Title

A Quantitative Hydrological Climate Classification Evaluated with Independent Streamflow Data

Authors

[1]Knoben, Wouter J. M.
[1]Woods, Ross A.
[2]Freer, Jim E.

Affiliations

[1]Department of Civil Engineering, University of Bristol, UK
[2]School of Geographical Sciences, University of Bristol, UK

Key points

- Dimensionless numbers that describe a location's aridity, seasonality of aridity and snowfall can define the global hydroclimate
- Seasonal streamflow regimes and values of hydrologic statistics are similar in locations with similar values for the dimensionless numbers
- This approach to hydrologic climate classification is more informative than Köppen-Geiger classes, especially in snow-dominated areas

Abstract

Classification is essential in the study of natural systems, yet hydrology has no formal way to structure the climatic forcing that underlies hydrologic response. Various climate classification systems can be borrowed from other disciplines but these are based on different organizing principles than a hydrological classification might need. This work presents a hydrologically-informed way to quantify global climates, explicitly addressing the shortcomings in earlier climate classifications. In this work, causal factors (climate) and hydrologic response (streamflow) are separated, meaning that our classification scheme is based only on climatic information and can be evaluated with independent streamflow data. Using gridded global climate data, we calculate three dimensionless indices per grid cell, describing annual aridity, aridity seasonality and precipitation-as-snow. We use these indices to create several climate groups and define the membership degree of 1103 catchments to each of the climate groups, based on each catchment's climate. Streamflow patterns within each group tend to be similar, and tend to be different between groups. Visual comparison of flow regimes and Wilcoxon two-sample statistical tests on 16 streamflow signatures show that this index-based approach is more effective than the often-used Köppen-Geiger classification for grouping hydrologically similar catchments. Climate forcing exerts a strong control on typical hydrologic response and we show that at the global scale both change gradually in space. We argue that hydrologists should consider the hydroclimate as a continuous spectrum defined by the three climate indices, on which all catchments are positioned and show examples of this in a regionalization context.

1 Introduction

Classification is an essential step in understanding natural phenomena, as evidenced by globally agreed-upon classification schemes in many different disciplines and a strong expressed need for a catchment classification scheme in hydrology (e.g. McDonnell and Woods, 2004; Wagener *et al.*, 2007). Well-known classification examples are the periodic table that chemistry uses to group elements with similar properties (e.g. Scerri, 2007) and Linnaean taxonomy as used in biology to group organisms based on similarity of their characteristics (e.g. de Queiroz & Gauthier, 1992). Classifying phenomena into groups with similar characteristics allows transfer of knowledge from well-observed members of the group to members about which less is known. In hydrology, defining similarity between catchments plays a crucial role in enabling predictions in ungauged basins (Wagener *et al.*, 2007).

In complex systems as are common in earth sciences, classification is not straightforward (McDonnell & Woods, 2004). Many different classification schemes are available, each with a different focus or underlying principles, and the choice for one is often motivated by a study's particular needs. All classification schemes however aim to group those elements of a system that are similar, and separate them from groups of other elements that are in some significant way different from the others. For example, soils can be classified with an international system based on their diagnostic horizons, properties and materials (IUSS Working Group WRB, 2015), but various national systems are used as well (e.g. Baize & Girard, 2009; Hewitt, 1992; Soil Classification Working Group, 1991). Lakes can be classified on a variety of characteristics; e.g. thermal properties (Forel, 1880, cited Hutchinson & Löffler, 1956), mixing properties (Lewis Jr, 1983), trophic status (Canfield Jr *et al.*, 1983) or a combination of hydrological, chemical and biological properties (Johnes *et al.*, 1994). Similarly, different schemes are available to classify vegetation, e.g. by using plants' survival strategy (Grime, 1974), a hierarchical scheme based on leaf cover area akin to Linnaean taxonomy (Viereck *et al.*, 1992) or as a function of dominant prevailing climate known as life zones (Holdridge, 1967).

Catchments are a common object of study within hydrology. The need for a catchment classification scheme (e.g. McDonnell & Woods, 2004; Wagener *et al.*, 2007) is usually interpreted as defining catchment similarity based on hydrological response, presumed drivers of the streamflow response, or a combination of both. Wagener *et al.* (2007) lists possible options for classification based on hydro-climatic region, catchment structure or functional catchment response. Hydrologic similarity (i.e. grouping similar catchments) follows from mapping the relationship between these aspects. An early example of a global classification of river regimes (Haines *et al.*, 1988) defines 15 different typical annual streamflow patterns across the globe. Increases in data availability have allowed more detailed regional studies covering e.g. Australia (Kennard *et al.*, 2010) and the US (Archfield *et al.*, 2014). Looking just at causal factors underlying streamflow, examples of regional classifications exist based on soil characteristics (Lilly, 2010) and climate (Berghuijs *et al.*, 2014). Many studies combine both approaches, using causal factors, such as average aridity, average catchment slope and land use, together with streamflow characteristics, often in the form of streamflow signatures such as mean flow and slope of the flow duration curve, to group similar catchments (e.g. Coopersmith *et al.*, 2012; Kuentz *et al.*, 2016; Sawicz *et al.*, 2011, 2014; Yadav *et al.*, 2007). Whereas a wide variety of metrics and models are used to describe catchment structure and functional response, there seems to be at least some consensus on how hydro-climatic aspects can be conceptualised: available water (precipitation) and energy (temperature, evaporation) interact within the catchment to control the water balance. Understanding of this principle has led to the Budyko-curve on an annual scale (Budyko, 1974) and shown the importance of within-year variation of climate (e.g. Milly, 1994).

In catchment classification studies, climate is often considered in a basic form (e.g. annual average aridity) or in direct relation to streamflow response (e.g. runoff ratio, streamflow elasticity) but recent work shows that a more nuanced approach that describes the influence of climatic input on typical flow regimes might be appropriate (Addor et al., 2017, 2018; Berghuijs et al., 2014). With three dimensionless numbers that summarize the climate's aridity, precipitation timing and snowiness, typical flow regimes in the US can be classified into 10 distinct groups (Berghuijs *et al.*, 2014). Addor *et al.* (2017) present an extended set of US catchments, including information about each catchment's climate (using three very similar indices), topography, soils and vegetation. In later work (Addor *et al.*, 2018), they correlate this information with streamflow signatures for each catchment and find that climate, as expressed by the three indices, has the strongest correlation with streamflow signature values for this set of USA data. Information about climate, even expressed as three simple numbers, can thus be used to explain broad streamflow patterns.

Several global climate classifications exist, but these are mostly bio-climatic in origin, and thus do not explicitly include those aspects of climate regimes that are important influences on hydrology. The original Köppen scheme (work by Köppen in the late 19th and early 20th century) is for an important part based on observations of vegetation, which could be used as a proxy for prevailing climate in times when large-scale climate data was unavailable (Peel *et al.*, 2007). Köppen's classification inspired several other classification schemes that tried to improve the correspondence between climate zones and observed global vegetation patterns (Geiger, 1954; Thornthwaite, 1948; Trewartha & Horn, 1968). These schemes use hierarchical rules, mainly based on temperature and to a lesser extent precipitation thresholds, to define climatic zones. They are still regularly updated with new data (e.g. Belda *et al.*, 2014; Kottek *et al.*, 2006; Peel *et al.*, 2007). Vahl's climatic divisions (Reumert, 1946) attempt to address the arbitrary nature of Köppen's thresholds and certain mismatches between the classification scheme and observations, by using fewer hierarchical divisions and introducing precipitation probabilities. Holdridge Life Zones (Holdridge, 1967) and the Thornthwaite classification (Thornthwaite, 1948) move away from using mainly temperature and precipitation for classification, although they are still bioclimatic in origin. Holdridge uses a combination of precipitation, potential evapotranspiration, humidity, altitude and latitude to define biomes. Thornthwaite attempts to address the perceived arbitrary nature of the Köppen-Geiger thresholds and to create a more rational classification scheme. Thornthwaite uses climate observations from the USA to create a classification approach that relies on a precipitation-effectiveness index, a moisture index, thermal efficiency index and the absolute value of potential evapotranspiration. However, despite these improved alternatives, the original Köppen-Geiger scheme remains widely used today.

Currently, the main available climate classifications suffer from significant shortcomings when applied to hydrology. Haines *et al.* (1988) tested the ability of the Köppen-Geiger classification to predict typical global runoff regimes and found some relationship between climate zones and flow regimes, but also considerable spread in the data: a flow regime might occur in many climate zones, and a single climate zone might contain many flow regimes. Based on recent work (Addor *et al.*, 2018; Berghuijs *et al.*, 2014) in the USA, we can hypothesise that this is likely because Köppen-like climate classification schemes lack hydrologically relevant detail, in the form of the interaction between water and energy availability, climate seasonality and snowpack formation. Thornthwaite's classification comes close to addressing this, but is only based on USA data and untested in its accuracy for predicting global hydrologic regimes. Additionally, Thornthwaite already noted that "*variations in the heat factor of climate do not generally result in the development of sharply defined boundaries between vegetation formations"* and that "*the boundaries separating tropical, mesothermal, microthermal and subpolar climates are vague and ill-defined"*(Thornthwaite, 1943),

as a point of potential improvement for classification schemes. Traditionally, classification maps include sharp, unrealistic, boundaries between different classes. More recently, advances in data sciences (e.g. Schwämmle & Jensen, 2010) have led to more nuanced classification schemes in hydrology where catchments can belong to several classes at the same time, but with differing degrees of membership to each class (e.g. Sawicz *et al.*, 2011).

This study addresses an identified need for a global hydrologically-informed climate classification scheme, that (i) corresponds to observed similarities and differences in observed hydrological response, (ii) avoids introducing artificial boundaries between classes. We choose to address climate alone, without consideration of catchment characteristics, as a first step to developing a more general catchment classification.

2 Data

This study first uses gridded climate data to summarize the world's climate with several climate indices and uses these to define different climate clusters. Then, 1103 catchments are associated with the appropriate climate clusters using the catchments' locations and boundaries, after which streamflow data from the catchments is used to evaluate the hydrological usefulness of the climate clusters.

2.1 Climate data

This study uses monthly average climate values from the CRU TS v3.23 data set (Harris *et al.*, 2014), for the climatic variables precipitation (P), number of rain days per month (N, defined as days with P $\geq$ 0.1mm), temperature (T) and potential evapotranspiration (Ep). These data are available at a 0.5° x 0.5° resolution for the Earth's land areas, excluding Antarctica. The data set offers so-called primary variables, which include P, N and T, that are a re-analysis of station observations and existing climatology. The secondary variables, such as Ep, are estimated from the primary variables. Ep is estimated with a variant of the Penman-Monteith formula (Allen *et al.*, 1998; Harris *et al.*, 2014).

Ep values are missing for approximately 7.3% of global land cells due to incomplete coverage of the wind speed data needed for Ep estimation. Ep values are highly spatially correlated (average correlation coefficient = 0.99 in latitude direction, average 0.72 in longitude direction) and most of the missing values are bordered by cells for which E_p values are available. Nearly all missing values can be filled with a weighted nearest-neighbour approach, apart from several small islands that are too isolated for correlations to be a useful approach.

For this study, P/T/N/Ep data for 1984-2014 are averaged per month to find a typical year (e.g. the typical January P is the average of all 30 January P values from 1984 to 2014), to approximate the typical annual variation in all four climate variables.

2.2 Streamflow data

The Global Runoff Data Centre (The Global Runoff Data Centre, 2017b) manages a large database of river discharge data. This study uses a subset of data known as Pristine River Basins that contains daily streamflow data for 1182 gauging stations world-wide for the study period 1984-2014. The catchments in this dataset are asserted to have minimal development and river regulations and diversions. In addition, records for each catchment cover at least 20 years (overview of record lengths in Supplementary Information S.1.3) and exceed a certain accuracy threshold (The Global Runoff Data Centre, 2017a). We applied quality assurance procedures to the data, and as a result 79 catchments were excluded from this study (details in Supplementary Information S.1.2), leaving 1103 stations for use in this study.

Catchment boundary information is available for 718 of the GRDC Pristine Basins (The Global Runoff Data Centre, 2011). The remaining 449 catchments in the Pristine Basins set vary in area from 0.69 km^2 to 4,680,000 km^2, with median 596 km^2. Larger catchments can cover many grid cells and without information on the catchments' boundaries it is impossible to tell how varied the climate within each catchment is. Therefore we include only those catchments with an area smaller than or equivalent to the approximate area of 9 grid cells, for which the climate at each gauge's location might reasonably be considered representative of the climate in the whole catchment. This limits both the uncertainty about the prevailing climate in these catchments, and the number of catchments that must be excluded from further analysis (details in Supplementary Information S.1.1).

