First audit of macroinvertebrate samples from an EU Water Framework Directive monitoring program: human error greatly lowers precision of assessment results

Authors: Haase, Peter, Pauls, Steffen U., Schindehütte, Karin, and Sundermann, Andrea

First audit of macroinvertebrate samples from an EU Water Framework Directive monitoring program: human error greatly lowers precision of assessment results

Peter Haase[1,3], Steffen U. Pauls[1,2,4], Karin Schindehütte[1,5], AND Andrea Sundermann[1,6]

[1] *Senckenberg, Research Institute and Natural History Museum, Department of Limnology and Conservation, Clamecy Strasse 12, 63571 Gelnhausen, Germany*
[2] *University of Minnesota, Department of Entomology, St Paul, Minnesota 55108 USA*

Abstract. Invertebrates are often used in biological monitoring of soil and water ecosystems. Because of the huge number of invertebrate species, sample processing (sorting and identification) is a labor-intensive and often difficult task that is prone to error. These errors can bias assessment results, which often are used by environmental managers to guide funding decisions for costly restoration measures. However, quality control of assessment results is not implemented in many freshwater monitoring programs. We conducted the first audit of an official European freshwater monitoring program based on 414 macroinvertebrate samples from streams and rivers in Germany. The samples were collected by personnel at 7 different commercial environmental laboratories using the European Union (EU) Water Framework Directive protocol. We audited 12% of all samples at 3 different levels: 1) a sorting audit, 2) an identification audit, and 3) a total audit based on both sorting and identification. The sorting audit revealed that 29% of the specimens and every 5th taxon (20.6%) had been overlooked by the primary analyst. Differences in sorting were correlated with taxon body size ($r = 0.61$, $p < 0.001$). The identification audit showed that >30% of taxa differed between the results of the primary analysts and auditors. Taxa considered difficult to identify were not more prone to error than were taxa considered easier to identify. Primary analysts and auditors assigned 34% of audited samples to different quality classes in ≥ 1 of 3 assessment modules (organic pollution, acidification, and general degradation). For 16% of the samples, these changes resulted in a different final ecological assessment. Such a high rate of differences between primary analysts and auditors could lead to ineffective allocation of several million Euros. Our results clearly illustrate the need for adequate quality control and auditing in freshwater monitoring.

Key words: stream assessment, methodology, quality control, quality assurance, benthic invertebrates, river, stream, ecological status, water management.

Use of benthic macroinvertebrates to monitor water quality and ecological status of freshwaters has a long history in Europe, North America, and Australia (Wright et al. 1984, Smith et al. 1999, Karr and Chu 2000). In Europe, the importance of monitoring increased significantly with the implementation of the European Union Water Framework Directive (EUWFD) in 2000 (European Union 2000). This directive requires that all surface water bodies achieve good or high ecological status based on structural, chemical, and biotic parameters. Where the current ecological status of a given body of water is not good or high, mitigation measures are demanded by the EUWFD. A first estimate of the ecological status of streams and rivers in Europe revealed that ~70% of surface water bodies currently fail to meet good or high ecological status (EU Commission 2007). Consequently, many restoration measures will be conducted throughout Europe in the next few years at a cost of billions of Euros. Therefore, assessment results from monitoring programs must be reliable so that water managers can make sound decisions concerning restoration measures.

Much research effort has been devoted to developing new or adapting established monitoring programs

[3] E-mail addresses: peter.haase@senckenberg.de
[4] steffen.pauls@senckenberg.de
[5] karin.schindehuette@munlv.nrw.de
[6] andrea.sundermann@senckenberg.de

and assessment systems throughout Europe (Haase et al. 2004a, Sandin et al. 2004, Vlek et al. 2004). All methods for monitoring benthic invertebrates are subject to variability derived from natural sources like patchiness (Pringle et al. 1988, Townsend 1989, Downes et al. 1993, Li et al. 2001, Olsen et al. 2007) or inherent to the method itself (Cao et al. 2003, Haase et al. 2004b, 2008, Clarke and Murphy 2006, Clarke et al. 2006, Friberg et al. 2006, Nichols et al. 2006, Sundermann et al. 2008). However, human error is a source of variability that has been largely neglected (Ostermiller and Hawkins 2004). Human error potentially affects all stages of freshwater biomonitoring, including site selection, field sampling, sorting, identification, data entry, analyses, and interpretation (Clarke and Hering 2006). However, we currently have very little understanding of the importance of this factor because few suitable auditing schemes exist. In Europe, regular quality control of macroinvertebrate sample processing for stream assessment is implemented only in Great Britain. Presumably, quality control is not widely implemented in macroinvertebrate-based stream assessment because it is costly. Moreover, identifying errors is often not as straightforward as it is for other analyses, e.g., in chemical water analysis, where analytical technologies are used to measure objective values. In biological quality control, differences in results are easy to detect, but true errors are difficult to evaluate because identifying species is often subjective. However, despite these difficulties, it is valuable to know how operator-dependent differences in sample processing and identification can affect outcomes of stream assessments. The few published studies of this issue show a considerable amount of operator-related sorting and identification error (Haase et al. 2006a, Stribling et al. 2008).

Our goal was to evaluate the effect of human error in sample processing on biomonitoring results. We audited macroinvertebrate samples collected from official EUWFD monitoring sites in Germany and analyzed the effect of sorting and identification errors on EUWFD assessment results. We present the first published quality-control audit of sample processing for official freshwater monitoring sites.

Methods

Audit design

In 2006, the authorities of the German federal state Hesse contracted 7 commercial environmental laboratories (hereafter referred to as primary analysts [PAs] 1–7) to collect 414 macroinvertebrate samples from EUWFD monitoring sites using the EUWFD standard sampling protocol in Germany (Haase et al. 2004a). The protocol calls for sampling microhabitats in proportion to their coverage at the survey site (multihabitat sampling). All microhabitats are recorded in 5% coverage intervals and noted on a field protocol. Each 5%-microhabitat sampling unit is sampled by kick sampling with a hand-held net (mesh size = 0.5 mm). A complete sample consists of 20 sampling units (total sampling area = 1.25 m^2), which are pooled for further treatment. The entire sample is sorted by the PA with the standardized protocol of Haase et al. (2004a). This protocol is analogous to the widely used **Sta**ndardized **R**iver Classifications/Assessment System for the Ecological Quality of Streams and Rivers throughout Europe using Benthic Macroinvertebrates (STAR/AQEM) protocol (Furse et al. 2006). Sample material consists of the organisms and some coarse particulate organic matter (e.g., leaves, twigs, filamentous algae, or moss).

The PAs were directed to sort all organisms from a sample and to store them in vials for later identification. The sorting protocol did not allow either the PA or the auditor to use magnification or stains. The PAs were aware that all 414 samples were potentially subject to audit. The audit consisted of a sorting audit, an identification audit, and a total audit of 50 of the 414 samples (~12%) randomly selected by the auditors. The number of processed samples was different for every PA, so the number of randomly selected samples varied between 4 and 18 and represented the relative contribution of each PA to the state-wide survey of 414 sites. Thus, the performance of individual PAs should be interpreted with caution because of the limited number of samples audited for many of the PAs.

Sorting audit.—The aim of the sorting audit was to detect specimens remaining in the sample residue. PAs were instructed to remove all individuals from the sample material and to retain the residue and sorted specimens separately. The auditors resorted the whole sample residue, removed any animals found, and placed them in a new, labeled vial. The auditors counted these specimens, identified them to the taxonomic level defined in the operational taxon list (Haase et al. 2006b), and added them to the corresponding taxon list generated by the PA. We compared the number of individuals, number of taxa, and assessment results for the original taxon list generated by the PA and the taxon list after the sorting audit. Percentages given in the text or tables refer to the taxon list after the sorting audit, which we assumed represented 100% of the organisms in the sample.

