

Heterogeneity in macroinvertebrate sampling strategy introduces variability in community characterization and stream trait-based biomonitoring: Influence of sampling effort and habitat selection criteria

Floriane Larras, Philippe Usseglio-Polatera

▶ To cite this version:

Floriane Larras, Philippe Usseglio-Polatera. Heterogeneity in macroinvertebrate sampling strategy introduces variability in community characterization and stream trait-based biomonitoring: Influence of sampling effort and habitat selection criteria. Ecological Indicators, 2020, 119, pp.106758. 10.1016/j.ecolind.2020.106758 . hal-03100945

HAL Id: hal-03100945 https://hal.univ-lorraine.fr/hal-03100945

Submitted on 30 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 Heterogeneity in macroinvertebrate sampling strategy introduces variability in

- community characterization and stream trait-based biomonitoring: influence of
 sampling effort and habitat selection criteria
- 4 5 Floriane Larras¹ & Philippe Usseglio-Polatera^{2*} 6 7 ¹ Helmholtz-Centre for Environmental Research UFZ, Department of Bioanalytical Ecotoxicology, 8 Permoserstrasse. 15, 04318 Leipzig (Germany) 9 10 ² Université de Lorraine, CNRS, UMR 7360, LIEC, Laboratoire Interdisciplinaire des Environnements 11 Continentaux, 57070 Metz (France) 12 13 *Corresponding author: philippe.usseglio-polatera@univ-lorraine.fr 14 Université de Lorraine, CNRS, UMR 7360, 15 16 Laboratoire Interdisciplinaire des Environnements Continentaux, 17 57070 Metz (France) 18
- 19

20 Abstract

21

22 Benthic macroinvertebrate assemblages have been of major importance in the biomonitoring of rivers 23 for many decades. To more closely fit with new regulatory purposes, standardized methods of 24 macroinvertebrate field sampling have changed over time both in terms of sampling effort and criteria of habitat type sampling priority (e.g. according to a decreasing gradient of "bottom surface coverage" 25 26 versus "substratum favorability"). Such methodological heterogeneity has rendered the exploitation of 27 biomonitoring data far more complex because they are indeed not fully comparable over time and may 28 induce bias in long term series studies. In this context, we assessed the influence of variation in field 29 sampling strategy on (i) the taxonomical characterization of benthic macroinvertebrate assemblages 30 (297 taxon abundances, taxonomic richness and diversity, EPT-related metrics), (ii) the response of 31 sensitivity descriptors (ASPT, SPEAR), (iii) the bio-ecological trait description of assemblages and (iv) the 32 resulting ecological diagnostic of river reaches, based on 10074 sampling events across France. Six 33 sampling strategies were compared, corresponding to three levels of sampling effort (4, 8 and 12 34 individual samples) and different criteria for substratum selection; e.g. marginal (= less than 5% coverage 35 at reach scale) vs. dominant; hosting capacity (= substratum favorability) vs. bottom coverage. A strong 36 influence of the sampling strategy was observed on the characterization of the macroinvertebrate 37 assemblages. Marginal substrata revealed a unique picture of the community at the reach scale and 38 presented a higher taxonomic diversity and richness than dominant ones, as well as a higher number of 39 sensitive EPT and SPEAR taxa. Both strategies based on dominant substratum sampling provided similar 40 pictures of the taxonomical and bio-ecological characteristics of macroinvertebrate assemblages. The 41 high hosting capacity of the most frequent marginal substrata has been identified as an important driver 42 of the observed responses. When using a tool calculating the probability that sampled 43 macroinvertebrate assemblages were impaired by several pressures related to water quality 44 degradation, hydromorphological alteration or land-use, differences in sampling strategy affected the 45 probability estimation. As a main trend, the results were especially reliable when they were obtained 46 from an 8-sample strategy, even more if they included the information provided by the marginal 47 substrata. If sampling effort definitely matters for environmental diagnostic, substratum selection 48 criteria play also a crucial role on (i) the description of local assemblages and (ii) the reliability of 49 calculated impairment risks, due to the strong relationship between stenotopic taxa and specific habitat 50 conditions.

51

52 Keywords: Assemblage structure; Bio-ecological traits; Habitat; Effort; Anthropogenic pressures; Risk 53 assessment

54

Abbreviations: AQEM: Assessment System for the Ecological Quality of Streams and Rivers throughout
 Europe using Benthic Macroinvertebrates; IBGN: Normalized Invertebrate-Based Index; I2M2:
 Multimetric Invertebrate Index; ASPT: Average Species per Taxon; SPEAR: SPEcies At Risk; ODI:
 Diagnostic tool based on macroinvertebrate metrics

59

60

61

63 1. Introduction

64

65 World-wide intensification of the use of both land and aquatic resources has highlighted the importance 66 of managing water quality as well as the critical role of innovative diagnostic tools for defining robust 67 management responses to aquatic ecosystem degradation (Davis et al., 2010; Dolédec & Statzner, 2010; 68 Hering et al., 2015; Navarro-Ortega et al., 2015; Sokolova, 2013; Van der Laan et al., 2013). Several 69 directives, including the EU Water Framework Directive (European Commission, 2000) or the Clean 70 Water Act (1977), aim at restoring and maintaining the integrity of water bodies. They have triggered the 71 development of many new bioassessment methods, which evaluate the ecological status of water 72 bodies, for example by comparing observed assemblages of a given Biological Quality Element (e.g., 73 fishes, macroinvertebrates or diatoms) to reference ones.

74 Benthic macroinvertebrates have been probably the most widely used group for the biological 75 assessment of human impacts in freshwater ecosystems (e.g., 27% of the 297 European methods listed 76 in Birk et al., 2012). Many of the macroinvertebrate-based assessment methods aggregate a variable 77 number of metrics (mean = 4 on 28 European countries, in Birk et al., 2012) describing the taxonomic or 78 trait structure of assemblages as well as the sensitivity of taxa. These multimetric indices are far more 79 consistent in the detection of reach impairment than single metrics (Hering et al., 2006; Urbanič, 2014). 80 For example, in France, the Multimetric Invertebrate Index (I₂M₂; Mondy et al., 2012) has improved the 81 detection of impaired river reaches by at least 17% for nitrogen compounds and up to 35% for organic 82 micropollutants and clogging risk, when compared to the previously normalized macroinvertebratebased index (IBGN; AFNOR, 1992, 2004). 83

Efficiently and robustly evaluating the global effect of anthropogenic pressures on river reach communities requires to aggregate a combination of generalist metrics (i.e., capable of efficiently detect a wide range of pressure categories), which are much more robust than specialist ones in stream ecological diagnostic especially when anthropogenic pressures interact (Dedieu et al., 2016; Mondy et

88 al., 2012; Touron-Poncet et al., 2014). As a result, new multimetric indices often provide a pertinent 89 evaluation of the global effect of anthropogenic pressures, but poorly feed the diagnostic in precise 90 information on the exact nature of the pressures really impairing local communities (Angradi et al., 2009; 91 Mondy and Usseglio-Polatera, 2013). Many authors have thus underlined the need for cost-effective 92 diagnostic tools, capable of disentangling the complex effects of interacting pressures (Brucet et al., 93 2013; Christensen et al., 2006; Darling & Côté, 2008; Downes, 2010). In this context, we have developed 94 a series of models individually able to estimate the risk of stream impairment by a given category of 95 anthropogenic pressure using trait-based and/or taxonomy-based characteristics of macroinvertebrate 96 assemblages (tool called in this study ODI=diagnostic tool based on macroinvertebrates metrics; Mondy 97 and Usseglio-Polatera, 2013). Several of these models have given good impairment risk assessment for 98 various pressures related to water quality (e.g., organic matter, nitrates, other nitrogen compounds, 99 phosphorous components, pesticides, metals) or hydromorphology and land-use alterations (e.g., 100 riverine vegetation alteration, urbanization at the reach scale, clogging risk or hydrological instability), 101 even at moderate risk level and in a multi-pressure context (Desrosiers et al., 2019; Mondy & Usseglio-102 Polatera, 2013). Existing biotic indices and diagnostic tools have been calibrated on a precise, 103 standardized sampling strategy. Variation in sampling can then induce detrimental effects on the 104 precision of the individual metrics used in indices and tools to assess the ecological status and provide an 105 ecological diagnostic of rivers (Clarke et al., 2006; Clarke & Hering, 2006; Wiederkehr et al., 2016). Such 106 variation can affect sampling effort or substratum selection.

