Integrated Catchment Classification Across China Based on Hydroclimatological and Geomorphological Similarities Using Self-Organizing Maps and Fuzzy C-Means Clustering for Hydrological Modeling

Jiefan Niu[1,2,3], Ke Zhang[1,2,3,4,5], Xi Li[1,2,3], and Hongjun Bao[4,6]

[1] The National Key Laboratory of Water Disaster Prevention, Hohai University, Nanjing, Jiangsu, 210098, China
[2] Yangtze Institute for Conservation and Development, Hohai University, Nanjing, Jiangsu, 210024, China
[3] College of Hydrology and Water Resources, Hohai University, Nanjing, Jiangsu, 210024, China
[4] China Meteorological Administration Hydro-Meteorology Key Laboratory, Hohai University, Nanjing, Jiangsu, 210024, China
[5] Key Laboratory of Water Big Data Technology of Ministry of Water Resources, Hohai University, Nanjing, Jiangsu, 210024, China
[6] National Meteorological Center, China Meteorological Administration, Beijing, 100081, China

Correspondence to: Ke Zhang (kzhang@hhu.edu.cn), Hongjun Bao (baohongjun@cma.gov.cn)

Abstract. Catchment classification supports regionalisation and runoff prediction in data limited regions by organising basins into hydrologically coherent classes. China spans strong gradients in moisture availability, temperature regime, snow influence, and terrain, yet discharge observations remain sparse. We develop and evaluate an integrated climate-landscape classification for 13 487 HydroBASINS catchments using a hierarchical Self-Organising Map and fuzzy c-means (SOM–FCM) framework. Six hydroclimatic indices delineate climate regions on a 0.25° grid, and catchments are classified within each region using geomorphological and drainage network descriptors. The framework yields six climate regions and 35 classes, with fuzzy memberships characterising transitional areas. Hydrological relevance is assessed using seasonal hydrographs and event scale flow duration curves (FDCs) for ten gauged headwater catchments, and 13 flow signatures for 722 headwater basins matched to a discharge reanalysis product. Seasonal regimes are organised mainly by climate regions, whereas event response and high flow behaviour are modulated by landscape classes. Flow magnitude and high flow frequency signatures discriminate classes most strongly, while duration metrics show weaker contrasts. The resulting typology provides a transferable basis for selecting donor basins and constraining model parameters, thereby improving runoff prediction and regionalisation in ungauged catchments across China.

1 Introduction

Reliable runoff prediction is essential for sustainable catchment management across a range of time scales, including flood defense design, water allocation, and environmental impact assessment (Zang et al., 2021; Wang et al., 2021; Ma et al., 2021). In most settings, hydrological models still rely on observed discharge records for calibration to achieve satisfactory

删除了: Accurately identifying similar catchments is crucial for transferring model parameters and improving hydrological modeling, especially in ungauged regions with varied climates and topographies. This study presents an integrated method for catchment classification by combining Self-Organizing Maps artificial neural network (SOM) and Fuzzy C-Means clustering (FCM), utilizing hydrometeorological and geomorphological data. We evaluated six climate indices and fifteen landscape characteristics for catchments across China, identifying key variables through correlation and principal component analyses. The optimal classification produced six distinct climate regions and 35 catchment types with unique streamflow patterns. Validation using ten catchments confirmed the effectiveness of the SOM-FCM approach. The study underscores the importance of considering both climate and landscape factors for a comprehensive classification of catchments, offering valuable insights for hydrological model predictions in ungauged areas and enhancing our understanding of hydrological processes at various timescales.

删除了: Runoff

删除了: watershed

删除了: at

删除了: various

删除了: (!!! INVALID CITATION !!! (Ma et al., 2021; Wang et al., 2021; Zang et al., 2021); Zang et al., 2021; Wang et al., 2021; Ma et al., 2021) (!!! INVALID CITATION !!! (Ma et al., 2021; Wang et al., 2021; Zang et al., 2021))

删除了: Currently, the models used for runoff prediction more or less…

删除了: hydrological data to be

删除了: calibrated

删除了: practically acceptable

performance (Yaseen et al., 2019; Liu et al., 2020a). However, streamflow observations remain unavailable for many catchments worldwide, which constrains hydrological modelling and forecasting in ungauged basins (Kratzert et al., 2019; Carozza and Boudreault, 2021). This challenge motivated the IAHS Decade on Predictions in Ungauged Basins (PUB), which highlighted the need for transferable approaches that relate hydrological response to climatic forcing and catchment properties (Wagener et al., 2010; Hrachowitz et al., 2013).

A common pathway to prediction in ungauged basins is regionalisation, in which information is transferred among catchments assumed to be hydrologically similar. Regionalisation methods include spatial proximity, regression based approaches, and physical similarity strategies (Tsegaw et al., 2019; Kittel et al., 2020; Guo et al., 2020). Physical similarity is particularly attractive because it can be implemented using climatic and catchment descriptors that are increasingly available from gridded datasets and remote sensing, and can therefore be applied directly to ungauged locations. This is especially relevant for China where many small and medium-sized basins are located in complex hilly and mountainous terrain. In these areas, monitoring is often limited, parameter calibration is difficult, and vulnerability to floods and droughts remains high (Zeng et al., 2021; Liu et al., 2020b). A robust similarity framework therefore has clear value for improving model parameterisation and forecasting skill in data limited regions.

Catchment classification provides an operational basis for similarity frameworks by organising climatic and landscape heterogeneity into homogeneous types, thereby supporting basin to basin transfer and systematic comparisons across regions (Wagener et al., 2010; Jehn et al., 2020). Existing approaches typically rely on three families of information: climatic descriptors (Pagliero et al., 2019; Knoben et al., 2018), catchment physical attributes (Tarasova et al., 2020; Loritz et al., 2019; Leibowitz et al., 2016), and hydrological signatures derived from observed discharge (Singh et al., 2016; Addor et al., 2018). Descriptor based approaches have the advantage of broad spatial availability, which makes them directly applicable to ungauged basins. Signature based approaches characterise hydrological behaviour more directly, but they require streamflow observations and therefore cannot be applied everywhere without additional modelling or reanalysis data.

Large sample studies consistently indicate that climate provides a first order control on streamflow regimes, particularly through aridity or moisture availability, snow influence, and seasonality (Kuentz et al., 2017; Jehn et al., 2020; Berghuijs et al., 2014). At the same time, landscape properties such as relief, soils, vegetation, and drainage network organisation can strongly modulate runoff variability and catchment response within similar climatic settings (Loritz et al., 2018; Addor et al., 2017). This highlights an important scale consideration. Broad climatic gradients often structure regional runoff regimes, whereas within a climatic envelope, landscape differences increasingly shape local response characteristics. However, many classification studies cluster climatic and landscape descriptors within a single feature space, which can obscure this scale dependence and may limit the interpretability and transferability of similarity groups for regionalisation (Yang et al., 2018; Ghotbi et al., 2020; Gao et al., 2019).

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

删除了: … ADDIN EN.CITE <EndNote><Cite><Author>Yaseen</Author><Year>2019</Year><RecNum>64</RecNum><DisplayText>(Yaseen et al., 2019; Liu et al., 2020a)</DisplayText><record><rec-number>64</rec-number><foreign-keys><key app="EN" db-id="rxpdep9dd5sws1edeauv5dacexrvrfxvzz55" timestamp="1607566778">64</key></foreign-keys><ref-type name="Journal Article">17</ref-type><contributors><authors><author>Yaseen, Zaher Mundher</author><author>Sulaiman, Sadeq Oleiwi</author><author>Deo, Ravinesh C</author><author>Chau, Kwok-Wing</author></authors></contributors><titles><title>An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction</title><secondary-title>Journal of Hydrology</secondary-title></titles><periodical><full-title>Journal of Hydrology</full-title></periodical><pages>387-408</pages><volume>569</volume><dates><year>2019</year></dates><isbn>0022-1694</isbn><urls></urls></record></Cite><Cite><Author>Liu</Author><Year>2020</Year><RecNum>89</RecNum><record><rec-number>89</rec-number><foreign-keys><key app="EN" db-

删除了: International Association of Hydrological Sciences (…AHS) launched the…Decade on Predictions in Ungauged Basins (PUB) … which highlighted the need for transferable approaches that relate hydrological response to climatic forcing and catchment properties (Wagener et al., 2010; Hrachowitz et al., 2013).in 2002 to explore prediction methods for ungauged basins using an improve

删除了: Over the past few years, the PUB community has proposed a series of regionalization methods, including…spatial proximity, regression based approaches, and physical similarity, and…strategies regression for data-sparse regions

删除了: (Tsegaw et al., 2019; Kittel et al., 2020; Guo et al., 2020)(Guo et al., 2021; Kittel et al., 2020; Tsegaw et al., 2019), essentially defining homogeneous zones with similar hydrological characteristics… Among these, the p…hysical similarity is particularly attractive because it can be implemented using climatic and catchment descriptors that are increasingly available from

设置了格式: 非突出显示

删除了: Regionalizing a particular hydrological characteristic and applying it to ungauged catchments is complicated as the behavior patterns in hydrology are the consequence of both climate and geomorphology (Gao et al., 2019). It is possible to organize climatological and geomorphological heterogeneity patterns and develop classification frameworks to categorize catchments with

删除了: The first two feature types offer the advantage of being available for all geographical locations, and they can be applied directly to ungauged catchments. …arge sample studies consistently indicate that Different characteristics are inclined to impact diverse hydrological behaviors (Mcmillan, 2020). According to recent research, …limate provides a first order control on streamflow

删除了: Another study supporting this conclusion focused on analyzing and clustering 35,215 catchments across Europe. They discovered that the flow signatures were primarily affected by climatic characteristics, particularly those corresponding to average and high flows. Furthermore, topography is also a major factor controlling flow variability (Kuentz et al., 2017). Different spatial

设置了格式: 非突出显示

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.

Machine learning offers practical tools for constructing similarity frameworks from high dimensional descriptor datasets (Yang et al., 2020). The Self-Organizing Map (SOM) is an efficient unsupervised method that projects multivariate data onto a 2D lattice while preserving neighbourhood structure (Kohonen, 1982). In many applications, SOM produces more neurons than the desired number of final groups and therefore requires an additional clustering step (Kiang, 2001). Most previous studies have used hard clustering after SOM, such as k-means or hierarchical clustering, which assigns each catchment to a single class (Nguyen et al., 2015; Boscarello et al., 2016). In contrast, fuzzy c-means (FCM) provides soft clustering by allowing partial memberships, which is better aligned with continuous environmental gradients and transitional catchments (Bezdek et al., 1984). Although SOM–FCM combinations have been applied in other environmental classification contexts (Lee et al., 2019) and have shown promise for heterogeneous feature spaces, they have not been systematically developed and validated as a national framework for catchment similarity and regionalisation in China.

In this study, we develop an integrated climate and landscape classification across China using a hierarchical SOM-FCM framework. Similarity is organised in two steps. First, we delineate homogeneous climate regions using hydroclimatic indices. Second, we classify catchments within each climate region using geomorphological and drainage network descriptors, which reflects the scale dependent roles of climatic forcing and landscape modulation. The framework is applied to 13 487 HydroBASINS catchments across China using six hydroclimatic indices and fifteen catchment descriptors compiled from national scale datasets. Hydrological validity is assessed using a two-stage strategy that combines gauge-based evaluation in ten headwater catchments with large sample, signature based statistical testing in 722 headwater basins matched to a discharge reanalysis product. In this way, the study provides a transferable national typology of climate and landscape similarity intended to support regionalisation and runoff prediction in ungauged basins across China.

The objectives of this paper are to (i) construct a hierarchical national framework for catchment similarity in China by integrating SOM-FCM; (ii) derive climate regions and within region catchment types based on hydroclimatic and geomorphological similarity; and (iii) evaluate the hydrological relevance of the resulting groups across time scales using both

删除了: (Liu et al., 2014)

删除了: At small spatial scales, catchments are typically in the same climatic region with similar temperatures and water conditions, and the most significant factor is the hillslope structure of the catchment (Loritz et al., 2018). Recent research on catchment classification has simply combined climate and landscape indices in different climate regions without considering the scale effects of these two types of variables (Ghotbi et al., 2020; Yang et al., 2018). Additionally, some countries have been conducted on hydrological similarity based on small-sample stations or catchments (Liu et al., 2019; Yang et al., 2020b; Zhai et al., 2021) achieving significant classification results. However, China has yet to integrate climatic and catchment physical characteristics with spatial scales to propose a comprehensive framework for similar basin classification.

删除了: With the advancement of computer technology in the 21st century, the widespread use of m

删除了: in regionalization studies has become an indisputable fact

删除了: (Yang et al., 2020)

删除了: self

删除了: organizing

删除了: map

删除了: machine learning

删除了: f

删除了: or visualizing complex high-dimensional data structures on a

删除了: surface

删除了: (Kohonen, 1982)

删除了: However, the number of output

删除了: exceeds the expected

删除了: in most applications

删除了: (Kiang, 2001)

删除了: As a result, previous

删除了: added

删除了: steps to

删除了: results

删除了: clustering algorithms and

删除了: g algorithms

删除了: to generate appropriate groups and facilitate quantitative analysis

删除了: (Boscarello et al., 2016; Nguyen et al., 2015)

删除了: Combining these algorithms has been used in various fields. Zang et al. (2021) used SOM with k-means to map future drought conditions in China, and produced relatively accurate classification results.

删除了: Kim et al. (2020) investigated deep thermal groundwater in South Korea using SOM combined with hierarchical clustering and found that five major clusters accounted for various bedrock ...

gauge records and reanalysis flow signatures. The remainder of this paper is organised as follows. Section 2 describes the datasets, indices and descriptors, and the hierarchical SOM-FCM methodology. Section 3 presents the derived climate regions and catchment classes and summarises validation results. Section 4 discusses implications for regionalisation in China and limitations of the proposed framework. Section 5 concludes the paper.

删除了: The objective of this study was to combine SOM and FCM algorithms to identify hydrologically similar catchments in China. This research contributes to a better understanding of the relationship between the similarities of hydroclimatological and geomorphological attributes, and the similarities of hydrological processes on the basin level across China.

2 Methodology and data

This study adopts a multi-step workflow that integrates multiple datasets, index calculations, classification, and validation analyses. The overall procedure is summarised in Fig. 1. The datasets, classification framework, and validation strategy are described below.

