Climate and landscape jointly control ~~Europe's hydrology~~European streamflow behaviour

Julia M. Rudlang[1], Thiago V.M. do Nascimento[2,3], Ruud van der Ent[1], Fabrizio Fenicia[2], and Markus Hrachowitz[1]

[1]Department of Water Management, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, the Netherlands
[2]Eawag: Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland
[3]Department of Geography, University of Zurich, Zurich, Switzerland

Correspondence: Julia M. Rudlang (J.M.Rudlang@tudelft.nl)

Abstract. The complex composition of hydrological systems, climates and landscapes makes it challenging to explain and predict hydrological streamflow response. Many previous large-sample studies, mostly focused on the United States, identified climate as the primary control, with landscape exerting only a minor role in shaping hydrological behaviour. Yet, a few other studies report ~~contradicting~~ contradictory results with landscape being a more dominant driver. In this study, we use an ~~unprecedentedly large~~ unprecedented sample of more than 7000 catchments in Europe from the EStreams dataset to identify and map functionally similar catchments, together with their spatially variable climate and landscape controls. The wide spatial and temporal gradient of the study catchments was used to identify hydrological response types (HRTs) based on 40 hydrological streamflow signatures related to long-term averages and inter-annual variability of magnitude, timing, duration, frequency, and seasonality. Overall, 10 HRTs could be identified. Several HRTs are well defined and well distinguishable, largely due to catchments with strongly seasonal or more extreme behaviour. Other HRTs remain difficult to distinguish, as these catchments represent more transitional conditions with increasingly overlapping characteristics between HRTs. The underlying drivers of the HRTs were identified by using 84 climate- and landscape attributes to predict catchment membership to their respective HRT with a Random Forest classification model. Climate emerges as the dominant driver of hydrological behaviour at the continental scale. However, landscape was found, in 4 out of 10 HRTs, to be at least as strong or even stronger a control on the hydrological streamflow response. These results highlight that the complex, integrated nature of hydrological response remains challenging to disentangle, even with extensive datasets and advanced modelling approaches, and therefore, climate and landscape ~~needs~~ need to be understood as joint drivers in a co-evolutionary perspective.

1 Introduction

In terrestrial hydrological systems, the landscape acts as a low-pass filter that buffers precipitation water input before it is released from the system as streamflow or evaporation. The magnitude and dynamic changes of streamflow over time are thus a manifestation of the combined and aggregated effects of the hydro-climatic characteristics and the physical properties of the landscape in a catchment. Any catchment is characterised by a heterogeneous mosaic of countless local landscape

features that each influence the movement of water in different ways ~~.~~ (Beven, 2000; Sivapalan, 2003a; Wagener et al., 2007) . In practice, these controls are inherently intertwined, as climatic inputs are mediated through landscape properties such as topography, soils, geology, vegetation and storage elements before contributing to streamflow. The combination of the hydro-climatic characteristics with these local landscape elements and their topology in any catchment is a multivariate fingerprint that may ~~plausibly even be assumed to be~~ be considered unique (Beven, 2000). In spite of this potential "uniqueness of place" (Beven, 2000) and the associated spatial complexity of the hydrological response, there is evidence that self-organisation leads to relatively simple hydrological response ~~pattern~~ patterns that emerge at larger scales (e.g. catchment-scale) and that can be relatively well quantified using various streamflow signatures. The latter are index values derived from streamflow observations that describe catchment characteristics such as magnitude, seasonality and flashiness. Common examples include mean flow, baseflow index or the slope of the flow duration curve (FDC) (McMillan et al., 2017). Hydrological signatures provide interpretable metrics that facilitate the identification of similarity, response patterns and dominant processes across catchments. Systematically mapping catchment-scale hydro-climatic and landscape attributes against catchment-scale hydrological response ~~pattern~~ patterns then holds the potential to identify hydrological (dis)similarities across catchments in diverse environments and to classify them into types with distinct hydrological functioning (Dooge, 1986; Blöschl, 2006; Wagener et al., 2007; McDonnell et al., 2007). Such a classification is critical to inform the design of architectures and parametrisations of hydrological models (Gupta et al., 2014), as demonstrated by several studies (e.g., Merz and Blöschl, 2004; Parajka et al., 2005; Ye et al., 2012; Fenicia et al., 2014; Beck et al., 2020; Dal Molin et al., 2020). The increasing number and volume of ~~large sample~~ large-sample datasets (e.g., Addor et al., 2017; Gudmundsson et al., 2018; Fowler et al., 2021; Kratzert et al., 2023; do Nascimento et al., 2024) in hydrology has over the past two decades facilitated ever more detailed comparison, classification and regionalisation studies which led to considerable progress in that direction ~~(e.g., Merz and Blöschl, 2003, 2009; Laaha and Blöschl, 200~~ (e.g., Merz and Blöschl, 2003, 2009; Laaha and Blöschl, 2006a, b; Carrillo et al., 2011; Sawicz et al., 2011; Coopersmith et al., 2012; Yae .

These studies exhibit substantial variation in terms of both the number of catchments analysed and the methodological choices employed in the classification. While some studies have classified on climatic characteristics (Knoben et al., 2018; Brunner et al., 2020), streamflow signatures (Kuentz et al., 2017; Almagro et al., 2024) or on a combination of the two (Kemter et al., 2023), others have considered landscape attributes only (Jones Jr. et al., 2021). For example, Berghuijs et al. (2014) used three elements of the seasonal water balance to group catchments in the ~~US.~~U.S. Based on the coherent similarity of streamflow signatures within the formed groups, they could conclude that climate is the dominant driver of streamflow variability. This broadly corresponds with the results of Kuentz et al. (2017), who ~~, however,~~ used 16 flow signatures to group ~~1366~~ 1366 catchments into 11 functionally distinct groups across Europe. Further analysis of these groups against 35 climate- and landscape attributes revealed that most signatures were controlled by climatic factors. Using Random Forest models to predict a range of hydrological signatures (Addor et al., 2018) and regime types (Brunner et al., 2020) of 671 ~~US~~ U.S. catchments, suggested that climate attributes were the most important predictor, with landscape attributes playing only a minor role.

Somewhat more nuanced results were reported by Almagro et al. (2024), who classified more than 700 Brazilian catchments into six groups based on their hydrological similarity and analysed which climate- and landscape attributes were the most influential on individual hydrological signatures within these groups. Although climate was, for the majority of the six groups, the main driver of hydrological behaviour, landscape ~~was~~ attributes related to topography, soil and land use were more dominant in some of them. In another study, Kerins et al. (2025) used the HBV model to simulate the hydrological processes in various catchments and found that climatic attributes, such as aridity and precipitation phase, were the dominant drivers of streamflow generation. Yet, landscape attributes, such as topography and soil texture, exert considerable influence as well. For example, Fenicia and McDonnell (2022) developed a perceptual model of the Moselle basin in Central Europe and found precipitation to dominate the streamflow variability, while lithology, topography and land use controlled the spatial pattern of multiple other aspects of the hydrograph within the basin.

Such recent efforts to disentangle climate and landscape influence on hydrological behaviour unsurprisingly suggest that more often than not, climate emerges as the overall driver of ~~pattern~~ patterns in hydrological behaviour. However, there is growing and somewhat contrasting evidence that landscape attributes may be more important controls on the hydrological response in some environments and at various spatial scales ~~(Fenicia and McDonnell, 2022; Almagro et al., 2024; do Nascimento et al., 202~~ (Fenicia and McDonnell, 2022; Almagro et al., 2024; do Nascimento et al., 2025a; Araki et al., 2025).

Many of the reported differences result from differences in the detailed objectives of previous studies and the related different choices in the design of the associated experiments. Most notably, many studies are based on either (1) a low-dimensional characterisation of the hydrological response with only a few streamflow signatures (e.g., Sawicz et al., 2011; Coopersmith et al., 2012; Jehn et al., 2020), (2) a small number of climate (e.g., Kuentz et al., 2017) or landscape attributes (e.g., Berghuijs et al., 2014; Kemter et al., 2023) as potential controls or (3) the intention to identify climate and landscape controls in a one-dimensional way for individual signatures (e.g., Addor et al., 2018; Almagro et al., 2024). Note that only a handful of studies have used climate- and landscape attributes to quantitatively predict membership to distinct classes of hydrological response types that encapsulate the multivariate and nuanced complexity of the hydrological response (e.g., McManamay et al., 2014; Kuentz et al., 2017; Brunner et al., 2020). Furthermore, Tarasova et al. (2024) pointed out that (4) most studies seeking to identify the dominant drivers of streamflow overlook inter-annual variability in signatures and attributes, unnecessarily reducing complex catchment processes to simple averages, and thus also reducing the discriminatory power of classification schemes (cf. McMillan et al., 2025). Finally, it must be conceded that (5) even many of the available ~~large sample~~ large-sample data sets, containing several hundreds of catchments, in fact only provide coarse spatial representations of hydrological, climatic and landscape contrasts. As such, they may contain insufficient information to describe the multivariate complexities of the hydrological response to allow a robust and generalisable distinction of different of hydrological response types and their drivers.

As a consequence and although the question has already been raised almost two decades ago (e.g., Blöschl et al., 2007), our ability to identify, distinguish and group catchments with (dis-)similar hydrological response types and to describe the spatially and temporally varying individual roles of climate and landscape to shape these (dis-)similarities remains fairly fragmented and incomplete (Blöschl et al., 2019). Let us illustrate this by postulating, in a thought-experiment, the validity of the uniqueness

of place conjecture (Beven, 2000). This conjecture further entails that the climate- and landscape attributes that control the hydrological response are a unique fingerprint of a catchment. By extension, the hydrological response then is, even in the presence of rather simple emergent catchment-scale processes (e.g., Sivapalan, 2003b), an equally unique fingerprint. It is therefore plausible to expect that these unique responses allow a distinction between a wide spectrum of classes of dissimilar hydrological response types. Yet, previous classification studies did largely not distinguish more than ~ 10 individual classes of response types (e.g., Berghuijs et al., 2014; Kuentz et al., 2017; Brunner et al., 2020; Jehn et al., 2020) with only a few exceptions (e.g., McManamay et al., 2014). It could be argued that these relatively few classes largely capture the major spatial pattern and differences in hydrological response types. However, the typically observed considerable variability within the classes (e.g., Berghuijs et al., 2014; Kuentz et al., 2017; Brunner et al., 2020; Jehn et al., 2020) together with the ambiguous results on the roles of climate- and landscape attributes as drivers (e.g., Almagro et al., 2024) suggests that the response types are much more varied than suggested by these classes. This is not only theoretically unsatisfying but seriously limits our ability to design suitable models for more reliable predictions in ungauged basins and under a changing climate.

Building on earlier studies and reflecting the ongoing community debate on the required sample size and the choice and number of hydrological signatures and catchment attributes that need to be considered for meaningful hydrological classification and identification of underlying drivers (e.g., McDonnell and Woods, 2004; McMillan, 2020; McMillan et al., 2022, 2023b; Tarasova et al., 2024), we here address several of the above unresolved issues. The overall aim of this study is to identify more detailed classes of distinct hydrological response types and to describe the spatially varying roles of climate and landscape as controls on the hydrological response. Using data from > 7000 catchments across Europe and a wide range of hydrological signatures as well as climate- and landscape attributes, for the first time including descriptors of inter-annual variability (Tarasova et al., 2024), thereby increasing discriminatory power and reducing the risk of expert bias associated with the use of narrow subsets of signatures (McMillan et al., 2017, 2023a; McMillan, 2021), the novelty of this study lies in both its scale and scope. Specifically, we test the hypotheses that (1) the use of a larger catchment sample and a wider range of different signatures, including descriptors of inter-annual variability, allows a more detailed classification of hydrological response types, thereby reducing intra-class variance and that (2) climate- rather than landscape attributes consistently emerge as dominant discriminatory elements that cause the dissimilarity between these classes of hydrological response types.

2 Data and study area

In this study, we used hydrological, hydro-meteorological and landscape data from ~~7175~~ 7175 catchments across Europe, available through the EStreams data portal (do Nascimento et al., 2024). The studied catchments cover a wide environmental gradient, capturing several climatic zones, varied weather patterns, diverse vegetation, land use, soils, geology and topography~~and~~ . The spatial representativeness of the dataset is reflected in its relatively high gauge density across Europe. While the distribution is not spatially uniform, with the highest density in central Europe and lower densities in southern and eastern Europe, the overall density corresponds to approximately one gauge per $1000\,\mathrm{km}^2$ (7175 gauges across $\sim 7 \times 10^6\,\mathrm{km}^2$). For

comparison, the CAMELS-US dataset (Addor et al., 2017) has a density of approximately one gauge per $14\,600\,\mathrm{km}^2$ (671 across $\sim 9.8 \times 10^6\,\mathrm{km}^2$) .

The catchments were selected as a subset from EStreams based on the following criteria: (1) at least 15 years of daily streamflow data within the period 1980–2022, (2) $< 5\,\%$ missing data in the total available time series, (3) no obvious violation of the long-term water balance of precipitation (P) and streamflow (Q), i.e. $\overline{P} > \overline{Q}$, under the assumption that the long-term storage ~~(S):~~ change is negligible, i.e. $\frac{dS}{dt} \sim 0$ and (4) catchment areas $< 25\,000\,\mathrm{km}^2$.

The median size of the selected catchments is $226\,\mathrm{km}^2$, and 4137 catchments thereof ($48\,\%$) are nested within larger basins included in the dataset, reflecting the hierarchical structure of river networks represented in EStreams. The majority of the catchments are unregulated or only weakly influenced by reservoirs, while more strongly regulated systems represent a relatively small subset (only $8\,\%$ have a reservoir storage exceeding 50 mm).

As next-to-complete annual time series years are important for seasonal signatures, individual years with > 15 days missing were excluded to avoid losing crucial information. For ~ 100 study catchments ($\sim 1\,\%$) ≤ 15 days of data were missing in individual years. In these cases, the years were completed by interpolation over the missing days. This was deemed acceptable as the subsequent analysis was based on long-term averages and the inter-variability around these averages.

3 Methods

The overall experiment in this study consists of four distinct steps as illustrated in Fig.1. In the first step, the streamflow time series were used to derive descriptors of the hydrological response for each catchment, hereafter referred to as hydrological signatures (1a), while climate- and landscape-related data were used to derive descriptors of catchment characteristics, hereafter referred to as climate- and landscape attributes (1b). In the second step, the set of hydrological signatures from all catchments was used to identify classes of distinct hydrological response types (HRT) based on k-means clustering (2). In the third step, membership to the individual classes of hydrological response types was predicted based on the set of climatic and landscape attributes using Random Forest classification (3). Finally, the roles of the individual climate- and landscape attributes as drivers of the hydrological response were identified based on their strengths as predictors in the Random Forest classification as quantified by their respective feature importances (4).

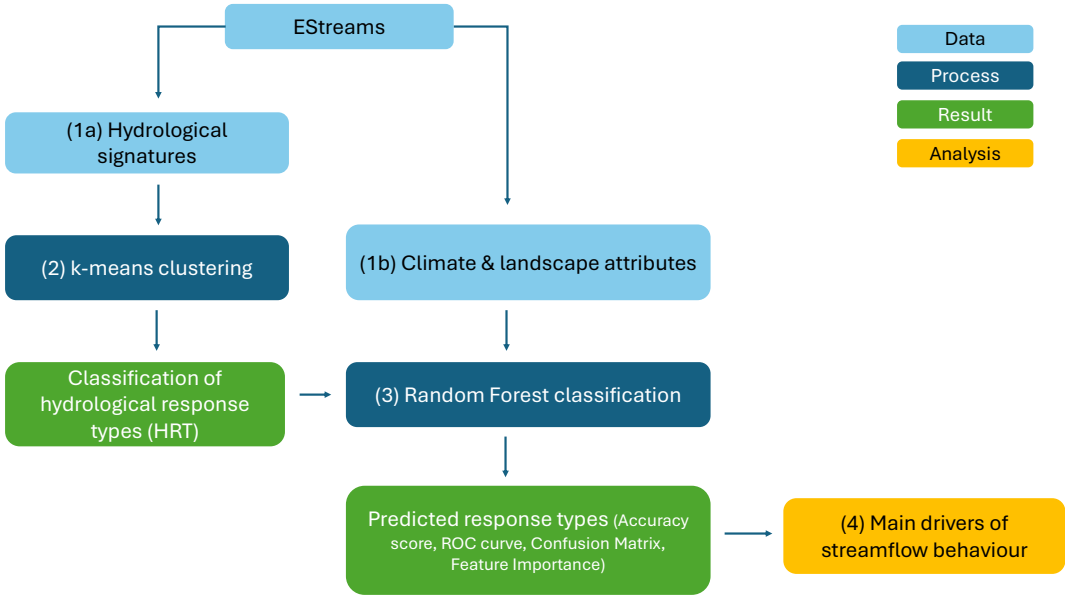


Figure 1. Workflow of the study. Streamflow time series and catchment attributes (1) were used to classify hydrological response types (HRT) to the individual catchments via k-means clustering (2) and predict HRT membership using a Random Forest classification model (3). The importance of individual attributes was quantified from the Random Forest classification model to identify the dominant controls on HRT membership and thus the main drivers of the hydrological response (4).

3.1 Signatures and attributes

The streamflow signatures and climate- and landscape attributes are compiled from the EStreams dataset (do Nascimento et al., 2024) and include long-term averages, as well as inter-annual variability around these averages (for derivation details and sources of the non-standard signatures and attributes, see Table S1 and S2 in the Supplementary Material). For variables that represent the intra-annual timing, the average timing signatures and attributes were calculated with circular statistics to account for the circularity of the year (Parajka et al., 2010; Blöschl et al., 2017; Berghuijs et al., 2025). Specifically, the normalised day of year, t_{norm} (ranging from 0 to 1) was mapped onto the unit circle using two dimensionless circular coordinates:

$t_{SA} = \cos(2\pi\, t_{norm})$ (spring-autumn axis),

$t_{WS} = \sin(2\pi\, t_{norm})$ (winter-summer axis)

This decomposition ensures that dates near the year boundary (e.g., December 31st and January 2nd) are represented as adjacent points on the circle, allowing for a correct computation of mean timing. The circular mean vector $(\overline{t_{SA}}, \overline{t_{WS}})$ was then converted back to day of year using $t_{doy} = (365.25/2\pi) * \text{atan2}(\overline{t_{SA}}, \overline{t_{WS}})$, producing an unambiguous average day of the year. In this mapping, $t_{WS} = -1$ corresponds approximately to January 1st and $t_{WS} = +1$ to July 1st, while $t_{SA} = +1$ corresponds approximately to April 1st and $t_{SA} = -1$ to October 1st. For subsequent analysis with k-means clustering and Random Forest

classification, the cosine and sine components ($\overline{\mathrm{t_{SA}}}, \overline{\mathrm{t_{WS}}}$) were used directly. However, throughout the manuscript, timing is for clarity mostly reported as the reconstructed day of year $\mathrm{t_{doy}}$ (noted as t). Further details on the circular-statistics approach for computing timing signatures and related attributes are provided in Supplementary Material S1.

