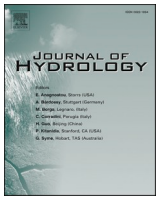
Journal of Hydrology 625 (2023) 130104

Contents lists available at ScienceDirect

Journal of Hydrology

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



Research papers

Geological controls of discharge variability in the Thames Basin, UK from cross-spectral analyses: Observations versus modelling

Graham P. Weedon [a,*,1], Emma L. Robinson [b], John P. Bloomfield [c], Stephen Turner [b], Emily J. Crane [b], Martin J. Best [d]

[a] Met Office, JCHMR, Maclean Building, Benson Lane, Crowmarsh Gifford, Wallingford, Oxfordshire OX10 8BB, UK
[b] UK Centre for Ecology & Hydrology, Maclean Building, Benson Lane, Crowmarsh Gifford, Wallingford, Oxfordshire, UK
[c] British Geological Survey, Maclean Building, Benson Lane, Crowmarsh Gifford, Wallingford, Oxfordshire, UK
[d] Met Office, Fitzroy Road, Exeter, UK

ARTICLE INFO

This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Roger Moussa, Associate Editor

Keywords:
Geological controls
Discharge variability
Spectral analysis
Transfer function
BFI
Lag1 autocorrelation
Thames Basin

ABSTRACT

Geological factors controlling daily- to multi-year discharge variability in 48 sub-catchments spanning 10–1000 km^2 in the Thames Basin were investigated using cross-spectral analysis. The analyses represent a 'transfer function approach' applied to daily observed streamflow (output) versus catchment-wide precipitation (input) for data spanning 1990–2014. Catchments dominated by high-permeability bedrock have significant attenuation of high-frequency precipitation variability and large delays at all frequencies with streamflow dominated by baseflow (high lag1 autocorrelation and high Base Flow Index, BFI). Catchments dominated by low-permeability rocks have little high-frequency attenuation and small delays and consequently 'flashy' behaviour. For all sub-catchments >300 km^2 in the Thames Basin, attenuation of the highest frequency precipitation variability caused by mixing of flow from upstream plus groundwater flow (representing 'older' variability) with direct surface flow ('younger' variability) constitutes real-world moving averaging as indicated by a roll-off in power at the highest frequencies.

The success of the JULES land surface model in simulating discharge (i.e. surface and sub-surface runoff routed between grid boxes) is also linked to the underlying geology. Larger catchments (>300 km^2) are modelled well because routing between numerous grid boxes leads to moving averaging that is a good analogue for the observations. Modelling was least successful (e.g. lowest Kling-Gupta Efficiency) for small catchments (<300 km^2) dominated by high-permeability bedrock - with far too little attenuation of high-frequency precipitation variability and insufficient delays at all frequencies. Experimentally switching the soil saturated hydraulic conductivity to that of the underlying bedrock for grid boxes dominated by aquifers significantly improves modelled discharge variability in small sub-catchments - confirming the importance of bedrock permeability in modelling.

For small catchments in data-sparse regions, knowledge of the relative proportions of different hydrogeological units (aquifers, aquitards) potentially could be used to predict and model discharge variability as characterised by BFI and lag1 autocorrelation.

1. Introduction

The land surface model (LSM) JULES (Joint UK Land Environment Simulator, Best et al., 2011) can be used to simulate discharge or streamflow using a non-calibrated distributed hydrological model. This takes into account factors such as precipitation inputs, evapotranspiration losses, diurnal and annual cycles of radiation, soil properties, sub-grid topographic heterogeneity (via TOPMODEL, Beven and Kirkby, 1979) and 'routing' of modelled surface-, and sub-surface-runoff between grid boxes. Compared to daily discharge observations in some catchments the performance is relatively good (judged using e.g. Nash-Sutcliffe Efficiency, NSE, or the modified Kling-Gupta Efficiency, KGE, of Gupta et al. 2009; Kling et al., 2012) whilst in others it is much less so (e.g. Blyth et al., 2011; Haddeland et al., 2011; Martínez-de la Torre et al., 2019; Chou et al., 2022). Similarly, the performance of various lumped hydrology models varies across the UK (Lane et al., 2019). Such results

* Corresponding author.
E-mail address: graham.p.weedon@gmail.com (G.P. Weedon).
[1] Note: Retired on 31st August 2023

https://doi.org/10.1016/j.jhydrol.2023.130104
Received 25 August 2022; Received in revised form 8 August 2023; Accepted 11 August 2023
Available online 30 August 2023

suggest there may be a missing, spatially-varying factor that might account for a significant proportion of localised problems with modelling streamflow.

A widely used index of discharge variability is the Base Flow Index, BFI (e.g. Gustard et al., 1992). This amounts to the ratio of the sum of a smoothed hydrograph to the sum of the observed hydrograph. Low BFI is associated with a high proportion of high-frequency (e.g. <14 day) variability (a 'flashy' catchment) whereas high BFI indicates relatively slowly varying streamflow. There are numerous ways to separate the low-frequency- from the high-frequency-variability and, since in reality there is a continuum across scales (Smakhtin, 2001; Stoelzle et al., 2020), there are numerous methods for calculating BFI (see comparison studies by e.g. Eckhardt, 2008; Chen and Teegavarapu, 2020). Variations in BFI have been explained in terms of factors such as climate and physiographical characteristics including topography, vegetation, soils and catchment area (e.g. Schneider et al., 2007; Longobardi and Villani, 2008; Beck et al., 2013; Singh et al. 2019).

The influence of geology on BFI – specifically bedrock permeability was suggested by Lacey and Grayson (1998), Longobardi and Villani (2008), Bloomfield et al. (2009, 2021) and Singh et al. (2019). Tague and Grant (2004) demonstrated that in the Willamette River catchment in western Oregon where basalt dominates the bedrock, streamflow variability is strongly correlated with the percentage of sub-catchment areas covered by young (<2 million year old) volcanic rocks. Bloomfield et al. (2009, 2021) demonstrated for the Thames Basin (used here to denote the geological structure rather than the river catchment) in southeast UK that BFI is strongly correlated with the proportions of rock permeability classes. Pfister et al. (2017) demonstrated using tracers that bedrock permeability is a factor involved with catchment storage and mean transit times. That study considered a limited number of nested catchments in the Alzette River basin (Luxembourg) with a correspondingly limited range of bedrock, climate, topographic settings and vegetation- and soil-coverage and the generality of their observations needs further testing.

A geological component to controlling streamflow variability is very important from the perspective of land surface modelling. Specifically, many LSMs currently lack a laterally moving groundwater component and consequently they fail to allow for any direct geological controls on modelled discharge ('modelled discharge' meaning the surface and sub-surface runoff 'routed' from grid box to grid box). Another modelling issue that influences simulated discharge is that the infiltration of precipitation into the sub-surface is typically determined using soil parameters derived from observed soil texture (e.g. from the Harmonized World Soil Database, Wieder et al., 2014) – especially soil saturated conductivity instead of allowing for the bedrock characteristics.

The objectives of the paper are to firstly improve understanding of the controls of discharge variability from spectral and cross-spectral analysis (Weedon et al., 2015) of observations from catchments of various sizes in the Thames Basin. Secondly, we apply the insights obtained to modify JULES. After a description of the data sources (Section 2) and the Methods (Section 3), we analyse the observations (Section 4) and then compare the observations with the model simulations and show how to improve the model for simulating discharge variability (Section 5).

2. Observations

The analysis period of both the observations and the modelling is for 1st Jan 1990 to 31st December 2014. This is constrained by the availability of gridded hourly precipitation at a 1 km^2 resolution as used for the modelling using JULES (Section 3.3).

2.1. Precipitation

Catchment-wide average precipitation in mm d^{-1} was obtained from the National River Flow Archive (NRFA) web site https://nrfa.ceh.ac.uk/data/search. These data represent the integration of precipitation upstream of each NRFA streamflow gauge as derived from the 1 km^2 gridded CEH-GEAR daily precipitation dataset (Tanguy et al., 2014; Keller et al., 2015). In southeast England there is a high density of precipitation gauges that were used in the generation of the gridded CEH-GEAR precipitation. For example, around 1000 precipitation gauges occur within the 9948.0 km^2 of the Thames catchment down to the tidal limit of the Thames estuary at Kingston - NRFA gauge 39001 (Fig. 1 of Weedon et al., 2015). In addition to the data from 39001, 48 sub-catchments were selected for analysis which all drain into the Thames itself or into the Thames estuary (Fig. 1 and Table 1).

The selected sub-catchments represent all those studied by Bloomfield et al. (2009), except for 6 where streamflow data were not available from the full study time interval, plus an additional 10. Mean annual rainfall ranges between 590 and 860 mm y^{-1} (Table 1) with highest values associated with the higher topography of the Cotswolds (dominated by the Great and Inferior Oolite groups), Chilterns, Berkshire Downs and North Downs (dominated by the Chalk Group). To allow comparison with the discharge variability in the cross-spectral analyses the precipitation data (in mm d^{-1}) were converted into total catchment precipitation rates (m^3 s^{-1}) by allowing for basin area as listed in Table 1. The order of the sub-catchments in Table 1 reflects the grouping by dominant rock formation and permeability class listed in Table 2 (Section 2.3).

2.2. Streamflow

The density of precipitation gauges in southeast England means that there are no missing data in the CEH-GEAR dataset in the region and period used in this study. However, there are numerous causes for observed daily streamflow data to be incomplete. Daily average streamflow in m^3 s^{-1} was obtained from the Environment Agency website http://environment.data.gov.uk/hydrology/explore. There were 13 gauges where data were not available from the website so they were obtained directly from the Environment Agency instead (for NRFA gauges: 39005, 39015, 39019, 39021, 39054, 39058, 39095, 39099, 39102, 39108, 39109, 39110, 40007).

Although the NRFA website used to obtain daily catchment precipitation also provides daily streamflow data, the EA website streamflow data are provided with quality flags and metadata so that time steps with missing or unreliable estimates can be removed prior to the spectral and cross-spectral analyses. Data were removed from analysis when the quality flags did not correspond to 'Good' and 'Complete' and/or when additional information indicated the values were labelled as 'Missing', 'Estimated', 'Suspect' or 'gap-filled'. Streamflow gauges were selected for cases with minimal anthropogenic influence and where not >25% of time steps are excluded from analysis in the 25 year time period under study (i.e. a minimum of 6848 days' retained data out of 9131). Nevertheless, there are only 4 gauges out of 49 with <90% complete data (Table 1). The Thames at Kingston (gauge 39001) includes significant anthropogenic influence so the spectral analyses were initially conducted on the reported discharge time series, but then repeated using 'naturalised' data which have slightly higher values compared to the measured data to account for abstraction (averaging 25.4 m^3 s^{-1} more than that observed).

2.3. Rock fractions

Bloomfield et al. (2009) showed that the numerous rock formations in the Thames Basin (Fig. 2a) can be meaningfully categorized (Fig. 2b) from lowest (0.0000001 m d^{-1}) to highest average permeability (1.0 to 200 m d^{-1}) as: (a) aquitards (clays and shales), (b) relatively low permeability superficial deposits (especially clays and Diamicton), and (c) intergranular aquifers (mainly sandstones) and (d) fractured aquifers (i.e. the fractured limestones). We have adopted this classification except that the superficial 'Clay-with-Flints' is now included with the

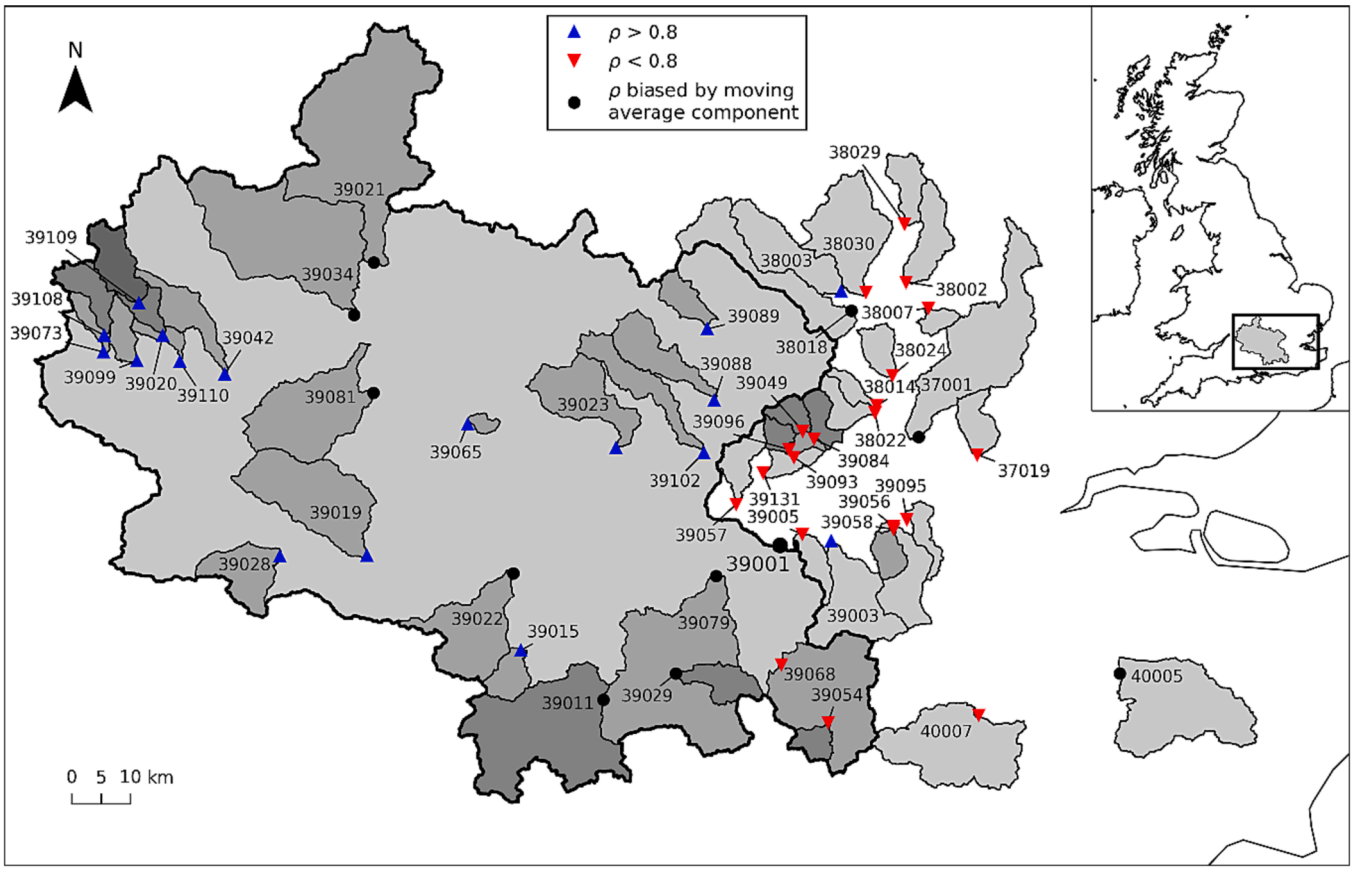


Fig. 1. Areas covered by the catchment Thames to Kingston monitored at NRFA gauge 39001 (thick outline) and additional sub-catchments investigated. Catchment gauge points are coloured according to the lag1 autocorrelation (ρ) of the observed discharge variability.

Chalk Group and the Weald Sand is now combined together with the Weald Clay, and these Lower Cretaceous sands and clays are classed as an aquitard (Fig. 2). The percentage fractions of these permeability classes based on 1:50,000 bedrock and superficial deposits geological maps were determined for each sub-catchment (Bloomfield et al., 2009) and are used within the multiple regression exercise (Section 4.3).

The majority of sub-catchments are dominated by rocks from a single permeability class as indicated in Table 2. Particularly important are those dominated by fractured aquifers (i.e. >50% Chalk Group or >85% the Inferior Oolite Group plus Great Oolite Group or, equivalently, >85% 'Oolite groups') or dominated by >70% relatively impermeable rocks – especially the aquitards represented by >50% Thames Group, >50% Weald Sand and Clay, or dominated by the low permeability superficial deposit: >50% Diamicton (i.e. glacial 'boulder clay'). Just one sub-catchment is dominated by an intergranular aquifer (>65% Lower Greensand). The remaining sub-catchments are not dominated by a single permeability class and so are described as 'Mixture' in Table 2.

