River Research and Applications

WILEY

RESEARCH ARTICLE OPEN ACCESS

A Kick in the Headwaters: Evaluating a Macroinvertebrate Sampling Method for Ecological Condition Monitoring in Small Streams

Rachel Stubbington[1] | Oliver Longstaffe[1] | Romain Sarremejane[1] | Phillipa Bates[1] | Kieran J. Gething[1] | J. Iwan Jones[2] | Mary Kelly-Quinn[3] | Alex Laini[4] | John Murray-Bligh[5] | Lesley Rippon[5] | Simon Rouen[1]

[1]School of Science and Technology, Nottingham Trent University, Nottingham, UK | [2]School of Biological and Behavioural Sciences, Queen Mary University of London, London, UK | [3]School of Biology and Environmental Science, UCD Earth Institute, University College Dublin, Dublin, Ireland | [4]Department of Life Sciences and Systems Biology, University of Turin, Torino, Italy | [5]Environment Agency, Bristol, UK

Correspondence: Rachel Stubbington (rachel.stubbington@ntu.ac.uk)

Received: 14 July 2024 | **Revised:** 13 November 2024 | **Accepted:** 14 November 2024

Funding: This research was funded by the Department for Environment, Food and Rural Affairs Natural Capital and Ecosystem Assessment Programme via the Environment Agency Small Streams Network Methods Development Project. Nottingham Trent University funded BSc research studentships.

Keywords: bioassessment | ecological health | ecological quality | ecological status | headwater | invertebrate | river health | sampling effort

ABSTRACT

Small streams dominate river networks and collectively support high biodiversity, but are rarely included in regulatory biomonitoring programmes. Macroinvertebrate communities are effective biomonitors of ecological condition and are routinely collected using 3-min 'kick' samples. However, this 3-min duration may not be suitable for small streams, which typically support fewer taxa at lower densities than larger rivers of equivalent condition. We evaluated the kick-sampling method at 30 sites representing a national small stream monitoring network. At each site, we collected three 5-min kick samples in 10 0.5-min component parts. We used the families collected in 15 min to represent 'total' site-scale taxonomic richness, then determined the duration needed to sample ≥ 65% of these taxa (a method and target comparable to those used in larger rivers). We also determined the sampling duration at which an average score per taxon (ASPT) biomonitoring index stabilized. Considering all streams, on average, 2.5-min durations captured ≥ 65% of taxa, but 3.5 min was required to reach this target in temporary streams, because numerous taxa occurred at low abundance. Only 54% of samples contained ≥ 65% of taxa after 2.5 min, compared to 70% after 3 min. In most streams, the ASPT stabilized after 2 min, whereas 3 min was required to meet this target in temporary streams. Considering the variation around any estimate of capture rates introduced by natural variability, taxonomic resolution and operator error, we suggest 3 min as the most robust sampling duration to enable condition monitoring in individual small streams and comparison with larger rivers.

1 | Introduction

Small streams are variously defined based on stream order (Kelly-Quinn et al. 2024; Minshall et al. 1983), distance from the source (Furse and Symes 1997), stream size (Biggs, von Fumetti, and Kelly-Quinn 2017) and upstream catchment area (European Commission 2000), and are commonly considered to be first-order to third-order streams. These small streams dominate the global river network length (Downing et al. 2012; Smith and Lyle 1979), collectively support high biodiversity (Callanan, Baars, and Kelly-Quinn 2014; Finn et al. 2011) including rare and specialist species (Aspin and House 2022; Kabir et al. 2024), and influence catchment-wide ecosystem functioning (Alexander et al. 2007; Datry et al. 2023). Small streams are closely linked to their riparian zones and wider catchments, are thus vulnerable to anthropogenic pressures (Biggs, von Fumetti, and Kelly-Quinn 2017; Riley et al. 2018), and can transfer the ecological impacts of local pressures to downstream waters (Alexander et al. 2007; Datry et al. 2023).

Despite their vulnerability to human pressures, small streams generally fall outside the scope of legislation such as the EU Water Framework Directive (WFD, which excludes streams with catchments $< 10\,km^2$; European Commission 2000). As a result, few are included in national networks designed to monitor river condition (considered herein as comparable to health, quality or status), and thus few long-term monitoring data document the condition of small streams (but see e.g., Baattrup-Pedersen et al. 2018; Dunbar et al. 2010). To improve understanding and effective management of small streams, ambitious new monitoring networks are needed (e.g., Kelly-Quinn et al. 2024; von Gönner et al. 2024). In England, the Environment Agency (the statutory regulator) has designed the Small Streams Network (SSN) to enable condition monitoring across sites representing England's small stream resource (Defra 2022).

Biomonitoring of SSN sites encompasses standard biotic groups including macroinvertebrates, which are ubiquitous, abundant and diverse in freshwater ecosystems, and have highly variable taxon-specific environmental tolerances, making them effective biomonitors (Gibbs, Cook, and Kulp 2023; Rosenberg and Resh 1993). Macroinvertebrate communities are routinely sampled from UK and other river ecosystems using the RIVPACS method of a 3-min 'kick' sample supplemented by a 1-min manual search (hereafter, kick sampling; Haase et al. 2004; Murray-Bligh and Griffiths 2022). Such samples capture a limited proportion of taxa; Furse et al. (1981) estimated a mean of 62% of the families present per 3-min sample. Such a proportion is sufficient to estimate ecological condition, including in WFD status assessments (Feeley et al. 2020; Majaneva et al. 2024). However, the proportion of the macroinvertebrate community captured by kick sampling has primarily been evaluated in mid-order rivers (i.e., Furse et al. 1981; but see e.g., Feeley et al. 2012).

Macroinvertebrate communities in small streams typically comprise fewer taxa than those in mid-order rivers of equivalent condition (Minshall, Petersen Jr, and Nimz 1985; Paller, Specht, and Dyer 2006), likely because many such streams are in isolated headwaters which experience more frequent disturbance by drying (Messager et al. 2021) as well as flooding (Scott et al. 2019), but have relatively slow post-disturbance recolonization rates (Clarke et al. 2008). As such, 3-min samples could capture a higher proportion of the taxa present in small streams, leading to overestimation of richness relative to larger, downstream reaches of equivalent condition. Equally, small streams including temporary headwaters (which sometimes dry out) may support communities in which many taxa occur at lower densities than in larger, perennial rivers, reducing capture rates and thus richness estimates (Arscott, Tockner, and Ward 2005; Aspin and House 2022). In either case, a 3-min kick-sampling duration could hamper robust estimation of condition in small streams and comparison with larger rivers.

We evaluated the performance of the kick-sampling method at sites representing spatial variability in habitat characteristics, and thus macroinvertebrate communities, in England's SSN. At each site, we collected three replicate 5-min kick samples in 10 0.5-min component parts, which enabled estimation of total site-scale taxa richness and thus quantification of how the number and percentage of captured taxa and associated biomonitoring metrics varied with sampling duration. Our aim was to inform the development of the kick-sampling method for use in small stream networks, and to aid interpretation of the macroinvertebrate assemblage data collected therein.

2 | Methods

2.1 | Site Selection

The SSN comprises 1280 sites across England, UK. Its upstream limit is defined by a catchment area of $0.4\,km^2$ and the downstream limit is defined, using an administrative criterion, as the upper limit of the Environment Agency's main river biomonitoring network. The SSN excludes artificial watercourses such as ditches and culverted stretches, as well as ephemeral (i.e., flashy, rainfed temporary) streams. We selected 26 SSN sites for which the Environment Agency had completed a preliminary survey and risk assessment. Small streams fed by the chalk aquifer are excluded from the SSN because they are included in England's main river biomonitoring network. Therefore, to evaluate the kick-sampling method across the national small stream resource including its chalk streams, we also selected three small chalk stream sites. Of the 29 sites, two were dry when visited. As such, we collected samples at 27 sites, including 24 SSN sites and 3 chalk stream sites (Table S1).

Sites spanned the English Midlands, northern England and East Anglia, but excluded southern England because of resource constraints (Figure 1). Collectively, sites represented the range of core abiotic characteristics (i.e., width, altitude and alkalinity) in England's small streams (Figures S1 and S2). Based on the Environment Agency data used to inform site selection, widths ranged from 0.3 to 6.2 m (mean ± standard deviation [SD], 1.4 ± 1.3 m), alkalinity from 19 to $282\,mg\,L^{-1}$ $CaCO_3$ ($162 \pm 77\,mg\,L^{-1}$) and altitudes from 4 to 347 m.a.s.l. (123 ± 101 m.a.s.l.; Table S1). Sites were exposed to a range of human pressure types and intensities, spanning least-disturbed upland sites to lowland sites impacted by agricultural land uses, but excluding urban streams.