In a deliberate decision to preserve and use as much information as possible to characterise the hydrological response, the k-means clustering and Random Forest classification analysis were based on a ~~rather extensive~~ comprehensive set of signatures and attributes. Note that all signatures and attributes used in the study contain independent information. Signatures and attributes that are consistently and strongly correlated with multiple other signatures and attributes were discarded after a preliminary scan. Some of the signatures and attributes retained for the analysis contain some overlapping information, but overall the level of correlation was considered low enough to be acceptable (Supplementary Material Fig.S1), in particular as k-means clustering and Random Forest classification are rather robust in the presence of correlated variables (Breiman, 2001).

In the following sections, the hydrological signatures and climate- and landscape attributes are referred to with a capital letter followed by the signature name in subscript to clearly distinguish between the different categories, the following notation is used: hydrological signatures are denoted with a H, the climate attributes with C, the landscape attributes are categorised into vegetation and landcover denoted by V, soil with S, geology with a G, topography with a T, and the anthropogenic attributes with an A.

3.1.1 Hydrological signatures

~~The~~ he hydrological signatures used to characterise the hydrograph included many commonly used signatures such as the long-term averages of annual mean flow (H_{Q}), high ~~flows (~~and low flow percentiles ($H_{\mathrm{Q_{95}}}$ ~~), low flows (~~and $H_{\mathrm{Q_5}}$), the baseflow index (H_{BFI}) or metrics that describe seasonality, such as the Gini coefficient (H_{GC}) (e.g., McMillan et al., 2017; Addor et al., 2018). Here, we define $H_{\mathrm{Q_{95}}}$ as high flows and $H_{\mathrm{Q_5}}$ as low flows, following the notation used in Addor et al. (2017) and do Nascimento et al. (2024). In addition, multiple signatures that characterise the average intra-annual timing of flows, such as the half-flow day ($H_{\mathrm{t(HFD)}}$) or the day of the annual maximum ($H_{\mathrm{t(Q_H)}}$) and minimum flows ($H_{\mathrm{t(Q_L)}}$) were included in this study. The long-term averages were complemented by further signatures that quantify the inter-annual variability around these averages, such as the variances of Q ($H_{\mathrm{V,Q}}$), the variance of high flows ($H_{\mathrm{V,Q_{95}}}$) or the variance of the half-flow day ($H_{\mathrm{V,t(HFD)}}$). A total of 40 hydrological signatures, constructed from $>90\,\mathrm{million}$ individual data points, were retained as sufficiently independent and used in the analysis (Table 1). To structure this high-dimensional description of the hydrological response, the signatures were loosely grouped into six categories that describe the magnitude (e.g. $H_{\mathrm{Q}}, H_{\mathrm{Q_{95}}}, H_{\mathrm{Q_5}}$), the frequency (e.g. $H_{\mathrm{f(Q_H)}}, H_{\mathrm{f(Q_L)}}, H_{\mathrm{f(Q_Z)}}$), the duration (e.g. $H_{D(Q_H)}, H_{\mathrm{D(Q_L)}}$), the timing (e.g. $H_{t(\mathrm{HFD})}, H_{\mathrm{t(Q_H)}}, H_{\mathrm{t(Q_L)}}$), the seasonality (e.g. Gini coefficient H_{GC} and baseflow index H_{BFI}) and the inter-annual variability (e.g. ($H_{\mathrm{V,Q}}$), ($H_{\mathrm{V,Q_{95}}}$), ($H_{\mathrm{V,f(Q_L)}}$)).To further assess the dimensionality and degree of redundancy within the selected hydrological signatures, a Principal Component Analysis (PCA) was performed on the full set of signatures to characterise the structure of variability across the components.

By covering such a wide range and level of detail in hydrological signatures, it is expected that a relatively detailed pattern with a high number of classes of different hydrological response types can be distinguished across Europe.

Table 1. Overview of the hydrological signatures derived from the EStreams dataset (do Nascimento et al., 2024). ~~For description~~ Detailed descriptions of the derivation ~~see~~ of each signature are provided in Supplementary Material Table S1.

Category	Symbol	Unit	Description
Magnitude			
	H_{Q}	mm d^{-1}	Mean daily streamflow.
	$H_{\mathrm{Q_5}}$	mm d^{-1}	$\mathrm{Q_5}$ is the 5 % flow quantile, representing low flows.
	$H_{\mathrm{Q_{95}}}$	mm d^{-1}	$\mathrm{Q_{95}}$ is the 95 % flow quantile, representing high flows.
Frequency			
	$H_{f(\mathrm{Q}_H)}$	d yr^{-1}	Frequency of high flows, when $\mathrm{Q} > 9$ the median daily flow.
	$H_{\mathrm{f(Q_L)}}$	d yr^{-1}	Frequency of low flows, when $\mathrm{Q} < 0.2$ times the median daily flow.
	$H_{\mathrm{f(Q_Z)}}$	d yr^{-1}	Frequency of zero flow, where days have $\mathrm{Q} = 0$.
Duration			
	$H_{\mathrm{D(Q_H)}}$	d	Average duration of high flow events of consecutive days > 9 times the median daily flow.
	$H_{\mathrm{D(Q_L)}}$	d	Average duration of flow ~~flow~~ events of consecutive days of < 0.2 times the median daily flow.
Timing			
	$H_{\mathrm{t(HFD)_{WS}}}$	-	The winter-summer component of the timing of the half-flow day, which is the day when the cumulative daily streamflow is half the annual streamflow.
	$H_{\mathrm{t(HFD)_{SA}}}$	-	The spring-autumn component of the timing of the half-flow day, which is the day when the cumulative daily streamflow is half the annual streamflow.
	$H_{\mathrm{t(Q_H)_{WS}}}$	-	The winter-summer component of the timing of high flow.
	$H_{\mathrm{t(Q_H)_{SA}}}$	-	The spring-autumn component of the timing of high flow.
	$H_{\mathrm{t(Q_L)_{WS}}}$	-	The winter-summer component of timing of low flow.
	$H_{\mathrm{t(Q_L)_{SA}}}$	-	The spring-autumn component of the timing of low flow.
Seasonality			
	$H_{\mathrm{N(Q_H\text{-}Q_L)}}$	d	Days between highest and lowest flow.
	H_{FDC}	-	Slope of the flow duration curve (FDC) between the log transformed 33$^{\mathrm{rd}}$ and 66$^{\mathrm{th}}$ streamflow percentiles.
	H_{RLD}	d^{-1}	Rising limb density, which is the number of rising peaks divided by the total time the hydrograph is rising.
	H_{FLD}	d^{-1}	Falling limb density, which is the number of falling peaks divided by the total time the hydrograph is falling.
	H_{RBI}	-	The Richard-Baker flashiness index, which measures oscillations in discharge relative to total discharge.
	H_{GC}	-	Mean Gini coefficient, which is a measure of dispersion or inequality of streamflow values. A value close to ~~1~~ 0 is a uniform distribution, and a value close to ~~0~~ 1 is a non-uniform distribution.
	$H_{\mathrm{R(Parde)}}$	-	The range of the Parde coefficients. Expressed as the mean monthly streamflow over annual streamflow, resulting in a sequence of Parde coefficients where the range is the largest minus the smallest coefficient.
	$H_{\mathrm{AC_1}}$	-	The autocorrelation with 1-day lag.
	$H_{\mathrm{AC_{30}}}$	-	The autocorrelation with 30-days lag.
	H_{BFI}	-	The baseflow index, which is the ratio of mean daily baseflow to mean daily discharge computed by hydrograph separation using Ladson et al. (2013) digital filter.
	H_{CV}	-	Coefficient of variation of the daily streamflow.
Inter-annual Variance			
	$H_{\mathrm{V,Q}}$	$(\mathrm{mm\ d^{-1}})^2$	Inter-annual variance of daily streamflow.
	$H_{\mathrm{V,Q_5}}$	$(\mathrm{mm\ d^{-1}})^2$	Inter-annual variance of $\mathrm{Q_5}$.
	$H_{\mathrm{V,Q_{95}}}$	$(\mathrm{mm\ d^{-1}})^2$	Inter-annual variance of $\mathrm{Q_{95}}$.
	$H_{\mathrm{V,f(Q_H)}}$	$(\mathrm{d\ yr^{-1}})^2$	Inter-annual variance of the frequency of high flow events.
	$H_{\mathrm{V,f(Q_L)}}$	$(\mathrm{d\ yr^{-1}})^2$	Inter-annual variance of the frequency of low flow events.
	$H_{\mathrm{V,f(Q_Z)}}$	$(\mathrm{d\ yr^{-1}})^2$	Inter-annual variance of frequency of zero flow.
	$H_{\mathrm{V,D(Q_H)}}$	d^2	Inter-annual variance of the duration of high flow events.
	$H_{\mathrm{V,D(Q_L)}}$	d^2	Inter-annual variance of the duration of low flow events.
	$H_{\mathrm{V,t(HFD)}}$	d^2	Inter-annual variance of the timing of the half-flow day.
	$H_{\mathrm{V,t(Q_H)}}$	d^2	Inter-annual variance of timing of the highest flow event.
	$H_{\mathrm{V,t(Q_L)}}$	d^2	Inter-annual variance of timing of the lowest flow event.
	$H_{\mathrm{V,N(Q_H\text{-}Q_L)}}$	d^2	Inter-annual variance of days between the highest and lowest flow.
	$H_{\mathrm{V,FDC}}$	-	Inter-annual variance of the slope of the flow duration curve.
	$H_{\mathrm{V,GC}}$	-	Inter-annual variance of the Gini coefficient.
	$H_{\mathrm{V,R(Parde)}}$	-	Inter-annual variance of the range of the Parde coefficients.

3.1.2 Climate and landscape attributes

In total, 84 climate- and landscape attributes were used in the subsequent analysis. The 47 climate attributes include the long-term averages of magnitudes of hydro-climatic variables such as precipitation (C_{P}), temperature (C_{T}), potential evaporation ($C_{\mathrm{E_P}}$) or actual evaporation ($C_{\mathrm{E_a}}$), but also of derived quantities such as the aridity index $C_{\mathrm{AI}} = \frac{E_P}{P}$ and the maximum annual water deficit C_{WD} as calculated from the running sum of $P - E_a$ for each year. In addition, multiple attributes that describe aspects of timing and duration, such as the average day of the year with the highest liquid water input $C_{\mathrm{t(P_{L,\,max})}}$, i.e. rainfall plus snowmelt, or the duration of periods with water deficit $C_{\mathrm{D(WD)}}$, were used (Table 2). The long-term average attributes were complemented by the inter-annual variance around these averages for multiple attributes, such as the variance of precipitation ($C_{\mathrm{V,\,P}}$), potential evaporation ($C_{\mathrm{V,E_P}}$) or the variance of the timing of the maximum annual water deficit ($C_{\mathrm{t(WD_{max})}}$). Note that the attributes involving snow processes were derived using a simple degree-day snow model to separate the available precipitation data into rainfall ($C_{\mathrm{P_R}}$), snowfall ($C_{\mathrm{P_S}}$), meltwater (C_{P_M}) and the corresponding liquid water input $C_{\mathrm{P_L}} = C_{\mathrm{P_R}} + C_{\mathrm{P_M}}$ (see Supplementary Material S4 for further information).

The 37 landscape attributes include subgroups of vegetation and landcover, soil, geology and topography, as well as anthropogenic attributes (Table 2). The vegetation and landcover attributes include, on the one hand, static attributes, which are either long-term averages in the case of vegetation indices such as LAI or attributes of specific years (1990, 2000, 2006, 2012, 2018) in the case of landcover from the CORINE database. On the other hand, dynamic vegetation attributes, such as the seasonality (i.e. range of the Parde coefficient) and inter-annual variances of LAI and NDVI (period from 2001-2022), were additionally distinguished as they may reflect short-term vegetation responses to hydro-climatic input.

Table 2: Climate- and landscape attributes derived from the EStreams dataset (do Nascimento et al., 2024). For description of derivation and original data sources, see Supplementary Material Table S2.

Category	Variable Name	Unit	Description
Climate			
Magnitude			
	$C_{\mathrm{E_p}}$	mm d^{-1}	Mean potential evaporation (Ep), estimated using the Hargreaves equation (Hargreaves and Samani, 1982).
	$C_{\mathrm{E_a}}$	mm d^{-1}	Mean actual evaporation (Ea), approximated from Ep.
	C_{T}	°C	Mean temperature.
	$C_{\mathrm{P_L}}$	mm d^{-1}	Mean liquid water as the sum of rainfall and snowmelt ($P_{\mathrm{L}} = P_{\mathrm{R}} + P_{\mathrm{M}}$).
	$C_{\mathrm{P_M}}$	mm d^{-1}	Mean water input from snowmelt.
	$C_{\mathrm{WD_{max}}}$	mm	Median annual maximum water deficit.
	$C_{\mathrm{F(Snow)}}$	-	Median annual fraction of precipitation falling as snow.

Category	Variable Name	Unit	Description
Frequency			
	$C_{\mathrm{f(P_{LH})}}$	d yr^{-1}	Frequency of high P_{L} events, when $P_{\mathrm{L}} > 5$ times the median daily P_{L}.
	$C_{\mathrm{f(P_{LL})}}$	d yr^{-1}	Frequency of low P_{L} events, when $P_{\mathrm{L}} < 1$ mm d^{-1}.
Duration			
	$C_{\mathrm{N(P_L)}}$	d	Median annual number of days when P_{L} occurs.
	$C_{\mathrm{D(P_{LH})}}$	d	Average duration of periods with consecutive high P_{L} events.
	$C_{\mathrm{D(P_{LL})}}$	d	Average duration of periods with consecutive low P_{L} events.
	$C_{\mathrm{D(WD)}}$	d	Median duration of water deficit.
	$C_{\mathrm{D(WD_{max})}}$	d	Median of the annual longest periods of water deficit.
	$C_{\mathrm{N(WD)}}$	d	Median annual days with water deficit.
	$C_{\mathrm{D(SnowC_{max})}}$	d	Median duration of the longest annual snow cover.
	$C_{\mathrm{D(SnowC)}}$	d	Median of the annual mean duration of snow cover.
Timing			
	$C_{\mathrm{t(P_{L,\,max})_{WS}}}$	-	The winter-summer component of the timing of the highest P_L event.
	$C_{\mathrm{t(P_{L,\,max})_{SA}}}$	-	The spring-autumn component of the timing of the highest P_L event.
	$C_{t(\mathrm{WD_{first}})_{WS}}$	-	The winter-summer component of the timing of the first annual water deficit.
	$C_{\mathrm{t(WD_{first})_{SA}}}$	-	The spring-autumn component of the timing of the first annual water deficit.
	$C_{\mathrm{t(WD_{last})_{WS}}}$	-	The winter-summer component of the timing of the last annual water deficit.
	$C_{\mathrm{t(WD_{last})_{SA}}}$	-	The spring-autumn component of the timing of the last annual water deficit.
	$C_{\mathrm{t(WD_{max})_{WS}}}$	-	The winter-summer component of the timing of the annual maximum water deficit.
	$C_{\mathrm{t(WD_{max})_{SA}}}$	-	The spring-autumn component of the timing of the annual maximum water deficit.
Seasonality			
	C_{AI}	-	Aridity Index.

Category	Variable Name	Unit	Description
	$C_{\mathrm{R(T)}}$	°C	Temperature range of long-term mean annual maximum and minimum temperature.
	$C_{\phi(\mathrm{E_P})}$	-	Seasonality and timing of potential evaporation, derived from long-term precipitation and temperature.
	$C_{\phi(\mathrm{P})}$	-	Seasonality and timing of precipitation.
	$C_{\Delta\phi(\mathrm{P,E_P})}$	-	Difference in seasonal timing between E_{P} and P.
	$C_{\mathrm{I(P_L)}}$	mm yr^{-1}	Median annual intensity of P_L.
	$C_{\mathrm{I(P_M)}}$	mm yr^{-1}	Median annual intensity of P_M.
	$C_{\mathrm{S_R(WD_{max})}}$	-	Median of the rising slope of the maximum water deficit.
	$C_{\mathrm{S_D(WD_{max})}}$	-	Median of the declining slope of the maximum water deficit.
Inter-annual Variance			
	$C_{\mathrm{V,E_P}}$	$(\mathrm{mm\ d^{-1}})^2$	Inter-annual variance of potential evaporation.
	$C_{\mathrm{V,E_A}}$	$(\mathrm{mm\ d^{-1}})^2$	Inter-annual variance of actual evaporation.
	$C_{\mathrm{V,P_L}}$	$(\mathrm{mm\ d^{-1}})^2$	Inter-annual variance of P_L.
	$C_{\mathrm{V,P_M}}$	$(\mathrm{mm\ d^{-1}})^2$	Inter-annual variance of P_{M}.
	$C_{\mathrm{V,WD_{max}}}$	mm^2	Inter-annual variance of the maximum water deficit.
	$C_{\mathrm{V,F(Snow)}}$	-	Inter-annual variance of fraction of precipitation falling as snow.
	$C_{\mathrm{V,P_S}}$	$(\mathrm{mm\ d^{-1}})^2$	Inter-annual variance of snowfall.
	$C_{\mathrm{V,N(P_L)}}$	d^2	Inter-annual variance of days with P_{L}.
	$\mathrm{C}_{\mathrm{V,N(WD)}}$	d^2	Inter-annual variance of the annual days with water deficit.
	$C_{\mathrm{V,t(P_{L,max})}}$	d^2	Inter-annual variance of the doy with maximum P_{L}.
	$C_{\mathrm{V,t(WD_{first})}}$	d^2	Inter-annual variance of doy timing of first water deficit.
	$C_{\mathrm{V,t(WD_{last})}}$	d^2	Inter-annual variance of doy timing of last water deficit.
	$C_{\mathrm{V,t(WD_{max})}}$	d^2	Inter-annual variance of circular doy timing of maximum water deficit.
Vegetation, Static			
Magnitude			
	$V_{\mathrm{F(Forest)}}$	-	Mean fraction of forest cover.
	$V_{\mathrm{F(Shrub)}}$	-	Mean fraction of shrub cover.