We used Spearman rank correlation analysis to test whether the probability of overlooking a taxon was

related to body size. Body size was determined based on trait descriptions in Usseglio-Polatera (1991) and Chevenet et al. (1994). We also used correlation analysis to test whether the number of overlooked specimens was related to the percentage of organic substrates (fine particulate organic matter, coarse organic matter, or submerged aquatic vegetation; calculated from micro-habitat records). We calculated the total number of differences in the sorting audit for all taxa. For taxon-specific calculations, we considered only taxa present in ≥10 of the 50 samples to avoid biasing results by taxa that occurred in very few samples.

Identification audit.—PAs were instructed to identify all specimens from the sorted sample and to retain up to 5 voucher specimens for each identified taxon, which is common practice in Germany. The voucher specimens for each sample were stored together in 1 vial. Voucher specimens were reidentified by the auditors (who were unaware of what the PA had determined) to the taxonomic level defined in the operational taxon list (Haase et al. 2006b), which is genus or species level for most taxa, and subfamily or family level for selected groups, such as Chironomidae or Oligochaeta. The auditors were taxonomic specialists for particular taxonomic groups, and they re-identified voucher specimens from their area of expertise in all samples. However, we do not consider the results of the auditors to be necessarily the correct identifications. Thus, we do not refer to *errors* in our results, but rather to differences between the taxon lists generated by PAs and auditors. Samples used in the identification audit were identical to those used in the sorting audit.

We examined whether differences in the identification audit were related to the generally perceived difficulty of identifying particular taxonomic groups. We based our analysis on the following identification categories: 1) taxon can be identified in the field with basic taxonomic knowledge, 2) taxon can be identified in the field with advanced taxonomic knowledge or in the laboratory with basic taxonomic knowledge, 3) taxon can be identified in the laboratory with advanced taxonomic knowledge, and 4) taxon cannot be identified or can be identified only by a taxonomic specialist (Bayerisches Landesamt für Wasserwirtschaft 2004). As in the sorting audit, we calculated the total number of differences in the identification audit for all taxa, but taxon-specific calculations were done only for those taxa that occurred in ≥10 of the 50 samples.

Data analyses and assessment results

The EUWFD macroinvertebrate assessment system for streams and rivers in Germany is based on a reference-condition approach, and class boundaries are defined specifically for each stream type. The system consists of 3 modules, and each addresses a different stressor: organic pollution, acidification, and general degradation (including morphology). The assessment score for each module is classified in 1 of 5 quality classes (high, good, moderate, poor, bad) in accordance with the EUWFD. The overall assessment result is the Ecological Quality Class (EQC) and is based on the worst quality class to which a sample is assigned among the 3 modules (worst-case principle).

The organic pollution module is based on a saprobic index that ranges from 1 to 4 and has stream-type-specific quality-class boundaries (Meier et al. 2006). The acidification module follows a similar principle as the saprobic index. In Germany, 278 taxa were listed according to their sensitivity to acidification and were classified into 5 classes (Braukmann and Biss 2004). The acidification index is calculated only for the 2 stream types (of 24 total types in Germany) potentially affected by acidification. However, these 2 stream types are relatively common in Germany. The general degradation module is based on a stream-type-specific multimetric index (MMI). Each stream-type-specific MMI is composed of 3 to 5 different metrics and is scaled to values between 0 (bad) and 1 (high). Class boundaries occur at 0.2-unit scoring intervals (Böhmer et al. 2004).

For each sample, assessment scores for all 3 modules were calculated from taxon lists that were adjusted to the standardized taxonomic level outlined in the operational taxon list (Haase et al. 2006b; see *Identification audit* above) to ensure that the taxonomic resolution of taxon lists were comparable. The taxonomic levels used in each module and metric generally are species or genus level except in the metrics % Ephemeroptera, Plecoptera, Trichoptera (EPT) taxa and number of Ephemeroptera, Plecoptera, Trichoptera, Colembola, Bivalvia, Odonata (EPTCBO) taxa of the general degradation module, which partially use lower resolution. We used the audit data to calculate all assessment results with the software ASTERICS (version 3.01; www.fliessgewaesserbewertung.de), which is used by water managers in Germany. The influence of single taxon on the EQC was derived empirically by considering the importance of the taxon in the assessment system.

Results

Sorting audit

On average, 340 specimens/sample (29%) were overlooked or not removed by the PA and were found

Table 1. Results of the sorting audit. Mean (±1 SD) and range of relative and absolute numbers of specimens and taxa found in the sample residue. New taxa = those taxa found only by the auditors. Relative numbers are given as percentages based on the taxon list after the sorting audit.

Primary analyst	N	Specimens		Taxa	
		Relative number (%)	Absolute number	Relative number (%)	Absolute number
All	50	29 ± 22.2 (0.3–88.6)	340 ± 593 (1–3624)	51 ± 21.5 (2.6–87.5)	18 ± 11.5 (1–45)
New taxa				20.6 ± 14.4 (0–53.1)	7.8 ± 7 (0–24)
1	4	22.1 ± 9.4 (9–31.5)	111 ± 45 (54–149)	51.6 ± 10.6 (36.4–60.7)	25.3 ± 6.1 (20–34)
New taxa				21 ± 8.6 (10.9–31.7)	10 ± 3.4 (6–13)
2	4	1.9 ± 1.4 (0.3–3.4)	11 ± 9 (1–22)	10.6 ± 5.5 (2.6–15.2)	4.3 ± 2.2 (1–6)
New taxa				1.3 ± 1.6 (0–3)	0.5 ± 0.6 (0–1)
3	5	48.4 ± 7.5 (40.2–57.6)	348 ± 79 (266–450)	65.7 ± 5.8 (58.5–72.2)	31.2 ± 7.3 (24–39)
New taxa				39.8 ± 10.1 (24.4–52.3)	19 ± 6.2 (10–24)
4	18	39 ± 21.1 (4.2–88.6)	513 ± 851 (22–3624)	60.1 ± 12.5 (31.3–87.5)	21.8 ± 10.2 (5–38)
New taxa				28.7 ± 10.8 (12.5–53.1)	10.6 ± 5.8 (2–18)
5	8	17.2 ± 13.4 (1.3–41.6)	92 ± 73 (10–219)	48.9 ± 19.6 (17.4–82.4)	11.1 ± 4 (4–15)
New taxa				12.1 ± 9.3 (0–23.5)	2.6 ± 1.8 (0–5)
6	5	44.8 ± 24.3 (6.3–67.1)	850 ± 578 (85–1658)	70.1 ± 8.6 (60–81.8)	19.2 ± 15.5 (6–45)
New taxa				20 ± 8.3 (9.1–28.6)	6.2 ± 6.4 (1–17)
7	6	7.8 ± 11.4 (1.1–30.7)	89 ± 180 (6–457)	24.8 ± 16 (9.3–52.5)	8.3 ± 6.5 (4–21)
New taxa				4.6 ± 3.5 (0–10)	1.5 ± 1.4 (0–4)

and removed by the auditor (Table 1). The number of overlooked specimens varied between 1.9% and 48.4% (11 to 850 specimens). On average, 18 different taxa/sample (51%) were overlooked or not removed by the PA (Table 1), and 20.6% of the taxa found in the residue by the auditors were new to the sample, i.e., were found only by the auditors and were added to the taxon list for the sample. The number of new taxa varied between 0.5 (1.3%) and 19.0 (39.8%) per sample. The number of specimens of these new taxa ranged from 4.3 (10.6%) to 31.2 (65.7%) per sample.

