

Invertebrates in pigeonholes: Evaluating pan-European freshwater typologies in light of their macroinvertebrate communities

Jonathan F. Jupke¹ Pepe Barquín² Núria Bonada³
Christian K. Feld^{4,6} Peter Haase^{5,6} Kaisa-Leena Huttunen⁷
Miguel Cañedo-Argüelles³ Timo Muotka^{7,8} Riku Paavola⁹
Edwin T. H. M. Peeters¹⁰ Leonard Sandin¹¹ Denes Schmera¹²
Philippe Usseglio-Polatera¹³ Ralf B. Schäfer¹

¹ iES Landau, Institute for Environmental Sciences, University Koblenz-Landau, Fortstrasse 7, 76829 Landau, Germany

² Environmental Hydraulics Institute, University of Cantabria, C/ Isabel Torres 15, Parque Científico y Tecnológico de Cantabria, 39011 Santander, Spain.

³ Grup de Recerca “Freshwater Ecology, Hydrology and Management” (FEHM), Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals, Facultat de Biologia, Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona (UB), Diagonal 643, 08028, Barcelona, Catalonia, Spain.

⁴ Centre for Water and Environmental Research, University of Duisburg-Essen, 45141 Essen, Germany

⁵ Department of River Ecology and Conservation, Senckenberg Research Institute and Natural History Museum Frankfurt, Gelnhausen 65371, Germany

⁶ Faculty of Biology, University of Duisburg-Essen, Essen 45141, Germany

⁷ Department of Ecology and Genetics, University of Oulu, P.O. Box 3000, FI-90014 Oulu, Finland

⁸ Finnish Environment Institute, Freshwater Centre, P.O. Box 413, FI-90014 Oulu, Finland

⁹ Oulanka Research Station, University of Oulu Infrastructure Platform, Liikasenvaarantie 134, Kuusamo 93900, Finland

¹⁰ Aquatic Ecology and Water Quality Management Group, Wageningen University, P.O. Box 47, 6700 AA Wageningen, The Netherlands

¹¹ Norwegian Institute for Water Research, Gaustadalléen 21, NO-0349 Oslo, Norway

¹² Centre for Ecological Research, Balaton Limnological Institute, Tihany, Hungary

Acknowledgements

This study was funded by the GETREAL (Incorporating spatial and seasonal variability in community sensitivity into chemical risk assessment) project, which was funded by the European Chemical Industry Council long-range research initiative (CEFIC-LRi project ECO 50).

We are grateful to the staff that collected and analyzed data for the AQUEM, STAR, WISER and ECOSURV projects and to the various water agencies that participated in the abiotic and biotic data collection.

1 Data Availability

Some of the monitoring data sets are freely available. Data from the Ebro Hydrographic Confederation is available under www.datossuperficiales.chebro.es:81/WCASF/?rvn=1, data from Nades under www.naiades.eaufrance.fr/france-entiere, data from the RIVPACS data base under www.ceh.ac.uk/services/rivpacs-reference-database and data from the Dutch monitoring agencies under waterkwaliteitsportaal.nl.

2 Abstract

Human actions are altering the composition and functioning of ecological communities globally. This transformation is especially rapid and intensive in freshwater ecosystems. Comprehensive and international regulatory frameworks are necessary to slow down and halt the current loss of freshwater biodiversity. These regulatory frameworks, such as the Water Framework Directive (WFD), typically reduce the inherent complexity of freshwater ecosystems through classification systems that use selected features to classify them into a typology. A recently proposed pan-European typology for freshwater systems (BRT) has unified the assortment of existing national typologies but lacks biological validation. Using cluster validity analyses, we evaluated whether freshwater macroinvertebrate communities are more similar within than among the BRT river types. To compare the typology's performance, we conducted the same analyses with three additional freshwater typologies. In addition, we derived lists of taxa that typically occur within the types of each typology and evaluated their similarities. Using the assayed typologies as surrogates, we also compared two fundamentally different approaches to classifying rivers: reach-based and region-based typologies. Though performance differed between validity metrics, the BRT typology rarely ranked among the best. This was confirmed by our finding that, macroinvertebrate faunas of different BRT types were more similar to each other than in alternative typologies. Overall, regional typologies achieved better scores than those based on river reaches. However, all typologies failed to achieve high cluster validity scores. Our findings indicate that superior alternatives to the new pan-European rivers typology already exist but also that none reliably delineate demarcations between internally homogeneous macroinvertebrate reference communities. Hence, it remains important to further improve freshwater typologies to improve biomonitoring and facilitate conservation efforts.

3 Introduction

Freshwater ecosystems provide humanity with clean drinking water and food, they ensure livelihoods for fisheries and farms, and they are places of high recreational and aesthetic value (Youn *et al.* 2014; Béné *et al.* 2016; Börger *et al.* 2021). Combined, these contributions to people, also called ecosystem services, are estimated to have an annual worth of more than four trillion dollars (Costanza *et al.* 2014), which is approximately a twentieth of the global GDP as of 2017 (World Bank 2021). Human activities have reduced biodiversity in freshwater ecosystems (Darwall *et al.* 2018; WWF 2020) and are thereby endangering these systems’ capacities to provide ecosystem services (e.g. Cardinale 2011; Duffy *et al.* 2017; Grizzetti *et al.* 2019).

The Living Planet Index indicates that populations of freshwater vertebrates on average declined by 84% relative to their 1970 levels (but see Leung *et al.* (2020)). Based on the same data, He *et al.* (2019) found a slightly stronger decline (88%) in freshwater megafauna. Declining abundances have also been recorded for invertebrates (e.g. Cumberlidge *et al.* 2009; Baranov *et al.* 2020), however, most studies find stable or ambiguous trends in abundance and richness (Floury *et al.* 2013; Outhwaite *et al.* 2020; van Klink *et al.* 2020) though often associated with strong taxonomic and functional turnover (Jourdan *et al.* 2018; Haubrock *et al.* 2020; Mouton *et al.* 2020). The main stressors to which this situation can be attributed are well known and can be summarized as overexploitation, water pollution, flow modification, habitat degradation and species invasions (Dudgeon *et al.* 2006; Vörösmarty *et al.* 2010). This set is complemented by emerging new stressors such as microplastics and light pollution (Reid *et al.* 2018) and complex interactions between co-occurring stressors (Schinegger *et al.* 2012; Birk *et al.* 2020; Comte *et al.* 2021). To prevent unacceptable harm to freshwater ecosystems and secure the supply of ecosystem services, internationally coordinated action and regulatory frameworks are necessary (Darwall *et al.* 2018; Rees *et al.* 2020; Tickner *et al.* 2020).

The Water Framework Directive (WFD) is among the major regulatory frameworks that aim to protect and restore freshwater systems in the European Union. It aims to reach or maintain a ‘good ecological status’ for all water bodies in the European Union (European Commission 2000). Since its implementation in 2000, the WFD has motivated improvements in methods for biomonitoring (Birk *et al.* 2012; Carvalho *et al.* 2019), advanced our knowledge of the ecological status of freshwater systems throughout Europe (EEA 2018) and has stimulated relevant research at the science-policy interface (Reyjol *et al.* 2014). It has not, however, reached its explicit aim, to achieve ‘good ecological quality’ for most surface and groundwater bodies by 2015. As of 2018, only 40% of surface water bodies had reached this status (EEA 2018). The member states of the EU use national methods to assess their water

bodies. To harmonize the national assessment methods and to ensure a coherent notion of ‘good ecological quality,’ an intercalibration process has been initiated. A recent output of the intercalibration process is a pan-European typology of lentic and lotic freshwater systems (broad river types (BRT), Lyche Solheim *et al.* 2019). While the WFD required its members to create river typologies for the assessment, these typologies differ widely in the number of river types they discern (between 1 and 367) and the variables used to define them (between 2 and 16). The recent effort by Lyche Solheim *et al.* (2019) combined the national typologies and connected the new, broad river types with the established intercalibration types (Birk *et al.* 2013; Poikane *et al.* 2014). This pan-European typology has already been used to compare nutrient threshold values across river types (Poikane *et al.* 2019), and in studies examining interaction types of multiple stressors (Birk *et al.* 2020) and the chemical pollution of European surface waters (Posthuma *et al.* 2020).

The BRT-typology adopts the type descriptors that are most commonly used in national typologies: catchments size, altitude, geology, region and flow. The first three descriptors are part of the System A approach outlined in the Annex II of the WFD and Lyche Solheim *et al.* (2019) largely follow the thresholds proposed there. While System A includes an ecoregion descriptor that recognizes 25 distinct regions throughout Europe, the region descriptor used by Lyche Solheim *et al.* (2019) only separates between the Mediterranean and the rest of Europe. Flow is also considered as a binary variable (perennial / intermittent) and was only applied to Mediterranean rivers. Whether this typology based on geographical and environmental attributes matches biological classifications remains unknown. Verdonshot (2006b) suggests that System A-type typologies only partially capture the major drivers or macroinvertebrate distributions in Europe, which he found to be temperature, slope/ flow velocity and river size. The status assessment of the WFD relies on stable type-specific reference conditions. If biological communities vary strongly between reference sites of the same river type erroneous assessments resulting in over- or underprotection are more likely. If no reliable type-specific reference conditions can be established, e.g. due to high natural variability, a river type might be excluded from the status assessment (European Commission 2000).

Following the BRT-typology, we classified observations from a data set of freshwater macroinvertebrate occurrences to find the taxa that commonly occur in each river type. These typical taxa can be used to define reference conditions or to compare the sensitivity of communities towards stressors between river types. In addition, we used cluster validity metrics to compare the BRT typology to three other freshwater typologies. While BRT classifies river reaches, earlier typologies classified ecoregions. Two of the typologies we evaluated were such regional typologies. They are simpler to use in practical application since their basic spatial units are larger and the water body type is easier to determine in the field. Because

practical applications is the main purpose of freshwater typologies, reach-based typologies need to have better classification performance to legitimate their added complexity.

4 Methods

4.1 Preparation of macroinvertebrate data

We compiled a data set of macroinvertebrate samples from rivers throughout Europe (Table 1). A visual summary of the steps described below is provided in Figure 1. The data included 152572 samples taken between 1968 and 2020. We assigned the BRT-type of the closest river segment to samples if said segments was within 500m of the sample. If a sample was further than 500m removed from the next segment, we omitted it. Removing samples with a distance larger than 500 m to the next segment reduced the number of samples to 76621 (50% of all samples). Next, we removed all samples from sites that were considerably impaired. The exclusion criterion depended on the data set, as some included information on environmental conditions while others solely included biological information. When environmental information was lacking, we identified impaired sites using the catchment land cover. Sampling sites from catchments where the combined area of urban and agricultural land exceeded 20% of the catchment area were considered impaired (Malaj *et al.* 2014; Schreiner *et al.* 2016; Szöcs *et al.* 2017; Wolfram *et al.* 2021). We removed these potentially impaired sites because differences between river types are more pronounced between least-impaired sites (Verdonschot 2006a) and because we wanted to compare the typologies' capability to delineate reference assemblages with low spatial variability. Removing impaired sites further decreased the number of samples to 23284 (15% of all samples). Lastly, we only considered the 21169 samples that were taken in or after the year 2000 (14% of all samples). The spatial distribution of the final data set is shown in Figure 2. We provide maps of the spatial distribution of removed sites in the supplementary information.

Table 1: The data sets that we used in the analysis. The provided number of samples is before we removed unsuitable sites. The exclusion criterion gives the variables and thresholds we used to evaluate whether a sites is impaired. For more detailed information on these criteria, we refer the interested reader to the supplementary information.