A fourth core abiotic characteristic—flow permanence (i.e., perennial or temporary)—was unknown for SSN sites at the time of sampling. Six of 24 SSN sites were subsequently recorded as dry during at least one monthly Environment Agency site visit between April and September 2023 and were thus classified as temporary (Table S1). Of the remaining 18 SSN sites, 13 were wet (flowing or ponded) during all monthly visits and were classified as perennial, and no observations were made at five sites. Based on long-term flow permanence data (e.g., Sefton et al. 2019), one chalk stream site is perennial, one is near-perennial (i.e., drying during drought) and one is temporary (Table S1).

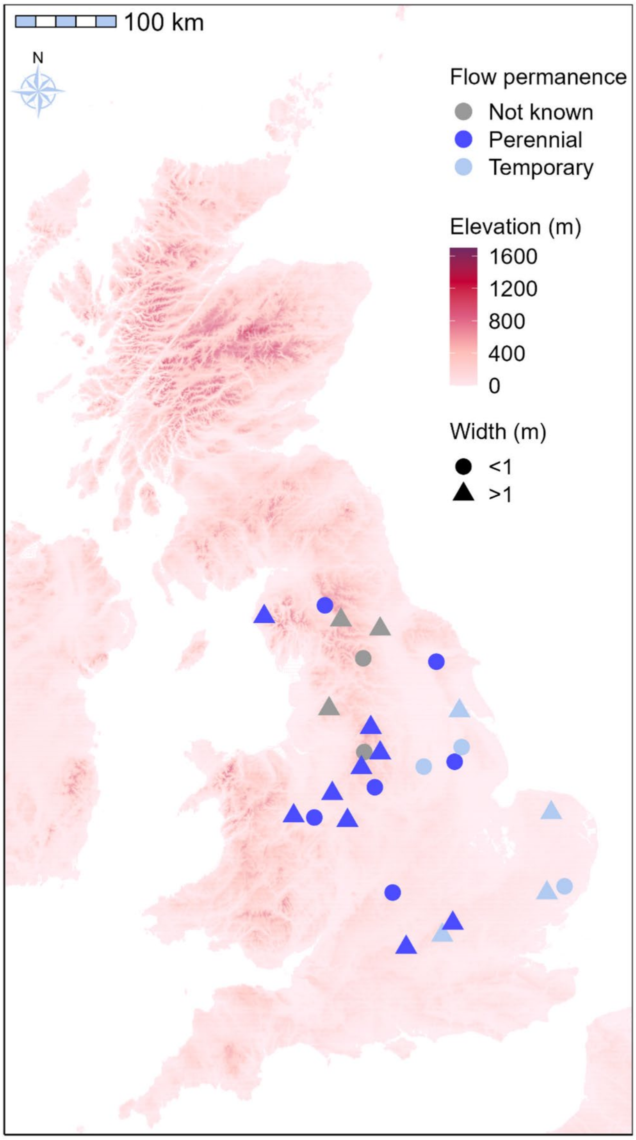

FIGURE 1 | Locations of the 27 macroinvertebrate sampling sites within England. [Color figure can be viewed at wileyonlinelibrary.com]

2.2 | Field Methods

We visited each site once in March–April 2023. This spring-season sampling campaign maximized the likelihood of characterizing stable communities present after long flowing phases at temporary sites. At each site, we collected kick samples following Murray-Bligh and Griffiths (2022), but with each of three replicate 5-min samples collected in 10 0.5-min component parts. The replicate samples characterized within-site variability in macroinvertebrate assemblages and their combined 15-min duration enabled estimation of each site's 'total' taxa richness. The 10 0.5-min parts generated sufficient data points to characterize taxa accumulation in relation to sampling duration. We sampled habitats in proportion to their occurrence for each sample, not each of its component parts. Two operators collected all samples, with each collecting all parts for two of three replicate samples at half the sites, to enable analysis of inter-operator variability (Davy-Bowker et al. 2008). We recorded water width in association with each sample. Samples were preserved in 70% industrial methylated spirits.

2.3 | Laboratory Methods

Samples were processed to remove macroinvertebrates and all processed material was checked by a second person. We identified aquatic macroinvertebrate taxa to family level, except for Oligochaeta, which we identified as such. We also identified certain semi-aquatic and/or river-associated taxa to family (Table S2). We counted each taxon represented by a few individuals and estimated the abundance of taxa present at higher densities.

2.4 | Data Analysis

All analyses were done in R version 4.2.3 (R Core Team 2023).

2.4.1 | Calculation of Metrics Representing Macroinvertebrate Assemblages

We used WHPT-scoring taxa (hereafter, NTAXA, WHPT being a UK biomonitoring index indicative of organic pollution and general environmental degradation; Paisley, Trigg, and Walley 2014) to calculate two biotic metrics to represent the macroinvertebrate assemblage at each site ($n=27$), in each replicate sample ($n=81$) and in each sample's 10 component parts ($n=810$): the number of NTAXA (#NTAXA, termed WHPT NTAXA in WFD-UKTAG 2021), and the WHPT-ASPT (average score per taxon) based on 'present-only' taxon scores (Paisley, Trigg, and Walley 2014), because abundance-related scores are based on 3-min sampling durations. #NTAXA and WHPT-ASPT are the two metrics used to assess ecological status in UK rivers (WFD-UKTAG 2021). In calculating #NTAXA, we used all WHPT composite taxa (i.e., Limoniidae and Pediciidae within Tipulidae; Paisley, Trigg, and Walley 2014) and included Helophoridae within the Hydrophilidae, as per standard UK regulatory practice.

Total and taxon-specific invertebrate densities influence the sampling effort needed to capture a given proportion of the taxa present and thus to robustly assess ecological condition, with greater effort being required where more taxa are present at low densities (Getachew et al. 2022). To investigate this influence, we calculated three additional biotic metrics (based on NTAXA): abundance (i.e., the number of individual macroinvertebrates counted or estimated per sample), and the number and proportion of singleton and doubleton taxa (i.e., taxa represented by 1 or 2 individuals per sample, respectively).

2.4.2 | The Number and Percentage of NTAXA

We used Furse et al. (1981) as a benchmark to determine the kick-sampling duration that captures a comparable percentage of taxa from small streams as is collected in 3 min in larger rivers. Comparison with other studies investigating kick-sampling durations was inappropriate due to factors including the taxonomic identification level (i.e., species/genus in Bradley and Ormerod 2002 and Mykrä, Ruokonen, and Muotka 2006) and method of estimating total richness (e.g., Feeley et al. 2012; Mackey, Cooling, and Berrie 1984).

Furse et al. (1981) estimated that, in larger (i.e., mid-order) rivers, a 3-min kick sample captured 62% of the families present. To produce this estimate, Furse et al. (1981) sampled for 18 min per site in six 3-min samples and took the taxa captured during the 18-min sampling duration to represent 100% of the taxa present. Similarly, we used the taxa within all three replicate samples (i.e., in 15 min) to represent the total taxa at a site. Furse et al. (1981) estimated that approximately 3% of families were caught in minutes 15–18. Therefore, to compensate for our shorter, 15-min total sampling duration (and thus our lower estimates of total taxa and thus higher percentage of taxa captured per unit time), we set 65% (i.e., Furse et al.'s 62% + 3%) as our target percentage of macroinvertebrate families to capture in a kick sample.

To promote the application of our findings to regulatory biomonitoring, we used NTAXA including Chironomidae and Oligochaeta. In contrast, Furse et al. (1981) identified Oligochaeta to family, recording five such families, and identified Chironomidae to subfamily or tribe, including six such taxa in their otherwise family-level analyses. To compensate for our coarser identification of Chironomidae and Oligochaeta, we treated our 65% target as a minimum. Furse et al.'s (1981) dataset is unpublished, preventing further investigation and mitigation of this difference between the two studies.