107 Many studies, grounded upon the hypothesis that the habitat is a flexible templet for the selection of 108 ecological strategies (Southwood, 1977; Southwood, 1988; Townsend & Hildrew, 1994), have indeed 109 demonstrated that macroinvertebrate assemblages exhibited quite different taxonomic composition 110 (Beisel et al., 1998a, 2000) or trait structure (Demars et al., 2012) at the mesohabitat (*sensu* Armitage & 111 Pardo, 1995) or biotope (*sensu* Demars et al., 2012) scale, as a result of contrasted local heterogeneity, habitat stability and trophic potentialities (e.g., see fig. 5 in Beisel et al., 1998a). However it is rather difficult to estimate the potential impact of sampling variation in local habitats on the ecological diagnostic of rivers over a large set of sites corresponding to a wide range of environmental conditions (e.g. Strahler ranks from 1 to 8) and anthropogenic pressures, especially when using diagnostic tools simultaneously working with several hundred of metrics (e.g., 216 for the ODI).

117 Examining how metrics, biotic indices and diagnostic tools respond to variation in sampling is also 118 important in the context of long-term ecological surveys, which are often suffering from abrupt 119 modifications in standardized sampling methods. In Europe, the WFD implementation has contributed to 120 variations in sampling methodologies over time in several countries, hampering the compilation of 121 common and homogeneous datasets to objectively analyze pressure-impact relationships over long-term 122 time series (Blocksom & Flotemersch, 2005; Cao & Hawkins, 2011; Tronstad & Hotaling, 2017). For 123 example, in France, the IBGN protocol was used at the National scale over the 1992-2005 period (norm 124 NF T 90-350 in AFNOR, 2004). It was based on eight sample units selected according to a decreasing 125 gradient of hosting capacity of predefined habitat types; *i.e.* distinct units of habitat within the stream, 126 easily recognizable from the bank and which are unique combinations of "substratum type x current 127 velocity range", hosting a specific faunal assemblage (Beisel et al., 1998a). Due to severe inconsistencies 128 of the IBGN with the WFD, a new multimetric index (I_2M_2) based on a new field sampling protocol (see 129 section 2.1) with a different sampling effort and habitat selection strategy (AFNOR, 2009, 2010) has been 130 defined for the macroinvertebrate-based ecological assessment of French shallow streams (Mondy et al., 131 2012).

The implementation of this new sampling strategy, gathering three sets of samples performed with different strategies of substratum selection, on about 2,000 sites of National routine survey networks, has offered the unique opportunity to test the impact of variation in macroinvertebrate sampling effort and substratum selection strategy on the results of metrics, indices and a diagnostic tool developed for

136 shallow rivers, at large spatial scale. Based on a large set of data, we aimed at evaluating the potential 137 impact of changes in field sampling strategy on (i) the taxon abundances, (ii) the trait-based and 138 taxonomy-based characterization of macroinvertebrate communities, (iii) assemblage sensitivity 139 descriptors to pesticides and organic pollution, and (iv) the reliability of risk assessment based on the 140 ODI. We compared six strategies that differ in terms of (i) substratum sampling priority rule (i.e., 141 differences in substratum type selection criteria) and (ii) sampled habitat number (i.e., variation in 142 sampling effort). We also analyzed the methodological and ecological rationale underlying the main 143 observed differences/deviations and provided guidance on the interpretation of biological quality 144 element data when different sampling protocols were successively used during the survey.

145

146 **2. Material and methods**

147 **2.1 Macroinvertebrate sampling design**

148 Benthic macroinvertebrate sampling was managed by French regional environmental agencies between 149 2005 and 2013 within the European WFD context, and consisted of 10074 sampling events. The fieldwork design included 1860 sites (Fig. A.1). These sites belonged to three major national networks 150 151 surveying (i) 404 reference sites (= Réseau de Référence, RCR), (ii) 185 sites severely impaired by specific 152 anthropogenic pressures and surveyed until reaching again the "good" ecological status (= Réseau de 153 Contrôle Opérationnel, RCO), and (iii) 1394 sites providing an objective idea of the mean ecological 154 quality of French shallow streams, routinely surveyed and used for periodic reporting on their ecological 155 status (Réseau de Contrôle de Surveillance, RCS; 123 of them are also belonging to the RCR), as required 156 of Member States by the WFD.

The assemblages were sampled following the I_2M_2 standardized method (Norm AFNOR XP T90-388; AFNOR, 2010). For each sampling event, 12 elementary samples (Surber net; sampled area: 0.05 m²; mesh size: 0.5 mm) were performed on benthic substrata (Tables A.1 and A.2). The sampling design

160 consisted of three different steps (= "phases" in the Norm) called "A", "B" and "C", taking into account 161 different combinations of criteria among (i) the relative benthic coverage of substratum types, (ii) their 162 "marginal" (i.e. with an individual share of less than 5% coverage at reach scale) or "dominant" (i.e. with 163 an individual share of at least 5% coverage) status, (iii) their hosting capacity (i.e. their ability to support 164 a rich and diverse macroinvertebrate assemblage) and (iv) the superficial current velocity range in front 165 of the substratum. Each phase was composed of 4 elementary samples. Phase A gathered samples 166 performed on the first four marginal substrata (from bryophytes to artificial substrata; cf. Table A.2) 167 selected according to a decreasing gradient of hosting capacity (criteria ii and iii). They were individually 168 performed in the current velocity range (among four pre-defined categories) in which each substratum 169 type was the most represented (criterion iv). If less than four marginal substratum types had been 170 identified in the site, additional samples were made on the marginal substratum types already sampled, 171 giving priority to the type with the highest individual share and the second most represented velocity 172 range for this type. Phase B consisted of samples performed on the first four dominant substrata 173 applying the same criteria than in phase A. Phase C was composed of four samples performed on 174 dominant substrata, but in proportion to their individual benthic coverages, considering samples already 175 performed on dominant substrata during phase B (AFNOR, 2009). The taxonomic levels of identification 176 corresponded to the standardized French I₂M₂ protocol (Norm AFNOR XP T90-388; AFNOR, 2010), i.e. 177 mainly the genus level, except for Diptera (family level) and Oligochaeta or Nematoda identified as such. 178 A total of 297 taxa were identified at this level.

179

180 **2.2 Ecological rationale of the investigated sampling strategies**

To summarize, we got at the beginning of the study three core taxonomic datasets (one from each of the phases "A", "B" and "C") providing the raw abundances of the 297 taxa over the 10074 sampling events. Then, organism abundances in individual phases were added to establish the assemblages sampled (i)

184 only in benthic substrata with the highest hosting capacity (A+B = "AB"), (ii) only in dominant substrata 185 (B+C = "BC") and (iii) in the whole river reach (A+B+C = "ABC"). AB and BC could be related in terms of 186 habitat selection to the philosophy of the IBGN and the AQEM strategies, respectively. The six resulting 187 taxonomic datasets (A, B, C, AB, BC and ABC) corresponded to six potential sampling strategies exhibiting 188 different sampling efforts and substratum selection strategies (Fig. 1). Strategy ABC, exhaustively 189 gathering all the sampled habitats in a given site and exhibiting the highest sampling effort (= 12 190 elementary samples), corresponded to the strict application of the French norm (i.e. the National 191 standardized strategy; AFNOR, 2009, 2010). Strategy AB corresponded to a moderate reduction of the 192 sampling effort (= 8) with substratum types (four marginal and four dominant substrata) selected only 193 according to a decreasing gradient of hosting capacity. Strategy BC corresponded to a moderate 194 reduction of the sampling effort (= 8) with samples performed exclusively on dominant substrata, taking 195 into account their bottom surface coverage. Finally, strategies A (only marginal substratum types), B 196 (dominant substratum types according to a decreasing gradient of hosting capacity) and C (dominant 197 substratum types according to a decreasing grading of bottom substratum coverage) were characterized by a high reduction of the sampling effort (= 4). We have to mention that, in this study, the effort effect 198 199 cannot be completely disentangle from the mesohabitat effect due to the sampling design settings.



Figure 1: Different combinations of the elementary phases (A, B and C) leading to the six sampling strategies tested and their specific characteristics regarding substratum sampling. The vertical location of

strategies within the triangle refers to sampling effort: bottom = 4 samples, middle = 8 samples, top = 12
 samples.

