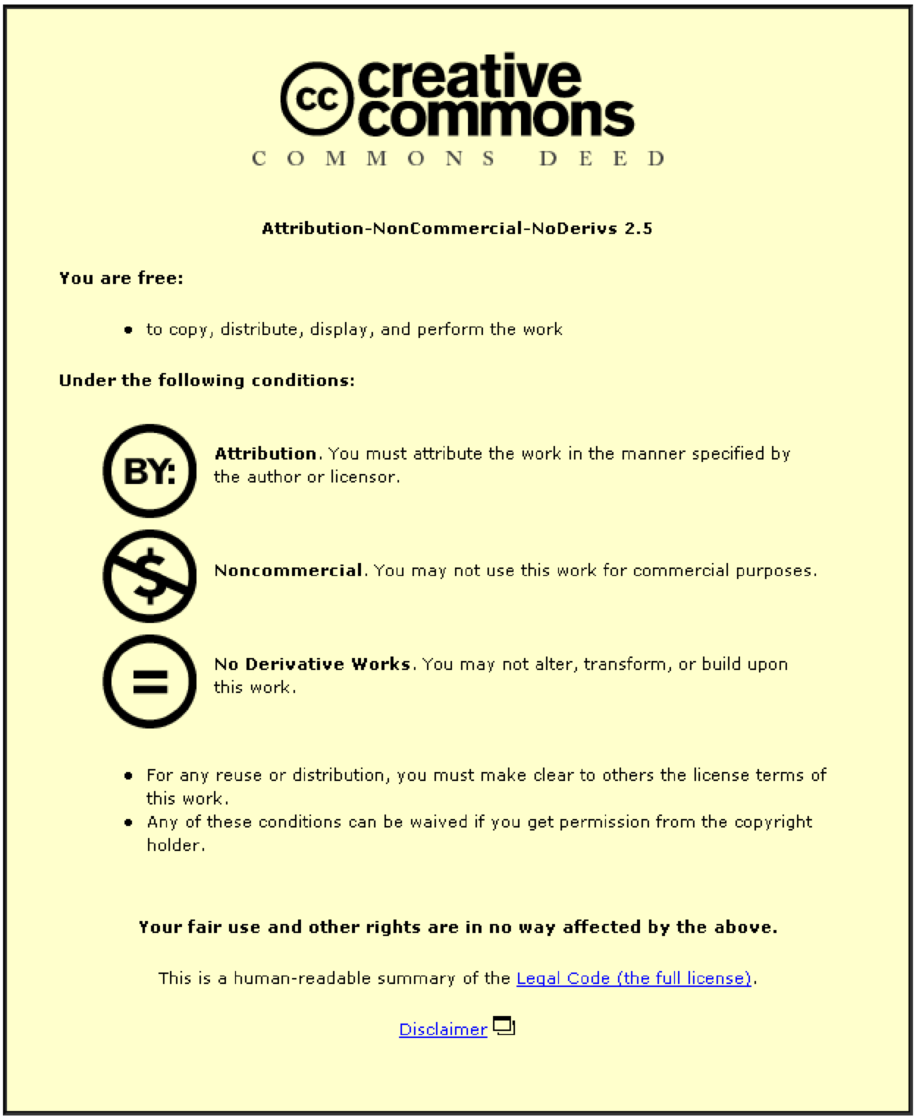


SHORT COMMUNICATION

Selection of river flow indices for the assessment of hydroecological change

Wendy A. Monk[1], Paul J. Wood[2], David M. Hannah[3] and Douglas A. Wilson[4]

1. Canadian Rivers Institute, Department of Biology, Bag Service #45111, University of New Brunswick, Fredericton, New Brunswick, E3B 6E1, Canada.
2. Department of Geography, Loughborough University, Loughborough, Leicestershire, LE11 3TU, UK.
3. School of Geography, Earth and Environmental Sciences, University of Birmingham. Edgbaston, Birmingham, B15 2TT, UK.
4. Water Resources, The Environment Agency of England and Wales, Kings Meadow House, Kings Meadow Road, Reading, RG1 8DQ, UK.

Author for Correspondence:

Wendy Monk, Canadian Rivers Institute, Department of Biology, Bag Service #45111, University of New Brunswick, Fredericton, New Brunswick, E3B 6E1, Canada.
Phone: +1-506-453-4845
Fax: +1-506-453-3583
Email: w.monk@unb.ca

Key Words: river flow regimes, Principal Components Analysis, Lotic-invertebrate Index for Flow Evaluation, UK, data redundancy.

