Dear Dr. Bogaard,

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*. We greatly appreciate the time and effort that both you and the reviewers have dedicated to providing feedback on our manuscript. We are also grateful for the insightful comments that have significantly contributed to improving our work.

In response to the reviewers' feedback, we have made several important revisions to the manuscript. The changes we made include:

Addressing Novelty: In line with the reviewers' suggestions, we have emphasized the novel aspects of our research, specifically detailing what distinguishes our findings from existing literature and what new knowledge our study provides.

Revising the Structure: In response to the reviewers' comments, we reorganised the manuscript to improve logical coherence and readability. The Results section now reports the key findings, while interpretation and implications are consolidated in the Discussion. We also strengthened the validation framework to provide clearer evidence for the applicability of the proposed approach and the robustness of our conclusions.

Clarifying Focus: We have revised the manuscript to focus more on the scientific implications of our results, rather than placing excessive emphasis on the mathematical techniques used. This shift strengthens the contribution of the paper to the field.

Further details of the changes can be found in the revised manuscript, where we have highlighted the modifications using tracked changes. Below, we also provide a point-by-point response to the reviewers' comments and concerns.

We hope that the changes we have made resolve all concerns and improve the overall quality of the manuscript. If any further revisions are necessary, we are more than happy to make additional adjustments to facilitate a successful publication.

Thank you and best regards.

Yours sincerely,

Ke Zhang

Hohai University

The National Key Laboratory of Water Disaster Prevention, Hohai University, Nanjing, Jiangsu, 210098, China

E-mail: kzhang@hhu.edu.cn

Responses to Reviewers' Comments

Dear reviewers,

Thank you very much for your comments and professional advice. These opinions help to improve the academic rigor of our article. Based on your suggestions and requests, we have made the corrected modifications to the revised manuscript. We hope that our work can be improved further. Below are the details of our responses to the reviewers' comments:

Responses to Comments from Reviewer 1:

Q1: Table 3 Why are the attributes in Table 3 selected based on the coefficient of variation?

A: Thank you for your valuable comments on our research. The attributes in Table 3 were selected based on the coefficient of variation (CV) because it serves as a measure of relative variability, which is particularly effective for identifying the most stable attributes within each catchment cluster. By focusing on the attributes with the lowest and second-lowest CVs, we aimed to highlight those that exhibit consistent behavior across different catchments within each climate region, making them more representative of the typical characteristics of the clusters. Additionally, using CV allows us to account for the inherent variability within the dataset, ensuring that the selected attributes are robust and not unduly influenced by outliers or extreme values. Scaling these attributes by the mean CV of the dataset further normalizes the variability, providing a clearer and more standardized comparison between the clusters.

Q2: Figure 2 It is suggested to include an explanation of the d-matrices in the methodology. Consider moving the statement "Vesanto (1999) suggested that SOM results can be expressed in the form of two types…" from L279 to section 2.1.2 and expand on it in more detail.

A: Thank you for this helpful suggestion. We agree that the distance matrix (d-matrix) should be introduced in the Methodology so that readers understand its meaning and how it supports the interpretation of the SOM results before the figure is presented (Fig. 2 in the original manuscript; Fig. 3 in the revised manuscript). Following your recommendation, we moved the statement citing Vesanto (1999) from its original location (former L279) to Sect. 2.3.1 (SOM–FCM combined algorithm) and expanded the accompanying text to clarify both the definition of the d-matrix and its role in our analysis.

In the revised manuscript, we explain that the d-matrix visualises the average Euclidean distance between neighbouring SOM neuron weight (codebook) vectors. Larger distances indicate sharper transitions in the mapped feature space and typically coincide with boundaries between clusters, whereas smaller distances indicate locally homogeneous areas. We also clarify how the d-matrix complements SOM component planes in this study: component planes display the spatial patterns of individual indices or descriptors across the SOM, while the d-matrix provides a diagnostic view of overall neighbourhood dissimilarity and helps assess whether the identified climate regions and catchment classes are consistent with the separation structure present in the feature space.

Revisions made in the manuscript:

Inserted/updated text (Sect. 2.3.1):

SOM component planes and distance matrices (d-matrices) are used diagnostically to examine variable relationships and to verify that the identified clusters are consistent with the underlying feature space (Vesanto, 1999).

Updated caption (Fig. 3 in the revised manuscript):

Figure 3. Self-organising map representation of the hydroclimatic feature space and fuzzy climate-region partition: (a) SOM component planes for the five climate indices: each hexagon represents a neuron summarising grid cells with similar standardised index values, colours show the corresponding codebook (weight) value, and black contour lines delineate the six SOM–FCM climate clusters that are later mapped to geographic space. (b) SOM distance matrix (d-matrix) giving the Euclidean distance between neighbouring neurons; cooler colours indicate smoothly varying hydroclimatic conditions, warmer colours mark sharper transitions in the climate-index space, and labels I–VI identify the six climate regions used in subsequent spatial analyses.

Q3: Figure 3 Consider adjusting the color band so that the color corresponding to 0.5 is set to white. This would better highlight basins belonging to a cluster with higher confidence.

A: Thank you for this helpful suggestion. We agree that emphasising higher membership values improves the interpretability of the fuzzy climate region maps and more clearly distinguishes core areas from transitional zones.

In the revised manuscript, we implemented a threshold-based visualisation, grid cells with membership values below 0.5 are set to white (masked), and only memberships greater than 0.5 are displayed using the colour scale. This directly highlights areas with higher confidence assignments and avoids visually over-interpreting low membership values. The figure caption has been updated accordingly, and the figure number has changed from Fig. 3 in the original manuscript to Fig. 4 in the revised manuscript.

Revisions made in the manuscript:

Updated Figure (Fig. 4 in the revised manuscript):

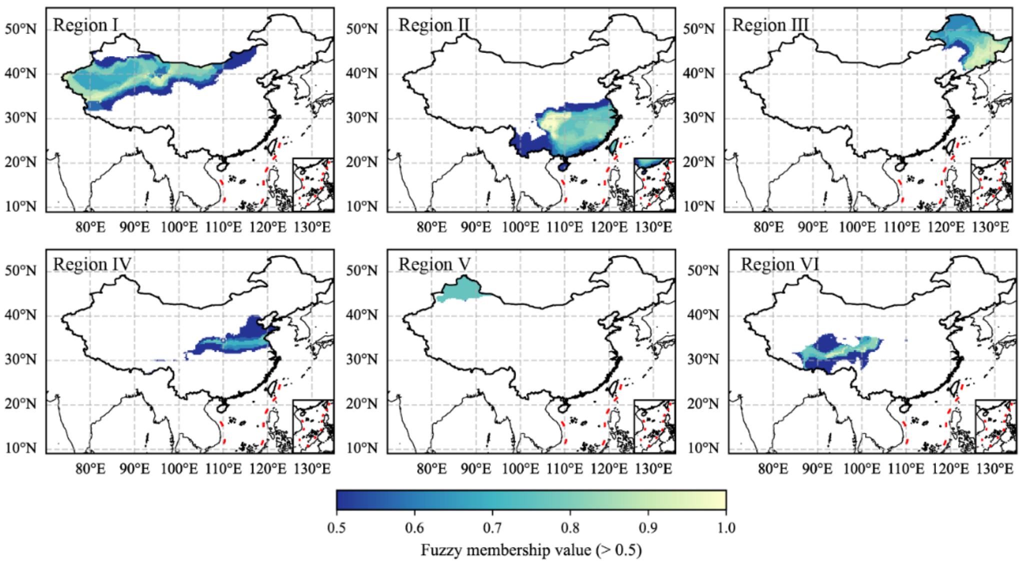

***Figure 4**. Spatial distribution of fuzzy membership to the six SOM-FCM climate regions across China; only grid cells with membership values greater than 0.5 are shown.*

Q4: Figure 6   It is recommended to use different color schemes for the third and fourth categories, as their current colors are too similar and not effective.

A: Thank you for this helpful suggestion. We agree that the original colours used for Categories III and IV were too similar, which reduced visual contrast and could hinder interpretation of the spatial patterns.

In the revised manuscript, we updated the colour scheme of this map by assigning clearly distinguishable colours to Categories III and IV, while keeping the remaining category colours unchanged. This adjustment improves the legibility of the figure, supports consistent interpretation across adjacent regions, and enhances readability in both digital and printed formats. The legend has been updated accordingly. Please note that this figure is now presented as Fig. 7 in the revised manuscript.

Revisions made in the manuscript:

Updated Figure (Fig. 7 in the revised manuscript):

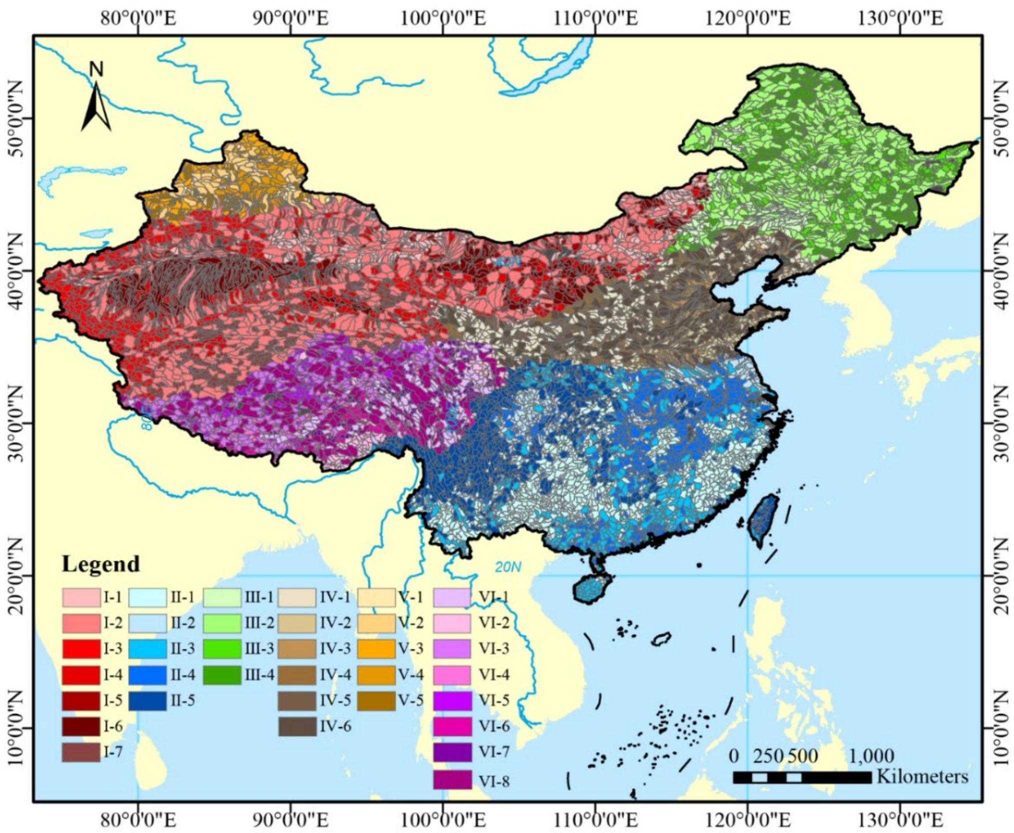

Figure 7. Spatial distribution of the climate-landscape catchment classes across China derived from the hierarchical SOM-FCM classification.

