20 December 2024

Dr Theresa Blume and **Professor Alberto Guadagnini**
Chief Executive Editors
Hydrology and Earth Systems Sciences

Ref: hess-2024-188

Dear Editors,

Following receipt of review comments and your invitation to revise our manuscript, I am pleased on behalf of our research team to submit a revised version of our paper. The title of the paper has been changed, following reviewers' comments, to *Integrating historical archives and geospatial data to revise flood estimation equations for Philippine rivers*.

The marked-up version of the paper shows where changes have been made in response to the comments received. The following table shows our responses to the specific comments made by the two reviewers.

Number	Comment	Response and revisions
Reviewer 1		
R1.1	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.	Comments noted and appreciated.
R1.2	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.	We note and largely agree with these comments – the aim of the paper is to undertake data analysis to generate equations that **can** be applied to ungauged catchments rather than demonstrating its effectiveness in such situations. The title, and abstract, did overstate this applicability and we have changed the title to 'Integrating historical archives and geospatial data to revise flood estimation equations for Philippine rivers' as well as editing the abstract. In addition, we have stressed the **potential** applicability of our results to ungauged catchments. We considered adding a further section to the paper to demonstrate how the method can work in ungauged catchments, but decided not to do this as it would extend an already lengthy paper and would have required considerable statistical work to provide robust evidence.

R1.3	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.	See replies below to specific points raised. We have made some clarifications and have corrected some erroneous values in the text.
R1.4	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).	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 and region – 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. Some editing of the abstract, introduction and discussion has been undertaken to try to clarify these points.
R1.5	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.	Comments below on specific issues raised, and the general points regarding clarity have been considered in making revisions to the manuscript.
R1.6	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	Noted – the steps by which the full data set was censored are explained in the text but has been be made more explicit and we note inconsistency between Tables 1 and 2 that has now been corrected. The '205' on line 167 in the original manuscript was an error. The table below shows

<table>
<tr><td></td><td>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?</td><td>the process that was followed, with each row representing a step in the analysis that resulted in the exclusion of some sites:
<table>
<tr><th></th><th>Number excluded</th><th>Number remaining</th></tr>
<tr><td>Total number of sites</td><td>-</td><td>842</td></tr>
<tr><td>Exclude sites with poor rating curves or indeterminate location</td><td>169</td><td>673</td></tr>
<tr><td>Exclude sites with short records (Table 1)</td><td>160</td><td>513</td></tr>
<tr><td>Exclude sites with invalid L-moments</td><td>2</td><td>511</td></tr>
<tr><td>Exclude sites with poor curve fit</td><td>45</td><td>466</td></tr>
</table></td></tr>
<tr><td>R1.7</td><td>The catchment area sizes are analyzed in this study, but this information is missing in the data section.</td><td>The method for calculating catchment area is noted in the supplementary data file and has been added to the manuscript.</td></tr>
<tr><td>R1.8</td><td>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.</td><td>We agree that this inconsistency is not helpful – Table 2 has been edited to match the groups used in the figures. Note that Table 2 does have 5 groups, as there is a <100km^2 group also.</td></tr>
<tr><td>R1.9</td><td>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.</td><td>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):
<table>
<tr><th>Catchment area km^2</th><th><25</th><th>25-50</th><th>50-250</th><th>250-2500</th><th>>2500</th></tr>
<tr><td>Number of sites</td><td>57</td><td>34</td><td>190</td><td>165</td><td>20</td></tr>
</table>
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</td></tr>
</table>

