General comments

Comments 1: The novelty of the study is unclear to me. Is this the first study of fuzzy classification for small and medium-sized watersheds in China? Compared with other classified results, what are the major differences (not methodology) or improvements, such as hydrologic signatures? This should be elaborated in the Introduction and Discussion sections.

Response 1: Thank you for your valuable comments on our research. We appreciate your insight into the novelty of our study. To clarify, while there have been previous studies on catchment classification in China, such as those based on climate and geomorphological characteristics, our study contributes in several key areas.

First, our research specifically integrates Self-Organizing Maps (SOM) with Fuzzy C-Means (FCM) clustering for small and medium-sized catchments in China, which has not been done before. This hybrid approach allows us to better capture the fuzzy boundaries between catchments with similar hydrological behaviors, an important improvement compared to traditional methods that tend to assign catchments to distinct, rigid groups. This flexibility is particularly useful for applications in ungauged basins where hydrological data may be sparse.

Furthermore, there have been relatively few comprehensive explorations of catchment classification within China. Previous studies, such as Luo (1954), divided China into three grades and nine regions based on basin boundaries, flow patterns, and sediment content, marking the first hydrological zoning scheme for China. Xiong and Zhang (1995) further divided China into 11 regions based on the annual average runoff depth. Liu (2014) also proposed three regions based on topography and climate patterns. More recently, Huan Xu (2024) utilized Fuzzy C-Means (FCM) clustering and Classification and Regression Tree (CART) methods to extend catchment classifications to ungauged basins, dividing China's basins into five clusters.

In comparison to these existing classification methods, which often focus on discrete groupings or specific climatic zones, our approach provides a more nuanced classification. We first perform climate zoning and then conduct finer-scale classification based on geomorphological features within these homogenous climate regions. This dual-layer classification system accounts for the gradual transitions between catchments, making it more adaptable and detailed. By incorporating both climate and geomorphological factors, our classification system offers a flexible framework that can be applied to ungauged catchments.

In summary, our study's novelty lies in its integrated approach, which combines SOM with FCM clustering for a multi-layered and more flexible classification of catchments. This innovative methodology provides a deeper understanding of hydrological behavior by accounting for both gradual transitions and the complex interaction between climatic and geomorphological factors, making it more suitable for hydrological modeling, especially in ungauged regions.

We sincerely thank the reviewer for this insightful comment, and we will provide a more detailed explanation of the novelty of our research in the Introduction and Discussion sections.

Comments 2: The structure of the paper needs to be reorganized. The Results section contains many texts that should be moved to Methods and Discussion. For example, Line 302-306 for how the optimal number of clusters is chosen should be moved to the Methods explaining FCM; Line 407-498 for the flow duration curve and hydrologic signature looks like a great point that should be moved to the Discussion section. Overall, the current organization, having discussions inside Results, makes the manuscript long and disruptive to read. The manuscript should be organized more neatly, where the Results should focus on presenting numbers, while moving and consolidating interpretations and implications in the Discussion.

Response 2: Thank you for your constructive suggestion. We greatly appreciate your feedback regarding the organization of our manuscript. We understand the importance of presenting the content in a clear and concise manner, and we agree that restructuring the paper would enhance its readability.

In response to your comment, we will reorganize the manuscript by focusing on adjusting the structure of the Results and Discussion sections to improve its flow. We believe this reorganization will make the manuscript more streamlined and easier to follow. Thank you again for your helpful suggestion.

Comments 3: The validation of classification is only performed for 10 watersheds in all entire China, which I think is insufficient. Based on the Figure 6, there are many same-class watersheds that are fairly distant from each other. However, the current selection of watersheds, though the similarity of FDC in each class is shown, might be insufficient to support the conclusion, as these watersheds in same classes are too spatially close to each other. Therefore, I am wondering how the similarity of FDC would be if watersheds that are more spatially distant are chosen for evaluation.

Response 3: Thank you for your valuable suggestion. We appreciate your concern regarding the validation of our classification, specifically the limited number of 10 watersheds across China. While these catchments are representative, their relatively close proximity may not fully capture the spatial variability within the same class.