Q5: Figure 7 Consider clearly marking the boundaries of each climate zone in the figure and labeling the basin class in the subplots.

A: Thank you for this helpful suggestion. We agree that explicitly delineating climate zone boundaries on the map and making the basin class information more prominent in each subplot improves figure readability and helps readers link the process based validation to the climate–landscape classification.

In the revised manuscript, the climate zone boundaries, which were shown in grey in the original figure, have been redrawn using a clearer blue outline. We also removed the river network elements that were not essential for interpreting the validation results, thereby reducing visual clutter. In addition, we ensured that each subplot explicitly reports the basin class by including the class code in the subplot title (for example, “Suide (IV-4)”). Here, the Roman numeral denotes the climate region and the Arabic numeral denotes the catchment class within that region. Please note that, due to figure renumbering during revision, this figure is now presented as Fig. 8 in the revised manuscript.

Revisions made in the manuscript:

Updated Figure (Fig. 7 in the revised manuscript):

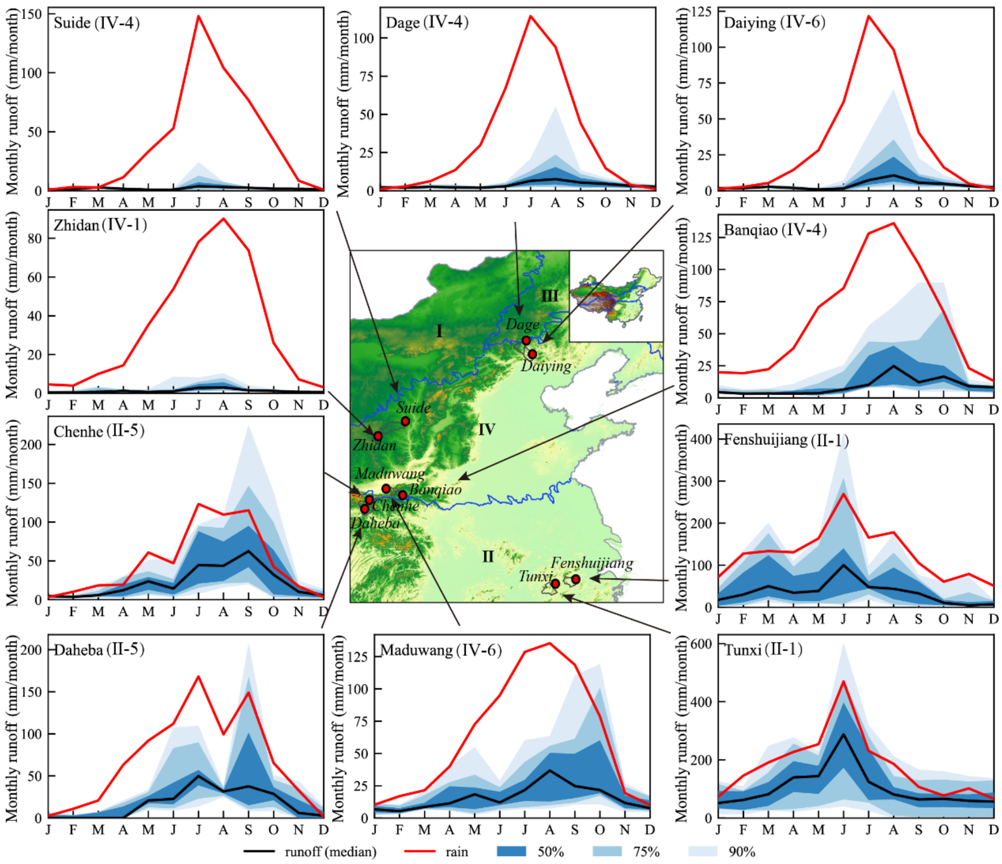

Figure 8. Seasonal runoff of the ten gauged catchments used for process-based validation.

Q6: Introduction The paper overlooks previous catchment classification studies conducted in China:

Luo, K. (1954) Draft of natural geography regionalization of China. (in Chinese)

罗开富,1954. 中国水文区划草案.

Xiong, Y., Zhang, J., et al. (1995) Hydrology Regionalization of China, Science Press, Beijing.(in Chinese)

熊怡,张家桢,等,1995. 中国水文区划. 科学出版社

Liu, C., Zhou, C., et al. (2014) Chinese Hydrological Geography, Science Press, Beijing

刘昌明，周成虎等，2014. 中国水文地理. 科学出版社

Xu, H., Wang, H., Liu, P. (2024). Identifying control factors of hydrological behavior through catchment classification in Mainland of China. Journal of Hydrology, 645, 132206. DOI: 10.1016/j.jhydrol.2024.132206

A: Thank you for highlighting this important point. We agree that the Introduction should more comprehensively acknowledge catchment classification and hydrological regionalisation efforts in China. In the revised manuscript, we added a dedicated paragraph summarising both early hydrological zoning schemes and recent large-sample classification studies in Mainland China, and we cited the references suggested by the reviewer (Luo, 1954; Xiong and Zhang, 1995; Liu et al., 2014; Xu et al., 2024). This addition also helps clarify how the present study builds on, and differs from, previous work.

Specifically, we added a concise overview of catchment classification and hydrological zoning studies in China, from early national schemes based on a limited set of indicators to recent large sample work by Xu et al. (2024) using flow signatures that highlights the joint role of climate and soil properties. We then clarified the remaining gap and our contribution: many studies cluster climatic and landscape descriptors in a single feature space without explicitly separating large scale climatic forcing from within region landscape modulation. To address this, we propose a hierarchical national framework that first delineates climate regions along continuous hydroclimatic gradients and then derives landscape based catchment classes within each climate region using SOM–FCM, improving interpretability and supporting regionalisation and runoff prediction in ungauged basins.

Revision made in the manuscript (revised manuscript, Lines 63–77):

China is an ideal study area for developing an integrated and scale aware classification framework because it spans strong regional contrasts, from cold and dry plateaus to warm and humid plains, and includes diverse terrains and climates. Early hydrological zoning in China relied on a limited set of indicators. Luo (1954) proposed one of the earliest national schemes based on basin boundaries, flow patterns, and sediment characteristics. Yi and Jiazhen (1995) delineated 11 regions using mean annual runoff depth as a primary indicator, and Liu et al. (2014) divided China into three broad regions based on topography and climate patterns. These foundational studies provided valuable national perspectives, but they were not designed to represent catchment scale hydrological behaviour in a multidimensional sense, which is important for similarity-based model transfer. More recently, improved datasets have enabled more data driven *classification efforts. For example, Xu et al. (2024) classified Chinese catchments using flow signature information and analysed the associated controls, highlighting the joint role of climate and soil properties in distinguishing basin groups. Such progress is highly valuable for advanceing large sample hydrology in China. However, their clustering strategy integrates climatic and landscape descriptors within a single feature space and therefore does not explicitly separate the effects of large-scale climatic forcing from within climate landscape modulation. As a result, a hierarchical national framework that systematically combines continuous hydroclimatic gradients with fuzzy landscape similarity, and remains directly applicable to ungauged basin prediction, still needs to be established.*

Q7: Methodology In L180, the author claims FCM has "low sensitivity to initialization." I am curious if this is the case, and it might be beneficial to demonstrate FCM results under multiple initializations.

A: Thank you for this important comment. We agree that the sensitivity of fuzzy c-means (FCM) to random initialisation should not be asserted without evidence. In the revised manuscript, we therefore removed the unqualified statement that FCM has "low sensitivity to initialization" and instead assessed the robustness of our FCM results empirically.

Specifically, the climate indices were first projected onto a two dimensional SOM with a 19 × 22 rectangular grid (418 neurons), selected based on low quantisation error (QE) and topological error (TE). FCM clustering was then applied to the SOM codebook vectors, and the optimal number of clusters was determined as six by jointly minimising the Davies–Bouldin index (DBI) and maximising the silhouette coefficient (SC). To evaluate sensitivity to initialisation, we repeated the FCM step for k = 6 a total of 50 times using different random initialisations. Across these runs, about 80 % of SOM neurons retained the same dominant cluster assignment on average, and differences were limited to a small number of boundary neurons where memberships are inherently mixed. We adopted the solution with the lowest FCM objective function among the 50 runs.

Revisions made in the manuscript (revised manuscript, Lines 327):

To assess robustness, the FCM step with k=6 was repeated 50 times with different random initialisations; on average, about 80 % of SOM neurons retained the same dominant cluster, and discrepancies were confined to a small fraction of boundary *neurons. The solution with the lowest FCM objective function was adopted.*

Q8: Methodology It is suggested that the methods used in the results section be introduced in the methodology, highlighting the logic and approach rather than just detailing the SOM and FCM algorithms. A flowchart would be helpful if possible.

A: Thank you for this constructive comment. We agree that the Methodology should emphasise the overall study logic and clearly introduce all analyses before the corresponding results are presented. In response, we substantially reorganised Sect. 2 and added a workflow flowchart (now Fig. 1) to summarise the end to end procedure. The revised Methodology now begins with an explicit overview of the multi-step workflow and directs readers to Fig. 1 , whereas the previous version placed substantial methodological explanation directly in the Results section (e.g. the description of SOM map diagnostics and interpretation) .

In the revised manuscript, Sect. 2 is structured to follow the analytical pipeline: data compilation (Sect. 2.1), derivation of climate indices and catchment descriptors (Sect. 2.2), hierarchical SOM–FCM classification (Sect. 2.3) , and hydrological validation and statistical analysis (Sect. 2.4) . Importantly, the validation methods that underpin Sect. 3 are now fully introduced in Sect. 2.4, including the seasonal and event scale analyses, the flow signature set, and the Games–Howell testing.

Revision made in the manuscript (revised manuscript, Sect. 2):

We added a workflow overview and flowchart at the start of Sect. 2 and reorganised Sect. 2 into Sects. 2.1-2.4 to present the full pipeline from data and index construction to hierarchical SOM-FCM classification and validation.

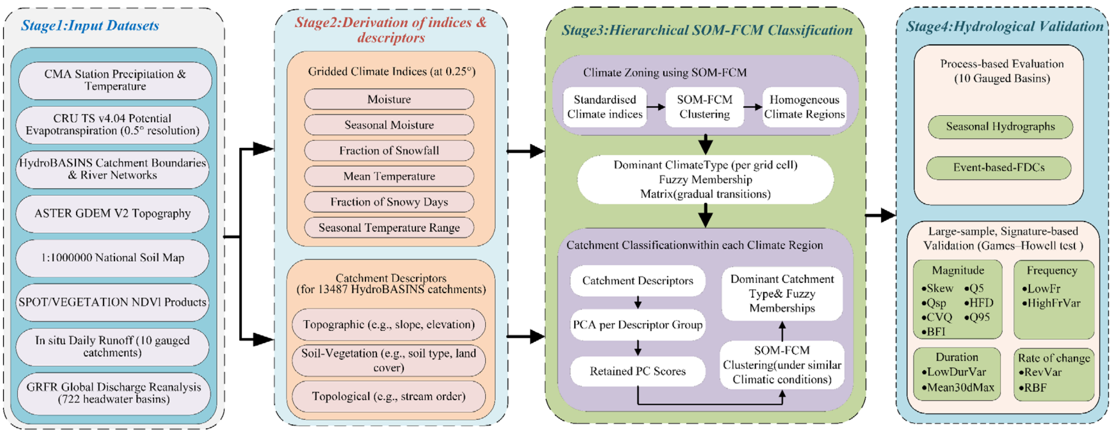

Figure 1. Flow chart of the different steps followed in the study.