		significant. A much more comprehensive assessment of the importance of catchment area is given by Figures 5 and 6, and Tables 5 and 6.
R1.10	The title of Section 5.1 is not coherent with its content.	Noted and has been changed.
R1.11	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.	We have used smaller markers. The plot style is the standard format for these results and so has been retained.
R1.12	Line 233: Figure 3 refers to area, not climate. Line 233: All the others should point to Figure S5, not Figure S6.	Noted and has been corrected.
R1.13	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.	We have added a definition of the regions of the Philippines (new Fig. S1) and an explanation of these regions has been added to section 3.
R1.14	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.	This is a fair comment. The FEH approach is now referenced at the end of section 2, and we have removed the reference to 'described earlier' – it is better to retain the outline of the method here in 4.2 than to explain it earlier and have to repeat key parts here.
R1.15	Improper structure: Lines 220-227 fit better in the data section rather than the methods section.	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.
R1.16	Line 245: Why are only 71 sites analyzed here instead of the full 466 sites?	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 have tried to clarify this statement "…for the 71 sites that had at least 20…"
R1.17	The figure captions should be more descriptive of the settings, but discussions of the results/patterns should not be included here.	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.
R1.18	Tables could be improved by adding delineation lines, clarifying headers (e.g., the last item in the first column of Table 2), and	The journal format has been followed for the Tables, although the reviewer's comment about delineation lines does make sense -some delineation lines have been added where these

	removing unnecessary information.	are most helpful, and we will rely on the editors' judgement as to whether these are allowable deviations from the journal's formatting guidelines. We have tried to keep tables concise and are unsure what 'unnecessary information' is being referred to here.
R1.19	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.	Some of the specific comments made by both reviewers have assisted us in clarifying the paper and we hope that we have met the reviewer's requirements.
Reviewer 2		
R2.1	The paper by Hoey et al., entitled 'Flood estimation for ungauged catchments in the Philippines' aims at delivering design equations to estimate flood magnitudes and frequency in ungauged catchments across the Philippines. The paper addresses an interesting topic, the design of equations to predict floods in data scarce region. The authors do enormous work to analyse the data. However, the methodology is not well defined/structured. Overall, the paper lacks clarity and needs to be restructured for better clarity.	See previous comments and responses.
R2.2	The title of the paper does not reflect the content of the paper appropriately. The authors propose and evaluate the index flood approach and multi-variate regression to estimate flood. The approach is evaluated only at catchments with data used to fit the regression equations. In a context of flood estimation in ungauged catchments, the paper is lacking the following steps: 1) a clear definition of the methodology used for regionalization to ungauged catchment, 2) a cross-validation to evaluate the performance of the designed equations in ungauged catchments. It is necessary to	These are all very good suggestions, based on the title of the paper. As noted in our response to Referee 1, this title is not accurate and we have amended the title to fit the contents of the paper. The reviewer's suggestions are exactly what we would have done if the paper retained its original title, and there is merit in presenting such an approach and flow chart to guide readers in the use of the equations. However, that would lengthen the current paper excessively and we consider that this is a separate piece of work.

	evaluate how well the proposed approach would perform in ungauged catchments. Therefore, I suggest the authors consider using some cross-validation approach (i.e., leave-one-out or k-cross validation) where a set of catchments are used to fit the design equations; then the remaining catchment(s) are used as pseudo-ungauged catchment(s) to evaluate the accuracy of the designed equations in ungauged catchments. In addition, the methodology used is not clearly described (see specific comments below).	
R2.3	The paper could benefit from a flow diagram that describes the steps followed to estimate floods at ungauged catchments.	We have added a new table (Table 8) and accompanying description to illustrate application of the method to two sites. We decided not to include a flow diagram as the only requirements for these calculations are measurements of catchment area and rainfall.
R2.4	**Specific comments:** The authors fit and compare the accuracy of the three distributions showing that the choice of the distribution influences flow estimates and no distribution performs well at all sites. Then, the GLO distribution is selected to predict high flow magnitude in section 4.2 with no justification of this choice.	There is an error in line 215, in that the Q_T values used were from the **best-fit curve** at each site, and not only from the GLO results. This has been corrected.
R2.5	In addition, Equation 5, the factorial standard error for the GLO distribution, is only applicable when number of records is at least 20 years. The study region has only 71 sites (line 245), with more than 20 years data, among the 466 sites retained for the analysis (Line 234) . Why choose the GLO distribution if it cannot be applicable to all sites? Why choose only GLO instead of using the best fitted distribution at each site as shown in table 2?	See previous comment (2.4) – clarifying the curves used will address all of these points.
R2.6	•The last paragraph in section 4.1, the authors lists three ways the construct growth curves by combining the curves fitted from each site by region, climate type or catchment areas. Therefore, I	The analysis reported in 4.2 is for all catchments individually, and does not used the regionalised growth curves. Section 4.2 does not follow from the final paragraph in section 4.1 as the reviewer assumed.