For each of the three replicate 5-min samples collected at each site, we calculated the #NTAXA and the percentage of the total (15-min) #NTAXA (hereafter, %NTAXA) captured after each cumulative 0.5-min sampling duration (i.e., including the NTAXA in a component part and all preceding parts). We estimated #NTAXA and %NTAXA for each cumulative duration using 100 random permutations of the component parts of each replicate sample, using the function *specaccum* in the R package *vegan* version 2.6–4 (Oksanen et al. 2022). This analysis enabled determination of the duration at which the mean %NTAXA reached the $\geq 65\%$ target. To determine if the %NTAXA differed between this target duration and other durations, we ran linear mixed-effects models (LMM) using the function *lme* in the R package *nlme* version 3.1–159 (Pinheiro et al. 2023) with duration as a fixed factor and the cumulative %NTAXA as a response variable. We included replicate samples nested within the site as a random factor, to account for the non-independence of samples from each site. We used the *r.squaredGLMM* function in the R package *MuMIn* version 1.48.4 (Bartoń 2024) to calculate R^2 statistics describing the variance explained by the fixed factor (marginal R^2; R^2M) and by both the fixed and random factors (conditional R^2; R^2C). We removed one site for which variation in #NTAXA among replicates (see Appendix S1) compromised model performance, retaining 26 sites in the LMM.

We repeated the NTAXA analyses for subsets of sites with different core abiotic characteristics (i.e., width, altitude and flow permanence—but not alkalinity, which preliminary analyses indicated as redundant). For width, we investigated patterns in the narrowest streams by analysing sites with a mean width ≤ 1 m, as measured during the sampling campaign ($n = 11$). For altitude, we analysed subsets of sites at a relatively high (> 100 m, $n = 14$) and low (< 100 m, $n = 13$) altitude. For flow permanence, we analysed sites with perennial ($n = 14$) and temporary ($n = 7$) flow. For each subset, we ran LMM as described for the all-site dataset, removing the same highly variable sample from the low-altitude ($n = 12$) and temporary ($n = 6$) subsets.

To quantify variation in sampled assemblages introduced by differences between operators, we ran LMM with operator (A, B) as a fixed factor, site as a random factor and the cumulative #NTAXA as a response variable, and calculated R^2 statistics as described above. Results are presented in Appendix S2.

2.4.3 | WHPT Average Score per Taxon

We calculated WHPT-ASPT in the R package biomonitoR version 0.9.3 (Laini et al. 2022). To avoid loss of information, we renamed two families which are not in the taxonomic database underpinning biomonitoR (Schmidt-Kloiber and Hering 2015) as families which have the same WHPT scores but which were absent from the dataset.

To identify the sampling duration at which WHPT-ASPT stabilized, we calculated the median WHPT-ASPT after each cumulative 0.5-min duration based on 100 random permutations of their order. We expected variation (as SD) to decrease as duration (and therefore the #NTAXA) increased, with each additional taxon reducing the influence of each taxon present on the WHPT-ASPT value. We calculated the mean ± SD of the WHPT-ASPT median values for each cumulative duration across all sites. We then identified the duration at which the mean and SD were both < 0.1 from the final (5-min) mean and SD. This 0.1 value is arbitrary but is likely to be conservative enough to avoid misinterpretation of ecological condition.

We also calculated the sampling duration at which the mean WHPT-ASPT stabilized in the narrow stream, high and low-altitude stream, and perennial and temporary stream subsets.

2.4.4 | Differences Between Stream Types

To determine if sampling duration altered characterization of assemblages in small streams with different core abiotic characteristics (i.e., width, altitude and flow permanence), we calculated the five biotic metrics (i.e., #NTAXA, WHPT-ASPT, abundance, and the number and proportion of singleton and doubleton taxa) after each of three durations (2.5 min, which was the duration taken to reach the target of $\geq 65\%$ NTAXA; 3 min, the standard duration; and 5 min, the maximum duration). For each duration, we modelled the response of each biotic metric to width (continuous, as measured during the sampling campaign), altitude (continuous) or flow permanence (categorical: perennial, temporary). For width, we used LMM with site as a random factor ($n = 81$). For altitude, we ran linear models (LM) because the predictor variable was the same for all replicates at a site ($n = 81$). For flow permanence, we used LM and accounted for the unbalanced sampling design (i.e., $n = 42$

and $n = 21$ for the 14 perennial and seven temporary sites, respectively; Table S1) by comparing metric values at the seven temporary sites with seven randomly selected perennial sites, with 100 iterations; reported p values are the mean of these iterations. We calculated R^2M and R^2C for the width LMM and R^2 for the altitude LM.

3 | Results

In total, we recorded 158,833 invertebrates (mean ± SD 1955 ± 1422, range 114–4967 individuals per sample) from 94 taxa (24 ± 7, 5–39 taxa per sample)—comprising 82 WHPT-scoring aquatic taxa (including two composite taxa of five families), seven non-WHPT-scoring semi-aquatic or river-associated families, and two non-WHPT-scoring aquatic families—in 81 replicate 5-min samples from 27 sites (Table S2). WHPT-ASPT ranged from 3.22–7.07 (5.54 ± 1.09) in 5-min samples. The number of singletons and doubletons ranged from 3 to 10 (6.7 ± 1.86) taxa per 5-min sample, which accounted for 0.31 ± 0.10 (0.17–0.55) of #NTAXA.

3.1 | The Number and Percentage of NTAXA

The mean ± SD %NTAXA (and #NTAXA) per sample increased from 38% ± 8.8% (12 ± 4.5 taxa) after 0.5 min to 78% ± 11% (23 ± 7.4 taxa) after 5 min (Figure 2, Table 1). On average, the percentage captured reached the ≥65% target after 2.5 min, but there was considerable variation around this mean (65.2% ± 10.2% NTAXA; Figure 2b): a minimum of 1 min and a maximum of > 5 min were required to capture ≥ 65% NTAXA. In total, 54.3% of samples contained 65% NTAXA after 2.5 min, increasing to 70.4% of samples after 3 min.

Considering the site subsets, the ≥ 65% target was also reached after 2.5 min in high-altitude and perennial streams (Table A1). In narrow and low-altitude streams, the %NTAXA captured was slightly < 65% after 2.5 min, and 3-min samples reached the target. In temporary streams, it took 3.5 min to capture ≥ 65% NTAXA, due to a high number and proportion of singleton and doubleton taxa in one sample (see Appendix S1).

Considering all sites, the %NTAXA captured after 2.5 min (LMM estimate 66.1%, 95% confidence intervals [CI] 63.8%–68.5%) differed from that captured after all other sampling durations, including 2 min (estimated 3.9% lower %NTAXA, 1.2 fewer #NTAXA, and Cl 60.0%–64.7%) and 3 min (3.2% higher %NTAXA, 1.0 more #NTAXA, and Cl 67.0%–71.7%; LMM, $p < 0.001$; Table 2); subset results are shown Table A2. Sampling duration explained far more variance in %TAXA than the random factor (i.e., sample nested within site), both for all sites and for each site subset (R^2M = 0.663–0.696, R^2C = 0.962–0.976).

3.2 | WHPT Average Score per Taxon

The mean ± SD of the median WHPT-ASPT increased from 5.22 ± 0.32 after 0.5 min to 5.54 ± 1.09 after 5 min (Figure 3), stabilizing (i.e., the mean and SD falling to < 0.1 from their 5-min values, 5.44 and 1.14) after 2 min. Considering all sites, WHPT-ASPT stabilized after a minimum of 0.5 min (in 20 samples from 15 sites) and a maximum of 4.5 min (in 1 sample). Results were largely comparable for the site subsets, except WHPT-ASPT stabilized after 3 min at temporary sites (see Appendix S1).

3.3 | Differences Between Stream Types

Water width did not affect any metric after any sampling duration (LMM: #NTAXA, R^2M ≤ 0.004, R^2C = 0.869–0.902, p = 0.363–0.521, Figure 4a; WHPT-ASPT, R^2M ≤ 0.004, R^2C = 0.967–0.978, p = 0.133–0.311, Figure 4b; abundance, R^2M ≤ 0.015, R^2C = 0.672–0.796, p = 0.270–0.515, Figure 4c; number of singletons and doubletons, R^2M ≤ 0.016, R^2C = 0.171–0.231, p = 0.305–0.599, Figure 4d; proportion of singletons and doubletons, R^2M ≤ 0.022, R^2C = 0.159–0.448, p = 0.218–0.645).

