Reviewer #2

This manuscript tackles the floodplain mapping at the global scale. The authors use a methodology to estimate floodplains using a geomorphic approach (integrating heterogeneity), based on past studies such as Nardi et al. (2019). This methodology involves applying a geomorphic descriptor such as HAND (Height Above the Nearest Drainage) and globally optimizing its parameters to delimit floodplains, resulting in a global map with a resolution of 250m. In this study, the authors take a further step by optimizing the parameters of the same geomorphic descriptor (HAND) for more than 200 basins (delimitation at level 3 with respect to HydroBASINS). They consider heterogeneity in the production of a new map on a global scale, which is provided with resolutions of approximately 90m and 1km.

For the calibration and validation of the applied methodology, they relied on 500-year return period maps (JRC, GAR flood maps) and on the 250m Nardi resolution map (GFPlain250m), which presented a general precision greater than 0.85. Additionally, they facilitate access to the results through the following links: available at https://zenodo.org/records/10440609 and the main code at https://github.com/Mostaaaaa/SHIFT_floodplain.

The study is of interest and may be worthy to be published, but some effort should be made to better emphasize the impact of the study. In the following, you will find my comments.

Reply: Thanks for your comment and positive evaluation of our work. You'll find our point-to-point Reply below.

Major comments

- The scaling of hydraulic depth is investigated at the global level, obtaining a very scattered graph. Data seem to be better aligned for larger basins, but some additional effort should be spent to explain the variability observed in other river basins. Climate cannot be the only variable controlling the scaling exponent. Other factors such as rainfall, river morphology, or land use could also impact the result.
    - Reply: Thank you for raising this reasonable concern, and we fully agree with your interpretation. In this revision, we have expanded our discussion on our analyses of possible relevant factors affecting the empirical parameters.

We've approached this expansion in the following ways:

- 1) Hypothesis Clarification: We have better articulated our hypothesis. We consider aridity to be the primary factor influencing the spatial variability of b due to the assumption that in a humid basin, rivers with larger upstream drainage areas (UPA) would have greater dominance over smaller river segments in shaping riverine floodplains. Vegetation should also be related since it is involved in the runoff generation process as well as modulating soil erosion that can be key to floodplain formation. Additionally, other factors such as terrain and soil composition in riverine areas might also influence the results, although the underlying mechanisms are not as intuitive. **It is important to clarify that in no way do we expect perfect relationships between these factors and our derived exponent b, because of the data uncertainties as well as the complex physical interpretations with b.** Based on this hypothesis, we have thoroughly tested the correlation with other possible factors, including Leaf Area Index (LAI), mean elevation, elevation standard deviation, and three types of soil components (clay, silt, and sand). Results indicate that only the Aridity Index exhibits significant correlations with our estimated parameters, while LAI shows significant correlations in the largest basins, and other factors do not present a uniform result. We have included the Aridity Index along with LAI in the revised manuscript as they are results of our primary hypothesis. Other results are provided in the Supplementary Materials. Please refer to the revised Figure 4 and Supplementary Table 1 for these results.

- 2) Scale Testing: We further tested our hypothesis on different scales. We estimated parameters on Level-4 and Level-5 basins to observe if the correlation changes or if any new patterns emerge. Results show that both the Aridity Index and LAI exhibit significant correlations, and terrain factors also show positive but weak correlations, suggesting that these factors might be influential, though the underlying mechanisms could be more complex. See Supplementary Table 1 for the results.

- 3) Discussion of Variability: We have discussed why other basins do not exhibit as strong correlations as the largest basins in the revised Section 4.4. The largest basins are hydrologically connected and thus are expected to have more internally consistent hydrological characteristics from upstream to downstream. In contrast, many other basins are aggregates of smaller basins, so the relationship between the basin-average estimate of parameter 'b' and the

Aridity Index might be affected by this aggregation process.

Overall, the comments helped us to perform some meaningful analyses. The insights were included in our revised manuscript and this response letter.

