Response to the reviewers

Editor Response

Associate Editor Evaluations:
Recommendation (Required): Return to author for major revisions
Accurate Key Points: Yes

Associate Editor (Remarks to Author):

Dear authors,

Your manuscript has been re-assessed by two of the original reviewers and one new reviewer. The two original reviewer acknowledge the improvements made with respect to the earlier version and point out a few additional areas where further improvements are needed. The new reviewer stresses that some further major revisions in terms of level of detail and storytelling are needed and requests some improvements with respect to model evaluation. I would like to give the authors another opportunity to address both the major and minor concerns and am looking forward to reading a revised version of the manuscript.
Best regards,
Your AE

Reply: We thank you, the editor, as well as the reviewers for carefully reading the manuscript and for providing valuable feedback to improve its quality. We have gone through the reviews and made every effort to address any specific concerns while staying within the scope of the paper. We hope this revised version of the manuscript meets the standards for publication and addresses the requests from the reviewers. In instances where we disagreed with the suggestions, we have provided a well-reasoned explanation for why we felt the change was not necessary.

Reviewer 1

Introductory Comment

Reviewer Point P 1.1 — Recommendation (Required): Return to author for minor revisions
Significant: Yes, the paper is a significant contribution and worthy of prompt publication.
Supported: Yes
Referencing: Yes
Quality: Yes, it is well-written, logically organized, and the figures and tables are appropriate.
Data: Yes
Accurate Key Points: Yes

Reviewer #1 (Formal Review for Authors (shown to authors)):

I have read the revised version and want to thank the authors for addressing all questions and concerns I raised in the first round. As a general note, it would be very helpful if authors could be specific as to where the manuscript has been changed (providing line numbers in the revised version).

The proposed approach (computing HAND based on level paths instead of nearest drainage) is now better clarified in the methodology. The newly added Fig. 7 helps readers understand the concept and novelty of the developed approach. The results and discussion sections are extended, and they comprehensively outline the method's limitations. Most of my comments were addressed, and I feel the revised manuscript is suitable for publication in WRR. I recommend, however, that the manuscript be returned to the authors for the following minor edits before publication.

Reply: Thanks for going through our paper for a second time and providing numerous suggestions for improvement. We were glad that you found that all of the original questions and concerns were properly addressed. We hoped to have adequately answered your final minor suggestions properly as well.

Comments

Reviewer Point P 1.2 — I believe the last part of the Fig. 9 caption, starting with "The BL stream network is also shown which is denser than ... This creates additional ..." should also be added to the discussion section as one of the sources of uncertainty or errors. Perhaps somewhere between lines 1025-1039.

Reply: Thank you for identify an important source of error that we have identified and suggesting it go within the Discussion section. Based on your suggestion, we have added a few sentences that better discusses the conflation errors that could occur by using BLE associated streamflows with the NWM network. These sentences we placed near the end of the paragraph you suggested the discussion be located around line 1030.

Reviewer Point P 1.3 — The level path approach leads to delineating catchments with larger sizes. The authors made this point discussed in the discussion section. However, there is less

discussion about the stream lengths in the discussion section. For example, what are the average, minimum, and maximum sizes for the stream length, L, within these large catchments? This is important because L impacts the stage-discharge relationship. As far as I know, previous work conducted some sensitivity analysis to show a proper range for L in HAND calculation (see references below).

- Godbout, L. D. 2018. "Error assessment for Height Above the Nearest Drainage Inundation 935 Mapping", Master of Science in Engineering thesis, The University of Texas at Austin, 936 https://repositories.lib.utexas.edu/handle/2152/68235.

- Zheng, X., Maidment D. R., Tarboton D. G., Liu Y. Y., & Passalacqua P. (2018a). GeoFlood: 1054 Large-Scale Flood Inundation Mapping Based on High-Resolution Terrain Analysis. 1055 Water Resources Research, 54. https://doi.org/10.1029/2018WR023457

Reply: Thank you for pointing out this important parameter in the behaviour of synthetic rating curves (SRCs). We agree that SRCs are sensitive to a variety of values and one of which as you have pointed out is reach length. The two citations you provided reviewed or mentioned this issue in detail so we did cite these references throughout the paper. While cited throughout the paper for various reasons, they were or are now used in the context of reach length on lines 149-150, 485-486, and 565. On line 485, we mention that the reach length is capped to 1.5 km which is consistent with the findings from the Godbout paper and also consistent with what Zheng used within GeoFlood. Additionally, we kept this constraint for all versions of HAND that we computed including FR, MS, and GMS, thus we did not observe a notable difference in reach length across the three versions. Due to this, we did not include it in the discussion of the paper. Nevertheless, we decided to add a sentence on 486 acknowledging that this parameter was held fixed at 1.5 km for all versions of HAND. This constraint is very important since we did not find a significant difference in the reach lengths. To demonstrate this we computed the averages and standard deviations for the two primary versions FR and GMS. We found GMS to have a mean derived reach length of 1398.1 km with a standard deviation of 50.2 km. The FR version had a mean of 1354.8 km with a standard deviation of 249.7 km. We found this difference in reach lengths across FR and GMS to be statistically insignificant. In accordance with this, we added these numbers specifically starting on line 983 to the discussion section to inform readers that reach length, albeit an important variable, does not contribute to variance in SRC values across HAND versions.

