Advances in Water Resources 40 (2012) 11–22

Contents lists available at SciVerse ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

A comparison of similarity indices for catchment classification using a cross-regional dataset

Geneviève Ali [a,*], Doerthe Tetzlaff [a], Chris Soulsby [a], Jeffrey J. McDonnell [a,b], René Capell [a]

[a] School of Geosciences, University of Aberdeen, Elphinstone Road, Aberdeen AB24 3UF, United Kingdom
[b] Department of Forest Engineering, Resources and Management, Oregon State University, Corvallis, OR, USA

ARTICLE INFO

Article history:
Received 28 May 2011
Received in revised form 11 January 2012
Accepted 25 January 2012
Available online 3 February 2012

Keywords:
Catchment classification
Catchment forcing
Catchment form
Catchment function
Affinity propagation
Scotland

ABSTRACT

While there is currently much research activity on catchment classification, there is no agreement on relevant measures of catchment similarity. Here we investigate whether the use of different catchment characteristics as similarity measures leads to convergent catchment classification results. We fed a clustering algorithm called affinity propagation (AP) with different combinations of catchment forcing, form and function indicators collected over 36 Scottish sites (0.44–1712.10 km^2). The AP algorithm was effective in determining the optimal number of groups needed to capture the most variability in each combination of variables. Catchment groupings obtained using physical properties only did not match those obtained using flow indices, mean transit times or storage estimates. The lack of correlation between flow-derived indicators and physical indicators was a surprising result. The combination of data which best approximated the interactions between catchment structural and functional properties included only topographic characteristics, soil properties and mean transit time estimates.

1. Introduction

Extensive hydrological studies conducted at experimental sites around the world have yielded a large amount of data often showing the idiosyncrasies of individual catchments (e.g., [1]). These data reflect highly complex hydrologic behaviours, hence the difficulty to come up with "*concise, easily understood explanations of different basin behaviours*" [2, p. 2]. *Uniqueness of place* [3] tends to limit our ability to create generalizable hypotheses about the overall functioning of hydrological systems, an issue which has been described as "*one of the most vexing*" in hydrology [4, p. 878]. Several authors have however suggested that organizational patterns might be discernable in the topography, soil, geology and vegetation of catchments [5,6,1]. This argument ensues from earlier bodies of work such as those on the catena concept [7], the hierarchy of stream tributaries [8] or the topographic wetness index [9]. While catenas are described as grouping of soils which are "*linked in their occurrence by conditions of topography and are repeated in the same relationships to each other wherever the same conditions are met with*" [7, p. 197], stream orders are used for the comparative analysis of drainage basins, and topographic index curves are often built to assess whether two catchments show a similar distribution of wetness. In all, the potential for relating specific aspects of catchment response to specific configurations of climate and landscape properties has generated some excitement in the field, and we believe that one of the most tangible manifestations of such excitement is the multiplication of catchment classification and regionalization studies (e.g. [10–15]).

Indeed, several authors have made pleas for a unified, broad-scale catchment classification system in hydrology (e.g. [2,16]). Besides providing a "*common language for discussions*" [16–18], catchment classification is a crucial step towards hydrologic synthesis [2,18] to better understand how the different levels of catchment complexity vary in space and time [2,6,19]. While all catchments are unique, they lie in a continuum of hydrological behaviours as a result of different interactions between climatic and physical characteristics. The use of catchment classification, where catchment characteristics are used as measures of similarity between different sites, can therefore be seen as a learning process where the particular controls on the hydrologic response of specific places are confronted with one another before establishing a tentative hierarchy of them. Classification efforts are not new in hydrology: we here refer to cases in which catchments are discriminated as humid versus arid, forested versus agricultural, fast versus slow-responding, groundwater-dominated versus surface-water-dominated, etc. [16]. The main drawback of these classifications is however their focus on individual catchment characteristics (i.e. climate, land use, catchment response, storage, etc.).

To date, no universally accepted metric or combination of metrics has been identified to quantify catchment similarity from the triple point of view of forcing, form and function; different

* Corresponding author. Address: School of Geosciences, University of Aberdeen, St. Mary's Building, Elphinstone Road, Room B35, Aberdeen AB24 3UF, Scotland, United Kingdom. Tel.: +44 1224 273728; fax: +44 1224 272331.
E-mail address: g.ali@abdn.ac.uk (G. Ali).

0309-1708/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.advwatres.2012.01.008

arguments have been made for what might constitute a useful similarity framework. For instance, Buttle [20] has argued that a catchment classification scheme should take into consideration the catchment typology, topography and topology to describe the partitioning of water inputs and their routing through various landscape elements to the river network. McDonnell and Woods [2] recommended classifying measures of fluxes, storages, and response timescales to enhance our ability to discriminate between alternative catchment behaviours. Wagener et al. [16] merged these ideas to suggest that static characteristics illustrating catchment forcing and structure should be combined with dynamic catchment response characteristics. In parallel to these hydrometric-based approaches, recent work has also led to catchment classifications from isotope-based mean transit time (MTT) estimates, that is estimates of the average time for a water molecule to travel through a catchment from rainfall to runoff (e.g. [14]). As a result, the quickly developing field of catchment classification has dealt with almost as many combinations of indicators to be used as similarity metrics as there are papers published on the subject (Table 1). It is interesting to note that while all the similarity metrics reported in Table 1 have a strong physical rationale, the reasons why some are included in or omitted from some papers are rarely mentioned, except perhaps for the obvious reason of availability. Hence, hydrologists have yet to evaluate the impacts of the chosen similarity indices on catchment classification results. A few studies have shown that physical catchment characteristics (e.g., land use, soil types and geology) do not always correlate well with catchment functional characteristics (e.g., runoff coefficient) across scale (e.g. [21,22]). These results therefore hint that catchment classification results might be strongly dependent upon the metrics that we feed into clustering algorithms.

Here we ask the question of whether climatic, topographic, pedologic and hydrologic similarity metrics lead to convergent catchment classification results. Here we do not restrict the definition of hydrologic similarity to runoff metrics but also consider their storage dynamics so that their hydrological behaviour can be compared over short and longer timescales. Our case study focuses on 36 catchments spread over seven different geomorphic provinces of Scotland and for which a whole suite of climatic indices, topographic properties, soil cover proportions, flow percentiles, streamwater mean transit times and storage estimates are available. We therefore test various combinations of these catchment characteristics so as to quantify catchment similarity from a solely structural or "static" point of view (i.e. topographic properties, soil cover proportions), a solely functional point of view (i.e. flow percentiles, streamwater mean transit times, storage estimates), or both. Our approach is different from that of previous catchment classification papers in that we choose to focus only on 36 long-term experimental sites and wish to use our field knowledge to assess the plausibility of the anticipated classification results. We also take this opportunity to introduce and test a relatively recent algorithm, affinity propagation [23], specially designed for clustering purposes but not yet exploited in hydrology. Our dual aim is therefore to investigate the existing or missing correlations between different sets of catchment properties while exploring the potential of a different clustering algorithm in hydrology.

2. Study catchments

The 36 study catchments drain areas ranging from 0.44 to 1712.10 km^2. They are located in seven different geomorphic provinces of Scotland (Fig. 1(B)) and are characterized by contrasting climate, topography, geology, soil cover and land use (see Table 2). All 36 sites have already been the subject of process-based hydrological studies and extensive descriptions of their attributes and behaviours can be found elsewhere (e.g. [14,24,25]).

In brief, our dataset accounts for the main West-East precipitation gradient across the Scottish territory; frontal systems from the

Table 1
Non-exhaustive overview of climatic, hydrologic, hydro-climatic and physical variables used in previous catchment classification and/or regionalization studies.

