ECOSYSTEMS

The use of genus-level determinations and biovolume classes as surrogates to indicate environmental drivers of stream diatom communities in the Brazilian Pampa

ANA PAULA TAVARES COSTA & FABIANA SCHNECK

Abstract: The demand for low-cost and time-efficient biomonitoring and ecological assessment strategies has increased due to threats to biodiversity. We evaluated the ability of surrogates (genera and biovolume classes) to access information provided by diatom species. Moreover, we investigated whether the relative importance of spatial and environment on diatom communities identified at the genus or classified at biovolume classes reflects the ecological responses at the species. We used biotic, environmental, and spatial data from Pampa streams. We employed Procrustes analysis to detect whether the spatial variation of communities at the species level was represented by the surrogates and variation partitioning to evaluate the relative importance of spatial and environment. We found high congruence between the surrogates and diatom species communities. Furthermore, whilst both spatial and environment were important to explain species composition, only the environment was important for both surrogates. We suggest that the use of genus-level determinations can provide information about the environmental relationships of species, whereas biovolume classes may be useful to detect changes in diatom communities. These results are a first step to seek for strategies aiming to simplify biological monitoring and assessments at pampean streams, especially considering the demand for the identification of organisms at the species.

Key words: congruence, grassland, lotic environment, metacommunity, Procrustes, variance partitioning

INTRODUCTION

Understanding freshwater biodiversity patterns is a challenge, especially in species-rich regions, such as the tropics and subtropics (e.g., Siqueira et al. 2012a, Trindade et al. 2018). This is because of the lack of reliable information about species identities (Hortal et al. 2015), in addition to the time spent on sample processing (Carneiro et al. 2010, de Oliveira et al. 2020). Such difficulties increase when anthropogenic impacts are part of the scenario as they may affect the composition of communities in different and unexpected ways, frequently resulting in biotic homogenisation (Siqueira et al. 2015, Petsch 2016) or differentiation (Vieira et al. 2022). As a consequence, biodiversity assessments and biomonitoring programmes of freshwater ecosystems tend to be expensive and time-consuming, which reduces the sampling effort in both space and time (Bevilacqua et al. 2015).

One way to overcome the absence of detailed information at the species level and to make biomonitoring programmes more cost- and time-effective is the use of surrogates (Margules & Pressey 2000). In general, a good surrogacy candidate needs to reflect the

patterns observed at the species level (de Oliveira et al. 2020). However, a high congruence between information at the species level and the surrogate is not sufficient as it is necessary that the surrogates also respond to spatial and environmental predictors similarly to species, especially for biomonitoring and ecological assessment purposes. In freshwater ecosystems, coarser taxonomic resolutions are good surrogates of species (e.g., Godoy et al. 2019, Riato et al. 2022), a relationship usually referred to as the higher taxon approach (see de Oliveira et al. 2020 for a meta-analysis). Furthermore, functional classifications have been widely used as surrogates (Gallego et al. 2012, Mueller et al. 2013). However, it is still necessary to advance further and test not only the higher taxon approach in less studied ecosystems around the world but also the surrogacy efficiency of simple and easily obtainable morphometric characteristics of species. One alternative is the classification of organisms into size or biovolume classes, which could drastically reduce the effort to represent community structure without losing information provided at the species level (e.g., Wunsam et al. 2002, Ribeiro et al. 2020).

Diatoms are among the most common bioindicators in freshwaters due to their wide distribution and ease to sample (Stevenson et al. 2010). Moreover, diatom species are characterised by large differences in their tolerances to chemical and physical characteristics of freshwater ecosystems, such as nutrient concentrations and flow velocity (Virtanen & Soininen 2012). However, it is a highly diverse group for which species identification is based on the analysis of small structures only distinguished by trained taxonomists, in addition to the need for a constant literature update since the nomenclature changes frequently (Rimet & Bouchez 2012a). In this sense, studies have evaluated the efficiency

of using coarse taxonomic resolutions and easily obtainable traits as surrogates for biomonitoring and assessment programmes (e.g., Rimet & Bouchez 2012a, Algarte et al. 2014, Riato et al. 2022). Moreover, non-taxonomical metrics, such as the number and volume of lipid bodies, the frequency of valve deformities and size classes, have gained prominence in the literature in recent years (Costa & Schneck 2022). This is because biomonitoring and ecological assessments based on such metrics are more time- and cost-effective when compared to the obtention of information at the species level for all communities (e.g., Pandey et al. 2018a, b).

