Hydrological Sciences Journal

ISSN: 0262-6667 (Print) 2150-3435 (Online) Journal homepage: www.tandfonline.com/journals/thsj20

Functional data analysis to investigate controls on and changes in the seasonality of UK baseflow

Kathryn A. Leeming, John P. Bloomfield, Gemma Coxon & Yanchen Zheng

To cite this article: Kathryn A. Leeming, John P. Bloomfield, Gemma Coxon & Yanchen Zheng (2025) Functional data analysis to investigate controls on and changes in the seasonality of UK baseflow, Hydrological Sciences Journal, 70:3, 522-534, DOI: 10.1080/02626667.2024.2434714

To link to this article: https://doi.org/10.1080/02626667.2024.2434714

HYDROLOGICAL SCIENCES JOURNAL
2025, VOL. 70, NO. 3, 522–534
https://doi.org/10.1080/02626667.2024.2434714

TECHNICAL NOTE

OPEN ACCESS Check for updates

Functional data analysis to investigate controls on and changes in the seasonality of UK baseflow

Kathryn A. Leeming[a], John P. Bloomfield[b], Gemma Coxon[c] and Yanchen Zheng[c]

[a]British Geological Survey, Keyworth, UK; [b]British Geological Survey, Wallingford, UK; [c]School of Geographical Sciences, University of Bristol, Bristol, UK

ABSTRACT

Continuous streamflow is critical for sustaining ecological systems and ensuring water resource security. Understanding controls on and changes in flows, including the seasonality of baseflow, is therefore an important task. Baseflow seasons have typically been investigated separately, potentially missing hydro-ecologically important timing changes. Instead, we apply a functional data analysis clustering approach to seasonal patterns of baseflow hydrographs for 671 catchments across Great Britain (GB). The baseflow clusters are characterized as early-, mid-, and late-season peaks. The spatial distribution of the baseflow seasonality clusters is closely connected to the baseflow index and a partition tree shows the influence of catchment topological, hydrogeological and soil factors. Changes in timing of baseflow seasonality are compared to climate seasonality. In GB there appears to be a small but systematic influence of a warming climate on baseflow seasonality via effective rainfall with a tendency for earlier seasonal baseflow peaks, with greater timing changes in snow-influenced catchments.

ARTICLE HISTORY
Received 21 December 2023
Accepted 29 October 2024

EDITOR
R. Singh

ASSOCIATE EDITOR
F. Tauro

KEYWORDS
seasonality; Functional Data Analysis (FDA); CAMELS-GB; snowmelt

1 Introduction

Baseflow is the delayed component of streamflow fed by subsurface storage between precipitation and/or snowmelt events (Tallaksen 1995, Price 2011, Zhang *et al.* 2017, Gnann *et al.* 2019, Singh *et al.* 2019). It is of interest for a number of reasons: baseflow can act to regulate the quality and temperature of streamflow (Jordan *et al.* 1997, Gomez-Velez *et al.* 2015, Hare *et al.* 2021), it supports ecological flows and ecosystem functioning (Poff *et al.* 1997, Boulton 2003), and, importantly for water resources, it sustains surface flows at times when there is a deficit in precipitation, and so baseflow is a significant component of streamflow during episodes of low flow and drought (Smakhtin 2001, Miller *et al.* 2016). Consequently, there is a need to understand the controls on and changes in baseflow, including the seasonality of baseflow – the focus of the current study.

Baseflow is a hydrological phenomenon that represents an integrated, whole-catchment response to meteorological and other environmental change signals (Bloomfield *et al.* 2009, 2011, Price 2011) as well as to water resource management practices, such as abstraction and discharge within catchments (Bloomfield *et al.* 2021). It typically exhibits catchment-specific responses over a wide range of spatio-temporal scales to variability or changes in driving climatology and longer-term climate change, and to changes in catchments, such as land-use and land-cover change.

A number of studies have considered baseflow and climate-influenced changes in many parts of the world, ranging from local to global spatial scales (e.g. Wang and Cai 2010, Ayers *et al.* 2021, Miller *et al.* 2021, Mo at al. 2021). Previous investigations of change in baseflow have typically used non-parametric methods such as the Mann-Kendall trend test, or slope-based methods (e.g. Ficklin *et al.* 2016, Mohammed and Scholz 2016, Ahiablame *et al.* 2017, Bosch *et al.* 2017, Rumsey *et al.* 2020, Tan *et al.* 2020) to characterize annual and/or seasonal trends. Consistent changes across time can be identified using these methods. However, a drawback to using trend tests such as these is that only monotonic changes can be identified and the tests are not particularly suitable to characterize changes in seasonality with time other than by simply segmenting the data into predefined "seasons" (as in Wang and Cai 2010, Ficklin *et al.* 2016, Bosch *et al.* 2017, Tan *et al.* 2020). In addition, care must be applied to the choice of test as autoregression and seasonal behaviour in a time series can affect the performance of trend tests (Hirsch and Slack 1984). Splitting the hydrographs by season allows for the investigation of higher or lower flows within each season, but more subtle changes (particularly timing changes), will be lost in the season-splitting approach. An alternative approach is to use functional data analysis (FDA) (Ramsey and Silverman 2005, Bouveyron *et al.* 2015). FDA treats the data (e.g. annual hydrographs) as a curve, rather than as discrete, sequential temporal observations in a time series, and allows for general seasonal shapes in the data to be identified and compared. Functional clustering enables groups of catchments with similar seasonal patterns of baseflow across time blocks (such as early, mid-, or

CONTACT Kathryn A. Leeming kle@bgs.ac.uk British Geological Survey, Keyworth NG12 5GG, UK
Supplemental data for this article can be accessed online at https://doi.org/10.1080/02626667.2024.2434714.

late seasonality) to be identified. Our approach does not consider seasons separately and instead considers the annual patterns that characterize seasonality. This is similar to how vegetation phenology changes are considered as alterations to seasonal timings (Kim *et al.* 2018, Piao *et al.* 2019, Geng *et al.* 2020, Chen *et al.* 2022). A number of other studies have applied FDA to hydrological data (e.g. Haggarty *et al.* 2015, Suhaila and Yusop 2016, Ternynck *et al.* 2017, Larabi *et al.* 2018, Alaya *et al.* 2020, Ghumman *et al.* 2020). However, to date, FDA has not been applied to the quantification of changes in hydrological seasonality.

To address the gap in observation-based understanding of changes in and controls on the seasonality of baseflow in catchments in temperate settings, a large-sample dataset for Great Britain (GB), Catchment attributes and hydro-meteorological time series for 671 catchments across Great Britain [CAMELS-GB] (Coxon *et al.* 2020a, 2020b) is analysed here using an application of FDA methods (Bouveyron *et al.* 2015, Ramsey and Silverman 2005). Data from the CAMELS-GB dataset catchments for the period 1976 to 2015 is split into two 20-year time blocks (1976–1995, 1996–2015) to characterize the distributions of average annual patterns of baseflow and how these vary over time. Following a description of the study area and analytical methods, the FDA approach is used to characterize the seasonality of the standardized baseflow hydrographs, and controls on the seasonal hydrographs and their changes are investigated.

2 Study area and data

2.1 Study area

The study area, GB, consists of England, Scotland and Wales (Fig. 1(a)) and includes a wide range of climate–landscape–water management features as described in Coxon *et al.* (2020b). Catchments in the north and northwest of the study area typically have higher mean elevations than those in the south and southeast of GB and the prevailing climatology reflects the broad gradient in catchment physiography. Wet and cooler conditions with reduced evapotranspiration are typically prevalent in the north and west of GB, compared with relatively dry and warmer conditions in the southeast (Fig. 1(b)).

Annual mean precipitation over England shows no systematic trends with time since records began in 1766 and there has been no attribution of changes in annual mean precipitation to anthropogenic factors (Jenkins *et al.* 2008, Watts *et al.* 2015). Precipitation in the UK is, however, seasonal and variable over a range of spatio-temporal scales, with a tendency towards drier summers in the southeast and wetter winters in the northwest (Jenkins *et al.* 2008, Watts *et al.* 2015) and towards showing significant inter-annual variations, including episodes of meteorological extremes (Bloomfield and Marchant 2013). Although annual average precipitation has not changed significantly in the observational record, there is a tendency for increasing winter rainfall and with more winter rain falling during intense events (Jenkins *et al.* 2008, Burt and Ferranti 2011, Jones *et al.* 2012). Air temperature has increased by about 1°C between 1980 and 2015 (Jenkins *et al.* 2008, Watts *et al.* 2015) consistent with long-term global warming trends. However, there have been few studies of historical changes in associated evapotranspiration. Kay *et al.* (2013) documented increased potential and actual evapotranspiration (PET and AE) across GB between 1961 and 2012 and Watts *et al.* (2015) speculated that it is reasonable to hypothesize that PET has increased in line with decadal-scale warming of air temperatures over GB.

