Derivation of Typical Assemblages

Jonathan Jupke

October 22, 2020

Contents

1 Introduction

In GetReal, we investigate whether selected freshwater and terrestrial assemblages vary spatially and seasonally in their sensitivity towards chemicals and draw conclusions for risk assessment. After we agreed on using the Pan-European river typology proposed by Lyche Solheim *et al.* (2019) to delineate the receiving freshwater systems, the next step was to derive their typical assemblages (TA). Here we describe the methods we used to derive TAs of macroinvertebrates and diatoms for selected river types across Europe. We also describe and briefly discuss the results.

2 Harmonizing taxa names

International diatom occurrence datasets require extensive harmonization because of the taxonomic resolution differing between datasets, different working groups using different nomenclatures, identification errors, and ongoing changes to the accepted nomenclature (Kahlert *et al.* 2020). Though harmonizing occurrence datasets can reduce their taxonomic resolution, it also facilitates the detection of large-scale spatio-temporal patterns (Lee *et al.* 2019). We compared all our datasets against six databases that contain accepted names, synonyms with links to the respective accepted names, and suggestions for grouping contentious taxa in larger complexes. If we found a taxon name in one of the databases we either accepted it, changed it into the accepted name in case of a synonym, or included it in a complex. We did not query taxa we found in earlier databases against later ones, but in case the name changed from the original one, we queried the new one against earlier databases. Lastly, we controlled the results visually for consistency. We used the following databases in the same order:

1. Table S2 from Kahlert *et al.* (2020)
2. The taxon list associated with the OMNIDA software (Lecointe *et al.* 1993)
3. The German list of freshwater organisms (Mauch *et al.* 2017)
4. The diat.barcode database (Rimet *et al.* 2019)
5. The website algaebase.org (Guiry 2020)
6. The global biodiversity information platform (gbif) (GBIF.org 2020)

We harmonized the macroinvertebrate data with gbif through the taxize R package (Chamberlain & Szöcs 2013).

3 What is the optimal taxonomic level?

One result of the last progress review for GetReal (29.04.2020) was that we should determine an optimal taxonomic level for each taxon separately instead of using one common level for all data. *Oligochaetes*, for example, are usually only determined to subclass level. In a setting

with one common level they would thus be omitted if this level would be higher than subclass, which would most likely be the case. By using taxon-specific levels that take low-resolved taxa into account, we can thus use more of the data. However this poses the challenge of finding an optimal level for each taxon given a dataset.
The following refers exclusively to the macroinvertebrate data, as the taxonomic resolution was not an issue with diatom datasets. We describe the procedure for diatom data at the end of the section. The optimal level was established with a hierarchical approach. First, we removed all observations from Phyla and Classes that were not present in all datasets. We assumed that these represented differences in sampling rather than in communities. The classes Clitellata (Annelida), Insecta, Malacostraca (Arthropoda), Bivalvia and Gastropoda (Mollusca) remained. In the following, a higher taxonomic level refers to levels with higher resolution, i.e. species is the highest taxonomic level and kingdom the lowest. For each taxon, we calculated the percentage of observations represented at each higher level. For example, 4.12% of observations from the order *Lepidoptera* are at the species level, 74.77% at the genus level, 7.75% at the family level, and 13.35% at the order level. Now given a threshold X, we hold a taxon to be optimally represented at a certain taxonomic level if less than X% are represented by higher levels. For example, *Lepidoptera* would be represented on order level if X > 4.12% + 74.77% + 7.75% = 86.64%. As there are no theoretical grounds on which to base such a threshold value we searched for noticeable patterns in the data (Figure 1). The most salient change occurs between 85% and 86%. It occurs because for X > 86% *Chironomidae* are represented at the family level. We used 85% as threshold. Observations that were missed by this procedure, e.g. observations of *Chironomidae* at the family level, were included at their respective level.

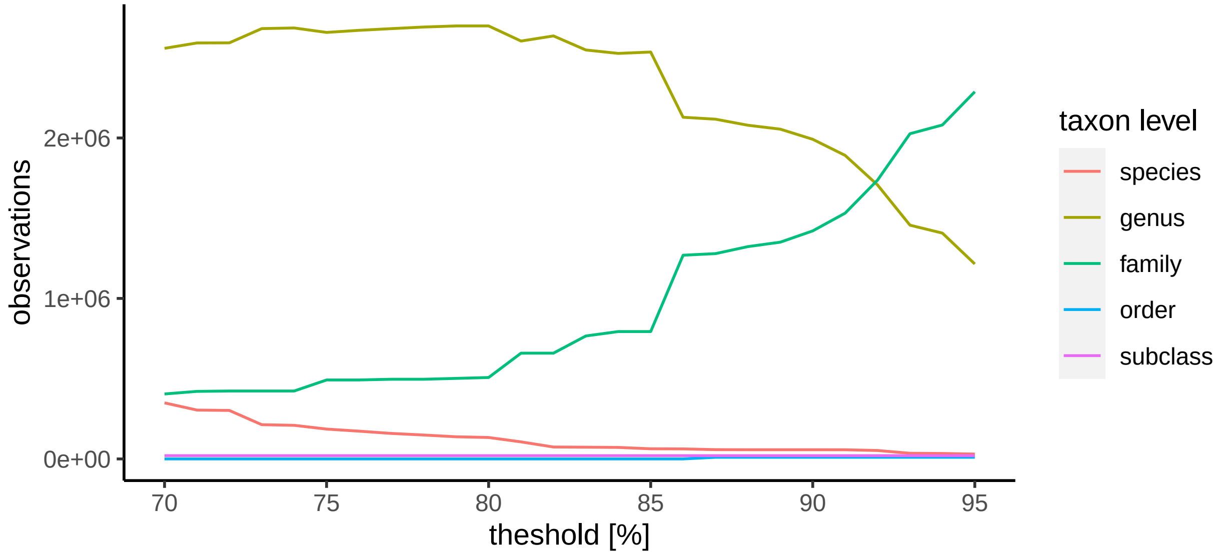

Figure 1: Number of macroinvertebrate observations at a given taxonomic level as function of the threshold value

For the diatoms, we employed 75% as a threshold, because for *Gomphonema*, which is the fourth most common genus in our dataset, 81.43% of the observations were at the species level. The taxonomic resolution was higher than in the macroinvertebrate data and the lowest resolution is the genus level. The equivalent of Figure 1 for diatoms can be found here.

4 Can we represent the stream types with our data?

We determined visually whether our dataset contained enough sampling sites in a given river type to derive meaningful TAs. The degree of representation for river types was graded in a three-tier system: high, medium, and low. A high degree of representation indicates, that we have many sampling locations, which are distributed across the instances of a river type that fall within the countries considered in GetReal. A low degree indicates the opposite, i.e. few and spatially clustered sites. A medium rating implies that we either have many sampling sites, but these only extend over parts of the countries or few sites that extend over most of the countries. The ratings for all river types for macroinvertebrates and diatoms are shown in table 1.

