided in this paper should be sufficient for readers to grasp the methodology and implement it. As a result, in the editorial review process, we had to revise this section.

To address this comment, we have added an appendix section to the revised manuscript, where we have included the major portion of methodology as you suggested. Please refer to lines 601-682 in the revised version.

Additionally, we have elaborated further on Figure 2 to help readers quickly grasp the underlying idea of the methodology and understand its advantages compared to other existing methods. Please refer to lines 335-347 in the revised version. The added text is also included below for your reference:

"Here, we refer readers to Appendix A, where we describe the implementation of the Evolutionary Particle Filter with Markov Chain (EPFM) data assimilation approach (Abbaszadeh et al., 2018). To facilitate the reproduction of HEAVEN, Figure 2 presents a schematic summarizing all the processes involved within this approach. Step 1 in this figure illustrates how the initial condition for the first window cycle is generated. As mentioned earlier, by minimizing the weak-constrained 4DVAR cost function, the optimal initial condition for the first cycle is obtained. Note that this is a deterministic value, which must be

reshaped into an ensemble for initialization of the sequential filtering process, as described in Step 2. In Step 3, the EPFM sequential filtering approach (explained in Appendix A) is used to calculate the posterior estimates of model states and parameters within the first assimilation cycle. Next, we use Eqs. A19 and A21 to calculate the dynamic error covariance matrix and the prior error covariance matrix. Finally, in Step 6, we use the updated error covariance matrix along with the expected values of the posterior estimates of model states and parameters to initialize the next assimilation cycle."

• If author decides to keep the mathematics, please make changes to add subtitles and improve the flow of the discussion. Also Figure 1 of Abbaszadeh 2019 + Steps outlined in the manuscript were great narratives.

We have addressed this comment in our response to the previous one.

• Some of the text is exact copy of previous papers, please revise to the extent possible to avoid plagiarism.

We have significantly revised and edited the text.

• A general comment is the fact that several time it was mentioned this framework addresses all the uncertainties, maybe the author could narrate how that is been done. It was attempted through the methodology section, however it could be beneficial to highlight how each uncertainty is addressed. In particular I do not know what was done to address the forcing uncertainty, could you elaborate on it?

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."

Model Calibration and validation:

• I do not follow this statement "This ensures the applicability of the calibrated model for predicting future events,"

To validate the calibrated model, we tested its ability to predict peak flow during a period that was not used for calibration. This helps assess the model's performance and generalizability to unseen data. To clarify this statement further, we have added the following text to the revised manuscript. Please refer to lines 368-370 in the revised version. The added text is also provided below for your reference:

"*To validate the calibrated model, we tested its ability to predict peak flow during a period that was not used for calibration. This helps assess the model's performance and generalizability to unseen data."*

• Line 404 to 409 is relevant to dataset section for ET and also the discussion of choice events used for this manuscript. I think these lines could be moved to earlier sections.

We have moved it to earlier section where we discussed the ET data and model configuration. Please refer to lines 266-270 in the revised version.

• It is not defined what gauges were used in the calibration, for example for Galveston basin has so many gauges, are all used in the model calibration? I am not clear how fine the lumped basins were, HUC8 or just one big basin. Could you clarify that? Is each gauge used to calibrated just a portion of basin, or they are all together?

The SAC-SMA model was calibrated separately for each drainage area contributing to the USGS stations. As a result, we did not use the entire basin for model calibration. Since the SAC-SMA is a lumped model, we calibrated and validated it at the USGS stations, considering their corresponding drainage areas. This approach is particularly important for calculating the model's forcing input variables, such as mean areal precipitation and mean areal PET.

To clarify this further, we have added the following text to the revised manuscript. Please refer to lines 379-392 in the revised version. The added text is also provided below for your reference:

"*The SAC-SMA model was calibrated separately for each drainage area associated with the USGS stations, rather than using the entire basin for the calibration process. This decision* *was based on the structure of the SAC-SMA model, which is a lumped model, meaning that it aggregates hydrological processes over a given area rather than considering them at individual sub-basins or locations. To ensure accurate representation of the hydrological processes, we calibrated and validated the model specifically at each USGS station, while carefully accounting for the unique drainage area contributing to each station's flow. This is important for accurately calculating the model's forcing input variables—such as mean areal precipitation and mean areal potential evapotranspiration—since these inputs depend on the spatial extent of the drainage area for each station. By focusing on the specific drainage area for each USGS station, we ensure that the model's inputs reflect the local conditions of the watershed, leading to more reliable and representative model calibration and validation results. This method also improves the model's ability to simulate hydrological processes at the station level while considering the variations in environmental factors across different parts of the basin."*

• What time step is used in the modeling and calibration? It was mentioned the model is running every 6 hours, did the calibration happen every 6 hour or daily? In Line 443 states the model output is hourly. I am confused on the time scale.

