Reviewer 1:

In this study, the authors present a Data Assimilation (DA)-hydrodynamic modeling framework where multiple gauge observations are integrated into the LISFLOOD-FP model to improve the performance of flood inundation mapping. The results indicate that the multivariate assimilation of point-source observations into hydrodynamic models can improve the accuracy and reliability of probabilistic flood inundation mapping by 5-7% while it also provides the basis for sequential updating and real-time flood inundation mapping. This paper is well written, well-organized and has practical meaning for inundation mapping, risk analysis and decision making. I have some minor comments for the authors to improve their paper.

Response: We thank the reviewer for the positive evaluation of our work.

Line 168: Xu et al., 2017 is used as a reference. However, this paper is not listed in the bibliography.

Response: This reference was added to the bibliography.

The text (e.g. names of counties and states) in Figure 1 is blurred. Please improve it

Response: The quality of the figure was improved.

In addition, the water discharge at gauge 1 and 2 as well as water stage at gauge 2 (n=3) are assimilated. The authors could also assimilate the water discharge and the water stage at gauge 1, 2 and 3 to further improve the performance. Is the data available?

Response: Thank you for the comment. The water stage at gauge 2 is not available for the peak of Harvey. We used all available data for two internal gauges. In this research, we explain the DA-hydrodynamic modeling framework in a generic format where n number of point source observations can be assimilated into a hydrodynamic model. This helps the potential users to follow this approach and apply it for different number of observations in future studies.

Figures 6 & 7: the two figures presented the comparison between OL and EnKF. It would be better to add the real inundation map in the third column of the figures.

Response: The reference flood inundation map we use in this study is representative of maximum flood maps during all days (The union of daily flood maps). We don't have daily real inundation maps to compare with simulated maps at a daily scale. However, we provided these two figures to explore the differences between OL and EnKF approaches for generating probabilistic inundation maps. In order to validate these two methods, we first evaluate the temporal behavior of these two approaches by comparing the simulated time series with observations at gauges (Figure 4) and then test the spatial behavior by taking the union of all simulated daily maps and comparing them with the reference map in Figure 8. Although we don't have daily real maps to update Figures 6 and 7, we improved Figure 8 by adding the comparison of maximum OL and ENKF with a reference map.

Line 542-545: Here, we used the EnKF data assimilation method in conjunction with a hydrodynamic model to account for different sources of uncertainties involved in different layers of model simulations, including the boundary conditions, model parameters, and initial condition, and generate real-time probabilistic flood inundation maps. It is nice to consider different sources of uncertainty here. How the uncertainty fluctuates with ensemble size and error settings?

Response: Analyzing the sensitivity of DA-hydrodynamic modeling results, to the error settings and ensemble size, is an interesting topic of research. However, to accurately address these two points, a multitude of DA-hydrodynamic models should be simulated within a sensitivity analysis framework which is not within the scope of this study. Here, we aimed to demonstrate the effectiveness of assimilating multi-source data into hydrodynamic models for improving inundation mapping skills. According to several past studies that explored the sensitivity of EnKF to the ensemble size, the model performance is highly sensitive to small ensemble sizes less than 50 but the sensitivity will be reduced by moving to larger numbers around 100 (e.g. see Figure 4 in Moradkhani et al., 2005 and Figure 2 in Gillijns et al., 2006). With access to a sufficient number of computing cores at the University of Alabama High-Performance Computing cluster, we were able to use a relatively large number of 100 ensemble members (It is larger than most of the studies that used the EnKF method). This ensures that our results will provide a reasonable range of uncertainties. Regarding the error settings, the synthetic experiment is typically used to tune the perturbation parameters. The synthetic experiment results together with the recommendations of past studies are the basis for designing the error settings in a DA framework.

Figure 4: it would be better to add the predictive interval of open loop

Response: Thanks for the great suggestion. Figure 4 was updated and the open-loop interval was added.

Line 112: the following paper may be a good reference for the SMAP soil moisture data assimilation. Xu et al. Continental drought monitoring using satellite soil moisture, data assimilation and an integrated drought index. Remote Sensing of Environment, 2020, 250:112028.

Response: The reference was added to the manuscript.

Reviewer 2:

The paper is interesting and presents a useful methodology for using data assimilation in flood modeling studies. A few assumptions need to be clarified. Limitations, uncertainties and implications need to be further discussed. The authors should also provide recommendations on future research and how this work can be applied to other study areas. I have recommended several edits and some comments in the PDF. Here are some additional comments:

Response: Thank you for the positive and detailed review of our work and for providing constructive comments. Please see our response and revisions below.

Introduction: The section can be shortened and be more concise.

Response: We shortened the introduction and revised the text according to the suggestion of the reviewer in the pdf file.

Past research on probabilistic flood modeling should be further acknowledged (e.g., Aronica et al. 2012, Savage et al. 2016, Papaioannou et al. 2017, Ahmadisharaf et al. 2018).

Response: The suggested references were added to the revised manuscript.

Please provide more information on the study catchment, particularly those that affect your model results. This includes computational area, soil type, channels' size, ground slope, land use, etc.

Response: More information regarding the study catchment was added to the revised manuscript. Please refer to lines 197-203.

How well do the 120 m DEM represent the watershed topography and bathymetry? What are potential errors and how they affect your findings?

Response: We added a new paragraph in the discussion section explaining the logic behind using a coarse DEM and its potential impacts on the results. Please refer to lines 619-627 and 639-649 in the revised manuscript.

More details on the hydrodynamic model setup is needed. How long was the warm-up period? What was the simulation time step? What was the Courant number?

Response: We added more information regarding the model setup. Please refer to lines 356-358 in the revised manuscript.