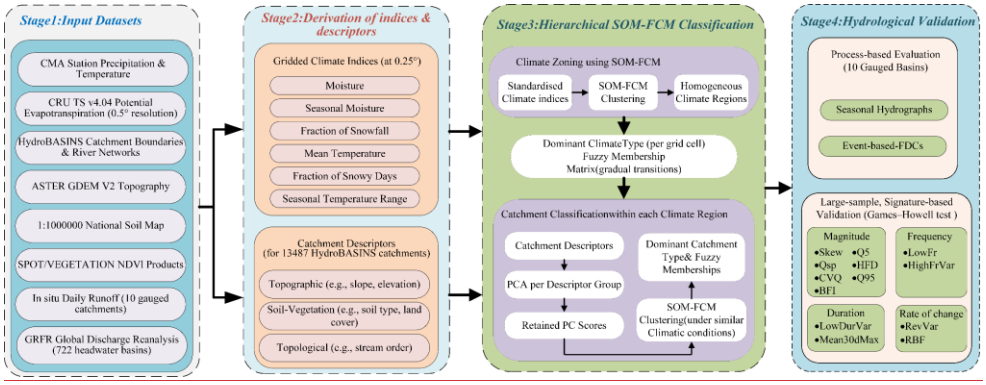


Figure 1. Flow chart of the different steps followed in the study.

设置了格式: 字体: 非加粗

2.1 Database

删除了:

Climatic forcing for the hydroclimatological indices was represented by precipitation (P), air temperature (T), and potential evapotranspiration (EP). Daily P and T for 1982-2015 were obtained from 613 meteorological stations operated by the China Meteorological Administration (CMA) and distributed by the National Meteorological Information Center (http://data.cma.cn/). 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). All climate variables were interpolated to a 0.25° × 0.25° grid over China. Missing values were infilled using a weighted nearest neighbour approach to improve spatial consistency.

Catchment boundaries and river networks were obtained from the HydroSHEDS and HydroBASINS products developed by the World Wildlife Fund (https://www.hydrosheds.org/page/overview). HydroBASINS provides a seamless, hierarchically nested set of sub basin polygons with consistent drainage topology (Lehner and Grill, 2013). These datasets have been widely

used in large sample hydrological studies and have been reported to outperform many existing global watershed and river maps (Yamazaki et al., 2014; Carozza and Boudreault, 2021). In this study, 13 487 catchments within China were selected from HydroBASINS as the basic spatial units for subsequent analyses.

To characterise geomorphological and landscape conditions, we compiled descriptors of topography, soils, vegetation, and drainage network structure for each catchment. Topographic and drainage network properties were derived from the ASTER GDEM V2 digital elevation model at 30m resolution, obtained from the Geospatial Data Cloud of the Computer Network Information Center, Chinese Academy of Sciences (http://www.gscloud.cn). Soil attributes were extracted from the 1:1 000 000 soil map of China produced by the Institute of Soil Science (http://www.issas.ac.cn), Chinese Academy of Sciences. Vegetation cover was characterised using SPOT/VEGETATION NDVI products from the Resource and Environmental Science and Data Center (https://www.resdc.cn). Together, these datasets provide consistent national coverage and enable the derivation of catchment descriptors that are suitable for regional classification and for application in ungauged basins.

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.

2.2 Selection of Climate Indices and Catchment Descriptors

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.

The spatial organisation of climate and landscape has long been recognised as a primary control on catchment hydrological response. 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). Previous large-sample studies have shown that a small set of climate indices can effectively summarise these controls. Betterle et al. (2019) identified five indices that are particularly relevant for hydrological processes: annual average aridity, seasonality of aridity, fraction of precipitation falling as snow, average rainfall intensity, and seasonality of rainfall intensity. Although precipitation intensity varies substantially across regions, its hydrological effects are strongly conditioned by local catchment characteristics. We therefore do not include rainfall-intensity-based indices in this study and adopt three indices (I_m, $I_{m,r}$ and fs) from this framework, while examining the role of landscape characteristics in modulating hydrological response in later sections.

Previous work has further demonstrated that aridity and snow-related indices are strongly correlated with streamflow patterns, even when rainfall intensity is not explicitly considered (Knoben et al., 2018). Temperature, as an indicator of both snow and evapotranspiration processes, provides additional information on climate similarity. Building on this reasoning, we introduce the three temperature-related indices (T_m, $T_{m,r}$ and Ds) to capture the seasonal and spatial variability of thermal controls on hydrological processes.

All six climate indices were computed for each 0.25°land grid cell over China based on the gridded climate fields and station observations described in Sect. 2.1. Some of these indices have previously been used to delineate homogeneous climate regions, but not in this specific combination. The indices were calculated using Eqs. (1)-(7).

$$MI(t) = \begin{cases} 1 - \frac{E_P(t)}{P(t)}, P(t) > E_P(t) \\ 0, P(t) = E_P(t) \\ \frac{P(t)}{E_P(t)} - 1, P(t) < E_P(t) \end{cases} \quad (1)$$

$$I_m = \frac{1}{12}\sum_{t=1}^{t=12} MI(t), \quad (2)$$

$$I_{m,r} = \max\big(MI(1,2,\dots,12)\big) - \min(MI(1,2,\dots,12)), \quad (3)$$

$$fs = \frac{\sum P(T(t)\le T_0)}{\sum_{t=1}^{t=12} P(t)}, \quad (4)$$

$$T_m = \frac{1}{12}\sum_{t=1}^{t=12} T(t), \quad (5)$$

$$T_{m,r} = \max\big(T(1,2,\dots,12)\big) - \min(T(1,2,\dots,12)), \quad (6)$$

$$Ds = \frac{\sum D(T(t)\le T_0)}{\sum_{t=1}^{t=12} D(t)}, \quad (7)$$

删除了: Betterle et al. (2019); Addor et al. (2017)

删除了:

$P(t)$, $E_P(t)$, and $T(t)$ are the mean monthly observed values of precipitation, potential evapotranspiration, and temperature, respectively; $D(t)$ is the number of days per month; T_0 is the threshold temperature, below which precipitation is presumed to occur in snow form, set at 0 °C.

2.2.2 Catchment Descriptors

To characterise geomorphological and drainage-network controls on hydrological response, we selected fifteen catchment descriptors that describe landscape properties at the basin scale, which are summarised in Table 1. These descriptors are grouped into three categories: (i) topographic characteristics, (ii) soil and vegetation characteristics, and (iii) drainage-network characteristics. Together, they provide complementary information on relief and surface form, subsurface properties, land cover, and drainage structure, all of which are known to influence runoff generation and flow routing. These variables have been widely used in regionalisation and prediction in ungauged basins and offer a physically interpretable basis for large-sample catchment classification (Jehn et al., 2020; Boscarello et al., 2016; Addor et al., 2017; Addor et al., 2018).

All descriptors were computed for the 13 487 catchments using the DEM, soil, land-cover, and hydrographic datasets introduced in Sect. 2.1. Topographic and drainage-network properties were derived from the digital elevation model and river network, whereas soil texture and vegetation indices were obtained from national soil maps and remote sensing products. This ensures that the selected variables are consistently defined across all basins and can, in principle, be transferred to ungauged regions where similar datasets are available.

Three key topographic descriptors are defined from the hypsometric curve f(x), which relates relative elevation to relative contributing area within a catchment. The hypsometric integral (HI) summarises the overall shape of the hypsometric curve and thus the gross surface form; the hypsometric gradient (AS) reflects the degree of topographic relief; and the mean topographic index (TI) characterises the tendency for water to accumulate in concave, low-slope areas. These descriptors are computed as follows:

$$HI = \int_0^1 f(x)dx, \quad (8)$$

$$AS = \frac{f(0.2)-f(0.8)}{0.8-0.2}, \quad (9)$$

$$TI = \frac{1}{n}\sum_{i=1}^{n} \ln\left(\frac{\alpha_i}{tan\beta_i}\right), \quad (10)$$

where f(0.2) and f(0.8) are the relative elevations at relative contributing areas 0.2 and 0.8, respectively; α_i and β_i indicate the upslope contributing area and local slope of grid cell i ; and n is the total number of raster cells within the catchment.

Table 1. Catchment descriptors used for geomorphological and network-related characterization in this study.

	Description	Variables	Unit	Mean	Range
Topographic characteristics	Mean elevation	H	m	1,795.52	-134.36-5,803.84
	Elevation range	ΔH	m	1,086.08	3-7,315
	Hypsometric curve integral	HI	-	0.64	0.24-1
	Gradient of hypsometric curve	AS	-	0.52	0-1.84
	Mean topographic index	TI	-	10.02	7.43-12.72
	Mean slope	β	degree	5.45	0.02-28.32
Soil and vegetation characteristics	Sand fraction	$Sand$	%	44.42	0-91.72
	Clay fraction	$Clay$	%	20.22	0-52.23
	Silt fraction	$Silt$	%	33.14	0-54.00
	NDVI	$NDVI$	-	0.56	0-0.9
Topological characteristics	Area	A	km^2	761.04	15-14,612.8
	Length	L	km	52.96	9.3-615.7
	Form factor	Rf	-	0.33	0.02-1.97
	Elongation ratio	Re	-	0.61	0.15-1.58
	Drainage density	Rd	km/km^2	0.29	0.06-1.40

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.

2.3.1 SOM-FCM combined algorithm

SOM is an unsupervised artificial neural network that projects high-dimensional inputs onto a low-dimensional array of neurons while preserving the topological structure of the input space (Kohonen, 1982). Each neuron is associated with a weight vector of the same dimension as the input vector. During training, the best-matching unit (BMU) for each input is identified by the minimum Euclidean distance between the input and neuron weight vectors, and the BMU and its neighbours are iteratively updated. As the learning rate and neighbourhood radius decrease, the SOM converges to a structured representation in which similar samples are mapped to neighbouring neurons. 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

删除了: 1

删除了: Methods

删除了: The methodology integrates hydroclimatological and geomorphological data to identify catchment similarities. This approach combines the Self-Organizing Maps (SOM) algorithm, an unsupervised artificial neural network, with Fuzzy C-Means (FCM) clustering, a soft clustering method rooted in fuzzy set theory. The process begins with the selection and preparation of relevant climate indices and geomorphological characteristics, followed by the unsupervised classification of catchments. Initially, SOM is used to group catchments based on spatial climate patterns, identifying regions of meteorological homogeneity. Subsequently, FCM refines the classification by clustering catchments within these regions, accounting for gradual transitions in climate and landscape. Finally, the classifications are validated using streamflow data from selected catchments to ensure the model's reliability.

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.

FCM is a soft clustering algorithm based on fuzzy set theory (Pal et al., 2005). It allows each sample to belong to multiple clusters with membership grades between 0 and 1 by minimising a weighted within-cluster squared-distance objective. The number of clusters is not fixed a priori but is selected using internal validity indices, here the Davies-Bouldin index (DBI) and the silhouette coefficient (SC), which jointly reflect within-cluster compactness and between-cluster separation (Rao and Srinivas, 2006; Pakhira et al., 2004; Halim et al., 2017).

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).

2.3.2 Climate Zoning Using SOM-FCM

The SOM-FCM framework is first used to delineate homogeneous climate regions on the 0.25° grid from the climate indices defined in Sect. 2.2.1. To reduce redundancy in the climate feature space, the six indices are screened for conceptual and statistical dependence; a subset is retained as clustering variables, and the remaining index is used only as an auxiliary descriptor when interpreting the resulting regions. For each land grid cell, the standardised values of the selected indices form the input vector to the SOM.

SOM and FCM are then applied as described in Sect. 2.3.1. A two-dimensional SOM is trained on the climate-index vectors, using QE and TE to select an appropriate grid size. The resulting codebook vectors provide a topology-preserving discretisation of the continuous climate space, to which FCM is applied to obtain a fuzzy partition. The number of climate regions is determined by jointly considering the DBI, SC, and hydrological interpretability. Each grid cell is assigned a dominant climate type according to the cluster with the highest membership value, while the full membership matrix is retained to represent gradual transitions between regions.

2.3.3 Catchment classification within climate zones

Building on the climatic partitioning, catchments are subsequently classified within each climate region using their landscape characteristics. Each of the 13 487 catchments is associated with a climate region via the dominant climate type at its outlet grid cell. Within each region, the fifteen catchment descriptors defined in Sect. 2.2.2 are used to describe geomorphological and drainage-network controls. To obtain a compact and non-redundant feature space, pairwise rank correlations are examined and principal component analysis (PCA) is applied separately to the three descriptor groups (topographic, soil and vegetation,

带格式的: 缩进: 左 0 字符

带格式的: 缩进: 左 0 字符

and drainage-network characteristics). 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 scores of the retained principal components, and FCM is applied to the corresponding SOM codebook vectors to derive catchment types under similar climatic conditions. The number of catchment clusters per region is selected using the same validity criteria (DBI, SC), together with hydrological interpretability. For each catchment, the cluster with the highest membership value is treated as its dominant type, and a membership threshold of 0.5 is adopted to distinguish clearly defined types from transitional catchments with overlapping properties. These catchment types provide the basis for subsequent analyses of spatial patterns of hydrological similarity and for model regionalisation.

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

带格式的: 缩进: 左 0 字符

设置了格式: 非突出显示

设置了格式: 非突出显示

设置了格式: 非突出显示

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.

Table 2. Definitions of the 13 hydrological flow signatures used for validating catchment classification.

Component of flow regime	Variable	Description
Magnitude	skew	Skewness = mean/median of daily flows
	Qsp	Mean specific flow
	CVQ	Coefficient of variation = standard deviation / mean of daily flows
	BFI	Base flow index: 7-day minimum flow divided by mean annual daily flow
	Q5	5th percentile of daily specific flow
	HFD	High-flow discharge: 10th percentile of daily flow divided by median daily flow
	Q95	95th percentile of daily specific flow in mm
Frequency	LowFr	Total number of low flow spells (threshold equal to 5 % of mean daily flow) divided by the record length
	HighFrVar	Coefficient of variation in annual number of high-flow occurrences (75th percentile)
Duration	LowDurVar	Coefficient of variation in annual mean duration of low flows (25th percentile)
	Mean30dMax	Mean annual 30-day maximum divided by median flow
Rate of change	RevVar	Coefficient of variation in annual number of reversals in flow direction
	RBF	Richards-Baker flashiness: sum of absolute values of day-to-day changes in mean daily flow divided by the sum of all daily flows

3 Results

3.1 Hydroclimatic patterns and climate regions

3.1.1 Spatial patterns of climate indices

Figure 2 shows pronounced spatial gradients in the six climate indices across China. At the large scale, patterns are mainly organised by latitude and elevation, whereas major mountain ranges (e.g. the Tianshan and Himalayan Mountains) introduce sharp local transitions. All indices were standardised to the range 0-1 and visualised using two ternary RGB composites: one for the moisture-related indices (I_m, $I_{m,r}$, and fs) and one for the temperature-related indices (T_m, $T_{m,r}$, and Ds) (Fig. 2g). These composites highlight coherent spatial structures and relationships among the indices. The moisture composite displays a clear northwest-southeast gradient in I_m, $I_{m,r}$, and fs, reflecting a progressive increase in wetness from arid to humid regions (Fig. 2g, left). Arid regions (red tones) occupy northwestern China and are characterised by extensive deserts, high potential evapotranspiration relative to precipitation, negligible seasonality in humidity, and virtually no snowfall. Humid regions (dark green) are concentrated along the middle and lower reaches of the Yangtze River and exhibit little snowfall, weak seasonality and abundant rainfall throughout the year. Transitional climate regions (bright green and yellow) lie between these extremes and show a strongly seasonal water-energy balance, most notably in the seasonal variation of precipitation and potential evapotranspiration. Areas where most precipitation falls as snow are highlighted by pink colours.

