



Environment
Agency



An analysis of national macroinvertebrate trends for England, 1991–2019

Chief Scientist's Group report

October 2021

We are the Environment Agency. We protect and improve the environment.

We help people and wildlife adapt to climate change and reduce its impacts, including flooding, drought, sea level rise and coastal erosion.

We improve the quality of our water, land and air by tackling pollution. We work with businesses to help them comply with environmental regulations. A healthy and diverse environment enhances people's lives and contributes to economic growth.

We can't do this alone. We work as part of the Defra group (Department for Environment, Food & Rural Affairs), with the rest of government, local councils, businesses, civil society groups and local communities to create a better place for people and wildlife.

Published by:

Environment Agency
Horizon House, Deanery Road,
Bristol BS1 5AH

www.gov.uk/environment-agency

© Environment Agency 2021

All rights reserved. This document may be reproduced with prior permission of the Environment Agency.

Further copies of this report are available from our publications catalogue: www.gov.uk/government/publications or our National Customer Contact Centre: 03708 506 506

Email: research@environment-agency.gov.uk

Author(s):

Emma Pharaoh, Steve J. Ormerod and Ian P. Vaughan

Keywords:

Macroinvertebrate, river, richness, trends, long-term, water quality, land use

Research contractor:

Water Research Institute and School of Biosciences, Cardiff University, Museum Avenue, Cardiff. CF10 3AX.

Environment Agency's Project Manager:
Mark Diamond and John Murray-Bligh

Research at the Environment Agency

Scientific research and analysis underpins everything the Environment Agency does. It helps us to understand and manage the environment effectively. Our own experts work with leading scientific organisations, universities and other parts of the Defra group to bring the best knowledge to bear on the environmental problems that we face now and in the future. Our scientific work is published as summaries and reports, freely available to all.

This report is the result of research commissioned by the Environment Agency's Chief Scientist's Group.

You can find out more about our current science programmes at <https://www.gov.uk/government/organisations/environment-agency/about/research>

If you have any comments or questions about this report or the Environment Agency's other scientific work, please contact research@environment-agency.gov.uk.

Professor Doug Wilson
Chief Scientist

Contents

Executive summary	6
Introduction	8
Overview	8
Project aim and objectives	9
Methods	10
Macroinvertebrate data	10
Environmental data	11
Trends in macroinvertebrate assemblages	12
Modelling the links between community structure and environmental covariates ..	13
Results	15
River representation through time	15
Temporal macroinvertebrate trends	15
Temporal macroinvertebrate trends across the urban land cover gradient	20
Links between community structure and environmental covariates	20
Discussion	27
Conclusions and recommendations	29
References	31
Acknowledgements	35
List of abbreviations	36
Appendix 1. Calculation of post-stratification weights for smoothed national trends in invertebrate assemblages.	37
Appendix 2. Correspondence analysis loadings of the 78 composite and family groups used in this study.	38
Appendix 3. Smoothed macroinvertebrate time series with and without post- stratification	42

Would you like to find out more about us or your environment?.....	43
incident hotline	43
floodline	43
Environment first.....	43

Executive summary

River macroinvertebrates are monitored routinely as recognised indicators of river quality. Understanding their national trends helps the Environment Agency to evaluate the effects environmental pressures such as climate change, increasing pressure for water resources and impacts of new chemicals whilst also showing where management actions, legislation and restoration brings benefit

This report presents national-scale trends in the river macroinvertebrate communities across England over 29 years (1991–2019). The current debate around the status and trends in England's rivers draws often on local studies or anecdotal evidence, and this report was commissioned to evaluate national-scale patterns using the Environment Agency's monitoring data. Data were collected following standard sampling protocols and subject to a quality assurance process. Sampled in this way, macroinvertebrates are a valuable indicator of the condition of river environments, being sensitive to a range of impacts (e.g. different pollutants), responding rapidly to changes in their environment and capturing the biological effects of short-term water quality issues that monthly water samples might miss.

This analysis updates a previous study, which looked at macroinvertebrate communities from 1991 to 2008 across England and Wales (Vaughan and Ormerod, 2012). It focuses on two characteristics: i) the number of macroinvertebrate families present in a typical sample (the richness), and ii) the types of families present. For the latter, each sample is given a score to describe the community in which higher values indicate the presence of more pollution-sensitive families such as mayflies and stoneflies.

Trends were derived using c. 3,500 locations across England. Since sampling locations were allocated for a range of different purposes, some types of water course (e.g. small headwater streams) are under-represented in the data. By comparing the macroinvertebrate sampling locations to the River Habitat Survey Baseline of England (2007–2008), which selected sites according to a stratified-random sampling strategy to ensure a representative sample of rivers, this bias was quantified and the national trends adjusted to make them more representative of English rivers as a whole. The changes were examined in relation to several environmental factors (e.g. water chemistry) – most notably whether catchments were urban or rural.

There were three main results:

1. The average number of macroinvertebrate families present increased by ~11%. This increase was restricted to the first half of the time series (1991–2005): rapid increase in the early 1990s slowed in the second half of the decade so that there was no overall increase after ~2005.

2. Rivers continued to gain pollution-sensitive macroinvertebrates between 1991 and 2019, despite richness plateauing earlier in the time series. Observed fluctuations around this trend are likely to be linked to variation in climate.
3. Consistent with previous work, gains in richness and sensitive macroinvertebrates were greatest in rivers draining urbanised catchments so that their composition converged on rural rivers. That said, increases in the richness of urban rivers appeared to stall after ~2010 (~ 5 years later than the overall national trend).

The increases in richness and prevalence of pollution-sensitive taxa are consistent with ongoing improvements in urban water quality up to 2019, although this study does not establish a causal relationship. These increases also suggest that improvements in water quality may have continued to offset some climate change impacts. The reason for the stalled increase in richness post 2005 is unclear: this could be a natural mechanism or indicate some form of degradation, such as invasive non-native species or novel pollutants. National scale improvements may also mask local declines in quality, and so further work is needed to disaggregate the national patterns, as well as to get a clearer idea about the factors driving change through time.

Introduction

Overview

River macroinvertebrates are a widely used indicator of ecological quality. Understanding long-term national trends in river macroinvertebrate communities will help the Environment Agency to contextualise more recent changes. Without an objective assessment it is difficult to evaluate the net effects of management actions, including changes to legislation, improved river quality standards and river restoration measures. This is especially important against a backdrop of shifting environmental pressures, such as changes in agricultural pollution, increasing impact of climate change, rising pressures on water resources and emerging contaminants.

Very few analyses of long-term change have been undertaken with the notable exception of a body of work undertaken at Cardiff University (e.g. Durance and Ormerod, 2009; Vaughan and Ormerod, 2012, 2014). This report is an update of these previous analyses looking at national trends in river macroinvertebrate communities based on Environment Agency data. In particular, it extends the work of Vaughan and Ormerod (2012) which looked at two summary measures of macroinvertebrate communities: family-level richness (i.e. the number of families found within a sample) and overall composition. The main findings from previous studies, covering England and Wales from 1991 to 2008 (or 2011) (Vaughan and Ormerod, 2012, 2014; Vaughan and Gotelli, 2019) were:

- Macroinvertebrate communities changed substantially, with increases in both richness and the prevalence of taxa sensitive to organic pollution. This is interpreted as biological recovery following improvements in water quality, including declines in polluting industries, and improved wastewater treatment and regulation.
- Recovery was detected across most of England and Wales, with the largest improvements in urbanised catchments. At the same time there were hints of declines in certain areas (e.g. upland catchments).
- Analysis at both family- and community-levels implicated water quality (especially biochemical oxygen demand; BOD) and to a lesser extent nutrient and climatic variables in these changes.
- The observed changes were the opposite of those expected with rising temperatures, suggesting that the improvements in water quality offset climate warming (a mean increase in water temperatures of $\sim 0.6^{\circ}\text{C}$ over 1991–2011) (Vaughan and Gotelli, 2019).
- Macroinvertebrate communities respond rapidly to environmental change, such that the observed biological changes related to contemporary changes in the environment (cf. delayed responses to past improvements).

Since 2008/2011, interest in invertebrate biodiversity trends in the UK and internationally has increased greatly. A range of studies has reported invertebrate declines, including for aquatic taxa (e.g. Sánchez-Bayo and Wyckhuys, 2019; Baranov *et al.*, 2020; Stepanian *et al.*, 2020), whilst others have found evidence of increases in the abundance or prevalence of aquatic invertebrates since 1990 (e.g. Outhwaite *et al.*, 2020; van Klink *et al.*, 2020). In the UK context, the analysis of Outhwaite *et al.* (2020), based on data from taxon-specific recording schemes (e.g. the Riverfly and Aquatic Hemiptera recording schemes), revealed trends in prevalence of several macroinvertebrate taxa that corresponded to Vaughan and Ormerod's (2012) analysis of Environment Agency data for the 1990s and early 2000s. Outhwaite *et al.*'s (2020) analysis extended to 2015 and suggested little change during the early 2010s. In the context of the invertebrate decline debate and concerns around both existing and future threats to freshwaters (Dudgeon *et al.*, 2006; Reid *et al.*, 2019), an updated analysis of Environment Agency data should allow a more comprehensive assessment of macroinvertebrate trends across English rivers and bring the assessment up to date.

Project aim and objectives

This report adds another decade of data to Vaughan and Ormerod's 2012 analysis, covering 29 years in total (1991–2019 inclusive), to see how English rivers have fared over this period. The sampling intensity (number of locations sampled per year) declined nationally over this period, but with developments in methodology compared to Vaughan and Ormerod (2012) it was possible to draw data from a larger array of locations, allowing national trends to be assessed. The main aims were to:

1. Provide a best estimate for how macroinvertebrate richness and composition have changed 1991–2019, and identify notable periods of change.
2. Carry out an analysis of the roles of water quality, climate, land cover and physical habitat in explaining macroinvertebrate community structure.

An assessment of the representativeness of the sampling locations – how well they reflect the diversity of rivers across England, is also undertaken.

Methods

Macroinvertebrate data

Environment Agency macroinvertebrate data covering 1991–2019 inclusive were collated for English rivers. Quality-assurance schemes ran throughout this period (J. Murray-Bligh, personal communication). The number of locations sampled per year declined nationally over this period and there were changes in the sampling locations, which could introduce bias to subsequent analyses and limit the ability to detect change. For this reason, the timeframe of the study was split into three near-equal windows (1991–2000, 2001–2010 and 2011–2019), and only sites sampled at least once in each of the three windows were retained: this minimised site turnover, making it easier to detect changes through time.