High-productivity aquifers are largely found in the southeast and east of GB (Allen *et al.* 1997, Bloomfield *et al.* 2009, Marchant and Bloomfield 2018), whereas less productive aquifers and non-aquifers are generally more extensive in the west and northwest. Catchments in which clay-dominated soils

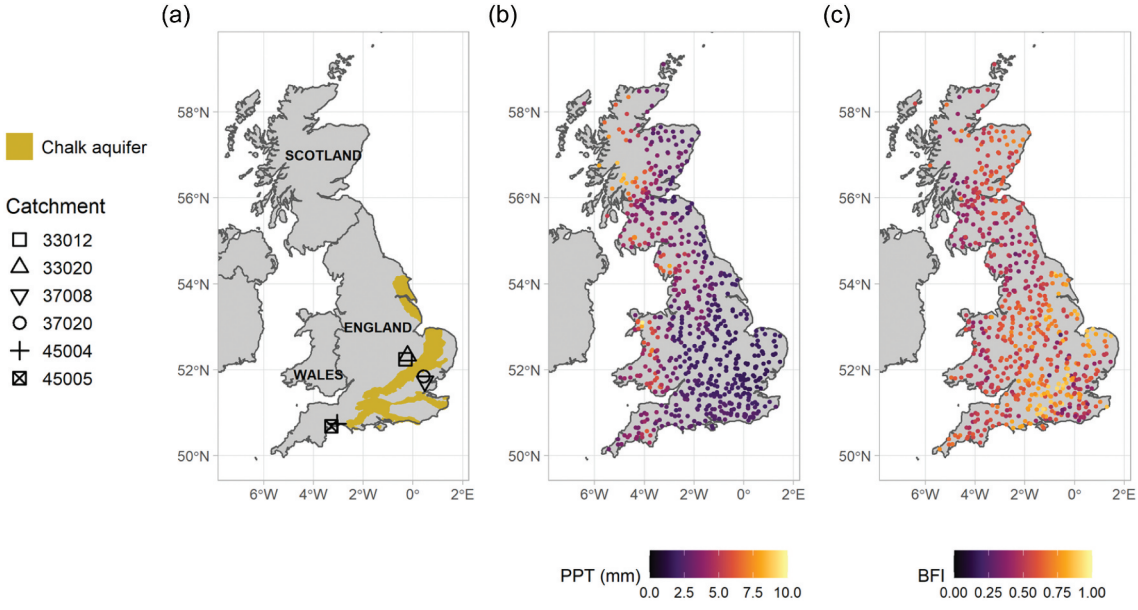

Figure 1. (a) Map of the study area with Chalk aquifer designation, country boundaries and paired catchments (described in Section 3.2). (b) Mean daily precipitation at CAMELS-GB catchments. (c) Baseflow index (BFI) at CAMELS-GB catchments.

overlie mudrock and clay bedrock formations and catchments with extensive glacial till deposits are typically present in central and eastern areas (Bloomfield *et al.* 2009, 2021, Bricker and Bloomfield 2014). These regional variations in underlying hydrogeological conditions are reflected in spatial variation in the baseflow index (BFI) across GB (Fig. 1(c)), with the highest BFI associated with streams flowing over the unconfined Chalk (a fractured microporous limestone of Cretaceous age and the main aquifer in GB) of southern, southeast and eastern England (Fig. 1(a)), and the lowest baseflow in northern and western catchments where low permeability–low storage bedrock and superficial deposits predominate (Allen *et al.* 1997, Bloomfield *et al.* 2021).

In a previous analysis of baseflow data from CAMELS-GB using multiple linear regression, Bloomfield *et al.* (2021) showed that, even though natural covariates, such as topography, aridity, and fractional area of highly productive fractured aquifers, provide the main explanatory power, BFI is also affected by groundwater abstraction and to a lesser extent discharges to rivers from sewage treatment works. Groundwater abstraction is focused on major aquifers, particularly the Chalk, in GB; however, discharges to rivers from sewage treatment works have no regional focus and are typically highest in catchments with large populations and high use of water (Bloomfield *et al.* 2021, fig. 2).

2.2 Data

Hydro-climatic time series for 671 catchments in GB are used from the CAMELS-GB dataset (Coxon *et al.* 2020a, 2020b). These data are a combination of UK National River Flow Archive (NRFA) and meteorological time series, provided at a daily resolution for the period 1970 to 2015. The streamflow series were collected by agencies including the Environment Agency, Natural Resources Wales and the Scottish Environmental Protection Agency and then compiled and quality checked by the NRFA. The data feature a good spatial coverage of GB, and over 80% of the locations have under 20% missing flow data, converted to mm d^{-1}.

The daily precipitation data in CAMELS-GB are derived from the gridded estimates of daily and monthly areal rainfall for the United Kingdom [CEH-GEAR] dataset (Keller *et al.* 2015). These data consist of observations from Met Office UK gauges, quality checked and converted to grid format using natural neighbour interpolation. Snow fraction is taken from the CAMELS-GB dataset (Coxon *et al.* 2020a) and given by the fraction of precipitation falling as snow for days colder than 0°C.

The CAMELS-GB temperature data used in this study are catchment daily averaged temperature from the climate hydrology and ecology research support system - meteorological [CHESS-met] dataset (Robinson *et al.* 2017), and CAMELS-GB PET data used in this study are catchment daily averaged PET for a well-watered grass based on the Penman–Monteith equation (Robinson *et al.* 2020).

The baseflow series are derived from the daily CAMELS-GB streamflow series using the Lyne-Hollick filter with standard settings (Ladson *et al.* 2013). The Lyne-Hollick digital filtering approach is chosen as this enables the separation of hundreds of baseflow series without requiring additional estimation of parameters. The monthly average baseflow values for each location are calculated, then the average seasonal shapes for each time block (A: 1976–1995 and B: 1996–2015) are formed by taking the median over the years in the time block. Any locations and months with less than 10 years' worth of data within a 20-year time block are removed from the dataset. This means that catchments may be present in one time block but not the other. These seasonal shapes are standardized to have a mean of 0 and standard deviation of 1, so that the FDA method considers different seasonal shapes rather than absolute levels or the magnitude of annual variation of baseflow. This treatment of the seasonal shapes allows a novel application of the FDA method to the seasonal distribution of the baseflow, focusing on the timing of annual patterns and changes across space and time.

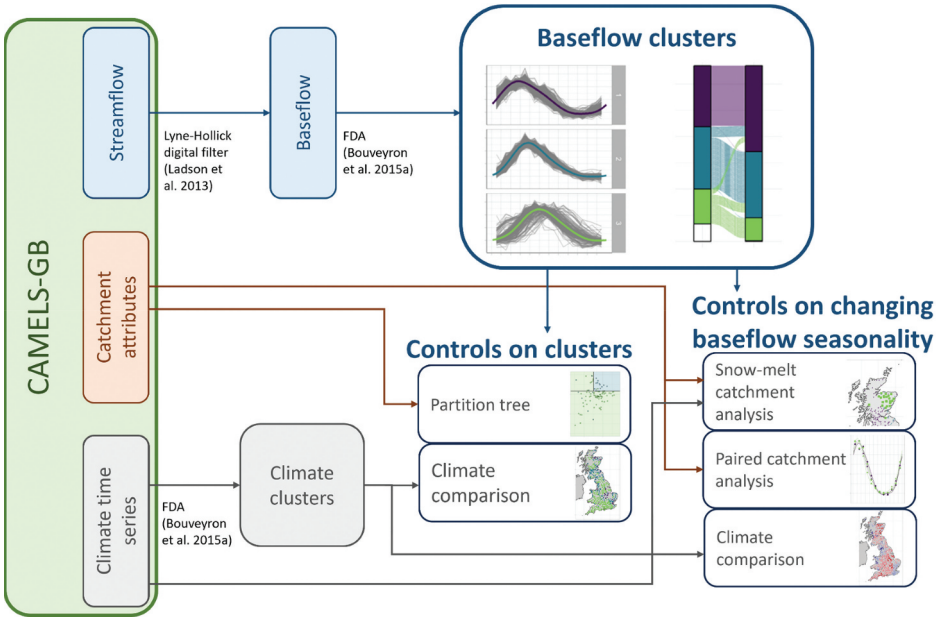

Figure 2. Graphical demonstration of the methods and data used within this study.

BFI is taken from CAMELS-GB (Coxon *et al.* 2020a) and is estimated from the ratio of mean daily baseflow to daily discharge, where hydrograph separation has been performed using the Ladson *et al.* (2013) digital filter.

To investigate changes over time in the hydro-climatic time series, the data are split into two 20-year time blocks: hydrological years (October–September) 1976–1995, and 1996–2015. This split was chosen to maximize the number of catchments and years within each time block and enable the comparison across time. The time series within each block and catchment are treated equally, so meaningful comparisons can be made over time and space. Peaks and troughs in the standardized functional baseflow curves are identified as days from the start of the hydrological year.