- In Section 4.1, The authors discuss the uncertainty associated to the parameter b. In this section, results are not clear or do not display a clear pattern. It is also surprising that the results obtained over the larger basins still have a large uncertainty even if the regression function works better.
    - Reply: Thank you for your comment. We have conducted a deeper investigation into the uncertainty related to parameter 'b', which has led us to restructure and clarify our discussion on this topic. Additionally, we have refined some technical details in the parameter 'b' estimation, resulting in generally smaller uncertainties. Our revisions are as follows:
    - Conceptually, we've better defined our metric and clarified its interpretation. The metric is now defined as residual uncertainty, which calculates the remaining uncertainty after our data filtering scheme. Physically, this metric assesses how well the data conforms to the power law: a better-conforming set of data result in a narrower range of the estimated *b* sequence and, consequently, lower standard deviation. By calculating the metric, we aim to see 1) how well the filtered data in these basins align with power law, and 2) the robustness of our parameters, as a lower standard deviation supports the application uniform filtering percentiles globally (see 2.2 Methods). Our goal in refining this section is to provide a clearer explanation of the uncertainty we are addressing and to give a quantitative assessment of our approach to managing it.
    - Technically, we have modified the technical details of parameter 'b' estimation, specifically the binning parameter in the estimated 'b' sequence generation. The binning parameter determines the number of bins when grouping 'b' values. It should be set higher for samples with high data noise to better filter out outliers. However, setting the binning parameter too high reduces the amount of data available in each bin, and could interfere with the results with few reference grids available. We have improved this process by adding a constraining mechanism to maintain a baseline level of data and clean out empty bins. We tested how the number of bins influences our estimated results. Sensitivity tests show that when the binning parameter exceeds a certain value (150), the estimated 'b' sequence becomes more statistically stable (median,

quantiles, and standard deviation). This parameter, which we previously thought was insignificant, has shown that it does not interfere with the overall 'b' values but narrows down the quantiles for large basins with numerous reference points (and subsequently more noise points, e.g., the Mississippi and the Amazon), leading to stabler estimates for these basins. As you pointed out, some large basins (e.g., the Amazon and the Mississippi) showed greater uncertainties in the previous version of our manuscript. It was possibly due to our prior setting of the binning parameter. Therefore, we've updated the process with the newly set binning parameter and the constraining mechanism, which helps us to rule out another possible source of data noise. Global residual uncertainty now shows a clearer pattern (see revised Fig. 9).

  - Based on these improvements, we have revised the original Section 4.1 on uncertainty. Please refer to Section 4.2 for the updated content, and Section 2.2 for the improvements above on parameter 'b' estimation.

- Results should be better described. For instance, it would be valuable to have floodplain patterns obtained from SHIFT with the river network layer and one image showing the differences between SHIFT and a reference map. Additionally, it would be good to enlarge the images in Figure 4.
  - Reply: Thank you for your advice on better presenting our results. Regarding figures, we've changed our figures on regional differences to better represent regional differences (see our changes in the revised Figure 5). We've also refined the interpretation of our results throughout the Results Section and elaborated on necessary details to better support our claims of optimizing local parameters, especially for Section 3.3.

Minor edits:

1. Line 120-123: 'overestimated floodplains in arid or semi-arid area as reported by existing assessments of geomorphic floodplains' (Dhote et al., 2023; Lindersson et al., 2021). In these references, only Lindersson et al. refers to arid areas and their difficulties. While Dhote et al. only highlights the overestimation and underestimation of the descriptors HAND and TWI respectively, but does not talk about the relationship with arid areas.
   a. Reply: Thanks for identifying our negligence and we've removed Dhote et al. in our revised manuscript.