	am assuming that the resulting curves are regionalized curves where catchments from within each group(region or climate type, catchment area bin) would be represented by a single curve. Correct In other words, catchments of each group would have a single QT value estimate from the regionalized growth curve. How the author use these results to perform a correlation analysis with catchment properties in Section 4.2 (Line 212: values of QT provided by the GLO analysis previously, were correlated with catchment properties")?	We have revised the wording at the end of 4.1 to make it clear why we are producing combined growth curves and how these can be used, and also edited the start of section 4.2 to clarify that this is concerned with individual catchments.
R2.7	The authors use precipitation dataset covering only 17 years (from 1998-2015) which is not sufficient to derive climatology of the catchments. In addition, this dataset are not available for a period that overlaps with flows records period. As suggestion, the author could use precipitation datasets from reanalysis products (e.g. ERA5) which are available from 1940 to present around the globe. It would cover the same period as flow records for most catchments and provide sufficient long timeseries to derive climatology for precipitation variables used in Table 3.	This is an interesting suggestion. We are aware that none of the available precipitation datasets is perfect for our analysis, as we are combining river flow data from different time periods. We used APHRODITE V2 which, as noted, covers the period 1988-2015. Compared with the other precipitation reanalysis products that are available, APHRODITE has the advantage for our study of being focused on Asia and so is likely to be more reliable in that region. The V2 product is regarded as an improvement over APHRODITE V1 in its representation of extreme events which is another benefit for using it in our study. For both of these reasons, we are satisfied that this data set is appropriate, but note that a comparison between different precipitation estimates would be an interesting area for future work. We note that robust validation of APHRODITE is only just starting (https://doi.org/10.1016/j.jhydrol.2024.132562 pre-print published December 2024).
R2.8	Why the fitted curves from individual sites are not used to fit a relationship between flood magnitude and catchment area?	Table 5 presents the results of these fits and compares them with equations that also include a precipitation term.
R2.9	It is unclear how the individual short historical records are combined to generate large dataset that produce consistent results.	This is covered in section 6.3, and we have tried to make this clearer. In summary, the absence of systematic differences between results from the different data sets (eg Figure 5a-c) provides confidence that we can treat the data from all data sources as a single data set for the analyses presented in the paper.
	Results presented in Section 5.3.3 (Eq6a-6c) would be better presented in a table with equation and associated R^2 and RMSE in the same way as in Table 5.	For consistency within the paper, we have converted equation (6) to table form (Table 6).

R2.10	Overall, the way the methodology and the results are presented is incoherent and lacks clarity.	Other than the specific comments above, it is difficult to know what the reviewer is referring to here. Both reviewers have made specific comments that will improve the clarity of the results. The mis-leading title of the paper may have caused some confusion, and clarifying this will hopefully also aid the clarity of the paper.

The reviewers' comments have been very valuable in guiding improvements to the manuscript, and we consider that this version is significantly improved as a result of their comments.

Once again, we affirm that this manuscript is original, has not been published previously, and is not under consideration for publication elsewhere. We have no conflicts of interest to disclose. All authors have read and approved the manuscript, and agree to its re-submission

Yours sincerely,

Pamela Louise M. Tolentino

On behalf of the first author, Trevor Hoey and the Catchment Project research team