Variables		Sample publications (bracketed numbers refers to publications in the reference list)
Climatic	Daily rainfall statistics	e.g., [35,14,24]
	Maximum annual daily precipitation	e.g., [21]
	Long-term mean annual rainfall	e.g., [10,11,21,36,13,22,15]
	Actual or potential evapotranspiration	e.g., [37,22,15]
	Ratio of annual precipitation to annual actual or potential evapotranspiration	e.g., [13,22]
Hydrologic	Mean daily flow and/or flow value exceeded 95% of the time	e.g., [11,15]
	Mean annual maximum flood date	e.g., [35]
	Slope of the flow duration curve	e.g., [38]
	Baseflow indices	e.g., [11,13,38]
	Long-term ratio of base flow to runoff	e.g., [22]
	Soil Conservation Service (SCS) curve number	e.g., [22]
	Baseflow chemistry or groundwater contributions from hydrograph separation	e.g., [39]
	Mean transit times or catchment storage estimates	e.g., [14,39,15,24]
Hydro-climatic	Ratio of mean long-term annual runoff to mean long-term annual rainfall	e.g., [38]
	Rainfall-runoff lag time	e.g., [11,35,38]
Physical	Drainage area and/or catchment perimeter	e.g., [10–12,21,36,13,14,39,15,24]
	Elevation	e.g., [21,22,36,13,14,39,15,24]
	Aspect	e.g., [13]
	Catchment slope	e.g., [12,37,21,22,39]
	Topographic and/or downslope index	e.g., [14,39,24]
	Longest flow path length	e.g., [13,14,24]
	Distance and/or gradient to stream	e.g., [39]
	Stream length	e.g., [11,22,36]
	Stream frequency	e.g., [10,11]
	Channel slope	e.g., [10,11,22,36]
	Drainage density	e.g., [21,22,14,39,24]
	Soil cover, vegetation types and/or land use	e.g., [11,37,36,13,22,14,39,15,24]
	Geologic and/or hydraulic properties	e.g., [12,22,15]
	Area covered by lakes, ponds or wetlands	e.g., [11,21,36]
	Soil runoff coefficient	e.g., [36]

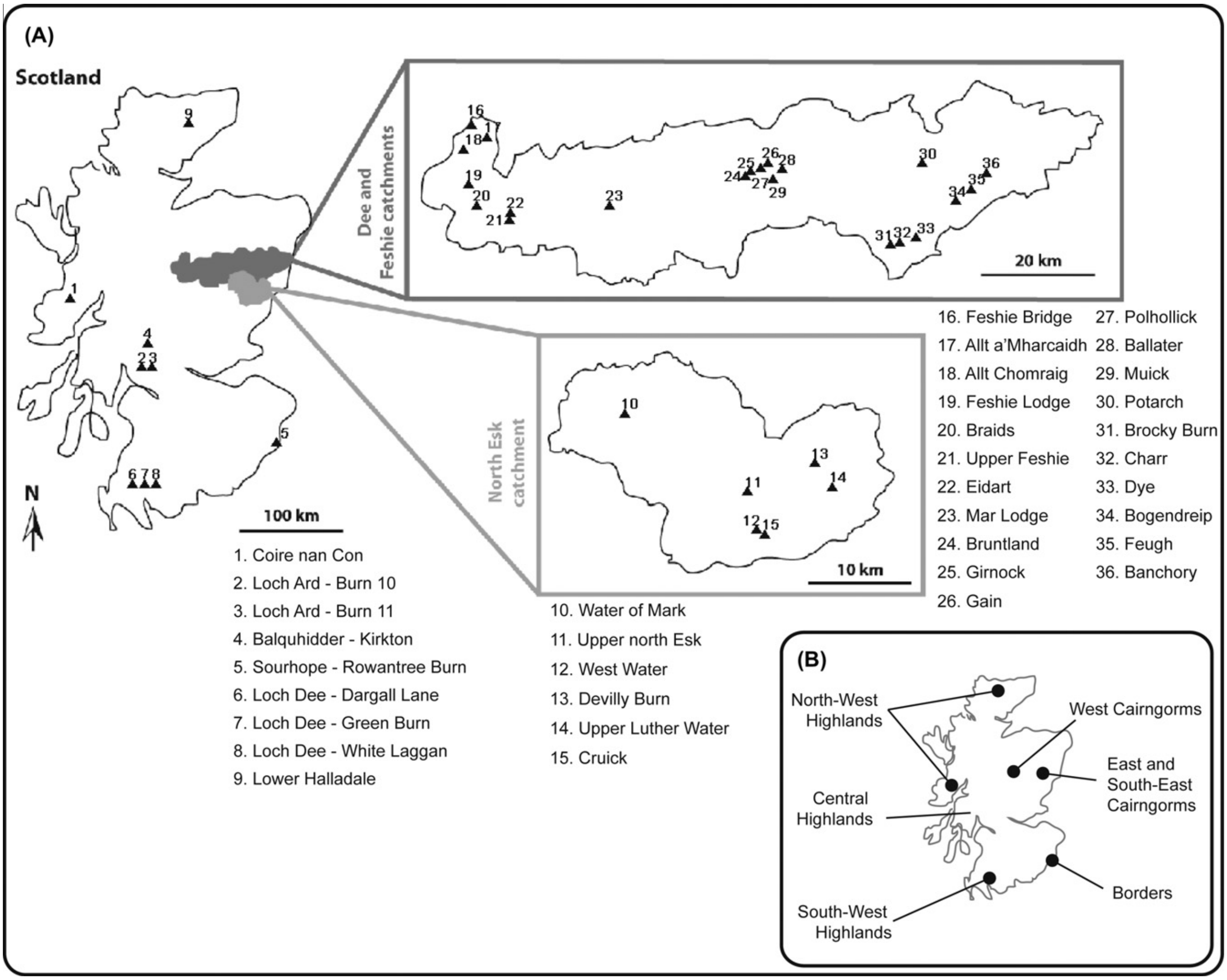

Fig. 1. Location of (A) the 36 study catchments, and (B) the seven geomorphic provinces in Scotland.

Atlantic are known to result in annual precipitation exceeding 2000 mm along the West coast whereas less than 1000 mm are recorded in the rain shadow to the East. Precipitation is more or less evenly distributed throughout the year, with higher-magnitude events usually occurring during the autumn and winter seasons. Mean annual temperature is mainly controlled by elevation differences and to a lesser extent by latitude. Elevation gradient are important as they help distinguish montane catchments (e.g. Coire nan Con, Loch Dee – Green Burn, Dargall Lane and White Lagan, Balquhidder – Kirkton, Allt a'Mharcaidh, refer to Fig. 1(A)) from lowland sites (<300 m in altitude, e.g. Loch Ard – Burn 10 and 11, Lower Halladale, Cruick). Headwater sites are generally alpine in character with steeper slopes while lowland regions consist of more gentle topography with undulating forms and large and flat valley bottoms. Most sites are dominated by low permeability igneous and metamorphic rocks which range from granite (e.g. Green Burn, Dargall Lane, Lower Halladale, Allt a'Mharcaidh), through to schist and gneiss (e.g. Coire nan Con) and other metamorphic rocks (e.g. White Laggan, Balquhidder – Kirkton, Loch Ard – Burn 10 and 11). Exceptions are the fractured volcanic rocks (e.g. Sourhope – Rowantree Burn) or sandstones (e.g. Cruick and other lowland North Esk sites). At most sites, variable assemblages of drift are superimposed on the solid geology. This ranges from compacted fine textured basement tills (e.g. Girnock) to freely draining fluvioglacial deposits (e.g. Allt Chomraig).

Soil cover proportions were extracted from the United Kingdom Hydrology Of Soil Types (HOST) classification [26] which groups soils into 29 different classes reflecting dominant runoff processes. The pedologic characteristics of the 36 catchments studied here reflect the dual influence of topography and geology within and across regions. For example, in valley bottoms and on gentle slopes where the superficial drifts are fine textured, peats and peaty gley soils are often present (e.g. Girnock, Feshie Lodge, Lower Halladale, Coire nan Con, Loch Ard – Burn 10 and 11, see Table 2); they are referred to as "highly responsive soils" as they are subjected to poor vertical drainage and remain close to saturation all year long. On the contrary, areas with steeper slopes or more permeable drifts are preferential locations for humus-iron Podzols, subalpine and alluvial soils (e.g. Allt a'Mharcaidh, Balquhidder – Kirkton, Upper Luther Water, Cruick, Dee at Banchory, see Table 2) and facilitate groundwater recharge, hence their denomination of "freely draining soils". Rankers (or regosols) also develop on the steepest mountain slopes. These hydropedological characteristics can be translated into contrasting dominant runoff generating flow paths: responsive soils are primarily associated with overland flow and shallow lateral subsurface flow whereas freely draining soils allow deeper subsurface flow and groundwater recharge to occur. As far as land use is concerned, most catchments are dominated by heather moorland or montane vegetation. Catchments with significant forest cover include Coire nan Con,

Table 2
Summary of study catchment characteristics. CV refers to "coefficient of variation". Physical characteristics were derived from DEM analyses while soil coverage proportions were computed from HOST maps. Flow indices reflect the streamflow dynamics at the outlet of each study catchment. Streamflow mean transit times were estimated by Hrachowitz et al. [14,24] using Gamma transit time distributions; "5%" and "95%" refer to the 5th and 95th percentiles of the behavioural subsets. Catchment storage values were approximated by multiplying MTT estimates by mean annual precipitation values.

