Designation and trend analysis of the updated UK Benchmark Network of river flow stations: the UKBN2 dataset

Shaun Harrigan, Jamie Hannaford, Katie Muchan and Terry J. Marsh

ABSTRACT

Observational trend analysis is fundamental for tracking emerging changes in river flows and placing extreme events in their longer-term historical context, particularly as climate change is expected to intensify the hydrological cycle. However, human disturbance within catchments can introduce artificial changes and confound any underlying climate-driven signal. The UK Benchmark Network (UKBN), designated in the early 2000s, comprised a subset of National River Flow Archive (NRFA) stations that were considered near-natural and thus appropriate for identification and interpretation of climate-driven hydrological trends. Here, the original network was reviewed and updated, resulting in the UKBN2 dataset consisting of 146 near-natural catchments. Additionally, the UKBN2 provides user guidance on the suitability of each station for the assessment of low, medium, and high flows. A trend analysis was performed on the updated UKBN2 dataset and results show that while the strength and direction of changes are dependent on the period of record selected, previously detected patterns of river flow change in the UK remain robust for longer periods (>50 years), despite the recent prevalence of extremes. Such a quality assured observational dataset will provide a foundation for future scientific efforts to better understand the changing nature of the hydrological cycle.

Key words | climate change, droughts, floods, hydrological trends, reference hydrologic networks, UK

Shaun Harrigan
Jamie Hannaford (corresponding author)
Katie Muchan
Terry J. Marsh
Centre for Ecology & Hydrology,
Wallingford, Oxfordshire OX10 8BB,
UK
E-mail: jaha@ceh.ac.uk

INTRODUCTION

It is expected that anthropogenic climate change will intensify the global hydrological cycle as the world continues to warm (IPCC 2013), thereby increasing the frequency and severity of extremes such as floods (Hirabayashi et al. 2013) and droughts (Prudhomme et al. 2014), although strong regional variability and uncertainties in projections exist (Arnell & Gosling 2016). The recent UK Climate Change Risk Assessment report (ASC 2016) identified both increased flooding and water scarcity among the UK's most important climate change risks. The notable hydrological volatility experienced in the early decades of the 21st century (Hannaford 2015; for a fuller description of these episodes see the National River Flow Archive (NRFA) website: http://nrfa.ceh.ac.uk/occasional-reports) has exposed the UK's vulnerability to hydrological extremes and thus there is a clear scientific and socio-economic need to understand the changing nature of these extremes.

There is a growing body of work using large ensemble modelling approaches suggesting extreme hydrological events can be attributable, in part, to a direct human

influence on climate (e.g., Schaller *et al.* 2016). Given inherent uncertainties introduced through the climate-hydrology modelling chain, and the complex and still poorly understood role of catchments in modifying climate signals, observations remain the foundation for any scientific understanding on climate change impacts on river flows, particularly when justifying costly adaptation plans. However, there are also many challenges involved with detection of a robust long-term climate change signal within observed river flow time-series (Hannaford 2015). For example, the climate change signal is much weaker than background natural decadal climate variability (DCV), especially in ocean influenced mid-latitude regions such as the UK (Wilby 2006). Additionally, artificial disturbances within catchments (such as urbanisation, deforestation, dam construction, and river engineering) have been shown to substantially alter flow regimes (e.g., Vorogushyn & Merz 2013; Harrigan *et al.* 2014; Prosdocimi *et al.* 2015) thus confounding trend detection and attribution. Disentangling the many interacting drivers of change in river flows is a major research challenge, but a first step is using river flow data that are sensitive to climate-driven changes.

Reference Hydrologic Networks (RHNs) provide such fit-for-purpose data as only catchments that can be considered 'near-natural' with long and good quality flow records are included. Whitfield *et al.* (2012) and Burn *et al.* (2012) review the development and status of national RHNs with the UK Benchmark Network (UKBN) being one of the most established of those contributing to the global RHN effort. The UKBN comprises a subset of gauging stations within the national hydrometric network that is most suited for identification and interpretation of long-term climate-driven hydrological variability and change. The UKBN is of fundamental importance in this regard, given the high population density and long history of settlement and water exploitation in the UK compared to many other countries; human influences on river flow regimes are pervasive, and in many catchments changes in long-term runoff patterns bear little relation to climate variability (Hannaford & Marsh 2006). Benchmark catchments can be considered reasonably free from human disturbances such as urbanisation, river engineering, and water abstractions, and hence can be used for detection of climate-driven changes in river flow. The first iteration of the UKBN,

henceforth UKBN1, was designated 15 years ago by Bradford & Marsh (2003) and included 122 catchments that met four primary criteria: (i) relatively natural flow regimes, (ii) good and consistent hydrometric data quality, (iii) relatively long records (ideally > 25 years) and (iv) were representative of UK hydroclimatic conditions with good geographical coverage. The core aim of the UKBN is to strengthen national capability to identify and quantify long-term trends and variability in runoff patterns and hydrological extremes. As well as being valuable for the national and international research community, this information is potentially useful for a wide range of practical applications including strategic water resources planning, environmental regulation, flood risk and engineering design, and climate change adaptation planning.

Since the designation of the network, it has been used extensively in trend studies on changes in UK runoff, low flows and droughts, high flows and floods, and seasonal flows (see the review of Hannaford (2015) and references therein). Although designated primarily to support hydrological change detection, RHNs are appropriate for a wide range of applications that require near-natural flow regimes – particularly understanding the climatic processes (Lavers *et al.* 2010, 2015) and catchment properties (Chiverton *et al.* 2015) influencing river flows. The network has also been used for quantifying trends in other variables such as water temperature (Orr *et al.* 2015). The network has fed into various international initiatives, including efforts to quantify river flow trends on the island of Ireland (Murphy *et al.* 2013), European (Stahl *et al.* 2010) and, more recently, intercontinental (Hodgkins *et al.* 2017) scales. UK and European studies using the network have also been cited in global assessments (e.g., IPCC 2013).

Given that 15 years have passed since the original UKBN, there is a clear need to review the stations within the network to ensure they still meet benchmark status and re-evaluate stations that were initially excluded from UKBN1 due to short record lengths. Additionally, Bradford & Marsh (2003) acknowledged in the original benchmark designation that compromises had to be made to ensure good spatial coverage, as relatively few gauged catchments in the UK have near-natural flow regimes and, of these, even fewer are gauged by stations with the ability to measure the full range of flows accurately. This is particularly an

issue in the densely populated areas of central, southern, and eastern England. Recognising this challenge, there is an opportunity for a more explicit classification of benchmark status of individual catchments according to suitability for low, medium, and high flows.

Furthermore, many of the previous UK trend studies reviewed within Hannaford (2015) use data ending in the early- to mid-2000s and so it is not known whether the previous reported patterns of change persist when more up-to-date flow data are used, especially given the prevalence of recent hydrological extremes. With the above in mind, the aim of this paper is two-fold:

1. Review and update the original UKBN of river flow stations, including a new classification of benchmark status at low, medium, and high flows.
2. Perform a standardised trend analysis on the updated UKBN dataset.

Accordingly, the Designation of the UKBN2 dataset section outlines the review process behind the designation of the updated Benchmark Network. The Trend analysis methods section describes the hydrological indicators, catchments, and trend tests used. The trend analysis is presented in the Results section and interpreted in the Discussion section before suggestions for further research are offered in the Concluding remarks section.

DESIGNATION OF THE UKBN2 DATASET

UKBN review process

Stations within the first iteration of the UKBN were selected from the national hydrometric network following a project beginning in the late-1990s. The benchmark review process involved evaluation of detailed station metadata, inspection of hydrographs, and consultation with those responsible for maintenance and collection of hydrometric data who have vital local knowledge of individual site conditions. The designation of the updated UKBN reported here, henceforth UKBN2, used the original benchmark criteria and UKBN1 stations of Bradford & Marsh (2003) as its foundation but employed a more extensive range of metadata now available on the NRFA, exploiting a number of developments over the

last 15 years, including: more detailed station metadata, incorporation of the NRFA Peak Flow database (http://nrfa.ceh.ac.uk/peak-flow-data) which includes gauging station rating curves, improved knowledge of artificial influences (AIs) such as water abstractions and discharges, and increased NRFA spatial and statistical analysis capabilities (Dixon *et al.* 2013). Given the extent and diversity of both qualitative and quantitative information, as well as the need to exercise expert judgement in many places, a completely objective application of the benchmark criteria was not possible, as was the case for the original UKBN1 designation. Nevertheless, the decision-making process was supported by a systematic framework underpinned by key evidence sources (Table 1). In step one of the appraisal, stations contained within UKBN1 were allocated one of three initial categories: endorse, review, or omit based on the review process shown in Table 1(A–E). An additional 54 stations that potentially met benchmark criteria were also considered, as they now have a record length >25 years. Of the original 122 stations, 67 were endorsed, 48 required further review, and seven were omitted, resulting in a total of 176 (including the 54 'Candidate' stations) considered in the overall UKBN2 appraisal.