Category	Variable Name	Unit	Description
	~~$V_{\mathrm{F(Open)}}$~~	-	Mean fraction of ~~open areas~~ areas with little or no vegetation.
	~~$V_{\mathrm{F(LowV)}}$~~		
	$V_{\mathrm{F(Wetland)}}$	-	Mean fraction of wetland.
	$V_{\mathrm{F(TC)}}$	-	Mean fraction of tree cover.
	V_{LAI}	-	Median LAI.
~~Seasonality~~			
	V_{NDVI}	-	Median NDVI.
Vegetation, Dynamic			
Seasonality			
	$V_{\mathrm{R(LAI_{Parde})}}$	-	Range of Parde coefficients calculated with LAI.
	$V_{\mathrm{R(NDVI_{Parde})}}$	-	Range of Parde coefficients calculated with NDVI.
Inter-annual Variance			
	$V_{\mathrm{V,LAI}}$	-	Inter-annual variance of the median LAI.
	$V_{\mathrm{V,NDVI}}$	-	Inter-annual variance of the median NDVI.
Soil, Geology, Topography			
Magnitude			
	$S_{\mathrm{F(SG)}}$	-	Mean sand and gravel fraction.
	$S_{\mathrm{F(SC)}}$	-	Mean silt and clay fraction.
	$S_{\mathrm{F(OC)}}$	-	Median soil organic content fraction.
	S_{TWAC}	mm	Median soil total water content.
	S_{SBD}	g cm^{-3}	Median soil bulk density.
	S_{RD}	cm	Median soil root depth.
	$G_{\mathrm{F(SS)}}$	-	Mean Sandstone fraction.
	$G_{\mathrm{F(MS)}}$	-	Mean Mudstones fraction.
	$G_{\mathrm{F(C)}}$	-	Mean Carbonates fraction.
	$G_{\mathrm{F(LP)}}$	-	Mean low permeable stones fraction.
	$G_{\mathrm{F(Perm_{VL})}}$	-	Mean Very Low Permeability fraction.

Category	Variable Name	Unit	Description

3.2 k-means clustering

To identify hydrological similarity between our catchments in the form of multiple HRTs, ~~a~~ k-means clustering~~was applied~~, a technique frequently used in comparable studies, was applied (Sawicz et al., 2011; Kuentz et al., 2017; Brunner et al., 2020; Almagro et al. . k-means clustering groups catchments by dividing the dataset into a predefined number of clusters (here: HRTs), each represented by a centroid. The algorithm iteratively assigns each catchment to the nearest centroid based on Euclidean distance, and then updates the centroid positions to minimise the within-cluster sum of squared distances. Because the number of clusters k must be specified beforehand, it is necessary to evaluate different choices of k to identify a solution that balances cluster compactness and interpretability. Two common evaluation metrics are the Elbow Method and the Silhouette Score. The Elbow Method assesses how the total within-cluster variance decreases with increasing k, with a breakpoint, i.e. an "elbow" indicating a point beyond which additional clusters do not provide substantial improvements. It is reported as inertia, which is the sum of the squared distances to the closest cluster centroid, for each catchment. The Silhouette Score, on the other hand, quantifies how similar each catchment is to its assigned cluster compared to other clusters, with higher values reflecting better-defined and more distinct clusters. The k-means clustering algorithm was run for a range of k from 2 to 20. The most suitable number of clusters and thus HRTs was determined using the Elbow Method and Silhouette Score metrics. To ensure that all hydrological signatures contributed equally to the clustering, the variables were normalised using min–max scaling to account for their different units and magnitudes. The k-means clustering resulted in a set of k hydrologically similar groups (HRTs), with each catchment assigned to one of these HRT classes (labelled 1 to k). These HRT class labels were subsequently used as the target variable in the Random Forest classification model.

3.3 Random Forest classification

The Random Forest classification model, implemented with the *scikit-learn* library, was used here with the climate- and landscape attributes as input features to predict the membership of each catchment to one of the classes of hydrological response types (HRT) (section 4.2) in a supervised multiclass classification problem. The idea behind this is that each HRT class represents a unique combination of hydrological characteristics as defined by the hydrological signatures. By predicting catchment membership to the class of HRT from climate- and landscape attributes, it can then be analysed to which extent hydrological classes can be predicted based on landscape and climate, as well as which climate- and landscape attributes are more relevant for shaping hydrological response pattern. ~~In contrast to predicting~~ While Random Forest regression models have been successfully used to predict individual hydrological signatures ~~with a Random Forest regression model~~ (e.g., Addor et al.,

2018; Almagro et al., 2024), predicting membership to a HRT class, ~~and thus to~~ representing a multi-dimensional composite of hydrological signatures, ~~holds potential to reveal details on the reasons for~~ provides a complementary perspective that may help reveal the drivers of overall hydrological behaviour and ~~spatial differences thereof~~its spatial variability.

As the choices made in the modelling process can substantially affect the results and interpretation thereof, a stepwise approach with many alternative model set-ups was applied here. In a first step, each model was tuned through iterative adjustment of hyperparameters in a Monte Carlo sampling strategy with 5000 realisations. Each realisation was evaluated based on its accuracy, i.e. the fraction of correctly classified catchments over the total sample (Acc = true positives/total sample) for both training and validation. From the set of all realisations, a solution was chosen that balances a good training ($Acc_t \geq 0.6$) with a robust validation performance Acc_{val} ($\Delta Acc_t - Acc_{val} \leq 0.1$) to limit overfitting (for the retained solution's model parameter set see Supplementary Material S7). In the second step, the retained solution was used to further train and evaluate the model using a 10-fold cross-validation, with 90 % of the data, i.e. catchments in a HRT, used for training and 10 % for testing in each fold. Due to the unequal number of catchments in the individual HRTs, the ~~HRT~~ HRTs were weighted according to their respective number of catchments in the model. After that initial model set-up that predicted membership to 10 individual HRT classes, the same sequence of steps was repeated using 6, 7, 8 and 9 HRT classes, respectively, to analyse the sensitivity of the model and to assess how well the HRTs are defined.

To evaluate in a last step, the relative influences of climate and landscape on hydrological behaviour, five experiments were conducted in addition to the initial experiment (1-CL). In these experiments, the Random Forest classification model was trained and tested using only climate- or different landscape-attributes (Table 3). In Experiment 2-C, exclusively the 47 climate attributes were used for model training and testing, while Experiment 3-VLC used the 11 vegetation and landcover attributes. As some of the vegetation signatures may directly reflect short-term climatic variability, Experiment 4-VLS used only the 7 static vegetation and landcover attributes. Experiment 5-SGT included the 21 signatures related to soil, geology and topography, and Experiment 6-A used the 5 anthropogenic attributes.

Table 3. Random Forest classification experiments

Experiment	Name	No. Input features
1-CL	Climate & landscape	84
2-C	Climate	47
3-VLC	Vegetation & landcover	11
4-VLS	Vegetation & landcover Static	7
5-SGT	Soil, geology & topography	21
6-A	Anthropogenic	5

For the analysis of the climate- and landscape-controls on the HRTs in the six experiments, model performances are reported as averages over the cross-validation folds. Besides the accuracy Acc, model performance of each experiment was evaluated

using receiver operating characteristic curves (ROC curves), the area-under-the-curve metric (A_{UC}) and the confusion matrix for a more nuanced picture of the true positive, false positive and false negative rates of the classification. The ROC curve illustrates the model performance across varying classification thresholds by plotting the true positive rate against the false positive rate, demonstrating the relative trade-off between the correctly and incorrectly classified instances. The classification threshold determines how predicted probabilities are turned into class labels. For example, if the threshold is 0.5, a catchment with a predicted probability of ≥ 0.5 of belonging to class A, will be classified as A. Changing this threshold shifts the balance between correct and incorrect classifications, and the average of all the different true and false positive rates makes up the ROC curve. An overall robust and well-performing model approaches the upper left corner (high true positive rate R_{TP} and low false positive rate R_{FP}), reaching high A_{UC} values, while poorer model performance with lower A_{UC} is reflected by curves closer to the diagonal, which represents random guessing.

However, dealing with a multiclass prediction, the ROC curves and A_{UC} are averaged across all HRT classes and folds, and therefore only provide partial information on the model performance. To assess the final classification outcomes, confusion matrices were used. These compare the true HRT class to the predicted HRT class, with correct predictions, i.e. true positive rates, for each HRT plotting along the diagonal. Each row represents the true HRT, and each column the predicted HRT. Off-diagonal rows therefore, indicate the false negative rates, while off-diagonal columns show the false positive rates. Together, this provides insights into which HRTs are more robustly predicted than others and which HRTs are more susceptible to misclassifications.

For each experiment, permutation feature importance scores from the associated Random Forest classification model were used to evaluate the relative contributions of the climate- and landscape attributes to the model performance. Feature importance reveals to which extent a feature (here: climate- or landscape attribute) influences the classification outcome. It is quantified for each feature as the decrease in predicted accuracy when this specific feature is removed from the model. To ensure that feature importance scores were robust, it was calculated 10 times for each of the folds. The reported feature importance is the average of the 10 calculations across all folds.

4 Results

4.1 Classes of hydrological response types across Europe

Preliminary k-means clustering runs with k ~~=~~ ranging from 2 to 20 (representing predefined numbers of HRTs ~~to identify the most robust number of HRTs warranted by the~~ based on the available hydrological signatures~~provides no unambiguous and clear number of well-defined~~) did not yield a clear or unambiguous optimal number of HRTs. While the Inertia metric of the Elbow Method indicates no apparent break but rather a smooth decline with an increasing number of HRTs (Fig.2a), the Silhouette Score metric suggests a major break at k = 4 and another minor break at 12 HRTs (Fig.2b). The generally low values of the Silhouette Score, i.e. < 0.25, further indicate that it is unlikely that the number of well-defined HRTs exceeds that range.

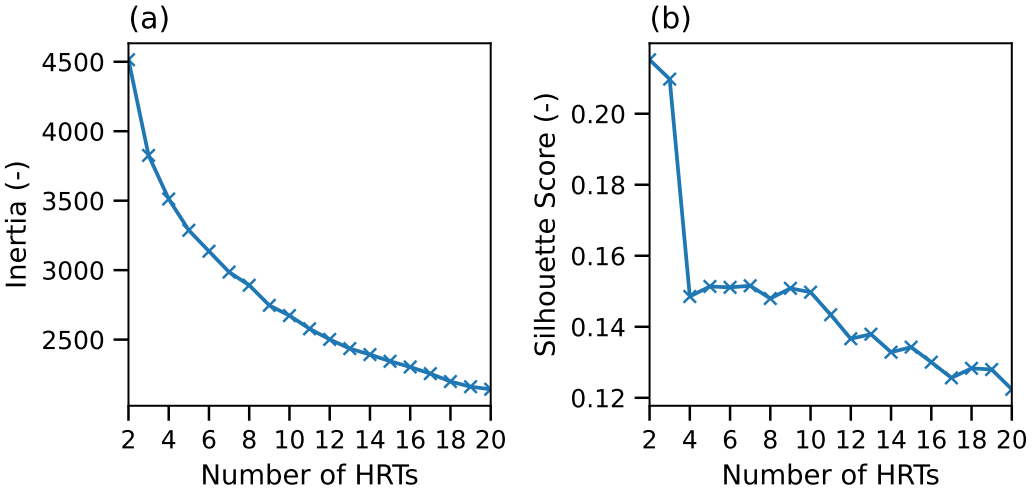


Figure 2. Assessment of cluster robustness using k-means clustering. (a) Elbow Method results for k = 2-20 HRTs, used to evaluate the optimal number of hydrological response types. (b) Silhouette Scores for the same range of HRTs, providing an additional metric of cluster cohesion and separation.

As a balanced choice to allow for some detail while not inflating the number of HRTs warranted by the data, the results are thus hereafter reported with a focus on 10 HRTs. However, to test the significance of the choice of HRTs numbers on the results, the results of k-means clustering with 6, 7, 8, and 9 HRTs were compared to the 10 HRTs, but no substantial difference was found between the different HRT numbers, see the Supplementary Material Fig.S3.

~~Overall, Figure~~ Overall, Fig.2 suggests that there is little evidence that the set of 40 hydrological signatures used for this analysis contains sufficient information to robustly distinguish more than 10 different types of HRTs for the 7175 study catchments across Europe. The results of a Principal Component Analysis of the signatures further support this interpretation. Indeed, as shown in ~~Figure~~ Fig.3a, the first two principal components (PC) account for merely $\sim 38\%$ of the variance in the signatures, and at least 19 principal components are required to explain $> 90\,\%$ of the variance (Fig.3c). This ~~entails~~ suggests that the signatures cannot be effectively compressed due to a low level of redundancy in their respective information contents and suggesting that signatures of individual catchments exhibit substantial scatter but not in a systematic enough manner. The latter implies that no detailed pattern of similarity that would go beyond ~ 10 HRTs could here be identified based on k-means clustering. As a further consequence, the within-HRT variability of individual signatures can be substantial. Showing the distributions of several selected signatures, ~~Figure 4~~ Fig.reffig:3 illustrates that some signatures, such as the baseflow index H_{BFI} (Fig.4e), the Gini coefficient H_{GC} (Fig.4f) or the flashiness index H_{RBI} (Fig.4i), exhibit a clear contrast across the HRTs and limited within-HRT variability which can also be seen by the standard deviations of the signatures around the normalised means of the individual HRTs in ~~Figure~~ Fig.5. In contrast, many other signatures such as the days between highest and lowest flow $H_{\mathrm{N(Q_H-Q_L)}}$ (Fig.4j) or the autocorrelation coefficient $H_{\mathrm{AC_{30}}}$ (Fig.4g) are characterised by substantial within-HRT variability, lower across-HRT contrast and thus little discriminatory power which can be seen by the limited differences

between the standard deviations of these signatures across the HRTs in Fig.5. Note that the distributions of all hydrological signatures in the individual HRTs is provided in the Supplementary ~~Figure~~ Material Fig.S2.

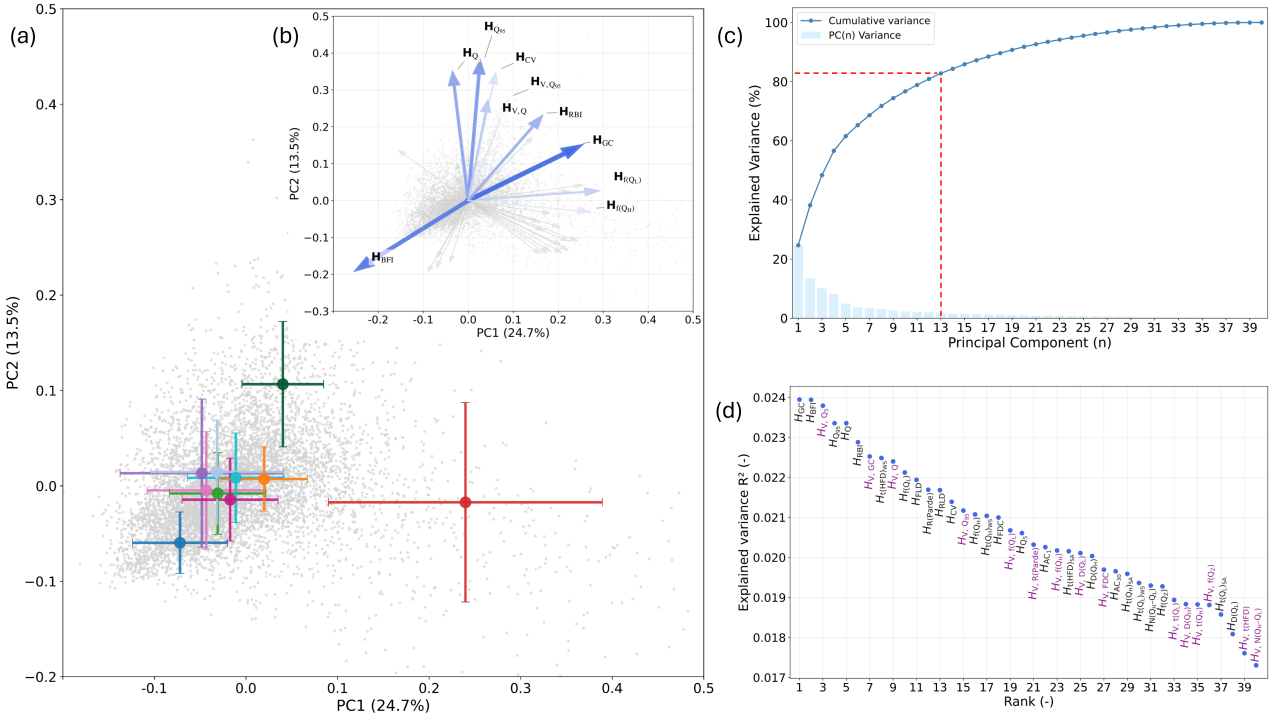


Figure 3. Principal component analysis (PCA) of the 40 hydrological signatures across the HRTs. (a) explained variance of PC1 (24.7%) and PC2 (13.5%) for each catchment (grey), with HRT means and standard deviations shown in their respective colour. (b) loadings of the hydrological signatures, shown for PC1 and PC2. Arrow thickness and colour intensity reflect each signature's ranked explained variance based on the first 13 PCs (see panel d), accounting for just above 80% of the cumulative explained variance. Thicker and more intense coloured arrows indicate more informative signatures. (c) cumulative explained variance across all principal components. The explained variance of the first 13 PCs is highlighted by the red dashed lines. (d) ranked explained variance of the first 13 PCs, which together account for approximately 80% of the total variance. Signatures related to variance are highlighted in purple.