206 2.3 Metric calculation

207 The influence of the sampling strategy on the community characterization was assessed on a set of 208 taxonomic and bio-ecological metrics. We investigated 101 trait-based metrics (gathered in 11 biological 209 and 10 ecological traits, Table A.3) expressed as the relative frequencies of trait category utilization 210 within each assemblage (in Mondy & Usseglio-Polatera, 2013). The effect of sampling strategy was also 211 assessed on pollution-sensitivity indices such as the Average Score Per Taxon (ASPT, Biological 212 Monitoring Working Party, 1978) and four metrics related to the "SPEcies At Risk" towards pesticides 213 (SPEAR, Liess & Von der Ohe, 2005). On the 297 taxa, 153 were retained for the ASPT score calculation. 214 The four SPEAR metrics designed to evaluate macroinvertebrate community sensitivity to pesticides 215 were the raw number and the raw abundance of SPEAR in the assemblage and the relative number and 216 the relative abundance of SPEAR in the assemblage. The total taxonomic richness (at the I₂M₂ level of 217 identification), the Shannon diversity index, the richness of Ephemeroptera, Plecoptera and Trichoptera 218 (EPT) taxa and their relative richness in the assemblage were also investigated.

219

220 2.4 Risk estimation for 15 pressure categories

We estimated the probability that each assemblage had been impaired by eight pressure categories related to water quality degradation (organic matter, nitrogen compounds (except nitrates), nitrates, phosphorous compounds, acidification, mineral micropollutants, pesticides and organic micropollutants excluding PAHs) and seven pressure categories related to hydromorphology and land use (transportation facilities, riverine vegetation degradation, clogging risk, hydrological instability, straightening, reach urbanization and watershed anthropization) using the ODI (Fig. A.2). This tool was run on each of the six taxonomic datasets (§ 2.2) and provided, for each macroinvertebrate assemblage (in a given dataset) and each pressure category, an impairment probability (IP) using a random forest model specifically designed and calibrated for this category of pressure (Mondy & Usseglio-Polatera, 2013). The provided probabilities range from 0 (no risk that the pressure category significantly impaired the assemblage) to 1 (maximal risk that the pressure category significantly impaired the assemblage).

232

233 2.5 Reliability of the risk estimation obtained for the different strategies

234 To assess the reliability of risk estimations, we compared the estimated IP obtained for each sampling 235 strategy and each pressure category with the observed chemical or hydromorphological quality classes 236 allocated to the sampling events (from field observations and measurements). The pressure categories 237 related to water quality degradation were described by one (i.e. [NO₃-] for "nitrates") to 71 chemical 238 parameters (i.e. 71 molecules for the pressure category "pesticides"). Each pressure category related to 239 habitat degradation was described by one or two elementary parameters (Mondy et al., 2012). The 240 water quality of a given reach at the sampling date was described by averaging available chemical 241 measures from this reach during the six months before macroinvertebrate sampling as suggested by 242 Mondy et al. (2012) for best describing environmental changes integrated by macroinvertebrate 243 communities. The quality class of each individual parameter was defined by comparing the mean value 244 of the parameter over this period of six months with the quality class boundaries specific to this 245 parameter ("high/good", "good/intermediate", "intermediate/poor", "poor/bad"), proposed in the 246 French water quality assessment system (Oudin & Maupas, 2003). The quality class allocated to a site for 247 a given pressure category at the sampling date was the worst quality class allocated over all the 248 individual parameters characterizing this pressure category.

If the observed quality class is "high" or "good", the model should detect a low risk of impairment and provide an IP value below 0.5. If the observed quality class is "intermediate", "poor" or "bad", the model should detect a significant risk and provide an IP value higher than 0.5.

253 2.6 Statistical analysis

254 The sampling strategy effect was assessed on (i) the taxonomic abundance of 297 taxa, (ii) the Shannon 255 diversity, total taxonomic richness and EPT metrics, (iii) the polluo-sensitivity indices (ASPT and SPEAR 256 indices) and (iv) the relative utilization of 101 modalities of traits within the assemblage. For the three 257 first categories of these metrics, the influence of the sampling strategy was investigated by comparing 258 metric value distributions between nine couples of strategies using Wilcoxon paired-tests with the R 259 software (R Core Team, 2018). From those nine couples, four exhibited the same sampling effort and 260 thus reflected the substratum effect (A vs. B, A vs. C, B vs. C, AB vs. BC). Five couples displayed a different 261 sampling effort and thus reflected both the sampling effort and the substratum effect (A vs. ABC, B vs. 262 ABC, C vs. ABC, AB vs. ABC, BC vs. ABC). As bio-ecological metrics were expressed as relative frequencies 263 of trait category utilization, it made no sense to test the effort effect. Thus, the Wilcoxon paired-tests 264 were performed on such metrics only between strategies with the same sampling effort.

265 To obtain more robust results, each Wilcoxon test was performed on 100 sub-sampled datasets 266 composed of 5000 sampling events randomly selected among the 10074 sampling events. Each subset of 267 data was composed of 20% of sampling events performed on least impacted river reaches and 80% on 268 impacted river reaches, mirroring the original dataset. In brief, "least impacted river reach" means that 269 no chemical pressure has been reported as belonging to the poor or bad quality class and that a 270 maximum of three pressure categories could be in the moderate quality class, only one with missing 271 information, and only one hydromorphological pressure in the poor quality class. At the end, we got for 272 each couple of sampling strategies and each tested metric, 100 p-values (one p-value per Wilcoxon test 273 performed on each subset of data). To synthetize the information obtained over the 100 tests, the 95th 274 percentile of the distribution of the 100 p-values was kept as the "representative p-value" of the 275 corresponding comparison. To counteract the problem of multiple statistical decisions (9 couples of sampling strategies over several hundred of metrics) on the alpha error risk, a Bonferroni correction wasused to adjust statistical significance for the number of tests that was performed.

Two heatmaps were built based on the arrays gathering the representative p-values obtained when comparing each couple of sampling strategies based on (i) the abundances of each of 297 taxa or (ii) the relative frequencies of utilization of each of 101 trait modalities within the macroinvertebrate assemblages.

A third heatmap was also built based on the relative frequencies of sampling of the 11 substratum types (cf. Table A.2) during the phases A, B and C of the sampling design, over the data available on reference sites (1275 sampling events) within the whole dataset. All the heatmaps were built using the Euclidian distance and the complete-linkage method with the gplot package (Warnes et al., 2016).

To assess if IPs showed significant differences between each of the five strategies A, B, C, AB, BC and ABC, a Friedman test was performed followed by a Nemenyi post-hoc test using the PMCMR package (Pohlert, 2014).

289

290 Results

291 **3.1 Substrata sampled with each strategy**

The heatmap separated the phase A from the cluster composed of the phases B and C (Fig. 2). Phase A was especially related to the hydrophytes, bryophytes, root mats and organic detritus. Phases B and C showed a low sampling frequency in those substrata. C exhibited a high sampling frequency in filamentous algae, sand, mud, rocks and helophytes. B appeared especially associated to gravel and boulders/cobbles/pebbles, but also to mud and helophytes in a lower extent.



Figure 2: Heatmap based on the relative frequencies of sampling of the 11 substratum-types during the phases A, B and C of the standardized sampling design, over the 404 reference sites of the whole dataset.

302 303

304 3.2 Abundance of taxa in assemblages

305 The heatmap based on taxon abundances highlighted two main clusters of taxa reflecting contrasted 306 responses to sampling strategy (Fig. 3). The first main cluster (cluster I), at the top of the heatmap, was 307 composed of 87 taxa (detailed in table B.1) for which no significant difference (representative p-value > 308 0.05) was observed at least among strategies characterized by 4 samples (A, B and C). The abundances of 309 the taxa belonging to Cluster I were not influenced by the substratum types selected during the three 310 phases of the field sampling. Within Cluster I, Cluster I.1 gathered taxa (e.g. Lestes, Nemurella, Hydroscapha, Ptilocolepus) showing also no significant differences between the other strategies 311 312 whatever their sampling effort (= no sampling effort effect). In contrast, Cluster I.2 was composed of taxa 313 (e.g. Mesovelia, Procambarus, Siphlonurus, Sympecma) exhibiting significant differences between several 314 strategies characterized by different sampling efforts, reflecting a potential higher affinity to A-related

substrata (see the yellow dots on the right part of fig. 3). The taxa of Cluster I were also characterized bylow abundances.