3. Methods

3.1. Power spectra

The spectral and cross-spectral procedures used are identical to those described with mathematical details in Appendix A of Weedon et al. (2015). Prior to spectral analysis the time series were linearly detrended to centre the data and to avoid power at the zero frequency that would bias the adjacent lowest frequency output. The first and last 10% of the detrended data were tapered using a split cosine taper to avoid periodogram leakage (Weedon et al., 2015).

When time series data include values inserted from interpolations designed to fill data gaps this inevitably leads to increased serial correlation and artificial steepening of the background of the associated power (or variance-) spectra (Schulz and Stattegger, 1997). The increase in serial correlation leads to an increase in the lag1 autocorrelation that is of interest here for characterising the spectral background. This problem can be avoided by using the Lomb-Scargle transform designed for unevenly spaced data (Press et al., 1992) instead of using a standard fast Fourier transform (FFT).

The PERIOD algorithm of Press et al. (1992) implements the Lomb-Scargle transform. It provides the average amplitude of the sine- and average amplitude of the cosine-components which, when squared and summed, give the periodogram spectral estimates. A 3-point Hanning window was applied to the spectral estimates three times to increase the degrees of freedom from 2 in the periodograms to 8 eight for the power spectra - doubling the signal to noise ratio (Priestley, 1981).

An advantage of the PERIOD algorithm is that the Rayleigh frequency (i.e. the first non-zero frequency position and spacing of the frequency estimates) can be controlled by the user. Consequently, the Rayleigh frequency was set to 0.04 cycles per year for all the spectral and cross-spectral analyses – allowing direct comparison of results between catchments regardless of the number of reliable (non-missing) time steps. Since 1.0 is an integer multiple of 0.04 this means the power associated with any variability at the scale of the annual cycle is also evaluated at precisely the right frequency – thereby avoiding 'Scalloping loss' (Ifeachor and Jervis, 1993).

Hence the power spectra of the observed streamflow data were analysed using only the time steps that are regarded as having reliable values (Section 2.2). Similarly, the observed precipitation data were analysed on the same time steps even though those datasets are complete. Analogously, when the daily discharge output from the JULES model was analysed, again the spectral (and cross-spectral) analyses were restricted to using only the times of the reliable observed

Table 1
Gauges numbers underlined indicate sub-catchments with independent streamflow (i.e. not receiving flow from other catchments that have gauge data being analysed). 'Number of values' indicates the number of days with reliable data (maximum 9131). Annual rainfall from the NRFA website is based on the averages for 1961–1990. Area, mean discharge and BFI are obtained from the NRFA website.

NRFA gauge	Stream or River	Gauge site	Latitude (°)	Longitude (°)	Area (km^2)	AnnualRainfall (mm)	Mean Discharge ($m^3 s^{-1}$)	BFI	No. of values
38003	Mimram	Panshanger Park	−0.1413	51.8034	133.9	660	0.52	0.93	9088
38018	Upper Lee	Howe Green, Water Hall	−0.1175	51.7731	150.0	660	1.20	0.82	9126
39003	Wandle	Connolly's Mill combined	−0.1813	51.4198	176.1	720	1.90	0.87	8733
39015	Whitewater	Lodge Farm	−0.9500	51.2640	44.6	780	0.39	0.94	8295
39019	Lambourn	Shaw	−1.3251	51.4111	234.1	740	1.80	0.97	9004
39023	Wye	Bourne End, Hedsor	−0.7086	51.5719	137.3	760	0.98	0.93	9033
39028	Dun	Hungerford	−1.5394	51.4148	101.3	790	0.69	0.95	9037
39065	Ewelme Brook	Ewelme	−1.0737	51.6197	13.4	700	0.05	0.97	8911
39088	Chess	Rickmansworth	−0.4613	51.6420	105.0	750	0.54	0.95	8971
39089	Gade	Bury Mill, Hemel Hempstead	−0.4754	51.7576	48.0	720	0.15	0.93	8918
39102	Misbourne	Denham Lodge	−0.4898	51.5691	94.8	750	0.26	0.90	8813
39020	Coln	Bibury	−1.8254	51.7646	106.7	820	1.40	0.93	9122
39042	Leach	Priory Mill, Lechlade	−1.6730	51.6922	76.9	740	0.76	0.80	8805
39073	Churn	Cirencester	−1.9723	51.7245	84.0	850	0.78	0.91	8982
39099	Ampney Brook	Ampney St Peter	−1.8899	51.7107	45.3	810	0.60	0.76	8289
39108	Churn	Perrott's Brook	−1.9703	51.7501	59.0	860	0.70	0.89	7299
39109	Coln	Fossebridge	−1.8852	51.7997	82.0	830	0.56	0.91	8611
39110	Coln	Fairford	−1.7840	51.7094	130.0	810	2.10	0.93	7709
39029	Tilling Bourne	Shalford	−0.5695	51.2209	59.0	810	0.54	0.88	8939
38014	Salmon Brook	Lower Edmonton	−0.0603	51.6261	20.5	670	0.16	0.29	9094
38022	Pymme's Brook	Edmonton, Silver Street	−0.0653	51.6157	42.6	670	0.45	0.48	9089
39005	Beverley Brook	Wimbledon Common	−0.2520	51.4317	43.5	630	0.54	0.66	9006
39049	Silk Stream	Colindeep Lane	−0.2445	51.5912	29.0	690	0.25	0.34	8947
39057	Crane	Cranford Park	−0.4126	51.4888	61.7	640	0.50	0.33	9086
39058	Pool	Winsford Bridge	−0.0273	51.4358	38.3	660	0.27	0.55	9087
39084	Brent	Brent Cross	−0.2176	51.5775	36.4	680	0.34	0.34	9118
39093	Brent	Monks Park	−0.2678	51.5516	118.0	680	0.94	0.23	9044
39095	Quaggy	Manor House Gardens	0.0062	51.4557	34.0	640	0.14	0.45	8691
39096	Wealdstone Brook	Wembley	−0.2795	51.5622	21.8	660	0.13	0.24	9128
39131	Brent	Costons Lane, Greenford	−0.3449	51.5274	142.2	670	1.10	0.29	9131
39054	Mole	Gatwick Airport	−0.1997	51.1437	31.8	820	0.35	0.23	7577
39068	Mole	Castle Mill	−0.3112	51.2384	316.0	780	3.80	0.42	8949
40005	Beult	Stilebridge	0.5160	51.2021	277.1	690	2.10	0.23	8788
40007	Medway	Chafford Weir	0.1685	51.1445	255.1	830	2.90	0.46	8547
37001	Roding	Redbridge	0.0406	51.5762	303.3	610	1.80	0.39	9118
38002	Ash	Warside Mardock	0.0203	51.8148	78.7	620	0.28	0.53	9079
38007	Canons Brook	Harlow, Elizabeth Way	0.0736	51.7739	21.4	600	0.17	0.39	9038
38029	Quin	Braughing, Griggs Bridge	0.0209	51.9042	50.4	630	0.14	0.45	9052
38030	Beane	Hertford Hartham Park	−0.0798	51.8012	175.1	630	0.56	0.76	9119
37019	Beam	Bretons Farm, Elm Park	0.1840	51.5465	49.7	590	0.33	0.38	8943
38024	Small River Lee	Enfield Lk, Ordnance Road	−0.0196	51.6719	41.5	640	0.29	0.47	8967
39021	Cherwell	Enslow Mill	−1.3014	51.8613	551.7	660	3.90	0.66	6880
39001	Thames	Kingston	−0.3077	51.4154	9948.0	710	90.40	0.63	8955
39011	Wye	Tilford	−0.7502	51.1820	396.3	860	3.30	0.72	9057
39022	Loddon	Sheepbridge	−0.9667	51.3818	164.5	740	2.03	0.77	9046
39034	Evenlode	Cassington Mill	-0.13517	51.7863	430.0	690	3.80	0.71	9011
39056	Ravensbourne	Catford Hill	−0.0269	51.4417	120.4	710	0.37	0.49	8889
39079	Wey	Weybridge	−0.4663	51.3725	1008.0	790	7.40	0.66	8876
39081	Ock	Abingdon	−1.3052	51.6667	234.0	640	1.60	0.65	8844

streamflow data. This allows a fair comparison of the cross spectral results from the modelling with the results from the observations for the same sub-catchment. Weedon et al. (2015) showed that variance/power spectra are minimally affected by analysing the logarithm of the hydrograph rather than the original values.

3.2. Cross spectra

Amplitude ratio spectra were obtained by dividing the square root of the power of the streamflow time series by the square root of the power of the precipitation time series (Eq. (A11) of Weedon et al., 2015). The amplitude ratio spectrum multiplied by the coherency spectrum corresponds to the gain spectrum (Eq. (A13) of Weedon et al., 2015).

Phase difference spectra are derived from smoothed versions of the coperiodogram and the quadrature periodograms (Priestley, 1981; Weedon et al., 2015). The ratio of the resulting quadrature spectrum and the cospectrum gives the phase difference in radians. Here the delay or lag of the discharge variability relative to the precipitation variability is indicated at each frequency (1/period) in units of days rather than radians or degrees and the results are shown on a logarithmic scale (i.e. delay(d) = period(d) × phase difference°/360°). Uncertainty in phase difference (±95% confidence interval) become very large (e.g. >±90°) when the coherency is low. To limit the data plotted to those values with phase uncertainties <±45°, data are only displayed and/or analysed when the coherency at the corresponding frequency exceeds its 95% confidence level (Weedon et al., 2015).

3.3. JULES modelling

Modelling used the standalone JULES land surface model at version 6.2 with the RAL3 configuration. This configuration is used with JULES coupled to an atmospheric model for forecasting within the Met Office's UK 1.5 km variable resolution operational forecast model (UKV, Bush et al., 2020). The RAL3 configuration uses TOPMODEL for representing

Table 2
Areal proportions of permeability classes by sub-catchment. F. = Formation, G. = Group.

NRFA gauge	Proportion of low permeability superficial	Proportion of aquitard	Proportion of fractured aquifer	Proportion of intergranular aquifer	Main permeability class	Character or main formation
38003	0.288	0.000	0.652	0.060	Fractured aquifer	>50% Chalk G.
38018	0.393	0.043	0.516	0.048	Fractured aquifer	>50% Chalk G.
39003	0.191	0.105	0.657	0.047	Fractured aquifer	>50% Chalk G.
39015	0.024	0.004	0.958	0.014	Fractured aquifer	>50% Chalk G.
39019	0.069	0.002	0.876	0.053	Fractured aquifer	>50% Chalk G.
39023	0.084	0.001	0.855	0.060	Fractured aquifer	>50% Chalk G.
39028	0.126	0.039	0.717	0.118	Fractured aquifer	>50% Chalk G.
39065	0.005	0.000	0.897	0.098	Fractured aquifer	>50% Chalk G.
39088	0.032	0.001	0.898	0.069	Fractured aquifer	>50% Chalk G.
39089	0.000	0.000	0.898	0.102	Fractured aquifer	>50% Chalk G.
39102	0.151	0.002	0.721	0.126	Fractured aquifer	>50% Chalk G.
39020	0.000	0.043	0.912	0.045	Fractured aquifer	>85% Oolite G.
39042	0.028	0.019	0.892	0.061	Fractured aquifer	>85% Oolite G.
39073	0.000	0.024	0.949	0.027	Fractured aquifer	>85% Oolite G.
39099	0.000	0.000	0.963	0.037	Fractured aquifer	>85% Oolite G.
39108	0.000	0.026	0.954	0.020	Fractured aquifer	>85% Oolite G.
39109	0.000	0.059	0.896	0.045	Fractured aquifer	>85% Oolite G.
39110	0.000	0.036	0.913	0.051	Fractured aquifer	>85% Oolite G.
39029	0.060	0.005	0.078	0.857	Intergranular aquifer	>65% Lower Greensand
38014	0.339	0.618	0.000	0.043	Aquitard	>70% Low permeability, >50% Thames F.
38022	0.332	0.610	0.000	0.058	Aquitard	>70% Low permeability, >50% Thames F.
39005	0.222	0.526	0.176	0.076	Aquitard	>70% Low permeability, >50% Thames F.
39049	0.052	0.931	0.000	0.017	Aquitard	>70% Low permeability, >50% Thames F.
39057	0.181	0.673	0.000	0.146	Aquitard	>70% Low permeability, >50% Thames F.
39058	0.047	0.833	0.004	0.116	Aquitard	>70% Low permeability, >50% Thames F.
39084	0.187	0.800	0.000	0.013	Aquitard	>70% Low permeability, >50% Thames F.
39093	0.094	0.882	0.000	0.024	Aquitard	>70% Low permeability, >50% Thames F.
39095	0.235	0.752	0.006	0.008	Aquitard	>70% Low permeability, >50% Thames F.
39096	0.003	0.980	0.000	0.017	Aquitard	>70% Low permeability, >50% Thames F.
39131	0.093	0.883	0.000	0.024	Aquitard	>70% Low permeability, >50% Thames F.
39054	0.056	0.897	0.000	0.047	Aquitard	>70% Low permeability, >50% Wealden G.
39068	0.056	0.733	0.034	0.177	Aquitard	>70% Low permeability, >50% Wealden G.
40005	0.036	0.852	0.000	0.112	Aquitard	>70% Low permeability, >50% Wealden G.
40007	0.003	0.939	0.000	0.058	Aquitard	>70% Low permeability, >50% Wealden G.
37001	0.630	0.227	0.000	0.143	Low permeability superficial	>70% Low permeability, >50% Diamicton
38002	0.807	0.015	0.034	0.144	Low permeability superficial	>70% Low permeability, >50% Diamicton
38007	0.666	0.226	0.000	0.108	Low permeability superficial	>70% Low permeability, >50% Diamicton
38029	0.838	0.000	0.071	0.091	Low permeability superficial	>70% Low permeability, >50% Diamicton
38030	0.708	0.000	0.199	0.093	Low permeability superficial	>70% Low permeability, >50% Diamicton
37019	0.433	0.437	0.000	0.130	Mixture	>70% Low permeability Mixture
38024	0.490	0.251	0.000	0.259	Mixture	>70% Low permeability Mixture
39021	0.031	0.728	0.183	0.058	Mixture	>70% Low permeability Mixture
39001	0.156	0.295	0.390	0.159	Mixture	Mixture
39011	0.047	0.102	0.337	0.514	Mixture	Mixture
39022	0.051	0.435	0.464	0.050	Mixture	Mixture
39034	0.094	0.275	0.557	0.074	Mixture	Mixture
39056	0.089	0.455	0.404	0.052	Mixture	Mixture
39079	0.096	0.234	0.182	0.488	Mixture	Mixture
39081	0.127	0.267	0.384	0.222	Mixture	Mixture