In response, we will expand the validation by including more spatially diverse watersheds within the same climatic and landscape categories. This will allow us to better assess how the similarity of flow duration curves (FDCs) holds for geographically distant catchments and further validate the robustness of our classification framework.

Our manuscript already discusses the spatial heterogeneity observed across different climate regions, as seen in the significant differences in seasonal flow regimes and FDC curves (Fig. 8). While catchments within the same region often exhibit similar hydrological behaviors, there is substantial variability within the same class. We will highlight how spatially distant watersheds, located within the same climatic homogeneity region, contribute to validating the effectiveness of our classification method.

However, it is important to note that the availability of continuous hydrological data, particularly in remote areas, is a limitation that restricts the ability to include a larger number of spatially diverse catchments in the validation. Despite this constraint, we believe the validation performed with the available data provides valuable insights into the classification's performance.

Thank you again for your insightful suggestion, which will help strengthen the validation of our study.

Comments 4: The application of the study is not thoroughly discussed. The section 4.1 focuses more on the advantages of the probabilistic approach of FCM over hard-boundary classification ones, and the potential of improving regional hydrological modeling. However, the potential of 1) transferring model parameters from calibrated to ungauged watersheds and 2) estimating floods under various design storms based on the similarity of flow duration curve could be discussed, and can improve the novelty and value of the research.

Response 4: We sincerely appreciate this constructive suggestion. We agree that explicitly linking our classification framework to practical hydrological applications would significantly enhance the study's impact, especially for ungauged catchment management.

In response, we will expand Section 4.1 to include a more detailed discussion on the following applications:

Transferring Model Parameters from Calibrated to Ungauged Watersheds: Hydrological modeling in ungauged basins is a major challenge, particularly due to the lack of observational data for calibration. Transferring model parameters from calibrated watersheds to ungauged catchments is crucial for improving hydrological model performance in such regions. Our classification system, which groups watersheds with similar climatic and hydrological behaviors, offers a promising solution for this challenge. Specifically, we will illustrate how the Xinanjiang Model, a widely used hydrological model in China, can benefit from this approach.

We will discuss how our classification system allows for the transfer of model parameters from well-calibrated basins to ungauged catchments by identifying hydrologically similar watersheds. Additionally, we will emphasize how this method improves the transferability of model parameters across larger regions. This enhances the applicability of hydrological models in areas where observational data are sparse, ultimately facilitating more accurate simulations and predictions.

This added discussion will highlight how our approach can significantly improve regional hydrological modeling, particularly in regions where traditional calibration is difficult due to data scarcity. We believe these additions will demonstrate the practical applications and the novelty of our classification approach, making it more relevant for ungauged catchment management and improving flood prediction and water resource planning.

Thank you again for your valuable suggestion, which will help strengthen the practical contribution of our research.

Comments 5: I think the references are not in the required style of HESS (https://www.hydrology-and-earth-system-sciences.net/submission.html#references)

Response 5: Thank you for pointing this out. We apologize for the formatting issues with the references. We will carefully review the reference list and ensure that all citations are formatted according to the HESS style guidelines, as provided in the submission instructions. This will be corrected in the revised manuscript to meet the journal's requirements.

Thank you again for your attention to detail.

Specific comments

Comments 1: Line 69: "an indisputable fact" looks like a strange statement. Do you mean machine learning is now widely used for regionalization studies?

Response 1: Thank you for pointing out this concern. We agree that the phrase "an indisputable fact" may sound overly definitive. We meant to convey that machine learning is now widely applied and increasingly recognized in regionalization studies. To clarify this, we will revise the sentence to:

"*With the advancement of computer technology in the 21st century, the widespread use of machine learning in regionalization studies has become increasingly common and well-recognized (Yang et al., 2020a).*"We believe this revision more accurately conveys the intended meaning and avoids any ambiguity.

Thank you again for your constructive feedback.

Comments 2: Line 102-103: need to add references.