The initial data set was filtered so that all remaining samples were: i) collected in spring (March-May inclusive), using a standardised protocol of 3-minute kick sampling; ii) sorted and identified in the laboratory (as opposed to bank-side sorting); and iii) taken from sites which had been sampled in at least four years in total, were not described as being downstream from an effluent outfall, and which contained site specific data on altitude, distance from source and channel slope. Catchments for individual sites were delineated from a 50-m DTM (OS Terrain 50, Ordnance Survey) and a 1:50,000 river network (Moore *et al.*, 1994) using ArcHydro tools (ESRI ArcGIS 10.7.1); where this was not possible (e.g. where the drainage network had been modified and included loops) sites were excluded from the analysis. This left a total of 3506 sites, with a mean of 12 years sampled per site (Figure 1a).

Data were reduced to 78 composite and family groups, using taxonomy outlined by Vaughan and Ormerod (2014) (Appendix 2). This ensured that samples could be compared across the time series, during which there were changes in invertebrate taxonomy and the resolution to which samples were identified. Taxon abundances were recorded as either individual counts or \log_{10} abundance classes (e.g. 1–9 and 10–99 individuals). To harmonise this, all data were converted to \log_{10} abundance classes. Where abundance data had been collected to a lower taxonomic rank than family, abundances were pooled to estimate family-level abundance. For analysis, abundance was converted to the midpoint for each abundance class (e.g. the abundance for taxa in the 1–9 class was recorded as 5). Secondly, the data were converted to presence-absence format, simply indicating whether each taxon was present or absent in a sample, and most analyses performed using both abundance and presence-absence data. This allows a conservative approach to the analysis, reflecting the relatively imprecise abundance estimates, and aiding comparisons with earlier studies that only used presence-absence data (Vaughan and Ormerod, 2012, 2014), whilst making use of the additional information contained in abundance estimates.

Presence-absence and abundance data were ordinated using correspondence analysis (CA) from the vegan package (Oksanen *et al.*, 2019) in R v4.0.1 (R Core Team, 2020) to extract the main axis of variation in macroinvertebrate communities (CA1 score). The resulting scores provide a straightforward way to quantify the main changes in macroinvertebrate composition within and among sampling sites, and is consistent with previous analysis of Environment Agency data by Vaughan and Ormerod (2012). CA1 scores alongside taxonomic richness were used as the main response variables throughout the analyses.

Environmental data

Within each of the 3506 river catchments, the percentages of urban or sub-urban landcover, improved grassland or arable agriculture were calculated from the 1 km resolution UK Land Cover Map 2000 (Fuller *et al.*, 2002). The proportion of the catchment underlain by calcareous geology was calculated from 1:625,000 scale maps (British Geological Survey/ Natural Environment Research Council). Mean annual precipitation in each catchment was calculated from 1961–1990 averages mapped at 5 km resolution (UK Meteorological Office, Exeter, UK; Perry and Hollis, 2005), allowing specific stream power at bankfull discharge ($W\ m^{-2}$ – which represents the ability of the river to erode and transport sediment) and the base flow index (BFI – the contribution of ground water to the river's flow) to be estimated following Vaughan *et al.* (2013). Stream power has been shown to be a good predictor of the physical habitat within the channel, including the predominant substratum (Vaughan *et al.*, 2013).

Water quality data to accompany the macroinvertebrate data, covering 1990–2019, were sourced from routine water quality monitoring data from the Environment Agency. Annual medians for a range of chemical determinands were calculated for the 12 months prior to the spring macroinvertebrate sampling period. Medians were calculated where at least nine out of the 12 calendar months were sampled and pH, temperature, BOD, orthophosphate and either total organic nitrogen (TON) or nitrate were available. Where sampling occurred more frequently than once a month, a maximum of two randomly selected samples per calendar month were retained to minimise bias towards seasons more frequently sampled. Where $\geq 50\%$ of values were below detection limits, regression-on-order-statistics was used to estimate medians using the NADA package in R (Lee and Helsel, 2005). Missing nitrate values were interpolated from TON by fitting a linear regression to predict nitrate from TON where both were recorded, and then generating predictions where only TON was recorded (Vaughan and Ormerod, 2012).

Discharge data for 1990–2019 were accessed from the UK National River Flow Archive (<https://nrfa.ceh.ac.uk/>). The median daily discharge for the 12 months prior to the spring sampling period was calculated. Discharge was divided by gauging

station catchment area to provide a measure of discharge that was independent of catchment area (units = $\text{m}^3 \text{s}^{-1} \text{km}^{-2}$).

Trends in macroinvertebrate assemblages

Temporal trends in macroinvertebrate assemblages were estimated for 1991–2019 using generalised additive models (GAMs) fitted using the `gam` function in R's `mgcv` package (Wood, 2011). The approach followed that of Vaughan and Ormerod (2012, 2014), combining Fewster *et al.*'s (2000) approach to estimating smoothed trends, with site weights estimated via a post-stratification process to minimise bias in national trends (Appendix 1). Post-stratification used the 2007–2008 River Habitat Survey (RHS) Baseline as the most representative sample available of rivers across England because RHS locations were allocated by stratified-random sampling, with each 10km National Grid square being a stratum within which river reaches were randomly sampled (Seager *et al.*, 2012). Adjusting for the length of rivers depicted on the 1:50,000 versus 1:250,000 scale river network within each 10km square (see Seager *et al.*, 2012 for details), each macroinvertebrate site was weighted to produce national trends that should simulate macroinvertebrate sampling across the RHS Baseline. For example, macroinvertebrate sites in relatively infrequently sampled headwater streams were more heavily weighted when calculating the national trend.

CA1 scores and richness from the 3506 sites were modelled as a function of year, using cubic regression splines. The models also included 12 site- and catchment-level environmental covariates that can account for variation in macroinvertebrate assemblages between river reaches (Fewster *et al.*, 2000; Vaughan and Ormerod, 2012). Covariates included in the models were: catchment area, mean annual rainfall, proportion of catchment underlain by calcareous bedrock geology, proportion of land cover classed as urban, arable or improved grassland, specific stream power, predicted base flow index (BFI) and site easting and northing, altitude and channel slope. The year term was modelled with nine degrees of freedom, as this is a good compromise between identifying long-term trends and shorter-term (multiple year) changes (Fewster *et al.*, 2000). Splines were used for each of the environmental covariates, ranging from 5 to 9 degrees of freedom. Bootstrapping was used to produce nonparametric 95% confidence limits of the trends, based on 399 bootstraps (including bootstrapping of post-stratification weights; Fewster *et al.*, 2000; Vaughan & Ormerod, 2012). Significant positive and negative 'change points' (inflections) in the trend were identified using bootstraps of the trends, following Fewster *et al.* (2000). Annual point estimates of richness and CA1 (i.e. unsmoothed values) throughout the time series were also calculated, by using the same covariates as the smoothed models, but modelling year as a factor instead of a spline.

Previous analysis of macroinvertebrate trends revealed larger changes in urbanised catchments (Vaughan & Ormerod, 2012). This analysis was updated using GAMs to model CA1 scores and richness against a full tensor product smooth of the percentage of catchment urbanisation and year, using cubic regression splines. The degrees of freedom (smoothness) of the full tensor product smooth was fixed at nine (Fewster *et al.*, 2000). The models included the same catchment level environmental covariates used in the temporal trend GAMs, except for catchment urban land cover which was included in the tensor product smooth with year.

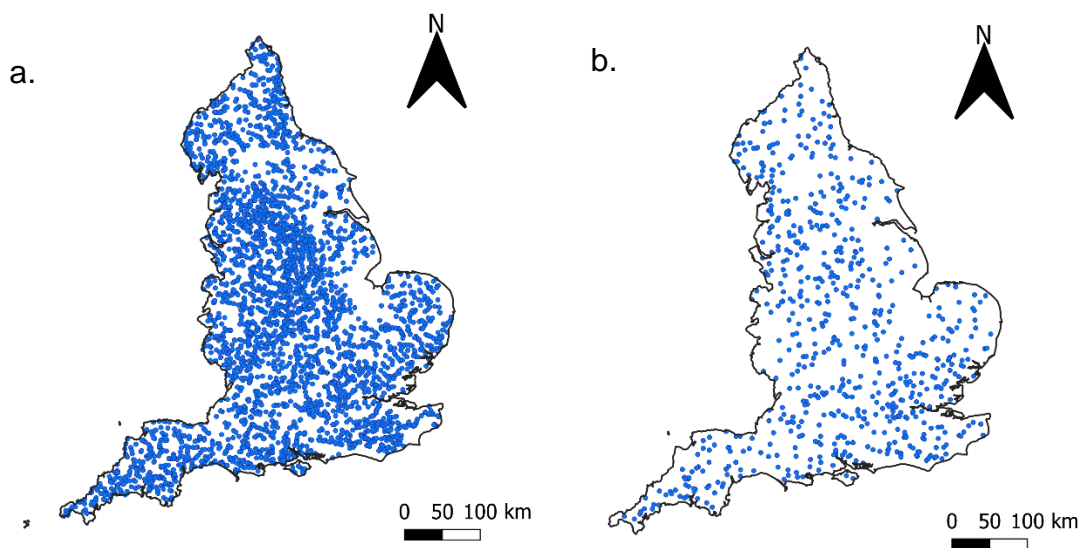


Figure 1: Locations of macroinvertebrate sites included in calculation of (a) temporal assemblage trends ($n = 3506$ sites) and (b) investigation of the extent to which macroinvertebrate community structure could be accounted for by water quality and temperature, discharge, stream power and land use ($n = 645$). Map contains OS data © Crown copyright and database right (2021).

Modelling the links between community structure and environmental covariates

Generalised additive mixed-effects models (GAMMs) were used to investigate the extent to which spatio-temporal variation in the macroinvertebrate communities could be accounted for by water quality and temperature, discharge, stream power and

urban land use. In a first step, a subset of macroinvertebrate locations was identified where water chemistry and discharge data were available in close proximity to the biological sampling point. Using maximum distances of 1km to the nearest water chemistry location and 5km to a gauging station (Vaughan and Ormerod, 2012), 645 macroinvertebrate locations were retained (Figure 1b). Data were filtered at each site so that only years with macroinvertebrate, chemistry and discharge data were used.