3 Methods

Figure 2 shows the data and method workflow used in the study. Following the preparation of the baseflow time series, FDA is used to identify functional clusters of seasonal baseflow curves. These are then explored to investigate spatial distributions of the clusters across GB and changes in cluster membership of individual catchments between the two time blocks. A range of graphical plots and other analytical methods, including partition tree analysis and paired catchments are used to quantify controls on cluster membership and changes in cluster membership between time blocks. The following section describes the methods used.

3.1 Functional data analysis

FDA allows smooth relationships between variables to be estimated statistically by fitting functions to curves (Ramsey and Silverman 2005). Our study considers seasonal baseflow patterns, which form a curve of length 1 year. Fitting a function to the seasonal baseflow curve allows us to characterize all of the seasonal distribution information simultaneously. These functions that represent the baseflow curves can then be compared across space and time to consider similarities and differences.

Functional data consist of representations of functions $\{X_1, \ldots, X_n\}$ observed on a domain $[t_1, t_2]$. In this work temporal domains are considered, but functional data can also be considered over other domains, such as spatial applications. The monthly data are considered to be observations from a curve on the domain $[0, 12)$ months. The functional approach allows discretely sampled data to be considered as observations from a process acting on a continuous domain. Unlike the previously cited work using FDA in hydrological contexts (Haggarty *et al.* 2015, Suhaila and Yusop 2016, Ternynck *et al.* 2017, Larabi *et al.* 2018, Alaya *et al.* 2020, Ghumman *et al.* 2020), we consider the seasonal shapes of baseflow, which are standardized to remove their mean and give unit variance for each curve. This allows the shapes defining the baseflow seasonality to be compared rather than the absolute values.

To identify similarities between the average annual baseflow curves at different locations and time blocks we apply the *funFEM* clustering method (Bouveyron 2015). The *funFEM* package runs the algorithm of the same name (Bouveyron 2015), which aims to find a discriminative functional subspace to separate the curves into different clusters. This involves starting with a matrix of coefficients containing a basis representation of each curve, then assuming a latent subspace exists such that the coefficients of the curves separate into distinct clusters. For full details see Bouveyron (2015). The clustering approach is different to testing methods such as Larabi *et al.* (2018) or Suhaila and Yusop (2016) as there is no explicit test of difference; instead, this method allows for characterization of the seasonal curves over space and time blocks.

In this application of *funFEM* to annual signals, the Fourier basis is chosen as it provides a periodic function space over the domain interval so the smoothed seasonal patterns can seamlessly repeat from year to year. The Fourier basis is constructed using a combination of sine and cosine functions defined over the year interval. Over the domain of one year, $[0, 12)$ months, seven basis functions are used as a balance between flexibility and complexity. The parameters of the latent subspace in the clustering algorithm are such that the mean and variance of each cluster can be different, with diagonal covariance matrices. Although information criteria such as the Akaike information criterion (AIC) or the Bayesian information criterion (BIC) could be applied to choose the number of clusters, it is noted by Bouveyron (2015) that these can be less efficient in real data scenarios than with simulated test data. Here, the number of clusters is set to three for the seasonal baseflow clustering. This number is chosen to allow for clear comparisons to be made between the shapes of baseflow seasonality and for comparison of these groupings spatially and across the time blocks.

3.2 Graphical and statistical methods

Following the functional clustering of the standardized seasonal hydrographs, we (i) explore and characterize the spatial and temporal distribution of the functional clusters, (ii) investigate controls on the functional clusters of seasonal baseflow, and (iii) investigate any changes in baseflow seasonality.

The spatial and temporal distribution of the functional clusters is characterized using colour-coded point maps. Changes (or not) in cluster membership between the time blocks are illustrated using graphical flow diagrams linking colour-coded cluster membership of a given catchment between time blocks.

To investigate controls on the functional clusters of seasonal baseflow, histograms of BFI associated with catchments in each functional cluster across the time blocks are investigated. In addition, the connections between key catchment attributes representing hydrogeological, topographical and soil features of catchments and the functional baseflow clusters are quantified using a partition tree analysis (Therneau and Atkinson 2022). Three representative catchment attributes from CAMELS-GB (Coxon *et al.* 2020a) were explored using a partition tree: the 90th percentile of elevation, volumetric porosity of soils, and the percentage of the catchment covered with high productivity fracture deposits. These attributes were chosen based

upon previously described relationships between topological, hydrogeological and soil attributes and BFI (Bloomfield *et al.* 2021) and were the attributes within each of those categories that had the highest magnitude of correlation with the cluster groupings, but correlation of magnitude under 0.4 across the three catchment attributes. The partition tree makes a series of splits of the input variables (catchment characteristics) in order to best predict the output variable (cluster), resulting in a rules-based label prediction. Default arguments for the *rpart* algorithm for classes were used (Therneau and Atkinson 2022), including only considering splits for nodes with more than 20 catchments, and having at least seven catchments in each node.

Two approaches have been used to investigate any changes in baseflow seasonality: a paired catchment approach and an analysis of changes in the timing of peak seasonal baseflow. Three pairs of catchments with different cluster allocations and changes in cluster allocation (one catchment in each pair moves to an earlier baseflow cluster and one does not) were selected to consider the controls on the annual shapes in baseflow and the changes over the time blocks. The three pairs of catchments are catchments 37008 and 37020, 33020 and 33012, and 45004 and 45005, and they are marked on Fig. 1(a). These were chosen based on several principles: (a) adjacent and nearby catchments were preferable; (b) each pair has rather similar catchment attributes in terms of climate (e.g. the differences of annual precipitation and PET should be within 10%), topography, land cover, geology, etc., yet with different cluster memberships over the time blocks; (c) catchment pairs with less human activity (no reservoir and urban percentage <10%) or under similar human influences are selected. The key catchment attributes of selected pairs and their cluster allocations over time blocks are presented in Supplementary material Table S6 and the respective pairs of standardized seasonal baseflow curves are shown in Supplementary material Fig. S7.

In a manner similar to Liebmann *et al.* (2012) and Dunning *et al.* (2018), the timing of seasonal baseflow peaks has also been identified, and maps are used to explore spatial distribution of changes in direction and magnitude of peak timing of seasonal curves for individual catchments between the time blocks.

4 Results

In the following sections, baseflow seasonality in GB is characterized by describing the spatio-temporal distribution of the functional seasonal baseflow clusters, and controls on those distributions are investigated by comparing functional seasonal baseflow cluster membership with BFI and catchment features. Controls on changes in cluster membership and baseflow seasonality are then explored.

4.1 Characterization of baseflow seasonality

Figure 3(a) shows the standardized median seasonal baseflow curves (in grey) grouped by cluster, with the cluster means overlaid (in bold colour). Figure 3(b) shows the spatial distribution of the catchments as a function of cluster memberships for each of the two time blocks. Figure 3(c) is a flow diagram showing how the cluster membership of individual catchments varies (or not) across the two time blocks, and Fig. 3(d) shows this information spatially. Table 1 gives the numbers of catchments assigned to each seasonal baseflow cluster and how they vary over time, and Table 2 contains the mean residual variance for each baseflow cluster and the timings of the peak and trough of the annual baseflow curves. Plots equivalent to Fig. 3 and Table 1 for precipitation, temperature and effective rainfall are given in the Supplementary material (Figs. S4 to S6 and Tables S3 to S5).

The three functional seasonal baseflow clusters have broadly similar annual shapes, as shown in Fig. 3(a). All are slightly asymmetrical with relatively sharper peaks in baseflow during late winter and broader, less well-defined troughs in summer. However, the timing of the peaks (and troughs) in seasonal baseflow varies between the three clusters. Baseflow cluster 1 peaks (troughs) earliest in December (July), with cluster 2 peaking (troughing) around a month later, and cluster 3 peaking later still, in February to March. In addition, there is a difference in the within-cluster variation in baseflow seasonality, with the smallest mean residual variance associated with cluster 2 and the greatest associated with cluster 3 (Table 2). The functional clustering, resulting in groups that we will describe as early-, mid- and late-season clusters, indicates that the main difference between the baseflow seasonality across catchments and time blocks is the timing of the peaks and troughs, rather than the seasonal shapes.

There are no spatial inputs to the clustering algorithm used to generate the clusters, the inputs are simply the average seasonal shapes for each location and time block. Consequently, any spatially coherent grouping of the clusters in Fig. 3(b) is a sign of similar seasonal behaviour in these catchments for a given time block. The maps of cluster membership (Fig. 3(b)) show consistent spatial relationships that persist across the two time blocks. Catchments in cluster 1 (associated with the earliest baseflow seasonality) are predominantly distributed throughout the west of GB across both time blocks, with only a few isolated, outlying catchments in the second time block (1996–2015) in this cluster found in southeast England. Catchments in cluster 2 are predominantly distributed along a band running from eastern Scotland down through central England to southwest England and a second, smaller spatially coherent region running from the easternmost area of England through southeast England to the southeast coast of England. Finally, cluster 3 catchments with the latest baseflow seasonality are predominantly situated in central, eastern and southern England (largely co-incident with the outcrop of the Chalk aquifer, Fig. 1(a)) with a small outlier of catchments distributed in the east Scottish Highlands. However, most of these latter catchments change to cluster 1 catchments by the second time block (Fig. 3(b)). As previously noted, cluster 3 has the greatest within-cluster variation in baseflow series (Fig. 3(a) and Table 2), which may result from the geographical diversity of regions in GB that contribute to this cluster (Fig. 3(b)) compared with clusters 1 and 2.

