ECOGRAPHY

Research article

Revealing hidden sources of uncertainty in biodiversity trend assessments

Martin A. Wilkes[1], Morwenna Mckenzie[2], Andrew Johnson[3], Christopher Hassall[4], Martyn Kelly[5], Nigel Willby[6] and Lee E. Brown[3]

[1]School of Life Sciences, University of Essex, Colchester, UK
[2]Geography and Environment, Loughborough University, Loughborough, UK
[3]School of Geography & water@leeds, University of Leeds, Leeds, UK
[4]School of Biology, University of Leeds, Leeds, UK
[5]Bowburn Consultancy, Durham, UK
[6]Department of Biological and Environmental Sciences, University of Stirling, Stirling, UK

Correspondence: Martin A. Wilkes (m.wilkes@essex.ac.uk)

Idiosyncratic decisions during the biodiversity trend assessment process may limit reproducibility, whilst 'hidden' uncertainty due to collection bias, taxonomic incompleteness, and variable taxonomic resolution may limit the reliability of reported trends. We model alternative decisions made during assessment of taxon-level abundance and distribution trends using an 18-year time series covering freshwater fish, invertebrates, and primary producers in England. Through three case studies, we test for collection bias and quantify uncertainty stemming from data preparation and model specification decisions, assess the risk of conflating trends for individual species when aggregating data to higher taxonomic ranks, and evaluate the potential uncertainty stemming from taxonomic incompleteness. Choice of optimizer algorithm and data filtering to obtain more complete time series explained 52.5% of the variation in trend estimates, obscuring the signal from taxon-specific trends. The use of penalized iteratively reweighted least squares, a simplified approach to model optimization, was the most important source of uncertainty. Application of increasingly harsh data filters exacerbated collection bias in the modelled dataset. Aggregation to higher taxonomic ranks was a significant source of uncertainty, leading to conflation of trends among protected and invasive species. We also found potential for substantial positive bias in trend estimation across six fish populations which were not consistently recorded in all operational areas. We complement analyses of observational data with in silico experiments in which monitoring and trend assessment processes were simulated to enable comparison of trend estimates with known underlying trends, confirming that collection bias, data filtering and taxonomic incompleteness have significant negative impacts on the accuracy of trend estimates. Identifying and managing uncertainty in biodiversity trend assessment is crucial for informing effective conservation policy and

practice. We highlight several serious sources of uncertainty affecting biodiversity trend analyses and present tools to improve the transparency of decisions made during the trend assessment process.

Introduction

Long-term, multi-species biodiversity monitoring datasets are potentially rich resources for ecosystem science and management (McCrea et al. 2023). Compiled time series covering a broad range of taxa and biomes are now openly available (Dornelas et al. 2018, Comte et al. 2021). This creates new opportunities for data-intensive approaches to ecosystem monitoring, conservation and restoration. A major application of ecological time series data is biodiversity trend assessment (BTA).

BTA focuses on trends in population level indicators such as species distribution (Outhwaite et al. 2020) or abundance (Powell et al. 2023) and community level indicators such as species richness (Dornelas et al. 2014) or composition (Pharaoh et al. 2023). Robust applications of BTA are key to classifying species conservation status (Finn et al. 2023), diagnosing pressures on ecosystems (Friberg 2014), formulating conservation and restoration plans (Tickner et al. 2020), and evaluating past interventions (Crouzeilles et al. 2016). However, best practice in the field is not fully established. Trend assessments undertaken using the same dataset can therefore reach different conclusions, generating significant debate in the literature (Seibold et al. 2019, 2021, Desquilbet et al. 2020, van Klink et al. 2020a, b, Daskalova et al. 2021, Jähnig et al. 2021, Gould et al. 2023, Boënnec et al. 2024, Johnson et al. 2024). This issue of reproducibility is echoed in a broad range of disciplines attempting to explain and predict complex phenomena through the synthesis of large datasets (Breznau et al. 2022).

Whilst consensus on BTA methodology is lacking, several themes have emerged from recent debates and model comparisons. These have overwhelmingly centred on model choice and formulation. A growing awareness of the importance of accounting for imperfect detection led to the development of species occupancy models which estimate occupancy and detection as separate processes (Mackenzie et al. 2002, Doser et al. 2022). However, simpler models can provide useful insights into biodiversity trends if the effects of imperfect detection are randomly distributed over time (Didham et al. 2020, Wenger et al. 2022). In particular, generalized linear mixed models (GLMMs) have emerged as popular tools for estimating taxon-level trends (Daskalova et al. 2021). Larger datasets may present challenges for fitting GLMMs, requiring exploration of alternative optimization algorithms to ensure convergence and model stability. In the most severe cases, GLMM coefficients may be optimized using penalized iteratively reweighted least squares (PIRLS) instead of a nonlinear optimizer, as done by Dunkley et al. (2020), Feng and Che-Castaldo (2021), Johnston et al. (2021) and Powell et al. (2023). PIRLS works by iteratively fitting a relationship using a series of weighted regressions and down weighting those points that are poorly predicted by the model, in contrast to nonlinear optimizers that seek to minimise a single cost function. In practice, model fitting by PIRLS is achieved by setting the number of points per axis for evaluating the adaptive Gauss–Hermite approximation (nAGQ) to zero rather than the more commonly used nAGQ = 1, with the latter corresponding to Laplace approximation (Bates et al. 2015). However, there is little understanding as to the veracity of trend estimates when nAGQ = 0. When nAGQ > 0, several alternative nonlinear optimizer algorithms may be selected, yet BTA studies have not previously examined how this choice affects the trend estimates reported.

Model choice, formulation, and optimizer algorithm selection are key decisions in the BTA process, yet uncertainty also arises from earlier stages, starting with exploratory data analysis and data processing (Desquilbet et al. 2020). Among the most serious yet often unresolved issues to be addressed at these stages are collection bias, taxonomic completeness, and taxonomic resolution (Fig. 1). Collection bias refers to systematic variation in the environmental and temporal structure of the observations (Fig. 1a, b), which are drawn from a biased sample of possible locations, habitat conditions and time steps (Hughes et al. 2021, McClure et al. 2023). Bias may also arise from random variation in time series length, sampling frequency and sample timing among monitoring sites. Some studies have therefore defined criteria (e.g. number of years sampled) to obtain more complete time series (Pharaoh et al. 2023, Powell et al. 2023, Qu et al. 2023). Others have suggested that strict data filtering is unnecessary for reliable estimation of population trends using GLMMs (Wenger et al. 2022) or, in contrast, that accurate determination of trend magnitudes is highly sensitive to sampling frequency (Wauchope et al. 2019). Taxonomic (in)completeness refers to the extent to which all focal taxa were recorded, whether detected or not, at the requisite taxonomic resolution at all locations and time steps (Wetzel et al. 2018) (Fig. 1c). Taxonomic resolution refers to the rank (e.g. family, genus, species) at which focal taxa are recorded, where detected, which may vary randomly from visit to visit and systematically by observer, time step, site characteristics, sampling conditions and taxonomic group (Guillera-Arroita 2017). When a focal taxon (e.g. a species) is described at multiple ranks within which other extant taxa are nested, simply excluding records at higher ranks (e.g. genus, family) risks exacerbating problems with imperfect detection if taxonomic resolution varies over time or along other dimensions of interest
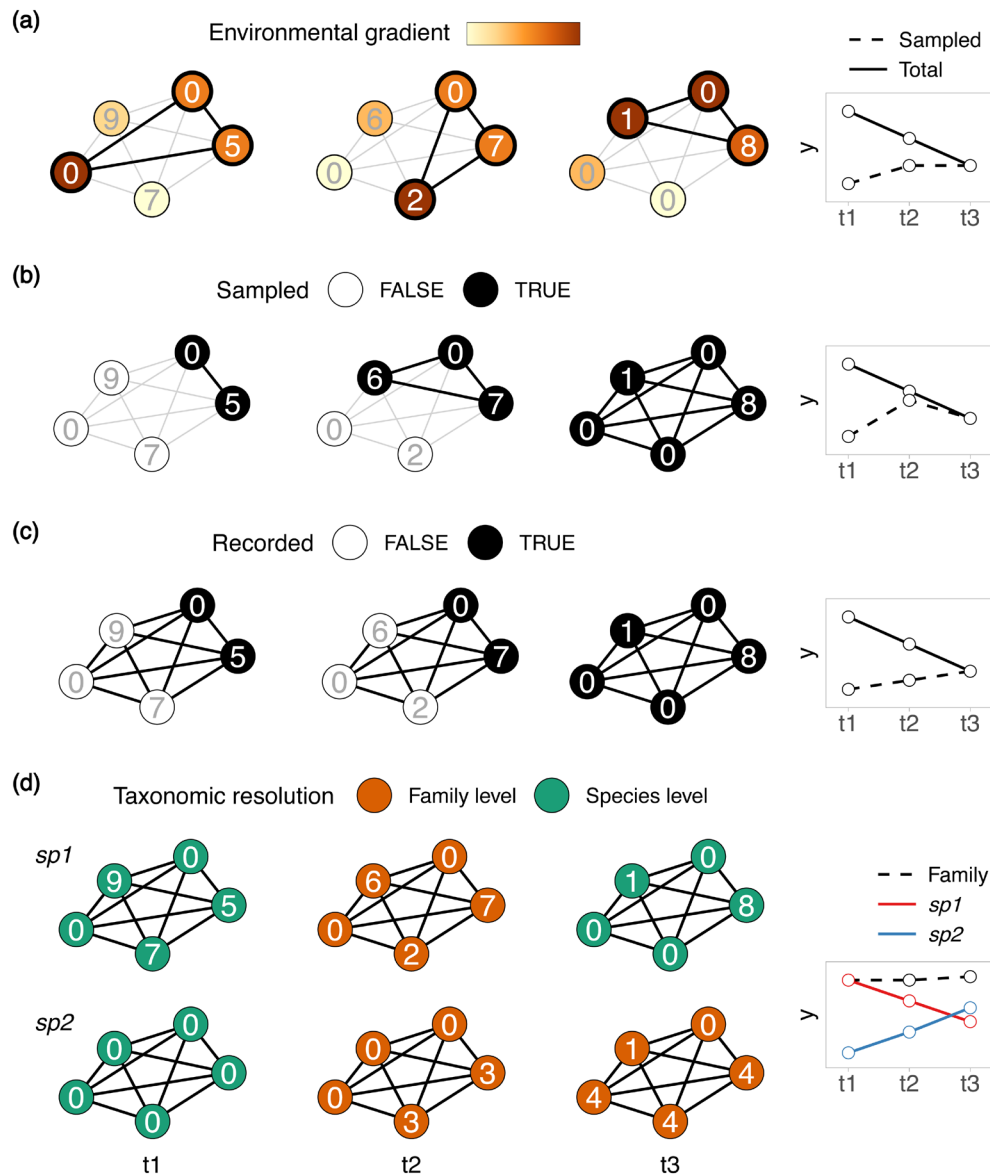
Figure 1. Potential biases in ecological monitoring datasets: environmental collection bias in which sampling does not capture the full range of an environmental gradient (a); temporal collection bias in which the number of sites sampled changes over time (b); taxonomic incompleteness in which a species population is not consistently recorded, where detected, at all sampled sites (c); and variable taxonomic resolution in which data at some time steps are recorded at family level and records for two confamilial species are therefore aggregated to family level (d). For each panel, counts for a single taxon are represented at the nodes of a small monitoring network consisting of five locations and three time steps (t1, t2, t3). Plots to the right illustrate the underlying (total) and sampled population abundance trends.

(Wilkes et al. 2024). On the other hand, aggregating records to higher ranks (e.g. family) risks conflating opposing trends at the species level (Fig. 1d). This taxonomic aggregation may confound BTA, such as when native and invasive non-native species with opposing trends are nested within the higher rank.

The reliability of BTA may be compromised if the issues of collection bias, taxonomic completeness and taxonomic resolution are not investigated fully during exploratory data analysis and addressed appropriately. Despite this, formal risk-of-bias assessment in the context of BTA is rare, and

there is little understanding of the uncertainty that data processing and model specification decisions introduce into trend estimation (Boyd et al. 2022). In this study, we evaluate 'hidden' uncertainty in population distribution and abundance trend estimates stemming from alternative analytical decisions using an 18-year statutory biological monitoring dataset covering freshwater fish, invertebrates, diatoms and macrophytes (higher aquatic plants) at a total of 26 730 sites on rivers in England (Fig. 2). The dataset represents those typical of statutory freshwater biological monitoring programmes worldwide in which records are generated
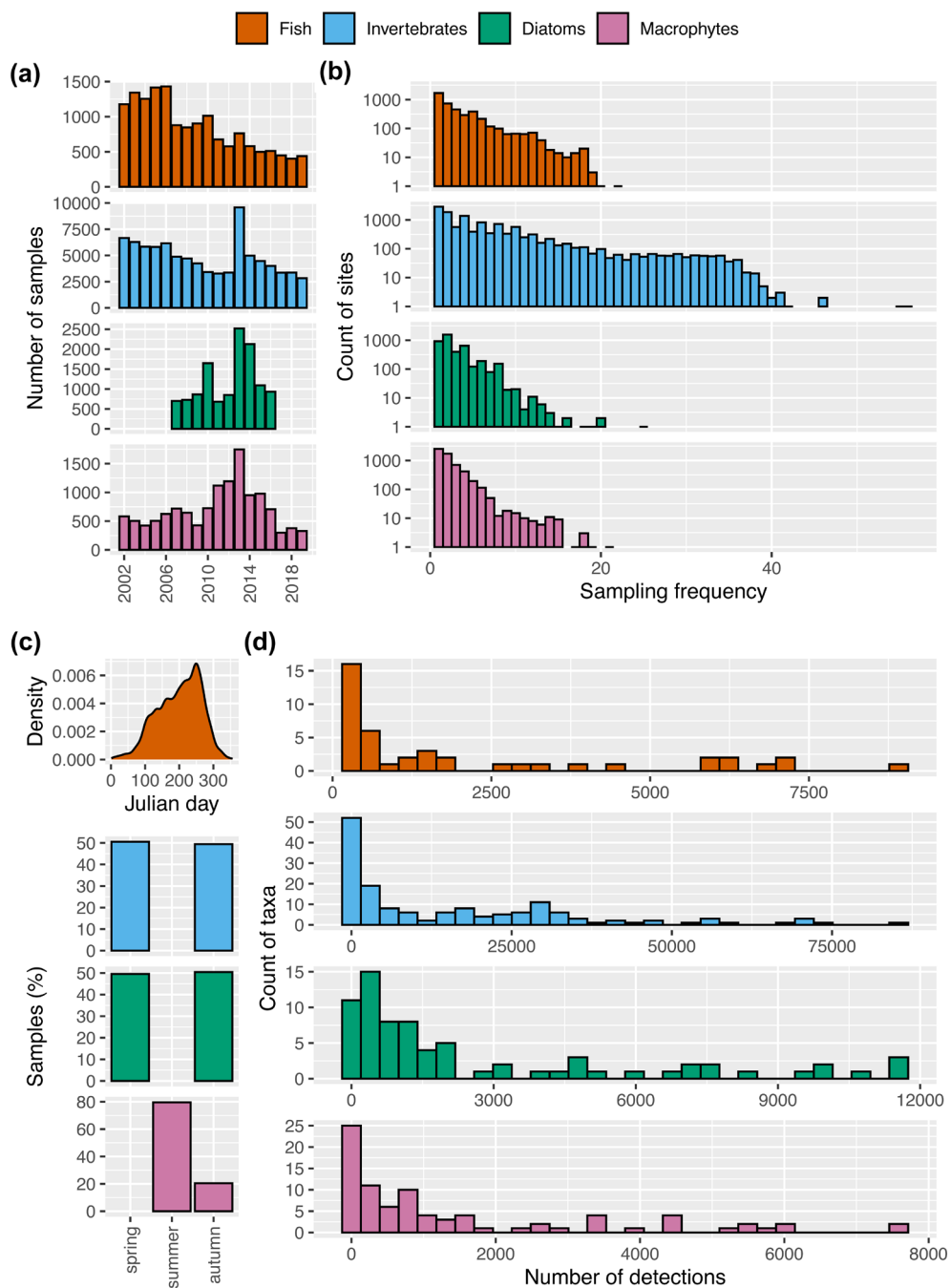
Figure 2. Summary of modelled data: total number of samples per year (a); sampling frequency per site (b); distribution of sample timing (c); and detection frequency per taxon (d). In (c), values are Julian day (day of the year) for fish and season for invertebrates, diatoms and macrophytes. The data in (d) are shown for records aggregated to a fixed rank. For fish, all records were retained at species level except for a single taxon (Petromyzontidae, family level). Invertebrate records were aggregated to family level. Diatom and macrophyte records were aggregated to genus level.

through unreplicated field survey techniques specific to focal taxonomic groups (Canadian Aquatic Biomonitoring Network 2012, Department of Water and Sanitation 2016, Environment Protection Authority Victoria 2021, Environmental Protection Agency Ireland 2023).