Here, we investigated the efficiency of a taxonomic and a morphometric approach as surrogates of information at the species level for stream diatom communities in the Brazilian Pampa region. Specifically, we tested whether the identification of genera and the use of biovolume classes are effective in representing the spatial variation in the structure of diatom communities. Moreover, we investigated the relative importance of environmental and spatial variables in determining this variation and whether the surrogates under evaluation respond similarly to diatom species.

MATERIALS AND METHODS
Study area
We sampled 25 low-order streams in the Ibicuí, Santa Maria and Vacacaí-Mirim hydrographic basins in southernmost Brazil (Figure 1) during austral summer (February 2020). The basins differ mostly in their land use and cover, with the Santa Maria and Vacacaí-Mirim basins dominated by agriculture, whereas the Ibicuí basin is mainly covered by natural grasslands. The climate is humid subtropical, with hot summers and without dry seasons ('Cfa' of Köppen classification) (Alvares et al. 2013). The sampling
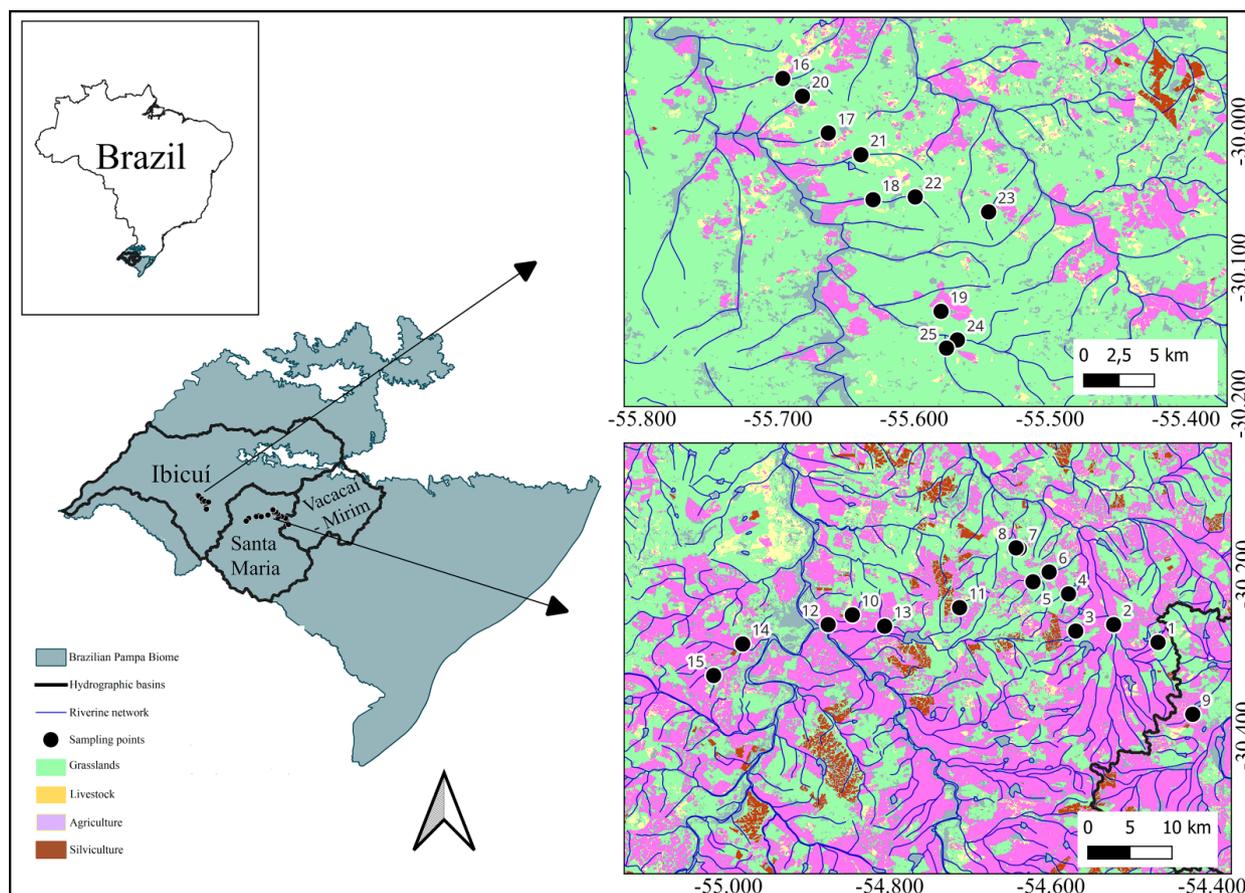
Figure 1. 25 streams sampled in the Brazilian Pampa Biome distributed in Vacacaí-Mirim, Santa Maria and Ibicuí hydrographic basins, Rio Grande do Sul state.