Running Title: Selecting flow indices for assessment of hydroecological change

Abstract

A wide range of ‘ecologically relevant’ hydrological indices (variables) have been identified as potential drivers of riverine communities. Recently, concerns have been expressed regarding index redundancy (i.e. similar patterns of variance) across the host of hydrological descriptors on offer to researchers and water resource managers. Some guiding principles are required to aid selection of the most statistically defensible and meaningful river flow indices for hydroecological analysis. In this short communication, we investigate the utility of a principal components analysis (PCA)-based method that identifies 25 hydrological variables to characterise the major modes of statistical variation in 201 hydrological indices for 83 rivers across England and Wales. The emergent variables, and all 201 hydrological variables, are used to develop regression models [for the whole data set and three river flow regime shape (i.e. annual hydrograph form) classes] for an 11-year macroinvertebrate community dataset (i.e. LIFE scores). The same ‘best’ models are produced using the PCA-based method and all 201 hydrological variables for two of the three river flow regime groups. However, weaker models are yielded by the PCA-based method for the remaining (flashy) river flow regime class and the whole data set (all 83 rivers). Thus, it is important to exercise caution when employing data reduction/ index redundancy approaches, as they may reject variables of ecological significance due to the assumption that the statistically dominant sources of hydrological variability are the principal drivers of, perhaps more subtle (sensitive), hydroecological associations.

Introduction

The ecological importance of river flow regime variability is increasingly well recognised (e.g., Clausen and Biggs, 1997; Wood and Armitage, 2004); and a wide range of potentially ‘ecologically relevant’ hydrological indices have been identified (e.g. Olden and Poff, 2003). However, such hydroecological analysis is limited by a general lack of paired long-term hydrological and ecological time-series (Wood *et al.*, 2001; Jackson and Füreder, 2006). The search for ‘ecologically relevant’ hydrological indices has been driven by the need to quantify variability in ecological communities and/or individual populations that may be sensitive to natural hydrological changes or anthropogenic modifications (Richter *et al.*, 1996). Some concerns have been raised regarding the large number of potential hydrological predictors available, since significant redundancy (multicollinearity) exists between many variables (Olden and Poff, 2003). Consequently, some guiding principles are required to aid researchers and water resource managers select the most ‘ecologically relevant’ hydrological variable(s).

Olden and Poff (2003) proposed a method using principal components analysis (PCA) for assessing redundancy between hydrological variables and identifying those indices which account for most variation in river flow regimes using long-term flow records for 420 locations across the continental USA. They suggested that the variables identified by this method may form the basis of future hydroecological analysis. However, to date, their redundancy methodology and the resulting variables have not been widely tested in terms of ecological prediction.

The aim of this short communication is to provide the first test of the PCA-based approach proposed by Olden and Poff (2003) in association with ecological data, and to compare its effectiveness against regression models developed using 201 potentially ‘ecologically relevant’ hydrological variables identified in previous research.

Data and methods

Hydrological and ecological data were employed for 83 sites in England and Wales (Figure 1). Prior to analysis, screening of raw data ensured a benchmark period of 20-years (1980 – 1999) of hydrological data and 11-years (1989 – 2000) ecological data. Hydrological indices were calculated from daily mean flows. Ecological data consisted of autumn (September – November) family-level macroinvertebrate data for each site collected using the semi-quantitative 3-minute kick sample method (Murray-Bligh, 1999). For each sample the LIFE (Lotic-invertebrate Index for Flow Evaluation; Extence *et al.*, 1999) score was derived and input as the dependent variable in subsequent analysis. The LIFE method has been developed by the Environment Agency of England and Wales to assess macroinvertebrate community response to ‘flow’ based upon known species- and family-level preferences for particular mean flow velocity conditions. The LIFE methodology is now routinely used by the Environment Agency to identify sites subject to ecological stresses associated with natural flow variability (e.g. floods or drought) and/or anthropogenic impacts (e.g. water abstractions).

A total of 201 hydrological indices used in 15 previously published articles were used in our analysis (Hughes and James, 1989; Poff and Ward, 1989; Richards, 1989; Biggs, 1990; Jowett and Duncan, 1990; Poff, 1996; Richter *et al.*, 1996; Clausen and Biggs, 1997; Richter *et al.*, 1997; Puckridge *et al.*, 1998; Richter *et al.*, 1998; Clausen and Biggs, 2000; Clausen *et al.*, 2000; Wood *et al.*, 2000; Wood *et al.*, 2001). For brevity, Appendix 1 lists those variables identified by the PCA method and/or utilised in regression models in this paper (for full details of all candidate variables see Monk *et al.*, 2006). These hydrological indices are grouped into five categories, as first proposed by Richter *et al.* (1996) and, subsequently, expanded by Poff *et al.* (1997) and Olden and Poff (2003). These categories include: (1) magnitude of flow events (n = 147); (2) duration of flow events (n = 31); (3) timing of flow events (n = 8); (4) frequency of flow events (n = 7); and (5) rate of change of flow conditions (n = 8).

An annual hydrograph classification technique was employed (devised Hannah *et al.*, 2000; adapted by Harris *et al.*, 2000; evaluated by Bower *at al.*, 2004) to group rivers with similar flow regime seasonality (i.e. timing of low/high flow periods). The method utilises hierarchical agglomerative cluster analysis (Ward's method) to classify annual flow regimes. This regime 'shape' (RS) classification identifies stations with a similar form of annual hydrograph, regardless of magnitude. Monthly averages across all record years for each basin (expressed in runoff mm month^{-1} to standardise for differences in basin area) were used as input data for the classification. Hence, this approach performs a similar function to the 'hydrogeographical' classification of Poff (1996) and so provides a more objective starting point for analysis of differences between river types.