We model alternative decisions during BTA through three case studies. The first focuses on collection bias and uncertainty due to data filtering and model specification decisions using records for 33 fish taxa (32 species, 1 family), 124 invertebrate families, 74 diatom genera, and 78 macrophyte genera (Fig. 2d). The second case study assesses the risk of conflating trends when aggregating data to higher ranks, focusing on five invertebrate families and one macrophyte genus within which native (including protected species)

and invasive non-native taxa are nested. The final case study focuses on uncertainty stemming from taxonomic incompleteness using records for six small-bodied fish species which were not consistently recorded, where detected, at all monitoring sites throughout the study period. We complement analyses of observational data with two in silico experiments to quantify the error associated with trend estimates produced using alternative sets of decisions. These experiments assess the effects of data filtering and taxonomic incompleteness, respectively, in the presence and absence of collection bias.

Material and methods

Modelled data

We downloaded the full statutory freshwater biological monitoring dataset for England from the Environment Agency's Ecology and Fish Data Explorer (Environment Agency 2023). The dataset comprises counts for fish, invertebrates and diatoms, as well as macrophyte cover in percentage bands. A full description of the dataset accompanies the source.

We filtered the raw dataset to obtain the most highly standardised subset of records available. Only records between 2002 and 2019 were considered as field and laboratory procedures were less standardised before this period. Furthermore, sampling in 2001 was disrupted by an epidemic affecting livestock which restricted access to riverbanks, whereas sampling after 2019 was affected by the Coronavirus pandemic. For diatoms, records were only available between 2007 and 2016. Records from the Dee River basin were removed as the Environment Agency covers only a small portion of this area, as opposed to other major river basins which are wholly or substantively covered by the organisation. We further filtered the dataset to contain only records collected from rivers using standardised protocols. For fish, we considered only data collected through electrofishing using a catch depletion sampling strategy where the fished area had been recorded. For invertebrates, we considered only data collected through standard three-minute kick sampling with taxon counts recorded as actual abundance, estimated abundance, or estimated log abundance (converted to a linear scale by the data owner). For diatoms, we considered only data enumerated through light microscopy and not through DNA metabarcoding. According to Environment Agency procedures, invertebrate and diatom sampling is undertaken within the defined seasons of spring (March, April, May) and autumn (September, October, November), although a small proportion of data was collected outside of these seasons. We therefore filtered data for these groups to consider only records from the defined sampling seasons to avoid any confounding effects of larger variations in sample timing.

Upon inspection, the taxonomic information accompanying the downloaded records was incomplete and contained instances of outdated taxonomy and incorrect taxonomic classifications. We therefore obtained independent taxonomic classifications using the *classification()* function from

the R package 'taxize' ver. 0.9.98 (Chamberlain and Szöcs 2013) followed by manual checks for accuracy. Before doing so, we checked the potential bias introduced by excluding 16 hybrid fish taxa, finding that records for all except a single hybrid (*Carassius auratus* × *Carassius carassius*) were accompanied by records of at least one parent species at the same site and sampling occasion. Furthermore, across all fish records, the total count of hybrid fish was < 0.7% of the total count of both parent taxa in all cases. We therefore excluded hybrid fish from further analysis.

For invertebrates, the raw taxon list for entering into taxonomic classifications extended to 1979 unique taxa. To make the taxon list more tractable, we ran tests to check for the effect of different filtering strategies, finding that removing taxa with < 100 detections and < 250 total counts reduced the list by 60% whilst excluding only 0.05% of total invertebrate counts within the dataset (Supporting information). We therefore excluded taxa below this threshold and classified the remaining 782 invertebrate taxa. Following classification, we further excluded non-target invertebrate taxa that were frequently recorded at phylum (Nematoda, Nematomorpha), class (Anthozoa, Arachnida, Collembola, Copepoda, Microturbellaria, Oligochaeta, Ostracoda, Tricladida), and order levels (Cladocera, Lepidoptera). Records for these taxa and all child taxa were excluded from further analysis, with the exception of the aquatic lepidopteran family Crambidae.

For diatoms, the raw taxon list contained 1163 taxa. Filtering tests determined that removing taxa with < 30 detections and < 10 000 total cell counts reduced the list by 60.0% but removed only 0.5% of the total cell count across the whole diatom dataset (Supporting information). We therefore excluded taxa below this threshold, as well as planktonic (e.g. *Cyclotella*) and marine taxa (e.g. *Tryblionella levidensis*), and classified the remaining 427 diatom taxa. Following classification, we excluded a small number of records at phylum (Ochrophyta) and class levels (Bacillariophyceae), each representing < 0.4% of the total cell counts recorded for all respective child taxa. For macrophytes, we classified all 799 unique taxa in the raw taxon list. We then filtered the resulting classification to include only those taxa listed in the LEAFPACS 2 tool for classifying river ecological status (UKTAG 2014). Records for genera represented in LEAFPACS 2 were checked manually and included only if all extant UK child species are considered aquatic.

Prior to fitting GLMMs, fish count records generated through a catch depletion electrofishing strategy were converted to densities by dividing the maximum weighted likelihood estimate (Carle and Strub 1978) by the fished area (expressed as number of individuals per 1000 m²). Invertebrate records were retained as counts. Diatom and macrophyte records were converted to a binary detection–nondetection response. For diatoms, this was necessary because a fixed number of total cells were counted for each sample, which would provide only a relative estimate of abundance. After filtering and taxonomic classification, records available for trend analysis covered a total of 127 381 samples (Fig. 2a) with a mean sampling frequency per site of 4.8 (Fig. 2b).

Generalized linear mixed models

Across all three cases studies, we fitted GLMMs to records for each taxon using the *glmer()* function from the R package 'lme4' ver. 1.1-31 (Bates et al. 2015). Both Poisson and negative binomial models were fitted to fish densities and invertebrate counts. For diatoms and macrophytes, binomial models were fitted to binary detection–nondetection records. For each model, a fixed effect of time was included as a continuous variable (time since the start of the study period, in years; Fig. 2a), as well as random intercepts and slopes on site to account for spatial pseudoreplication and within-site trend variation. Following guidance on BTA model formulation (Daskalova et al. 2021), we included a random intercept on year. For fish and invertebrate models, observation level random effects were also added to account for non-zero-inflated overdispersion (Harrison 2014). We included further fixed effects to account for variation in sample timing. For fish, this was achieved through the addition of a quadratic effect on Julian day (day of the year). For other taxonomic groups which were sampled in defined seasons, season (as a factor) was the additional fixed effect (Fig. 2c). Formulaic representations of each model are presented in the Supporting information. Models for all taxa were fitted using six alternative nonlinear optimizer algorithms with nAGQ = 1: the 'glmer()' default (combination of bobyqa and Nelder_Mead), bobyqa, Nelder_Mead, nlminbwrap, nloptwrap, and L-BFGS-B from the R package 'optimx' ver. 2022-4.30 (Nash and Varadhan 2011, Nash 2014). A further model variant for each taxon was fitted using PIRLS by setting nAGQ to zero (hereafter 'nAGQ = 0'). Count models were tested for overdispersion and zero inflation using the *testDispersion()* and *testZeroInflation()* functions, respectively, from the R package 'DHARMa' ver. 0.4.6 (Hartig 2022).

GLMMs produced a range of warning messages (Supporting information). Across all optimizer algorithms and distributional families, 71.8% of models generated a warning or error. In 22.6% of cases an error indicated that no model was found due to the objective function returning infinite or NA values, or in the case of nAGQ = 0, the PIRLS loop returning an NA value. Models fitted with nAGQ = 0 more frequently returned a model and less frequently returned a warning message. Convergence failures were indicated for 15.1% of all models, due to either a failure to converge within the maximum number of iterations (10.3%) or a degenerate Hessian with negative eigenvalues (4.8%). A further 3.9% of models produced extreme estimates identified through diagnostic checks. These models were excluded from further analysis along with models that failed to converge. For 26.2% of models, a singular fit warning indicated that random effects variances were close to zero, suggesting that random effects structures could be simplified but not presenting an issue in terms of convergence. Dispersion ratios for invertebrate Poisson models were < 1, indicating that this distributional family was an appropriate choice for all invertebrate taxa (Supporting information). The warning 'gradient contains NAs' exclusively affected Poisson

models for fish, being returned for models fitted with all six nonlinear optimizer algorithms, suggesting that this distributional family may not be an appropriate choice for fish. We therefore considered negative binomial models to be appropriate for all fish taxa.

Model specification and data filtering uncertainty

To test for uncertainty introduced by model specification and data filtering decisions, we fitted GLMMs to the full dataset containing all 26 730 sites across the four taxonomic groups, and another set of GLMMs to the same dataset filtered to contain only sites sampled for three, six, nine and 12 years within the study period (Fig. 3). For invertebrates and diatoms, which were monitored in defined seasons, we considered only sites sampled in both seasons in at least the corresponding number of years to meet the threshold. Due to the limited data available for diatoms and macrophytes, only the three-year filter was applied to these groups.

Due to the frequent recording of taxa at genus and family levels, we aggregated records to a higher rank than species where necessary (Supporting information). For fish, densities for all except a single taxon (Petromyzontidae, aggregated to family level) were retained at species level. Lampreys (Petromyzontidae) were frequently recorded at family level due to the difficulties of identifying juveniles to species level (G. Peirson, unpubl.). For invertebrates, all counts were aggregated to family level. Records for diatoms and macrophytes were aggregated to genus level. We performed an unbalanced hierarchical analysis of variance (ANOVA) on GLMM results for each taxonomic group to determine the proportion of variance in trend estimates explained by taxon identity, optimizer algorithm (including nAGQ = 0) within taxon, and sampling frequency filter (number of years) within optimizer algorithm within taxon. ANOVAs were fitted using the *aov()* function from the R package 'stats' ver. 4.2.2 (www.r-project.org).

To assess environmental collection bias, we gathered data from a range of sources to characterise the abiotic environment and human pressures surrounding the monitoring sites, as compared to the wider river network. Human population density in 2020 (1 km resolution) was obtained from WorldPop (WorldPop 2023). Data on river and catchment physiography (50 m resolution) were obtained from the River Invertebrate Classification Tool (RICT) (SEPA 2023). Riparian land cover data were obtained from UK Centre for Ecology and Hydrology Land Cover Maps (1 km resolution) relating to the years 2000, 2007, 2015, 2017, 2018, 2019 and 2020 (UKCEH 2023). We created a land cover time series for each grid cell through linear interpolation of each land cover class between consecutive land cover maps, before standardising back to 100% total land cover. From these gridded data we extracted values for each record in the modelled dataset. To represent the wider river network, we extracted values at the centroid of each RICT grid cell. We quantified collection bias by comparing observed environmental conditions against conditions within the wider river network using
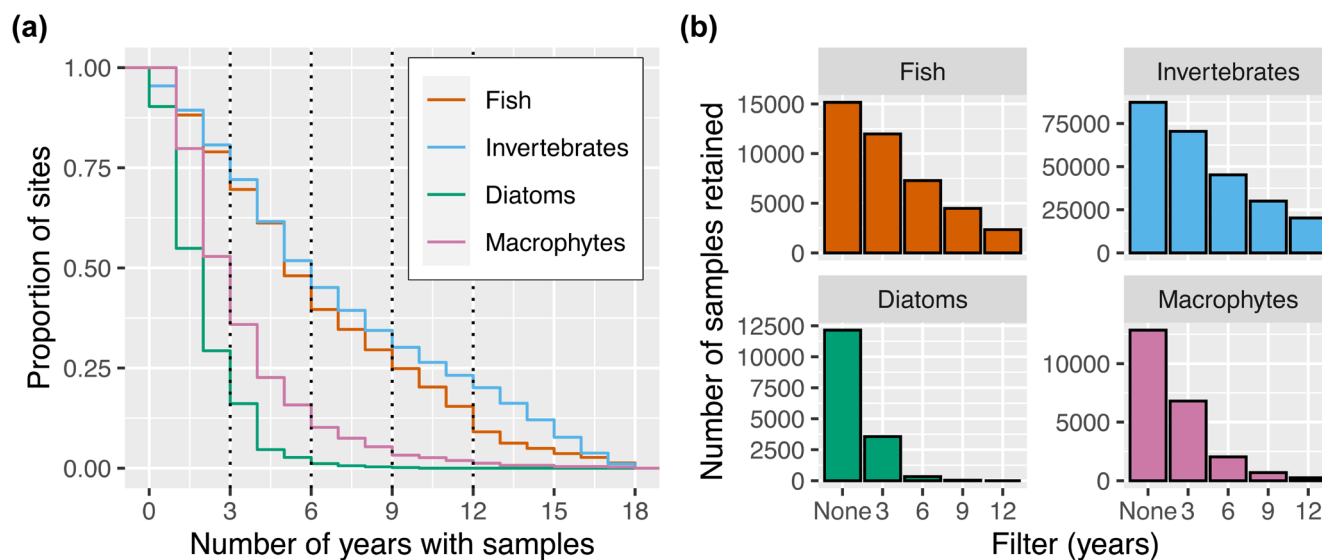
Figure 3. Summary of data filtering based on sampling frequency: empirical cumulative distribution function showing the proportion of sites sampled for at least the corresponding number of years (a); number of individual samples retained under four increasingly harsh filtering strategies based upon the minimum sampling frequency (b). In (a), dashed vertical lines denote the sampling frequency thresholds used in the analysis. In (b), 'None' refers to cases where data were not filtered by sampling frequency.

the D statistic from a two-sample, two-sided Kolmogorov–Smirnov test with the *ks.test()* function from the R package 'stats' ver. 4.2.2 (www.r-project.org).