317 The second main cluster was composed of the remaining 210 taxa and was shared in three sub-clusters 318 exhibiting different response profiles. Cluster II.1 included 63 taxa such as Stenelmis, Ephemera, 319 Dreissena and Libellula showing in most cases no difference between strategies A, B and C but strong 320 significant differences between almost all the strategies with different sampling efforts. Cluster II.2 321 gathered 137 taxa for which the abundance in assemblages was influenced by the sampling strategy in 322 terms of substratum type selection as well as sampling effort. Most of them exhibited no significant 323 difference between B and C, meaning that either 1/ they are more abundant in A than in B and C (most 324 of the cases as shown by the dot plot at the right of the heatmap) like for example Platycnemis, 325 Limnebius, Leptocerus, Hydrometra or 2/ they are more abundant in B and C than in A like for example 326 Psychomyia, Glossosoma, Ephoron, Choroterpes, and Agriotypus. Whatever the trend, the criterion 327 (hosting capacity vs. bottom coverage) applied to select the substratum types to sample has a clear 328 effect on the abundance of these taxa in the three phases of the sampling design.

Finally, cluster II.3 was composed of 10 taxa such as *Triaenodes*, *Ylodes*, *Plea* and *Somatochlora* mainly caught in marginal substrata (i.e., for which the A-related substrata highly contributed to the taxon abundance at the site scale as demonstrated by the dot plot).



Figure 3: Responses of the abundances of 297 taxa to the sampling strategy. On the left, the heatmap reflects the p-values obtained from the Wilcoxon paired-tests for each taxon (in row). On the right, the dot plot is based on the total abundance of each taxon over the 10074 sampling events. In heatmap, grey = p.value > 0.05, red = p.value ≤ 0.05 but ≥ 0.0005 , dark-red = p.value < 0.0005. In dot plot, yellow = strategy A, purple = strategy B, black = strategy C. Horizontal lines indicate cluster limits. The taxonomic composition of each cluster is detailed in table B.1.

339 3.3 Taxonomy-related metrics

For Shannon diversity, total taxonomic richness and the richness of EPT taxa in assemblages (Fig. 4.A, 4.B and 4.C, respectively), an effect of substratum type selection and sampling effort was observed. The values mainly increased over the sampling strategies in the order C < B < A < BC < AB <ABC. In addition, the Wilcoxon paired-tests showed significant differences in these metric values between all the investigated strategies (Table B.2). The relative richness of EPT taxa in the assemblage (Fig. 4D) showed an opposite response with A, AB and ABC displaying lower values than B, C and BC with no significant difference observed for "B vs. C".



347

Figure 4: Distribution of the values of A) total taxonomic richness, B) Shannon diversity index, C) richness

of EPT taxa and D) relative richness (%) of EPT taxa, calculated over the 10074 sampling events, for each
 sampling strategy (A, B, C, AB, BC and ABC). Boxplots represent the

351 minimum/Q25/median/Q75/maximum of the distributions of values over the 10074 sampling events.

352 Outliers are out of the 1.5 interquartile range.

353 3.4 Sensitivity-related metrics

354 The number and the abundance of species considered at risk towards pesticides (Fig. 5A and 5B, respectively) exhibited also an increasing gradient of values in the order C < B < A < BC < AB < ABC. For 355 356 both metrics, significant differences were observed between all the couples of sampling strategies (Table B.3). The relative number of species at risk showed similar values over strategies (Fig. 5C) and no 357 significant difference was detected between strategies of similar sampling efforts (A vs. B, A vs. C, B vs. C 358 359 and AB vs. BC). In contrast, the relative abundance of species at risk in the assemblage (Fig. 5D) as well as 360 the ASPT (Fig. 5E) exhibited an opposite response with A, AB and ABC displaying lower metric values than 361 B, C and BC. Only B and C exhibited no significant difference in ASPT values.



Figure 5: Distributions of the values of A) the number of "SPEcies At Risk" (SPEAR), B) the abundance of SPEAR, C) the relative number of SPEAR, D) the relative abundance of SPEAR and E) the ASPT values; for each sampling strategy. Boxplots represent the minimum/Q25/median/Q75/maximum of the distributions of values over the 10074 sampling events. Outliers are out of the 1.5 interquartile range.

368 3.5 Bio-ecological metrics

369 The influence of the sampling strategy on the utilization of 101 bio-ecological trait categories by 370 macroinvertebrate assemblages was assessed by comparing trait category utilization on the basis of the 371 representative p-value obtained for each of the four couples of strategies corresponding to similar 372 sampling effort. The data were summarized by a heatmap (based on the p-values of the Wilcoxon-paired 373 tests) and a dot plot representing, for each trait category and each phase, the median of the distribution of its relative frequencies of use over the 10074 sampling events (Fig. 6). Cluster I was composed by two 374 375 sub-clusters gathering 88 metrics (detailed in table B.4) influenced by the sampled substrata. Cluster I.1 376 included 66 metrics with no-significant difference only between B and C, and exhibited a high number of 377 trait-based metrics (i) more frequently used in A such as 'ovoviviparity' (reproduction trait), 'shredder' (feeding habits) and 'outside river system' (longitudinal distribution) or (ii) more related to B and C such 378 379 as 'living macrophytes' (food), 'scraper' (feeding habits) and 'flag/boulder/cobble/pebble' (substrate 380 preferences). Cluster I.2 gathered 22 metrics exhibiting mostly significant differences between all the 381 couples of strategies. Some metrics such as 'null' (current velocity trait) and 'clutches/terrestrial/in 382 vegetation' (reproduction) appeared more related to A, in contrast to the metrics 'medium' and 'fast' 383 (current velocity) and 'river channel' (transversal distribution). The second cluster was composed by 13 384 metrics divided in two sub-clusters related to metrics such as 'asexual reproduction' (reproduction), 'nymph' and 'larva' (life-stage) or 'polysaprobic' (tolerance to organic contamination) for which mostly 385 no significant differences were observed among A, B and C, but frequent significant differences between 386 387 AB and BC in Cluster II.1 in contrast to Cluster II.2.

388

389



Figure 6: Responses of 101 trait categories to the sampling strategy. On the left, the heatmap reflects the p-values obtained from the Wilcoxon paired-test results for each trait category (in row). On the right, the dot plot represents for each trait category and each phase the median of the distribution of its relative frequencies of use over the 10074 sampling events. In heatmap, grey = p.value > 0.05, red = p.value \leq 0 .05 but \geq 0.0005, dark-red = p.value < 0.0005. In dot plot, yellow = strategy A, purple = strategy B, black = strategy C. Horizontal lines indicate cluster limits. The composition of each cluster of trait categories is detailed in table B.4.

401 **3.6 Predicted risk vs. observed chemical status**

402 For each of the 15 pressure categories, the IP values derived from the A, B, C, AB, BC and ABC sampling 403 strategies were compared to the corresponding observed chemical quality classes (i.e. high, good, 404 moderate, poor, bad) allocated according to the French SEQ-eau2 system (Oudin & Maupas, 2003) (Fig. 7 405 & B.1). In most of the cases, IP values remained in their range of expected values taking into account the 406 corresponding quality class of water or habitat (i.e. IP < 0.5 for high and good status, IP \ge 0.5 for 407 moderate, poor and bad status). A few models like 'organic matter', 'nitrogen compounds (except 408 nitrates)', 'nitrates' and 'phosphorous compounds' displayed contrasted IP values among quality classes 409 (low IP values in high and good quality classes, high IP values for the others) and low IP median variability 410 among strategies. For these models, the six sampling strategies provided rather similar risk estimation. 411 That was confirmed by the main linear relationship between the IPs obtained with each sampling 412 strategy and the IPs obtained with ABC (Fig. B.2). Other models like 'acidification', 'pesticides', 'mineral' 413 and 'organic micropollutants', 'riverine vegetation', 'clogging risk' and 'straightening' displayed also 414 contrasting IP values among a few or all the quality classes but mostly between ABC and the other 415 sampling strategies; while A, B and C provided similar IP value, around 0.5, whatever the quality class. 416 For most of those models, the relationship between IPs obtained by each strategy against ABC presented 417 a Z-shape (Fig. B.2). The basis and the top of each Z shape correspond to assemblages for which one 418 strategy provides an IP value highly above 0.5 while the other strategy provides an IP value clearly lower 419 than 0.5. That means that different lists coming from the same site but obtained with different sampling 420 strategies provided contrasted risk estimation leading either to impairment risk underestimation or 421 overestimation.