Data set	Area	Time span	Number of Samples	Exclusion criterion	Reference
Monitoring data from German federal agencies	Germany	1968 - 2013	65211	anthropogenic land use > 20%	Bhowmik & Schäfer (2015), Berger et al. (2018), Le et al. (2021)
Ebro Hydrographic Confederation	Spain	2004 - 2019	3668	IASPT* < 4.5 & Land use	Escribano et al (2018)
Naiades	France	2002 - 2020	27052	BDI† < 14.5 & Land use	Alvarez et al. (2010), Alvarez et al. (2011)
Cantabria	Spain	2005	55	anthropogenic land use > 20%	
Picos de Europa	Spain	2015 - 2017	24	anthropogenic land use > 20%	
Ecosurv	Hungary	2005	491	anthropogenic land use > 20%	Schmera & Baur (2011), Schmera et al. (2012)
RIVPACS database	UK	1978 - 2004	2504	original site selection	Clark et al. (2003), Turley et al. (2014)
STAR	Europe	2002 - 2003	91	classification in data set & anthropogenic land use > 20	Verdonschot (2006a), Johnson et al. (2007)
WISER	Europe	2000 - 2008	2565	anthropogenic land use > 20%	Lyche-Solheim et al. (2013)
AQUEM Sweden	Sweden	2000	150	anthropogenic land use > 20%	Verdonschot (2006a)
Koutajoki drainage basin	Finnland	2000 - 2013	322	original site selection	Huttunen et al. (2017)
Monitoring data from the RCS national network managed by the French water agencies	France	2007 - 2013	2694	qualitative ratings of multiple anthropogenic stressors	Mondy & Usseglio-Polaterra (2013), Alric et al. (2021)
Monitoring data from Dutch regional water authorities	Netherlands	1978 - 2017	44702	anthropogenic land use > 20%	Peters et al. (2013)
Monitoring data from the German Federal state of Hesse	Germany	2005 - 2018	3044	anthropogenic land use > 20%	Haubrock et al. (2020)

* Iberian Average Score per Taxon; see Leunda et al. (2009) or Munné & Prat (2009)

† Biological Diatom Index; see Lenoir & Coste (1996)

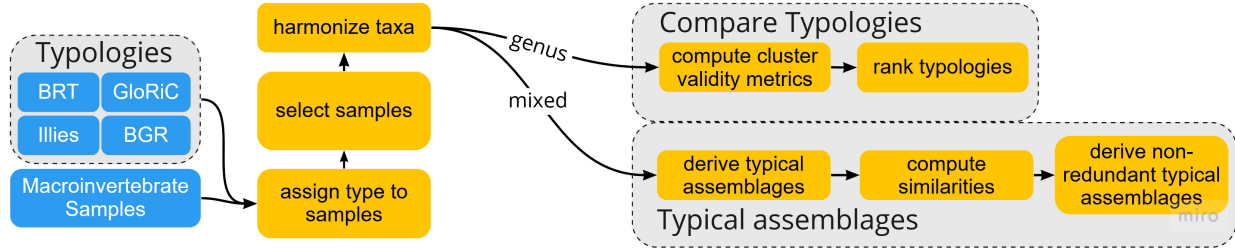


Figure 1: Schema of the data analysis. Blue boxes represent data sets and yellow boxes analyses. Names of typologies have been abbreviated to BRT (broad river typology), GloRiC (Global River Classification), Illies (Illies freshwater ecoregions) and BGR (Biogeographic Regions).

The resulting data set had sufficient data to represent 13 of 20 broad river types. We considered our samples representative for a river type, if we had more than 20 samples from the river type and the samples covered most of the area covered by the river type. Maps of the sampling locations for the individual river types are provided in the supplementary information. We had insufficient samples of the river types: RT6, 7, 12, 13, 17, 19 and 20, consequently, these were omitted from further analyses.

The data originated from different sources and required adjustments to ensure taxonomic consistency. We used the Global Biodiversity Information Facility (GBIF) to replace synonyms with accepted names and to assign a taxonomic level to each observation. To integrate the different data types, we transformed all data to presence-absence. Next, we restricted observations to the classes Clitellata, Insecta, Malacostraca, Bivalvia and Gastropoda, as they occurred in all data sets. When harmonizing taxonomies across data sets, there is a

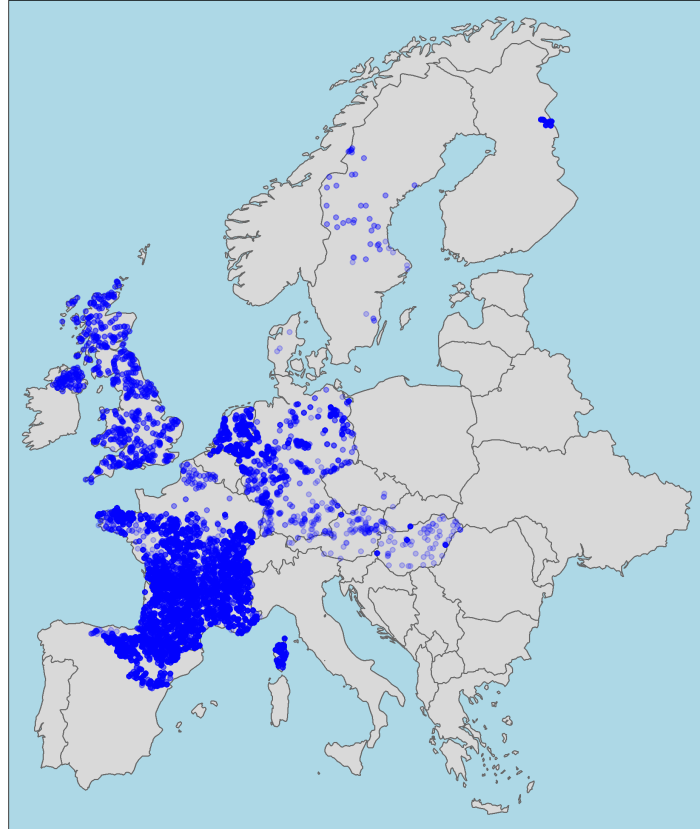


Figure 2: Map of samples used in the analyses. Each dot represents one sampling event. Darker dots indicate several overlapping sampling events which can be due spatially close sampling events or multiple sampling events at one site (e.g. yearly samples).

trade-off between comparability among data sets and taxonomic resolution. If we employ a single taxonomic level (e.g. genus) for all taxa across river types, we have to agglomerate observations with finer resolutions and omit those with coarser resolutions. The resulting data set would be appropriate for comparisons between river types because taxa have consistent representations across river types. If we vary the taxonomic resolution between river types and taxa, we artificially increase the dissimilarity between river types but attain the best available representation for the individual river types. To compare freshwater typologies (see Comparison of typologies), we used the first approach and harmonized all observations at the genus level, omitting all observations with coarser taxonomic levels. This resulted in a data set with 786 genera. To established typical assemblages (see Typical assemblages), we used the second approach. Typical assemblages were derived independently for each type, which allowed us to set an appropriate taxonomic level per river type and taxon. For each taxon, we calculated the percentage of observations represented at each finer taxonomic resolution. If this value was below 50 %, we harmonized all observations to this level. If it was above 50%, we repeated the procedure with the next finer level. We also included all observations with coarser than optimal resolutions. The resulting data set contained 370 taxa represented at the species level, 558 taxa represented at genus level and 186 taxa represented at family or coarser taxonomic levels.

4.2 Comparison of typologies

4.2.1 General approach

Lyche Solheim *et al.* (2019) proposed two versions of the BRT typology. BRT20 differentiates between 20 different river types, some of which are very rare. To reduce the number of broad river types, Lyche Solheim *et al.* (2019) merged rare types with common types that matched the rare type in at least two type descriptors. The resulting typology (BRT12) has twelve types. Previous publications have used both (e.g. Posthuma *et al.* (2020) used BRT12 and Poikane *et al.* (2019) used BRT20) so we consider both.

To evaluate the two versions of BRT, we compared them with the three other freshwater typologies: i) the k-means typology of the global river classification framework (GloRiC, Ouellet Dallaire *et al.* 2019); ii) the freshwater ecoregions proposed by Illies (1978); and iii) the biogeographical regions proposed by the European Environmental Agency (BGR, EEA 2016). These typologies can be categorized into: reach-based typologies (BRT and GloRiC) and regional typologies (Illies and BGR). Reach-based typologies classify single river reaches. The instances of individual types are not spatially contiguous and can be far apart. They are commonly very close to instances of different types. By contrast, regional typologies

classify large contiguous areas and there is only one instance per type. Different types are only adjacent at the boundaries between two types. As with BRT before, we retained those types for which we had more than 20 samples. See table 2 for the remaining types in each typology.

For each typology, we computed four cluster validity metrics, each emphasizing a different aspect of cluster quality: the average silhouette width, the Calinski-Harabasz index, the indicator value score and the classification strength. We used the Jaccard distance between sites as a distance metric.

As a reference frame for these for metrics, we created two additional typologies. To have an upper bound of what to expect, we clustered the sampling sites with an agglomerative clustering method. This typology is constrained neither by environment nor by space but only by the biological data. Thus, we expect it to delineate more clearly between different biological assemblages than the other typologies and to achieve better cluster validity scores. We will refer to this typology as the biological typology. All agglomerative clustering methods begin with as many clusters as observations and then they iteratively combine the most similar observations into clusters until only one cluster remains. The approaches differ in how this similarity is computed. Lance & Williams (1967) clarified that many approaches can be expressed as different parametrizations of one method: flexible beta clustering. We computed biological typologies with different number of types and a β parameters of 0.25. Based on two cluster validity metrics, we decided to use nine types. We refer the interested reader to the supplementary information for more detailed information.

As a lower bound of what to expect, we created 100 random typologies. Differences between the types of these typologies are purely random. There should be no systematic differences between the types. We hence expect these random typologies to receive lower cluster validity scores than all of the other typologies. For each random typology, we first drew the number of types as a random variable from the interval between the lowest number of types in any of the typologies (6 in BGR) and the highest number (14 in GloRiC). Then we assigned each sample randomly to one of the groups.

4.2.2 Cluster validity metrics

Here, we provide a short primer on each cluster validity metric we used. More extended information is provided in the supplementary information. The average silhouette width (ASW, Kaufmann & Rousseeuw 1990) is the average of all silhouette widths. A silhouette width is the difference between the average distance of sample i to all samples from the most similar type and the average distance of sample i to all other samples from the same

type. The silhouette width is scaled so it lies between -1 and 1. Values close to 1 indicate that most samples are more similar to samples from their own type than to those of other types. Values around zero suggest that the sample lies between two types and values close to -1 signal that the sample is more similar to samples from another type than to those from its own. In addition to the commonly applied ASW, we also computed two modified versions, which summarize the distance in a_i and b_i with the generalized mean instead of the arithmetic mean (Lengyel & Botta-Dukát 2019). The generalized mean has one modifiable parameter, p , and for $p = 1$ the modifiable mean is equal to the arithmetic mean. As we increase p the weight of outlying samples increases. We used $p = -\infty$ which corresponds to the minimum distance to compute ASW_{min} and $p = \infty$ which returns the maximum distance to compute ASW_{max} .

The Calinski-Harabasz Index (CH, Caliński & Harabasz 1974) is the scaled ratio of the distances between types and the distances within types. Higher values indicate that distances between types are larger than those within types and thus that samples are more similar to samples from their own type than to samples from other types.

The indicator value score (IVS) is based on the indicator value (IndVal) proposed by Dufrêne & Legendre (1997). IVS is the fraction of taxa that are statistically significant indicators (at a significance level of 0.01) for some type of a typology. High scores show that the commonness and specificity of taxa differ between types. Therefore higher scores indicate better typologies.

Lastly, we computed the classification strength (CS, Van Sickle 1997). Classification strength is the difference between mean within cluster similarity (\overline{W}) and mean between cluster (\overline{B}) similarity. We calculated similarities as 1 - Jaccard distance. The CS ranges between 0 ($\overline{W} = \overline{B}$) and 1 ($\overline{B} = 0$), where higher values indicate a stronger classification. CS has been used in similar analyses (e.g. Gerritsen *et al.* 2000; Van Sickle & Hughes 2000; Vasconcelos *et al.* 2013), which enables us to directly compare our results with those of previous studies.

We ranked the typologies based on each of these four cluster validity metrics, from 1 to 5 relating to the worst and best performances respectively. Differences smaller than 5% of the range between the biological and the mean of the random typologies were regarded as ties. When typologies were tied, they all received the mean of their untied scores. For example, if three typologies are tied for the first place, instead of 5, 4 and 3 points, each receives a score of 4.

4.3 Typical assemblages

We established typical assemblages for the types of all five freshwater typologies using the data set with mixed taxonomic levels. We define a typical taxon as one that frequently occurs in a type and typical assemblages as the lists of these taxa. Typical taxa differ from indicator taxa which usually need to be common and type-specific. As a measure of commonness, we used the percentage of samples from a type that included a taxon. We considered species as typical if they occurred in more than 20% of samples. Coarser taxonomic levels needed to be more common to be typical. Genera needed to occur in at least 33% of samples and families or coarser taxonomic levels in at least 66%. Having derived typical assemblages that way, we assessed how similar they were within typologies. Low similarities show that the common taxa differ between types, whereas high similarities imply that the same taxa frequently occur irrespective of type. As similarity metric, we used the Jaccard similarity. When the similarity between type A and type B is 0.5, half of the taxa in type A's typical assemblage also occur in type B's. Note that these similarities are asymmetric since they relate to the number of taxa in a typical assemblage. If type B's typical assemblage has more taxa, the similarity between B and A will be lower than that between A and B.