Please present the computational time of your simulations. In practice and with the existing resources, how practical it is to use the 2D unsteady models for real-time flood forecasting? In particular, high resolution analyses might be desired. Is the presented DA-hydrodynamic modeling framework computationally efficient to be applied in practice?

Response: Thank you for the great questions. We addressed these questions by adding more information regarding the computational efficiency of the proposed DA-hydrodynamic modeling framework and its benefits in practice. Please refer to lines 619-635 in the revised manuscript.

Please italicize all variables/parameters in the text.

Response: Done.

Any rational for considering a normal distribution for the model boundary conditions with a mean of zero and relative error of 20%? If it is derived from previous studies, please point this out the first place you present this assumption. You should also discuss the validity of your assumption by discussing how close the previous studies are to your study.

Response: We added two main references that explain the error of observed flows varies in the range 8-25%. However, the basis for choosing the best error term in the assimilation works is the trial and error and the manual tuning of the values to achieve the most reliable predictions. Please refer to lines 324-327 and lines 368-370 in the revised manuscript.

Any rational for considering a normal distribution for the model initial conditions with a mean of zero and relative error of 1 m? If it is derived from previous studies, please point this out the first place you present this assumption. You should also discuss the validity of your assumption by discussing how close the previous studies are to your study.

Response: The error term for the bathymetry is chosen from expert judgment, trial and error and tuning to achieve the most reliable predictions. It is quite common in DA studies that we choose errors based on manual tuning if there is very little a priori knowledge about the physically correct error terms. We added more information about the logic behind our selections. Please refer to lines 338-340 in the revised manuscript.

Any rational for considering uniform distributions for the channel roughness and bathymetry? Where are the ranges—(0,0.1) and (39,42)—coming from? If it is derived from previous studies, please point this out the first place you present this assumption. You should also discuss the validity of your assumption by discussing how close the previous studies are to your study.

Response: For the channel roughness, we already provided the references in the manuscript. The range of bathymetry parameters is again estimated from expert judgment and trial and error. It is quite common in DA studies that we choose errors based on tuning if there is very little a priori knowledge about the physically correct error terms. Please refer to lines 365-370 in the revised manuscript.

To perturb the uncertainty of parameters via Latin Hypercube Sampling, how many samples were settled? How did you ensure that the samples are sufficient for the numerical convergence?

Response: The number of samples is equal to the ensemble size of the experiment. Therefore we generated 100 samples from the uniform distributions. According to several past studies that explored the sensitivity of EnKF to the ensemble size, the model performance is highly sensitive to small ensemble sizes less than 50 but the sensitivity will be reduced by moving to larger numbers around 100 (e.g. see Figure 4 in Moradkhani et al., 2005 and Figure 2 in Gillijns et al.,

2006). With access to the sufficient number of computing cores in the University of Alabama High-Performance Computing cluster, we were able to use a relatively large number of 100 ensemble members (It is higher than most of the studies that used the EnKF method). This ensures that our results will provide a reasonable range of uncertainties.

The three experiments presented under the Results section, should be clearly defined and described under the Methodology section. As of now, it is hard to follow the difference between the three experiments and details of each.

Response: Thanks for making this point. We revised the manuscript by adding required texts, and changing the headings of both method and results sections.

Figures 6 and 7 should be merged.

Response: Using one single figure representing inundation maps for all 6 days reduces the size of subplots. We believe that using smaller subplots for each inundation map will degrade the quality and decrease the usefulness of the figure as the details will not be observable anymore. Thus, we feel it will be better to keep this format as one figure shows the upper limb and the other one highlights the lower limb of the flood hydrograph.

In addition to the inundation extent, it will be useful to compare the performance of EnKF with the OL on other flood characteristics such as depth.

Response: We already compared the EnKF and OL performance for simulating water depth and discharge in Figure 4. In fact, our validation strategy includes two components: 1. Validating the dynamics of the DA-hydrodynamic modeling framework at point sources (Figure. 4): Here, we focus on water depth and discharge simulated by OL and EnKF and compare their performances. 2. Validating the spatial behavior of the DA-hydrodynamic modeling framework (Figure. 8): Here, we compare the flood extents simulated by the OL and EnKF.

Broader impacts need to be discussed. The authors should discuss what implications these results have for urban planners and floodplain managers etc. and what existing programs in the US may benefit from this research.

Response: The broader impacts of the proposed DA-hydrodynamic modeling framework for applying in practice were further discussed. Please refer to lines 619-635 in the revised manuscript.

Study limitations and potential areas for future research need to be expanded.

Response: The final paragraph of the manuscript was expanded by adding limitations while suggesting more areas of research for future studies. Please refer to lines 667-672 in the revised manuscript.

Please discuss how your presented DA-hydrodynamic modeling framework can be used in other study areas. What considerations should be taken to do so?

Response: Thanks for the great suggestion. We added a new paragraph to the end of the manuscript where the required considerations for applying the proposed framework to other studies are further discussed. Please refer to lines 675-686 in the revised manuscript.

References:

Gillijns, S., Mendoza, O.B., Chandrasekar, J., De Moor, B.L.R., Bernstein, D.S., Ridley, A., 2006. What is the ensemble Kalman filter and how well does it work?, in: 2006 American Control Conference. Presented at the 2006 American Control Conference, p. 6 pp.-. https://doi.org/10.1109/ACC.2006.1657419

Moradkhani, H., Sorooshian, S., Gupta, H.V., Houser, P.R., 2005b. Dual state–parameter estimation of hydrological models using ensemble Kalman filter. Adv. Water Resour. 28, 135–147. https://doi.org/10.1016/j.advwatres.2004.09.002