A weak positive relationship between altitude and #NTAXA increased in strength and significance as sampling duration increased from 2.5 min (LM, $R^2 = 0.077$, $p = 0.012$) to 3 min ($R^2 = 0.083$, $p = 0.009$) then 5 min ($R^2 = 0.091$, $p = 0.006$; Figure 5a). This relationship may reflect the low #NTAXA at lowland sites in poorer ecological condition, but we lack the data to substantiate this suggestion. A positive relationship between

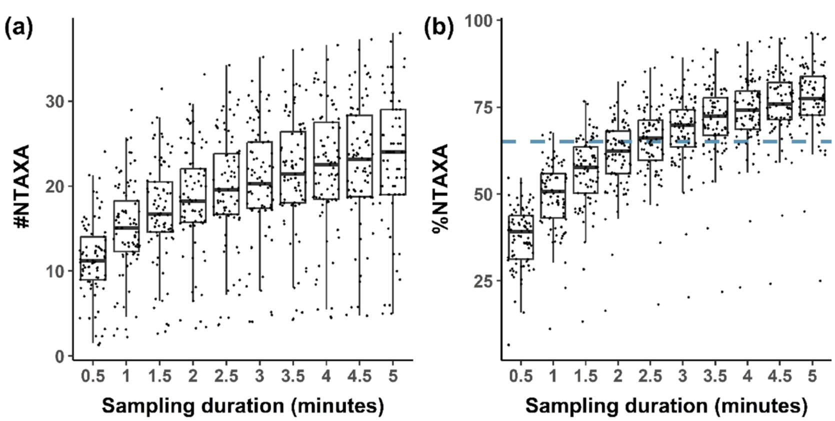

FIGURE 2 | The (a) number of WHPT-scoring taxa (#NTAXA) and (b) percentage of the total (i.e., 15-min) number of WHPT-scoring taxa (%NTAXA) captured after each cumulative 0.5-min sampling duration in 5-min kick samples ($n = 81$). The dashed line on (b) represents the ≥ 65% target. In each box, the horizontal line is the median, the box area indicates the first and third quartiles, and whiskers represent 95% confidence intervals. Points are jittered to aid visualization. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 1 | Mean and standard deviation (SD) number and percentage of WHPT-scoring taxa (#NTAXA, %NTAXA) captured after each cumulative 0.5-min sampling duration. Percentages are in relation to the total (i.e., 15-min) #NTAXA captured at each site.

	#NTAXA		%NTAXA	
Duration (min)	**Mean**	**SD**	**Mean**	**SD**
0–0.5	11.5	4.5	37.7	8.8
0.5–1	15.0	5.5	49.3	9.8
1–1.5	17.1	6.0	56.3	10.1
1.5–2	18.6	6.3	61.3	10.1
2–2.5	19.7	6.5	65.2	10.2
2.5–3	20.6	6.8	68.4	10.2
3–3.5	21.4	7.0	71.1	10.2
3.5–4	22.1	7.1	73.6	10.3
4–4.5	22.8	7.3	75.7	10.5
4.5–5	23.3	7.4	77.6	10.6

TABLE 2 | Linear mixed-effects model results comparing the estimated number and percentage of WHPT-scoring taxa (#NTAXA, %NTAXA) captured after each cumulative 0.5-min sampling duration with the 2.5-min target duration required to capture ≥ 65% NTAXA (highlighted).

	#NTAXA		%NTAXA	
Duration (min)	**Estimate**	***p***	**Estimate**	***p***
0–0.5	−8.4	< 0.001	−28	< 0.001
0.5–1	−4.8	< 0.001	−15.8	< 0.001
1–1.5	−2.7	< 0.001	−8.8	< 0.001
1.5–2	−1.2	< 0.001	−3.9	< 0.001
2–2.5	20.1	< 0.001	66.1	< 0.001
2.5–3	+1.0	< 0.001	+3.2	< 0.001
3–3.5	+1.8	< 0.001	+5.9	< 0.001
3.5–4	+2.5	< 0.001	+8.3	< 0.001
4–4.5	+3.1	< 0.001	+10.4	< 0.001
4.5–5	+3.6	< 0.001	+12.3	< 0.001

WHPT-ASPT and altitude was moderate ($R^2 = 0.365–0.374$) and significant ($p < 0.001$) after all durations (Figure 5b). Abundance was highly variable at altitudes < 200 m.a.s.l. and was consistently low at altitudes > 250 m.a.s.l., resulting in a weak negative relationship that became significant and increased in strength with sampling duration ($R^2 = 0.047$, 0.052 and 0.059, and $p = 0.052$, 0.041 and 0.030, after 2.5, 3 and 5 min, respectively; Figure 5c). Consistent with the increase in #NTAXA, the number of singletons and doubletons increased with altitude after all durations ($p = < 0.001–0.009$; Figure 5d), this relationship becoming weaker with duration

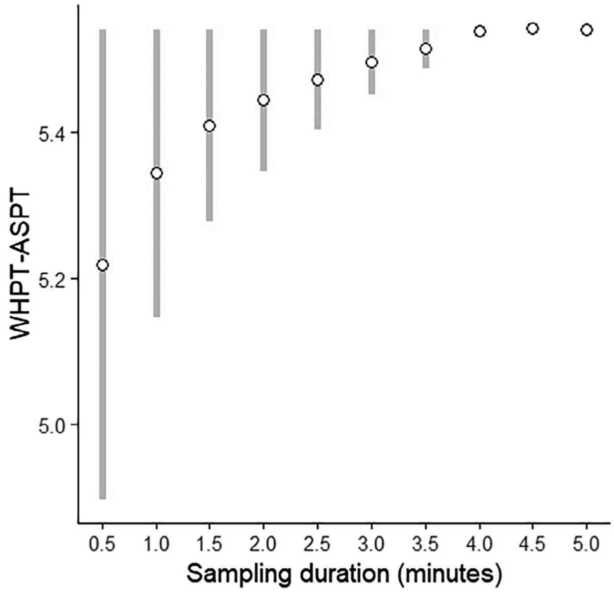

FIGURE 3 | Change in the mean of the median cumulative WHPT-ASPT over the 5-min sampling duration, based on 100 permutations of their order. Vertical bars represent the difference between the mean WHPT-ASPT at that duration and the WHPT-ASPT value at 5 min.

($R^2 = 0.15$, 0.12 and 0.08 after 2.5, 3 and 5 min, respectively). The proportion of these taxa became increasingly comparable across the altitude gradient as duration increased from 2.5 min ($p = 0.370$) to 5 min ($p = 0.975$).

Flow permanence had no significant effect on any metric ($p = 0.182–0.508$), but #NTAXA and WHPT-ASPT, in particular, were moderately higher at perennial compared with temporary sites (Figure 6), for all sampling durations. The #NTAXA was 4.6 taxa higher at perennial than temporary sites after 2.5 min, increasing to 4.9 taxa higher after 5 min (Figure 6a). Accordingly, the number of singletons and doubletons was 0.26–0.91 higher at perennial sites (Figure 6d) but did not change with sampling duration. The proportion of singletons and doubletons was comparable across site types and durations. WHPT-ASPT was 1.10 higher at perennial sites after 2.5 min, decreasing to 1.03 after 5 min (Figure 6b). Whereas WHPT-ASPT remained stable at 5.7 ± 1.1 at 2.5–5-min durations at perennial sites, it increased from 4.5 ± 0.6 after 2.5 min to 4.7 ± 0.6 after 5 min at temporary sites. Abundance was 387 individuals per sample higher at perennial sites after 2.5- and 3-min durations, increasing to 628 individuals per sample after 5 min (Figure 6c).

4 | Discussion

Integrating small streams into biomonitoring programmes is crucial to generate holistic catchment-scale understanding of river condition. Biomonitoring programmes routinely collect macroinvertebrate assemblages using 3-min kick samples (Friberg et al. 2006; Murray-Bligh and Griffiths 2022) which, in larger, mid-order rivers, are estimated to capture 62% of the taxa present—sufficient to indicate condition (Furse et al. 1981; WFD-UKTAG 2008). However, the extent to which assemblages