Q9: Methodology How to classify catchment from climate region to basin class? FCM?

If so, are the inputs to FCM the features in Table 1 or their principal components?

A: Thank you for raising this important point. In the revised manuscript, Sect. 2.3 (Catchment classification method) has been rewritten to clarify the complete classification workflow from climate regions to catchment classes and to specify the inputs used for SOM–FCM at each stage.

Briefly, SOM–FCM is applied in a hierarchical two stage procedure. First, SOM–FCM is applied to the climate indices on the 0.25° grid to delineate homogeneous climate regions (Sect. 2.3.2). Each catchment is then linked to a climate region using the dominant climate type at its outlet grid cell (Sect. 2.3.3). Second, catchments are classified within each climate region using landscape information (Sect. 2.3.3). In this within region step, the clustering is not performed on the raw descriptors in Table 1. Instead, we reduce redundancy by examining rank correlations and applying PCA separately to the three descriptor groups (topographic, soil and vegetation, and drainage network). Principal components are retained until the cumulative explained variance stabilises, and the retained component scores are used as clustering variables. For each climate region, a SOM is trained on the standardised retained principal component scores, and FCM is applied to the corresponding SOM codebook vectors. Catchment memberships are inherited from the best matching SOM neuron, and the dominant class is defined by the highest membership value, with a membership threshold of 0.5 used to identify clearly defined versus transitional catchments.

Revision made in the manuscript (revised manuscript, Sect. 2.3):

We rewrote Sect. 2.3 (Catchment classification method) to clearly describe the hierarchical SOM–FCM workflow and to clarify how catchments are assigned from climate regions to catchment classes.

2.3 Catchment classification method

The aim of the catchment classification is to delineate hydrologically similar regions by jointly accounting for large-scale climatic controls and local landscape characteristics. To this end, a hierarchical two-stage procedure is used that combines Self-Organising Maps (SOM) and Fuzzy C-Means (FCM). In the first stage, the SOM-FCM framework is applied to the climate indices to identify homogeneous climate regions. In the second stage, SOM-FCM is applied to the catchment descriptors within each climate region to derive catchment types under similar climatic conditions…

Q10: Results Were the selected 10 small watersheds affected by human activities, such

as agricultural water use or urban consumption? Would this impact the results?
Using 10 small watersheds for validation might be insufficient. If the author is willing, more runoff data can be found in NESSDC (https://www.geodata.cn), such as:
DOI: 10.12041/geodata.30184613892738.ver1.db
DOI: 10.12041/geodata.69811525443157.ver1.db
DOI: 10.12041/geodata.31258482188424.ver1.db

A: Thank you for this thoughtful comment. We agree that human influences and sample size are important considerations for process-based validation.

For the ten gauged headwater catchments, we screened the candidates to minimise anthropogenic disturbance. As stated in Sect. 2.1, the selected catchments satisfy three criteria, including no major upstream regulation or abstractions. This selection reduces the likelihood that reservoir operation or large withdrawals dominate the observed seasonal regimes or event responses. We recognise that minor water use may still occur in some basins, but the validation in this study focuses on broad differences in seasonal runoff regimes and event scale response shapes, and these patterns remain consistent with the intended near natural setting of the selected catchments.

We also agree that ten catchments alone cannot represent all climate landscape classes. For this reason, the manuscript uses a two-stage validation framework: the ten gauged catchments provide process-based evidence across time scales, and this is complemented by a large sample statistical validation using 722 headwater basins matched to GRFR, retaining only basins with area mismatch below 10 % and without documented upstream modifications. This combined design strengthens the robustness of the hydrological evaluation beyond the limited gauge sample.

Finally, we acknowledge this limitation explicitly in the revised manuscript. We note that the process-based evaluation relies on a limited set of gauged catchments and that the framework focuses on near natural basins, while human influences such as reservoir operation, irrigation, and urbanisation are not represented explicitly. We also state that expanded gauging networks and alternative discharge products would be valuable for further testing.

Revisions made in the manuscript:

We clarified the near natural screening criteria for the ten gauged catchments and the complementary large sample validation using 722 headwater basins in Sect. 2.1, and strengthened the discussion of limitations and human influence in the revised Discussion and Conclusions.

Updated text (added in Sect. 2.1 Database):

The selected catchments satisfied three criteria: (i) 10-15 years of continuous daily rainfall and runoff records with 10 -35 documented flood events, (ii) no major upstream regulation or abstractions, and (iii) representation of a broad range of climate regions and catchment classes (Li et al., 2018). In addition, flow signatures were derived from the Global Reach Level Flood Reanalysis (GRFR) dataset (Yang et al., 2021b), which provides 3 hourly discharge time series for river reaches worldwide from 1980-2019 and has been evaluated against daily discharge records from more than 14 000 gauging stations. HydroBASINS catchments were matched to GRFR river reaches based on upstream drainage area. Headwater basins with a relative area mismatch below 10 % and without documented upstream modifications were retained, yielding 722 representative basins for large sample validation based on flow signatures.

Updated text (added in Sect. 4.3 Methodological limitations and future research directions):

Hydrological validation in this study combines limited in situ records with the GRFR discharge reanalysis. The ten gauged catchments used for process-based evaluation provide detailed insight into seasonal regimes and event responses, but they represent only a subset of the climate landscape classes. In contrast, the 722 headwater basins used for flow signature analysis provide broad spatial coverage, but they inherit uncertainty from the reanalysis and from the matching between river reaches and catchment polygons (Yang et al., 2021a). Although the coherent and physically plausible patterns in signatures and the Games Howell results suggest that these uncertainties do not dominate our conclusions, future work should test robustness using alternative discharge products and expanded gauging networks.

The framework is essentially static and assumes approximate stationarity over the analysis period. Many Chinese catchments have experienced substantial human intervention and climate change, which can alter hydrological signatures and weaken the correspondence between climatic forcing, landscape structure, and runoff response (Xu et al., 2024; Guo et al., 2020). Incorporating indicators of human influence, such as reservoir storage, irrigation extent, or urbanisation, would allow a clearer distinction between natural and managed regimes. Time varying classifications that track changes in catchment behaviour could also help identify emerging hydrological classes and diagnose regime shifts.

Q11: Discussion The discussion needs to emphasize the connection with the results. Currently, the discussion section seems to introduce existing knowledge within the basin. Perhaps discussing similarities and differences with similar studies, limitations, and potential applications would be more effective.

A: Thank you for this constructive recommendation. We agree that the Discussion should be more tightly anchored to the key findings of Sect. 3 and should focus on what we learn from the results in relation to existing classification studies, as well as the limitations and implications for applications.

In the revised manuscript, we substantially restructured Sect. 4 to strengthen the result driven narrative. First, we revised Sect. 4.1 to interpret the observed patterns in seasonal runoff regimes, event scale FDCs, and flow signatures in terms of climate versus landscape controls, directly reflecting the validation results. Second, we added a dedicated subsection on implications for modelling and prediction in ungauged basins (Sect. 4.2), which translates the two-level control identified in the Results into guidance for parameter regionalisation, donor basin selection, and data driven modelling strategies. Third, we expanded Sect. 4.3 to explicitly discuss methodological limitations, including uncertainties in the validation datasets, sensitivity to clustering choices, and the omission of explicit human influence and non stationarity, and we outlined priority directions for future work.

We also strengthened the comparison with related studies by relating our findings to large sample classification efforts in other regions and in China, highlighting both consistent conclusions (for example the dominant role of aridity and snow metrics) and differences arising from our hierarchical design that separates large scale climatic forcing from within region landscape modulation (e.g. Knoben et al., 2018; Kuentz et al., 2017; Jehn et al., 2020; Xu et al., 2024). Overall, these revisions ensure that the Discussion now synthesises the main results, situates them within the literature, and articulates limitations and practical relevance.

Revisions made in the manuscript:

Sect. 4 has been reorganised into three result oriented subsections (Sects. 4.1–4.3), covering interpretation of climate and landscape controls based on Sect. 3, implications for modelling and prediction in ungauged basins, and methodological limitations and future research directions.

Q12: Discussion The features used in this paper do not consider any human activities. How might this affect the results of catchment classification? Given the significant human activities in many regions of China, how should we interpret or use the classification results obtained without considering human activities?

A: Thank you for raising this important point. We agree that the omission of explicit human activity indicators is a key limitation for catchment classification in China, where regulation, irrigation, and urbanisation can substantially alter streamflow regimes.

Our classification is intentionally based on hydroclimatic indices and landscape descriptors that represent catchment form and long term forcing, because these variables are widely available for both gauged and ungauged basins and are suitable for similarity based regionalisation. As a result, the framework primarily characterises the "natural tendency" of catchment behaviour under given climate and landscape conditions. In basins with strong human modification, observed hydrological signatures may deviate from those expected from climate and landscape alone, which can weaken the correspondence between class membership and measured streamflow behaviour. This effect is most relevant where reservoir operations reshape seasonality and extremes, irrigation withdrawals reduce low flows, or urbanisation increases flashiness, potentially shifting signatures in directions not represented by the present descriptor set. In the revised Discussion, we therefore clarify how the classification should be interpreted and used. First, the typology provides a physically grounded baseline for organising catchments and for identifying where hydrological behaviour is likely to be climate and landscape controlled. Second, for strongly regulated basins, the classes should be interpreted as a reference condition rather than a direct predictor of current flow regime, and applications should incorporate additional information on human influence when selecting donor basins or transferring model parameters. Third, we highlight that the present framework can be extended by adding human disturbance indicators, such as reservoir storage, irrigation extent, and urban land fraction, either as additional descriptors within each climate region or as an overlay that separates near natural and managed regimes. Such extensions would allow a clearer distinction between natural and human modified hydrological types and improve applicability for water management contexts.

Revisions made in the manuscript:

We added an explicit limitation statement in Sect. 4.3 noting that the framework is essentially static and focuses on near natural basins, and that human influences are not represented explicitly. We also outlined how the classification can be interpreted as a baseline typology and how future work could incorporate human activity indicators and time varying classification to better represent managed regimes.

Updated text (added in Sect. 4.3 Methodological limitations and future research directions):

L583: Despite these strengths, several limitations of the proposed framework should be acknowledged. The classification depends on the choice and number of indices and descriptors. Although the selected set is supported by previous large sample studies and physical reasoning (Kuentz et al., 2017; Knoben et al., 2018; Addor et al., 2017), additional variables could reveal further dimensions of hydrological control. Examples include explicit indicators of subsurface permeability, glacier cover, land use change, and within catchment climatic variability (Xu et al., 2024; Jehn et al., 2020).