Principal components analysis facilitated the examination of data structure and dominant modes of intercorrelation amongst hydrological indices. PCA was used to identify those variables that accounted for the major sources of variation within the dataset, thus minimising redundancy. PCA was undertaken using hydrological data for all rivers and individually for each regime shape group (above). The 25 flow indices with the highest loadings on the first 4 PC axes were selected for each regime shape class and for the whole data set (all 83 rivers), following the procedures outlined in Olden and Poff (2003). The number of hydrological indices selected for each axis was weighted by the proportion of the variance explained by that PC relative to all PCs retained (e.g., based on all streams the first PC explained 41.6% of the total 71.53% variance explained by the first 4 PCs - resulting in 15 of 25 indices being selected from PC1). The variables identified by the PCA-based method were used as independent variables in the development of stepwise multiple linear regression models to predict LIFE scores. In addition, all 201 variables, including those initially rejected as a result of the PCA, were used to build regression models for comparison with the results yielded for the 25 PCA-selected variables.

Results

Three distinct regime shape classes were identified, which grouped basins with similar patterns of annual runoff timing, which have a clear geographical expression (Figure 1 and Figure 2). Regime shape A (RS_A) exhibited multiple high flow periods with a dominant peak in December and secondary peaks in October and March. All RS_A sites were located on impermeable geologies and concentrated in the wetter northwest of England and Wales. Regime shape B (RS_B) sites were characterised by a single peak in January, with relatively steep rising and falling limbs. RS_B sites were located throughout north-eastern, central and southern England and across a range of geologies. Regime shape C (RS_C) sites were characterised by a prolonged rising limb to a March peak and were mainly located in eastern and southern England associated with major groundwater aquifers.

PCA indicated up to eight significant PCs for some of the shape classes and across all sites. The percentage of variance explained by axes 1 – 4 varied between 71.52% for RS_B up to 73.00% for RS_C; and 73.88% for all sites (Table 1). A total of 42 variables were identified across the three regime shape classes (from a total of 201 candidate variables) using the PCA method, with 13 variables common to all regime shape classes (Table 2). Detailed examination indicated the majority of the 42 variables identified were from the category representing the magnitude of average flow conditions (MA – 24 variables in Table 2) followed by magnitude of low flow conditions (ML – 6 variables); low flow duration (DL – 5 variables), high flow duration (DH – 4 variables), magnitude of high flow conditions (MH – 2 variables), and frequency of high flow events (FH – 1 variable). A similar pattern was observed when all 83 sites were considered: magnitude of average flow conditions (MA – 15 variables); magnitude of low flow conditions (ML – 2 variables); low flow duration (DL – 2 variables); high flow duration (DH – 3 variables); magnitude of high flow conditions (MH – 1 variable); and frequency of high flow events (FH – 2 variables).

Stepwise multiple regression models generated for the LIFE scores using hydrological variables (predictors) identified by PCA and for all 201 candidate variables were identical for two regime shape classes (RS_B and RS_C; Table 3). The PCA-based method produced a weaker model than when all variables were used for regime shape A (RS_A: DAY30MAX for PCA model cf. QFEB for 201 candidate variable model), and when all 83 sites were considered together (DFMEDMAX for PCA model cf. SMED for 201 candidate variable model; Table 3). Only one hydrological variable was incorporated into any of the regression models. The specific median flow (SMED; for definition of variables see Appendix 1) was identified as the 'best' variable for two regime shape classes (RS_B and RS_C) using both methods, and for all 83 sites when all 201 indices were offered as candidates.

Discussion

The results of this study indicate the methodology proposed by Olden and Poff (2003) was effective in identifying hydrological variables that may influence instream ecology for two of the three river flow regime types in England and Wales. Six different 'hydrogeographical' stream types were identified for the continental USA (Poff, 1996; Olden and Poff,, 2003), which included two intermittent and two snowmelt driven stream types. In contrast to the research of Olden and Poff (2003), all of the sites used in this study had perennial flow and none have a significant snowmelt contribution. The regime shape classification identified herein reflects known regional climatic and geological differences across England and Wales (Bower *et al.*, 2004), the temperate-maritime climate of the region and the small number of upland sites within the dataset.

The PCA methodology identified 42 variables, from a total of 201 candidate variables. Most of these were from the ‘magnitude of flow events’ group (147 candidate variables) (Richter *et al.,* 1996; Olden and Poff, 2003) and specifically the sub-group representing the ‘magnitude of average conditions’ (MA – 24 of 42 variables for the three regime shape classes, 15 of 25 for all sites). This sub-group contains the largest number of candidate indices (92 variables) and includes a diversity of hydrological measures including monthly and annual mean values, as well as indices derived from specific points (percentiles) on the flow duration curve (high and low). Therefore, it was not unexpected that MA indices describe the dominant modes of variability in the hydrological series for perennial temperate rivers in England and Wales.