The largest numbers of overlooked specimens in the sample residue belonged to the taxa Oligochaeta (Enchytraeidae 95.9%), Trichoptera (Limnephilidae 86.2%, *Hydropsyche* sp. 65.5%), Coleoptera (*Hydraena* sp. 69.9%, *Oulimnius* sp. 68.0%), and Diptera (Psychodidae 66.0%). The number of overlooked specimens was not correlated with the percentage of organic substrates at the sampling sites (Spearman rank correlation, $p > 0.05$). Thus, specimens were not more likely to be overlooked in samples from sites with higher proportions of depositional (e.g., fine sediments) or organic substrates (e.g., coarse organic matter or submerged aquatic vegetation) than in samples from clean substrates (e.g., sites dominated by gravel/cobble). No correlation was found between differences in the sorting audit and occurrence ($p = 0.21$) or abundance ($p = 0.21$) of taxa. However, the probability of being overlooked increased significantly with decreasing body size ($r = 0.61$, $p < 0.001$; Fig. 1).

Identification audit

Differences between taxon lists occurred when PAs and auditors came to different identification results. Differences also occurred when the auditors could not confirm identification of certain taxa, e.g., because the voucher specimens were not deposited by the PA or the voucher specimens lacked necessary structures for identification. On average, PA and auditor taxon lists differed by 9.3 taxa (33.8%). The number of differences between taxon lists varied by PA. The largest average number of taxon differences for a single PA
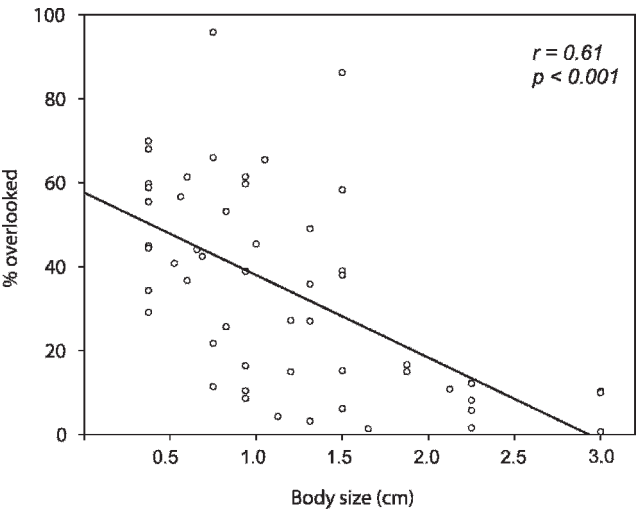

Fig. 1. Scatterplot for the percentage of overlooked specimens in sorting audit vs taxon body size.

TABLE 2. Results of the identification audit. Means (±1 SD) and ranges of relative and absolute differences in identification results of primary analysts and auditors. New taxa = number of those taxa that were found only by the auditors. Differences between taxon lists occurred when the primary analyst and auditors came to different results. Relative and absolute numbers are based on the taxon list after the sorting audit.

Primary analyst	N	Relative number (%)	Absolute number
Total	50	33.8 ± 16.8 (0–80)	9.3 ± 4.8 (0–20)
1	4	31.3 ± 15.8 (16.7–53.6)	11.8 ± 4.3 (6–15)
2	4	24.6 ± 3.5 (20.5–28.9)	9.8 ± 1.5 (8–11)
3	5	44.5 ± 20.7 (29–80)	13.8 ± 4.7 (9–20)
4	18	40.5 ± 16.8 (11.1–69.2)	10.2 ± 5.1 (1–20)
5	8	33.1 ± 13.2 (12.1–53.8)	7.3 ± 2.5 (4–11)
6	5	30.5 ± 18.5 (0–50)	7.2 ± 6.5 (0–14)
7	6	15.8 ± 4.9 (9.1–22.6)	5.2 ± 1.2 (4–7)

was 13.8 taxa (44.5%), whereas the smallest difference was 5.2 taxa (15.8%) (Table 2).

The largest number of differences in the identification audit involved *Hydropsyche* sp. (58.3%), followed by *Rhyacophila* sp. (47.4%), and *Baetis* sp. (40.4%) (Table 3). Among different taxonomic levels, 20.4% of the differences were at species level, 40.3% at genus level, and 39.4% at family level. The largest number of identification differences in higher taxonomic groups involved Trichoptera (17.6%), followed by Turbellaria (17.1%), Mollusca (15.2%), and Hirudinea (15.2%).

Of the taxa that were identified differently by the PA and the auditor, 14.7% belonged to identification category 1, 26.5% to category 2, 58.8% to category 3, and 0% to category 4. Taxa in category 4 generally were rare in the operational taxon list, and none of these taxa were found in ≥20% of all samples. However, mean differences in identification of all taxa did not differ significantly among categories (Kruskal–Wallis, $p = 0.97$; 19.1% in category 1, 13.1% in category 2, and 14.8% in category 3).

Effects on assessment results

Sorting and identification audits revealed substantial differences between PA and auditor results. We evaluated how these differences were reflected in scores for each of the 3 assessment modules (organic pollution, acidification, general degradation) and the resulting EQC by calculating and comparing assessment scores for the PA taxon list and the corresponding taxon lists obtained after the sorting and identification audits (Table 4).

Organic pollution module.—The sorting and identification audits yielded similar results for the organic pollution module. Average absolute differences in scores were 0.11 (in both the sorting and identification audits) and 0.12 (total audit). The metric is scaled from 1 to 4, so these differences reflect a relative change of 3.6% and 3.9%, respectively. On average,

sites scored worse based on auditor taxon lists than when based on PA results (Table 4). Samples were classified in different quality classes in 4 (8%), 3 (6%), and 5 (10%) cases in sorting, identification, and total audits, respectively. No clear tendency toward better or worse quality-class assignments was found for classification based on post-audit results.

Acidification module.—Only 2 of the 24 stream types in Germany are affected by acidification, but these types are quite common. Twenty-five samples of our audit required assessment for this module. Of those 25 samples, 5 (20%), 3 (12%), and 5 (20%) samples were classified in a different quality class after sorting, identification, and total audits, respectively (Table 4).

General degradation module.—The MMI values based on PA and auditor taxon lists were quite similar for several samples, but differed greatly for others (cf. samples 15 and 18 of PA No. 4; marked in Fig. 2). The mean difference in MMI scores based on PA and on auditor results was 0.03 (2.9%) in both the sorting and identification audits. This mean difference was 0.04 (3.8%) for the total audit (Table 4). The absolute difference in MMI values was positively correlated with the number of new taxa found by the auditors in the sorting audit across the entire data set (Spearman rank correlation, $r = 0.56$, $p < 0.01$) mainly because additional taxa increased the values of most of the metrics used in the MMI (for example number of EPTCBO taxa).

The differences in MMI values observed in each step of the audit caused some samples to score into a different degradation-module quality class. After the sorting, identification, and total audit, 8 (16%), 9 (18%), and 10 (20%) of 50 samples, respectively, scored into different quality classes. Most (6) of these 10 samples had better assessment results after the total audit (Table 4).