Reviewer Point P 1.4 — The first paragraph of the conclusion can be removed as it is more like an introduction of the flood inundation mapping and HAND method.

Reply: Based on your suggestion, we completely removed the opening paragraph of the conclusion. Thanks for providing specific feedback to help reduce some of the superflous content.

Reviewer 2

Introductory Comment

Reviewer Point P 2.1 — Reviewer #2 Evaluations:
Recommendation (Required): Publish in present form
Significant: Yes, the paper is a significant contribution and worthy of prompt publication.
Supported: Yes
Referencing: Yes
Quality: The organization of the manuscript and presentation of the data and results need some improvement.
Data: Yes
Accurate Key Points: Yes

Reviewer #2 (Formal Review for Authors (shown to authors)):

The main concerns from my last round of review have been addressed in this revision. As I mentioned last time, there is significant value to expose operational research approaches adopted by federal agencies to general academic audiences for broad discussion and rigorous review. Therefore, I recommend accepting this paper. Some issues that still need to be fixed are listed below:

Reply: We appreciate the time and energy dedicated in producing your reviews as they have greatly enhanced the quality of our manuscript. We have gone through your latest round of suggestions and feel we have addressed them as best as possible.

Comments

Reviewer Point P 2.2 — Figure 5 It's better to pick a headwater basin for illustration. The current spot is still a little misleading: the upstream purple link with a label of 1667 is assigned as a part of the mainstem LP. However, in this figure, it's shorter than the yellow branch above with labels like 15 and 7. If this is the full network, I think that the yellow branch should be chosen as a part of the mainstem LP, instead of the current one.

Reply: Based on your advice, we elected to highlight a small headwater subset of the stream network in the HUC8 12020002. This subset avoids inheritance issues from using nested HUCs.

Reviewer Point P 2.3 — Figure 6 (a) The symbology needs some extra work. The two demo LPs should be highlighted with a thicker line.

Reply: Based on your comment, we increased the line weights of the stream lines and catchments to make them more visible. We hope this thickness stands out better to the readers.

Reviewer Point P 2.4 — Figure 8 Remove the extra axis stickers on the upper boundary.

Reply: Thanks for the attention to detail. The extra tick marks have been removed.

Reviewer Point P 2.5 — Figure 8 9 12 It's better to present results at the same site throughout all the sections, instead of zooming in and out to different parts of your study area. The authors show cross-sections at one site and then jump to another location or zoom level to show the inundation extent. There is no intrinsic connection between these plots in this case.

Reply: Please note for the response below that Figures 8, 9, and 12 correspond to 7, 8 and 10 now, respectively, as some changes were made in the course of this review. The original Figure numbers are referenced below.

After giving this some thought, we have elected to maintain the same locations for the three suggested figures. We do appreciate your concern to maintain consistency across plots which could, in a way, facilitate understanding for readers as their context would stay more fixed. However, we believe these three figures convey different concepts and ideas, thus, are best represented by utilizing separate sites and scales. To be specific, Figure 8 is meant to convey the spatial extents of the study domain. In order to do this, we must take a very high-level view of an entire region of the US. In this large scale, the central tenants of Figures 9 and 12 would not be able to be properly conveyed.

Figure 9 is meant to illustrate a cross-walking technique for assigning recurrence flows at two intervals, 100 and 500 year, from HEC-RAS derived BLE cross-sections. These cross-sections are intersected with the NWM reaches and the intersecting flow values are aggregated utilizing a median filter. This site was selected because it illustrates two common sources of error with this technique. One is the lack of certain headwaters represented in the BLE but not the NWM stream network. This leads to FNs in the HAND inundation when compared to BLE extents. Another source of error is the fact that cross-sections could span stream segments of varying stream order, thus have large differences in flow rates. While most of these are addressed by taking the median value, some may suffer from errors which could violate mass conservation assumptions. We wanted to pick a site that graphically conveyed our method that does reduce hydro-climatic uncertainties but also be true to the limitations of the technique. We don't feel that representing this in the same map frames as Figures 8 or 12 would convey this message to the readers.

Figure 12 is meant to convey the impact of computing HAND with overlapping, reduced stream order spatial units on the extents of the resulting FIM. Due to the nature of only accounting for a single fluvial source within HAND, this problem is most evident in areas with very large mainstems with high flows that are directly fed by lower order and flow tributaries. When this occurs, it creates a significant bottle-necking effect in the inundation because the inundation from the mainstem cannot influence the inundation within the neighboring catchments tributaries. For this reason, we wanted to select this region due to its high flow mainstem and many lower flow tributaries that impede inundation with the HAND method. By computing HAND independently for each level path and mosaicing the inundations together, we are able to account for inundation from multiple fluvial sources. This region conveys that concept well in terms of inundation extents which is the main focus of the paper. Conveying this idea within the map frame of Figure 8 would not give the reader a close enough look as to what is actually happening. The mapframe in Figure 9 would, indeed, illustrate a positive change in the spatial extents but due to the lower stream order nature of this region it would not have the same impact as utilizing the higher order region in the current Figure 12 map frame would.