We create a typical streamflow year for each catchment from daily streamflow data, by taking the median flow for each Julian day (e.g. the typical Jan-1 flow refers to the median of all available Jan-1 for a particular place). We also align all flow records in time so that t=1 coincides with the start of the hydrological/water year for each location. Using the median flow decreases the influence of extreme events in the data and is preferable to the mean because of the skewed nature of flow variability. Catchments are spread across both hemispheres, so hydrological years are preferable to calendar years for comparison purposes. It is easier to visually assess similarities between flow patterns when distinctive features such as the seasonal flow peaks are aligned. By convention the water year in the Northern Hemisphere runs from October to September (e.g. Beck, de Roo and van Dijk, 2014; U.S. Geological Survey, 2016) however for the Southern Hemisphere both April to March (Beck *et al.*, 2014) and July to June (Utah State University, 2017) are in use. While conventions such as these can be useful on a small scale, on a global scale these are too general. Therefore, we use a 61-day moving window to find the period of maximum flow in a typical year for each catchment, and assume that the water year has started 120 days before this point (Figure 1). Both numbers are determined through trial-and-error and were found to give the best results (in terms of ease of visual comparison of flow similarity) for the data used in this study, but should be revisited if a more comprehensive data set is available.

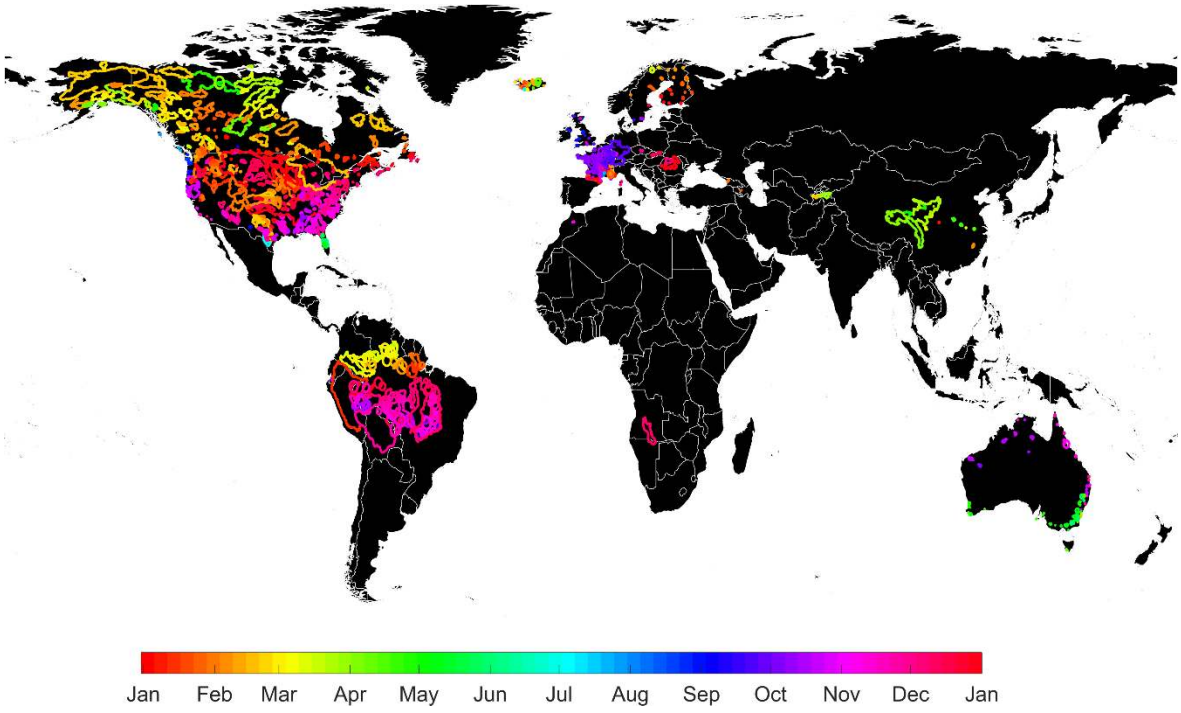

Figure 1: Location and boundaries (if available, circles with size relative to catchment area are used when not) of GRDC Pristine catchments. Catchments for which no boundary data is available are only used if the aproximate catchment length is smaller or equal to a climate correlation threshold length and removed from the analysis if larger. Colouring indicates approximate start of the hydrological year, here defined as the 120 days before the time of the 61-day average maximum flow.

3 Method

This study creates a climate classification scheme that summarises global climate patterns as a causal factor of global streamflow response. Causal factors (climate) and response (streamflow) are separated, meaning that our classification scheme is based on only climatic information and can be independently evaluated with streamflow data. First, we summarize the global climate with several gridded dimensionless indices (section 3.1). These climatic indices are clustered into fuzzy groups with a fuzzy c-means clustering algorithm (section 3.2) to define several climates that are representative of the land surface. We then evaluate the correspondence of the climatic clusters with global streamflow response, testing the hypothesis that locations within a cluster experience similar flow patterns while locations in different clusters show different streamflow regimes. We evaluate this both qualitatively through comparing typical seasonal flow patterns within and between climate clusters, and quantitatively through streamflow signature values and statistical tests. We compare the effectiveness of our climatic clustering with that of the Köppen-Geiger classification, testing the hypothesis that our scheme improves on another often-used method (section 3.3). Last, we investigate the potential of abandoning the idea of climate classes or clusters and show the benefits of viewing the global hydro-climate as a continuum rather than a patchwork of different classes (section 3.4).

3.1 Dimensionless climate indices

The climate at any given location influences the processes near the land surface and those concerning precipitation and evaporation. The balance between available water and energy determines whether water will remain on land or be returned to the atmosphere. Periods with lower temperatures can lead to snow pack formation, and precipitation intensity can influence whether water will infiltrate into the soil or become surface runoff. However, precipitation and temperature

(and by extension potential evapotranspiration) patterns are variable throughout the year and precipitation and temperature peaks are not necessarily in phase. It is thus plausible that our indices need to cover not only annual averages, but also provide a measure of the seasonal variability of climate variables. This leads to the hypothesis that, in addition to the total annual precipitation, five different climate aspects might be hydrologically relevant (e.g. Addor et al., 2017, 2018; Berghuijs et al., 2014; Milly, 1994; Woods, 2003, 2009): (i) the annual average aridity, specifying the ratio of available energy and water; (ii) the seasonality in aridity, indicating if seasonal water and energy distributions are in or out of phase; (iii) the fraction of precipitation that falls as snow, indicating whether precipitation will be (temporarily) stored on the land surface; (iv) the average rainfall intensity, showing whether rainfall will exceed infiltration rates and thus produce surface runoff; and (v) the seasonality of rainfall intensity, indicating whether infiltration excess runoff is more likely to occur in certain parts of the year.

We limit this work to aridity and snow indices for several reasons. First, although precipitation intensity can vary significantly across the world, its impact on local hydrology (i.e. whether rain infiltrates or becomes surface flow) depends on local catchment characteristics. Accounting for global differences in soil types and other catchment characteristics is considered beyond the scope of this work. Second, the CRU TS climate data set lacks information on the sub-monthly time scale, and precipitation intensity can thus only be quantified by dividing the monthly precipitation totals by the number of rain days per month (days with P ≥ 0.1mm). Both the annual average and seasonality of this approximate intensity are strongly inversely correlated with the annual average aridity (Spearman rank correlation coefficient $R < -0.8$ across all land cells) and thus are unlikely to add any significant new information at the global scale. Similarly, we considered using the absolute annual average precipitation [mm/y] as a metric, but this is strongly correlated with annual average aridity ($R = 0.74$). Several tests during clustering (not shown for brevity) confirm that these metrics indeed add very little independent information at the global scale. Third, earlier work (Addor et al., 2017, 2018; Berghuijs et al., 2014) shows that average and seasonal aridity indices and one snow index are strongly related to seasonal streamflow patterns, without considering rainfall intensity or absolute precipitation totals.

Using CRU TS climate data averaged into a typical year (section 2.1), we calculate three climate indices for each 0.5° land cell. We use a version of Thornthwaite's moisture index MI (Willmott & Feddema, 1992) to express average aridity (I_m) and its seasonality ($I_{m,r}$), and a numerical implementation of the fraction of annual precipitation that occurs as snowfall f_s (Woods, 2009). These indices have been used for climate classification before but not in this particular combination (e.g. Willmott & Feddema, 1992, for MI; Berghuijs *et al.*, 2014, for f_s). These indices describe the processes of interest using bounded intervals, which is useful for interpretation and clustering analysis.

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

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

$$I_{m,r} = \max MI(t) - \min MI(t) \quad (3)$$

$$f_s = \frac{\sum P(T(t) \leq T_0)}{\sum_{t=1}^{t=12} P(t)} \quad (4)$$

P(t), Ep(t) and T(t) are mean monthly observations of precipitation, potential evapotranspiration and temperature in the CRU TS data set. T_0 is a threshold temperature below which precipitation is assumed to occur as snow, here set at 0°C. The annual average moisture index I_m has range [-1, 1] where -1 indicates the most arid (water-limited) conditions and 1 indicates the most humid (energy-limited) conditions. The moisture index seasonality $I_{m,r}$ has range [0, 2] where 0 indicates that there are no intra-annual changes in the water/energy budget and 2 indicates that the climate switches between fully arid (I_m = -1) and fully saturated (I_m = 1) within a single year. f_s has range [0,1] where 0 indicates no snowfall in a year and 1 that all precipitation falls as snow. Note that f_s = 0 does not imply that the temperature does not go below the threshold temperature T_0, but merely that during this period no precipitation occurs. The indices rely on similar information and express phenomena with similar underlying causes (e.g. seasonality of aridity might be caused by a strong summer-winter contrast, which may also increase the likelihood of snowfall) so some correlation between the indices is unavoidable. The Spearman rank correlation between I_m and $I_{m,r}$ is 0.27, between I_m and f_s 0.27, and between $I_{m,r}$ and fs 0.37. These are considered to be sufficiently independent for use in this study, because each index has a different physical interpretation.

3.2 Selecting representative climates for comparison with the Köppen-Geiger classification

Traditional climate classification schemes use distinct boundaries between climate classes (e.g. Geiger, 1954; Kottek *et al.*, 2006; Peel *et al.*, 2007; Trewartha & Horn, 1968), but Thornthwaite already pointed out that climates change gradually in space and distinct boundaries do not do this justice (Thornthwaite, 1943). However, sharp boundaries are a logical and inescapable result of the classification method that underlies Köppen-like classifications. In this work, we argue that the global hydro-climate should be seen as a continuous spectrum and that imposing boundaries on this spectrum should generally be avoided. However, for illustration purposes we use an automated fuzzy c-means clustering algorithm (Bezdek, 1981) to select several representative points in the climate space described by our three indices. Each location (grid cell in the global data) belongs with a certain degree of membership to each representative climate, based on the similarity of each location's climate index values to the climate in each representative point. Memberships can vary from 0 (the location does not belong to this representative climate at all) to 1 (the location's climate is the same as the representative climate), with the possibility for a location to belong simultaneously to several representative climates. Using these representative climates, it is straightforward to compare how similar the hydrologic regimes are for locations with the same Köppen-Geiger class compared to locations with the same representative climate.

While the fuzzy c-means algorithm can objectively create clusters from data, it does require human input in finding the appropriate settings and determining the appropriate number of clusters. We use Matlab's c-means implementation (function *fcm*) in a multi-start framework to account for the inherent randomness resulting from its use of random initial cluster centroids. Before clustering, we standardize the values of our climate indices so that each has a range [0,1], to avoid biasing the clustering procedure towards the index with the largest range. The fuzzy c-means procedure uses a so-called fuzzifier parameter to allow data points to belong to different clusters through fuzzy membership. This parameter can be used to decrease the influence of data points that are near the boundaries between two clusters when determining the cluster centroid positions (Schwämmle &

Jensen, 2010). This value is kept at its default value of 2. The number of representative climates was determined through trial-and-error, by performing the clustering procedure with 2 to 30 clusters and analysing the resulting climate clusters. We did not use any river flow data to either create or help choose the number of climate clusters. We chose 18 clusters for communication purposes in this study, because this provides an adequate amount of detail but does not create overly specific geographically-focussed clusters. However, we emphasise that our key goal is the identification of climate indices for global hydrology, rather than the set of 18 clusters.

3.3 Effectiveness of hydrologic grouping based on representative climates versus Köppen-Geiger classes

We use GRDC river flow data for 1103 catchments to compare how well hydrologic regimes can be grouped based on our representative climates. We also group the same catchments based on their Köppen-Geiger climate class, to assess whether our approach improves upon this alternative. The success of this grouping exercise is determined with a qualitative approach to investigate typical streamflow patterns per group and a quantitative approach to investigate differences between streamflow signatures in each cluster. First, we define the membership degree of 1103 catchments to all 18 representative climates, using the catchment-averaged values of our three climate indices. For catchments without boundary information we assume that the outlet location is representative of the whole catchment. We can then show the typical flows per representative climate, using every catchment's membership degree to determine how closely the climate in each catchment resembles that of each representative climate. We assess the typical flows in a qualitative way.

We also assess the differences between flows per representative climate quantitatively through streamflow signatures and statistical tests. Olden and Poff (2003) categorize 171 streamflow signatures into five main types relating to flow event magnitude, frequency, duration, timing and rate of change, distinguishing between high and low flow conditions within the first three categories. This study uses 16 signatures that cover these 5 categories (Table 1), mainly following recommendations from Kuentz *et al.* (2016) and Addor *et al.* (2017). For each catchment, we calculate a signature's value per hydrological year and then take the average of these yearly signature values. We repeat this for all 16 signatures. Correlation analysis (not shown here for brevity) indicates that each signature contains some independent information although there is duplication of information as well. We consider this acceptable for our purposes because the signatures are only used to evaluate the two classification schemes and are not part of the classification methods themselves. The classification thus remains unbiased by potential duplicate information in the streamflow signatures.