Table 1: The ratings for all river types for macroinvertebrates and diatoms

Rating	Taxon	River.Types
high	macroinvertebrates	4, 5, 9, 10, 11, 12, 13, 16
high	diatoms	
medium	macroinvertebrates	1, 2, 3, 8, 14, 15, 18
medium	diatoms	1, 2, 3, 4, 5, 6, 8, 9, 12, 14, 16, 17, 18, 19
low	macroinvertebrates	6, 7, 17, 19, 20
low	diatoms	7, 10, 11, 13, 15, 20

For each river type we provide maps with the associated sampling sites for macroinvertebrates and for diatoms.

Further analyses were conducted for all stream types with a high or medium degree of representation. More information on the river types is available in Lyche Solheim *et al.* (2019). We have fewer sampling sites for diatoms than for macroinvertebrates which entails that the representation of stream types is mostly lower.

5 What is a typical assemblage?

We derived TAs based on a rule that considered:

1. The probability of site x belonging to stream type z given that species y is present (a measure of specificity, henceforth **A**)

2. The probability of species y being present given that site x belongs to stream type z (a measure of commonness, henceforth **B**)

3. The Species Indicator Value

The Species Indicator Value (Dufrêne & Legendre 1997; Cáceres & Legendre 2009) is the weighted product of **A** and **B** (see Equation (1))

$$\sqrt{Indval} = \sqrt{A_g \times B} = \sqrt{\frac{\frac{n_p}{N_p}}{\Sigma_{k=1}^{K} \frac{n_k}{N_k}} \times \frac{n_p}{N_p}} \tag{1}$$

where N_p is the number of sites that belong to river type p and n_p the number of occurrences of the focal species in sites of type p. K is the number of river types. **A** is weighted by the total number of occurrences to account for unequal sample sizes. The statistical significance of the Species Indicator Value can be assessed with permutation-based pseudo-p-values, which we did with 999 permutations. Here, we are not interested in indicator species for each community, but TAs. Hence, continuing with those species that have a pseudo-p-value below some significance level would not serve our purpose. A species that occurs at each site, across all stream types, highlights the difference: while it would not be indicative of any stream type (low specificity) it should be part of each TA. Hence, we need additional criteria to derive the TAs which can be based on **A**, **B**, and the pseudo-p-value of the Species Indicator Value. We used the following rules:
For macroinvertebrates, species were considered typical if **B** > 0.25 or **B** > 0.20 and $p < 0.05$ or **A** > 0.80.
Genera were considered typical if **B** > 0.50 or **B** > 0.33 and $p < 0.05$ or **A** > 0.95. Families or lower taxonomic levels were considered typical if **B** > 0.95 or **B** > 0.80 and $p < 0.01$ or **A** > 0.99.
For diatoms, species where considered typical if **B** > 0.4 or **B** > 0.3 and $p < 0.05$ or **A** > 0.7. Genera where considered typical if **B** > 0.8 or **B** > 0.6 and $p < 0.05$ or **A** > 0.95.

Note that we did not systematically optimize these thresholds. Such procedures would require optimization criteria, but we are not aware of a criterion that would work in this context. As TA can be very similar in composition or harbor strongly differing numbers of taxa, neither criterion would optimize what we would consider a TA. We think the use of subjectively defined thresholds is justified, as long as they are clearly and openly communicated, to be what we define as "typical" assemblages.

However, we conducted a sensitivity analysis to see how varying the parameters would alter the results. We altered the threshold values of **A** and **B**. The rules above contain two distinct **B** thresholds: B_1 which does not consider the pseudo-p-value (**B** > 0.25 for macroinvertebrate species and **B** > 0.40 for diatom species) and B_2 which does take the pseudo-p-value into account ($p < 0.05$ and **B** > 0.2 macroinvertebrate species and **B** > 0.30 for diatom species

). In the following simulations, the B_2 was always taken to be 25% below B_1. Henceforth, when referring to the threshold for B, we refer to B_1. For species, we varied the threshold for B in ten steps between 0.10 and 0.75 and that for **A** in ten steps between 0.5 and 1.0. For lower taxonomic levels these thresholds were raised. For genera, the threshold values of **A** and **B** were raised by a factor of 1.25 and 2 respectively. For families and lower taxonomic levels, the thresholds were raised by factors of 1.5 and 3 respectively. The taxon richness and uniqueness scores (see Figure 5 and preceding text) of each TA were computed for all 100 combinations of these parameters and each taxonomic level. Please note that results are only shown and discussed for the non-redundant TAs (see section 6).

Taxa richness decreased with increasing **A** and **B** threshold in macroinvertebrates and diatoms (Figure 2A and Figure 3A), while the uniqueness scores increased with **B** thresholds but decreased with **A** thresholds ((Figure 2B and Figure 3B)). Uniqueness scores decreased noticeably with very high **A** thresholds (> 0.9), indicating that taxa that are specific to certain river types are an important driver of TA differentiation. Note that macroinvertebrate graphs are only shown for all taxa levels combined while the diatom plot only shows species. Plots for each taxon level separately are available for macroinvertebrates. However, the general patterns visible in Figure 2 and Figure 3, hold for them as well.