2. Line 314-315: ‘This iterative process stops either when every data point fits within all moving windows, or if the procedure fails to converge towards a stable solution’. It could explain what is meant by a stable solution, for a better understanding.

    a. Reply: "Fails to converge towards a stable solution" refers to the situation where, when dealing with highly noisy or unevenly distributed data, the iterative process fails to reach a stable state within a finite number of steps, resulting in extensive data filtering. The ideal iterative denoising process should filter out fewer and fewer points each time, eventually keeping all points within the 3-sigma range of the sliding window. This assumes that the data is primarily composed of a majority of valid data fitting an assumed overall distribution, combined with a small amount of noise. Under such ideal circumstance, the final sliding window's mean and STD should be an unbiased representation of the population. However, if the valid data is scarce and noise is abundant, it may lead to high natural variability of the data, or the data distribution may be significantly skewed or non-normal, preventing the sliding window's mean and STD from representing the main population, thus leading to a fail of convergence. In this study, non-convergence occurred in a few watersheds with limited reference data points, to address which we established this termination condition. Consequently, we believe that the parameters fitted in these highly noisy watersheds may come with uncertainty. This is further discussed in the revised Section 4.1. We've added a short explanation in the corresponding paragraph to explain it further (see our revision at xxx), but since it's not a major concern we didn't expand in the method section.

3. Line 343: change ‘as the as the’ for ‘as the’.

    a. Reply: Thanks for pointing out. We’ve removed the typo.

4. Line 385: only the range of values obtained for the coefficient 'a', For what reason is not presented a graph as in fig.3 of parameter 'b'? If it is possible to provide the values of both parameters, so that this method can be studied at smaller scales focusing future studies in a single basin or a single region and its sub-basins. It would be ideal to base the importance given to the parameter 'b' on the parameter 'a'.

    a. Reply: Thank you for your constructive comment. We do consider parameter 'b' to be more important in our study. In the broader sense of hydraulic geometry, while the value of 'a' is also a determining factor in the floodplain delineation process, the physical interpretation and research focus have historically been more concentrated on the parameter 'b' (since Leopold, 1953). In most cases, the understanding of parameter 'a' is dependent on 'b', and 'b' could be more clearly interpreted as the sensitivity to scale. Thus, in cases where the actual mechanism of the hydraulic geometry relation is not clear, it would be more common to dig into the possible influencing factors of parameter 'b'. In the context of FHG, previous researchers have also primarily focused on 'b'. For example, in a study by Annis et al. (2019), when evaluating the performance of varying FHG parameters across different stream orders, the emphasis was mainly on 'b'.

b. In our study, the two parameters have different impacts on delineating floodplains. Parameter 'b' determines whether the river with larger upstream drainage area dominates the floodplain. The difference in impact between large and small rivers is greater when 'b' takes a larger value. This is related to our assumption in our paper that in humid areas, the difference between small and large rivers would be more significant, leading to a stronger dominance of those with larger upstream drainage areas. In comparison, 'a' lacks a unified unit and a clear physical interpretation, and it is highly dependent on 'b'. To provide better understanding, we have provided the spatial distribution of parameter 'a' (see supplementary Figure 1). Unlike the clearer pattern of 'b' of generally better aligning with the Aridity Index, values of 'a' vary largely. For instance, small 'a' values appear in some of the largest river basins (e.g., the Amazon and Yangtze River Basin), possibly balancing out the influence brought by larger 'b' values. While there are possible physical interpretations, it is challenging to interpret 'a' when the underlying mechanism related to 'b' is not clear. Our discussion on the interpretation of parameters was further added to Section 4.1 in Discussion.

c. Thank you for suggesting that we provide parameters for all basins. We have uploaded all the parameters, along with our confidence levels of all the parameters, in a shapefile our Zenodo repository. We will provide a link in our revised manuscript.

1. 3b: include a legend.
    a. Reply: Thanks for pointing out. The original Figure 3b was revised to Figure 4 now to expand our discussion on relevant factors. In each of the sub-graphs we’ve added a legend correspondingly.
2. 3a: change 'estimatio' for 'estimation' in the description.
    a. Reply: Thanks for pointing out. We’ve removed the typo.

References:

1. Leopold, L. B. and Maddock, T.: The Hydraulic Geometry of Stream Channels and Some Physiographic Implications, U.S. Government Printing Office, 68 pp., 1953.

2. Annis, A., Nardi, F., Morrison, R. R., and Castelli, F.: Investigating hydrogeomorphic floodplain mapping performance with varying DTM resolution and stream, Hydrol. Sci. J., 64, 515–538, 2019.