Olden and Poff’s (2003) PCA-based approach implicitly assumes that the hydrological variables identified following redundancy analyses are the dominant influence on instream ecology. This study indicates that the LIFE score for two regime shape classes can be modelled by the same variable (specific median discharge – SMED) using both methodologies. This clearly demonstrates that a small number of variables can describe/ model instream community response to flow regime variability. The model for one regime shape class (RS_A) and the model for all sites were weaker for the PCA-based approach. As a result, careful consideration of the candidate hydrological indices is required since the ecological response may not simply reflect the dominant modes of statistical variation but more subtle changes in the flow regime. In addition, many of the variables used in this and other published studies only differ subtly (e.g., based on monthly or daily mean values) and there is a need for consistency in both the way indices are derived and the names they are given to avoid confusion.

Even when all 201 hydrologic variables were used as candidate variables, only one hydrological index was included within any of the resultant regression models. This suggests the presence of a limited number of key drivers of hydroecological variability. The selection and derivation of hydrological indices is time consuming and not always simple due to the large number of candidate variables and inevitable redundancy that exists between many. However, variables that reflect the range of hydrograph characteristics, as proposed within the 'Indicators of Hydrologic Alteration' (IHA) methodology (Richter *et al.*, 1996), are clearly appropriate.

In this investigation, the ecological data took the form of family-level macroinvertebrate community data recorded in abundance classes. The LIFE methodology has been developed based on known species and family preferences for particular mean flow velocities (Extence *et al.*, 1999); and the response of the LIFE score to regime variability has been examined in association with other macroinvertebrate community metrics (Monk *et al.*, 2006). However, utilising family-level data to derive LIFE scores is not without problems since some families, such as the mayfly Baetidae, include taxa with variable flow requirements. The effect that differences in taxonomic resolution have on the LIFE score and the resultant models is not currently quantified and further research should consider this, and the use of individual taxa and other organisms (e.g., fish, periphyton and macrophytes).

The specific median discharge (SMED – which incorporates median flow and basin area) was found to be the 'best' descriptor of the macroinvertebrate community for the two largest regime shape classes (RS_B = 52 rivers; RS_C = 20 rivers) and for all sites when all 201 hydrological variables were considered. This suggests that the size/ area of the river basin may be a particularly important scaling factor that strongly influences the ecological response. However, this variable has only been used in one previous investigation (Biggs, 1990), where it was found to be a good discriminator between the taxonomic composition of periphyton communities and periphyton biomass. The relatively weak models produced for regime shape C (RS_C) were surprising given that all of the rivers receive a significant groundwater contribution and, as a result, have very stable flow regimes, similar to *"superstable or stable groundwater"* (Olden and Poff, 2003, p.103). Previous research on groundwater-dominated rivers in England has indicated that the ecology responds strongly to changes in flow regime associated with floods and droughts (Wood and Armitage, 2004; Wright *et al.*, 2004). However, these studies were confined to single catchments and, at a broader scale, it may be necessary to consider other hydrological indices for these rivers, such as groundwater level or residence time of the water within the aquifer, to accurately model these rivers using this approach.

This study demonstrates that the PCA-based method proposed by Olden and Poff (2003) is effective for two of the three river regime shape types identified for England and Wales. However, it is important to exercise caution when employing data reduction/ redundancy approaches, as they may reject variables of ecological significance due to the assumption that the statistically dominant sources of hydrological variability are the principal drivers of perhaps more subtle (sensitive) hydroecological associations. Hence, future research should, where practicable, employ a refined number of clearly defined hydrological indices based on the IHA methodology (Richter *et al.*, 1996), where known duplication of hydrological information has been removed/ minimised using hydrological understanding rather than relying upon statistical approaches. This should ensure that the full range of the hydrological regime variability is considered and, thus, maximise the potential for modelling instream community response.

Acknowledgements

WAM acknowledges the support of a Loughborough University development fund studentship and an Environment Agency CASE award. The Environment Agency of England and Wales are thanked for provision of the LIFE paired dataset. The views expressed in this paper are those of the authors and not necessarily those of the Environment Agency.

References

Biggs BJF. 1990. Periphyton communities and their environments in New Zealand rivers. *New Zealand Journal of Marine and Freshwater Research* **24**: 367 – 386.

Bower D, Hannah DM and McGregor GR. 2004. Techniques for assessing the climatic sensitivity of river flow regimes. *Hydrological Processes* **18**: 2515 – 2543.

Clausen B and Biggs BJF. 1997. Relationships between benthic biota and hydrological indices in New Zealand streams. *Freshwater Biology* **38**: 327 – 342.

Clausen B and Biggs BJF 2000. Flow variables for ecological studies in temperate streams: groupings based on covariance. *Journal of Hydrology* **237**: 184 – 197.

Clausen B, Iversen HL and Ovesen NB. 2000. Ecological flow variables for Danish streams. In *XXI Nordic Hydrology Conference, Nordic Association for Hydrology. June 26-30, 2000. NHP Report 46*, Nilsson T (ed.). Uppsala, Sweden: 3 – 10.

Extence CA, Balbi DM and Chadd RP. 1999. River flow indexing using British benthic macroinvertebrates: A framework for setting hydroecological objectives. *Regulated Rivers: Research and Management* **15**: 543 – 574.