The majority of catchments have unchanged cluster membership between the two time blocks (Fig. 3(d)). Even though over time some catchments change cluster allocations, as shown by Table 1 and in the maps in

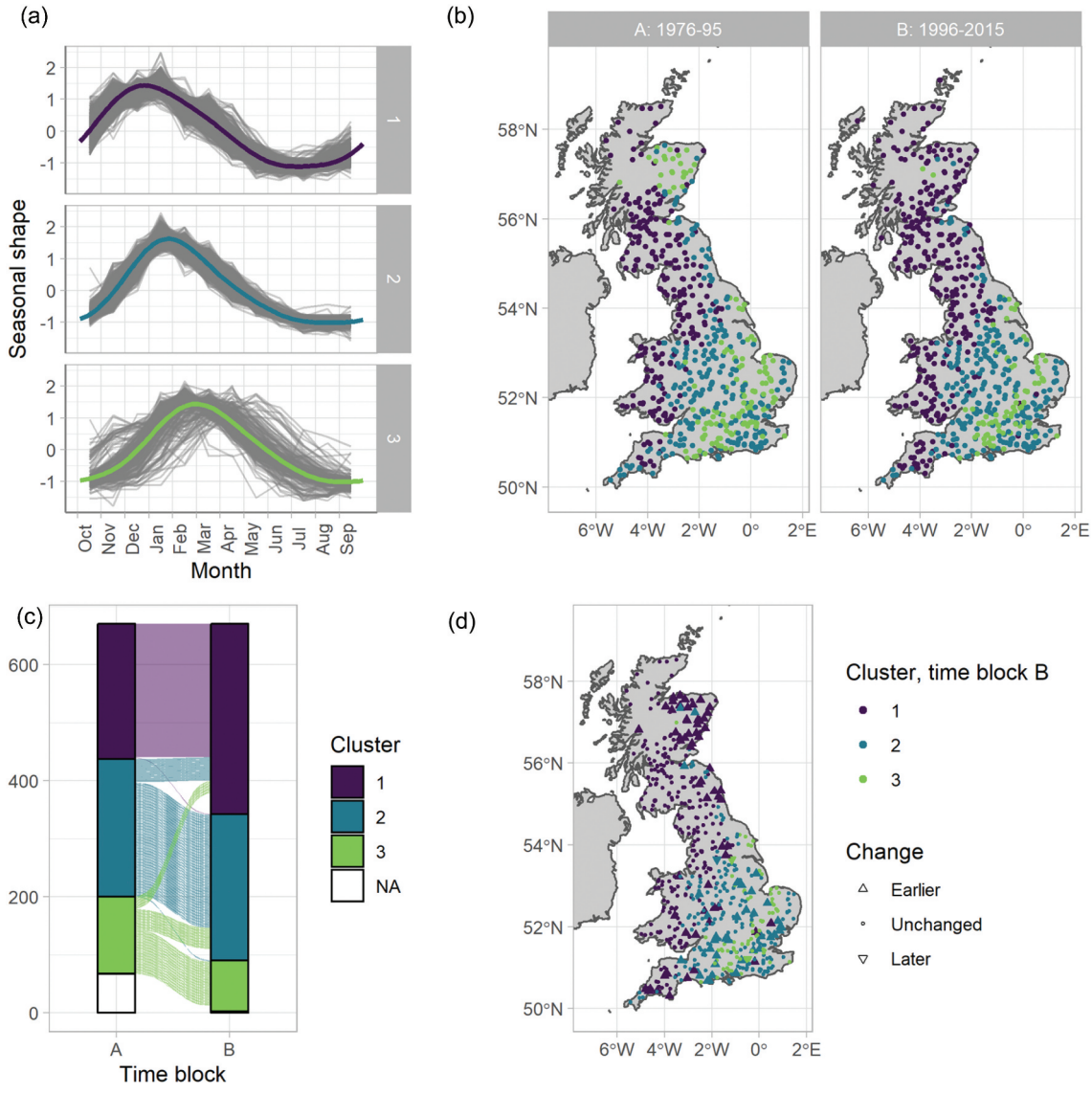

Figure 3. (a) Median annual baseflow for each location and time block plotted by cluster (grey lines), with the cluster means overlaid (in cluster colour and bold). (b) Cluster membership of each location within each time block. (c) Flow diagram of each location showing cluster membership over the time blocks. Each location is a thin line within the plot. (d) Map showing the cluster membership of each location in time block B (shown by colour), with triangular symbols denoting those locations in an earlier or later cluster compared with time block A.

Table 1. Number of locations assigned to each baseflow cluster within each time block.

Cluster	Block A	Block B
1	233	328
2	237	252
3	133	88

Table 2. Mean residual variance per baseflow cluster (variance is calculated per time series and averaged over each cluster), approximate timing of cluster peak and trough (presented as days through the hydrological year).

Cluster	Mean residual variance	Peak timing	Trough timing
1	0.083	85	283
2	0.050	116	319
3	0.182	152	338

Fig. 3(b,d), the overall spatial disposition of the seasonal baseflow clusters remains broadly similar between the two time blocks. The most noticeable changes over time are the increase in membership of cluster 1 (the cluster with the earliest peak in seasonal baseflow) from 233 catchments in 1976–1995 to 328 catchments in 1996–2015, and the decrease in membership of cluster 3 (the cluster with the latest peak in seasonal baseflow) from 133 catchments to 88 catchments over the same time (Table 1). The increase in membership of cluster 1 is due to movement from clusters 2 and 3, and some catchments that were not included in the first time block due to missing data. In addition, some catchments move from cluster 3 to cluster 2 between the time blocks. Of the catchments included in both time blocks, 97 catchments move to an earlier functional baseflow seasonality cluster in the last time block and 501 catchments do not change cluster. In comparison, only three move to a cluster with later baseflow seasonality.

4.1.1 Controls on baseflow seasonality

Given the strong spatial association between catchments with the latest baseflow seasonality (cluster 3, Fig. 3(b)) and the

outcrop of the unconfined Chalk aquifer with associated high BFI (Fig. 1(a,c)), the association between the timings of baseflow seasonality and the BFI of the catchments is considered here.

Figure 4 shows histograms of BFI for the catchments plotted by time block, with different colours used for the different seasonal baseflow clusters. There is a broad alignment between the baseflow clusters and BFI. In general, the catchments in clusters with earlier (later) baseflow seasonality have lower (higher) BFI. In the later time block there is a wider spread of BFI in cluster 2 and a narrower spread of BFI in cluster 3, corresponding to the movement of many of the cluster 3 catchments to cluster 2. Bloomfield *et al.* (2021) found that the BFI of a catchment is influenced by a combination of factors including drainage path slope, underlying aquifers, soil composition and human influences such as abstraction and land cover, so links between catchment factors and baseflow seasonality are considered next.

Here we demonstrate the connections between some of the same key catchment attributes representing hydrogeological, topographical and soil features and the clusters using a partition tree analysis (Therneau and Atkinson 2022). A partition tree involving seven splits correctly categorizes the first cluster allocation of 76% (403/603) of the catchments (see Supplementary material, Fig. S1). Early baseflow seasonality catchments (cluster 1) are associated with higher elevation (correlated with steeper slopes) (Supplementary material, Fig. S1(a)), high porosity and no high-productivity fracture deposits (Supplementary material, Fig. S1(b)). Cluster 3 is associated with low elevation and the presence of high-productivity deposits (Supplementary material, Fig. S1(c)), and cluster 2 is between these, with lower porosity than cluster 1 (Supplementary material, Fig. S1(a), (b)) and higher elevation than cluster 3 (Supplementary material, Fig. S1(c)). This demonstrates that the clusters are closely linked to the catchment characteristics that relate the supply of groundwater, the speed of water transition through the soil, and the height and slopes within the catchment. Further details are available in the Supplementary material (S1).