设置了格式: 字体: 小五
带格式的: 缩进: 首行缩进: 1 字符, 行距: 2 倍行距
带格式的: 行距: 固定值 12 磅
设置了格式: 字体: 小五
带格式的: 行距: 固定值 12 磅
设置了格式: 字体: 小五
带格式的: 行距: 固定值 12 磅
设置了格式: 字体: 小五
带格式的: 行距: 固定值 12 磅
设置了格式: 字体: 小五
带格式的: 行距: 固定值 12 磅
设置了格式: 字体: 小五
带格式的: 行距: 固定值 12 磅

上移了 [2]:

$$MI(t)=\begin{cases}1-\frac{E_P(t)}{P(t)}, P(t)>E_P(t)\\ 0, P(t)=E_P(t)\\ \frac{P(t)}{E_P(t)}-1, P(t)<E_P(t)\end{cases}, \quad (1)$$

$$I_m=\frac{1}{12}\sum_{t=1}^{t=12}MI(t), \quad (2)$$

$$I_{m,r}=\max\left(MI(1,2,\dots,12)\right)-\min(MI(1,2,\dots,12)), \quad (3)$$

$$fs=\frac{\sum P(T(t)\le T_0)}{\sum_{t=1}^{t=12}P(t)}, \quad (4)$$

$$T_m=\frac{1}{12}\sum_{t=1}^{t=12}T(t), \quad (5)$$

$$T_{m,r}=\max\left(T(1,2,\dots,12)\right)-\min(T(1,2,\dots,12)), \quad (6)$$

$$Ds=\frac{\sum}{\sum_{t=1}^{t=12}D(t)}, \quad (7)$$

$P(t)$, $E_P(t)$, and $T(t)$ are the mean monthly observed values of precipitation, potential evapotranspiration, and temperature, respectively; $D(t)$ is the number of days per month; T_0 is the threshold temperature, below which precipitation is presumed to occur in snow form, set at 0 °C.

2.1.2 Self-organizing map clustering algorithm

The function of catchment classification is to map the complex spatial structure of hydrological patterns onto regional patterns that are immediately recognizable and interpretable. We recast similar hydrological patterns through two-step clustering to first identify the meteorological homogeneity regions and then classify the

移动了(插入) [1]

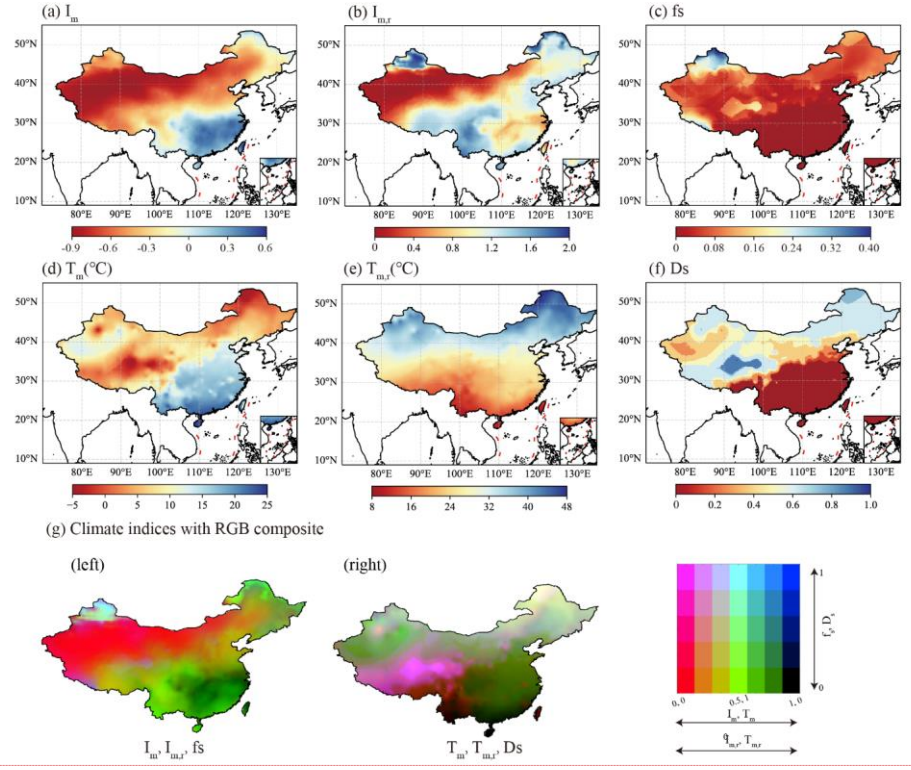
删除了: 2.1.1 Selection of climate indices preparation

The spatial patterns of climate and landscape have been evaluated as causal factors in determining the hydrological response of catchments. Climate not only directly influences runoff generation processes at the event scale but also indirectly affects the hydrological cycle by acting on longer-term soil moisture availability and the co-evolution of landscape and vegetation. Climate patterns produce significant differences in the long-term balance between available water and energy. Catchments in arid regions are generally considered to have a limited water supply. It is characterized by sparse precipitation with high evaporation, resulting in less recharge of the aquifer. In addition, the stream in these regions is usually lost and reaches a certain position, with th ...

删除了: Classification of regions based on climate factors

删除了: Spatial distributions of climate indices

In contrast, the temperature composite primarily varies along latitudinal bands and is modulated by orography (Fig. 2g, right). Warm regions in southern China (dark green) display high mean temperatures and weak intra-annual variability, whereas cold regions in northern China (yellow) are characterised by low mean temperatures, pronounced seasonality and frequent snow processes. Transitional temperature regimes (dark yellow to green) occur between these zones, with increasing temperature and decreasing seasonality from north to south. A distinct cold region in southwestern China (pink), associated with the Tibetan Plateau, is marked by persistently low temperatures, long-lasting snow cover and limited seasonal variation. Overall, the climate indices reveal a strong regional organisation of meteorological conditions across China. The spatial pattern of snowy days is strongly and negatively correlated with mean annual temperature (Spearman rank correlation -0.82). To reduce redundancy in the climate feature space, the *Ds* index was therefore not used as a clustering variable and was retained only for descriptive analysis.



删除了： Figure 1 shows that the spatial distribution of six individual climate indicators gradually changed. Climate features are primarily influenced by latitude and altitude; however, abrupt changes in topography (e.g., the Tianshan and Himalayan Mountains) can result in relatively sharp climate transitions in some regions. The six indicators were standardized to between 0 and 1, and were then applied to show a single overview of the moisture (I_m, $I_{m,r}$, and fs) and temperature indices (T_m, $T_{m,r}$, and Ds) with RGB color scales combined to a map, respectively (Fig. 1g). This visualizes the spatial distribution of each climate indicator and describes them according to their corresponding relationships. Overall, there was a trend from northwest to southeast in the first three indices, indicating the degree of wetness in China. (Fig. 1g, left). Specifically, arid regions (red areas) are located in the northwest part of China and are characterized by a large desert with high PET compared to available precipitation, almost no seasonal changes, and no snowfall. The wet regions (dark green) are concentrated along the middle and lower reaches of the Yangtze River. . Generally, these areas have no snowfall, little seasonal variation, and continuous rainfall. The transitional climate regions (bright green and yellow) are located between arid and humid zones, and the climate in this region experiences strong seasonality in their water-energy balance, most notably in the seasonal variation of precipitation and the seasonal pattern of PET. The pink regions indicate areas where the majority of the precipitation falls as snow. Unlike the regional distribution of the wetness indices, the temperature feature gradient from the indices varies gradually along the latitudinal band and mutates with abrupt changes in topography (Fig. 1g, right). There are regions with high temperatures in southern China (dark green), where the temperature remains constant throughout the year.The low-temperature regions (yellow) in northern China have relatively strong seasonal temperature changes and snow processes. The transitional climate regions (dark yellow and green) are located between two zones, increasing in temperature and decreasing in seasonality from north to south. Furthermore, there is a low-temperature region (pink) in the southwest, influenced by the Tibetan Plateau, where snow accumulates for years and temperatures remain extremely low with little seasonal variation.
According to the climate spatial distribution characteristics of each individual indicator, the spatial distribution of meteorological variables has a strong regional component in China. Meanwhile, the spatial distribution of snow days was strongly inversely correlated with the average temperature (Spearman rank correlation coefficient was -0.82). To avoid a classification procedure that could introduce redundancy, the snow day index was removed, and five different climate indices were selected for climate partitioning.

Figure 2: Spatial patterns of the six climate indices over China for 1982–2015: **(a)** average moisture index (I_m), **(b)** seasonal moisture index ($I_{m,r}$), **(c)** fraction of precipitation falling as snow (fs), **(d)** annual average temperature (T_m), **(e)** seasonal temperature range ($T_{m,r}$), and **(f)** fraction of snowy days (Ds). Panel **(g)** shows two ternary RGB composites that summarise the joint spatial organisation of the indices, with the moisture-related indices (left) encoded as I_m (red), $I_{m,r}$ (green), and fs (blue), and the temperature-related indices (right) encoded as T_m(red), $T_{m,r}$ (green), and Ds (blue).

3.1.2 Climate regions derived from SOM-FCM

The climate indices were first projected onto a two-dimensional SOM with a 19 × 22 rectangular grid (418 neurons), selected based on low QE and TE. FCM clustering was then applied to the SOM codebook vectors. The optimal number of clusters was identified as six by jointly minimising the Davies-Bouldin index (DBI) and maximising the silhouette coefficient (SC). 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.

Figure 3 summarises the SOM-FCM classification in climate-index space. The component planes (Fig. 3a) display weight values of the five clustering indices and reveal distinct yet partly coherent patterns: for example, low I_m and $I_{m,r}$ often coincide with low snowfall fractions, whereas high I_m frequently co-occurs with high T_m, reflecting physically interpretable regimes such as arid climates with low moisture availability and warm, humid monsoonal climates. The black contour lines indicate the six FCM clusters on the SOM grid. The distance matrix (Fig. 3b) shows mostly smooth transitions between neighbouring neurons, confirming that the hydroclimatic space is continuous but structured rather than sharply segmented.

When projected back into geographic space, the six SOM-FCM clusters correspond to major climate regions across China (Fig. 4). In the climate feature space, the cluster centroids are separated along moisture, temperature, and snow-related gradients (Fig. A1). Region I covers much of the north-western desert belt and is extremely arid, with very low moisture indices, almost no snowfall and only modest seasonality in temperature. Region II occupies the middle and lower reaches of the Yangtze River in south-eastern China and exhibits a warm, humid monsoon climate with high temperatures and weak intra-annual variability in humidity and temperature. Region III spans the Northeast China Plain and is cold and snow-influenced, with low mean temperatures, frequent snowfall and strong seasonality in both temperature and precipitation. Region IV, covering much of the North China Plain and the middle and lower Yellow River basin, is warmer and more seasonal than Region III, showing higher temperatures and more pronounced variability in the moisture indices. Region V comprises the foreland basins north of the Tianshan Mountains and is characterised by a high fraction of snowfall and marked seasonal contrasts in both temperature and moisture. Region VI corresponds to the Tibetan Platea; like Region III it is cold and snow-affected, but it features more uniform temperatures and somewhat weaker seasonality. Because the underlying climate fields vary smoothly, boundaries between regions are diffuse and grid cells near the interfaces exhibit mixed memberships. 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.

删除了: 1

删除了: .

删除了: Maps of average values of climate indices from 1982 to 2015 across China

设置了格式: 字体: 加粗

设置了格式: 字体: 加粗

删除了: that falls

设置了格式: 字体: 加粗

设置了格式: 字体: 非加粗

设置了格式: 字体: 加粗

设置了格式: 字体: 加粗

设置了格式: 字体: 非加粗

设置了格式: 字体: 加粗

设置了格式: 字体: 非加粗

设置了格式: 字体: 加粗

删除了: provides an overview of the climate indices combination using an RGB color scale, where the moisture indices (left) are represented by

删除了: temperature indices (right) by

设置了格式: 字体: 非加粗

删除了: **The clustering results of SOM**

上移了 [1]: Vesanto (1999) suggested that SOM results can be form of two types of maps: component planes and distance matrices (d-matrices). On the component planes, the individual neuron weight vector values are shown using color coding; blue and red correspond to low and high values, respectively. This allows the recognition of the mutual dependence relationship among variables when comparing the patterns of the component planes. For instance, opposite gradients in the component planes indicate a negative correlation between the variables. In the median d-matrix, the median Euclidean distances between neighboring neurons are indicated by the color scale. Consequently, we can obtain an indication of the relative distances between neurons; neurons with high similarity (blue) may be considered as clusters.
Five features are apparent in the component planes (Fig. 2a). First, the vector values did not exhibit a horizontal or diagonal gradient distribution in color. Overall, the five climate indicators had different patterns, and the weight vectors in the component planes did not follow a uniform distribution, indicating that the indicators were relatively independent and could represent a range of hydroclimatic characteristics. Second, the low-value regions of I_m, $I_{m,r}$, as well as the high-value regions of I_m and T_m, have a consistent distribution on the component plane map. The consistency between climate indices indicates that several climate features develop synergistically in local regions. For example, the arid regions exhibit stable humidity with little seasonal variation, while humid regions have high temperatures and precipitation predominantly in the form of rainfall. This aligns with existing studies on the synchrony of rain and heat in China's climate characteristics (Hao et al., 2018).

删除了: Using standardized climate index raster data from China, SOM technology was used to cluster climate data, and 418 output ...