area is within the Pampa biome, also known as the Rio de la Plata Grasslands (Baeza et al. 2022). The region is characterised by the predominance of native grasslands (*campos*) (Overbeck et al. 2007), being the largest temperate grassland area in South America, covering approximately 75 million hectares in Uruguay, central-eastern Argentina and southernmost Brazil (Soriano 1991). The conversion of the natural landscape into agricultural areas, especially for the cultivation of soy, rice and silviculture, has been intensified in recent years (Overbeck et al. 2007, Oliveira et al. 2017) threatening terrestrial and freshwater ecosystems. Moreover, conservation policies do not adequately contemplate the Pampa, which is considered the most neglected biome in Brazil (Overbeck et al. 2007) and

among the most altered ecoregions of the world (Watson et al. 2016).

Sampled streams were dominated by grasslands within the riparian area, with sparse shrub formations and less commonly with riparian vegetation dominated by trees. At the Santa Maria and Vacacaí-Mirim hydrographic basins the streambeds are characterised by the predominance of sand and silt, whereas at the Ibicuí basin, it is common to observe pebbles, cobbles, and boulders, in addition to sand and silt.

Biological and environmental variables

At each stream site, we sampled biological and environmental variables from one stretch of 100 m. We sampled three subsamples of benthic diatoms from the surface of streambed sediment

by using a syringe. The material from the three subsamples was pooled and preserved in 4% formalin. In the laboratory, the material was acid-cleaned according to Metzeltin & García-Rodríguez (2012). We prepared permanent slides of diatoms using Naphrax® (Brunel Microscopes Ltd., Chippenham, UK) and counted 500 valves from each stream, using a light microscope with 1,000× magnification. We identified diatoms at the lowest possible taxonomic level, mostly species, using specialised bibliography such as Metzeltin et al. (2005) and Metzeltin & García-Rodriguez (2012). After identification, we used published biovolume-based classes to classify the taxa into five biovolume classes: 1 ($\leq$ 99 μm^3), 2 (100–299 μm^3), 3 (300–599 μm^3), 4 (600–1,499 μm^3), 5 (> 1,500 μm^3) (Rimet & Bouchez 2012b).

We measured pH, dissolved oxygen (mg L^{-1}), electrical conductivity (μS cm^{-1}) and total dissolved solids (g L^{-1}) using a Horiba U-50 multiparameter probe. At each stream, we made 10 measurements for each variable. We measured current velocity (cm s^{-1}), stream width (m) and depth (cm) at five random locations. We also sampled 500 mL of surface water at each stream for the analysis of total nitrogen (Allen et al. 1974) and total phosphorus (Valderrama 1981; Baumgarten et al. 1996). For further analyses, we used the mean of each variable per stream (Table I).

Data analyses

We performed principal components analysis (PCA; Legendre & Legendre 1998) to summarise the environmental variation among streams. For this, all variables were standardized to zero mean and unit variance. The interpretation of the results was based on the axes retained using the Broken-Stick criterion (Jackson 1993).