This connection to catchment characteristics is supported by the observation that when the FDA methodology is applied to precipitation (6-month smoothed precipitation), temperature, and effective rainfall time series for the same set of CAMELS-GB catchments (see Supplementary material, Figs. S4 to S6), there is no clear spatial correlation between the resulting clusters for the seasonality of the climatological variables and the equivalent baseflow seasonality clusters (Fig. 3(b)). For example, the annual patterns in precipitation (Supplementary material, Fig. S4) appear to be moving earlier between the two time blocks, but spatially the eastern locations exhibit earlier peaks in baseflow. There are barely discernible differences between the three clusters of temperature seasonality (Supplementary material, Fig. S5), indicating a process with shared annual distribution across this study area. For effective rainfall (Supplementary material, Fig. S6), in the first time block relatively late effective rainfall seasonality dominates across much of GB (with the exception of the northwestern and northern GB that shows relatively early seasonality), but at the second time block both the early- and late-season effective rainfall cluster membership switches to the mid-seasonality cluster. Given that the majority of catchments across GB exhibited late seasonality in the first time block, the effect is that the majority of sites across GB move to earlier effective rainfall seasonality (with the exception of those in northwestern and northern GB that move to a slightly later seasonality). Based on these observations, the first-order control on baseflow seasonality is inferred to be catchment

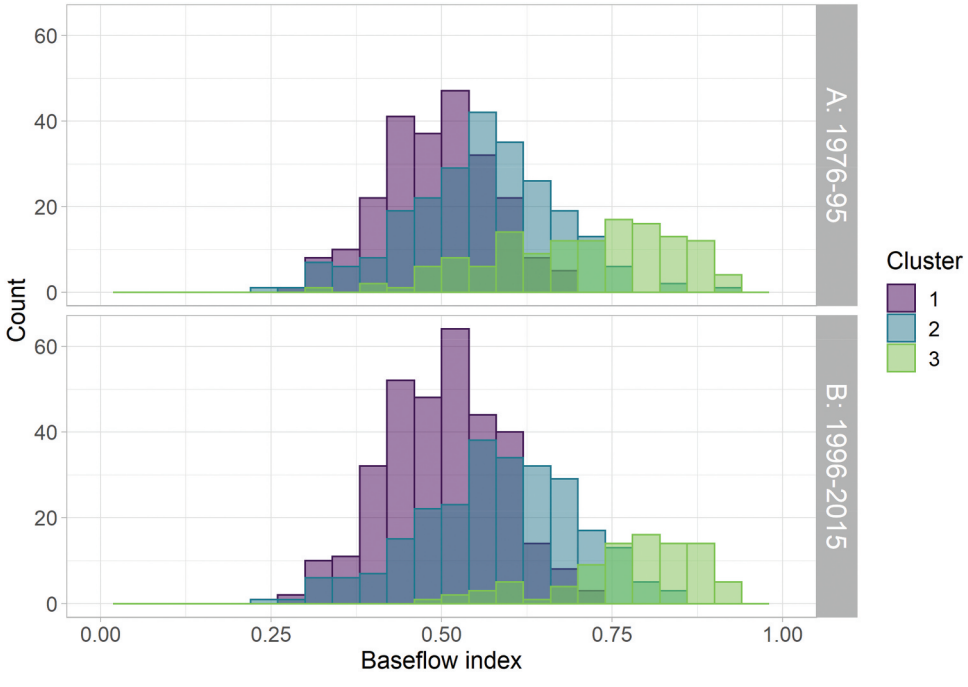

Figure 4. Histogram of baseflow index (BFI) for the cluster allocations in each time block (sub-plots). Histogram bars for different clusters are overlaid (not stacked).

characteristics (combinations of geological, hydrological, topographical and landcover features that are in large part reflected by the BFI) rather than the driving climatology.

4.2 Controls on changes in baseflow seasonality

Although the study area is predominantly temperate in character, there are a few catchments in the north of GB, in the mountains of eastern Scotland, where snow accumulates during winter. Given previously documented links between the effects of warming on snowmelt-influenced flow regimes (Barnett *et al.* 2005, 2008, Leppi *et al.* 2011, Kormos *et al.* 2016, Rumsey *et al.* 2020, Tan *et al.* 2020, Ayers *et al.* 2021), changes in baseflow seasonality in snow-influenced GB catchments are first explored. Then we explore the evidence for potential controls on changes in baseflow seasonality more generally across the study area.

4.2.1 Changes in baseflow seasonality associated with snowmelt-influenced catchments

One geographically distinct group of catchments that start in cluster 3 (late season baseflow) in the earlier time block (Fig. 5(a)) are 24 catchments in Scotland. The BFI of these catchments is shown in Fig. 5(b). These catchments have BFI values typically in the range 0.4 to 0.7, rather than the higher BFI associated with the majority of the groundwater-dominated catchments in cluster 3 located in the southeast of England primarily on the Chalk (BFI typically >0.9). These Scottish catchments are mostly in locations with relatively high elevation (mean elevation greater than the 80th percentile of GB catchments, 315 m, for all but four of these catchments) and have a correspondingly high fraction of precipitation falling as snow (greater than or equal to the 80th percentile for GB catchments).

Of the 24 Scottish catchments, 20 move to cluster 1 (the earliest seasonal baseflow), three move to cluster 2 and one stays in cluster 3 in the later time block, and it is inferred that the catchments that move from cluster 3 to earlier clusters are associated with earlier snowmelt associated with long-term warming in these high catchments. Figure S2 in the Supplementary material shows that warming January temperatures for the selected catchments over the two time blocks are consistent with the snow melting at an earlier time in the year: Fig. S2(a) shows an increase in median January temperature and Fig. S2(b) shows an increase in the proportion of days with temperature over 0°C between the two time blocks. The only catchment to stay in cluster 3 (with the latest timed peak) in the later time block is the catchment with the lowest temperature, highest elevation and highest proportion (17%) of precipitation as snowfall. The movement from cluster 3 to cluster 1 for many of the selected catchments indicates a large change in the timing of the peak in annual baseflow (median change ~87 days), consistent with higher temperatures contributing to earlier melting of the snowpack. This observation is consistent with similar observations from North America (Barnett *et al.* 2005, 2008, Leppi *et al.* 2011, Kormos *et al.* 2016) and with the findings of Pohle *et al.* (2019) who have shown that, over the period of analysis of this study, increased air temperature due to climate warming has led to earlier snowmelt across the Scottish Highlands.

4.2.2 Other controls on changes in baseflow seasonality

In addition to the Scottish catchments described above, 73 catchments show a change to earlier baseflow seasonality between the time blocks (Fig. 3(d)) and, unlike the

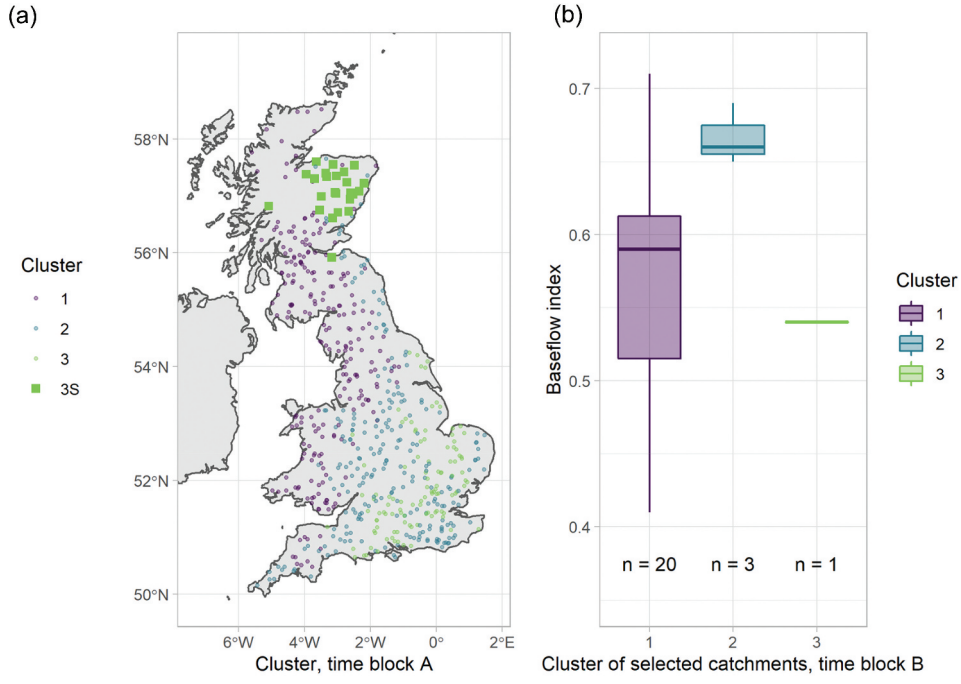

Figure 5. (a) Map denoting the selected Scottish catchments that start in cluster 3 in time block A (marked as cluster 3S). (b) Box plot of baseflow index (BFI) of the selected catchments, split according to their cluster allocation in time block B.

Scottish catchments, these catchments that move to earlier baseflow seasonality have a wide spatial distribution across large parts of GB. Two approaches have been used to investigate what may be controlling this systematic change to earlier baseflow seasonality in these catchments across GB: a paired catchment approach and an analysis of changes in the timing of peak seasonal baseflow.