The Friedman statistical tests (per pressure category and quality class) revealed most often significant
 differences in IPs between strategies except for the pressure acidification (good quality class), pesticides

424 (bad quality class) and organic micropollutants (bad and poor quality classes) (data not shown). The
425 Nemenyi post-hoc tests confirmed this observation and highlighted the most frequent, statistically
426 significant differences in IP values between the sampling strategy ABC and the five other strategies
427 (Table B.5).

428

429



430

431 Figure 7: Distribution of impairment probabilities (IPs) assigned to macroinvertebrate assemblages 432 provided by the A, B, C, AB, BC and ABC sampling strategies for 10074 sampling events against their 433 respective observed quality classes (based on chemical characteristics of water or hydromorphological 434 characteristics of habitat) for the pressure categories: A) 'Nitrogen compounds' and B) 'Riverine 435 vegetation'. The horizontal line (IP = 0.5) represents the threshold between a 'low' impairment risk and a 436 'significant' impairment risk. The bow and whisker plots represent the 437 minimum/Q25/median/Q75/maximum values, respectively. Outliers (open circles) are outside the 1.5 438 interquartile range of the corresponding distribution. The number of observations for each sampling 439 strategy and quality class is specified as N.

441 4. Discussion

442 We compared six invertebrate sampling strategies differing (i) only in substratum selection criteria, or (ii) 443 in both substratum selection criteria and sampling effort. Two of these strategies (i.e., AB and ABC) could 444 be related to field-sampling protocols currently used to estimate the ecological status of river reaches 445 based on standardized methods. The AB strategy has been associated to the French IBGN index applied 446 since 1992 for evaluating the ecological status of French shallow river reaches (AFNOR, 1992; Verneaux 447 et al., 1982). The ABC strategy is the current standardized method (AFNOR, 2009) explicitly related to the 448 I_2M_2 index (Mondy et al., 2012), which will be used for the invertebrate-based reporting on the 449 ecological status of the French shallow rivers during the third European water body management cycle 450 (2021-2027). Finally, the BC strategy mimics the philosophy of the AQEM sampling protocol in terms of 451 habitat selection criteria (Hering et al., 2004) giving priority to the benthic coverage of substrata and 452 focusing on the dominant ones (Haase et al., 2008). However, the reference list of substratum types, the 453 benthic mosaic description (i.e., including or not mineral substrata covered by organic ones in this 454 description), the sampling effort (i.e., eight versus 20 sampling units), and the sampling device (Surber 455 net versus hand net) are different.

456

457 Because sampling strategies have often progressed over time (Moog et al., 2018), e.g. within a given 458 country due to legislative compliance requirements, the characterization of macroinvertebrate 459 assemblages may also vary over time independently of variations caused by changes in habitat condition. 460 Variations in sampling effort and/or in the selection of sampled substrata could indeed seriously alter 461 the interpretation of long time-series in terms of taxonomy- or trait-based assemblage description as 462 well as river ecological assessment (Friberg et al., 2006; Haase et al., 2008; Vlek et al., 2006). Moreover, 463 the different strategies used over the different European countries rely not necessarily on the same 464 habitat sampling priority criteria, sampling effort, or sampling devices, inducing variability in biological

assessment (Birk et al., 2012; Clarke et al., 2006; Friberg et al., 2006). The following sections will point
out how our results converge towards this statement.

467

468 **4.1 Influence of sampling strategy on invertebrate community description**

469 In bioassessment, macroinvertebrate communities are commonly characterized on the basis of 470 taxonomic or functional metrics. Our study pointed out that sampling strategies not sharing the same 471 substratum sampling priority criteria but also the same sampling effort can provide clearly different 472 descriptions of the community due to different distributions of substratum types in marginal and 473 dominant substrata of benthic mosaics, even when observed at National scale (cf. Fig. 2), leading to 474 different compositions of invertebrate assemblages at mesohabitat scale. Our study has highlighted that 475 marginal substrata appeared to host a greater diversity and richness of macroinvertebrate taxa than the 476 dominant substrata (e.g., when comparing A to B, A to C and AB to BC). This result was in line with the 477 significantly higher number of individuals obtained by the French IBGN method over the eleven 478 European methods compared by Friberg et al. (2006) and despite its lowest normalized sampling effort. 479 This greater taxonomic diversity was accompanied by a higher number of EPT and SPEAR taxa. Variation 480 in taxonomic richness and taxonomic diversity of macroinvertebrate assemblages has been already 481 observed between substrate-types (Beisel et al., 1998a, 2000) or biotopes (Demars et al., 2012; White et 482 al., 2017) with higher alpha (within-habitat) taxonomic diversity rather supported by organic (e.g., 483 bryophytes, marginal plants, root mats, organic litter) than mineral (e.g., sand, silt, mud, flagstone) 484 substrata or biotopes. These results are in accordance with the composition of habitat clusters and the 485 related patterns of taxonomic metric values observed in our study, allowing to assume the strong driving 486 role of marginal substrata on the taxonomic richness and diversity of macroinvertebrate assemblages at 487 reach scale (Beisel et al., 1998b; Haase et al., 2008). These results are in agreement with the highest 488 hosting capacity of the most frequent marginal substrata (i.e., bryophytes, hydrophytes, root mats and

489 organic detritus in Fig. 2; table A.2), maximizing taxonomic richness and diversity - but also taxon 490 sensitivity - in local assemblages (Beisel et al., 1998a, 2000). When the macroinvertebrate sampling 491 protocol ignores the marginal substrata (e.g., the AQEM protocol), a lower taxonomic diversity and 492 number of EPT taxa (at least at the family level) have been obtained in comparison to communities 493 sampled using other methods on the same sites (Friberg et al., 2006). A similar trend has been observed 494 for taxa sensitive to pesticides (SPEAR) and reflecting water quality (ASPT). However, when the EPT and 495 SPEAR metrics were expressed as relative richness and abundance in the whole assemblage respectively 496 (Fig. 4D and 5D), between-sampling strategies differences in metrics values were lower.

Another assumption could be that marginal – mainly organic - substrata are characterized by greater
beta (between-habitats) diversity than the dominant ones (Verdonschot et al., 2016; White et al., 2017).
However, even if the I₂M₂ sampling protocol provides a good overview of reach habitats, it gathers a
maximum of four substrata per phase, not allowing to rigorously test this assumption on the whole
range of habitat combinations in river reaches including often more than four marginal ones.

502

503 Most of the 101 investigated trait-based metrics were influenced by the substratum priority criterion of 504 the selected sampling strategy. The three main patterns highlighted that (i) only a minority of trait-based 505 metrics (e.g., >1-2cm (size), flier (locomotion), twigs/roots (substrate preference)) were independent of 506 the substratum priority criterion and thus may provide the same information whatever the substrata sampled, (ii) several metrics (e.g. macrophytes (substrate preference), slow (current velocity 507 preference)) depended on this criterion, providing a different picture of assemblage trait composition 508 509 according to the sampled substratum types, and (iii) most of metrics indicated that marginal substrata 510 with high hosting capacity have not the same functional characteristics than dominant ones. The two last 511 patterns are in coherence with the habitat templet concept which has linked organisms with specific 512 traits to specific habitat characteristics (Southwood, 1977; Townsend & Hildrew, 1994). The

513 macroinvertebrate trait structure in an assemblage is known to be related to the substratum type 514 (Demars et al., 2012; Usseglio-Polatera et al., 1999). Thus, it appears crucial to identify, prior to the 515 sampling, the different habitats present at the reach scale in order to capture cautiously the diversity of 516 traits of *in-situ* communities. A reliable definition of community trait-based metrics is especially 517 important when they are included in index calculation for evaluating the ecological status of streams or 518 for supporting the stakeholder decision-making process.