It was apparent in the early stages of the review that compromises were needed in particular regions to achieve an adequate density of benchmark catchments. This primarily reflects both the ubiquitous nature of AIs on flow regimes, and the inherent difficulties of hydrometric measurement in the extreme flow ranges at many UK gauging stations; very few gauging stations can be considered truly 'full range' (Marsh 2002). For example, at low flows, hydrometric uncertainty arises due to insensitivity of measuring structures, or wide scatter in spot flow measurements (gaugings) used to derive rating curves, e.g., due to summer weed growth. Low flows are also the most heavily impacted by substantial surface and/or groundwater abstractions within the catchment. For high flows, common issues include unmeasured bypass flow and non-modularity (drowning) at gauging structures (Herschy 2008), or simply an insufficiency of gaugings to accurately define the high flow rating curve.

Given these challenges, the original aspiration (Bradford & Marsh 2003) of full-range benchmark catchments was a major constraint on the network. Recognising

Table 1 | Details of systematic benchmark review process with key sources of information consulted and how they were applied to benchmark criteria

Review process/Source of information	Details	How it was applied to benchmark criteria
A General NRFA[a] station metadata[b]	Station description (gauge type, changes in gauging methods/structure over time); Hydrometric description (indicative hydrometric quality of gauge and indication of issues at extreme flow ranges, e.g., high flows bypassing gauging structure); Flow record description (particular measurement issues over time); Flow regime description (highlight AIs that affect runoff); Site photographs (assessment of site conditions, often during past extreme events)	Station failed/given caution if evidence of serious impacts from AIs/performance issues, or, query raised with MAs for further assessment/information
B NRFA Peak Flow[c] data and metadata	Over 85% of UKBN2 stations are also peak flow stations so have access to rating curves and gauging schedule to assess high flow hydrometric performance/issues; AMAX and POT hydrographs were assessed	Station failed/given caution if site had too few gaugings/too much scatter in gaugings at high flow range, or, query raised with MAs for further assessment/information
C Hydrometric Data Quality (HDQ) scores	Hannaford *et al.* (2013b) created a Hydrometric Data Quality (HDQ) score for catchments in England and Wales based on Lamb *et al.* (2003) Gauging Station Data Quality classifications (GSDQs). These metrics reflect hydrometric performance and data quality including modelled impact of AIs on low flows (i.e., impact of known abstractions, discharges and impoundments at $Q95^d$)	Station failed/given caution if substantial evidence of AIs and clear impact on low flow regime, or, query raised with MAs for further assessment/information
D Visual assessment of GDF and peak flow hydrographs	Assessment of GDFs and peak flows for evidence of hydrometric issues (e.g., high flow truncation, artificial patterns during low flows, effect of urbanisation on flashiness, and temporal homogeneity issues)	Station failed/given caution if non-natural flow response or clear temporal homogeneity issues, if supported by metadata, or, query raised with MAs for further assessment/information
E Quantitative assessment of GDFs and peak flow time-series	Statistical tests for screening evidence of gradual (using the Mann–Kendall test and the Theil–Sen approach) and abrupt changes (using the Pettitt test) in river flow time-series for low (minimum flow and Q95), medium ($Q50^d$ and AMF), and high flows ($Q05^d$ and maximum flow)	Station failed/given caution if non-natural flow response or clear temporal homogeneity issues, if supported by metadata, or, query raised with MAs for further assessment/information
F Expert consultation with MAs	Query sheet compiled for each MA region based on questions and issues identified in A–E	Station failed/given caution if query confirmed by MA or new information brought to light during this process
G Synthesis: Identification of *benchmark score* and *benchmark qualifier*	Finally, information from A–F was collated along with maps on catchment representativeness/spatial coverage and reviewed together by the project team exercising expert judgement to arrive at the final selection of 146 stations. A *benchmark score* was given to low, medium, and high flow ranges for each station along with a brief *benchmark qualifier* to explain why *not-suitable* or *caution* flags were warranted[e]	Balanced most natural, best quality records at extremes, longest record length, and hydrological representativeness and spatial coverage. Application of criteria were more strict in regions with many stations and necessarily relaxed, within reason, for regions where few stations met criteria

[a]http://nrfa.ceh.ac.uk/.
[b]http://nrfa.ceh.ac.uk/basic-station-information.
[c]http://nrfa.ceh.ac.uk/peak-flow-data.
[d]Qn is the flow equalled or exceeded n% of the time.
[e]information available for each station within the UKBN2 station list file (http://nrfa.ceh.ac.uk/benchmark-network).
AIs, artificial influences; AMAX, 15-minute annual maximum flows; AMF, annual mean flow; GDFs, gauged daily (mean) flows; MAs, Measuring Authorities; NRFA, National River Flow Archive; POT, 15-minute peaks-over-threshold flows.

this limitation, and the often different uses and user communities for low flow and high flow assessments (e.g., Hannaford *et al.* 2013b), the UKBN2 model advocates a classification system that allows 'sub-networks' to be defined. To facilitate this, and help the user community assess the utility of individual benchmark station records in the presence of these hydrometric challenges, their suitability for analysis at low, medium, and high flow was evaluated.

Any evaluation of the ability of a station to effectively measure extreme flows requires local knowledge of site and catchment conditions. Step two in the benchmark review process (Table 1F) engaged personnel within each of the four UK Measuring Authorities (Environment Agency for England, Natural Resources Wales, the Scottish Environment Protection Agency, and the Rivers Agency for Northern Ireland). A query questionnaire was compiled for each station within the 'review' or 'candidate' benchmark categories in stage one that required deeper expert local knowledge on a site's capability of capturing low and/or high flows.

Bringing knowledge together from steps one and two, the final step (Table 1G) assigned each station a *benchmark score* based on suitability for analysis of low, medium, and high flows (2 = suitable, 1 = caution, and 0 = not-suitable). Thus a station scoring a maximum of 6 means it is suitable for use across the full flow regime. Where a station scores 1 or 0 for a category, a brief *benchmark qualifier* is provided to help end users understand why the time-series might not be suitable for analysis or requires caution, if for example, water abstractions, poor high flow performance/bypassing, or artificial regulation of flows from hydroelectric power schemes were particularly prevalent.

The new UKBN2 dataset

The UKBN2 appraisal identified 146 of the 176 stations under review as qualifying for benchmark status (Figure 1 and Table 2). Of these, 80 are considered benchmark across the full flow regime. However, these full range benchmark stations are distributed mainly in the less densely populated western and upland areas of the UK (Figure 1), leaving some strategically important network gaps in
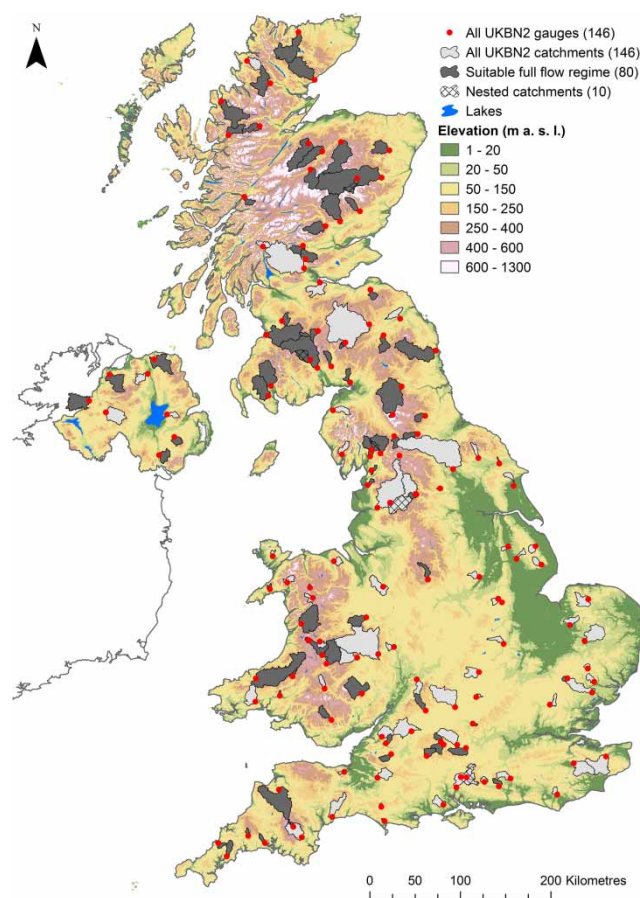

Figure 1 │ Catchment boundaries and gauging station locations for all 146 UKBN2 stations (gauges dots and grey catchments), 80 stations suitable for analysis across the full flow regime (i.e., *benchmark score* = 6) (dark grey catchments), with 10 nested catchments hatched.

central, southern, and eastern England mainly due to the larger impact of water abstractions and discharges. However, adequate spatial coverage is maintained when using 132 stations classified *suitable* or *caution* for low flows and the 133 stations classified as *suitable* or *caution* for high flows. A primary objective of the UKBN2 *benchmark scores* and *benchmark qualifiers* is to guide users to the most appropriate sub-network of stations that meet their

Table 2 │ Number of UKBN2 stations within each benchmark classification

Benchmark score	Low flow	Medium flow	High flow	Full flow regime
2 (Suitable)	112	141	110	80
1 (Caution)	20	5	23	–
0 (Not-suitable)	14	0	13	–

specific study needs, while highlighting where due care must be exercised when interpreting results from stations flagged as *caution* or *not-suitable*.