Given the first-order control of BFI and catchment characteristics on baseflow seasonality (Section 4.1.1), we first used three sets of paired catchments with different cluster allocations and contrasting changes (or not) in cluster allocation to investigate whether there is evidence for catchment controls on the move to earlier baseflow seasonality (see Supplementary material Table S6 and Fig. S7 for details of the paired catchments and results). From the paired catchment study, there was no evidence of systematic control of catchment characteristics on cluster allocations and changes in cluster membership. Instead, we found that all catchments exhibit an earlier seasonality at the second time block regardless of whether a site changes cluster membership with time when details of the monthly baseflow curves are considered. Thus increased temporal granularity (beyond early-, mid-, and late-season functional clusters) in the seasonal analysis may be useful in exploring controls on changes in baseflow seasonality.

As the paired catchment analysis has indicated that changes in baseflow seasonality without a corresponding change in cluster allocation may be affected by factors other than catchment characteristics and to provide increased temporal granularity to the analysis, the timing of the peaks in the baseflow curves is identified for each location and time block and compared with comparable plots for precipitation, temperature, and effective rainfall to investigate whether changes in the seasonality of climate variables may be influencing changes in baseflow seasonality. This is shown in Fig. 6 where the direction of change in the seasonal peak timing and magnitude for each of the variables is mapped across GB.

The majority of the symbols have a red outline in Fig. 6(d), indicating that most catchments across GB are exhibiting an earlier peak in baseflow in the second time block compared to the first, by up to a month. Conversely, there are a few catchments, indicated by light blue triangles mostly situated in western, northwestern and northern GB, along with a few catchments near London that exhibit later baseflow peaks (by up to a month) in the second time block compared with the first. Most of these catchments with later peaks in baseflow in the second time block are in an area predominantly allocated to cluster 1 in both time blocks. This could indicate that there is a limit to how early the peaks of the baseflow curves can be, as curves that are already early do not see a shift to an earlier peak in annual baseflow. Note that the Scottish catchments that have been inferred to be affected by snowmelt processes (Section 3.2) show peak seasonal baseflow more than 1 month earlier in the second time block compared with the first time block (Fig. 6(d)).

The predominant change in precipitation is to earlier peaks (Fig. 6(a)), except for locations in western Scotland where there is evidence for some later peaks in precipitation seasonality. This tendency to earlier peaks in precipitation is strongest (indicated by dark red filled triangles) in the east of GB. There is very little change shown in the timing of peaks in temperature seasonality (Fig. 6(b)): either the peaks are slightly later or there is no change. The map of change in effective rainfall (Fig. 6(c)) shows regions of later seasonal peaks in the western, northwestern and northern GB and in the far southeast of GB, with a band of earlier peaks in lowland southern, central and eastern England and up to the east of Scotland. There is a strong similarity in the overall patterns of changes in peak timings of the effective rainfall and baseflow seasonal curves in Fig. 6(c,d) (with the exception of a greater tendency for a change to later effective rainfall in the far southeast of England compared with generally earlier baseflow seasonality). Lin's concordance coefficient for the respective changes in peak timing of the baseflow compared with the effective rainfall is 0.40, whereas comparing the changes in peak timing of baseflow to precipitation and temperature yields coefficients of 0.00 and −0.01.

In summary, unlike Tan *et al.* (2020), we have found no evidence for an association between changes in the seasonality of baseflow with changes in the seasonality of either precipitation or temperature. However, from Fig. 6(c,d), and given Lin's concordance coefficients, there is some evidence that changes in the seasonality of effective rainfall may be associated with changes in baseflow seasonality across the study area. Changes in the seasonality of PET are driven at least in part by long-term warming across the UK (Kay *et al.* 2013, Watts *et al.* 2015). This association is consistent with the vegetation phenology-mediated changes in PET proposed by Chen *et al.* (2022) under climate warming.

5 Discussion and conclusion

The baseflow separation method used in this work is the digital filtering approach (Ladson *et al.* 2013), which suits the task of performing separation for over 600 series using the same methodology. There are many different baseflow separation methods available, with different merits. As our work starts with the daily separated baseflow, then averages over the month, and finally averages over the years within the time block, the differences in baseflow separation results should not have a great influence on the results of the analysis. This is demonstrated in the Supplementary material Fig. S3 and Table S2, which include the clustering results calculated using the 5-day minima method of baseflow separation (Gustard *et al.* 1992). Those results show the same patterns as those detailed within this article.

Within this study, the data were split by year into two time blocks. This allowed for the comparison of the seasonal patterns over time. Other choices of split could be applied to these data, such as shorter or overlapping time blocks. As the time blocks used are relatively long, and the median was used to summarize the values within the time blocks, the results should not be affected by occasional extreme years.

One of the key challenges we found when conducting this study was identifying clear links between changes in baseflow

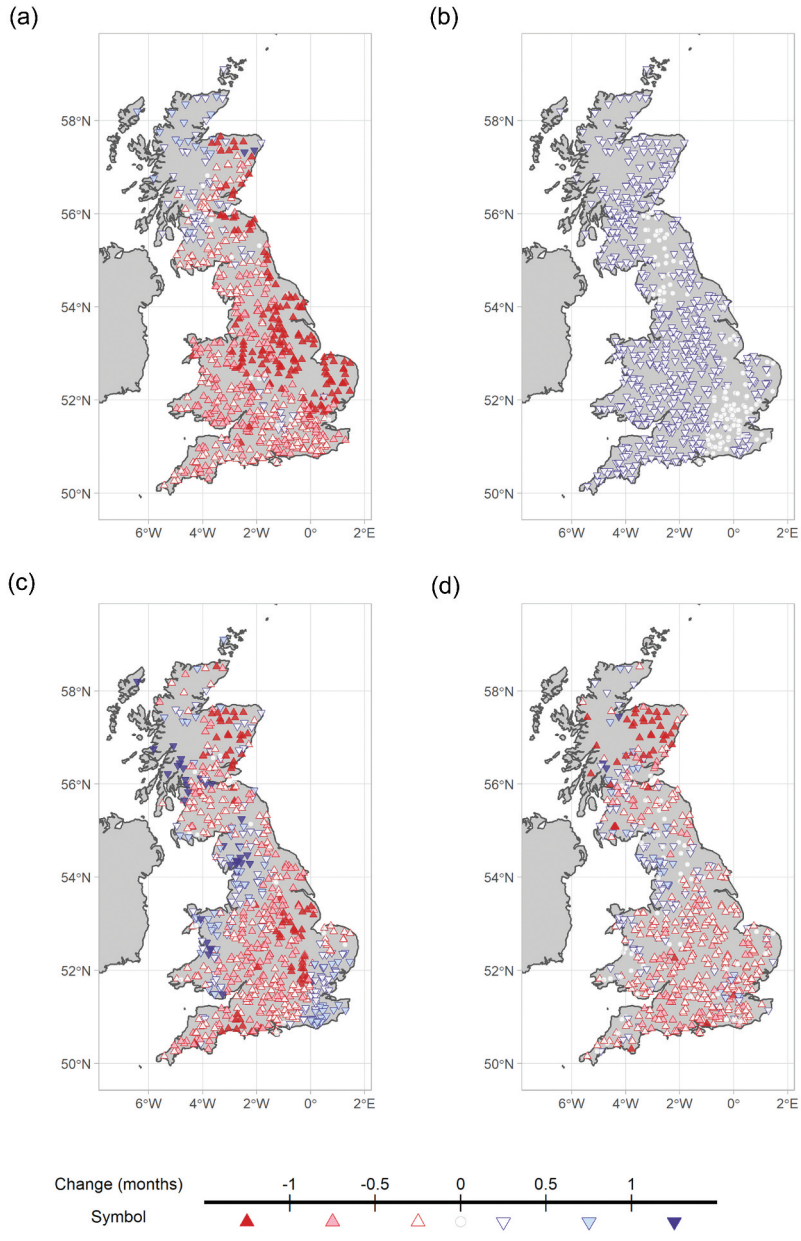

Figure 6. Changes in peak timing between the two time blocks: (a) six-month smoothed precipitation, (b) temperature, (c) effective rainfall (PPT − PET), (d) baseflow. Red-outlined upward-pointing triangles indicate locations with earlier peaks, white circles indicate no change, and blue-outlined downward-pointing triangles indicate later seasonal peaks. The shapes are filled with colour according to the magnitude of change (in months).

seasonality and climatic, geophysical and water management catchment attributes. Catchment attributes included in large-sample datasets such as the CAMELS-GB dataset (Coxon *et al.* 2020a) typically represent a snapshot (i.e. a specific year) or an average over time, rather than changes or trends. This hinders their application in large-sample studies when identifying controls on changes in time. Changes to water management schemes and other changes to catchments over the analysis period that increased the responsiveness of catchments to precipitation, such as changes in land cover (including, for example, increased urban coverage) or stream conveyance, may also have contributed to the switch to earlier seasonal peaking of baseflows and should also be the focus for future research. This study highlights the need for future large-sample datasets such as the CAMELS family of datasets (Addor *et al.* 2017, Chagas *et al.* 2020, Coxon *et al.* 2020a), the related LamaH-CE large-sample dataset for Central Europe (Klingler *et al.* 2021), and the Caravan meta-dataset (Kratzert *et al.* 2023).