Hannah DM, Smith BPG, Gurnell AM and McGregor GR. 2000. An approach to hydrograph classification. *Hydrological Processes* **14**: 317 – 338.

Harris NM, Gurnell AM, Hannah, DM and Petts GE. 2000. Classification of river regimes: a context for hydroecology. *Hydrological Processes* **14**: 2831 – 2848.

Hughes JMR and James B. 1989. A hydrological regionalization of streams in Victoria, Australia with implications for stream ecology. *Australian Journal of Marine and Freshwater Research* **40**: 303 – 326

Jackson JK and Füreder L. 2006. Long-term studies of freshwater macroinvertebrates: a review of the frequency, duration and ecological significance. *Freshwater Biology* **51**: 591-603.

Jowett IG and Duncan MJ. 1990. Flow variability in New Zealand rivers and its relationship to in-stream habitat and biota. *New Zealand Journal of Marine and Freshwater Research* **24**: 305 – 317.

Monk WA, Wood PJ, Hannah DM, Wilson DA, Extence CA and Chadd RP. 2006. Flow variability and macroinvetebrate community response within riverine systems. *River Research and Application* **22**: ***-***.

Murray-Bligh J. 1999. *Procedures for Collecting and Analysing Macroinvertebrate Samples – BT0001*. The Environment Agency, Bristol; 176.

Olden JD and Poff NL. 2003. Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. *River Research and Applications* **19**: 101 – 121.

Poff NL. 1996. A hydrogeography of unregulated streams in the United States and an examination of scale-dependence in some hydrological descriptors. *Freshwater Biology* **36**: 71 – 91.

Poff NL, Allan JD, Bain MB, Karr JR, Prestegaard KL, Richter BD, Sparks RE and Stromberg JC. 1997. The natural flow regime: A paradigm for river conservation and restoration. *BioScience* **47**: 769-784.

Poff NL and Ward JV. 1989. Implications for streamflow variability and predictability for lotic community structure: a regional analysis of streamflow patterns. *Canadian Journal of Fisheries and Aquatic Sciences* **46**: 1805 – 1817.

Puckridge JT, Sheldon F, Walker KF and Boulton AJ. 1998. Flow variability and the ecology of large rivers. *Marine and Freshwater Research* **49**: 55 – 72.

Richards RP. 1989. Measures of flow variability for Great Lakes tributaries. *Environmental Monitoring and Assessment* **12**: 361 – 377.

Richter BD, Baumgartner JV, Powell J and Braun DP. 1996. A method for assessing hydrologic alteration within ecosystems. *Conservation Biology* **10**: 1163 – 1174.

Richter BD, Baumgartner JV, Wigington R and Braun DP. 1997. How much water does a river need? *Freshwater Biology* **37**: 231 – 249.

Richter BD, Baumgartner JV, Braun DP and Powell J. 1998. A spatial assessment of hydrologic alteration within a river network. *Regulated Rivers: Research and Management* **14**: 329 – 340.

Wood PJ, Agnew MD and Petts GE. 2000. Flow variations and macroinvertebrate community response in a small groundwater-dominated stream. *Hydrological Processes* **14**: 3133 – 3148.

Wood PJ and Armitage PD. 2004. The response of the macroinvertebrate community to low-flow variability and supra-seasonal drought within a groundwater dominated stream. *Archiv fur Hydrobiologie* **161**: 1 – 20.

Wood PJ, Hannah DM, Agnew MD and Petts GE. 2001. Scales of hydroecological variability within a groundwater-dominated stream. *Regulated Rivers: Research and Management* **17**: 347 – 367.

Wright JF, Clarke RT, Gunn RJM, Kneebone NT and Davy-Bowker J. 2004. Impact of major changes in flow regime on the macroinvertebrate assemblages of four chalk stream sites, 1997 – 2001. *River Research and Applications* **20**: 775 – 794.

List of Tables

Table 1. Summary of the percentage variance explained on axes 1-4 by principal component analysis (PCA) of the correlation matrix of 201 hydrological indices for the three river regime shapes (RS_{A-C}) and all 83 sites (All sites).

Table 2. Hydrological indices (in descending order) exhibiting the greatest loadings on the first 4 principal components. MA = magnitude of average flow conditions, ML = magnitude of low flow conditions, MH = magnitude of high flow conditions, DL = low flow duration, DH = high flow duration and FH = frequency of high flow events. See Appendix I for definition of variables.

Table 3. Stepwise multiple linear regression models for the LIFE score using hydrological variables identified by the principal components analysis method (PCA) and for all 201 hydrological variables (RAW) for (A) the three regime shape classes and (B) all sites. See Appendix I for definition of variables.

List of Figures

Figure 1. Location of the 83 river sites across England and Wales and their classification into river flow regime shape (RS) groups.

Figure 2. Standardised long-term (1980-1999) annual river flow regimes for three shape (RS) classes.