L598: The framework is essentially static and assumes approximate stationarity over the analysis period. Many Chinese catchments have experienced substantial human intervention and climate change, which can alter hydrological signatures and weaken the correspondence between climatic forcing, landscape structure, and runoff response (Xu et al., 2024; Guo et al., 2020). Incorporating indicators of human influence, such as reservoir storage, irrigation extent, or urbanisation, would allow a clearer distinction between natural and managed regimes. Time varying classifications that track changes in catchment behaviour could also help identify emerging hydrological classes and diagnose regime shifts.

Q13: L460-471 This part is not easy to understand. Especially, I didn't understand this sentence: L464"The flow regime in climate region II presented multiple peaks following multiple peaks in precipitation in June and July during the same period."

A: Thank you for pointing this out. We agree that the original wording was confusing because it repeated "multiple peaks" and did not clearly specify the timing relationship between rainfall and runoff. In the revised manuscript, we rewrote this sentence to state the intended meaning more explicitly: catchments in climate region II exhibit a pronounced monsoonal seasonal cycle, and monthly runoff shows one or two peaks that occur in response to the summer precipitation maxima (typically June to July).

Revision made in the manuscript:

The sentence has been rephrased in Sect. 3.3.1 L450 to: *In climate region II, monthly* *runoff displays a pronounced monsoonal pattern with one or two peaks following the summer rainfall maxima and substantial interannual variability in high-flow months (e.g. Fenshuijiang and Tunxi).*

Q14: L495-498 What do “combined indicators” refer to? What does “at different scales” mean? Basin area? Time?

A: Thank you for noting this ambiguity. We agree that the terms “combined indicators” and “at different scales” were not defined clearly in the original text. In the revised manuscript, we have removed the vague phrase “combined indicators” and clarified both what is being combined and what “different scales” refers to.

Specifically, “combined indicators” refers to the two complementary process-based diagnostics used in the validation, namely seasonal hydrographs and event-scale flow-duration curves (FDCs). We also clarify that “at different scales” refers to different temporal scales (seasonal versus event), rather than basin area or spatial scale. This is now stated explicitly in the Results section: the revised text explains that seasonal behaviour is primarily organised by climate regions, while event-scale response is modulated by landscape characteristics. The methodological framing of the validation across time scales (seasonal and event) is also clearly described in Sect. 2.4 .

Revision made in the manuscript (revised manuscript, around Lines 475):

Climate regions primarily control the seasonal timing and magnitude of runoff, whereas landscape characteristics modulate event-scale response through their influence on storage and routing.

Q15: L560-561 What does “There is no particular classification for one catchment that allows greater flexibility in the selection of a catchment for comparative studies or parameter transplantation in ungauged catchments” mean?

A: Thank you for pointing out that this sentence was unclear. Our intended meaning is that, under a fuzzy (soft) classification, a catchment is not forced into a single, rigid class membership. Instead, each catchment can belong to multiple classes with different membership degrees, which provides flexibility when selecting donor catchments for comparative analysis or parameter transfer in ungauged basins.

Revision made in the manuscript:

We rewrote this sentence to make the meaning explicit. The revised wording (Lines 622) is: *The classification provides a practical basis for regionalisation and prediction in* *ungauged basins. Climate regions can constrain parameters governing water balance and snow processes, whereas landscape classes within regions can inform parameters related to soil storage, groundwater connectivity, and routing. Fuzzy memberships quantify transitional behaviour and can be used to weight donor basins and express uncertainty across sharp hydroclimatic and geomorphological gradients.*

Q16: L556-557 The statement "Moreover, climate-homogeneous regions respond to hydrological behaviors at medium- or longtime scales, whereas catchment classification regulates hydrological processes at the flood event scale" needs to be strengthened in the results to support this conclusion.

A: Thank you for this helpful suggestion. In addition to strengthening the supporting evidence in the Results, we also expanded Sect. 4.1 to explicitly synthesise the scale dependent roles of climate regions and within region catchment classes and to anchor the interpretation in our validation outcomes.

In Sect. 4.1, we added four result linked arguments. First, we state that the six climate indices reproduce the major gradients in water and energy availability across China and therefore provide an interpretable backbone for the classification, consistent with previous large sample findings on aridity and snow controls (Addor et al., 2017; Knoben et al., 2018; Kuentz et al., 2017). Second, we interpret the within region SOM–FCM classes in terms of differences in relief, soils, vegetation, and drainage organisation, and we connect these directly to the observed event scale behaviour from the process based validation, contrasting buffered responses in low relief catchments with flashier responses in steep uplands. Third, we reinforce this interpretation using the large sample flow signature results, noting that magnitude and high flow frequency signatures (e.g. Qsp, Q95, HFD, HighFrVar) discriminate classes most strongly, whereas duration related metrics show weaker contrasts, which aligns with our conclusion that landscape modulation is most evident in event scale and high flow behaviour. Finally, we added an explicit synthesis paragraph stating that climatic and landscape controls operate on distinct but complementary time scales: seasonal hydrographs are organised mainly by climate region, whereas event based FDCs within a given climate region are strongly modulated by landscape properties (Sawicz et al., 2011).

Revisions made in the manuscript:

We expanded Sect. 4.1 to explicitly connect the Discussion to the Results by (i) interpreting the climate regions and within region classes using the validation evidence, (ii) summarising which flow signatures provide the strongest discrimination among classes, and (iii) stating clearly the time scale partitioning of controls (seasonal versus event scale).

4.1 Climate-landscape controls on hydrological behaviour

The proposed climate landscape classification reveals a strongly organised hydroclimatological structure across China. The six climate indices indicate that aridity, moisture seasonality, snow fraction, and the mean level and seasonality of temperature are sufficient to reproduce the major gradients in water and energy availability. This is consistent with previous large sample studies that identified aridity and snow related metrics as first order controls on long term streamflow patterns (Addor et al., 2017; Knoben et al., 2018; Kuentz et al., 2017). By combining moisture and temperature indices explicitly, our framework better differentiates cold, snow affected regimes on the Tibetan Plateau and in north eastern China from warm, humid monsoon regions and from persistently arid interiors. The resulting climate regions therefore provide an interpretable backbone for the subsequent landscape-based classification.

Within each climate region, the SOM-FCM analysis of geomorphological and drainage network descriptors identifies coherent catchment classes that reflect differences in relief, soil texture, vegetation cover, and drainage organisation. The process--based validation indicates that these classes are closely associated with event scale runoff response. Low relief catchments with finer soils and denser vegetation tend to exhibit buffered event hydrographs with sustained high flows and gentle recessions. In contrast, steep upland catchments with coarser soils show flashier behaviour, characterised by rapid drainage and limited baseflow contribution. These patterns are consistent with conceptual expectations from hydrologic landscape theory and with empirical analyses based on FDCs in other regions, where combinations of climate, relief, and soil storage jointly shape high and low flow regimes (Coopersmith et al., 2012; Ghotbi et al., 2020; Yaeger et al., 2012).

The large sample flow signature analysis reinforces this interpretation. Signatures related to flow magnitude, particularly specific discharge (Qsp) and high flow indices (Q95, HFD, HighFrVar), show the strongest discrimination among classes, whereas duration related metrics such as Mean30dMax exhibit more muted contrasts. This suggests that the proposed classification is especially effective at separating catchments according to overall runoff production and the frequency and intensity of *high flows. In comparison, prolonged high flow episodes appear to be less tightly constrained by the combined climate and landscape factors represented here. Similar results have been reported for continental scale classifications in Europe and North America, where signatures linked to mean flow and high flow variability carried most of the hydrological signal, while some low flow duration measures were comparatively redundant (Jehn et al., 2020; Kuentz et al., 2017; Xu et al., 2024).*

A further insight is that climatic and landscape controls operate on distinct but complementary time scales. Seasonal hydrographs primarily reflect the balance and timing of precipitation and potential evapotranspiration, and are therefore organised mainly by climate region with weaker sensitivity to local geomorphology. In contrast, event based FDCs within a given climate region are strongly modulated by landscape properties, translating similar meteorological forcing into markedly different runoff responses. This partitioning of control across time scales supports the view that hydrological similarity cannot be described by climate or catchment structure alone, and that explicit combinations of forcing and form are required to capture functional behaviour (Sawicz et al., 2011).

Responses to Comments from Reviewer 2:

Q1: 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.

A: Thank you for this important comment. We agree that the novelty should be stated more explicitly and grounded in hydrological outcomes rather than in the clustering technique itself. We therefore strengthened both the Introduction and Discussion to clarify (i) what is new in the scientific framing of similarity in China and (ii) what the proposed classes reveal about hydrological behaviour.

Clarifications and additions to novelty (Introduction). We added a concise overview of catchment classification and hydrological zoning efforts in China, from early indicator-based national schemes to recent large-sample, data-driven studies such as Xu et al. (2024), and then articulated the remaining gap and our contributio. In particular, we emphasised that many existing approaches, including recent data-driven work, commonly cluster climatic and landscape descriptors in a single feature space, which makes it difficult to separate large-scale climatic forcing from within-region landscape modulation . Our study addresses this gap by proposing a hierarchical national framework that first delineates climate regions along continuous hydroclimatic gradients and then derives landscape-based catchment classes within each climate region, with fuzzy memberships representing transitions.

Major differences and improvements in hydrological terms (Discussion). We revised Sect. 4.1 to link directly to the validation results and to make clear what is gained hydrologically. The strengthened discussion highlights that seasonal runoff regimes are primarily organised by climate regions, whereas event-scale response, diagnosed using event-based FDCs and high-flow-related signatures, is strongly modulated by landscape classes within the same climatic setting. We also explicitly report which signatures provide the strongest class discrimination, namely specific discharge and several high-flow indices (Qsp, Q95, HFD, HighFrVar), while some duration-related metrics show weaker contrast. This directly addresses the requested "differences or improvements" in terms of hydrologic signatures rather than methodology.

Positioning relative to Xu et al. (2024) and prior schemes. We explicitly recognise Xu et al. (2024) as an important benchmark for large-sample hydrology in China and use it to motivate the need for a framework that remains directly applicable to ungauged basins. Unlike signature-only classifications that require discharge information, our classes are derived from widely available climate indices and catchment descriptors, and the within-region landscape classification is performed using dimension-reduced descriptor information, with fuzzy memberships retained for transitional behaviour . This design supports regionalisation and donor-basin selection while maintaining a clear hydrological interpretation across time scales.