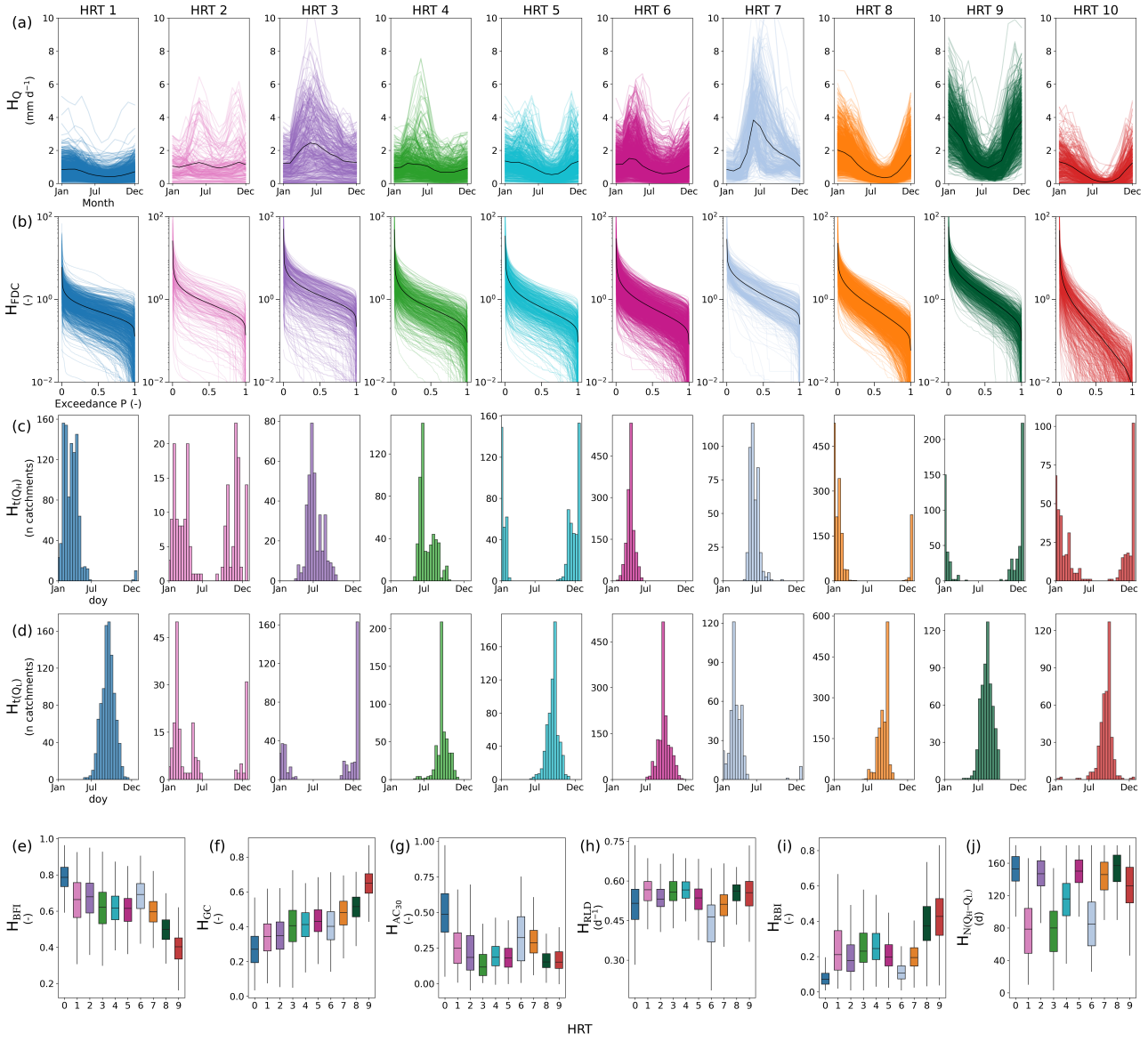


Figure 4. Visualisation of 10 out of the 40 hydrological signatures across hydrological response types (HRTs), the remaining 30 signatures can be seen in the Supplementary Material Fig.S2. Panels (a)-(d) show four selected signatures for each HRT (one row per signature, one column per HRT), (a) streamflow (H_{Q}), (b) flow duration curves (H_{FDC}), (c) high flow timing ($H_{\mathrm{t(Q_H)}}$) and (d) low flow timing ($H_{\mathrm{t(Q_L)}}$), highlighting within- and between-cluster variability. In panels (a) and (b), the black line indicates the mean value of mean discharge (Q) and mean flow duration curve (FDC), respectively. Panels (e)-(j) display boxplots of six additional hydrological signatures, (e) baseflow index (H_{BFI}), (f) Gini coefficient (H_{GC}), (g) autocorrelation ($H_{\mathrm{AC_{30}}}$), (h) rising limb density (H_{RLD}), (i) flashiness index (H_{RBI}) and (j) days between highest and lowest flow ($H_{\mathrm{N(Q_H\text{-}Q_L)}}$), summarising their distribution across HRTs.

The most informative hydrological signatures, ranked by their cumulative variance across the first 13 PCs (~~~ 80%~~ ~ 80% of variance explained) as shown in ~~Figure~~ Fig.3b and d, included the Gini coefficient H_{GC} as metric of flow seasonality (1st rank), the baseflow index H_{BFI} (2nd) as metric of flow stability, the mean flow H_{Q} (5th) as metric of overall flow magnitude but also descriptors of inter-annual variance, such as $H_{\mathrm{V,Q_5}}$ as metric of variance of low-flow magnitude (3rd), and the variance of the Gini coefficient $H_{\mathrm{V,GC}}$, as metric of inter-annual differences in flow seasonality (7th). The picture that emerges illustrates that while a few signatures of inter-annual variability are among the top-ranked ones, the majority of them, i.e. 10 out of 15, ranks in the bottom half of all signatures. HRTs are thus dominantly defined by signatures of averages, in particular flow seasonality, stability and magnitude (Fig.3d).

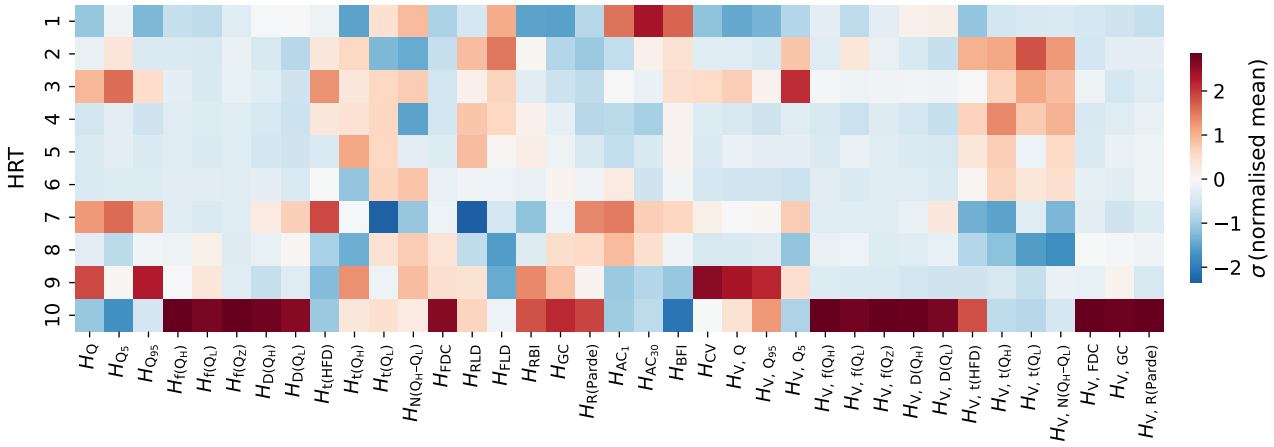


Figure 5. Standard deviation of the normalised mean for each hydrological signature shown for each HRT. Positive deviations are shown in red and negative deviations in blue, with colour intensity indicating the magnitude of the deviation.

Building on these defining signatures, the 10 HRTs exhibit a wide range of hydrological behaviours, with varying degrees of distinctness (Fig.4~~).~~, Table 4).

For example, HRTs 1 ~~,~~ (stable, weak, winter regime), 7 ~~,~~ (wet, strong, spring regime), 8 ~~,~~ (wet, strong, late-winter regime), 9 ~~and~~ (wet, strong, early-winter regime) and 10 ~~show strongor extreme seasonal and flow-regime characteristics, which makes them~~ (dry, intermittent, winter regime) are readily distinguishable (Fig.4a–d) due to their stable, strong or extreme flow-regime characteristics. HRT 1 represents a highly stable regime with low mean flows, high baseflow (H_{BFI}) contributions and the lowest flashiness index (H_{RBI}) (Fig.4a, e, i). In contrast, HRT 10 shows an extreme regime characterised by very steep flow duration curves (H_{FDC}) and strong seasonality as illustrated by the highest Gini coefficients (H_{GC}) (Fig.4b, f), while HRTs 7–9 exhibit pronounced seasonal peaks and lows and highly consistent timing of high flows ($H_{\mathrm{t(Q_H)}}$) (Fig.4a, c, d). In contrast, less distinct HRTs, such as HRT 2 ~~,~~ (transitional, low, multi-seasonal regime), show a wide spread in the timing of high and low flows ($H_{\mathrm{t(Q_H)}}$, $H_{\mathrm{t(Q_L)}}$) and in the slopes of the flow duration curves (H_{FDC}), whereas HRT 5 (transitional, moderate, winter regime) remains close to the overall average for most signatures (Figs.4, 5). However, the within-HRT variability is not uniform, each HRT shows limited variability in some signatures and substantial spread in others. For example, HRT 1 shows

low variability in the baseflow (H_{BFI}) and flashiness (H_{RBI}) indices but high variability in autocorrelation($H_{\mathrm{AC}_{30}}$) and rising limb density (H_{RLD}). Similarly, HRT 7 has tightly grouped flashiness (H_{RBI}) and high-flow timing ($H_{\mathrm{t(Q_H)}}$), but a wide spread in rising limb density (H_{RLD}) and the days between high and low flows ($H_{\mathrm{N(Q_H\text{-}Q_L)}}$). Less distinct classes, such as HRT 2, show a similar variation in signature values, with relatively low variability in rising limb density (H_{RLD}) but high variability in baseflow (H_{BFI}) and flashiness (H_{RBI}) indices and the timing of flows ($H_{\mathrm{t(Q_H)}}$, $H_{\mathrm{t(Q_L)}}$) (Fig.4c, d, e, i). These heterogeneous patterns illustrate why detailed descriptions of the HRTs beyond the major contrasts remain challenging.

Despite the difference in within-HRT hydrological signature variation, broad hydrological patterns emerge across the HRTs and form the basis for their classification into flow-regime types (Table 4). Key characteristics, reflecting the dominant signatures from PCA analysis (Fig.3d), include the magnitude of average discharge, the strength and shape of the seasonal cycle, and the predominant season of high flows. Using these criteria, the HRTs range from stable to transitional, wet, and dry regimes, and from weakly seasonal to strongly seasonal systems with winter-, spring-, summer-, or multi-seasonal peak-flow timing. The geographical distribution of the HRTs as shown in ~~Figure~~ Fig.6, reflects some of these large-scale hydrological differences but does not show consistent spatial clustering. Certain HRTs are geographically concentrated: for example, HRT 7 (wet, strong, spring regime) is largely confined to the Alps, northern Scandinavia and Iceland (Fig.6h), while HRT 9 (wet, strong, early-winter regime) is mainly located in the UK and Ireland (Fig.6j). Other well-defined classes, such as HRT 10 (dry, intermittent, winter regime) and HRT 8 ~~,~~ (wet, strong, late-winter regime), are more widely distributed across Europe. Likewise, HRT 1 (stable, weak, winter regime) and HRT 6 (transitional, moderate, spring regime) span a broad latitudinal gradient from southern Spain to Finland, while HRT 3 ~~and~~ (wet, moderate, summer regime) and HRT 4 (transitional, moderate, summer regime) are clustered in transitional regions between high and low elevation regions. Thus, spatial proximity may explain some, but far from all patterns in the geographical distribution of HRTs.

Table 4. Overview of the 10 hydrological response type classes. The labels focus on hydrological signatures H_{Q} and $H_{\mathrm{t(Q_H)}}$ (Fig.4a and c). The characteristics focus further on those most distinct within each HRT, like the contribution of baseflow, timing of highest and low flows and flashiness, while geography can be visually checked in Fig.6.

HRT Class	No of catchments	Label	Characteristics	Geography
HRT 1	976	Stable, weak, winter regime	Stable regime type with a large baseflow contribution and low average streamflow. Flow seasonality is low, with limited variation throughout the year.	Mostly low-lying and flat areas in central Europe, with some catchments in northern and western Europe.
HRT 2	184	Transitional, low, multi-seasonal regime	Moderate average streamflow with low seasonality. High and low flows often occur in consecutive seasons and vary between catchments. In several catchments, there are two distinct high-flow peaks throughout the year: spring and winter.	Found both in higher as well as lower elevation areas, spanning from southern to northern Europe.
HRT 3	369	Wet, moderate, summer regime	Relatively high average streamflow and baseflow index. Distinct summer flow peaks and moderate seasonality.	The majority of catchments are located in mid- to high- elevation areas in the Alps, the Carpathians and Spain.
HRT 4	522	Transitional, moderate, summer regime	Similar to HRT 3, but with lower average streamflow. The highest and lowest annual flows occur in close temporal proximity.	The majority of catchments are located in mid- to high- elevation areas in the Alps, the Carpathians and the Pyrenees.
HRT 5	658	Transitional, moderate, winter regime	Moderate average streamflow with winter-dominated flow peaks and low flows in autumn. Catchments with multiple high-flow peaks during the year also occur, with peaks at the end of the year and in spring.	Spread across central, western and southern Europe.
HRT 6	1460	Transitional, moderate, spring regime	Moderate average streamflow, with very distinct spring high-flows and autumn low-flows, causing pronounced seasonality.	Concentrated in central Europe, with catchments also in northern and southern Europe, spanning both high and low elevation areas.
HRT 7	421	Wet, strong, spring regime	Distinct spring flow peaks, with high average streamflow throughout the year and low winter flows sustained by baseflow.	Found in cold and high-elevation regions of northern Europe, Iceland and the Alps.
HRT 8	1557	Wet, strong, late-winter regime	High winter flows and low autumn flows, with high Q_{95} values indicating sustained high flow conditions and strong seasonality.	Located in western Europe, with additional groups in southern Europe and southern Scandinavia, covering variable elevations.
HRT 9	599	Wet, strong, early-winter regime	Highest average streamflow and coefficient of variation, exhibiting flashy behaviour, with the highest Q_{95} (high-flows) and relatively high Q_5 (low-flows).	Predominantly coastal catchments in Ireland and the UK, with a few groups in central Europe.
HRT 10	429	Dry, intermittent, winter regime	Extreme seasonal regime with high winter flows and low flows in summer and autumn. Flashy hydrological behaviour, with the highest frequency of both high and low flow events. Only cluster with a significant number of catchments with intermittent flow.	Catchments are mostly found in southern Europe, with some located in western Europe.

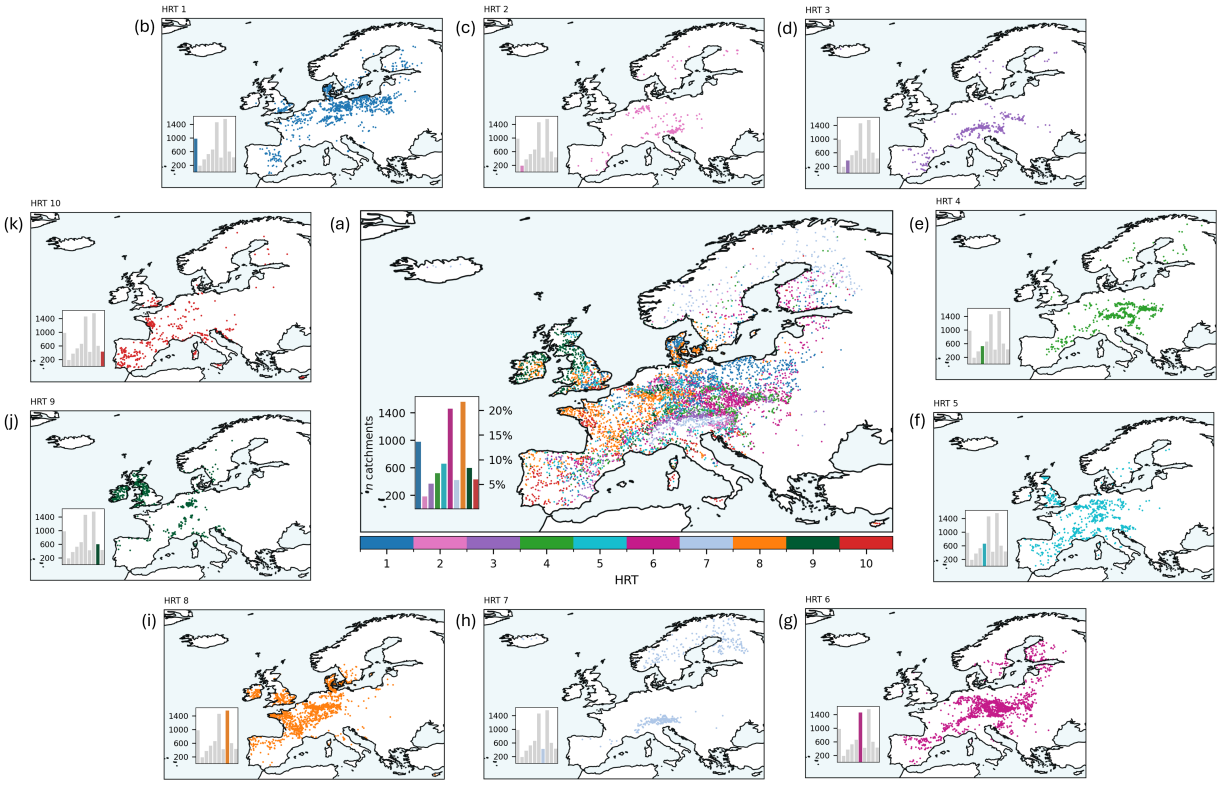


Figure 6. Classification of 7175 catchments into 10 HRT based on 40 hydrological signatures. The HRTs are ordered by increasing slope of the flow duration curves. (a) presents all HRTs together, (b) to (k) shows the spatial distribution of each HRT individually.

4.2 Prediction of HRTs by climate- and landscape attributes and drivers of streamflow behaviour

The initial hyperparameter tuning approach for model selection, based on 10 classes of HRTs and all 84 climate- and landscape attributes as input features of the Random Forest classification model to predict membership of catchments to individual HRTs, resulted in training accuracies $Acc_t \sim 0.45$–1 (Supplementary Material Fig.S4a). In contrast, the validation accuracy Acc_{val} remains < 0.70 for all realisations. These overall reductions of accuracies between calibration and validation indicate substantial model overfitting, and thus limited predictive power as Acc_t increases. As expected, validation accuracies increase for model set-ups based on a lower number of HRTs (k = 6-9), as shown in ~~Supplementary Material~~ Fig.S4b-e. However, with Acc_{val} not exceeding 0.75 for any of the set-ups, the improvements are very minor, while the overall pattern remains unchanged. The opposite, with partly substantial declines in predictive power, was observed with model set-ups using higher numbers of HRTs (not shown). To thus preserve both, at least moderate levels of detail in the distinction of HRTs as well as at least moderate levels of model predictive power, 10 HRTs were eventually used as a balanced choice for the subsequent analysis. More specifically, the hyperparameter set resulting in $Acc_t = 0.70$ and $Acc_{val} = 0.60$ for these initial runs (~~Supplementary Material~~ Fig.S4a) was selected for further analysis.