We hope that this explanation is satisfying to you as we are just trying to convey different concepts with each figure and felt that utilizing different regions that best convey the messages would be most

appropriate. We do appreciate that switching across regions could be cognitively expensive for the reader but we hope that this price is rewarded by ensuring our explanations are complete.

Reviewer 3

Introductory Comment

Did not repeat review. Skip ahead to reviewer #4 below.

Reviewer 4

Introductory Comment

Reviewer Point P 4.1 — Reviewer #4 Evaluations:
Recommendation (Required): Return to author for major revisions
Significant: The paper has some unclear or incomplete reasoning but will likely be a significant contribution with revision and clarification.
Supported: Yes
Referencing: Yes
Quality: The organization of the manuscript and presentation of the data and results need some improvement.
Data: Yes
Accurate Key Points: Yes

Reviewer #4 (Formal Review for Authors (shown to authors)):

This manuscript proposes a new version of HAND-based FIM that accounts for multiple sources of fluvial inundation and results in an improvement in overall FIM skill. The topic of this manuscript for improving the FIM skill at such a large scale is of interest to the flood community and adds value to the literature. However, I have some main concerns mostly regarding how it has been written for a scientific journal and the validation parts. Since I was not the reviewer in the previous round, I decided to only read the revised version and add my comments. Later on, I checked the reviewer comments in the previous round and interestingly I found several common concerns that have not been fully addressed yet. Please see my two major comments and the following minor ones below:

Reply: Thank you for reviewing the paper and for providing feedback to help improve its content. We hope we have interpreted your requests correctly and assumed the correct solution to many of the problems you noted. Every effort has been made to make sure your corrections have been well addressed.

Major Comments

Reviewer Point P 4.2 — The language of this manuscript, especially the method section is too detailed and technical, it includes many jargons, and is long for a scientific journal. More specifically, I would like to point to sections 2.3 and 2.4.

Reply: We have gone through your sub-comments below and have made substantial changes to reduce detailed jargon. Please note that some of these details were added in response to comments from previous reviewers looking for specific, technical clarifications on detailed issues. We hope these changes streamline the paper while also maintaining the details necessary to communicate the differences of our approach compared to previous efforts.

Reviewer Point P 4.3 — Section 2.3.1 includes too much detailed information that may not be that much beneficial for the majority of the Journal audience who are not experts in this area. I don't think these detailed technical guidelines are required to be in the main body of the manuscript as it may make the readers reluctant to follow the rest of the manuscript. In addition, since the main novelty and contribution of this manuscript is section 2.5, it would be more reasonable if the authors focus on this section as the core of the method section and make other parts as concise as possible. My suggestion is that the authors move additional information to the appendix so those specific audiences who are interested can follow details such as DEM preprocessing, etc. Similarly, sub-sections 2.4.1, 2.4.2, and 2.4.3 can be joined under a concise section and additional information can be moved to the appendix. Figure 3 can be also moved to the appendix.

Reply: Thank you for suggesting specific changes that could potentially lighten the paper. As per your suggestion, we have greatly reduced the content of Section 2.3.1 and moved most of the details to the Appendix. We also felt the same about Section 2.4.1 as most of the details are only of concern to a specific subset of the community that deals with the computational aspects of hydro-conditioning operations. We greatly condensed this section and moved the remainder to the Appendix.

As per Section 2.4.2, the other reviewers were very keen as to how the stream network was defined and how the reach lengths were computed. These concerns were rightfully brought up as previous research has shown how reach averaged SRCs are sensitive to sampling which is the reach length parameter in this context. Much effort was taken to ensure these concerns by these experts were properly addressed and wish to ensure these details are stated in the manuscript. Several other points in the paper including some in the Discussion section allude to this important parameter that the community has identified. Due to this reason, we wish to maintain this section in the paper.

Lastly with Section 2.4.3, derivation of the catchments is a simple procedure and not much background is stated here. Most of the section is dedicated to explaining pieces of information that make the rest of the manuscript understandable. This includes the fact that two sets of catchments are derived including the pixel scale and reach scale catchments. Without this, the reader would be lost in other parts of the paper. This section is also very short so moving a few sentences to the appendix would not make much sense. For these reasons, we also elected to keep this section in place.

Reviewer Point P 4.4 — Using too much jargon make the paper hard to understand and follow by the majority of readers who are not a technical expert in this area. I always encourage the authors to avoid these terminologies as they adversely affect the quality of the manuscript. Some examples are "etch or burn flowlines" (line 210), "BurnlinEvents are pruned" (line 351), and "The elevations found in the NLD are burned onto the DEM" (line 399).