Three GAMMs – for richness and the CA1 scores for presence-absence and abundance data – were fitted using the mgcv package in R (Wood, 2017). All covariates were fitted using regression splines, with the degree of smoothing chosen using the default generalised cross-validation process (Wood 2017): this allowed nonlinear relationships to be modelled where appropriate. The following variables were included in the models: i) site-level covariates: proportion of catchment with urban land cover and specific stream power; ii) site location: OS National Grid eastings and northings, modelled as a two-dimensional tensor product smooth; and iii) time-varying covariates: median BOD, pH, nitrate, phosphate, water temperature and discharge. Sampling site was included as a random term in the model to account for repeat sampling through time, and a first-order autoregressive function used to model residual temporal autocorrelation (Pinheiro and Bates, 2000).

Results

River representation through time

Within the entire pool of Environment Agency macroinvertebrate monitoring sites, locations selected for this study were among those that had the greatest sampling effort in the years 1991–2019, with little evidence of a major drop in sampling intensity (number of samples per year) through time across the 3506 locations (Figure 2f). The number of locations sampled per year was variable: notable features included years of extensive coverage (1995, 2000 and 2013) and low coverage in 2001, which was impacted by foot-and-mouth disease restrictions. Nevertheless, there was some evidence of changes in sampling emphasis, with a slight broadening of coverage across altitudes, catchment area and distances from river source in recent years (flattening of the distributions in Figure 2a-c). Sites with more extensively urbanised catchments were sampled less frequently later in the time series (Figure 2e).

Although a similar set of macroinvertebrate locations was sampled in each year, the overall distribution of the 3506 biological sites was quite different from the second RHS Baseline of England (Figure 2). Smaller catchments with steeper channels in particular were underrepresented in the macroinvertebrate data, consistent with limited sampling of headwater streams. There was a more similar spread of altitude and catchment urbanisation in the two data sets, although macroinvertebrate sampling locations tended to be slightly lower altitude and more extensively urbanised on average (Figure 2).

Temporal macroinvertebrate trends

The CA1 scores (presence-absence and abundance based) captured similar overall patterns in the macroinvertebrate community. Negative values of both scores represented communities with greater proportions of taxa tolerant of poorer water quality and siltier substrata, such as oligochaetes and molluscs (Appendix 2). Communities with positive CA1 scores were characterised by taxa typical of faster flows, higher water quality and better-oxygenated conditions, such as families within the Ephemeroptera (mayflies), Plecoptera (stoneflies) and Trichoptera (caddisflies).

Smoothed national time series for CA1 scores for macroinvertebrate abundance (eigenvalue = 0.50) and presence-absence (eigenvalue = 0.25) data showed trends towards higher scores, with total increases of 0.3 units and 0.2 units respectively (Figure 3; for reference, equivalent plots are shown for simple smooths of the data,

without post-stratification in Appendix 3, to demonstrate the effect that the post-stratification procedure had on the outputs). This indicates trends towards a greater frequency or abundance of taxa connected to more turbulent rivers of higher water quality and well oxygenated conditions. Both time series also showed fluctuations around the overall trend: these formed three relatively clear peaks for abundance data and two clear peaks, followed by a less distinct third, for presence-absence data. The timings of these fluctuations were different across the two-time series, with CA1 abundance scores lagging behind those from presence-absence data and possibly having a slightly longer wavelength. CA1 abundance scores had one major inflection in the trend in 1996, where rate of decline increased significantly, whereas significant turning points – representing both upturns and downturns of the trend line – were much more frequent for the presence-absence data (Figure 3).

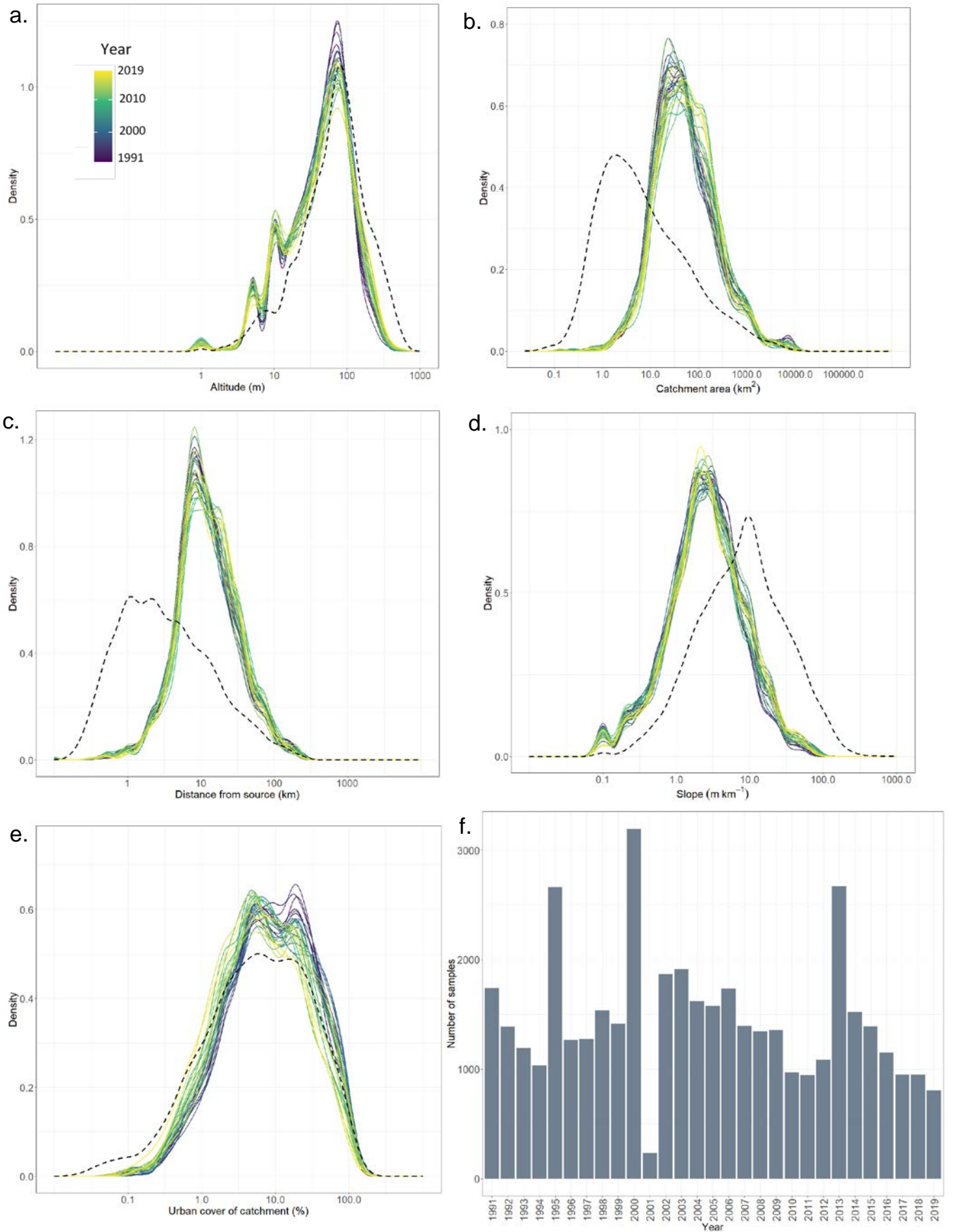


Figure 2: The distribution of macroinvertebrate sites sampled in each year of the study (solid lines, with greens and yellows representing the most recent years) and English sites sampled during the second RHS Baseline (dashed line), compared to (a) altitude, (b) catchment area, (c) distance downstream from the source, (d) channel slope and (e) percentage of catchment attributed to urban land cover. Panel (f) indicates the number of macroinvertebrate samples per year.