The first of the six full experiments, predicting membership of individual catchments to the 10 different HRTs based on the complete set of 84 climate- and landscape attributes, i.e. 1-CL, resulted in rather stable results across the 10 cross-validation folds, with $Acc_t = 0.68 - 0.69$ and $Acc_{test} = 0.58 - 0.60$ (Fig.7a, Supplementary Material Fig.S5). A more nuanced analysis of the results reveals that, at the chosen classification threshold, the true positive rates, i.e. the fraction of catchments whose membership to a specific HRT is correctly predicted, reaches with R_{TP} = 0.57-0.61 comparable values across all 10 folds as illustrated by the individual dots in Fig.7b. It also illustrates that the false positive rates, i.e. the fraction of catchments with a specific HRTs whose membership to other HRTs is incorrectly predicted, remains very low with R_{FP} = 0.04-0.05. These values describe the model's performance when turning predicted probabilities into hard class labels, and are also reflected in the confusion matrices (Fig.8). Independently of this threshold, the ROC analysis yields an A_{UC} = 0.93, indicating that the combined climate- and landscape attributes used in the 1-CL experiment have predictive power that is in any case substantially higher than a random guess, represented by the 1:1 line in ~~Figure~~ Fig.7b and the associated A_{UC} = 0.5.

A further stratification of the results exposes pronounced differences between the individual 10 HRTs. Membership to some HRTs, such as HRTs 7 (wet, strong, spring regime) and 9 ~~,~~ (wet, strong, early-winter regime), is with Acc_{test} of up to 0.89 and A_{UC} of up to 0.99 very well predicted (Fig.8a, Supplementary Material Fig.S5). These well-predicted HRTs are generally well defined by several hydrological signatures that are distinct to those of other HRTs (Fig.4, ~~Supplementary Material~~ Fig.S5). While HRT 7 is, for example, characterised by elevated flow magnitudes, in particular H_{Q} and $H_{\mathrm{Q_5}}$, as well as markedly different timing characteristics, including the latest half-flow day $H_{\mathrm{t(HFD)}}$ and the earliest low flows $H_{\mathrm{t(Q_L)}}$, HRT 9 equally stands out with high H_{Q}, but also exhibits the highest $H_{\mathrm{Q_{95}}}$, the largest flow variability H_{CV} and the earliest half-flow day $H_{\mathrm{t(HFD)}}$ (~~Supplementary Material~~ Fig.S2). This is contrasted by membership to other HRTs that are less well predicted. This is in particular the case for HRT 5 (transitional, moderate, winter regime) with Acc_{test} = 0.34 (Fig.8a) and A_{UC} = 0.83 (~~Supplementary Material~~ Fig.S6). As a transitional HRT, none of its hydrological signatures is clearly discernible from the other HRTs, making an effective discrimination difficult (Fig.4, ~~Supplementary Material~~ Fig.S6).

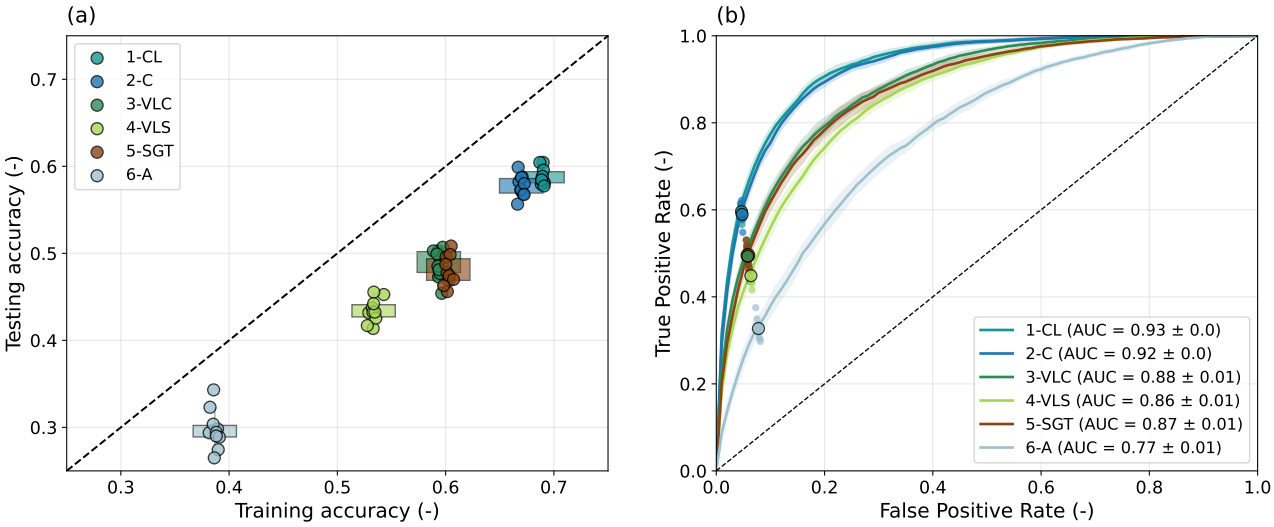


Figure 7. (a) testing versus training accuracy for each experiment. Each circle represents one of the 10 cross-validation folds. The axes are extended slightly beyond the data range to show the 1:1 reference line. Boxplots beneath the points summarise the distribution of testing accuracies across the 10-fold cross-validation for each experiment. (b) the averaged ROC curves for each experiment are displayed, computed over the 10 cross-validation folds. For each fold, the true positive rate (T_{PR}) and false positive rate (T_{PR}) at the actual classification threshold are shown as small dots. The larger dot represents the average T_{PR} and T_{PR} across all folds, corresponding to the aggregated performance reported in the confusion matrices.

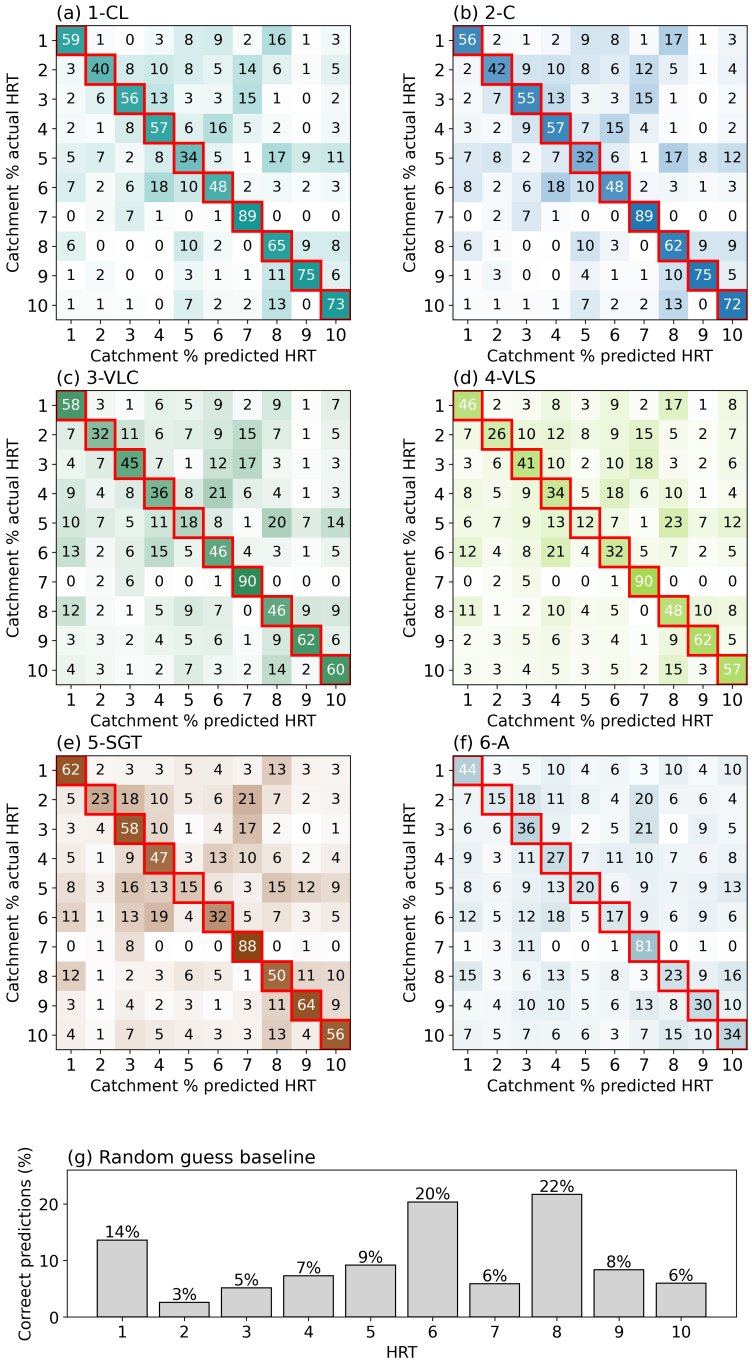


Figure 8. Confusion matrices for each hydrological response type (HRT), normalised by row (i.e., by true HRT). The diagonal elements show the percentage of catchments correctly predicted as their true HRT. Off-diagonal elements in each row represent the percentage of that HRT misclassified as other HRTs (false negatives), while off-diagonal elements in each column indicate false positives. All values are rounded to whole percentages; therefore, individual rows may not sum to exactly 100 %. For example, (a) shows the confusion matrix for experiment 1-CL: the first cell in row 1 (row 1, column 1) gives the percentage of HRT 1 correctly predicted as HRT 1, whereas the entry in row 1, column 2 indicates the proportion of HRT 1 predicted as HRT 2. Similarly, the value in row 2, column 1 represents the percentage of HRT 2 incorrectly predicted as HRT 1. (g) baseline scenario showing the percentage of catchments expected in each HRT class under random guessing, based on the class distribution.

From a spatial perspective, no clear pattern emerges in the correctly predicted HRTs (Fig.9). However, some regions show an overrepresentation of misclassifications, notably northern Spain, the Paris Basin, South-East England, and southern Poland (Fig.9). While HRT 10 (dry, intermittent, winter regime) is predicted fairly homogeneously across its domain (Fig.9j), the misclassified catchments span nearly all other HRTs, with the largest share belonging to HRT ~~8.~~ 8 (wet, strong, late-winter regime). Other well-predicted classes include HRTs 7 ~~,~~ (wet, strong, spring regime), 8, and 9 ~~,~~ (wet, strong, early-winter regime), with HRTs 7 and 9 being the most spatially homogeneous. These HRTs occupy relatively compact hydro-climatic domains, in contrast to HRT 8, which is generally well predicted but has misclassifications distributed widely across Europe (Fig.9g, h, i).

HRTs 2 ~~,~~ (transitional, low, multi-seasonal regime), 5 ~~,~~ (transitional, moderate, winter regime), and 6 (transitional, moderate, spring regime) display more spatially variable predictive performance (Fig.9b, e, f). For HRT 2 and HRT 6, correctly classified catchments occur in dense clusters, whereas the misclassified ones are scattered across Europe. In contrast, both correct and incorrect predictions of HRT 5 are widespread. The most frequent confusions are HRT 2 being misclassified as HRT 7 ~~,~~ (wet, strong, spring regime), HRT 5 as HRT 8 ~~,~~ (wet, strong, late-winter regime), and HRT 6 as HRT 4 ~~,~~ (transitional, moderate, summer regime), consistently across all experiments (Fig.9b, e, f; Supplementary Material Figs.S7-S11). HRT 5 and HRT 8 share similarities in timing (Fig.4c, d), but otherwise show limited resemblance (Fig.5). HRT 6 and HRT 4 also overlap in the timing of their high and low flows (Fig.4c, d) and exhibit similar seasonal patterns (Fig.4a). In contrast, a comparison of signatures between HRT 2 and HRT 7 (Fig.5) reveals no striking similarities, suggesting that their mutual misclassification may result from similarities in climatic or landscape controls rather than hydrological behaviour alone.

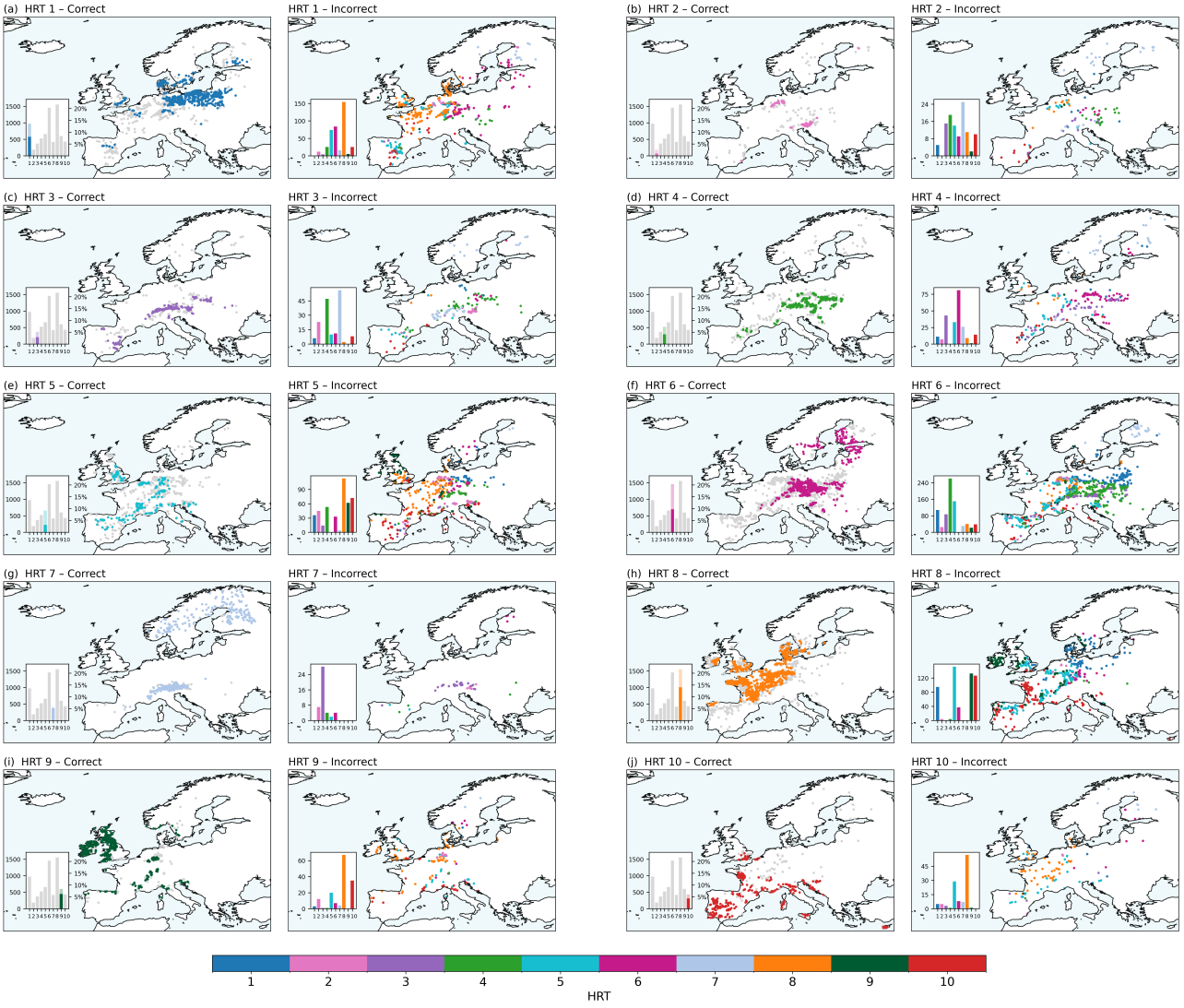


Figure 9. Maps showing correct and incorrect classification results for experiment 1-CL. In the "correct" panels, only correctly classified catchments are shown in their HRT colours, while incorrectly classified catchments appear in grey. In the "incorrect" panels, only the misclassified catchments are displayed and coloured according to their *true* HRT, allowing visual assessment of which clusters are most frequently misclassified.

The relative roles of the individual climate- and landscape attributes as drivers of streamflow behaviour were then analysed by comparing their feature importance ~~.~~(Fig.10. Overall, the seasonality of precipitation $C_{\phi(\mathrm{P})}$ (1^{st} rank), the seasonality of potential evaporation $C_{\phi(\mathrm{E_P})}$ (2^{nd}) as well as the fraction of precipitation falling as snow $C_{\mathrm{F(Snow)}}$ (3^{rd}) emerge as the strongest predictors of membership to classes of HRTs and thus as most dominant controls on streamflow behaviour (Fig.10a). Next to these three climate attributes, almost half of the top 20 ranking predictors are landscape attributes. Most notably, several

descriptors of vegetation attributes such as the average NDVI V_{NDVI} (4^{th}), its inter-annual variance $V_{\mathrm{V,NDVI}}$ (7^{th}) or average tree cover fraction $V_{\mathrm{F(TC)}}$ (8^{th}) stand out, while the fraction of carbonate rocks $G_{\mathrm{F(C)}}$ (11^{th}) is the highest-ranking attribute from the soil, geology and topography group. In addition, the fraction of agricultural land $A_{\mathrm{F(Agri)}}$) (9^{th}) appears as most relevant from the group of anthropogenic attributes.