Revisions made in the manuscript:

Introduction (revised manuscript, Lines 63–77): Added China-specific hydrological zoning and catchment classification context (Luo, 1954; Xiong and Zhang, 1995; Liu et al., 2014; Xu et al., 2024), and clarified the remaining gap and our contribution :

China is an ideal study area for developing an integrated and scale aware classification framework because it spans strong regional contrasts, from cold and dry plateaus to warm and humid plains, and includes diverse terrains and climates. Early hydrological zoning in China relied on a limited set of indicators. Luo (1954) proposed one of the earliest national schemes based on basin boundaries, flow patterns, and sediment *characteristics. Yi and Jiazhen (1995) delineated 11 regions using mean annual runoff depth as a primary indicator, and Liu et al. (2014) divided China into three broad regions based on topography and climate patterns. These foundational studies provided valuable national perspectives, but they were not designed to represent catchment scale hydrological behaviour in a multidimensional sense, which is important for similarity-based model transfer. More recently, improved datasets have enabled more data driven classification efforts. For example, Xu et al. (2024) classified Chinese catchments using flow signature information and analysed the associated controls, highlighting the joint role of climate and soil properties in distinguishing basin groups. Such progress is highly valuable for advanceing large sample hydrology in China. However, their clustering strategy integrates climatic and landscape descriptors within a single feature space and therefore does not explicitly separate the effects of large-scale climatic forcing from within climate landscape modulation. As a result, a hierarchical national framework that systematically combines continuous hydroclimatic gradients with fuzzy landscape similarity, and remains directly applicable to ungauged basin prediction, still needs to be established.*

Discussion (Sect. 4.1): Strengthened the result-based interpretation, including the time-scale separation of controls and the specific flow signatures that most clearly discriminate classes:

4.1 Climate-landscape controls on hydrological behaviour

The proposed climate landscape classification reveals a strongly organised hydroclimatological structure across China. The six climate indices indicate that aridity, moisture seasonality, snow fraction, and the mean level and seasonality of temperature are sufficient to reproduce the major gradients in water and energy availability. This is consistent with previous large sample studies that identified aridity and snow related metrics as first order controls on long term streamflow patterns (Addor et al., 2017; Knoben et al., 2018; Kuentz et al., 2017). By combining moisture and temperature indices explicitly, our framework better differentiates cold, snow affected regimes on the Tibetan Plateau and in north eastern China from warm, humid monsoon regions and from persistently arid interiors. The resulting climate regions therefore provide an interpretable backbone for the subsequent landscape-based classification.

Within each climate region, the SOM-FCM analysis of geomorphological and drainage network descriptors identifies coherent catchment classes that reflect differences in relief, soil texture, vegetation cover, and drainage organisation. The process--based *validation indicates that these classes are closely associated with event scale runoff response. Low relief catchments with finer soils and denser vegetation tend to exhibit buffered event hydrographs with sustained high flows and gentle recessions. In contrast, steep upland catchments with coarser soils show flashier behaviour, characterised by rapid drainage and limited baseflow contribution. These patterns are consistent with conceptual expectations from hydrologic landscape theory and with empirical analyses based on FDCs in other regions, where combinations of climate, relief, and soil storage jointly shape high and low flow regimes (Coopersmith et al., 2012; Ghotbi et al., 2020; Yaeger et al., 2012).*

The large sample flow signature analysis reinforces this interpretation. Signatures related to flow magnitude, particularly specific discharge (Qsp) and high flow indices (Q95, HFD, HighFrVar), show the strongest discrimination among classes, whereas duration related metrics such as Mean30dMax exhibit more muted contrasts. This suggests that the proposed classification is especially effective at separating catchments according to overall runoff production and the frequency and intensity of high flows. In comparison, prolonged high flow episodes appear to be less tightly constrained by the combined climate and landscape factors represented here. Similar results have been reported for continental scale classifications in Europe and North America, where signatures linked to mean flow and high flow variability carried most of the hydrological signal, while some low flow duration measures were comparatively redundant (Jehn et al., 2020; Kuentz et al., 2017; Xu et al., 2024).

A further insight is that climatic and landscape controls operate on distinct but complementary time scales. Seasonal hydrographs primarily reflect the balance and timing of precipitation and potential evapotranspiration, and are therefore organised mainly by climate region with weaker sensitivity to local geomorphology. In contrast, event based FDCs within a given climate region are strongly modulated by landscape properties, translating similar meteorological forcing into markedly different runoff responses. This partitioning of control across time scales supports the view that hydrological similarity cannot be described by climate or catchment structure alone, and that explicit combinations of forcing and form are required to capture functional behaviour (Sawicz et al., 2011).

Q2: 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.

A: Thank you for this constructive recommendation. We agree that the earlier version mixed methodological detail and interpretation into the Results. In the revised manuscript, we have reorganised the structure to more cleanly separate Methodology, Results, and Discussion.

Methodological details on how SOM and FCM are implemented and how the number of clusters is selected using internal validity indices are now presented in the Methodology (Sect. 2.3), including the use of the Davies–Bouldin index and silhouette coefficient for determining cluster numbers. This content was previously described in the Results in the original submission (for example, the statement that six clusters were selected by jointly considering DBI and SC).

We also moved and consolidated methodological descriptions for the validation analyses into Sect. 2.4, including how seasonal runoff and event scale FDCs are used, how flow signatures are defined, and how the Games–Howell test is applied for large sample statistical validatio. The Results section now focuses on reporting the classification outcomes and the validation evidence (figures and quantitative discrimination patterns). Interpretations, comparisons with related studies, limitations, and applications are consolidated in the Discussion. For example, the scale dependent interpretation that seasonal regimes are organised mainly by climate region while event response is modulated by landscape class is synthesised in Sect. 4.1 , and the modelling implications are developed in Sect 4.2.

Revisions made in the manuscript:

Reorganised the manuscript to ensure Methods describe the workflow, clustering logic, and cluster selection criteria (Sect. 2.3), and Results focus on reporting outcomes and validation evidence (Sect. 3.3).

Consolidated validation methodology (seasonal runoff, event based FDCs, flow signatures, Games–Howell test) in Sect. 2.4, while relocating broader interpretation and implications to the Discussion (Sects. 4.1-4.2).

Q3: 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.

A: Thank you for this important comment. We agree that using only ten gauged catchments cannot fully represent the national diversity of classes, and that spatial distance within the same class is a key aspect of hydrological similarity. In the revised manuscript, we clarified the role of the ten gauged catchments and strengthened the validation logic to address this concern.

First, the ten catchments are used for process-based validation because they provide event-scale hydrographs needed to construct event-based FDCs. Their selection was constrained by the limited availability of long, quality-controlled discharge observations for small and medium headwater basins across China. Consequently, we did not further expand the number of gauged catchments for event-scale diagnostics. Instead, we designed the validation as a two-stage framework that combines (i) process-based evidence from the ten gauged catchments and (ii) a complementary large-sample statistical validation using discharge reanalysis to provide broader spatial coverage.

Second, we clarified that the gauged catchments are not intended to be a geographically dense local sample. Table 5 now reports river systems, record lengths, flood-event counts, and class labels, showing that the gauged basins span multiple climate regions and major river systems. Importantly, we also highlighted an explicit spatially non-local within-class comparison. For example, Daiying (Hai River) and Maduwang (Yellow River) are both classified as IV-6, providing a same-class pair from different river systems, and their event-based FDCs exhibit comparable response shapes. This directly demonstrates that event-scale similarity can hold for catchments that are geographically distant but belong to the same climate–landscape class.

Third, to address the broader question of similarity across distant basins nationally, we relied on the second validation stage. We used 722 headwater basins matched to the GRFR discharge reanalysis and computed 13 flow signatures, followed by Games–Howell testing to quantify between-class differences. This large-sample validation covers a much wider spatial extent than the ten gauged catchments and provides quantitative support that the derived classes are hydrologically distinct across China.

Revisions noted in the manuscript (revised manuscript):

Expanded the description of validation data availability and the rationale for the two-stage validation design (Sect. 2.4):

2.4 Validation and Analysis of Catchment Classification.

The hydrological validity of the proposed climate-landscape classification was assessed using a two-stage framework based on in situ gauge records and a global discharge reanalysis product. The framework combines (i) process-based validation using seasonal runoff and event scale flow duration curves (FDCs) and (ii) statistical validation based on flow signatures of the classified catchments.

For the process-based validation, daily runoff records from ten gauged catchments were used to assess hydrological similarity across time scales. At the seasonal scale, mean monthly runoff and its interannual variability were computed for each catchment and compared within and across climate regions to test whether catchments assigned to the same region exhibit consistent seasonal flow regimes, following previous work that uses seasonal runoff characteristics to analyse climate controls on streamflow behaviour and catchment similarity (Kuentz et al., 2017; Berghuijs et al., 2014). At the event scale, high frequency discharge data for identified flood events were used to construct event-based FDCs, which provide an integrated diagnostic of rainfall-runoff response and of the relative contributions of fast surface runoff and slower subsurface or baseflow components (Kuentz et al., 2017). Comparisons of seasonal hydrographs and event-based FDC shapes across climate regions and catchment classes were then used to evaluate whether the classification delineates groups with similar runoff dynamics and hydrological signatures.

To extend the analysis beyond this small set of gauged basins, a large-sample statistical assessment was conducted using flow signatures (FS) derived from daily discharge time series. We adopted 13 FS following Kuentz et al. (2017) and Xu et al. (2024). These signatures (Table 2) summarise key aspects of hydrological behaviour, including flow magnitude, the frequency and duration of high and low flows, and the rate of change in discharge. The FS were computed from daily discharge records for all basins with sufficiently long and consistent time series.

Differences in hydrological behaviour between catchment classes were evaluated using the Games-Howell test (Games and Howell, 1976). For each FS, pairwise comparisons *were performed between all classes under the null hypothesis of no difference in mean FS between class pairs. The Games-Howell procedure does not assume equal variances or equal sample sizes among groups and is therefore well suited to the present application. Statistically significant differences in FS between classes were interpreted as evidence that the classification separates catchments with distinct hydrological regimes, providing a quantitative, large-sample validation of the climate-landscape-based grouping. Results of the process-based and large-sample statistical validations are presented in Sect. 3.3.*

Added explicit text and examples to show that within-class similarity is not limited to spatially close basins (e.g., Daiying and Maduwang, both IV-6, in different river systems)(Sect. 3.3).

Strengthened and clearly positioned the large-sample validation using 722 GRFR-matched basins and signature-based statistics as national-scale support for class robustness (Sect. 2.4 and Sect. 3.3).

Q4: 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.

A: Thank you for this constructive comment. We agree that the practical value of a national similarity framework should be discussed more explicitly. In the revised manuscript, we strengthened Sect. 4.2 (Implications for modelling and prediction in ungauged basins) to clarify how the proposed classification can be used for parameter transfer, regionalisation, and donor catchment selection. We now emphasise that climate regions provide a robust scaffold for seasonal water balance behaviour, while catchment classes within each climate region primarily modulate event scale response, implying that model structures and parameters should be conditioned first on climatic regime and then refined using catchment class information . We further specify that climate regions can constrain parameters related to water balance and snow processes, whereas catchment classes can inform parameters linked to soil storage, groundwater connectivity, and routing , and we explicitly note that the framework supports a process informed basis for selecting donor catchments beyond simple geographic proximity.

In addition, we clarified in Sect. 2.4 that event based flow duration curves provide an integrated diagnostic of rainfall runoff response and the relative contributions of fast and slow components, which supports the interpretation of class level similarity in event scale behaviour. While this paper does not perform design storm based flood estimation explicitly, the strengthened discussion explains how the class structure and demonstrated event scale similarity can support donor basin selection and parameter regionalisation for flood oriented modelling applications.

Revisions made in the manuscript (revised manuscript):

Expanded application oriented discussion of parameter transfer, regionalisation, and donor catchment selection in Sect. 4.2:

The proposed climate landscape classification has direct implications for hydrological modelling and prediction in ungauged basins. A central goal of the Predictions in Ungauged Basins (PUB) initiative is to develop strategies for parameter transfer and regionalisation that reflect the dominant physical controls on runoff generation (Hrachowitz et al., 2013). Our results indicate that climate regions provide a robust framework for seasonal water balance behaviour, whereas landscape based classes within those regions primarily modulate event scale response. This suggests that model structures and parameter sets should first be conditioned on climatic regime and then refined using catchment class information, rather than relying on similarity measures based only on climate or only on catchment attributes.