Yes, the model runs every 6 hours, consistent with the time scale of the input data. However, the model output is provided at a daily time scale. The term "hourly" was used incorrectly in the text, and we have corrected it in response to one of your earlier comments. Thank you again for your careful review and constructive feedback, which we believe has significantly improved the quality of the manuscript.

• Has there been a warm up period before the calibration, could you comment on it?

Yes, a 3-month warm-up period was used at the beginning of the calibration and validation periods. This spin-up period helped establish reliable initial conditions for the model's state variables.

To further address this comment, we have added the following text to the revised manuscript. Please refer to lines 393-396 in the revised version. The added text is also provided below for your reference:

"To ensure reliable initial conditions for the model's state variables, a 3-month spin-up period was used at the beginning of both the calibration and validation periods. This warm-up period allowed the model to stabilize prior to the actual calibration and validation processes."

• How about including what parameters were calibrated? Maybe refer to the previous papers table if there is no need for repeating it.

We have added the following text to the revised manuscript and referred readers to our previous publication. Please refer to lines 422-423 in the revised version. The added text is provided below for your reference:

"The model parameters are those listed in Table 1 of our previous paper (Abbaszadeh et al., 2018)."

Results:

• Figure 4. Why not displaying the NSE in the figures? NSE was used as the objective function.

That's a great suggestion. We have incorporated this information into the revised text. To further address your comment, we have also included the following figure in the updated manuscript. Please refer to lines 458-470 in the revised version. The added text is provided below for your reference:

"The NSE values for the calibration period were 0.80, 0.78, and 0.69 for Galveston, Mobile, and Savannah, respectively. Similarly, for the validation period, the NSE values for these regions were 0.68, 0.71, and 0.65, respectively. Figure 6 also illustrates the NSE for the USGS stations across the Galveston and Mobile watersheds. In this figure, positive NSE values are shown with black circles, and negative NSE values are shown with red circles. The regulated USGS stations are marked with black dots to facilitate interpretation of model performance at those specific locations. The results indicate that, in general, the model performance is lower at the regulated USGS stations. For example, the NSE for USGS station 8074000 (Figure 6) is negative; this station is located downstream of Addicks and Barker Dams."

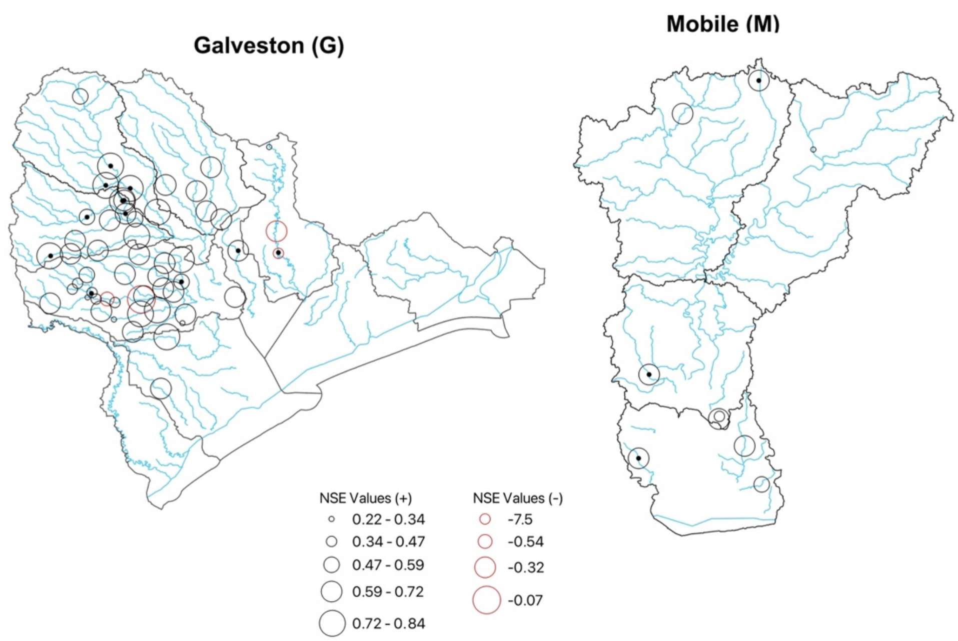


Figure 6 also illustrates the NSE for the USGS stations across the Galveston and Mobile watersheds.