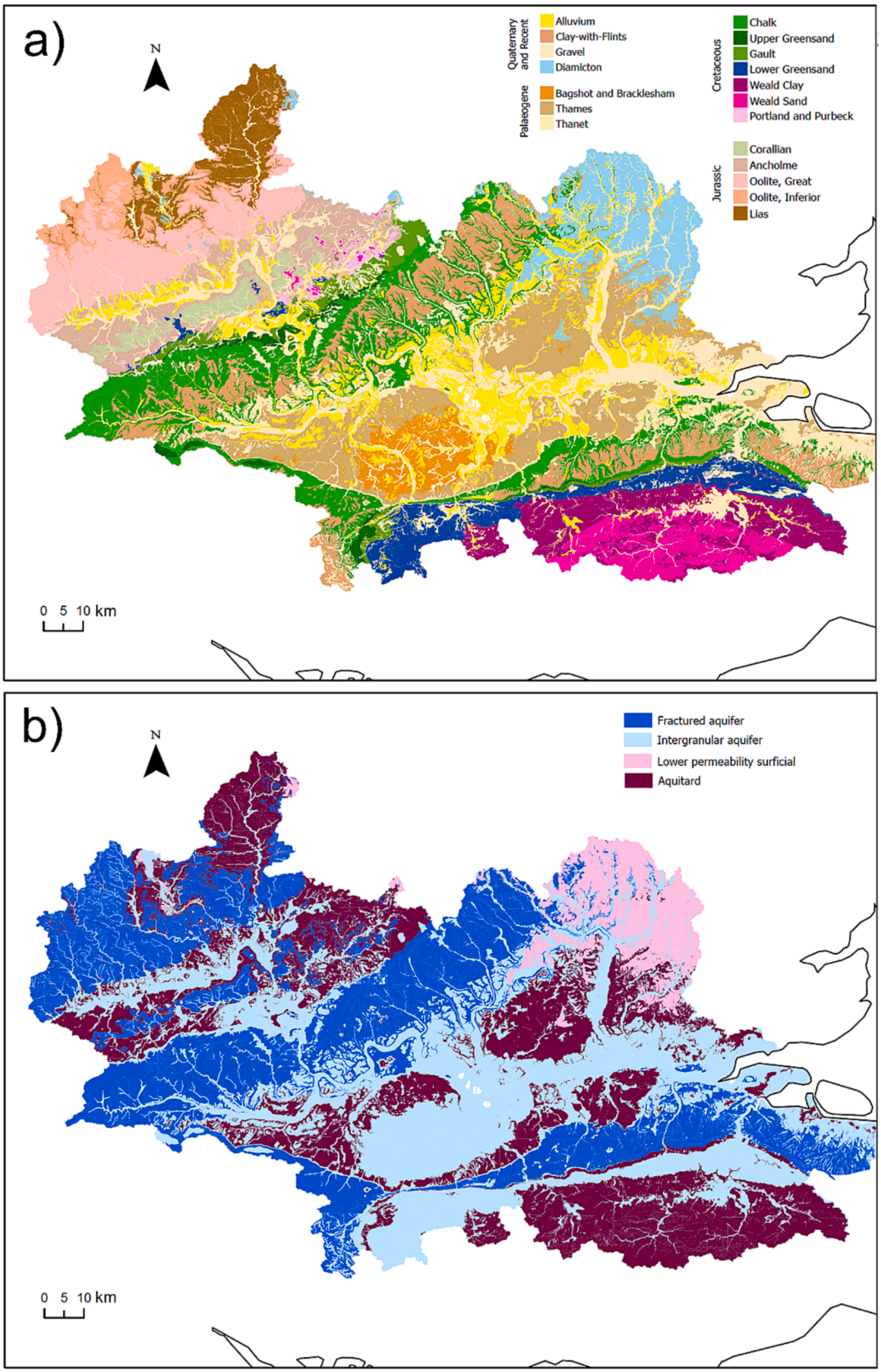


Fig. 2. A) Geology of the Thames Basin. British Geological Survey © UKRI 2023. b) Rock formations aligned to four permeability classes of Bloomfield et al. (2009) except that 'Clay-with-Flints' is assigned to Chalk/Fractured aquifer and both Weald Clay and Weald Sand are assigned to Aquitard. British Geological Survey © UKRI 2023.

subgrid heterogeneity in runoff production and diagnosing the water table (Gedney and Cox, 2003). Soil hydrological parameters are based on Brooks and Corey (1964) and Cosby et al. (1984) soil hydraulics. Soil ancillary data were copied at 1 km^2 directly from 2.2 km^2 values obtained from Harmonized World Soil database (Wieder et al., 2014). The model was run at an hourly time step and at a grid resolution of 1 km^2.

Meteorological data for forcing the model (near-surface air temperature and specific humidity, wind speed, pressure, downwards longwave radiation and downwards shortwave radiation) were provided by the daily CHESS data (Robinson et al., 2017, 2020) after disaggregation to hourly steps within JULES using representative diurnal variation (Williams and Clark, 2014). However, to provide more realistic precipitation data than can be obtained by disaggregating the CHESS daily average values, the 1 km^2 resolution, hourly average precipitation 'CEH-GEAR1hr' data of Lewis et al. (2018; 2019) were utilized. These gridded hourly precipitation observations were only available from 1st Jan 1990 to 31st December 2014 during the analysis.

The model was run following 'spin-up' such that soil moisture and soil temperature remained unchanged, within the tolerances set, from the start of one spin-up cycle (spanning 1990 to 1995 inclusive) to the next. The output surface and subsurface runoff was routed using the RFM kinematic wave model (Bell et al., 2007) according to a UK-wide 1 km^2 routing map (Davies et al., 2022). The modelling is for natural flow – i.e. it excludes representing anthropogenic influences on discharge (e.g. due to abstraction, damns, weirs and locks).

4. Analysis of the observations

4.1. Discharge variability, lag1 autocorrelation and BFI

Fig. 3 illustrates at top left, examples of discharge time series (in black) from a sub-catchment with a high BFI (from NRFA gauge 39088, River Chess at Rickmansworth) and from one with a relatively low BFI (gauge 39096, Wealdstone Brook at Wembley). These are plotted below the corresponding daily catchment-wide precipitation (in grey). The associated power spectra (top right) show, as expected, that the background power of precipitation (in grey) is almost flat – i.e. the distribution of variance is almost the same as would be expected from a pure random number or 'white noise' process. Conversely the spectrum of observed discharge from 39088 (in black) slopes steeply towards the Nyquist (i.e. highest calculated-) frequency. On the other hand, the spectrum of discharge from 39096 (also black) has a shape that is much more similar to the associated precipitation spectrum.

Note that a significant difference between the spectra of precipitation and discharge are that the latter contain strong concentrations of variance (i.e. spectral peaks) representing the annual cycle in evapotranspiration. In southeast England there is a relatively even distribution of rainfall throughout the year so the precipitation spectra contain relatively weak evidence for an annual cycle in rainfall amounts.

The lag1 autocorrelation (ρ) can provide a useful alternative to BFI to describe the shape of at least some spectra of discharge. ρ is simply the Pearson's correlation of the time series with itself offset by one time step

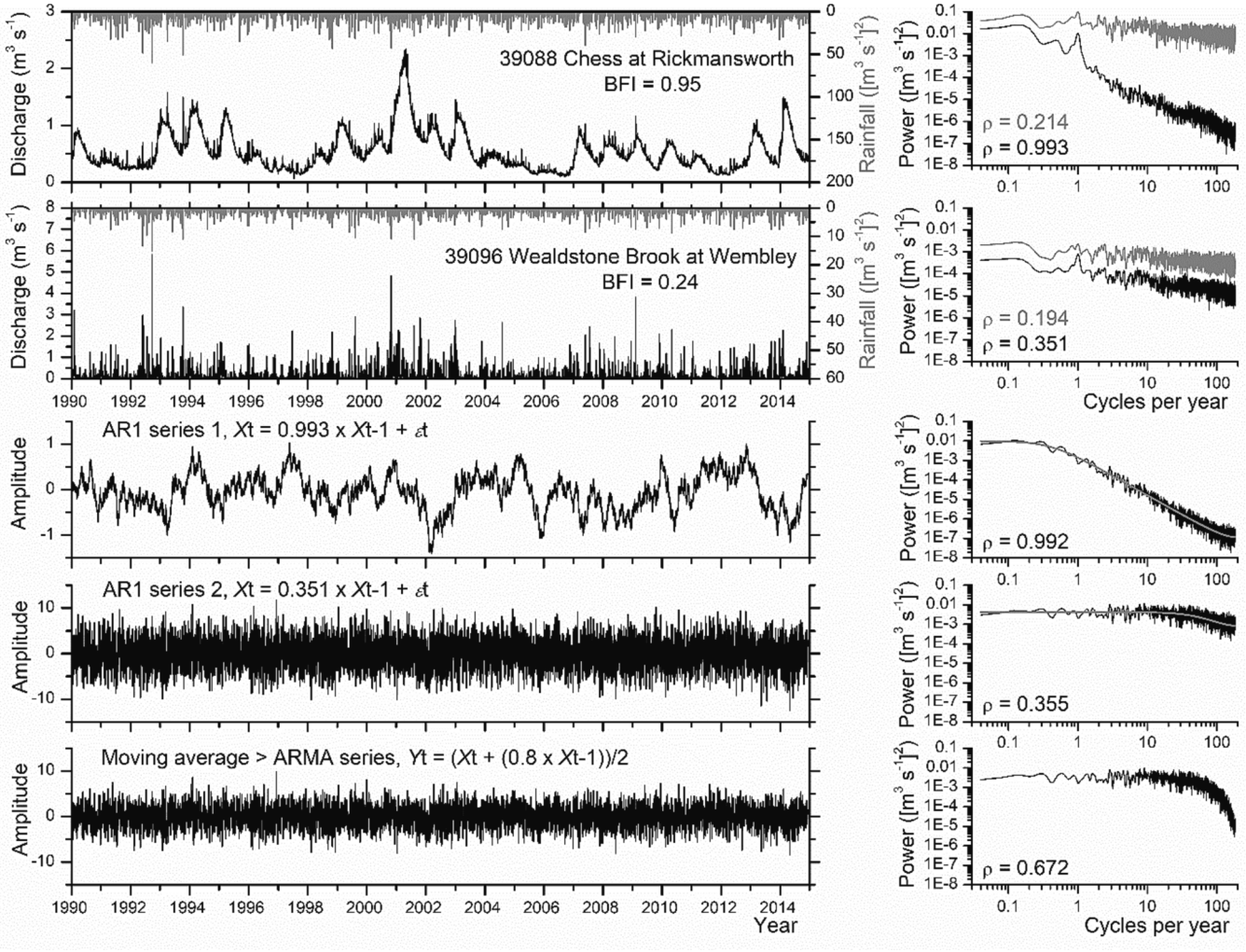


Fig. 3. Time series of rainfall (grey) and observed discharge (black) and associated power spectra and examples of synthetic time series and their power spectra.

(or lag hence 'lag1 autocorrelation'). As shown in Fig. 3 ρ for the precipitation time series in the Thames Basin are close to 0.2, but range between 0.35 and 1.0 for the observed discharge data (Table 3). Steeper power spectra are associated with a higher value of ρ.

ρ provides a measure of the average serial correlation of the time series. Modelling of the spectral backgrounds using artificial time series leads to a physical understanding of the significance of ρ. A discrete time, first order autoregressive (AR1) time series can be constructed (Gilman et al., 1963; Priestley, 1981) from a time series of stationary Gaussian distributed random numbers (ε_t) as:

$$X_t = \rho X_{t-1} + \varepsilon_t \quad (1)$$

where X_t is the value of the time series at time step t, X_{t-1} is the value at the previous time step. If ρ is zero the time series corresponds to random numbers and the spectral background is flat. When ρ is unity this corresponds to a 'random walk' and the spectrum is a power-law with a constant slope down to the Nyquist frequency if both power and frequency are plotted with logarithmic scales. Most relevant here is when ρ lies between 0.0 and 1.0. In that case the spectrum slopes to the right ('red noise', Gilman et al., 1963) with a shape determined by ρ itself with the form (e.g. Mann and Lees, 1996):

$$P_f = P_0 + \left(1 - \rho^2\right)/\left(1 - -2\rho\cos\pi(f/f_N) + \rho^2\right) \quad (2)$$

where P_f is the power at frequency f, P_0 is the average value of the power spectrum and f_N is the Nyquist frequency. Weedon et al. (2015) discuss why, even in the largest catchment studied here (represented by the Thames to Kingston at gauge 39001) the areas involved are too small to produce discharge power spectra exhibiting a power-law and corresponding to the Hurst phenomenon (cf Mudelsee, 2007). The small catchment and sub-catchment areas mean that the tributaries even in the largest catchment have correlated rather than independent variations in water inputs (precipitation).

The lower part of Fig. 1 illustrates AR1 time series constructed using equation 1 with ρ shown for the spectra of discharge from 39088 and

Table 3
Cross spectral results for observations and JULES model output. Mov. Ave. = Moving average, pre. = precipitation, Ann. = Annual, AR = Amplitude ratio. 39001* = values for naturalised discharge.

NRFA gauge	Mov. Ave. ?	ρ pre.	ρ dis. obs.	Ann. AR obs.	HF AR obs.	Ave. delay obs. (d)	ρ1 dis. JULES	Ann. AR JULES	Ave. delay JULES (d)	KGE
38003	No	0.214	0.956	0.299	0.017	3.011	0.414	0.264	1.894	-0.610
38018	Yes	0.217	0.902	0.724	0.085	3.862	0.602	0.252	2.049	-0.214
39003	No	0.227	0.828	0.270	0.063	1.822	0.426	0.338	1.650	-1.325
39015	No	0.253	0.976	0.489	0.031	6.458	0.293	0.236	1.103	-1.464
39019	No	0.232	0.996	0.657	0.017	15.139	0.481	0.408	2.063	-0.216
39023	No	0.217	0.972	0.289	0.017	3.128	0.345	0.317	1.906	-1.251
39028	No	0.242	0.987	0.557	0.020	7.118	0.343	0.325	1.494	-1.432
39065	No	0.202	0.999	0.334	0.007	28.636	0.143	0.092	0.664	-1.996
39088	No	0.214	0.993	0.327	0.014	7.532	0.410	0.367	1.928	-0.312
39089	No	0.212	0.991	0.337	0.011	4.413	0.281	0.253	1.377	-0.255
39102	No	0.213	0.990	0.235	0.013	7.841	0.518	0.404	1.910	-0.305
39020	No	0.239	0.998	1.362	0.025	8.123	0.577	0.441	1.678	+0.126
39042	No	0.223	0.989	1.705	0.775	11.249	0.344	0.336	1.515	+0.068
39073	No	0.236	0.998	1.286	0.032	15.780	0.419	0.406	1.769	+0.303
39099	No	0.234	0.989	1.802	0.094	9.765	0.287	0.338	1.072	-0.195
39108	No	0.233	0.992	1.534	0.049	15.546	0.425	0.380	1.538	+0.268
39109	No	0.233	0.997	0.973	0.018	8.660	0.439	0.3969	1.662	+0.376
39110	No	0.233	0.996	1.546	0.046	16.358	0.610	0.438	1.824	+0.063
39029	Yes	0.238	0.865	0.260	0.053	3.695	0.397	0.236	1.972	-1.304
38014	No	0.203	0.576	1.447	0.357	1.508	0.191	0.363	0.550	+0.470
38022	No	0.203	0.479	0.881	0.310	0.766	0.237	0.460	0.580	+0.102
39005	No	0.197	0.469	0.399	0.202	0.846	0.230	0.310	0.606	-1.168
39049	No	0.194	0.533	1.007	0.294	1.072	0.191	0.404	0.395	+0.204
39057	No	0.192	0.514	0.952	0.334	1.639	0.263	0.407	0.760	+0.472
39058	No	0.200	0.475	0.451	0.177	0.981	0.189	0.129	0.693	-0.118
39084	No	0.200	0.476	1.299	0.403	1.120	0.212	0.408	0.465	+0.456
39093	No	0.196	0.476	0.791	0.294	1.249	0.223	0.422	0.646	+0.395
39095	No	0.199	0.542	0.370	0.125	1.193	0.228	0.229	0.496	+0.285
39096	No	0.194	0.351	0.585	0.273	0.612	0.210	0.479	0.348	+0.130
39131	No	0.199	0.509	0.874	0.294	1.331	0.262	0.440	0.858	+0.374
39054	No	0.246	0.625	0.120	0.448	1.388	0.266	0.267	0.951	+0.430
39068	No	0.242	0.747	1.113	0.367	1.954	0.483	0.393	1.468	+0.572
40005	Yes	0.247	0.776	1.216	0.427	2.205	0.432	0.297	1.554	+0.531
40007	No	0.251	0.697	0.995	0.310	2.632	0.458	0.331	1.768	+0.503
37001	Yes	0.202	0.879	1.362	0.232	3.981	0.638	0.170	2.042	+0.515
38002	No	0.191	0.672	0.794	0.175	5.619	0.320	0.151	1.351	+0.467
38007	No	0.197	0.492	0.990	0.304	0.962	0.193	0.239	0.627	+0.088
38029	No	0.193	0.512	0.654	0.193	5.506	0.195	0.149	1.243	+0.463
38030	No	0.197	0.647	0.350	0.709	4.467	0.293	0.155	1.625	+0.004
37019	No	0.203	0.538	1.003	0.290	1.847	0.212	0.292	0.593	+0.293
38024	No	0.197	0.694	0.937	0.207	2.521	0.210	0.197	0.702	+0.066
39021	Yes	0.174	0.953	0.511	0.064	17.593	0.482	0.082	1.910	-0.315
39001	Yes	0.259	0.976	1.002	0.086	4.900	0.946	1.034	7.697	+0.556
39001*	Yes	0.259	0.976	1.041	0.087	4.762	0.946	1.034	7.697	+0.711
39011	Yes	0.266	0.790	0.403	0.110	3.183	0.661	0.407	1.927	+0.712
39022	Yes	0.247	0.836	1.024	0.186	3.247	0.408	0.359	1.482	-0.346
39034	Yes	0.218	0.967	1.395	0.100	7.052	0.583	0.358	2.250	+0.338
39056	No	0.212	0.488	0.203	0.087	1.091	0.357	0.254	1.528	+0.111
39079	Yes	0.262	0.891	0.509	0.114	2.950	0.859	0.475	2.893	+0.833
39081	Yes	0.206	0.944	1.379	0.114	8.713	0.291	0.153	1.646	-0.392

39096 (i.e. 0.993 and 0.351 respectively). The general characteristics of the artificial time series match those of the observed discharge quite well though there is no attempt to produce values that always exceed 0.0 and there is no superimposed regularity in the form of annual cycles (and thus no spectral peaks at the annual scale).