The UKBN2 catchments are mainly relatively small headwater catchments with a median area of 100 km² (ranging from 3 to 1,500 km²) and median altitude of 182 m a.s.l. (ranging from 20 to 650 m a.s.l.). Over 92% of the catchments can be considered 'essentially rural' in terms of the Flood Estimation Handbook (FEH) degree of urbanisation criteria (i.e., <2.5% of catchment area urbanised (Institute of Hydrology 1999)). The number of UKBN2 stations active in each year is shown in Figure 2. The majority of stations were opened during the 1960s and 1970s with only four stations with data before 1950. The mean record length is 46 years, with a minimum of 21 years and maximum of 85 years. Gauged daily mean flow records have high completeness with a mean per cent missing value of 1.4%. However, five stations have records with >10% missing (the highest has 30% missing as the station was not operational for a 12-year gap), but have strategic value so the decision whether to exclude these in an analysis will depend on the context. Ten catchments have records with some degree of ephemeral behaviour (presence of zero flows) and a further ten are nested within a larger parent catchment (Figure 1). While these catchments are not appropriate for some applications, there is merit in including them in the network as some users may be particularly interested in the differential responses of headwater to lower catchment locations.

Users are directed to the UKBN section of the NRFA website: http://nrfa.ceh.ac.uk/benchmark-network for the vUKBN2.0 station list (includes basic metadata as well as

benchmark *scores* and benchmark *qualifiers* for each station), additional user guidance, instructions for downloading the UKBN2 dataset, and for tracking future updates to the network. A version control system has been implemented to ensure reproducibility of subsequent analyses through time and we envisage that on each major update, a routine trend analysis using the methodology outlined below will be undertaken.

TREND ANALYSIS METHODS

The second aim of this paper is to develop a standardised trend analysis procedure to apply routinely to the Benchmark Network, based on established methods within the hydroclimatic literature, with a first application on the newly designated UKBN2 dataset. Various trend assessment methods have been applied to UKBN1 previously. Here, we set out the following as a rigorous, standardised approach focusing on three components aimed at understanding spatio-temporal changes in river flow:

1. Trend analysis using two fixed periods (short and long) to identify the spatial nature of changes in river flows.
2. Assessment of temporal variability of changes in light of the known influence of DCV.
3. Investigation of persistence of trends for the full available time-series.

Hydrological indicators and catchment selection

For each year, a set of 12 hydrological indicators used in Hannaford & Buys (2012) were extracted from gauged daily flow data (last retrieved from the NRFA on 2nd February 2017) covering the full flow regime:

- annual low flow: Q95, Q90;
- annual medium flow: Q70, Q50 (median), annual mean flow (AMF), Q30;
- annual high flow: Q10, Q05;
- seasonal mean flow: winter (DFJ), spring (MAM), summer (JJA), autumn (SON).

Qn is the flow threshold exceeded n% of the time in each year. We acknowledge that Q95 (and Q90)/Q10 (and

Figure 2 │ Number of UKBN2 stations operational in each calendar year over 1930–2015.

Q05) do not necessarily characterise drought/flood events but are, nonetheless, useful indicators for assessing the tendency for changing extremes based on daily flow data. Applications of trend analysis to peak flow (e.g., AMAX, POT) and drought indicators (e.g., the Standardised Streamflow Index) are different cases with particular requirements (i.e., censored data, high numbers of zeros, high spatial and temporal persistence). These are already the focus of other initiatives and so are not considered part of the standard benchmark trend testing methodology advocated here.

Indices were computed on time-series for the full period of record for each of the 146 UKBN2 gauges. Missing data can lead to spurious values of indicators and hence misleading trends. While gap filling is desirable (Harvey *et al.* 2012) and is part of the NRFA quality control process (Dixon *et al.* 2013), in practice, it has not been extensively carried out for historic time-series. Missing data were handled by applying a strict rule that less than 10% of data could be missing in any year or season for a flow index to be returned, otherwise the particular year/season was given a missing value flag.

It is not appropriate to analyse all stations for all indices given some stations are flagged as *not-suitable* for analysis at particular flow ranges. For stations given a *benchmark score* of 0 (*not-suitable*) for a range, the indicator was excluded from analysis in the remainder of the paper (i.e., no high flow indicators were calculated for stations with a high flows score of 0). In addition to the full period of record analysis, two set periods (short and long) were chosen, optimising spatio-temporal distribution of stations, for a relative comparison. A 30-year (short) period was selected from calendar years 1985–2014 and a 50-year (long) period from calendar years 1965–2014. As missing values can affect the resulting trends in various ways depending on the extent of missing values and position within a series (Slater & Villarini 2017), a further missing data criterion was applied at this stage whereby a maximum of 10% of indicator values could be missing in either fixed period (i.e., five (three) years for the long (short) period). To allow for as many stations as possible to be included in both periods, particularly in the long period where data are more sparse, stations with start years within two years of the target 1965 and 1985 start years, and/or within one year of the target 2014 end year, were also included in the analysis. However, the combined number of missing values in a series and number of years from relaxed start and end years did not exceed a total 10% limit of the respective long and short period length.

Finally, to avoid potentially 'double counting' statistically significant trends, only non-nested catchments were used. In cases where the larger parent catchment is flagged as *not-suitable* for analysis of either high or low flows, but the nested catchment was, the nested catchment was used instead. Application of the above criteria resulted in 116 (short period) and 42 (long period) stations for low flow, 125 (short period) and 46 (long period) for medium flow, and 113 (short period) and 43 (long period) for high flow trend analysis.

Trend analysis tests

Evidence for monotonic trends was assessed using the Mann–Kendall (MK) test (Mann 1945; Kendall 1975), a non-parametric rank-based method that is widely applied in analyses of streamflow (e.g., Hannaford & Marsh 2008; Villarini *et al.* 2011; Murphy *et al.* 2013). The standardised MK statistic (MKZs) follows the standard normal distribution with a mean of zero and variance of one. A positive (negative) value of MKZs indicates an increasing (decreasing) trend. Statistical significance was evaluated with probability of Type 1 error set at the 5% significance level. A two-tailed MK test was chosen, hence the null hypothesis of no trend (increasing or decreasing) is rejected when |MKZs| >1.96 using traditional statistical testing.

The MK test requires data to be independent (i.e., free from serial correlation or temporal autocorrelation) as positive serial correlation increases the likelihood of Type 1 errors or incorrect rejection of a true null hypothesis (Kulkarni & von Storch 1995). All indicators were checked for positive lag-1 serial correlation at the 5% level using the autocorrelation function (ACF) on detrended series. The linear trend β used to detrend the original time-series was estimated using the robust Theil–Sen approach (TSA) (Theil 1950; Sen 1968). Block bootstrapping (BBS) was used to overcome the presence of serial correlation and involves application of the MKZs statistic to block resampled series that preserve any short-term autocorrelation structure. Following guidance from Önöz & Bayazit (2012) regarding the optimal block length given the sample

size and magnitude of temporal autocorrelation coefficient, a block length $L = 4$ was chosen and applied only when a series had statistically significant serial correlation. A robust estimate of the significance of the MKZs statistic was generated from a distribution of 10,000 resamples where the null hypothesis of no trend is rejected when MKZs calculated from original data is higher than the 9,750th largest (statistically significant increasing trend) or lower than the 250th smallest (statistically significant decreasing trend) MKZs value from the resampled distribution under a two-tailed test at the 5% level (Murphy *et al.* 2013). Results are presented in Tables 3 and 4 for

both traditional and BBS MK tests to highlight the impact serial correlation plays, if any, on the statistical significance of trend results. Note that the BBS column in Tables 3 and 4 counts both statistically significant results from the traditional MK test (for non-significant serially correlated series) and for significantly serially correlated series using BBS with $L = 4$ and is also used for reporting statistically significant trends in the maps in Figures 3 and 4.