Data aggregation uncertainty

The modelled dataset includes records at a broad range of taxonomic ranks (www.r-project.org). Thus, a common approach to BTA using this dataset is to aggregate records to a higher rank (e.g. family) and estimate the trend for the higher rank taxon (Pharaoh et al. 2023, Powell et al. 2023, Qu et al. 2023). However, this decision risks conflating opposing trends at lower ranks (e.g. species), which is problematic for BTA, particularly in cases where native (including protected species) and invasive non-native taxa are nested within the higher rank. Thus, for each lower rank taxon nested within a higher rank that included both native and non-native taxa, we assessed this risk by removing samples in which the higher rank taxon was detected but not the lower rank taxon. In cases where both the lower and higher rank taxon were recorded in the same sample, we assumed that the higher rank taxon was not an individual of the lower rank taxon.

We considered lower rank taxa (genus, species) down to the coarsest taxonomic resolution necessary to ensure that all nested taxa within the lower rank had the same designation (i.e. no designation, protected, non-native). The sample removal procedure retained between 64.9% and 99.7% of available samples across 25 lower rank taxa nested within six higher ranks that contain both native and non-native taxa (Fig. 4). We fitted GLMMs to the remaining lower rank records and compared the resulting trend estimates with those previously returned for the corresponding higher rank taxon. Among models fitted with alternative nonlinear optimizer

algorithms, the model variant with the minimum Akaike information criterion (AIC) was selected for each taxon.

Taxonomic completeness uncertainty

Among the fish taxa observed within the modelled dataset, six small-bodied species were not recorded consistently, where detected, across all operational areas at all time steps. Although an Environment Agency mandate to record small-bodied fish was given to all operational areas in 2007, it was not consistently implemented until 2012 at the earliest. Only in the Anglian operational area were these species recorded throughout the entire study period (G. Peirson, unpubl.). However, it would not be possible to determine this through exploratory data analysis alone (Fig. 5). This a clear example of taxonomic incompleteness (Fig. 1c). To understand the potential uncertainty stemming from a failure to account for this, we fitted two alternative models to density records for these small-bodied fish species, one at the national scale and the other within the Anglian operational area only. Among models fitted with alternative nonlinear optimizer algorithms, the model variant with the minimum AIC was selected for each taxon and alternative model (national, Anglian).

Summary trend estimates

To provide a summary of trends whilst addressing uncertainty stemming from model specification, data filtering, data aggregation and taxonomic incompleteness, we filtered trend estimates by selecting the model variant with the minimum AIC for each taxon. In the first instance, trend estimates were taken from models fitted to aggregated records using the full dataset. Estimates for higher rank taxa (e.g. families) with nested native and non-native taxa (genus, species) were
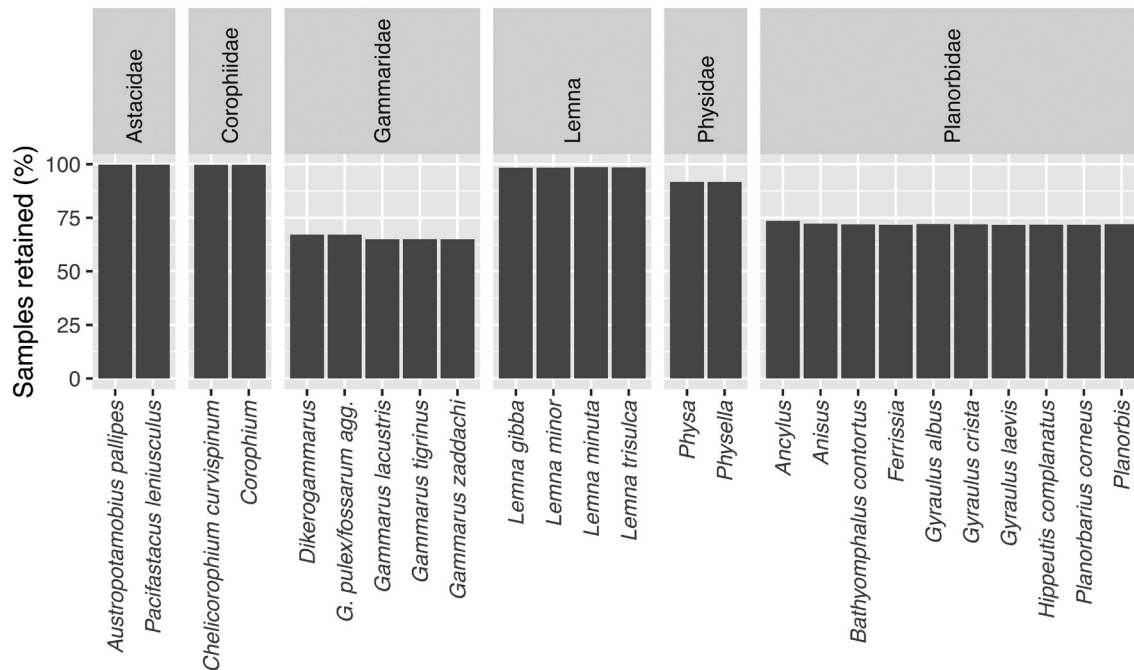
Figure 4. Samples retained for nested taxa after removing samples associated with data aggregation uncertainty.

replaced with estimates for the lower rank taxa. Estimates for small-bodied fish species were replaced with estimates from Anglian-only models. Details of selected models are provided in the Supporting information. Summary trends were reported as the annual growth rate (AGR):

$$\Psi = \left(\frac{y_n - y_1}{y_1} \right) \times 100$$

$$AGR = \frac{\Psi}{n-1}$$

(1)

where ψ is the overall percentage change, y_1 and y_n are the fitted values (density, count or occurrence probability) for the first and last year of the time series respectively, and n is the number of years in the time series.

In silico experiments

To assess the error associated with trend estimates produced using alternative sets of decisions with reference to known underlying population trends, we generated realistic river habitat configurations using optimal channel networks (OCNs) and simulated distributions and abundances of multiple species within these synthetic river networks using a process-based metacommunity model under two sets of experimental conditions. The first experiment crossed five levels of data filtering (unfiltered data and locations sampled for three, six, nine and 12 years under simulated monitoring) with two collection bias levels (no bias in the placement of simulated monitoring sites, and a bias towards monitoring on larger rivers). The second experiment crossed five levels of

taxonomic incompleteness, corresponding to the proportion of records outside of the single largest simulated river basin removed in the first half of the time series (0%, 25%, 50%, 75% and 100%), with the same two levels of the collection bias treatment as the first experiment.

We generated OCNs using the R package 'OCNet' ver. 1.2.2 (Carraro et al. 2020), synthesising 100 different OCNs on a 100 × 100 lattice with a cell size of 50 to give a total of 5×10^5 pixels, an initial state corresponding to a hip roof configuration, and a 'hot' cooling schedule (rate = 0.5, initial no cooling phase = 0.1) using the *create_OCN()* function. We then passed the resulting OCNs to the *rivergeometry_OCN()* function to calculate hydraulic properties at the nodes of the networks using Leopold's scaling relationships (Leopold and Maddock 1953). See the Supporting information for an example of the resulting river network topology and associated environment for a single example OCN, and for a comparison of six example OCNs. On each OCN, we simulated a 100-time step burn-in period for the process-based metacommunity model by repeating each row of the OCN's node data 100 times. We then generated a period of environmental change for 18 time steps. To perturb the system in a realistic way, we considered two simultaneous environmental changes. First, to simulate local environmental change at each time step, we increased flow velocity by up to 20% by multiplying the flow velocity at each node in the previous time step with a random deviate between 1 and 1.2 using the *runif()* function from the R package 'stats' ver. 4.2.2 (www.r-project.org). Second, to simulate large-scale environmental change, we took the y coordinate of nodes as a proxy for mean annual temperature and increased the value by 2.5% at each time step across the whole OCN.
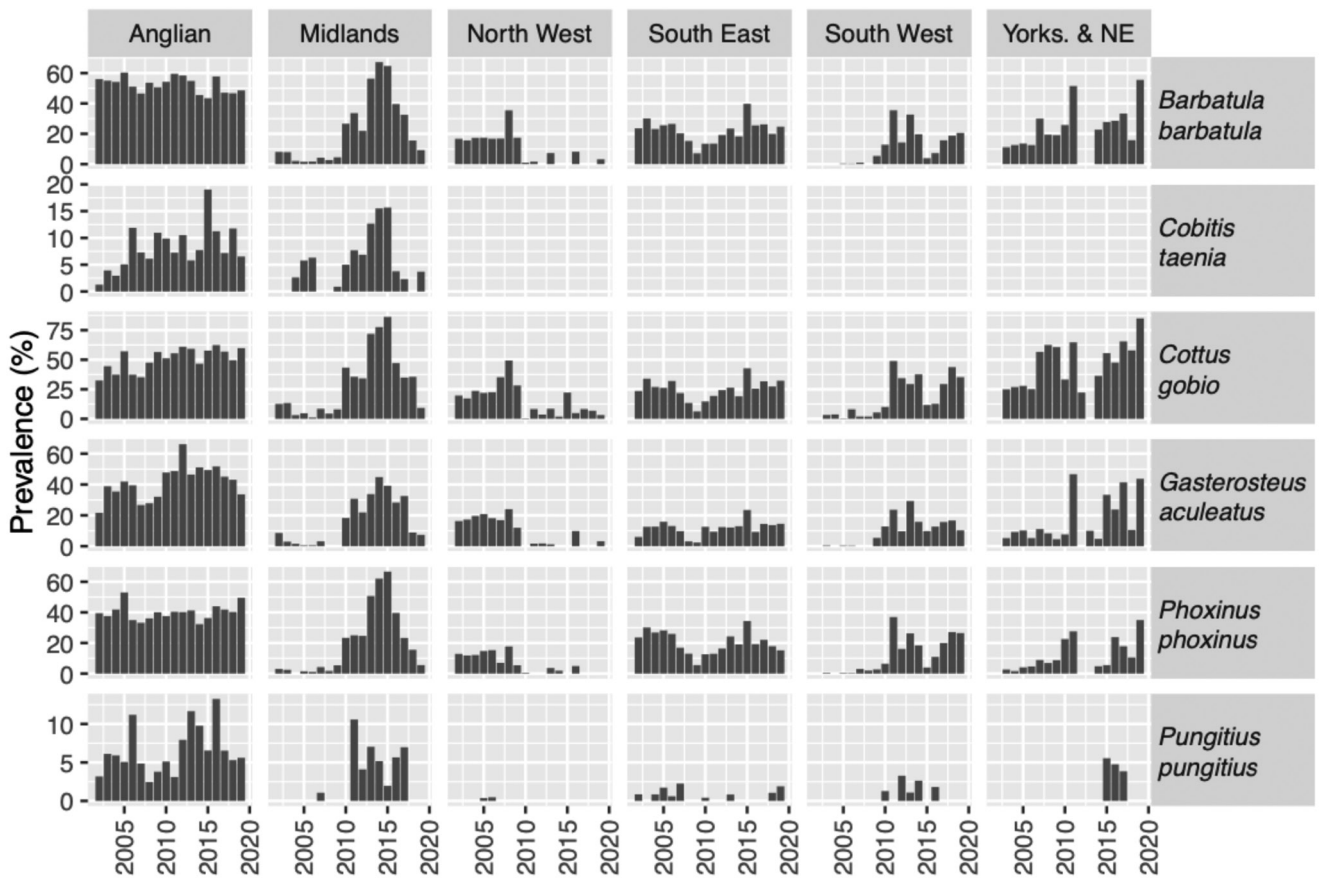
Figure 5. Records for small-bodied fish species (rows) over time in different operational areas (columns). Values shown correspond to prevalence, i.e. the percentage of total samples within each operational area and year in which the corresponding species was recorded. 'Yorks. & NE' is the Yorkshire and North East operational area.

The time series data for velocity and temperature were standardised to between 0 (minimum of the respective gradients) and 1 (maximum of the respective gradients). See the Supporting information for an example of the environmental changes simulated for a single OCN.

Using the resulting time series and topologies from each OCN, we fitted a process-based metacommunity model adapted from Thompson et al. (2020):

$$N_{ix}(t+1) = N_{ix}(t) \frac{r_{ix}(t)}{1 + \sum_{j=1}^{S} \alpha_{ij} N_{jx}(t)} + I_{ix}(t) - E_{ix}(t) \quad (2)$$

where $N_{ix}(t)$ is the abundance of species i in patch x at time t, $r_{ix}(t)$ is its density-independent growth rate (Eq. 3), α_{ij} is the per capita effect of species j on species i, $I_{ix}(t)$ is the number of individuals that arrive from elsewhere in the metacommunity via immigration, and $E_{ix}(t)$ is the number of individuals that leave via emigration.

Our simulations involved first generating 100 synthetic species with a range of realistic traits to be seeded in the model. The dispersal kernel trait (describing the probability,

at time step t, that an organism moves to any position in 2-dimensional space relative to its starting position) was generated using a uniform distribution between 0.01 and 0.1. The maximum population growth rate (intrinsic growth rate in the optimal environment) was sampled from a Poisson distribution with $\lambda = 5$. Niche optima for velocity and temperature were sampled from the respective standardised gradients in the time series generated on the OCN. Niche breadths for velocity and temperature were sampled from a beta distribution with $\alpha = 0.5$ and $\beta = 0.5$. The Supporting information shows an example of generated trait distributions. Further settings were specified to simplify simulations: The probability that an individual disperses at each time step was held constant across all species at 0.01; the probability of local extirpation for each population was set to zero; the intraspecific competition coefficient was set to 1; and interspecific competition coefficients were generated using a uniform distribution between 0 and 1, multiplied by a scaling value of 0.05 (Thompson et al. 2020). As well as adding functionality for interspecific trait variation, we built upon the model of Thompson et al. (2020) (Eq. 2) by calculating the density-independent growth rate in the presence of two environmental gradients such that:

$$r_{ix}(t) = \min_{j \in \{1,2\}} f_j \left(r_{max,i} e^{\left(\frac{z_{ij} - env_{jx}(t)}{2\sigma_{ij}} \right)^2} \right) \quad (3)$$

where $r_{ix}(t)$ is a Gaussian function of the two environmental gradients, $r_{max,i}$ is the maximum density-independent growth rate for species i, z_{ij} and σ_{ij} are the optimum and niche breadth respectively of species i for environmental gradient j, and $env_{jx}(t)$ is the value in node x at time t for environmental gradient j.