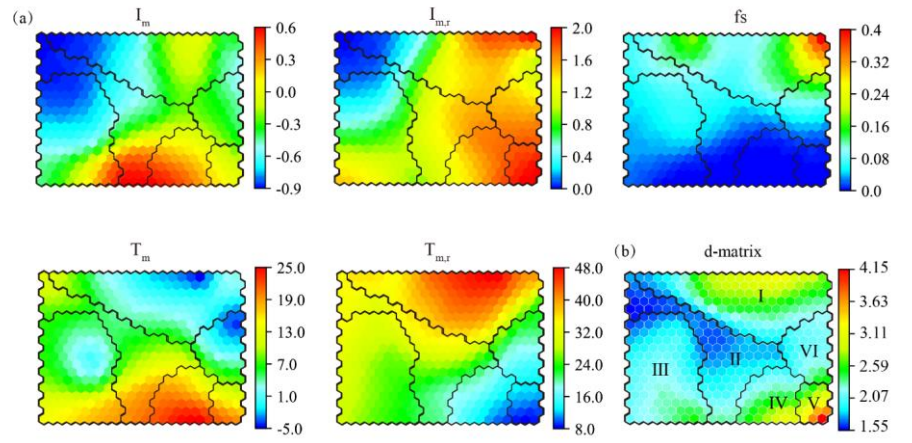
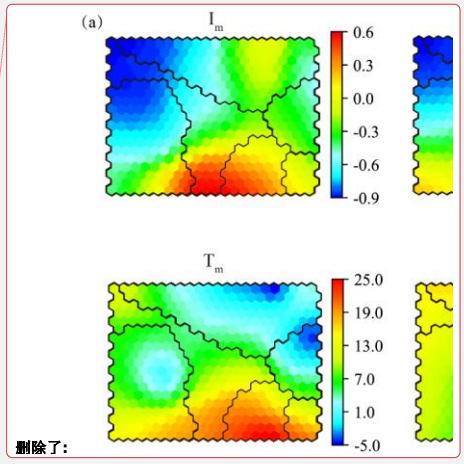


Figure 3. Self-organising map representation of the hydroclimatic feature space and fuzzy climate-region partition: **(a)** SOM component planes for the five climate indices: each hexagon represents a neuron summarising grid cells with similar standardised index values, colours show the corresponding 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.



删除了:

删除了: 2

删除了: .

删除了: Results of

设置了格式: 字体: (中文) +中文正文 (宋体), (中文) 简体中文(中国大陆)

设置了格式: 字体: (中文) +中文正文 (宋体), 非加粗, (中文) 简体中文(中国大陆)

设置了格式: 字体: 加粗

带格式的: 行距: 单倍行距

设置了格式: 字体: 加粗

删除了: SOM clustering: (a) component planes showing weight vector values of climate variables, and (b) clustering distance matrix (d-matrix).

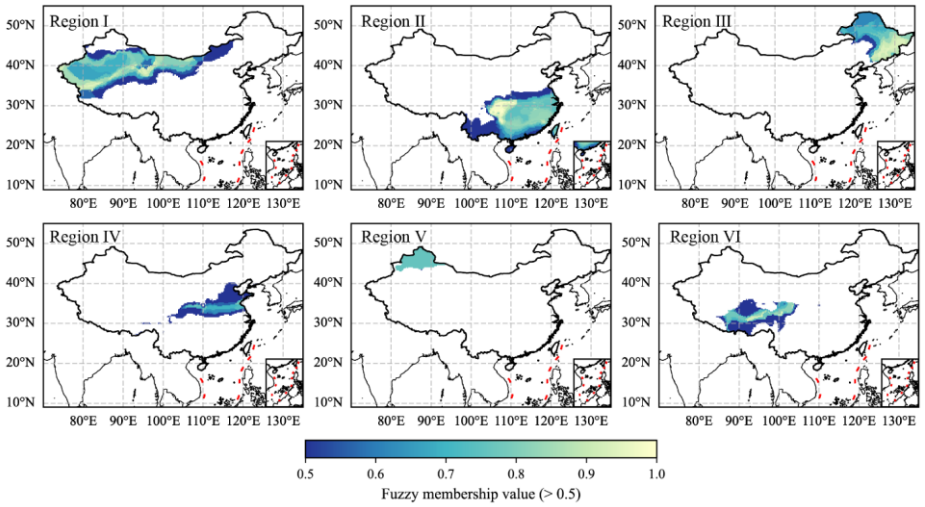


Figure 4. Spatial distribution of fuzzy membership to the six SOM-FCM climate regions across China; only grid cells with membership values greater than 0.5 are shown.

3.2 Catchment types within climate regions

3.2.1 Correlation structure of catchment descriptors

Within each climate region, we first examined how catchment descriptors covary to assess redundancy and to guide dimensionality reduction. Spearman rank correlation were computed for all pairs of descriptors in each climate region, and the resulting regional correlation patterns were compared with the national-scale correlations (Fig. 5). Descriptors belonging to the same class generally showed strong correlations (e.g. β and TI, both topographic characteristics, correlation coefficient >0.9), whereas descriptors from different classes were only weakly correlated (e.g. β and L, which belong to topographic characteristics and topological characteristics, respectively, with a correlation coefficient <0.1). Correlation structures were broadly similar across climate regions (Fig. A2), for more than 80 % of descriptor pairs, the difference between regional and national correlation coefficients was < 0.4. Some notable regional deviations occurred, particularly for the relationship between $NDVI$ and H. At the national scale, vegetation cover tends to decrease with elevation; but in the high-elevation Tibetan Plateau (Region Ⅵ) NDVI and H are positively correlated (correlation coefficient =0.7), whereas in lowland plains (Regions Ⅱ and Ⅴ), the correlation is weak or negative. This pattern reflects the stronger control of elevation on

设置了格式: 字体: (中文) Times New Roman, 小五, 加粗
带格式的: 正文, 行距: 单倍行距
设置了格式: 字体: (中文) +中文正文 (宋体), 小五, 加粗, (中文) 简体中文(中国大陆)
删除了: 3.1.3 FCM clustering results
Before clustering the output neurons from the SOM competitive layer using the FCM algorithm, two validation metrics were calculated (i.e., DBI and SC) to determine the optimal number of clusters. An experiment was conducted to test the number of clusters from two to the maximum number determined by the AP algorithm. The optimal number of clusters was theoretically determined by minimizing DBI and maximizing SC. They showed the best values when the number was up to six, and then DBI increased with an increase in the number of clusters, and SC changed slowly. Therefore, we defined six representative climates in continuous climate index space. The spatial distribution of these six climate groups in China is shown in Fig.3, where blue and red ...
删除了: Results of catchment classification
删除了: Correlations of catchment attributes in the different ...
删除了: The catchments in terms of climate homogeneity were ...
删除了: 's
删除了: coefficients between catchment attributes across climate ...
删除了: similar types of catchment feature indicators have high ...
删除了: (
删除了: belonging to
删除了: different types of features have low correlation
删除了: (
删除了: Additionally,
删除了: catchment attribute correlations varied little between ...
删除了: S2
删除了: and the Spearman rank correlation coefficient for most ...
删除了: Nonetheless, some catchment attributes differed by up to ...
删除了: were the most noticeable
删除了: From a
删除了: perspective
删除了: vegetation and elevation are negatively correlated
删除了: as elevation increases, vegetation becomes sparser. There ...
删除了: while in the
删除了: r
删除了: 2
删除了: 5
删除了: ,
删除了: direction of
删除了: shifted
删除了: is because

temperature, and hence vegetation growth, in mountainous areas, while in plains temperature varies mainly with latitude. Overall, despite such local differences, the descriptors exhibit broadly comparable correlation structures across climate regions.

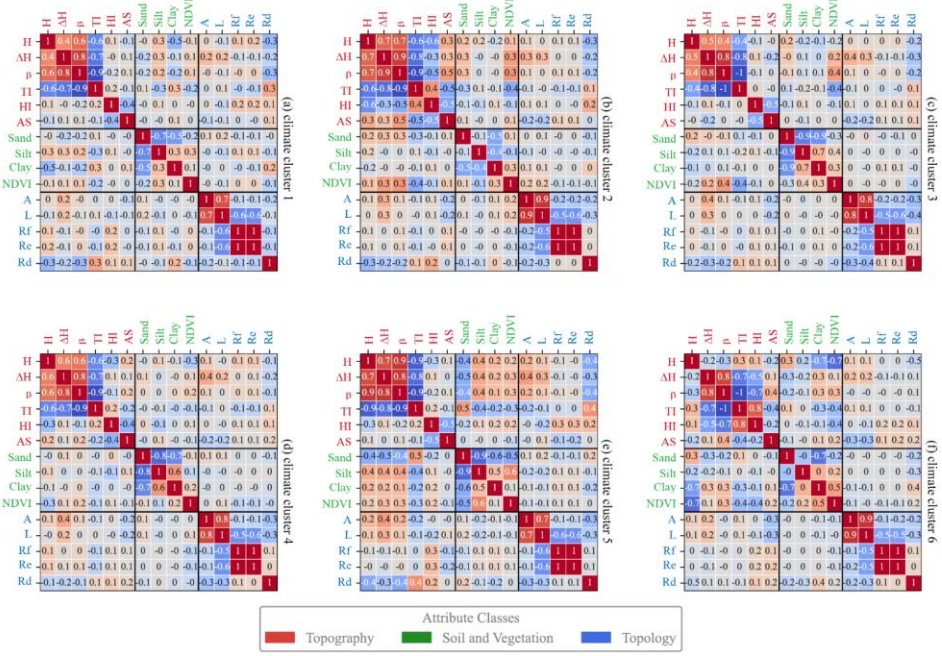


Figure 5. Spearman rank correlation coefficients for catchment attributes across different climate regions.

To reduce redundancy, PCA was performed for each of the three types of catchment features to remove correlations between individual signatures of the same type. The results showed that catchments within the same climate region exhibited widely disparate topographic and topological characteristics in the principal component space (Fig. 6), implying the necessity for identifying similar catchments in the climate regions. Eigenvalues and explained-variance ratios are reported in Table 3. 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 %

删除了: often varies significantly with elevation in highland areas

删除了: thus,

删除了:

删除了: is influenced by temperature gradient

删除了: . In contrast, temperature in plain areas is more closely tied to latitude, and therefore, vegetation is less correlated to elevation

删除了: Considering that region 6 is located on the Tibetan Plateau and is characterized by complex topography, the correlation between part of the elevation and topographic features is inverse to that of the other five regions (for instance, elevation range and slope do not increase significantly as elevation increases). Although the correlation coefficient has some local differences across climate regions, they still have comparable relationships between catchment attributes.

删除了: 4

删除了: eliminate

删除了: a principal component analysis

删除了: 5

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.

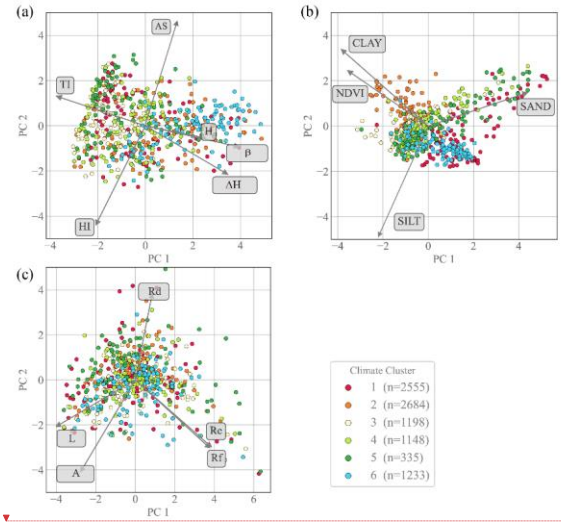


Figure 6. Biplots of the principal components: **(a)** topographic characteristics, **(b)** soil and vegetation, and **(c)** topological characteristics. Colors represent the climate clusters, while n indicates the number of catchments within each climate region.

Table 3. Eigenvalues and proportions of principal components for three types of landscape characteristics.

Descriptor classes	Component	Eigenvalue	Proportion	Cumulative Proportion
Topographic characteristics	1	3.24	54.02	54.02
	2	1.17	19.44	73.46
	3	0.68	11.37	84.83
	4	0.54	8.92	93.75
	5	0.31	5.14	98.89
	6	0.07	1.11	100.00
Soil and vegetation characteristics	1	2.31	57.71	57.71
	2	0.91	22.83	80.54
	3	0.54	13.43	93.98
	4	0.24	6.02	100.00
Topological characteristics	1	1.94	38.89	38.89
	2	1.71	34.22	73.11
	3	0.96	19.23	92.34
	4	0.26	5.15	97.49

删除了： Additionally, the soil and vegetation properties of the catchments in the same climate region were similar. While most catchments were characterized by a high percentage of sand and poor vegetation cover in regions 1 and 6, the catchments in region 2 had a high percentage of clay and dense vegetation cover. The remaining catchments fell between these two extremes. Each catchment feature type was analyzed using principal component analysis, and the principal component eigenvalues and proportions are shown in Table 2. Principal components were determined by eigenvalues greater than 1. As a result, two principal components were selected for each type of feature, which encompassed the most information in which the cumulative proportion exceeded 70%. In terms of topographic attributes, the first principal component had the highest correlation with the elevation and slope. The second most strongly correlated with AS; and the first principal component of soil and vegetation characteristics had the highest correlation with sand content. The second correlated most strongly with $NDVI$; the first principal component of topological characteristics had a strong correlation with Re and Rf, while the second had the highest correlation with river network density.

删除了： 5

设置了格式： 字体: 加粗

设置了格式： 字体: 加粗

设置了格式： 字体: 加粗

删除了： 2

5	0.13	2.51	100.00

3.2.2 Catchment types and spatial organization

Within the six climate regions, the SOM-FCM procedure partitions the 13 487 catchments into 35 landscape-based clusters, with 7, 5, 6, 6, 5 and 8 clusters in Regions I-VI, respectively. Each catchment has fuzzy membership values to all clusters; for subsequent analyses, catchments with a maximum membership ≥0.5 are treated as clearly assigned to a dominant type (Schwämmle and Jensen, 2010), whereas lower maximum memberships indicate transitional behaviour between clusters. The proportion of clearly assigned catchments differs among climate regions (e.g. a relatively large share of transitional basins in Region VI versus predominantly well-defined types in Region V), reflecting contrasts in landscape complexity. Typical attributes and spatial distribution patterns of all clusters are summarized in Table 4 and Fig. 7, with additional descriptor boxplots in Fig. A3.

In Region I (northwestern arid belt), clusters separate low-relief basins with fine soils in the northern forelands from strongly dissected mountain catchments with large elevation ranges and steep slopes in the west, and from very flat, sand-dominated basins with sparse vegetation in the Tarim Basin. Together, these types capture a gradient from highly developed mountainous terrain to weakly developed desert plains.