This application of FDA to seasonal data has shown differences over time in annual baseflow (shifting to earlier patterns) that might not have been identified using other seasonal approaches such as trend identification following seasonal averaging. The clustering part of the methodology allows for discrete categorization of the time series, and catchments with large changes in seasonal baseflow can be identified as those changing cluster. However, the discrete nature of the clustering analysis can mask smaller temporal changes that are present in the functional representation of the annual series.

The FDA methodology is well suited to annual patterns as the fitted curves can be defined as cyclic so that the end of the year continues seamlessly into the start of the next. Here the data are averaged over months before applying the functional method; however, the daily data could also be used, as in Ternynck *et al.* (2017). The monthly averaging provides smoother starting curves for analysis. The months were also considered to be equally spaced through the year, which is an approximation. However, whilst the ability to compare the seasonal distributions of the data is a key factor of this work, standardizing the curves also limits the information available to the distribution over the year and not the absolute values. This is related to the application of the FDA method to standardized shapes in this work and is not a feature of FDA approaches in general. Means and different variances can be included in FDA but would have distracted from the seasonal patterns in this work. It is also noted that for this application, the cluster means are of similar shapes but different timing so the clusters are described as earlier and later versions of the annual pattern, but this will not always be the case for the resulting clusters. In other applications where the curves (seasonality or other series) are of different shapes, the categorization of clusters would focus on shape rather than timing.

Our application of FDA has shown that membership of functional clusters of similar seasonal baseflow curves is strongly linked to catchment characteristics. Here we have shown that elevation and geological and soil attributes can be used within a partition tree to correctly classify 76% of the catchments into clusters in the first time block. However, change in climate, and specifically warming, is the first-order effect on changes in baseflow seasonality. For snow-influenced catchments in GB, there has been a shift towards earlier baseflow seasonality inferred to be due to earlier snowmelt associated with global warming. For other catchments, there is a geographical association between a shift to earlier seasonal effective rainfall and earlier seasonal baseflow.

This FDA approach to identifying patterns of seasonal distribution of hydrological variables is not specific to the GB catchment data in CAMELS-GB. This method could also be applied to other hydrological or climate time series to identify similarities and changes over space and time. A natural extension to this work would be application to further CAMELS (Addor *et al.* 2017) and related datasets in other countries and regions. It has been shown that the FDA approach described here is a potentially powerful data-driven analytical tool to identify and quantify hydrological changes as a precursor to designing subsequent research and models to address specific process-based questions, such as elements of the causal cascade of process from changing phenology, through changing evapotranspiration and associated effective rainfall to changes in recharge, flow and discharge within catchments.

Acknowledgements

KAL and JPB publish with the permission of the Executive Director of the British Geological Survey, part of UK Research and Innovation (UKRI).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

KAL and JPB were funded on this study by the NERC Climate change in the Arctic-North Atlantic region and impact on the UK (CANARI) project (NE/W004984/1). GC and YZ were supported by a UKRI Future Leaders Fellowship award (MR/V022857/1).

Data availability statement

All figures contain/are derived from CAMELS-GB data (Coxon *et al.* 2020a). The CAMELS-GB dataset contains data supplied by the UK Centre for Ecology & Hydrology © UK Centre for Ecology & Hydrology; © University of Bristol; contains Environment Agency information © Environment Agency and/or database right; derived from UK bedrock hydrogeological mapping and UK superficial productivity mapping, BGS © UKRI. https://doi.org/10.5285/8344e4f3-d2ea-44f5-8afa-86d2987543a9.

References

Addor, N., *et al.*, 2017. The CAMELS data set: catchment attributes and meteorology for large-sample studies. *Hydrology and Earth System Sciences*, 21 (10), 5293–5313. doi:10.5194/hess-21-5293-2017.

Ahiablame, L., *et al.*, 2017. Annual baseflow variations as influenced by climate variability and agricultural land use change in the Missouri River Basin. *Journal of Hydrology*, 551, 188–202. doi:10.1016/j.jhydrol.2017.05.055.

Alaya, M.A.B., *et al.*, 2020. Change point detection of flood events using a functional data framework. *Advances in Water Resources*, 137, 103522. doi:10.1016/j.advwatres.2020.103522.

Allen, D.J., *et al.*, 1997. *The physical properties of major aquifers in England and Wales*. Nottingham UK: British Geological Survey.

Ayers, J.R., *et al.*, 2021. On the statistical attribution of changes in monthly baseflow across the U. S. Midwest. *Journal of Hydrology*, 592, 125551. doi:10.1016/j.jhydrol.2020.125551.

Barnett, T.P., Adam, J.C., and Lettenmaier, D.P., 2005. Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature*, 438 (7066), 303–309. doi:10.1038/nature04141.

Barnett, T., *et al.*, 2008. Human-induced changes in the hydrology of the Western United States. *Science*, 319 (5866), 1080–1083. doi:10.1126/science.1152538.

Bloomfield, J.P., *et al.*, 2021. How is Baseflow Index (BFI) impacted by water resource management practices? *Hydrology and Earth System Sciences*, 25 (10), 5355–5379. doi:10.5194/hess-25-5355-2021.

Bloomfield, J.P., Allen, D.J., and Griffiths, K.J., 2009. Examining geological controls on baseflow index (BFI) using regression analysis: an illustration from the Thames Basin, UK. *Journal of Hydrology*, 373, 164–176. doi:10.1016/j.jhydrol.2009.04.025

Bloomfield, J.P., Bricker, S.H., and Newell, A.J., 2011. Some relationships between lithology, basin form and hydrology: a case study from the Thames basin, UK. *Hydrological Processes*, 25 (16), 2518–2530. doi:10.1002/hyp.8024.

Bloomfield, J.P. and Marchant, B.P., 2013. Analysis of groundwater drought building on the standardised precipitation index approach. *Hydrology and Earth System Sciences*, 17 (12), 4769–4787. doi:10.5194/hess-17-4769-2013.

Bosch, D.D., *et al.*, 2017. Temporal variations in baseflow for the Little River experimental watershed in South Georgia, USA. *Journal of Hydrology: Regional Studies*, 10, 110–121.

Boulton, A.J., 2003. Parallels and contrasts in the effects of drought on stream macroinvertebrate assemblages. *Freshwater Biology*, 48 (7), 1173–1185. doi:10.1046/j.1365-2427.2003.01084.x.

Bouveyron, C., 2015. funFEM: clustering in the discriminative functional subspace, R package version 1.1.

Bouveyron, C., Côme, E., and Jacques, J., 2015. The discriminative functional mixture model for a comparative analysis of bike sharing systems. *The Annals of Applied Statistics*, 9 (4). doi:10.1214/15-AOAS861.

Bricker, S.H. and Bloomfield, J.P., 2014. Controls on the basin-scale distribution of hydraulic conductivity of superficial deposits: a case study from the Thames Basin, UK. *Quarterly Journal of Engineering Geology and Hydrogeology*, 47 (3), 223–236. doi:10.1144/qjegh2013-072.

Burt, T.P. and Ferranti, E.J.S., 2011. Changing patterns of heavy rainfall in upland areas: a case study from northern England. *International Journal of Climatology*, 32 (4), 518–532. doi:10.1002/joc.2287.

Chagas, V.B., *et al.*, 2020. CAMELS-BR: hydrometeorological time series and landscape attributes for 897 catchments in Brazil. *Earth System Science Data*, 12 (3), 2075–2096. doi:10.5194/essd-12-2075-2020.

Chen, S., *et al.*, 2022. Vegetation phenology and its ecohydrological implications from individual to global scales. *Geography and Sustainability*, 3 (4), 334–338. doi:10.1016/j.geosus.2022.10.002.

Coxon, G., *et al.*, 2020a. Catchment attributes and hydro-meteorological timeseries for 671 catchments across Great Britain (CAMELS-GB). NERC Environmental Information Data Centre.

Coxon, G., *et al.*, 2020b. CAMELS-GB: hydrometeorological time series and landscape attributes for 671 catchments in Great Britain. *Earth System Science Data*, 12 (4), 2459–2483. doi:10.5194/essd-12-2459-2020.

Dunning, C.M., Black, E., and Allan, R.P., 2018. Later wet seasons with more intense rainfall over Africa under future climate change. *Journal of Climate*, 31 (23), 9719–9738. doi:10.1175/JCLI-D-18-0102.1.

Ficklin, D.L., Robeson, S.M., and Knouft, J.H., 2016. Impacts of recent climate change on trends in baseflow and stormflow in United States watersheds. *Geophysical Research Letters*, 43, 5079–5088. doi:10.1002/2016GL069121

Geng, X., *et al.*, 2020. Extended growing season reduced river runoff in Luanhe River basin. *Journal of Hydrology*, 582, 124538. doi:10.1016/j.jhydrol.2019.124538.

Ghumman, A.R., *et al.*, 2020. Functional data analysis of models for predicting temperature and precipitation under climate change scenarios. *Journal of Water and Climate Change*, 11 (4), 1748–1765. doi:10.2166/wcc.2019.172.

