Review of egusphere-2023-1969

First of all we want to thank the reviewers for their further recommendations on our work. We did our best to incorporate all the proposed suggestions. Below our detailed response to each comment.

egusphere-2023-1969-referee-report-3

Issues remain between the study's aims, methods, and conclusions. Specifically, the second aim of the paper is not supported by the methods. Representative bias in the gage network prevents adequately "mapping the spatial variability". However, Figure 10 and the associated discussion provide valuable insight. Rather, consider re-evaluating the information provided by the methods relative to the study's aims. For this reason and other minor edits, revisions are required prior to publication. Please see the attached document for more detailed and additional comments.

Issues remain between the study's aims, methods, and conclusions. This type of dataset is not appropriate to "map the spatial variability" (line 100 -101). As stated in previous comments, the gaging network does not provide a representative sample to map the spatial variability. The authors have added some discussion acknowledging some of these limitations. However, it does not remedy the disconnect between the aim and methodology. For instance, smaller streams are vastly under-represented; and some individual larger rivers contain larger portions of gages than those surrounding them. These rivers and streams comprise a population of geomorphic characteristics, that if biased in representation, impacts the results of spatial variability in geomorphic response.

Response: *We want to thank the reviewer for this comment. We agree with the limitations of the gaging sites' distribution as we highlighted in the manuscript. Please consider that the 3101 gages in this study provide a vast dataset covering varying landscapes and climate zones. For our model, we added further information for the clustering, pertaining to storm properties and landscape properties, as well as watershed characteristics (e.g. Stream density, Strahler stream order). At this stage, the Gage dataset is the best available dataset we can get our hands on target shifts in stage discharge, by accepting the limitations on its wide but 'constrained' variability.*

Considering the concerns regarding the disconnect between the aim and the methodology, we have revised the manuscript, also following the other reviewer's suggestions. As a consequence, we rephrased and added the following paragraphs.

Line 87-111: "*For this purpose, the availability of a large dataset representing a wide range of flood-inducing storm characteristics and channel morphology under different boundary conditions, such as underlying climatic, hydrologic, and geomorphologic settings, is crucial. This set of information forms a complex interacting system. The processes underlining these boundary conditions vary in spatial and temporal scale, and this calls for the use of improved analysis methods, able to draw predictions interlocking data of varying nature. In this context, machine learning (ML) is gaining popularity in the field of hydrology, geomorphology, and climate studies (Bergen et al., 2019; Schlef et al., 2019; Valentine and Kalnins, 2016), thanks to its ability to tackle coupled processes across space and time. Despite some limitations (Karpatne et al., 2019), and provided that the benchmark data used for the training are of high quality (Bergen et al. 2019), ML offers a valuable tool to gain new data-driven insights with high accuracy, transferability, and scalability (Houser et al., 2022; Sarker, 2021; Schlef et al., 2019; Sofia, 2020a).*

In the context of river morphology, specifically, in the last few years recent studies highlighted their capability to predict channel types (Guillon et al., 2020), providing a geomorphological characterization of *channels (Rabanaque et al., 2022), quantifying below-water (Woodget et al., 2019), or spatiotemporal changes (Boothroyd et al., 2021) in rivers, and guiding discharge estimation building from river morphology (Brinkerhoff et al., 2020), These works highlight how ML, when properly guided by field-based interpretation, can offer a valuable potential to push geomorphology into an increasingly predictive science (Fryirs and Brierley, 2022; Brierley et al., 2021). Tackling on the opportunities offered by ML potential, in this study, we sought to understand and predict the effects of flood-inducing storms on channel conveyance and, consequently, flood hazards. To achieve this, we have utilized stage-discharge "Residual" as a proxy of the channel capacity change, and we introduced an ML framework (section 2.3) that characterizes the interdependence of flood drivers, including atmospheric drivers (precipitation), hydrologic drivers (flow, stage), and geomorphologic drivers (channel width, depth, drainage area, geophysical characteristics). Overall, the analysis aims to: (1) highlight the variability of geomorphic response to flood-inducing storms across various climatic and geomorphologic regions in the Contiguous US (CONUS), and (2) understand the impact of these storms on the stage-discharge relationships at gaged sites as a proxy for changes in flood hazard. The study provided an independent test of discharge-based results and produced a tool for generating timely short-term updates of flood hazard estimates for dynamic rivers.*

While mapping the predicted residuals at streams is valuable and inferences can be made by comparing results across different regions, the authors should reconsider the information provided by the methods relative to the study's aims. The authors fail to mention spatial patterns in the conclusions, except perhaps, a slight alluding to at line 695.

