Responses to comments: Anonymous Referee 1

This study evaluates 11 physical variables for index flood estimation across catchments in the Philippines, aiming to enhance flood estimation for ungauged catchments. The authors show significant effort in data collection and selection, and they present extensive analyses in this manuscript.

Response: Comments noted and appreciated.

Notice that the authors claim their study is applicable to ungauged catchments. However, if my understanding is correct, the analyses presented do not show anything regarding this applicability. While they propose using more local information to improve flood estimation—a common approach in many studies—they suggest this could benefit ungauged sites. Although this suggestion might be correct, it is overstated in the title since there are no relevant analyses or validation to support this claim.

Response: We note and largely agree with these comments – the aim of the paper is to provide a methodology that **can** be applied to ungauged catchments rather than demonstrating its effectiveness in such situations. The title does overstate this applicability, and we propose to revise this to 'Integrating historical archives and geospatial data to revise flood estimation equations for Philippine rivers' or something similar. In addition, we will ensure that the potential applicability to ungauged catchments is stressed rather than illustrating this – we could add a further section to the paper to demonstrate how the method can work in ungauged catchments, but that would extend an already lengthy paper and would require considerable statistical work to provide robust evidence.

In addition, the manuscript has several critical issues regarding the quality: 1) Unclear Critical Information - There is confusion regarding the number and details of study sites, as well as incomplete or unclear descriptions of methodologies.

Response: See replies below to specific points raised

2) Lack of Novelty and Significant Findings - The framework lacks innovation and the findings are not particularly groundbreaking (as noted by the authors in line 527).

Response: This comment is to some extent true, but we contend that the novelty in the paper comes from: (i) combining data from multiple sources to extend the database in what is a relatively data-poor country – this is the first analysis adding data from the early 20th century (SWS in Table 1) that covers sites for which no later data are available; (ii) utilising modern databases, for topography, rainfall and land-cover, to significantly extend previous analyses of flood magnitude in the Philippines; and, (iii) assessing the results against those form a largely independent hydrological modelling exercise. The consistency of our results with previous analyses reveals that the noise observed by previous authors (Meigh, 1995; Meigh et al., 1997) is not due to excluding precipitation and land-use variables from the analysis, but rather is a consequence of catchment and/or climate characteristics that remain unknown and worthy of further study. We contend that this finding is itself novel and represents progress in our understanding of flood generation in the Philippines.

3) Quality and Clarity - The structure of the manuscript, along with its figures and tables (including captions), lacks quality and clarity. There are numerous mistakes throughout the document. I found it challenging to understand the authors' main points, both from the text and the figures. While the authors' efforts in conducting numerous analyses are commendable, they are strongly encouraged to improve the manuscript by enhancing its accuracy, clarity, and focus.

Response: Comments below on specific issues raised, and the general points regarding clarity will be considered when revising the manuscript.

Some specific comments (but not all) for improvement are listed below for reference:

How many catchments are analyzed exactly? Is it 513 or 466? The abstract states 513, but other parts of the manuscript (e.g., Figure 1, Table 2, line 172) suggest it is 466. Lines 164-172 are particularly confusing: how does 513 minus 205 result in 466 sites?

Response: Noted – the steps by which the full data set was censored are explained in the text but this can be made more explicit and we note inconsistency between Tables 1 and 2 that can be corrected. The '205' on line 167 is an error. The table below shows the process that was followed, with each row representing a step in the analysis that resulted in the exclusioin of some sites:

	Number excluded	**Number remaining**
Total number of sites	-	842
Exclude sites with poor rating curves or indeterminate location	169	673
Exclude sites with short records	160	513
Exclude sites with invalid L-moments	2	511
Exclude sites with poor curve fit	45	466

The catchment area sizes are analyzed in this study, but this information is missing in the data section.

Response: The method for calculating catchment area is noted in the supplementary data file and will be added to the manuscript.