We used principal coordinates analysis (PCoA) to summarise the variation in diatom community composition for species, genus and biovolume classes, using presence-absence and abundance data. For this, we used the Sørensen (presence-absence) and the Bray-Curtis (logarithmised abundance + 1) dissimilarity coefficients (Legendre & Legendre 1998). To evaluate the congruence between the species dataset and the surrogates, we compared the ordination sample scores from PcoAs obtained with the species dataset with those obtained with each of the surrogates, using Procrustes analysis (Jackson 1995, Peres-Neto & Jackson 2001). The Procrustean superimposition is represented by the sum of squared differences (m^2) between matrices. In this sense, the lower the value of m^2, the greater the similarity between ordination matrices. We transformed the values of m^2 to r (Procrustes correlation).

We used partial redundancy analysis (pRDA; Legendre & Legendre 1998) to evaluate the relative importance of spatial and environmental factors on the community composition of diatoms for species, genus and biovolume class matrices. Before pRDA, we applied the Hellinger transformation on the biotic matrices (Legendre & Gallagher 2001). We generated the spatial predictor matrix using Moran eigenvector maps (MEM), with eigenvectors extracted from a Principal Coordinate of Neighbouring Matrices analysis (PCNM; Borcard & Legendre 2002). We used the minimum spanning tree criteria to truncate the distance matrix. In this analysis, the MEMs with the lowest values represent fine spatial scales, and those with higher values represent large spatial scales. The environmental matrix was determined after the exclusion of collinear variables with a variance inflation factor (VIF) higher than 3. The final environmental matrix contained eight variables: pH, dissolved oxygen, electrical conductivity, total phosphorus, total nitrogen, width, depth and current velocity. We then used redundancy analyses to test the significance of each global

Table I. Mean of environmental variables from 25 streams sampled in the Brazilian Pampa. Cond = electrical conductivity; DO = dissolved oxygen; TDS = total dissolved solids; TP = total phosphorus; TN = total nitrogen; Vel = current velocity.

Stream	pH	DO (mg L⁻¹)	Cond (µS cm⁻¹)	TDS (g L⁻¹)	TP (µg L⁻¹)	TN (mg L⁻¹)	Vel (cm s⁻¹)	Width (m)	Depth (cm)
1	8.2	8.4	0.24	0.23	85.0	1.2	<1.0	2.4	48.1
2	7.1	7.6	0.12	0.07	591.7	1.3	26.5	5.1	30.6
3	7.3	7.3	0.07	0.04	244.8	1.2	<1.0	2.2	19.3
4	7.8	11.1	0.22	0.14	72.6	0.7	<1.0	5.2	49.1
5	7.4	5.6	0.10	0.07	52.1	0.6	<1.0	5.9	49.8
6	6.9	4.3	0.06	0.04	105.6	1.3	<1.0	2.2	34.9
7	6.9	7.3	0.04	0.02	94.1	0.8	<1.0	2.5	23.9
8	7.2	7.9	0.06	0.03	38.8	0.6	<1.0	1.9	25.5
9	7.3	4.2	0.45	0.29	215.2	1.6	<1.0	3.9	30.3
10	8.0	8.7	0.30	0.19	102.3	1.5	<1.0	3.1	16.9
11	7.5	7.8	0.03	0.02	134.4	0.9	<1.0	1.2	20.2
12	7.7	8.6	0.04	0.02	108.1	0.8	<1.0	7.7	24.0
13	7.4	6.9	0.04	0.02	52.9	0.6	<1.0	5.4	56.0
14	6.6	5.0	0.03	0.02	57.0	0.6	<1.0	2.3	47.2
15	7.1	7.2	0.05	0.03	178.1	0.9	<1.0	2.2	45.0
16	7.3	6.6	0.15	0.09	43.0	0.4	21.0	1.1	17.5
17	7.9	7.9	0.07	0.04	47.1	0.6	23.9	3.0	14.6
18	7.1	8.0	0.04	0.02	30.6	0.5	66.4	2.2	21.9
19	7.5	6.2	0.08	0.06	49.6	0.7	<1.0	1.0	9.5
20	7.6	7.6	0.04	0.03	11.4	0.5	<1.0	5.5	5.4
21	7.4	7.0	0.06	0.04	25.7	0.6	9.7	4.1	16.8
22	8.4	7.8	0.12	0.08	29.0	0.8	<1.0	1.3	9.8
23	7.5	6.4	0.11	0.07	27.3	0.8	<1.0	1.9	39.8
24	7.5	8.0	0.07	0.05	33.1	0.7	30.5	4.1	7.3
25	7.3	6.7	0.07	0.04	96.5	0.6	38.5	4.7	11.1

model (i.e., a model using the complete set of spatial predictors and another model using the complete set of environmental predictors, separately for each biological matrix).