The ρ values shown with the power spectra of the artificial time series are a good match for the values used in the generating equations (i.e. 0.992 v 0.993 and 0.355 v 0.351). This demonstrates that directly measuring ρ from time series is only modestly biased (Mudelsee, 2010). However, there are several factors that have a much more significant impact on measured ρ.

The first factor was indicated in the Methods (Section 3.1). Interpolation of data leads to increased serial correlation, increased spectral slopes and higher ρ. This problem has been avoided here by using the Lomb-Scargle algorithm for data transformation instead of an FFT by excluding values from time steps where the data are missing or unreliable.

The second factor is the presence of superimposed regular components such as, in this case, the effects of annual cycles of evapotranspiration that can affect streamflow. Regular components, identified as spectral peaks on top of the spectral background, also cause an increase in measured ρ (Mann and Lees, 1996). For this reason the spectra of the artificial spectra – especially the second case – are not a perfect match for the spectra of observed discharge since the regular components of the real data have slightly increased the measured ρ that was used for constructing the artificial series as a purely AR1 process.

The third factor is the occurrence of moving averaging. A time series of a moving average (MA) process can be constructed from Gaussian-distributed random numbers as, for example (Priestley, 1981):

$$X_t = a_0\varepsilon_t + a_1\varepsilon_{t-1} \tag{3}$$

where a_0 and a_1 represent constants. The critical point here is that moving average processes 'mix' data from different time steps. This can be seen at the bottom of Fig. 3 where instead of constructing the time series from random numbers, a new time series has been constructed by using weighted averages of successive pairs of values from the second AR1 series – thereby creating an ARMA series. This causes intense smoothing at the shortest time periods and consequently an abrupt 'roll off' in power at the highest frequency part of the power spectrum (bottom spectrum in Fig. 3). This smoothing also leads to a much higher ρ. In the example shown the ρ measured for the second AR1 series of 0.355 has increased to 0.672. Therefore measured ρ cannot be interpreted without inspection of the associated power spectrum. If there is a clear roll-off in power at the highest frequencies then the measured lag1 autocorrelation will not provide a good characterization of the spectral background. Weedon et al. (2015) illustrated additional examples of power spectra with moving average roll off. Those examples relate to models where, although the models were run with hourly or half-hourly time steps, the routing of runoff between grid boxes was only applied at daily steps.

4.2. Observed discharge versus observed precipitation

4.2.1. Power spectral analysis

Fundamentally the background variability in streamflow or discharge is inherited from precipitation variability (e.g. Weedon et al., 2015). The increased spectral slope of the discharge spectrum is indicative of suppression of the highest frequency precipitation variability. Numerically this is represented by the increase in lag1 autocorrelation – i.e. the increase in serial correlation and thus short term memory. Physically this results from infiltration of the precipitation into soils and bedrock (aquifers) when the water table is well below the surface, whereby short-term variability is attenuated due to the filtering effects of water movement through pores and macro-pores including fractures over intervals of days to months (Milly and Wetherald, 2002; Li and Zhang, 2007; Little and Bloomfield, 2010). In general terms the higher the proportion of streamflow that is fed by direct surface flow the lower the ρ and the lower the BFI. With increasing proportion of streamflow fed by water from soil, and particularly from groundwater from rocks, the higher the suppression of the original short-term precipitation variability and consequently the greater the baseflow, ρ and BFI.

Groundwater response to precipitation variability depends on hydraulic diffusivity – i.e. hydraulic conductivity divided by specific storage. In general, when directly compared, specific storage varies by less than hydraulic conductivity and for most lithologies they are not correlated (Kuang et al., 2020) and the former depends upon the time scale of measurement and hence analytical method (Worthington et al., 2019). Nevertheless, the idea that the variability of streamflow is linked to the proportion of water obtained from the bedrock is consistent with tracer studies of catchments where bedrock permeability is considered to be the main control of subsurface storage and consequently of mean transit times (Hale and McDonnell, 2016; Hale et al., 2016; Pfister et al., 2017). However, for long transit times (years) differences related to bedrock permeability do not necessarily translate to differences in hydrologic response (Hale and McDonnell, 2016).

As already mentioned, superimposed on daily discharge variability in the Thames Basin is a regular (frequency limited) component due to the annual cycle of evapotranspiration (e.g. the spectral peaks at the annual scale in the discharge spectra in Fig. 3). Hourly measurements of discharge would be expected to have associated power spectra that include spectral peaks related to the diurnal cycle of evapotranspiration.

As discussed in Section 4.1 the regular components lead to a slight increase in measured ρ. A further complication is the evidence for moving averaging in the power spectra of discharge from the larger sub-catchments – i.e. spectra of observed discharge showing high-frequency roll-off in power. This is interpreted as an indication that mixing of water from upstream and from soils and groundwater (exhibiting earlier variability) with direct surface flow (with current/younger variability) leads to a moving-average attenuation of the shortest period variability. Longer streams and rivers will be expected to have more evidence for moving averaging – explaining why in this study spectral roll-offs have not been seen in sub-catchments that are <150 km^2, but are always seen in catchments >300 km^2 (see 'Mov. Ave.?' in Table 3). Very large rivers with catchments $>10{,}000$ km^2 (not studied here) exhibit the Hurst phenomenon, 'long-term memory' and thus power-law spectra (Hurst, 1951; Mesa and Povenda, 1993; Mudelsee, 2007; Fleming, 2014).

Power spectra of groundwater levels are typically power-laws with ρ equal to unity and long-memory characteristics rather than the short memory of the discharge analysed here (Zhang and Schilling, 2004; Little and Bloomfield, 2010; Habib et al. 2017). However, Zhang and Schilling (2004) illustrate a spectrum from their well 6B, that they acknowledge does not correspond to a power-law, but rather has the low frequency interval indicative of a short memory process with high ρ. These observations can be interpreted as indicating that so much attenuation of precipitation variability occurs at depth within aquifers that the lag1 autocorrelation rises to unity and power-law behaviour arises in groundwater levels. The degree of subsurface attenuation is known to rise with distance from rivers (Li and Zhang, 2007; Little and Bloomfield, 2010). The power-law spectral properties of groundwater levels could also be explained as due to an aggregation of numerous independent contributions, each of which have short memory characters (Hosking, 1984). This has been argued for the downstream generation of the Hurst phenomenon in large catchments from the contribution of runoff from numerous tributaries (Mudelsee, 2007).

4.2.2. Cross spectral analysis

Fig. 4a shows, in black, power spectra and cross spectra (amplitude ratio and phase difference or delay spectra) for observed discharge and precipitation for a sub-catchment (gauge 39019, Lambourn at Shaw) where >50% of the bedrock is fractured aquifer (Chalk Group) and therefore highly permeable. This is contrasted in Fig. 4b with a sub-

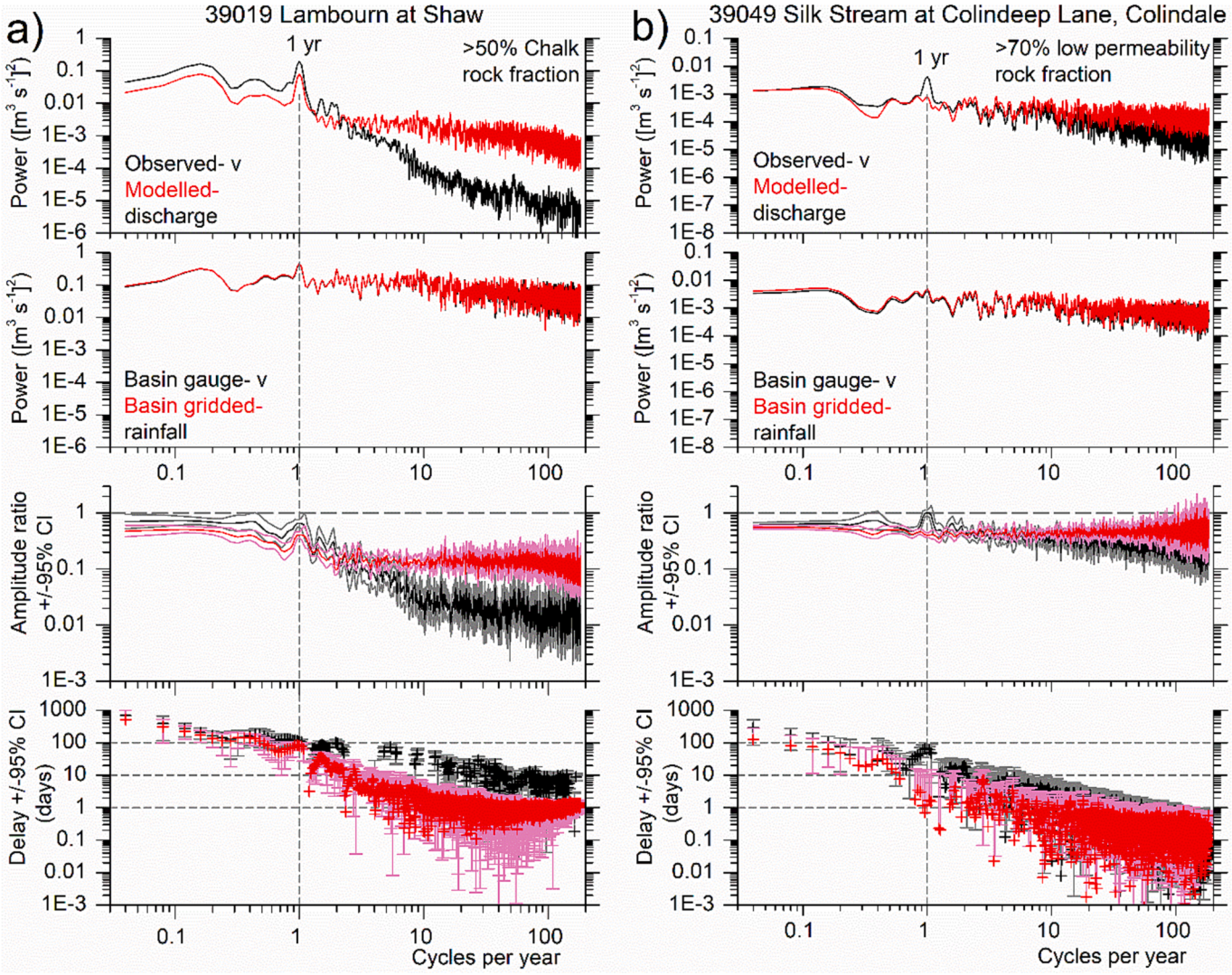


Fig. 4. A) Spectral results for data at gauge 39089. b) Spectral results for 38022. Top: Power spectra of discharge time series as observed (black) and as modelled in JULES (red). Top middle: Power spectra of catchment-wide precipitation from CEH-GEAR data (black) and for daily average catchment-wide cCEH-GEAR1hr data (red). Bottom Middle: Amplitude ratio spectra for observed (black) and modelled (red) discharge versus precipitation. Bottom: Delay (phase difference) spectra for observed (black) and modelled (red) discharge. Vertical bars around crosses indicate 95% confidence intervals).

catchment (gauge 39049, Silk Stream at Colindeep Lane, Colindale) where >70% of the bedrock is classed as low permeability.

Here the amplitude ratio- and phase difference-spectra are used to investigate the 'transfer function' associated with the environmental change of the input precipitation variations into the output streamflow variations due to filtering by catchment and groundwater flow processes (cf. Weedon et al., 2015). In formal systems analysis used in engineering and physics the equivalent 'Bode plots' involve plotting the gain spectrum in decibels and the phase difference spectrum in radians (e.g. Jenkins and Watts, 1969).

It has been assumed that the 25-year long time series provide representative average, frequency-specific distributions of variance and phase. Furthermore, it is assumed that the systems being studied are linear and time-invariant without feedbacks so that the cross spectra can be interpreted directly (Jenkins and Watts, 1969, Priestley, 1981). Though the sub-catchments were selected as having minor or minimal anthropogenic influence, the analysis does not take account of possible changes in conveyance, catchment storage or abstraction (Bloomfield et al., 2021).

The amplitude ratio spectrum is typically very similar in shape to the gain spectrum (Weedon et al., 2015), but the latter has larger 95% confidence intervals since it combines the uncertainty in estimating the amplitude ratio with the uncertainty in estimating the coherency (which is essentially the frequency-specific correlation between time series). The power spectra and amplitude ratio spectra indicate the frequency distribution of the average variability – non-stationary baseflow and extreme high-flow (flood) events are not distinguishable in the analyses (unlike wavelet-based methods e.g. Towler and McCreight, 2021). Note that the phase-difference spectra, expressed as delay in days, provide average values for each frequency. In some aquifers including the Chalk it is known that delays can vary through the year according to water table depth and rainfall intensity (e.g. Lee et al., 2006). Such non-stationary behaviour cannot be assessed via the cross spectral analysis used here.

For 39019, a sub-catchment with high permeability rocks, in Fig. 4a the power for the observed discharge in black drops steeply towards the Nyquist frequency and as a result the amplitude ratio spectrum (square root of discharge spectrum/precipitation spectrum) decreases from around 0.7 to about 0.01. This large attenuation of high-frequency precipitation variability is consistent with a substantial contribution of streamflow from the bedrock in the form of groundwater. Additionally, the delay of discharge variability relative to precipitation variability

(black crosses) is relatively large (>1 day) especially at the highest frequencies. This is also consistent with considerable infiltration of water into the aquifer (i.e. more delay in variability is associated with more infiltration).

By contrast 39049, with low permeability rocks, in Fig. 4b has a relatively flat power spectrum for the observed discharge (black) and only minor attenuation of the highest frequency precipitation variability (amplitude ratio about 0.3 at the Nyquist frequency). The delay is also much less at all frequencies and <1 day at the Nyquist frequency. These results are consistent with only minor infiltration into the low permeability bedrock so that on average there is minimal attenuation of high-frequency precipitation variability and little delay at the highest frequencies.

Fig. 5 shows amplitude ratio spectra of observed discharge versus observed precipitation, in black, for sub-catchments that have a variety of bedrock characteristics (fractured aquifers, aquitards) and those with a spectral shape indicating a moving average component. The sub-catchments dominated by fractured aquifers (>50% Chalk Group and >85% Oolite groups) have strong attenuation at high frequencies and, as recorded in Table 3, long delays (i.e. large delay averaged across all frequencies). The sub-catchments dominated by low permeability rocks (>70% low permeability – i.e. Diamicton, Thames Group) consistently show minimal high-frequency attenuation and short delays (i.e. small delay averaged across all frequencies).

40007 (River Medway at Chafford Weir) has bedrock consisting of 23% Weald Clay and 71% Weald Sand (Fig. 5a). With such a high proportion of sandstone it would be expected that it would have the characteristics of an intergranular aquifer. Nevertheless, the amplitude ratio of the observed discharge is very similar to the sub-catchments with >70% low permeability rocks. Despite the high proportion of Weald Sand it appears that this unit is behaving as an aquitard - like the Weald Clay - perhaps due to relatively fine-grained sand, plus porosity occluded by diagenetic minerals and/or by the presence of thin relatively low permeability clays within the sand sequence (Jones et al., 2000).

4.3. Lag1 autocorrelation versus fraction of rock formation permeability class

The coloured triangles in Fig. 1 show the gauges where the lag1 autocorrelation exceeds 0.8 (upright blue triangle) and where it is <0.8 (inverted red triangle). Sub-catchments where the ρ is compromised by the presence of a moving average component are shown as black circles. The coloured triangles clearly form spatial clusters as determined by the underlying geology. For example, in the west is a cluster of sub-catchments with high ρ linked to predominantly the Inferior Oolite and Great Oolite groups (fractured aquifers), with another fractured aquifer cluster along the Berkshire Downs and Chilterns (formed of Chalk Group rocks) in the centre and top of Fig. 1. Clusters associated with low permeability rocks include the Diamicton (glacial 'boulder-clay) in the northeast, the London Clay (Thames Group) along the Thames itself and the Weald Sand and Clay in the south.

Most sub-catchments are underlain by a variety of rock formations representing more than one of the four permeability classes. Bloomfield et al. (2009, 2021) showed using multiple regression that BFI is correlated with the proportion of the different permeability classes. This exercise is repeated here and a comparison made with ρ. Fig. 6 shows that both BFI and ρ are indeed inversely correlated with the proportion of low permeability superficial deposit (P <0.05 and P <0.02 respectively) and the proportion of aquitard (P <0.001) and directly correlated with the proportion of fractured aquifer (P <0.001). There are generally relatively small proportions of intergranular aquifers so there is no significant correlation with either BFI or ρ (Fig. 6).