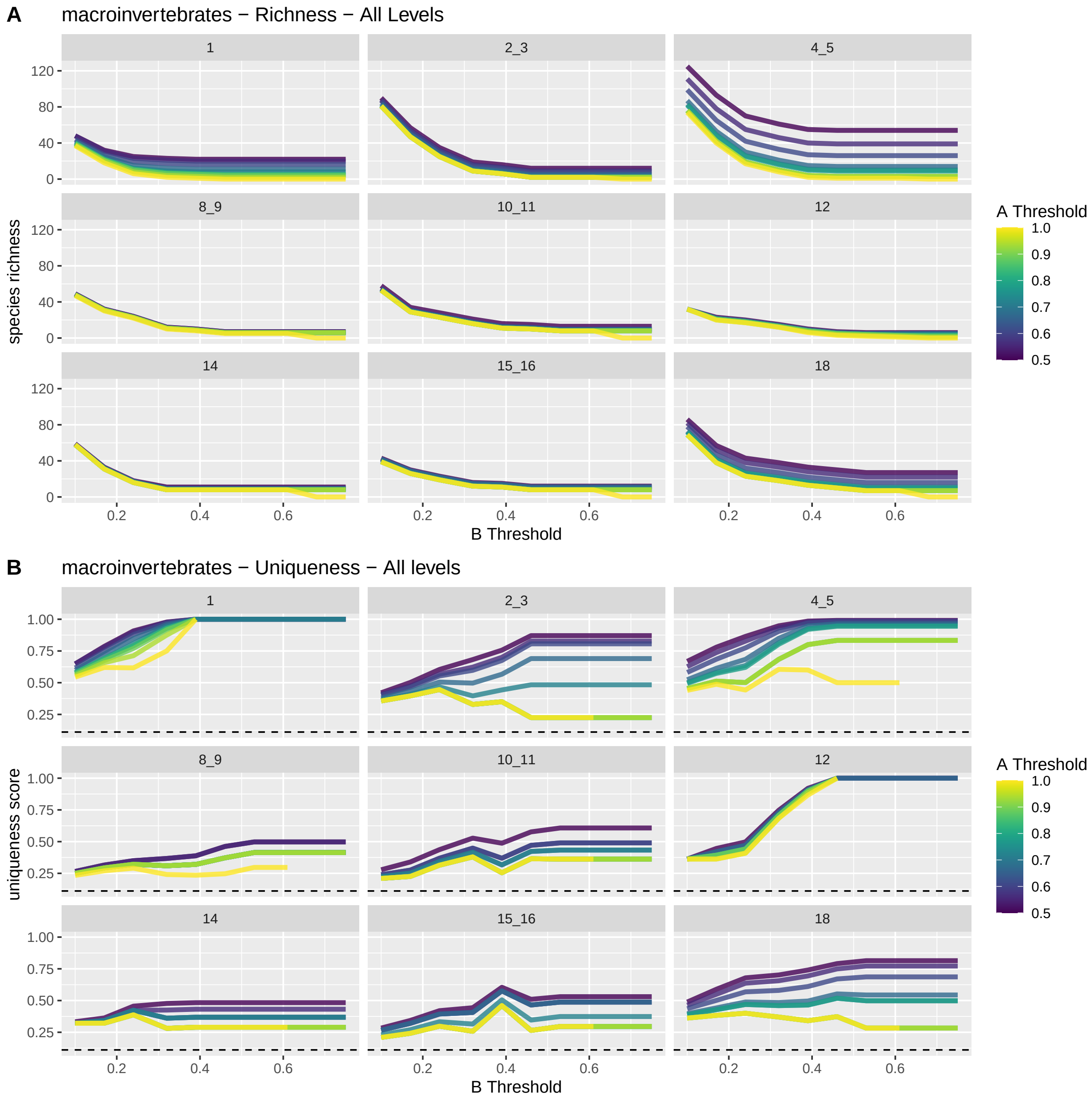

Figure 2: Changes in taxon richness along a changing B threshold. Line color indicates the A threshold

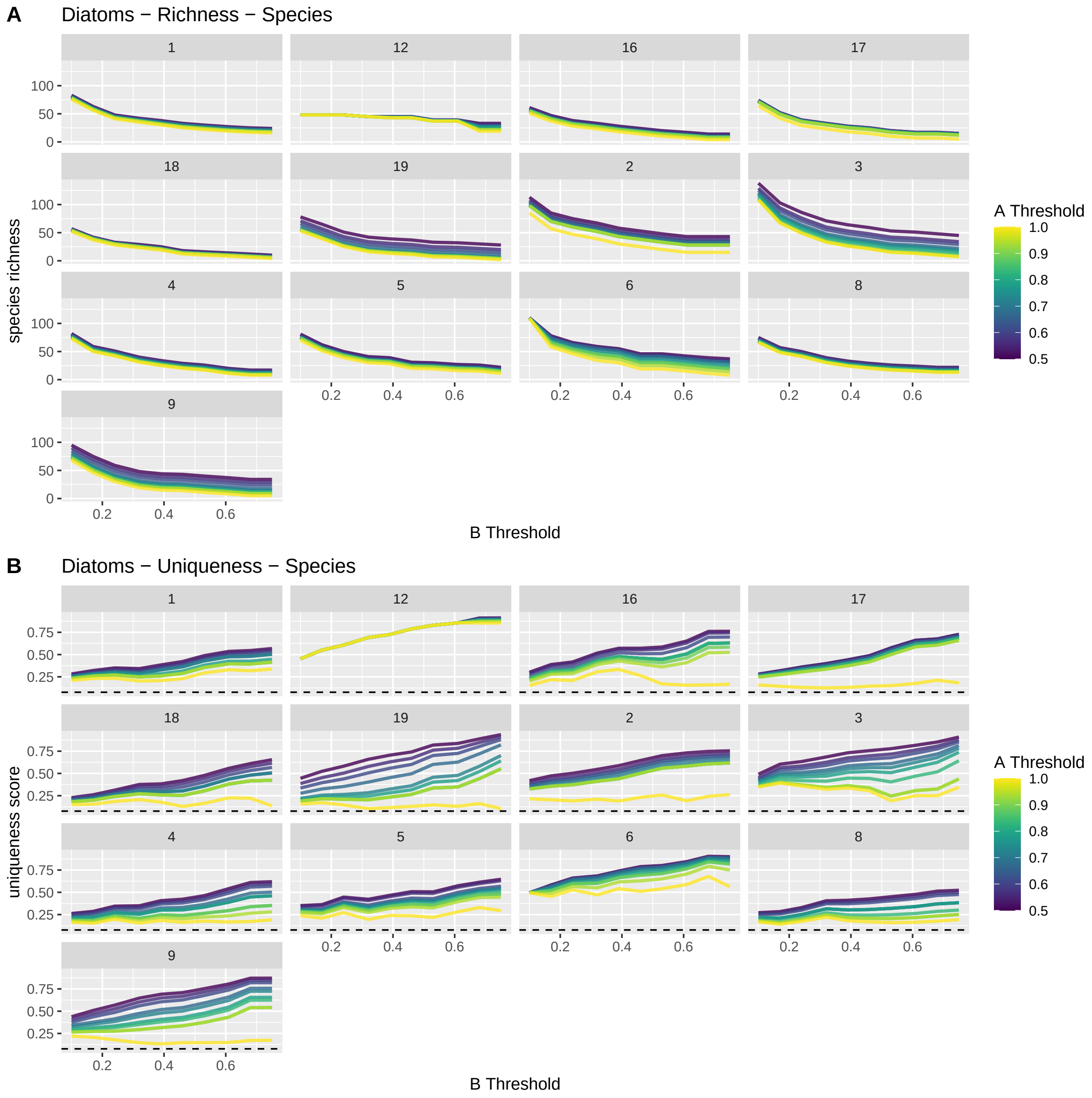

Figure 3: Changes in uniqueness score along a changing B threshold. Line color indicates the A threshold

6 Redundancy between typical assemblages

We assessed the degree to which the different TAs overlap (Table 2). The degree of overlap is the percentage of taxa in a TA that is also present in the most similar (largest overlap) TA. Again, choosing a threshold above which we consider two assemblages to be redundant is somewhat arbitrary. We proceeded with 75% but are open to other suggestions. This threshold leads to five redundant assemblages in macroinvertebrates and none in diatoms.