Lastly, we derived non-redundant typical assemblages. Two typical assemblages were considered redundant when their similarities exceeded 0.7 both ways. We iteratively combined the most similar river types until no redundant types remained. For example, the BRT20-types RT8 and 9 (mid-altitude siliceous rivers) were redundant and combined into RT8_9. Subsequently, RT8_9 was found to be redundant with RT16 (glacial rivers) so these types were combined to RT8_9_16. This way, we can evaluate whether the combinations of types from BRT20 to BRT12 are justified by biological homogeneity. In the supplementary information, we provide lists of all complete and non-redundant typical assemblages.

4.4 Data and Software

As geospatial representation of the BRT typology, we used the data provided by Globevnik (2019) (Accessed 11.12.2019). The version of the GloRiC typology we used is available under <https://www.hydrosheds.org/page/gloric> (Accessed 30.07.2019). Illies freshwater ecoregions were queried from <https://www.eea.europa.eu/data-and-maps/data/ecoregions-for-rivers-and-lakes> and the EEA biogeographical regions from <https://www.eea.europa.eu/data-and-maps/data/biogeographical-regions-europe-3> (both Accessed 08.01.2020). We used the Catchment Characterization and Modeling database (Vogt *et al.* 2007) and the Corine Landcover data from 2018 (CLC 2018) to derive the relative area of land use types in catchments.

All computations were conducted in the R Statistical Environment v. 4.0.3 (R Core Team 2020). Data were prepared using data.table 1.14.0 (Dowle & Srinivasan 2021), tidyverse packages (Wickham *et al.* 2019) and taxize 0.9.98 (Scott Chamberlain & Eduard Szocs 2013; Chamberlain *et al.* 2020). Geospatial analyses were conducted using sf (Pebesma 2018). Clusters were created and evaluated with fpc (Hennig 2020), indicpecies (Caceres & Legendre 2009). Generalized silhouette widths were computed with the R functions provided in the supplementary information of Lengyel & Botta-Dukát (2019). Figures and maps were created with ggplot2 (Wickham 2016) and tmap (Tennekes 2018).

5 Results

5.1 Comparison of classifications

Overall Illies received the highest score, with 18 of 20 possible points, followed by BGR (14.5), BRT12 (12.5), GloRiC (11) and BRT20 (7.5). The CH was higher in regional typologies than in reach-based typologies and higher in BGR than in Illies' ecoregions (Figure 3). BRT12 had a higher score than BRT20 and GloRiC had the lowest CH. Most typologies had similar CS, approximately 0.04. Only Illies' ecoregions exceeded 0.05 with a score of 0.07. Illies' ecoregions had the highest IVS closely followed by GloRiC. Their scores were similar to that of the biological typology. The three other typologies had similar IVS around 0.09. Increasing the emphasis on compactness (i.e. p) decreased the silhouette widths. All typologies had positive ASW_{min} , the regional typologies had scores similar to that of the biological typology. Reach-based typologies had lower ASW_{min} than regional ones and GloRiC's was higher than those of the BRT-typologies. With $p = 1$, only the ASW of BGR exceeded zero and hence the random typology. The difference between region- and reach-based typologies narrowed. Illies' ecoregions had a slightly higher ASW than BRT12, which had a higher ASW than BRT20. GloRiC had the lowest ASW. The ASW_{max} was negative for all typologies, including the biological. It declined most noticeably for BGR, Illies and GloRiC. BRT12 had the highest ASW_{max} . It was followed by BGR, BRT20, Illies and GloRiC.

5.2 Typical assemblages

The mean similarity in BRT-type typical assemblages was higher than in the other three typologies (Figure 4).

For both BRT-type typical assemblages the mean similarity was above 0.5, approaching 0.75 in BRT12. The mean similarity for the other three typical assemblages centered around 0.2.

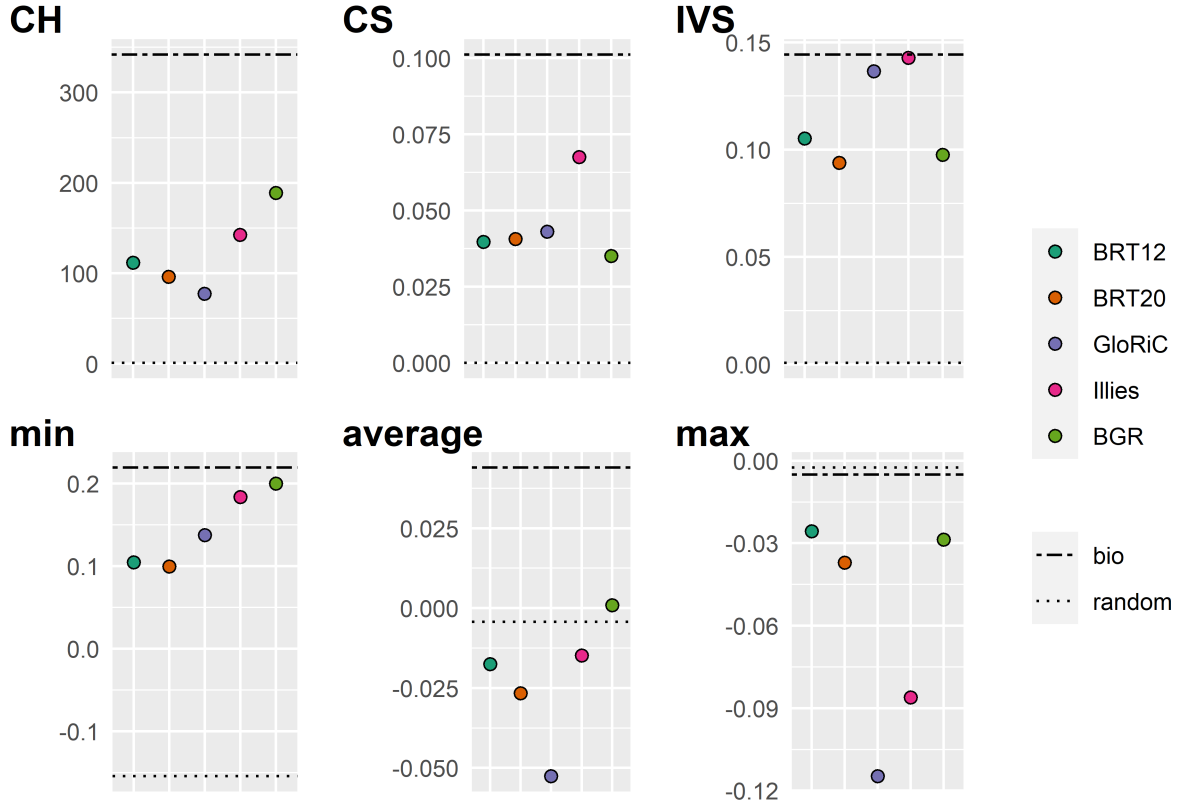


Figure 3: Cluster validity criteria for freshwater typologies. The points represent the different typologies which are indicated by their color. The biological typology and the mean of random typologies are represented by lines. Each plot shows a different cluster validity criterion: Calinski Harabasz (CH), Classification Strength (CS), Indicator Value score (IVS), average minimal silhouette width (min), average silhouette width (ASW) and average maximal silhouette width.

While we found high similarities (>0.75) in the typical assemblages of all five typologies, only for the BRT-type typologies did we find complete overlap (a similarity of 1).

The mean size of typical assemblages was similar in all typologies. The mean size over all typologies was 28.4 taxa. BGR had the smallest typical assemblages on average with 25.3 taxa and Illies the largest with 30.8 taxa. The differences are more pronounced in the sizes of individual typical assemblages. The two smallest typical assemblage were from the high-altitude calcareous/ mixed rivers (RT15) of BRT20 with ten taxa and the Mediterranean Region of BGR with eleven. The two largest typical assemblages were both from types of Illies' freshwater ecoregions: Ireland and Northern Ireland as well as England with 66 and 62 taxa respectively. Note that the two types were also similar to one another and were combined for the non-redundant typical assemblages.

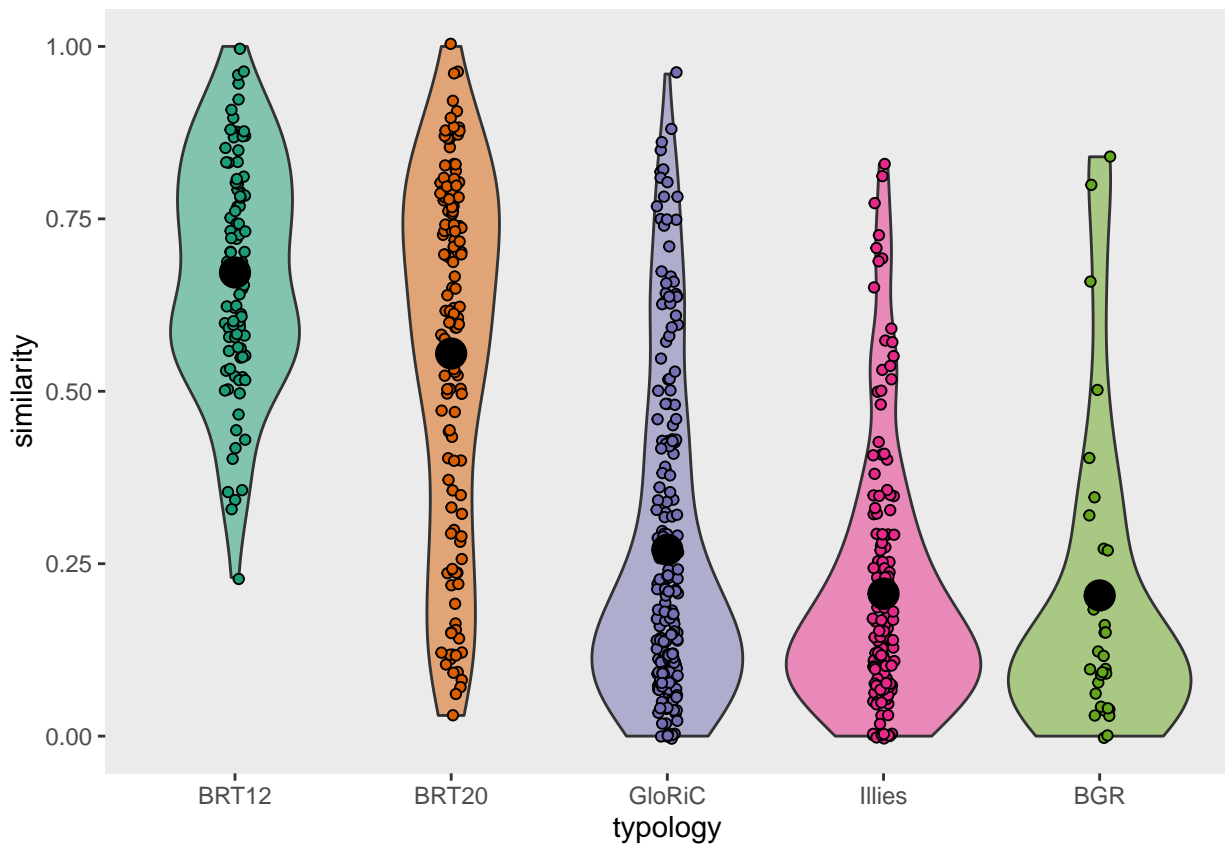


Figure 4: Similarity between typical assemblages for five freshwater typologies. The colored dots are similarities between two types and the violins represent their frequency distributions. The bigger black dots are the arithmetic means.

The number of taxa across typical assemblages of a typology varied considerably between typologies. The typical assemblages included 5.7% of taxa for BRT12, 7% for BRT20, 12.5%

for GloRiC, 11.1% for Illies and 8.3% for BGR. On average, taxa occurred in 4.4 typical assemblages in BRT12, 4.1 in BRT20, 2.6 in GloRiC, 2.2 in Illies and 1.5 in BGR. The two taxa that occurred in the most typical assemblages were both families: Chironomidae (49 of 56 typical assemblages) and Simuliidae (40). The most common genera were the mayfly *Baetis* (38), the caddisfly *Hydropsyche* (37) and the gammarid *Gammarus* (35). The most common species were two mayflies *Serratella ignita* (34) and *Baetis rhodani* (14).

Based on these typical assemblages, we found redundant types in all typologies except for BGR. In BRT12, we combined the river types 2, 3, 4, 5, 6 and 7 which include very small to large lowland rivers as well as calcareous mid-altitude rivers. We also combined the two river types 8 and 9 - siliceous mid-altitude rivers. In BRT20, we found the river types 8, 9 and 16 (mid-altitude siliceous and glacial rivers) to be redundant. In addition, we combined the river types 2, 3, 4, 5, 10 and 11. They represent very small to large lowland rivers and mid-altitude calcareous and mixed rivers. In GloRiC, we combined two river types, 4 and 6. Unfortunately, Ouellet Dallaire *et al.* (2019) do not provide an interpretation for their types beyond these numbers so neither can we.