A BFI multiple regression based on proportions of the variables: a) proportion of aquitard formations and b) proportion of fractured aquifer formations explains 85% of the variance (n = 49, r^2 = 0.85, P <0.0001). On the other hand, for ρ the proportion of explained variance is lower but still highly significant (n = 49, r^2 = 0.69, P <0.0001). For these multiple regressions, if a third explanatory is added: c) the proportion of low permeability superficial deposits, the explained variance increases by just a few percent.

The multiple regression for ρ versus the proportion of aquitard and fractured aquifers shown in Fig. 6 includes data from all sub-catchments, but in 11 cases there is spectral evidence for a moving average component (Table 3) and therefore the ρ does not characterise the spectral background across the full range of frequencies (Section 4). Excluding these 11 sub-catchments (black squares in Fig. 6) leads to a much higher proportion of variance explained (n = 38, r^2 = 0.86, P <0.0001).

Arguably these analyses include data that are not fully independent because in 7 cases the sub-catchments drain into adjacent examples. Therefore, using only sub-catchments where the streamflow is truly independent (indicated by underlining of gauge numbers in Table 1 to 3), and also excluding places where a moving average component is present, leaves a total of 31 cases. For these independent sub-catchments the ρ is still very highly significantly correlated with the proportion of the two permeability classes (n = 31, r^2 = 0.83, P <0.0001).

Alternative spectral metrics were also investigated. The average amplitude ratio at high frequencies is useful for distinguishing the sub-catchments with little attenuation of the precipitation variability (predominantly low-permeability rocks) from those with high attenuation (with predominantly high permeability rocks). To avoid the effects on the amplitude ratio spectrum of moving average components the average amplitude ratio was determined between 30 and 40 cycles per year. The result is lower, but still has highly significant correlation with the two key permeability classes (n = 49, r^2 = 0.75, P <0.0001). The average delay was determined across the full width of the spectra, but only at frequencies where the coherency exceeded the 95% confidence level so that the phase difference is reasonably well constrained (Section 4). The regression of the log(average delay) v proportion of aquitard and fractured aquifers is again significant but with lower variance explained (n = 49, r^2 = 0.59, P <0.0001).

These results demonstrate that, as Bloomfield et al. (2009, 2021) found for BFI, in the Thames Basin the spectral characteristics of the streamflow: a) lag1 autocorrelation, b) high frequency amplitude ratio of discharge versus precipitation and c) delay in discharge variability relative to precipitation variability, are influenced by the permeability of the underlying rocks. Note that geology determines several physiological factors that are linked to discharge variability (Bloomfield et al., 2009, 2011). These include higher elevation and more relief rainfall over the Oolite-dominated and Chalk-dominated hills and greater likelihood of shallow slopes with easily saturated soils and meandering channels over the more readily weathered low-permeability rocks such as the Thames Group and Weald formations.

5. Modelling of discharge variability

5.1. Modelled discharge versus gridded observed precipitation

Within JULES discharge is modelled using the surface and sub-surface runoff, as routed using a kinematic wave model (RFM) between the 1 km^2 grid boxes (Section 3.3). Fig. 4 shows the spectral and cross spectral results from the modelling in red. As expected, for frequencies analysed (<1/(2 d)), the power spectra of the daily basin-wide gauge precipitation in black are almost identical to the spectra of hourly basin-wide gridded precipitation in red used to drive JULES. However, in both cases shown the discharge from JULES generates a flatter power spectrum than that observed – very markedly so for the high permeability case. The examples in Fig. 4 come from sub-catchments that have very different areas (234.1 km^2 for 39019 versus 29.0 km^2 for 39049, Table1). Even so, Fig. 5 with a wide range of sub-catchment sizes demonstrates very similar modelling results dependent on the dominant bedrock permeability.

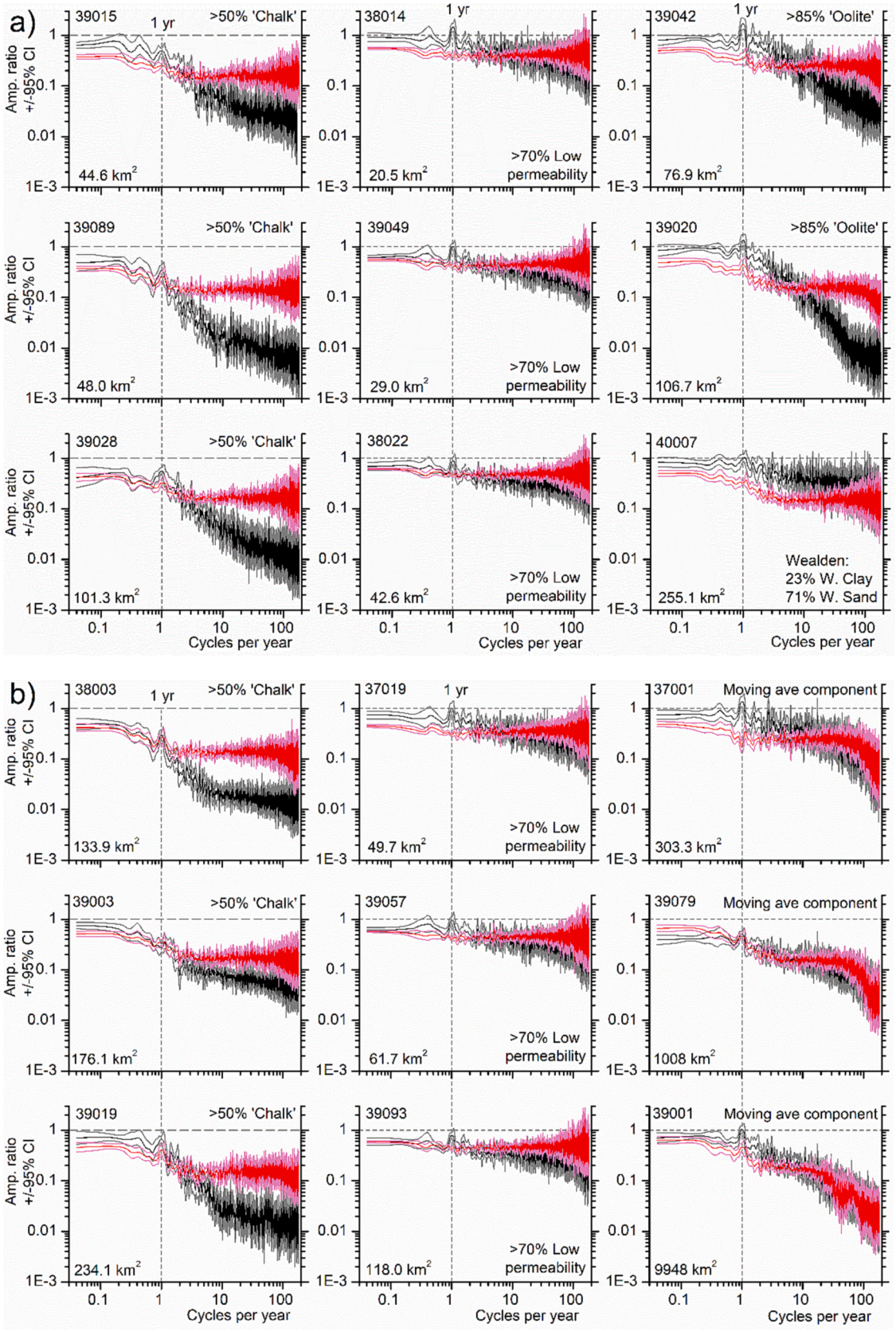


Fig. 5. A) Examples of amplitude ratio spectra for observed discharge v. precipitation (black) and modelled discharge v. precipitation (red). b) Further examples of amplitude ratio spectra for 39001 the amplitude ratio spectrum is for the naturalised observed flow (virtually identical to result for measured flow).

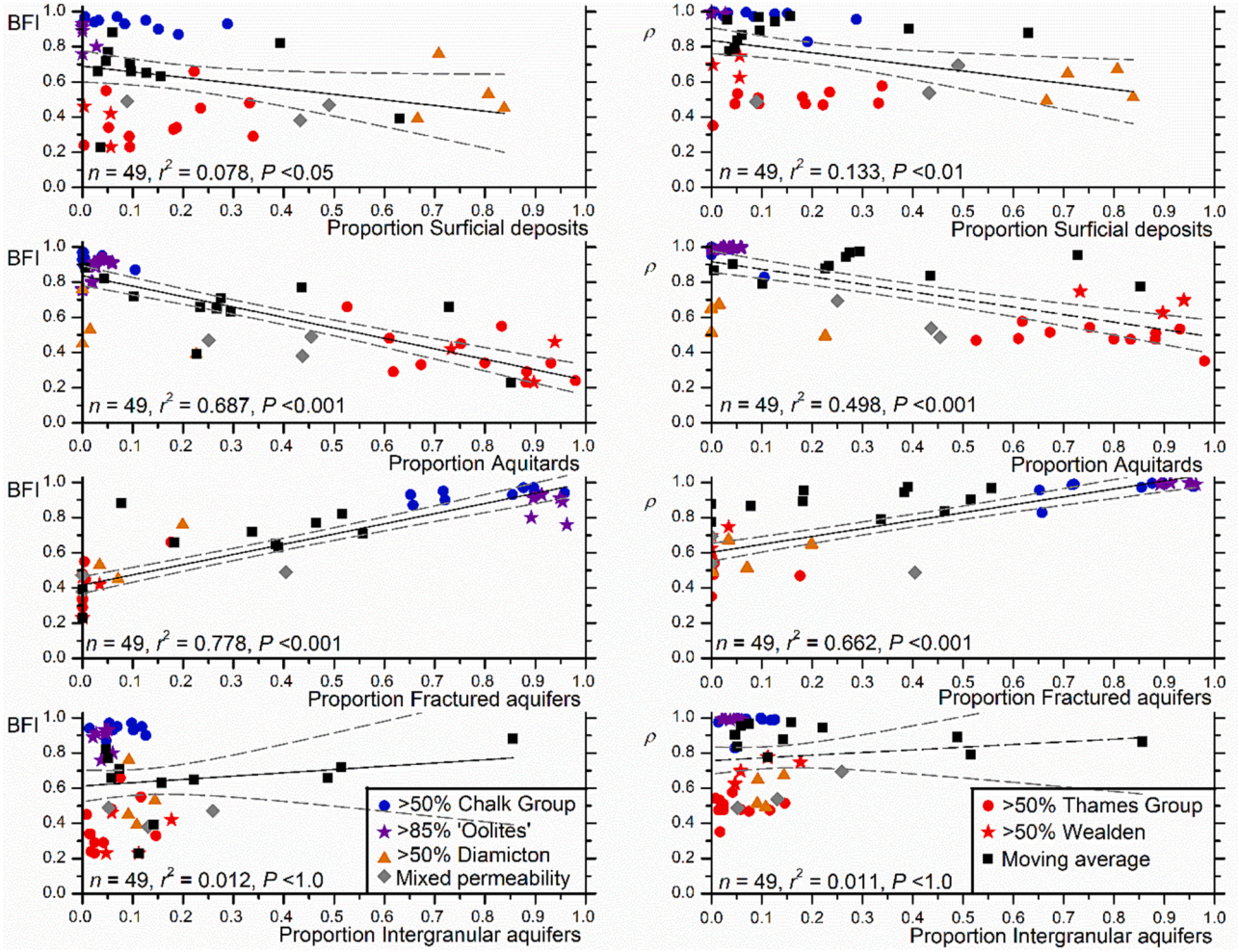


Fig. 6. Left: Multiple regression of (left) BFI and (right) ρ v. the proportion of the four permeability classes (Fig. 2b).

Table 3 lists the KGE when comparing JULES discharge to the observed discharge. Knoben et al. (2019) showed that values of KGE as low as -0.41 can be considered to represent a model performance better than the observational benchmark. Of the sub-catchments dominated by >50% Chalk Group, 6 out of 11 are less than this threshold (down to -1.996), and of the remainder the highest KGE is -0.214 (Fig. 7a). Of the 7 cases with >85% Oolite groups, the KGE values are all better than for Chalk (-0.195 to +0.376). The single sub-catchment with >75% Lower Greensand, i.e. an intergranular aquifer (at gauge 39209) has a KGE of -1.304.

Cases with >70% low permeability bedrock have one case below the threshold (39005 at -1.168), but the remaining 18 range from -0.118 to +0.572. Sub-catchments with a mixture of permeability classes range from -0.346 to +0.833. There is almost no difference in KGE for the Thames at Kingston (gauge 39001) comparing JULES discharge with that measured (+0.711) or with the observations naturalised (+0.712) despite the better match in mean flow with the naturalised data (mean bias error contributes to the calculation of KGE). Sub-catchments exhibiting moving average characteristics (Table 3) are mostly dominated by a mixture of formations as they generally relate to the large basins. Therefore, in general the worst overall performance relates to sub-catchments dominated by Chalk (Fig. 7a).

The preponderance of modelled discharge with spectra that are too flat, reflecting too much 'flashiness', is shown by the lag1 autocorrelation of the JULES discharge versus that observed (Fig. 7b). In the largest sub-catchments a moving average component is indicated by a roll off in power on the spectra of both the observations and the JULES output (Fig. 5b). The two cases with the largest catchment area (39001, 39079) both involve a moving average component and the lag1 autocorrelation of the modelled discharge is a reasonable match to that observed (Fig. 7b, Table 3). This apparently indicates that the routing used in the modelling, by combining the variability of discharge from upstream grid boxes (older variability) with the current variability within a grid box (due to current precipitation events and direct surface runoff), replicates the moving averaging of different aged variability caused by the mixing of upstream water and groundwater with water entering from direct surface runoff in the real channel.

Fig. 7c shows that, at the annual scale, allowing for the 95% confidence intervals of the amplitude ratios for the JULES output relative to the observations, in most cases there is too little baseflow variability modelled. This is especially the case for the gauge sites where the Oolite groups dominate the bedrock. None of the sub-catchments dominated by low permeability rocks (Thames and Wealden Groups and Diamicton) have the correct amplitude of baseflow variability at the annual scale.

Fig. 7d shows the delay averaged across the whole spectrum, at frequencies where the uncertainty of the phase difference uncertainty is $<\pm 45°$ (i.e. where coherency >95%, Section 3.2), for JULES output versus observations. Allowing for the 95% confidence intervals (barely visible in Fig. 7d), one sub-catchment dominated by the Chalk Group and one with a moving average component have the right delay (i.e. values sitting on the 1:1 dashed line). In two cases there is too much delay, but for the remaining 45 cases the delay is too small (Fig. 7d).