There is much debate in the field of hydroclimatology around trend significance testing such as the existence of long-term persistence, which could introduce a statistically significant trend when none is present (Cohn & Lins 2005;

Table 3 │ Direction of change and proportion (% of N stations) of statistically significant (5% level) trends for 30-year short fixed period (1985–2014) for annual and seasonal indices

	Indicator	N	Increasing (sig.; BBS sig.) %	Decreasing (sig.; BBS sig.) %	Magnitude (± bounds) %	Sig. serial correlation %
Low	Q95	116	71.6 (6.9; 5.2)	27.6 (0.9; 0.9)	11.2 (−2.6, 27.6)	12.9
	Q90	116	75.0 (4.3; 2.6)	24.1 (0.9; 0.9)	13.8 (−0.2, 23.5)	8.6
Medium	Q70	125	72.8 (3.2; 2.4)	27.2 (0.0; 0.0)	11.2 (−2.5, 22.3)	9.6
	Q50	125	71.2 (1.6; 1.6)	28.8 (0.0; 0.0)	9.0 (−1.8, 21.5)	10.4
	AMF	125	83.2 (8.0; 7.2)	16.8 (0.0; 0.0)	10.7 (1.8, 20.4)	14.4
	Q30	125	74.4 (4.0; 3.2)	25.6 (0.0; 0.0)	10.0 (−0.4, 21.8)	14.4
High	Q10	113	77.0 (7.1; 7.1)	23.0 (0.0; 0.0)	11.1 (0.0, 28.6)	6.2
	Q05	113	79.6 (5.3; 5.3)	20.4 (0.0; 0.0)	13.0 (3.3, 24.9)	0.9
Season	DFJ	125	88.8 (8.0; 8.0)	11.2 (0.0; 0.0)	14.3 (5.5, 25.1)	0.0
	MAM	125	20.8 (0.0; 0.0)	79.2 (16.0; 16.0)	−20.1 (−33, −2.9)	4.8
	JJA	125	75.2 (1.6; 1.6)	24.8 (0.0; 0.0)	13.3 (−0.1, 22.2)	1.6
	SON	125	92.0 (4.0; 4.0)	8.0 (0.0; 0.0)	23.2 (11.6, 33.6)	0.0

Direction and significance from Mann–Kendall (MKZs) and magnitude calculated with the relative Theil–Sen approach (TSA$_{rel}$). Magnitude of change is based on the median TSA$_{rel}$ with spread (± bounds) given by interquartile range. Per cent of stations statistically significant using block-bootstrapping (BBS) are also shown along with the proportion of series with statistically significant serial correlation.

Table 4 │ As for Table 3 but for 50-year long fixed period (1965–2014)

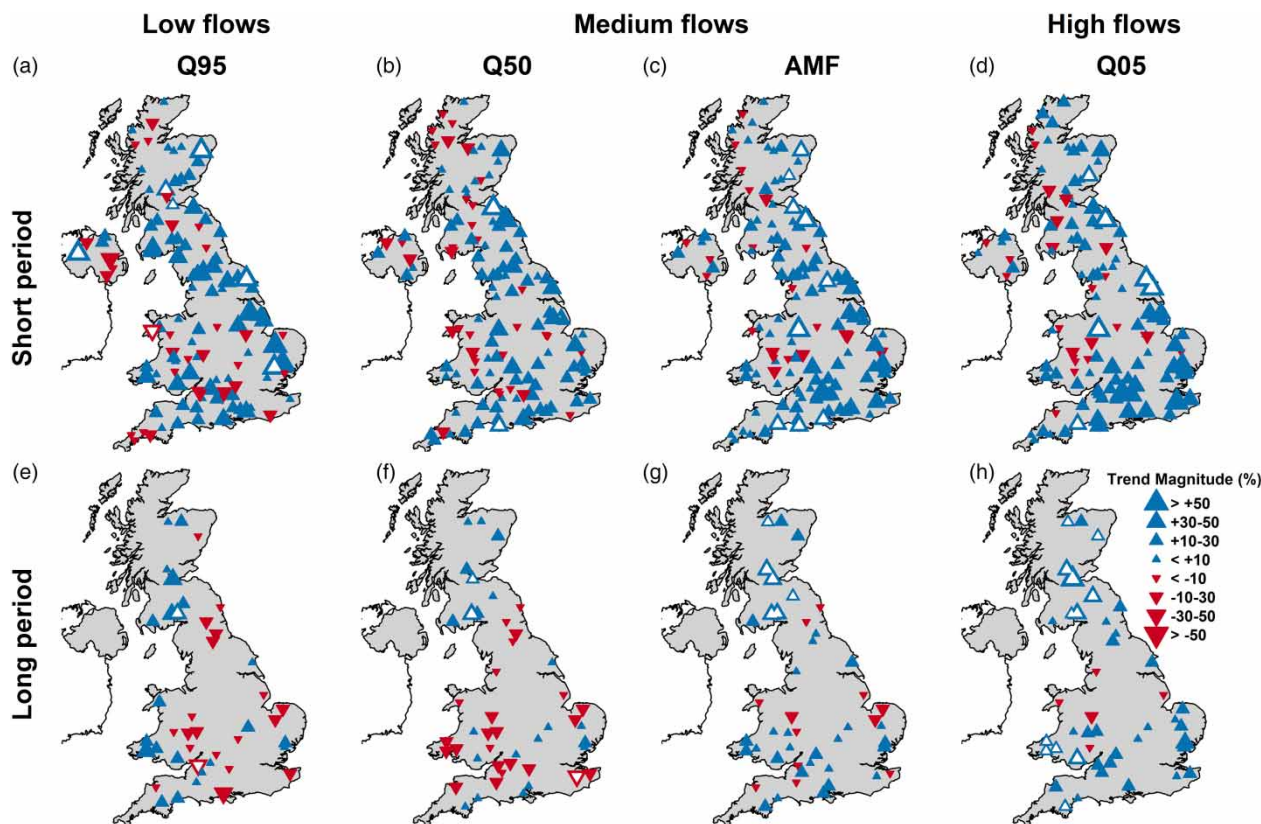|	Indicator	N	Increasing (sig.; BBS sig.) %	Decreasing (sig.; BBS sig.) %	Magnitude (± bounds) %	Sig. serial correlation %
Low	Q95	42	52.4 (4.8; 2.4)	47.6 (2.4; 2.4)	1.2 (−9.4, 14.8)	31.0
	Q90	42	52.4 (2.4; 0.0)	47.6 (0.0; 0.0)	1.7 (−8.6, 15.5)	28.6
Medium	Q70	46	63.0 (8.7; 6.5)	37.0 (0.0; 0.0)	1.5 (−8.5, 13.5)	32.6
	Q50	46	47.8 (4.3; 4.3)	52.2 (2.2; 2.2)	−0.9 (−12.5, 6.2)	26.1
	AMF	46	73.9 (13.0; 13.0)	26.1 (0.0; 0.0)	6.8 (0.1, 13.8)	21.7
	Q30	46	56.5 (10.9; 8.7)	43.5 (0.0; 0.0)	1.8 (−7.6, 10.2)	19.6
High	Q10	43	86.0 (16.3; 16.3)	14.0 (0.0; 0.0)	11.3 (5.2, 21.1)	14.0
	Q05	43	88.4 (27.9; 27.9)	11.6 (0.0; 0.0)	13.5 (7.9, 23.8)	11.6
Season	DFJ	46	87.0 (19.6; 19.6)	13.0 (0.0; 0.0)	12.7 (4.1, 26.2)	0.0
	MAM	46	30.4 (0.0; 0.0)	69.6 (0.0; 0.0)	−10.7 (−19.0, 1.6)	0.0
	JJA	46	54.3 (2.2; 2.2)	45.7 (2.2; 2.2)	1.4 (−11.2, 18.5)	4.3
	SON	46	82.6 (4.3; 4.3)	17.4 (0.0; 0.0)	16.7 (5.5, 25.5)	0.0

Figure 3 | Magnitude and direction of trends for the short fixed period (1985–2014) (top row) and long fixed period (1965–2014) (bottom row) for selected low, medium, and high flow indices. Upward-pointing triangles represent increasing trends and downward-pointing decreasing trends, with magnitude proportional to size. Trend magnitude is based on TSA$_{rel}$. Block-bootstrapped significant trends (5% level) shown by white triangles and derived from MKZs.