For Illies, we merged the two types England and Ireland and Northern Ireland as well as Western plains and Western highlands.

6 Discussion

We used four cluster validity metrics to compare the partitioning of macroinvertebrate samples proposed by five different freshwater typologies. Three typologies were reach-based (BRT12, BRT20 and GloRiC) and two were regional (Illies and BGR). As a reference frame for the cluster validity metrics, we also computed a biological typology of the samples and 100 random typologies. As we expected, the cluster validity scores of the five freshwater typologies lay in the interval between biological and random typologies in most cases. The only exceptions were silhouette widths (Figure 3).

6.1 Comparison of typologies

Overall, the freshwater ecoregions proposed by Illies (1978) performed best in our assessment followed by BGR, BRT12, GloRiC and BRT20. Illies' freshwater ecoregions had the highest IVS and classification strength. While GloRiC is a close second in the IVS, it has the lowest scores in CH and ASW. Between the two BRT typologies, BRT12 always received better or comparable scores and had the highest ASW_{max} of all typologies.

The average silhouette widths were remarkably low for all typologies. Only BGR had a positive ASW and a higher score than the random typologies. The negative ASWs indicate that most communities are on average more similar to communities from other types than to those from their own. ASW_{min} gives the largest weight to the separation of clusters without considering compactness. It clearly shows a stronger separation between the types of regional typologies than between those of reach-based ones. It is also at this level that all typologies have positive ASWs and that the difference to random partitions is largest. The differences between regional- and reach-based typologies and between all typologies and random typology decrease with increasing p . The negative ASW_{max} indicates that for each of the typologies, there is some community within their type that is more dissimilar to it, than the most dissimilar community from the next closest type. Overall, we see that the typologies discriminate between different macroinvertebrate faunas. They all have CH, CS and IVS larger than purely random typologies. However, most types seem to have very similar, neighboring types. The assignment of individual communities to a specific type is therefore far from clear, as communities vary as strongly within types as among neighboring ones.

This result is confirmed by the high similarity between some typical assemblages, especially in the BRT typologies. In BRT12, the non-redundant typical assemblages only differentiate between four of the ten types and in BRT20 between six of thirteen. While we also found redundancies between typical assemblages in other typologies, they were never between more than two types. In both BRT-typologies, we combined low-land river types of different sizes and geologies with calcareous or mixed-geology mid-altitude streams of different sizes.

It is possible, that more finely resolved taxonomic data would be necessary to show differences between these types. An alternative explanation is, that the BRT typology has too few types. Our analysis identified mostly those taxa as common that are sufficiently euryoecious that their distribution is largely independent from the features that delineate the different river types. Within the individual river types only few type-specific taxa were common. This could indicate that even within a single river type environmental conditions vary sufficiently to limit the distribution of more stenooecious species. It is also possible that common anthropogenic stressors homogenized the assemblages (Mondy & Usseglio-Polatera 2014). We removed samples from impaired sites to prevent this but the definition of reference conditions and their identification after the fact and at large spatial scales, are difficult (Herlihy *et al.* 2008; Pardo *et al.* 2012; McNellie *et al.* 2020). However, this would have increased redundancies in all typologies not just the BRT-typologies.

We had insufficient data to evaluate some of the combinations made to arrive at BRT12 from BRT20. In most cases, Lyche Solheim *et al.* (2019) merged river types, because one of them was rare (e.g. RT7 or RT20). Consequently, these were also the river types that had to

be omitted from the analysis due to the low number of available samples. The high altitude river types in RT14 to 16 are an exception to the rule above. They were combined into RT10 (Highland and glacial rivers) in BRT12, on the grounds that altitude was assumed to be more important than geology or size. However, our results do not unequivocally support the combination to a single high-altitude river type. The typical assemblages of highland river types were not redundant in BRT20. The corresponding types in Illies (Pyrenees, Alps, Caparthiens and Borealic uplands) or GloRiC (8, 15, 22, 30) were also not redundant. However, BGR also only has one high-altitude-type (Alpine) and while it did receive a lower overall score, it received higher scores than Illies for CH and all three variants of ASW. As BRT12 generally performed better than BRT20 in this study, the combinations made by Lyche Solheim *et al.* (2019) seem to be justified. We suggest that future studies should rather employ the aggregated than the full typology. However, given the low performance of reach-based typologies in general, we advise caution when assuming spatially homogeneous reference communities within them.

6.2 Typical assemblages

In the following the typical assemblages of the BRT-type typologies are discussed. The other typical assemblages are available in the supplementary information but as this paper focuses on the BRT-typologies, we will discuss them in more detail. Several taxa are tramp taxa (*sensu* McGeoch *et al.* (2002)). They are part of all typical assemblages. These taxa are: the snail *Ancylus*, the mayfly *Baetis*, the dipteran family Chironomidae, the caddisfly *Hydropsyche* and the beetle *Limnius*.

The typical assemblages of very large river (RT1 in BRT12 and BRT20) harbor many genera that include invasive species like *Dreissena polymorpha*, *Dikerogammarus villosus*, *Corbicula fluminea* and *Potamogyrus antipodarum*. The presence of these taxa in the typical assemblages highlights the prominent role that neozoa play very large rivers (Leitner *et al.* 2020). The typical assemblages of very large rivers are also the only ones that do not include the beetle *Elmis*, the caddisfly *Rhyacophila* and the mayfly *Serratella ignita*. Elmidae breathe through plastrons and thus require high oxygen saturation which is uncommon in very large rivers. The genus *Rhyacophila* encompasses a diverse array of species that occur from large lowland stream (e.g. *Rhyacophila dorsalis*, Dohet *et al.* 2008) to high-altitude streams (e.g. *Rhyacophila torrentium* and *Rhyacophila vulgaris*, Lorenz *et al.* 2004) but seldomly in very large streams. The typical assemblages of very large rivers contain the most specific taxa, i.e. taxa that do not occur in other typical assemblages. On one hand, these are most of the neozoa listed above. Besides *Potamogyrus*, which occurs in most typical assemblages, these taxa are only typical for very large rivers. On the other hand, there are taxa such as

the mayflies *Cleon*, *Procleon* and *Heptagenia* as well as the caddisfly *Oecetis*. These taxa all include taxa that are known to be potamophil, e.g. *Cleon dipterum*, *Heptagenia flava* and *Oecetis notata* (Nagell *et al.* 1978; Westermann 1997; Dohet *et al.* 2008). The siliceous lowland river types of RT4/5 (BRT12) and RT2/3 (BRT20) are mostly made up of caddisfly, mayfly and beetle genera. These genera include several species that are typical for the rhithral, such as *Esolus parallelepipedus*, *Elmis aenea* and *Hydropsyche pellucidula*. In addition, all four contained *Pisidium* clams and *Calopteryx* damselflies. In the typical assemblages of calcareous or mixed lowland river of RT2/3 (BRT12) and RT 4/5 (BRT20) we can find many of the same taxa as for the siliceous types. The most notable differences are the presence of *Riolus* beetles, the damselfly *Platycnemis* and the gammarid *Echinogammarus* as well as the absence of the mayfly *Ecdyonurus* and the stonefly *Leuctra*. *Riolus* species, such as *Rioulus cupreus*, are known to prefer calcareous streams (Hebauer 1986; Lorenz *et al.* 2004; Elliott 2008). The differences are more pronounced when we move towards mid-altitude river types. The typical assemblages of siliceous mid-altitude rivers in RT 8/9 (BRT12 and BRT20) have unique stonefly taxa (*Nemoura* and *Perla*), a unique beetle taxon (*Dupophilus*) and two unique mayfly taxa (*Ephemerella* and *Habrophlebia*). *Ecdyonurus*, which are included in the lowland siliceous river types but absent from the calcareous ones, are also included. The typical assemblages of calcareous and mixed mid-altitude rivers (RT6/7 in BRT12 and RT10/11 in BRT20) are similar to those of lowland calcareous streams. Both have the mayfly *Centropilum* and the snail *Radix* in the large but not the small rivers. The most noticeable differences are that the gammarid *Echinogammarus* is part of both lowland typical assemblages and absent from the mid-altitude types while the opposite holds for the dipteran family *Simuliidae*. The high altitude types RT10 (BRT12) as well as RT14, 15 and 16 (BRT20) contain no unique taxa. If at all, these typical assemblages are distinct through the taxa they lack. The mussel *Pisidium*, the gammarid *Gammarus*, the mayfly *Ephemera* and the damselfly *Calopteryx* occur in all typical assemblages except for these.

6.3 Ecology of regional and reach-based typologies

Overall, regional approaches that assign large contiguous areas to a single type were a better summary of our extensive data set of invertebrate occurrences than reach-scale typologies. The constraints of our study may have played in favor of regional approaches. The taxonomic resolution of our data only allowed inferences at the genus level. Potentially, species-level data would have highlighted bigger differences at smaller scales and hence been better summarized by typologies with smaller mapping units. Along the same lines, Moog *et al.* (2004) showed that higher taxonomic resolution was necessary to delineate smaller ecoregions and the same logic could expand to reach-scale typologies. Similarly, Verdonshot (2006a) showed

that differences in classification strength were noticeable between ‘best-available’ and family-level data on a pan-European scale. However, the difference was small and their analysis should not be extrapolated to infer if this difference also exists between species- and genus-level data. Hawkins *et al.* (2000) note that, across taxa groups and typologies, higher taxonomic resolution does not always lead to greater classification strength.

Additionally, the likelihood of assigning an observation to the wrong type due to inaccuracies in the spatial data are lower for regional typologies, where such mistakes can only happen toward the borders between types. We reduced this error by using a more stringent distance-to-river criterion for the selection of sites than similar studies. For example, Irving *et al.* (2018) used a cutoff of 3 km to assign streamflow gauging stations to rivers. Very small rivers are largely missing from the river network representations used by GloRiC and the BRT as provided in Globevnik (2019). Hence, samples from these rivers are either omitted from the analysis if they are too far removed from the next river or assigned to the next river segment. In most cases, this is likely but not certainly a reach of a similar type. New and more highly resolved river networks (e.g. Lin *et al.* 2021) might provide more accurate mapping of river types and ameliorate these problems.

It is worthwhile to interpret the patterns we observed in light of metacommunity theory. Our analysis of generalized silhouette widths showed that the separation between communities of different BRT types decreases when we give more weight to outliers. The separation declines up to a point where it is lower than in the random typologies. This pattern could be explained by strong dispersal limitation leading to divergent communities in comparable environments. The spatial distance between the instances of different types in reach-scale typologies is often small. Typically, landscapes and catchments are made up of mosaics of different river types. In positing that assemblages vary between instances of different reach-scale types, we implicitly assume that dispersal is strong enough for species to track variation in local conditions (Leibold *et al.* 2004) but not so strong that mass effect can overrule environmentally induced patterns (Mouquet & Loreau 2003). If taxa are strongly limited by dispersal e.g. through historical and anthropogenic dispersal barriers (Leibold *et al.* 2010; Belletti *et al.* 2020) or through the sheer distance between sites, they are unable to reach potentially favorable sites. In our case, stronger dispersal limitation would lead to larger differences between sites of the same type. Dispersal limitation leads to a pattern known as distance decay, which is often found at large spatial scales (Nekola & White 1999; Morlon *et al.* 2008) but varies in its magnitude between taxa, realms and degrees of latitude (Graco-Roza *et al.* 2021). Studies have found both, invertebrate communities that are predominantly structured by environment (e.g. Heino *et al.* 2012; Landeiro *et al.* 2012) and those that are structured by space (Mykrä *et al.* 2007; Astorga *et al.* 2012). However, even for microbes, which under the Bass-Becking hypothesis (“everything is everywhere but the

environment selects”) were long considered to have cosmopolitan distributions (e.g. Finlay *et al.* 2002), the role of dispersal limitation, at least at large spatial scales, is becoming evident (Telford *et al.* 2006; Lindström & Langenheder 2012; Soininen 2012).