Our null hypothesis is that there are no significant differences between signature values calculated for flows in different representative climates. The alternative is that there are differences between signature values of flows per representative climate, which indicates that our climate classification scheme can tell us something informative about the hydrologic response. The Wilcoxon two-sample test (Wilcoxon, 1945, cited Walpole, 1968, p. 232) is a suitable statistical test to compare a signature's values between two climate clusters, because the test assumes no knowledge of the distribution and parameters of the total population, and allows comparing samples with very different sizes. It allows testing of distributions (e.g. the values of a signature calculated for 70 catchments in climate I and 115 catchments in climate II) with H_0: $\mu_1 = \mu_2$. We apply this test to all climate cluster pairs and for all signatures. The sheer number of tests makes it likely that we will find significant differences through chance alone. We therefore investigate the number of signatures for which a climate pair is statistically different: if a pair is different for 16 out of 16 signatures, we can assume that typical streamflow for these pairs is different. If a statistical difference is only found for

1 out 16 signatures, it is more likely that we have found this result through chance. We repeat this analysis on the catchment grouping created based on the Köppen-Geiger class of each catchment and comment on the differences.

Table 1: Overview of the hydrological signatures used in this study. Signatures are calculated for every hydrological year available for each catchment, after which we take the mean for each signature across all hydrological years available for a catchment. Numbering in the leftmost column refers to ***Error! Reference source not found.****7d and* ***Error! Reference source not found.****.*

	Signature	Unit	Description	References
Magnitude				
1	Mean flow	[mm/d]	Mean of daily flow	-
11	Q5	[mm/d]	5^{th} percentile of daily flow	Kuentz *et al.* (2016)
12	Q95	[mm/d]	95^{th} percentile of daily flow	Kuentz *et al.* (2016)
14	Skewness	[-]	Mean divided by median of daily flow	Kuentz *et al.* (2016)
2	Baseflow index	[-]	Baseflow fraction of total flow	Gustard, Bullock and Dixon (1992)
4	High flow discharge	[-]	90^{th} percentile divided by median flow	Kuentz *et al.* (2016)
Frequency				
10	No flow frequency	[-]	Normalized average frequency of no flow (number of days with 0 flow)	-
8	Low flow frequency	[-]	Normalized average frequency of low flow (number of days with flow < 0.2*mean)	Olden and Poff (2003); Westerberg and McMillan (2015)
6	High flow frequency	[-]	Normalized average frequency of high flow (number of days with flow > 9*median)	Clausen and Biggs (2000); Westerberg and McMillan (2015)
Duration				
9	No flow duration	[-]	Normalized average duration of no flow (number of consecutive days with 0 flow)	-
7	Low flow duration	[-]	Normalized average duration of low flow (number of consecutive days < 0.2*mean)	Olden and Poff (2003); Westerberg and McMillan (2015)
5	High flow duration	[-]	Normalized average duration of high flow (number of consecutive days > 9*median)	Clausen and Biggs (2000); Westerberg and McMillan (2015)
Timing				
16	Half flow date	[-]	Fraction of year when 50% flow occurs	Court (1962)
15	Half flow interval	[-]	Fraction of year in which 25^{th} to 75^{th} percentile flow occurs	Court (1962)
Rate of change				
3	Flow duration curve slope	[-]	FDC slope between 33^{rd} and 66^{th} percentile in log space	(Yadav et al., 2007)
13	Rising limb density	[d^{-1}]	Number of rising limbs divided by time that hydrograph is rising	Sawicz *et al.* (2014)

3.4 Beyond catchment grouping and towards climatic assessment on a continuous spectrum

In addition to being a quantified way to communicate the climate of hydrological systems, these indices can be used as a rational way to transfer hydrological information from gauged to ungauged basins. This can also be a starting point to define more powerful hydrological similarity metrics, eventually resulting in a hydrological catchment classification scheme. In the second part of this paper we briefly explore the predictive power of the three climate indices. Each catchment is treated as ungauged in turn, and we use climatic similarity as a very preliminary flow prediction method. Climatic similarity is expressed as those catchments that (1) belong to the same Köppen-Geiger class,

(2) belong to the same climate cluster, or (3) are nearby based on standardized Euclidean distance in climate index space (so that every index has range [0,1]) expressed by the I_m, $I_{m,r}$ and f_s indices. In the latter case, we investigate both (3a) distance-based weighting of all catchments and (3b) distance-based weighting of the five catchments that are climatically the most similar to the "ungauged" catchment. We estimate both the flow regime of each "ungauged" catchment and values for the 16 signatures. The accuracy metric used to compare estimated and observed flow regimes is the Kling-Gupta Efficiency (KGE, Gupta *et al.*, 2009). The metric used to compare estimated and observed signature values is the absolute error.

4 Results

4.1 Approximating climatic gradients with representative climates

Figure 2 shows that values for the three climate indices (annual average aridity, I_m; the seasonal change in aridity, $I_{m,r}$; and the fraction of precipitation as snowfall, f_s) generally change gradually in space (Figure 2c-e for individual indices, 2b for a map combining all three indices into a single global overview). The presence of mountain ranges leads to relatively sharp transitions in climate (e.g. Canadian Rockies, Andes, European Alps, Himalayas). Large areas of deserts are visible in red. These are arid locations with a high potential evapotranspiration compared to available precipitation, only small seasonal changes in this ratio and no snowfall. Very wet regions (dark green) are centred mostly around the equator. These are areas with a continual water surplus and low snowfall. Traditionally this climate is associated with tropical rain forests but other areas (e.g. Scotland, Japan, northern New Zealand) show similar index values, even if the underlying climatic drivers are different in absolute terms. Regions in bright green and yellow show transitional zones between constantly arid and constantly wet regions. The transitional zones experience strong seasonality in their water-energy balance, either through clearly defined wet and dry seasons (seasonal rain), through summer and winter patterns (seasonal changes in potential evapotranspiration) or a combination of both. Blue and pink regions indicate places where nearly all precipitation occurs as snowfall. Figure 2a further shows that climates with low seasonality concentrate near both ends of the aridity (I_m) axis (bright red, dark green) and that annual average aridity is not necessarily an accurate representation of month-to-month aridity, especially in cases where the annual water and energy budgets are approximately balanced ($I_m = 0$).

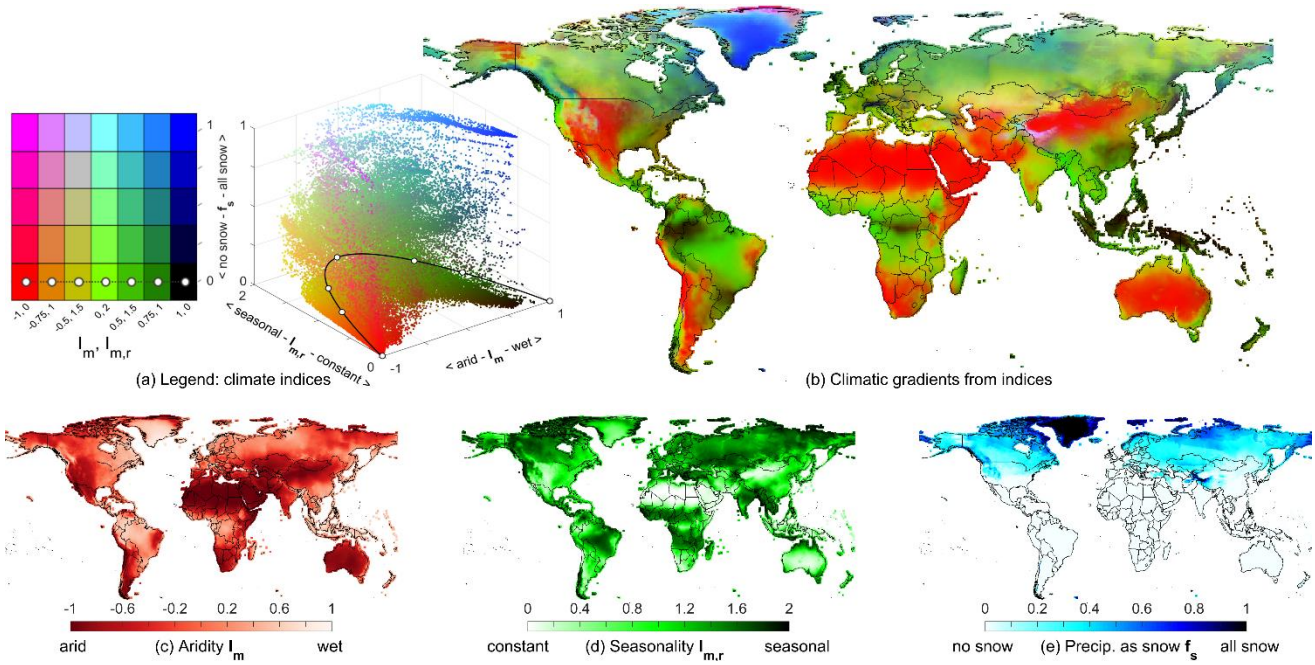

Figure 2: overview of average climate index values calculated for 1984-2014. (a) Climate index legend to help interpret figure 2b, showing how values on the three climate index axes determine the final RGB colour. The 3D-plot includes all land cells shown in 2b. The coloured square shows the colour scheme at 7 pre-determined points in 5 different I_m,$I_{m,r}$ planes along the f_s axis. (b) World map with each 0.5° resolution grid cell with local average aridity (Red), aridity seasonality (Green) and fraction precipitation as snowfall (Blue) determining the RGB colour scale. (c-e) Plots of average aridity I_m (red), aridity seasonality $I_{m,r}$ (green) and fraction of precipitation as snow f_s (blue) respectively, showing how each index varies globally.

In this part of the paper, we investigate whether our index-based classification is better suited for grouping hydrologically similar regimes than the Köppen-Geiger classification is. For a straightforward comparison with the Köppen-Geiger classes, we define 18 representative climates in our continuous climate-index space. These give a representative sample of the climate on the land surface. Figure 3a shows that 18 clusters approximate the climatic gradients in Figure 2b well, but the continuous variation of climate in space makes it impossible to create completely homogeneous classes where every location has a climate that strongly resembles that of the representative point it belongs to. Each grid cell is coloured based on the climate cluster that the cell belongs to with the highest degree of membership, here called the "main cluster" for each cell. Figure 3c shows how high this main membership degree is. A membership threshold of 0.5 is commonly seen as the cell belonging exclusively to its main cluster (Schwämmle & Jensen, 2010). Large areas of high membership degree values are visible (blue) and mainly occur away from cluster boundaries. However, the gradual nature of the changes in climate indices makes it difficult to classify all cells in homogenous clusters, as evidenced by the large number of cells that have membership degrees <0.5 for their main cluster. These cells can be thought of as belonging to multiple clusters simultaneously. With 18 clusters, slightly over half (50.4%) of all land cells have membership degrees >0.5 for their main cluster.

The position of climate cluster centroids (Figure 3b) shows that they are not distributed uniformly in climate index space and the centroid marker size (larger size indicates that a higher number of land cells have that cluster as their main cluster) shows that certain climates are more prevalent than others. The centroids approximate the pattern of all individual cells in climate index space (Figure 2b), showing where this pattern is dense and comparatively sparse. This is a result of the clustering procedure trying to maximise within-cluster similarity and between-cluster differences. In the absence of clearly defined clusters/groups in the data, as is the case with the gradual changes in

climate, the algorithm will struggle to draw appropriate boundaries between clusters and reverts to positioning the cluster centroids in response to point density. Figure 3d quantifies the number of cells for which each climate cluster is the main cluster and the degree of membership to the main cluster. Hot and very arid deserts (clusters 1 and 2) are both common and well-defined. Clusters 16 and 17 are on the other extremes (being very wet and snow-dominated respectively) and are also well-defined but contain fewer cells. In most clusters however, membership degrees are generally lower (< 0.5, red shading), because locations tend to lie between several representative climate points. Clusters 1, 2, 16 and 17 are relatively well-defined because their climates can be roughly approximated with terms as "always" and "no" (e.g. climate 1: always arid, no seasonality and no snow). The other clusters are all positioned at some non-extreme point on each climate index axis, and this makes it impossible to draw distinct appropriate boundaries between different climatic zones in these cases.