	Min	Max	Mean	Median	Std	CV
Area (km^2)	0.44	1712.10	194.09	49.00	385.31	1.99
Climatic index						
Mean annual precipitation (mm)	876.00	3400.00	1543.70	1256.00	714.58	0.46
Terrain characteristics						
Minimum elevation (m)	18.00	518.00	205.06	224.90	128.69	0.63
Maximum elevation (m)	220.00	1305.00	864.94	851.50	326.38	0.38
Mean elevation (m)	143.00	865.00	464.95	452.50	178.37	0.38
Maximum slope (deg)	25.80	78.30	51.94	50.10	16.42	0.32
Mean slope (deg)	4.00	19.00	10.74	10.90	3.45	0.32
Drainage density (km/km^2)	0.65	3.88	2.14	2.35	1.01	0.47
Mean flow path length (km)	1.41	128.99	23.04	12.86	29.96	1.30
Median topographic index (ln(m))	5.10	7.80	6.51	6.46	0.74	0.11
Soil coverage proportions (decimal fractions)						
Alluvial soils	0.00	0.11	0.02	0.00	0.03	1.40
Humus-iron Podzols; subalpine soils	0.00	0.74	0.18	0.12	0.19	1.06
Brown forest soils	0.00	0.65	0.04	0.00	0.14	3.82
Rankers	0.00	0.82	0.14	0.02	0.21	1.48
Peaty Podzols and peaty gleys	0.00	1.00	0.38	0.32	0.28	0.73
Peat	0.00	0.66	0.21	0.12	0.21	1.00
Eroded peat	0.00	0.31	0.02	0.00	0.06	3.75
Gleysols	0.00	0.09	0.01	0.00	0.02	2.17
Open water	0.00	0.02	0.00	0.00	0.00	3.18
Freely draining soils	0.00	0.78	0.30	0.28	0.24	0.80
Responsive soils	0.22	1.00	0.70	0.72	0.24	0.35
Flow indices						
Mean daily discharge ($l\ s^{-1}\ km^{-2}$)	10.42	99.39	37.69	31.39	21.83	0.58
Q95 (exceeded 95% of time) ($l\ s^{-1}\ km^{-2}$)	1.49	10.15	5.84	6.14	2.37	0.41
Q5 (exceeded 5% of time) ($l\ s^{-1}\ km^{-2}$)	43.98	381.65	131.49	99.00	86.84	0.66
Mean annual flow (sum, mm)	478.00	2974.00	1159.90	975.50	601.90	0.52
MTT (mean transit time) indices						
Median MTT (days)	61.00	2285.00	811.42	662.00	640.32	0.79
MTT, 5% (days)	12.00	1172.00	406.17	354.00	301.18	0.74
MTT, 95% (days)	117.00	3866.00	1247.80	992.00	1029.90	0.83
Storage indices						
Median storage (mm)	300.00	5787.00	1953.80	1699.00	1222.00	0.63
Median storage, 5% (mm)	72.00	2652.00	1003.30	896.50	596.55	0.59
Median storage, 95% (mm)	491.00	8808.00	3039.50	2517.00	1970.80	0.65

Balquhidder – Kirkton, Allt a'Mharcaidh, and the Loch Dee and Loch Ard sites. Lowland areas are often used for arable agriculture and pasture. Very limited anthropogenic influences can be observed; the largest settlement in all our 36 catchments is Banchory with a population of ca. 6000 inhabitants.

Mean daily discharges among all 36 sites range from 10.42 $l\ s^{-1}\ km^{-2}$ to 99.39 $l\ s^{-1}\ km^{-2}$ (Table 2). Nonparametric Spearman rank correlation coefficients between the median MTT and the mean daily discharge and between the median MTT and Q5 are, respectively, $r_{Spearman} = -0.75$ (p-value < 0.0001) and $r_{Spearman} = -0.68$ (p-value < 0.0001). This shows that in catchments with the longest streamwater mean transit times, the damped hydrograph responses are perceptible through the lower values of the mean daily flow and the flow levels that are exceeded 5% of the time (Table 2). Transit time estimates range from about two months to five or six years (Table 2): shorter ones are encountered in the Coire con Nan, Loch Ard (Burn 10 and 11), Lower Halladale and the Loch Dee (Green Burn, Dargall Lane and White Lagan) while longer ones are present in the North Esk sub-catchment and at Sourhope (Rowantree).

3. Classifying catchments using affinity propagation

Affinity propagation (AP) was introduced by Frey and Dueck [23] in the field of computer science and is becoming increasingly popular in physical sciences as a powerful clustering tool. It is different from standard clustering algorithms as it has the double aim of (i) partitioning the objects of a dataset into groups of apparently similar objects, and (ii) identifying, for each group, a single object or "exemplar" that is the most representative of that group. The AP algorithm has the ability of "*greatly compressing a potentially massive dataset very efficiently while identifying and retaining its most representative elements*" ([27], p. 2). The rationale behind AP is different from standard clustering methods as the algorithm does not need the number of clusters to be specified by the end-user prior to the classification. Each object in a dataset is considered as a node in a network; real-value messages are recursively transmitted along the edges of the network until a good set of exemplars gradually emerges. At each iteration, the magnitude of each message reflects the current affinity that one object has for choosing another object as its exemplar, hence the term "affinity propagation" [23].

Full algorithmic details for AP can be found in [23]. For the purpose of this paper, objects are in fact catchments with various combinations of attributes. As input to the algorithm, a square similarity matrix s is used:

$$s = -d^2 \tag{1}$$

where d is an Euclidean distance matrix computed from data vectors. Each object must be supplied with a preference value that specifies *a priori* how likely each of them is to become an exemplar. When no prior knowledge is available and all objects are considered to be equally suitable as exemplars, as is the case with our dataset,

Table 3
Combinations of catchment properties used for different classification runs. Refer to Table 2 for detailed information about each property or group of properties. The "level of complexity" varies from 1 to 3 and refers to whether the dataset used for the classification includes only one type of information (e.g. topographic or pedologic or flow data) or rather multiple types of information.

Combination name	Properties included							
	Area	Climatic index	Terrain properties	Soil coverage proportions	Flow indices	MTT indices	Storage indices	Complexity level
CLIMATIC		√						1
TOPOGRAPHIC			√					1
SOIL				√				1
FLOW					√			1
MTT						√		1
STORAGE							√	1
PHYSICALwithAREA	√	√						2
PHYSICALwithoutAREA		√	√	√				2
HYDROLOGIC					√	√	√	2
ALLwithAREA	√	√	√	√	√	√	√	3
ALLwithoutAREA		√	√	√	√	√	√	3

the preference values are set to the median of the input similarities [23], thus allowing the clustering procedure to "learn" the appropriate number of exemplars (and clusters) from the data. Upon publication of their paper, Frey and Dueck released a MATLAB (Mathworks, Inc.) code for affinity propagation which was later adapted for use in R [28]. The resulting R package, *apcluster* [29], was used for our analyses.

In our application, we applied the AP algorithm using different combinations of catchment characteristics as similarity metrics, and later we compared the classification results obtained. Table 2 shows the suite of characteristics which were compiled for each catchment while Table 3 shows the different combinations of catchment properties used as similarity metrics and fed into the AP algorithm. It should be noted here that in order to be consistent with the objective of comparing "purely structural" and "purely functional" similarity metrics, it would have been preferable to use non-processed soil data, namely information on soil cover which had not be transformed and/or interpreted in light of dominant runoff processes as is the case with HOST maps. Unfortunately the HOST data were the only soil-related information available. While the 29 HOST classes are assumed to reflect dominant runoff processes, the conceptual models of runoff generation on which they rely are very simplistic. Hence, as a secondary research objective, we also wish to test whether relative proportions of HOST classes are correlated with traditional indicators of catchment behaviour or catchment response such as flow percentiles or transit time estimates. Each combination of catchment properties was also assigned a level of complexity (Table 3). These levels of complexity ranged from 1 to 3 to reflect the different data types included in each combination. For example, the "TOPOGRAPHIC" combination included terrain characteristics only, hence its complexity level of 1, while the "PHYSICALwithoutAREA" combination incorporated terrain characteristics and soil cover proportions (complexity level of 2) and the "ALLwithAREA" combination (complexity level of 3) covered all variables listed in Table 2. In total, 11 different classification runs were achieved using the 11 different combinations of variables reported in Table 3.