Response: *We thank the reviewer for this comment. We have added the following to the conclusion-*

Line 696-699: *The gages used in the study although distributed across CONUS have intrinsic limitations in terms of stream size representations and spatial coverage of the river network. Therefore, careful considerations should be applied while considering the model for predicting the impact of flood-inducing storms on abrupt loss of channel capacity outside the basins used in our study. As for ML in general, successful translation is expected given that new sites under investigation possess similarities to the training sites.*

The introduction should be more specific about the type of response considered in this analysis: abrupt loss of channel capacity following flood inducing storm events. This will provide context and clarity.

Response: *Thank you for this comment. We have mentioned "abrupt loss of channel capacity" not only in the introduction but in the conclusions too to be consistent and specific about this.*

Minor Comments:

- Line 57 remove parenthesis around citation.

Response: *We have corrected this.*

- Line 131. Revise sentence by removing "and" and capitalizing first word in sentence.

Response: *We have corrected this.*

- Line 197: Greater than "the" 80th percentile. And 80th percentile of what? The annual flow duration curve?

Response: *We have corrected this. Please see below:*

Line 201-2-3: The reader should consider that while the median characteristic per se is not a 'severe' value, given the sample of data in Shen et al., 2017, it is a value representative of the typical event, for *storms which in general encompass events having peak flows greater than 80th percentile of the entire flow series.*

- Line 372: To me, the relative magnitude of the RMSE is unclear. The authors state that it was close to 0, which is good. But the values range from 0.09 – 0.14 m. What would be considered unacceptable? While zero is the target; I do not have a frame of reference to know the relative magnitude of this error. This is important, if I understand correctly, because it influences confidence in the predicted residuals. What confidence can we have in the predicted residuals?

Response: *We thank the reviewer for this comment, and we agree with this. We have revised the text. Please see below-*

Line 376-380: *Table 2 also represents the correlation distance and RMSE between the measured and predicted residuals for each cluster of the validation datasets. The average correlation was close to 1 for all N values, suggesting the performance of the SOM model was satisfactory. The average RMSE was in the range of 0.09 – 0.14 m, which indicates a low random error relative to the dynamic range (-3 to 3) of the predicted variable. Both the unsupervised correlation distances and the average correlation showed the best results for N- 365 days. The RMSE diminished with the increase in the interval.*

- Line 605: Remove the text "In the revised manuscript, we will incorporate these comments."

Response: *We have corrected this.*

- Line 609: I am unsure about translating the model to gaged sites. In my previous comment, I was referring to a limitation in data availability. Wouldn't the model require a stage-discharge curve to relate the residual to? I a stage-discharge curve would not be determined from gage extrapolation.

Response: *We thank the reviewer for this comment. We added the following lines to the discussion*

"The model predicts a shift in the discharge at the flood stage (residuals), as a proxy for flood hazard changes, implying that a certain discharge, expected to produce floodings, will be reached for lower stages than expected (residual shifting from positive to negative, at a specific gage). The approach starts from the concept that, typically, discharge time series are derived from water level measurements through an existing stage-discharge relationship. This is the general case for most gaging sites in the US, as well as other realities in other countries. As rating changes often happen during episodic storms, the proposed model can be adapted for other gage datasets, in different parts of the world, by assuming the operational existence of a similar approach". (Line 692-699)

- Line 619: Was the ML model trained using geomorphic properties as well?