In practice, climate regions can be used to define prior ranges for parameters controlling water balance and snow processes, while catchment classes inform parameters related to soil storage, groundwater connectivity, and routing. This hierarchical strategy is consistent with recommendations that regionalisation should account for both climatic and physiographic similarity, and that multi step approaches often outperform simple nearest neighbour transfer based on geographic proximity (Qi et al., 2022; Gupta et al., 2014; Guo et al., 2021). Our finding that hydrological behaviour can be more similar between distant basins within the same climate landscape class than between neighbouring basins in different classes further highlights the limitations of purely spatial regionalisation (Prieto et al., 2019) and supports a process informed basis for selecting donor catchments.

The classification is also relevant for data driven prediction. Recent studies show that machine learning models benefit from conditioning on catchment descriptors and *hydrological signatures, either through explicit grouping or through feature based regionalisation (Rasheed et al., 2022; He et al., 2024). In such approaches, the SOM-FCM classes can be used as categorical inputs that guide model design, training objectives, or ensemble configuration. For example, separate models can be trained for individual climate regions or catchment classes, or a single model can include class specific parameters that enable information sharing among similar basins while limiting adverse pooling across fundamentally different regimes. Compared with purely data driven clustering, the present framework has the advantage that classes are defined using physically interpretable indices and descriptors. This supports clearer diagnosis of model performance and facilitates the incorporation of expert knowledge.*

Q5: I think the references are not in the required style of HESS (https://www.hydrology-and-earth-system-sciences.net/submission.html#references)

A: Thank you for your feedback. We have reformatted the in-text citations and the reference list to comply with the HESS reference style guidelines.

Q6: Line 69: "an indisputable fact" looks like a strange statement. Do you mean machine learning is now widely used for regionalization studies?

A: Thank you for pointing this out. We agree that "an indisputable fact" is inappropriate in academic writing. We revised the sentence to a neutral and precise statement (revised manuscript, Line 78):

"Machine learning offers practical tools for constructing similarity frameworks from high dimensional descriptor datasets (Yang et al., 2020)."

Q7: Line 102-103: need to add references.

A: Thank you for this comment. We have added an appropriate citation to support this statement. In the revised manuscript (Line 154), it now reads:

"Climate directly affects runoff generation at the event scale and indirectly shapes the hydrological cycle through its influence on long-term soil moisture storage and the co-evolution of landscape and vegetation (Jehn et al., 2020)."

Q8: 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).

A: Thank you for this helpful suggestion. We agree that the original paragraph could confuse readers because it discussed "three indices" before explicitly stating what they were. In the revised manuscript, we improved the logical flow by first listing the six selected climate indices and then presenting the reasoning for their selection.

Reordered the paragraph in Sect. 2.2.1 (Climate indices):

In this study, six climate indices were selected to represent moisture availability, thermal conditions, and snow influence that are relevant to catchment hydrological response: the average moisture index (I_m), the seasonal moisture index ($I_{m,r}$), the fraction of precipitation falling as snow (fs), the annual average temperature (T_m), the seasonal temperature range ($T_{m,r}$), and the fraction of snowy days (Ds). The first three indices are derived from a modified version of Thornthwaite's moisture index MI (Willmott and Feddema, 1992) and describe the availability and seasonality of climatic water, as well as the partitioning between rainfall and snowfall. The three temperature related indices summarise the mean thermal regime, its seasonal variability, and the occurrence of conditions conducive to snowfall and snow storage. Together, these six indices provide a compact description of the dominant climatic controls on water and energy availability at the catchment scale.

Q9: Line 174: reference for FCM?

A: Thank you for your feedback.We have added the appropriate reference when introducing FCM in the revised manuscript (Line 225):

"FCM is a soft clustering algorithm based on fuzzy set theory (Pal et al., 2005)."

Q10: Line 178: "may be the most" to "is a".

A: Thank you for this suggestion. We agree that "may be the most suitable" is too strong and speculative. In addition to revising this sentence, we reviewed the manuscript to avoid overly strong or absolute wording and replaced such expressions with more neutral, academically appropriate phrasing where needed.

Q11: Line 207: reference for Penman-Monteith equation.

A: Thank you for your feedback. We added the appropriate reference in the revised manuscript (Line 115) when describing the Penman–Monteith formulation used in CRU TS v4.04:

"EP was taken from CRU TS v4.04 at a 0.5° × 0.5° resolution produced by the Centre for Environmental Data Analysis (https://www.ceda.ac.uk/), where EP is estimated using a Penman-Monteith type formulation (Moratiel et al., 2020)."

Q12: Line 214: What are the average size and the range of catchment sizes?

A: Thank you for this comment. For the 13 487 HydroBASINS catchments used in this study, the mean catchment area is 761.04 km^2, and the area ranges from 15 km^2 to 14 612.8 km^2.

Revisions made in the manuscript (revised manuscript):

We added these catchment size statistics to Table 1 (Topographic characteristics), where catchment area and its summary statistics are reported as part of the descriptors used for geomorphological and network-related characterisation.

Q13: 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

A: Thank you for this constructive comment. We agree that the physical meaning of the SOM visualisations and their link to geographic space should be explained more explicitly before the figure is introduced. In the revised manuscript, the former Fig. 2 is now presented as Fig. 3, and we revised both the Methodology and the figure caption to clarify interpretation.

First, we clarified that the SOM plots represent a two dimensional feature space rather than geographic space. In the updated caption, we explain that each hexagon is a SOM neuron that summarises grid cells with similar standardised climate index values, and the colour represents the corresponding codebook (weight) value for each index (Fig. 3a). We also clarify that the black contour lines delineate the six SOM–FCM climate clusters on the SOM lattice. For the d-matrix (Fig. 3b), we now state that it shows the Euclidean distance between neighbouring neurons, where larger distances indicate sharper transitions in the underlying hydroclimatic feature space and typically coincide with cluster boundaries, while smaller distances indicate locally homogeneous areas.

Second, to support spatial interpretation, we explicitly point readers to the geographic projection of the fuzzy climate partition in Fig. 4, which maps membership back to

China and provides the intended spatial context. Following the revision, only grid cells with membership values greater than 0.5 are shown, so the mapped patterns emphasise higher confidence assignments.

Finally, as suggested, we moved the explanatory text (previously located around Lines 276–285 in the original manuscript) into the Methodology, and we now introduce component planes and d-matrices as diagnostic tools within Sect. 2.3.1, including the citation to Vesanto (1999)

Revisions made in the manuscript (revised manuscript):

Added and expanded the methodological explanation of SOM component planes and d-matrices in Sect. 2.3.1:

In this study, SOM and FCM are combined as follows. A SOM is first trained on the standardised input variables. FCM is then applied to the SOM codebook vectors (neuron weight vectors), rather than directly to all samples, to obtain a fuzzy partition in a reduced and smoothed feature space. The resulting cluster membership of each neuron is assigned to all samples mapped to that neuron. Prior to all SOM-FCM analyses, climate indices and catchment descriptors are standardised to zero mean and unit variance to avoid scale effects. SOM component planes and distance matrices (d-matrices) are used diagnostically to examine variable relationships and to verify that the identified clusters are consistent with the underlying feature space (Vesanto, 1999).

Revised the caption of Fig. 3 (formerly Fig. 2) to explain what hexagons and colours represent, and how the d-matrix should be interpreted

Revised the caption of Fig. 3 (formerly Fig. 2) to explain what the hexagons and colours represent and how the d-matrix should be interpreted. We also added the climate-region labels (I–VI) on the SOM map and the d-matrix to facilitate interpretation and to help readers relate distinct areas of the SOM feature space to the corresponding climate regions used in subsequent analyses.

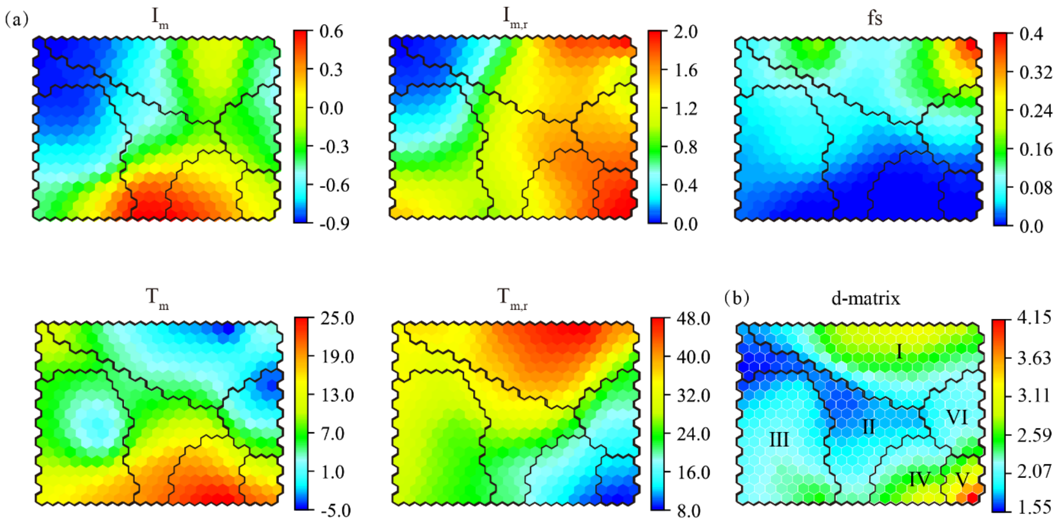

Figure 3. Self-organising map representation of the hydroclimatic feature space and fuzzy climate-region partition: (a) SOM component planes for the five climate indices: each hexagon represents a neuron summarising grid cells with similar standardised index values, colours show the corresponding codebook (weight) value, and black contour lines delineate the six SOM-FCM climate clusters that are later mapped to geographic space. (b) SOM distance matrix (d-matrix) giving the Euclidean distance between neighbouring neurons; cooler colours indicate smoothly varying hydroclimatic conditions, warmer colours mark sharper transitions in the climate-index space, and labels I-VI identify the six climate regions used in subsequent spatial analyses.

Q14: Line 304: What is the AP algorithm? I don't think this was mentioned in Methods, and should add the reference

A: Thank you for this comment. AP refers to the Affinity Propagation clustering algorithm. In the revised manuscript, we simplified the cluster-number selection procedure and removed the use of AP, so AP is no longer used or discussed and the abbreviation does not appear in the updated text. Instead, as described in Sect. 2.3, the number of clusters is determined by jointly considering the Davies–Bouldin index (DBI) and the silhouette coefficient (SC), together with hydrological interpretability. This streamlining avoids introducing an additional algorithm that is not essential for the proposed SOM–FCM framework and improves the readability and consistency between Methods and Results.

Revisions made in the manuscript:

We deleted the AP-related description and revised the text to describe cluster-number selection based on DBI and SC within Sect. 2.3.

Q15: 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.

A: Thank you for the detailed comment. We addressed both points in the revised manuscript.