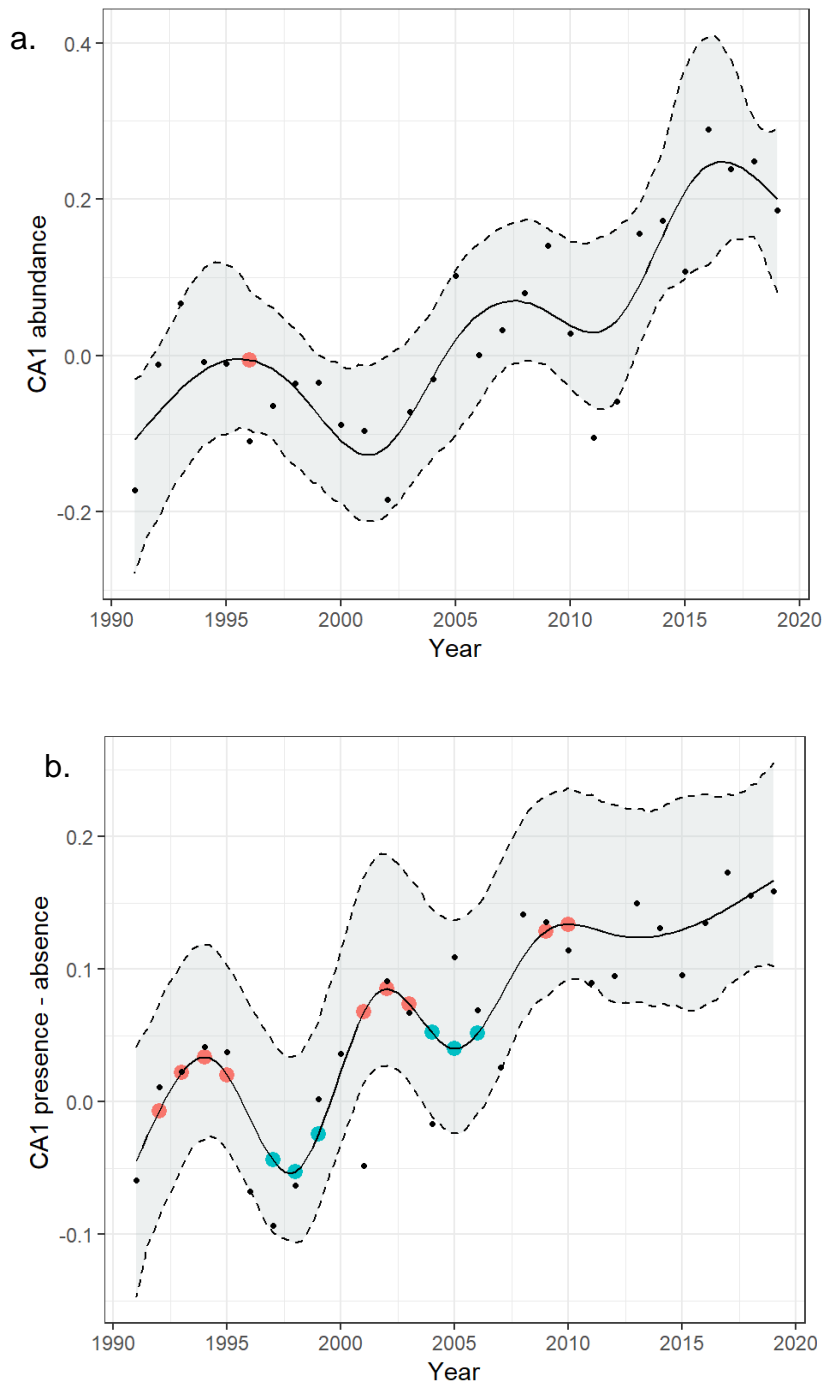


Figure 3: Smoothed macroinvertebrate trends for England in the years 1991–2019, represented by CA1 scores from (a) abundance and (b) presence-absence data. Solid lines represent the post-stratified estimates, with dashed lines indicating bootstrapped 95% confidence limits. Dots along the solid line represent statistically significant inflections in the gradient of the curve: red indicating reduced rate of increase or greater rate of decline, and blue indicating greater rate of increase or reduced rate of decrease. Black points denote annual (unsmoothed) point estimates.

Richness increased by around 11% in the years 1991–2019, with a gain of just over 1.5 taxa (Figure 4). The increase was concentrated in the first part of the time series (1991–2005), with a significant reduction in the rate of increase in the mid-1990s, and reaching its highest point in 2005. From the late 1990s there was little net change in richness, with just a series of fluctuations around the trend (Figure 4).

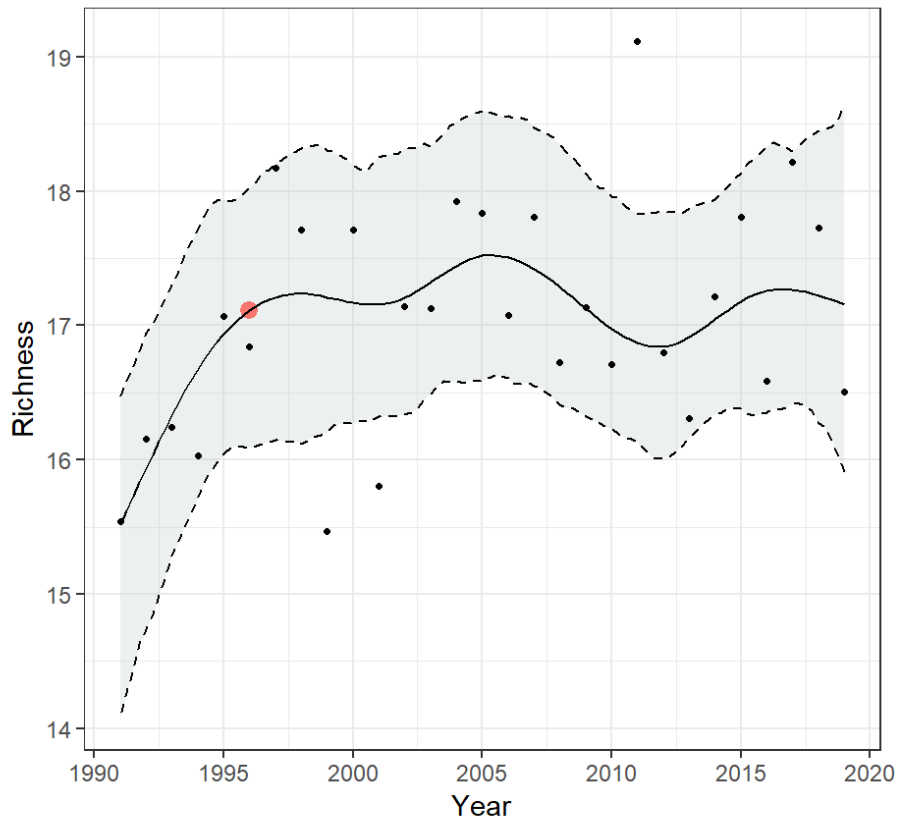


Figure 4: The trend of macroinvertebrate richness in England in the years 1991–2019. The solid line represents the post-stratified estimate, with dashed lines indicating bootstrapped 95% confidence limits. The red dot represents the timing of a statistically significant reduction in the rate of increase. Black points denote annual (unsmoothed) point estimates.

Temporal macroinvertebrate trends across the urban land cover gradient

Temporal changes in macroinvertebrate communities varied across the catchment urbanisation gradient. More heavily urbanised catchments had lower CA1 scores and richness at the start of the time series, compared to more rural catchments. Richness increased to the greatest degree in more urbanised catchments, narrowing the gap to rural rivers (Figure 5a). Increases were most rapid early in the time period (~1991–2005) and levelled off over time: this happened earlier in rural catchments and was later with increased catchment urbanisation – biological recovery in the most urbanised catchments appeared to plateau after ~2010 (Figure 5a).

Increases in CA1 abundance score were most marked in the more heavily urbanised catchments, with the greatest increases in the second half of the time series (Figure 5b). The increase in CA1 scores from presence-absence data was of a similar magnitude across the land cover gradient (Figure 5c). Fluctuations around the trend were clearest in the most rural catchments, similar to the patterns in Figure 3.

Links between community structure and environmental covariates

The environmental covariates explained 20–65% of the variation in macroinvertebrate communities (Tables 1 and 2). Presence-absence based CA1, solely reflecting the composition of the community, was the most predictable (R^2 -adjusted = 0.65), with abundance-based CA1 (R^2 -adjusted = 0.20) and richness (0.33) being less so.

CA1 scores for abundance and presence-absence data showed highly significant relationships with urban land cover, discharge, pH and BOD ($p \leq 0.001$; Table 1), whilst CA1 presence-absence also showed strong evidence for relationships with phosphate and specific stream power. CA1 scores declined with increasing urbanisation, BOD and phosphate concentrations, and increased with discharge (both abundance and presence-absence; Figure 6). There was evidence for a nonlinear relationship between CA1 presence-absence scores and phosphate, with a consistent decrease above 0.1 mg l^{-1} , whereas no evidence for nonlinearity was found with CA1 abundance. Both CA1 scores decreased with higher catchment urbanisation, most rapidly before 25% urban land cover – above which the relationship became less pronounced. They also showed a nonlinear relationship with BOD, with both CA1 scores declining up to around 6 mg l^{-1} , with little further change above that. Trends in CA1 score with pH contrasted, with CA1 abundance scores increasing with pH, while CA1 presence-absence scores reflected the reverse

trend, which was less pronounced. Both CA1 scores increased with discharge, up to around $0.01 \text{ m}^3 \text{ s}^{-1} \text{ km}^{-2}$ (abundance data) and $0.04 \text{ m}^3 \text{ s}^{-1} \text{ km}^{-2}$ (presence-absence). Presence-absence CA1 scores increased monotonically with specific stream power, whereas there was some evidence of abundance CA1 scores being highest at intermediate values (c. $30\text{--}40 \text{ W m}^{-2}$).

Relationships between presence-absence CA1 scores and phosphate, temperature, and specific stream power and CA1 abundance score and temperature were also identified, but at a lower level of statistical significance ($p \leq 0.01$ and 0.05 ; Table 1).

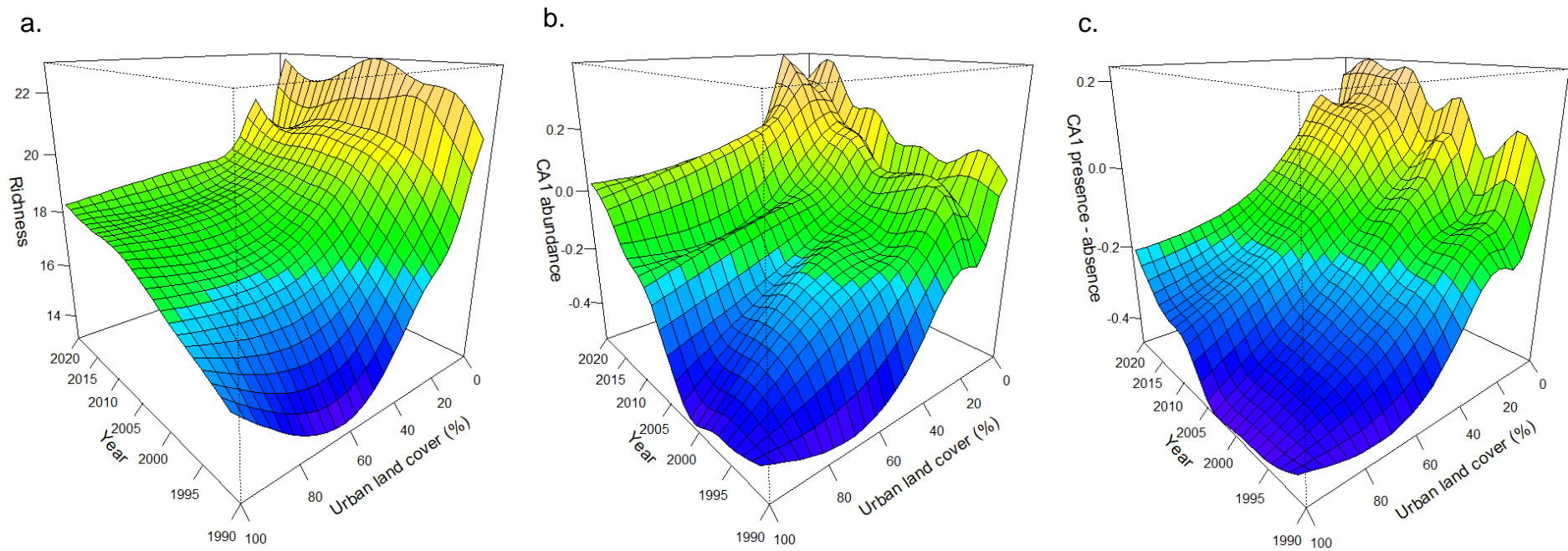


Figure 5: Smoothed temporal trends in invertebrate (a) richness, (b) CA1 score from abundance data and (c) CA1 score from presence-absence data across a gradient of catchment urbanisation. Smooths were generated from generalised additive models, fixing all other covariates in the models at their median values.