Response 2: Thank you for raising this important point regarding the need for references. We agree that supporting the statement with relevant literature will strengthen the argument. In response, we will add appropriate references to support the claim that climate patterns significantly influence the hydrological response of catchments, including their effects on soil moisture availability and the co-evolution of landscape and vegetation.

Thank you for your insightful feedback, which has helped us improve our work.

Comments 3： Line 117-126: The organization of this paragraph needs improvement. The six indices should be stated before reasoning why they are selected. It would make the flow more logical, rather than making readers wonder what indices are chosen (line 119, three indices but not stating what they are).

Response 3: Thank you for your valuable suggestion. We appreciate your feedback regarding the organization of the paragraph. We agree that stating the six climate indices before explaining the reasoning for their selection would improve the flow and clarity of the discussion.

In response, we will restructure the paragraph to first list the six selected indices and then explain the rationale behind their selection. This will help readers understand the indices from the outset and follow the reasoning more logically. The revised paragraph will be as follows:

"*To capture the seasonal and spatial variability of temperature, we selected six indices for climate classification: average moisture index (I_m), seasonal moisture index ($I_{m,r}$), fraction that falls as snow (fs), annual average temperature (T_m), seasonal temperature ($T_{m,r}$), and the fraction of snowy days (Ds). The first three indices are expressed using a version of Thornthwaite's moisture index MI (Willmott and Feddema, 1992). These indices were chosen based on their ability to represent key climate factors, including moisture availability, snow distribution, and temperature, which are critical for understanding climate variability and hydrological responses. The indices were calculated for each 0.25° land cell using the CRU TS V4.04 dataset and meteorological station data. Although some of these indices have been previously used to map climate homogeneity regions, they have not been combined in this specific way.*"

This revision will make the paragraph more logical and easier to follow. Thank you again for your helpful suggestion.

Comments 4: Line 174: reference for FCM?

Response 4: Thank you for your question regarding the reference for fuzzy c-means clustering (FCM). We will include the appropriate reference for FCM to support the discussion. We will revise the sentence as follows: "*Fuzzy c-means clustering (FCM), based on fuzzy set theory, is one of the most widely used soft clustering algorithms (Bezdek, 1984). Unlike hard clustering algorithms, such as k-means and hierarchical clustering, the FCM cluster procedure uses a fuzzy parameter to create overlapping cluster boundaries.*"

We will update the manuscript to include this reference. Thank you again for your helpful suggestion.

Comments 5: Line 178: “may be the most” to “is a”.

Response 5: Thank you for your suggestion. We agree that changing "may be the most" to "is a" will make the statement more definitive and clear. We will make this revision to improve the strength of the claim.

Comments 6: Line 207: reference for Penman-Monteith equation.

Response 6: Thank you for your comment regarding the reference for the Penman-Monteith equation. We will include the appropriate citation to support the use of this formula. We will revise the sentence to include the reference as follows:

"*The EP is estimated using a variant of the Penman-Monteith formula (Moratiel, 2020). The climate data were interpolated at a 0.25° × 0.25° spatial resolution, and missing data were filled using the weighted nearest-neighbor approach.*"

We will update the manuscript to include these references. Thank you again for your valuable suggestion.

Comments 7: Line 214: What are the average size and the range of catchment sizes?

Response 7: Thank you for your comment regarding the size of the catchments. Based on the HydroBASINS dataset, the average catchment size is 761.04 km^2, with the sizes ranging from 15 km^2 to 14,612.8 km^2. This information will be added to the manuscript to provide a clearer understanding of the spatial extent of the catchments analyzed.

Comments 8: Figure 2: Though the interpretations of hexagons values are provided, I still don’t quite get the physical meanings of these plots and wonder how they should be interpreted spatially on maps (if a basemap can be added, it would be helpful). In the figure’s caption, briefly explain the legends and how readers should interpret the figure. Also, line 276-285 should be moved to Methods

Response 8: Thank you for your valuable feedback. We appreciate your suggestion to enhance the interpretation of Figure 2. We will make the following improvements in response to your comments:

We will revise the figure caption to provide a clearer explanation of the physical meaning of the hexagonal values shown in Figure 2. Each hexagon represents a unit of the self-organizing map (SOM) grid, with its color corresponding to the weight vector values of the climate indices. The color scale ranges from blue (low values) to red (high values), and we will explain that these color values represent the relative intensity of the climate indices across the SOM grid. We will also briefly mention how the component planes show spatial patterns of different climate variables, and how these patterns relate to the underlying catchment characteristics.