In Region II (humid monsoon belt of southeastern China), clusters distinguish low-elevation, clay-rich and densely vegetated hilly catchments in the southeast and southwest from gently sloping alluvial basins with high hypsometric integrals along the middle-lower Yangtze, and from steeper, geomorphically more developed mountain basins in the Hengduan Mountains. Differences are primarily expressed in relief, soil texture and vegetation cover, while overall elevation remains moderate.

In Region III (cold, snow-influenced Northeast China), clusters separate low-slope basins with coarse soils and high topographic index values in the south from steeper, silt-rich mountain catchments in the Greater Khingan and Changbai ranges, and from more weakly drained lowland basins with low drainage density or clay-rich soils. The resulting types reflect contrasts between flat, accumulation-prone landscapes and more dissected uplands.

In Region IV (North China Plain and surrounding mountains), clusters differentiate steep, low-drainage-density basins in the western Taihang Mountains from gently sloping lowland basins with high hypsometric integrals along the lower Yellow River, and from fine-textured, silt-rich plains east of the Taihang. Additional clusters represent intermediate upland types with moderate hypsometric gradients and varying sand content and drainage density, highlighting a pronounced topographic and soil-texture gradient from mountains to plains.

In Region V (Junggar-Tianshan region), one cluster is associated with high-elevation mountain catchments in the Tianshan and Altay ranges, combining steep slopes, high relief and dense vegetation with relatively low drainage density. The remaining clusters occupy the Junggar Basin and share broadly similar low-relief morphologies, but differ in hypsometric gradient, drainage density, basin size and soil texture, capturing subtle yet systematic variations in the organisation of the lowland drainage network.

删除了：3.2.2 Catchment classification results using SOM-FCM algorithm

The SOM-FCM combined algorithm was used to cluster 13,487 catchments within six climate regions separately in China, and catchments were divided into 35 classifications. The membership values of the catchments within meteorologically homogeneous zones can be determined for various surface patterns. However, some catchment properties are spatially continuous, so catchments can belong to more than one mode. A membership threshold of 0.5 is generally interpreted as the cell belonging exclusively to its main cluster (Schwämmle and Jensen, 2010). Therefore, only membership values above 0.5 are regarded as belonging to a certain cluster, and below 0.5 are considered an indication that the catchment does not belong to a significant cluster and may have overlapping hydrological properties between multiple groups. The proportion of catchments with a single main cluster varied across meteorological regions. There were 44 percent of catchments in climate region 6 with insignificant main clusters and more complicated basin conditions, whereas 78 percent of catchments in meteorological region 5 had a delineated main cluster and low uncertainty in catchment mode. Catchment clusters are discussed in terms of their characteristics (Fig. S3) and location (Fig. 6); the crucial components of this description are presented in Table 3.

The catchments in climate region I were divided into seven classifications. Clusters 1 and 5 were mainly in the northern climate region 1 with relatively low elevation and gentle topography, defined by a high clay fraction and high hypsometric curve integral, respectively. Nevertheless, the soil texture in cluster 5 was coarser than that in cluster 1. Cluster 3 covered the majority of the west side of the climate region and was characterized by a high elevation range. This region showed high elevation and slope, low hypsometric curve integral, and low level of geomorphic development. The catchments in cluster 6 were in the Tarim Basin. They were characterized by a high sand fraction, flat terrain, low elevation, and sparse vegetation. Most of cluster 7 was located in the southern region. They possessed a relatively high level of geomorphic development, high silt fraction, and high elevation. Cluster 2, like cluster 4, was dispersed in space with a low drainage density. The two clusters had similar topography and soil characteristics, but the catchments in cluster 2 had larger areas and longer rivers.

There were five clusters in climate region II. Cluster 1 encompassed hilly areas in the southeast and southwest of climate region 2. The catchments were defined by high *NDVI* and had relatively low terrain and elevation, high clay fraction, and abundant vegetation. Cluster 2 consisted of catchments located in the Sichuan Basin and Lower Yangtze Coast. They were distinguished by a high gradient of hypsometric curves and have low elevation, moderate slope, and high silt fraction. Cluster 3, indicated by a moderate elevation range, demonstrated dispersed spatial distribution. The catchments here were very similar in their geographical characteristics to cluster 1, but cluster 3 shows a coarser soil texture. The majority of the catchments in cluster 4 were concentrated in the middle and lower reaches of the Yangtze River. The region had a gentle topography, with high hypsometric curve integra and low gradient of hypsometric curve, and the degree of landform development was modest. Cluster 5, characterized by high elevation, was mostly located in the Hengduan Mountains, and the catchments had high elevation and slope, as well as a high degree of geomorphic development.

The catchments in climate region III were grouped into six clusters. Cluster 1 was concentrated in the southern climatic region and was defined by a high topographic index. The catchments had an undulating topography, low slope degree, and coarse soil texture. Cluster 4 was identified in the Greater Khingan Range and Changb ...

In Region VI (Tibetan Plateau), clusters separate high-elevation basins with limited hypsometric development from lower-elevation eastern and southern basins with finer soils, higher NDVI and moderate drainage density, and from elongated, low-drainage-density catchments with large areas and long main channels. The relatively large share of transitional catchments (low maximum membership) in this region reflects the complex, multi-scale variability of plateau topography and surface conditions.

Across all climate regions, the resulting catchment types represent coherent combinations of relief, soil texture, vegetation cover and drainage structure. This physically interpretable typology provides the basis for the subsequent hydrological validation (Sect. 3.3).

Table 4. Properties of catchment clusters within climate regions. "Typical attribute" and "Second attribute" refer to the attributes of the cluster with the lowest and second lowest coefficient of variation, which were scaled by the mean coefficient of variation of the dataset.

Climate region	Cluster	Catchment numbers	Typical attribute	Second attribute
I	I-1	662	High $Clay$	High TI
	I-2	568	Low Rd	High A
	I-3	651	high ΔH	High H
	I-4	361	Low Rd	High HI
	I-5	365	High HI	High AS
	I-6	605	High $Sand$	High AS
	I-7	578	High $Silt$	High HI
II	II-1	1110	High $NDVI$	High $Clay$
	II-2	615	High AS	Low $Clay$
	II-3	661	Mid ΔH	Mid Rd
	II-4	690	High HI	Mid Rd
	II-5	825	High H	Mid $Sand$
III	III-1	218	High TI	High AS
	III-2	403	Low Rd	High $Clay$
	III-3	538	High $Clay$	High $Silt$
	III-4	603	High $Silt$	High $NDVI$
IV	IV-1	234	Low Rd	High ΔH
	IV-2	211	High HI	Mid Rd
	IV-3	256	High $Silt$	High TI
	IV-4	221	Mid AS	Mid H
	IV-5	388	High β	High H
	IV-6	226	Low Rd	High A
V	V-1	105	High AS	Mid $Silt$
	V-2	69	Mid Rd	High A
	V-3	129	Low Rd	High ΔH
	V-4	93	High TI	High Rd
	V-5	102	High $Silt$	High TI
VI	VI-1	131	High $Silt$	Mid $NDVI$
	VI-2	150	High $NDVI$	Mid Rf
	VI-3	197	Low Rd	High H
	VI-4	247	High H	High $Sand$

删除了:

VI-5	317	High H	High TI
VI-6	143	Mid Rd	High β
VI-7	322	High $NDVI$	High $Silt$
VI-8	171	Low Rd	Mid $Silt$

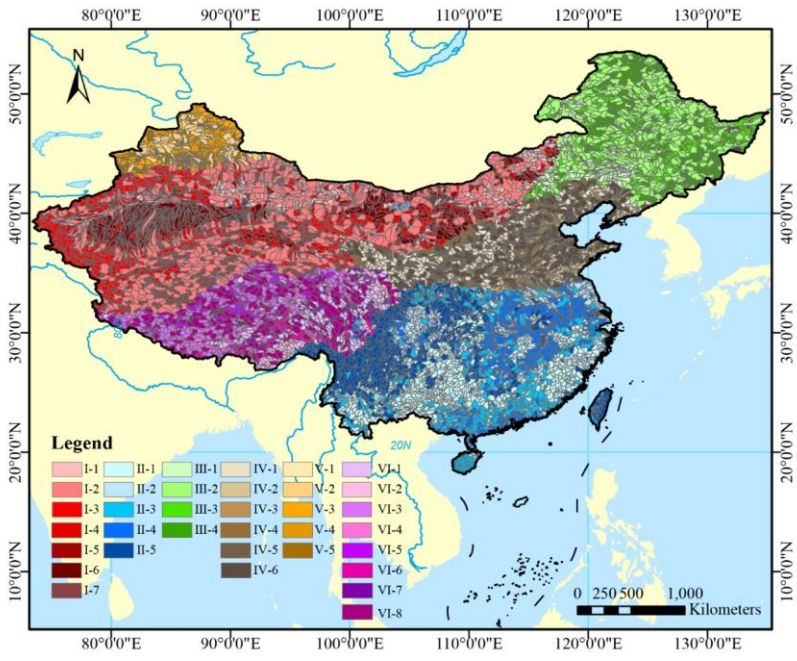


Figure 7. Spatial distribution of the climate-landscape catchment classes across China derived from the hierarchical SOM-FCM classification.

3.3 Hydrological validation of the classification

3.3.1 Process-based validation across seasonal and event scales

The process-based evaluation focuses on the ten small headwater catchments, which represent a range of climate-landscape classes and are mainly located in climate regions II and IV (Table 5).

删除了: ...

删除了: 6

删除了: Map of the derived catchment classification across China

删除了: Validation results for small catchments in China

带格式的: 标题 3

删除了: Small catchments from various clusters were considered to validate the hydrological similarity with catchment classification. The catchments selected for this study were based on the following criteria: (1) available hydrological data, (2) unregulated catchments, and (3) containing a range of climate types and catchment classes(Li et al., 2018). Each catchment had 10–15 years of available daily continuous rainfall and runoff data with 10-35 flood events. The catchment areas range from 441 to 4,321 km^2, and the climate types span different regions containing various small catchment classes. The period of the data record and small catchment classes for each catchment are listed, and the catchments are primarily located in climate regions II and IV, which contain a variety of small basins (Table 4).

Table 5. Information and characteristics of the study catchments. Basin class indicates the class types of the subbasins in catchment.

Station	River System	Area(km^2)	Number of flood events	Record period	Basin classes
Fenshuijiang	Qiantang River	2619	10	2003-2014	II-1
Tunxi	Qiantang River	2680	35	2008-2017	II-1
Chenhe	Yellow River	1429	18	2003-2012	II-5
Daheba	Yangtze River	2198	13	2013-2017	II-5
Banqiao	Yangtze River	441	13	2000-2010	IV-4
Suide	Yellow River	3897	22	2010-2017	IV-4
Daiying	Hai River	4321	16	1990-2002	IV-6
Maduwang	Yellow River	1605	12	2000-2010	IV-6
Dage	Hai River	1864	17	1990-2008	IV-4
Zhidan	Yellow River	779	15	2000-2010	IV-1

Seasonal hydrographs indicate that catchments assigned to the same climate region exhibit broadly similar seasonal runoff regimes, despite differences in landscape properties (Fig. 8). 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). In contrast, catchments in climate region IV show lower overall runoff and weaker interannual variability, with a single, delayed peak after July that reflects both smaller precipitation totals and a clearer lag between rainfall and runoff (e.g. Daiying and Dage). These patterns support the notion that medium- to long-term hydrological signatures, such as seasonal runoff, are primarily controlled by the relative seasonality of precipitation and potential evaporation and are therefore more strongly linked to climate region than to local landscape characteristics.

删除了: 4

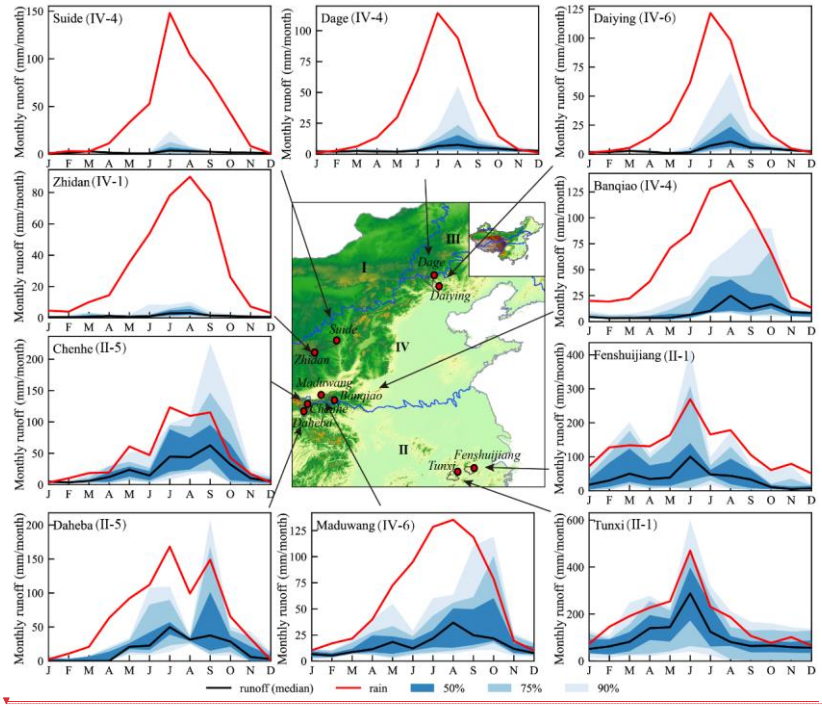


Figure 8. Seasonal runoff of the ten gauged catchments used for process-based validation. Monthly precipitation (red) and median monthly runoff (black) are shown, with shaded bands indicating the central 50 %, 75 %, and 90 % ranges of monthly runoff across years.

Event-scale flow-duration curves (FDCs) derived from hourly discharge during individual flood events reveal additional structure in runoff response that is closely related to the landscape classes (Fig. 9). The shapes of the event-based FDCs allow the ten catchments to be grouped into five hydrologically homogeneous response types: (1) Tunxi and Fenshuijiang, (2) Chenhe and Daheba, (3) Daiying and Maduwang, (4) Dage, Suide and Banqiao, and (5) Zhidan. Tunxi and Fenshuijiang (class II-1) are low-elevation, fine-soil, densely vegetated basins in a humid monsoon climate. Their FDCs are relatively flat, with high discharges sustained over a wide range of exceedance probabilities and gentle lower tails, indicating buffered event responses with substantial groundwater and subsurface contributions. Chenhe and Daheba (class II-5) drain steeper terrain at higher elevations; their FDCs show high event flows that are only slightly lower than those of class II-1 but still relatively gradual recessions, suggesting strong, but not excessively flashy, storm responses underlain by considerable storage. Zhidan (class IV-1), by contrast, represents a highly dissected, coarse-soil upland basin with low precipitation in climate region IV. Its FDC lies well below those of the other catchments over most of the exceedance range and is much steeper, with event flows dropping

删除了：...