Why do the catchment area groupings differ between Table 2 and Figure 3? The former uses four groups (100-200, 200-400, 400-800, <800), whereas the latter uses five groups (<25, 25-50, 50-250, 250-2500, >2500), with so different ranges.

Response: We agree that this inconsistency is not helpful – Table 2 will be edited to match the groups used in the figures. Table 2 does have 5 groups, as there is a <100km^2 group also.

Is the area grouping range rational? Comparing catchments across such varied groupings (<25, 25-50, 50-250, 250-2500, >2500) seems to be strange for representing hydrological responses.

Response: Given the shape of the frequency distribution of catchment areas, a non-linear grouping is appropriate. We tested a strictly logarithmic division of catchment areas and found no appreciable differences to the results when using the proposed grouping. The total number of catchments within each group is as follows (total = 466):

Catchment area km^2	**<25**	**25-50**	**50-250**	**250-2500**	**>2500**
Number of sites	57	34	190	165	20

The presentation of results in catchment area groups allows readers to assess visually if there are catchment area related effects in the data. It is intended only as a visual device and so the choice of group boundaries is not in itself significant. A much more comprehensive assessment of the importance of catchment area is given by Figures 5 and 6, and Tables 5 and 6.

The title of Section 5.1 is not coherent with its content.

Noted and this will be changed.

Moreover, it is difficult to discern the patterns the authors aim to show (lines 232-234) because Figure 3 is unclear. Improving the color and marker settings or changing the plot type (e.g., stacked bar plots) might be helpful.

Response: We will experiment with smaller markers for clarity. The plot style is the standard format for these results.

Line 233: Figure 3 refers to area, not climate.

Response: Noted and will be corrected.

Line 233: All the others should point to Figure S5, not Figure S6.

Response: Noted and will be corrected.

There are references to regions 1-13 in many analyses, but no introduction or proper definition of these regions is provided. The first mention appears on line 78, without any definition.

Response: We will add a definition of the regions of the Philippines and an appropriate reference to section 3.

Line 213 claims that the authors apply the FEH approach described earlier, but there is no earlier description. The term is introduced on line 50, but without sufficient detail.

Response: This is a fair comment – we have referenced several FEH studies but do not provide a brief summary of the FEH approach. The best place to do this is in section 4.2, where our methods are related to the FEH approach.

Improper structure: Lines 220-227 fit better in the data section rather than the methods section.

Response: We disagree with this suggestion as the description of variables in 220-7 needs to follow Table 3. The title of section 4.2 is "Predicting high magnitude floods from catchment properties" which is entirely appropriate for defining data sources for these variables.

Line 245: Why are only 71 sites analyzed here instead of the full 466 sites?

Response: As it says in line 248 "…for 71 sites with at least 20 annual maxima and for which the GLO distribution provided the best fit to the data." We could perhaps try to clarify this statement with some minor re-wording for emphasis "…for the 71 sites that had at least 20…"

The figure captions should be more descriptive of the settings, but discussions of the results/patterns should not be included here.

Response: It is not clear what 'more descriptive of the settings' means. The figure captions include very little discussion of the results presented within them, and where this is presented we think that it is valuable to guide the reader's interpretation of the figure. Readers may look at figures in isolation and the captions have been prepared to allow figures to be understood without having to read all of the related text.

Tables could be improved by adding delineation lines, clarifying headers (e.g., the last item in the first column of Table 2), and removing unnecessary information.

Response: The journal format has been followed for the Tables, although the reviewer's comment about delineation lines does make sense. We have tried to keep tables concise and are unsure what 'unnecessary information' is being referred to here.

In summary, I acknowledge that such a study is needed for the selected country, as the authors claim there are no other similar studies to such an extent. While the analyses are comprehensive, the manuscript lacks sufficient clarity in its structure and critical information, which hampers its transferability and overall readability.

Response: Some of the specific comments made by both reviewers will definitely assist in this regard.