From the modelled dynamics of each extant species population with a non-zero abundance variance during the 18-time step period of environmental change, we calculated the AGR according to Eq. 1. See the Supporting information for simulated distributions and abundances of four example species over time on a single OCN. We then modelled AGRs for each set of experimental conditions using a virtual ecologist approach (Zurell et al. 2010). For the data filtering experiment, we first took a random sample of 5% of nodes (without replacement) as the monitored locations. The number of visits per monitored location was sampled from a Poisson distribution with $\lambda = 10$, truncated to a maximum of 18 visits. We then randomly allocated the sampled number of visits per monitored location across the 18-time step period of environmental change. For unbiased monitoring, we retained the full random sample of monitored locations for the unfiltered case. For the four levels of data filtering, we subsetted the full sample to include only locations monitored in at least three, six, nine and 12 time steps, respectively (Supporting information). For biased monitoring, we followed the same procedure but weighted the random sample of 5% of nodes by the Strahler stream order to simulate collection bias towards larger rivers (Supporting information).

For the taxonomic completeness experiment, we again took a random sample of 5% of nodes (without replacement) as the monitored locations, weighted by Strahler stream order in the case of biased monitoring. We then identified the largest river basin in the OCN and, for samples from monitored locations in all other river basins during the first half of the period of environmental change, we reduced a randomly sampled proportion of nonzero abundances to zero to simulate taxonomic incompleteness. The proportions of nonzero abundances reduced to zero were 0%, 25%, 50%, 75% and 100%, reflecting five levels of the taxonomic incompleteness treatment (Supporting information). Hence, at all levels of taxonomic incompleteness, monitored locations in the largest river basin retained all nonzero abundances.

We fitted Poisson GLMMs to each extant species represented in the simulated monitoring data from each experimental run using the *glmer()* function from the R package 'lme4' ver. 1.1-31 (Bates et al. 2015). These GLMMs were fitted with a fixed effect and random intercept on time step, random intercepts, and slopes on monitoring location, and an observation level random effect. Each GLMM was fitted using all seven alternative optimizers listed above and,

among these, the model variant with the minimum AIC was selected. For each species, we calculated the modelled AGR using Eq. 1, thus enabling us to compute the error associated with modelled AGRs. This experimental procedure was repeated on 100 unique OCNs. Across these OCNs, the number of species (mean ± SD) for which modelled AGRs were available was 78.9 ± 4.9.

We used a linear mixed effects model *(lmer()* function from the R package 'lmerTest' ver. 3.1-3; Kuznetsova et al. 2017) followed by a type II Wald test *(Anova()* function from the R package 'car' ver. 3.1-1; Fox and Weisberg, 2019) to test for the effects of data filtering and bias, as well as their interaction, on the AGR error whilst controlling for non-independence by including a random effect on OCN. For significant terms from the Wald test, we used a Tukey's honestly significant difference (HSD) test using the *emmeans()* function from the R package 'emmeans' ver. 1.8.5 (Lenth 2023), approximating the degrees of freedom using the Satterthwaite method (Satterthwaite 1946). We repeated this procedure to test for the effects of taxonomic incompleteness, bias, and their interaction. In both cases, extreme positive skew in the AGR error distribution was addressed by Yeo-Johnson transforming the AGR error prior to analysis using the *powerTransform()* function in the 'car' package. To further analyse the experimental results, we fitted binomial GLMMs to the occurrence of sign switching (a binary variable), i.e. when modelled AGRs were negative and known underlying AGRs were positive, or vice versa. These binomial GLMMs again included fixed effects on data filtering (or taxonomic completeness), bias, and their interaction, along with a random effect on OCN. This was followed by type II Wald tests and Tukey's HSD post hoc tests in the same way as for the linear mixed effects models described above. Finally, we also fitted a linear mixed effects model and a binomial GLMM to test for the effects of optimizer as well as experimental treatments on AGR error and sign switching, respectively.

Results

Data filtering and model specification uncertainty

A hierarchical ANOVA showed that data filtering and model specification explained considerable proportions of the variance in trend estimates (Fig. 6). If these decisions had no effect on model estimates, taxon identity would explain 100% of the variance. However, across all optimizer algorithms (including nAGQ = 0), the mean variance explained at the taxon level was only 47.5% across the four taxonomic groups (Fig. 6a). Data filtering explained an average of 28.3% of the variance in trend estimates within taxon and optimizer algorithm. For fish, taxon identity explained only 9.6% of the variance in trend estimates, with data filtering the dominant source of variation (72.6%). Excluding models fitted with nAGQ = 0 reduced the mean variance explained by optimizer algorithm from 24.2% to 2.5%, whilst the variance explained by taxon increased to a mean of 85.6% across
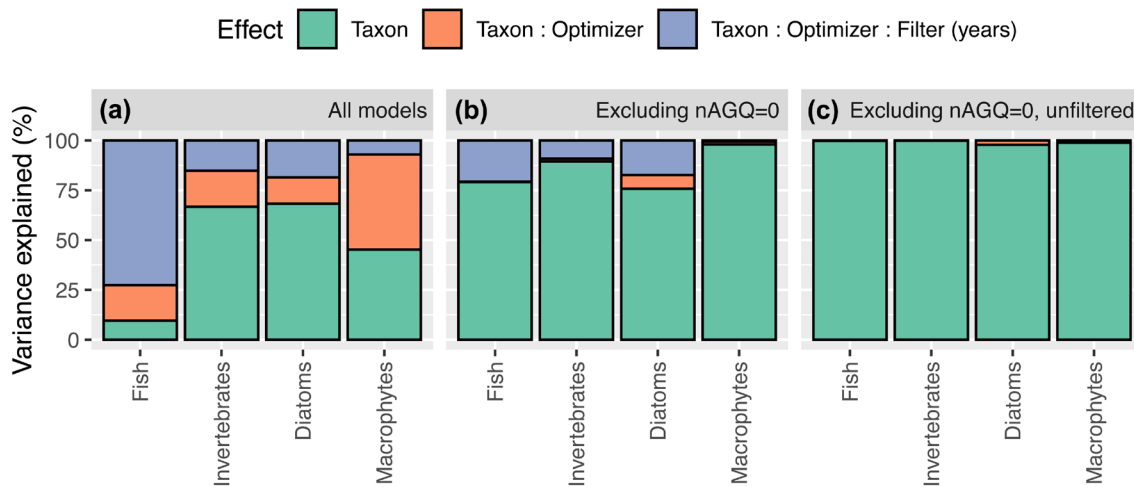
Figure 6. Variance in mean trend estimates from generalized linear mixed models explained by taxon, optimizer algorithm within taxon, and data filtering within optimizer algorithm within taxon: considering all models (a), excluding models fitted with nAGQ = 0 (b), and only models fitted to unfiltered data whilst excluding nAGQ = 0 (c).

the four taxonomic groups (Fig. 6b). Considering only models fitted to the full dataset without filtering, and excluding models fitted with nAGQ = 0, the variance explained by data filtering and model specification decisions decreased further to a negligible proportion (Fig. 6c). The dominant effect of data filtering was to reduce the trend estimate, with estimates for 61.3% of taxa decreasing with the harshness of the filter (Supporting information). For fish, trend estimates decreased with the harshness of the filter for 84.8% of taxa, compared to 62% for invertebrates, 48% for diatoms and 63% for macrophytes.

Assessment of environmental collection bias demonstrated that sampling across all four taxonomic groups was poorly representative of the wider river network within the study region, with macrophytes typically showing the greatest bias and diatoms the least (Fig. 7). The full dataset overrepresents lower elevation rivers (Fig. 7a) with larger catchment areas (Fig. 7b) and shallower longitudinal gradients (Fig. 7c) in areas with higher human population densities (Fig. 7d), lower arable land cover (Fig. 7e), and higher urban land cover (Fig. 7f). Across a broad suite of river environmental variables including land cover, hydrography, and catchment geology, in most cases data filtering exacerbated already severe collection bias (Fig. 7g, Supporting information).

Data aggregation uncertainty

Trend estimates at higher ranks were often poorly representative of estimates for nested lower rank taxa (Fig. 8). Family level estimates for Astacidae and Corophiidae (invertebrates) and the genus level estimate for *Lemna* (macrophyte) were greater than the mean estimates for all nested native species (Fig. 8a, b, d). For Gammaridae (invertebrate), the trend estimate for the invasive *Dikerogammarus* did not contribute substantially to the family level trend estimate. This is because the mean abundance of *Dikerogammarus* by the end of the time series (1.2 individuals) remained low relative to

its most abundant confamilial taxon, *Gammarus pulex/fossarum* agg. (212.3 individuals). For Physidae and Planorbidae (invertebrates), family level trend estimates were negative despite positive estimates for all nested species. This indicates that, within these families, the aggregate trend of other lower rank taxa not represented in the modelled data was negative. These lower rank taxa include *Menetus* (44 records in the raw dataset), *Planorbella* (0 records), and *Segmentina* (42 records) within Planorbidae, and *Aplexa* (24 records) within Physidae. These genera did not have sufficient records to model the trend at the lower rank but contributed disproportionately to trend estimates for the corresponding family, plausibly because individuals of these rarer genera were more frequently recorded at family level.

Taxonomic completeness uncertainty

Trend estimates returned for six small-bodied fish species were consistently higher from national models than from models fitted to records only from the Anglian operational area, in which these species were consistently recorded throughout the study period (Fig. 9). In the case of *Barbatula barbatula* and the protected *Cobitis taenia*, a strongly positive national estimate contrasted with estimates that were negative and close to zero, respectively, from the Anglian-only model.

Summary trend estimates

Across all taxonomic groups, mean trend estimates were non-negative (i.e. positive or stable) for 59.5% of taxa (Fig. 10). Within taxonomic groups, trends for the majority of fish (66.7%; Fig. 10a), invertebrate (59.5%; Fig. 10b), and diatom taxa (67.6%; Fig. 10c) were non-negative, whereas negative trends were reported for 54.3% of macrophyte genera (Fig. 10d). Reported AGRs ranged from −5.9% (macrophyte: *Littorella*, *Groenlandia*) to > 100%. AGR > 100% was reported for several non-native invertebrate taxa:
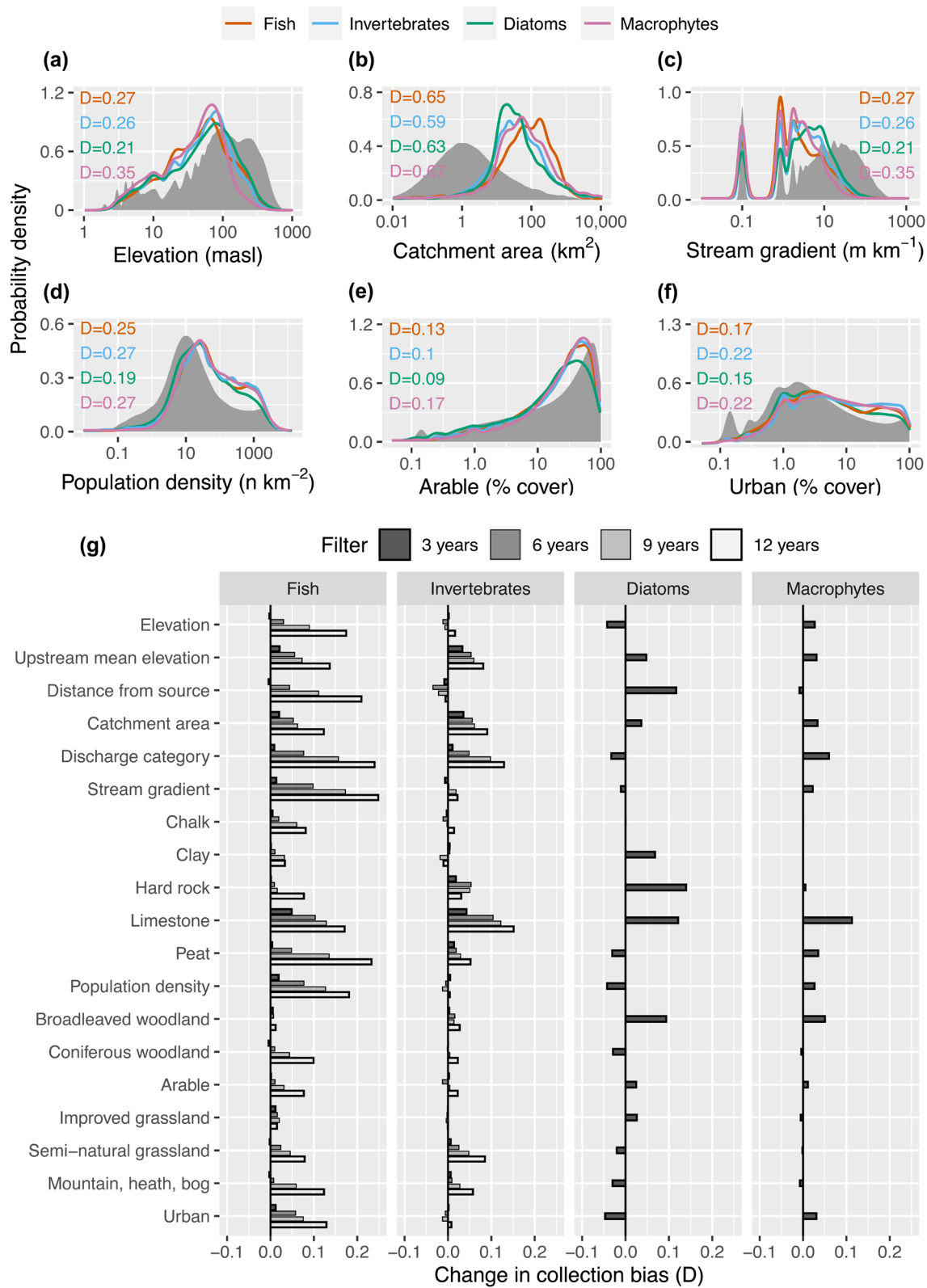
Figure 7. Collection bias in the modelled dataset: probability densities of summary variables (a–f); change in collection bias with increasing harshness of sampling frequency filter compared to the full dataset (g). In (a–f), grey-filled areas denote the probability density function for the wider river network, whereas coloured lines show the distribution of the full dataset for each taxonomic group. Text within the plot area reports collection bias as the D statistic from a two-sample Kolmogorov–Smirnov test. Larger values of D correspond to greater collection bias.
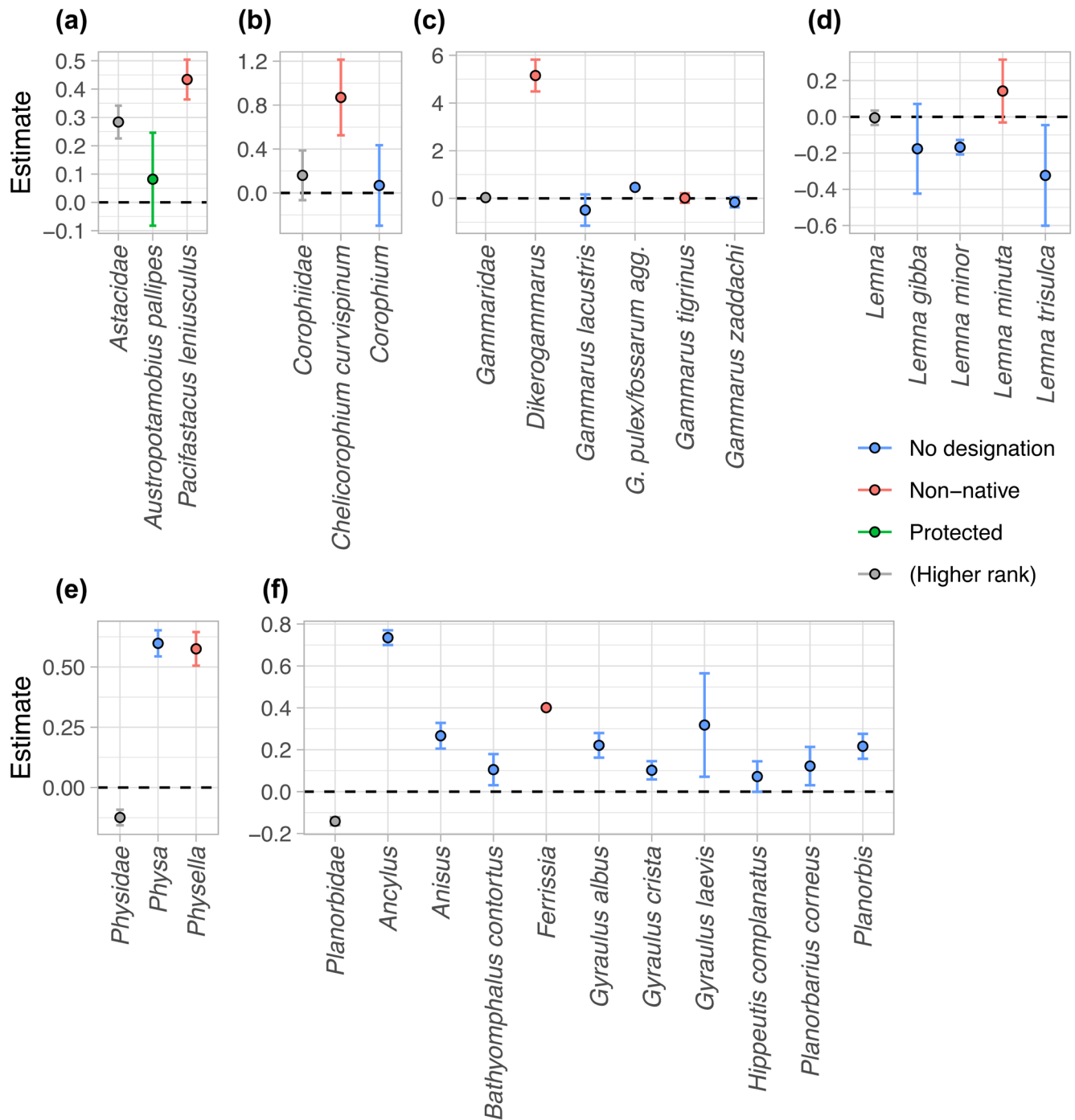
Figure 8. Trend estimates from generalized linear mixed models for lower rank taxa nested within higher ranks: Astacidae (a); Corophiidae (b); Gammaridae (c); *Lemna* (d); Physidae (e); and Planorbidae (f). Estimates are from Poisson models except for estimates in (d) which are from binomial models. Symbols denote mean estimates. Whiskers denote standard errors. Symbols and whiskers are coloured according to the designation of the lower rank taxon. 'No designation' refers to native species except for protected species, which are designated separately.