The better performance of regional typologies might also be taken to support the notion, that aquatic macroinvertebrate communities are adapted to changes in large-scale rather than small-scale variables. Many studies have investigated this question before. Among them are studies that have studied the explanatory potential of ecoregions, which might be considered as categorical combinations of large-scale environmental conditions like climate, geology and altitude. Verdonschot & Nijboer (2004) used the data compiled during the AQEM project to investigate potential drivers of macroinvertebrate community composition. They found that the distribution of macroinvertebrates follows Illies’ freshwater ecoregions and that large-scale variables like geology explain most of the variation. However, using an extended data set, that included data from both AQUEM and STAR, Verdonschot (2006b) found ecoregions to be a rather weak predictor of invertebrate community composition. Similarly, ecoregions explained a negligible amount of variance in community composition in Swedish boreal rivers (Johnson *et al.* 2004). The review of Hawkins *et al.* (2000) finds that ecoregion-based approaches usually outperform catchment-scale classifications. However, the classification strength was generally low and lay between 0.07 and 0.16, which slightly exceeds the ones we found (between 0.04 to 0.07) but is in agreement with several other studies (e.g. Snelder *et al.* 2004; Heino & Mykrä 2006; Mykrä *et al.* 2009; Vasconcelos *et al.* 2013). The interplay between local and regional control cannot be considered detached from the previous discussion of dispersal limitation since the degree to which assemblages are regulated regionally or locally likely depends on the magnitude of dispersal (Ryabov & Blasius 2011). Synthesis of these findings is impeded by the fact that many of the studies relied on combinations of variance partitioning and eigenfunction analyses that have been shown to be flawed (Gilbert & Bennett 2010; Smith & Lundholm 2010; Tuomisto *et al.* 2012). An alternative method to evaluate the importance of dispersal processes for community assembly was recently proposed (Vilmi *et al.* 2020) but has yet to be extensively tested and applied.

Typologies based on environmental variables cannot account for the effects of biotic interactions on species distributions. While these received much attention in the early days of macroecology (Diamond 1975; Connor & Simberloff 1983) the focus shifted towards environmental filters (e.g. Whittaker 1970; Pearson & Dawson 2003) assuming that biotic interactions are only important at local scales. Their role for macroecology has been revisited (e.g. Wisz *et al.* 2013) partly due to the realization that many studies can not differentiate between environmental filtering *sensu stricto* and biotic interactions (Kraft *et al.* 2015). Propelled by a series of new methods (e.g. Popovic *et al.* 2019; Ovaskainen &

Abrego 2020) several studies have recently investigated the effect of biotic interactions on larger biogeographical patterns. While some indicate, that the role of environmental conditions is more important than that of biotic interactions (Elo *et al.* 2021), others suggest that biotic interactions modulate the response to the abiotic environment (Abrego *et al.* 2021).

6.4 Conclusions and outlook

While there were clear and systematic differences between typologies, all cluster validity metrics had comparatively low scores. It is beyond the scope of the current study to assess whether this is a problem of the implementation or a general issue, i.e. if it is theoretically possible to create environmental typologies that capture large amounts of variation between communities or if niche processes or environmental filters *sensu stricto* (Kraft *et al.* 2015) generally do not explain enough variance. The latter would constitute a serious problem for the reference state approach as this would entail that spatial stability of reference communities (Statzner *et al.* 2001) should generally not be assumed. However, since several studies have found typologies to explain a significant amount of variation (e.g. Lorenz *et al.* 2004; Johnson *et al.* 2007) we deem issues with the implementation more likely.

The river typology of the WFD must be optimized not only for macroinvertebrates but for all four biological quality elements (macroinvertebrates, diatoms, fish and macrophytes) (Dodkins *et al.* 2005). Future studies should assess whether the conclusions drawn from this study are supported by results for the other taxa groups. River classifications must be understood as models (Goodwin 1999; Loveland & Merchant 2004). Alongside our growing knowledge about freshwater systems, we should strive to further improve them, as to increase their value for biomonitoring and to render them more effective tools for conservation. Our study has shown that the BRT typologies should not be the end of this effort.

7 Supplementary Information

7.1 Exclusion criteria

In this section the exclusion criteria from Table 1 of the main text are described in more detail.

anthropogenic land use > 20% catchment area. We used the Corine Land Cover data from 2018 to assign relative areas of land cover categories to each catchment in the Catchment Characteristics and Modelling data base of watersheds. On the first level there are five different categories in Corine Land Cover: 1. Artificial surfaces, 2. Agricultural areas, 3. Forest and semi natural areas, 4. Wetlands and 5. Water bodies. We combined the fractions of category 1 and 2 to obtain a measure of anthropogenic land use. If the relative catchment area of anthropogenic land use exceeded 20% all sites in that catchment were categorized as impaired.

IASPT The Iberian average score per taxon was introduced by Alba-Tercedor & Sánchez-Ortega (1988) who developed the original metric proposed by Party (1978) and Armitage *et al.* (1983). The score that the name refers to is the biological monitoring working party (BMWP) score. Many freshwater invertebrate families were assigned scores between 1 (indicator of strongly impaired conditions) and 10 (indicator of least-impaired conditions). For example, Oligochaetes have a low score and Heptageniidae, a caddisfly family, a high score. The BMWP is the sum of scores for all taxa that occur at a site. Since the absolute score depends on the number of taxa, general thresholds for good or bad conditions can not be derived. This problem is solved by taking the average score of a sites (ASPT). Scores above 4.5 are generally taken to indicate good conditions.

BDI The Biological Diatom Index was proposed by Lenoir & Coste (1996) and is a standardized method used in France to monitor the biological quality of running waters. The original score was based on 209 diatom taxa and their ecological profiles. These profiles are probability distributions over the occurrence of a given species in the seven quality classes. By multiplying the presence probabilities for all taxa at a site one can derive the quality class which a site belongs to. The score is then mapped on a scale from 0 to 20 where 20 represents the highest ecological quality. Coste *et al.* (2009) further improved the index by extending the taxon pool to 838 species and updating the ecological profiles of the available taxa. Scores above 14.5 are generally taken to indicate good conditions.

original site selection. Two of the data sets we used resulted from studies that only took samples from least-impacted sites. We did not exclude any sites from these sites on the basis of anthropogenic impairment.

classification in data set The STARS data set provided a qualitative rating of ecological conditions that ranged from high over good, moderate and poor to bad. We rated sites with a good or high rating as least-impaired.

qualitative ratings of multiple anthropogenic stressors. The data from the RCS monitoring program provided qualitative rating of several anthropogenic stressors. They included organic matter, nitrogen compounds (except nitrates), nitrates (mg/l NO₃), phosphorous compounds, suspended matter, organic pollutants, mineral micropollutants, pesticides, and polycyclic aromatic hydrocarbon. Each stressor received a rating as high, good, intermediate, poor or bad. We considered all sites where two or more variables were rated as intermediate or one received a rating of poor or bad as impaired.

7.2 Distribution of removed sampling sites

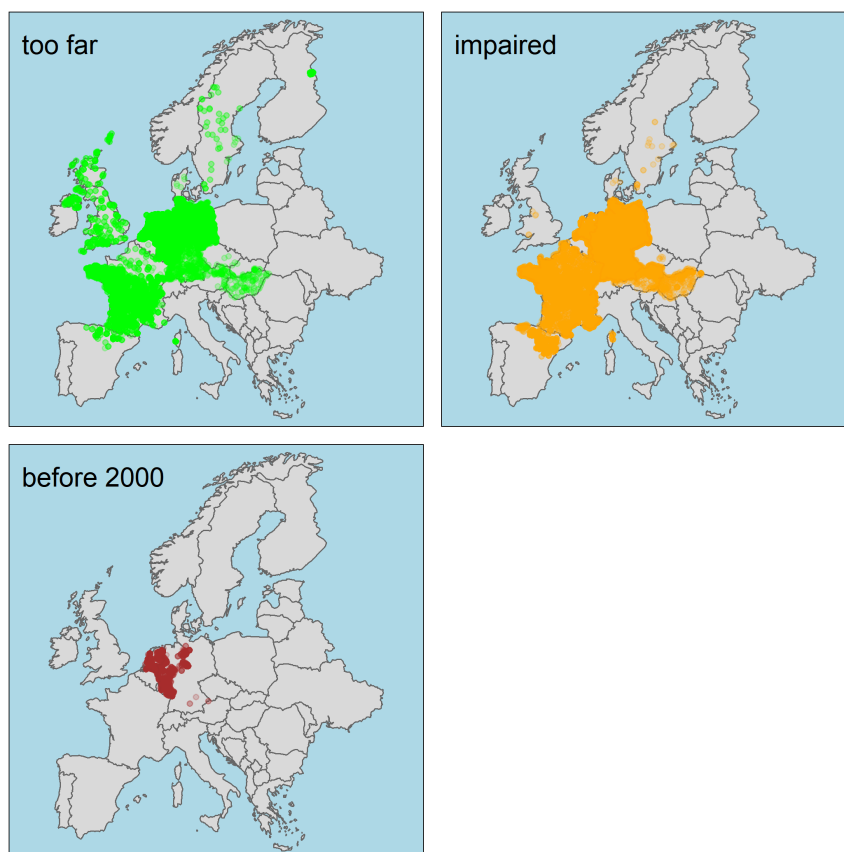


Figure 5: Maps of removed sampling sites. The first map (too far) shows sites that were more than 500 meters removed from the closeded BRT river segment. The second map (impaired) shows sites that were categorized as impaired. The third map (before 2000) shows sites that were sampled before 2000.

7.3 Flexible beta clustering

This section describes in greater detail how we computed the flexible beta clustering of the sampling sites. We used flexible beta clustering as it has been shown to be robust to various potential problems like outliers or noisy data (Milligan 1989). Flexible beta clustering has four parameters but it is common to only specify one, β , while the other three parameters are derived from it. We employed the common parameterization of ($\alpha_1 = \alpha_2 = \alpha, \beta = (1 - \alpha)/2, \gamma = 0$) and used a beta of 0.25. We cut the resulting dendograms at each split between 4 and 30 final nodes. For each of the resulting clusterings we calculated the average silhouette width and the cluster membership entropy. The average silhouette width is already explained in the main text as well as in the following section on cluster validity metrics. The cluster membership entropy measures how similar the number of observation within different groups is. Clusterings with only few clusters tend to produce high silhouette widths by singling out individual outliers. Such solutions have very low entropies. To correct this tendency, we looked for a solutions that had high scores for both entropy and silhouette width. While entropy increased monotonically with the number of clusters (Figure 6) without obvious jumps or dips, the average silhouette width started very low (4 clusters) then increased to a stable higher level (6-8 clusters) and lastly decreased sharply in three steps.

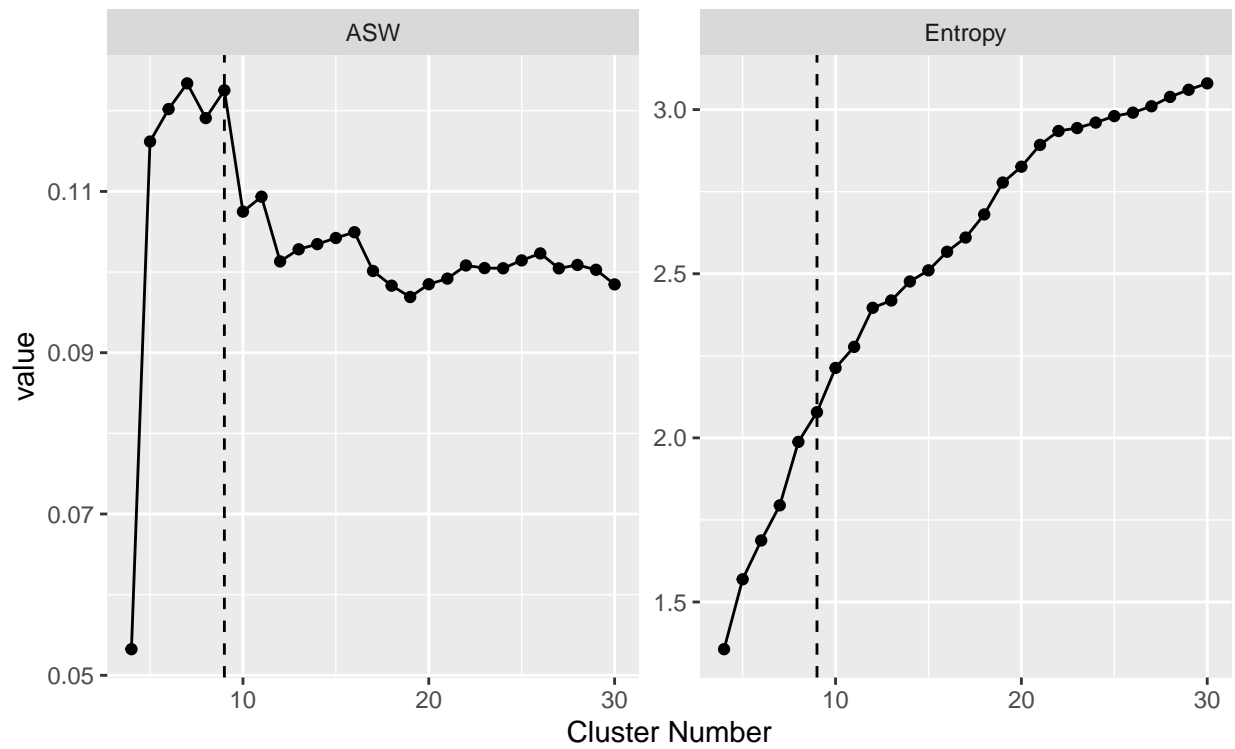


Figure 6: Average silhouette width and entropy of flexible beta clustering of the macroinvertebrate data with cluster numbers between 4 and 30. The black dashed horizontal line highlights the number of clusters we chose.