EQC.—The final EQC resulting from all 3 modules was calculated using a worst-case principle (Meier et

TABLE 3. Mean (±1 SD) percentages of all cases where individual taxa were missed (sorting audit) or determined differently (identification audit). Taxa are listed if they were present in ≥20% of all samples. Identification category represents the generally perceived difficulty in identifying a taxon based on Bayerisches Landesamt für Wasserwirtschaft (2004), where category 1 is easiest and category 4 is the most difficult (see text for definitions). Occurrence in sorting audit gives the number of samples in which the taxon was present. Abundance of a taxon is calculated as a mean of all samples. Importance is the relevance of a taxon in assessment and increases as the number of metrics in which it is used increases. Importance ranges from no (−) to high (+++) relevance. gr. = group, ad. = adult, lv. = larvae.

Taxonomic group	Family	Taxon	Differences in Sorting (%)	Differences in Identification (%)	Identification category	Occurrence	Abundance	Body size (cm)	Importance
Bivalvia	Sphaeriidae	*Pisidium* sp.	42.5 ± 46.0	18.2 ± 40.5	1	20	6	0.7	++
Coleoptera	Elmidae	*Elmis aenea* gr. ad.	55.5 ± 34.5	12.6 ± 33.8		28	29	0.4	+
Coleoptera	Elmidae	*Elmis* sp. lv.	45.0 ± 37.8	6.7 ± 25.4	2	34	30	0.4	++
Coleoptera	Elmidae	*Limnius perrisi* ad.	34.3 ± 45.7			10	20	0.4	+
Coleoptera	Elmidae	*Limnius* sp. lv.	29.1 ± 62.5	14.9 ± 34.8	2	23	18	0.4	++
Coleoptera	Elmidae	*Limnius volckmari* ad.	44.4 ± 50.3	38.5 ± 50.6	3	14	9	0.4	+++
Coleoptera	Elmidae	*Oulimnius* sp. lv.	68.0 ± 44.8	27.3 ± 46.7	2	10	6	0.4	+
Coleoptera	Elmidae	*Oulimnius tuberculatus* ad.	58.9 ± 37.8	0.0 ± 0.0	3	21	9	0.4	+
Coleoptera	Hydraenidae	*Hydraena gracilis* ad.	59.8 ± 44.1			14	8	0.4	++
Coleoptera	Hydraenidae	*Hydraena* sp. ad.	69.9 ± 44.4			16	6	0.4	++
Crustacea	Asellidae	*Asellus aquaticus*	15.3 ± 24.9	9.1 ± 29.4	3	24	30	1.5	+++
Crustacea	Gammaridae	*Gammarus fossarum*	5.8 ± 17.4	9.1 ± 29.4	3	21	84	2.3	+++
Crustacea	Gammaridae	*Gammarus pulex*	1.6 ± 7.3	4.2 ± 20.4	3	32	31	2.3	++
Crustacea	Gammaridae	*Gammarus roeselii*	8.2 ± 9.4	4.0 ± 20.0	1	24	163	2.3	+++
Crustacea	Gammaridae	*Gammarus* sp.	45.4 ± 45.2	30.4 ± 45.4	2	39	43	1.0	+
Diptera	Ceratopogonidae	Ceratopogoninae/Palpomyiinae	49.1 ± 44.5	7.1 ± 26.7		23	12	1.3	−
Diptera	Chironomidae	Chironomidae	53.1 ± 33.3	12.1 ± 30.1	1	46	102	0.8	−
Diptera	Chironomidae	Chironomini	27.2 ± 31.5	12.5 ± 33.6	3	32	57	1.2	−
Diptera	Chironomidae	*Prodiamesa olivacea*	25.8 ± 40.4	20.8 ± 41.5	3	25	24	0.8	+
Diptera	Chironomidae	Tanypodinae	16.4 ± 29.0	10.0 ± 30.5	3	37	20	0.9	−
Diptera	Chironomidae	Tanytarsini	61.4 ± 39.5	25.8 ± 44.5	3	38	31	0.6	+
Diptera	Limoniidae	Limoniidae	15.0 ± 33.7			10	6	1.2	−
Diptera	Pediciidae	*Dicranota* sp.	15.0 ± 26.3	9.1 ± 29.4	2	22	10	1.9	+
Diptera	Psychodidae	Psychodidae	66.0 ± 43.3			17	6	0.8	−
Diptera	Simuliidae	*Prosimulium* sp.	11.4 ± 13.8	0.0 ± 0.0	2	19	51	0.8	++
Diptera	Simuliidae	*Simulium* sp.	36.7 ± 36.0	20.3 ± 39.3	2	35	18	0.6	+
Ephemeroptera	Baetidae	*Baetis muticus*	59.7 ± 46.3			13	9	0.9	+++
Ephemeroptera	Baetidae	*Baetis rhodani*	8.6 ± 11.3	3.3 ± 18.3	3	31	105	0.9	+++
Ephemeroptera	Baetidae	*Baetis* sp.	61.4 ± 45.0	40.4 ± 49.7	3	34	62	0.9	+
Ephemeroptera	Ephemerellidae	*Serratella ignita*	21.7 ± 28.9	9.7 ± 28.6	2	11	68	0.8	+
Ephemeroptera	Ephemeridae	*Ephemera danica*	10.4 ± 19.4	0.0 ± 0.0	3	13	6	3.0	+++
Ephemeroptera	Heptageniidae	*Ecdyonurus venosus* gr.		26.7 ± 43.9	3				+++
Ephemeroptera	Heptageniidae	*Rhithrogena semicolorata* gr.	27.1 ± 28.6	6.7 ± 25.8	3	15	102	1.3	+++
Gastropoda	Hydrobiidae	*Potamopyrgus antipodarum*	40.8 ± 44.3			10	49	0.5	+++
Gastropoda	Planorbidae	*Ancylus fluviatilis*	38.9 ± 38.9	2.8 ± 11.8	1	24	7	0.9	+++
Hirudinea	Erpobdellidae	Erpobdellidae	58.3 ± 51.5			12	6	1.5	++
Hirudinea	Glossiphoniidae	*Glossiphonia complanata*	12.2 ± 29.9	7.7 ± 27.7	3	14	6	2.3	++

TABLE 3. Continued.

Taxonomic group	Family	Taxon	Differences in			Sorting audit			
			Sorting (%)	Identification (%)	Identification category	Occurrence	Abundance	Body size (cm)	Importance
Oligochaeta	[Kl:Oligochaeta]	Naididae/Tubificidae	56.7 ± 38.5	25.8 ± 44.5		36	57	0.6	+
Oligochaeta	Enchytraeidae	Enchytraeidae	95.9 ± 13.7			11	3	0.8	−
Oligochaeta	Lumbricidae	Eiseniella tetraedra	0.7 ± 2.5			13	12	3.0	+
Oligochaeta	Lumbriculidae	Lumbriculidae	38.0 ± 31.4			17	12	1.5	−
Plecoptera	Leuctridae	Leuctra sp.	10.4 ± 19.1			10	18	0.9	++
Plecoptera	Nemouridae	Nemoura sp.	44.0 ± 42.4	15.0 ± 33.7	3	13	12	0.7	++
Plecoptera	Taeniopterygidae	Brachyptera risi	3.2 ± 8.3			10	28	1.3	+++
Trichoptera	Hydropsychidae	Hydropsyche siltalai	1.4 ± 3.3	0.0 ± 0.0	2	21	28	1.7	++
Trichoptera	Hydropsychidae	Hydropsyche sp.	65.5 ± 44.0	58.3 ± 48.4	1	31	30	1.1	+
Trichoptera	Lepidostomatidae	Lasiocephala basalis	39.0 ± 44.6			10	6	1.5	+++
Trichoptera	Limnephilidae	Chaetopterygini/Stenophylacini sp.	16.7 ± 31.4	25.0 ± 45.2		15	12	1.9	−
Trichoptera	Limnephilidae	Halesus sp.	10.0 ± 31.6	18.2 ± 40.5	3	10	6	3.0	++
Trichoptera	Limnephilidae	Limnephilidae	86.2 ± 32.1			18	18	1.5	−
Trichoptera	Rhyacophilidae	Rhyacophila (Rhyacophila) sp.	35.9 ± 44.5	47.4 ± 48.1	3	19	24	1.3	+
Trichoptera	Rhyacophilidae	Rhyacophila sp.	4.4 ± 9.2			10	12	1.1	++
Trichoptera	Sericostomatidae	Sericostoma sp.	6.2 ± 15.7			14	7	1.5	+++
Turbellaria	Dugesiidae	Dugesia gonocephala	10.9 ± 23.8	0.0 ± 0.0	3	11	12	2.1	+++