Table 1

	Principal component (% variance explained)				Total
	I	II	III	IV	
(A) RUNOFF SHAPE CLASSES					
RS_A	41.46	15.40	7.81	6.86	71.53
RS_B	43.06	16.10	6.63	5.72	71.52
RS_C	40.35	19.26	8.08	5.31	73.00
(B) ALL SITES	43.82	19.21	5.96	4.89	73.88

Table 2

		RS_A		RS_B		RS_C		ALL SITES
PC I	*MA*	MADQ	*MA*	TOTALVOL	*MA*	TOTALVOL	*MA*	TOTALVOL
	MA	TOTALVOL	*MA*	MDF	*MA*	MDF	*MA*	MDF
	MA	MDF	*MA*	MADQ	*MA*	Q50DF	*MA*	MADQ
	MA	Q10DF	*MA*	Q25DF	*MA*	MADQ	*DH*	DAY90MAX
	DH	DAY90MAX	*MA*	Q20DF	*MH*	DFMEDMAX	*MA*	Q10DF
	MA	Q20DF	*MA*	Q50DF	*ML*	MMID	*MA*	Q20DF
	MA	Q25DF	*ML*	MMID	*ML*	DFMEDMIN	*MA*	Q5DF
	MA	Q5DF	*DH*	DAY90MAX	*MA*	Q75DF	*MA*	Q1
	ML	MMID	*MA*	Q10DF	*MA*	Q80DF	*MA*	Q25DF
	MA	Q1	*MA*	Q1	*MA*	Q25DF	*DH*	DAY30MAX
	MA	Q50DF	*MA*	Q75DF	*MA*	MINAPR	*MA*	MMAD
	DH	DAY30MAX	*ML*	DFMEDMIN	*MA*	MMAD	*MH*	DFMEDMAX
	ML	DFMEDMIN	*MA*	Q5DF	*MA*	Q20DF	*MA*	Q1DF
	MA	Q75DF	*MA*	Q80DF	*DL*	DAY90MIN	*DH*	DAY7MAX
	MA	Q80DF	*DH*	DAY30MAX			*MA*	STDEVDF
PC II	*MA*	Q1090DF	*DL*	DFQ95MEAN	*DL*	Q95MEAN	*DL*	DFQ95MEAN
	DL	Q95MEAN	*ML*	BASEFLOW	*DL*	DFQ95MEAN	*ML*	BASEFLOW
	MA	Q2080DF	*MA*	Q1090DF	*ML*	BFI	*ML*	DFBFI
	DL	DFQ95MEAN	*MA*	Q2080DF	*MA*	Q1090DF	*MA*	Q1090DF
	MA	Q2575DF	*ML*	DFBFI	*MA*	CVDF	*DL*	Q95MEAN
			DL	Q95MEAN	*MA*	S80	*MA*	CVANNQ
PC III	*DH*	D3MAX50	*MA*	SMED	*MA*	SMED	*FH*	FRE1
	MH	AMAXDF	*FH*	FRE1	*MA*	MAR	*FH*	FRE1YR
	DH	D7MAX50			*MA*	SMIN		
PC IV	*DL*	D30MIN50	*MA*	Q5Q50	*ML*	AMINDF	*MA*	SK2
	MA	Q95DF50	*MA*	Q10Q50	*DL*	D3MIN50	*MA*	Q5Q50

Table 3

Model	Adjusted R^2	*F*	Number of rivers (samples)	Predictor variables plus sign of relationship
(A) RUNOFF SHAPE CLASSES				
$RS_{A\ PCA}$	0.240	22.750 ***	11 (71)	– DAY30MAX
$RS_{A\ RAW}$	0.300	30.544***	11 (71)	– QFEB
$RS_{B\ PCA}$	0.410	333.020 ***	52 (478)	+ SMED
$RS_{B\ RAW}$	0.410	333.020 ***	52 (478)	+ SMED
$RS_{C\ PCA}$	0.104	20.568 ***	20 (170)	+ SMED
$RS_{C\ RAW}$	0.104	20.568 ***	20 (170)	+ SMED
(B) ALL SITES				
PCA	0.111	90.423 ***	83 (719)	+ DFMEDMAX
RAW	0.381	442.622***	83(719)	+SMED