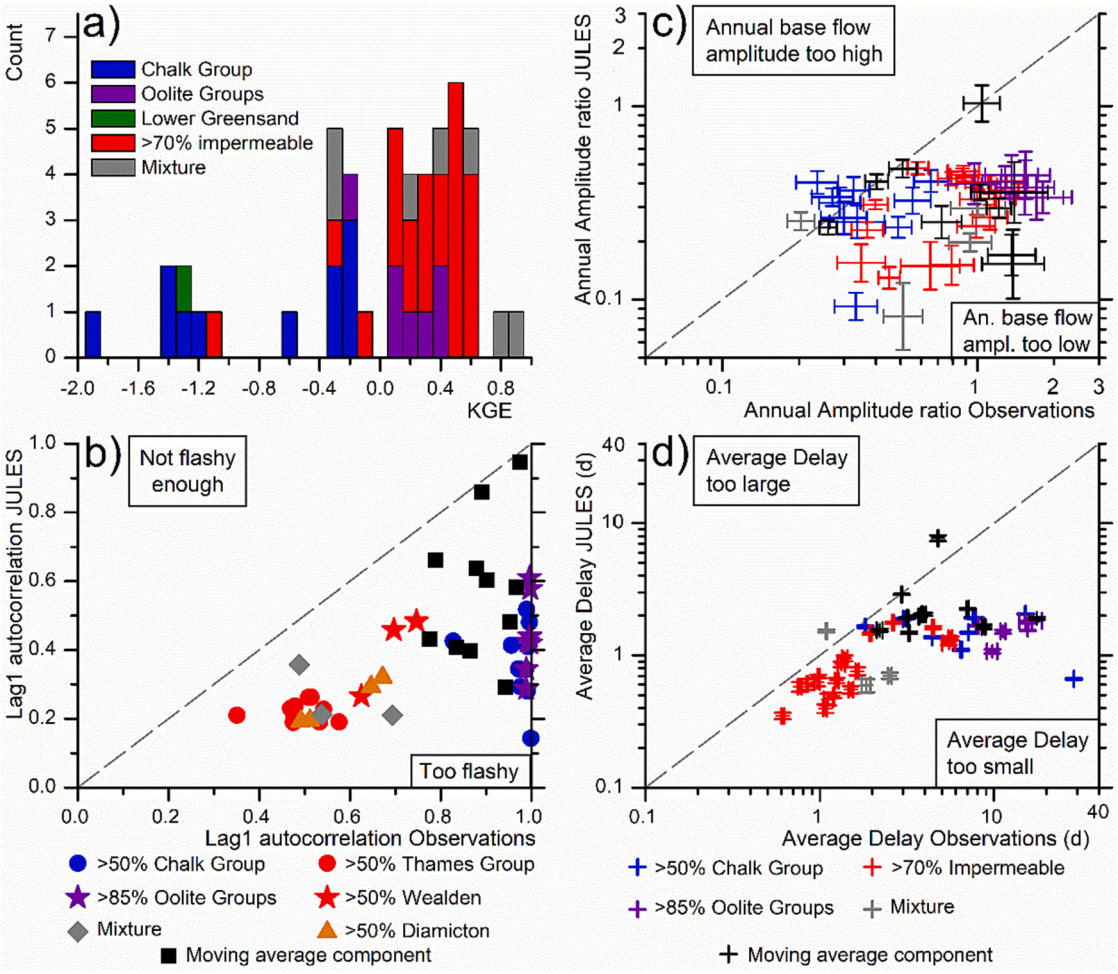


Fig. 7. Metrics of JULES performance versus observations. Grey dashed lines in b), c) and d) indicate 1:1 lines. The errors bars in c) and d) show the 95% confidence intervals.

5.2. Saturated hydraulic conductivity

The results in Section 5.1 are consistent with JULES generally representing: a) too little attenuation of high frequency variability, especially in sub-catchments dominated by high permeability bedrocks, b) too little baseflow variability at the annual scale and c) too little delay in variability over all time scales. This suggests insufficient infiltration of precipitation into the subsurface due to the model being parameterized with permeability that is too low. The spatially varying hydraulic conductivity at saturation set via the parameter *satcon* in JULES is designed to represent the behaviour of the soils. The parameterization derives from a pedotransfer function (Cosby et al., 1984) according to gridded observations of the percentage sand and clay within the soils. This limits the range of saturated conductivities applied in the model to those for 100% clay up to those for 100% sand, or a range from about 0.035 to about 3.0 m d^{-1}. The *satcon* values by grid box as used in the JULES runs in the sub-catchments studied are shown in Fig. 8a.

Bloomfield et al. (2009) compiled typical ranges and average hydraulic conductivity for the rock formations in the Thames Basin (their Table 5). This indicates aquitards (Diamicton, Thames and Wealden groups) with a range of about 0.0000001 and 1.0 m d^{-1} and aquifers (Chalk Group, Oolite groups, Lower Greensand) with range of about 1.0 to 200 m d^{-1} (Fig. 8b). Fig. 8b shows the average hydraulic conductivity of rocks in the Thames Basin and that of the much more permeable alluvium (Allen et al., 1997; Jones et al., 2000). Although the *satcon* values corresponding to the Oolite Groups (Cotswolds in the northwest of Fig. 8a), and the Chalk Group outcrops (e.g. Berkshire Downs and Chilterns band running southwest to northeast in the centre of Fig. 8a) at about 0.5 m d^{-1} are higher than adjacent areas – they are far lower than the average hydraulic conductivity of these units (>10.0 m d^{-1}, Fig. 8b).

A simple approach to influencing the modelled discharge variability would be to examine the effects of changing the *satcon* values to match the hydraulic conductivity of the underlying rock. Theoretically this could compromise the behaviour of the model since it might produce physical inconsistencies between *satcon* and the other parameters that are also determined by the pedotransfer function from the proportion of sand and clay. In practice, we experimented for individual grid boxes where *satcon* (Fig. 8a) is less than the bedrock saturated conductivity (*ksat*) as shown in Fig. 8b, by setting *satcon* = *ksat* for all four soil layers in JULES. Such a change has a very substantial influence on the relative proportion of surface and subsurface runoff. On the other hand, it was found that soil moisture is little changed where *satcon* = *ksat* and that evapotranspiration is hardly affected (<2% change in annual total latent heat). In many sub-catchments dominated by Chalk these changes to *satcon* had the effect of increasing the diagnosed water table depth (*zw*) to the pre-set maximum *zw_max*. In JULES when *zw* = *zw_max* subsurface runoff stops. Hence, it was necessary to reset *zw_max* to 15 m (instead of 6 m in the RAL3 configuration).

For catchments predominantly underlain by chalk and other aquifer limestones such as the Oolites soil thicknesses are frequently <0.5 m according to the 1 km^2 resolution MySoil map of soil thicknesses (Lawley, 2012). However, the RAL3 configuration of JULES uses total soil thickness of 3 m. In these settings, by setting *satcon* = *ksat* for all four soil layers, the permeability of the soil is substantially increased –

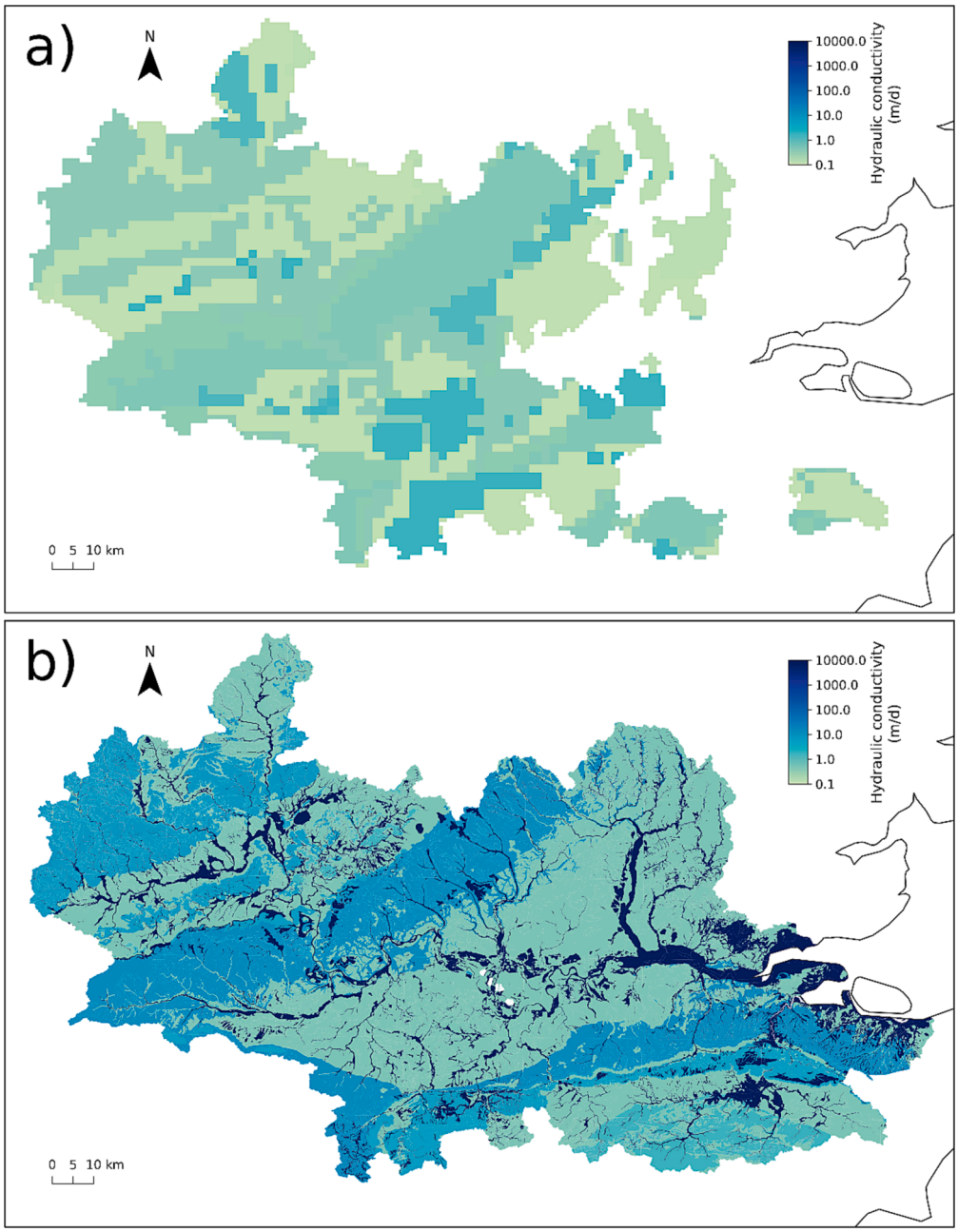


Fig. 8. A) *satcon* values used for soils in JULES. b) Average hydraulic conductivity of bedrock and overlying alluvium in the Thames Basin. British Geological Survey © UKRI 2023.

mimicking the effect of very shallow soils and macropores extending down to the bedrock.

The effects of increasing *satcon* are most pronounced for the catchments dominated by chalk and are shown in Fig. 9: there is far greater attenuation of high frequency precipitation variability than before (compare the blue to the red lines). In general, increasing *satcon* to match the bedrock *ksat* leads to a substantially better match with the observed amplitude ratio spectrum at medium and high frequencies. At the lower frequencies for the sub-catchments >50 km^2 the attenuation by JULES is about right.

However, for very small catchments the RAL3 configuration has too much attenuation at low frequencies including at the annual scale. This problem is exaggerated in the *satcon* = *ksat* results especially in the grid boxes corresponding to gauges 39065, 39015 and 39089. This apparently indicates that in such small sub-catchments with very few model grid boxes between which water is exchanged/routed (catchment sizes ranging from 13.4 to 48.0 km^2), there is insufficient opportunity to accumulate discharge variability from upstream and therefore insufficient accumulation of the low frequency/baseflow variability.

Fig. 10 illustrates examples of observed and modelled discharge hydrographs (time series) for a variety of sub-catchments (i.e. varying in terms of bedrock permeability, BFI and area). For a Chalk-dominated sub-catchment, Lambourn at Shaw with gauge 39019, the *satcon* = *ksat* experiment (shown in blue) results in a large decrease in discharge variability compared to the original JULES run (in red). The KGE improved substantially, rising from -0.216 to +0.644, demonstrating that hydraulic conductivity plays a key role when modelling discharge variability. Note that there remain important deficiencies in this catchment (though not others in the experiment) in that there is insufficient delay at the annual scale and, probably linked to too much

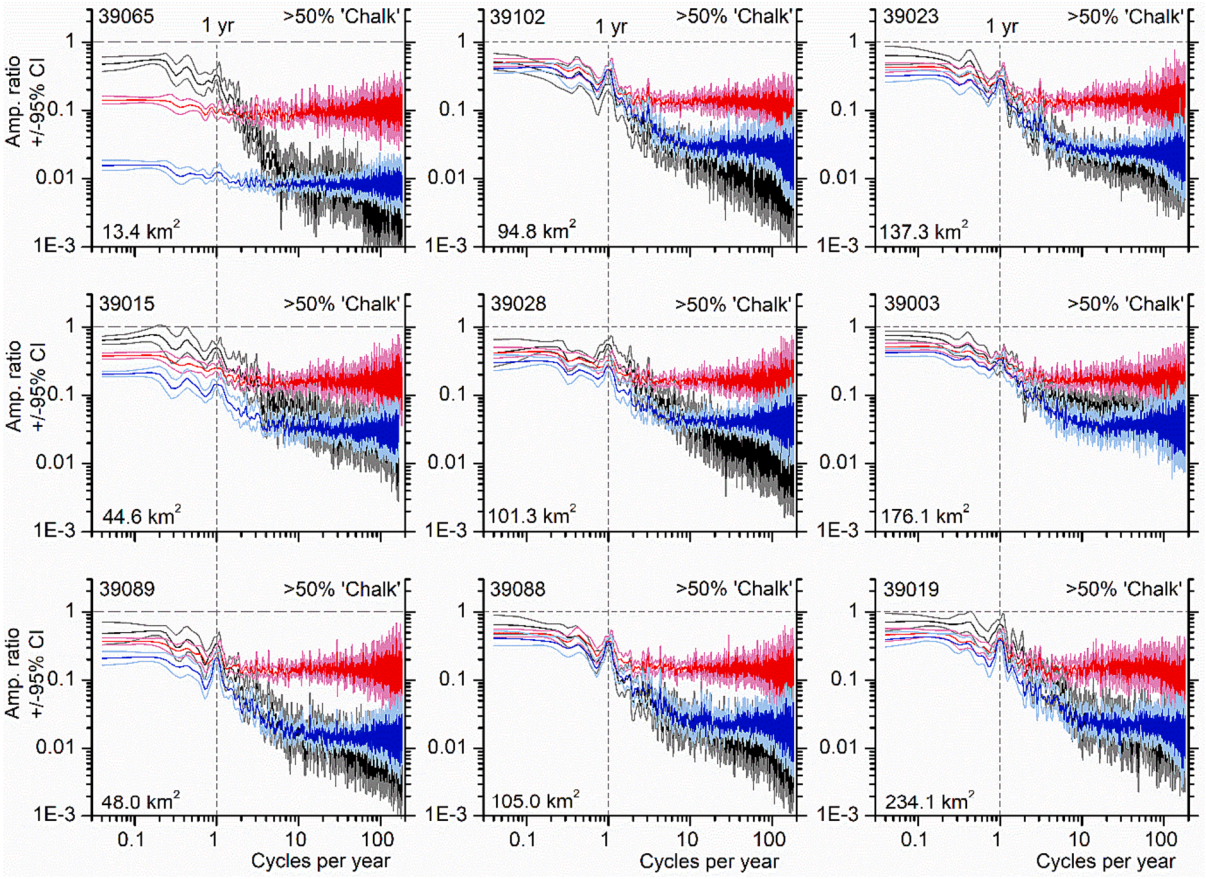


Fig. 9. Amplitude ratio spectra of discharge versus precipitation for: black = observations, red = JULES, blue = JULES using bedrock saturated conductivity (*ksat*) in place of soil *satcon* for grid boxes where *satcon* < *ksat*.

evapotranspiration, the mean flow is too small.

For an example of a sub-catchment with >85% Oolite groups (Leach at Priory Mill, Lechlade, gauge 39042), Fig. 10 shows how the original JULES run is far too flashy as for the Chalk example. Using *satcon* = *ksat* for grid boxes underlain by Oolite leads to substantial improvements by supressing the very short-term variability: KGE rises from +0.068 to +0.322. Nevertheless, there is insufficient variation in discharge at the monthly to annual scales and the mean flow is too low. The Oolite groups consist of interbedded units of fractured aquifers (oolitic limestone) and thin aquitards (mudrocks) so that, as found here, using the values for *satcon* selected do not account for the longer-term discharge variability.

Fig. 10 also shows that, even for a sub-catchment with a low BFI (Silk Stream at Colindeep Lane, gauge 39049) linked to >70% low permeability bedrock the original JULES run is too flashy though at +0.204 the KGE is far higher than for the Chalk and Oolite examples. Despite the small proportion of grid boxes with bedrock having higher permeability than the soil *satcon*, the KGE increases a little to +0.242 for the *satcon* = *ksat* experiment. A very similar result is found for a sub-catchment underlain by approximately equal proportions (Table 2) of aquitard and fractured aquifer (Ravensbourne at Catford Hill, gauge 39056).

The model does best for the Thames to Kingston (gauge 39001) with a KGE of +0.711 and a lag1 autocorrelation that is about right (Table 3). This is the largest basin modelled (9,948 km^2) where routing between the numerous grid boxes dominates the attenuation of precipitation variability – fairly successfully reproducing the real-world moving-average attenuation caused by mixing of upstream with downstream water. As expected, the modelled JULES discharge, lacking anthropogenic effects, is a better match to the 'naturalised' discharge for 39,001 than the observed discharge for the amplitude ratio at the annual scale and for the average delay even though the KGE is hardly affected (Table 3). However, during low flow the original model-run indicates an excessive response to rainfall events in this relatively large basin that is not seen in the observations. When *satcon* is increased to *ksat* of the bedrock for the approximately half of the grid cells dominated by aquifers (Table 2) KGE drops to +0.571. In that case at times (e.g. mid 1999 and mid 2002) the small-amplitude short-term variability matches the observations better than the original model run, but at other times (e.g. 2000 to 2001 and 2002 to 2003) the short-term variability is too low and the baseflow is too high. Hence, for this catchment to implement *satcon* = *ksat* over aquifer-dominated grid boxes, model improvements require both an increase in flashiness over grid boxes dominated by aquitards (to improve variability) and a drop in the total evapotranspiration (to improve the mean bias error).