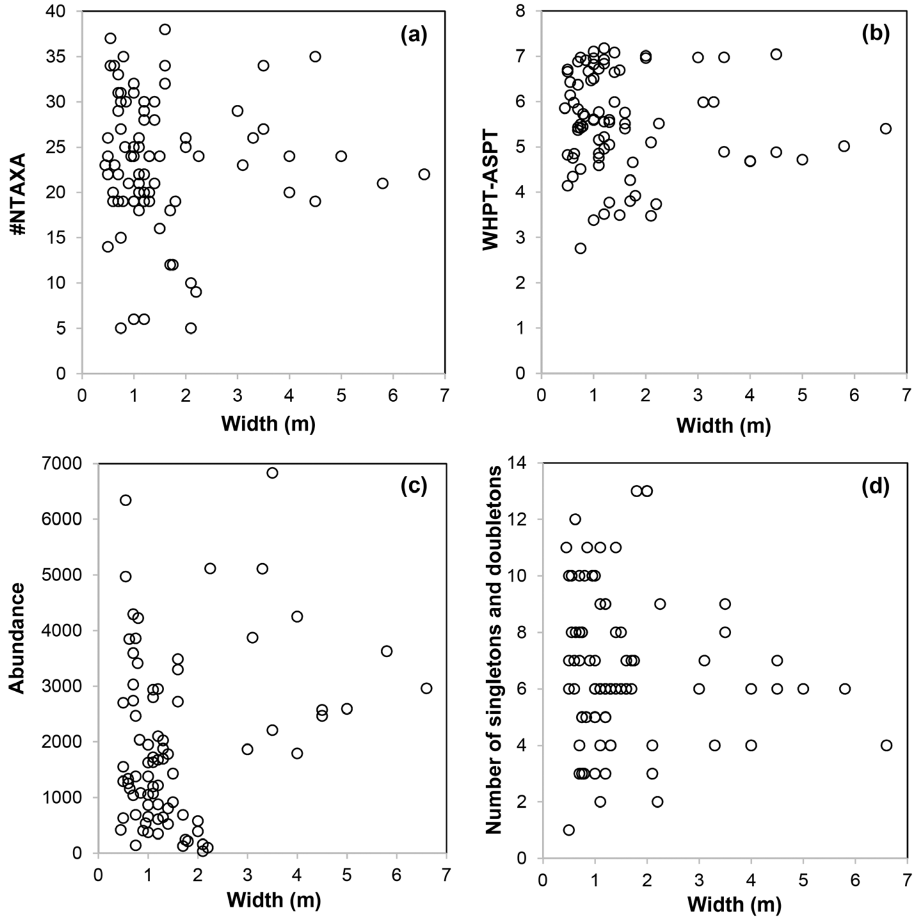

FIGURE 4 | Relationships between biotic metrics based on WHPT-scoring taxa and water widths: the (a) number of taxa (#NTAXA), (b) WHPT-ASPT (average score per taxon), (c) total abundance and (d) number of singletons and doubletons, in each replicate 5-min kick sample ($n = 81$).

collected using this method provide comparable representation of small stream communities has not previously been investigated, and systematic differences in capture rates could alter conclusions regarding condition, hampering holistic, river-scale and catchment-scale assessments.

We kick sampled sites representing England's SSN, using Furse et al.'s (1981) capture rates as a benchmark. Considering all sites, sampling for 2.5 min captured a comparable percentage of WHPT-scoring taxa (NTAXA) to that sampled from larger river sites in 3 min (Furse et al. 1981), and WHPT-ASPT stabilized after 2 min. Considering sites with different characteristics—namely narrow widths, high and low altitudes, and perennial and temporary flow permanence—we observed one notable deviation from this pattern: it took 3.5 min to capture ≥ 65% NTAXA and 3 min for WHPT-ASPT to stabilize at sites with temporary flow regimes. Below, we argue that these results—alongside the considerable variability in estimated capture rates and the value of aligning with methods used in larger rivers—suggest 3-min kick samples as appropriate in small stream biomonitoring.

4.1 | Three Minutes: The Best Duration to Characterize Small Stream Condition?

Considering all sites, kick sampling for 2.5 min captured a comparable mean %NTAXA as was sampled in 3 min in larger rivers by Furse et al. (1981), while WHPT-ASPT stabilized after 2 min. These results might suggest 2.5 min as the ideal kick-sampling duration in small streams. However, Furse et al.'s (1981) inclusion of Chironomidae and Oligochaeta at lower taxonomic levels than we achieved means that our 65% is the minimum value in a target range, and the 68% we captured after 3 min could also match Furse et al.'s (1981) findings, depending on the richness within these two ubiquitous taxa. In addition, only 54% of samples contained ≥ 65% NTAXA after 2.5 min, compared to 70% after 3 min. Given that taxa richness and densities can be lower in small streams compared with mid-order and larger rivers (Clarke et al. 2008; Minshall, Petersen Jr, and Nimz 1985; Paller, Specht, and Dyer 2006), a 3-min duration would also promote capture of sufficient NTAXA to confidently assess condition. Moreover, adopting a 3-min duration would facilitate integration of new small stream biomonitoring programmes into existing river monitoring networks and associated analytical models, enabling holistic, river-wide and catchment-wide comparisons of condition. Thus, we suggest that a 3-min duration can effectively represent small stream assemblages and enable comparable assessment of their condition to that achieved in larger rivers.

Considerable variation around any estimate of capture rates is inevitable, due to natural spatial and temporal variability in macroinvertebrate communities as well as error introduced by operators and procedures. Evidencing considerable

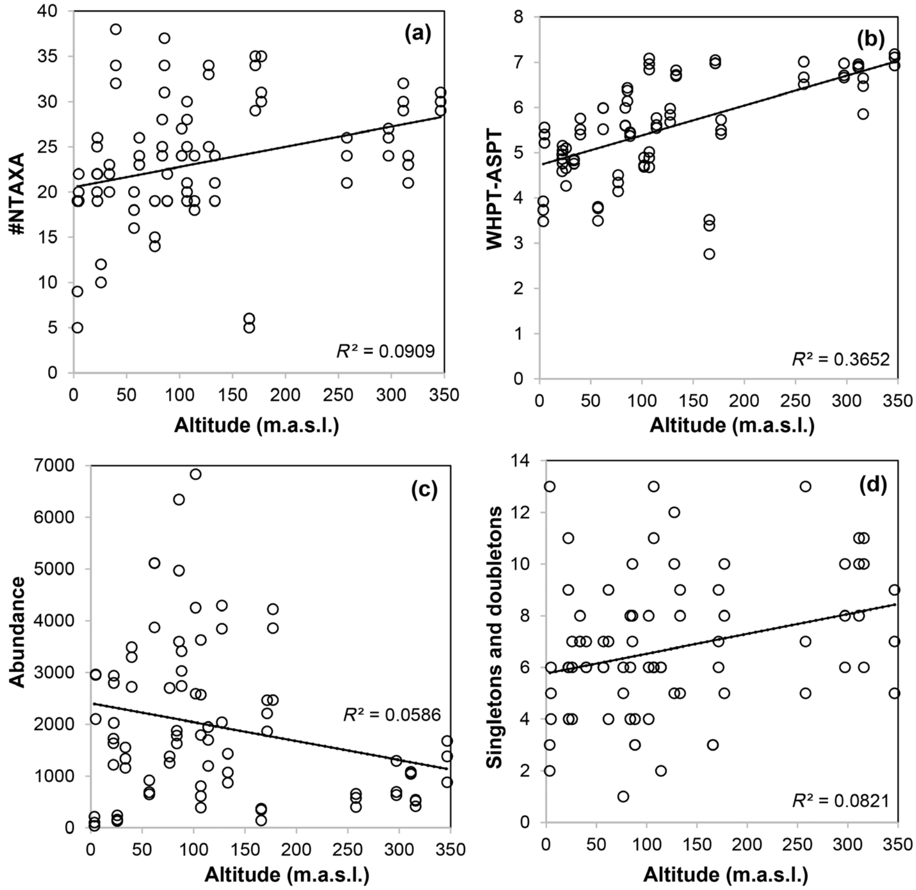

FIGURE 5 | Relationships between metrics based on WHPT-scoring taxa and site altitude: the (a) number of taxa (#NTAXA), (b) WHPT-ASPT (average score per taxon), (c) total abundance, and (d) number of singletons and doubletons, in each replicate 5-min kick sample ($n = 81$).

spatial variability in macroinvertebrate communities (and thus sampling efficiency), we captured ≥ 65% NTAXA in as little as 1 min, and more often, < 65% NTAXA were recorded after 5 min. Similarly, Feeley et al. (2012) required a mean of five (range 3–7) 1-min kick samples to capture ≥ 70% of the taxa present in small headwater streams. Spatial variability in macroinvertebrate communities may reflect the high habitat heterogeneity of small stream sites (Clarke et al. 2008) and, in the case of headwaters, their isolation (Sarremejane et al. 2017). Characterizing temporal variation was beyond our scope: our findings are specific to spring-season communities. In terms of seasonal variability, Feeley et al. (2012) found that the kick-sampling effort needed to capture ≥ 70% of taxa was comparable in spring and summer, suggesting that—although #NTAXA varies among seasons (Davy-Bowker et al. 2008)—seasonal variation in the percentage captured may be limited. Equally, we do not envisage significant interannual variation in the percentage captured.

Variability is also introduced by operators and procedures (Clarke et al. 2002). We quantified variation in the #NTAXA collected by the two field operators, which was non-significant but increased over time, potentially reflecting differences in operator fatigue (Feeley et al. 2012). These results suggest that, across river types, shorter (i.e., 2.5–3-min) kick-sampling durations may produce more consistent estimates of metrics representing macroinvertebrate assemblages (Furse et al. 1981; Mackey, Cooling, and Berrie 1984).