Table 1: Summaries of GAMMs of CA1 scores of abundance and presence-absence data with smoothed terms for water quality, catchment, and environment.

Abundance				Presence-absence				
	Smoothed terms			Smoothed terms				
Variables	edf	<i>F</i>	<i>p</i>	edf	<i>F</i>	<i>p</i>		
Urbanisation	1.000	61.748	<0.001	1.000	160.750	<0.001		
pH	1.000	12.973	<0.001	1.000	15.028	<0.001		
BOD	3.585	25.045	<0.001	4.739	11.683	<0.001		
Nitrate	1.000	1.294	0.255	1.000	1.562	0.212		
Phosphate	1.000	6.682	0.010	4.357	13.352	<0.001		
Water temperature	1.000	4.547	0.033	1.000	4.432	0.035		
Discharge	4.087	4.805	<0.001	5.549	26.038	<0.001		
Specific stream power	3.590	3.749	0.012	2.799	84.053	<0.001		
	Parametric coefficients				Parametric coefficients			
	Estimate	Std. Error	<i>t</i>	<i>p</i>	Estimate	Std. Error	<i>t</i>	<i>p</i>
Intercept	0.088	0.081	1.089	0.276	0.244	0.055	4.432	<0.001
Easting, Northing	-4.47e-7	-1.798e-7	-2.487	0.013	-8.64e-7	-1.23e-7	-7.024	<0.001
	$R^2 = 0.204$				$R^2 = 0.648$			

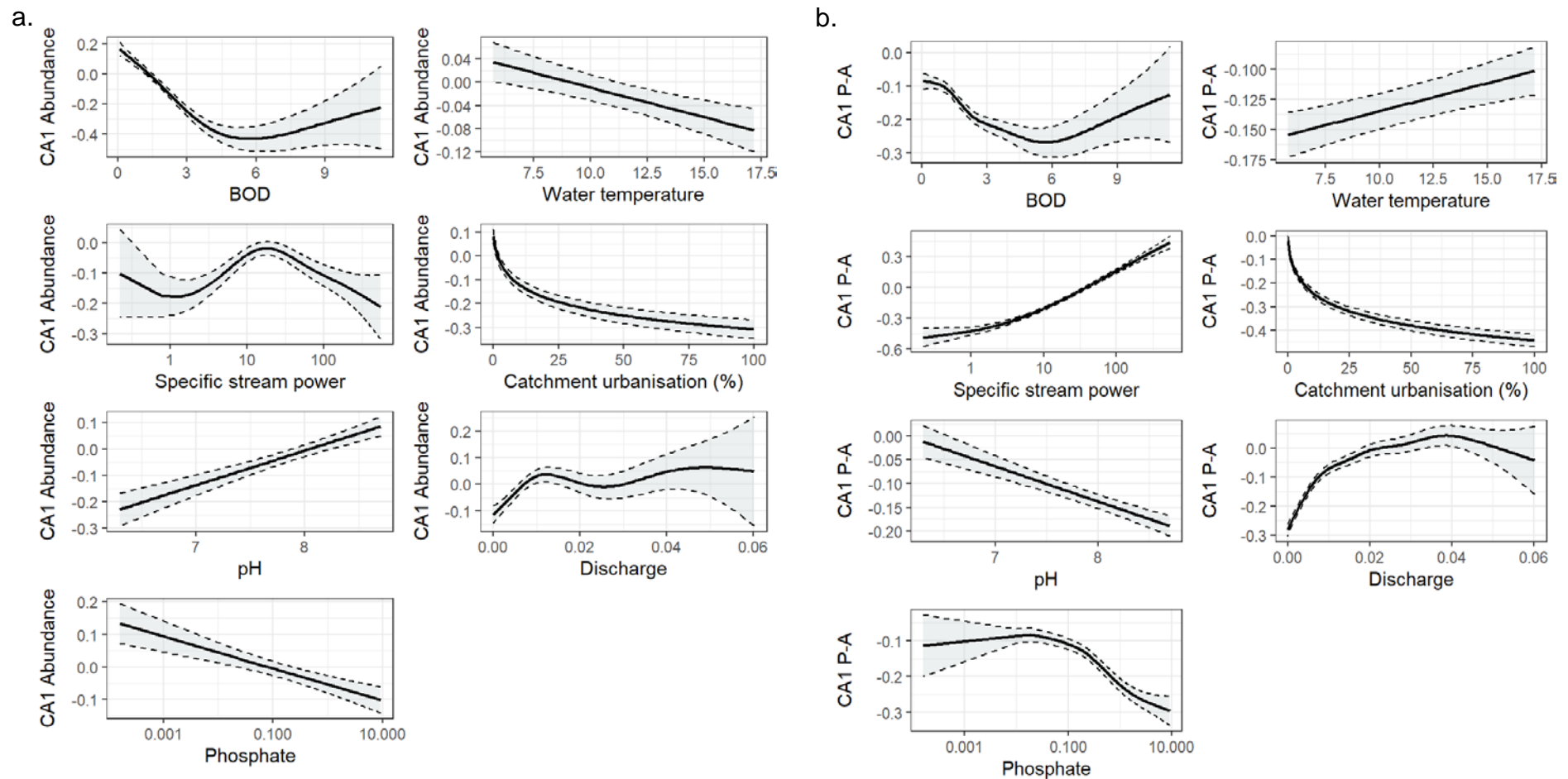


Figure 6: Smoothed predicted responses of macroinvertebrate CA1 scores from generalised additive mixed models (GAMM) for (a) abundance and (b) presence-absence data for each significant model term (BOD, water temperature, specific stream power, catchment urbanisation, pH, discharge, and phosphate). Dotted lines denote \pm standard error. All variables in the model apart from the one depicted in each panel were held at their median values for plotting.

Macroinvertebrate richness was significantly related to catchment urbanisation, BOD, and specific stream power ($p \leq 0.001$; Table 2). Richness increased with catchment urbanisation until around 1% of the catchment was attributed to urban land cover, after which it declined (Figure 7). Macroinvertebrate richness and BOD were largely negatively associated, with richness reaching its lowest point at around 7 mg l⁻¹. Similar to the CA1 abundance model, the trend between specific stream power and richness showed an inverted U-shape where lower macroinvertebrate richness was associated with the extremes of stream power (Figure 7).

Table 2: Summaries of GAMMs of macroinvertebrate richness with smoothed water quality, catchment, and environment terms.

Smoothed terms				
Variables	edf	<i>F</i>	<i>p</i>	
Urbanisation	4.725	56.853	<0.001	
pH	1.000	9.743	0.002	
BOD	3.535	19.928	<0.001	
Nitrate	1.000	5.537	0.019	
Phosphate	1.000	0.242	0.623	
Water temperature	1.872	1.393	0.212	
Discharge	1.000	2.157	0.142	
Specific stream power	3.062	5.970	<0.001	
Parametric coefficients				
	Estimate	Std. Error	<i>t</i>	<i>p</i>
Intercept	24.363	1.004	24.281	<0.001
Easting, Northing	-8.60e-6	-2.24e-6	-3.846	<0.001
$R^2 = 0.332$				

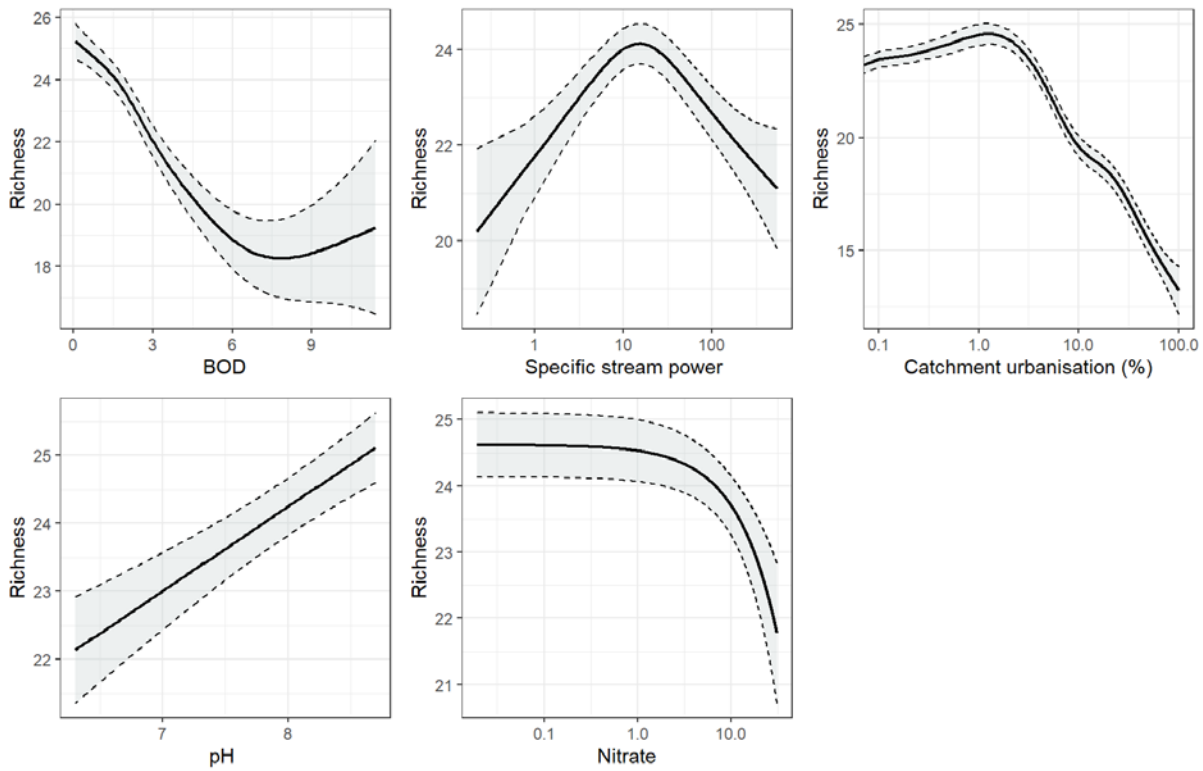


Figure 7: Smoothed predicted responses of macroinvertebrate richness from generalised additive mixed models (GAMM) for each significant model term (BOD, specific stream power, catchment urbanisation, pH, and nitrate). Dotted lines denote standard errors. All variables apart from the one depicted in each panel were held at their median values for plotting.