Classification results were examined in several ways:

- We mapped the AP results by showing the partitioning of the catchments into different groups and the location of catchment exemplars. In addition to the maps, the regional dependence of the identified groups was also investigated. Cramér's V [30] measure of association was used to assess whether each cluster was self-contained into a given geomorphic province. Cramér's V is a numerical index that describes the strength of the relationship between two nominal variables, in our case the names of the seven geomorphic provinces of Scotland (Fig. 1(B)) and the cluster memberships in a given classification run. Cramér's V comes from contingency table analysis and is computed by taking the square root of the chi-square statistic χ^2 divided by the sample size N and the length of the minimum dimension of the crosstab K:

$$\text{Cramér's V} = \sqrt{\frac{\chi^2}{N(K-1)}} \tag{2}$$

K is the smallest of the number of rows or columns in the contingency table. Cramér's V values lie between 0 and 1, with a minimum value of zero indicating that the groups show no regional dependence and a maximum value of one rather indicating that the groups are strongly associated with one of the seven geomorphic provinces. For simplification purposes, we qualitatively assessed the regional dependence of groups as null, mediocre, moderate or strong when Cramér's V values were less than 0.3, between 0.3 and 0.5, between 0.5 and 0.7, and above 0.7 respectively.
- We built cluster plots to compare the individual characteristics of different clusters in a given classification run. Each cluster was defined by its catchment exemplar and associated with a qualitative label. These qualitative labels (e.g. low, moderately low, moderate, moderately high, and high) were defined using the 10th, 25th, 50th and 75th percentile values of each catchment characteristic.
- We defined an exemplar propensity function to quantify the propensity of each catchment to become a group exemplar in one or many of the 11 classification runs. The exemplar propensity function was defined as follows:

$$P_{k\leftarrow i} = \frac{\sum_{nclassifruns=1}^{11} Exemplarity_{nclassifruns} \times ComplexityLevel_{nclassifruns}}{\sum_{nclassifruns=1}^{11} ComplexityLevel_{nclassifruns}} \tag{3}$$

where $P_{k\leftarrow i}$ is the propensity of catchment i to be an exemplar k across all classification runs, $Exemplarity_{nclassifruns}$ is a binary index (1 or 0) indicating whether or not catchment i is an exemplar k in the classification run *nclassifruns*, and $ComplexityLevel_{nclassifruns}$ is the complexity level associated with the combination of catchment properties used in the current classification run. The exemplar propensity $P_{k\leftarrow i}$ therefore varies between 0 and 1 and gives a greater importance to catchments which tend to become exemplars when higher complexity datasets are used for the classification.
- We employed the Adjusted Rand Index (ARI, see [31] for details) to determine whether "physically similar" catchments were also "hydrologically similar". The standard Rand Index measures the agreement between two classifications C1 and C2 using the following formula:

$$\text{Rand Index} = \frac{a+b}{a+b+c+d} \tag{4}$$

Table 4
Number of groups determined using the traditional AP algorithm (negative squared Euclidean distance matrix and preferences set to the median of input similarities, see details in text). The regional dependence of identified groups is also assessed via Cramér's V measure of association (i.e. association between the spatial spread of the group members and the seven Scottish physiographic regions investigated).

Combination of catchment properties	Number of groups	Cramér's V
CLIMATIC	6	Strong
TOPOGRAPHIC	6	Moderate
SOIL	8	Moderate
FLOW	8	Moderate
MTT	5	Strong
STORAGE	6	Strong
PHYSICALwithAREA	6	Strong
PHYSICALwithoutAREA	6	Strong
HYDROLOGIC	9	Moderate
ALLwithAREA	8	Strong
ALLwithoutAREA	8	Strong

where a is the number of catchment pairs that are in the same group in classification C1 and in the same group in classification C2; b is the number of catchment pairs that are in different groups in C1 and in different groups in C2; c is the number of catchment pairs that are in the same group in C1 but in different groups in C2; and d is the number of catchment pairs that are in different groups in C1 but in the same group in C2.

The adjusted form of the Rand Index simply corrects the formula for chance, taking into account the fact that randomness may cause some catchments to pertain to the same group:

$$\text{Adjusted Rand Index} = \frac{\binom{N}{2}(a+d) - [(a+b)(a+c)+(c+d)+(b+d)]}{\binom{N}{2}^2 - [(a+b)(a+c)+(c+d)+(b+d)]} \quad (5)$$

where $\binom{N}{2}$ is the total number of possible combinations of pairs. The closer the ARI is to one, the better the agreement between the two classifications C1 and C2.

4. Results

4.1. Group patterns

The differences in group patterns and exemplar locations were highly dependent upon the combination of catchment properties fed into the AP algorithm. Table 4 shows that most combinations of catchment properties led to classifications with 6 or 8 different groups. The "HYDROLOGIC" classification resulted in the highest number of groups (i.e. 9) while the "MTT" classification gave the lowest number of clusters (i.e. 5) and the "FLOW" classification was associated with 8 catchment groups. This suggests that indicators of short-term hydrologic functioning were highly heterogeneous between catchments while MTT estimates effectively dampened this variability and presented a more uniform clustering across many sites. According to the values of Cramér's V, it was assessed that all classifications but those based on the "FLOW", "SOIL", "TOPOGRAPHIC" and "HYDROLOGIC" combinations of

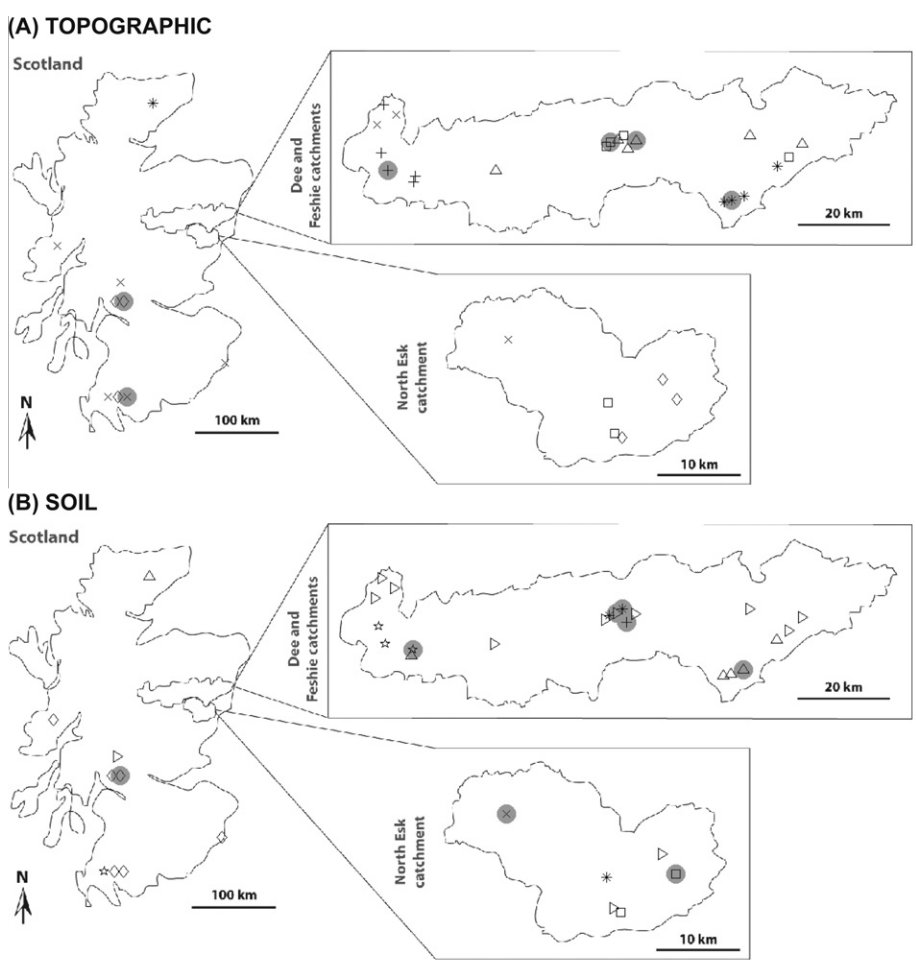

Fig. 2. Classification results according to the TOPOGRAPHIC and SOIL combinations of variables. Each black symbol illustrates a different group. Group exemplars are flagged with a grey-shaded circle. Note that while symbols are re-used in panels (A) and (B), they do not identify the same groups.