Table 4. Absolute (Abs) and relative (Rel) differences in assessment scores based on primary analyst (PA) taxon lists and auditor taxon lists after a sorting audit, an identification audit, and a total audit. SD = standard deviation, deviation = number of positive/negative changes in assessment results, quality-class changes = number of improvements/deteriorations in quality class resulting from the differences in assessment results, MMI = multimetric index.

Audit step	Organic pollution N = 50		Acidification N = 25		MMI N = 50		Ecological quality class N = 50	
	Abs	Rel (%)	Abs	Rel (%)	Abs	Rel (%)	Abs	Rel (%)
Sorting audit								
Average	0.11	3.6	0.16	3.2	0.03	2.9	0.14	2.8
SD	0.14	4.8	0.37	7.5	0.03	3.4	0.35	7.0
Minimum	0.00	0.0	0.00	0.0	0.00	0.0	0.00	0.0
Maximum	0.74	24.7	1.00	20.0	0.13	13.0	1.00	20.0
Deviation	15/34	30/68	5/0	20/0	28/10	56/20	4/2	8/4
Quality-class changes	3/1	6/2	5/0	20/0	6/2	12/4	4/2	8/4
Identification audit								
Average	0.11	3.6	0.08	1.6	0.03	2.8	0.18	3.6
SD	0.12	4.0	0.28	5.5	0.03	3.0	0.39	7.8
Minimum	0.00	0.0	0.00	0.0	0.00	0.0	0.00	0.0
Maximum	0.51	17.0	1.00	20.0	0.13	13.0	1.00	20.0
Deviation	19/30	38/60	1/2	4/8	19/18	38/36	4/4	8/8
Quality-class changes	0/3	0/6	1/2	4/8	5/4	10/8	4/4	8/8
Total audit								
Average	0.12	3.9	0.16	3.2	0.04	3.8	0.18	3.6
SD	0.14	4.6	0.37	7.5	0.04	4.0	0.39	7.8
Minimum	0.00	0.0	0.00	0.0	0.00	0.0	0.00	0.0
Maximum	0.74	24.7	1.00	20.0	0.21	21.0	1.00	20.0
Deviation	15/34	30/68	4/1	16/4	26/12	52/24	4/4	8/8
Quality-class changes	2/3	4/6	4/1	16/4	6/4	12/8	4/4	8/8

al. 2006). EQC for 6 of 50 samples (12%) differed after the sorting audit. EQC of 8 samples (16%) differed after both the identification audit and the total audits (Table 4). Lower and higher EQCs were calculated equally as often after the identification and total audits.

Changes in the EQC arise from differences that affect all 3 modules. However, the EQC does not summarize the total number of differences in the 3 modules. A change in quality class in a particular module does not necessarily lead to a change in EQC because the EQC is derived from a worst-case principle. We counted and summarized all changes in the total audit of assessment results from each of the 3 modules to get a more detailed overview of the total number of changes in quality-class assignments for all assessment modules. Audits resulted in a change in quality class in ≥1 module for 17 of 50 samples (34%). This number is >2× the number of changes in EQC.

Two *Baetis* species, *Ephemera*, *Rhithrogena* (Ephemeroptera), 2 *Gammarus* species and *Assellus* (Crustacea), as well as taxa of other taxonomic groups (e.g., *Leuctra* [Plecoptera], *Dugesia* [Tricladia], and *Sericostoma* [Trichoptera]) had the greatest influence on changing the EQC (Table 3). This is mainly because these taxa were used in more metrics than other species and thus had a greater impact.

Discussion

Sorting audit

The sorting protocol required PAs to remove and sort all specimens from a sample. We assumed that sorting an entire sample under laboratory conditions would be a relatively simple task that should generally lead to nearly complete removal of specimens from a sample. However, the audit revealed that close to ⅓ of all specimens and ⅕ of all taxa were either overlooked or ignored and remained in the sample residue. Small and slender taxa, such as Elmidae, were particularly affected (Table 3, Fig. 1). Other sorting techniques, such as live-sorting in the field and all sorting methods based on the estimation of the number of specimens, should have an even higher potential for error (Haase et al. 2004b). In an audit of macroinvertebrate samples done within the EU-funded STAR project (Furse et al. 2006), sorting error was <½ for STAR/AQEM samples than for samples from other large assessment programs, such as River Invertebrate Prediction and Classification System (RIVPACS; Wright et al. 2000)
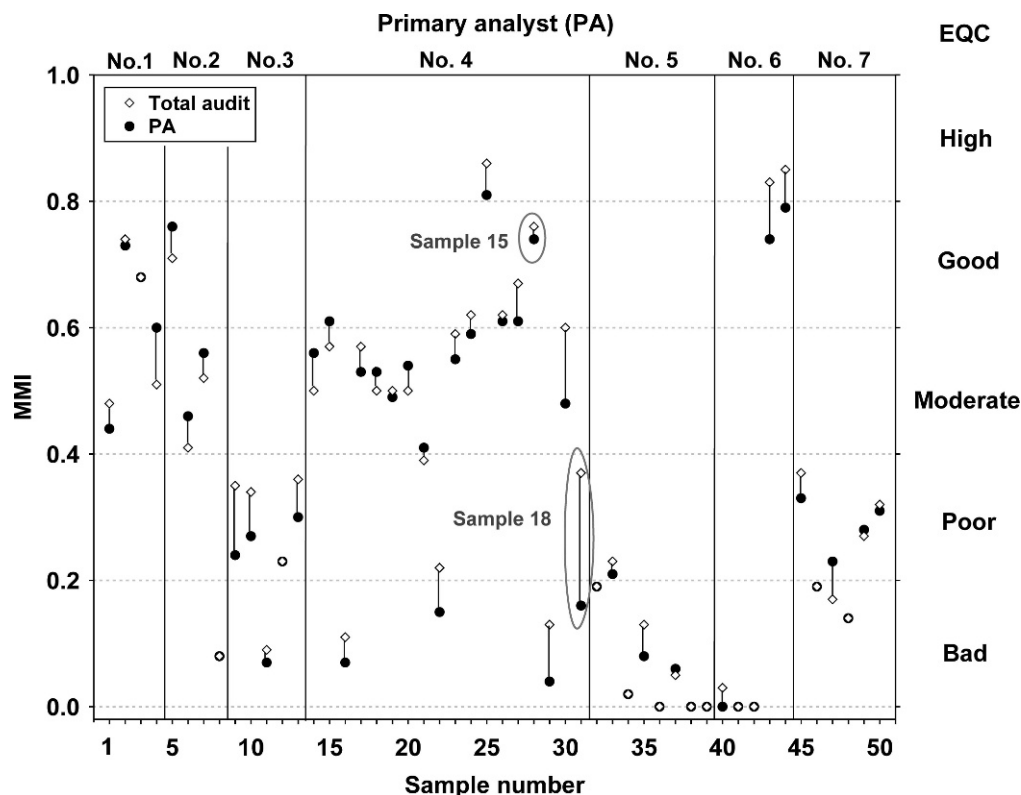
Fig. 2.   Assessment results (multimetric index [MMI] values) based on primary analyst (PA) taxon lists and auditor taxon lists after the total audit. Both taxon lists for each sample are joined by vertical lines. If only an open circle is present, the closed circle is invisible behind it (identical MMI values). Dotted horizontal lines indicate Ecological Quality Class (EQC) boundaries. Vertical lines separate results for each PA.