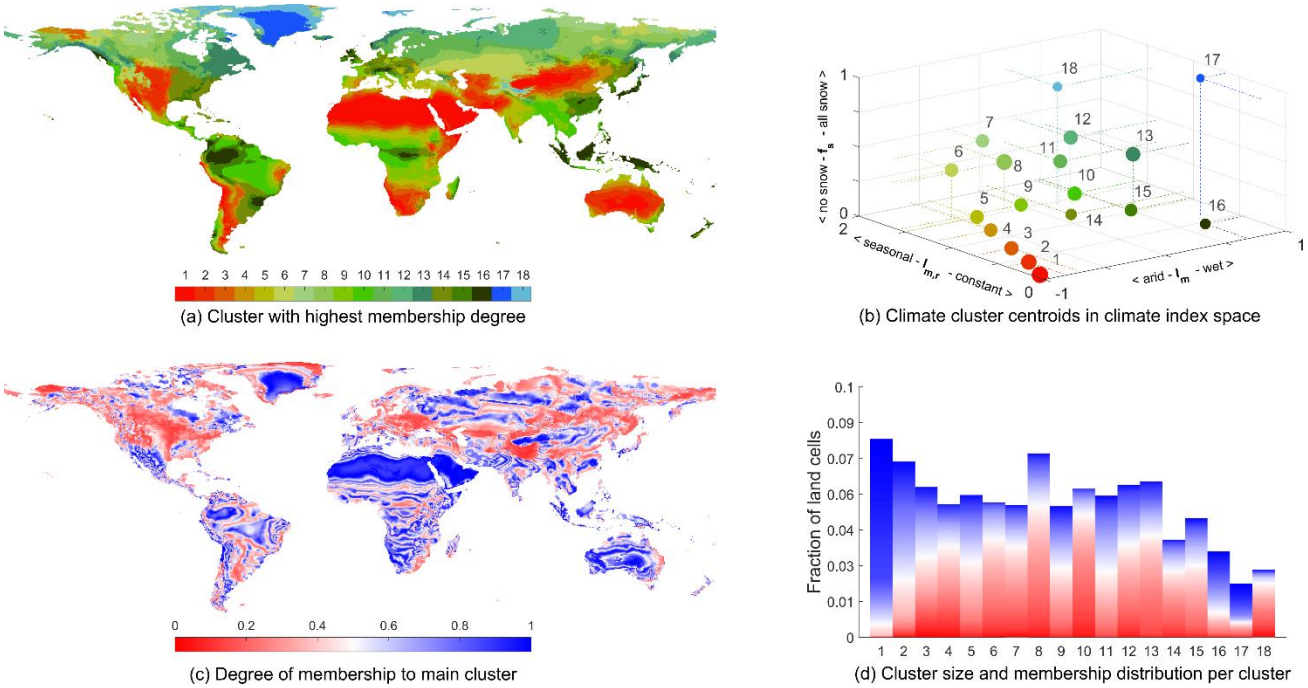

Figure 3: Results of fuzzy c-means clustering performed on climatic indices. (a) The cluster to which a cell belongs with the highest degree of membership (here called main cluster). (b) Location of climate cluster centroids in climate index space, with marker size corresponding to the number of cells for which a cluster is the main cluster (circle size is indicative of the number of cells having each cluster as main cluster). (c) The degree of membership with which each cell belongs to its main cluster; membership of each cell to the remaining 17 non-main clusters is by definition lower than its membership to the main cluster. (d) Number of cells for which a cluster is the main cluster (bar height) and degree of membership distribution per cluster (bar shading, legend in 4c).

4.2 Effectiveness of hydrologic grouping

4.2.1 Comparison of climatic gradients and Köppen-Geiger classes

The proposed new climate indices do not map directly onto Köppen-Geiger classes. The subclasses of the tropical (A) and arid (B) Köppen-Geiger main classes are relatively distinct from one another in the climate space defined by indices I_m, $I_{m,r}$ and f_s, whereas the subclasses of the colder temperate (C), continental (D) and polar (E) classes cover relatively similar regions in climate index space (*Figure 4*). This can be seen around the equator, in North-Africa, the Middle-East and most of Australia, where the Köppen-Geiger map (*Figure 4*b) is similar to the climate index map (*Figure 4*a). These regions are either very dry (through a combination of high temperatures and low precipitation) or very wet (resulting from very high precipitation) and see virtually no snowfall. These characteristics are captured well through the threshold approach in the Köppen-Geiger classification scheme. The

hydrologically relevant nuances of precipitation differences in colder climates are not well captured in the Köppen-Geiger scheme. This can be seen in e.g. the Eastern USA, Alaska, Greenland, most of Northern Europe and Russia, where the Köppen-Geiger boundaries are nearly exclusively determined by temperature thresholds. Different degrees of relative water availability and snow pack formation are lost in this classification. While the thresholds are an appropriate choice to define vegetation zones, as is the original goal of the Köppen-Geiger scheme, this approach is less relevant from a hydrological point of view. The climate indices contain more hydrologically relevant information, as the following sections will show.

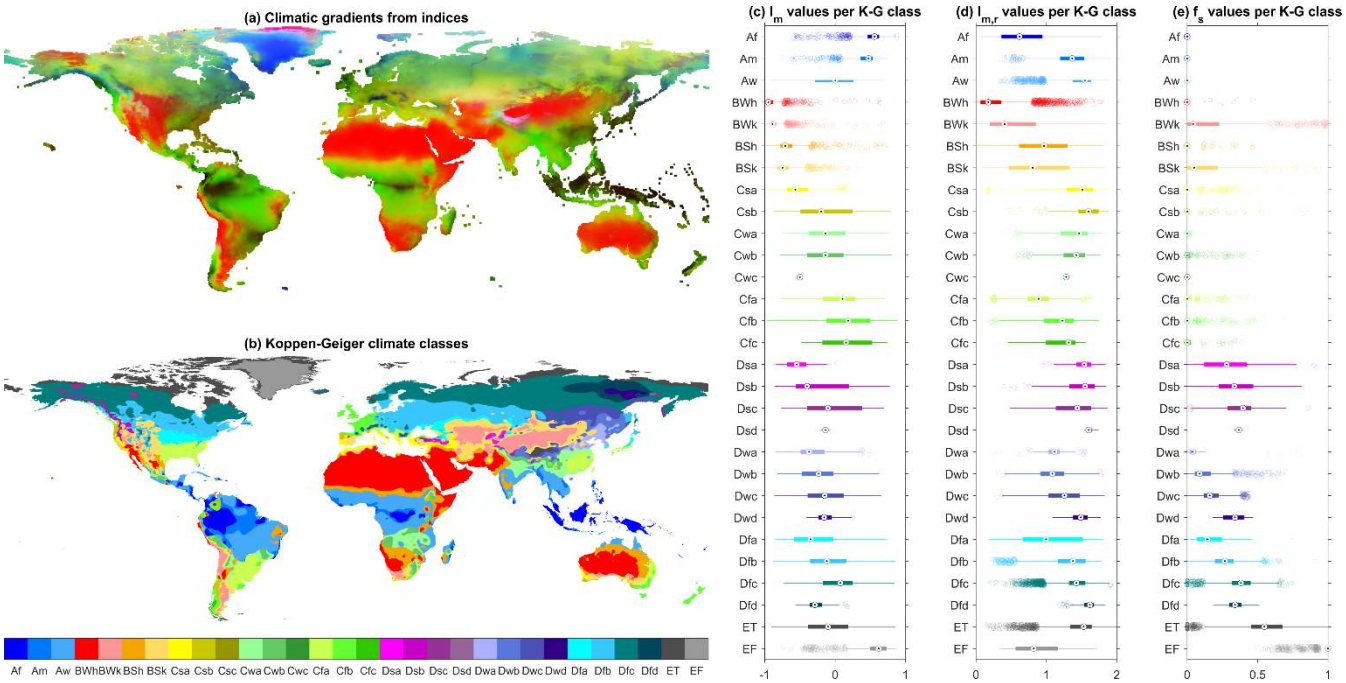

Figure 4: Comparison of the Köppen-Geiger climate classification and the global distribution of climate index values. (a) Global distribution of climate index values, as shown in Figure 2a. (b) Köppen-Geiger climate classification (Peel et al., 2007). (c-e) Boxplots of average aridity (Im), seasonality of aridity (Im,r) and fraction of precipitation as snow (fs) values per Köppen-Geiger class

4.2.2 Qualitative comparison of grouped flow regimes

Grouping the typical flow regime of all catchments according to the catchments' climate indices (Figure 5) shows that seasonal flow patterns gradually evolve along climate gradients. Clusters 4, 14 and 15 are similar with respect to the aridity seasonality $I_{m,r}$ and snow f_s metrics but are progressively less arid (I_m metric). As a result of this increased water availability, the clusters' typical flow patterns look similar but average flows become progressively higher. Clusters 1, 2, 3, 4, 5 and 6 are similarly arid (I_m) and low on snow (f_s) but their aridity is progressively more seasonal ($I_{m,r}$). The latter clusters thus occasionally experience a water-surplus, even if on average these places are severely water-limited. As a result, the average flow is low for all clusters, but a progressively higher seasonal flow peak can be seen. Clusters 6, 8, 11, 12 and 13 have similar values for the snow (f_s) and seasonality ($I_{m,r}$) metrics but are progressively less arid. As a result of this increased water availability, average flows become progressively higher and the main flow peak (likely resulting from snow melt since $f_s > 0$ at the cluster centroids) becomes progressively more pronounced.

Typical flow features such as average flow magnitude and flow peak height and shape are distinctly different between clusters, but climate can only inform us about average seasonal patterns. For example, the flow peak shape in snow dominated climates (e.g. clusters 13, 12, 11) shows a much sharper rise and decline than elsewhere, presumably due to snow storage and melt processes. In

warmer but not water-limited climates (e.g. 16, 15, 10) the flow peak rises and declines gradually, presumably as a result of seasonal changes in water surplus. However, within each cluster a wide variety of flows are included and what is true on average for the cluster, is not necessarily true for a single specific catchment. In a catchment classification context, climate is an important driver of hydrologic processes but the influence of the catchment itself (e.g. topography, vegetation, anthropogenic influence) cannot be ignored. This is however considered beyond the scope of this work.

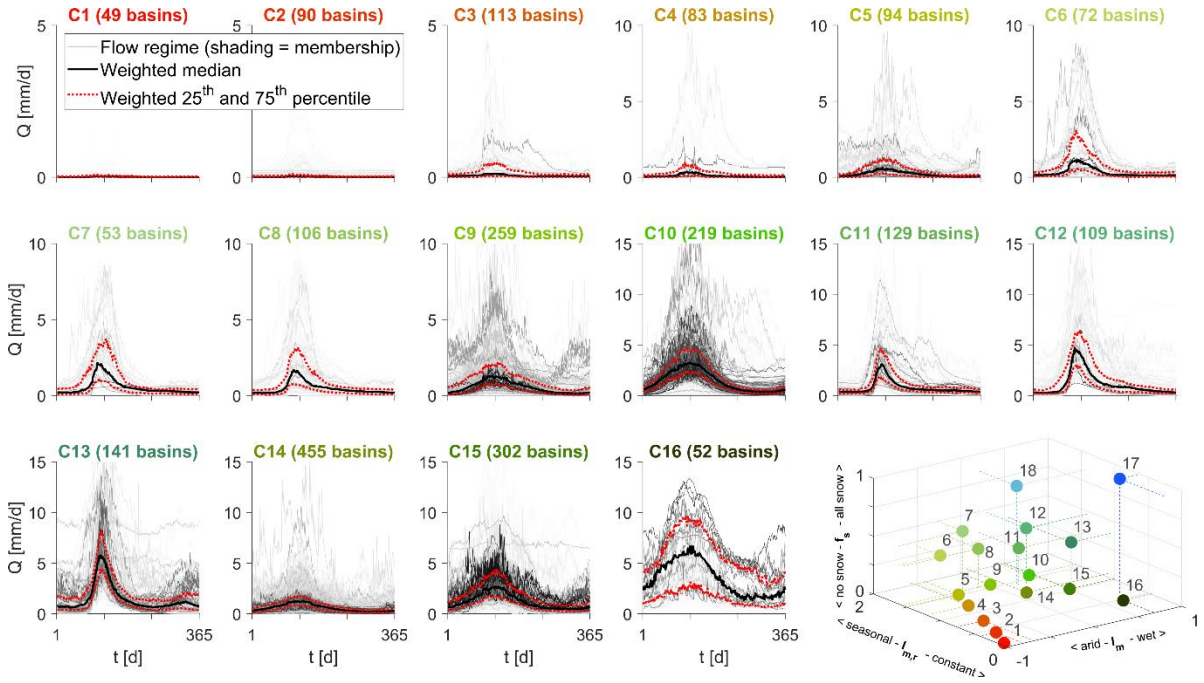

Figure 5: Typical flow regime for catchments grouped by climate cluster with the membership-weighted weighted median in black and the weighted 25th and 75th percentiles in red. Only catchments with a minimum membership of 0.10 or higher are shown, with darker lines corresponding to higher membership degrees. Includes all 1103 unique catchments, although catchments may appear in multiple climate plots. Title colouring corresponds to climate cluster centroids (figure 4a, 4b). Clusters 8 and 12 are not shown because the data lacks climate-specific flow records for these clusters.