Koutsoyiannis & Montanari 2007), or the exaggerated importance of use of the 5% significance level (Clarke 2010). As the focus here is on the direction and strength of changes and not entirely on statistical significance relative to arbitrary *p*-value thresholds (Nicholls 2001), trend magnitudes were also estimated from the TSA in order to corroborate and map the strength and regional coherence of trends. To facilitate a relative comparison between sites, the trend magnitude TSA$_{rel}$ (%) for each time-series was expressed as a percentage of the long-term mean flow μ over the period of record of *n* years where β is the TSA slope, given by Stahl *et al.* (2012) as:

$$\text{TSA}_{\text{rel}} = \left(\frac{\beta \times n}{\mu}\right) \times 100$$

Hannaford & Buys (2012) found this approach preferable compared to expressing trend magnitude as a simple percentage change over the full record, which can yield larger

changes in the presence of anomalously large start or end values. Trend magnitude, serial correlation, and standard and block-bootstrapped (for serially correlated series) MK tests were calculated for all 12 hydrological indicators across each station for both the short and long fixed periods.

RESULTS

Fixed period trends

In low, medium, and high flow indices for the 1985–2014 short period, positive trends are prominent (Table 3). Over 70% of stations report an increasing trend in 11 of the 12 indices, with spring (MAM) mean flow the exception, showing strong and statistically significant decreases (16% of stations under BBS). Eight per cent of stations show statistically significant increases in winter (DJF) mean flow
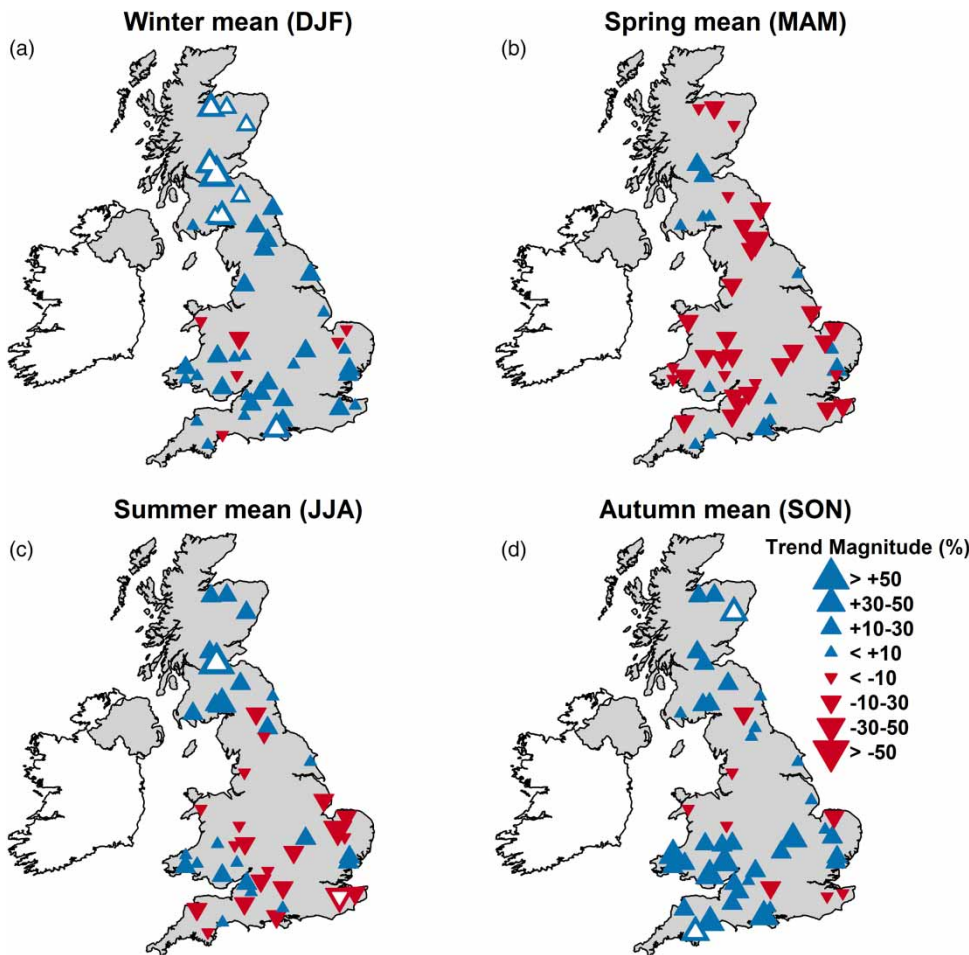
Figure 4 │ As for Figure 3 but for seasonal indices for the long period only.

resulting in a median trend magnitude of +14.3% (+5.5%, +25.1%) across the network. For the 1965–2014 long period, increasing trends continue to dominate the majority of medium and high flow indices (Table 4). Similarly to the short period, almost 70% of stations show a decrease in MAM, and while none are statistically significant the UK-wide trend magnitude is −10.7% (−19.0%, +1.6%). The number of stations with increasing and decreasing trends is more even for low flows (Q95, Q90), as well as Q70, Q50, Q30 and JJA; thus, overall increasing and decreasing trend magnitudes tend to cancel each other out resulting in UK-wide median trend magnitudes in low flow indices of just +1 to 2% over 1965–2014. However, statistically significant increases are found in AMF, Q30, Q10, Q05 and DJF, ranging from 8.7% of stations (Q30) to the highest 27.9% of stations (Q05) under BBS, with median trend

magnitudes ranging from +1.8% (−7.6%, +10.2%) to +13.5% (+7.9%, +23.8%), respectively.

Trends are mapped for selected low (Q95), medium (Q50, AMF), and high flow (Q05) indices for both fixed periods in Figure 3. Spatial patterns of trends in the short period (top row) show a spatially consistent increase across the UK, although few of these are statistically significant. On the other hand, for the long period, low flow (Q95) and median flow (Q50) trends (bottom row) show a marked spatial gradient with increases in the north and west and decreases in the south and east of Britain. AMF appears to follow a similar pattern to that of high flows (Q05), but with fewer statistically significant trends, with strongest increasing trends for catchments in Scotland over 1965–2014. Patterns in long period seasonal mean trends (Figure 4) for summer (JJA) reflect the north-west/south-east gradient

found in low lows (Q95), whereas decreasing trends in spring (MAM) flows occur across the majority of Britain. The wetter winter (DFJ) and autumn (SON) seasons follow broadly the pattern of observed strong increasing trends in AMF and Q05, particularly in Scotland. Overall, the pattern of changes found in AMF are clearly biased towards patterns in wetter seasons and high flows.

It was found that accounting for serial correlation was important. Almost all indices in both short and long periods had stations with statistically significant serially correlated series (at the 5% level). These were most prominent in low flow indices (~30% of stations in the long period for Q95). There are several cases where the number of statistically significant increasing/decreasing trends was reduced when block-bootstrapping was applied to serially correlated series (Tables 3 and 4).

Temporal variability analysis and persistence of trends

While it is necessary to analyse trends using fixed periods for a relative comparison of direction, magnitude, and spatial patterns, these are just snapshots of the temporal evolution of changes over time – as demonstrated by the marked differences in trends between the two fixed periods (i.e., Figure 3). Apparent from the standardised and smoothed series in Figure 5 (left column) is widespread consistency of decadal scale variability across the flow regime from Q95 (top left) to Q05 (bottom left). There is a marked transition from low to high flows in the 1970s as well as an increase in flows in the early 2000s. These have consequences in terms of placing results from the two fixed periods in context of the overall variability (the start years of the two fixed periods are marked as vertical dotted black lines). The dependency of trends on period of record is captured in the trend persistence analysis (Figure 5, right column) whereby MKZs values for series are highly variable through time. For low flows (Q95), trends with start years in the early 1970s result in strong increases, but stations with longer records show this is an artefact of the period used and instead longer-term trends are not increasing strongly in low flows. This is in contrast to high flows (Q05) where longer records tend to show stronger increasing trends.