(Haase et al. 2006a). STAR/AQEM samples are sorted in the laboratory following a similar procedure to the one applied in our study. RIVPACS also uses a laboratory sorting technique, but allows estimation of number of individuals in abundant taxa. Thus, sorting error in more complex or less standardized sorting procedures or protocols allowing for an estimation of number of individuals, such as live sorting, is greater than in simple laboratory sorting procedures. If the putatively most robust sorting technique (laboratory sorting) leads to error rates as high as those observed in our audit, then error rates for more complex sorting techniques are likely to be even higher. Sorting error seems to be much more likely in sample processing and more likely to affect assessment outcomes than expected. Therefore, we recommend using only laboratory sorting techniques and carefully trained personnel in important surveys.

Identification audit

The mean number of differences between PA and auditor taxon lists was 33.8% of the recorded taxa in

our audit. Thus, identification results differed for ~⅓ of our records. Stribling et al. (2008) found identification differences in 22.1% of the taxa investigated. Results of both studies demonstrate the risk of inaccurate identification in sample processing. Moreover, PAs might have chosen the least ambiguous individuals of a species as voucher specimens. Thus, the results of our identification audit might have been biased toward a smaller error rate. It is reasonable to assume that a fully quantitative identification audit based on all individuals of a sample would lead to more frequent and even greater differences in assessment results.

In both studies, differences in identification (and sorting) were related to the taxonomic unit (order, family) to which specimens belonged. However, comparing differences in identification (and sorting) among higher taxonomic units is difficult because these units are composed of species that differ greatly in attributes like body size, difficulty of identification, or ecological preference. We expected greater differences in identification for taxa known to be difficult to identify. However, differences were independent of

identification difficulty category. This result could indicate that operators were unconsciously aware of the different identification difficulty categories and paid less attention to taxa that are easier to identify. Thus, identification error might depend less on how difficult a taxon is to identify than on how much attention an operator pays to identification of different taxa. These results correspond to the results of the sorting audit where operator attention to detail affected the error rate. Differences in identification could be caused by limited attention to detail, deficits in taxonomic expertise, or both. Identification courses and workshops are critically needed to help overcome these problems. Those courses should, thus, not focus on "difficult" taxa, but rather on common and higher-level taxa and should promote operator awareness of attention to detail when working with easier-to-identify taxa.

Our identification audit assessed only qualitative differences because the PAs were asked to deposit ≥5 individuals/taxon as voucher specimens. Restricting the number of voucher specimens is a common practice mandated by German water authorities to save time and money. In light of our results, we think it reasonable to consider changing the procedure to at least retain all voucher specimens combined in 1 vial and the remainder of the individuals from a sample in a separate vial. This change would enable better, quantitative quality assurance in the future with a minimal increase in overall costs for routine sampling.

Effects on assessment results

The EQC of 6 monitoring sites (12%) changed in response to the sorting audit, and the EQC of 8 monitoring sites (16%) changed in response to the identification audit. However, the EQC of only 8 monitoring sites (16%) changed after the total audit. The effects of sorting and identification differences were not cumulative because the EQC of only 2 of the 50 samples used in the audit changed in both the sorting and the identification audit.

About every 3rd monitoring site (34%) scored into a different quality class in 1 to 3 of the assessment modules (organic pollution, acidification, and general degradation) after the total audit. This relatively high number of reclassifications clearly demonstrates the effect of human sorting and identification differences on assessment results. However, the effects of sorting, identification, and total audit on scoring did not differ among assessment modules (Wilcoxon, all $p > 0.25$), indicating no difference in sensitivity of the modules to error.

The general degradation module is the most important module in the German assessment system because scores for ~90% of German rivers are lowest in this module. Thus, based on the worst-case principle, scores in the general degradation module are responsible for 90% of the final EQC assessments in Germany. Difference rates in the sorting audit were high for several taxa that have high relevance in this module (indicated by ++ and +++ in Table 3), but difference rates in the identification audit were high for only a few of them (Table 3). However, the relative and total number of changes in EQC was slightly higher after the identification audit than after the sorting audit. Thus, taxa that are generally not expected to influence assessments (e.g., *Gammarus* sp. or *Baetis* sp.) must have contributed to post-audit changes in the EQC. This effect could arise if noninfluential taxa were very abundant and differences in identification led to changes in a relatively large proportion of samples.

Across the 3 modules and the final EQC, the total number of post-audit quality-class changes did not differ significantly among sorting (18 better/5 worse), identification (10/13), and total (16/12) audits (Table 4). However, additional specimens and taxa found in the sorting audit led to more positive than negative changes, whereas positive and negative changes were equally common in the identification and total audits. Thus, overlooking or ignoring specimens during sorting led to underestimation of the EQC.

The EUWFD requires use of the worst-case principle when calculating the final EQC. We compared the effect of sorting and identification audits on this worst-case approach to effects on an approach based on a mean score calculated across all 3 assessment modules. The mean-score approach leads to better assessment results than the worst-case approach, but the effect of the 3 audits did not differ significantly between the 2 approaches (Wilcoxon, all $p > 0.46$).

The value of audits and quality control

To our knowledge, the UK is the only country with a routine auditing scheme for macroinvertebrate biomonitoring samples. In the 1st year after implementation of the UK auditing scheme, differences between PAs and auditors were large (Dines and Murray-Bligh 2000, Murray-Bligh et al. 2005), probably because most investigators were convinced that they worked accurately and, therefore, failed in self-evaluation. However, in the 2nd year after implementing the UK auditing scheme, the differences between PAs and auditors decreased rapidly. This quick response is probably attributable to increased awareness of operators that sample processing was more prone to error than previously assumed.

In our study, the differences in sorting and identification between PAs and auditors equally affected assessment results. Sample sorting is conceptually very simple, and one might reasonably expect the task to be left to the most junior and inexperienced biologists. However, our audit results and the results from Haase et al. (2006a) demonstrate that sorting is, in fact, a task that requires more skill than has been recognized in the past. Our results demonstrate the need for formal training of sorting personnel and for improvement of internal laboratory quality-control procedures.