To improve the spatial interpretation, we will add labels for climate regions in Figure 2. This will help readers better understand the geographic distribution of the climate data and the relationship between climate regions and catchment characteristics. The figure caption will be updated to include a more detailed explanation of the legends and how to interpret the color-coded values for each hexagon. Specifically, we will explain how the color scale corresponds to the values of the climate indices and their spatial distribution across the SOM grid.

We agree that the explanation of the component planes and d-matrix in lines 276-285 would be more appropriate in the Methods section. We will move this content to ensure the Results section focuses on the analysis, while the Methods section provides the necessary background on the SOM and FCM algorithms and how they were applied.

Thank you again for your constructive suggestions, which will enhance the presentation and interpretation of the results.

Comments 9: Line 304: What is the AP algorithm? I don't think this was mentioned in Methods, and should add the reference

Response 9: Thank you for your valuable comment. We apologize for the confusion regarding the AP algorithm. The AP algorithm refers to the Affinity Propagation algorithm, a clustering algorithm that does not require the specification of the number of clusters in advance. It identifies cluster centers and assigns data points to the closest exemplar based on similarity.

We will add a brief explanation of the AP algorithm in the Methods section and include a citation to the relevant reference for this algorithm. The revised text in the Methods section will read as follows:

"The Affinity Propagation (AP) algorithm (Frey & Dueck, 2007) was used to determine the maximum number of clusters. This algorithm identifies exemplar points in the

dataset and assigns other points based on similarity, without requiring the user to specify the number of clusters in advance."

Additionally, we will update Section 3.1.3 to clarify this point:

"Before clustering the output neurons from the SOM competitive layer using the FCM algorithm, two validation metrics were calculated (i.e., DBI and SC) to determine the optimal number of clusters. An experiment was conducted to test the number of clusters from two to the maximum number determined by the Affinity Propagation (AP) algorithm (Shang, 2016)."

Thank you again for pointing this out. We will make these changes in the manuscript to provide clarity and ensure proper referencing.

Comments 10: Line 369: Two questions here: 1) For soil & veg characteristics, the second PC has an eigenvalue of 0.91. Why do you choose this PC below one? 2) Improve the writing: Instead of using semicolons, clearly state which class (topographic, soil & veg, and tolological) you are discussing first (For topographic, XXX. For soil & veg, XXX.), then state how each PC is correlated to the input indices.

Response 10: Thank you for your valuable feedback. We appreciate your questions and suggestions for improving the clarity of the manuscript. Here's how we will address your concerns:

According to the standard rule for PCA, we typically retain components with eigenvalues greater than 1. However, in this case, we kept the second principal component due to its significant contribution to the variance explained in the data (it helps explain a portion of the soil and vegetation characteristics, especially those that were not fully captured by the first component). While the eigenvalue is slightly below 1, the cumulative proportion of variance explained by the two components combined is greater than 70%, justifying the inclusion of both components. This will be clarified in the revised manuscript.

To improve clarity, we will revise the text to explicitly identify which feature class is being discussed before presenting the principal components' correlations. Instead of using semicolons, we will structure the sentences to state which class (topographic, soil & vegetation, and topological) is being discussed first, followed by the correlation of each principal component with the relevant indices.

Thank you again for your insightful suggestions, which will help improve the clarity and robustness of our study.

Comments 11: Line 449: space between class and (Li et al).

Response 11: Thank you for pointing out the formatting issue. We will add a space between "class" and the citation "(Li et al., 2018)" to ensure proper formatting.