519

520 The ecological rationale of a sampling strategy and the way for describing a benthic mosaic at the reach 521 scale have appeared as two very critical points. In order to prevent important bias in data interpretation, 522 several general guidelines can be given. First, the hosting capacity of substrata is clearly an important 523 determinant of the taxonomic and functional macroinvertebrate assemblage composition, confirming 524 the important role of the rationale underpinning the sampling strategy (hosting capacity vs dominance). 525 Some structural aspects of the community are indeed captured differently by different sampling 526 strategies (Friberg et al., 2006; Rossini et al., 2016), potentially leading to deliver to stakeholders 527 different - or even divergent - messages depending on the selected substrata in the benthic mosaic and 528 the selected metrics. In streams, the hydromorphological conditions are one of the drivers of the benthic 529 mosaic (Villeneuve et al., 2018) which influence the biological characteristics of local macroinvertebrate 530 assemblages as well as the ecological status assessment (Buffagni et al., 2019). Thus, the definition of the 531 benthic mosaic appears as a critical step for the ecological diagnostic. Second, the marginal substrata in a 532 given reach are clearly important for macroinvertebrate assemblages and significantly influence 533 macroinvertebrate community characterization at the reach scale. Third, the compositional structure of 534 marginal substrata is more specific than that of dominant ones, as already observed by Leitão et al. 535 (2014). As a result, including or not marginal substrata in the reach macroinvertebrate sample could lead 536 to a more or less important selectivity that has to be kept in mind when defining the stream sampling

537 design with the major objective of surveying long-term effects of anthropogenic pressures or when 538 analyzing data achieved with different successive sampling protocols. Potentially observed tipping points 539 could be indeed not only related to objective variations in the taxonomy- or trait-based structure of local 540 macroinvertebrate assemblages subjected to anthropogenic pressures but could be also linked to 541 potential modifications in sampling strategies especially if successive selected strategies differently 542 consider marginal substrata. Fourth, macroinvertebrate compositional metrics aggregated in multimetric 543 indices should preferably be expressed as relative abundance or richness of the targeted taxa than with 544 raw values of abundance or richness, for providing a more robust ecological assessment of river reaches. 545 Fifth, the highest richness of species sensitive to pesticides and organic pollution in marginal substrata 546 should allow to explore the development of rapid, cost-effective methods especially focusing on 547 macroinvertebrate assemblages of a limited number of marginal substrata, used as early warning 548 indicators of corresponding categories of anthropogenic pressures, as already proposed by Piscart et al. 549 (2006) for pebbles, a sentinel substratum reflecting salinization.

550

551 **4.2 Influence of sampling strategy on risk assessment**

552 In Europe, different methods have been successively used over time to sample benthic 553 macroinvertebrate assemblages. In France, the IBGN (1992-2004) has been replaced by the I₂M₂ since 554 2005 (Mondy et al., 2012). Similarly, the European AQEM protocol has been routinely applied in many 555 European countries since 2004, in place of older bioassessment methods (Buffagni et al., 2004; Hering et 556 al., 2004; Hering et al., 2006). Various other macroinvertebrate sampling methods presenting different 557 rationales have been also used across European countries (Clarke et al., 2006; Friberg et al., 2006). 558 Consequently, a huge amount of information has been gained over years, but often heterogeneously. 559 Retrospectively investigating ecological risk assessment at large spatial and temporal scales using data

obtained with different sampling strategies should be made cautiously to avoid potential bias related to
between-sampling methods heterogeneity (this is especially obvious when the sampling effort is low).

562

563 The tool we developed to estimate the risk of a significant impact of different anthropogenic pressure 564 categories on the local macroinvertebrate communities have clearly shown some limits when it was run 565 on faunal assemblages not obtained with the sampling protocol (i.e., ABC) on which it had been 566 calibrated. Both sampling effort and substratum selection criteria matter for impairment probability 567 estimation by the ODI. Above all, a loss of reliability in risk estimation (Fig. 7B, B.1) was observed when 568 the models were not running on the ABC list but this loss was more or less dramatic depending on the 569 pressure type. In most of the cases, the models were not powerful enough to support the loss of 570 biological information induced by the reduction of the sampling effort and/or the restricted biological 571 picture of the assemblage (e.g., 'acidification', 'mineral' and 'organic' micropollutants, 'hydrological 572 instability'). This can be due to the (only) fair accuracy of the model, a lack of homogeneity in the 573 gradient pressure description in the training dataset or because the model relies on habitat-specific 574 metrics, and then is sensitive to sampling strategy. For example, even if both marginal and dominant 575 substrata are taken into account by the sampling strategy (e.g., the strategy AB), only slight differences 576 in trait-based metric values calculated on the assemblages provided by the sampling strategies AB and 577 ABC can be sufficient for inducing strong differences in estimated IP values (i.e., at the origin of the Z-578 shapes on Fig. B.2), leading to divergent diagnostics (one sampling strategy detecting a clear site 579 community impairment, the other not).

For other models, the loss of biological information induced by lower sampling effort and variation in substratum sampling priority induced only a moderate loss of reliability and the IP values remained in the expected range of values (IP < 0.5 for the high quality and good quality classes; IP \ge 0.5 for the moderate, poor and bad quality classes) when they run on lists obtained via other strategies than ABC (organic matter, nitrogen compounds, nitrates and phosphorous compounds). Another trend is that the sampling strategies using hosting capacity as the main criterion of substratum selection tend to provide the IP values most similar to those provided by ABC in the moderate, poor and bad quality classes. Only a few models (i.e. those that not exhibited a Z-shape when comparing IPs values between sampling strategies) could be potentially used, but very cautiously, with data provided by sampling protocols slightly different from the calibration sampling protocol (ABC).

590

591 Conclusion

592 Differences in taxonomic and trait structure characterization as well as sensitivity descriptors of 593 macroinvertebrate communities due to differences in sampling strategies clearly influence the 594 macroinvertebrate-based environmental diagnostic of streams. If both sampling effort and sampled 595 habitat selection criteria have an impact on community characterization and impairment risk 596 assessment, an important component is undoubtedly the presence/absence of marginal habitats in the 597 sampling design. Such methodological confounding effect undoubtedly represents a critical issue for the 598 ecological interpretation of the provided results, for example in the context of long-term series surveys 599 when the sampling strategy has changed over time. However, the effect is globally less important on 600 trait-based metrics than on taxonomy-based metrics. Moreover, robust models assessing the risk of 601 impairment by a given pressure category (e.g., organic matter, nitrogen compounds (except nitrates) and 602 phosphorous compounds; AUC > 0.85) are not only able to run on data achieved with the sampling 603 strategy on which they were calibrated (i.e., ABC), but also on data slightly differing in the number of 604 samples performed or in the rules of sampled habitat selection (e.g., strategy AB (IBGN-like) or BC 605 (AQEM-like)). Even if the results provided by such diagnostic tools should be interpreted cautiously (in 606 particular when pressure effects are moderate), they could clearly facilitate the ecological interpretation

- of macroinvertebrate community dynamics over long periods, especially when sampling strategy hasvaried over time.
- 609
- 610
- 611 References
- 612 AFNOR (1992). Détermination de l'indice biologique global normalisé (I.B.G.N.)—9 p. Agence de l'Eau
- 613 R.M.C., 1990: Application de l'indice biologique global (I.B.G.) aux grands cours d'eau du
 614 bassin Rhône Méditerranée Corse, 14 p.
- AFNOR (2004). Qualité de l'eau—Détermination de l'indice biologique global normalisé (IBGN). NF T90350.
- AFNOR (2009). Qualité de l'eau Prélèvement des macro-invertébrés aquatiques en rivières peu
 profondes. XP T90-333, 22 pp.
- 619 AFNOR (2010). Qualité de l'eau Traitement au laboratoire d'échantillons contenant des macro-
- 620 invertébrés de cours d'eau. XP T90-388, 21 pp.
- Angradi, T. R., Pearson, M. S., Bolgrien, D. W., Jicha, T. M., Taylor, D. L., & Hill, B. H. (2009). Multimetric
- 622 macroinvertebrate indices for mid-continent US great rivers. Journal of the North American
- 623 Benthological Society, 28(4), 785–804. https://doi.org/10.1899/09-003.1
- 624 Armitage, P. D., & Pardo, I. (1995). Impact assessment of regulation at the reach level using
- 625 macroinvertebrate information from mesohabitats. Regulated Rivers: Research & Management,
- 626 10(2–4), 147–158. https://doi.org/10.1002/rrr.3450100210
- 627 Beisel, J.-N., Usseglio-Polatera, P., & Moreteau, J.-C. (2000). The spatial heterogeneity of a river bottom:
- 628 A key factor determining macroinvertebrate communities. In M. Jungwirth, S. Muhar, & S.
- 629 Schmutz (Eds.), Assessing the Ecological Integrity of Running Waters (pp. 163–171). Springer
- 630 Netherlands. https://doi.org/10.1007/978-94-011-4164-2_13