- Line 458: SAC-SMA does not have any regulation/reservoir/lakes represented so it is a deficiency of the model as you stated, and it would be beneficial to differentiate between natural and regulated basins.

To fully address this comment, we decided to update all the figures in the manuscript and incorporate the impacted stations by regulation/reservoir/lakes in the analyses.

- Line 495: I cannot connect the dot why this methodology is specifically more important in peak flow estimation. Could you elaborate on that? Is it because the model is not decent in estimating the peak flow?

Thank you for the comment. To address this comment, we added the following text to the revised manuscript. Please refer to lines 495-499 in the revised version. The added text is provided below for your reference:

"In summary, the capability of the developed hydrologic data assimilation method to estimate peak flows stems from its automatic process of inflating and deflating the background error covariance matrix. This process enables the model to achieve a more realistic ensemble spread, leading to a more accurate expected value compared to the observed value."

- Line 510: Elaborate what you mean by "sub optimal"

The model performance was not at the best possible level.

- Line 513: could you elaborate what should be done instead for hydrodynamical modeling?

Flood inundation modeling using hydrodynamic models is part of our future research. In this context, we want to highlight that for stations with heavy regulation, streamflow forecasts can be used to condition the upstream boundaries of the hydrodynamic model.

- Lines 517-519. Repetitive, could be dropped.

Done.

- Lines 252-531. Do you think having a more frequent cycling could have helped Rita? Keeping the Assimilation window as a day, however, recycling more frequently? This would be kind of like using a smoother. Does this suggest this is a limitation pf HEAVEN in improving the accuracy of short-lived events such as sub daily scale thunderstorms?

Thank you for the suggestion. Since the assimilation is performed at a daily time scale, HEAVEN requires a few days as a window to solve the 4DVAR objective function and find the optimal initial condition for sequential assimilation (particle filtering). Therefore, using an assimilation window shorter than one day is not feasible, unless we had observational data at a sub-daily scale and could perform the assimilation at that resolution— which is not the case here.

• Figure 6 and 7 could be combined, also adding the major river network would help the visual inspection. Maybe you could combine the size and color to covey more information just the POI as the figures are not busy.

We updated this figure and incorporated the suggested comments. Please refer to lines 527-530 in the revised version.

• What is POI? It is probably a commonly used metric by the author, however, it is not clear to me what it stands for. Could you please define the metric before using it?

POI represents the percentage of improvement achieved by performing the assimilation. In other words, it shows how much the streamflow forecast improved (in percentage) by using data assimilation compared to the open-loop model simulation (without assimilation).

To further address this comment, we added the following text to the revised manuscript. Please refer to lines 515-518 in the revised version. The added text is provided below for your reference:

"In Figure 7, POI represents the percentage of improvement achieved by performing the assimilation. In other words, it shows how much the streamflow forecast improved (in percentage) by using data assimilation compared to the open-loop model simulation (without assimilation)."

• Line 541. Nicely said!

Thank you.

• Figure 8. What is the green band? It is not in the legend of the figure. Is it referring to the time period that POI is based on it? It was not clear to me, whether the verification is based just the peak value or an even containing peak value. The green band here suggest the latter, however, it would be great if clarified and clearly stated.

In Figure 8, the green band represents the time window that contains the peak flow value, not just the peak itself. This band is intended to show the broader period during which the peak flow occurs, rather than the single peak value. The verification process was based on this entire time window, including the period before and after the peak. We apologize for any confusion caused and have updated the figure legend to more clearly indicate this distinction.

To further address this comment, we added the following text to the revised manuscript. Please refer to lines 551-555 in the revised version. The added text is provided below for your reference:

“In Figure 8, the green band represents the time window that contains the peak flow value, not just the peak itself. This band is intended to show the broader period during which the peak flow occurs, rather than the single peak value. The verification process (POI reported in Figure 7) was based on this entire time window, including the period before and after the peak.”