Most of the redundant TAs belong to two river types that only differ in river size: RT2 and 3, 4 and 5, 8 and 9, as well as 10 and 11. The only exception is the combination of RT15 and 16. Both are high altitude river types that occur mainly in southern Europe, which differentiates them from the northern high altitude rivers in RT14. RT13 is also redundant with RT2 and 3 however joining it with these two river types led to a drastically reduced number of taxa in the TA, when compared to that of the combined river type RT2_3. Since RT13 represents an exceedingly rare river type we decided to omit it from the analysis and proceed with RT2_3 instead of RT2_3_13. The new TAs resulted in overall lower degrees of overlap, none of which exceeds the 75% threshold. The largest overlaps were between RT8_9 and RT10_11, with 70%.

Table 2: Overlap between different typical assemblages. Pluses indicate that the overlap was equally high for serveral other typical assemblages all of which are listed with a "+" inbetween.

River.Type	Macroinvertebrate	Diatoms
RT1	RT2+4 (25%)	RT2 (62.9%)
RT2	RT3 (88.2%)	RT1 (46.8%)
RT3	RT2 (68.2%)	RT2 (48.7%)
RT4	RT3 (45.8%)	RT2 (74.1%)
RT5	RT4 (83.3%)	RT2 (62.5%)
RT6		RT12 (38.3%)
RT8	RT10 (77.8%)	RT2 (71.4%)
RT9	RT8 (76.5%)	RT5+8 (56%)
RT10	RT11+18 (65.2%)	
RT11	RT10 (88.2%)	
RT12	RT9 (50.0%)	RT6 (40.9%)
RT13	RT2 (87.5%)	
RT14	RT16+18 (69.2%)	
RT15	RT16 (85.7%)	
RT16	RT9+10+11+15 (57.1%)	RT18 (56.5%)
RT17		RT1 (63%)
RT18	RT10 (55.6%)	RT1 (71.4%)
RT19		RT17 (68.4%)

7 Characteristics of typical assemblages

In all macroinvertebrate TAs, genus is the prevalent taxonomic level (figure 4A). The numbers of species and families or lower levels are similar with both exceeding the other in four assemblages. The mean number of species was 3.2, mean number of genera 14.3, and the mean number of families or lower 2.4. RT01 and RT14 were the smallest assemblages with 13 taxa and RT18 was the most taxa rich assemblage with 28 taxa. For diatoms, species is the prevalent taxonomic level in all TAs (figure 4B). Some assemblages consist entirely of species (i.e. RT4, 6, 9, 16, 17, and 19). The mean number of species per TA is 30 and the mean number of genera 0.5. RT3 has the most taxa rich TA with 49 taxa and RT19 has the least taxa in its TA with 18.

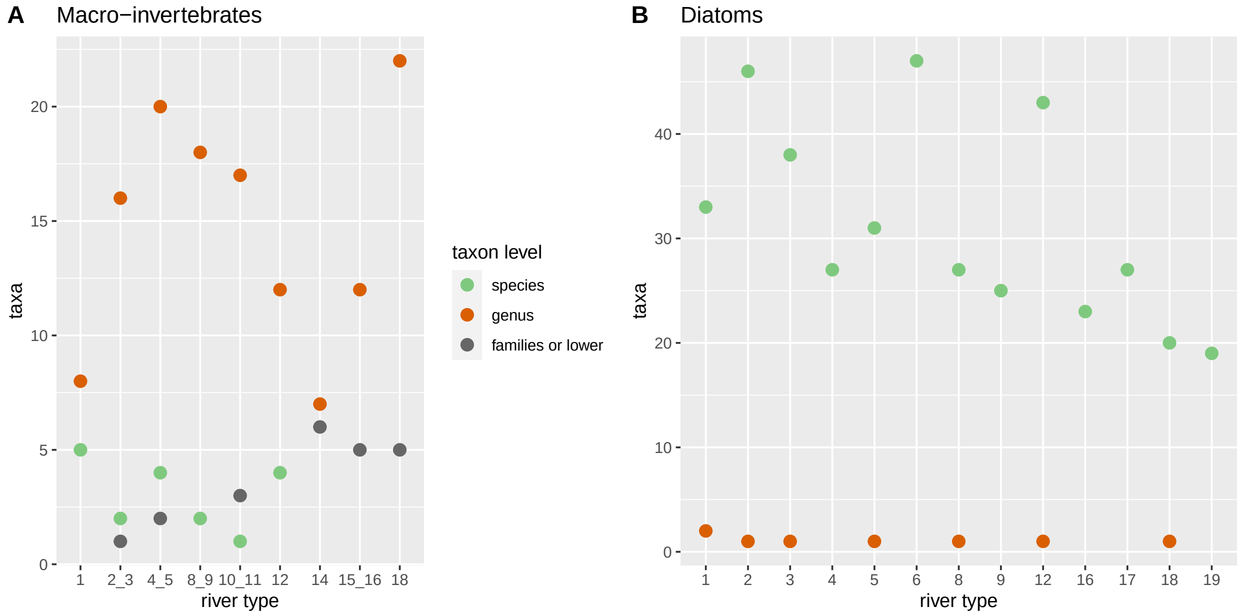

Figure 4: Numbers of taxa on each taxonomical level for all typical assemblages.

We can express the uniqueness of a TA with the following score: Each taxon receives a taxon uniqueness score that is one divided by the number of TAs it occurs in. For each river type, we sum the taxon scores of all taxa up and divide it by the number of taxa in the river type's TA. If all taxa in the TA are unique to that TA the score is one. If all species occur in one other TA the score is 0.5. The minimal score depends on the number of TAs, as it is 1 divided by that number and it signals that all species in that TA occur in all other TAs. These scores are shown in figure 5. The dashed horizontal lines indicate the minimum scores.