Reply: Based on your feedback, we have elected to remove the terms "etch" and "burn" in favor of the term "enforce" which we feel has a broader understanding and less likely to be taken as geospatial jargon. The term "pruned" was either removed or replaced with more generally acceptable words. Lastly, "BurnLineEvents" was maintained as it is a term inherited from the NHD to differentiate it from other stream networks contained within the NHDPlusHR. We did wish to maintain technical consistency with the source datasets.

Reviewer Point P 4.5 — The amount of text provided under the method section is too much. It is highly recommended to keep a balance among the lengths of different sections of a manuscript. As an example, the manuscript has 12 figures while the first 9 figures belong to the method section

and only 3 figures have been developed for the result section. Typically we expect to see more figures in the result section.

Reply: Thanks for your observation noting the imbalances in the sections of the paper. Due to your previous comments suggesting specific changes, we have greatly reduced the amount text and figures within the methods sections. Due to your suggestions, we have moved two figures out of the methods and into the introduction or the appendix. With those figures a considerable amount of text has been removed from this section which we believe better balances the paper out. Our methodology section is now greatly reduced albeit still dense in order to adequately explain the many techniques employed in developing continental scale FIMs, addressing fundamental limitations with HAND, and validating the proposed enhancements. Thanks for your specific suggestions addressing ways to reduce information within the methods section and the paper overall.

Reviewer Point P 4.6 — The evaluation approach used to validate the effectiveness of the proposed HAND-based technique for FIM is missing some important components:

Reply: Please see responses to specific points below.

Reviewer Point P 4.7 — The reference data used in this manuscript is FEMA flood extent maps corresponding to synthetic storms (100-year and 500-year floods). However, the authors claim that the application of their method is for real-time flood inundation mapping. In general, the mechanisms and approaches used for detecting flood extent corresponding to synthetic storms differ from those used for real-time flood inundation mapping. If the authors want to show that their approach is beneficial for real-time flood inundation mapping, they need to demonstrate this by simulating real flood events (not synthetic cases). Following this concern, I found a meaningful portion of the introduction, discussion and conclusion refer to the operational forecasting by NWM streamflow forecasts and AHPS systems. However, the manuscript has not used any of these forecasts provided by NWM or other operational agencies. To me, this is an inconsistency between what has been implemented (synthetic storm simulations) and what is mentioned as the goal of this method (forecasting real flood events by NWM).

Reply: We appreciate your concerns regarding the use and evaluation of the specific FIM model. We would like to take the opportunity to clearly articulate our reasoning and provide a more comprehensive explanation. We are confident that our clarifications and proposed revisions to the manuscript will effectively address your issues, while preserving the core objectives of the paper.

As outlined in the sub-section titled "OWP FIM" in the Introduction, there are two main objectives of the paper. The first objective is to introduce, explain, and assess a novel HAND implementation that is being utilized for operational flood forecasting alongside the NWM. The second objective is to identify a limitation with HAND-based FIMs, present a new solution, and evaluate its effectiveness by quantifying and analyzing the improvements. Providing a detailed explanation of the first objective provides the reader with the necessary background to understand the motivation and context behind this model's creation. It is important to note for the hydrology community that this model is considered operational, as they may find it relevant for future work. However, we acknowledge that referring to the model as operational and then evaluating it using risk-based, synthetic data may cause confusion, and we aim to clarify this.

It is correct to state that the process of creating a synthetic FIM differs from the process of producing a FIM for operational forecasting when the entire hydro-climatic process is taken into account. The key

difference is that HAND operates somewhat independently of this process, as it couples with hydro-climatic processes through catchment scale depths. By coupling HAND with SRCs, we have the ability to accept streamflows, which serve as the main point of connection with the NWM and any other streamflow sources we may use for retrospective, monitoring, forecasting, or evaluation purposes. This makes HAND somewhat exclusive of the hydro-climatic processes modeled "upstream" and focuses solely on computing inundation depths and extents. HAND is agnostic to the nature of the streamflow values it receives, be it synthetic or operational, and only concerns itself with the magnitude of the streamflow values.

It's crucial to examine multiple sources of flood extent information for evaluations as each has its own limitations and level of uncertainty. The sources examined include using high water marks, remote sensing, and hyper-local flood inundation models. It's important to acknowledge the limitations and advantages of each source. For instance, agencies like the USGS create FIMs by combining local HWMs with a DEM. HWMs and FIMs based on a cluster of HWMs can serve as a reliable source of validation data for large-scale, operational models. RS-based FIMs can also provide valuable validation information, depending on factors such as the type of sensor, the flooded environment, atmospheric conditions, and more. FIM models are often compared to RS-based inundation extents in academic literature as they provide information where no other source of observed or modeled data could be available. Additionally, hyper-local flood inundation models, such as those produced by the AHPS mentioned in the paper and by the USGS' Flood Inundation Mapper program, also provide high-quality information that can be used to validate large-scale inundation proxies like the OWP FIM. All of these methods are potentially suitable for operational evaluation as they reflect real-world flood events but also subject to limitations themselves.