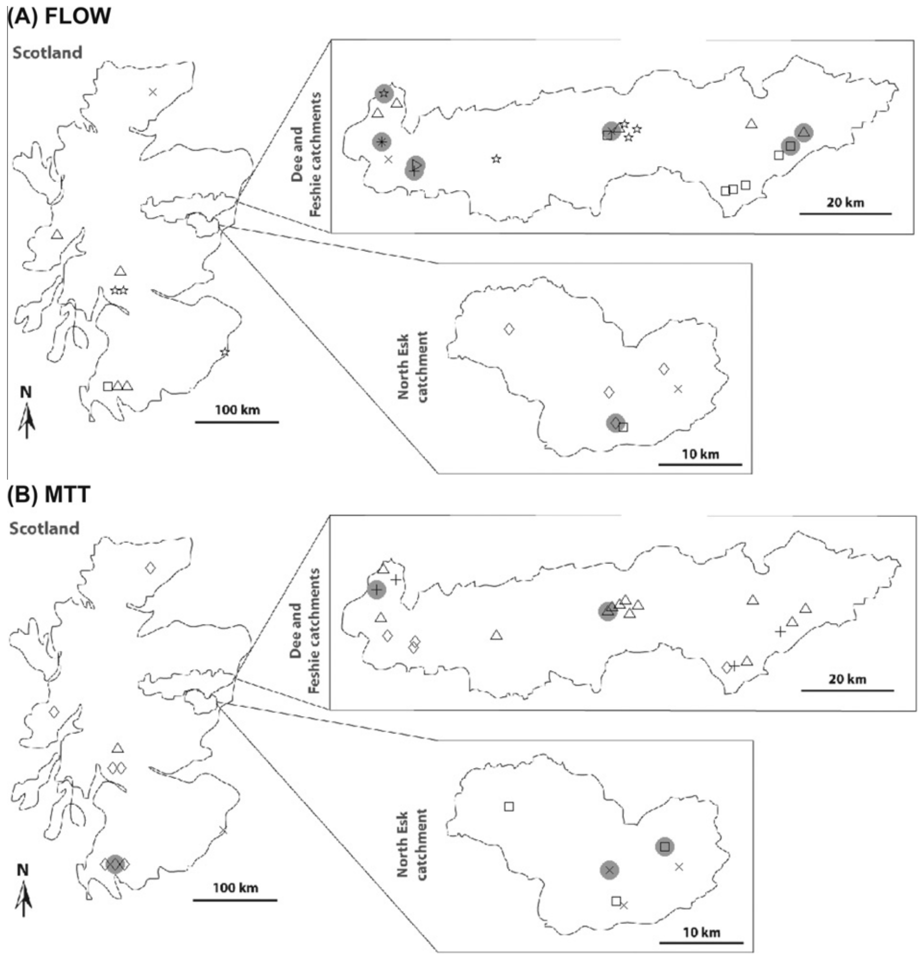

Fig. 3. Classification results according to the FLOW and MTT combinations of variables. Symbology is the same as in Fig. 2.

catchment characteristics showed relatively strong patterns of regional dependence (Table 4). This is supported by the group mappings shown in Figs. 2–5. Indeed, some groups associated with the ''MTT'' and the ''HYDROLOGIC'' classifications were very self-contained within the limits of a geomorphic province (e.g. square symbols are exclusively located in the North Esk region in Figs. 4(B) and 5(B)). On the other hand, the spatial dependence was moderate with the ''FLOW'' classification as some groups clearly spanned over multiple geomorphic provinces (e.g. upward-pointing triangles in Fig. 3(A)). It should be noted that the classification map associated with the ''TOPOGRAPHIC'' combination of variables (Fig. 2(A)) was the only one where no exemplar sites were selected within the North Esk region. Also, when looking at the Dee and Feshie catchments, the most Eastern, lowland sites were never identified as exemplars except in the ''FLOW'' classification.

For some combinations of variables with a complexity level of 1 (refer to Table 3), it was possible to discern patterns among the defined groups. For instance, when considering the ''TOPOGRAPHIC'' classification, the mean elevation values associated with the exemplar catchments were strongly correlated with mean slope values (Spearman rank correlation coefficient: $\rho = -0.83$, 5% statistical significance level) and mean flow path length values ($\rho = -0.88$), thus meaning that we could observe consistent ''lower to higher elevation'', ''higher to lower slope'' and ''shorter to longer flow path lengths'' patterns. Similarly, when considering the ''SOIL'' classification, some sites showed a strong association between alluvial soils and humus-iron podzols ($\rho = -0.88$) while some others were characterized by high proportions of rankers and gleysols ($\rho = -0.79$). When aggregating properties and considering dataset complexity levels of 2 and 3, however, patterns were more difficult to discern and rather complex. With the ''PHYSICALwithAREA'' classification, gradients in mean elevation and proportion of responsive soils could be perceived but they were associated with complex group patterns when it came to drainage area, mean annual precipitation, mean slope and mean flow path length. Such was also the case with the ''HYDROLOGIC'' and the ''ALLwithAREA'' classifications. As an example, Fig. 6 highlights some individual characteristics of the 8 catchments exemplars in the ''ALLwithAREA'' classification. This figure shows that the three groups with the highest mean daily discharges (i.e. cross, asterisk and right-pointing triangle symbols) were also associated with the largest levels of mean annual precipitation, moderately low MTT values and moderately high to high proportions of responsive soil cover. The three groups with the longest MTT values (i.e. diamond, square and upward-pointing triangle symbols) were however associated to both lowland (i.e. square and upward-pointing triangle symbols) and upland (i.e. diamond symbol) sites, gently sloping (i.e. square and upward-pointing triangle symbols) and steep (i.e. diamond symbol) topographies, and moderate (i.e. diamond symbol) and high proportions (i.e. square and upward-pointing triangle symbols) of responsive soil cover, thus making the identification of dominant physical controls on hydrologic functioning difficult and the group patterns somehow unpredictable.

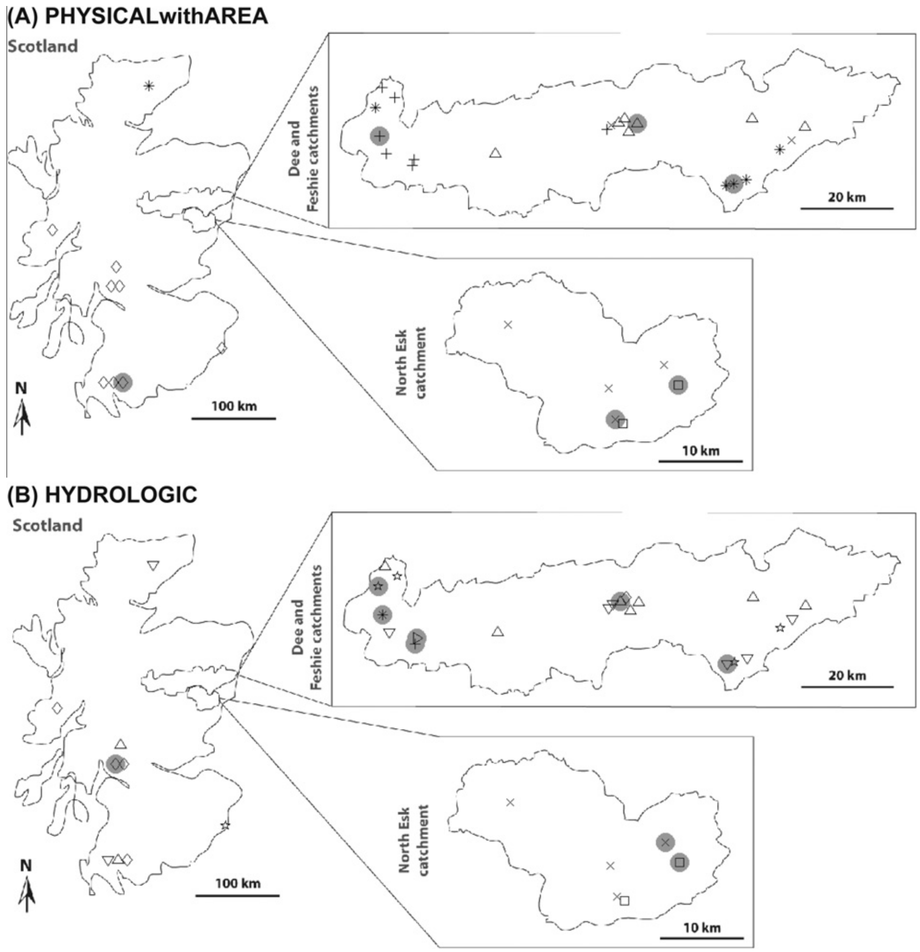

Fig. 4. Classification results according to the PHYSICALwithAREA and HYDROLOGIC combinations of variables. Symbology is the same as in Fig. 2.