- 631 Beisel, J.-N., Usseglio-Polatera, P., Thomas, S., & Moreteau, J.-C. (1998a). Stream community structure in
- 632 relation to spatial variation: The influence of mesohabitat characteristics. Hydrobiologia,

633 389(1/3), 73–88. https://doi.org/10.1023/A:1003519429979

- Beisel, J.-N., Usseglio-Polatera, P., Thomas, S., & Moreteau, J.-C. (1998b). Effects of sampling on benthic
- 635 invertebrate assemblages in assessment of biological quality of running water. Annales de
- 636 Limnologie International Journal of Limnology, 34(4), 445–454.
- 637 https://doi.org/10.1051/limn/1998035
- 638 Biological Monitoring Working Party. (1978). Final Report: Assessment and Presentation of the Biological

639 Quality of Rivers in Great Britain. Unpublished report, Dept. Of Envir., Water Data Unit.

- Birk, S., Bonne, W., Borja, A., Brucet, S., Courrat, A., Poikane, S., Solimini, A., van de Bund, W.,
- 641 Zampoukas, N., & Hering, D. (2012). Three hundred ways to assess Europe's surface waters: An
- almost complete overview of biological methods to implement the Water Framework Directive.

643 Ecological Indicators, 18, 31–41. https://doi.org/10.1016/j.ecolind.2011.10.009

- Blocksom, K. A., & Flotemersch, J. E. (2005). Comparison of macroinvertebrate sampling methods for
- 645 nonwadeable streams. Environmental Monitoring and Assessment, 102(1–3), 243–262.
- 646 https://doi.org/10.1007/s10661-005-6025-3
- 647 Brucet, S., Poikane, S., Lyche-Solheim, A., & Birk, S. (2013). Biological assessment of European lakes:

648 Ecological rationale and human impacts. Freshwater Biology, 58(6), 1106–1115.

- 649 https://doi.org/10.1111/fwb.12111
- 650 Buffagni, A., Barca, E., Erba, S., & Balestrini, R. (2019). In-stream microhabitat mosaic depicts the success
- of mitigation measures and controls the Ecological Potential of benthic communities in heavily
- 652 modified rivers. Science of The Total Environment, 673, 489–501.
- 653 https://doi.org/10.1016/j.scitotenv.2019.04.124

- Buffagni, A., Erba, S., Cazzola, M., & Kemp, J. L. (2004). The AQEM multimetric system for the southern
- 655 Italian Apennines: Assessing the impact of water quality and habitat degradation on pool
- 656 macroinvertebrates in Mediterranean rivers. Hydrobiologia, 516(1–3), 313–329.
- 657 https://doi.org/10.1023/B:HYDR.0000025273.15958.6a
- 658 Cao, Y., & Hawkins, C. P. (2011). The comparability of bioassessments: A review of conceptual and
- 659 methodological issues. Journal of the North American Benthological Society, 30(3), 680–701.
 660 https://doi.org/10.1899/10-067.1
- 661 Christensen, M. R., Graham, M. D., Vinebrooke, R. D., Findlay, D. L., Paterson, M. J., & Turner, M. A.
- 662 (2006). Multiple anthropogenic stressors cause ecological surprises in boreal lakes. Global
 663 Change Biology, 12(12), 2316–2322. https://doi.org/10.1111/j.1365-2486.2006.01257.x
- 664 Clarke, R. T., Davy-Bowker, J., Sandin, L., Friberg, N., Johnson, R. K., & Bis, B. (2006). Estimates and
- 665 comparisons of the effects of sampling variation using 'national' macroinvertebrate sampling
- 666 protocols on the precision of metrics used to assess ecological status. Hydrobiologia, 566(1),
- 667 477–503. https://doi.org/10.1007/s10750-006-0076-5
- 668 Clarke, R. T., & Hering, D. (2006). Errors and uncertainty in bioassessment methods major results and
- 669 conclusions from the STAR project and their application using STARBUGS. Hydrobiologia, 566(1),
- 670 433–439. https://doi.org/10.1007/s10750-006-0079-2
- 671 Clean Water Act. (1977). P.L. 95-217.
- Darling, E. S., & Côté, I. M. (2008). Quantifying the evidence for ecological synergies. Ecology Letters,
- 673 11(12), 1278–1286. https://doi.org/10.1111/j.1461-0248.2008.01243.x
- Davis, J., Sim, L., & Chambers, J. (2010). Multiple stressors and regime shifts in shallow aquatic
- 675 ecosystems in antipodean landscapes. Freshwater Biology, 55, 5–18.
- 676 https://doi.org/10.1111/j.1365-2427.2009.02376.x

677 Dedieu, N., Clavier, S., Vigouroux, R., Cerdan, P., & Céréghino, R. (2016). A Multimetric

- 678 Macroinvertebrate Index for the Implementation of the European Water Framework Directive in
- 679 French Guiana, East Amazonia: A biotic index for East-Amazonian streams. River Research and
- 680 Applications, 32(3), 501–515. https://doi.org/10.1002/rra.2874
- 681 Demars, B. O. L., Kemp, J. L., Friberg, N., Usseglio-Polatera, P., & Harper, D. M. (2012). Linking biotopes to
- 682 invertebrates in rivers: Biological traits, taxonomic composition and diversity. Ecological
- 683 Indicators, 23, 301–311. https://doi.org/10.1016/j.ecolind.2012.04.011
- 684 Desrosiers, M., Usseglio-Polatera, P., Archaimbault, V., Larras, F., Méthot, G., & Pinel-Alloul, B. (2019).
- 685 Assessing anthropogenic pressure in the St. Lawrence River using traits of benthic
- 686 macroinvertebrates. Science of The Total Environment, 649, 233–246.
- 687 https://doi.org/10.1016/j.scitotenv.2018.08.267
- Dolédec, S., & Statzner, B. (2010). Responses of freshwater biota to human disturbances: Contribution of
- 689 J-NABS to developments in ecological integrity assessments. Journal of the North American

690 Benthological Society, 29(1), 286–311. https://doi.org/10.1899/08-090.1

- 691 Downes, B. J. (2010). Back to the future: Little-used tools and principles of scientific inference can help
- 692 disentangle effects of multiple stressors on freshwater ecosystems. Freshwater Biology, 55, 60–
- 693 79. https://doi.org/10.1111/j.1365-2427.2009.02377.x
- European Commission (2000). Directive 2000/60/EC of the European Parliament and of the Council of 23
- 695 October 2000 Establishing a Framework for Community Action in the Field of Water Policy. Office 696 for official publications of the European Communities, Brussels.
- 697 Friberg, N., Sandin, L., Furse, M. T., Larsen, S. E., Clarke, R. T., & Haase, P. (2006). Comparison of
- 698 macroinvertebrate sampling methods in Europe. Hydrobiologia, 566(1), 365–378.
- 699 https://doi.org/10.1007/s10750-006-0083-6

700	Haase, P., Pauls, S. U., Engelhardt, C. H. M., & Sundermann, A. (2008). Effects of sampling microhabitats
701	with low coverage within the STAR/AQEM macroinvertebrate sampling protocol on stream
702	assessment. Limnologica, 38(1), 14–22. https://doi.org/10.1016/j.limno.2007.06.002
703	Hering, D., Carvalho, L., Argillier, C., Beklioglu, M., Borja, A., Cardoso, A. C., Duel, H., Ferreira, T.,
704	Globevnik, L., Hanganu, J., Hellsten, S., Jeppesen, E., Kodeš, V., Solheim, A. L., Nõges, T.,
705	Ormerod, S., Panagopoulos, Y., Schmutz, S., Venohr, M., & Birk, S. (2015). Managing aquatic
706	ecosystems and water resources under multiple stress—An introduction to the MARS project.
707	Science of The Total Environment, 503–504, 10–21.
708	https://doi.org/10.1016/j.scitotenv.2014.06.106
709	Hering, D., Feld, C. K., Moog, O., & Ofenböck, T. (2006). Cook book for the development of a Multimetric
710	Index for biological condition of aquatic ecosystems: Experiences from the European AQEM and
711	STAR projects and related initiatives. Hydrobiologia, 566(1), 311–324.
712	https://doi.org/10.1007/s10750-006-0087-2
713	Hering, D., Moog, O., Sandin, L., & Verdonschot, P. F. M. (2004). Overview and application of the AQEM
714	assessment system. Hydrobiologia, 516(1–3), 1–20.
715	https://doi.org/10.1023/B:HYDR.0000025255.70009.a5
716	Leitão, F., Hughes, S. J., Máximo, I., Atanasova, N., Furtado, A., & Chicharo, L. (2014). Habitat-oriented
717	sampling of macroinvertebrates affects the determination of ecological status in temporary
718	mediterranean river systems. River Research and Applications, 30(10), 1233–1247.
719	https://doi.org/10.1002/rra.2800
720	Liess, M., & Von der Ohe, P. C. (2005). Analyzing effects of pesticides on invertebrate communities in
721	streams. Environmental Toxicology and Chemistry, 24(4), 954–965. https://doi.org/10.1897/03-
722	652.1