DISCUSSION

Our results from the trend analysis of the updated UK Benchmark Network (UKBN2) using recent UK-wide flow records show no fundamental discrepancies with previously published flow patterns in annual low, medium, and high, or seasonal mean flow indices, despite the prevalence of notable hydrological extremes in the most recent decade. Overall indicators of water availability (AMF, Q50, and seasonal mean flows) for benchmark stations are consistent with Hannaford & Marsh (2006) and the UK national outflow series (Marsh & Dixon 2012). AMF has increased across the UK but mostly in Scotland, and follows a similar pattern to winter and autumn mean flow. There is a marked spatial north-west to south-east gradient for summer trends (Figure 4(c)) with much of England showing decreases and increases in the north-west. The decreasing trend in spring found by Hannaford & Buys (2012) for Britain and Murphy *et al.* (2013) for the Island of Ireland was also found here and we echo calls for improved understanding of the drivers of these seasonal changes in river flow, which may have important implications for water management and ecology. Generally, the results reinforce earlier findings but strengthen them given the use of a more rigorous, updated Benchmark Network.

One of the most societally relevant impacts of climate change is the expected increase in flooding due to increased precipitation intensity in a warming climate. Compelling evidence from the literature is emerging for flood-related variables in maritime-influenced upland areas in the north and west, including detected increases in observed winter precipitation (Dadson *et al.* 2017), extreme precipitation (Jones *et al.* 2014), and high flow and flood indices (Hannaford & Marsh 2008). While it is acknowledged the high flow indices used here do not explicitly characterise flooding, detected changes support the conclusion of a tendency for an increase in high flows over the past 50 years, and this signal is robust when longer records were considered (Figure 5(h)). Nevertheless, there is also remarkable evidence of DCV in the flow series. Trends are not part of a simple linear increase, but form a multitude of flood-rich and flood-poor episodes over time. The increase in high flows in the past decade or so (Figure 5(g)) appears to be
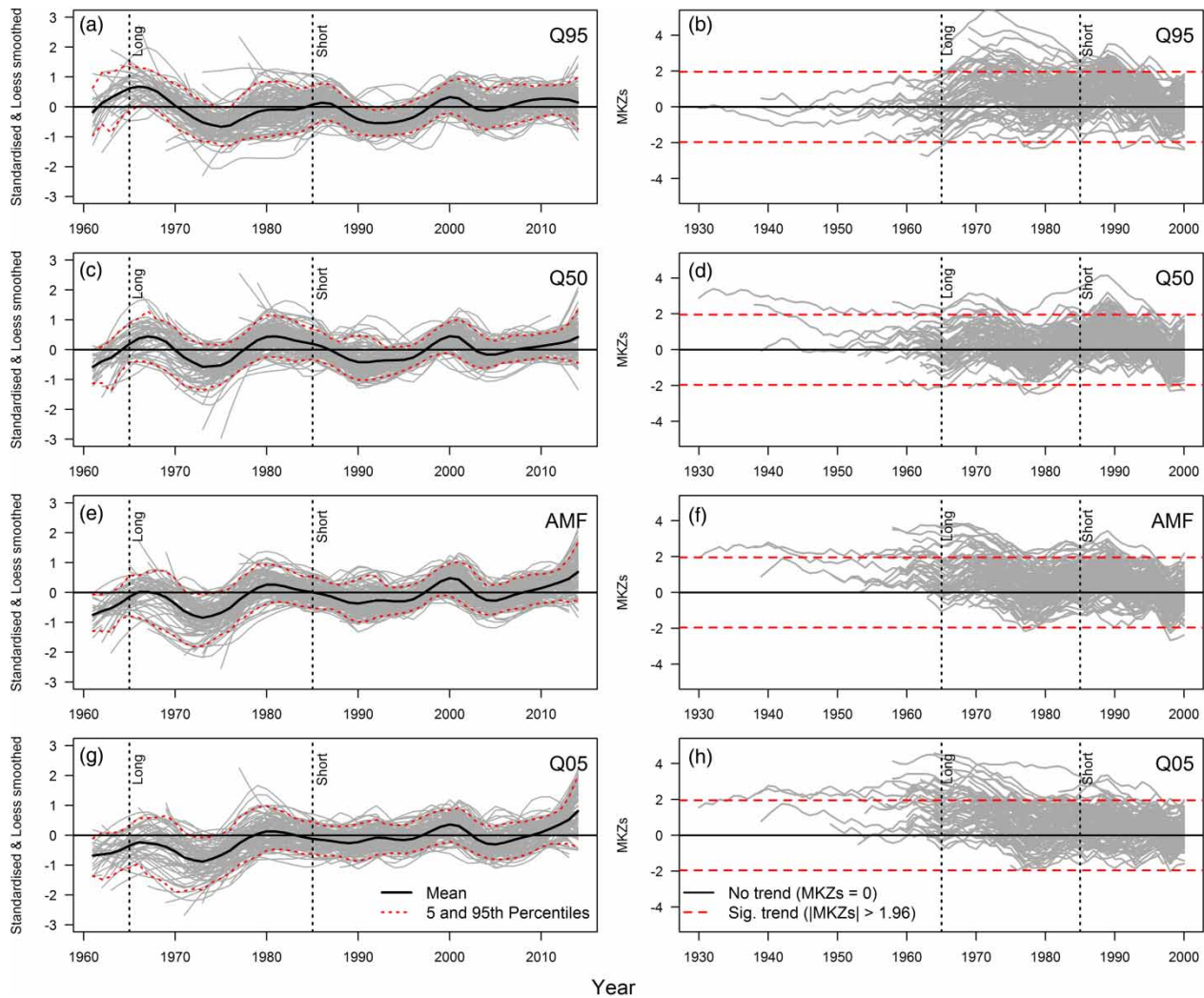
Figure 5 | Left panel shows standardised (by mean and standard deviation calculated over common 1985–2014 base period) and smoothed (by locally weighted regression smoothing (Loess) using a span =15 years) series (between 1961 and 2014) following Hannaford *et al.* (2013a) for low (Q95), medium (Q50 and AMF), and high flow (Q05) indices defined in the Hydrological indicators and catchment selection section. Solid grey lines are individual stations; dotted lines are the 5th and 95th percentiles across all stations, and solid black line the mean (1961 deemed earliest year to reliably calculate percentiles). Right panel shows persistence of trends in the same indices alternating all available start years between 1930 and 2014, 1931 and 2014, and so on up to 2000–2014 using MKZs following Wilby (2006). Solid grey lines represent MKZs statistics for varying start years for individual stations across the UK. Dashed horizontal lines are the threshold for statistically significant trends at the 5% level with MKZs values above (below) these indicating statistically significant increasing (decreasing) trends since the corresponding start year. In both panels, black dotted vertical lines mark start years of the two fixed periods used in the present analysis for context.

part of a flood-rich period from the late-1990s onward (see Wilby & Quinn 2013), largely driven by decadal-scale clustering of flood-generating cyclonic and westerly weather types, which have been linked mostly to changes in the North Atlantic Oscillation (NAO) (Hannaford & Marsh 2008; Svensson *et al.* 2015). Understanding the drivers and evolution of these periods of flood propensity should be a research priority, especially quantifying the role climate change might play in altering the dynamics as well as

interactions with catchment properties; the UKBN2 provides a climate sensitive dataset to contribute to this and results from such analyses as performed here will help inform current science policy-making discourse (e.g., Dadson *et al.* 2017).

Low flows show few statistically significant decreasing trends, as found in previous studies (Hannaford & Marsh 2006; Hannaford & Buys 2012). However, while not *statistically* significant, decreasing 50-year trend magnitudes in the

English lowlands are in the −10–30% range for several catchments (Figure 3(e)), especially for summer (Figure 4(c)), and might be important for water management. Wilby (2006) showed the signal-to-noise ratio for basins in the UK is low, particularly in summer, and that robust statistically detectable trends are not expected for several decades yet. This is further highlighted in Figure 5(a) with strong evidence of DCV, and hence trends are sensitive to the period of record analysed (Figure 5(b)). This is most prevalent for records beginning in the 1960s and 1970s, which is the case for the majority of UK trend studies as hydrometric network expansion coincided with a period of a particularly high degree of natural variability.

In addition to the previous limitations, few catchments in the densely populated region of southern and eastern England can be considered pristine in the strictest sense, so caution must be exercised in interpreting changes in low flows in this region. Nonetheless, the consistent temporal and spatial pattern across low, medium, and high flow indices (Figure 5, left column) is encouraging and suggests that even in the English lowlands river flows are generally reflecting changes driven by climate, rather than from artificial sources (e.g., from groundwater and/or surface water abstraction) which, while controlled as far as possible in the benchmark designation, cannot be ruled out in the catchments flagged as 'caution'. However, it is challenging based on these results alone to provide clear guidance regarding potential long-term implications for water resources management, so future work that combines innovative observational and modelling approaches using several lines of hydroclimatic inquiry is still needed. It is also noted that the majority of studies examine changes in low flows, rather than actual drought 'events', and so such event-based analyses should be another research priority.