删除了：7

删除了：Regional

删除了：differences in seasonality of precipitation and runoff.

rapidly from moderate peaks to very low discharges. This pattern indicates the dominance of fast surface-runoff pathways, limited event water storage and a weak baseflow component. The remaining catchments occupy intermediate positions between these end-members, with FDC shapes that reflect progressive changes in relief, soil texture, and drainage density.

Taken together, the seasonal hydrographs and event-based FDCs demonstrate that the proposed climate-landscape classification captures hydrological similarity at multiple temporal scales. 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. 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. This process-based validation therefore supports the use of the combined climate and landscape classification as a meaningful basis for regional analyses of hydrological behaviour and for model regionalisation.

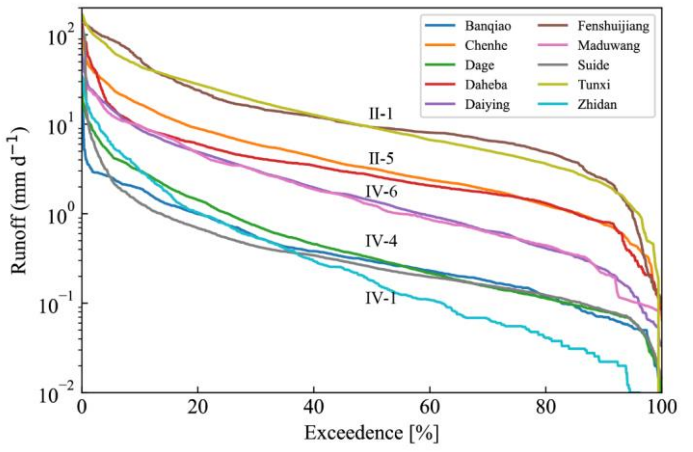


Figure 9. Event-scale flow-duration curves of the ten gauged catchments.

3.3.2 Statistical validation based on flow signatures

The 722 selected headwater catchments used for the large-sample validation are distributed across all major climate regions; their locations and assigned climate-landscape classes are shown in Fig. 10. These basins span 35 climate-landscape classes defined in Sect. 3.2. However, within climate region V only class V-3 contains a sufficient number of catchments to support statistically meaningful analysis; the remaining classes in this region were therefore not evaluated because of limited sample size, with fewer than five basins per class. The flow-signature analysis thus covers 31 of these 35 classes. The distributions of the 13 flow signatures (FS) for the 31 classes are shown in Fig. 11. For most signatures, median values and interquartile ranges

删除了： The flow duration curve (FDC) is a valuable tool for diagnosing rainfall-runoff responses in gauged catchments at a holistic level. The hourly FDC in flood event periods from different catchments located across several cluster regions showed significant differences in the runoff regime (Fig. 8), illustrating the classification of the 10 catchments into five homogeneous regions: (1) Tunxi, Fenshuijiang, (2) Chenhe, Daheba, (3) Daiying, Maduwang, (4) Dage, Suide, Banqiao, and (5) Zhidan. Among them, the Tunxi and Fenshuijiang catchments belong to II-1, which has low elevation, fine soil, and dense vegetation. These areas experience relatively stable humid climates with abundant precipitation, resulting in similar surface runoff and stable groundwater flow processes.. The FDC curves represent a gentle flow, with high flow values obtained in the upper portion, and stable and gentle flow in the lower flow portion. The Chenhe River and Daheba catchments belong to category II-5, characterized by relatively high elevations and steep terrain. These areas experience significant high flows during precipitation periods, with the flow becoming more gradual as the exceedance percentage increases, indicating higher base flow levels. Zhidan is located in IV-1, which is distinguished by a hazardous terrain with a high degree of geomorphic development, high elevation, and coarse soils. This climate is reflected in seasonal climate changes and low precipitation. The FDC in Zhidan indicated a steeper curve with rapid regional surface runoff as the major flow, but essentially no subsurface runoff activity. FDCs in the other catchments showed a situation between these two states. These results indicate that catchment classification based on climate and landscape can reflect the performance of runoff characteristics and differences. Overall, climate appears to be the most important factor for medium- or long-term hydrological signatures, such as seasonal runoff, whereas landscape features are more essential to hydrological features at the flood event scale. The results and comparison show that the hydrological similarity of the catchment is more difficult to identify than the spatial proximity. Using spatial proximity may prove effective only in areas where hydrological behavior gradually changes. However, in regions with high spatial variability, spatial proximity cannot be used to establish hydrological similarities (Knoben et al., 2018). Additionally, catchments with large spatial distances are capable of exhibiting similar hydrological characteristics (Maduwang and Daiying, year). Hydrological complexity is reflected in various climates and individual signatures. The combined indicators appear to depict a dynamic that is climatic in origin, but are influenced by catchment characteristics (Berghuijs et al., 2014b; Jehn et al., 2020). The validation experiment demonstrated that the hydrological similarity of catchments at different scales can be reflected through ...

带格式的： 缩进: 左 0 字符

differ markedly between classes, whereas within-class variability is comparatively small, indicating that catchments assigned to the same class tend to share similar flow regimes. Magnitude-related signatures provide the clearest separation, with Qsp being the most discriminative in terms of the proportion of significantly different class pairs, followed by Q95 and CVQ. By contrast, duration-related signatures show the weakest response to the classification. In particular, Mean30dMax exhibits nearly uniform distributions with only minor shifts in median values, suggesting that prolonged high-flow episodes are less sensitive to the combined climate-landscape controls captured by our classification than other aspects of the flow regime.

While Fig. 11 illustrates the magnitude of between-class contrasts for each signature, Fig. 12 summarises their statistical significance across all class pairs. Fig. 12a shows that the climate-landscape classes are clearly distinguishable in terms of most flow signatures: for the majority of indices, more than half of all class pairs differ significantly in their mean values, even at the stricter threshold of $p < 0.05$. Signatures related to flow magnitude (Qsp, CVQ, Q9) and to the frequency of high flows (HFD, HighFrVar) achieve the highest discrimination, with a large proportion of significantly different class pairs. By contrast, skew and several low-flow or flashiness metrics (e.g. LowFr, RBF) display a smaller, but still substantial, fraction of significant contrasts. The small difference between the $p < 0.05$ and $p < 0.10$ bars for most signatures suggests that the main conclusions are robust and not overly sensitive to the chosen significance level. Although no formal correction for multiple comparisons was applied, the consistently high fractions of significant class pairs across related signatures and the coherent patterns in Fig. 10 indicate that the overall discrimination is unlikely to be an artefact of random sampling variability.

The Games-Howell matrices for Qsp, BFI, and Q95 in Fig. 12b further illustrate how hydrological behaviour varies among individual climate-landscape classes. For Qsp, significant differences (dark cells) are widespread, especially between classes belonging to contrasting climate regions (e.g. arid versus humid or snow-dominated environments), providing strong evidence that the classification separates basins with distinct overall runoff production. The BFI matrix exhibits a more clustered pattern, with strong contrasts between climate regions and between clearly different types of basins, whereas several neighbouring classes within the same region show statistically indistinguishable baseflow behaviour, consistent with the smoother spatial variation expected for groundwater-dominated processes. For Q95, the spatial pattern of significant differences closely resembles that of Qsp but emphasises high-flow conditions and highlights pronounced contrasts in flood response among the catchment classes. These patterns align with the anticipated influence of aridity, snow fraction, relief, and soil texture on runoff generation, storage, and release, and they support the hydrological relevance of the proposed climate-landscape classification.

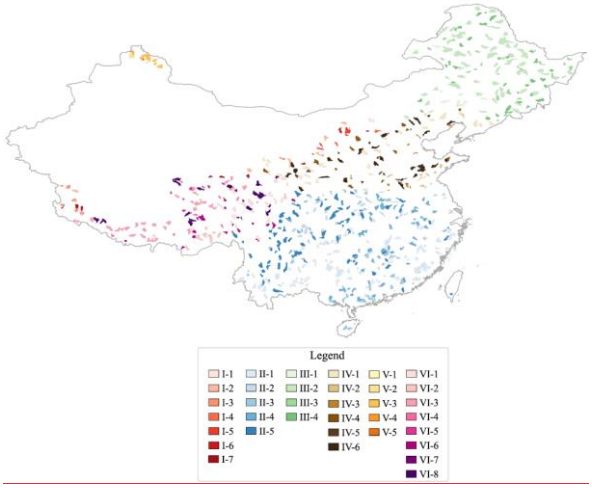


Figure 10. Spatial distribution of the 722 validation catchments used for flow-signature-based validation of the climate-landscape classification.

设置了格式: 字体: 小五, 加粗

设置了格式: 字体: 小五

设置了格式: 字体: 小五, 加粗

设置了格式: 字体: 小五

设置了格式: 字体: 小五, 加粗

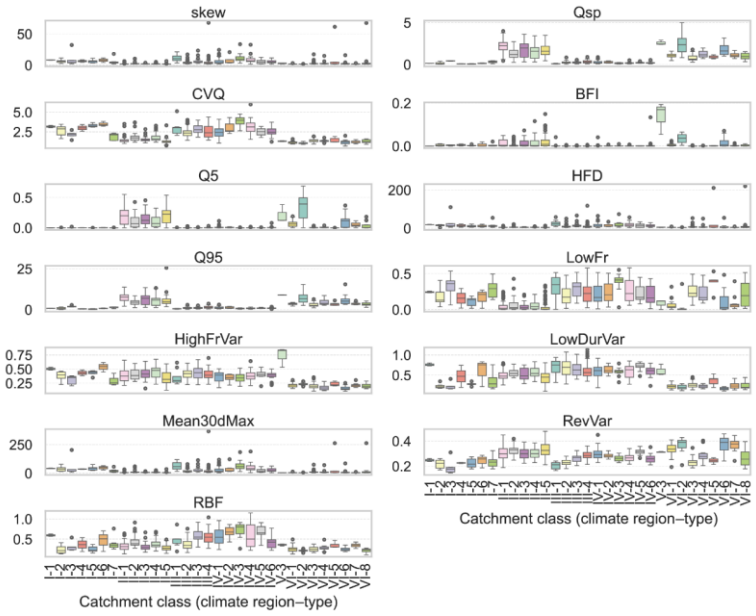


Figure 11. Distributions of 13 flow signatures for the catchment classes.

设置了格式: 字体: 小五, 加粗

设置了格式: 字体: 小五, 加粗

设置了格式: 字体: 小五

设置了格式: 字体: 小五, 加粗

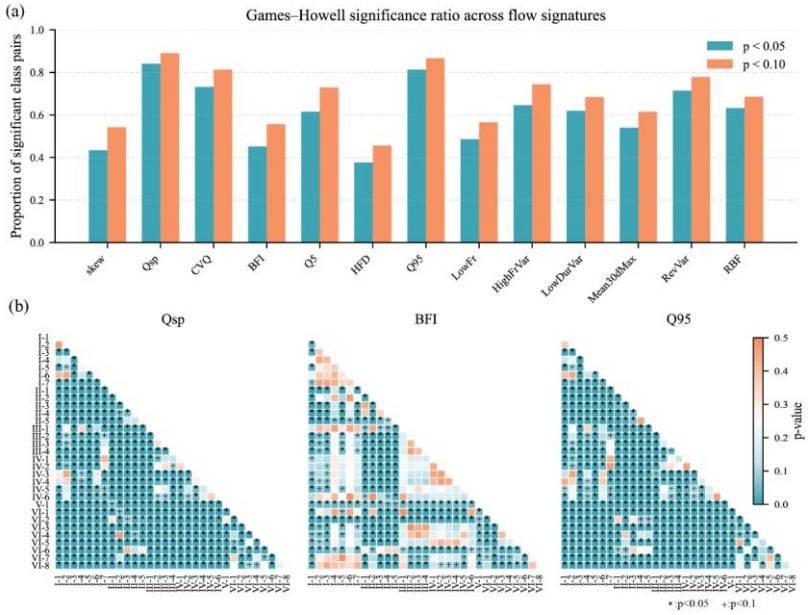


Figure 12. Validation of catchment classification using hydrological flow signatures. **(a)** Proportion of significantly different class pairs ($p < 0.05$ and $p < 0.10$) for 13 flow signatures based on the Games-Howell test. **(b)** Pairwise Games-Howell p-value matrices for selected flow signatures (Qsp, BFI, and Q95), showing the degree of hydrological distinctiveness among catchment classes.

带格式的：正文, 缩进: 左 0.9 字符

4 Discussion

删除了：Figure 8. Flow duration curves for the test catchments.

4.1 Climate-landscape controls on hydrological behaviour

删除了：The necessity and boundary effects of dividing meteorologically homogeneous zones

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).

4.2 Implications for modelling and prediction in ungauged basins

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

删除了： (Coopersmith et al., 2012; Ghotbi et al., 2020a; Yaeger et al., 2012)

删除了： Dividing small catchments based on meteorologically homogeneous zones involves grouping regions with similar climatic characteristics to understand hydrological processes. This approach is based on the premise that climate, as the primary driver of hydrological behavior, significantly influences water resource availability, flow patterns, and overall watershed dynamics (Kuentz et al., 2017; Zhang et al., 2022). We believe that climatic variables such as precipitation and temperature play a critical role in shaping hydrological processes. Differences in climatic conditions lead to significant variations in the hydrological responses of watersheds (Addor et al., 2018; West et al., 2022; Wu et al., 2020). Studies have demonstrated that regionalization techniques are more effective when applied to areas with similar climatic conditions (Bharath and Srinivas, 2015; Hazarika and Sarma, 2021; Samantaray et al., 2021). Classifying small catchments with similar climatic conditions enables the development of more accurate regional hydrological models, thereby improving water resource management and planning (Liu et al., 2021; Pagliero et al., 2019). However, the application of meteorologically homogeneous zones encounters boundary effect issues in practice. This issue arises because climatic variables, such as temperature and precipitation, generally change gradually across space. The introduction of artificial boundaries can result in inaccurate classifications, particularly at the edges of these zones. Although climate change is continuous (Viviroli et al., 2011), creating abrupt boundaries can disrupt natural gradients, potentially leading to errors in hydrological modelling. An effective solution to this issue is the use of fuzzy clustering techniques, which allow each catchment to belong to multiple clusters with varying degrees of membership (Bharath and Srinivas, 2015; Sreeparvathy and Srinivas, 2022). Some studies have introduced transition zones or buffer areas between clusters to create gradients in climatic and hydrological characteristics, thereby reducing the impact of abrupt boundaries (Cantidio and Souza, 2019; Wen et al., 2017). Hierarchical clustering methods also offer a multi-scale approach, starting with broader climatic zones and then making more detailed subdivisions, thus maintaining the continuity of climatic variables at various spatial resolutions (Sreeparvathy and Srinivas, 2022). Finally, fuzzy clustering was adopted to preserve the continuity of climatic variables, providing a more accurate representation of hydrological regions. This approach resulted in well-classified similar catchments.