5.3. Previous hydrological modelling using JULES over Chalk

Since Chalk-dominated sub-catchments are associated with the poorest performance by JULES it is worth quickly reviewing chalk hydrology and prior modelling efforts. Chalk matrix consists primarily of calcite nannoplankton coccoliths resulting in a high porosity (20 – 45%), but because the majority of pore throats are <1 μm wide (Price et al, 2000), the matrix permeability is very low (<0.01 m d^{-1}, Price et al., 1993). The presence of cracks and fissures in the Chalk results in a low bulk porosity (<2%, Worthington et al., 2019; Price et al., 1993), but much higher bulk permeability or hydraulic conductivity in the unsaturated zone (>0.1 m d^{-1}, Price et al., 2000).

Conventional recharge estimation underestimates the rates observed in Chalk since this typically occurs throughout the summer despite the presence of soil moisture deficits (Rushton and Ward, 1979). Chalk

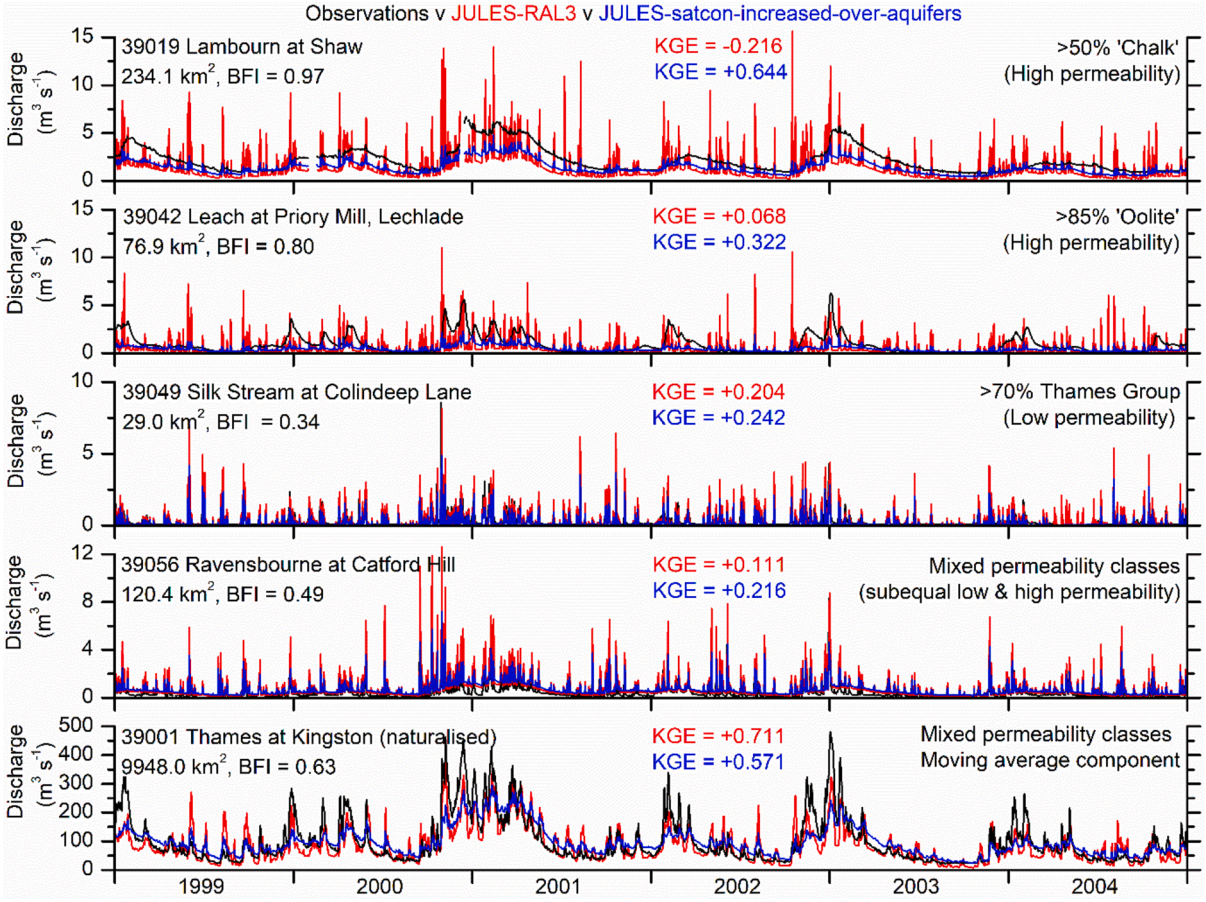


Fig. 10. Examples of observed (black) and modelled (red or blue) discharge for 1999 to 2004. Note 39,001 shows naturalised observations (slightly higher at all times than observed).

exhibits dual-porosity and dual-permeability in the unsaturated zone above the water table, with slow, albeit varying, aquifer recharge from the matrix throughout the year plus episodic fast recharge via fractures that have been enlarged by dissolution (e.g. Ireson et al., 2009; Ireson and Butler, 2011; Lee et al., 2006; Price et al., 2000; Worthington et al., 2019).

There have been two studies representing the unusual nature of Chalk recharge within JULES both targeting the Kennett sub-catchment in the west of the Thames Basin. Le Vine et al. (2016) used JULES (plus PDM, Moore, 1985, rather than TOPMODEL) with two intersecting Brooks-Corey curves of soil moisture retention to represent the two distinct flow domains combined with a three-dimensional groundwater flow model. Their study was intended to understand the problems of producing more realistic Chalk hydrology, but the hydrographs of 10-day average flow values for the Lambourn at Shaw are far more 'flashy' than the observations (their Fig. 8).

Rahman and Rosolem (2017) introduced the Bulk Conductivity (BC) modification to JULES. With the soil interval limited to 0.3 m total depth instead of 3 m, combined with a 'bulk saturated conductivity' (intended to be representative of Chalk) imposed above a threshold soil moisture content, leads to increased infiltration. Both the soil moisture threshold and increased bulk hydraulic conductivity are obtained by calibration. Although the BC modification produced a seasonal cycle in recharge to the aquifer that was considered to be much more realistic than the standard JULES performance, in detail it shows no drainage through the summer (their Fig. 8) – contrary to observations in the same catchment (Fig. 13 of Ireson et al., 2009; Ireson and Butler, 2011). The annual cycle of discharge, estimated using monthly averages, is far better in terms of amplitude and phase using the BC modification, but daily modelled discharge is required to allow a full spectral assessment as used here.

The results described in this paper indicate that a change is needed to JULES in the overall hydraulic conductivity particularly for sub-catchments dominated by high-permeability bedrock. Early on Rushton and Ward (1979) considered the possibility of representing groundwater recharge within Chalk as including a proportion of rainfall moving as bypass flow directly into the bedrock. A physically-based model for flow in the unsaturated zone of the Chalk based on observations successfully represents short-term variations in water flow in unsaturated Chalk between 0.1 and 35 m depth, as well as the water table depths, in the Pang and adjacent Lambourn sub-catchments (e.g. Ireson et al., 2009; Ireson and Butler 2011). However, this model requires setting 17 parameters whereas direct comparison with a simple bypass model showed water table variations that are almost as good (Ireson and Butler, 2013). The simple bypass model gets the water table variation correct (due to the bypassing) but observations suggest that the summer recharge is mainly due to prolonged slow recharge from the matrix even during soil moisture deficits rather than from the rather rare episodic recharge and preferential flow down fractures following intense rain events (Ireson and Butler, 2011; 2013).

The *satcon* = *ksat* experiment demonstrates that increases in hydraulic conductivity to well above the ancillary soil values in grid boxes dominated by aquifers leads to better simulation of discharge variability (Figs. 9 and 10). Note that by changing all four soil layers extending to 3 m in the model, resetting *satcon* to the bedrock *ksat* cause large increases in permeability (mimicking soils much thinner than 3 m and soil macropores) so that a much higher proportion of precipitation is transferred to sub-surface runoff. Despite the improvements in terms of variability due to increased bypass flow, as found for the previous studies the

changes in *satcon* do not lead to improvements in delay mismatches.

The simple expedient of increasing *satcon* to bedrock *ksat* values over aquifers in JULES leads to substantial improvements in discharge variability for the basins less than about 300 km^2 where attenuation of precipitation variability is dominated by infiltration processes instead of by moving averaging due to routing between grid boxes. Potentially future implementation of LEAF2-Hydro (Fan et al., 2007; Miguez-Macho et al., 2007) will lead to improvements in the water table and soil moisture representation as it permits lateral groundwater movement and hydraulic conductivity in the bedrock independent of that in the soil (pers. comm. with A. Martínez-de la Torre, 2022). This could lead to further improvements in the representation of discharge variability.

6. Conclusions

Lag1 autocorrelation (ρ) provides a useful first-order description of spectral shape when the data are dominated by a first-order autoregressive process - for example with discharge time series from relatively small catchments (<300 km^2). In this situation the serial correlation can be related physically to the high-frequency attenuation of precipitation variability in streamflow that is derived from soil and particularly groundwater. Hence a greater proportion of streamflow from groundwater is associated with a higher ρ and a larger BFI.

Spurious increases in ρ due to data interpolation can be avoided by using a transform algorithm designed for processing missing and/or irregularly spaced data. The presence of regular components, in this case an annual cycle in discharge due to the seasonal cycle in evapotranspiration, causes a slight increase in ρ. However, a moving average component, recognisable from high-frequency roll-off, causes a large increase in ρ and invalidates its use as a description of spectral background shape. In the Thames Basin a moving average spectral character is observed in some sub-catchments >150 km^2 and in all sub-catchments >300 km^2. A moving average component in observed streamflow can be explained in terms of longer rivers and larger catchments where streamflow from direct runoff (exhibiting current precipitation variability) is mixed with streamflow from upstream and from groundwater (exhibiting preceding precipitation variability).

Lag1-autocorrelation of observed streamflow, the average delay, and the high-frequency amplitude ratio of discharge versus precipitation variability are all correlated with the proportion of the various bedrock permeability classes. Sub-catchments dominated by high permeability rocks (especially fractured aquifers like the Chalk) have, from observations, high ρ, substantial attenuation of high-frequency variability independent of moving averaging, long average delays at all frequencies, and high BFI.

For large catchments the moving average component is reproduced relatively well in JULES due to the effect of routing of surface and subsurface runoff between model grid boxes. In terms of discharge variability JULES performs least well for small sub-catchments dominated by high permeability bedrock. Experimentation confirms that in aquifer-dominated catchments increased saturated hydraulic conductivity leads to improved discharge variability – particularly with improved high-frequency attenuation of precipitation variability.

The results explain how despite reasonably good representation of discharge variability in large basins (>300 km^2) a model may not correctly represent variability in small catchments without allowing for bedrock permeability for example if groundwater is not modelled. For small catchments in data-sparse regions, knowledge of the relative proportions of different hydrogeological units (aquifers, aquitards) potentially could be used to predict and model discharge variability as characterised by BFI and lag1 autocorrelation.

CRediT authorship contribution statement

Graham P. Weedon: Conceptualization, Methodology, Data curation, Software, Formal analysis, Visualization, Writing – original draft, Writing – review & editing. **Emma L. Robinson:** Methodology, Software, Formal analysis, Visualization, Writing – review & editing. **John P. Bloomfield:** Methodology, Writing – review & editing. **Stephen Turner:** Data curation, Writing – review & editing. **Emily Crane:** Visualization, Writing – review & editing. **Martin J. Best:** Conceptualization, Methodology, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work has been supported by the Natural Environment Research Council award number NE/S017380/1 as part of the Hydro-JULES programme. JPB and EC publish with permission of the Executive Director, British Geological Survey, (NERC). Data from the UK National River Flow Archive. Thanks to three anonymous reviewers for their suggestions for improvements to the manuscript.

Data and code availability.

The observations of daily precipitation and discharge are freely available under the relevant NFRA gauge code at https://nrfa.ceh.ac.uk/data/search .

The code used for the cross-spectral analysis, csagan1.1.f, is freely available for anyone to use from the Met Office Science Repository Service at https://code.metoffice.gov.uk/trac/lmed/browser%23main/trunk/benchmarking. Access requires registration for an account and this will be supported by a member of the JULES group. Requests for new accounts can be made by emailing Jules-Support@metoffice.gov.uk with details of the user's name, email address, institution and purpose for requiring access.

References

Allen, D.J., Brewerton, L.J., Coleby, L.M., Gibbs, B.R., Lewis, M.A., MacDonald, A.M., Wagstaff, S.J., Williams, A.J., 1997. The physical properties of minor aquifers in England and Wales. British Geological Survey Technical Report WD/97/34, Environment Agency R&D Publ. 8, 312 pp.

Beck, H.E., van Dijk, A.I.J.M., Miralles, D.G., de Jeu, R.A.M., Bruijnzeel, L.A.S., McVicar, T.R., Schellekens, J., 2013. Global patterns in base flow index and recession based on streamflow observations from 3394 catchments. Water Resour. Res. 49, 7843–7863. https://doi.org/10.1002/2013WR013918.

Bell, V.A., Kay, A.L., Jones, R.G., Moore, R.J., 2007. Development of a high resolution grid-based river flow model for use with regional climate model output. Hydrol. Earth Syst. Sci. 11, 532–549. https://doi.org/10.5194/hess-11-532-2007.

Best, M.J., Pryor, M., Clark, D.B., Rooney, G.G., Essery, R.L.H., Ménard, C.B., Edwards, J. M., Hendry, M.A., Porson, A., Gedney, N., Mercado, L.M., Sitch, S., Blyth, E., Boucher, O., Cox, P.M., Grimmond, C.S.B., Harding, R.J., 2011. The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes. Geosci. Mod. Dev. 4 (3), 677–699.

Beven, K.J., Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. Hydrol. Sci Bull. 24, 43–69. https://doi.org/10.1080/02626667909491834.

Bloomfield, J.P., Allen, D.J., Griffiths, K.J., 2009. Examining geological controls on baseflow index (BFI) using regression analysis: an illustration from the Thames Basin. UK. J. Hydrol. 373, 164–176 https://doi.org/j.hydrol.2009.04.025.

Bloomfield, J.P., Bricker, S.H., Newell, A.J., 2011. Some relationships between lithology, basin form and hydrology: a case study from the Thames Basin. UK. Hydrol. Proc. 25, 2518–2530. https://doi.org/10.1002/hyp.8024.

Bloomfield, J.P., Gong, M., Marchant, B.P., Coxon, G., Addor, N., 2021. How is Baseflow Index (BFI) impacted by water resource management practices? Hydrol. Earth Syst. Sci. 25, 5355–5379. https://doi.org/10.5194/hess-25-5355-2021.

Blyth, E., Clark, D.B., Ellis, R., Huntingford, C., Los, S., Pryor, M., Best, M., Sitch, S., 2011. A comprehensive set of benchmark tests for a land surface model of simultaneous fluxes of water and carbon at both the global and seasonal scale. Geosci. Model Dev. 4, 255–269. https://doi.org/10.5194/gmd-4-255-2011.

Brooks, R.H., Corey, A.T., 1964. Hydraulic properties of porous media. Colorado State University Papers 3, 37 pp. https://mountainscholar.org/bitstream/handle/10217/197982/CERF_63_24.pdf?sequence=1.

Bush, M., Allen, T., Bain, C., Boutle, I., Edwards, J., Finnenkoetter, A., Franklin, C., Hanley, K., Lean, H., Lock, A., Manners, J., Mittermaier, M., Morcrette, C., North, R., Petch, J., Short, C., Vosper, S., Walters, D., Webster, S., Weeks, M., Wilkinson, J., Wood, N., Zerroukat, M., 2020. The first Met Office Unified Model-JULES Regional Atmosphere and Land configuration, RAL1. Geosci. Model Dev. 13, 1999–2029. https://doi.org/10.5194/gmd-13-1999-2020.

Chen, H., Teegavarapu, R.S.V., 2020. Comparative analysis of four baseflow separation methods in the South Atlantic-Gulf Region of the U.S. Water 12, 17 pp. https://doi.org/10.3390/w12010120.