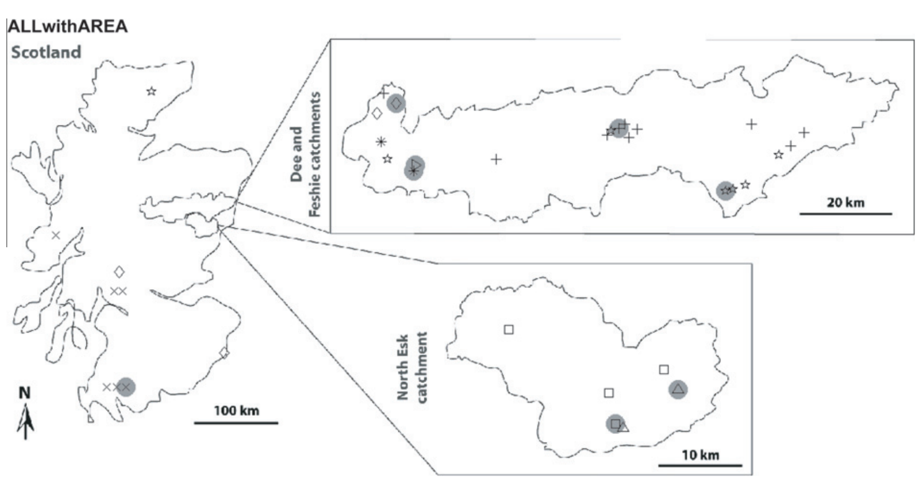

Fig. 5. Classification results according to the ALLwithAREA combination of variables. Symbology is the same as in Fig. 2.

4.2. Comparison of similarities

The quantification of the agreement between the different classification runs revealed that these physically and climatically similar Scottish catchments were not necessarily hydrologically similar. Adjusted Rand Index (ARI) values reported in Table 5 show that in general, there was a very weak agreement between classifications based on physical catchment characteristics (e.g. "CLIMATIC", "TOPOGRAPHIC", "SOIL" combinations of variables) and classifications based on proxies for hydrological behaviour (e.g. "FLOW", "MTT", "STORAGE" combinations of variables). The "PHYSICALwithAREA" classification shared an ARI value of only 0.28 with the "HYDROLOGIC" classification, 0.24 with the "MTT" classification and 0.16 with the "FLOW" classification (Table 5). The comparison of the "ALLwithAREA" and the "ALLwithoutAREA" classifications, as well as the comparison of the "PHYSICALwithAREA"

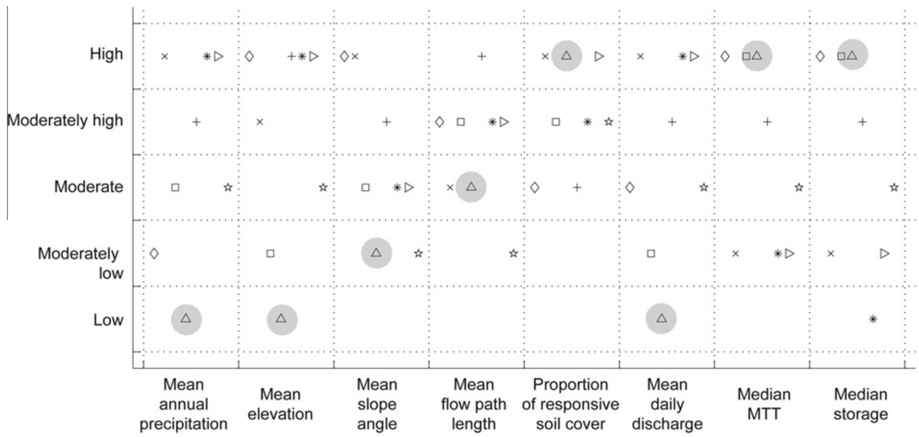

Fig. 6. Cluster plot showing the main characteristics of individual groups following the "ALLwithAREA" classification run. Symbols used to identify groups are the same as in Fig. 5(C). Each group is defined by its exemplar and associated with a qualitative label (e.g. low to high) defined using the 10th, 25th, 50th and 75th percentile values of each characteristic. As an example of how to read this diagram, upward-pointing triangles are flagged with a grey-shaded circle. It can be observed that catchments represented by these triangles in Fig. 5(C) are usually characterized by low values of mean annual precipitation, mean elevation and mean daily discharge, moderately small slopes, moderate flow path lengths, high proportions of responsive soil, and high values of MTT and storage.

Table 5
Agreement between the different classification runs assessed using the Adjusted Rand Index (ARI). The closer to 1 the value of the ARI is, the better the agreement between two classification runs. As an example, the shaded area illustrates the comparison between the "ALLwithAREA" and the "ALLwithoutAREA" classifications with an ARI value of 0.9.

	CLIMATIC	TOPOGRAPHIC	SOIL	FLOW	MTT	STORAGE	PHYSICALwithAREA	PHYSICALwithoutAREA	HYDROLOGIC	ALLwithAREA	ALLwithoutAREA
CLIMATIC	1.00	0.07	0.16	0.00	0.03	0.02	0.06	0.09	0.04	0.13	0.15
TOPOGRAPHIC		1.00	0.25	0.12	0.11	0.19	0.49	0.47	0.20	0.36	0.39
SOIL			1.00	0.14	0.25	0.09	0.30	0.32	0.22	0.40	0.39
FLOW				1.00	0.03	0.11	0.16	0.15	0.31	0.24	0.21
MTT					1.00	0.21	0.24	0.34	0.32	0.44	0.51
STORAGE						1.00	0.22	0.19	0.25	0.26	0.26
PHYSICALwithAREA							1.00	0.85	0.28	0.49	0.49
PHYSICALwithoutAREA								1.00	0.25	0.55	0.63
HYDROLOGIC									1.00	0.44	0.41
ALLwithAREA										1.00	**0.90**
ALLwithoutAREA											1.00

and the "PHYSICALwithoutAREA" classifications yielded high ARI values of 0.90 and 0.85 respectively, thus suggesting that the catchment drainage area does not have a significant influence on the way the classification groups were defined during AP.

The "ALLwithAREA" combination of properties allowed us to assess overall catchment similarity rather than focusing on physical (i.e. structural) or hydrological (i.e. functional) aspects in isolation. Both the 8 groups and the 8 exemplar catchments that best represented each group can be seen as the most representative yet distinct models of catchment dynamics that can be found amongst our 36 study sites. In the absence of overlapping results between the different classification runs, however, one may ask what replacement data might be used to approximate the "ALLwithAREA" similarity patterns when all the characteristics listed in Table 2 are not available. Table 5 shows that the two best compromises with that regards would be the "PHYSICALwithoutAREA" and the "MTT" classifications which shared ARI values of 0.55 and 0.44, respectively, with the "ALLwithAREA" grouping results. The ARI value between the "ALLwithAREA" and the "FLOW" classifications was only 0.24, thus highlighting the fact that streamflow characteristics alone could not approximate the interactions between physical and hydrological characteristics in the catchments studied.