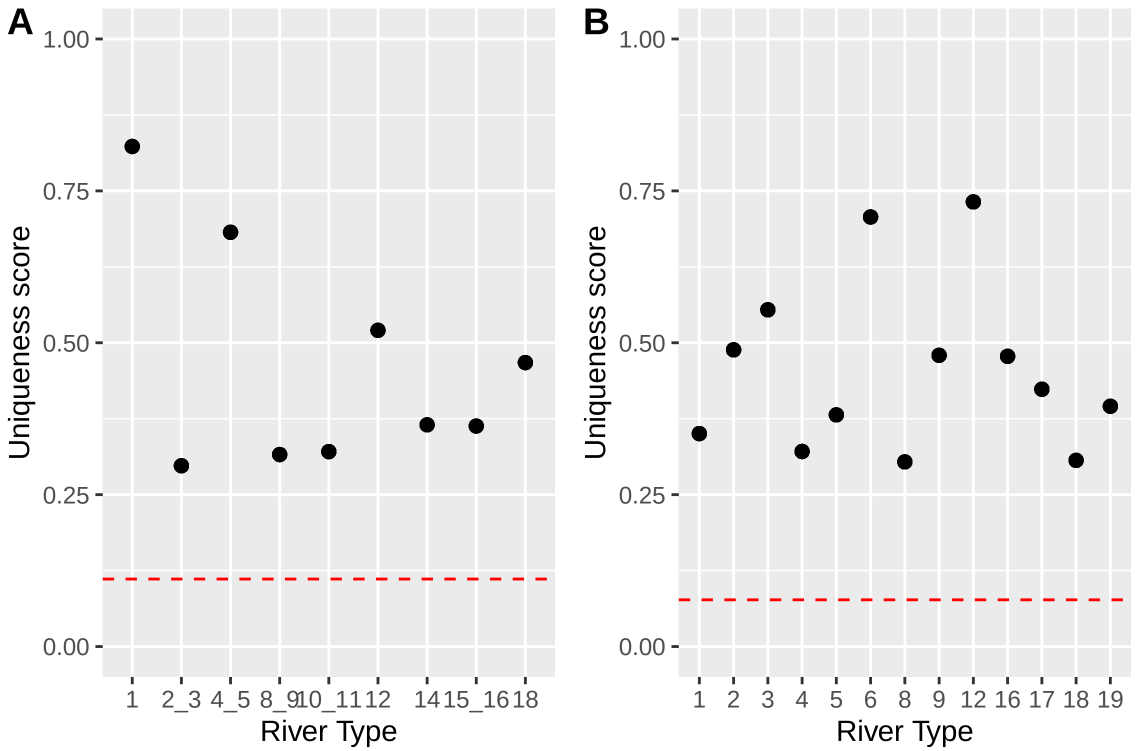

Figure 5: Uniqueness scores for typical assemblages of macroinvertebrates(A) and diatoms (B). The red dashed line indicates the lowest possible score.

We used Principal Coordinate Analysis (PCoA, Gower (1966)) to visualize the similarity of TAs, based on Jaccard distance matrices (Figure 6). The first axis in the macroinvertebrate PCoA represents an altitude gradient going from lowland (RT1, RT2_3, RT4_5), over mid-altitude (RT8_9, RT10_11, RT12, RT18), to high altitude (RT14, RT15_16). Mid-altitude streams are between 200-800 meters above sea level and the other two groups are lower or higher respectively.

RT8_9 and RT10_11 are the most similar which agrees with the considerable overlap (70%) we found before. Both river types represent mid-altitude rivers of all sizes, with RT8_9 containing siliceous rivers and RT10_11 calcareous rivers. The difference in geology does not to lead to strongly diverging TAs. None of the taxa, that the two TAs differ in, is considered indicative of geology. However several are typical for mid- to high-altitude streams (e.g. *Amphinemura* and *Protonemura*). The difference between siliceous (RT14) and calcareous(15_16) becomes more pronounced with increasing altitude. The TA of mid-altitude streams from the Mediterranean (RT18) is distinct from RT8_9 and RT10_11 but similar to that of RT12 which are mid-altitude streams with many (>20% Catchment area) histosol soils in their catchment. The lowland TAs are distinct from the mid- and high-altitude TAs and each other. This agrees with the high uniqueness scores of RT1 and RT2_3. The similarity between RT12 and RT15_16 is quite surprising given the large geographic distance between

sampling sites (see sampling maps).
In the PCoA of diatom TAs, the river types RT16 to 19 form one group. RT17 to 19 are Mediterranean streams and RT16 are glacial streams that also fall mainly within the region (the Pyrenees, the Western Alps, and the Central Alps. We do not have sampling sites in the later). RT17 (Mediterranean, lowland, medium-Large, perennial) is further removed from the core of the cluster. Potentially because they are the only lowland stream type in that group. The second cluster consists of river types RT1 to 5 as well as RT8 and 9. The river types RT1, 4, 8 are one subcluster as are RT2 and 3. RT5 connects the first to the second subcluster and RT9 is more similar to the first but noticeably removed from it. The structuring can be read as a size gradient. RT1,4, and 8 are very similar RT1 are very large rivers and RT4 and 8 are both medium to large, while RT5 and 9 are the smaller versions of RT4 and 8. Interestingly the imprint of the non-size-related characteristics of the rivers are more pronounced in smaller streams (i.e. RT5 and RT9 are more distinct than RT4 and RT8). The two TAs RT6 and RT12 are further removed from the formerly described clusters. Both represent rivers that are highly influenced by organic matter, which is known to impact diatom communities (Hering *et al.* 2006).

Online, we provide the complete taxa lists for all macroinvertebrate and diatom TAs.

8 Seasonal typical assemblages

In addition to the spatially defined TAs, we derived seasonal TAs (sTA) for a subset of river types. The four seasons were defined as follows: spring is March to May, Summer is June to August, Fall is September to November, and Winter is December to February. To avoid strong spatial signals in the sTAs, we only considered those river types (RT) in which samples were evenly distributed between seasons. In most cases, we had to omit parts of the data (e.g. certain seasons or datasets) to achieve an even spatio-temporal distribution. Online, we provide maps for all RT with all available seasons as well as the respective subsets that we used in the further analyses for macroinvertebrates and diatoms.
As an example, the map of macroinvertebrate samples for the combined RT 10_11 is shown in Figure 7.

To visualize differences between the seasons we used Non-metric multidimensional scaling (NMDS) on Jaccard dissimilarity matrices. The resulting NMDS plots are available for macroinvertebrates and diatoms. Figure 8 shows the NMDS plot for invertebrate samples in RT10_11. In most river types the differences between seasons are not large and the NMDS stress values are typically above 0.2.