Figure 4 showed that Köppen-Geiger main classes A and B show strong correspondence with our more arid and wet representative climates (e.g. climates 1-4, and 10, 15, 16 respectively). This pattern repeats with respect to grouping flow regimes by Köppen-Geiger classes (Figure 6): grouped flows for subclasses in the tropical (A) zone are very similar to the flows in representative climates 16, 15 and 10 (compare Figure 5), which have low aridity and no snowfall. The flows in the arid (B) subclasses are similar to our arid clusters 1, 2, 3 and 4. However, subclasses of C, D and E climates do not seem to group flow patterns in any meaningful way. To aid in this comparison, each flow record is coloured according to our catchment-averaged climatic index values and within main classes A and B the colouring seems relatively consistent. In climates C, D and E however, catchments with very different hydro-climates are lumped in each subclass and don't reveal any obvious typical flow pattern. E.g. subclass ET (polar tundra) contains flow patterns ranging from being nearly zero all-year round (orange), to very high, snowmelt-dominated regimes (blue-green). The snowmelt regimes are not as obviously grouped in the Köppen-Geiger classes as they are in the climate-index clusters (compare Figure 5 and Figure 6).

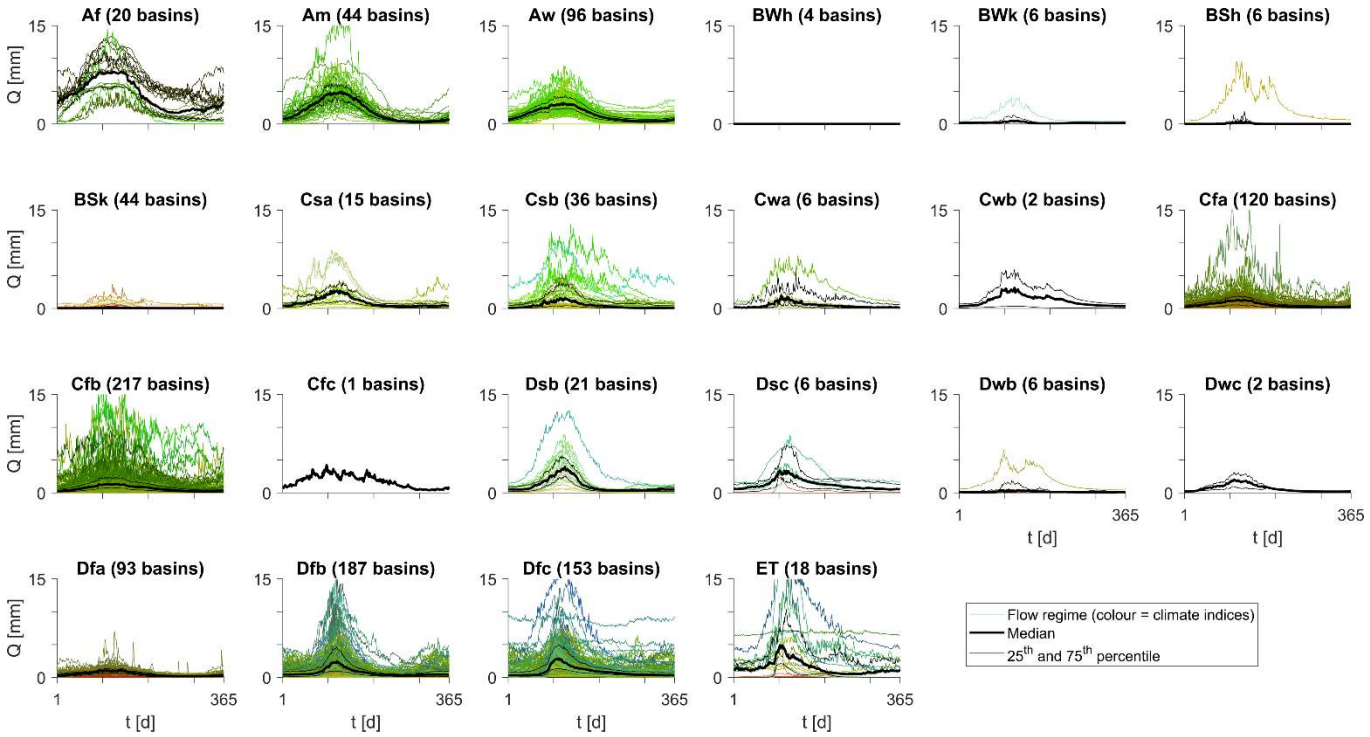

Figure 6: Typical annual flow for all 1103 catchments per Köppen-Geiger climate class with the median (black) and 25th and 75th percentiles (red). Typical flows from individual catchments are coloured by the catchment's climate index values as used with the climate index approach.

4.2.3 Quantitative comparison of grouped streamflow signatures

We use catchment membership degrees to create a weighted average streamflow signature value for 16 different streamflow signatures for each of the 18 representative climates. Statistical tests show that 145 out of the 153 possible combinations of two representative climates are statistically different at a 0.01 significance level. Another 7 out of 153 pairs are different at a 0.1 significance level and only a single pair shows no significant difference (p-value of 0.28; Figure 7d). Figure 7a shows an example of weighted average signature values per representative climate, here showing results for the average_flow signature (overview of all signatures is given in Supplementary Information S.2.2). A clear gradient is visible in the climate space, with the signature value increasing primarily as aridity decreases and secondarily as seasonality increases. Figure 7b shows the results of an empirical Wilcoxon test to determine the statistical significance of the differences in average_flow signature values between all representative climates. This procedure uses the average_flow signature value for each catchment, coupled with the catchment's membership degree (Figure 7c) to each representative climate, to estimate an empirical p-value (details in Supplementary Information S.2.1). Most representative climates have statistically different average_flow signature values at a 0.01 level (dark blue shading), but not all climate pairs are significantly different based on this single signature (white and red shades). Figure 7d shows the lowest p-value per climate pair across all 16 signatures and shows that 148 out of 153 climate pairs are different at the 0.05 significance level (bottom-left section of the figure). The top-right part of the figure shows the number of signatures for which the empirical p-value is below 0.05. The prevalence of darker shades indicates that climate pairs are statistically different for multiple signatures, indicating that our clustering approach can indeed group catchments with similar flow characteristics.

Climate pairs 6-8, 7-8, 6-7, 4-5 and 17-18 are not statistically different on any of the signatures, possibly due to a lack of climate-specific flow records. If there are statistical differences to be found,

either these differences manifest in flow characteristics not captured in the chosen signatures, or we lack climate-specific (high membership) flow records to construct an image of how a typical flow pattern for each representative climate looks. It is unlikely that the signatures are poorly chosen because they are adequate to distinguish between all other climate pairs (Figure 7d). Lack of climate-specific flow records is a likely explanation in the case of representative climate 18 (only 4 catchments with membership > 0.1; Figure 7c) and 17 (only 1 catchment has membership > 0.1; Figure 7c). Similarly, climates 6, 7 and 8 are close together in climate space and membership degrees of all catchments to each of these three representative climates are quite similar (Figure 7c). It is likely that the 1103 catchments lack the diversity that would allow the signatures to distinguish better between these three representative climates. This same explanation might be applied to climates 4 and 5. The alternative to these explanations is that there are no statistical differences between the typical flows of these representative climates; i.e. our assumption that the typical flow regime should be different between these representative climates, because the catchments associated with each climate have different hydro-climates, is false. However, given the success of the method with other climate pairs that are close together in climate space (e.g. 10-14, 11-12, 1-2), lack of climate-specific flow records seems the more likely explanation.

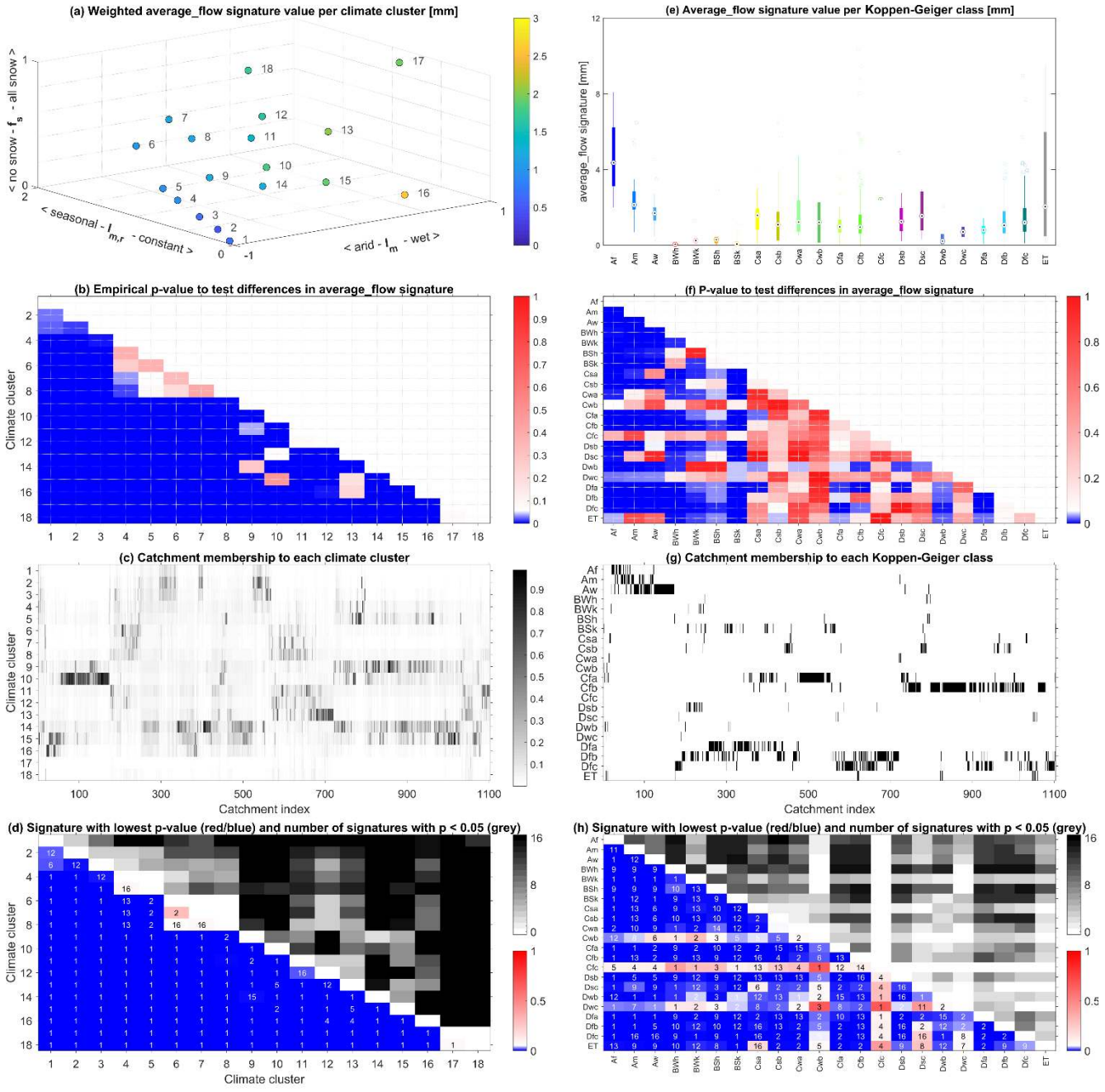

Figure 7: (a-d): quantitative differences between grouped flows regimes using climate clusters, (e-h) quantitative differences between grouped flows regimes using Köppen-Geiger classes. (a, e) Value of the average_flow signature per climate cluster/Köppen-Geiger class. Similar plots for all signatures in S.2.2 [a] Values are calculated as a weighted average from all 1103 catchments, with weights being each catchment's membership to a cluster. Numbers refer to climate clusters. [e] Boxplot colour refers to the legend in Figure 4. (b, f) Statistical tests to determine whether values of the average_flow signature per cluster/Köppen-Geiger class are statistically different. Blue shades show p ≤ 0.05, white shades show 0.05 ≤ p ≤ 0.10 and red shades show p > 0.10. [b] Results of an Empirical Wilcoxon test (S.2.1) used with cluster grouping. [f] Regular Wilcoxon test used with Köppen-Geiger grouping. (c, g) Membership degree of catchments (x-axis) to climate cluster/Köppen-Geiger classes (y-axis). [c] Darker shades show that a catchment belongs more strongly to a given cluster, and thus contributes more to the average cluster signature value. [g] All memberships to Köppen-Geiger classes are 1. (d, h) Bottom-left shows the lowest p-value from all 16 signatures, i.e. the largest significant difference between two groups. The number in each cell shows for which signature this lowest p-value is found (see Table 1 for numbering). Top-right grey shading shows for how many out of 16 signatures we find a significant difference (p ≤ 0.05). [d] Results cluster grouping. [h] Results Köppen-Geiger grouping.

Even though Köppen-Geiger has more climate classes than our climate-index method, analysis of signature values shows that grouping catchments by their dominant Köppen-Geiger climate class

leads to fewer distinguishable differences in typical flow patterns (Figure 7). Using the average_flow signature as an example, catchments in classes A and B seem to be sorted well according to their signature values (Figure 7e.) This is not the case in classes C, D and E, where the boxplots for subclasses tend to overlap (e.g. for Cfa, Cfb, Dfb and Dfc, all of which include 90 catchments or more). A Wilcoxon test confirms that statistically significant differences occur less frequently for the classes C, D and E than for classes A and B (Figure 7f). Using all signatures, we can find statistically significant differences in signature values between most Köppen-Geiger classes (Figure 7h). Where we don't, we likely lack enough catchments in our dataset for that subclass to make proper statistical inferences (i.e. Cwb, Cfc, Dwc; Figure 7g). However, in many cases in classes C, D and E we only find statistically significant differences in a few out of all 16 signatures (compare grey shades in Figure 7d and Figure 7h). This supports the idea that the temperate (C), continental (D) and polar (E) Köppen-Geiger classes are not well suited to grouping hydrological flow regimes.

