

The authors present a study that implements an additional environmental variable into a commonly used approach for RFFA through a series of statistical methods. The objective is to incorporate the spatial heterogeneity of basins into the approach, which is interesting in its nature. The results demonstrate improvement, and various variables are utilized in statistical methods to produce more rational results. I have two major comments regarding the description that I believe the authors could address. Furthermore, I have listed several minor comments in order to maximize the impact of this study.

Major comments:

1. The methods section of the current manuscript is somewhat scattered, with several critical details not being explicitly stated and only being supplemented in the results section. As a result, the reader may find it challenging to understand the entire procedures used in this study and would require a considerable amount of effort to piece together all the information, while some crucial details may still be missing. I have listed some examples below, but there may be more. Addressing these concerns would aid in maximizing the reproducibility of this method.
  - L210: It is not clear how the authors define "the theoretical association." I believe it refers to the Pearson correlation, which is mentioned much later in the paper.
  - Section 3.4 consists of only one sentence, which indicates that several details are missing.
  - L218-220: It is unclear how the spatial patterns of 45 sub-basins were incorporated into the analysis of 207 gauges. Additionally, it is not clear whether "sub-watersheds" and "sub-basin units" refer to the same concept or different concepts.
  - L221-222: This information should have been included in Section 3.1 because it pertains to the fundamental information regarding the hydrological data used in the study.
  - L224-226: This information should also have been included in Section 3.1 because it pertains to a strategy used in the study for applying the model selection criteria. The first part of this information is already included in Section 3.1, while the second part is in the results section, leading to confusion. Additionally, the authors should provide relevant references.
  - L235: This sentence is unclear and difficult to comprehend.
  - L258: It is unclear how the authors define "strongly significant."
  - Section 4.3.1 is a method, rather than results.
2. The authors aimed to improve the RFFA method by incorporating variables that consider the spatial heterogeneity of sub-basins and investigated the influence of 28 variables on the RFFA. While the final results indicated an overall 5% improvement in adjusted R<sup>2</sup>, with the value increasing from approximately 0.85 to 0.9, the authors employed several

statistical methods to eliminate 27 out of the 28 variables. The backward elimination process was used to identify the most critical variables and remove those that did not significantly contribute to the model. However, the authors' justification for discarding the 27 variables is not entirely convincing, especially considering that 27 is a significant portion of the total 28 variables investigated (e.g., the relatively confusing statement in lines 347-370). To reinforce the validity of this analysis, a clear and systematic procedure should be presented (as discarding such a large portion of variables without sufficient justification may raise questions about the validity and robustness of the proposed method), which currently appears to be in a relatively random order. The inclusion of tables or hierarchical diagrams could potentially aid in clarifying the methodology. Additionally, the authors should explain how they selected the different but parallel statistical methods, such as correlation analysis and PCA, and how the elimination process was objectively determined. By reorganizing the description in a more structured and coherent way, the authors could strengthen the justification for the proposed model.

Minor comments:

1. It is not clear by which method the authors used for comparing the 3(or 4 in some cases) model selection criteria (i.e., AIC, BIC, ADC, AICc) for determining the best fit distributions.
2. Concerning the Pearson correlation: (1) why use a standard of 0.6 (instead of a common one, e.g., 0.5)? any references? (2) Does the normality been checked? If not, would using the Spearman correlation be a better option
3. L327-331 stated that WS\_Area and Stream\_Length are removed whereas WS\_Perimeter and Basin\_Compactness are retain due to the intension to focus on basin shape rather than basin size for concentrating on the efficiency of the fluvial system transport. However, it should be aware that the length-area relationship has been found as a function of basin shape (Sassolas-Serrayet et al., 2018). Additionally, several studies have shown that basin size plays a critical role in catchment hydrological response and determines the curve of the flood frequency analysis (Merz & Blöschl, 2009). While it is recognized that the drainage size appears in the final model based on its origin as a one-variable model, the description provided by the authors may be misleading.
4. It is unclear why drainage size still appears in the 28 variables if it has already been recognized as a prominent variable in previous RFFA.
5. The link between the interpretation of the PCA results and Figure 6 needs improvement, and it is unclear what procedure/standard the authors used to objectively eliminate variables based on the results of the Pearson correlations and PCA
6. "the observed correlation" in L326 should be stated in a quantitative way.
7. It is stated in L227-233 and L428-436 that the results of the best-fit distributions show an extreme low percentage for GEV distributions and the authors refer (or hint?) this to the