We proceeded with the analysis only if the global model for a given set of predictors was significant ($P < 0.05$). In this case, we used a forward selection with two stopping criteria:

the significance of each explanatory variable smaller than 0.05 and the adjusted coefficient of multiple determination (adjusted R^2) of the reduced model smaller than the adjusted R^2 of the global model, as recommended by Blanchet et al. (2008). The retained explanatory variables were then used in pRDA to partition biological variation into four fractions: [E]

pure environmental component, [E-S] spatially structured environmental component, [S] pure spatial component and [UN] residual unexplained variation. Finally, we tested the significance of the pure environmental and pure spatial components using 999 permutations, and the explained variation was estimated using adjusted R^2 (Peres-Neto et al. 2006). Finally, to visualise the relationships between each biological dataset and the environment we constructed RDA plots for each approach using the selected environmental variables. We used the R environment for all analyses (R Core Team 2020), with the packages *car* (Fox & Weisberg 2011), *vegan* (Oksanen et al. 2017), *adespatial* (Dray et al. 2016) and *ggplot2* (Wickham 2016).

RESULTS

The sampled streams varied in their environmental characteristics (Table I; Figure 2).

According to the Broken-Stick criterion, the first two axes of the PCA were retained and explained 52% of the variation in environmental data. The first principal component explained 32% of data variation, being positively related to current velocity and negatively to electrical conductivity, total dissolved solids, total nitrogen and stream width (Figure 2). The second principal component explained 20% of the variation and was positively related to dissolved oxygen and pH and negatively related to depth and total phosphorus (Figure 2).

We found 245 diatom species from 50 genera (Supplementary Material - Table SI). Diatom species richness ranged from 37 to 81 per stream. The more abundant genera were *Navicula* (15%), *Nitzschia* (14%), *Encyonema* (8%) and *Achnanthidium* (6%). All other genera comprised less than 5% of abundance. Regarding diatom biovolume classes, the most abundant
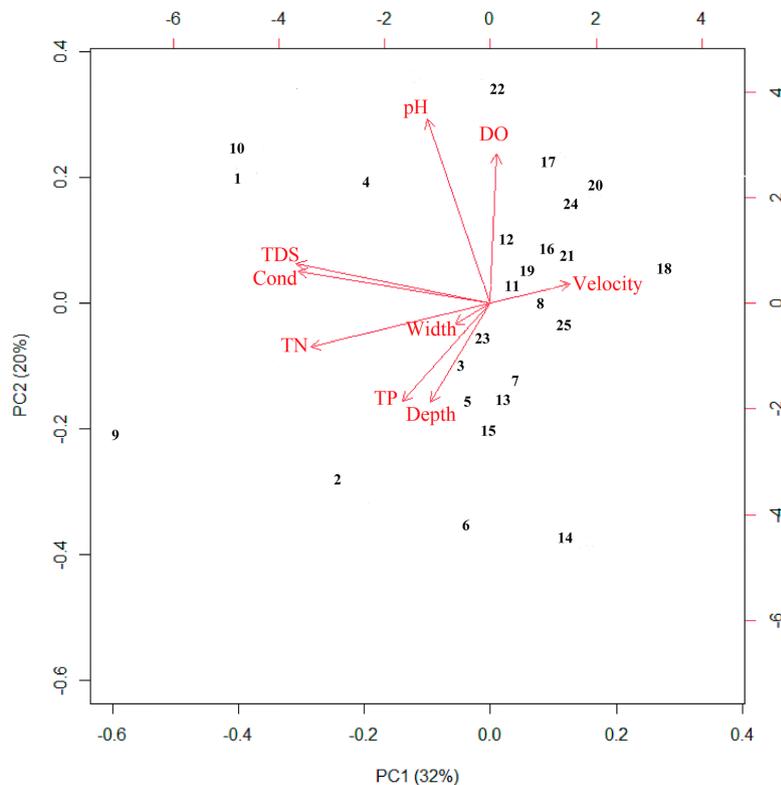
Figure 2. Principal Component Analysis (PCA) using environmental variables from 25 streams in the Brazilian Pampa. Numbers represent the streams sampled; TP = total phosphorus; TN = total nitrogen; Cond = electrical conductivity; TDS = total dissolved solids; DO = dissolved oxygen.