Macroinvertebrate specimen identification is generally regarded to be more difficult than sample sorting. However, carelessness and lack of formally trained and experienced staff are presumably the predominant sources of error in both sorting and identification. Causes of human errors include incorrect interpretation of technical literature; transcription or recording errors; coarse definitions of terminology, nomenclature, and standard procedures; differences in optical equipment; and sample handling and preparation techniques (Stribling et al. 2003, 2008, Dalcin 2004, Chapman 2005). Unfortunately, biologists often receive little formal training in sample sorting, specimen identification, or other steps in sample processing. Thus, the large differences between PAs and auditors in the identification audit are not surprising. Our audit results clearly indicate that formal training is of utmost importance in macroinvertebrate sample processing and could presumably reduce assessment errors. Basic taxonomic skills are being taught less and less in tertiary education programs around the globe (Holzenthal et al. 2010). This trend has been acknowledged in the benthological community. Efforts like the North American Benthological Society's Taxonomist Certification Program (www.nabstcp.com/) provide professional taxonomic certification to ensure high-quality taxonomy and, thus, credible ecological and reliable bioassessment studies, and support graduate training of next generation taxonomic experts. Such efforts should be promoted even more strongly, perhaps even mandated, for operators of official bioassessments.

Our data also show that human error in sample processing has a great effect on assessment results. Implementation of a quality-management system is vital to overcome these shortcomings. At minimum, such a system must incorporate standardized methods and protocols, formal training procedures and evaluations, and an auditing scheme for randomly selected samples. Experience in the UK clearly shows that implementing a routine audit scheme has a strong positive effect on data quality. Therefore, quality control will increase the accuracy and precision of biological data and strengthen confidence of water managers in assessment results. This confidence is of utmost importance because assessment results are used to guide and direct cost-intensive mitigation measures in river rehabilitation. According to BMU (2005), 60% of German streams and rivers will fail the good or high ecological status required by the EUWFD. An additional 26% of streams and rivers are at risk of failing to meet this status. If we estimate a cost of 400,000 Euro/km of river length for morphological improvements (Interwies et al. 2004), restoration will cost billions of Euros in Germany alone. An error rate of 16% (as measured in our study) would lead to inefficient allocation of several million Euros. The costs of misplaced remedial actions will outweigh costs caused by the implementation of quality control by several orders of magnitude. Thus, a quality-control system should be implemented in freshwater monitoring programs to help avoid the high costs of unnecessary or incorrectly guided restoration measures based on inaccurate assessment results.

Conclusions

Our results are already reaching water managers in Germany and are causing them to rethink current and past practices. They realize that human error is more prominent than previously thought and have started to develop a quality-control system. This new course of action is encouraged by the positive effect of quality-control application in marine monitoring programs (Ranasinghe et al. 2003). The UK, German, and marine examples should encourage water managers in other countries to establish quality-management systems where they are presently not in place. Furthermore, we assume that high error rates in sorting and identification of invertebrate samples is a general pattern that is not restricted to riverine systems but also occurs during processing of samples from lakes and terrestrial ecosystems (e.g., soil fauna). The large number of individuals or taxa in invertebrate samples makes invertebrate-based monitoring more prone to human error than monitoring based on other groups of organisms (e.g., vertebrates or higher plants). We still know little about the effect of human error on monitoring outcomes, but we do know that human error is a serious problem that affects applied science and any kind of ecological research dealing with invertebrate taxon lists. Therefore, implementation of quality-management systems is vital for both applied and basic ecology.

Acknowledgements

Literature Cited

BAYERISCHES LANDESAMT FÜR WASSERWIRTSCHAFT. 2004. Qualitätssicherung in der biologischen Gewässeranalyse - Bestimmungsliteratur und Bestimmbarkeitskategorien. Materialien Nr. 112. Bayerisches Landesamt für Wasserwirtschaft, Munich, Germany. (Available from: http://www.bestellen.bayern.de)

BMU (BUNDESMINISTERIUM FÜR UMWELT, NATURSCHUTZ UND REAKTORSICHERHEIT). 2005. Die Wasserrahmenrichtlinie – Ergebnisse der Bestandsaufnahme 2004 in Deutschland. Bonifatius, Paderborn, Germany. (Available from: http://www.umweltdaten.de/publikationen/fpdf-l/3043.pdf)

BÖHMER, J., C. RAWER-JOST, A. ZENKER, C. MEIER, C. FELD, R. BISS, AND D. HERING. 2004. Assessing streams in Germany with benthic invertebrates: development of a multimetric invertebrate based assessment system for German rivers. Limnologica 34:416–432.

BRAUKMANN, U., AND R. BISS. 2004. Conceptual study: an improved method to assess acidification in German streams by using benthic macroinvertebrates. Limnologica 34:433–450.

CAO, Y., C. P. HAWKINS, AND M. R. VINSON. 2003. Measuring and controlling data quality in biological assemblage surveys with special reference to stream benthic macroinvertebrates. Freshwater Biology 48:1898–1911.

CHAPMAN, A. D. 2005. Principles of data quality. Version 1.0. Report for the Global Biodiversity Information Facility (GBIF). Global Biodiversity Information Facility, Copenhagen, Denmark. (Available from: http://www2.gbif.org/DataQuality.pdf)

CHEVENET, F., S. DOLÉDEC, AND D. CHESSEL. 1994. A fuzzy coding approach for the analysis of long-term ecological data. Freshwater Biology 31:295–309.

CLARKE, R. T., AND D. HERING. 2006. Errors and uncertainty in bioassessment methods: major results and conclusions from the STAR project and their application using STARBUGS. Hydrobiologia 566:433–439.

CLARKE, R. T., A. LORENZ, L. SANDIN, A. SCHMIDT-KLOIBER, J. STRACKBEIN, N. T. KNEEBONE, AND P. HAASE. 2006. Effects of sampling and sub-sampling variation using the STAR-AQEM sampling protocol on the precision of macroinvertebrate metrics. Hydrobiologia 566:441–459.

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.

DALCIN, E. C. 2004. Data quality concepts and techniques applied to taxonomic databases. PhD Thesis, University of Southampton, Southampton, UK. (Available from: http://www.dalcin.org/eduardo/downloads/edalcin_thesis_submission.pdf)

DINES, R. A., AND J. A. D. MURRAY-BLIGH. 2000. Quality assurance and RIVPACS. Pages 71–78 in J. F. Wright, D. W. Sutcliffe, and M. T. Furse (editors). Assessing the biological quality of fresh waters: RIVPACS and other techniques. Freshwater Biological Association, Ambleside, UK.

DOWNES, B. J., P. S. LAKE, AND E. S. G. SCHREIBER. 1993. Spatial variation in the distribution of stream invertebrates: implication of patchiness for models of community organisation. Freshwater Biology 30:119–132.

EU COMMISSION. 2007. Towards sustainable water management in the European Union. First stage in the implementation of the Water Framework Directive 2000/60/EC. Commission staff working document. Accompanying document to the communication forum from the commission to the European Parliament and the council. COM (2007) 128 final. Commission to the European Communities, Brussels, Belgium. (Available from: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2007:0128:FIN:EN:PDF)

EUROPEAN UNION. 2000. Directive 2000/60/EC. Establishing a framework for community action in the field of water policy. European Commission PE-Cons 3639/1/100 Rev 1. Commission to the European Communities, Luxembourg City, Luxembourg.

FRIBERG, N., L. SANDIN, M. FURSE, S. E. LARSEN, R. CLARKE, AND P. HAASE. 2006. Comparison of macroinvertebrate sampling methods in Europe. Hydrobiologia 566: 365–378.

FURSE, M., D. HERING, O. MOOG, P. VERDONSCHOT, R. K. JOHNSON, K. BRABEC, K. GRITZALIS, A. BUFFAGNI, P. PINTO, N. FRIBERG, J. MURRAY-BLIGH, J. KOKES, R. ALBER, P. USSEGLIO-POLATERA, P. HAASE, R. SWEETING, B. BIS, K. SZOSZKIEWICZ, H. SOSZKA, G. SPRINGE, F. SPORKA, AND I. KRNO. 2006. The STAR project: context, objectives and approaches. Hydrobiologia 566:3–29.