Conclusion:

• Please consider improving the conclusion section. It is not summarizing the study and the findings very well. Given this study is a use of HEAVEN for peak flow estimate and hurricane induced flooding, you could elaborate on it. Any discussion of the limitations? For example the Rita case and what could be done to resolve it.

To address the reviewer’s comment and improve the conclusion section, we have rewritten this part of the manuscript to ensure that all the reviewer’s concerns/comments are fully incorporated. Please refer to lines 562-599 in the revised manuscript. The revised text is provided below for your reference:

“This study investigates the application of the HEAVEN data assimilation technique to improve the forecasting of extreme river flow during hurricane-induced flooding in the SEUS. By integrating HEAVEN with the SAC-SMA hydrologic model, this research aimed to address the various sources of uncertainty in hydrologic simulations, particularly during extreme events such as hurricanes. The results show that HEAVEN effectively enhances the SAC-SMA model’s streamflow forecasting capabilities by incorporating uncertainty from multiple sources, including meteorological forcing, model parameters, and structural errors.

The key findings highlight that data assimilation through HEAVEN significantly improved streamflow forecasts during peak flow conditions, especially in cases where extreme river discharge occurred. For example, in the Galveston watershed during Hurricane Harvey, data assimilation led to substantial improvements in the forecasting of peak flows, with forecasted values much closer to observed streamflow than those from the open-loop (non-assimilated) model. However, in cases like Hurricane Rita, where the streamflow increased abruptly within a very short time window, the assimilation approach was less effective. Despite accurate initial conditions, the model struggled to capture the rapid onset of extreme flow, highlighting a limitation of the current approach. This was due to the inability of the model to anticipate such a drastic shift in flow within a single day, an event that may require further refinement of the assimilation process to account for sudden, large fluctuations in discharge.

The HEAVEN technique proved capable of addressing model structural uncertainties by inflating and deflating the background error covariance matrix, ensuring a more reliable posterior distribution of streamflow forecasts. The assimilation process also facilitated the quantification and interaction of multiple sources of uncertainty, improving the overall robustness of the predictions. While the model performed well across most stations, some locations remained challenging, particularly those influenced by water releases from upstream dams. These locations, which significantly alter the natural flow dynamics, may *require more specialized modeling approaches in future work, such as hydrodynamic modeling.*

This study also discusses the computational limitations associated with optimizing the 4DVAR cost function using the Nelder-Mead algorithm, as the tangent linear or adjoint models were not available. While this approach is effective, it remains computationally intensive. With the growing use of Machine Learning (ML) emulators in hydrologic modeling, future work may focus on incorporating these techniques to accelerate optimization and further enhance the efficiency of data assimilation in large-scale hydrologic forecasting.

In summary, the HEAVEN data assimilation method offers a promising advancement in the accurate prediction of extreme river flow during hurricane events, but challenges remain in addressing sudden and large fluctuations in streamflow. Future developments may focus on refining the assimilation process for such events and incorporating additional modeling techniques, such as hydrodynamic models for regulated river systems, to further improve forecasting accuracy and reliability."

Editorial and Technical Corrections:

• Line 50: Literature is misspelled.

Corrected.

• Line 51: there is an extra space after parenthesis.

Corrected.

• Line 52: there is an extra ":" at the end of the citations.

Corrected.

• Line 53: how about using "evaporation" instead of "water evaporating"?

Revised.

• Line 100: Extra "." Before citations.

Corrected.

• Lien 104: How about using a different word instead of "layers", maybe components?!

Revised.

• Line 107: How about changing "were designed" to "are commonly used".

Revised.

• Line 119: Please state the full name and put HAVEN in the paratheses. Similar format is used everywhere in the text for other abbreviations also, for example line 208 and 209. I have seen mostly defining the name and then putting the abbreviation in the parentheses.

We have corrected this issue throughout the text.

• Line 162 and 165: There is first name of the authors in the citations. I think the HESS format required only the last name, please double check the policy. Same as line 206, Corrected.

• Line 227: The sentence is being cut into two. Replace the "." With ",".

Corrected.

• Line 255: Please change the "in the above study" to "Abbaszadeh et al 2019"

Revised.

• Line 318: Here is being misspelled.

Corrected.