Retention of the second soil and vegetation PC (eigenvalue = 0.91). We clarified that principal components were retained to achieve an adequate cumulative explained variance and to preserve hydrologically meaningful, largely independent information within each descriptor group. In particular, we retained the first two components for each group because they yield cumulative explained variances above 70%. For the soil and vegetation group, we explicitly justified keeping PC2 despite its eigenvalue being slightly below unity (0.91), because it raises the cumulative explained variance above 70% and captures an interpretable contrast linked to vegetation cover (NDVI) beyond soil texture.

Improved writing and clearer structure (removal of semicolons and explicit grouping). We rewrote the paragraph to clearly separate the three descriptor groups and describe the interpretation of PC1 and PC2 for each group in a consistent "For ⋯, ⋯" format.

Revisions made in the manuscript (revised manuscript):

Added a clear retention criterion and group-wise interpretation of PCs:

For each group, the first two principal components were retained, yielding cumulative explained variances above 70 %. For topographic descriptors, PC1 primarily reflects overall elevation and slope, whereas PC2 is most strongly associated with the hypsometric gradient AS. For soil and vegetation descriptors, PC1 is dominated by soil texture, especially sand fraction, while PC2 captures independent variation in NDVI. For topological descriptors, PC1 is mainly related to catchment shape metrics (Re and Rf), and PC2 is strongly associated with drainage density. Although the second soil-vegetation component has an eigenvalue slightly below unity (0.91), it was retained because it increases the cumulative explained variance to above 70 % and represents a hydrologically meaningful contrast between vegetation cover and soil texture. These *six components thus provide a compact yet interpretable representation of landscape variability for subsequent catchment classification within each climate region.*

Q16: Line 449: space between class and (Li et al).

A: Thank you for pointing out the formatting issue. We corrected the missing space and checked all in-text citations throughout the manuscript to ensure consistent formatting.

Q17: 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.

A: Thank you for these helpful suggestions. We implemented all three points in the revised manuscript. Due to figure renumbering, this figure is now presented as Fig. 8. Catchment classification shown in each panel. We added the class code to every subplot title (e.g. Suide (IV-4)), where the Roman numeral denotes the climate region and the Arabic numeral denotes the catchment class within that region. Improved colour scheme for variability ranges. We replaced the previous grey shading with clearly distinguishable blue percentile bands to better convey interannual variability. Ranges explained in caption/legend. We updated the caption and the legend to specify that the shaded envelopes represent the central 50 %, 75 %, and 90 % ranges of monthly runoff, while the black line shows median monthly runoff and the red line shows monthly precipitation.

Revisions made in the manuscript: Updated Fig. 8 and its caption accordingly.

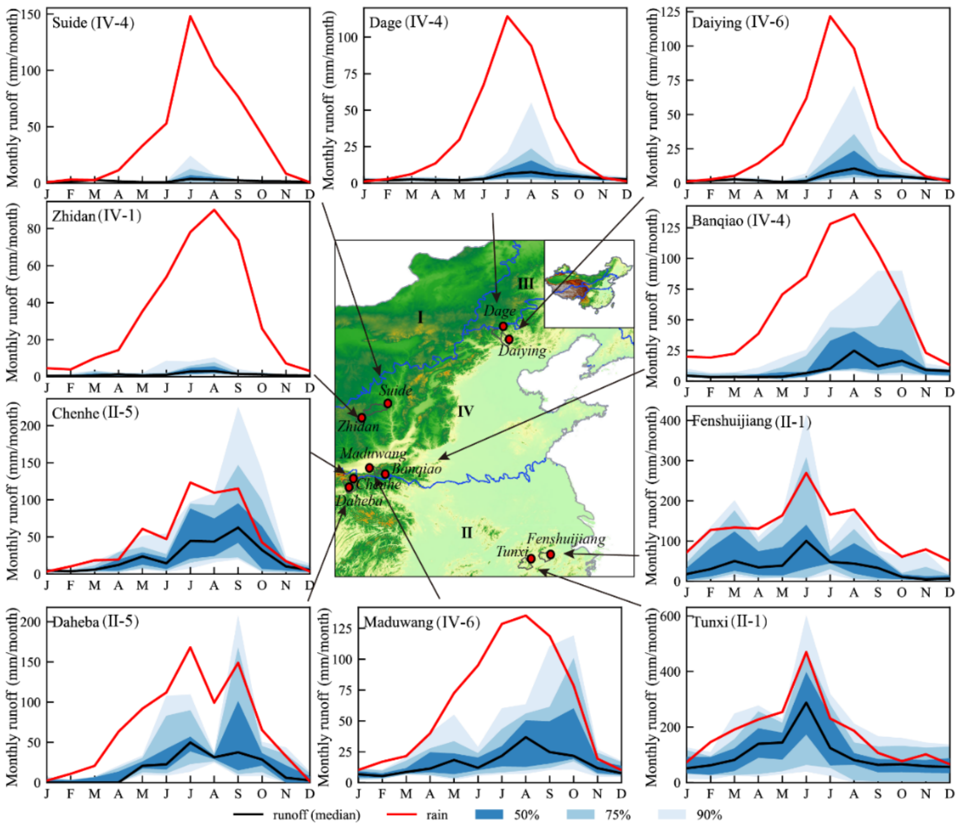

Figure 8. Seasonal runoff of the ten gauged catchments used for process-based validation. Monthly precipitation (red) and median monthly runoff (black) are shown, with shaded bands indicating the central 50 %, 75 %, and 90 % ranges of monthly runoff across years.

Q18: Line 495: correct the citation

A: Thank you for pointing out the citation issue. We apologise for the oversight and have corrected the citation in the revised manuscript:

The results also highlight that hydrological similarity does not necessarily coincide with spatial proximity: catchments that are geographically distant but belong to the same climate-landscape class (e.g. Maduwang and Daiying) can exhibit comparable flow regimes, whereas neighbouring basins in different classes may behave quite differently.

Q19: 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.

A: Thank you for this constructive suggestion. We agree that normalising discharge by drainage area improves the comparability of flow-duration curves across catchments with different sizes. In the revised manuscript, the event scale FDCs are therefore presented as runoff depth (mm d⁻¹), obtained by converting discharge to a drainage-area-normalised metric. This is shown directly by the y-axis unit in Fig. 9 (formerly Fig. 8). Presenting FDCs in mm d⁻¹ preserves the curve shape while reducing the influence of catchment area on magnitude, and it is consistent with our signature-based validation where magnitude-related metrics are also expressed as specific discharge (e.g. Qsp).

Revisions made in the manuscript (revised manuscript):

Fig. 9 (formerly Fig. 8) now reports event scale FDCs in runoff units (mm d⁻¹), and the associated description is aligned with the signature-based validation framework.

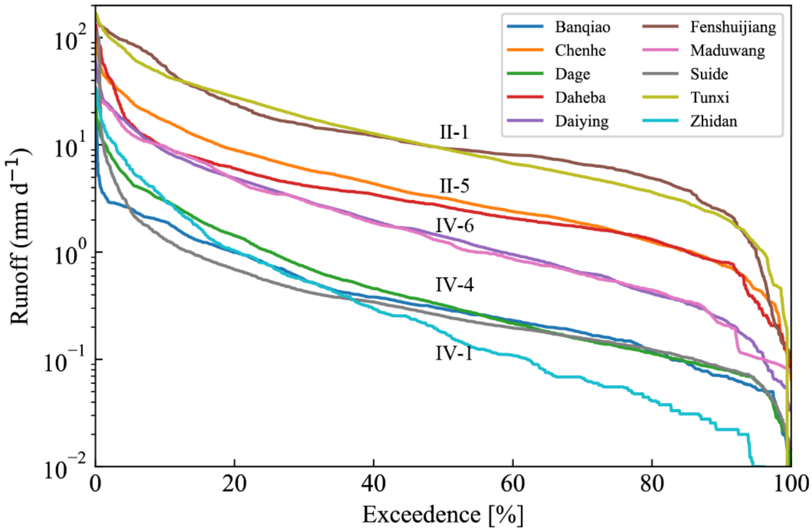

Figure 9. Event-scale flow-duration curves of the ten gauged catchments.

Q20: 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.

A: Thank you for this comment. We agree that the earlier wording was too general and not sufficiently supported by evidence. During revision, we restructured the Discussion and removed this paragraph (formerly in Sect. 4.1 on "boundary effects") to avoid making claims that are not directly demonstrated by our results. The revised Discussion now focuses on (i) result-based climate and landscape controls on hydrological behaviour (Sect. 4.1), (ii) implications for modelling and prediction in ungauged basins (Sect. 4.2), and (iii) limitations and future research directions (Sect. 4.3), without statement about "errors" caused by artificial boundaries.

Q21: Line 546: Is this a possible reason causing the challenge that some watersheds are hard to be classified with one dominant group?

A: Thank you for this question. Yes, human interference can be a plausible contributing factor. Anthropogenic influences such as reservoir regulation, abstractions, and land-use change can cause hydrological behaviour to deviate from that expected based on natural climate and landscape controls. As a result, catchments that appear similar in climate and topography may exhibit different runoff behaviour and can show lower maximum fuzzy memberships, indicating mixed similarity rather than a single dominant type. At the same time, even in near-natural conditions, continuous environmental gradients and complex physiographic transitions can also yield diffuse boundaries and mixed memberships.

Responses to Comments from Reviewer 3:

Q1: The Results section should focus on reporting experimental outcomes and data interpretation only. Methodological explanations, such as the description of the Affinity Propagation (AP) algorithm (Line 304), should be moved to the Methods section for better structural consistency.

A: Thank you for this suggestion. We agree that methodological descriptions should not appear in the Results section. In the revised manuscript, we streamlined the presentation and removed the AP-related methodological explanation from the Results, with all clustering procedures and criteria now described in the Methods section (Sect. 2.3).

Q2: Please justify the representativeness of these catchments or consider expanding the validation dataset to reinforce the robustness of your conclusions.

A: Thank you for this important comment. In the revised manuscript, we strengthened the justification of the ten gauged catchments and reinforced the validation with an additional large-sample dataset.

First, we clarified that the ten catchments were selected to be (i) near natural headwater basins with 10 to 15 years of continuous daily rainfall and runoff records and 10 to 35 documented flood events, (ii) without major upstream regulation or abstractions, and (iii) representative of a broad range of climate regions and basin classes. We also provided their key information (river system, drainage area, record length, and assigned basin class) in Table 5, showing that they span multiple river systems and cover a wide range of areas (441 to 4321 km²) and classes (II-1, II-5, IV-1, IV-4, IV-6).

Second, because discharge observations with consistent national coverage are limited, we complemented the process-based evaluation with a large-sample statistical validation using 722 headwater basins matched to GRFR, which provides broader spatial coverage and tests class distinctiveness using 13 flow signatures and Games–Howell statistics .