Dikerogammarus (AGR = 5.3 × 10^9 %), Tateidae (5.9 × 10^4 %), Ampharetidae (5.8 × 10^3 %), *Hemimysis anomala* (439%), and *Chelicorophium curvispinum* (185%). Several native taxa were also associated with AGR > 100%, including the macrophytes *Equisetum* (1.9 × 10^{15} %) and *Bolboschoenus* (1.0 × 10^{11} %), the diatoms *Adlafia* (3.8 × 10^4 %) and *Stauroneis*

(3.2 × 10^4 %), and the invertebrates Glossosomatidae (136%), Pediciidae (120%), and *Ancylus* (105%). Within the invertebrates, declines were more frequently reported for non-arthropods (40.5% of taxa declining), as well as the insect orders Odonata (66.7%), Hemiptera (77.8%) and Coleoptera (53.8%), compared to other arthropods (27.1%).
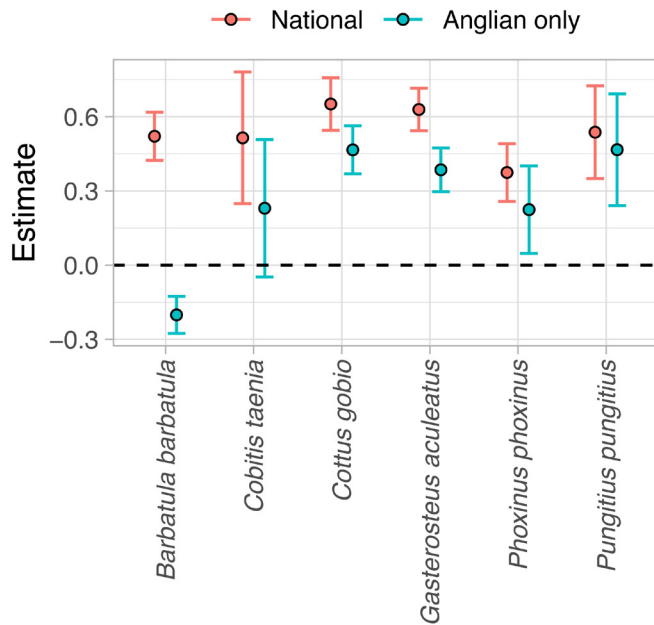
Figure 9. Trend estimates from alternative generalized linear mixed models for six small-bodied fish species. Symbols denote mean estimates. Whiskers denote standard errors.

Declines were reported across all major diatom orders, whereas macrophyte declines were more frequently found among bryophytes (67% of taxa declining) than angiosperms (47% declining).

Among non-native fish species, positive trends were reported for *Leucaspius delineatus* (AGR = 25.1%) and *Carassius auratus* (5.9%), whereas *Leuciscus idus* (−3.8%), *Oncorhynchus mykiss* (−2.4%) and *Sander lucioperca* (−3.2%) had negative trends. Among protected fish taxa, positive trends were reported for Petromyzontidae (90.8%), *Cottus gobio* (32%), *Thymallus thymallus* (10.0%), *Cobitis taenia* (8.9%), *Salmo trutta* (4.0%) and *Barbus barbus* (2.0%), whereas *Salmo salar* (−3.9%) and *Anguilla anguilla* (−2.1%) had negative trends. Of 11 non-native invertebrates, negative trends were reported for two taxa: Dreissenidae (−1.7%) and *Gammarus tigrinus* (−0.4%). A single protected invertebrate taxon, *Austropotamobius pallipes*, had a small positive mean trend estimate (AGR = 2.3%). Among the non-native macrophyte genera, *Acorus* (−1.1%) and *Lemna minuta* (4.5%) had positive trends, whereas *Azolla* (−5.6%), *Elodea* (−4.0%), and *Mimulus* (−0.7%) had negative trends. Goodness-of-fit for models summarised in Fig. 10 is illustrated in the Supporting information.

In silico experimental results

Both underestimation and overestimation, including extreme overestimation (Supporting information) and sign switching, were frequent outcomes of all treatments under both experiments (Fig. 11). In the data filtering experiment there was a significant main effect of bias on the AGR error (Wald $\chi^2(1) = 4.562$, p = 0.033; Supporting information). The post

hoc analysis showed that mean AGR error was significantly greater in the unbiased condition compared to the biased condition (Tukey's HSD: estimate = 0.212, SE = 0.095, t = 2.237, p = 0.025; Supporting information). This effect is equal to 0.24% AGR on the original scale. There was no significant main effect of data filtering (Wald $\chi^2(4) = 4.859$, p = 0.302) or interaction effect between data filtering and bias (Wald $\chi^2(4) = 2.393$, p = 0.664).

In the taxonomic incompleteness experiment there were significant main effects of bias (Wald $\chi^2(1) = 46.125$, p < 0.001) and the level of incompleteness (Wald $\chi^2(4) = 7511.905$, p < 0.001; Supporting information) on AGR error. The post hoc analysis showed a significantly greater AGR error in the unbiased condition (Tukey's HSD: estimate = 0.6, SE = 0.088, t = 6.486, p < 0.001), equal to 0.81% AGR on the original scale. There were also significantly greater AGR errors for higher levels of the taxonomic incompleteness treatment under both unbiased and biased sampling (Supporting information). This effect was greatest for the contrast between the 25% and 100% taxonomic incompleteness levels under unbiased sampling (Tukey's HSD: estimate = −10.373, SE = 0.198, t = −52.448, p < 0.001; Supporting information), with an effect equal to 9400.7% AGR confirming the abrupt increase in extreme overestimation of trends to the 100% taxonomic incompleteness level (Fig. 11b). There was no significant interaction between taxonomic incompleteness and bias (Wald $\chi^2(4) = 4.230$, p = 0.376).

Considering the probability of sign switching, under the data filtering experiment there were significant main effects of both bias (Wald $\chi^2(1) = 11.842$, p < 0.001) and data filtering (Wald $\chi^2(4) = 72.237$, p < 0.001; Supporting information). The post hoc analysis showed a significantly greater probability of sign switching in the biased condition (Tukey's HSD: estimate = −0.053, SE = 0.016, z = -3.417, p < 0.001). The probability of sign switching also increased significantly under harsher data filtering thresholds, with the greatest effect reported for the contrast between the unfiltered and 12-year scenarios (Tukey's HSD: estimate = −0.248, SE = 0.035, z = −7.032, p < 0.001; Supporting information). Under the second experiment, there was a significant main effect of taxonomic incompleteness only (Wald $\chi^2(4) = 400.816$, p < 0.001; Supporting information), with post hoc analysis showing significantly a greater probability of sign switching under higher levels of taxonomic incompleteness for most contrasts. The effect reached a maximum under the contrast between 0% and 100% incompleteness (Tukey's HSD: estimate = −0.437, SE = 0.024, z = −18.126, p < 0.001; Supporting information).

Considering the performance of each individual optimizer algorithm, using the data filtering experiment as an example, models fitted with nAQQ = 0 had significantly greater absolute AGR error than all other optimizers with the exception of bobyqa (Supporting information). The probability of sign switching was less sensitive to the optimizer used, although models fitted with nAGQ = 0 had a significantly lower probability of sign switching than bobyqa and optimx optimizers (Supporting information).
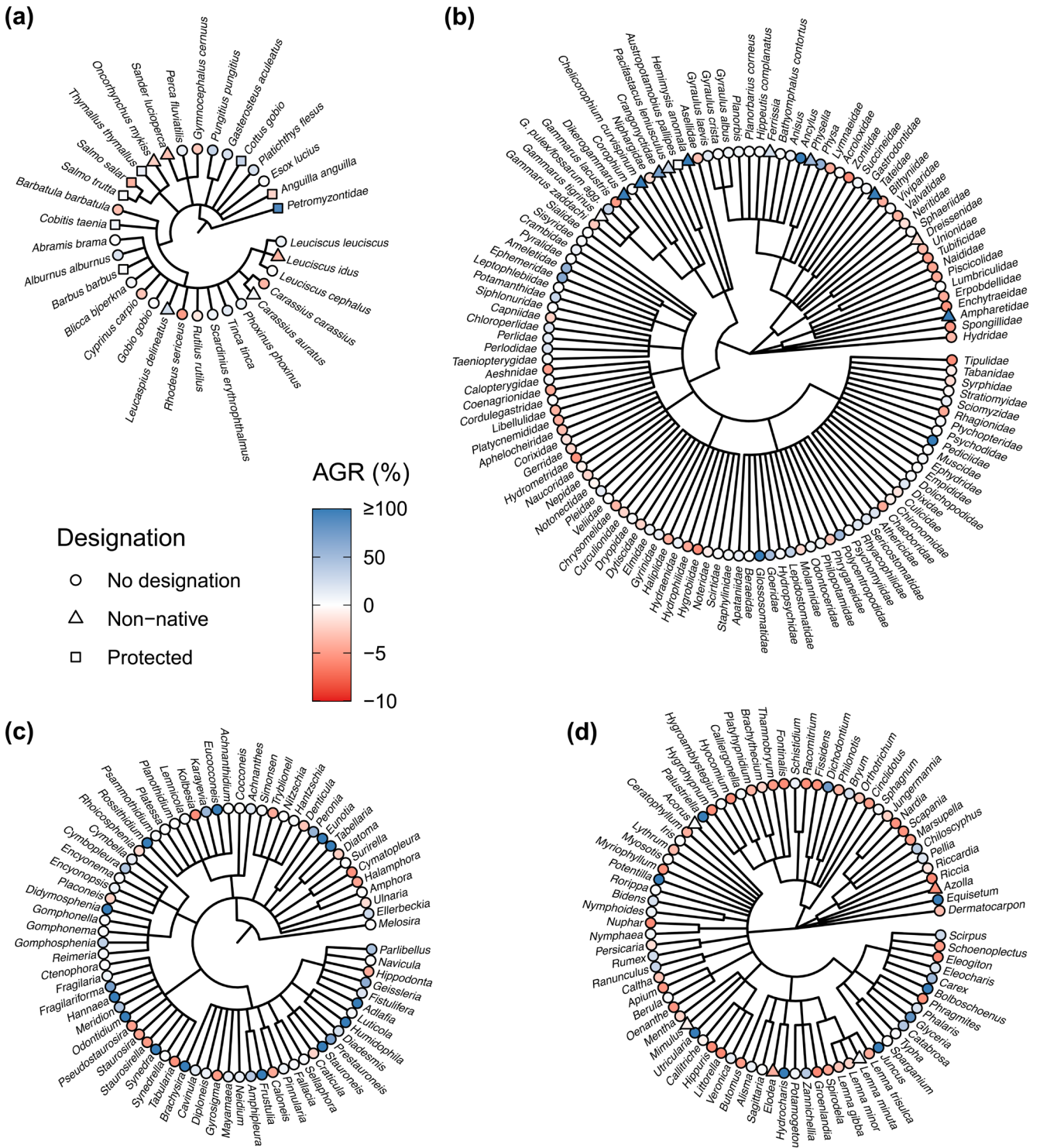
Figure 10. Summary trend estimates for fish (a), invertebrates (b), diatoms (c), and macrophytes (d). Values reported correspond to the annual growth rate (AGR). Positive AGR values are truncated to 100% for visualisation. Trend estimates are taken from models fitted to the full dataset except for six small-bodied fish species (*Barbatula barbatula*, *Cobitis taenia, Cottus gobio, Gasterosteus aculeatus, Phoxinus phoxinus*, *Pungitius pungitius*). More information on the model from which each of these summary trends is taken is specified in the Supporting information.