7.4 Cluster validity metrics

7.4.1 Average silhouette width

The silhouette width $s(i)$ is defined as follows:

1. Compute a_i , the average dissimilarity between sample i and all other samples from the same type
2. Among all other clusters C , find the cluster than minimizes $d(i, C)$ the average distance of i to all observations of C . This cluster is b ; $b_i = \min_C(d(i, C))$.
3. $s(i) = \frac{b_i - a_i}{\max(a_i, b_i)}$
4. The average silhouette width is the arithmetic mean of all silhouette widths. $ASW = \frac{1}{n} \sum_{i=1}^n s(i)$

Lengyel & Botta-Dukát (2019) recently proposed a generalized version of the ASW. The classic ASW employs the arithmetic average to compute a_i and b_i , which assumes that spherical clusters are optimal. Using a generalized mean instead, we can flexibly adjust our validity metric to put a stronger emphasis on compactness (a_i) or separation (b_i). The generalized mean of degree p (M^p) is computed as:

$$M^p(\mathbf{x}) = \left(\frac{1}{n} \sum_{i=1}^n x_i^p \right)^{1/p}$$

The generalized mean can take the value of common summary statistics such as the minimum ($p = -\infty$), the maximum ($p = \infty$) or the harmonic mean ($p = -1$). For example, for $p = -\infty$ the silhouette width is the difference between the minimum distance of observation i to any other observation from the same type and the minimum distance from that observation to any observation from the next closest type. This perspective excludes outliers and values separation over compactness. The weighting shifts towards compactness as we increase p .

7.4.2 Calinski Harabasz Index

The Calinski-Harabasz Index (CH, Caliński & Harabasz 1974) is computed as

$$CH = \frac{BGSS}{WGSS} \times \frac{n - k}{k - 1}$$

where $BGSS$ is the squared sum of distances between group centroids and the overall centroid (between group sum-of-squares), $WGSS$ is sum of squares of distances between observations

of one group (within group sum-of-squares), k is the number of clusters. High values indicate that variation within types is smaller than between types. As the second term controls for the degrees of freedom, here determined by the number of clusters, it can be understood as an analog to the F-Statistic. The algorithm assumes Euclidean data but good performance with a similar metric was shown for binary data in the context of fMRI-scans (Dimitriadou *et al.* 2004).

7.4.3 Indicator Value Statistic

The indicator value score (IVS) is based on the Indicator value (IndVal) proposed by Dufrêne & Legendre (1997). The IndVal can be understood as the product of the two quantities A and B . For our purposes, A is the relative number of observations of taxon i that are within type j . It was initially described as specificity (Dufrêne & Legendre 1997) but is better understood as concentration (Podani & Csányi 2010) because it is independent of the total number of types. B is the relative frequency with which species i occurs in type j . The maximum score is assigned to a species that only occurs in one type ($A = 1$) and occurs in all samples of that type ($B = 1$). We used the group-equalized version of the IndVal which accounts for the fact that the number of samples differs between types. The statistical significance of the IndVal statistic was assessed with a permutation test that computes IndVal values for random permutations of sites and types and compares the observed IndVal against this empirical distribution. We determined statistical significance with 999 permutations and a significance level of 0.01. The IndVal score is the fraction of statistically significant taxa X type combinations. A score of 1 would indicate that every taxon tested is a statistically significant indicator for every type and a score of 0 that no taxon is an statistically significant indicator for any type.

7.4.4 Analysed river types/ ecoregions

Table 2: Types per typology that were analysed. We had insufficient data for the types that are not shown in this table.

Typology	Types
BRT12	Very large rivers (RT1); Lowland, calcareous or mixed, medium-large (RT2); Lowland, calcareous or mixed, very small-small (RT3); Lowland, siliceous incl. organic, medium-large (RT4); Lowland, siliceous incl. organic, medium-large (RT5); Mid-altitude, calcareous incl. organic, medium-large (RT6); Mid-altitude, calcareous or mixed, very small-small (RT7); Mid-altitude, siliceous incl. organic, medium-large (RT8); Mid-altitude, siliceous incl. organic, very small-small (RT9); Highland and glacial (RT10)
BRT20	Very large rivers (RT1); Lowland, siliceous, medium-large (RT2); Lowland, siliceous, very small-small (RT3); Lowland, calcareous or mixed, medium-large (RT4); Lowland, calcareous or mixed, very small-small (RT5); Mid-altitude, siliceous, medium-large (RT8); Mid-altitude, siliceous, very small-small (RT9); Mid-altitude, calcareous or mixed, medium-large (RT10); Mid-altitude, calcareous or mixed, very small-small (RT11) Highland (all Europe), siliceous, incl. organic (humic) (RT14); Highland (all Europe), calcareous/mixed (RT15); Glacial rivers (all Europe) (RT16); Mediterranean, mid altitude, medium-large, perennial (RT18)
GloRiC	3, 4, 6, 7, 8, 14, 15, 17, 18, 22, 24, 26, 29, 30
Illies	Alps, Central highlands, Central plains, England, Fenno-scandian shield, Hungarian lowlands, Ibero-Macaronesian region, Ireland and Northern Ireland, Italy and Corsica, Pyrenees, Western highlands, Western plains
BGR	Alpine Bio-geographical Region, Atlantic Bio-geographical Region, Boreal Bio-geographical Region, Continental Bio-geographical Region, Mediterranean Bio-geographical Region, Pannonian Bio-geographical Region

References

- Abrego, N., Roslin, T., Huotari, T., Ji, Y., Schmidt, N.M., Wang, J., *et al.* (2021). Accounting for species interactions is necessary for predicting how arctic arthropod communities respond to climate change. *Ecography*, 1–12.
- Alba-Tercedor, J. & Sánchez-Ortega, A. (1988). Un método rápido y simple para evaluar la calidad biológica de las aguas corrientes basado en el de hellawell (1978). *Limnetica*, 4, 1–56.
- Alric, B., Dézerald, O., Meyer, A., Billoir, E., Coulaud, R., Larras, F., *et al.* (2021). How diatom-, invertebrate- and fish-based diagnostic tools can support the ecological assessment of rivers in a multi-pressure context: Temporal trends over the past two decades in France. *Science of the Total Environment*, 762, 143915.
- Alvarez-Cabria, M., Barquin, J. & Juanes, J.A. (2010). Spatial and seasonal variability of macroinvertebrate metrics: Do macroinvertebrate communities track river health? *Ecological Indicators*, 10, 370–379.
- Alvarez-Cabria, M., Barquin, J. & Juanes, J.A. (2011). Macroinvertebrate community dynamics in a temperate european atlantic river. Do they conform to general ecological theory? *Hydrobiologia*, 658, 277–291.
- Armitage, P., Moss, D., Wright, J. & Furse, M. (1983). The performance of a new biological water quality score system based on macroinvertebrates over a wide range of unpolluted running-water sites. *Water research*, 17, 333–347.
- Astorga, A., Oksanen, J., Luoto, M., Soininen, J., Virtanen, R. & Muotka, T. (2012). Distance decay of similarity in freshwater communities: Do macro- and microorganisms follow the same rules? *Global Ecology and Biogeography*, 21, 365–375.
- Baranov, V., Jourdan, J., Pilotto, F., Wagner, R. & Haase, P. (2020). Complex and nonlinear climate-driven changes in freshwater insect communities over 42 years. *Conservation Biology*, cobi.13477.
- Belletti, B., Garcia de Leaniz, C., Jones, J., Bizzi, S., Börger, L., Segura, G., *et al.* (2020). More than one million barriers fragment Europe’s rivers. *Nature*, 588, 436–441.
- Berger, E., Haase, P., Schäfer, R.B. & Sundermann, A. (2018). Towards stressor-specific macroinvertebrate indices: Which traits and taxonomic groups are associated with vulnerable and tolerant taxa? *Science of the Total Environment*, 619, 144–154.
- Béné, C., Arthur, R., Norbury, H., Allison, E.H., Beveridge, M., Bush, S., *et al.* (2016). Contribution of fisheries and aquaculture to food security and poverty reduction: Assessing

- the current evidence. *World Development*, 79, 177–196.
- Bhowmik, A.K. & Schaefer, R.B. (2015). Large scale relationship between aquatic insect traits and climate. *PLoS one*, 10, e0130025.
- Birk, S., Bonne, W., Borja, A., Brucet, S., Courrat, A., Poikane, S., *et al.* (2012). Three hundred ways to assess Europe’s surface waters: An almost complete overview of biological methods to implement the Water Framework Directive. *Ecological Indicators*, 18, 31–41.
- Birk, S., Chapman, D., Carvalho, L., Spears, B.M., Andersen, H.E., Argillier, C., *et al.* (2020). Impacts of multiple stressors on freshwater biota across spatial scales and ecosystems. *Nature Ecology & Evolution*.
- Birk, S., Willby, N.J., Kelly, M.G., Bonne, W., Borja, A., Poikane, S., *et al.* (2013). Inter-calibrating classifications of ecological status: Europe’s quest for common management objectives for aquatic ecosystems. *Science of the Total Environment*, 454-455, 490–499.
- Börger, T., Campbell, D., White, M.P., Elliott, L.R., Fleming, L.E., Garrett, J.K., *et al.* (2021). The value of blue-space recreation and perceived water quality across europe: A contingent behaviour study. *Science of The Total Environment*, 145597.
- Caceres, M.D. & Legendre, P. (2009). Associations between species and groups of sites: inindices and statistical inference. *Ecology*, 90, 3566–3574.
- Caliński, T. & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3, 1–27.
- Cardinale, B.J. (2011). Biodiversity improves water quality through niche partitioning. *Nature*, 472, 86–89.
- Carvalho, L., Mackay, E.B., Cardoso, A.C., Baattrup-Pedersen, A., Birk, S., Blackstock, K.L., *et al.* (2019). Protecting and restoring Europe’s waters: An analysis of the future development needs of the Water Framework Directive. *Science of the Total Environment*, 658, 1228–1238.
- Chamberlain, S., Szoecs, E., Foster, Z., Arendsee, Z., Boettiger, C., Ram, K., *et al.* (2020). *Taxize: Taxonomic information from around the web*.
- Clarke, R.T., Wright, J.F. & Furse, M.T. (2003). RIVPACS models for predicting the expected macroinvertebrate fauna and assessing the ecological quality of rivers. *Ecological modelling*, 160, 219–233.
- CLC. (2018). CORINE land cover. <https://land.copernicus.eu/pan-european/corine-land-cover>.

- Comte, L., Grantham, T. & Ruhi, A. (2021). Human stabilization of river flows is linked with fish invasions across the USA. *Global Ecology and Biogeography*, 725–737.
- Connor, E.F. & Simberloff, D. (1983). Interspecific competition and species co-occurrence patterns on islands: Null models and the evaluation of evidence. *Oikos*, 455–465.
- Costanza, R., De Groot, R., Sutton, P., Van der Ploeg, S., Anderson, S.J., Kubiszewski, I., *et al.* (2014). Changes in the global value of ecosystem services. *Global environmental change*, 26, 152–158.
- Coste, M., Boutry, S., Tison-Rosebery, J. & Delmas, F. (2009). Improvements of the Biological Diatom Index (BDI): Description and efficiency of the new version (BDI-2006). *Ecological Indicators*, 9, 621–650.
- Cumberlidge, N., Ng, P.K., Yeo, D.C., Magalhães, C., Campos, M.R., Alvarez, F., *et al.* (2009). Freshwater crabs and the biodiversity crisis: Importance, threats, status, and conservation challenges. *Biological Conservation*, 142, 1665–1673.
- Darwall, W., Bremerich, V., De Wever, A., Dell, A.I., Freyhof, J., Gessner, M.O., *et al.* (2018). The Alliance for Freshwater Life: A global call to unite efforts for freshwater biodiversity science and conservation. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 28, 1015–1022.
- Diamond, J.M. (1975). Ecology and evolution of communities. In: *Ecology and evolution of communities* (ed. Cody, J.M., M. L. & Diamond). Harvard University Press, pp. 342–444.
- Dimitriadou, E., Barth, M., Windischberger, C., Hornik, K. & Moser, E. (2004). A quantitative comparison of functional MRI cluster analysis. *Artificial Intelligence in Medicine*, 31, 57–71.
- Dodkins, I., Rippey, B., Harrington, T.J., Bradley, C., Chathain, B.N., Kelly-Quinn, M., *et al.* (2005). Developing an optimal river typology for biological elements within the Water Framework Directive. *Water Research*, 39, 3479–3486.
- Dohet, A., Ector, L., Cauchie, H.M. & Hoffmann, L. (2008). Identification of benthic invertebrate and diatom indicator taxa that distinguish different stream types as well as degraded from reference conditions in Luxembourg. *Animal Biology*, 58, 419–472.
- Dowle, M. & Srinivasan, A. (2021). *Data.table: Extension of ‘data.frame’*.
- Dudgeon, D., Arthington, A.H., Gessner, M.O., Kawabata, Z.I., Knowler, D.J., Lévêque, C., *et al.* (2006). Freshwater biodiversity: Importance, threats, status and conservation challenges. *Biological Reviews*, 81, 163–182.