Further, we evaluated whether the Jaccard dissimilarity between sites would be better explained by spatial distance or by season. To this end, we employed generalized dissimilarity modeling (GDM, Ferrier *et al.* (2007)). In GDMs, the response variable is the ecological dissimilarity between two sites (expressed in some *a priori* chosen dissimilarity metric, here

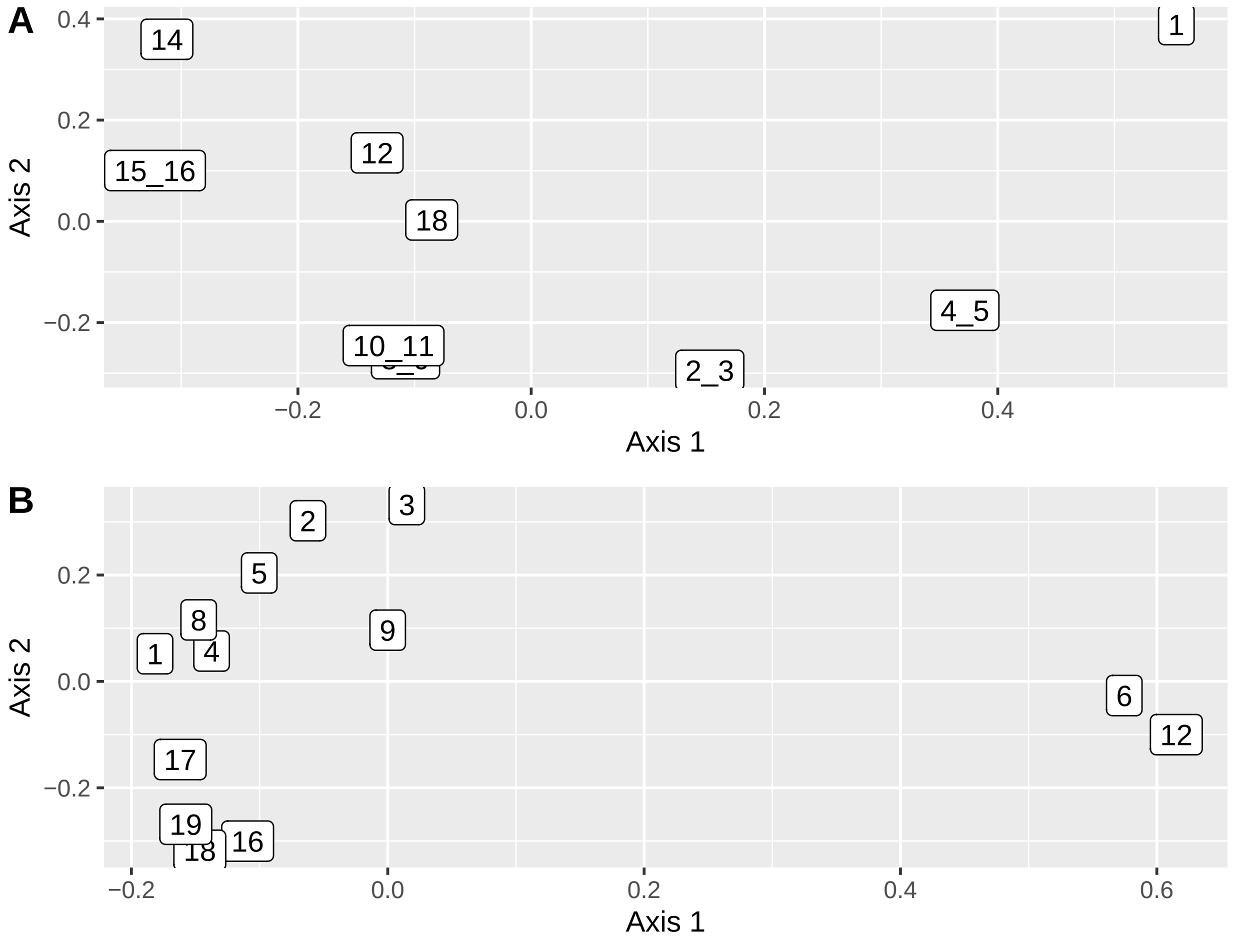

Figure 6: Principal Coordinate Analysis ordinations of typical assemblages based on Jaccard distance matrices. A shows the typical asseblages of macroinvertebrates and B those of diatoms.

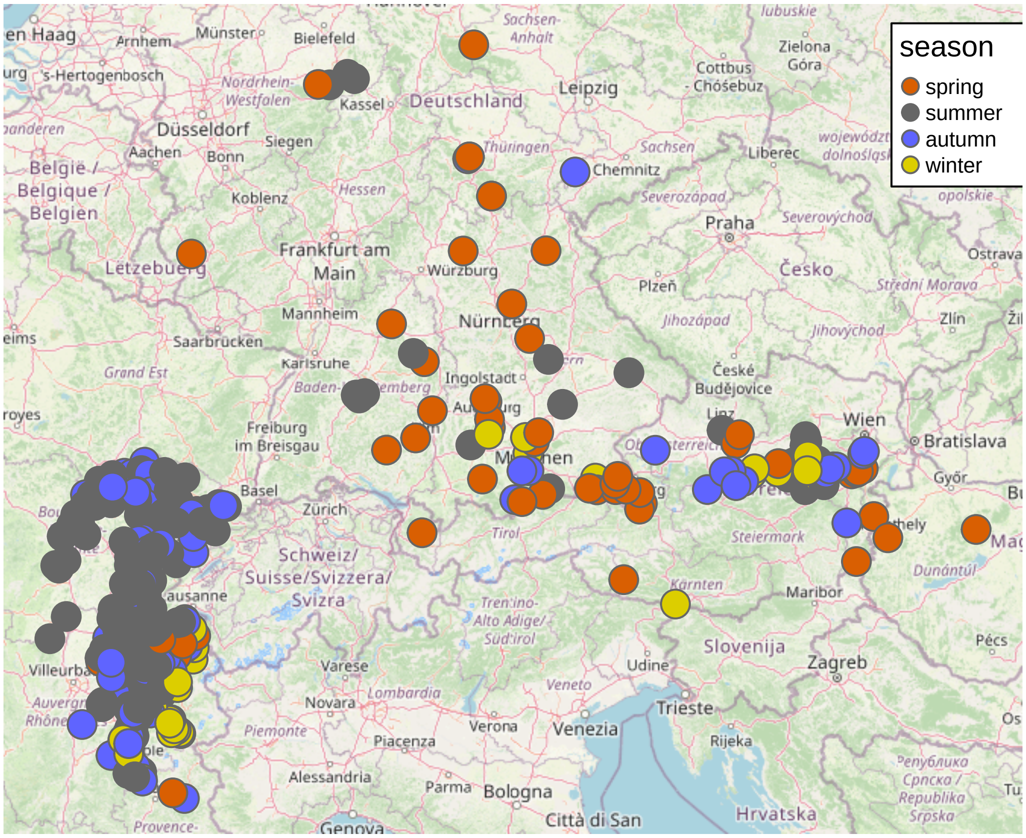

Figure 7: Map of sampling sites for the combined River Type 10_11. The color of the points shows the season of sampling.