While RHNs are vital in hydroclimatology, there are growing calls for the need to also improve our understanding of how the hydrological cycle is responding to rapidly changing human systems (Montanari *et al.* 2013; Van Loon *et al.* 2016). By the very nature of RHNs, such impacts are removed or controlled, by definition. This can be seen as an inherent limitation of RHNs, of which end users must be aware when designing their analyses: RHNs typically quantify changes in small, headwater catchments away from the downstream population centres that are most

likely to be affected, socio-economically, by any changes in hydrological extremes. For example, there are no UKBN2 catchments >1,500 km^2 so any study, including the trend analysis here, will be biased towards medium and small catchments, particularly in the south and east of England as abstractions and discharges are less prevalent in headwater catchments. There is also a dearth of very small catchments; only four catchments within the UKBN2 dataset have areas <10 km^2 and only one of those can be considered upland (elevation >300 m a.s.l.). Therefore, processes operating only at these scales would not be captured. On the other hand, RHNs can provide a near-natural baseline for comparing with human-influenced sites (e.g., using paired 'impacted' catchments as in Prosdocimi *et al.* 2015) or for modelling studies, so can play a vital role even in efforts to quantify human disturbances on the hydrological cycle.

The second iteration of the UKBN has made several improvements since UKBN1, but there are many potential further improvements that could be made to future iterations of the dataset and to how users access it. For example, we anticipate future analytical efforts will undertake comprehensive homogeneity testing and infilling, while a particular focus will be efforts to improve the assessments of AIs. One of the most challenging aspects of the UKBN update was the fragmented quality and availability of information on AIs, especially access to water abstractions and discharges. While some datasets were consulted (e.g., Hannaford *et al.* 2013b), information is typically based on model estimates of impacts, and not available widely across the UK. Hence, *benchmark qualifiers* in UKBN2 are necessarily brief and qualitative. Finally, we hope to improve access to benchmark data and analyses through an online NRFA trend facility – following the example of the Australian Bureau of Meteorology's data and trend explorer (Zhang *et al.* 2016).

CONCLUDING REMARKS

The first designation of the UKBN has proven a valuable dataset that has fed into many national and international scientific studies, several of which are relied upon for making policy and water management decisions on future

flood design and long-term drought planning. Results from the trend analysis of the updated UKBN2 have reinforced previous findings. We recognise the UKBN will always remain a work-in-progress as new information about gauging stations and the catchments they drain comes to light, or new techniques for assessing benchmark suitability developed. A benchmark version control system has been instigated to ensure minor and major network changes are recorded in a transparent way, the datasets are easily accessible, and studies using previous versions reproducible. Further information about the UKBN2 and how to access the data can be found here: http://nrfa.ceh.ac.uk/benchmark-network.

A community effort involving both those who collect the data (Measuring Authorities) and those who use it (e.g., researchers and practitioners), would make the process of UKBN evolution and updating more efficient and comprehensive. We hope by releasing the UKBN2 we present an opportunity for the hydrological community to provide ideas, novel methods, and feedback on the current version. The metadata holdings of the NRFA and the knowledge of NRFA and Measuring Authority experts are only one set of performance criteria; there is no doubt a wealth of other local knowledge, and a wide range of initiatives generating useful information about these catchments and the gauging stations that monitor them (e.g., ongoing studies of rating uncertainty, e.g., Coxon *et al.* (2015); national-scale modelling studies that could potentially provide naturalised data estimates and degrees of influences, e.g., Rudd *et al.* (2017)). We therefore invite users to provide information on these catchments, or others that may be candidate benchmark catchments, via contacting the NRFA (nrfa@ceh.ac.uk). We are also interested to understand the range of uses of the network, and invite users to engage with the NRFA team about current and future applications of the dataset.

ACKNOWLEDGEMENTS

REFERENCES

Arnell, N. W. & Gosling, S. N. 2016 The impacts of climate change on river flood risk at the global scale. *Climatic Change* **134** (3), 387–401.

ASC 2016 *UK Climate Change Risk Assessment 2017 Synthesis Report: Priorities for the Next Five Years*. Adaptation Sub-Committee of the Committee on Climate Change, London.

Bradford, R. B. & Marsh, T. J. 2003 Defining a network of benchmark catchments for the UK. *Proceedings of the Institution of Civil Engineers – Water and Maritime Engineering* **156** (2), 109–116.

Burn, D. H., Hannaford, J., Hodgkins, G. A., Whitfield, P. H., Thorne, R. & Marsh, T. 2012 Reference hydrologic networks II. using reference hydrologic networks to assess climate-driven changes in streamflow. *Hydrological Sciences Journal* **57** (8), 1580–1593.

Chiverton, A., Hannaford, J., Holman, I., Corstanje, R., Prudhomme, C., Bloomfield, J. & Hess, T. M. 2015 Which catchment characteristics control the temporal dependence structure of daily river flows? *Hydrological Processes* **29** (6), 1353–1369.

Clarke, R. T. 2010 On the (mis)use of statistical methods in hydro-climatological research. *Hydrological Sciences Journal* **55** (2), 139–144.

Cohn, T. A. & Lins, H. F. 2005 Nature's style: naturally trendy. *Geophysical Research Letters* **32** (23), L23402.

Coxon, G., Freer, J., Westerberg, I. K., Wagener, T., Woods, R. & Smith, P. J. 2015 A novel framework for discharge uncertainty quantification applied to 500 UK gauging stations. *Water Resources Research* **51** (7), 5531–5546.

Dadson, S. J., Hall, J. W., Murgatroyd, A., Acreman, M., Bates, P., Beven, K., Heathwaite, L., Holden, J., Holman, I. P., Lane, S. N., O'Connell, E., Penning-Rowsell, E., Reynard, N., Sear, D., Thorne, C. & Wilby, R. 2017 A restatement of the natural science evidence concerning catchment-based 'natural' flood management in the UK. *Proceedings of the Royal Society A* **473** (2199), 20160706.

Dixon, H., Hannaford, J. & Fry, M. J. 2013 The effective management of national hydrometric data: experiences from the United Kingdom. *Hydrological Sciences Journal* **58** (7), 1383–1399.

Hannaford, J. 2015 Climate-driven changes in UK river flows: a review of the evidence. *Progress in Physical Geography* **39** (1), 29–48.

Hannaford, J. & Buys, G. 2012 Trends in seasonal river flow regimes in the UK. *Journal of Hydrology* **475**, 158–174.

Hannaford, J. & Marsh, T. J. 2006 An assessment of trends in UK runoff and low flows using a network of undisturbed catchments. *International Journal of Climatology* **26** (9), 1237–1253.

Hannaford, J. & Marsh, T. J. 2008 High-flow and flood trends in a network of undisturbed catchments in the UK. *International Journal of Climatology* **28** (10), 1325–1338.

Hannaford, J., Buys, G., Stahl, K. & Tallaksen, L. M. 2013a The influence of decadal-scale variability on trends in long European streamflow records. *Hydrology Earth System Sciences* **17** (7), 2717–2733.

Hannaford, J., Holmes, M. G. R., Laizé, C. L. R., Marsh, T. J. & Young, A. R. 2013b Evaluating hydrometric networks for prediction in ungauged basins: a new methodology and its application to England and Wales. *Hydrology Research* **44** (3), 401–418.

Harrigan, S., Murphy, C., Hall, J., Wilby, R. L. & Sweeney, J. 2014 Attribution of detected changes in streamflow using multiple working hypotheses. *Hydrology Earth System Sciences* **18** (5), 1935–1952.

Harvey, C. L., Dixon, H. & Hannaford, J. 2012 An appraisal of the performance of data-infilling methods for application to daily mean river flow records in the UK. *Hydrology Research* **43** (5), 618–636.

Herschy, R. W. 2008 *Streamflow Measurement*, 3rd edn. Taylor & Francis, Abingdon, UK.

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H. & Kanae, S. 2013 Global flood risk under climate change. *Nature Climate Change* **3** (9), 816–821.

Hodgkins, G. A., Whitfield, P. H., Burn, D. H., Hannaford, J., Renard, B., Stahl, K., Fleig, A. K., Madsen, H., Mediero, L., Korhonen, J., Murphy, C. & Wilson, D. 2017 Climate-driven variability in the occurrence of major floods across North America and Europe. *Journal of Hydrology* **552**, 704–717.

