Dear Editor,

Thank you for giving us the opportunity to submit a revised draft of our manuscript titled "*Integrated Catchment Classification Across China Based on Hydroclimatological and Geomorphological Similarities Using Self-Organizing Maps and Fuzzy C-Means Clustering for Hydrological Modeling*" for publication in *Hydrology and Earth System Sciences*.

In this revision, we have specifically addressed the request for clearer justification and clarification of the validation and statistical evidence. First, we have clarified the roles and limitations of the two-stage validation design, making a clear distinction beteween the process-oriented Stage 1 evaluation and the large-sample Stage 2 statistical validation. Second, we have refined the interpretation of the Games-Howell results to avoid overstatement, providing a more thorough explanation of why high p-values may occur for cerntain within-region subclass pairs, while the overall discrimination remains robust across complementary flow signatures. Third, we have improved the presentation of figures and tables, ensuring consistent inclusion of units in Table 2 and Figure 11. Finally, we have added an Author contributions section as requested.

We believe these targeted revisions enhance the transparency and interpretability of our validation and statistical analysis, while preserving the original objectives and conclusions of the study. A detailed point-by-point response to the referees, along with a marked-up manuscript, is provided alongside the revised manuscript. If any further revisions are necessary, we are more than happy to make additional adjustments to support the successful publication of the paper.

Yours sincerely,

Ke Zhang
Hohai University
State Key Laboratory of Water Disaster Prevention, Hohai University, Nanjing, Jiangsu, 210098, China
E-mail: kzhang@hhu.edu.cn

Responses to Reviewers' Comments

Responses to Comments from Reviewer 1:

We sincerely thank the reviewer for the positive assessment and for recommending acceptance of the manuscript as is.

Responses to Comments from Reviewer 2:

Major comment 1： While I acknowledge the addition of the large-sample statistical validation using 722 basins matched to the GRFR reanalysis, the process-based validation (Stage 1) still relies on only 10 in-situ gauged catchments. My concern is, only 10 gauges are used to represent the diversity of a national-scale classification across China. While I understand the data limitation, could you please justify why it was not possible to include one or two additional gauges for the unrepresented regions? What are the selection criteria? Are they too strict at present, and can they be loosened to include more field observations to strengthen your classification?

Response: We thank the reviewer for raising this important point. We agree that more catchments would be better to statistically represent the full diversity of China. However, our Stage 1 validation is process-oriented, aimed at illustrating whether the climate-landscape classes capture mechanistic similarities at seasonal and event time scales. This requires high-quality observations, particularly sub-daily discharge data for event-scale flow-duration curves (FDCs) and sufficient documented flood events. The 10 Stage-1 catchments were therefore selected primarily based on data availability and quality, with the following criteria: (i) long, continuous paired rainfall-runoff records, (ii) sufficient documented flood events to support event-based FDC construction, and (iii) minimal upstream regulation/abstraction. In practice, when reviewing candidate stations listed in the available hydrological yearbooks and related station metadata, we found that only a small subset of headwater basins can meet all these requirements simultaneously. In certain climate regions (e.g., arid interiors and the Tibetan Plateau), potential basins were often constrained by regulation, data gaps, or the lack of sub-daily discharge needed for event-based diagnostics.

We also acknowledge the reviewer's suggestion that adding one or two additional gauges is feasible and could enchance the process-based illustration. However, we emphasize that the key issue here is national representativeness, which would require a comprehensive, harmonized nationwide screening of gauges, as well as consistent

quality control across regions and data holders. Implementing such a fully systematic nationwide screening and harmonization would necessitate substantial additional data compilation and coordination, which falls outside the scope of the present study.

Our primary objective is to develop and evaluate a hydroclimatological-geomorphological classification framework and demonstrate its hydrological relevance through a complementary two-stage validation design. Importantly, the national-scale representativeness and statistical support are provided by Stage 2, which evaluates 722 headwater basins matched to the GRFR discharge reanalysis across all major climate regions.

To address the reviewer's concern without overstating the evidence from Stage 1, we revised the manuscript to (1) explicitly state the role, selection criteria, and limitations of Stage 1 as a data-availability-driven, process-based illustration (please see lines 135-139; 454-455), (2) clarify that Stage 2 provides the national-scale representativeness(lines 268-271), and (3) add a forward-looking note that Stage 1 can be expanded as more consistent gauge and event-scale discharge data become available (lines 629-632).

Major comment 2： You have identified the three significant variables (Qsp, Q95, BFI) for discriminating between catchment classes. However, I noticed that the Games-Howell test results yield high p-values (indicating a lack of statistical significance) for certain pairs, such as Region IV-1 and IV-2. Specifically, if the statistical evidence for distinctness is weak for some subclasses, does this suggest that a simpler classification might be sufficient for regionalization in those specific regions? Please provide a brief clarification on how you reconciled the high p-values for certain subclasses with your claim that the classification identifies "hydrologically distinct" groups.

Response: We appreciate this careful reading. We agree that not all subclass pairs exhibit significant differences for every flow signature, especially for within-region subclasses that are close in the feature space or have smaller sample sizes. Our intent is not to claim universal pairwise separation for each metric, but rather to demonstrate that the classification offers meaningful discrimination across a set of complementary signatures. The strongest separation is consistently observed for magnitude and high-flow related metrics (as summarized in Fig. 12a), while some duration/flashiness

metrics show weaker contrasts.

High p-values for specific subclass pairs (e.g., within Region IV) indicate that these subclasses may have similar values for certain signatures. This is plausible, given gradual environmental, and aligns with the fuzzy membership concept inherent in our SOM-FCM framework. To prevent overstatement, we have revised the wording around "hydrologically distinct" to clarify that distinctness is supported in aggregate across signatures (see lines 290-292; 523-524). We also added a brief note indicating that, in cases where parsimony is prioritized over fine-scale landscape stratification, users can choose to aggregate weakly separated subclasses into a coarser grouping (see lines 599-602).

We have retained the finer subclasses in the manuscript because they capture interpretable landscape contrasts, which remain valuable for applications such as parameter regionalization and donor-basin selection, even when some signatures do not show significant differences. To address the concerns, we also provide a concise discussion in the revised manuscript (see lines 599-604).

Specific comments: Table 2 and Figure 11: The unit of each variable should be labeled in the plot (on y axis or title) and table (e.g., Qsp in unit of mm).

A: We have revised Table 2 to included units for each flow signature and updated Figure 11 to consistently label units in the panel titles (please see the revised Table 2 and Figure 11).