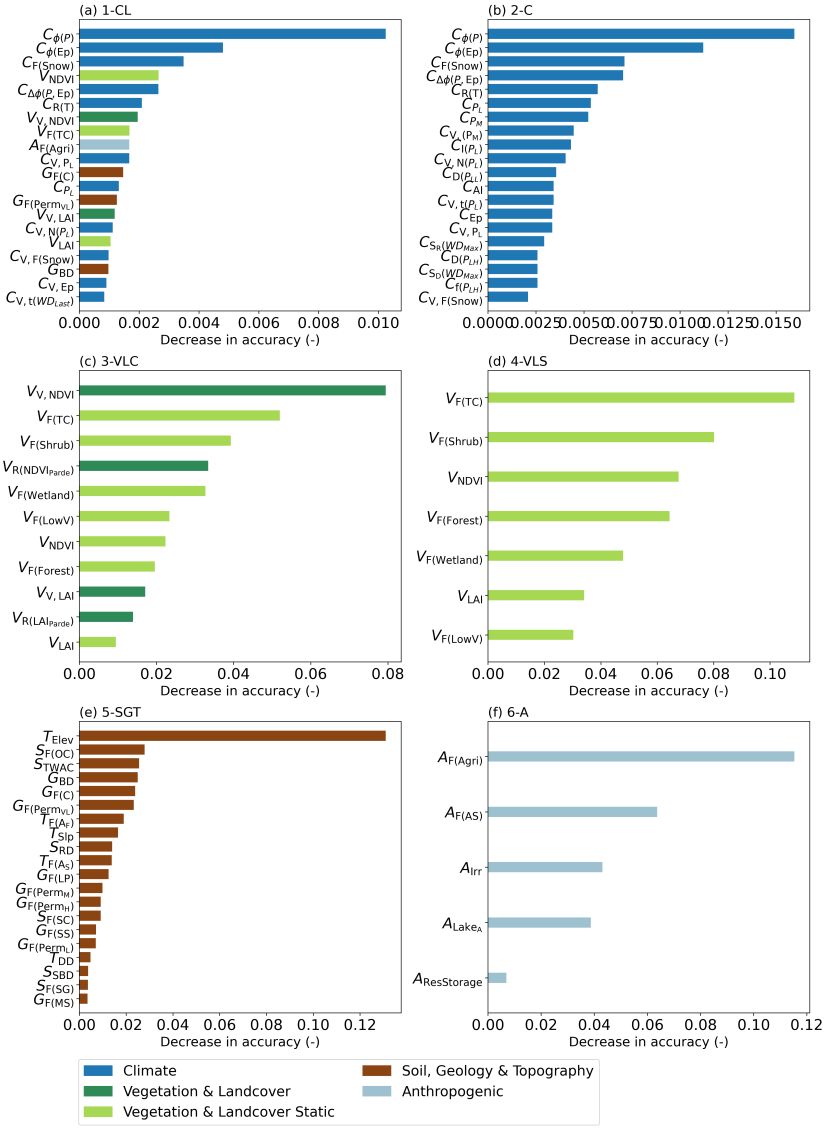


Figure 10. Median feature importance across the 10-fold cross-validation for all six experiments. Features are ranked by importance (y-axis), which is dictated by the decrease in model accuracy (Acc_{test}) by removing the corresponding attribute (x-axis). (a), (b) and (e) display only the 20 most important features. Dots represent the median importance for each feature, and black lines show the standard deviation across folds. Feature importance values are not directly comparable across experiments because they depend on the model type and the number and composition of input features; experiments with fewer predictors typically yield higher importance values per feature.

In the second experiment 2-C, predicting membership to the 10 HRTs based exclusively on the reduced set of 47 climate attributes resulted in model performances $Acc_t = 0.66 - 0.67$ and $Acc_{test} = 0.56 - 0.60$ that are not considerably different to the 1-CL experiment using the complete set of 84 climate- and landscape attributes (Fig.7a, Supplementary Material Fig.S5). Similarly, the ROC curve is with $A_{UC} = 0.92$ practically indiscernible from the one of the 1-CL experiment (Fig.7b). Not only is the overall model performance of the 2-C experiment comparable to 1-CL, but also the model's skill to predict membership to the individual HRTs is largely equivalent between the two experiments with $Acc_{test} = 0.32 - 0.89$ for the individual HRTs but also preserving the hierarchy of which HRTs are more robustly predicted than others (Fig.8b, ~~Supplementary Material~~ Fig.S6). It can further be observed that the ranking of the feature importance of, in particular, the highest ranking climate attributes is equally preserved with precipitation $C_{\phi(\mathrm{P})}$, $C_{\phi(\mathrm{E_P})}$, $C_{\mathrm{F(Snow)}}$ as well as the phase shift between P and E_P $C_{\Delta\phi(\mathrm{P,E_P})}$ and the temperature range $C_{\mathrm{R(T)}}$ also being the dominant climate controls in the 2-C experiment (Fig.10b, ~~Supplementary Material~~ Fig.S12a).

The prediction of membership to individual HRTs only based on the 11 vegetation and landcover attributes of this study in experiment 3-VLC resulted in $Acc_t = 0.59 - 0.60$ and $Acc_{test} = 0.45 - 0.51$ (Fig.7a, Supplementary Material Fig.S6) and consequently a slightly reduced $A_{UC} = 0.87 - 0.89$ (Fig.7b) across the 10 cross-validation folds. Despite these slightly lower overall accuracies, several individual HRTs could be predicted at least as well or even better with vegetation and landcover attributes here in 3-VLC than when also considering climate attributes in 1-CL and 2-C (Fig.8c, ~~Supplementary Material~~ Fig.S6). More specifically, it was found that HRTs 1 ~~,~~(stable, weak, winter regime), 6 (transitional, moderate, spring regime) and 7 (wet, strong, spring regime) are characterised by average Acc_{test} = 0.58, 0.46 and 0.90, respectively. They are thus all predicted within a margin of ± 0.02 as compared to 2-C, which is not a significant difference ($p > 0.05$). This is contrasted with other HRTs that could be less well predicted in the 3-VLC experiment, most notably HRT 4 (transitional, moderate, summer regime) ($Acc_{test} = 0.36$) and HRT 5 (transitional, moderate, winter regime) ($Acc_{test} = 0.18$). While the two top-ranking vegetation and landcover attributes in terms of feature importance, i.e. the inter-annual variance of NDVI $V_{\mathrm{V,NDVI}}$ and the tree cover fraction $V_{\mathrm{F(TC)}}$ (Fig.10c), also appear among the top 3 vegetation and landcover controls in the 1-CL experiment (Fig.10a), the overall coherence of the ranking compared to 1-CL is rather modest (~~Supplementary Material~~ Fig.S12b). The latter is reflected, amongst others, by the marked rise of the range of the Parde coefficient of NDVI $V_{\mathrm{R(NDVI_{Parde})}}$ from the lowest ranking in 1-CL to the 4$^{\mathrm{th}}$ ranking vegetation and landcover attribute. As $V_{\mathrm{R(NDVI_{Parde})}}$ is a metric of seasonality of vegetation activity, it also reflects a direct imprint of seasonal climatic conditions on vegetation, and is thus, to some extent a proxy of $C_{\phi(\mathrm{P})}$ and $C_{\phi(\mathrm{E_P})}$, the dominant climate attributes in 1-CL and 2-C.

This reasoning is further supported by the 4-VLS experiment, which uses only static vegetation and ~~land cover~~ landcover attributes. The omission of any information on temporal dynamics, i.e. seasonal as well as inter-annual, in this experiment led to further reduced prediction performances, i.e. $Acc_t = 0.53 - 0.54$, $Acc_{test} = 0.41 - 0.46$ and $A_{UC} = 0.85 - 0.87$ (Fig.7a, b, Fig.8d, Supplementary Material Fig.S6), with tree cover fraction $V_{\mathrm{F(TC)}}$ remaining the most dominant static vegetation and landcover attributes (Fig.10d). In contrast, basing the predictions exclusively on soil, geology and topography attributes in experiment 5-SGT, results in accuracies $Acc_t = 0.60 - 0.61$ and $Acc_{test} = 0.46 - 0.51$ (Fig.7a, ~~Supplementary Material~~ Fig.S6) with associated $A_{UC} = 0.86 - 0.88$ (Fig.7b). This suggests that these 21 static landscape attributes can predict membership

to individual HRTs to the same degree as the vegetation and landcover attributes (3-VLC) and only slightly less accurately than when using climate attributes (1-CL, 2-C). At the level of individual HRTs, it was found that soil, geology and topography attributes are the strongest predictors on HRT 1 ($Acc_{test} = 0.62$) and HRT 3 (wet, moderate, summer regime) ($Acc_{test} = 0.58$), exceeding climate as well as vegetation and landcover as dominant controls (Fig.8e). In addition, HRT 7 (wet, strong, spring regime) is predicted with $Acc_{test} = 0.88$ in 5-SGT, which is effectively equivalent to the first four experiments. Further, the overall ranking of feature importances is broadly consistent with 1-CL. The top-ranking soil, geology and topography attributes of the 1-CL experiment, i.e. fraction of carbonate rock $G_{\mathrm{F(C)}}$, fraction of very low permeability rock $G_{\mathrm{F(Perm_{VL})}}$ and bedrock depth G_{BD}, also appear among the stronger predictors here in 5-SGT (Fig.10e, ~~Supplementary Material~~ Fig.S12d). Notwithstanding this general level of stability in the hierarchy of controls, the by far most dominant soil, geology and topography attribute emerging as control from 5-SGT is the catchment elevation T_{Elev}. Comparable to the role of dynamic vegetation and landcover attributes in the 3-VLC experiment, T_{Elev} here reflects climatic gradients and thus correspondingly acts as proxy for climate attributes such as for the fraction of $C_{\mathrm{F(Snow)}}$.

In the final experiment 6-A, using only anthropogenic attributes as predictors of membership to HRTs, the lowest model performances among all six experiments were found with $Acc_{test} = 0.27$–0.34 and $A_{UC} = 0.76 - 0.78$ (Fig.7). Similarly, the membership to the individual HRTs was also much less well predicted (Fig.8e). The most dominant control on HRT membership from this group of attributes was the fraction of agricultural areas $A_{\mathrm{F(Agri)}}$ (Fig.10e), which consistently also appeared as the highest ranking from this group of attributes in the 1-CL experiment (9^{th} rank) (Fig.10a).

Viewing these results in a broader context, we can spatially map the HRTs according to the attribute group that serves as their strongest predictor. In total, 4 out of the 10 HRTs were predicted equally well or better by vegetation and landscape attributes than by climate attributes (Fig.11). Together, this amounts to 3226 catchments, which is 45% of the total dataset. Soil, geology and topography attributes emerge as the dominant drivers of hydrological response in low to middle-elevation regions in central-eastern Europe, corresponding to HRTs 1 ~~and 3.~~ (stable, weak, winter regime) and 3 (wet, moderate, summer regime). Other catchments in the same geographic region, but belonging to HRT 6 ~~,~~ (transitional, moderate, spring regime), are more strongly influenced by vegetation attributes. Vegetation also emerges as the strongest driver for catchments in HRT 7 ~~,~~ (wet, strong, spring regime), in the higher elevation region of the Alps and at high-latitude regions.

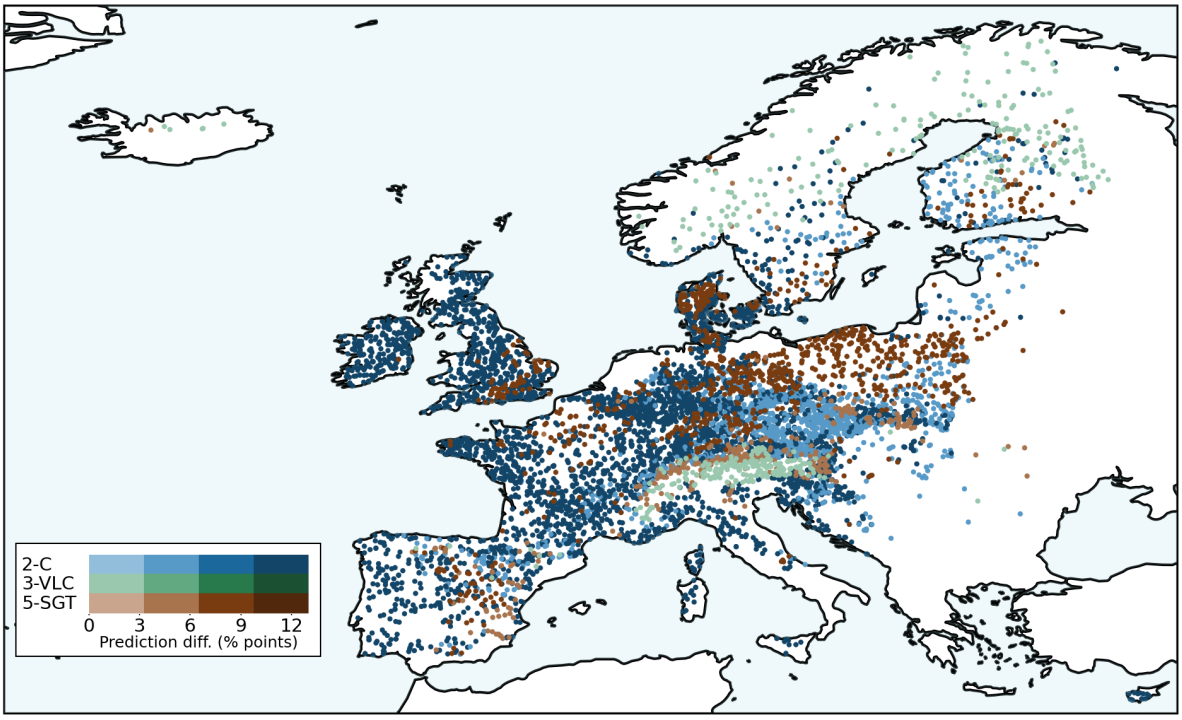


Figure 11. The map shows, for each HRT, which experiment achieved the highest proportion of correct predictions (indicated by colour). The colour intensity represents the magnitude of the difference between the highest and second-highest prediction accuracy: lighter colours indicate a small difference, whereas saturated colours indicate a larger difference. For example, the lighter blue catchments in central and eastern Europe and southern Finland indicate that the difference between the two best-performing experiments is small, whereas the more saturated colours in the UK indicate a larger performance gap. The colour bar units represent the percentage-point difference between the best and second-best prediction accuracies.

5 Discussion

5.1 Streamflow behaviour in Europe

With the aim to distinguish the rich, multi-dimensional intricacies of the hydrological response and to identify coherent groups of hydrological similarity across 7175 river catchments in Europe at a much higher level of detail than what is currently known (Kuentz et al., 2017), individual HRTs were defined by 40 signatures of averages and inter-annual variability of ~~stream flow~~ streamflow magnitudes, timing and dynamics. In contrast to the initial expectations and somewhat surprisingly given the amount of data underlying the analysis, the k-means clustering method used here did not provide sufficient evidence to distinguish more than 10 individual classes of HRTs. These individual HRTs effectively only capture rather broad systematic differences between different response types at the continental scale. As the study domain is limited to Europe, these patterns

reflect the variability of climatic and landscape conditions present in the region and may differ in areas with other environmental gradients. Within-HRT variability of the individual signatures remains pronounced (Fig.4, Supplementary Material Fig.S2) and comparable to previous studies (e.g., Berghuijs et al., 2014). Each HRT shows low variability for some signatures and high variability for others, indicating that the HRTs represent certain aspects of hydrological behaviour well, while remaining broad or diffuse in others.

For example, HRT 7 (wet, strong, spring regime) is one of the most distinct classes, yet it exhibits large variability in the 30-day autocorrelation ($H_{\mathrm{AC}_{30}}$), rising limb density (H_{RLD}), and days between highest and lowest flow ($H_{\mathrm{N(Q_H\text{-}Q_L)}}$) (Fig. 4g, h, j), while showing very little variability in the slope of the flow duration curve (H_{FDC}) (Fig. 4b). HRT 10 ~~,~~ (dry, intermittent, winter regime), another well-defined HRT, has, on the other hand, one of the largest variability in the slope of the flow duration curve (H_{FDC}), but less variability in timing of low flow ($H_{\mathrm{t(Q_L)}}$) and 30-day autocorrelation ($H_{\mathrm{AC}_{30}}$). Similar variation in the hydrological signatures is also seen in the less distinct classes such as HRT 2 ~~,~~ (transitional, low, multi-seasonal regime), which exhibit less variability in rising limb density (H_{RLD}) and high variability for the baseflow index (H_{BFI}) (Fig.4e, h). These heterogeneous patterns of within-class variability illustrate that general descriptions of the hydrological response characteristics that go beyond the major contrasts between the 10 HRTs, remains challenging. The key hydrological signatures with the strongest contrasts and thus the highest discriminatory power between HRTs included the Gini coefficient H_{GC}, the baseflow index H_{BFI}, as well as H_{Q} and $H_{\mathrm{Q_{95}}}$ but also the half-flow day $H_{\mathrm{t(HFD)}}$ as measure of timing and the variances of the Gini coefficient $H_{\mathrm{V,GC}}$ and flow $H_{\mathrm{V,Q}}$ as measures of inter-annual variability (Figs.4, 3d, Supplementary ~~Figure~~ Material Fig.S2). The combined importance of metrics of intra-annual timing, dynamics, magnitude and inter-annual variability for the identification of HRTs underlines the multi-faceted nature of streamflow, which requires all these different aspects for a meaningful characterisation of (dis)similarity of the hydrological response. In line with these findings, it was found that several HRTs were well-defined and clearly distinct from others. This is in particular the case for HRTs 1 (stable, weak, winter regime) and 10, which represent the respective extremes of many of the hydrological signatures. While HRT 1 is characterised by a stable, groundwater-driven system with little seasonality and generally low magnitudes of flow, HRT 10 is a highly dynamic, intermittent regime with strong seasonal variations in flow. HRTs 6-9 are similarly well defined by contrasts in multiple signatures. Other HRTs are less clearly defined. For example, HRT 5 (transitional, moderate, winter regime) closely resembled the overall sample average across most signatures, making it challenging to interpret, a challenge also noted by Jehn et al. (2020). Attempting to solve this issue, we analysed the effect of reducing the number of HRTs from 10 to 6-9. However, this did not eliminate the less distinct groups but instead increased the within-cluster variability of the remaining HRTs (not shown). Illustrating the complexity of classification, this suggests that certain HRTs may represent transitional or mixed hydrological behaviours that cannot be systematically and clearly enough distinguished with the available data, as it is in these cases difficult to attribute cluster boundaries to a single or even a few dominant signatures. The challenges in classifying HRTs are similar to those reported by Knoben et al. (2018), who found that some clusters are difficult to discern because catchments share climatic conditions and lack between-cluster diversity, whereas in other cases, differences in climate do not translate into statistically distinct flow regimes. The overall spatial pattern of the 10 HRTs identified in our study are broadly consistent with Kuentz et al. (2017), who identified 11 clusters of hydrological similarity across Europe. Some of the

most distinct HRTs are similar, such as the high average flow, dynamic (i.e. low H_{BFI}) spring flow regime clustering on the west coast of the UK and Ireland (HRT 9 (wet, strong, early-winter regime)), the summer high flow regime of the Alps and wide parts of Scandinavia (HRT 7) or the drier, more extreme regime with frequently intermittent flow in the Mediterranean regions (HRT 10). However, differences in catchment sample size and signature selection lead to differences in other HRTs and limit further direct comparison.