*** $p<0.001$

Figure 1

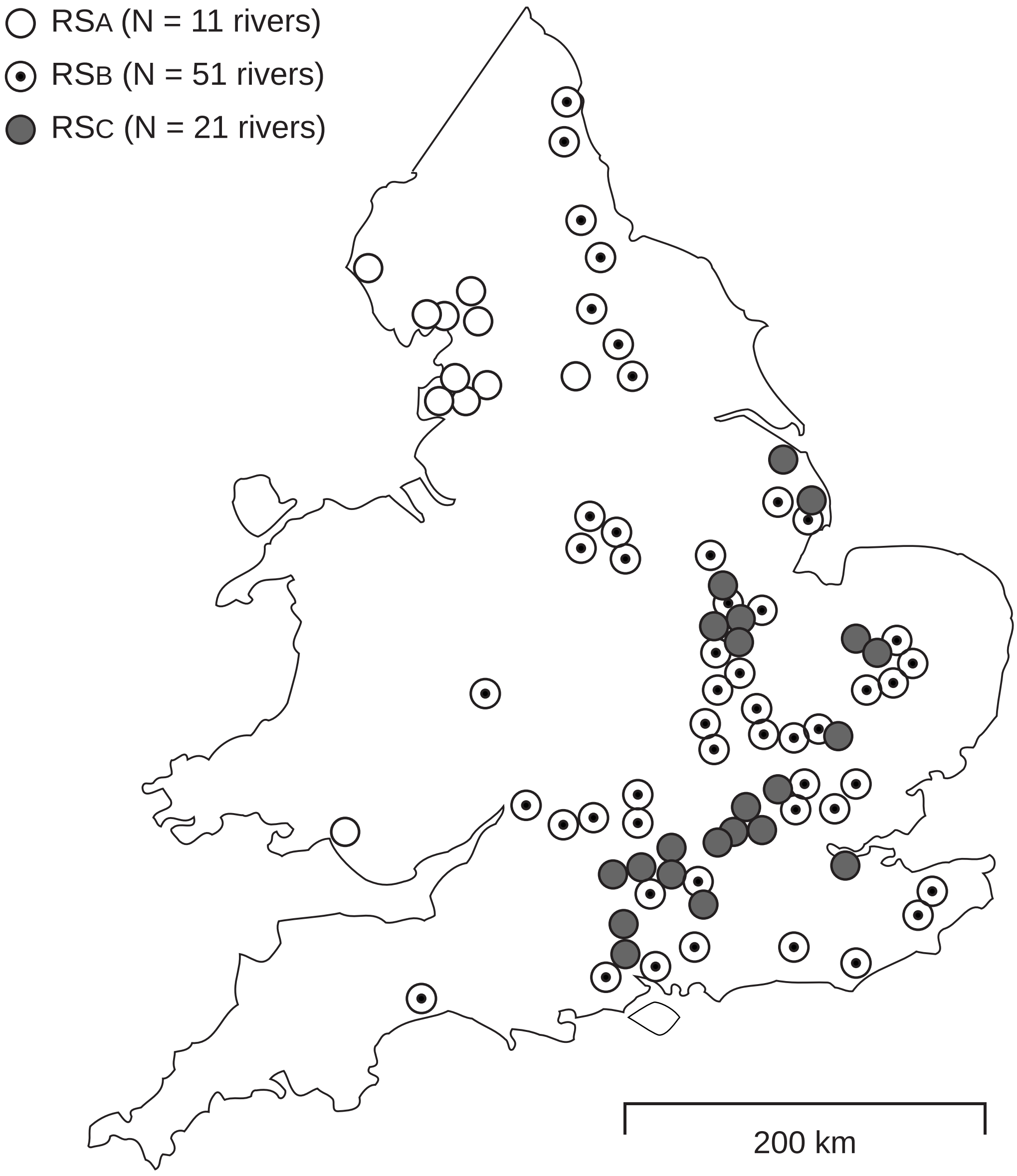
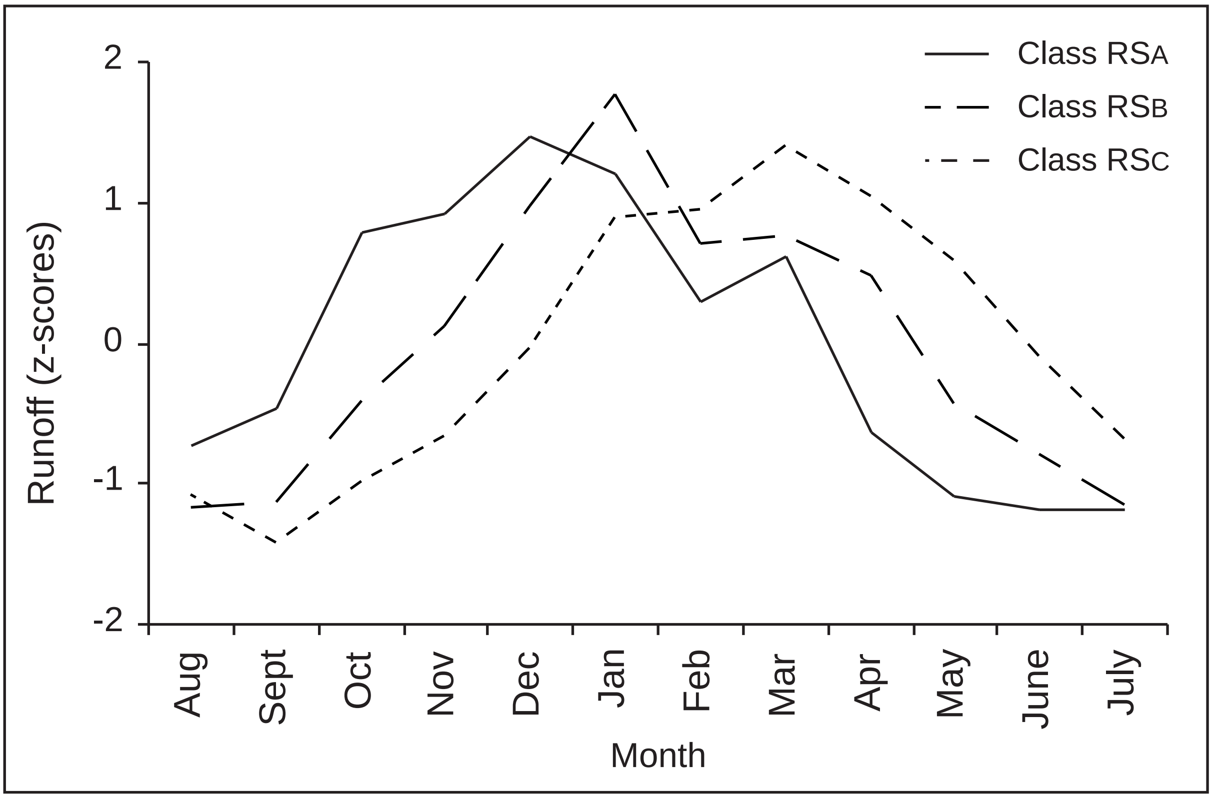