Discussion

England's rivers have been monitored extensively over recent decades, enabling detailed study of trends in macroinvertebrate communities, environmental conditions and their interactions (e.g. Burt *et al.*, 2008; Dunbar *et al.*, 2010; Kernan *et al.*, 2010; Vaughan and Ormerod, 2012, 2014; Jourdan *et al.*, 2018; Vaughan and Gotelli, 2019; Outhwaite *et al.*, 2020). The data provide a valuable resource for studying the effects of multiple stressors and biological recovery across large spatial and temporal scales, in turn informing evidence-based decision making and ecosystem management (Lindenmayer *et al.*, 2012; Sutherland *et al.*, 2004; Worrall *et al.*, 2009). Despite sampling intensity declining across the Environment Agency sampling network, careful site selection enabled temporal trends to be studied from 1991 to 2019 in the current study. Further declines in sampling intensity have the potential to reduce the scope for future analysis of this type, where improved methodology cannot offset reduced data availability.

The current analysis confirmed that there have been large changes in river macroinvertebrate communities in England over the last three decades, with an increase in richness of around 11% and a shift towards families characteristic of higher quality, less polluted conditions. This extends the analysis conducted on earlier years (1991–2008) of English and Welsh macroinvertebrate data by Vaughan and Ormerod (2012), and the results are broadly consistent. The increase in richness in the current study was less pronounced than in the 2012 analysis, where richness increased by 20%. However, the results were the same in terms of the most rapid increase in the early- to mid-1990s, slowing significantly in the late-1990s, and with little change thereafter: the current analysis found no evidence for an overall increase in richness over the additional 2009–2019 period. The magnitude of the increases in richness cannot be compared directly between the two studies primarily due to the different geographical scope (i.e. Wales is not included within the current analysis), but also due to some methodological refinements. Key amongst the latter was that post-stratification in the current study was based on the second RHS Baseline of England (cf. the first RHS Baseline for the 2012 study), providing better coverage of headwater streams (Seager *et al.*, 2012) – this should make the current estimated trends more representative of English rivers. Furthermore, many more macroinvertebrate sampling locations were included (3506 versus 1083), primarily because more recent research has shown that specific stream power, estimated for virtually any location using GIS, is an effective proxy for aspects of RHS data (Vaughan *et al.*, 2013) – these physical habitat data were required in the post-stratification process.

The changes in richness observed in this study are consistent with an independent analysis of freshwater data (1970–2015), based on records from national recording schemes and societies in the UK (Outhwaite *et al.*, 2020). Outhwaite *et al.* (2020) found a large decrease from the baseline in 1970, before a rapid increase in the prevalence of many freshwater macroinvertebrates in the early 1990's, which appeared to stall around 2005. Whilst the environmental correlates of the biological changes were not investigated, Outhwaite *et al.* (2020) suggested that the introduction of the European Urban Wastewater Treatment Directive in 1991 and regulatory changes in the water industry around this time (Saal and Parker, 2000) could have driven the changes.

The increasing CA1 scores observed through time indicate an ongoing shift towards taxa considered to be more sensitive to both poor water quality and climate warming (Durance and Ormerod, 2007). In contrast to richness, the increases continued across the complete time period (1991–2019): this implies that there was increasing turnover in the second half, with sensitive taxa replacing tolerant ones – hence no net change in richness – rather than sensitive taxa augmenting existing tolerant ones. The importance of turnover is consistent with the analysis of family-level trends for England and Wales 1991–2011 (Vaughan and Ormerod 2014). Fluctuations in the CA1 score around the overall increasing trend may be a consequence of climatic variability, such as those captured by the North Atlantic Oscillation (Bradley and Ormerod 2001; Vaughan and Ormerod 2014). However, fluctuations in CA1 scores based on abundance data appeared to lag behind those based on presence-absence data, and had a longer wavelength: this observation requires further investigation.

As expected, urban areas had lower CA1 scores and richness than more rural catchments throughout the time series. Urban rivers are often subjected to greater pressures than their rural counterparts, with the species present reflective of this (Brown *et al.*, 2005; Chadwick *et al.*, 2006; Roy *et al.*, 2003). The largest increases in richness and abundance-based CA1 scores were observed in heavily urbanised catchments, consistent with recovery from point-source pollution (Vaughan and Ormerod, 2012). Based on richness and correspondence analysis scores, urban rivers narrowed the gap with rural rivers across the study period (1991–2019), but were still characterised by lower richness and a higher proportion of pollution-tolerant taxa in 2019. Several notable features warrant further investigation, including: i) the apparent stalling of the increase in richness in urban rivers after ~2010, albeit with an indication that improvements may have resumed in the final few years (~2015–2019); ii) the dynamics of macroinvertebrate community change, with increases in richness tending to precede an overall shift towards cleaner-water taxa (see Figure 5); and iii) the larger fluctuations around the trend observed for CA1 score and richness within the least urbanised catchments, perhaps suggesting that these waterbodies may be more responsive to climatic variation.

A substantial proportion of the variation in macroinvertebrate communities could be explained by the combination of water quality, climatic variables, land cover, physical habitat and geographic location ($R^2 = 0.20\text{--}0.65$). In particular, urban land cover and BOD were clearly linked with reductions in richness and CA1, consistent with previous work (Vaughan and Ormerod, 2012). In the case of BOD, this suggests that organic pollution and subsequent de-oxygenation could be playing a key role in driving macroinvertebrate trends at the low- to mid-range concentrations in this study (Hynes, 1974). Phosphate was the only modelled nutrient measure to explain any variation in CA1 scores, with higher concentrations correlated with lower scores, whilst richness was weakly correlated with nitrate concentration. The linkages between nutrient concentrations and macroinvertebrates may be complex, as responses are largely indirect (Everall *et al.*, 2019; Nessel *et al.*, 2021; Wang *et al.*, 2007), so further work would be needed to explore this more fully.

Conclusions and recommendations

Changes in macroinvertebrate communities across English rivers over the years 1991–2019 are consistent with improving water quality, especially in urbanised catchments. Although this study does not establish a causal relationship, macroinvertebrate richness increased by around 11%, with changes concentrated in 1991–2005, and rivers gained pollution-sensitive taxa across the time series. The facts that improvements were greatest in more urbanised catchments suggests that these are closing the gap on rural rivers. The analysis brings the national trends of Vaughan and Ormerod (2012) up to date, reaffirming the links between macroinvertebrate communities and key environmental variables, whilst also affording new insights, such as the apparent increase in the importance of taxon turnover through time.

Recommendations for future work:

1. Separating spatial and temporal variation in macroinvertebrate communities, and their links to environmental variables. Obtaining a clearer understanding of the causal factors driving the observed changes through time, and distinguishing them from changes across the country, is both a research priority and a major challenge especially with the changing suite of stressors to which river systems are exposed. The modelling presented here highlighted strong relationships between the environment and community composition (presence-absence CA1 score), but further work is needed to establish how much of the change through time (cf. space) can be explained by changing water quality, temperature and discharge. Previous analysis indicated that the majority of variation in macroinvertebrate communities ($\geq 70\%$) was observed among sampling locations despite the large changes through time (Vaughan and Ormerod 2012).
2. Investigating the differences between presence-absence and abundance CA1 results. Incorporating simple abundance information resulted in different short- to medium-term temporal patterns being detected around the overall increase in CA1 through time, and variation in the presence-absence CA1 score was much more predictable ($R^2 = 0.65$ v. 0.20 for abundance CA1). This may be an artefact, with the crude abundance data used here resulting in a lower signal-to-noise ratio. Alternatively, it may be that even simple, categorical abundance responds more rapidly or in different ways to environmental variation, making it a more sensitive indicator of environmental change and/or suggesting that greater diagnostic power may be obtained by using the abundance and presence-absence scores in tandem.
3. A power analysis to simulate the effects of declining sample sizes and changes to the types of rivers sampled upon the ability to detect biodiversity changes at national and regional scales. Whilst it was possible to overcome reductions in sampling frequency through time in this study, it is important to know the point where meaningful analysis would no longer be possible to ensure that this threshold is not crossed.
4. Disaggregate national trends, as national scale improvements could mask local or regional declines in quality.

There is a more medium-term need to understand other factors that influence macroinvertebrates such as invasive non-native species, combined sewer outflows,

emerging pollutants and more subtle changes in land use (cf. broad land cover changes based on the UK-CEH Landcover Map).