These sources of natural variability and operator error collectively suggest that the estimated 3.2% difference in NTAXA between 2.5-min and 3-min samples, although significant, is minor. In addition, metric relationships with width, altitude and flow permanence were generally comparable at 2.5, 3 and 5-min sampling durations, and for altitude and both #NTAXA and abundance, these comparable relationships increased in strength and significance with duration. In addition, at temporary sites, WHPT-ASPT scores increased with sampling duration. Collectively, these three findings support our suggestion that 3-min rather than 2.5-min durations may best enable robust macroinvertebrate-based assessment of small stream condition.

4.2 | It Takes Longer to Characterize Temporary Stream Communities

Our analysis of narrow streams, high and low-altitude streams, and perennial and temporary streams identified one notable deviation from the all-site patterns (i.e., the mean of 2.5 min required to capture ≥ 65% NTAXA and 2 min for WHPT-ASPT to stabilize): it took longer to reach these targets in temporary streams, that is, 3.5 min to capture ≥ 65% NTAXA and 3 min for WHPT-ASPT to stabilize. This difference was driven by one sample in which 13 of 19 NTAXA were singletons or doubletons and all other taxa also occurred at low abundance. Although all metrics representing macroinvertebrate assemblages were

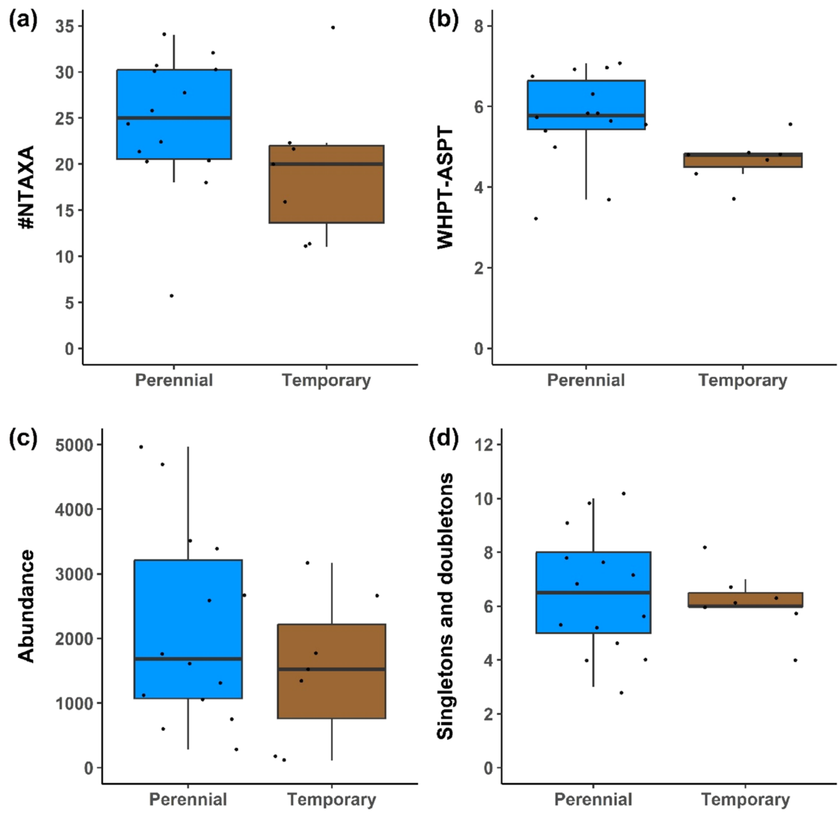

FIGURE 6 | The (a) number of WHPT-scoring taxa (#NTAXA), (b) WHPT-ASPT (average score per taxon), (c) total abundance, and (d) number of singletons and doubletons, in each replicate 5-min kick sample from perennial ($n=14$) and temporary ($n=7$) sites. The Figure 2 legend provides further details. [Color figure can be viewed at wileyonlinelibrary.com]

statistically comparable at temporary and perennial sites, abundance and taxa richness are typically lower in temporary than perennial streams, especially during the period of recolonization and community assembly after flow returns (Hill et al. 2019). As such, despite our spring-season sampling campaign, recent flow resumption may explain the low metric values at this site. In our temperate study area, and such assemblages may be more common in the autumn sampling season (Murray-Bligh and Griffiths 2022), during which flow may resume at temporary sites.

As such, the above suggestion—that the percentage of taxa captured by kick sampling is comparable across seasons—may not apply to temporary small streams. Here, assemblages can have lower densities, lower richness and higher proportions of singletons and doubletons after flow resumes (Aspin and House 2022), extending the duration required to sample a sufficient percentage of the community to robustly estimate #NTAXA and WHPT-ASPT (Mackey, Cooling, and Berrie 1984). Including rare (i.e., infrequently occurring) taxa can cause considerable change in both #NTAXA and WHPT-ASPT (Clarke and Murphy 2006); for example, these metrics increased from 13 to 19 and from 3.48 to 3.92, respectively, when singletons and doubletons were excluded from/included in analysis of the highly variable sample discussed above. Despite introducing variability among samples, including rare taxa can enhance estimation of condition (Clarke and Murphy 2006). Spring sampling thus maximizes the likelihood of capturing a sufficient, and sufficiently stable, assemblage to enable robust assessment of temporary small stream condition—and regardless of season, a 3-min sampling duration would promote capture of sufficient taxa.

4.3 | Towards Better Biomonitoring of Small Streams

At the scale of an individual sample, the relatively low taxonomic richness of small stream macroinvertebrate communities and the low densities of many taxa (Arscott, Tockner, and Ward 2005; Minshall, Petersen Jr, and Nimz 1985; Paller, Specht, and Dyer 2006) make it crucial to capture a sufficient number and percentage of taxa to enable robust assessment of stream condition. We characterized assemblages using the UK's two standard biomonitoring metrics, #NTAXA (i.e., WHPT NTAXA; WFD-UKTAG 2021) and WHPT-ASPT (Paisley, Trigg, and Walley 2014), recording 21 ± 6.7 scoring taxa per 3-min sample. Although no minimum number of taxa is required for metric calculation, additional taxa can enhance estimation of condition. First, we recorded seven semi-aquatic and river-associated terrestrial taxa, and observed many other terrestrial organisms that we did not identify. Second, we excluded two non-scoring aquatic families, Corduliidae and Thaumaleidae. Third, we included Hydrophilidae, Limoniidae and Pediciidae within composite families (Paisley, Trigg, and Walley 2014). As such, at least 12 captured taxa were excluded from analysis. Additional insight could be gained by incorporating such taxa into small stream biomonitoring, with particular benefits for temporary streams (England et al. 2019). Statistical analysis and expert judgement could both contribute to scoring a greater range of aquatic, semi-aquatic and river-associated taxa as new data become available. In addition, although not required to achieve our study aims, species-level identification greatly enhances understanding of ecological condition, and is common practice at regulatory agencies (Hering et al. 2004).

Beyond the individual sample scale, our study would ideally have considered communities both at reference sites (including least-disturbed sites, sensu Stoddard et al. 2006) and human-impacted sites, to determine the percentage of taxa captured by kick sampling in the near-absence and presence of anthropogenic influences. However, as elsewhere, the types and intensities of human pressures influencing sites in England's SSN have yet to be surveyed, and thus reference sites have yet to be identified. As such, the low taxa richness we recorded at some sites likely reflects human pressures as well as stream size. Our findings and recommendations will enable robust data collection in future work to characterize reference communities in small streams, then to characterize deviations therefrom. Both the values and variability of biomonitoring metrics (here, #NTAXA and WHPT-ASPT) representing small stream reference communities also require robust analysis, in particular given the potential for taxa present at low densities and thus rarely captured to alter metric values (Clarke and Murphy 2006). Metrics may benefit from evaluation and potentially from adaptation to incorporate the greater range of taxa discussed above, if their inclusion promotes consistent metric performance. Both characterization of reference conditions and testing of metric performance are urgently needed to support future biomonitoring of small stream condition.