Gnann, S.J., Woods, R.A., and Howden, N.J.K., 2019. Is there a baseflow Budyko curve? *Water Resources Research*, 55 (4), 2838–2855. doi:10.1029/2018WR024464.

Gomez-Velez, J.D., *et al.*, 2015. Denitrification in the Mississippi River network controlled by flow through river bedforms. *Nature Geoscience*, 8 (12), 941–945. doi:10.1038/ngeo2567.

Gustard, A., Bullock, A., and Dixon, J.M., 1992. *Low flow estimation in the United Kingdom*. Wallingford, UK: Institute of Hydrology.

Haggarty, R.A., Miller, C.A., and Scott, E.M., 2015. Spatially weighted functional clustering of river network data. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 64 (3), 491–506. doi:10.1111/rssc.12082.

Hare, D.K., *et al.*, 2021. Continental-scale analysis of shallow and deep groundwater contributions to streams. *Nature Communications*, 12 (1). doi:10.1038/s41467-021-21651-0.

Hirsch, R.M. and Slack, J.R., 1984. A nonparametric trend test for seasonal data with serial dependence. *Water Resources Research*, 20 (6), 727–732. doi:10.1029/WR020i006p00727.

Jenkins, G.J., Perry, M.C., and Prior, M.J., 2008. *The climate of the United Kingdom and recent trends*. Exeter, UK: Met Office Hadley Centre.

Jones, M.R., *et al.*, 2012. An assessment of changes in seasonal and annual extreme rainfall in the UK between 1961 and 2009. *International Journal of Climatology*, 33 (5), 1178–1194. doi:10.1002/joc.3503.

Jordan, T.E., Correll, D.L., and Weller, D.E., 1997. Relating nutrient discharges from watersheds to land use and streamflow variability. *Water Resources Research*, 33 (11), 2579–2590. doi:10.1029/97WR02005.

Kay, A.L., *et al.*, 2013. A hydrological perspective on evaporation: historical trends and future projections in Britain. *Journal of Water and Climate Change*, 4 (3), 193–208. doi:10.2166/wcc.2013.014.

Keller, V.D.J., *et al.*, 2015. CEH-GEAR: 1 km resolution daily and monthly areal rainfall estimates for the UK for hydrological and other applications. *Earth System Science Data*, 7 (1), 143–155. doi:10.5194/essd-7-143-2015.

Kim, J.H., *et al.*, 2018. Warming-induced earlier greenup leads to reduced stream discharge in a temperate mixed forest catchment. *Journal of Geophysical Research: Biogeosciences*, 123 (6), 1960–1975. doi:10.1029/2018JG004438.

Klingler, C., Schulz, K., and Herrnegger, M., 2021. LamaH-CE: large-sample data for hydrology and environmental sciences for Central Europe. *Earth System Science Data*, 13 (9), 4529–4565. doi:10.5194/essd-13-4529-2021.

Kormos, P.R., *et al.*, 2016. Trends and sensitivities of low streamflow extremes to discharge timing and magnitude in Pacific Northwest mountain streams. *Water Resources Research*, 52 (7), 4990–5007. doi:10.1002/2015WR018125.

Kratzert, F., *et al.*, 2023. Caravan - a global community dataset for large-sample hydrology. *Scientific Data*, 10 (1). doi:10.1038/s41597-023-01975-w.

Ladson, A.R., *et al.*, 2013. A standard approach to baseflow separation using the Lyne and Hollick filter. *Australasian Journal of Water Resources*, 17 (1), 25–34. doi:10.7158/13241583.2013.11465417.

Larabi, S., St-Hilaire, A., and Chebana, F., 2018. A new concept to calibrate and evaluate a hydrological model based on functional data analysis. *Journal of Water Management Modeling*. doi:10.14796/JWMM.C442.

Leppi, J.C., *et al.*, 2011. Impacts of climate change on August stream discharge in the Central-Rocky Mountains. *Climatic Change*, 112 (3–4), 997–1014. doi:10.1007/s10584-011-0235-1.

Liebmann, B., *et al.*, 2012. Seasonality of African precipitation from 1996 to 2009. *Journal of Climate*, 25 (12), 4304–4322. doi:10.1175/JCLI-D-11-00157.1.

Marchant, B.P. and Bloomfield, J.P., 2018. Spatio-temporal modelling of the status of groundwater droughts. *Journal of Hydrology*, 564, 397–413. doi:10.1016/j.jhydrol.2018.07.009

Miller, M., *et al.*, 2016. The importance of base flow in sustaining surface water flow in the Upper Colorado River Basin. *Water Resources Research*, 52 (5), 3547–3562. doi:10.1002/2015WR017963.

Miller, O.L., *et al.*, 2021. How will baseflow respond to climate change in the Upper Colorado River Basin? *Geophysical Research Letters*, 48 (22). doi:10.1029/2021GL095085.

Mo, C., *et al.*, 2021. Impact of climate change and human activities on the baseflow in a typical karst basin, Southwest China. *Ecological Indicators*, 126, 107628. doi:10.1016/j.ecolind.2021.107628.

Mohammed, R. and Scholz, M., 2016. Impact of climate variability and streamflow alteration on groundwater contribution to the base flow of the Lower Zab River (Iran and Iraq). *Environmental Earth Sciences*, 75 (21). doi:10.1007/s12665-016-6205-1.

Piao, S., *et al.*, 2019. Plant phenology and global climate change: current progresses and challenges. *Global Change Biology*, 25 (6), 1922–1940. doi:10.1111/gcb.14619.

Poff, N.L., *et al.*, 1997. The natural flow regime. *BioScience*, 47 (11), 769–784. doi:10.2307/1313099.

Pohle, I., *et al.*, 2019. Citizen science evidence from the past century shows that Scottish rivers are warming. *Science of the Total Environment*, 659, 53–65. doi:10.1016/j.scitotenv.2018.12.325.

Price, K., 2011. Effects of watershed topography, soils, land use, and climate on baseflow hydrology in humid regions: a review. *Progress in Physical Geography: Earth and Environment*, 35 (4), 465–492. doi:10.1177/0309133311402714.

Ramsey, J.O. and Silverman, B.W., 2005. *Functional data analysis*. 2nd ed. New York: Springer.

Robinson, E.L., *et al.*, 2017. *Climate hydrology and ecology research support system meteorology dataset for Great Britain (1961–2015) [CHESS-met] v1.2*. NERC Environmental Information Data Centre.

Robinson, E.L., *et al.*, 2020. *Climate hydrology and ecology research support system potential evapotranspiration dataset for Great Britain (1961–2017) [CHESS-PE]*. NERC Environmental Information Data Centre.

Rumsey, C.A., Miller, M.P., and Sexstone, G.A., 2020. Relating hydroclimatic change to streamflow, baseflow, and hydrologic partitioning in the Upper Rio Grande Basin, 1980 to 2015. *Journal of Hydrology*, 584, 124715. doi:10.1016/j.jhydrol.2020.124715.

Singh, S.K., *et al.*, 2019. Towards baseflow index characterisation at national scale in New Zealand. *Journal of Hydrology*, 568, 646–657. doi:10.1016/j.jhydrol.2018.11.025.

Smakhtin, V.U., 2001. Low flow hydrology: a review. *Journal of Hydrology*, 240 (3–4), 147–186.

Suhaila, J. and Yusop, Z., 2016. Spatial and temporal variabilities of rainfall data using functional data analysis. *Theoretical and Applied Climatology*, 129 (1–2), 229–242. doi:10.1007/s00704-016-1778-x.

Tallaksen, L.M., 1995. A review of baseflow recession analysis. *Journal of Hydrology*, 165 (1–4), 349–370. doi:10.1016/0022-1694(94)02540-R.

Tan, X., Liu, B., and Tan, X., 2020. Global changes in baseflow under the impacts of changing climate and vegetation. *Water Resources Research*, 56 (9). doi:10.1029/2020WR027349.

Ternynck, C., *et al.*, 2017. Streamflow hydrograph classification using functional data analysis. *Journal of Hydrometeorology*, 17 (1), 327–344. doi:10.1175/JHM-D-14-0200.1.

Therneau, T. and Atkinson, B. 2022. *Rpart: recursive partitioning and regression trees, R package version 4.1.16*.

Wang, D. and Cai, X., 2010. Comparative study of climate and human impacts on seasonal baseflow in urban and agricultural watersheds. *Hydrology and Land Surface Studies*, 37 (6), L06406.

Watts, G., *et al.*, 2015. Climate change and water in the UK – past changes and future prospects. *Progress in Physical Geography: Earth and Environment*, 39 (1), 6–28. doi:10.1177/0309133314542957.

Zhang, M., *et al.*, 2017. A global review on hydrological responses to forest change across multiple spatial scales: importance of scale, climate, forest type and hydrological regime. *Journal of Hydrology*, 546, 44–59. doi:10.1016/j.jhydrol.2016.12.040.