A key feature of our results is that while a few HRTs are characterised by rather sharply defined geographic associations, such as the Alps and Scandinavia (HRT 7 (wet, strong, spring regime)) or the British Isles (HRT 9 (wet, strong, early-winter regime)), others extend over wide parts of Europe, some even spanning from southern Spain to northern Finland (e.g. HRT 6 (transitional, moderate, spring regime)). As a major implication, this demonstrates that spatial proximity, even across larger regional scales, is a poor predictor of streamflow behaviour. It also contrasts with earlier studies (Sawicz et al., 2011; Knoben et al., 2018; Brunner et al., 2020), where spatial proximity was reported to play a stronger role, and clusters tended to correspond to regions with similar climatic conditions or physical likeness. One potential reason is that those studies included climatic descriptors for clustering hydrological response types, whereas our classification approach was exclusively based on hydrological signatures.

5.2 Drivers of streamflow behaviour

In six individual experiments, membership of individual catchments to one of the 10 HRTs was predicted based on climate- and/or landscape attributes with the aim to identify the dominant drivers of streamflow behaviour. While this allowed to capture the major continental pattern, the overall prediction performances were rather modest, with accuracies not exceeding 0.69 for training and 0.59 for testing, respectively (Fig.7). These levels of performance are comparable to previous studies. For example, Kuentz et al. (2017) reported that 60% of catchments in Europe were correctly classified in their approach. In contrast, McManamay et al. (2014) and Brunner et al. (2020) reported lower errors ($\sim 10\%$) for predictions of correct regime class, although based on much smaller samples of catchments (< 850) in the United States.

At the level of the individual HRTs, the prediction accuracies varied across all experiments. HRT 7 was throughout the best predicted with $Acc_{test} = 0.81$–0.90, due to its distinct snowmelt-driven regime. HRTs 1 and 8-10 were also consistently well predicted with climate- and landscape signatures, mirroring the same HRTs that were most clearly defined by hydrological signatures. In contrast, HRT 5 is the worst-predicted cluster in all the experiments and at the same time the least well-defined cluster in terms of contrasts in hydrological signatures. Similar difficulties were seen for HRTs 2-4 and 6, which appear to be transitional classes, with mixing features of other HRTs. Confusion matrices for the experiments further revealed systematic misclassification among these less distinct HRTs (Fig.8). This is consistent with earlier work suggesting that weak between-HRT diversity complicates the predictions (Knoben et al., 2018; Jehn et al., 2020), making it challenging to capture the essence of what is driving those clusters, at least at a continental scale. On the other hand, these HRTs have catchments distributed across multiple climate regions (from southern Spain to northern Nordics), indicating that similar hydrological response patterns can emerge under differing climatic conditions. This does not necessarily imply that landscape characteristics are more important than climate but rather suggests that similar hydrological behaviour can arise from different combinations of climatic and landscape controls. This is consistent with the concept of equifinality and uniqueness of place (Beven, 2000), where different

processes or conditions can lead to comparable hydrological responses. Our Random Forest results support this interpretation, as both climate and landscape variables contribute to predictive skill across experiments. It could be reasoned that reducing the number of total HRTs would also limit the choice of the model, resulting in these "bland" catchments being lumped together into a more distinguishable and thus better predictable group. However, this was not the case, as in all experiments run with a reduced number of 6-9 HRTs, the less distinct clusters were preserved and the prediction accuracies were not substantially increased (Supplementary Material Fig.S4), highlighting the inherent ambiguity in these HRTs.

The overall prediction accuracy $Acc_{test} = 0.55 - 0.61$ (Fig.7a) of experiment 2-C based on climate attributes underlines the important role of climate as driver of streamflow behaviour. However, the experiments based on landscape attributes resulted in only slightly lower accuracies with $Acc_{test} < 0.52$. In addition, it was also found that landscape attributes can predict membership to HRTs at least as good or even better than climate attributes for 4 out of 10 HRTs (Fig.11), with a total of 3226 catchments (45%). In low- to mid-elevation regions of central-east Europe, soil, geology and topography related attributes are the most dominant controls of the hydrological response (HRTs 1, 3). Similarly, vegetation attributes emerge as strong controls in these regions (HRT 6), but notably also in the higher elevation region of the Alps and at higher latitudes (HRT 7). The importance of landscape as driver of the hydrological response is further exacerbated by the ranked importances of the attributes in experiment 1-CL (Fig.10a). While climate attributes rank the highest, around half of the top 20 predictors are landscape attributes. In contrast to several studies that conclude that climate controls the hydrological response to a large extent with only limited landscape influence (Kuentz et al., 2017; Beck et al., 2015; Addor et al., 2018), our results provide broad empirical evidence supporting other previous studies that argue that landscape exerts a stronger influence than often assumed, in particular in specific regions ~~(Knoben et al., 2018; Dal Molin et al., 2020; Almagro et al., 2024)~~(Knoben et al., 2018; Dal Molin et al., 2020; Almagr
.

In light of our results, it can be argued that the frequently adopted climate vs. landscape perspective may lead to an overly simplistic perception of the system. This is for example illustrated by the dominant attributes that emerge in the landscape-only experiments 3-VLC (Fig.10c) and 5-SGT (Fig.10e). In 3-VLC, metrics of both inter-annual variability ($V_{\mathrm{V,NDVI}}$) and seasonality ($V_{\mathrm{R(NDVI_{Parde})}}$) were in the top-ranking attributes. Both are direct imprints of climatic conditions on the landscape over rather short time scales. Similarly, the fraction of tree cover ($V_{\mathrm{F(TC)}}$) as well as the elevation (T_{Elev}) as further key attributes in 3-VLC and 5-SGT, respectively, reflect aggregate effects of different temperature regimes across elevational gradients, although correlation with individual climate attributes such as the fraction of snow $C_{\mathrm{F(Snow)}}$ is very limited ($R^2 = 0.09$). These examples illustrate that landscape attributes are to a certain extent shaped by climate, thus also containing information on climatic conditions. It is therefore argued that the climate vs. landscape dichotomy is false and should be replaced by a climate and landscape perspective that explicitly accounts for the interactions in and co-evolution of the overall system (e.g., Savenije, 2010; Troch et al., 2013, 2015).

5.2.1 Spatial autocorrelation

When using a large dataset spanning Europe, neighbouring catchments are more likely to exhibit similar behaviour in terms of hydrology, but also in climate and landscape. This, in turn, could have an influence on the hydrological clustering as well as

potentially yielding more optimistic Random Forest classification results. Moran's I is a measure for spatial autocorrelation, and was calculated for each Random Forest experiment as a first step in quantifying the spatial autocorrelation. Values of $I \sim 0.3$ were found, indicating a degree of spatial correlation consistent with previous studies (e.g., Addor et al., 2018). The spatial autocorrelation is difficult to meaningfully deal with because it is inherent in any type of hydrological clustering, as most of the signatures are affected by climate and landscape, which have very local and regional similarities. To further disentangle the effect of spatial autocorrelation, a 10-fold spatial block cross-validation was run in the Random Forest analysis. In this approach, catchments were partitioned into contiguous spatial blocks (e.g. training set: catchments in the south-west, testing: catchments in the north-east of the domain) rather than randomly sampled. While training accuracies remained largely the same, testing accuracies decreased as compared to the original analysis (Fig.S13 in the Supplementary Material). However, the hierarchy of drivers remained largely the same, as for example HRT 1 performs best in the soil, geology and topography experiment (5-SGT), while HRT 7 is still largely driven by vegetation and landcover attributes (Fig.S14). These results suggest that spatial autocorrelation contributes to inflated performance under random cross-validation, and that spatial proximity has a considerable effect on model predictive skill. This, in turn, indicates that even the wide spectrum of climate and landscape attributes considered in the analysis carries less information than what is suggested by a random analysis, further suggesting that location itself is at least as informative to predict membership to HRTs and thus hydrological response patterns as the climate and landscape attributes. Overall, this highlights the challenge of disentangling hydrological drivers across regions with strong spatial structure, and underlines that currently available data are largely insufficient to resolve the "uniqueness of place" (Beven, 2000), in a meaningful way.

5.3 Limitations and future work

The spatial distribution and quality of the underlying data represent a source of uncertainty in this study. Streamflow observations are unevenly distributed across Europe, with a high density of gauged catchments in central Europe and lower coverage in the eastern and southern regions (do Nascimento et al., 2024). This imbalance may lead to an overrepresentation of hydrological characteristics of well-monitored regions, potentially limiting the representation of more data-scarce areas. In addition, the meteorological forcing is derived from the E-OBS dataset, whose accuracy similarly depends on station density and varies spatially. In regions with lower station density, particularly in parts of southern and eastern Europe, precipitation may be underestimated (Clerc-Schwarzenbach and do Nascimento, 2026). This may influence the derived climatic attributes and, consequently, their relationship with hydrological response. Despite the limitation in station density and uncertainty in the meteorological data, the dataset is considered sufficiently robust for identifying large-scale patterns across Europe (Clerc-Schwarzenbach and do Nascimen

.

It should be noted that there exists potential multicollinearity among the hydrological signatures and climate and landscape attributes. As the aim of the study was to retain a broad set of signatures and attributes, no extensive variable selection was applied to minimise redundancy. Instead, a Pearson correlation-based pre-filtering step was used to remove highly correlated variables ($|r| > 0.75$). This resulted in 50% of the hydrological signatures and climate and landscape attributes exhibiting $r < 0.12$ (equivalent to explaining $\sim 1\%$ of the variance) and only 2.5% of the attributes being correlated with $r > 0.6$

(explaining 36% of the variance; see Fig.S1 in the Supplementary Material). Some correlation remains, and to explicitly test the robustness of the model interpretation, the consistency of the feature importance ranking of the climate and landscape attributes across experiments was evaluated (Fig.S12). The most influential variables show broadly consistent rankings across model configurations, suggesting that the main conclusions are not sensitive to multicollinearity effects. Nevertheless, correlated predictors may still affect the attribution of importance and comparisons between experiments with different predictor sets.

The presence of nested catchments may introduce dependencies that influence the cross-validation in the Random Forest experiments. To assess the impact of nested catchments on the results, the analysis was repeated with a subset of non-nested catchments (n = 4137). The results show that both the clustering into hydrological response types and the Random Forest classification performance remain largely unchanged (Supplementary Material Figs S15, S16 and S17), with only a small decrease in testing accuracy (1–5% across experiments, Fig.S17). These results indicate that, while nesting introduces some dependence, its influence on the overall clustering and classification results is sufficiently limited for the purpose of this analysis.

There are also methodological limitations related to the use of k-means clustering to group catchments into hydrological response type groups (HRTs) based on 40 hydrological signatures. k-means relies on rigid cluster boundaries and assigns equal weight to all signatures, which may not reflect the actual complexity of hydrological systems. This limitation is reflected in the weak elbow (Fig.2) and consistently low Silhouette Scores across the range of cluster numbers, indicating limited separability in the data. The PCA results also show that a relatively large number of principal components are required to explain most of the variance, suggesting that the signatures do not form clearly distinct groups in feature space. This may partly reflect the system complexity, where hydrological behaviour varies continuously and overlaps in multidimensional space, rather than forming discrete groups (McManamay et al., 2014; Knoben et al., 2018).

Several directions for future research can be derived from the above reflections. First, further work should aim to disentangle the relative roles of landscape by analysing subsets of catchments within more homogeneous climatic regions. Second, the present analysis is based on long-term aggregated statistics and does not explicitly account for temporal dynamics beyond inter-annual variances. Future studies should investigate changes in hydrological response over time, for example, by analysing different climatic periods or assessing the impacts of climate change. Third, alternative clustering approaches that allow for more flexible or overlapping group structures (e.g. fuzzy or hierarchical clustering) could be explored to better capture the continuous nature of hydrological variability. Finally, continued development of spatially explicit modelling and validation approaches may help to better account for spatial autocorrelation in large-sample hydrological studies.

5.4 Wider implications

Overall, this study demonstrates that even the use of an unprecedented range of data does not allow to systematically characterise and robustly identify more than 10 broad types of hydrological response classes across Europe, and to narrow down the within-class variability. The question of what *constitutes* the hydrological response and (dis)similarity therein thus remains largely unresolved. Similarly, the modest accuracies to predict membership to the 10 HRTs based on a wide spectrum of climate- and landscape attributes illustrates that the question what *drives* the hydrological response remains equally unresolved. In

other words, a satisfactory and generalisable answer to the question "Why do we in a specific catchment observe a specific flow magnitude at a specific moment, and why is it different to the flow in other catchments at other moments?" is still lacking. Ultimately, the results of this study underline that available data are still insufficient to describe the complexity of the hydrological system, ~~suggesting~~ highlighting the need for considerably more data with higher spatial and temporal resolution and observational detail. This further suggests that Beven's "uniqueness of place" (Beven, 2000) is a larger obstacle to generalisation in hydrology than previously thought and that this postulate is more relevant than ever.

6 Conclusions

The aim of this study was to (1) realise a detailed classification of hydrological response types (HRTs) with minimal intra-class variance by using a large catchment sample and a wide range of hydrological signatures, including descriptors of inter-annual variability, and (2) to identify the dominant drivers of streamflow behaviour, and in particular whether climate, rather than landscape exert more discriminatory power to explain the dissimilarity between the identified HRT classes.

Our results showed that, classifying 7175 catchments across Europe, using 40 hydrological signatures describing averages, inter-annual variability of streamflow magnitudes, timing, duration, frequency, and seasonality, no more than 10 HRTs could be meaningfully distinguished. Even with this level of detail, the HRTs could only capture broad patterns of the hydrological response. The most distinct HRTs typically contained catchments with strong seasonal or extreme flow regimes. Others were more difficult to define, as they overlapped in terms of average characteristics with the other HRTs.

Climate consistently emerged as the dominant driver of streamflow behaviour when 84 climate- and landscape attributes were used to predict catchment membership to the 10 HRTs. However, landscape attributes represent half of the 20 most important drivers of the hydrological response. In addition, in 4 out of the 10 HRTs, landscape was found to be as strong or even stronger a control than climate. This highlights that the landscape characteristics, often reported as having little influence on the differences in hydrological behaviour at continental scales, plays a much more important role than frequently thought.

Overall, the results of this study demonstrate that the "uniqueness of place" concept (Beven, 2000) remains highly relevant, and provide further evidence that more consideration needs to be given to the perspective of co-evolution between climate and landscape in shaping hydrological behaviour at continental scales.

Code and data availability. EStreams data and code are available at: https://doi.org/10.5281/zenodo.14778580 (do Nascimento et al., 2025b) and at https://estreams.eawag.ch/ (last access: 18 December 2025) and described in detail by do Nascimento et al. (2024). The code used for the analysis and additional signatures and attributes calculations is available at the Zenodo repository: https://doi.org/10.5281/zenodo.~~17987885~~ 17987884 (Rudlang, 2025).

Author contribution. MH and FF had the original idea and conceptualised the study together with JR and RE. JR carried out the formal analysis with input from MH, RE, TH and FF. JR prepared the manuscript with contributions from all authors.

Competing interests. At least one of the (co-)authors is a member of the editorial board of *Hydrology and Earth System Sciences*.

Acknowledgements. The use of DelftBlue computing facility at the Delft High Performance Computing Centre (DHPC) is acknowledged.

Financial support. This project was funded by the Netherlands Organisation for Scientific Research (NWO), project no. OCENW.M.21.230.

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. J., 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.

Addor, N., Do, H. X., Alvarez-Garreton, C., Coxon, G., Fowler, K., and Mendoza, P. A.: Large-sample hydrology: recent progress, guidelines for new datasets and grand challenges, Hydrological Sciences Journal, 65, 712–725, https://doi.org/10.1080/02626667.2019.1683182, 2020.

Almagro, A., Meira Neto, A. A., Vergopolan, N., Roy, T., Troch, P. A., and Oliveira, P. T. S.: The Drivers of Hydrologic Behavior in Brazil: Insights From a Catchment Classification, Water Resources Research, 60, https://doi.org/10.1029/2024WR037212, 2024.

Araki, R., Holt, A., Hammond, J. C., Husic, A., Coxon, G., and Mcmillan, H. K.: Continental-scale prediction of hydrologic signatures and processes, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2025-6156, 2025.

Beck, H. E., de Roo, A., and van Dijk, A. I.: Global Maps of Streamflow Characteristics Based on Observations from Several Thousand Catchments*, Journal of Hydrometeorology, 16, 1478–1501, https://doi.org/10.1175/JHM-D-14-0155.1, 2015.

Beck, H. E., Pan, M., Lin, P., Seibert, J., van Dijk, A. I., and Wood, E. F.: Global Fully Distributed Parameter Regionalization Based on Observed Streamflow From 4,229 Headwater Catchments, Journal of Geophysical Research: Atmospheres, 125, e2019JD031 485, https://doi.org/10.1029/2019JD031485, 2020.

Berghuijs, W. R., Sivapalan, M., Woods, R. A., and Savenije, H. H.: 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.

Berghuijs, W. R., Hale, K., and Beria, H.: Technical note: Streamflow seasonality using directional statistics, Hydrology and Earth System Sciences, 29, 2851–2862, https://doi.org/10.5194/hess-29-2851-2025, 2025.

Beven, K. J.: Uniqueness of place and process representations in hydrological modelling, Hydrology and Earth System Sciences, 4, 203–213, https://doi.org/10.5194/hess-4-203-2000, 2000.

Blöschl, G.: Hydrologic synthesis: Across processes, places, and scales, Water Resources Research, 42, https://doi.org/10.1029/2005WR004319, 2006.

Blöschl, G., Ardoin-Bardin, S., Bonell, M., Dorninger, M., Goodrich, D., Gutknecht, D., Matamoros, D., Merz, B., Shand, P., and Szolgay, J.: At what scales do climate variability and land cover change impact on flooding and low flows?, Hydrological Processes, 21, 1241–1247, https://doi.org/10.1002/HYP.6669, 2007.

Blöschl, G., Hall, J., Parajka, J., Perdigão, R. A. P., Merz, B., Arheimer, B., Aronica, G. T., Bilibashi, A., Bonacci, O., Borga, M., Čanjevac, I., Castellarin, A., Chirico, G. B., Claps, P., Fiala, K., Frolova, N., Gorbachova, L., Gül, A., Hannaford, J., Harrigan, S., Kireeva, M., Kiss, A., Kjeldsen, T. R., Kohnová, S., Koskela, J. J., Ledvinka, O., Macdonald, N., Mavrova-Guirguinova, M., Mediero, L., Merz, R., Molnar, P., Montanari, A., Murphy, C., Osuch, M., Ovcharuk, V., Radevski, I., Rogger, M., Salinas, J. L., Sauquet, E., Šraj, M., Szolgay, J., Viglione, A., Volpi, E., Wilson, D., Zaimi, K., and Živković, N.: Changing climate shifts timing of European floods, Science, 357, 588–590, https://doi.org/10.1126/science.aan2506, 2017.