- Duffy, J.E., Godwin, C.M. & Cardinale, B.J. (2017). Biodiversity effects in the wild are common and as strong as key drivers of productivity. *Nature*, 549, 261–264.
- Dufrêne, M. & Legendre, P. (1997). Species assemblages and indicator species: The need for a flexible asymmetrical approach. *Ecological monographs*, 67, 345–366.
- EEA. (2016). Biogeographical Regions.
- EEA. (2018). European waters assessment of status and pressures 2018. *EEA Report*.
- Elliott, J.M. (2008). The Ecology of Riffle Beetles (Coleoptera: Elmidae). *Freshwater Reviews*, 1, 189–203.
- Elo, M., Jyrkänkallio-Mikkola, J., Ovaskainen, O., Soininen, J., Tolonen, K.T. & Heino, J. (2021). Does trait-based joint species distribution modelling reveal the signature of competition in stream macroinvertebrate communities? *Journal of Animal Ecology*.
- Escribano, N., Oscoz, J., Galicia, D., Cancellario, T., Durán, C., Navarro, P., *et al.* (2018). Freshwater macroinvertebrate samples from a water quality monitoring network in the Iberian peninsula. *Scientific data*, 5, 1–6.
- European Commission. (2000). DIRECTIVE 2000/60/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 23 October 2000 establishing a framework for Community action in the field of water policy. *Official Journal of the European Union*, L327.
- Finlay, B.J., Monaghan, E.B. & Maberly, S.C. (2002). Hypothesis: The rate and scale of dispersal of freshwater diatom species is a function of their global abundance. *Protist*, 153, 261–273.
- Floury, M., Usseglio-Polatera, P., Ferreol, M., Delattre, C. & Souchon, Y. (2013). Global climate change in large European rivers: Long-term effects on macroinvertebrate communities and potential local confounding factors. *Global Change Biology*, 19, 1085–1099.
- Gerritsen, J., Barbour, M.T. & King, K. (2000). Apples, oranges, and ecoregions: On determining pattern in aquatic assemblages. *Journal of the North American Benthological Society*, 19, 487–496.
- Gilbert, B. & Bennett, J.R. (2010). Partitioning variation in ecological communities: Do the numbers add up? *Journal of Applied Ecology*, 47, 1071–1082.
- Globevnik, L. (2019). Broad typology for rivers and lakes in Europe for large scale analysis.
- Goodwin, C.N. (1999). Fluvial classification: Neanderthal necessity or needless normalcy. *Wildland Hydrology*, 5230, 229–236.

- Graco-Roza, C., Aarnio, S., Abrego, N., Acosta, A.T., Alahuhta, J., Altman, J., *et al.* (2021). Distance decay 2.0—a global synthesis of taxonomic and functional turnover in ecological communities. *bioRxiv*.
- Grizzetti, B., Liqueste, C., Pistocchi, A., Vigiak, O., Zulian, G., Bouraoui, F., *et al.* (2019). Relationship between ecological condition and ecosystem services in European rivers , lakes and coastal waters. *Science of the Total Environment*, 671, 452–465.
- Haubrock, P.J., Pilotto, F. & Haase, P. (2020). Do changes in temperature affect EU Water Framework Directive compliant assessment results of central European streams? *Environmental Sciences Europe*, 32.
- Hawkins, C.P., Norris, R.H., Gerritsen, J., Hughes, R.M., Jackson, S.K., Johnson, R.K., *et al.* (2000). Evaluation of the Use of Landscape Classifications for the Prediction of Freshwater Biota : Synthesis and Recommendations. *Journal of the North American Benthological Society*, 19, 541–556.
- He, F., Zarfl, C., Bremerich, V., David, J.N.W., Hogan, Z., Kalinkat, G., *et al.* (2019). The global decline of freshwater megafauna. *Global Change Biology*, 25, 3883–3892.
- Hebauer, F. (1986). Käfer als bioindikatoren, dargestellt am Ökosystem bergbach. *Laufener Seminarbeiträge: Ausgewählte Referate zum Arteschutz*, 7, 55–65.
- Heino, J., Grönroos, M., Soininen, J., Virtanen, R. & Muotka, T. (2012). Context dependency and metacommunity structuring in boreal headwater streams. *Oikos*, 121, 537–544.
- Heino, J. & Mykrä, H. (2006). Assessing physical surrogates for biodiversity: Do tributary and stream type classifications reflect macroinvertebrate assemblage diversity in running waters? *Biological Conservation*, 129, 418–426.
- Hennig, C. (2020). *Fpc: Flexible procedures for clustering*.
- Herlihy, A.T., Paulsen, S.G., Van Sickle, J., Stoddard, J.L., Hawkins, C.P. & Yuan, L.L. (2008). Striving for consistency in a national assessment: The challenges of applying a reference-condition approach at a continental scale. *Journal of the North American Benthological Society*, 27, 860–877.
- Huttunen, K.-L., Mykrä, H., Oksanen, J., Astorga, A., Paavola, R. & Muotka, T. (2017). Habitat connectivity and in-stream vegetation control temporal variability of benthic invertebrate communities. *Scientific Reports*, 7, 1–9.
- Illies, J. (1978). *Limnofauna europaea*. Fischer Stuttgart.
- Irving, K., Kuemmerlen, M., Kiesel, J., Kakouei, K., Domisch, S. & Jähnig, S.C. (2018). A high-resolution streamflow and hydrological metrics dataset for ecological modeling using

- a regression model. *Scientific data*, 5, 1–14.
- Johnson, R.K., Furse, M.T., Hering, D. & Sandin, L. (2007). Ecological relationships between stream communities and spatial scale: Implications for designing catchment-level monitoring programmes. *Freshwater Biology*, 52, 939–958.
- Johnson, R.K., Goedkoop, W. & Sandin, L. (2004). Spatial scale and ecological relationships between the macroinvertebrate communities of stony habitats of streams and lakes. *Freshwater Biology*, 49, 1179–1194.
- Jourdan, J., O'Hara, R.B., Bottarin, R., Huttunen, K.L., Kuemmerlen, M., Monteith, D., *et al.* (2018). Effects of changing climate on European stream invertebrate communities: A long-term data analysis. *Science of the Total Environment*, 621, 588–599.
- Kaufmann, L. & Rousseeuw, P. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons.
- Kraft, N.J.B., Adler, P.B., Godoy, O., James, E.C., Fuller, S. & Levine, J.M. (2015). Community assembly, coexistence and the environmental filtering metaphor. *Functional Ecology*, 29, 592–599.
- Lance, G.N. & Williams, W.T. (1967). A general theory of classificatory sorting strategies: 1. Hierarchical systems. *The computer journal*, 9, 373–380.
- Landeiro, V.L., Bini, L.M., Melo, A.S., Pes, A.M.O. & Magnusson, W.E. (2012). The roles of dispersal limitation and environmental conditions in controlling caddisfly (trichoptera) assemblages. *Freshwater Biology*, 57, 1554–1564.
- Le, T.D.H., Schreiner, V.C., Kattwinkel, M. & Schäfer, R.B. (2021). Invertebrate turnover along gradients of anthropogenic salinisation in rivers of two German regions. *Science of the Total Environment*, 753, 141986.
- Leibold, M.A., Economo, E.P. & Peres-Neto, P. (2010). Metacommunity phylogenetics: Separating the roles of environmental filters and historical biogeography. *Ecology letters*, 13, 1290–1299.
- Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes, M.F., *et al.* (2004). The metacommunity concept: A framework for multi-scale community ecology. *Ecology letters*, 7, 601–613.
- Leitner, P., Borgwardt, F., Birk, S. & Graf, W. (2020). Multiple stressor effects on benthic macroinvertebrates in very large European rivers – A typology-based evaluation of faunal responses as a basis for future bioassessment. *Science of The Total Environment*, 143472.
- Lengyel, A. & Botta-Dukát, Z. (2019). Silhouette width using generalized mean—A flexible method for assessing clustering efficiency. *Ecology and Evolution*, 9, 13231–13243.

- Lenoir, A. & Coste, M. (1996). Development of a practical diatom index of overall water quality applicable to the french national water board network. In: *International symposium, volksbildungsheim grilhof vill, AUT, 17-19 september 1995*. Universität Innsbruck, pp. 29–43.
- Lenoir, A. & Coste, M. (1996). Development of a practical diatom index of overall water quality applicable to the french national water board network. In: *International symposium, volksbildungsheim grilhof vill, AUT, 17-19 september 1995*. Universität Innsbruck, pp. 29–43.
- Leunda, P.M., Oscoz, J., Miranda, R. & Ariño, A.H. (2009). Longitudinal and seasonal variation of the benthic macroinvertebrate community and biotic indices in an undisturbed Pyrenean river. *Ecological Indicators*, 9, 52–63.
- Leung, B., Hargreaves, A.L., Greenberg, D.A., McGill, B., Dornelas, M. & Freeman, R. (2020). Clustered versus catastrophic global vertebrate declines. *Nature*.
- Lin, P., Pan, M., Wood, E.F., Yamazaki, D. & Allen, G.H. (2021). A new vector-based global river network dataset accounting for variable drainage density. *Scientific Data*, 1–9.
- Lindström, E.S. & Langenheder, S. (2012). Local and regional factors influencing bacterial community assembly. *Environmental Microbiology Reports*, 4, 1–9.
- Lorenz, A., Feld, C.K. & Hering, D. (2004). Typology of streams in Germany based on benthic invertebrates: Ecoregions, zonation, geology and substrate. *Limnologica*, 34, 379–389.
- Lorenz, A., Feld, C.K. & Hering, D. (2004). Typology of streams in Germany based on benthic invertebrates: Ecoregions, zonation, geology and substrate. *Limnologica*, 34, 379–389.
- Loveland, T.R. & Merchant, J.M. (2004). Ecoregions and ecoregionalization: geographical and ecological perspectives. *Environmental management*, 34 Suppl 1, 1–13.
- Lyche Solheim, A., Austnes, K., Globevnik, L., Kristensen, P., Moe, J., Persson, J., *et al.* (2019). A new broad typology for rivers and lakes in Europe: Development and application for large-scale environmental assessments. *Science of the Total Environment*, 697, 134043.
- Lyche-Solheim, A., Feld, C.K., Birk, S., Phillips, G., Carvalho, L., Morabito, G., *et al.* (2013). Ecological status assessment of european lakes: A comparison of metrics for phytoplankton, macrophytes, benthic invertebrates and fish. *Hydrobiologia*, 704, 57–74.