Figure 11. Summary of annual growth rate (AGR) errors, including sign switching, under the data filtering (a) and taxonomic completeness (b) experiments. Results for each experiment show the proportion of modelled AGRs, within each band of absolute erro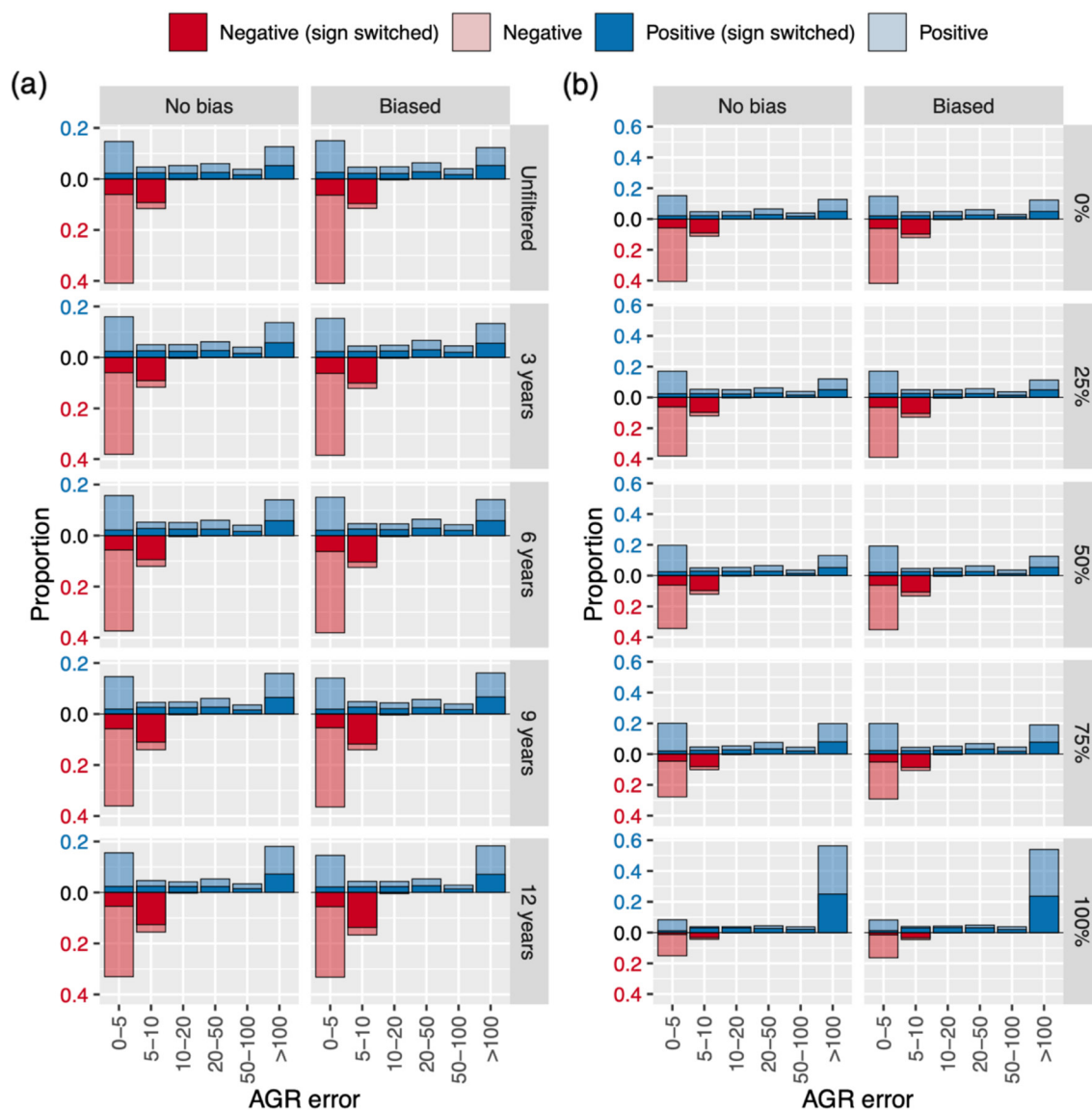r, for combinations of collection bias (columns) and data filtering (a) or taxonomic incompleteness (b) treatment levels (rows). Opaque colours denote the proportion of all AGRs within that band for which the modelled AGR was positive (blue) or negative (red) and the known underlying AGR was of the opposite sign.

Discussion

Hidden uncertainty in biodiversity trend assessments

Realising the promise of 'big data' in ecology requires consideration of uncertainty stemming from the whole analytic pipeline, from exploratory data analysis to model selection and inference (McCrea et al. 2023). In the context of BTA, our results show that idiosyncratic decisions made during data preparation and modelling stages have the potential to introduce substantial uncertainty, mirroring findings in diverse fields of science now grappling with the synthesis of large datasets (Breznau et al. 2022, Gould et al. 2023).

In our analysis of observational data, the choice of optimizer algorithm explained a considerable proportion of variation in trend estimates at the taxon level. Approximation of the likelihood using nAGQ = 0 was the dominant source of model specification-related variation. Considering only the six nonlinear optimizer algorithms implemented with nAGQ = 1, variation explained by choice of optimizer was negligible (Fig. 6b, c). This is to be expected given that alternative nonlinear optimizers should converge on the same solution, with only small variations due to numerical imprecisions that accumulate during the optimization process, although it should be noted that the bobyqa algorithm was associated with significantly greater error in the in silico

analysis. When nAGQ > 0, the random effects are integrated out during estimation of the fixed effects. On the other hand, when nAGQ = 0, random effects only influence estimation of the fixed effects through their conditional modes (Bates et al. 2015), which can result in a suboptimal solution despite no model convergence warnings. This finding was supported by results from our in silico experiments, showing that models fitted with nAGQ = 0 were associated with significantly greater absolute error than five of the six nonlinear optimizers (Supporting information). For these reasons, setting nAGQ = 0 for BTA should be seen as a last resort.

The decision to focus on more complete time series by filtering on the basis of site-level sampling frequency was also an important source of variation, generally leading to a reduction in the magnitude of trends and exacerbating already severe collection bias in the modelled dataset. Our in silico experiments showed that data filtering significantly increases the risk that true negative trends are misreported as positive trends, and vice versa (Fig. 11, Supporting information), although there was no support for the hypothesis that filtering increased the error associated with resulting trend estimates. Experimental results also revealed that, in the presence of data filtering, collection bias introduced a significant increase in the risk of sign switching. Recent studies have highlighted sampling biases that distort our view of global biodiversity trends (Saunders et al. 2020, Hughes et al. 2021). Our findings indicate that biases can also be important to consider at the national level. In particular, the over-representation of larger, lowland rivers in highly populated urban areas, together with underrepresentation of agricultural landscapes, biases reported trends towards those driven by changes in wastewater management and other features of the built environment. Data filtering exacerbated this effect, which was particularly strong for fish because electrofishing sites with higher sampling frequencies were concentrated in more densely populated urban areas (Supporting information). Sampling frequency filters are commonly applied in BTA (Haase et al. 2023, Pharaoh et al. 2023, Powell et al. 2023, Qu et al. 2023). We show that this introduces unnecessary bias when assessing trends using GLMMs since mixed models are inherently suited to estimating temporal trends at the statistical population level by incorporating data from all sites (with their random intercepts and slopes), regardless of sampling frequency. The exception to this is for models with autoregressive terms if missing count data are not imputed as part of the modelling process (Wenger et al. 2022).

The common practice of aggregating records to higher taxonomic ranks obscured trends among nested taxa, with potentially serious consequences for understanding the status of protected and invasive species. Crayfish (Astacidae) represented the clearest example of this issue in our findings (Fig. 8a). Our estimated AGR for the endangered white-clawed crayfish *Austropotamobius pallipes* was 2.3%, whereas for the invasive signal crayfish *Pacifastacus leniusculus* it was 27.5%. The family level AGR of 12.4% lay between the two species level estimates. Similar outcomes were found for several other higher rank taxa, demonstrating the need

to exercise caution in the interpretation of aggregated trend estimates, particularly when protected and invasive species are nested together. For bladder snails (Physidae; Fig. 9e) and ramshorn snails (Planorbidae; Fig. 9f), we inferred that negative trends among lower rank taxa with few (or zero) records at the species or genus level nevertheless contributed strongly to family level estimates. In these cases, negative family level trends were not representative of any individual lower rank taxon modelled, emphasising the complexities of interpreting the results of BTA on datasets with variable taxonomic resolution.

Our findings on taxonomic completeness showed that failure to detect and address inconsistencies in the recording of certain taxa through time and space confounds BTA. A comparison of alternative models for six small-bodied fish species suggested that national level models may have overestimated trends due to zero inflation earlier in the time series (Fig. 5, 9). This was directly confirmed through in silico experiments, with extreme positive bias and elevated risk of sign switching introduced at higher levels of taxonomic incompleteness (Fig. 11, Supporting information). This could bring severe consequences for the conservation of protected species. For example, in the case of the spined loach *Cobitis taenia*, we estimated an AGR of 40.2% at the national level compared to 8.9% within the only operational area (Anglian) where this species was consistently recorded, where detected, throughout the study period. For the native stone loach *B. barbatula*, a strong positive trend (AGR = 41.4%) at the national level contrasted with a strong negative trend from the Anglian-only model (AGR = −3.3%). This is consistent with experimental results showing a higher probability of positive sign switching for taxa affected by incomplete recording during the time series. Whilst there remains the possibility that fish populations within the Anglian basin are undergoing different population trends to those elsewhere, overall these findings highlight the need for robust BTA to ensure that conservation efforts are not misdirected.

Many biological recording schemes were conceived of as a way to indicate the state of the abiotic environment (Kelly and Whitton 1995, Hawkes 1998) or to track the status of populations with commercial or recreational value (Radinger et al. 2019). However, increasing awareness of biodiversity loss has shifted the focus of biological recording towards the consideration of a wider range of species than the original schemes were designed to detect. As species previously considered out of scope are increasingly recorded, such as the six small-bodied fish species modelled here, there is stronger potential for uncertainty stemming from taxonomic incompleteness. Such uncertainties may be reduced through improved surveyor training and awareness, but these attempts may themselves introduce discontinuities that increase risk of bias in future BTA studies. Indeed, efforts to improve taxonomic identification skills among operators have been ongoing throughout the study period, which has also been a time of unprecedented upheaval in diatom systematics in particular (Spaulding et al. 2021). Whilst unconfirmed, these factors may mean that the

extreme positive trend reported for the diatom genus *Adlafia* is an artefact. Before a training exercise for diatom taxonomists in 2013, the most commonly encountered species within this genus (*A. suchlandtii*) may frequently have been misidentified as *Achnanthidium*, potentially driving a dramatic increase in the number of records for *Adlafia* later in the study period. This is a potential example of the discontinuities we refer to. It is important that work to improve the taxonomic accuracy of monitoring, and more generally to increase species detection rates, is well documented so that future BTA studies may account for, or avoid, such issues.

Although we have attempted to minimise the impact of our own idiosyncratic decisions, the design of our analytic pipeline may have contributed uncertainty in addition to those sources we have explicitly evaluated. These include our decisions to remove fish hybrids, filter invertebrate and diatom taxa using total detection and total count thresholds, and indeed our choice of study period, response variables (population versus community level; count versus binary), and model type (i.e. GLMM versus alternatives, linear versus non-linear models). Furthermore, BTA is associated with a range of other challenges in drawing robust inference which we have not directly addressed here. These include the false baseline, missing zero, detection bias, and other effects (Didham et al. 2020).

We do not follow previous studies which have employed Bayesian techniques to explicitly address the concerns of sampling bias and uncertainty propagation (Cressie et al. 2009, Yen et al. 2019) as our primary aim has been to model alternative decisions made in previous trend assessments, particularly those which used the same dataset as our study (Pharaoh et al. 2023, Powell et al. 2023, Qu et al. 2023). Whilst we do account for temporal and spatial structures in the data via random slopes and intercepts on site, as well as random intercepts on year (Daskalova et al. 2021), we do not account for phylogenetic structure which Johnson et al. (2024) recently showed was critical to producing valid inference when modelling trends in total abundance across a whole dataset. Further research should explore if such correlative effect models (Johnson et al. 2024) could be adapted to improve other BTA applications such as those we implement here (i.e. distribution or abundance trends for individual taxa).

Historical sampling biases may affect the modelled dataset and other similar datasets worldwide in ways which we have not fully captured in our analysis of environmental collection bias (McClure et al. 2023). Many observation networks were established in the 1970s and 1980s to monitor sites exposed to perceived threats at the time, yet these sites may have improved by the early 21st century whilst the nature and spatial distribution of the threats changed. For example, it is known that threats to freshwater biodiversity in Britain shifted in the late 20th century, with improvements in sanitary water quality in urban areas coinciding with increases in concentrations of sewage-associated emerging contaminants and the deterioration of water quality in rural areas (Whelan et al. 2022).

River biodiversity trends in England, 2002–2019

Underpinned by rigorous evaluation of data filtering, model specification, and data aggregation uncertainty, as well as environmental collection bias and the effects of taxonomic incompleteness, our assessment of river biodiversity in England represents the most comprehensive picture yet of 21st century trends in the region. Despite this, we urge caution in their interpretation given our experimental findings which indicate that realistic monitoring network designs are often inadequate for reliable trend estimation, even in the absence of bias, data filtering or taxonomic incompleteness (Fig. 11).

Together, our findings on fish population trends suggest mixed success for fisheries conservation efforts in the face of environmental change. Among migratory species requiring access to both marine and freshwater habitats, we report abundance declines for Atlantic salmon *Salmo salar* and European eel *Anguilla anguilla* but strong increases for lampreys (Petromyzontidae; *Lampetra fluviatilis*, *L. planeri*, *Petromyzon marinus*). Stable or positive trends for other, strictly freshwater protected species (*C. taenia*, *C. gobio*, *T. thymallus*) and declines for three of five invasive fish species again suggest some success for species conservation and invasive species management. Declines or near-zero trends among several coarse fish species commonly targeted by recreational anglers, notably chub *Leuciscus leuciscus*, dace *Leuciscus cephalus*, bream *Abramis brama*, roach *Rutilus rutilus* and barbel *Barbus barbus*, are surprising given substantial long-term investment in coarse fish stocking programmes in England (Environment Agency 2021). It must be noted, however, that our reports of linear trends in the period 2002–2019 should be interpreted in the context of both longer-term and shorter-term changes in regional freshwater fish populations (Nunn et al. 2023). For example, despite modelling a strong increase in the abundance of lampreys over the study period, these species remain considerably less abundant in Britain than they were prior to the mid-twentieth century (Maitland et al. 2015). On shorter timescales, populations of both Atlantic salmon (63% reduction over three generations; ICES 2021) and European eel (> 80% reduction; ICES 2022) in Britain are known to have undergone steeper abundance declines than reported here for 2002–2019.