References

- BARANOV, V., JOURDAN, J., PILOTTO, F., WAGNER, R. AND HAASE, P., 2020. Complex and nonlinear climate-driven changes in freshwater insect communities over 42 years. *Conservation Biology*, 34 (5), 1241–1251.
- BROWN, L. R., GRAY, R. H., HUGHES, R. M. AND MEADOR, M. R., 2005. Introduction to Effects of Urbanization on Stream Ecosystems. American Fisheries Society Symposium, 47, 1–8.
- BURT, T. P., HOWDEN, N. J. K., WORRALL, F., AND WHELAN, M. J., 2008. Importance of long-term monitoring for detecting environmental change: Lessons from a lowland river in south east England. *Biogeosciences*, 5 (6), 1529–1535.
- CHADWICK, M. A., DOBBERFUHL, D. R., BENKE, A. C., HURYN, A. D., SUBERKROPP, K. AND THIELE, J. E., 2006. Urbanization affects stream ecosystem function by altering hydrology, chemistry, and biotic richness. *Ecological Applications*, 16(5), 1796–1807.
- DUDGEON, D., ARTHINGTON, A. H., GESSNER, M. O., KAWABATA, Z. I., KNOWLER, D. J., LÉVÊQUE, C., NAIMAN, R. J., PRIEUR-RICHARD, A. H., SOTO, D., STIASSNY, M. L. J., AND SULLIVAN, C. A., 2006. Freshwater biodiversity: Importance, threats, status and conservation challenges. *Biological Reviews*, 81 (2), 163–182.
- DUNBAR, M. J., WARREN, M., EXTENCE, C., BAKER, L., CADMAN, D., MOULD, D. J., HALL, J., AND CHADD, R., 2010. Interaction between macroinvertebrates, discharge and physical habitat in upland rivers. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 20 (1), 31–44.
- DURANCE, I., AND ORMEROD, S. J., 2007. Climate change effects on upland stream macroinvertebrates over a 25-year period. *Global Change Biology*, 13 (5), 942–957.
- EVERALL, N. C., JOHNSON, M. F., WOOD, P., PAISLEY, M. F., TRIGG, D. J., AND FARMER, A., 2019. Macroinvertebrate community structure as an indicator of phosphorus enrichment in rivers. *Ecological Indicators*, 107.
- FEWSTER, R. M., BUCKLAND, S. T., SIRIWARDENA, G. M., BAILLIE, S. R., AND WILSON, J. D., 2000. Analysis of population trends for farmland birds using generalized additive models. *Ecology*, 81 (7), 1970–1984.
- FULLER, R. M., SMITH, G. M., SANDERSON, J. M., HILL, R. A., AND THOMSON, A. G., 2002. The UK Land Cover Map 2000: Construction of a parcel-based vector map from satellite images. *Cartographic Journal*, 39 (1), 15–25.
- HYNES, H. B. N., 1974. *The Biology of Polluted Waters*. Liverpool University: Liverpool University Press.
- JOURDAN, J., O'HARA, R. B., BOTTARIN, R., HUTTUNEN, K. L., KUEMMERLEN, M., MONTEITH, D., MUOTKA, T., OZOLIŅŠ, D., PAAVOLA, R., PILOTTO, F., SPRINGE, G.,

- SKUJA, A., SUNDERMANN, A., TONKIN, J. D., AND HAASE, P., 2018. Effects of changing climate on European stream invertebrate communities: A long-term data analysis. *Science of the Total Environment*, 621, 588–599.
- KERNAN, E. M., BATTARBEE, R. W., CURTIS, C. J., MONTEITH, D. T., SHILLAND, E. M., AHERNE, J., DRIVE, W. B., KJ, C., COLLEN, P., CURTIS, C., EVANS, C., FLOWER, R., KERNAN, M., MONTEITH, D., MURPHY, J., AND NISBET, T. 2010. Recovery of lakes and streams in the UK from the effects of acid rain UK Acid Waters Monitoring Network 20 Year Interpretative Report. Report to DEFRA. Environmental Change Research Centre.
- LEE, L., AND HELSEL, D., 2005. Statistical analysis of water-quality data containing multiple detection limits: S-language software for regression on order statistics. *Computers and Geosciences*, 31 (10), 1241–1248.
- LINDENMAYER, D. B., LIKENS, G. E., ANDERSEN, A., BOWMAN, D., BULL, C. M., BURNS, E., DICKMAN, C. R., HOFFMANN, A. A., KEITH, D. A., LIDDELL, M. J., LOWE, A. J., METCALFE, D. J., PHINN, S. R., RUSSELL-SMITH, J., THURGATE, N., AND WARDLE, G. M., 2012. Value of long-term ecological studies. *Austral Ecology*, 37 (7), 745–757.
- MOORE, R. V., MORRIS, D. G., AND FLAVIN, R. W., 1994. Sub-set of UK digital 1:50,000 scale river centre-line network. NERC, Institute of Hydrology, Wallingford.
- NESSEL, M. P., KONNOVITCH, T., ROMERO, G. Q., AND GONZÁLEZ, A. L., 2021. Nitrogen and phosphorus enrichment cause declines in invertebrate populations: a global meta-analysis. *Biological Reviews*. [online]. Available from: <https://doi.org/10.1111/brv.12771> [Accessed 28 September 2021].
- OUTHWAITE, C. L., GREGORY, R. D., CHANDLER, R. E., COLLEN, B., AND ISAAC, N. J. B., 2020. Complex long-term biodiversity change among invertebrates, bryophytes and lichens. *Nature Ecology and Evolution*, 4 (3), 384–392.
- PERRY, M., AND HOLLIS, D., 2005. The generation of monthly gridded datasets for a range of climatic variables over the UK. *International Journal of Climatology*, 25 (8), 1041–1054.
- PINHEIRO, J., AND BATES, D., 2000. *Mixed-Effects Models in S and S-PLUS*. New York: Springer.
- R CORE TEAM (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- REID, A. J., CARLSON, A. K., CREED, I. F., ELIASON, E. J., GELL, P. A., JOHNSON, P. T. J., KIDD, K. A., MACCORMACK, T. J., OLDEN, J. D., ORMEROD, S. J., SMOL, J. P., TAYLOR, W. W., TOCKNER, K., VERMAIRE, J. C., DUDGEON, D., AND COOKE, S. J., 2019. Emerging threats and persistent conservation challenges for freshwater biodiversity. *Biological Reviews*, 94 (3), 849–873.

- ROSENBAUM, P. R., AND RUBIN, D. B., 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70 (1), 41–55.
- ROY, A. H., ROSEMOND, A. D., PAUL, M. J., LEIGH, D. S., AND WALLACE, J. B., 2003. Stream macroinvertebrate response to catchment urbanisation (Georgia, U.S.A.). *Freshwater Biology*, 48 (2), 329–346.
- SAAL, D. S., AND PARKER, D., 2000. The impact of privatization and regulation on the water and sewerage industry in England and Wales: A translog cost function model. *Managerial and Decision Economics*, 21 (6), 253–268.
- SÁNCHEZ-BAYO, F., AND WYCKHUYS, K. A. G., 2019. Worldwide decline of the entomofauna: A review of its drivers. *Biological Conservation*, 232, 8–27.
- SEAGER, K., BAKER, L., PARSONS, H., RAVEN, P. J., AND VAUGHAN, I. P., 2012. The Rivers and Streams of England and Wales: An Overview of their Physical Character in 2007-2008 and Changes Since 1995-1996. BOON, P. J. AND RAVEN, P. J. (eds). *River Conservation and Management*. Chichester: Wiley.
- STEPANIAN, P. M., ENTREKIN, S. A., WAINWRIGHT, C. E., MIRKOVIC, D., TANK, J. L., AND KELLY, J. F., 2020. Declines in an abundant aquatic insect, the burrowing mayfly, across major North American waterways. *Proceedings of the National Academy of Sciences of the United States of America*, 117 (6), 2987–2992.
- SUTHERLAND, A. B., CULP, J. M., AND BENOY, G. A., 2010. Characterizing deposited sediment for stream habitat assessment. *Limnology and Oceanography: Methods*, 8 (1), 30–44.
- VAN KLINK, R., BOWLER, D. E., GONGALSKY, K. B., SWENGEL, A. B., GENTILE, A., AND CHASE, J. M., 2020. Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances. *Science*, 368 (6489), 417–420.
- VAUGHAN, I. P., AND GOTELLI, N. J., 2019. Water quality improvements offset the climatic debt for stream macroinvertebrates over twenty years. *Nature Communications*, 10, 1956.
- VAUGHAN, I. P., MERRIX-JONES, F. L., AND CONSTANTINE, J. A., 2013. Successful predictions of river characteristics across England and Wales based on ordination. *Geomorphology*, 194, 121–131.
- VAUGHAN, I. P., AND ORMEROD, S. J., 2012. Large-scale, long-term trends in British river macroinvertebrates. *Global Change Biology*, 18 (7), 2184–2194.
- VAUGHAN, I. P., AND ORMEROD, S. J., 2014. Linking interdecadal changes in British river ecosystems to water quality and climate dynamics. *Global Change Biology*, 20 (9), 2725–2740.

WANG, L., ROBERTSON, D. M., AND GARRISON, P. J., 2007. Linkages between nutrients and assemblages of macroinvertebrates and fish in wadeable streams: Implication to nutrient criteria development. *Environmental Management*, 39 (2), 194–212.

WOOD, S. N., 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society, Series B*, 73 (1), 3–36.

WOOD, S. N., 2017. *Generalized additive models: An introduction with R*, second edition London: CRC Press.

WORRALL, F., SPENCER, E., AND BURT, T. P., 2009. The effectiveness of nitrate vulnerable zones for limiting surface water nitrate concentrations. *Journal of Hydrology*, 370 (1–4), 21–28.

Acknowledgements

Our thanks to Mark Diamond and John Murray-Bligh for comments and advice during the project, and Judy England for comments on a previous draft.

List of abbreviations

BFI - base flow index

TON - total organic nitrogen

GAM - generalised additive model

RHS - River Habitat Survey

GAMM - generalised additive mixed-effects model

P-A - presence-absence

GIS - geographic information system

Appendix 1. Calculation of post-stratification weights for smoothed national trends in invertebrate assemblages.

Post-stratification was carried out by river habitat character using data from the second River Habitat Survey Baseline (2007–8; Seager *et al.*, 2012). Environmental and site level variables from the 3506 invertebrate sites were combined with 3503 RHS sites located in England, and a logistic regression GAM was used to predict the likelihood of a river reach being sampled for macroinvertebrates based on these environmental variables. Variables used as predictors in the model were: catchment area, mean annual rainfall, proportion of catchment underlain by calcareous bedrock geology, proportion of land cover classed as urban, arable, and improved grassland, stream power, predicted base flow index (BFI) along with site easting, northing, altitude, and channel slope. Predicted probabilities from the model acted as propensity scores for sites, with RHS sites divided into five adjustment cells based upon the quintiles of predicted probabilities (Rosenbaum and Rubin, 1983; Little, 1993). The proportions of English river lengths within the five adjustment cells were calculated, thereby taking the stratification of the RHS Baseline by 1:50,000 and 1:250,000 river networks into account (Seager *et al.*, 2012). Post-stratification weights, w , were then calculated for the macroinvertebrate sites in each of the five adjustment cells, as $w_h = rP_h / r_h$ where r was the number of invertebrate sites (3506), P_h was the proportion of English rivers in adjustment cell h and r_h was the number of invertebrate sites in adjustment cell h (Little 1993). The result of this was that data from rivers underrepresented in the macroinvertebrate data (e.g. headwaters) had a greater weighting than river types that were more frequently sampled.

Appendix 2. Correspondence analysis loadings of the 78 composite and family groups used in this study.