- Malaj, E., Ohe, P.C. von der, Grote, M., Kühne, R., Mondy, C.P., Usseglio-Polatera, P., *et al.* (2014). Organic chemicals jeopardize the health of freshwater ecosystems on the continental scale. *Proceedings of the National Academy of Sciences*, 111, 9549–9554.
- McGeoch, M.A., Van Rensburg, B.J. & Botes, A. (2002). The verification and application of bioindicators: A case study of dung beetles in a savanna ecosystem. *Journal of Applied Ecology*, 39, 661–672.
- McNellie, M.J., Oliver, I., Dorrough, J., Ferrier, S., Newell, G. & Gibbons, P. (2020). Reference state and benchmark concepts for better biodiversity conservation in contemporary ecosystems. *Global Change Biology*, 26, 6702–6714.
- Milligan, G.W. (1989). A Study of the Beta-Flexible Clustering Method. *Multivariate Behavioral Research*, 24, 163–176.
- Mondy, C.P. & Usseglio-Polatera, P. (2013). Using conditional tree forests and life history traits to assess specific risks of stream degradation under multiple pressure scenario. *Science of the Total Environment*, 461–462, 750–760.
- Mondy, C.P. & Usseglio-Polatera, P. (2014). Using fuzzy-coded traits to elucidate the non-random role of anthropogenic stress in the functional homogenisation of invertebrate assemblages. *Freshwater Biology*, 59, 584–600.
- Moog, O., Schwalb, A.N., Ofenböck, T. & Gerritsen, J. (2004). Does the ecoregion approach support the typological demands of the EU ‘Water Framework Directive’? *Hydrobiologia*, 516, 21–33.
- Morlon, H., Chuyong, G., Condit, R., Hubbell, S., Kenfack, D., Thomas, D., *et al.* (2008). A general framework for the distance–decay of similarity in ecological communities. *Ecology letters*, 11, 904–917.
- Mouquet, N. & Loreau, M. (2003). Community patterns in source-sink metacommunities. *The american naturalist*, 162, 544–557.
- Mouton, T.L., Tonkin, J.D., Stephenson, F., Verburg, P. & Floury, M. (2020). Increasing climate-driven taxonomic homogenization but functional differentiation among river macroinvertebrate assemblages. *Global Change Biology*, 0–3.
- Munné, A. & Prat, N. (2009). Use of macroinvertebrate-based multimetric indices for water quality evaluation in Spanish Mediterranean rivers: An intercalibration approach with the IBMWP index. *Hydrobiologia*, 628, 203–225.
- Mykrä, H., Aroviita, J., Hämäläinen, H., Karjalainen, S.M., Visuri, M., Riihimäki, J., *et al.* (2009). Utility of a single a priori river typology for reference conditions of boreal macroinvertebrates and diatoms. *Fundamental and Applied Limnology*, 175, 269–280.

- Mykrä, H., Heino, J. & Muotka, T. (2007). Scale-related patterns in the spatial and environmental components of stream macroinvertebrate assemblage variation. *Global Ecology and Biogeography*, 16, 149–159.
- Nagell, B., Fagerström, T., Nagell, B. & Fagerstrom, T. (1978). Adaptations and Resistance to Anoxia in *Cloeon dipterum* (Ephemeroptera) and *Nemoura cinerea* (Plecoptera). *Oikos*, 30, 95.
- Nekola, J.C. & White, P.S. (1999). The distance decay of similarity in biogeography and ecology. *Journal of biogeography*, 26, 867–878.
- Ouellet Dallaire, C., Lehner, B., Sayre, R. & Thieme, M. (2019). A multidisciplinary framework to derive global river reach classifications at high spatial resolution. *Environmental Research Letters*, 14, 024003.
- Outhwaite, C.L., Gregory, R.D., Chandler, R.E., Collen, B. & Isaac, N.J.B. (2020). Complex long-term biodiversity change among invertebrates, bryophytes and lichens. *Nature Ecology & Evolution*, 4, 384–392.
- Ovaskainen, O. & Abrego, N. (2020). *Joint species distribution modelling: With applications in r*. Cambridge University Press.
- Pardo, I., Gómez-Rodríguez, C., Wasson, J.G., Owen, R., Bund, W. van de, Kelly, M.G., *et al.* (2012). The European reference condition concept: A scientific and technical approach to identify minimally-impacted river ecosystems. *Science of the Total Environment*, 420, 33–42.
- Party, B.M.W. (1978). Final report: Assessment and presentation of the biological quality of rivers in great britain. *Unpublished report*.
- Pearson, R.G. & Dawson, T.P. (2003). Predicting the impacts of climate change on the distribution of species: Are bioclimate envelope models useful? *Global ecology and biogeography*, 12, 361–371.
- Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, 10, 439–446.
- Peeters, E.T.M., Zuidam, J.P. van, Zuidam, B.G. van, Van Nes, E.H., Kosten, S., Heuts, P.G., *et al.* (2013). Changing weather conditions and floating plants in temperate drainage ditches. *Journal of Applied Ecology*, 50, 585–593.
- Podani, J. & Csányi, B. (2010). Detecting indicator species: Some extensions of the IndVal measure. *Ecological Indicators*, 10, 1119–1124.
- Poikane, S., Kelly, M.G., Salas Herrero, F., Pitt, J.A., Jarvie, H.P., Claussen, U., *et al.* (2019). Nutrient criteria for surface waters under the European Water Framework Di-

- rective: Current state-of-the-art, challenges and future outlook. *Science of the Total Environment*, 695, 133888.
- Poikane, S., Zampoukas, N., Borja, A., Davies, S.P., Bund, W. van de & Birk, S. (2014). Intercalibration of aquatic ecological assessment methods in the european union: Lessons learned and way forward. *Environmental Science & Policy*, 44, 237–246.
- Popovic, G.C., Warton, D.I., Thomson, F.J., Hui, F.K.C. & Moles, A.T. (2019). Untangling direct species associations from indirect mediator species effects with graphical models. *Methods in Ecology and Evolution*, 2019, 1571–1583.
- Posthuma, L., Zijp, M.C., De Zwart, D., Van de Meent, D., Globevnik, L., Koprivsek, M., *et al.* (2020). Chemical pollution imposes limitations to the ecological status of European surface waters. *Scientific Reports*, 1–20.
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rees, C.B. van, Waylen, K.A., Schmidt-Kloiber, A., Thackeray, S.J., Kalinkat, G., Martens, K., *et al.* (2020). Safeguarding Freshwater Life Beyond 2020: Recommendations for the New Global Biodiversity Framework from the European Experience. *Conservation Letters*, preprints202001.0212.v1.
- Reid, A.J., Carlson, A.K., Creed, I.F., Eliason, E.J., Gell, P.A., Johnson, P.T., *et al.* (2018). Emerging threats and persistent conservation challenges for freshwater biodiversity. *Biological Reviews*.
- Reyjol, Y., Argillier, C., Bonne, W., Borja, A., Buijse, A.D., Cardoso, A.C., *et al.* (2014). Assessing the ecological status in the context of the european water framework directive: Where do we go now? *Science of the Total Environment*, 497, 332–344.
- Ryabov, A.B. & Blasius, B. (2011). A graphical theory of competition on spatial resource gradients. *Ecology Letters*, 14, 220–228.
- Schinegger, R., Trautwein, C., Melcher, A. & Schmutz, S. (2012). Multiple human pressures and their spatial patterns in European running waters. *Water and Environment Journal*, 26, 261–273.
- Schmera, D. & Baur, B. (2011). Testing a typology system of running waters for conservation planning in Hungary. *Hydrobiologia*, 665, 183–194.
- Schmera, D., Baur, B. & Erős, T. (2012). Does functional redundancy of communities provide insurance against human disturbances? An analysis using regional-scale stream invertebrate data. *Hydrobiologia*, 693, 183–194.

- Schreiner, V.C., Szöcs, E., Bhowmik, A.K., Vijver, M.G. & Schäfer, R.B. (2016). Pesticide mixtures in streams of several European countries and the USA. *Science of the Total Environment*, 573, 680–689.
- Scott Chamberlain & Eduard Szocs. (2013). Taxize - taxonomic search and retrieval in r. *F1000Research*.
- Smith, T.W. & Lundholm, J.T. (2010). Variation partitioning as a tool to distinguish between niche and neutral processes. *Ecography*, 33, 648–655.
- Snelder, T.H., Cattaneo, F., Suren, A.M. & Biggs, B.J.F. (2004). Is the River Environment Classification an improved landscape-scale classification of rivers? *Journal of the North American Benthological Society*, 23, 580–598.
- Soininen, J. (2012). Macroecology of unicellular organisms—patterns and processes. *Environmental microbiology reports*, 4, 10–22.
- Statzner, B., Bis, B., Dolédec, S. & Usseglio-Polatera, P. (2001). Perspectives for biomonitoring at large spatial scales: A unified measure for the functional composition of invertebrate communities in European running waters. *Basic and Applied Ecology*, 2, 73–85.
- Szöcs, E., Brinke, M., Karaoglan, B. & Schäfer, R.B. (2017). Large Scale Risks from Agricultural Pesticides in Small Streams. *Environmental Science and Technology*, 51, 7378–7385.
- Telford, R.J., Vandvik, V. & Birks, H.J.B. (2006). Dispersal limitations matter for microbial morphospecies. *Science*, 312, 1015.
- Tennekes, M. (2018). tmap: Thematic maps in R. *Journal of Statistical Software*, 84, 1–39.
- Tickner, D., Opperman, J.J., Abell, R., Acreman, M., Arthington, A.H., Bunn, S.E., *et al.* (2020). Bending the Curve of Global Freshwater Biodiversity Loss: An Emergency Recovery Plan. *BioScience*, XX, 1–13.
- Tuomisto, H., Ruokolainen, L. & Ruokolainen, K. (2012). Modelling niche and neutral dynamics: On the ecological interpretation of variation partitioning results. *Ecography*, 35, 961–971.
- Turley, M.D., Bilotta, G.S., Extence, C.A. & Brazier, R.E. (2014). Evaluation of a fine sediment biomonitoring tool across a wide range of temperate rivers and streams. *Freshwater Biology*, 59, 2268–2277.
- van Klink, R., Bowler, D.E., Gongalsky, K.B., Swengel, A.B., Gentile, A. & Chase, J.M. (2020). Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances. *Science*, 368, 417–420.

- Van Sickle, J. (1997). Using Mean Similarity Dendrograms to Evaluate Classifications. *Journal of Agricultural, Biological and Environmental Statistics*, 2, 370–388.
- Van Sickle, J. & Hughes, R.M. (2000). Classification strengths of ecoregions, catchments, and geographic clusters for aquatic vertebrates in Oregon. *Journal of the North American Benthological Society*, 19, 370–384.
- Vasconcelos, M.C., Melo, A.S. & Schwarzbald, A. (2013). Comparing the performance of different stream classification systems using aquatic macroinvertebrates. *Acta Limnologica Brasiliensia*, 25, 406–417.
- Verdonschot, P.F.M. (2006a). Data composition and taxonomic resolution in macroinvertebrate stream typology. *Hydrobiologia*, 566, 59–74.
- Verdonschot, P.F.M. (2006b). Evaluation of the use of Water Framework Directive typology descriptors, reference sites and spatial scale in macroinvertebrate stream typology. *Hydrobiologia*, 566, 39–58.
- Verdonschot, P.F.M. & Nijboer, R.C. (2004). Testing the European stream typology of the Water Framework Directive for macroinvertebrates. *Hydrobiologia*, 516, 35–54.
- Vilmi, A., Gibert, C., Escarguel, G., Happonen, K., Heino, J., Jamoneau, A., *et al.* (2020). Dispersal–niche continuum index: a new quantitative metric for assessing the relative importance of dispersal versus niche processes in community assembly. *Ecography*, 1–10.
- Vogt, J., Soille, P., De Jager, A., Rimavičiūtė, E., Mehl, W., Foisneau, S., *et al.* (2007). *A pan-European river and catchment database*.
- Vörösmarty, C.J., McIntyre, P.B., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., *et al.* (2010). Global threats to human water security and river biodiversity. *Nature*, 467, 555.
- Vörösmarty, C.J., McIntyre, P.B., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., *et al.* (2010). Global threats to human water security and river biodiversity. *Nature*, 467, 555.
- Westermann, F. (1997). Bemerkenswerte funde potamobionter elmidae und ephemeroptera in rheinland-pfalz. *Lauterbornia*, 31, 67–72.
- Whittaker, R.H. (1970). Communities and ecosystems. *Communities and ecosystems*.
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., *et al.* (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4, 1686.

- Wisn, M.S., Pottier, J., Kissling, W.D., Pellissier, L., Lenoir, J., Damgaard, C.F., *et al.* (2013). The role of biotic interactions in shaping distributions and realised assemblages of species: Implications for species distribution modelling. *Biological Reviews*, 88, 15–30.
- Wolfram, J., Stehle, S., Bub, S., Petschick, L.L. & Schulz, R. (2021). Water quality and ecological risks in European surface waters – Monitoring improves while water quality decreases. *Environment International*, 152, 106479.
- World Bank. (2021). <https://data.worldbank.org/indicator/ny.gdp.mktc.cd>.
- WWF. (2020). *Living planet report 2020-bending the curve of biodiversity loss*. World Wildlife Fund.
- Youn, S.-J., Taylor, W.W., Lynch, A.J., Cowx, I.G., Beard Jr, T.D., Bartley, D., *et al.* (2014). Inland capture fishery contributions to global food security and threats to their future. *Global Food Security*, 3, 142–148.