Amid reports of widespread and catastrophic insect declines globally (Wagner et al. 2021), our results are consistent with a growing body of evidence showing that the distributions and abundances of many freshwater insect populations in temperate zones have increased in recent decades (Outhwaite et al. 2020, van Klink et al. 2020a). In common with several previous studies from the same region (Vaughan and Ormerod 2012, Outhwaite et al. 2020, Haase et al. 2023, Pharaoh et al. 2023, Powell et al. 2023, Qu et al. 2023), we report increases within insect groups traditionally associated with better water quality, including the 'EPT taxa' of mayflies (Ephemeroptera), stoneflies (Plecoptera), and caddisflies (Trichoptera). However, these positive trends contrasted with frequent declines among other invertebrate groups, including

16000587, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1111/ecog.07441 by Test, Wiley Online Library on [21/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

molluscs and annelids typically associated with poorer water quality, but also insect taxa considered of conservation value – most notably dragonflies (Odonata) and beetles (Coleoptera). Along with the strong increases reported for several invasive non-native invertebrate populations, conservation managers may find these results concerning, despite recent reports of increases in 'overall family richness' (Qu et al. 2023; p.11) within invertebrate communities in the region.

It is important to place our results within the context of longer-term changes in invertebrate biodiversity. There is a growing consensus that the recovery of European freshwater invertebrate communities ascribed to water quality improvements in the 1990s (Vaughan and Ormerod 2012) has slowed or even come to a halt since the early twentieth century (Outhwaite et al. 2020, Haase et al. 2023, Pharaoh et al. 2023, Qu et al. 2023). Our findings on invertebrate trends between 2002 and 2019 are consistent with this emerging consensus but we also highlight substantial heterogeneity in the trends of individual taxa, with continuing increases among EPT taxa reported alongside decreases within several other groups. In the face of multiple stressors, including climate change and shifting water quality threats, the presence and abundance of invertebrate families classically associated with good water quality may no longer be adequate indicators of freshwater ecological status (Simmons et al. 2021, Whelan et al. 2022).

With over 30 years of European legislation specifically targeting phosphorus inputs to rivers, it would be reasonable to expect diatom genera associated with low phosphorus concentrations (e.g. *Achnanthidium*, *Fragilaria*, *Hannaea*) to have expanded their distributions. Yet we did not observe consistent positive trends among these taxa. As a result of the Urban Waste Water Treatment Directive and, latterly, the Water Framework Directive, the UK has seen a substantial fall in the total phosphorus flux from 120 to 16 kt a^{-1} between 1990 and 2012 (Worrall et al. 2016) and has stringent boundaries for river phosphorus concentrations (Kelly et al. 2022). However, reduction efforts have largely focused on point sources, whereas diffuse sources are significant (Charlton et al. 2018, Bell et al. 2021). Thus, phosphorous concentrations remain above what is considered necessary to support good ecological status in 55% of rivers in England (GOV.UK 2023). Some of the more striking shifts in diatom assemblages have been observed in upland areas recovering from acidification (Battarbee et al. 2014) but these changes would affect regions underrepresented in the present study (Fig. 7).

Among macrophytes, we report largely negative trends mirroring those seen in the flora of Britain's aquatic habitats more generally (Stroh et al. 2023). Indeed, some of the genera we report as showing the strongest negative trends in English rivers (*Littorella* and *Groenlandia*) are known to have experienced severe declines in southern Britain that extend back several decades (Preston and Croft 1997, Stroh et al. 2023). Key macrophyte trends are consistent with the increasingly dynamic nature of river flows in recent decades, with increasing winter or annual flow maxima (Hannaford et al. 2021), often coupled with fine sediment loading in lowland agricultural catchments, interspersed with several prolonged droughts during the study period. These trends include increases in amphibious ruderal genera of river margins (e.g. *Alisma*, *Bidens*, *Veronica*, *Eleocharis*) and medium-large emergent plants (e.g. *Typha*, *Glyceria*, *Equisetum*, *Carex*), alongside declines of many desiccation-sensitive bryophytes. In general, free-floating or weakly anchored genera of low energy habitats that are sensitive to scouring and washout have also declined (*Ceratophyllum*, *Elodea*, *Lemna*, *Spirodela*), whilst those that include some widespread fluviatile species (*Ranunculus*, *Potamogeton*) have increased. As with the diatoms, there is no clear evidence of net declines among eutrophication-sensitive macrophyte genera, or, indeed, of their recovery, as might be expected in the wake of nutrient management and reduced phosphorus loading to rivers (Worrall et al. 2016). Analysis at the species level would likely be needed to detect such trends. Among bryophytes there is additional evidence of declines in more acid-tolerant genera (e.g. *Marsupella*, *Nardia*, *Scapania*, *Hyocomium*, *Racomitrium*), though whether this reflects recovery from historic acidification is unclear as several bryophyte genera associated with circum-neutral conditions have also declined.

Supporting reproducibility in biodiversity trend assessments

To support the reproducibility of BTA, we make several recommendations and present a set of functions for exploratory analysis and data processing via a GitHub repository (Table 1, https://github.com/wilkesma/rbta).

Table 1. Summary of functions available in the GitHub repository accompanying this study.

Function name	Description
filterSites()	Apply a sampling frequency filter to a dataset and optionally plot the empirical cumulative distribution function over multiple sampling frequency thresholds
collectionBias()	Assess the environmental collection bias associated with a dataset
exploreTaxa()	Given a taxonomic classification, plot the number of detections and/or total abundance recorded at successive taxonomic ranks
aggregateTaxa()	Given a taxonomic classification and a target rank, for each taxon at the target rank, aggregate the data to the target rank and optionally remove samples in which each higher rank taxon was recorded but not the target rank
prepareData()	Combines the above functions into a single function prompting user inputs to decide sampling frequency filter and target rank informed by outputted graphics, including optional plotting of environmental collection bias at specified sampling frequency thresholds

Based on our finding that national level models often severely overestimated trends for those taxa which were not consistently recorded, where detected (Fig. 9, 11b), we recommend that those undertaking BTA consult with data owners to build a narrative of how the data-generating process changed over time to address taxonomic incompleteness. The observational data we modelled were associated with severe collection bias (Fig. 7), and this may be the case for many other biodiversity datasets. BTA studies should therefore quantify the environmental collection bias and not assume that modelled trends are representative of the wider landscape. Consultation with data owners will also be useful here to better understand how historical sampling biases may affect findings. Filtering on the basis of site-level sampling frequency exacerbated collection bias (Fig. 7g), generally reduced the magnitude of trend estimates (Supporting information), and significantly increased the risk of sign switching (Fig. 11a, Supporting information). Thus, we recommend not to filter input data based on sampling frequency unless there is a strong justification for doing so.

Trends estimated at higher ranks (e.g. family, genus) were unrepresentative of individual native species (Fig. 8), leading to our recommendation that trends should be modelled at the finest practicable taxonomic resolution. Furthermore, caution should be exercised when interpreting genus or family level trends as 'biodiversity trends', particularly when native and non-native species are nested within the genus or family. Our findings that nAGQ = 0 was the dominant source of model specification-related uncertainty in the observational analysis (Fig. 6), and was associated with a significantly greater absolute error than all but one nonlinear optimizer in the in silico experiments (Supporting information), suggests that the use of nAGQ = 0 should be avoided wherever possible. In the specific context of modelling regional taxon-level distribution and abundance trends, our recommendations are supported by previous studies demonstrating the need for inclusion of temporal random effects (Daskalova et al. 2021), addition of observation level random effects to control for non-zero-inflated overdispersion (Harrison 2014), and application of dispersion and zero-inflation tests to inform model specification (Hartig 2022).

Conclusions

To inform effective biodiversity policy, conservation and restoration decisions, it is vitally important that we learn how to detect and manage uncertainty when modelling biodiversity trends, thus avoiding misdirection of public funding and private finance for conservation. In the context of estimating population level trends using GLMMs, we show that data filtering, model specification, collection bias, data aggregation and taxonomic incompleteness are all serious sources of uncertainty. Robust biodiversity trend analysis depends on investigating and appropriately addressing these concerns, yet many studies reporting biodiversity trends fail to include an assessment of the uncertainty introduced by idiosyncratic decisions made during data preparation and modelling stages (Boyd et al. 2022). We reveal how these decisions lead to 'hidden' sources of uncertainty in biodiversity trend analysis and provide tools for investigating their contribution to uncertainty in future studies.

Author contributions

Martin A. Wilkes: Conceptualization (lead); Data curation (lead); Formal analysis (equal); Funding acquisition (equal); Investigation (lead); Methodology (lead); Project administration (lead); Resources (equal); Software (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing - original draft (equal); Writing - review and editing (equal). **Morwenna Mckenzie**: Conceptualization (supporting); Data curation (supporting); Methodology (supporting). **Andrew Johnson**: Conceptualization (supporting); Data curation (supporting); Methodology (supporting); Writing - review and editing (equal). **Christopher Hassall**: Conceptualization (equal); Formal analysis (supporting); Methodology (equal); Visualization (equal); Writing - review and editing (equal); **Martyn Kelly**: Conceptualization (supporting); Data curation (equal); Methodology (supporting); Writing - review and editing (equal). **Nigel Willby**: Data curation (equal); Methodology (supporting); Writing - review and editing (equal). **Lee E. Brown**: Conceptualization (equal); Formal analysis (equal); Funding acquisition (equal); Methodology (equal); Writing - review and editing (equal).

Transparent peer review

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/ecog.07441.

Data availability statement

Data are available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.np5hqc034 (Wilkes et al. 2025).

Supporting information

The Supporting information associated with this article is available with the online version.

References

Bates, D., Mächler, M., Bolker, B. M. and Walker, S. C. 2015. Fitting linear mixed-effects models using lme4. – J. Stat. Softw. 67: 1–48.

Battarbee, R. W., Simpson, G. L., Shilland, E. M., Flower, R. J., Kreiser, A., Yang, H. and Clarke, G. 2014. Recovery of UK lakes from acidification: an assessment using combined palaeo-ecological and contemporary diatom assemblage data. – Ecol. Indic. 37: 365–380.

Bell, V. A., Naden, P. S., Tipping, E., Davies, H. N., Carnell, E., Davies, J. A. C., Dore, A. J., Dragosits, U., Lapworth, D. J., Muhammed, S. E., Quinton, J. N., Stuart, M., Tomlinson, S., Wang, L., Whitmore, A. P. and Wu, L. 2021. Long term simulations of macronutrients (C, N and P) in UK freshwaters. – Sci. Total Environ. 776: 145813.

Boënnec, M., Dakos, V. and Devictor, V. 2024. Sources of confusion in global biodiversity trends. – Oikos 2024: e10732.

Boyd, R. J., Powney, G. D., Burns, F., Danet, A., Duchenne, F., Grainger, M. J., Jarvis, S. G., Martin, G., Nilsen, E. B., Porcher, E., Stewart, G. B., Wilson, O. J. and Pescott, O. L. 2022. ROBITT: A tool for assessing the risk-of-bias in studies of temporal trends in ecology. – Methods Ecol. Evol. 13: 1497–1507.

Breznau, N. et al. 2022. Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty. – Proc. Natl Acad. Sci. USA 119: e2203150119.

Canadian Aquatic Biomonitoring Network 2012. Canadian aquatic biomonitoring network, field manual – wadeable streams. – Environment Canada, https://publications.gc.ca/collections/collection_2012/ec/En84-87-2012-eng.pdf.

Carle, F. L. and Strub, M. R. 1978. A new method for estimating population size from removal data. – Biometrics 34: 621.

Carraro, L., Bertuzzo, E., Fronhofer, E. A., Furrer, R., Gounand, I., Rinaldo, A. and Altermatt, F. 2020. Generation and application of river network analogues for use in ecology and evolution. – Ecol. Evol. 10: 7537–7550.

Chamberlain, S. A. and Szöcs, E. 2013. taxize: taxonomic search and retrieval in R. – F1000Research 2: 191.

Charlton, M. B., Bowes, M. J., Hutchins, M. G., Orr, H. G., Soley, R. and Davison, P. 2018. Mapping eutrophication risk from climate change: future phosphorus concentrations in English rivers. – Sci. Total Environ. 613–614: 1510–1526.

Comte, L. et al. 2021. RivFishTIME: a global database of fish time-series to study global change ecology in riverine systems. – Global Ecol. Biogeogr. 30: 38–50.

Cressie, N., Calder, C. A., Clark, J. S., Ver Hoef, J. M. and Wikle, C. K. 2009. Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. – Ecol. Appl. 19: 553–570.

Crouzeilles, R., Curran, M., Ferreira, M. S., Lindenmayer, D. B., Grelle, C. E. V. and Rey Benayas, J. M. 2016. A global meta-analysis on the ecological drivers of forest restoration success. – Nat. Commun. 71: 11666.

Daskalova, G. N., Phillimore, A. B. and Myers-Smith, I. H. 2021. Accounting for year effects and sampling error in temporal analyses of invertebrate population and biodiversity change: a comment on Seibold et al. 2019. – Insect Conserv. Divers. 14: 149–154.

Department of Water and Sanitation 2016. REMP (River Eco-status Monitoring Programme). – https://www.dws.gov.za/iwqs/rhp/default.aspx

Desquilbet, M., Gaume, L., Grippa, M., Céréghino, R., Humbert, J. F., Bonmatin, J. M., Cornillon, P. A., Maes, D., Dyck, H. V. and Goulson, D. 2020. Comment on "Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances." – Science 370: eabd8947.

Didham, R. K., Basset, Y., Collins, C. M., Leather, S. R., Littlewood, N. A., Menz, M. H. M., Müller, J., Packer, L., Saunders, M. E., Schönrogge, K., Stewart, A. J. A., Yanoviak, S. P. and Hassall, C. 2020. Interpreting insect declines: seven challenges and a way forward. – Insect Conserv. Divers. 13: 103–114.

Dornelas, M., Gotelli, N. J., McGill, B., Shimadzu, H., Moyes, F., Sievers, C. and Magurran, A. E. 2014. Assemblage time series reveal biodiversity change but not systematic loss. – Science 344: 296–299.

Dornelas, M. et al. 2018. BioTIME: a database of biodiversity time series for the Anthropocene. – Global Ecol. Biogeogr. 27: 760–786.

Doser, J. W., Finley, A. O., Kéry, M. and Zipkin, E. F. 2022. spOccupancy: an R package for single-species, multi-species, and integrated spatial occupancy models. – Methods Ecol. Evol. 13: 1670–1678.