one was class 3 (300–599 μm³). Only the smallest biovolume class (1) did not occur in all streams.

Both surrogates showed significant but low congruence with the dataset at the species level when presence-absence data were used, with identification at the genus-level and biovolume class data showing correlations of 0.57 and 0.56 ($P < 0.001$) with the species dataset, respectively. As expected, when using abundance data, congruence with the species-level data increased for both the genus-level ($r = 0.97$; $P < 0.001$) and the biovolume-class ($r = 0.84$; $P < 0.001$) datasets.

The variation in diatom species composition was partially explained by pure environmental, pure spatial and spatially structured environmental components, which accounted for similar variation for both presence-absence or abundance data. However, those three components together explained 14% of the variation for abundance data but only 5% for presence-absence data (Table II; Figure 3). Furthermore, whilst only total nitrogen significantly explained the variation for presence-absence data, stream width, pH

and total nitrogen explained the variation in composition based on abundance data. The selected spatial variables were PCNMs 1 and 3 for presence-absence data and 1 and 5 for abundance data.

Regarding both surrogates, variation partitioning showed similar patterns. In all cases (i.e., genus-level identification and biovolume-class approaches for both presence-absence and abundance data), the pure spatial fraction did not significantly explain the variation in diatom composition. However, both approaches presented similar results as those found for species in relation to the environmental component (Table II; Figure 3). The dataset based on biovolume classes better explained variation in composition, with the pure environmental and the spatially structured environmental components accounting together for 39% and 25% of the variation for presence-absence and abundance data, respectively (Table II; Figure 3). More important is that the selected environmental variables were mostly the same as those selected for the species dataset: total nitrogen for presence-absence data and stream

Table II. Results of the partial redundancy analysis (pRDA) of environmental and spatial models explaining the variation in stream diatom community composition for datasets identified at the species level, genus level, or cell biovolume classes. Environmental and spatial variables (PCNM) selected in the final model. Values of significance ($P < 0.05$) for the global environmental (pG[E]) and global spatial (pG[S]) models. Adjusted R^2 values for pure environmental (Adj R^2 [E]), spatially-structured environmental (Adj R^2 [E–S]) and pure spatial (Adj R^2 [S]) fractions. Values of significance ($P < 0.05$) for the environmental (p[E]) and spatial (p[S]) fractions.

|	Environmental	Spatial	pG[E]	pG[S]	Adj R²[E]	Adj R²[E-S]	Adj R²[S]	p[E]	p[S]
Presence-absence									
Species	TN	1,3	**0.03**	**0.002**	0.02	0.01	0.02	**0.003**	**0.04**
Genus	TN	1	**0.02**	**0.040**	0.06	0.01	0.02	**0.001**	0.05
Size class	TN	1	**0.02**	**0.030**	0.27	0.12	0.00	**0.008**	0.96
Abundance									
Species	Width, pH, TN	1, 5	**0.001**	**0.006**	0.06	0.05	0.03	**0.007**	**0.04**
Genus	Width, pH, TN	1	**0.006**	**0.020**	0.11	0.04	0.01	**0.001**	0.13
Size class	Width, TN	1	**0.01**	**0.040**	0.16	0.09	0.01	**0.008**	0.26

width and total nitrogen for abundance data, for all three approaches (Table II). The exception was the relationship of species and genus-level datasets with pH, which was not selected when biovolume classes were used.