Figure 2



Class RSA
Class RSB
Class RSC
2
1
0
-1
-2
Runoff (z-scores)
Aug
Sept
Oct
Nov
Dec
Jan
Feb
Mar
Apr
May
June
July
Month

Appendix 1 - Variables identified using principal components analysis and/or incorporated within multiple regression models in this study.

Identification code	Hydrological variables	Units
Magnitude of flow events		
Average flow conditions		
CVDF	Coefficient of variation of daily discharges.	–
Q1090DF / Q2080DF / Q2575DF	Ratios of daily discharges of Q_{10}/Q_{90}, Q_{20}/Q_{80} and Q_{25}/Q_{75} percentile flows	–
Q1DF / Q5DF / Q10DF / Q20DF / Q25DF / Q75DF / Q80DF	Percentile flow with the daily discharge exceeded 1/ 5/ 10/ 20/ 25/ 75/ 80% of the time.	m^3s^{-1}
MDF	Mean daily discharge.	m^3s^{-1}
Q50DF	Median daily discharge.	m^3s^{-1}
Q95DF50	Daily Q_{95} percentile flow divided by median daily discharge.	–
STDEVDF	Standard deviation of daily discharge.	m^3s^{-1}
QFEB	Mean February discharge.	m^3s^{-1}
CVANNQ	Coefficient of variation of annual discharges.	–
MADQ	Mean annual discharge.	m^3s^{-1}
MAR	Mean annual discharge divided by catchment area.	$m^3s^{-1}km^{-2}$
Q1	Percentile flow with the annual discharge exceeded 1% of the time.	m^3s^{-1}
Q5Q50 / Q10Q50	Percentile flows with the annual discharge exceeded 5% / 10% divided by median annual discharge.	–
S80	S80 = $(Q_{90} - Q_{10})/ Q_{50}$ calculated from monthly discharge.	m^3s^{-1}
SK2	Skewness = (mean annual discharge - median annual discharge)/ median annual discharge.	m^3s^{-1}
SMED	Median annual discharge divided by catchment area	$m^3s^{-1}km^{-2}$
TOTALVOL	Total discharge volume for that hydrological year.	m^3s^{-1}
High flow conditions		
AMAXDF	Maximum annual daily discharge divided by median annual daily discharge.	–
DFMEDMAX	Median of the highest annual daily discharge divided by the median annual daily discharge.	–
MMAD	Maximum annual monthly discharge.	m^3s^{-1}

Low flow conditions		
AMINDF	Minimum annual daily discharge divided by median annual daily discharge.	–
DFBFI	Baseflow index, i.e. mean of the ratio of the lowest annual daily discharge to the mean daily discharge.	–
DFMEDMIN	Median of the lowest annual daily discharge divided by median annual daily discharge.	m^3s^{-1}
MINAPR	Minimum April discharge.	m^3s^{-1}
MMID	Minimum annual monthly discharge.	m^3s^{-1}
BASEFLOW	Seven-day annual minimum discharge divided by the mean annual discharge.	–
BFI	Baseflow index, i.e. ratio of the lowest annual monthly discharge to the mean annual discharge.	–
SMIN	Annual minimum monthly discharge divided by catchment area.	$m^3s^{-1}km^{-2}$
Frequency of flow events		
High flow conditions		
FRE1	Number of high flow events per year above the median.	–
FRE1YR	Mean number of high flow events per year above the median.	yr^{-1}
Duration of flow events		
High flow conditions		
D3MAX50 / D7MAX50	Average annual 3-day/7-day maximum discharge divided by median annual discharge.	–
DAY7MAX / DAY30MAX / DAY90MAX	Average annual 3-day/7-day/30-day/90-day maximum discharge.	m^3s^{-1}
Low flow conditions		
DFQ95MEAN	Daily Q_{95} percentile flow divided by mean daily discharge.	–
D3MIN50 / D30MIN50	Average annual 3-day/30-day minimum divided by median annual discharge.	–
DAY90MIN	Average annual 90-day minimum.	m^3s^{-1}
Q95MEAN	Monthly Q_{95} percentile flow divided by mean annual discharge.	–