We considered the watershed properties included in the GAGE II dataset, such as slope and aspect, drainage properties, geological information, and information on the physiographic region. After the variable importance, only some of the parameters were retained as reported in Figure 6.

egusphere-2023-1969-referee-report-2

The authors have addressed all of my previous comments. For future review responses, I strongly recommend that the authors clearly specify in the response document the locations (e.g., line numbers) in the manuscript where changes have been made, to streamline the review process.

Response: *We want to thank the reviewer for the time and effort. We apologize for any inconvenience caused by the way we produced our response. We appreciate the suggestions and will keep this in mind for future review processes.*

egusphere-2023-1969-referee-report-1

The revised version of the manuscript still needs additional work in the format of the introduction and discussion sections.

Major Comments

A separate discussion section is now included in the manuscript. However, the authors have unnecessarily complicated the section by adding subsections and being lengthy. I will try to summarize this section since most of the text has been repeated throughout the manuscript. For example, the text on Line 719-720 should be in the results sections since it describes what is being seeing in Figure 8. I assume this happen when the authors just move the exact same text from the results sections, instead of creating a new discussion section from they findings interpretation. Furthermore, the discussion section should not present any new figures like Figure 10, as this is done in the results section.

Similarly, I will make the limitation subsection its own section (instead of a subsection within the discussion) since it has enough length and has been drafted properly. Thus, I urge the authors to review and rewrite the discussion section to be efficient.

Response: *We want to thank the reviewer for this comment. As per previous comments stating the results section is too long, we have created a discussion section. Yes, we have taken some of the inferences from the previous results section so that the readers can relate the discussion to the figures. Considering the reviewer's comment, we rephrased this part, to be more consistent in the discussion. Please note that we prefer to keep Figures 10 and 11 in the discussion because we believe that these figures strengthen the discussions rather than be a result. As per the reviewer's suggestion, we have created a section for the "advantages and limitations" rather than keeping that as a subsection of the discussion.*

While the author improves the introduction section, there are still some portions that read choppy, mainly because the authors have small paragraphs (2-3 sentences) with their own topic but failed to connect them properly (L92-112). Thus, I ask the authors to revise the later part of the introduction.

We thank the reviewer for this comment. We revised the later part of the introduction as per the reviewer' suggestion.

Overall, most of my comments/questions from the revised version were answer by referencing Slater's 2015 workflow. Thus, meaning that my question could be answer in that publication. This represents that the authors need to include more information about Slater's work in this manuscript so the reader does not have to be back and forward between this manuscript and Slater.

Response: *We want to thank the reviewer for this comment. As the work by Slater et al is public and provides also access to the codes used to retrieve and calculate the residuals, we believe that the revised manuscript* *clarified sufficiently the methods used. We prefer not to add further information regarding Slater's work, to avoid lengthening even further the manuscript.*

Minor Comments

- L43: combine that small paragraph (two sentences) with the previous one at L36.

Response: *We combined the first two paragraphs in the introduction.*

- L776: The section index should be 4.2 and not 4.3.

Response: *We have corrected this.*

- L855-860: If you are focusing in the amount of stream gauges within the east and west region of CONUS do not starting comparing with Alaska and Hawaii since they are not part of CONUS. If you do, then you need to talk about Puerto Rico and American Samoa (both US territories).

Response: *Thank you for this comment. We have revised this in the manuscript. Please see below-*
Line 606- 609: Broadly speaking, the Eastern United States has better coverage compared to its Western counterpart. Particularly, the arid Southwestern United States shows notably lacking spatial coverage. Discrepancies in hydrology contribute to variations in the statistical uncertainty calculated across various parts of the country (Kiang et al., 2013).

- Figure 1: Please add a description of the arrow and red circle on the figure caption.

Response: *We have corrected this. Please see below:*
Line 69: Figure 1: Change in channel width in Boulder Creek, Colorado, before (2012) and after (2013-2015-2019) a flash flood in 2013 (Google Earth imagery). The Discharge reported here is the daily discharge measured at USGS 06730200 Boulder Creek at north 75th St. near Boulder, co. The red circles denote the section of the channel that has changed over the years and the blue arrow shows the missing channel from the year 2012 to 2019.