5 | Conclusions

Small streams, in particular those comprising the headwaters of river networks, play vital roles in supporting catchment-wide biodiversity and river health (Alexander et al. 2007). Initiatives such as England's SSN as well as Ireland's small stream network (Kelly-Quinn et al. 2024) reflect the long-overdue incorporation of small streams into regulatory biomonitoring programmes. To maximize the benefits of such initiatives, robust sampling approaches are needed to assess ecosystem condition, thus promoting timely and accurate identification of sites at which action is required either to safeguard valued biodiversity or to reverse damage caused by human activities. Our results inform the design of such approaches. For macroinvertebrate-based biomonitoring programmes, we recommend a kick-sampling duration of 3 min, to facilitate collection of sufficient taxa across a diverse range of small streams including temporary streams, and—as the standard duration also used in larger rivers—to enable holistic assessment of catchment-wide river condition.

Acknowledgments

We thank Abi Halder, Billy Kane and Shana Hosty for their contribution to laboratory work. We thank Martin Winter and Chloe Hayes for help with invertebrate identification. We thank Judy England and Glen Meadows for information about, and arranging access to, chalk streams. We thank two anonymous reviewers whose constructive comments improved the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are openly available in the Supporting Information.

References

Alexander, R. B., E. W. Boyer, R. A. Smith, G. E. Schwarz, and R. B. Moore. 2007. “The Role of Headwater Streams in Downstream Water Quality.” *Journal of the American Water Resources Association* 43: 41–59. https://doi.org/10.1111/j.1752-1688.2007.00005.x.

Arscott, D. B., K. Tockner, and J. V. Ward. 2005. “Lateral Organization of Aquatic Invertebrates Along the Corridor of a Braided Floodplain River.” *Journal of the North American Benthological Society* 24: 934–954. https://doi.org/10.1899/05-037.1.

Aspin, T., and A. House. 2022. “Alpha and Beta Diversity and Species Co-Occurrence Patterns in Headwaters Supporting Rare Intermittent-Stream Specialists.” *Freshwater Biology* 67: 1188–1202. https://doi.org/10.1111/fwb.13910.

Baattrup-Pedersen, A., S. E. Larsen, D. K. Andersen, N. Jepsen, J. Nielsen, and J. J. Rasmussen. 2018. “Headwater Streams in the EU Water Framework Directive: Evidence-Based Decision Support to Select Streams for River Basin Management Plans.” *Science of the Total Environment* 613: 1048–1054. https://doi.org/10.1016/j.scitotenv.2017.09.199.

Bartoń, K. 2024. “MuMIn: Multi-Model Inference.” R Package Version. 1.48.4. https://cran.r-project.org/web/packages/MuMIn/index.html.

Biggs, J., S. von Fumetti, and M. Kelly-Quinn. 2017. “The Importance of Small Waterbodies for Biodiversity and Ecosystem Services: Implications for Policy Makers.” *Hydrobiologia* 793: 3–39. https://doi.org/10.1007/s10750-016-3007-0.

Bradley, D. C., and S. J. Ormerod. 2002. “Evaluating the Precision of Kick-Sampling in Upland Streams for Assessments of Long-Term Change: The Effects of Sampling Effort, Habitat and Rarity.” *Archiv für Hydrobiologie* 155: 199–221. https://doi.org/10.1127/archiv-hydrobiol/155/2002/199.

Callanan, M., J. R. Baars, and M. Kelly-Quinn. 2014. “Macroinvertebrate Communities of Irish Headwater Streams: Contribution to Catchment Biodiversity.” *Biology and Environment: Proceedings of the Royal Irish Academy* 114: 143–162. https://doi.org/10.3318/bioe.2014.30.

Clarke, A., R. Mac Nally, N. Bond, and P. S. Lake. 2008. “Macroinvertebrate Diversity in Headwater Streams: A Review.” *Freshwater Biology* 53: 1707–1721. https://doi.org/10.1111/j.1365-2427.2008.02041.x.

Clarke, R. T., M. T. Furse, R. J. M. Gunn, J. M. Winder, and J. F. Wright. 2002. “Sampling Variation in Macroinvertebrate Data and Implications for River Quality Indices.” *Freshwater Biology* 47: 1735–1751. https://doi.org/10.1046/j.1365-2427.2002.00885.x.

Clarke, R. T., and J. F. Murphy. 2006. “Effects of Locally Rare Taxa on the Precision and Sensitivity of RIVPACS Bioassessment of Freshwaters.” *Freshwater Biology* 51: 1924–1940. https://doi.org/10.1111/j.1365-2427.2006.01611.x.

Datry, T., A. J. Boulton, K. Fritz, et al. 2023. “Non-Perennial Segments in River Networks.” *Nature Reviews Earth and Environment* 4: 815–830. https://doi.org/10.1038/s43017-023-00495-w.

Davy-Bowker, J., R. Clarke, T. Corbin, et al. 2008. *River Invertebrate Classification Tool: Final Report: Project WFD72c*. Edinburgh, UK: Scotland and Northern Ireland Forum for Environmental Research. http://eprints.bournemouth.ac.uk/16550/2/SNIFFER_WFD72C_RICT_Final_Report_-_Davy-Bowker%2C_Clarke_et_al_2008.pdf.

Defra (Department for Environment Food & Rural Affairs). 2022. *Policy Paper: Natural Capital and Ecosystem Assessment Programme*. London, UK: Defra. https://www.gov.uk/government/publications/natural-capital-and-ecosystem-assessment-programme/natural-capital-and-ecosystem-assessment-programme.

Downing, J. A., J. J. Cole, C. M. Duarte, et al. 2012. "Global Abundance and Size Distribution of Streams and Rivers." *Inland Waters* 2: 229–236. https://doi.org/10.5268/IW-2.4.502.

Dunbar, M., J. Murphy, R. Clarke, R. Baker, C. Davies, and P. Scarlett. 2010. *Countryside Survey: Headwater Streams Report From 2007, CS Technical Report No. 8/07*. Lancaster, UK: Centre for Ecology & Hydrology. https://nora.nerc.ac.uk/id/eprint/9434/.

England, J., R. Chadd, M. J. Dunbar, et al. 2019. "An Invertebrate-Based Index to Characterize Ecological Responses to Flow Intermittence in Rivers." *Fundamental and Applied Limnology* 193: 93–117. https://doi.org/10.1127/fal/2019/1206.

European Commission. 2000. "Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 Establishing a Framework for Community Action in the Field of Water Policy." *Official Journal of the European Communities* L327, no. 64: 1–73.

Feeley, H. B., C. Bradley, G. Free, et al. 2020. "A National Macroinvertebrate Dataset Collected for the Biomonitoring of Ireland's River Network, 2007–2018." *Scientific Data* 7: 280. https://doi.org/10.1038/s41597-020-00618-8.

Feeley, H. B., M. Woods, J. R. Baars, and M. Kelly-Quinn. 2012. "Refining a Kick Sampling Strategy for the Bioassessment of Benthic Macroinvertebrates in Headwater Streams." *Hydrobiologia* 683: 53–68. https://doi.org/10.1007/s10750-011-0940-9.

Finn, D. S., N. Bonada, C. Múrria, and J. M. Hughes. 2011. "Small but Mighty: Headwaters Are Vital to Stream Network Biodiversity at Two Levels of Organization." *Journal of the North American Benthological Society* 30: 963–980. https://doi.org/10.1899/11-012.1.

Friberg, N., L. Sandin, M. T. Furse, S. E. Larsen, R. T. Clarke, and P. Haase. 2006. "Comparison of Macroinvertebrate Sampling Methods in Europe." *Hydrobiologia* 566: 365–378. https://doi.org/10.1007/s10750-006-0083-6.

Furse, M. T., and K. L. Symes. 1997. *Operational Mechanisms for the Protection and Enhancement of Headwaters. R&D Technical Report E25*. Bristol, UK: Institute of Freshwater Ecology and the Environment Agency. https://nora.nerc.ac.uk/id/eprint/507657/1/N507657CR.pdf.

Furse, M. T., J. F. Wright, P. D. Armitage, and D. Moss. 1981. "An Appraisal of Pond-Net Samples for Biological Monitoring of Lotic Macro-Invertebrates." *Water Research* 15: 679–689. https://doi.org/10.1016/0043-1354(81)90160-3.

Getachew, M., W. L. Mulat, S. T. Mereta, G. S. Gebrie, and M. Kelly-Quinn. 2022. "Refining Benthic Macroinvertebrate Kick Sampling Protocol for Wadeable Rivers and Streams in Ethiopia." *Environmental Monitoring and Assessment* 194: 196. https://doi.org/10.1007/s10661-021-09594-x.

Gibbs, W. K., S. B. Cook, and M. A. Kulp. 2023. "Comparing Passive and Active Macroinvertebrate Sampling Gear Efficacy During Biomonitoring in Southern Appalachian Mountain Streams." *River Research and Applications* 39: 2009–2018. https://doi.org/10.1002/rra.4202.