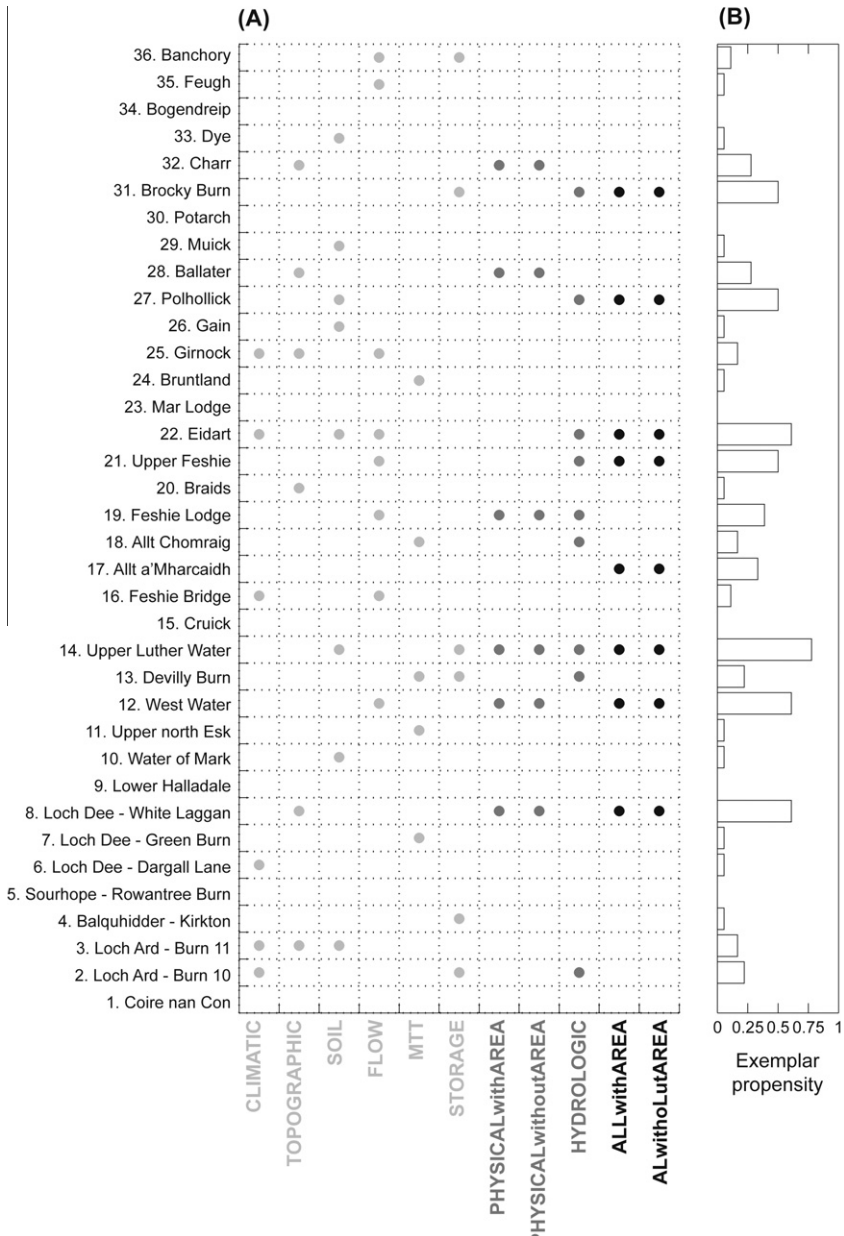

Fig. 7. (A) Catchments endorsing the role of group exemplars in each classification run. Dot colours from light grey to black show datasets with increasing levels of complexity (refer to Table 3). (B) Value of the exemplar propensity function for each catchment. Refer to Fig. 1 for the location of the study sites.

The exemplar propensity function previously defined in Section 3 was also useful as it revealed that the suitability of a given site to become the ambassador of a group was highly dependent upon the data fed into the AP algorithm. Fig. 7 shows that catchments such as the Dee at Banchory, Girnock and Loch Ard – Burn 11 tended to be group exemplars only in the classifications involving combinations of characteristics with a complexity level of 1. When datasets with complexity levels of 2 or 3 were used, however, very few catchments consistently endorsed the role of group exemplars except for Brocky Burn, Polhollick, Eidart, Upper Luther Water, West Water and Loch Dee – White Laggan. This observation adds to the evidence that many “catchment similarities” exist and classification exercises should be undertaken with caution when different types of catchment characteristics are not available.

5. Discussion

5.1. On the equivalence of catchment similarity indices

Our AP-based classification exercise showed the usefulness of the chosen similarity metrics for the resulting groupings. By comparing various combinations of catchment characteristics so as to quantify catchment similarity, we sought to test the hypothesis that classifications based on low-complexity combinations of variables would be significantly different to the groupings obtained using higher-complexity datasets. Figs. 2–5 together with Tables 4 and 5 support this hypothesis and thus highlight the lack of correlation between physical (i.e. forcing and form) and hydrologic (i.e. function) similarity indices. While looking for the best compromises to approximate the “ALLwithAREA” grouping results, the “PHYSICALwithoutAREA” combination of catchment properties was ranked first, followed by the “MTT” dataset. The way these two alternatives were ranked is especially interesting as it reveals the first-order control of catchment physical attributes and the usefulness of MTT estimates as a proxy of hydrological behaviour, especially in terms of inferring the dynamics of water storage and release.

It is surprising that no combinations of forcing and form indicators (i.e. “PHYSICALwithAREA”) tested in this paper correlated well with the streamflow characteristics (Q95, Q5 and mean daily discharge). While previous catchment classification studies have relied on streamflow indicators (Table 1) our results suggest that the choice of relevant similarity metrics may be region- or context-dependent, at least for our 36 sites. It is likely that flow regime properties are useful to characterize the quickly-responding catchments in which near-surface flow paths are often activated; flow properties appear to be less useful for more groundwater-dominated catchments in which deeper mixing processes occur. These dual dynamics are captured well by streamwater transit time distributions, and this might explain why MTT estimates performed better as catchment functioning surrogates in our analyses. The fact that streamflow-derived indices were not well correlated with catchment physical properties in our analyses might also be linked to differences in landscape evolution histories among our study catchments. The landscape of many parts of the Scottish Highlands reflects the ancient geological history and the effects of selective glacial erosion. Thus, the relatively short post-glacial period means that recent hydrological and fluvial processes have had a secondary influence on catchment characteristics such as the presence of wide and deep valleys. Also, relict paraglacial features influence the distribution of wetlands and zones of internal drainage [32] which can be topographically isolated from the drainage network. Runoff generated on these areas might therefore not reach the stream or does so only via deeper groundwater pathways. Thus, in ancient glaciated landscapes such as Scotland, it is the combination of complex drift distributions and topography together that determines soil hydrology, hence the importance of MTTs. In regions where limited topographic variations and relatively uniform soils are encountered, however, it is rather the topology of landscape features adjacent to the river channel network which are strong hydrological determinants [20].

5.2. On the potential of affinity propagation for catchment classification

In light of the analyses reported in this paper, we believe that the AP algorithm has some potential for catchment classification as it is highly computationally efficient, and has the advantage of determining “on its own” the optimal degree of partitioning (number of groups) needed for a specific dataset. This later aspect could be interesting from a process understanding point of view as we believe that the optimal number of groups determined by AP could be interpreted as the different levels of catchment organizing principles along a given continuum. The AP algorithm allows one to go further than the simple definition of catchment classes by helping one identify “benchmark” or “exemplar” sites and contextualizing existing “iconic” sites that are implicitly assumed to be more generally representative. Mézard [33] notes that “*detecting exemplars goes beyond simple clustering, as the exemplars themselves store compressed information*” (p. 949). The identification of such exemplars could be helpful in rationalizing sampling efforts in hydrology, especially as it is possible to modify the AP algorithm so that it can differentiate outliers from exemplars [34]. This small distinction might be useful when dealing with very large datasets, namely datasets larger than the one we relied on in this paper. However, exemplars do not help understand what the response of a particular catchment might be if it lies at the boundary between two classification groups. With that regards, the AP algorithm is very similar to other NP-hard clustering methods where each object is assigned to a unique group even though it might seems more sensible for each object to pertain to different groups with different degrees of membership (e.g., fuzzy clustering).

While beyond the scope of the present paper, we briefly evaluated the sensitivity of the classification results to the distance matrix and the preference value used (data not shown). In addition to the negative squared Euclidean distance matrix (common rule), other distance matrices were also tested (i.e. maximum, Manhattan, Canberra, Minkowski ($p = 3$), radial basis function (Gaussian) kernel and Laplace kernel). We observed that for datasets with a complexity level of 1, all distance matrices, except Canberra and Minkowski, led to similar classifications. For datasets with complexity levels of 2 and 3, the agreement between classifications using different distance matrices was usually poor. The chosen distance matrix was then highly important as it controlled not only the number of groups but also the catchments that would likely be chosen as group exemplars. We also ran the algorithm by setting the preference value to the minimum, rather than the median, of input similarities and this caused the number of groups in all classifications to be consistently equal to or higher than the number obtained while using the median of input similarities. Hence, even though the AP algorithm has clear advantages, it does share some of the same drawbacks as common clustering methods when it comes to the results dependency on the chosen distance matrix.