In order to use the three alternative sources of benchmark data mentioned previously in an operational capacity with HAND, one must utilize operational streamflow values. These streamflow values would come from the NWM if we are looking to evaluate operational forecasts. If operational observation data were considered, we could have used gage information. Utilizing the NWM offers a dense availability of streamflow data in the forecasting context that OWP FIM would be intended to be used for in an operational setting. Using NWM streamflow data would allow us to account for multiple fluvial sources of inundation which would be required to evaluate our technique. However, the limitation of using NWM streamflows would be introduction of hydro-climactic uncertainties that would add error and variations in our analysis outputs. This could modify, dilute, or add noise to the effects of stream order reduction in order to account for multiple fluvial sources with HAND. Another aspect to take into account is that relying on observational data from gages to obtain streamflows would reduce the number of streamflow values accessible in a specific area, thereby spatially restricting the observations. A similar problem would arise by using the point stages or streamflows that hyper-local FIMs offer. These single point samples of streamflow data would lack the network density required to assess the performance of a proposed technique interested in accounting for multiple fluvial sources of inundation with HAND. This would likely lack sufficient information to properly account for cases presented throughout the paper where low flow tributaries converge with high flow, higher order streams. These two issues brought about by using NWM streamflows and single point streamflows from modeled or observed gages deterred us from using these methods to evaluate our version of HAND and its proposed improvements. Even though we found these three sources to be less than optimal for our purpose, we don't advocate ignoring them. We intend to use them for on-going and future evaluations of forecasting capabilities as they offer advantages themselves, but for the time being we deemed the BLE to be a good evaluation dataset for HAND based FIMs and the principal modification proposed.

Based on our analysis, we concluded that the BLE, a synthetic dataset, was the most suitable option for our study. By using the BLE and its accompanying streamflow data, we were able to exclude the hydro-climatic uncertainties that would have been present if we used modeled streamflow values from other sources. Moreover, the ample supply of streamflow data from the BLE enabled us to accurately evaluate the impact of our suggested modifications for multi-fluvial sourced flooding with HAND. After conducting a literature review, we discovered several studies that employed synthetic datasets to evaluate models for operational purposes (Wing et al., 2017; Hu & Demir, 2021; Li, Mount, & Demir, 2022; Li & Demir, 2022; Zheng, Maidment, Tarboton, Liu, & Passalacqua, 2018; Wing et al., 2017; Afshari et al., 2018). Consequently, we decided to utilize the BLE for our study, however, it is crucial to acknowledge the requirement for continual evaluations using a range of benchmark datasets across broad geographic areas. This is a challenging matter that calls for sustained efforts from various viewpoints, and we therefore appreciate and consider your feedback as valuable.

We hope we alleviated any concerns about the suitability of using a synthetic dataset, the BLE, for evaluating a model intended for operational forecasting, OWP HAND based FIM. We concur that the paper did not provide a logically sound explanation for the conclusion to employ a synthetic evaluation for a model intended for operational use. It appeared that the appropriate location to offer this explanation was in the first paragraph of the Evaluation section. The current explanation appeared insufficient, so we significantly strengthened it to ensure that the reader fully understands the benefits of using the BLE and the limitations avoided by not using alternative benchmark datasets. Hopefully this revised paragraph conveys our justification in a concise manner.

We conducted a thorough review of the manuscript and looked for references to operational forecasting especially in the sections you mentioned which were the Introduction, Discussion, and Conclusions. The only reference to operational forecasting in the Conclusions section was found in the opening paragraph, which was removed upon request from another reviewer as it was considered redundant in summarizing the paper. In the Discussion section, we found no explicit references to operational forecasting as the focus of the section was to examine the impact of the primary modification proposed for HAND. Some room was allocated in the section to address other sources of error and to present future possibilities for improvement in this field. In the Introduction, two sub-sections, titled "Operational Forecasting" and "National Water Model," discuss operational capabilities in detail. As previously mentioned, the purpose of this implementation of HAND is for operational forecasting, making it crucial to provide context on past operational systems and the NWM. We defend the presence of the sub-sections titled "Operational Forecasting" and "National Water Model" by highlighting the rationale behind them:

- Introduce legacy forecasting systems including their limitations in the aspects of spatial and temporal coverage.

- Convey the fact that the NWM fills in these gaps by greatly expanding the forecasting points in both space and time.

- Acknowledge the value of assimilating legacy forecasts into NWM forecasts.

- Introduce the term hydrofabric which is the collection of geospatial datasets used for NWM purposes including catchments, stream network, and reservoirs.

- Mention the use of the two stream networks with varying density including the full resolution (FR) and its subset mainstems (MS) which is defined as all FR streamlines at or downstream of AHPS forecasting points.

- This network also allowed for computing HAND with a stream network of lower drainage density which demonstrated our multiple fluvial inundation technique for 4% of the modeling domain.