Haase, P., S. Lohse, S. Pauls, et al. 2004. "Assessing Streams in Germany With Benthic Invertebrates: Development of a Practical Standardised Protocol for Macroinvertebrate Sampling and Sorting." *Limnologica* 34: 349–365. https://doi.org/10.1016/S0075-9511(04)80005-7.

Hering, D., O. Moog, L. Sandin, and P. F. M. Verdonschot. 2004. "Overview and Application of the AQEM Assessment System." *Hydrobiologia* 516: 1–20. https://doi.org/10.1023/B:HYDR.0000025255.70009.a5.

Hill, M. J., K. L. Mathers, S. Little, T. Worrall, J. Gunn, and P. J. Wood. 2019. "Ecological Effects of a Supra-Seasonal Drought on Macroinvertebrate Communities Differ Between Near-Perennial and Ephemeral River Reaches." *Aquatic Sciences* 81: 62. https://doi.org/10.1007/s00027-019-0659-7.

Kabir, J., G. Biondi, K. J. Gething, T. Aspin, T. Sykes, and R. Stubbington. 2024. "Spring Forth Diversity: Specialist Species Contribute to the Conservation Value of Headwater Springs and Streams at the Landscape Scale." *River Research and Applications* 40: 863–874. https://doi.org/10.1002/rra.4275.

Kelly-Quinn, M., M. Bruen, M. Bowes, et al. 2024. *Managing Small Stream Networks for Improved Water Quality, Catchment Biodiversity and Ecosystem Services Protection*. Wexford, Ireland: Environmental Protection Agency. https://www.epa.ie/publications/research/water/research-468-managing-small-stream-networks-for-improved-water-quality-catchment-biodiversity-and-ecosystem-services-protection.php.

Laini, A., S. Guareschi, R. Bolpagni, et al. 2022. "BiomonitoR: An R Package for Managing Ecological Data and Calculating Biomonitoring Indices." *PeerJ* 10: e14183. https://doi.org/10.7717/peerj.14183.

Mackey, A. P., D. A. Cooling, and A. D. Berrie. 1984. "An Evaluation of Sampling Strategies for Qualitative Surveys of Macro-Invertebrates in Rivers, Using Pond Nets." *Journal of Applied Ecology* 21: 515–534. https://doi.org/10.2307/2403426.

Majaneva, M., L. E. Sundt-Hansen, H. Brandsegg, R. Sivertsgård, T. Bongard, and F. Fossøy. 2024. "Comparing Methods and Indices for Biodiversity and Status Assessment in a Hydropower-Regulated River." *Hydrobiologia* 851: 2139–2157. https://doi.org/10.1007/s10750-023-05448-4.

Messager, M. L., B. Lehner, C. Cockburn, et al. 2021. "Global Prevalence of Non-perennial Rivers and Streams." *Nature* 594: 391–397. https://doi.org/10.1038/s41586-021-03565-5.

Minshall, G. W., R. C. Petersen Jr., and C. F. Nimz. 1985. "Species Richness in Streams of Different Size From the Same Drainage Basin." *American Naturalist* 125: 16–38. https://doi.org/10.1086/284326.

Minshall, G. W., R. C. Petersen, K. W. Cummins, et al. 1983. "Interbiome Comparison of Stream Ecosystem Dynamics." *Ecological Monographs* 53: 1–25. https://doi.org/10.2307/1942585.

Murray-Bligh, J., and M. Griffiths. 2022. *Freshwater Biology and Ecology Handbook*. Marlow, UK; Ulverston, UK: Foundation for Water Research; Freshwater Biological Association. https://www.fba.org.uk/shop/p/freshwater-biology-and-ecology-handbook.

Mykrä, H., T. Ruokonen, and T. Muotka. 2006. "The Effect of Sample Duration on the Efficiency of Kick-Sampling in Two Streams With Contrasting Substratum Heterogeneity." *Internationale Vereinigung Für Theoretische Und Angewandte Limnologie: Verhandlungen* 29: 1351–1355. https://doi.org/10.1080/03680770.2005.11902901.

Oksanen, J., G. L. Simpson, F. G. Blanchet, et al. 2022. "Vegan: Community Ecology Package." R Package Version 2.6-4. https://CRAN.R-project.org/package=vegan.

Paisley, M. F., D. J. Trigg, and W. J. Walley. 2014. "Revision of the Biological Monitoring Working Party (BMWP) Score System: Derivation of Present-Only and Abundance-Related Scores From Field Data." *River Research and Applications* 30: 887–904. https://doi.org/10.1002/rra.2686.

Paller, M. H., W. L. Specht, and S. A. Dyer. 2006. "Effects of Stream Size on Taxa Richness and Other Commonly Used Benthic Bioassessment Metrics." *Hydrobiologia* 568: 309–316. https://doi.org/10.1007/s10750-006-0208-y.

Pinheiro, J., D. Bates, and R Core Team. 2023. "Nlme: Linear and Nonlinear Mixed Effects Models." R Package Version 3.1-159. https://CRAN.R-project.org/package=nlme.

R Core Team. 2023. "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing. https://www.Rproject.org/.

Riley, W. D., E. C. Potter, J. Biggs, et al. 2018. "Small Water Bodies in Great Britain and Ireland: Ecosystem Function, Human-Generated Degradation, and Options for Restorative Action." *Science of the Total Environment* 645: 1598–1616. https://doi.org/10.1016/j.scitotenv.2018.07.243.

Rosenberg, D. M., and V. H. Resh. 1993. *Freshwater Biomonitoring and Benthic Macroinvertebrates*. New York, NY: Chapman & Hall.

Sarremejane, R., H. Mykrä, N. Bonada, J. Aroviita, and T. Muotka. 2017. "Habitat Connectivity and Dispersal Ability Drive the Assembly Mechanisms of Macroinvertebrate Communities in River Networks." *Freshwater Biology* 62: 1073–1082. https://doi.org/10.1111/fwb.12926.

Schmidt-Kloiber, A., and D. Hering. 2015. "An Online Tool That Unifies, Standardises and Codifies More Than 20,000 European Freshwater Organisms and Their Ecological Preferences." *Ecological Indicators* 53: 271–282. https://doi.org/10.1016/j.ecolind.2015.02.007.

Scott, D. T., J. D. Gomez-Velez, C. N. Jones, and J. W. Harvey. 2019. "Floodplain Inundation Spectrum Across the United States." *Nature Communications* 10: 5194. https://doi.org/10.1038/s41467-019-13184-4.

Sefton, C. E. M., S. Parry, J. England, and G. Angell. 2019. "Visualising and Quantifying the Variability of Hydrological State in Intermittent Rivers." *Fundamental and Applied Limnology* 193: 21–38. https://doi.org/10.1127/fal/2019/1149.

Smith, J. J., and A. Lyle. 1979. *Distribution of Freshwaters in Great Britain*. Cambridge, UK: Institute of Terrestrial Ecology. https://nora.nerc.ac.uk/id/eprint/5238/1/Distribution_freshwaters.pdf.

Stoddard, J. L., D. P. Larsen, C. P. Hawkins, R. K. Johnson, and R. H. Norris. 2006. "Setting Expectations for the Ecological Condition of Streams: The Concept of Reference Condition." *Ecological Applications* 16: 1267–1276. https://doi.org/10.1890/1051-0761(2006)016[1267:SEFTEC]2.0.CO;2.

von Gönner, J., J. Gröning, V. Grescho, et al. 2024. "Citizen Science Shows That Small Agricultural Streams in Germany Are in a Poor Ecological Status." *Science of the Total Environment* 922: 171183. https://doi.org/10.1016/j.scitotenv.2024.171183.

WFD-UKTAG (Water Framework Directive–United Kingdom Technical Advisory Group). 2008. "UKTAG River Assessment Methods: Benthic Invertebrate Fauna: River Invertebrate Classification Tool (RICT)." WFD–UKTAG, Edinburgh, UK. https://www.wfduk.org/sites/default/files/Media/Characterisation%20of%20the%20water%20environment/Biological%20Method%20Statements/river%20invertebrates.pdf.

WFD-UKTAG. 2021. "UKTAG River Assessment Method: Benthic Invertebrate Fauna: Invertebrates (General Degradation): Walley, Hawkes, Paisley & Trigg (WHPT) Metric in River Invertebrate Classification Tool (RICT)." Environment Agency, Bristol, UK. https://wfduk.org/sites/default/files/River%20Invertebrates%20WHPT%20UKTAG%20Method%20Statement%20-%20updated%20May%202021.pdf.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.