4.3 Beyond climate grouping and towards a continuous representation of climates

Figure 8a shows the results of treating each catchment as ungauged in turn and using a climatic similarity approach to estimate the flow regime of this "ungauged" catchment. Comparing the effectiveness of Köppen-Geiger classes and our climate clusters, the clusters are somewhat more effective for estimating typical flow regimes. However, it is not our intent to advocate replacing one set of climate groups with another. Avoiding groups/clusters and using hydro-climatic similarity only, can have strong benefits compared to the classes/clusters approach. Using just 5 climatically similar basins to estimate the "ungauged" flows shows a significant increase in the number of basins where KGE values of the estimated regime exceeds 0. Quality of the flow estimates does depend on the number of samples used: using climate-weighted records from 1102 catchments to estimate each "ungauged" catchment leads to worse results than using either Köppen-Geiger classes or climate clusters. In these cases, most catchments are dissimilar from the "ungauged" one and these flow records dilute the estimate through sheer numbers (even if any individual catchment has a low weight). Using a small number of climatically similar basins overcomes this issue (within this data set, climatic similarity consistently outperforms climate clusters when fewer than 150 catchments are used to estimate any "ungauged" regime – the best results are obtained when 3-10 climatically similar catchments are used) .

A similar pattern is revealed when climatic similarity is used to estimate signature values for "ungauged" catchments (Figure 9). Climatic similarity of a few catchments to the target catchment generally gives the lowest errors across the ensemble of catchments. However, there is considerable spread in performance between the various signatures. Estimates of signatures associated with the magnitude of various parts of the water balance (e.g. average flow, q5, q95) and duration and frequency of high/low/no flow events seem to benefit the most from using index-based similarity over other options. Improvements are smaller for signatures related to timing (t50, t25_75) and rate of change (rising limb density, slope of the flow duration curve). Using climatic similarity of a few basins to estimate signatures also occasionally results in a higher occurrence of larger errors than other methods (e.g. for the skew signature). There is a delicate balance between using all available catchments (weighted by climatic similarity) and using just a few climatically very similar catchments to create estimates. Using more catchments decreases the risk of selecting a small number of climatically similar but structurally (in terms of vegetation, geology, etc) different catchments as donors, but also has the potential to dilute the quality of estimate through the sheer number of dissimilar basins included. Using fewer catchments can be very accurate due to the absence of dissimilar basins in the estimate, but leaves one vulnerable for differences in the catchment structure which this approach does not account for. In general, results seem to indicate that using

climatic similarity expressed through indices is a promising avenue for catchment classification. With refinement and introduction of catchment characteristics into the procedure, this approach to transfer knowledge between gauged and ungauged catchments can potentially be a powerful tool for prediction in ungauged basins.

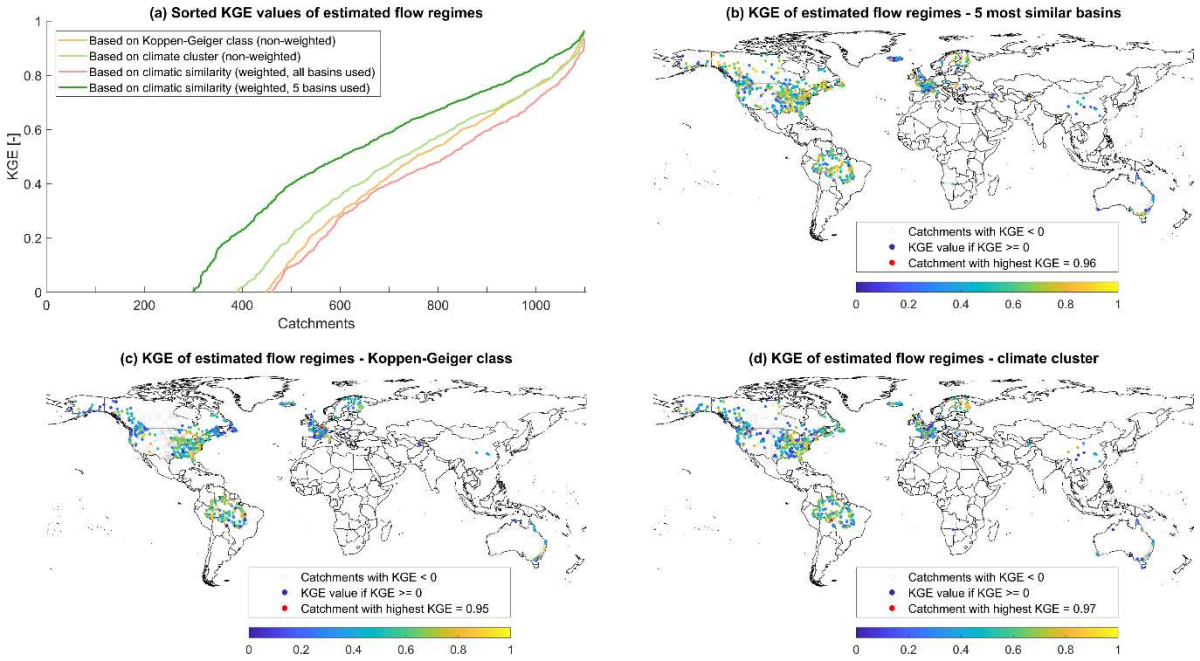

Figure 8: Estimating flow regimes through various definitions of climatic similarity, treating all 1103 catchments as if they were ungauged in turn. (a) Overview of sorted KGE values for estimated typical flow regimes, based on various similarity metrics. KGE values below 0 are not shown for clarity. (b-d) Geographical location of catchments and KGE value for estimated flow regime for each. Catchments with KGE < 0 shown in grey. [b] Estimation of "ungauged" catchment as the weighted daily mean of the regimes in the 5 climatically most similar catchments. [c] Estimation of "ungauged" catchment as the daily mean of the regimes in the same Köppen-Geiger class. [d] Estimation of "ungauged" catchment as the daily mean of the regimes in the same climatic cluster.

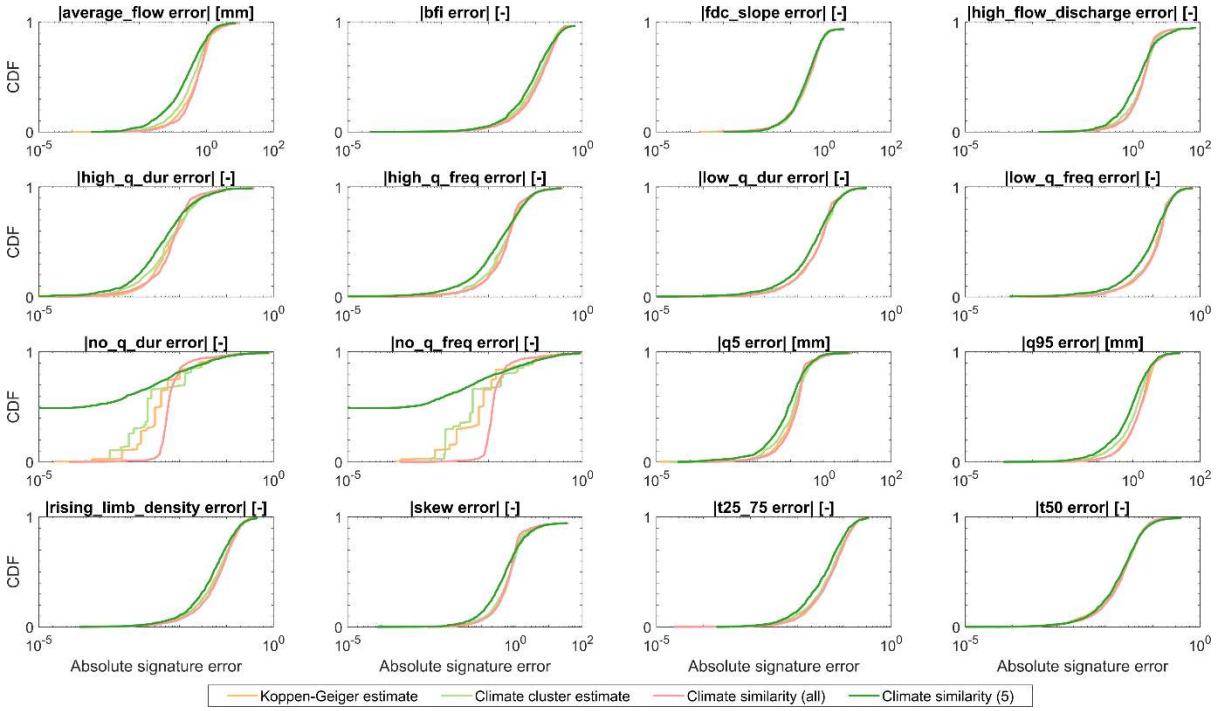

Figure 9: Cumulative Distribution Functions of absolute errors in signature value estimation. Signatures are estimated by treating each catchment as ungauged in turn and using climatic similarity to find donor catchments. Signature values from the donor catchments are averaged to estimate the signature's value in the "ungauged" catchment. Climatic similarity is expressed as belonging to the same Köppen-Geiger class (orange), the same climate cluster (light green), or as Euclidean distance in climate index space expressed by the I_m, $I_{m,r}$ and f_s indices. In the latter two cases, the estimated signature value for the "ungauged" catchments are based on distance-based weighting of all catchments (red) or the five catchments that are climatically the most similar to the "ungauged" catchment (dark green).

5 Discussion

This work presents a hydrologically-motivated alternative to traditional climate classification schemes, accounting for gradual changes in climate and the influence that has on flow regimes and streamflow signatures. This addresses two criticisms of traditional classification schemes used for hydrology, namely that their underlying motivation is not hydrological and the subjective nature of the number of classes and their distinct boundaries. Although we define 18 representative climates, these are intended as a communication device only, to enable straightforward comparison between our method and the Köppen-Geiger classification. Section 4.3 shows that clear benefits can be gained by using a continuous hydroclimatic spectrum instead.

We find that three simple climate indices, that quantify a location's average aridity (I_m), the seasonal range of water-versus-energy availability ($I_{m,r}$) and the fraction of precipitation that occurs as snowfall (f_s), are good indicators for finding similar hydrological regimes on a continuous scale. To further illustrate this, Figure 1010 shows the degree to which all catchments belong to representative climates 16, 15, 9 and 8 respectively, and how the typical flows that are strongly associated with each cluster look. From climate 16 to 15 to 9, the index values indicate progressively more arid climates, with increasing aridity seasonality and constant (nearly zero) snowfall. The corresponding streamflow regimes become lower on average as a result of increasing average aridity, with lower low flows resulting from the increase in aridity seasonality. From cluster 9 to 8, aridity and seasonality remain constant, but snowfall increases. The corresponding streamflow regimes in climate 8 are similar to those in climate 9 (both on average and during low flows) but have a much sharper high flow peak as a result of snow accumulation and melt processes. This reinforces the hypothesis that gradual changes in climatic conditions lead to gradual changes in

seasonal streamflow patterns and can be of importance during catchment classification and catchment similarity studies. Most catchment characteristics can be described on a continuous scale (e.g. area, elevation, slope, porosity, conductivity, degree of vegetation cover, leaf area index) and these results suggest that climate should be treated in the same way, rather than using discrete classes.

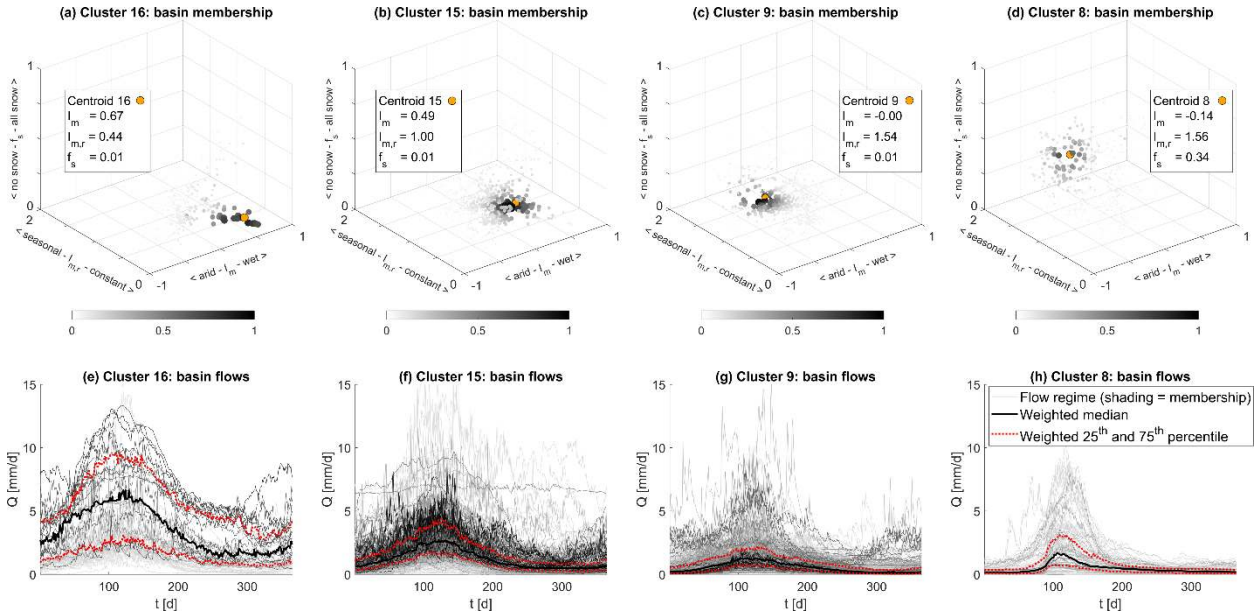

Figure 10: (a-d) Membership degree [0,1] of catchments to clusters 16, 15, 9 and 8 respectively (shading) and the catchment-averaged values for the three climate indices that describe aridity (I_m), seasonality of aridity ($I_{m,r}$) and fraction of precipitation as snowfall (f_s). (e-h) Typical streamflow in catchments with climates similar to representative points 16, 15, 9 and 8 respectively (flows shaded by their degree of membership to the cluster) with the weighted mean (black) and 25th and 75th percentile (red).