Institute of Hydrology 1999 *Flood Estimation Handbook* (five volumes). Centre for Ecology & Hydrology, Wallingford, UK.

IPCC 2013 *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, NY, USA.

Jones, M. R., Blenkinsop, S., Fowler, H. J. & Kilsby, C. G. 2014 Objective classification of extreme rainfall regions for the UK and updated estimates of trends in regional extreme rainfall. *International Journal of Climatology* **34** (3), 751–765.

Kendall, M. G. 1975 *Rank Correlation Methods*, 4th edn. Charles Griffin, London, UK.

Koutsoyiannis, D. & Montanari, A. 2007 Statistical analysis of hydroclimatic time series: uncertainty and insights. *Water Resources Research* **43** (5), W05429.

Kulkarni, A. & von Storch, H. 1995 Monte Carlo experiments on the effect of serial correlation on the Mann-Kendall test of trend. *Meteorologische Zeitschrift* **4** (2), 82–85.

Lamb, R., Zaidman, M. D., Archer, D. R., Marsh, T. J. & Lees, M. L. 2003 *River Gauging Station Data Quality Classification (GSDQ), R&D Technical Report W6-058/TR*. Environment Agency, p. 136.

Lavers, D., Prudhomme, C. & Hannah, D. M. 2010 Large-scale climate, precipitation and British river flows: identifying hydroclimatological connections and dynamics. *Journal of Hydrology* **395** (3–4), 242–255.

Lavers, D. A., Hannah, D. M. & Bradley, C. 2015 Connecting large-scale atmospheric circulation, river flow and groundwater levels in a chalk catchment in southern England. *Journal of Hydrology* **523**, 179–189.

Mann, H. B. 1945 Nonparametric tests against trend. *Econometrica* **13** (3), 245–259.

Marsh, T. J. 2002 Capitalising on river flow data to meet changing national needs – a UK perspective. *Flow Measurement and Instrumentation* **13** (5–6), 291–298.

Marsh, T. J. & Dixon, H. 2012 The UK water balance: How much has it changed in a warming world? In: *Proceedings of the Eleventh National BHS Symposium*, British Hydrological Society, Dundee, UK.

Montanari, A., Young, G., Savenije, H. H. G., Hughes, D., Wagener, T., Ren, L. L., Koutsoyiannis, D., Cudennec, C., Toth, E., Grimaldi, S., Blöschl, G., Sivapalan, M., Beven, K., Gupta, H., Hipsey, M., Schaefli, B., Arheimer, B., Boegh, E., Schymanski, S. J., Di Baldassarre, G., Yu, B., Hubert, P., Huang, Y., Schumann, A., Post, D. A., Srinivasan, V., Harman, C., Thompson, S., Rogger, M., Viglione, A., McMillan, H., Characklis, G., Pang, Z. & Belyaev, V. 2013 'Panta Rhei–everything flows': change in hydrology and society – the IAHS scientific decade 2013–2022. *Hydrological Sciences Journal* **58** (6), 1256–1275.

Murphy, C., Harrigan, S., Hall, J. & Wilby, R. L. 2013 Climate-driven trends in mean and high flows from a network of reference stations in Ireland. *Hydrological Sciences Journal* **58** (4), 755–772.

Nicholls, N. 2001 Commentary and analysis: the insignificance of significance testing. *Bulletin of the American Meteorological Society* **82** (5), 981–986.

Önöz, B. & Bayazit, M. 2012 Block bootstrap for Mann–Kendall trend test of serially dependent data. *Hydrological Processes* **26** (23), 3552–3560.

Orr, H. G., Simpson, G. L., des Clers, S., Watts, G., Hughes, M., Hannaford, J., Dunbar, M. J., Laizé, C. L. R., Wilby, R. L., Battarbee, R. W. & Evans, R. 2015 Detecting changing river temperatures in England and Wales. *Hydrological Processes* **29** (5), 752–766.

Prosdocimi, I., Kjeldsen, T. R. & Miller, J. D. 2015 Detection and attribution of urbanization effect on flood extremes using nonstationary flood-frequency models. *Water Resources Research* **51** (6), 4244–4262.

Prudhomme, C., Giuntoli, I., Robinson, E. L., Clark, D. B., Arnell, N. W., Dankers, R., Fekete, B. M., Franssen, W., Gerten, D., Gosling, S. N., Hagemann, S., Hannah, D. M., Kim, H., Masaki, Y., Satoh, Y., Stacke, T., Wada, Y. & Wisser, D. 2014 Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment. *Proceedings of the National Academy of Sciences* 111 (9), 3262–3267.

Rudd, A. C., Bell, V. A. & Kay, A. L. 2017 National-scale analysis of simulated hydrological droughts (1891–2015). *Journal of Hydrology* 550, 368–385.

Schaller, N., Kay, A. L., Lamb, R., Massey, N. R., van Oldenborgh, G. J., Otto, F. E. L., Sparrow, S. N., Vautard, R., Yiou, P., Ashpole, I., Bowery, A., Crooks, S. M., Haustein, K., Huntingford, C., Ingram, W. J., Jones, R. G., Legg, T., Miller, J., Skeggs, J., Wallom, D., Weisheimer, A., Wilson, S., Stott, P. A. & Allen, M. R. 2016 Human influence on climate in the 2014 southern England winter floods and their impacts. *Nature Climate Change* 6 (6), 627–634.

Sen, P. K. 1968 Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association* 63 (324), 1379–1389.

Slater, L. & Villarini, G. 2017 On the impact of gaps on trend detection in extreme streamflow time series. *International Journal of Climatology* 37 (10), 3976–3983.

Stahl, K., Hisdal, H., Hannaford, J., Tallaksen, L. M., van Lanen, H. A. J., Sauquet, E., Demuth, S., Fendekova, M. & Jódar, J. 2010 Streamflow trends in Europe: evidence from a dataset of near-natural catchments. *Hydrology Earth System Sciences* 14 (12), 2367–2382.

Stahl, K., Tallaksen, L. M., Hannaford, J. & van Lanen, H. A. J. 2012 Filling the white space on maps of European runoff trends: estimates from a multi-model ensemble. *Hydrology Earth System Sciences* 16 (7), 2035–2047.

Svensson, C., Brookshaw, A., Scaife, A. A., Bell, V. A., Mackay, J. D., Jackson, C. R., Hannaford, J., Davies, H. N., Arribas, A. & Stanley, S. 2015 Long-range forecasts of UK winter hydrology. *Environmental Research Letters* 10 (6), 064006.

Theil, H. 1950 A rank-invariant method of linear and polynomial regression analysis. *Proceedings of the Royal Netherlands Academy of Sciences* 53, 386–392.

Van Loon, A. F., Stahl, K., Di Baldassarre, G., Clark, J., Rangecroft, S., Wanders, N., Gleeson, T., Van Dijk, A. I. J. M., Tallaksen, L. M., Hannaford, J., Uijlenhoet, R., Teuling, A. J., Hannah, D. M., Sheffield, J., Svoboda, M., Verbeiren, B., Wagener, T. & Van Lanen, H. A. J. 2016 Drought in a human-modified world: reframing drought definitions, understanding, and analysis approaches. *Hydrology Earth System Sciences* 20 (9), 3631–3650.

Villarini, G., Smith, J. A., Serinaldi, F. & Ntelekos, A. A. 2011 Analyses of seasonal and annual maximum daily discharge records for central Europe. *Journal of Hydrology* 399 (3–4), 299–312.

Vorogushyn, S. & Merz, B. 2013 Flood trends along the Rhine: the role of river training. *Hydrology Earth System Sciences* 17 (10), 3871–3884.

Whitfield, P. H., Burn, D. H., Hannaford, J., Higgins, H., Hodgkins, G. A., Marsh, T. & Looser, U. 2012 Reference hydrologic networks I. The status and potential future directions of national reference hydrologic networks for detecting trends. *Hydrological Sciences Journal* 57 (8), 1562–1579.

Wilby, R. L. 2006 When and where might climate change be detectable in UK river flows? *Geophysical Research Letters* 33 (19), L19407.

Wilby, R. L. & Quinn, N. W. 2013 Reconstructing multi-decadal variations in fluvial flood risk using atmospheric circulation patterns. *Journal of Hydrology* 487, 109–121.

Zhang, X. S., Amirthanathan, G. E., Bari, M. A., Laugesen, R. M., Shin, D., Kent, D. M., MacDonald, A. M., Turner, M. E. & Tuteja, N. K. 2016 How streamflow has changed across Australia since the 1950s: evidence from the network of hydrologic reference stations. *Hydrology Earth System Sciences* 20 (9), 3947–3965.