Table A1: Correspondence analysis loadings for composite and family groups of macroinvertebrates, calculated from abundance data.

Taxon	Abundance CA loading	Taxon	Abundance CA loading
Oligochaeta	-0.99504	Caenidae	0.053268
Asellidae	-0.77743	Neritidae	0.059217
Chironomidae	-0.68485	Leptoceridae	0.068053
Libellulidae	-0.63208	Leptophlebiidae	0.08059
Viviparidae	-0.59724	Planariidae.Dugesiidae	0.101567
Coenagrionidae	-0.5588	Hydrobiidae.Bithyniidae	0.109944
Aeshnidae	-0.55101	Polycentropodidae	0.151976
Corophiidae	-0.54152	Siphonuridae	0.155995
Platycnemididae	-0.5291	Hydrophilidae.Hydraenidae	0.158995
Naucoridae	-0.50383	Piscicolidae	0.166783
Notonectidae	-0.46477	Hydropsychidae	0.174671
Valvatidae	-0.45526	Simuliidae	0.181168
Lymnaeidae	-0.43894	Cordulegastridae	0.183944
Physidae	-0.42867	Aphelocheiridae	0.201731
Nepidae	-0.4152	Chloroperlidae	0.228813
Erpobdellidae	-0.41448	Gyrinidae	0.240801
Phryganeidae	-0.40523	Capniidae	0.240909

Pleidae	-0.39177	Lepidostomatidae	0.245122
Molannidae	-0.29373	Brachycentridae	0.246791
Haliplidae	-0.29183	Astacidae	0.254416
Sphaeriidae	-0.28436	Baetidae	0.267272
Sialidae	-0.26645	Potamanthidae	0.272379
Glossiphoniidae	-0.23779	Leuctridae	0.274564
Gerridae	-0.2286	Limnephilidae	0.275324
Corixidae	-0.21591	Nemouridae	0.284573
Hirudinidae	-0.21303	Ephemeridae	0.299055
Paelobiidae	-0.21047	Perlodidae	0.34244
Hydrometridae	-0.19543	Taeniopterygidae	0.356703
Mesoveliidae	-0.19226	Perlidae	0.368521
Dytiscidae.Noteridae	-0.16058	Elmidae	0.389958
Planorbidae	-0.10457	Sericostomatidae	0.396611
Calopterygidae	-0.08961	Heptageniidae	0.442369
Hydroptilidae	-0.02992	Philopotamidae	0.47027
Unionidae	-0.00594	Goeridae	0.508397
Psychomyiidae.Ecnomidae	-0.00038	Odontoceridae	0.512633
Dendrocoelidae	0.032278	Ephemerellidae	0.586559
Dryopidae	0.032801	Scirtidae	0.656409
Tipulidae	0.049474	Rhyacophilidae.Glossosomatidae	0.784958
Beraeidae	0.051229	Gammaridae	1.258977

Table A2: Correspondence analysis loadings for composite and family groups of macroinvertebrates, calculated from presence-absence data.

Taxon	Presence-absence CA loading	Taxon	Presence-absence CA loading
Naucoridae	-2.00419	Leptoceridae	-0.12277
Aeshnidae	-1.81689	Psychomyiidae.Ecnomidae	-0.10917
Viviparidae	-1.80369	Oligochaeta	-0.09623
Notonectidae	-1.66123	Chironomidae	-0.06771
Libellulidae	-1.56147	Hydroptilidae	-0.06657
Corophiidae	-1.49041	Astacidae	-0.06277
Phryganeidae	-1.48725	Limnephilidae	-0.00579
Pleidae	-1.48131	Dryopidae	0.037072
Coenagrionidae	-1.41752	Baetidae	0.083354
Nepidae	-1.39073	Tipulidae	0.092794
Platycnemididae	-1.27015	Ephemeridae	0.100519
Molannidae	-1.21236	Simuliidae	0.18103
Corixidae	-1.16724	Elmidae	0.182279
Paelobiidae	-1.09285	Hydropsychidae	0.242325
Unionidae	-1.09106	Potamanthidae	0.278505
Hydrometridae	-1.04594	Brachycentridae	0.283086
Halplidae	-1.00849	Leptophlebiidae	0.333345
Physidae	-0.92865	Polycentropodidae	0.340745
Gerridae	-0.92406	Ephemerellidae	0.391222
Valvatidae	-0.91475	Goeridae	0.412341
Hirudinidae	-0.88522	Gyrinidae	0.452612

Sialidae	-0.76991	Scirtidae	0.459957
Calopterygidae	-0.66772	Beraeidae	0.552396
Dendrocoelidae	-0.65788	Sericostomatidae	0.597495
Neritidae	-0.54057	Rhyacophilidae.Glossosomatidae	0.625926
Lymnaeidae	-0.52392	Hydrophilidae.Hydraenidae	0.659059
Asellidae	-0.50678	Lepidostomatidae	0.694096
Piscicolidae	-0.49195	Heptageniidae	0.923758
Glossiphoniidae	-0.45767	Nemouridae	1.004671
Erpobdellidae	-0.4128	Cordulegastridae	1.028504
Dytiscidae.Noteridae	-0.37769	Leuctridae	1.077125
Mesoveliidae	-0.31783	Odontoceridae	1.104871
Sphaeriidae	-0.28124	Capniidae	1.159775
Hydrobiidae.Bithyniidae	-0.22684	Perlodidae	1.172539
Planorbidae	-0.21801	Chloroperlidae	1.410579
Aphelocheiridae	-0.18629	Taeniopterygidae	1.451569
Caenidae	-0.13219	Philopotamidae	1.580775
Planariidae.Dugesiidae	-0.12968	Siphonuridae	1.719365
Gammaridae	-0.12938	Perlidae	1.789232

Appendix 3. Smoothed macroinvertebrate time series with and without post-stratification.

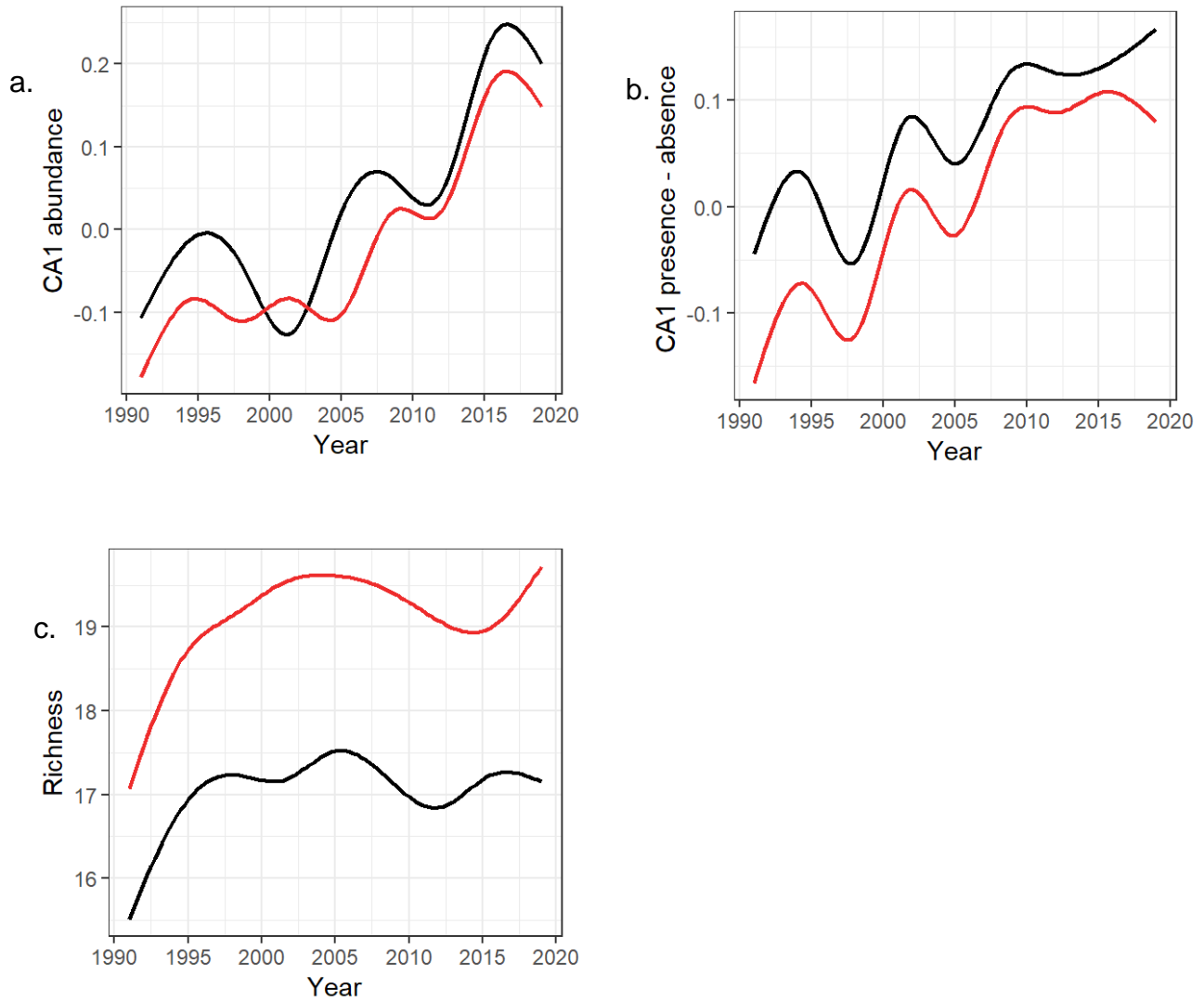


Figure A1: Smoothed time series for (a) abundance CA1 scores, (b) presence-absence CA1 scores and (c) richness. Black lines denote post-stratified trends and red lines denote estimates without post-stratification.

Would you like to find out more about us or your environment?

Then call us on

03708 506 506 (Monday to Friday, 8am to 6pm)

Email: enquiries@environment-agency.gov.uk

Or visit our website

www.gov.uk/environment-agency

incident hotline

0800 807060 (24 hours)

floodline

0345 988 1188 (24 hours)

Find out about call charges (<https://www.gov.uk/call-charges>)

Environment first

Are you viewing this onscreen? Please consider the environment and only print if absolutely necessary. If you are reading a paper copy, please don't forget to reuse and recycle.