Our findings are in line with earlier work on the relation between seasonal streamflow patterns and climate (Addor et al., 2017, 2018; Berghuijs et al., 2014) and with work on the suitability of the Köppen-Geiger classification for mapping global flow regimes (Haines *et al.*, 1988). An important difference is that both Berghuijs *et al.* (2014) and Haines *et al.* (1988) create climatic classes by grouping flow regimes, whereas this work uses streamflow data only to evaluate the appropriateness of our climate indices in relation to hydrologic regimes. The specific climate indices we have chosen are slightly different from those used in Berghuijs *et al.* (2014) and Addor *et al.* (2017), but they are intended to capture the same climatic aspects (aridity, seasonality and snow). Both those studies are regional, focussing on the contiguous USA and our results indicates that their general findings (i.e. that 3 climate indices can be used in defining hydrologic similarity) might be applicable on the global scale as well. Haines *et al.* (1988) present a global classification of river flows based on monthly streamflow data and find considerable spread in how their proposed regimes relate to Köppen-Geiger classes, similar to our results. Haines *et al.* (1988) also compare their result to an earlier study by Beckinsale (1969), who adapted the Köppen-Geiger classification to apply to river regimes, and find that *"many of the 'different' regimes proposed by Beckinsale are in practice found not to be significantly different at the world scale"* (Haines *et al.*, 1988). This is a consequence of their choice to cap the number of possible regime classes, such that all classes contained a significant (but unspecified) number of observed flows and were consistent with known geographic features. Analyzing river flows on a continuous spectrum, such as proposed in this work, rather than using discrete groups would avoid this problem and allow rarely occuring regimes to be somewhere on this spectrum as well. However, we emphasize that our work is not intended as a river regime

classification scheme (which would necessarily involve accounting for a catchment's characteristics as well), but rather presents a hydrologically-relevant way of accounting for the influence of global climates in such a catchment classification.

5.1.1 On geographical proximity of the catchments

Geographical proximity of catchments could explain similarity between typical flows as well as climatic similarity, but the GRDC catchments are spread out enough in climatic and geographical space that this plays only a small role. Typical correlation lengths for hydrologic similarity are in the order of 100 to 200 km (e.g. Castiglioni *et al.*, 2011; Gottschalk *et al.*, 2011; Skøien *et al.*, 2003). Within the GRDC data set, approximately 1.3% of catchment pairs are within this distance from one another. This can explain certain similarities in flow patterns per climate cluster (Figure 5), because geographically close catchments are likely to have high membership degrees to the same representative climate(s). However, nearly all representative climates include catchments with high degrees of membership from at least two continents and all climates contain catchments that are far enough apart to ignore spatial correlation (see Supplementary Information S.3). In Figure 11, the typical flows in representative climate 15 are separated by continent (columns) and degree of membership (rows). Within a column, flows are relatively similar in pattern and size, which could be explained by relative geographical proximity (although the catchments still span several 100s of kilometres). Across columns however, especially above 0.50 membership degree, the flows on each of the four continents are remarkably similar. This reinforces the idea that similar climatic conditions lead to relatively similar flow patterns.

5.1.2 On the choice of climate indices

Our climate indices express the annual average water and energy budget, the seasonality of water and energy availability and the fraction of precipitation that occurs as snowfall. These indices relate to similar climatic attributes as earlier regional studies in the US (Addor et al., 2017, 2018; Berghuijs et al., 2014) have used but we use different equations for aridity and the seasonality of aridity. Our choices are motivated by both practical concerns and a need to find indices that confer relevant information on a global scale. Indices on a bounded interval are easy to visualise (in the case of our 3D climate index space, Figure 2a) and straightforward to normalize to a [0,1] interval. The latter is useful for clustering analysis and regionalization both within this study and for potential later work. Traditionally, aridity is often given as a dryness index PET / P (e.g. Addor et al., 2017; Berghuijs et al., 2014; Budyko, 1974) with range [0,∞>. We adopt a moisture index (Feddema, 2005; Thornthwaite, 1948) instead that expresses the same information but on a bounded interval [-1,1] and thus fits our criteria better. We've chosen to use the term "seasonality of aridity" over "precipitation seasonality" as is used in Berghuijs *et al.* (2014) and Addor *et al.* (2017). The precipitation seasonality metric is based on an expression of local P and PET time series as sinusoidal functions and finding the difference between the timing of the P and PET peak. This approach was originally developed in the context of snow modelling (Woods, 2009) and thus assumes that PET follows a distinct summer-winter pattern due to temperature seasonality. This assumption is appropriate in the context of the US but less so towards the equator. Furthermore, precipitation seasonality only confers information about timing and not relative volumes of P and PET. Therefore, we have opted to use the within-year range of our monthly moisture index as a seasonality metric instead. This metric conveys information about the possible states of water availability a location can go through, which is relevant information on a global scale (although it has its own limitations, see section 5.1.3).

5.1.3 On study limitations

This study has several limitations which can be improved upon in later work. First, we investigate the relation between climate and streamflow patterns by comparing averaged monthly climate values

747 over a 30-year period and median daily streamflow values. This approach smooths out outliers in
748 both climate and streamflow data, but ignores interannual variability. The results shown in this work
749 form a good basis to investigate interannual variability from. Second, the number of catchments
750 could be increased. The GRDC catchments were selected for their global nature and availability of
751 daily flow records, but this still leads to an underrepresentation of African and Asian river systems.
752 Additionally, to keep as many catchments for the analysis as possible, we have not set an upper
753 bound to catchment size (provided that the catchment boundaries are known). It is possible that in
754 large catchments not every part of the catchment contributes equally to overall runoff. The
755 catchment-average climate that we use might thus not be representative of the climate in the runoff
756 generating part of large catchments. A larger database of river basins would allow us to restrict the
757 analysis to smaller catchments where this issue is unlikely to play a role. Third, the seasonality of
758 aridity metric can be improved. The metric measures the range between the lowest and highest
759 monthly aridity value (eq. 3) but this range is not necessarily symmetrical around the average aridity
760 (I_m, eq. 2) value. There is thus a certain amount of non-uniqueness for each combination of I_m and $I_{m,r}$
761 values. For example, $I_m = 0$ and $I_{m,r} = 1$ can theoretically mean "this location is on average neither
762 arid nor wet, but reaches a very arid state at some point during the year", "it's neither arid nor wet
763 on average, but has a large water-surplus at some point" and everything in between. Extremely
764 asymmetrical occurrences are unlikely though, because this would require nearly balanced
765 precipitation and potential evapotranspiration all-year round, apart from a single extremely
766 dry/rainy month. The impact of this effect is currently hard to judge but might be investigated
767 through an increased number of catchments. Another limitation of the I_m and $I_{m,r}$ indices is that
768 (unlike the sine curve approach of Milly, 1994) they do not allow us to reconstruct the monthly times
769 series of climate. Other choices of climate indices could lead to improvements, but the results
770 already look promising: by comparing just three simple indices, we are able to locate catchments
771 with similar seasonal flow patterns and flow regimes. The climate index values can be used to define
772 a quantitative measure of "climatic similarity" between catchments in an easier, more succinct way
773 than is possible with earlier climate classification schemes.

774

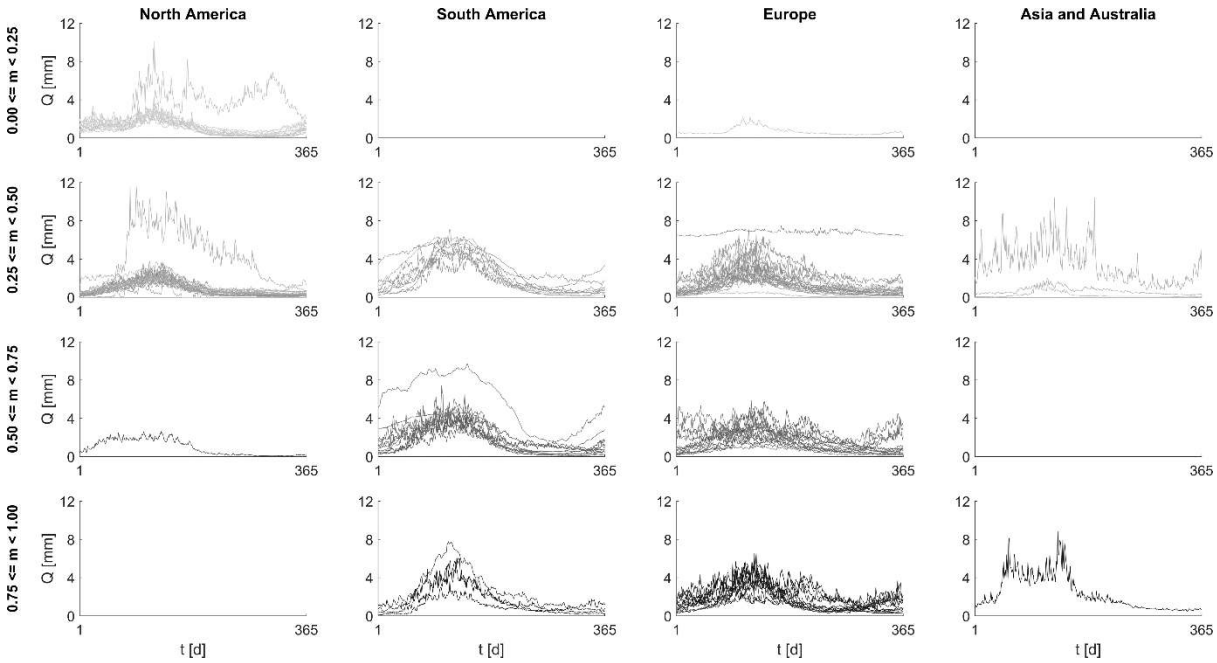

Figure 11: flows from GRDC catchments that have climate cluster 15 as their main cluster (i.e. their highest degree of membership to any cluster is to cluster 15) separated by continent (columns) and degree of membership (rows). Flow shading corresponds to degree of membership of each individual catchment to cluster 15.

6 Conclusions

Hydrology needs its own structured way to quantify climates, acknowledging that climates vary gradually on a global scale, that distinct boundaries between climate classes do not represent reality well, and that climate descriptors should explicitly including those climate aspects that drive changes in hydrologic regimes. Until now, climate classification in hydrology has either used classification schemes from other disciplines (e.g. the Köppen-Geiger scheme) or used ad-hoc methods (e.g. a within-study selection of metrics such as aridity or streamflow elasticity). In this work, causal factors (climate) and streamflow response are intentionally separated, meaning that the classification scheme presented here is based on only climatic information and can be evaluated with independent streamflow data. We define the hydro-climate on a global scale, using three dimensionless indices that describe each location's aridity, the seasonal changes in aridity and the fraction of precipitation that occurs as snowfall. Using 1103 catchments, we show that typical streamflow regimes and streamflow signature values correlate strongly with the local hydro-climate. Gradual spatial changes in climatic conditions are accompanied by gradual changes in flow regimes. In a climate classification context, using these three indices is a better way to identify hydrologically similar catchments than the Köppen-Geiger classification. This is partly because the Köppen-Geiger scheme is not hydrologically based and does not capture the hydrologically relevant nuances of colder climates properly, and partly because the Köppen-Geiger scheme uses discrete climate classes. The gradual changes in climatic and streamflow conditions are not adequately captured using discrete classes. Instead, it is more useful to view the global hydro-climate as a continuous spectrum on which every catchment is located. Regionalization of typical streamflow patterns and streamflow signature values tends to be better when a small number of climatically similar basins (i.e. close together in the climate space described by our climate indices) is used instead of donors chosen based on either Köppen-Geiger or climate cluster grouping. Using the work shown here, a catchment's climate can be described with three simple numbers, which allows easier knowledge transfer between catchments and can form the basis of a catchment classification method.