Dunkley, K., Cable, J. and Perkins, S. E. 2020. Consistency in mutualism relies on local, rather than wider community biodiversity. – Sci. Rep. 10: 21255.

Environment Agency 2021. Fisheries annual report, 2019 to 2020 – GOV.UK, https://www.gov.uk/government/publications/fisheries-annual-report-2019-to-2020/fisheries-annual-report-2019-to-2020.

Environment Agency 2023. Ecology & fish data explorer. – https://environment.data.gov.uk/ecology/explorer/.

Environment Protection Authority Victoria 2021. 604.2: guideline for environmental management (GEM) – Rapid bioassessment methodology for rivers and streams. – https://www.epa.vic.gov.au/about-epa/publications/604-2.

Environmental Protection Agency Ireland 2023. Ireland's National Water Quality Monitoring Programme. – https://www.epa.ie/publications/monitoring--assessment/freshwater--marine/irelands-national-water-quality-monitoring-programme-20222027.php.

Feng, M. E. and Che-Castaldo, J. 2021. Comparing the reliability of relative bird abundance indices from standardized surveys and community science data at finer resolutions. – PLoS One 16: e0257226.

Finn, C., Grattarola, F. and Pincheira-Donoso, D. 2023. More losers than winners: investigating Anthropocene defaunation through the diversity of population trends. – Biol. Rev. 98: 1732–1748.

Fox, J. and Weisberg, S. 2019. An R companion to applied regression. – Sage Publications.

Friberg, N. 2014. Impacts and indicators of change in lotic ecosystems. – WIREs Water 1: 513–531.

Gould, E. et al. 2023. Same data, different analysts: variation in effect sizes due to analytical decisions in ecology and evolutionary biology. – EcoEvoRxiv, https://doi.org/10.32942/x2gg62.

GOV.UK 2023. State of the water environment indicator B3: supporting evidence. – https://www.gov.uk/government/publications/state-of-the-water-environment-indicator-b3-supporting-evidence/state-of-the-water-environment-indicator-b3-supporting-evidence.

Guillera-Arroita, G. 2017. Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. – Ecography 40: 281–295.

Haase, P. et al. 2023. The recovery of European freshwater biodiversity has come to a halt. – Nature 620: 582–588.

Hannaford, J., Mastrantonas, N., Vesuviano, G. and Turner, S. 2021. An updated national-scale assessment of trends in UK

peak river flow data: how robust are observed increases in flooding? – Hydrol. Res. 52: 699–718.

Harrison, X. A. 2014. Using observation-level random effects to model overdispersion in count data in ecology and evolution. – Peer J. 2: e616.

Hartig, F. 2022. Dharma: residual diagnostics for hierarchical (multi-level / mixed) regression models. – https://CRAN.R-project.org/package=DHARMa.

Hawkes, H. 1998. Origin and development of the biological monitoring working party score system. – Water Res. 32: 964–968.

Hughes, A. C., Orr, M. C., Ma, K., Costello, M. J., Waller, J., Provoost, P., Yang, Q., Zhu, C. and Qiao, H. 2021. Sampling biases shape our view of the natural world. – Ecography 44: 1259–1269.

ICES 2021. Working group on North Atlantic salmon. – https://hal.inrae.fr/hal-03228611v1.

ICES 2022. European eel (*Anguilla anguilla*) throughout its natural range. – https://doi.org/10.17895/ices.data.10268.

Jähnig, S. C. et al. 2021. Revisiting global trends in freshwater insect biodiversity. – WIREs Water 8: e1506.

Johnston, A., Hochachka, W. M., Strimas-Mackey, M. E., Ruiz Gutierrez, V., Robinson, O. J., Miller, E. T., Auer, T., Kelling, S. T. and Fink, D. 2021. Analytical guidelines to increase the value of community science data: an example using eBird data to estimate species distributions. – Divers. Distrib. 27: 1265–1277.

Johnson, T. F., Beckerman, A. P., Childs, D. Z., Webb, T. J., Evans, K. L., Griffiths, C. A., Capdevila, P., Clements, C. F., Besson, M., Gregory, R. D., Thomas, G. H., Delmas, E. and Freckleton, R. P. 2024. Revealing uncertainty in the status of biodiversity change. – Nature 628: 788–794.

Kelly, M. G. and Whitton, B. A. 1995. The trophic diatom Index: a new index for monitoring eutrophication in rivers. – J. Appl. Phycol. 7: 433–444.

Kelly, M. G., Phillips, G., Teixeira, H., Várbíró, G., Salas Herrero, F., Willby, N. J. and Poikane, S. 2022. Establishing ecologically-relevant nutrient thresholds: a tool-kit with guidance on its use. – Sci. Total Environ. 807: 150977.

Kuznetsova, A., Brockhoff, P. B. and Christensen, R. H. B. 2017. lmerTest package: tests in linear mixed effects models. – J. Stat. Softw. 82: 1–26.

Lenth, R. V. 2023. emmeans: estimated marginal means, aka least-squares means. – https://CRAN.R-project.org/package=emmeans.

Leopold, L. and Maddock, T. 1953. The hydraulic geometry of stream channels and some physiographic implications. – US Government Printing Office, https://pubs.usgs.gov/pp/0252/report.pdf.

Mackenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Andrew Royle, J. A. and Langtimm, C. A. 2002. Estimating site occupancy rates when detection probabilities are less than one. – Ecology 83: 2248–2255.

Maitland, P. S., Renaud, C. B., Quintella, B. R., Close, D. A. and Docker, M. F. 2015. Conservation of native lampreys. – Lampreys Biol. Conserv. Control. 1: 375–428.

McClure, C. J. W., Rolek, B. W., Christopher, C. and Mcclure, J. W. 2023. Pitfalls arising from site selection bias in population monitoring defy simple heuristics. – Methods Ecol. Evol. 14: 1489–1499.

McCrea, R., King, R., Graham, L. and Börger, L. 2023. Realising the promise of large data and complex models. – Methods Ecol. Evol. 14: 4–11.

Nash, J. C. 2014. On best practice optimization methods in R. – J. Stat. Softw. 60: 1–14.

Nash, J. C. and Varadhan, R. 2011. Unifying optimization algorithms to aid software system users: optimx for R. – J. Stat. Softw. 43: 1–14.

Nunn, A. D., Ainsworth, R. F., Walton, S., Bean, C. W., Hatton-Ellis, T. W., Brown, A., Evans, R., Atterborne, A., Ottewell, D. and Noble, R. A. A. 2023. Extinction risks and threats facing the freshwater fishes of Britain. – Aquat. Conserv. Mar. Freshw. Ecosyst. 33: 1460–1476.

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. – Nat. Ecol. Evol. 4: 384–392.

Pharaoh, E., Diamond, M., Ormerod, S. J., Rutt, G. and Vaughan, I. P. 2023. Evidence of biological recovery from gross pollution in English and Welsh rivers over three decades. – Sci. Total Environ. 878: 163107.

Powell, K. E., Oliver, T. H., Johns, T., González-Suárez, M., England, J. and Roy, D. B. 2023. Abundance trends for river macroinvertebrates vary across taxa, trophic group and river typology. – Global Change Biol. 29: 1282–1295.

Preston, C. D. and Croft, J. M. 1997. Aquatic plants in Britain and Ireland. – Harley Books.

Qu, Y., Keller, V., Bachiller-Jareno, N., Eastman, M., Edwards, F., Jürgens, M. D., Sumpter, J. P. and Johnson, A. C. 2023. Significant improvement in freshwater invertebrate biodiversity in all types of English rivers over the past 30 years. – Sci. Total Environ. 905: 167144.

Radinger, J., Britton, J. R., Carlson, S. M., Magurran, A. E., Alcaraz-Hernández, J. D., Almodóvar, A., Benejam, L., Fernández-Delgado, C., Nicola, G. G., Oliva-Paterna, F. J., Torralva, M. and García-Berthou, E. 2019. Effective monitoring of freshwater fish. – Fish Fish. 20: 729–747.

Satterthwaite, F. E. 1946. An approximate distribution of estimates of variance components. – Biometrics 2: 110.

Saunders, M. E., Janes, J. K. and O'Hanlon, J. C. 2020. Moving on from the insect apocalypse narrative: engaging with evidence-based insect conservation. – BioScience 70: 80–89.

Seibold, S., Gossner, M. M., Simons, N. K., Blüthgen, N., Müller, J., Ambarlı, D., Ammer, C., Bauhus, J., Fischer, M., Habel, J. C., Linsenmair, K. E., Nauss, T., Penone, C., Prati, D., Schall, P., Schulze, E. D., Vogt, J., Wöllauer, S. and Weisser, W. W. 2019. Arthropod decline in grasslands and forests is associated with landscape-level drivers. – Nature 574: 671–674.

Seibold, S., Hothorn, T., Gossner, M. M., Simons, N. K., Blüthgen, N., Müller, J., Ambarlı, D., Ammer, C., Bauhus, J., Fischer, M., Habel, J. C., Penone, C., Schall, P., Schulze, E. D. and Weisser, W. W. 2021. Insights from regional and short-term biodiversity monitoring datasets are valuable: a reply to Daskalova et al. 2021. – Insect Conserv. Divers. 14: 144–148.

SEPA 2023. River invertebrate classification tool. – https://www.sepa.org.uk/environment/water/aquatic-classification/river-invertebrate-classification-tool/.

Simmons, B. I., Blyth, P. S. A., Blanchard, J. L., Clegg, T., Delmas, E., Garnier, A., Griffiths, C. A., Jacob, U., Pennekamp, F., Petchey, O. L., Poisot, T., Webb, T. J. and Beckerman, A. P. 2021. Refocusing multiple stressor research around the targets and scales of ecological impacts. – Nat. Ecol. Evol. 5: 1478–1489.

Spaulding, S. A., Potapova, M. G., Bishop, I. W., Lee, S. S., Gasperak, T. S., Jovanoska, E., Furey, P. C. and Edlund, M. B. 2021.

Diatoms.org: supporting taxonomists, connecting communities. – Diatom Res. 36: 291–304.

Stroh, P. A., Walker, K. J., Humphrey, T. A., Pescott, O. L. and Burkmar, R. J. 2023. Plant atlas 2020 : mapping changes in the distribution of the British and Irish Flora. – Princeton Univ. Press.

Thompson, P. L., Guzman, L. M., De Meester, L., Horváth, Z., Ptacnik, R., Vanschoenwinkel, B., Viana, D. S. and Chase, J. M. 2020. A process-based metacommunity framework linking local and regional scale community ecology. – Ecol. Lett. 23: 1314–1329.

Tickner, D. et al. 2020. Bending the curve of global freshwater biodiversity loss: an emergency recovery plan. – BioScience 70: 330–342.

UKTAG 2014. UKTAG Lake Assessment MethodMacrophytes and Phytobenthos. – https://www.wfduk.org/sites/default/files/Media/Characterisation%20of%20the%20water%20environment/Biological%20Method%20Statements/Lake%20Macrophytes%20UKTAG%20Method%20Statement.pdf.

UKCEH 2023. UKCEH land cover maps. – https://www.ceh.ac.uk/data/ukceh-land-cover-maps.

van Klink, R., Bowler, D. E., Gongalsky, K. B., Swengel, A. B., Gentile, A. and Chase, J. M. 2020a. Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances. – Science 368: 417–420.

van Klink, R., Bowler, D. E., Gongalsky, K. B., Swengel, A. B. and Chase, J. M. 2020b. Response to comment on "Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances." – Science 370: p.eabe0760.

Vaughan, I. P. and Ormerod, S. J. 2012. Large-scale, long-term trends in British river macroinvertebrates. – Global Change Biol. 18: 2184–2194.

Wagner, D. L., Grames, E. M., Forister, M. L., Berenbaum, M. R. and Stopak, D. 2021. Insect decline in the Anthropocene: death by a thousand cuts. – Proc. Natl Acad. Sci. USA 118: e2023989118.

Wauchope, H. S., Amano, T., Sutherland, W. J. and Johnston, A. 2019. When can we trust population trends? A method for quantifying the effects of sampling interval and duration. – Methods Ecol. Evol. 10: 2067–2078.

Wenger, S. J., Stowe, E. S., Gido, K. B., Freeman, M. C., Kanno, Y., Franssen, N. R., Olden, J. D., Poff, N. L. R., Walters, A. W., Bumpers, P. M., Mims, M. C., Hooten, M. B. and Lu, X. 2022. Simple statistical models can be sufficient for testing hypotheses with population time-series data. – Ecol. Evol. 12: e9339.

Wetzel, F. T., Bingham, H. C., Groom, Q., Haase, P., Kõljalg, U., Kuhlmann, M., Martin, C. S., Penev, L., Robertson, T., Saarenmaa, H., Schmeller, D. S., Stoll, S., Tonkin, J. D. and Häuser, C. L. 2018. Unlocking biodiversity data: prioritization and filling the gaps in biodiversity observation data in Europe. – Biol. Conserv. 221: 78–85.

Whelan, M. J., Linstead, C., Worrall, F., Ormerod, S. J., Durance, I., Johnson, A. C., Johnson, D., Owen, M., Wiik, E., Howden, N. J. K., Burt, T. P., Boxall, A., Brown, C. D., Oliver, D. M. and Tickner, D. 2022. Is water quality in British rivers 'better than at any time since the end of the Industrial Revolution'? – Sci. Total Environ. 843: 157014.

Wilkes, M. A., Mungee, M., Naura, M., Bell, V. A. and Brown, L. E. 2024. Predicting nature recovery for river restoration planning and ecological assessment: a case study from England, 1991–2042. – River Res. Appl. 41: 68-81.

Wilkes, M. A., Mckenzie, M., Johnson, A., Hassall, C., Kelly, M., Willby, N. and Brown, L. E. 2025. Data from: Revealing hidden sources of uncertainty in biodiversity trend assessments. – Dryad Digital Repository, https://doi.org/10.5061/dryad.np5hqc034.

WorldPop 2023. Population counts. – https://hub.worldpop.org/project/categories?id=18.

Worrall, F., Jarvie, H. P., Howden, N. J. K. and Burt, T. P. 2016. The fluvial flux of total reactive and total phosphorus from the UK in the context of a national phosphorus budget: comparing UK river fluxes with phosphorus trade imports and exports. – Biogeochemistry 130: 31–51.

Yen, J. D. L., Tonkin, Z., Lyon, J., Koster, W., Kitchingman, A., Stamation, K. and Vesk, P. A. 2019. Integrating multiple data types to connect ecological theory and data among levels. – Front. Ecol. Evol. 7: 428303.

Zurell, D., Berger, U., Cabral, J. S., Jeltsch, F., Meynard, C. N., Münkemüller, T., Nehrbass, N., Pagel, J., Reineking, B., Schröder, B. and Grimm, V. 2010. The virtual ecologist approach: simulating data and observers. – Oikos 119: 622–635.