删除了： **The complexity of catchment hydrological behavior**

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 (Prieto et al., 2019). 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.

4.3 Methodological limitations and future research directions

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

删除了: (Qi et al., 2022; Gupta et al., 2014; Guo et al., 2021)

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.

5 Conclusions

We developed and hydrologically evaluated an integrated climate-landscape classification for 13 487 catchments across China using a hierarchical SOM-FCM framework. Six hydroclimatic indices (moisture, temperature, snow) and 15 geomorphological and drainage-network descriptors were used to delineate six climate regions and 35 landscape-based catchment classes. The resulting typology is physically interpretable and captures major gradients in aridity, seasonality, snow influence, relief, soils, vegetation, and drainage organisation.

Two complementary evaluations indicate that the classification is hydrologically meaningful. Analyses of ten gauged headwater catchments and 13 flow signatures from 722 GRFR basins show coherent within-class behaviour and systematic between-class contrasts. Seasonal hydrographs primarily separate climate regions, reflecting differences in the magnitude and timing of precipitation relative to potential evapotranspiration. By contrast, event-scale flow duration curves and high-flow signatures (e.g. Qsp, Q95, HFD, HighFrVar) provide added discrimination among landscape classes, highlighting the role of relief, soils, vegetation, and network structure in shaping flood frequency and intensity and the balance between fast and slow runoff components. Overall, similarity in hydrological behaviour aligns more closely with climate – landscape classes than with geographic proximity, and cannot be represented adequately by climate or catchment structure alone.

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. Because the required indices and descriptors are derived from widely available datasets, the framework is transferable and supports comparative hydrology, model benchmarking, hydrometric network design, and data-driven prediction conditioned on physically interpretable classes.

Limitations remain. Results depend on the accuracy of gridded climate fields, elevation- and soil-based descriptors, and GRFR-derived discharge estimates, and the process-based evaluation uses a limited set of gauges that does not cover all classes. The framework is static and focuses on near-natural basins; human influences (e.g. reservoir operations, irrigation, and urbanisation)

删除了： Watersheds are complex systems resulting from the interaction of climatic and landscape processes, leading to the co-evolution of hydrological processes (Addor et al., 2018). Therefore, identifying the primary drivers of changes in hydrological response is challenging (Jehn et al., 2020). This watershed study encompasses mountainous, plateau, and plain regions but does not account for the impact of human activities. Due to data limitations, hydrological characteristics were not used as a zoning factor; instead, the study was based on a fundamental assumption: watersheds with similar climatic and basin characteristics exhibit comparable hydrological behavior (Dallaire et al., 2019; Jehn et al., 2020). This assumption is based on the premise that hydrological behavior in watersheds is largely influenced by climatic factors (e.g., precipitation and evaporation) and basin characteristics (e.g., topography, soil type, and land use). For example, Jehn et al. (2020) pointed out that watersheds under different climatic conditions typically exhibit different hydrological responses because precipitation and evaporation patterns directly determine the inputs and outputs of the hydrological cycle. Additionally, topography and soil type affect water infiltration and runoff pathways, thereby influencing the hydrological response of the watershed (Kuentz et al., 2017). On the other hand, clustering analyses using catchment attributes and flow signatures showed significant consistency in clustering results, further demonstrating the significant control catchment attributes have on hydrological behavior (Du et al., 2023).
the hydrological behavior of a watershed is not determined solely by external factors like climate and topography, but also by complex interactions within internal hydrological processes (Mcdonnell et al., 2007). Co-evolutionary processes within a watershed, such as dynamic changes in vegetation, soil, and geomorphology, can lead to watersheds with similar climatic and topographical features exhibiting different hydrological behaviors (Bogaart et al., 2016). Second, anthropogenic factors such as land-use changes and water resource management practices can significantly alter a watershed's hydrological behavior (Dwarakish and Ganasri, 2015). Human activities, such as urbanization and agriculture, can exert a greater influence on a watershed's hydrological processes than natural characteristics (Neupane and Kumar, 2015). Consequently, even watersheds with similar climatic and topographical features may exhibit significantly different hydrological behaviors due to human interference. Therefore, watershed hydrological behavior results from the interaction of multiple factors, necessitating a comprehensive analysis to accurately understand watershed similarities and predict hydrological responses.

and non-stationarity associated with climate and land-use change are not represented explicitly. Future work should incorporate indicators of human disturbance and additional signatures, explore time-varying classifications that track shifts in class membership, and couple the typology with process-based and data-driven models to test class-informed regionalisation in predictive applications.

Data availability. The observation-driven datasets analyzed in this study are publicly available as referenced within the article. Meteorological and land surface products datasets utilized in this study can be accessed through the following sources: Precipitation and Temperature datasets (http://data.cma.cn/), Potential Evapotranspiration dataset (https://www.ceda.ac.uk/), watershed boundary HydroSHEDS dataset (https://www.hydrosheds.org/page/overview), ASTER GDEMV2 digital elevation model (http://www.gscloud.cn), Soil and Vegetation characteristics (http://www.issas.ac.cn), Spot/vegetation NDVI dataset (https://www.resdc.cn), the Global Reach Level Flood Reanalysis dataset (https://www.reachhydro.org/home/records/grfr). The self-organizing map clustering methodology used in this study is available online (https://github.com/sevamoo/SOMPY). Python code for data computation, analysis, and graphical visualization can be obtained from the respective authors upon reasonable request.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgments. This study was supported by the National Key Research and Development Program of China (2023YFC3006505), Fundamental Research Funds for the Central Universities of China (B240203007), and the fund of National Key Laboratory of Water Disaster Prevention (524015222).

删除了： Hydrologically homogeneous catchments were clustered in China based using both SOM and FCM algorithms. As runoff is the result of the interaction between climate and catchment processes, an index system for climate and landscape characteristics was constructed. We assembled six climatic regions and 35 watershed classifications that fully reflect the regional hydrological characteristics of China. Furthermore, 10 catchments belonging to different classifications were selected to verify and analyze homogeneous regions. The results indicated that hydrological behavior is better characterized through climate and landscape characteristics in catchments. Moreover, climate-homogeneous regions respond to hydrological behaviors at medium- or long-time scales, whereas catchment classification regulates hydrological processes at the flood event scale. Combining the SOM and FCM algorithms provides a comprehensive quantitative evaluation of complex catchment structures. SOM enables complex, high-dimensional input data to be converted into intuitive 2D output surfaces. FCM utilizes membership values to address identifying catchments with fuzzy boundaries. 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.
The issue of flood simulation and forecasting in ungauged catchments has been a challenge owing to the lack of effective observational data. The development of a better hydrological similarity classification is critical for transferring hydrological model parameters and runoff simulations. In a high-dimensional heterogeneous feature space, the proposed method for identifying similar catchments at the basin scale can provide a guide for selecting similar basins. However, the hydrological behavior of catchments is not only determined by external characteristics such as climate and topography but also influenced by the complex interactions of internal hydrological processes, making it challenging to provide an in-depth description of catchment hydrological characteristics. Future research should focus on finer scales and consider including additional hydrological features, taking into account the impact of scale and feature selection on hydrological classification. Open data sources enable new regionalization studies, showing great potential for generating new knowledge and hydrological insights across various environmental conditions. For the Chinese region, however, there is a lack of homogeneous datasets on runoff characteristics and human impacts. Therefore, it is crucial to make more public sector data available and to construct standardized datasets for research purposes. This is of great importance for understanding the causes of catchment hydrological processes and for conducting regional studies.

删除了： ()

设置了格式： 默认段落字体, 到齐到网格

References

Addor, N., Newman, A. J., Mizukami, N., and Clark, M. P.: The CAMELS data set: catchment attributes and meteorology for large-sample studies, Hydrology and Earth System Sciences, 21, 5293-5313, https://doi.org/10.5194/hess-21-5293-2017, 2017.

Addor, N., Nearing, G., Prieto, C., Newman, A., Le Vine, N., and Clark, M. P.: A ranking of hydrological signatures based on their predictability in space, Water Resources Research, 54, 8792-8812, https://doi.org/10.1029/2018WR022606, 2018.

Berghuijs, W. R., Sivapalan, M., Woods, R. A., and Savenije, H. H. G.: Patterns of similarity of seasonal water balances: A window into streamflow variability over a range of time scales, Water Resources Research, 50, 5638-5661, https://doi.org/10.1002/2014WR015692, 2014.

Betterle, A., Schirmer, M., and Botter, G.: Flow dynamics at the continental scale: Streamflow correlation and hydrological similarity, Hydrological Processes, 33, 627-646, https://doi.org/10.1002/hyp.13350, 2019.

Bezdek, J. C., Ehrlich, R., and Full, W.: FCM: The fuzzy c-means clustering algorithm, Computers & geosciences, 10, 191-203, https://doi.org/10.1016/0098-3004(84)90020-7, 1984.

Boscarello, L., Ravazzani, G., Cislaghi, A., and Mancini, M.: Regionalization of flow-duration curves through catchment classification with streamflow signatures and physiographic–climate indices, Journal of Hydrologic Engineering, 21, 05015027, https://doi.org/10.1061/(ASCE)HE.1943-5584.0001307, 2016.

Carozza, D. A. and Boudreault, M.: A Global Flood Risk Modeling Framework Built With Climate Models and Machine Learning, Journal of Advances in Modeling Earth Systems, 13, https://doi.org/10.1029/2020MS002221, 2021.

Coopersmith, E., Yaeger, M. A., Ye, S., Cheng, L., and Sivapalan, M.: Exploring the physical controls of regional patterns of flow duration curves–Part 3: A catchment classification system based on regime curve indicators, Hydrology and Earth System Sciences, 16, 4467-4482, https://doi.org/10.5194/hess-16-4467-2012, 2012.

Games, P. A. and Howell, J. F.: Pairwise multiple comparison procedures with unequal n's and/or variances: a Monte Carlo study, Journal of Educational Statistics, 1, 113-125, https://doi.org/10.3102/10769986001002113, 1976.

Gao, H., Birkel, C., Hrachowitz, M., Tetzlaff, D., Soulsby, C., and Savenije, H. H. G.: A simple topography-driven and calibration-free runoff generation module, Hydrology and Earth System Sciences, 23, 787-809, https://doi.org/10.5194/hess-23-787-2019, 2019.

Ghotbi, S., Wang, D., Singh, A., Mayo, T., and Sivapalan, M.: Climate and Landscape Controls of Regional Patterns of Flow Duration Curves Across the Continental United States: Statistical Approach, Water Resources Research, 56, https://doi.org/10.1029/2020WR028041, 2020.

Guo, Y., Zhang, Y., Zhang, L., and Wang, Z.: Regionalization of hydrological modeling for predicting streamflow in ungauged catchments: A comprehensive review, Wiley Interdisciplinary Reviews: Water, 46, https://doi.org/10.1002/wat2.1487, 2020.

Halim, Z., Waqas, M., Baig, A. R., and Rashid, A.: Efficient clustering of large uncertain graphs using neighborhood information, International Journal of Approximate Reasoning, 90, 274-291, https://doi.org/10.1016/j.ijar.2017.07.013, 2017.

删除了:

删除了: Addor, N., Newman, A. J., Mizukami, N., and Clark, M. P.: The CAMELS data set: catchment attributes and meteorology for large-sample studies, Hydrology and Earth System Sciences, 21, 5293-5313, 10.5194/hess-21-5293-2017, 2017.
Addor, N., Nearing, G., Prieto, C., Newman, A., Le Vine, N., and Clark, M. P.: A ranking of hydrological signatures based on their predictability in space, Water Resources Research, 54, 8792-8812, 2018.
H. H. G.: Patterns of similarity of seasonal water balances: A window into streamflow variability over a range of time scales, Water Resources Research, 50, 5638-5661, 10.1002/2014wr015692, 2014.
Betterle, A., Schirmer, M., and Botter, G.: Flow dynamics at the continental scale: Streamflow correlation and hydrological similarity, Hydrological Processes, 33, 627-646, 10.1002/hyp.13350, 2019.
Bezdek, J. C., Ehrlich, R., and Full, W.: FCM: The fuzzy c-means clustering algorithm, Computers & geosciences, 10, 191-203, 1984.
Boscarello, L., Ravazzani, G., Cislaghi, A., and Mancini, M.: Regionalization of flow-duration curves through catchment classification with streamflow signatures and physiographic–climate indices, Journal of Hydrologic Engineering, 21, 05015027, 2016.
Carozza, D. A. and Boudreault, M.: A Global Flood Risk Modeling Framework Built With Climate Models and Machine Learning, Journal of Advances in Modeling Earth Systems, 13, 10.1029/2020ms002221, 2021.
S., Cheng, L., and Sivapalan, M.: Exploring the physical controls of regional patterns of flow duration curves–Part 3: A catchment classification system based on regime curve indicators, Hydrology and Earth System Sciences, 16, 4467-4482, 2012.
Howell, J. F.: Pairwise multiple comparison procedures with unequal n's and/or variances: a Monte Carlo study, Journal of Educational Statistics, 1, 113-125, 1976.
Tetzlaff, D., Soulsby, C., and Savenije, H. H. G.: A simple topography-driven and calibration-free runoff generation module, Hydrology and Earth System Sciences, 23, 787-809, 10.5194/hess-23-787-2019, 2019.
Sivapalan, M.: Climate and Landscape Controls of Regional Patterns of Flow Duration Curves Across the Continental United States: Statistical Approach, Water Resources Research, 56, 10.1029/2020wr028041, 2020.
Wang, Z.: Regionalization of hydrological modeling for predicting streamflow in ungauged catchments: A comprehensive review, Wiley Interdisciplinary Reviews: Water, 46, 2020.
Baig, A. R., and Rashid, A.: Efficient clustering of large uncertain graphs using neighborhood information, International Journal of Approximate Reasoning, 90, 274-291, 2017.
L., Cui, H., Qin, T., Du, S., Zhu, Y., Fang, X., and Xu, C.-Y.: Streamflow prediction in ungauged catchments through use of catchment classification and deep learning, Journal of Hydrology, 639, 131638, 2024.
McDonnell, J., Sivapalan, M., Pomeroy, J., Arheimer, B., Blume, T., Clark, M., and Ehret, U.: A decade of Predictions in Ungauged Basins (PUB)—a review, Hydrological sciences journal, 58, 1198-1255, 2013.
Using hydrological and climatic catchment clusters to explore drivers of catchment behavior, Hydrology and Earth System Sciences, 24, 1081-1100, 2020.
Jeong, K.-S., Hong, D.-G., Byeon, M.-S., Jeong, J.-C., Kim, H.-G., Kim, D.-K., and Joo, G.-J.: Stream modification patterns in a river basin: Field survey and self-organizing map (SOM) application, Ecological Informatics, 5, 293-303, 2010.
Kiang, M. Y.: Extending the Kohonen self-organizing map network

He, M., Jiang, S., Ren, L., Cui, H., Qin, T., Du, S., Zhu, Y., Fang, X., and Xu, C.-Y.: Streamflow prediction in ungauged catchments through use of catchment classification and deep learning, Journal of Hydrology, 639, 131638, https://doi.org/10.1016/j.jhydrol.2024.131638, 2024.