Chou, H.-K., Avila, A.M.H., Bray, M., 2022. Evaluating the Atibaia River hydrology using JULES6.1. Geosci. Model Dev. 15, 5233–5240. https://doi.org/10.5194/gmd-15-5233-2022.

Cosby, B.J., Hornberger, G.M., Clapp, R.B., Ginn, T.R., 1984. A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. Water Resour. Res. 20, 682–690. https://doi.org/10.1029/WR020i006p00682.

Davies, H.N., Rameshwaran, P., Bell, V.A., 2022. Gridded (1km) physical river characteristics for the UK. NERC EDS Environmental Information Data Centre. https://nora.nerc.ac.uk/id/eprint/532361.

Eckhardt, K., 2008. A comparison of baseflow indices, which were calculated with seven different baseflow separation methods. J. Hydrol. 352, 168–173. https://doi.org/10.1016/j.hydrol.2008.01.005.

Fan, Y., Miguez-Macho, G., Weaver, C.P., Walko, R., Robock, A., 2007. Incorporating water table dynamics in climate modelling: 1. Water table observations and equilibrium water table simulations. J. Geophys. Res. 112, D102125. htpps://doi.org/10.1029/2006JD008111.

Fleming, S.W., 2014. A non-uniqueness problem in the identification of power-law scaling for hydroclimatic time series. Hydrol. Sci. J. 59, 73–84. https://doi.org/10.1080/02626667.2013.851384.

Gedney, H., Cox, P.M., 2003. The sensitivity of global climate model simulations to the representation of soil moisture heterogeneity. J. Hydrometeorol. 4, 1265–1275. https://doi.org/10.1175/1525-7541(2003)004<1265:TSOGCM>2.0.CO;2.

Gilman, D.L., Fuglister, F.J., Mitchell, J.M., 1963. On the power spectrum of 'red noise'. J. Atmos. Sci. 20 (2), 182–184.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. J. Hydrol. 377, 80–91. https://doi.org/10.1016/j.hydrol.2009.08.003.

Gustard, A., Bullock, A. and Dixon, J.M., 1992. Low flow estimation in the United Kingdom. Institute of Hydrology Report No, 108, 292 pp, https://nora.nerc.ac.uk/id/eprint/6050/1/IH_108.pdf.

Habib, A., Sorensen, J.P.R., Bloomfield, J.P., Muchan, K., Newell, A.J., Butler, A.P., 2017. Temporal scaling phenomena in groundwater-floodplain systems using robust detrended fluctuation analysis. J. Hydrol. 549, 715–730 https://doi.org/10.2017.04.034.

Haddeland, I., Clark, D.B., Franssen, W., Ludwig, F., Voβ, F., Arnell, N., Bertrand, N., Best, M., Folwell, S., Gerten, D., Gomes, S., Gosling, S.N., Hagemann, S., Hanasaki, N., Harding, R., Heinke, J., Kabat, P., Koirala, S., Oki, T., Polcher, J., Stacke, T., Viterbo, P., Weedon, G.P., Yeh, P., 2011. Multimodel estimate of the global terrestrial water balance: setup and first results. J. Hydrometeorol. 12, 869–884. https://doi.org/10.1175/2011JHM1324.1.

Hale, V.C., McDonnell, J.J., 2016. Effect of bedrock permeability on stream base flow mean transit time scaling relations: 1. A multiscale catchment intercomparison. Water Resour. Res. 52, 1358–1374. https://doi.org/10.1002/2014WR016124.

Hale, V.C., McDonnell, J.J., Stewart, M.K., Solomon, D.K., Doolittle, J., Ice, G.G., Pack, R. T., 2016. Effect of bedrock permeability on stream base flow mean transit time scaling relationships: 2. Process study of storage and release. Water Resour. Res. 52, 1375–1397. https://doi.org/10.1002/2015WR017660.

Hosking, J.R.M., 1984. Modeling persistence in hydrological time series using fractional differencing. Water Resour. Res. 20 (12), 1898–1908.

Hurst, H.E., 1951. Long-term storage capacity of reservoirs. Trans. Amer. Soc. Civ. Eng. 116 (1), 770–799.

Ifeachor, E.C., Jervis, B.W., 1993. Digital Signal Processing: A Practical Approach. Addison-Wesley, Harlow, England, p. 760.

Ireson, A.M., Butler, A.P., 2011. Controls on preferential recharge to Chalk aquifers. J. Hydrol. 398, 109–123 https://doi.org/j.jhydrol.2010.12.015.

Ireson, A.M., Butler, A.P., 2013. A critical assessment of simple recharge models: application to the UK Chalk. Hydrol. Earth Syst. Sci. 17, 2083–2096. https://doi.org/10.5194/hess-17-2083-2013.

Ireson, A.M., Mathias, S.A., Wheater, H.S., Butler, A.P., Finch, J., 2009. A model for flow in the chalk unsaturated zone incorporating progressive weathering. J. Hydrol. 365, 244–260 https://doi.org/j.jhydrol.2008.11.043.

Jenkins, G.M., Watts, D.G., 1969. Spectral Analysis and its Applications. Holden Day, Cambridge, UK, p. 525.

Jones, H.K., Morris, B.L., Cheney, C.S., Brewerton, L.J., Merrin, P.D., Lewis, M.A., MacDonald, A.M., Coleby, L.M., Talbot, J.C., McKenzie, A.A., Bird, M.J., Cunningham, J. and Robinson, V.K., 2000. The Physical Properties of Major Aquifers in England and Wales. British Geological Survey Technical Report WD/00/4, Environment Agency R&D Publ. 68, 234 pp, https://nora.nerc.ac.uk/id/eprint/13137/1/WD97034.pdf.

Keller, V.D.J., Tanguy, M., Prosdocimi, I., Terry, J.A., Hitt, O., Cole, S.J., Fry, M., Morris, D.G., Dixon, H., 2015. CEH-GEAR: 1 km resolution daily and monthly areal rainfall estimates for the UK for hydrological and other applications. Earth Syst. Sci. data 7 (1), 143–155.

Kling, H., Fuchs, M., Paulin, M., 2012. Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. J. Hydrol. 424–425, 264–277. https://doi.org/10.1016/j.hydrol.2012.01.011.

Knoben, W.J.M., Freer, J.E., Woods, R.A., 2019. Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores. Hydrol. Earth Syst. Sci. 33, 4322–4331. https://doi.org/10.5194/hess-23-3423-2019.

Kuang, X., Jiao, J.J., Zheng, C., Cherry, J.A., Li, H., 2020. A review of specific storage in aquifers. J. Hydrol. 581, 124383 https://doi.org/10.1016/j.jhydrol.2019.124383.

Lacey, G.C., Grayson, R.B., 1998. Relating baseflow to catchment properties in south-eastern Australia. J. Hydrol. 204, 231–250. https://doi.org/10.1016/S0022-1694(97)00124-8.

Lane, R.A., Coxon, G., Freer, J.E., Wagener, T., Johnes, P.J., Bloomfield, J.P., Greene, S., Macleod, C.J.A., Reaney, S.M., 2019. Benchmarking the predictive capability of hydrological models for river flow and flood peak predictions across over 1000 catchments in Great Britain. Hydrol. Earth Syst. Sci. 23, 4011–4032. https://doi.org/10.5194/hess-23-4011-2019.

Lawley, R., 2012. Soil parent material 1 kilometre dataset. British Geological Survey Internal Report OR/14/025, 20 pp.

Le Vine, N., Butler, A., McIntyre, N., Jackson, C., 2016. Diagnosing hydrological limitations of a land surface model: application of JULES to a deep-groundwater chalk basin. Hydrol. Earth Syst. Sci. 20, 143–159. https://doi.org/10.5194/hess-20-143-2016.

Lee, L.J.E., Lawrence, D.S.L., Price, M., 2006. Analysis of water-level responses to rainfall and implications for recharge pathways in the Chalk aquifer, SE England. J. Hydrol. 330, 604–620 https://doi.org/j.jhydrol.2006.04.025.

Lewis, E., Quinn, N., Blenkinsop, S., Fowler, H.J., Freer, J., Tanguy, M., Hitt, O., Coxon, G., Bates, P., Woods, R., 2018. A rule based quality control method for hourly rainfall data and a 1 km resolution gridded hourly rainfall dataset for Great Britain: CEH-GEAR1hr. J. Hydrol. 564, 930–943. https://doi.org/10.1016/j.jhydrol.2018.07.034.

Lewis, E., Quinn, N., Blenkinsop, S., Fowler, H.J., Freer, J., Tanguy, M., Hitt, O., Coxon, G., Bates, P., Woods, R., 2019. Gridded estimates of hourly areal rainfall for Great Britain (1990–2014) [CEH-GEAR1hr] v2. NERC Environmental Information Centre. https://doi.org/10.5285/fc9423d6-3d54-467f-bb2b-fc7357a3941f.

Li, Z., Zhang, Y.-K., 2007. Quantifying fractal dynamics of groundwater systems with detrended fluctuation analysis. J. Hydrol. 336, 139–146. https://doi.org/10.1016/j.jhydrol.2006.12.017.

Little, M.A., Bloomfield, J.P., 2010. Robust evidence for random fractal scaling of groundwater levels in unconfined aquifers. J. Hydrol. 393, 362–369. https://doi.org/10.1016/j.jhydrol.2010.08.031.

Longobardi, A., Villani, P., 2008. Baseflow index regionalization analysis in a Mediterranean area and data scarcity context: role of the catchment permeability index. J. Hydrol. 355, 63–75. https://doi.org/10.1016/j.hydrol.2008.03.011.

Mann, M.E., Lees, J.M., 1996. Robust estimation of background noise and signal detection in climatic time series. Clim. Change 33, 409–445. https://doi.org/10.1007/BF00142586.

Martínez-de la Torre, A., Blyth, E.M., Weedon, G.P., 2019. Using observed river flow data to improve the hydrological functioning of JULES land surface model (vn4.3) used for regional coupled modelling in Great Britain (UKC2). Geosci. Mod. Dev. 12, 765–784. https://doi.org/10.5194/gmd-12-765-2019.

Mesa, O.J., Povenda, G., 1993. The Hurst effect: the scale fluctuation approach. Water Resour. Res. 29, 3995–4002. https://doi.org/10.1029/93WR01686.

Miguez-Macho, G., Fan, Y., Weaver, C.P., Walko, R., Robock, A., 2007. Incorporating water table dynamics in climate modelling: 2. Formulation, validation, and soil moisture simulation. J. Geophys. Res. 112, D13108 https://doi.org/10.1029/2006JD008112.

Milly, P.C.D., Wetherald, R.T., 2002. Macroscale water fluxes 3. Effects of land processes on variability of monthly river discharge. Water Resour. Res. 38 (11), 17-1–17-12.

Moore, R.J., 1985. The probability-distributed principle and runoff production at point and basin scales. Hydrol. Sci. J. 30, 273–297. https://doi.org/10.1080/02626668509490989.

Mudelsee, M., 2007. Long memory of rivers from spatial aggregation. Water Resour. Res. 43, W01202. https://doi.org/10.1029/2006WR005721.

Mudelsee, M., 2010. Climate Time Series Analysis. Springer, London, UK, p. 474.

Pfister, L., Martínez-Carreras, N., Hissler, C., Klaus, J., Carrer, G.E., Stewart, M.K., McDonnell, J.J., 2017. Bedrock geology controls on catchment storage, mixing, and release: a comparative analysis of 16 nested catchments. Hydrol. Proc. 31, 1828–1845. https://doi.org/10.1002/hyp.11134.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1992. Numerical Recipes in FORTRAN. Cambridge University Press, Cambridge, UK, The Art of Scientific Computing.

Price, M., Downing, R.A., Edmunds, W.M., 1993. The Chalk as an aquifer. In: Downing, R.A., Price, M. and Jones, G.P. (eds), The Hydrogeology of the Chalk in North-West Europe, Clarendon Press, Oxford, UK.

Price, M., Low, R.G., McCann, C., 2000. Mechanisms of water storage and flow in the unsaturated zone of the Chalk aquifer. J. Hydrol. 233, 54–71. https://doi.org/10.1016/S0022-1694(00)00222-5.

Priestley, M.B., 1981. Spectral Analysis and Time Series. Academic Press, London, UK, p. 890.

Rahman, M., Rosolem, R., 2017. Towards a simple representation of chalk hydrology in land surface modelling. Hydrol. Earth Syst. Sci. 21, 459–471. https://doi.org/10.5194/hess-21-459-2017.

Robinson, E.L., Blyth, E.M., Clark, D.B., Finch, J., Rudd, A.C., 2017. Trends in atmospheric evaporative demand in Great Britain using high-resolution meteorological data. Hydrol. Earth Syst. Sci. 21, 1189–1224. https://doi.org/10.5194/hess-21-1189-2017.

Robinson, E.L., Blyth, E.M., Clark, D.B., Comyn-Platt, E., Rudd, A.C., 2020. Climate hydrology and ecology research support system meteorology dataset for Great Britain (1961–2017) [CHESS-met]. NERC Environmental Information Data Centre. https://nora.nerc.ac.uk/id/eprint/516156/.

Rushton, K.R., Ward, C., 1979. The estimation of groundwater recharge. J. Hydrol. 41, 345–361. https://doi.org/10.1016/0022-1694(79)90070-2.

Schneider, M.K., Brunner, F., Hollis, J.M., Stamm, C., 2007. Towards a hydrological classification of European soils: preliminary test of its predictive power for the base flow index using river discharge data. Hydrol. Earth Syst. Sci. 11, 1501–1513. https://doi.org/10.5194/hess-11-1501-2007.

Schulz., M. and Stattegger, K., 1997. Spectrum: spectral analysis of unevenly spaced palaeoclimatic time series. Computers & Geosci. 23, 929–945. https://doi.org/10.1016/S0098-3004(97)00087-3.

Singh, S.K., Pahlow, M., Booker, D.J., Shankar, U., Chamorro, A., 2019. Towards baseflow index characterisation at national scale in New Zealand. J. Hydrol. 568, 646–657. https://doi.org/10.1016/j.hydrol.2018.11.025.

Smakhtin, V.U., 2001. Low-flow hydrology: a review. J. Hydrol. 240, 147–186. https://doi.org/10.1016/S0022-1694(00)00340-1.

Stoelzle, M., Schuetz, T., Weiler, M., Stahl, K., Tallaksen, L.M., 2020. Beyond binary baseflow separation: a delayed-flow index for multiple streamflow contributions. Hydrol. Earth Syst. Sci. 24, 849–867. https://doi.org/10.5194/hess-24-849-2020.

Tague, C., Grant, G.E., 2004. A geological framework for interpreting the low-flow regimes of Cascade streams, Willamette River Basin. Oregon. Water Resour. Res. 40, W04303. https://doi.org/10.1029/2003WR002629.

Tanguy, M., Dixon, H., Prosdocimi, I., Morris, D.G., Keller, V.D.J., 2014. Gridded estimates of daily and monthly areal rainfall for the United Kingdom (1890–2012) [CEH-GEAR]. NERC Environmental Information Data Centre. https://doi.org/10.5285/33604ea0-c238-4488-813d-0ad9ab7c51ca.

Towler, E., McCreight, J.L., 2021. A wavelet-based approach to streamflow event identification and modelled timing error evaluation. Hydrol. Earth Syst. Sci. 25, 2599–2615. https://doi.org/10.5194/hess-25-2599-2021.

Weedon, G.P., Prudhomme, C., Crooks, S., Ellis, R.J., Folwell, S.S., Best, M.J., 2015. Evaluating the performance of hydrological models via cross-spectral analysis: case study of the Thames Basin, United Kingdom. J. Hydrometeorol. 16 (1), 214–231.

Wieder, W.R., Boehnert, J., Bonan, G.B. and Langseth, M., 2014. Regridded Harmonized Soil Database v1.2. ORNL DAAC, Oak Ridge, Tennessee, USA, https://daac.ornl.gov/cgi-bin/dnsviewer.pl?ds_id=1247.

Williams, K., Clark, D., 2014. Disaggregation of daily data in JULES. Met Office Hadley Centre Technical Note 96, 27 pp. https://nora.nerc.ac.uk/id/eprint/508002/.

Worthington, S.R.H., Foley, A.E., Soley, R.W.N., 2019. Transient characteristics of effective porosity and specific yield in aquifers. J. Hydrol. 578, 124129. https://doi.org/10.1016/j.jhydrol.2019.124129.

Zhang, Y.-K., Schilling, K., 2004. Temporal scaling of hydraulic head and river base flow and its implications for groundwater recharge. Water Resour. Res. 40, W03504. https://doi.org/10.1029/2003WR002094.