We acknowledge that some readers would not be interested in learning details about these authoritative forecasting systems but many readers would. One of your review questions below asked for some details on the AHPS for which we re-included previously removed information. These sections also included significant background on the hydrofabric which serves as the geospatial backbone of both the NWM and OWP HAND FIM. We wish to pay respect to these key items that provide the background for the community while hopefully not taking away too much from the main scope of the paper. In lieu of this, we opted not to remove any more information concerning references to operational forecasting. We hope we have offered sufficient justification as to why.

Reviewer Point P 4.8 — The authors should clearly explain if their method is only used for flood extent mapping (binary mapping) or if it is useful for flood depth mapping. According to Eq. 4, this method can be used to estimate the flood depth as well. Thus, some metrics (such as RMSE) that show the effectiveness of this method for flood depth mapping are required and should be added to the manuscript. The reference data for this evaluation could be High Water Marks (HWMs). The current validation approach only tests the effectiveness of the proposed method for flood extent mapping.

Reply: According to previous work, HAND is technically able to produce both inundation extents and depths (Nobre et al., 2016; Maidment, 2017; Garousi-Nejad, Tarboton, Aboutalebi, & Torres-Rua, 2019; Li, Mount, & Demir, 2020; Li et al., 2022; Li, Duque, Grout, Bates, & Demir, 2023). We want to acknowledge the capabilities of HAND by pointing out that it can produce both, and by thresholding the depths, we can obtain extents. However, the OWP is only producing FIM extents as a product with only plans in the future to offer depths as well. For this reason, FIM extents is of principal interest to OWP and of primary concern in this study. We acknowledge the importance of flood depths and understand that extents and depths have a nuanced relationship. We certainly value the need to evaluate depths moving forward but do to constraints we have opted to ensure that the methods proposed here were evaluated for extents. Additionally, there are a number of studies with HAND and other FIM techniques that are able to produce both spatially distributed extents and depths or elevations but only chose to evaluate extents (Garousi-Nejad et al., 2019; Li et al., 2023, 2022; Papaioannou, Loukas, Vasiliades, & Aronica, 2016; McGehee, Li, & Poston, 2016; Zhang et al., 2018; Shastry et al., 2019; Follum, Tavakoly, Niemann, & Snow, 2017; Wing et al., 2017; Aristizabal & Judge, 2021; Rahimzadeh, Bahremand, Noura, & Mukolwe, 2019; Amarnath, Umer, Alahacoon, & Inada, 2015; Wood et al., 2016; Hooker, Dance, Mason, Bevington, & Shelton, 2022; Tansar, Babur, & Karnchanapaiboon, 2020). In the future, we hope to produce more evaluations to cover depths and probabilistic information but for now constrain the scope of this study to only include extents for evaluation purposes.

In terms of the content of the manuscript, we wish to ensure that your concern is properly accounted for. There is a need to convey to the readers that we are only evaluating our methods on FIM extents and not depths. It is also important to convey that although related effects on FIM extents don't translate the same to FIM depths. Evaluating FIM depths would make the study more robust but was simply left out of scope due to the many other topics to cover and for organizational priorities. To account for these key points, we have removed any superfluous mentions of FIM depths that don't tie into the fact that they are computed prior to achieving extents. Additionally, we have added some notes in the Discussion section that acknowledge the limitation in our evaluation methodology by not including depths.

Minor Comments

Reviewer Point P 4.9 — Lines 84 and 109: What are the hydrologic/hydrodynamic models used by AHPS for streamflow forecasting and flood inundation mapping? Are AHPS more accurate and reliable than NWM forecasts?

Reply: We previously had a deeper explanation on the AHPS in Section 1.1 and removed most of the content due to it being only related to the scope of flood inundation mapping with HAND. This was partly done in response to requests to remove irrelevant content from previous reviewers. In lieu of your questions, we have elected to maintain the current explanation while adding a shorter sentence to use the references we cited originally. This should give the reader adequate information to pursue if more information is required on their behalf. With regards to your second question, yes, the Office of Water Prediction does indeed consider AHPS to be of high quality. This is alluded to on L108-112 of the original document that mentions that AHPS forecasts are assimilated in the NWM by replacing the NWM at those respective locations and routing that flow downstream. We termed this assimilation to be incorporating the best available regional scale forecasts. We hope that this sufficiently answers your questions while also not adding too much more to the paper.

Reviewer Point P 4.10 — Figure 1: AHPS Location and NWM FR Streams provided in the legend is hardly visible and distinguishable in the figure. Please give a better illustration.

Reply: This figure is challenging because it attempts to convey many pieces of fine scale information across very large spatial scales. We have iterated on this a few times trying to optimize the illustration to best account for the shear scale and density of information. Previous reviews have stated that the underlying networks were too thin and needed to be thickened. This has possibly drowned some of the points. We value your feedback and make an attempt to revise the figure to see if we can best account for everything just a little bit better. In our latest effort which is motivated by your feedback, we have elected to expand the image as to take the maximum advantage of limited space in the horizontal direction. Hopefully, this gives the reader more scale to operate with. The NWM MS stream lines were transitioned to green while the AHPS locations without FIM were transitioned to red. This is likely to contrast with one another more. Additionally, we amplified the sizes of the legend objects so the readers can clearly identify what the symbols should be on the map. We hope this representation best communicates all the datasets better given the steep constraints.