Comments 12: Figure 7: some recommendations: 1) List each site/catchment's classification in the line charts, and 2) maybe consider another color scheme to present the variation range. The grey colors are hard to differentiate. Also, briefly describe the ranges in the legend within the caption for people to understand.

Response 12: Thank you for your valuable feedback and suggestions regarding Figure 7. Here's how we will address your comments:

We will add labels to the line charts to clearly indicate the classification of each catchment. This will help readers easily identify the site corresponding to each line, improving the clarity of the figure. We agree that the grey colors are difficult to differentiate. We will consider using a more distinct and visually accessible color scheme to present the variation range. This will make it easier for readers to distinguish between the different percentiles (e.g., 10%, 50%, 90%). We will modify the figure caption to briefly describe the ranges represented by the colors in the legend. This will help readers better understand the variation in runoff across the different catchments and regions.

Thank you again for your helpful suggestions, which will enhance the figure's clarity and overall presentation.

Comments 13: Line 495: correct the citation

Response 13: Thank you for pointing out the issue. We apologize for the oversight. We will revise the sentence as follows:

"Additionally, catchments with large spatial distances are capable of exhibiting similar hydrological characteristics (e.g., Maduwang and Daiying)."

We will ensure that the correct context is clearly presented in the manuscript.

Thank you again for your attention to detail.

Comments 14: Figure 8: What will FDCs look like if using discharge in mm/day (normalized by drainage area)? I am wondering this because these watersheds vary significantly in drainage size, and normalizing discharge by size may allow for expanding the validation to more gauged watersheds.

Response 14: Thank you for your insightful suggestion regarding the normalization of discharge by drainage area. We agree that normalizing discharge by the catchment area could be a valuable approach. By normalizing the discharge in units of mm/day (precipitation depth per day), the FDCs would become more comparable across watersheds of different sizes, allowing for a more consistent validation across a broader range of gauged catchments.

We will update the manuscript to include a discussion of this potential modification. Specifically, we will mention that normalizing discharge by drainage area could allow for extending the validation of our classification method to more gauged watersheds, especially those with different catchment sizes.

Thank you again for your valuable suggestion, which will help enhance the robustness of our analysis.

Comments 15: Line 517: The statement, “leading to errors”, should be more evidence-based. What specific errors could inaccurate classification result in? What consequences/risks will these errors cause? Provide references of previous studies showing so.

Response 15: Thank you for your valuable comment. We agree that the statement, “leading to errors,” should be more evidence-based. Inaccurate classification due to the introduction of artificial boundaries can lead to several errors in hydrological modeling:

Artificial boundaries may cause misclassification of hydrological regions, leading to inaccurate parameterization of models. For instance, a catchment classified in the wrong climatic zone may have model parameters (e.g., runoff coefficients, baseflow, etc.) that do not properly reflect its hydrological processes, resulting in less accurate predictions of runoff and streamflow. Misclassification at the boundaries of climatic zones could lead to incorrect predictions of flood events, as the flow regimes and seasonal patterns may be poorly represented. This could significantly affect flood risk assessment and the design of flood protection measures.

We will revise the manuscript to incorporate these explanations, along with relevant references, to strengthen the evidence supporting our claim. Thank you again for your insightful comment, which will improve the clarity and depth of our manuscript.

Comments 16: Line 546: Is this a possible reason causing the challenge that some watersheds are hard to be classified with one dominant group?

Response 16: Human activities, such as urbanization and agriculture, can make it more challenging to classify some watersheds into one dominant group. These human influences can cause significant deviations in hydrological behavior, even among watersheds with similar climatic and topographical characteristics. As a result, catchments with similar natural features may exhibit different hydrological responses due to anthropogenic factors, complicating the classification process.

This variability can contribute to the challenge of grouping certain watersheds into a single, dominant cluster. We emphasize the importance of considering both natural and anthropogenic factors in the classification process, which requires a more comprehensive analysis of watershed characteristics. We will revise the manuscript to include this explanation, highlighting the importance of integrating human factors into the classification process in future studies.

Thank you again for your valuable suggestion, which will enhance the clarity and depth of our discussion.