HAASE, P., S. LOHSE, S. PAULS, K. SCHINDEHÜTTE, A. SUNDERMANN, P. ROLAUFFS, AND D. HERING. 2004a. Assessing streams in Germany with benthic invertebrates: development of a practical standardised protocol for macroinvertebrate sampling and sorting. Limnologica 34: 349–365.

HAASE, P., J. MURRAY-BLIGH, S. LOHSE, S. PAULS, A. SUNDERMANN, R. GUNN, AND R. T. CLARKE. 2006a. Assessing the impact of errors in sorting and identifying macroinvertebrate samples. Hydrobiologia 566:505–521.

HAASE, P., S. U. PAULS, C. H. M. ENGELHARDT, AND A. SUNDERMANN. 2008. Effects of sampling microhabitats with low coverage within the STAR/AQEM macroinvertebrate sampling protocol on stream assessment. Limnologica 38:14–22.

HAASE, P., S. PAULS, A. SUNDERMANN, AND A. ZENKER. 2004b. Testing different sorting techniques in macroinvertebrate samples from running waters. Limnologica 34: 366–378.

HAASE, P., A. SUNDERMANN, AND K. SCHINDEHÜTTE. 2006b. Informationstext zur Operationellen Taxaliste als Mindestanforderung an die Bestimmung von Makrozoobenthosproben aus Fließgewässern zur Umsetzung der EU-Wasserrahmenrichtlinie in Deutschland. Fliessgewaesserbewertung, Essen, Germany. (Available from: www.fliessgewaesserbewertung.de/download/bestimmung/)

HOLZENTHAL, R. W., D. R. ROBERTSON, S. U. PAULS, AND P. K. MENDEZ. 2010. Taxonomy and systematics: contributions to benthology and future directions. Journal of the North American Benthological Society 29:147–169.

INTERWIES, E., A. KRAEMER, N. KRANZ, B. GÖRLACH, T. DWORAK, D. BORCHARDT, S. RICHTER, AND J. WILLECKE. 2004. Handbook for the selection of cost-effective measures according to Art. 11 Water Framework Directive. Report Federal Environmental Agency Germany 02/04. Federal Environmental Agency, Dessau-Roßlau, Germany. (Available from: http://www.umweltbundesamt.de/index-e.htm)

KARR, J. R., AND E. W. CHU. 2000. Sustaining living rivers. Hydrobiologia 422/423:1–14.

LI, J., A. HERLIHY, W. GERTH, P. KAUFMANN, S. GREGORY, S. URQUHART, AND D. P. LARSEN. 2001. Variability in stream macroinvertebrates at multiple spatial scales. Freshwater Biology 46:87–97.

MEIER, C., J. BÖHMER, R. BISS, C. FELD, P. HAASE, A. LORENZ, C. RAWER-JOST, P. ROLAUFFS, K. SCHINDEHÜTTE, F. SCHÖLL, A. SUNDERMANN, A. ZENKER, AND D. HERING. 2006. Weiterentwicklung und Anpassung des nationalen Bewertungssystems für Makrozoobenthos an neue internationale Vorgaben. Fliessgewaesserbewertung, Essen, Germany. (Available from: www.fliessgewaesserbewertung.de/download/berechnung/)

MURRAY-BLIGH, J., J. VAN DER MOLEN, AND P. VERDONSCHOT. 2005. STAR deliverable No. 7: audit of performance incorporating results of the La Bresse sampling and analysis workshop. Unpublished report to the Standardisation of River Classifications Project, Commission to the European Communities, Brussels, Belgium. (Available from: http://www.eu-star.at/pdf/Deliverable7.pdf)

NICHOLS, S. J., W. A. ROBINSON, AND R. H. NORRIS. 2006. Sample variability influences on the precision of predictive bioassessment. Hydrobiologia 572:215–233.

OLSEN, D. A., C. D. MATTHAEI, AND C. R. TOWNSEND. 2007. Patch history, invertebrate patch dynamics and heterogeneous community composition: perspectives from a manipulative stream experiment. Marine and Freshwater Research 58:307–314.

OSTERMILLER, J. D., AND C. P. HAWKINS. 2004. Effects of sampling error on bioassessments of stream ecosystems: application to RIVPACS-type models. Journal of the North American Benthological Society 23:363–382.

PRINGLE, C. M., R. J. NAIMAN, G. BRETSCHKO, J. R. KARR, M. W. OSWOOD, J. R. WEBSTER, R. L. WELCOMME, AND M. J. WINTERBOURN. 1988. Patch dynamics in lotic systems: the stream as a mosaic. Journal of the North American Benthological Society 7:503–524.

RANASINGHE, J. A., D. E. MONTAGNE, S. B. WEISBERG, M. BERGEN, AND R. G. VELARDE. 2003. Variability in the identification and enumeration of marine benthic invertebrate samples and its effect on benthic assessment measures. Environmental Monitoring and Assessment 81:199–206.

SANDIN, L., J. DAHL, AND R. K. JOHNSON. 2004. Assessing acid stress in Swedish boreal and alpine streams using benthic macroinvertebrates. Hydrobiologia 516:129–148.

SMITH, M. J., W. R. KAY, D. H. D. EDWARD, P. J. PAPAS, K. S. RICHARDSON, J. C. SIMPSON, A. M. PINDER, D. J. GALE, P. H. J. HORWITZ, J. A. DAVIS, F. H. YUNG, R. H. NORRIS, AND S. A. HALSE. 1999. AusRivAS: using macroinvertebrates to assess ecological condition of rivers in Western Australia. Freshwater Biology 41:269–282.

STRIBLING, J. B., S. R. MOULTON, AND G. L. LESTER. 2003. Determining the quality of taxonomic data. Journal of the North American Benthological Society 22:621–631.

STRIBLING, J. B., K. L. PAVLIK, S. M. HOLDSWORTH, AND E. W. LEPPO. 2008. Data quality, performance, and uncertainty in taxonomic identification for biological assessments. Journal of the North American Benthological Society 27:906–919.

SUNDERMANN, A., S. U. PAULS, R. T. CLARKE, AND P. HAASE. 2008. Within-stream variability of benthic invertebrate samples and EU Water Framework Directive assessment results. Fundamental and Applied Limnology 173:21–34.

TOWNSEND, C. R. 1989. The patch dynamics concept of stream community ecology. Journal of the North American Benthological Society 8:36–50.

USSEGLIO-POLATERA, P. 1991. Représentation graphique synthétique de la signification écologique d'un peuplement. Application aux macroinvertébrés du Rhône à Lyon. Bulletin d'Ecologie 22:195–202.

VLEK, H. E., P. F. M. VERDONSCHOT, AND R. NIJBOER. 2004. Towards a multimetric index for the assessment of Dutch streams using benthic macroinvertebrates. Hydrobiologia 516:173–189.

WRIGHT, J. F., D. MOSS, P. D. ARMITAGE, AND M. T. FURSE. 1984. A preliminary classification of running water sites in Great Britain based on macro-invertebrate species and the prediction of community type using environmental data. Freshwater Biology 14:221–256.

WRIGHT, J. F., D. W. SUTCLIFFE, AND M. T. FURSE. 2000. Assessing the biological quality of freshwaters: RIVPACS and similar techniques. Freshwater Biological Association, Ambleside, UK.