7 Acknowledgements and data

This work was funded by the EPSRC WISE CDT, grant reference number EP/L016214/1. CRU TS climate datasets are freely available from https://crudata.uea.ac.uk/cru/data/hrg/. This study uses version 3.23 (Harris et al., 2014), downloaded on 02-07-2016. GRDC streamflow data are available on request from http://www.bafg.de/GRDC/. This study uses a sub set known as "Climate Sensitive Stations Dataset (Pristine River Basins)" (The Global Runoff Data Centre, 2017b) downloaded on 16-05-2017. We gratefully acknowledge the input from the editor and three anonymous reviewers, whose insightful comments have helped clarify and improve this manuscript.

8 References

Addor, N., Newman, A. J., Mizukami, N., & Clark, M. P. (2017). 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-2017-169

Addor, N., Nearing, G., Prieto, C., Newman, A. J., Le Vine, N., & Clark, M. P. (2018). Selection of hydrological signatures for large-sample hydrology. *Water Resources Research*, *pre-print*. https://doi.org/10.17605/OSF.IO/2EM53

Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). *Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56*. Rome.

Archfield, S. A., Kennen, J. G., Carlisle, D. M., & Wolock, D. M. (2014). An Objective and Parsimonious Approach for Classifying Natural Flow Regimes at a Continental Scale. *River Research and Applications*, *30*, 1085–1095. https://doi.org/10.1002/rra.2710

Baize, D., & Girard, M.-C. (2009). *Référentiel pédologique 2008*. Editions Quae.

Beck, H. E., de Roo, A., & van Dijk, A. I. J. M. (2014). Global Maps of Streamflow Characteristics Based on Observations from Several Thousand Catchments*. *Journal of Hydrometeorology*, *16*(4), 1478–1501. https://doi.org/10.1175/JHM-D-14-0155.1

Beckinsale, R. P. (1969). River regimes. In R. J. Chorley (Ed.), *Water, earth, and man* (pp. 455–471). London: Methuen.

Belda, M., Holtanová, E., Halenka, T., & Kalvová, J. (2014). Climate classification revisited: from Köppen to Trewartha. *Climate Research*, *59*(1961), 1–13. https://doi.org/10.3354/cr01204

Berghuijs, W. R., Sivapalan, M., Woods, R. A., & Savenije, H. H. G. (2014). Patterns of similarity of seasonal water balances: A window into streamflow variability over a range of time scales. *Water Resources Research*, *50*(7), 5638–5661. https://doi.org/10.1002/2014WR015692

Budyko, M. I. (1974). *Climate and life*. New York: Academic Press.

Canfield Jr, D. E., Langeland, K. A., Maceina, M. J., Haller, W. T., Shireman, J. V., & Jones, J. R. (1983). Trophic state classification of lakes with aquatic macrophytes. *Canadian Journal of Fisheries and Aquatic Sciences*, *40*(10), 1713–1718.

Castiglioni, S., Castellarin, A., Montanari, A., Skøien, J. O., Laaha, G., & Blöschl, G. (2011). Smooth regional estimation of low-flow indices: physiographical space based interpolation and top-kriging. *Hydrology and Earth System Sciences*, *15*, 715–727. https://doi.org/10.5194/hess-15-715-2011

Clausen, B., & Biggs, B. J. F. (2000). Flow variables for ecological studies in temperate streams: Groupings based on covariance. *Journal of Hydrology*, *237*(3-4), 184–197. https://doi.org/10.1016/S0022-1694(00)00306-1

Coopersmith, E., Yaeger, M. A., Ye, S., Cheng, L., & Sivapalan, M. (2012). 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

Court, A. (1962). Measures of streamflow timing. *Journal of Geophysical Research*, *67*(11), 4335–4339. https://doi.org/10.1029/JZ067i011p04335

Davison, A. C., & Hinkley, D. V. (1997). Bootstrap methods and their application. (R. Gill, B. D. Ripley, S. Ross, M. Stein, & D. Williams, Eds.). Cambridge: Cambridge University Press.

Feddema, J. J. (2005). A Revised Thornthwaite-Type Global Climate Classification. *Physical Geography*, (November). https://doi.org/10.2747/0272-3646.26.6.442

Geiger, R. (1954). Klassifikation der Klimate nach W. Köppen" [Classification of climates after W. Köppen]. In *Landolt-Börnstein – Zahlenwerte und Funktionen aus Physik, Chemie, Astronomie, Geophysik und Technik, alte Serie* (pp. 603–607). Berlin: Springer.

Gottschalk, L., Leblois, E., & Skøien, J. O. (2011). Distance measures for hydrological data having a support. *Journal of Hydrology*, *402*(3-4), 415–421. https://doi.org/10.1016/j.jhydrol.2011.03.020

Grime, J. P. (1974). Vegetation classification by reference to strategies. *Nature*, *250*. https://doi.org/10.1038/250026a0

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1-2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Gustard, A., Bullock, A., & Dixon, J. . (1992). *Low flow estimation in the United Kingdom*.

Haines, A. T., Finlayson, B. L., & McMahon, T. A. (1988). A global classification of river regimes. *Applied Geography*, *8*(4), 255–272. https://doi.org/10.1016/0143-6228(88)90035-5

Harris, I., Jones, P. D., Osborn, T. J., & Lister, D. H. (2014). Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. *International Journal of Climatology*, *34*(3), 623–642. https://doi.org/10.1002/joc.3711

Hewitt, A. E. (1992). Soil classification in New Zealand - Legacy and lessons. *Australian Journal of Soil Research*, *30*(6), 843–854. https://doi.org/10.1071/SR9920843

Holdridge, L. R. (1967). *Life zone ecology*. San Jose, Costa Rica: Tropical Science Center. https://doi.org/Via 10.1046/j.1365-2699.1999.00329.x

Hutchinson, G. E., & Löffler, H. (1956). The Thermal Classification of Lakes. *Proceedings of the National Academy of Sciences*, *42*(2), 84–86.

IUSS Working Group WRB. (2015). *World Reference Base for Soil Resources 2014, update 2015 - International soil classification system for naming soils and creating legends for soil maps.* (World Soil). Rome: FAO.

Johnes, P., Moss, B., & Phillips, G. (1994). *Lakes - Classification & Monitoring: A strategy for the classification of lakes*.

Kennard, M. J., Pusey, B. J., Olden, J. D., Mackay, S. J., Stein, J. L., & Marsh, N. (2010). Classification of natural flow regimes in Australia to support environmental flow management. *Freshwater Biology*, *55*, 171–193. https://doi.org/10.1111/j.1365-2427.2009.02307.x

888 Kottek, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F. (2006). World map of the Köppen-Geiger
889 climate classification updated. *Meteorologische Zeitschrift*, *15*(3), 259–263.
890 https://doi.org/10.1127/0941-2948/2006/0130

891 Kuentz, A., Arheimer, B., Hundecha, Y., & Wagener, T. (2016). Understanding Hydrologic Variability
892 across Europe through Catchment Classification. *Hydrology and Earth System Sciences*
893 *Discussions*, *Under revi*(August), 1–28. https://doi.org/10.5194/hess-2016-428

894 Lewis Jr, W. M. (1983). A revised classification of lakes based on mixing. *Canadian Journal of*
895 *Fisheries and Aquatic Sciences*, *40*(10), 1779–1787.

896 Lilly, A. (2010). A hydrological classification of UK soils based on soil morphological data. In
897 *Proceedings of the 19th World Congress of Soil Science: Soil solutions for a changing world* (pp.
898 1–4).

899 McDonnell, J. J., & Woods, R. A. (2004). On the need for catchment classification. *Journal of*
900 *Hydrology*, *299*, 2–3. https://doi.org/10.1016/j.jhydrol.2004.09.003

901 Milly, P. C. D. (1994). Climate, soil water storage, and the average annual water balance. *Water*
902 *Resources Research*, *30*(7), 2143–2156.

903 North, B. V, Curtis, D., & Sham, P. C. (2002). A note on the calculation of empirical P values from
904 Monte Carlo procedures. American Journal of Human Genetics, 71(2), 439–41.
905 http://doi.org/10.1086/341527

906 Olden, J. D., & Poff, N. L. (2003). Redundancy and the choice of hydrologic indices for characterizing
907 streamflow regimes. *River Research and Applications*, *19*(2), 101–121.
908 https://doi.org/10.1002/rra.700

909 Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen-Geiger
910 climate classification. *Hydrology and Earth System Sciences*, *11*(5), 1633–1644.
911 https://doi.org/10.5194/hess-11-1633-2007

912 de Queiroz, K., & Gauthier, J. (1992). Phylogenetic Taxonomy. *Annual Review of Ecology and*
913 *Systematics*, *23*(1), 449–480. https://doi.org/10.1146/annurev.es.23.110192.002313

914 Reumert, J. (1946). Vahls climatic divisions. An explanation. *Geografisk Tidsskrift*, *48*.

915 Sawicz, K. A., Wagener, T., Sivapalan, M., Troch, P. A., & Carrillo, G. (2011). Catchment classification:
916 Empirical analysis of hydrologic similarity based on catchment function in the eastern USA.
917 *Hydrology and Earth System Sciences*, *15*(9), 2895–2911. https://doi.org/10.5194/hess-15-
918 2895-2011

919 Sawicz, K. A., Kelleher, C., Wagener, T., Troch, P., Sivapalan, M., & Carrillo, G. (2014). Characterizing
920 hydrologic change through catchment classification. *Hydrology and Earth System Sciences*,
921 *18*(1), 273–285. https://doi.org/10.5194/hess-18-273-2014

922 Scerri, E. S. (2007). *The periodic table: its story and its significance*. New York: Oxford University
923 Press.

924 Schwämmle, V., & Jensen, O. N. (2010). A simple and fast method to determine the parameters for
925 fuzzy c-means cluster analysis. *Bioinformatics*, *26*(22), 2841–2848.
926 https://doi.org/10.1093/bioinformatics/btq534

927 Skøien, J. O., Blöschl, G., & Western, A. W. (2003). Characteristic space scales and timescales in
928 hydrology. *Water Resources Research*, *39*(10). https://doi.org/10.1029/2002WR001736

929 Soil Classification Working Group. (1991). *Soil classification: a taxonomic system for South Africa*.

930 Pretoria: Department of Agricultural Development.

931 The Global Runoff Data Centre. (2011). Watershed Boundaries of GRDC Stations. 56068 Koblenz,
932 Germany: Global Runoff Data Centre.

933 The Global Runoff Data Centre. (2017a). Climate Sensitive Stations Dataset (Pristine River Basins).
934 Retrieved July 19, 2017, from
935 http://www.bafg.de/GRDC/EN/04_spcldtbss/46_CSS/css.html?nn=201574

936 The Global Runoff Data Centre. (2017b). GRDC pristine catchments data set, 1984-2014. 56068
937 Koblenz, Germany.

938 Thornthwaite, C. W. (1943). Problems in the Classification of Climates. *American Geographical*
939 *Society*, *33*(2), 233–255. https://doi.org/10.2307/209776

940 Thornthwaite, C. W. (1948). An Approach toward a Rational Classification of Climate. *Geographical*
941 *Review*, *38*(1), 55–94.

942 Trewartha, G. T., & Horn, L. H. (1968). *An introduction to climate 5th edition*. New York, NY, USA:
943 McGraw-Hill.

944 U.S. Geological Survey. (2016). What is a Water Year? Retrieved September 15, 2017, from
945 https://water.usgs.gov/nwc/explain_data.html

946 Utah State University. (2017). Water year. Retrieved May 24, 2017, from
947 https://climate.usurf.usu.edu/reports/waterYear.php

948 Viereck, L. A., Dyrness, C. T., Batten, A. R., & Wenzlick, K. J. (1992). *The Alaska Vegetation*
949 *Classification*.

950 Wagener, T., Sivapalan, M., Troch, P., & Woods, R. (2007). Catchment Classification and Hydrologic
951 Similarity. *Geography Compass*, *1*(4), 901–931. https://doi.org/10.1111/j.1749-
952 8198.2007.00039.x

953 Walpole, R. E. (1968). *Introduction to Statistics*. New York, NY, USA: The MacMillan Company.

954 Westerberg, I. K., & McMillan, H. K. (2015). Uncertainty in hydrological signatures. *Hydrology and*
955 *Earth System Sciences*, *19*(9), 3951–3968. https://doi.org/10.5194/hess-19-3951-2015

956 Willmott, C. J., & Feddema, J. J. (1992). A more Rational Climatic Moisture Index. *The Professional*
957 *Geographer*, *44*(1), 84–87. https://doi.org/10.1111/j.0033-0124.1992.00084.x

958 Woods, R. A. (2003). The relative roles of climate, soil, vegetation and topography in determining
959 seasonal and long-term catchment dynamics. *Advances in Water Resources*, *26*(3), 295–309.
960 https://doi.org/10.1016/S0309-1708(02)00164-1

961 Woods, R. A. (2009). Analytical model of seasonal climate impacts on snow hydrology: Continuous
962 snowpacks. *Advances in Water Resources*, *32*(10), 1465–1481.
963 https://doi.org/10.1016/j.advwatres.2009.06.011

964 Yadav, M., Wagener, T., & Gupta, H. (2007). Regionalization of constraints on expected watershed
965 response behavior for improved predictions in ungauged basins. *Advances in Water Resources*,
966 *30*(8), 1756–1774. https://doi.org/10.1016/j.advwatres.2007.01.005

967