Reviewer Point P 4.11 — Lines 148-152: In one of the recent comparisons, Hocini et al., (2020) compared the performance of the HAND method with 1D and 2D hydrodynamic models and showed that HAND performance is weaker in many cases.

- Hocini, N., Payrastre, O., Bourgin, F., Gaume, E., Davy, P., Lague, D., Poinsignon, L., Pons, F., 2020. Performance of automated flood inundation mapping methods in a context of flash floods: a comparison of three methods based either on the Height Above Nearest Drainage (HAND) concept, or on 1D/2D shallow water equations. Hydrol Earth Syst Sci Discuss Httpsdoi Org105194hess-2020-597

Reply: Thank you for furnishing this important and relevant reference. Given that this method evaluates and compares the use of HAND with models that have higher order physics, we found it necessary to include in this paper. It is now cited in multiple locations.

14

Reviewer Point P 4.12 — Lines 444-445: Following my major comment, I think the majority of readers would like to know what HAND, SRCs, and the cross-walk table are. However, the main focus is on detailed information about geospatial analysis used to get DEM, streamlines and etc.

Reply: Due to the ambiguity of these terms as they are stated on the lines you referenced, we have added a few sentences there to help explain these three geospatial objects before they are deep-dived later in the section.

Reviewer Point P 4.13 — Lines 512-515: This is not a straightforward approach to explaining HAND calculation. There are many better ways to explain it so that everyone can understand it. Please reword it.

Reply: We have reworded large portions of this paragraph hoping to more clearly explain how HAND is computed while also staying true to the level of detail requested by previous authors. Specifically it was asked to elaborate on the use of the native elevations outside of the AGREE conditioned regions. We did our best here so if you have any specific suggestions those would be appreciated and help us better address your concerns.

Reviewer Point P 4.14 — Lines 539-543: what is known and what is unknown in this equation? Are we trying to estimate Q or y? Please add more information.

Reply: Thank you for asking for clarification on this. In practice, we pick discretized stage values and compute the discharge for the synthetic rating curve. The last sentence of this section describes that we elect 84 samples between 0 and 25 meters in increments of 1/3 m. In order to avoid ambiguity, we have elected to move this sentence to the end of the paragraph immediately proceeding after equation 1. We hope this better conveys how we constrain the equation to compute discharge for SRCs.

Reviewer Point P 4.15 — Lines 569-571: I think several parameters of Manning's equations are dependent on the value of stage so depending on any of these 84 stage values we have different parameters. Please clarify it in the text here.

Reply: Based on your suggestion, we have elected to add a sentence at the end of this paragraph that explicitly states that the parameters V(y) and B(y) which are functions of the stage, y, also change when the stage changes.

Reviewer Point P 4.16 — Lines 609-612: This is the main novelty and contribution of this manuscript. I would recommend the authors move this information (texts and figures) to the introduction and specifically mention that this is the problem they are seeking to address. It is better that transfer motivation and objectives to the beginning of the manuscript and only explain the method in this section.

Reply: Due to your recommendation and to better motivate the study, we moved most of the discussion you elected to the last subsection of the Introduction titled "Office of Water Prediction Flood Inundation Mapping." The paragraph you alluded to introducing our scale reduction techniques has been greatly reduced and, in doing so, removed a Figure from the already dense methods section.

Reviewer Point P 4.17 — Line 619: Please specify what unary stream order networks are. It seems this is a key point.

Reply: Good catch. Based on the fact that this is now the only use of the term unary, we removed its use in lieu of using "stream networks of unit stream order" which is more generally accepted.

Reviewer Point P 4.18 — Lines 629-631: Please reword this part. It is not clear whether this method aims to improve the computational demand or improve the accuracy by reducing the negative effects of nearest drainage or both.

Reply: Thanks for catching this ambiguity. We have adjusted to specify that we intend these techniques to enhance the quality of FIM extents not improve computational performance.

Reviewer Point P 4.19 — Line 761: Associated streamflow data.... Like what data?

Reply: We decided to remove the term "associated" in favoring the phrase "corresponding to the flow peak" which adds specificity to the term "streamflow data."

Reviewer Point P 4.20 — Line 931: Figure 12 is referenced before Figure 11 so please replace their order.

Reply: Another good catch. We have swapped the order of the figures thus Figure 12 is now Figure 11 and vice versa. Please note that the original Figure numbers on these have changed.

References

Afshari, S., Tavakoly, A. A., Rajib, M. A., Zheng, X., Follum, M. L., Omranian, E., & Fekete, B. M. (2018). Comparison of new generation low-complexity flood inundation mapping tools with a hydrodynamic model. *Journal of Hydrology*, *556*, 539–556.

Amarnath, G., Umer, Y. M., Alahacoon, N., & Inada, Y. (2015). Modelling the flood-risk extent using lisflood-fp in a complex watershed: case study of mundeni aru river basin, sri lanka. *Proceedings of the International Association of Hydrological Sciences*, *370*(370), 131–138.