6. Conclusion

This paper aimed at comparing a range of similarity indices for catchment classification using a cross-regional dataset. Our focus was on 36 catchments, some of them partly nested, ranging in size and spread over seven different geomorphic provinces of Scotland. We fed a relatively new clustering algorithm called affinity propagation with various combinations of catchment characteristics to assess whether climatic, topographic, pedologic and hydrologic similarity indices lead to convergent catchment classification results. Affinity propagation provided an objective means to quantify the optimal number of groups needed to capture the most variability within our dataset. This application also allowed us to identify exemplar catchments that were the most representative of their respective groups. While the idea of exemplars is quite popular in other sciences (e.g., [33]), further work might be needed before we can assess the usefulness of such exemplars towards catchment process understanding and hydrologic synthesis. Our results showed that neighbouring catchments were usually but not always more similar than distant catchments. Also, catchment groupings obtained on the basis of topographic properties did not always match those obtained using flow indices, mean transit times or storage estimates. The lack of correlation between flow-derived and physical similarity indices was particularly surprising as such indices have been used indifferently in previous catchment classification studies. While we do not claim that such a conclusion would hold in another environment, we hypothesize that for our Scottish regional context, the combination of data which best approximates the complex interactions between catchment structural and functional properties only included topographic characteristics, soil properties and mean transit time (MTT) estimates. We therefore conclude that while there have been calls for a unified, broad-scale classification framework, our results seem to imply that the choice of relevant catchment similarity metrics should be region- or context-dependent.

Acknowledgements

The contributions of many organisations and individuals to the data sources synthesised in this paper are gladly acknowledged. We are particularly grateful to staff at SEPA for river flow data, the James Hutton Institute for soils data, Iain Malcolm at Marine Scotland, and Mark Speed at the University of Aberdeen for tracer data. Markus Hrachowitz (Delft Technical University) carried out GIS analysis and transit time estimates for many sites when working on the Leverhulme project. F/00 152/U: Water Movement in Large River Basins.

References

[1] McDonnell JJ, Sivapalan M, Vaché K, Dunn S, Grant G, Haggerty R, et al. Moving beyond heterogeneity and process complexity: a new vision for watershed hydrology. Water Resour Res 2007;43:W07301. doi:10.1029/2006WR005467.
[2] McDonnell JJ, Woods R. On the need for catchment classification. J Hydrol 2004;299:2–3.
[3] Beven KJ. Uniqueness of place and process representations in hydrological modelling. Hydrol Earth Syst Sci 2000;4:203–13.
[4] Poor CJ, McDonnell JJ, Bolte J. Testing the hydrological landscape unit classification system and other terrain analysis measures for predicting low-flow nitrate and chloride in watersheds. Environ Manage 2008;42:877–93.
[5] Dooge JCI. Looking for hydrologic laws. Water Resour Res 1986;22:S46–58.
[6] Sivapalan M. Pattern, process and function: elements of a unified theory of hydrology at the catchment scale. In: Anderson MG, editor. Encyclopedia of hydrological sciences. Chichester, UK: John Wiley & Sons; 2005. p. 193–220.
[7] Milne G. Some suggested units of classification and mapping, particularly for East African soils. Soil Res 1935;4:183–98.
[8] Horton RE. Erosional development of streams and their drainage basins: hydro-physical approach to quantitative morphology. Geol Soc Am Bull 1945;56:275–370.
[9] Beven KJ, Kirkby MJ. A physically based variable contributive area model of basin hydrology. Hydrol Sci Bull 1979;24:43–69.
[10] Acreman MC, Sinclair CD. Classification of drainage basins according to their physical characteristics: an application for flood frequency analysis in Scotland. J Hydrol 1986;84:365–80.
[11] Burn DH, Boorman DB. Catchment classification applied to the estimation of hydrological parameters at ungauged catchments. IH Report No. 118, Institute of Hydrology, Wallingford; 1992.
[12] Winter TC. The concept of hydrologic landscapes. J Am Water Resour Assoc 2001;37:335–49.
[13] Yadav M, Wagener T, Gupta H. Regionalization of constraints on expected watershed response. Adv Water Resour 2007;30:1756–74.
[14] Hrachowitz M, Soulsby C, Tetzlaff D, Dawson JJC, Malcolm IA. Regionalization of transit time estimates in montane catchments by integrating landscape controls. Water Resour Res 2009;45:W05421. doi:10.1029/2008WR007496.
[15] Carey SK, Tetzlaff D, Seibert J, Soulsby C, Buttle J, Laudon H, et al. Inter-comparison of hydro-climatic regimes across northern catchments: synchronicity, resistance and resilience. Hydrol Proc 2010;24:3591–602.
[16] Wagener T, Sivapalan M, Troch P, Woods R. Catchment classification and hydrologic similarity. Geog Compass 2007;1:901–31.
[17] Sivapalan M, Takeuchi K, Franks SW, Gupta VK, Karambiri H, Lakshmi V, et al. IAHS decade on predictions in ungauged basins (PUB), 2003–2012: shaping an exciting future for the hydrological sciences. Hydrol Sci J 2003;48:857–80.
[18] Sivakumar B. Dominant processes concept, model simplification and classification framework in catchment hydrology. Stoch Environ Res Risk Assess 2008;22:737–48.
[19] Schröder B. Pattern, process, and function in landscape ecology and catchment hydrology – how can quantitative landscape ecology support predictions in ungauged basins? Hydrol Earth Syst Sci 2006;10:967–79.
[20] Buttle JM. Mapping first-order controls on streamflow from drainage basins: the T^3 template. Hydrol Proc 2006;20:3415–22. doi:10.1002/hyp.6519.
[21] Merz R, Blöschl G. Flood frequency regionalisation – spatial proximity vs. catchment attributes. J Hydrol 2005;302:283–306.
[22] Merz R, Blöschl G. A regional analysis of event runoff coefficients with respect to climate and catchment characteristics in Austria. Water Resour Res 2009;45:W01405. doi:10.1029/2008WR007163.
[23] Frey BJ, Dueck D. Clustering by passing messages between data points. Science 2007;315:372–6.
[24] Hrachowitz M, Soulsby C, Tetzlaff D, Speed M. Catchment transit times and landscape controls-does scale matter? Hydrol Proc 2010;24:117–25.
[25] Capell R, Tetzlaff D, Soulsby C, Hartley AL. Linking metrics of hydrological function and transit times to landscape controls in a heterogeneous mesoscale catchment. Hydrol Proc 2011. doi:10.1002/hyp.8139.
[26] Boorman DB, Hollis JM, Lilly A. Hydrology of soil types: a hydrological classification of the soils of the United Kingdom. Institute of Hydrology Report 126, Institute of Hydrology, Wallingford, UK; 1995.
[27] Cardille JA, Lambois M. From the redwood forest to the Gulf Stream waters: human signature nearly ubiquitous in representative US landscapes. Front Ecol Environ 2010;8:130–4. doi:10.1890/080132.
[28] R Development Core Team.: R. A language and environment for statistical computing. Dissertation, R Foundation for Statistical Computing, Vienna, Austria. ISBN: 3-900051-07-0. <http://www.R-project.org/:2010>.
[29] Bodenhofer U, Kothmeier A. An R package for affinity propagation clustering. R package version 1.0.3. Dissertation, Institute of Bioinformatics, Johannes Kepler University, Linz, Austria; 2010.
[30] Cramér H. Mathematical methods of statistics. Princeton: Princeton University Press; 1946.
[31] Hubert L, Arabie P. Comparing partitions. J Classif 1985;2:193–218.
[32] Gordon J, Wignall P. Geology and geomorphology. In: Shaw J, Thompson D, editors. The cairngorms. Edinburgh: Scottish Natural Heritage; 2006. p. 13–43.
[33] Mézard M. Where are the exemplars? Science 2007;315:949–51.
[34] Thavikulwat P. Affinity propagation: a clustering algorithm for computer-assisted business simulations and experimental exercises. Develop Bus Simul Exp Learn 2008;5:220–4.
[35] Castellarin A, Burn DH, Brath A. Assessing the effectiveness of hydrological similarity measures for flood frequency analysis. J Hydrol 2001;241:270–85.
[36] Rao AR, Srinivas VV. Regionalization of watersheds by fuzzy cluster analysis. J Hydrol 2006;318:57–79.
[37] Wolock DM, Winter TC, McMahon G. Delineation and evaluation of hydrologic-landscape regions in the United States using geographic information system tools and multivariate statistical analyses. Environ Manage 2004;34:71–88.
[38] Oudin L, Kay A, Andréassian V, Perrin C. Are seemingly physically similar catchments truly hydrologically similar? Water Resour Res 2010;46:W11558. doi:10.1029/2009WR008887.
[39] Tetzlaff D, Seibert J, Soulsby C. Inter-catchment comparison to assess the influence of topography and soils on catchment transit times in a geomorphic province; the cairngorm mountains. Scotland Hydrol Proc 2009;23:1874–86.