Revisions made in the manuscript (revised manuscript):

expanded validation-data description in Sect. 2.1:

Because discharge observations with consistent coverage across China are limited, we evaluated the hydrological validity of the classification using both gauge records and a global discharge reanalysis product. Daily runoff data for ten gauged catchments were collected from national hydrological yearbooks and used for process-oriented evaluation. The selected catchments satisfied three criteria: (i) 10-15 years of continuous daily rainfall and runoff records with 10 -35 documented flood events, (ii) no major upstream regulation or abstractions, and (iii) representation of a broad range of climate regions and catchment classes (Li et al., 2018). In addition, flow signatures were derived from the Global Reach Level Flood Reanalysis (GRFR) dataset (Yang et al., 2021b), which provides 3 hourly discharge time series for river reaches worldwide from 1980-2019 and has been evaluated against daily discharge records from more than 14 000 gauging stations. HydroBASINS catchments were matched to GRFR river reaches based on upstream drainage area. Headwater basins with a relative area mismatch below 10 % and without documented upstream modifications were retained, yielding 722 representative basins for large sample validation based on flow signatures.

clarified the two-stage validation logic in Sect. 2.4 :

The hydrological validity of the proposed climate-landscape classification was assessed using a two-stage framework based on in situ gauge records and a global discharge reanalysis product. The framework combines (i) process-based validation using seasonal runoff and event scale flow duration curves (FDCs) and (ii) statistical validation based on flow signatures of the classified catchments…

Q3: Please include a paragraph discussing the limitations of the proposed method and outline possible future research directions to improve applicability and generalization.

A: Thank you for this suggestion. We agree that it is important to explicitly discuss limitations and future directions. In the revised manuscript, we added a dedicated subsection Sect. 4.3 (Methodological limitations and future research directions). This section summarises key limitations related to (i) the dependence on the selected indices and descriptors, (ii) the reliance on a combination of limited in situ records and GRFR discharge reanalysis for validation, and (iii) the static nature of the framework and the lack of explicit representation of human influences. We also outline future work directions, including extending the descriptor set (e.g. human disturbance indicators), testing robustness using alternative discharge products, exploring dynamic classifications that track changes through time, and coupling the typology with process-based and data-driven models to evaluate class-based regionalisation strategies in predictive settings.

Revisions made in the manuscript (revised manuscript, Sect. 4.3):

Despite these strengths, several limitations of the proposed framework should be acknowledged. The classification depends on the choice and number of indices and descriptors. Although the selected set is supported by previous large sample studies and physical reasoning (Kuentz et al., 2017; Knoben et al., 2018; Addor et al., 2017), additional variables could reveal further dimensions of hydrological control. Examples include explicit indicators of subsurface permeability, glacier cover, land use change, and within catchment climatic variability (Xu et al., 2024; Jehn et al., 2020). The SOM-FCM approach also requires decisions on map size, fuzziness, and the number of clusters. While QE, TE, DBI, and SC provide objective guidance, the final solution inevitably involves judgement when balancing simplicity, stability, and interpretability. Alternative clustering strategies, including model based and graph based methods, may yield different partitions and warrant further evaluation.

Hydrological validation in this study combines limited in situ records with the GRFR discharge reanalysis. The ten gauged catchments used for process based evaluation provide detailed insight into seasonal regimes and event responses, but they represent only a subset of the climate landscape classes. In contrast, the 722 headwater basins used for flow signature analysis provide broad spatial coverage, but they inherit uncertainty from the reanalysis and from the matching between river reaches and catchment polygons (Yang et al., 2021a). Although the coherent and physically plausible patterns in signatures and the Games Howell results suggest that these uncertainties do not dominate our conclusions, future work should test robustness using *alternative discharge products and expanded gauging networks.*

The framework is essentially static and assumes approximate stationarity over the analysis period. Many Chinese catchments have experienced substantial human intervention and climate change, which can alter hydrological signatures and weaken the correspondence between climatic forcing, landscape structure, and runoff response (Xu et al., 2024; Guo et al., 2020). Incorporating indicators of human influence, such as reservoir storage, irrigation extent, or urbanisation, would allow a clearer distinction between natural and managed regimes. Time varying classifications that track changes in catchment behaviour could also help identify emerging hydrological classes and diagnose regime shifts.

Q4: I think you need check the English grammar and sentence carefully. Please revise the English description of your manuscript.

A: Thank you for your feedback. We have thoroughly revised and polished the English throughout the entire manuscript to improve grammar, clarity, and readability, and to ensure consistency with HESS writing style.

Q5: Please carefully review and standardize all references to conform with the HESS reference style guide.

A: Thank you for your feedback. We have carefully reviewed and standardised all references and in-text citations to conform to the HESS reference style guidelines.

Q6: Line 276: When describing the SOM neuron grid (19×22), briefly justify why this specific grid size was selected (e.g., based on data dimensionality, heuristic optimization, or quantization error minimization).

A: Thank you for this suggestion. In the revised manuscript, we now provide an explicit methodological justification for the SOM neuron grid size. We clarify that the SOM map dimensions were selected by testing candidate grids and evaluating their performance using the quantisation error (QE) and topological error (TE). The final grid size is chosen as a compromise that achieves low QE and TE while maintaining a stable topology . We also state the selected grid used for the climate zoning (19 × 22, 418 neurons) and note that it was chosen based on low QE and TE , with the grid-size selection procedure described in Sect. 2.3.

Revisions made in the manuscript (revised manuscript, Sect. 2.3):

The performance of a SOM is evaluated using the quantisation error (QE) and topological error (TE), which measure the average distance between samples and their BMUs and the degree of topology preservation, respectively (Park et al., 2003; Jeong et al., 2010). Among candidate grid sizes, the final SOM structure is selected as a compromise between low QE and TE and a stable topology.

Q7: Figure 2: The description of the SOM component planes is clear but would benefit from a short explanatory note in the figure caption clarifying the color scale meaning (e.g., "red indicates high values, blue indicates low values").

A: Thank you for this helpful suggestion. In the revised manuscript, we updated the caption of Fig. 3 (formerly Fig. 2) to explicitly explain the colour scale on the SOM component planes. The caption now states that colours represent the standardised codebook (weight) values of each climate index across the SOM neurons, with warmer colours indicating higher values and cooler colours indicating lower values.

Figure updated (now Fig. 3):

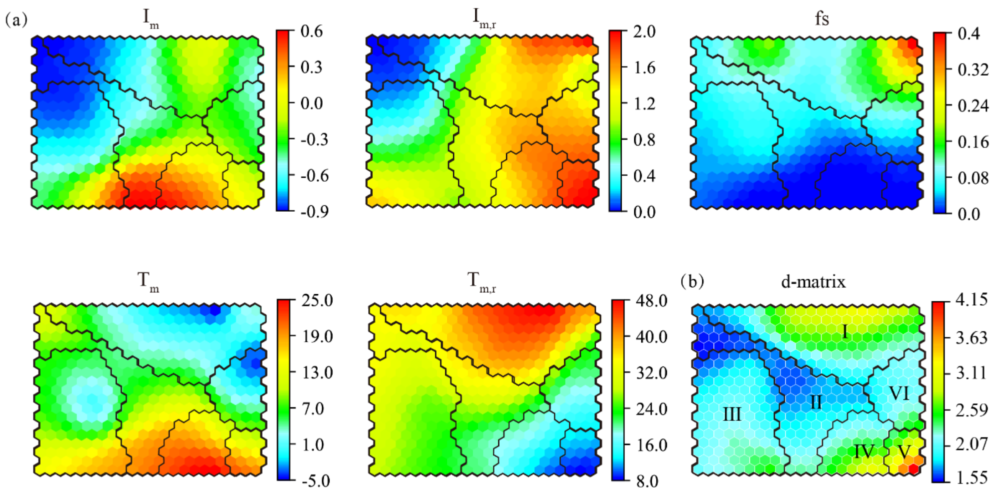

Figure 3. Self-organising map representation of the hydroclimatic feature space and fuzzy climate-region partition: (a) SOM component planes for the five climate indices: each hexagon represents a neuron summarising grid cells with similar standardised index values, colours show the corresponding codebook (weight) value, and black contour lines delineate the six SOM-FCM climate clusters that are later mapped to geographic space. (b) SOM distance matrix (d-matrix) giving the Euclidean distance between neighbouring neurons; cooler colours indicate smoothly varying hydroclimatic conditions, warmer colours mark sharper transitions in the climate-index space, and labels I-VI identify the six climate regions used in subsequent spatial

analyses.

Q8: Ensure consistent use of "" or "Figure" throughout the manuscript according to the journal's style guide.

A: Thank you. We have reviewed the entire manuscript and standardised the terminology to be consistent with the journal style, using "Fig." for figure references throughout the text and "Figure" only in figure captions where appropriate.

Q9: The transition between Section 3.1.3 (FCM clustering results) and Section 3.2 (Results of catchment classification) is abrupt. Consider adding a bridging sentence such as:"Based on the derived climate clusters, we further classified catchments with similar landscape attributes within each climate region."

A: Thank you for this helpful suggestion. In the revised manuscript, we smoothed the transition by adding an explicit bridging statement at the end of Sect. 3.1.3 (now Sect. 3.1.2) that clarifies how the derived climate regions are carried forward into the catchment classification. Specifically, we now state that, because climate memberships are diffuse near boundaries, each grid cell is assigned its dominant climate type for subsequent catchment classification, while retaining the full membership matrix for interpreting transitional areas . This leads directly into Sect. 3.2 ("Catchment types within climate regions"), where the within region landscape based classification is presented.

Revisions made in the manuscript (revised manuscript):

Added a brief bridging statement at the end of Sect. 3.1.2: *For subsequent catchment classification, each grid cell is therefore assigned to its dominant climate type (maximum membership), while the full fuzzy membership matrix is retained to support the interpretation of transitional areas.*

Q10: Lines 460–471:The sentence "The flow regime in climate region II presented multiple peaks following multiple peaks in precipitation in June and July during the same period." is ambiguous. Please revise or clarify its intended meaning.

A: Thank you for pointing this out. We agree that the original wording was confusing because it repeated "multiple peaks" and did not clearly specify the timing relationship between rainfall and runoff. In the revised manuscript, we rewrote this sentence to state the intended meaning more explicitly: catchments in climate region II exhibit a pronounced monsoonal seasonal cycle, and monthly runoff shows one or two peaks that occur in response to the summer precipitation maxima (typically June to July).

Revision made in the manuscript:

The sentence has been rephrased in Sect. 3.3.1 L450 to: *In climate region II, monthly runoff displays a pronounced monsoonal pattern with one or two peaks following the summer rainfall maxima and substantial interannual variability in high-flow months (e.g. Fenshuijiang and Tunxi).*

Q11: Line 494:The expression "Additionally, catchments with large spatial distances are capable of exhibiting similar hydrological characteristics (Maduwang and Daiying, year)" needs revision for clarity and proper citation formatting.

A: Thank you for pointing out the citation issue. We apologise for the oversight and have corrected the citation in the revised manuscript:

The results also highlight that hydrological similarity does not necessarily coincide with spatial proximity: catchments that are geographically distant but belong to the same climate-landscape class (e.g. Maduwang and Daiying) can exhibit comparable flow regimes, whereas neighbouring basins in different classes may behave quite differently.

Q12: Please revised your figure and make it clearly, specially the size and format of figures.

A: Thank you for this comment. We revised figures to improve clarity and consistency, including increasing font sizes, adjusting line widths and symbol sizes, improving colour contrast, and standardising layouts and formatting across figures. We also updated figure captions accordingly to ensure the figures are clear and readable.