number of parameters. Although it is not incorrect, it should be noted that the two two-parameter distributions used in this study (EV1 and LN) have lighter-tailed behavior than the two three-parameter distributions (GEV and LP3) (El Adlouni et al., 2008). Based on their results, it implies that the estimates of the extreme flood quantiles are relatively small in this region. Such a result may be due to the shorter dataset used in this study, which should be roughly between 37-48 years long based on the informative mean and standard deviation of the record length provided in this paper. As a result, there may be considerable uncertainty in estimating extreme quantiles for GEV (e.g., Papalexiou and Koutsoyiannis, 2013), which should be properly referenced in the discussion.

8. The title suggests a potential linkage or interaction between these conditions and flood frequency analysis, so it is expected that more discussion of this topic would be included (which is still weak, e.g., 450-452, and supporting references are expected). Otherwise, it may be advisable to reconsider the title to avoid any potential for confusion.
9. In L463-464, the authors imply that they intend to implement another modified model to properly capture spatial variability in a glacial region, but the resulting model obtained in this study eventually excluded all these variables. Please comment on this, and it is not clear whether the scope of this study is "glacial conditions" or "land use" or both from the current title.
10. Please provide more information on the dataset used in Sec. 3.3, such as whether the same basin boundary was used for the analyses of the one-variable model and the two-variable model, the duration of streamflow data, and whether the analysis is representative if the duration of each variable does not overlap (as shown, the precipitation data is from 1981-2000, whereas the land use data is based on the dataset from 2014-2017).
11. L160-161: Since the authors explain the reason for using annual maxima instead of partial duration series (PDS), it would be appropriate to acknowledge other commonly used metrics, e.g., the peak-over-threshold (POT) analysis.
12. There is no clear discussion in the previous section concerning the statement in L537-538, which appears as the closing sentence of the conclusion.
13. It could be beneficial for the authors to consider providing some additional discussion on how the findings of this study may be transferable to other regions. Do the authors suggest directly including naturalized variables in RFFA for other regions, or do they recommend conducting an entire eliminating process based on different environmental variables obtained in the study area?
14. Throughout:
  - There appear to be a relatively large number of minor mistakes, such as unclear meanings of italics and special font sizes (e.g., L83, L86, L93, L310-312, L365, L366), using abbreviations before their definitions (e.g., L16, L87), mistakes in figures (e.g.,

the legend of Fig. 2: HYDAT is not a GIS dataset; the y-ticks of Fig. 7 are missing; Dim1 explains 27.7% in L318 but 28% in Fig. 6), and typos and grammatical errors. It may be beneficial to have someone do a thorough proofreading to address these issues.

- It is suggested that several statements should provide supporting references (e.g., L155, L225 (n/p<40), L469-471).

## Reference

El Adlouni, S., Bobée, B., and Ouarda, T. B. M. J.: On the tails of extreme event distributions in hydrology, *J. Hydrol.*, 355, 16–33, <https://doi.org/10.1016/j.jhydrol.2008.02.011>, 2008.

Merz, R. and Blöschl, G.: Process controls on the statistical flood moments - a data based analysis, *Hydrol. Process.*, 23, 675–696, <https://doi.org/10.1002/hyp>, 2009.

Papalexiou, S. M., Koutsoyiannis, D., and Makropoulos, C.: How extreme is extreme? An assessment of daily rainfall distribution tails, *Hydrol. Earth Syst. Sci.*, 17, 851–862, <https://doi.org/10.5194/hess-17-851-2013>, 2013.

Sassolas-Serrayet, T., Cattin, R., and Ferry, M.: The shape of watersheds, *Nat. Commun.*, 9, 1–8, <https://doi.org/10.1038/s41467-018-06210-4>, 2018.