Blöschl, G., Bierkens, M. F., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., Kirchner, J. W., McDonnell, J. J., Savenije, H. H., Sivapalan, M., Stumpp, C., Toth, E., Volpi, E., Carr, G., Lupton, C., Salinas, J., Széles, B., Viglione, A., Aksoy, H., Allen, S. T., Amin, A.,

Andréassian, V., Arheimer, B., Aryal, S. K., Baker, V., Bardsley, E., Barendrecht, M. H., Bartosova, A., Batelaan, O., Berghuijs, W. R., Beven, K., Blume, T., Bogaard, T., Borges de Amorim, P., Böttcher, M. E., Boulet, G., Breinl, K., Brilly, M., Brocca, L., Buytaert, W., Castellarin, A., Castelletti, A., Chen, X., Chen, Y., Chen, Y., Chifflard, P., Claps, P., Clark, M. P., Collins, A. L., Croke, B., Dathe, A., David, P. C., de Barros, F. P., de Rooij, G., Di Baldassarre, G., Driscoll, J. M., Duethmann, D., Dwivedi, R., Eris, E., Farmer, W. H., Feiccabrino, J., Ferguson, G., Ferrari, E., Ferraris, S., Fersch, B., Finger, D., Foglia, L., Fowler, K., Gartsman, B., Gascoin, S., Gaume, E., Gelfan, A., Geris, J., Gharari, S., Gleeson, T., Glendell, M., Gonzalez Bevacqua, A., González-Dugo, M. P., Grimaldi, S., Gupta, A. B., Guse, B., Han, D., Hannah, D., Harpold, A., Haun, S., Heal, K., Helfricht, K., Herrnegger, M., Hipsey, M., Hlaváčiková, H., Hohmann, C., Holko, L., Hopkinson, C., Hrachowitz, M., Illangasekare, T. H., Inam, A., Innocente, C., Istanbulluoglu, E., Jarihani, B., Kalantari, Z., Kalvans, A., Khanal, S., Khatami, S., Kiesel, J., Kirkby, M., Knoben, W., Kochanek, K., Kohnová, S., Kolechkina, A., Krause, S., Kreamer, D., Kreibich, H., Kunstmann, H., Lange, H., Liberato, M. L., Lindquist, E., Link, T., Liu, J., Loucks, D. P., Luce, C., Mahé, G., Makarieva, O., Malard, J., Mashtayeva, S., Maskey, S., Mas-Pla, J., Mavrova-Guirguinova, M., Mazzoleni, M., Mernild, S., Misstear, B. D., Montanari, A., Müller-Thomy, H., Nabizadeh, A., Nardi, F., Neale, C., Nesterova, N., Nurtaev, B., Odongo, V. O., Panda, S., Pande, S., Pang, Z., Papacharalampous, G., Perrin, C., Pfister, L., Pimentel, R., Polo, M. J., Post, D., Prieto Sierra, C., Ramos, M. H., Renner, M., Reynolds, J. E., Ridolfi, E., Rigon, R., Riva, M., Robertson, D. E., Rosso, R., Roy, T., Sá, J. H., Salvadori, G., Sandells, M., Schaefli, B., Schumann, A., Scolobig, A., Seibert, J., Servat, E., Shafiei, M., Sharma, A., Sidibe, M., Sidle, R. C., Skaugen, T., Smith, H., Spiessl, S. M., Stein, L., Steinsland, I., Strasser, U., Su, B., Szolgay, J., Tarboton, D., Tauro, F., Thirel, G., Tian, F., Tong, R., Tussupova, K., Tyralis, H., Uijlenhoet, R., van Beek, R., van der Ent, R. J., van der Ploeg, M., Van Loon, A. F., van Meerveld, I., van Nooijen, R., van Oel, P. R., Vidal, J. P., von Freyberg, J., Vorogushyn, S., Wachniew, P., Wade, A. J., Ward, P., Westerberg, I. K., White, C., Wood, E. F., Woods, R., Xu, Z., Yilmaz, K. K., and Zhang, Y.: Twenty-three unsolved problems in hydrology (UPH)–a community perspective, Hydrological Sciences Journal, 64, 1141–1158, https://doi.org/10.1080/02626667.2019.1620507, 2019.

Brêda, J. P. L., Melsen, L. A., Athanasiadis, I., Van Dijk, A., Siqueira, V. A., Verhoef, A., Zeng, Y., and van der Ploeg, M.: Predictor Importance for Hydrological Fluxes of Global Hydrological and Land Surface Models, Water Resources Research, 60, e2023WR036 418, https://doi.org/10.1029/2023WR036418, 2024.

Breiman, L.: Random forests, Machine Learning, 45, 5–32, https://doi.org/10.1023/A:1010933404324/METRICS, 2001.

Brunner, M. I., Melsen, L. A., Newman, A. J., Wood, A. W., and Clark, M. P.: Future streamflow regime changes in the United States: assessment using functional classification, Hydrology and Earth System Sciences, 24, 3951–3966, https://doi.org/10.5194/hess-24-3951-2020, 2020.

Carrillo, G., Troch, P. A., Sivapalan, M., Wagener, T., Harman, C., and Sawicz, K.: Catchment classification: hydrological analysis of catchment behavior through process-based modeling along a climate gradient, Hydrology and Earth System Sciences, 15, 3411–3430, https://doi.org/10.5194/hess-15-3411-2011, 2011.

Clerc-Schwarzenbach, F. and do Nascimento, T. V. M.: Evaluating E-OBS forcing data for large-sample hydrology using model performance diagnostics, Hydrology and Earth System Sciences, 30, 119–140, https://doi.org/10.5194/hess-30-119-2026, 2026.

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.

Dal Molin, M., Schirmer, M., Zappa, M., and Fenicia, F.: Understanding dominant controls on streamflow spatial variability to set up a semi-distributed hydrological model: The case study of the Thur catchment, Hydrology and Earth System Sciences, 24, 1319–1345, https://doi.org/10.5194/hess-24-1319-2020, 2020.

do Nascimento, T. V. M., Rudlang, J., Höge, M., van der Ent, R., Chappon, M., Seibert, J., Hrachowitz, M., and Fenicia, F.: EStreams: An integrated dataset and catalogue of streamflow, hydro-climatic and landscape variables for Europe, Scientific Data, 11, 879, https://doi.org/10.1038/s41597-024-03706-1, 2024.

do Nascimento, T. V. M., Rudlang, J., Gnann, S., Seibert, J., Hrachowitz, M., and Fenicia, F.: How do geological map details influence the identification of geology-streamflow relationships in large-sample hydrology studies?, Hydrology and Earth System Sciences, 29, 7173–7200, https://doi.org/10.5194/hess-29-7173-2025, 2025a.

do Nascimento, T. V. M., Rudlang, J., Höge, M., van der Ent, R., Chappon, M., Seibert, J., and Hrachowitz, M.: EStreams: An Integrated Dataset and Catalogue of Streamflow, Hydro-Climatic Variables and Landscape Descriptors for Europe (1.2), https://doi.org/10.5281/zenodo.14778580, 2025b.

Dooge, J. C.: Looking for hydrologic laws, Water Resources Research, 22, 46S–58S, https://doi.org/10.1029/WR022I09SP0046S, 1986.

Fenicia, F. and McDonnell, J. J.: Modeling streamflow variability at the regional scale: (1) perceptual model development through signature analysis, Journal of Hydrology, 605, 127 287, https://doi.org/10.1016/j.jhydrol.2021.127287, 2022.

Fenicia, F., Kavetski, D., Savenije, H. H., Clark, M. P., Schoups, G., Pfister, L., and Freer, J.: Catchment properties, function, and conceptual model representation: Is there a correspondence?, Hydrological Processes, 28, 2451–2467, https://doi.org/10.1002/HYP.9726, 2014.

Fowler, K. J., Acharya, S. C., Addor, N., Chou, C., and Peel, M. C.: CAMELS-AUS: Hydrometeorological time series and landscape attributes for 222 catchments in Australia, Earth System Science Data, 13, 3847–3867, https://doi.org/10.5194/ESSD-13-3847-2021,, 2021.

Gudmundsson, L., Do, H. X., Leonard, M., and Westra, S.: The Global Streamflow Indices and Metadata Archive (GSIM) – Part 2: Quality control, time-series indices and homogeneity assessment, Earth System Science Data, 10, 787–804, https://doi.org/10.5194/essd-10-787-2018, 2018.

Gupta, H. V., Perrin, C., Blöschl, G., Montanari, A., Kumar, R., Clark, M., and Andréassian, V.: Large-sample hydrology: A need to balance depth with breadth, Hydrology and Earth System Sciences, 18, 463–477, https://doi.org/10.5194/HESS-18-463-2014,, 2014.

Hargreaves, G. H. and Samani, Z. A.: Estimating Potential Evapotranspiration, Journal of the Irrigation and Drainage Division, 108, 225–230, https://doi.org/10.1061/JRCEA4.0001390, 1982.

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, https://doi.org/10.1016/j.jhydrol.2024.131638, 2024.

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.

Jones Jr., C. E., Leibowitz, S. G., Sawicz, K. A., Comeleo, R. L., Stratton, L. E., Morefield, P. E., and Weaver, C. P.: Using hydrologic landscape classification and climatic time series to assess hydrologic vulnerability of the western U.S. to climate, Hydrology and Earth System Sciences, 25, 3179–3206, https://doi.org/10.5194/hess-25-3179-2021, 2021.

Kemter, M., Marwan, N., Villarini, G., and Merz, B.: Controls on Flood Trends Across the United States, Water Resources Research, 59, https://doi.org/10.1029/2021WR031673, 2023.

Kerins, D., Knapp, A. S., Liu, F. S., Smykalov, V. D., Berzonsky, M. P., Vierbicher, A., Sadayappan, K., Stewart, B., Andrews, E. M., Sullivan, P. L., Barnard, H. R., Seibert, J., McPhillips, L. E., Singha, K., and Li, L.: Controls From Above and Below: Snow, Soil, and Steepness Drive Diverging Trends of Subsurface Water and Streamflow Dynamics, Hydrological Processes, 39, https://doi.org/10.1002/HYP.70120, 2025.

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.

Kratzert, F., Nearing, G., Addor, N., Erickson, T., Gauch, M., Gilon, O., Gudmundsson, L., Hassidim, A., Klotz, D., Nevo, S., Shalev, G., and Matias, Y.: Caravan - A global community dataset for large-sample hydrology, Scientific Data 2023 10:1, 10, 1–11, https://doi.org/10.1038/s41597-023-01975-w, 2023.

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.

Laaha, G. and Blöschl, G.: A comparison of low flow regionalisation methods—catchment grouping, Journal of Hydrology, 323, 193–214, https://doi.org/10.1016/J.JHYDROL.2005.09.001, 2006a.

Laaha, G. and Blöschl, G.: Seasonality indices for regionalizing low flows, Hydrological Processes, 20, 3851–3878, https://doi.org/10.1002/HYP.6161, 2006b.

Ladson, A. R., Brown, R., Neal, B., and Nathan, R.: A Standard Approach to Baseflow Separation Using The Lyne and Hollick Filter, Australasian Journal of Water Resources, 17, 25–34, https://doi.org/10.7158/13241583.2013.11465417, 2013.

McDonnell, J. J. and Woods, R.: On the need for catchment classification, Journal of Hydrology, 299, 2–3, https://doi.org/10.1016/j.jhydrol.2004.09.003, 2004.

McDonnell, J. J., Sivapalan, M., Vaché, K., Dunn, S., Grant, G., Haggerty, R., Hinz, C., Hooper, R., Kirchner, J., Roderick, M. L., Selker, J., and Weiler, M.: Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology, Water Resources Research, 43, https://doi.org/10.1029/2006WR005467, 2007.

McManamay, R. A., Bevelhimer, M. S., and Kao, S. C.: Updating the US hydrologic classification: an approach to clustering and stratifying ecohydrologic data, Ecohydrology, 7, 903–926, https://doi.org/doi.org/10.1002/eco.1410, 2014.

McMillan, H.: Linking hydrologic signatures to hydrologic processes: A review, Hydrological Processes, 34, 1393–1409, https://doi.org/10.1002/HYP.13632, 2020.

McMillan, H.: A review of hydrologic signatures and their applications, WIREs Water, 8, e1499, https://doi.org/10.1002/wat2.1499, 2021.

McMillan, H., Westerberg, I., and Branger, F.: Five guidelines for selecting hydrological signatures, Hydrological Processes, 31, 4757–4761, https://doi.org/https://doi.org/10.1002/hyp.11300, 2017.

McMillan, H., Gnann, S. J., and Araki, R.: Large Scale Evaluation of Relationships Between Hydrologic Signatures and Processes, Water Resources Research, 58, e2021WR031 751, https://doi.org/https://doi.org/10.1029/2021WR031751, 2022.

McMillan, H., Araki, R., Gnann, S., Woods, R., and Wagener, T.: How do hydrologists perceive watersheds? A survey and analysis of perceptual model figures for experimental watersheds, Hydrological Processes, 37, https://doi.org/10.1002/HYP.14845, 2023a.

McMillan, H., Coxon, G., Araki, R., Salwey, S., Kelleher, C., Zheng, Y., Knoben, W., Gnann, S., Seibert, J., and Bolotin, L.: When good signatures go bad: Applying hydrologic signatures in large sample studies, Hydrological Processes, 37, https://doi.org/10.1002/HYP.14987, 2023b.

McMillan, H., Araki, R., Bolotin, L., Kim, D. H., Coxon, G., Clark, M., and Seibert, J.: Global patterns in observed hydrologic processes, Nature Water, 3, 497–506, https://doi.org/10.1038/s44221-025-00407-w, 2025.

Merz, R. and Blöschl, G.: A process typology of regional floods, Water Resources Research, 39, 1340, https://doi.org/10.1029/2002WR001952, 2003.

Merz, R. and Blöschl, G.: Regionalisation of catchment model parameters, Journal of Hydrology, 287, 95–123, https://doi.org/10.1016/J.JHYDROL.2003.09.028, 2004.

Merz, R. and Blöschl, G.: A regional analysis of event runoff coefficients with respect to climate and catchment characteristics in Austria, Water Resources Research, 45, https://doi.org/10.1029/2008WR007163, 2009.

Parajka, J., Merz, R., and Blöschl, G.: A comparison of regionalisation methods for catchment model parameters, Hydrology and Earth System Sciences, 9, 157–171, https://doi.org/10.5194/HESS-9-157-2005, 2005.

Parajka, J., Kohnová, S., Bálint, G., Barbuc, M., Borga, M., Claps, P., Cheval, S., Dumitrescu, A., Gaume, E., Hlavčová, K., Merz, R., Pfaundler, M., Stancalie, G., Szolgay, J., and Blöschl, G.: Seasonal characteristics of flood regimes across the Alpine–Carpathian range, Journal of Hydrology, 394, 78–89, https://doi.org/10.1016/J.JHYDROL.2010.05.015, 2010.

Rudlang, J. M.: Climate and landscape jointly control Europe's hydrology, https://doi.org/10.5281/zenodo.17987884, 2025.

Savenije, H. H.: HESS opinions "topography driven conceptual modelling (FLEX-Topo)", Hydrology and Earth System Sciences, 14, 2681–2692, https://doi.org/10.5194/HESS-14-2681-2010, 2010.

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, Hydrol. Earth Syst. Sci, 15, 2895–2911, https://doi.org/10.5194/hess-15-2895-2011, 2011.

Sivapalan, M.: Prediction in ungauged basins: a grand challenge for theoretical hydrology, Hydrological Processes, 17, 3163–3170, https://doi.org/10.1002/hyp.5155, 2003a.

Sivapalan, M.: Process complexity at hillslope scale, process simplicity at the watershed scale: is there a connection?, Hydrological Processes, 17, 1037–1041, https://doi.org/10.1002/HYP.5109, 2003b.

Slater, L., Coxon, G., Brunner, M., McMillan, H., Yu, L., Zheng, Y., Khouakhi, A., Moulds, S., and Berghuijs, W.: Spatial Sensitivity of River Flooding to Changes in Climate and Land Cover Through Explainable AI, Earth's Future, 12, https://doi.org/10.1029/2023EF004035, 2024.

Tarasova, L., Gnann, S., Yang, S., Hartmann, A., and Wagener, T.: Catchment characterization: Current descriptors, knowledge gaps and future opportunities, Earth-Science Reviews, 252, 104 739, https://doi.org/10.1016/j.earscirev.2024.104739, 2024.

Troch, P. A., Carrillo, G., Sivapalan, M., Wagener, T., and Sawicz, K.: Climate-vegetation-soil interactions and long-term hydrologic partitioning: Signatures of catchment co-evolution, Hydrology and Earth System Sciences, 17, 2209–2217, https://doi.org/10.5194/HESS-17-2209-2013, 2013.

Troch, P. A., Lahmers, T., Meira, A., Mukherjee, R., Pedersen, J. W., Roy, T., and Valdés-Pineda, R.: Catchment coevolution: A useful framework for improving predictions of hydrological change?, Water Resources Research, 51, 4903–4922, https://doi.org/10.1002/2015WR017032, 2015.

van Oorschot, F., Hrachowitz, M., Viering, T., Alessandri, A., and van der Ent, R. J.: Global patterns in vegetation accessible subsurface water storage emerge from spatially varying importance of individual drivers, Environmental Research Letters, 19, https://doi.org/10.1088/1748-9326/ad8805, 2024.

Vu, T., Kiesel, J., Guse, B., Domisch, S., and Fohrer, N.: Disentangling Spatio-Temporal Impacts of Multiple Environmental Factors on the Global Discharge Regime, Earth's Future, 12, https://doi.org/10.1029/2023EF004267, 2024.

Wagener, T., Sivapalan, M., Troch, P., and Woods, R.: Catchment Classification and Hydrologic Similarity, Geography Compass, 1, 901–931, https://doi.org/10.1111/j.1749-8198.2007.00039.x, 2007.

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.

Ye, S., Yaeger, M., Coopersmith, E., Cheng, L., and Sivapalan, M.: Exploring the physical controls of regional patterns of flow duration curves - Part 2: Role of seasonality, the regime curve, and associated process controls, Hydrology and Earth System Sciences, 16, 4447–4465, https://doi.org/10.5194/hess-16-4447-2012, 2012.