723	Mondy, C. & Usseglio-Polatera, P. (2013). Using conditional tree forests and life history traits to assess
724	specific risks of stream degradation under multiple pressure scenario. Science of The Total
725	Environment, 461–462, 750–760. https://doi.org/10.1016/j.scitotenv.2013.05.072
726	Mondy, C., Villeneuve, B., Archaimbault, V., & Usseglio-Polatera, P. (2012). A new macroinvertebrate-
727	based multimetric index (I2M2) to evaluate ecological quality of French wadeable streams
728	fulfilling the WFD demands: A taxonomical and trait approach. Ecological Indicators, 18, 452–
729	467. https://doi.org/10.1016/j.ecolind.2011.12.013
730	Moog, O., Schmutz, S., & Schwarzinger, I. (2018). Biomonitoring and Bioassessment. In S. Schmutz & J.
731	Sendzimir (Eds.), Riverine Ecosystem Management (pp. 371–390). Springer International
732	Publishing. https://doi.org/10.1007/978-3-319-73250-3_19
733	Navarro-Ortega, A., Acuña, V., Bellin, A., Burek, P., Cassiani, G., Choukr-Allah, R., Dolédec, S., Elosegi, A.,
734	Ferrari, F., Ginebreda, A., Grathwohl, P., Jones, C., Rault, P. K., Kok, K., Koundouri, P., Ludwig, R.
735	P., Merz, R., Milacic, R., Muñoz, I., Barceló, D. (2015). Managing the effects of multiple
736	stressors on aquatic ecosystems under water scarcity. The GLOBAQUA project. Science of The
737	Total Environment, 503–504, 3–9. https://doi.org/10.1016/j.scitotenv.2014.06.081
738	Oudin, L., & Maupas, D. (2003). Assessment system of streams quality (Système d'évaluation de la
739	qualité de l'eau des cours d'eau (Seq-Eau)) Ministère de l'Ecologie et du Développement Durable
740	– Agences de l'Eau, Paris, France.
741	Piscart, C., Moreteau, JC., & Beisel, JN. (2006). Salinization consequences in running waters: Use of a
742	sentinel substrate as a bioassessment method. Journal of the North American Benthological
743	Society, 25(2), 477–486. https://doi.org/10.1899/0887-3593(2006)25[477:SCIRWU]2.0.CO;2
744	Pohlert, T. (2014). The Pairwise Multiple Comparison of Mean Ranks Package (PMCMR). R package.
745	Http://CRAN.R-project.org/package=PMCMR.

- R Core Team. (2018). R: A language and environment for statistical computing. R Foundation for
 Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.
- 748 Rossini, R. A., Fensham, R. J., & Walter, G. H. (2016). Determining optimal sampling strategies for
- 749 monitoring threatened endemic macro-invertebrates in Australia. Marine and Freshwater
- 750 Research, 67(5), 653. https://doi.org/10.1071/MF15023
- 751 Sokolova, I. M. (2013). Energy-Limited Tolerance to Stress as a Conceptual Framework to Integrate the
- 752 Effects of Multiple Stressors. Integrative and Comparative Biology, 53(4), 597–608.
- 753 https://doi.org/10.1093/icb/ict028
- Southwood, T. (1977). Habitat, the templet for ecological strategies. Journal of Animal Ecology, 46, 337–
 365.
- 756 Southwood, T. R. E. (1988). Tactics, Strategies and Templets. Oikos, 52(1), 3.
- 757 https://doi.org/10.2307/3565974
- 758 Touron-Poncet, H., Bernadet, C., Compin, A., Bargier, N., & Céréghino, R. (2014). Implementing the
- 759 Water Framework Directive in overseas Europe: A multimetric macroinvertebrate index for river
- 760 bioassessment in Caribbean islands. Limnologica, 47, 34–43.
- 761 https://doi.org/10.1016/j.limno.2014.04.002
- Townsend, C., & Hildrew, A. (1994). Species traits in relation to a habitat templet for river systems.
- 763 Freshwater Biology, 31, 265–275.
- 764 Tronstad, L. M., & Hotaling, S. (2017). Long-term trends in aquatic ecosystem bioassessment metrics are
- 765 not influenced by sampling method: Empirical evidence from the Niobrara River. Knowledge &
- 766 Management of Aquatic Ecosystems, 418, 28. https://doi.org/10.1051/kmae/2017020
- 767 Urbanič, G. (2014). Hydromorphological degradation impact on benthic invertebrates in large rivers in
- 768 Slovenia. Hydrobiologia, 729(1), 191–207. https://doi.org/10.1007/s10750-012-1430-4

- 769 Vander Laan, J. J., Hawkins, C. P., Olson, J. R., & Hill, R. A. (2013). Linking land use, in-stream stressors,
- and biological condition to infer causes of regional ecological impairment in streams. Freshwater

771 Science, 32(3), 801–820. https://doi.org/10.1899/12-186.1

- Verdonschot, R. C. M., Kail, J., McKie, B. G., & Verdonschot, P. F. M. (2016). The role of benthic
- 773 microhabitats in determining the effects of hydromorphological river restoration on
- 774 macroinvertebrates. Hydrobiologia, 769(1), 55–66. https://doi.org/10.1007/s10750-015-2575-8
- 775 Verneaux, J., Galmiche, P., Janier, F., & Monnot, A. (1982). Une nouvelle méthode pratique d'évaluation
- de la qualité des eaux courantes. Un indice biologique de qualité générale (IBG). Ann. Sc. Univ.
- 777 Franche-Comté Besançon , Biol. Anim. , 4: 11-21.
- 778 Villeneuve, B., Piffady, J., Valette, L., Souchon, Y., & Usseglio-Polatera, P. (2018). Direct and indirect
- effects of multiple stressors on stream invertebrates across watershed, reach and site scales: A
 structural equation modelling better informing on hydromorphological impacts. Science of The
 Total Environment, 612, 660–671. https://doi.org/10.1016/j.scitotenv.2017.08.197
- 782 Vlek, H. E., Šporka, F., & Krno, I. (2006). Influence of macroinvertebrate sample size on bioassessment of
- 783 streams. In M. T. Furse, D. Hering, K. Brabec, A. Buffagni, L. Sandin, & P. F. M. Verdonschot (Eds.),
- 784 The Ecological Status of European Rivers: Evaluation and Intercalibration of Assessment Methods
- 785 (pp. 523–542). Springer Netherlands. https://doi.org/10.1007/978-1-4020-5493-8_35
- 786 Warnes, G., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M.,
- 787 Magnusson, A., Moeller, S., Schwartz, M., & Venables, B. (2016). Gplots: Various R Programming
 788 Tools for Plotting Data. Version 3.0.1. Https://CRAN.R-project.org/package=gplots.
- 789 White, J. C., Hill, M. J., Bickerton, M. A., & Wood, P. J. (2017). Macroinvertebrate Taxonomic and
- 790 Functional Trait Compositions within Lotic Habitats Affected By River Restoration Practices.
- 791 Environmental Management, 60(3), 513–525. https://doi.org/10.1007/s00267-017-0889-1

792	Wiederkehr, J., Grac, C., Fontan, B., Labat, F., Le Ber, F., & Trémolières, M. (2016). Experimental study of
793	the uncertainty of the intrasubstrate variability on two French index metrics based on
794	macroinvertebrates. Hydrobiologia, 779(1), 59–73. https://doi.org/10.1007/s10750-016-2799-2
795	