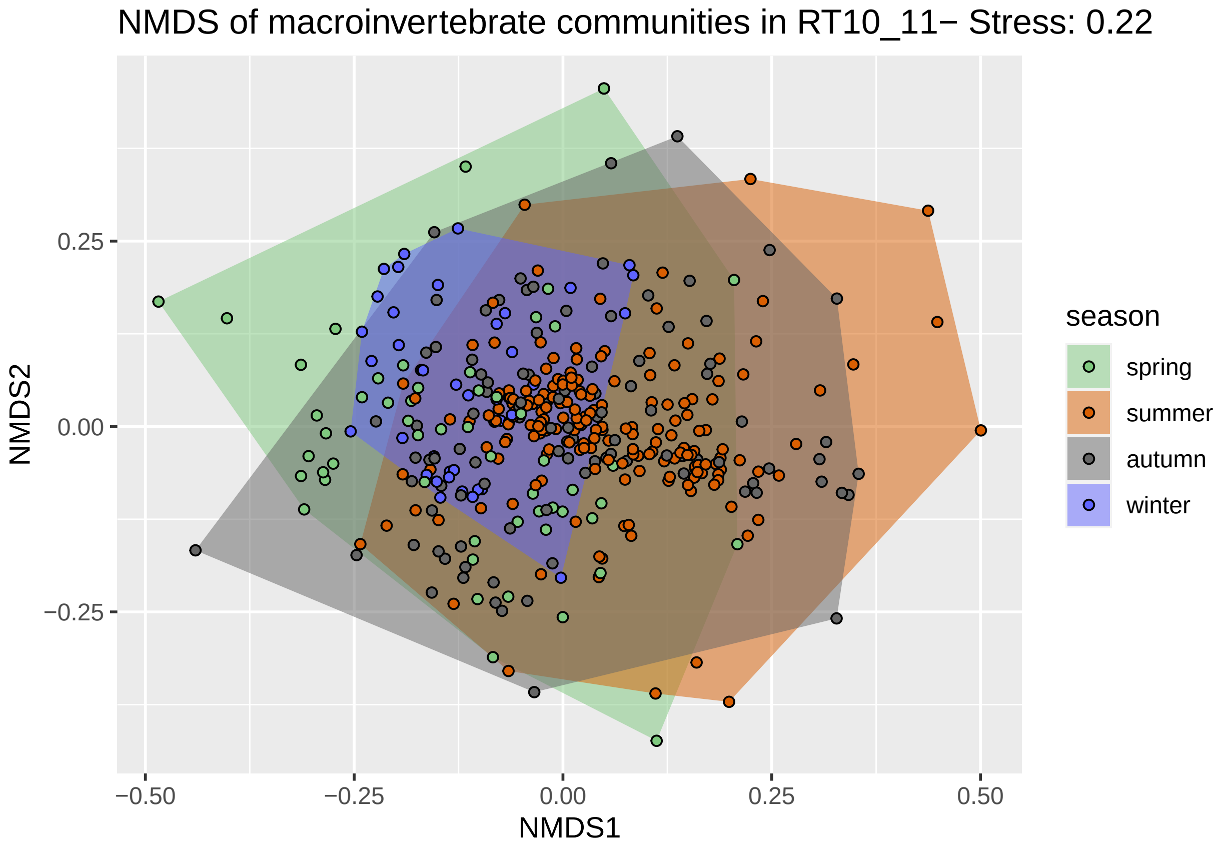

Figure 8: Nonmetric multidimensional scaling plot of Jaccard dissimilarity matrices for macroinvertebrate communities in RT10_11. The color of the points shows the season. Convex hulls surround all sampling points from one season.

Jaccard). Smooth functions are fitted to the environmental data and the differences between the values of these functions at the two sites of interest are used as explanatory variables. By using a generalized modeling framework we can account for the bounded nature of dissimilarity metrics (between 0-1) and the smooth functions allow for variation in the rate of compositional turnover along gradients. Plots that show the effect of spatial distance and that of season for all GDMs are available for macroinvertebrates and diatoms. The plot for invertebrates in RT10_11 is shown in Figure 9.

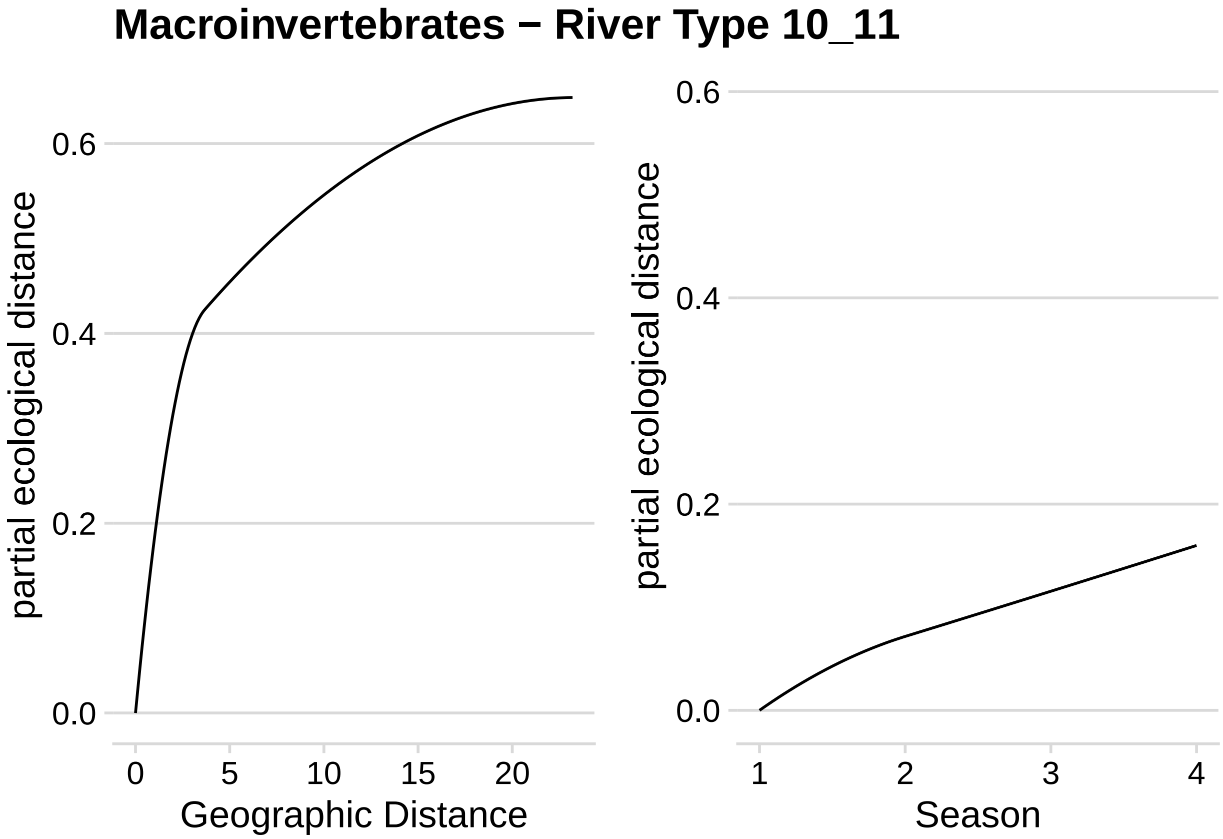

Figure 9: Partial ecological distance between sites with increasing geographic distance (A) or chaning season (B, 1 = spring, 2 = summer, 3 = autumn, and 4 = winter) predicted with Generalized Dissimilarity Models

We selected RT 10_11 and RT 15_16 for invertebrates and RT11 and RT15 for diatoms because they showed the strongest effect of season in GDM and NMDS. For these four river types sTA were derived in the same way as the non-seasonal TAs.