Hrachowitz, M., Savenije, H., Blöschl, G., McDonnell, J., Sivapalan, M., Pomeroy, J., Arheimer, B., Blume, T., Clark, M., and Ehret, U.: A decade of Predictions in Ungauged Basins (PUB)—a review, Hydrological sciences journal, 58, 1198-1255, https://doi.org/10.1080/02626667.2013.803183, 2013.

Jehn, F. U., Bestian, K., Breuer, L., Kraft, P., and Houska, T.: Using hydrological and climatic catchment clusters to explore drivers of catchment behavior, Hydrology and Earth System Sciences, 24, 1081-1100, https://doi.org/10.5194/hess-24-1081-2020, 2020.

Jeong, K.-S., Hong, D.-G., Byeon, M.-S., Jeong, J.-C., Kim, H.-G., Kim, D.-K., and Joo, G.-J.: Stream modification patterns in a river basin: Field survey and self-organizing map (SOM) application, Ecological Informatics, 5, 293-303, https://doi.org/10.1016/j.ecoinf.2010.04.005, 2010.

Kiang, M. Y.: Extending the Kohonen self-organizing map networks for clustering analysis, Computational Statistics & Data Analysis, 38, 161-180, https://doi.org/10.1016/S0167-9473(01)00040-8, 2001.

Kittel, C. M. M., Arildsen, A. L., Dybkjær, S., Hansen, E. R., Linde, I., Slott, E., Tøttrup, C., and Bauer-Gottwein, P.: Informing hydrological models of poorly gauged river catchments – A parameter regionalization and calibration approach, Journal of Hydrology, 587, 124999, https://doi.org/10.1016/j.jhydrol.2020.124999, 2020.

Knoben, W. J., Woods, R. A., and Freer, J. E.: A quantitative hydrological climate classification evaluated with independent streamflow data, Water Resources Research, 54, 5088-5109, https://doi.org/10.1029/2018WR022913, 2018.

Kohonen, T.: Self-organized formation of topologically correct feature maps, Biological cybernetics, 43, 59-69, 1982.

Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S., and Nearing, G. S.: Toward Improved Predictions in Ungauged Basins: Exploiting the Power of Machine Learning, Water Resources Research, 55, 11344-11354, https://doi.org/10.1029/2019WR026065, 2019.

Kuentz, A., Arheimer, B., Hundecha, Y., and Wagener, T.: Understanding hydrologic variability across Europe through catchment classification, Hydrology and Earth System Sciences, 21, 2863-2879, https://doi.org/10.5194/hess-21-2863-2017, 2017.

Lee, K.-J., Yun, S.-T., Yu, S., Kim, K.-H., Lee, J.-H., and Lee, S.-H.: The combined use of self-organizing map technique and fuzzy c-means clustering to evaluate urban groundwater quality in Seoul metropolitan city, South Korea, Journal of Hydrology, 569, 685-697, https://doi.org/10.1016/j.jhydrol.2018.12.031, 2019.

Lehner, B. and Grill, G.: Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems, Hydrological Processes, 27, 2171-2186, https://doi.org/10.1002/hyp.9740, 2013.

Leibowitz, S. G., Comeleo, R. L., Wigington, P. J., Weber, M. H., Sproles, E. A., and Sawicz, K. A.: Hydrologic Landscape Characterization for the Pacific Northwest, USA, JAWRA Journal of the American Water Resources Association, 52, 473-493, https://doi.org/10.1111/1752-1688.12402, 2016.

Li, Q., Li, Z., Zhu, Y., Deng, Y., Zhang, K., and Yao, C.: Hydrological regionalisation based on available hydrological information for runoff prediction at catchment scale, Proceedings of the International Association of Hydrological Sciences, 379, 13-19, https://doi.org/10.5194/piahs-379-13-2018, 2018.

Liu, C., Zhou, C., and Yu, J.: Chinese Hydrological Geography, 2014.

Liu, J., Yang, L., Jiang, J., Yuan, W., and Duan, Z.: Mapping diurnal cycles of precipitation over China through clustering, Journal of Hydrology, 125804, https://doi.org/10.1016/j.jhydrol.2020.125804, 2020a.

Liu, Y., Zhang, K., Li, Z., Liu, Z., Wang, J., and Huang, P.: A hybrid runoff generation modelling framework based on spatial combination of three runoff generation schemes for semi-humid and semi-arid watersheds, Journal of Hydrology, 590, 125440, https://doi.org/10.1016/j.jhydrol.2020.125440, 2020b.

Loritz, R., Gupta, H., Jackisch, C., Westhoff, M., Kleidon, A., Ehret, U., and Zehe, E.: On the dynamic nature of hydrological similarity, Hydrology & Earth System Sciences, 22, 3663-3684, https://doi.org/10.5194/hess-22-3663-2018, 2018.

Loritz, R., Kleidon, A., Jackisch, C., Westhoff, M., Ehret, U., Gupta, H., and Zehe, E.: A topographic index explaining hydrological similarity by accounting for the joint controls of runoff formation, Hydrology and Earth System Sciences, 23, 3807-3821, https://doi.org/10.5194/hess-23-3807-2019, 2019.

Luo, K.: Draft of natural geography regionalization of China, Acta Geographica Sinica, 20, 379-394, 1954.

Ma, K., Feng, D., Lawson, K., Tsai, W. P., Liang, C., Huang, X., Sharma, A., and Shen, C.: Transferring hydrologic data across continents–Leveraging data-rich regions to improve hydrologic prediction in data-sparse regions, Water Resources Research, 57, e2020WR028600, https://doi.org/10.1029/2020WR028600, 2021.

Moratiel, R., Bravo, R., Saa, A., Tarquis, A. M., and Almorox, J.: Estimation of evapotranspiration by the Food and Agricultural Organization of the United Nations (FAO) Penman–Monteith temperature (PMT) and Hargreaves–Samani (HS) models under temporal and spatial criteria–a case study in Duero basin (Spain), Natural Hazards and Earth System Sciences, 20, 859-875, https://doi.org/10.5194/nhess-20-859-2020, 2020.

Nguyen, T. T., Kawamura, A., Tong, T. N., Amaguchi, H., Nakagawa, N., Gilbuena Jr, R., and Du Bui, D.: Identification of spatio-seasonal hydrogeochemical characteristics of the unconfined groundwater in the Red River Delta, Vietnam, Applied geochemistry, 63, 10-21, https://doi.org/10.1016/j.apgeochem.2015.07.009, 2015.

Pagliero, L., Bouraoui, F., Diels, J., Willems, P., and McIntyre, N.: Investigating regionalization techniques for large-scale hydrological modelling, J Hydrol (Amst), 570, 220-235, https://doi.org/10.1016/j.jhydrol.2018.12.071, 2019.

Pakhira, M. K., Bandyopadhyay, S., and Maulik, U.: Validity index for crisp and fuzzy clusters, Pattern recognition, 37, 487-501, https://doi.org/10.1016/j.patcog.2003.06.005, 2004.

Pal, N. R., Pal, K., Keller, J. M., and Bezdek, J. C.: A possibilistic fuzzy c-means clustering algorithm, IEEE transactions on fuzzy systems, 13, 517-530, https://doi.org/10.1109/TFUZZ.2004.840099, 2005.

Park, Y.-S., Céréghino, R., Compin, A., and Lek, S.: Applications of artificial neural networks for patterning and predicting aquatic insect species richness in running waters, Ecological modelling, 160, 265-280, https://doi.org/10.1016/S0304-3800(02)00258-2, 2003.

Prieto, C., Le Vine, N., Kavetski, D., García, E., and Medina, R.: Flow prediction in ungauged catchments using probabilistic random forests regionalization and new statistical adequacy tests, Water Resources Research, 55, 4364-4392, https://doi.org/10.1029/2018WR023254, 2019.

Rao, A. R. and Srinivas, V.: Regionalization of watersheds by hybrid-cluster analysis, Journal of Hydrology, 318, 37-56, https://doi.org/10.1016/j.jhydrol.2005.06.003, 2006.

Rasheed, Z., Aravamudan, A., Sefidmazgi, A. G., Anagnostopoulos, G. C., and Nikolopoulos, E. I.: Advancing flood warning procedures in ungauged basins with machine learning, Journal of Hydrology, 609, 127736, https://doi.org/10.1016/j.jhydrol.2022.127736, 2022.

Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., and Carrillo, G.: Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA, Hydrology and Earth System Sciences, 15, 2895-2911, https://doi.org/10.5194/hess-15-2895-2011, 2011.

Schwämmle, V. and Jensen, O. N.: A simple and fast method to determine the parameters for fuzzy c–means cluster analysis, Bioinformatics, 26, 2841-2848, https://doi.org/10.1093/bioinformatics/btq534, 2010.

Singh, S. K., McMillan, H., Bárdossy, A., and Fateh, C.: Nonparametric catchment clustering using the data depth function, Hydrological Sciences Journal, 61, 2649-2667, https://doi.org/10.1080/02626667.2016.1168927, 2016.

Tarasova, L., Basso, S., Wendi, D., Viglione, A., Kumar, R., and Merz, R.: A Process-Based Framework to Characterize and Classify Runoff Events: The Event Typology of Germany, Water Resources Research, 56, https://doi.org/10.1029/2019WR026951, 2020.

Tsegaw, A. T., Alfredsen, K., Skaugen, T., and Muthanna, T. M.: Predicting hourly flows at ungauged small rural catchments using a parsimonious hydrological model, Journal of Hydrology, 573, 855-871, https://doi.org/10.1016/j.jhydrol.2019.03.090, 2019.

Vesanto, J.: SOM-based data visualization methods, Intelligent data analysis, 3, 111-126, https://doi.org/10.3233/IDA-1999-3203, 1999.

Wagener, T., Sivapalan, M., Troch, P. A., McGlynn, B. L., Harman, C. J., Gupta, H. V., Kumar, P., Rao, P. S. C., Basu, N. B., and Wilson, J. S.: The future of hydrology: An evolving science for a changing world, Water Resources Research, 46, https://doi.org/10.1029/2009WR008906, 2010.

Wang, S., Zhang, K., Chao, L., Li, D., Tian, X., Bao, H., Chen, G., and Xia, Y.: Exploring the utility of radar and satellite-sensed precipitation and their dynamic bias correction for integrated prediction of flood and landslide hazards, Journal of Hydrology, 603, 126964, 2021.

Willmott, C. J. and Feddema, J. J.: A more rational climatic moisture index, The Professional Geographer, 44, 84-88, https://doi.org/10.1111/j.0033-0124.1992.00084.x, 1992.

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

Yaeger, M., Coopersmith, E., Ye, S., Cheng, L., Viglione, A., and Sivapalan, M.: Exploring the physical controls of regional patterns of flow duration curves–Part 4: A synthesis of empirical analysis, process modeling and catchment classification, Hydrology and Earth System Sciences, 16, 4483-4498, https://doi.org/10.5194/hess-16-4483-2012, 2012.

Yamazaki, D., Sato, T., Kanae, S., Hirabayashi, Y., and Bates, P. D.: Regional flood dynamics in a bifurcating mega delta simulated in a global river model, Geophysical Research Letters, 41, 3127-3135, https://doi.org/10.1002/2014GL059744, 2014.

Yang, X., Magnusson, J., Rizzi, J., and Xu, C.-Y.: Runoff prediction in ungauged catchments in Norway: comparison of regionalization approaches, Hydrology Research, 49, 487-505, https://doi.org/10.2166/nh.2017.071, 2018.

Yang, X., Magnusson, J., Huang, S., Beldring, S., and Xu, C.-Y.: Dependence of regionalization methods on the complexity of hydrological models in multiple climatic regions, Journal of Hydrology, 582, 124-357, https://doi.org/10.1016/j.jhydrol.2019.124357, 2020.

Yang, Y., Liu, J., Yang, S., and He, R.: Understanding the hierarchical controls of geographical features on hydrological responses in humid mountainous areas through a stepwise clustering scheme, Stochastic Environmental Research and Risk Assessment, https://doi.org/10.1007/s00477-021-02080-3, 2021a.

Yang, Y., Pan, M., Lin, P., Beck, H. E., Zeng, Z., Yamazaki, D., David, C. H., Lu, H., Yang, K., and Hong, Y.: Global reach-level 3-hourly river flood reanalysis (1980–2019), Bulletin of the American Meteorological Society, 102, E2086-E2105, https://doi.org/10.1175/BAMS-D-20-0057.1, 2021b.

Yaseen, Z. M., Sulaiman, S. O., Deo, R. C., and Chau, K.-W.: An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction, Journal of Hydrology, 569, 387-408, https://doi.org/10.1016/j.jhydrol.2018.11.069, 2019.

Yi, X. and Jiazhen, Z.: Hydrology Regionalization of China, 1995.

Zang, S., Li, Z., Zhang, K., Yao, C., Liu, Z., Wang, J., Huang, Y., and Wang, S.: Improving the flood prediction capability of the Xin'anjiang model by formulating a new physics-based routing framework and a key routing parameter estimation method, Journal of Hydrology, 603, 126867, https://doi.org/10.1016/j.jhydrol.2021.126867, 2021.

Zeng, P., Sun, F., Liu, Y., Wang, Y., Li, G., and Che, Y.: Mapping future droughts under global warming across China: A combined multi-timescale meteorological drought index and SOM-Kmeans approach, Weather and Climate Extremes, 31, https://doi.org/10.1016/j.wace.2021.100304, 2021.