Aristizabal, F., & Judge, J. (2021). Mapping fluvial inundation extents with graph signal filtering of river depths determined from unsupervised clustering of synthetic aperture radar imagery. In *2021 ieee international geoscience and remote sensing symposium igarss* (pp. 6124–6127).

Follum, M. L., Tavakoly, A. A., Niemann, J. D., & Snow, A. D. (2017). Autorapid: a model for prompt streamflow estimation and flood inundation mapping over regional to continental extents. *JAWRA Journal of the American Water Resources Association*, *53*(2), 280–299.

Garousi-Nejad, I., Tarboton, D. G., Aboutalebi, M., & Torres-Rua, A. F. (2019). Terrain analysis enhancements to the height above nearest drainage flood inundation mapping method. *Water Resources Research*, *55*(10), 7983–8009.

Hooker, H., Dance, S. L., Mason, D. C., Bevington, J., & Shelton, K. (2022). Spatial scale evaluation of forecast flood inundation maps. *Journal of Hydrology*, *612*, 128170.

Hu, A., & Demir, I. (2021). Real-time flood mapping on client-side web systems using hand model. *Hydrology*, *8*(2), 65.

Li, Z., & Demir, I. (2022). A comprehensive web-based system for flood inundation map generation and comparative analysis based on height above nearest drainage. *Science of The Total Environment*, *828*, 154420.

Li, Z., Duque, F. Q., Grout, T., Bates, B., & Demir, I. (2023). Comparative analysis of performance and mechanisms of flood inundation map generation using height above nearest drainage. *Environmental Modelling & Software*, *159*, 105565.

Li, Z., Mount, J., & Demir, I. (2020). Evaluation of model parameters of hand model for real-time flood inundation mapping: Iowa case study. *EarthArXiv. July*, *1*.

Li, Z., Mount, J., & Demir, I. (2022). Accounting for uncertainty in real-time flood inundation mapping using hand model: Iowa case study. *Natural Hazards*, *112*(1), 977–1004.

Maidment, D. R. (2017). Conceptual framework for the national flood interoperability experiment. *JAWRA Journal of the American Water Resources Association*, *53*(2), 245-257. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1111/1752-1688.12474` doi: 10.1111/1752-1688.12474

McGehee, R., Li, L., & Poston, E. (2016). The modified hand method. In D. R. Maidment, A. Rajib, P. Lin, & E. P. Clark (Eds.), *National water center innovators program summer institute report 2016* (Vol. 4).

Nobre, A. D., Cuartas, L. A., Momo, M. R., Severo, D. L., Pinheiro, A., & Nobre, C. A. (2016). Hand contour: a new proxy predictor of inundation extent. *Hydrological Processes*, *30*(2), 320–333.

Papaioannou, G., Loukas, A., Vasiliades, L., & Aronica, G. (2016). Flood inundation mapping sensitivity to riverine spatial resolution and modelling approach. *Natural Hazards*, *83*, 117–132.

Rahimzadeh, O., Bahremand, A., Noura, N., & Mukolwe, M. (2019). Evaluating flood extent mapping of two hydraulic models, 1d hec-ras and 2d lisflood-fp in comparison with aerial

imagery observations in gorgan flood plain, iran. *Natural resource modeling*, *32*(4), e12214.

Shastry, A., Egbert, R., Aristizabal, F., Luo, C., Yu, C.-W., & Praskievicz, S. (2019). Using steady-state backwater analysis to predict inundated area from national water model streamflow simulations. *JAWRA Journal of the American Water Resources Association*, *55*(4), 940–951.

Tansar, H., Babur, M., & Karnchanapaiboon, S. L. (2020). Flood inundation modeling and hazard assessment in lower ping river basin using mike flood. *Arabian Journal of Geosciences*, *13*, 1–16.

Wing, O. E., Bates, P. D., Sampson, C. C., Smith, A. M., Johnson, K. A., & Erickson, T. A. (2017). Validation of a 30 m resolution flood hazard model of the conterminous u nited s tates. *Water Resources Research*, *53*(9), 7968–7986.

Wood, M., Hostache, R., Neal, J., Wagener, T., Giustarini, L., Chini, M., . . . Bates, P. (2016). Calibration of channel depth and friction parameters in the lisflood-fp hydraulic model using medium-resolution sar data and identifiability techniques. *Hydrology and Earth System Sciences*, *20*(12), 4983–4997.

Zhang, J., Huang, Y.-F., Munasinghe, D., Fang, Z., Tsang, Y.-P., & Cohen, S. (2018). Comparative analysis of inundation mapping approaches for the 2016 flood in the brazos river, texas. *JAWRA Journal of the American Water Resources Association*, *54*(4), 820–833.

Zheng, X., Maidment, D. R., Tarboton, D. G., Liu, Y. Y., & Passalacqua, P. (2018). Geoflood: Large-scale flood inundation mapping based on high-resolution terrain analysis. *Water Resources Research*, *54*(12), 10–013.