9 Patterns and overlap in seasonal assemblages

In RT11, the number of diatom taxa in the sTAs did not vary strongly between the seasons (Table 3). The summer and autumn sTAs were more similar to each other than either to the winter sTA. The latter was most similar to the summer sTA, as they share some Gomphonema species (*Gomphonema olivaceum olivaceoides* and *Gomphonema parvulum* Complex) which are absent from the autumn sTA with exception of *Gomphonema pumilum* Complex.

Table 3: Overlap between seasonal typical assemblages (sTA) of diatoms in river type 11 expressed in percent of taxa in row sTA also present in column sTA. N is the number of taxa in the respective sTA.

	summer	autumn	winter	N
summer	100.0	46.4	35.7	28
autumn	61.9	100.0	38.1	21
winter	45.5	36.4	100.0	22

For diatom in RT15, the winter sTA is considerably larger than the summer and autumn sTAs (Table 4).
Both, the summer and the autumn sTAs, overlap 81.2% with the winter sTA.
Therefore, they cross the threshold of 75% overlap we used to delineate redundant TAs. The overlap between the winter sTA and either summer or autumn sTA is of a similar size (41.4% and 44.8%). In general, the overlaps in RT15 are larger than those in RT11 which might indicate a weaker seasonal turnover in these ecosystems.

Table 4: Overlap between seasonal typical assemblages (sTA) of diatoms in river type 15 expressed in percent of taxa in row sTA also present in column sTA. N is the number of taxa in the respective sTA.

	summer	autumn	winter	N
summer	100.0	57.9	63.2	19
autumn	68.8	100.0	81.2	16
winter	41.4	44.8	100.0	29

For the macroinvertebrates, the number of taxa in the sTAs is lower than for diatoms. In both river types, the number of taxa in the autumn sTA is also markedly higher than for all other macroinvertebrate sTAs. In the combined type RT10_11, the spring sTA was nested in the winter sTA and had no overlap with the summer sTA (Table 5). The summer sTA was most similar to the autumn TA (71.4% overlap) and vice versa (29.4% overlap). Half of the taxa in the winter TA are also part of the autumn TA which is the highest overlap for the winter TA. In the other combined river type considered here, RT 15_16, the summer sTA is nested within the spring sTA and the winter sTA is almost nested within the autumn sTA

(Table 6). *Limnoidae* is the only taxon that occurs in the winter sTA but not the autumn sTA. Across this divide, the sTAs only share the two taxa which are common to all four: *Baetis* and *Chironomidae*.

Table 5: Overlap between seasonal typical assemblages (sTA) of macroinvertebrates in the combined river type 10+11 expressed in percent of taxa in row sTA also present in column sTA. N is the number of taxa in the respective sTA.

	spring	summer	autumn	winter	N
spring	100.0	0.0	50.0	100.0	2
summer	0.0	100.0	71.4	28.6	7
autumn	5.9	29.4	100.0	17.6	17
winter	33.3	33.3	50.0	100.0	6

Table 6: Overlap between seasonal typical assemblages (sTA) of diatoms in river type 15 expressed in percent of taxa in row sTA also present in column sTA. N is the number of taxa in the respective sTA.

	spring	summer	autumn	winter	N
spring	100.0	75.0	50.0	50.0	4
summer	100.0	100.0	66.7	66.7	3
autumn	12.5	12.5	100.0	18.8	16
winter	50.0	50.0	75.0	100.0	4

Online, we provide the complete taxa lists for macroinvertebrate and diatom sTAs.

References

Cáceres, M.D. & Legendre, P. (2009). Associations between species and groups of sites: inindices and statistical inference. *Ecology*, 90, 3566–3574.

Chamberlain, S. & Szöcs, E. (2013). Taxize - taxonomic search and retrieval in r. *F1000Research*.

Dufrêne, M. & Legendre, P. (1997). Species Assemblages and Indicator Species: The need for a flexible asymmetrical Approach. *Ecological Monographs*, 67, 345–366.

Ferrier, S., Manion, G., Elith, J. & Richardson, K. (2007). Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions*, 13, 252–264.

GBIF.org. (2020). *GBIF home page.*

Gower, J.C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53, 325–338.

Guiry, G.M., M. D. & Guiry. (2020). *AlgaeBase. World-wide electronic publication, national university of ireland, galway.*

Hering, D., Johnson, R.K., Kramm, S., Schmutz, S., Szoszkiewicz, K. & Verdonschot, P.F.M. (2006). Assessment of European streams with diatoms, macrophytes, macroinvertebrates and fish: A comparative metric-based analysis of organism response to stress. *Freshwater Biology*, 51, 1757–1785.

Kahlert, M., Rühland, K.M., Lavoie, I., Keck, F., Saulnier-Talbot, E. & Bogan, D. *et al.* (2020). Biodiversity patterns of Arctic diatom assemblages in lakes and streams: Current reference conditions and historical context for biomonitoring. *Freshwater Biology*, 1–25.

Lecointe, C., Coste, M. & Prygiel, J. (1993). “Omnidia”: Software for taxonomy, calculation of diatom indices and inventories management. *Hydrobiologia*, 269, 509–513.

Lee, S.S., Bishop, I.W., Spaulding, S.A., Mitchell, R.M. & Yuan, L.L. (2019). Taxonomic harmonization may reveal a stronger association between diatom assemblages and total phosphorus in large datasets. *Ecological Indicators*, 102, 166–174.

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.

Mauch, E., Schmedtje, U., Maetze, A. & Fischer, F. (2017). Taxaliste der Gewässerorganismen Deutschlands. *Informationsberichte des Bayerischen Landesamtes für Wasserwirtschaft*, 1.

Rimet, F., Gusev, E., Kahlert, M., Kelly, M.G., Kulikovskiy, M. & Maltsev, Y. *et al.* (2019). Diat.barcode, an open-access curated barcode library for diatoms. *Scientific Reports*, 9, 1–12.