In fact, the directions of the relationships with environmental gradients were similar for all the three approaches (Figure 4). There was a strong gradient of total nitrogen at the first axis, in which species of *Cocconeis* (Figure 4a) and the genera *Cocconeis* and *Diadesmis* (Figure 4b) were positively associated with nitrogen-rich streams, mostly located at agricultural areas (see Figure 1), as was the case for biovolume class 5, the largest one (Figure 4c). The environmental gradient represented by the second axis of the RDAs showed that *Achnanthidium* and *Encyonema* were mainly related to narrow streams with high pH (Figure 4a and b), most of them occurring in grassland-dominated areas (see Figure 1), whereas *Navicula* was related to wide streams (Figure 4a and b) in agricultural areas. Moreover, the smallest biovolume classes (1, 2, and 3) were related to narrow streams with low total nitrogen concentrations (Figure 4c).

DISCUSSION

We found that both surrogate approaches, identification to the genus level and classification of species in biovolume classes, showed high congruence with the dataset identified to the species level when abundance data were used, but not for presence-absence data. Moreover, surrogates partly reflected the environmental responses of species but not the effects of space. These results indicate that both approaches can be good candidates to represent the composition of diatoms in Brazilian Pampa streams.

The efficiency of freshwater surrogates has long been investigated as an alternative for describing biodiversity patterns and defining more cost- and time-efficient biomonitoring programmes (e.g., Bevilacqua et al. 2015, Godoy et al. 2019, Riato et al. 2022). Here, we found that the identification of diatoms at a coarse taxonomic level and the use of a morphometric non-taxonomical approach present high congruence with the original species matrix based on abundance data. Moreover, we found that the use of surrogates reduces the
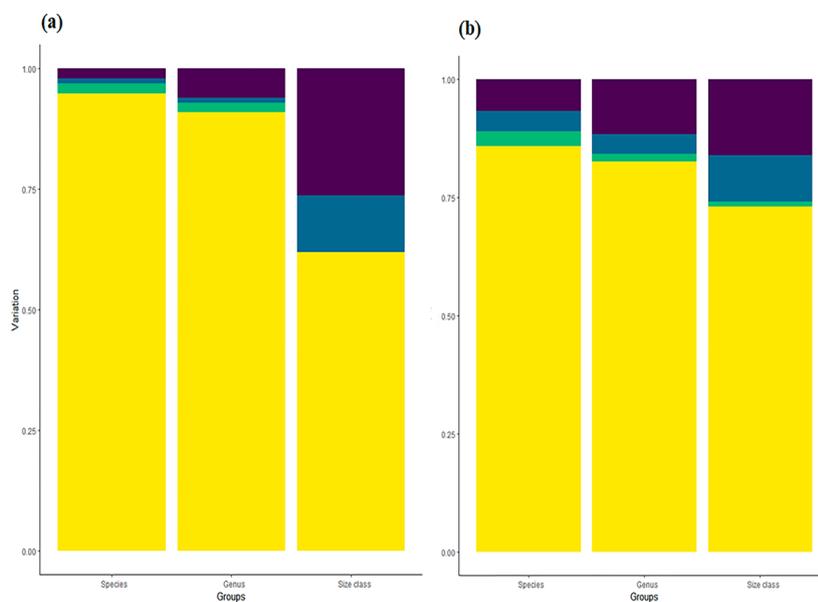

Figure 3. Partition of the variance of the diatom community composition (species and genus) and diatom biovolume classes for presence-absence (a) and abundance data (b). E = environmental; E+S = shared variation between space and environment; S= spatial; UN= Unexplained variation.

unexplained variation from partitioning analysis, which may be related to a reduction in the noise derived from multiple species in the analysis of community-environment relationships (e.g., Godoy et al. 2019). In this sense, our results are of great relevance, especially considering the demand of time and the effort required to assess communities at the species level for biomonitoring and ecological assessment programmes and considering the Linnean shortfall for microorganisms (e.g., Marrone et al. 2023), which may hinder the correct identification of species. This is true especially for less studied freshwater ecosystems, as is the case of Pampa streams in South America.

In addition to a high congruence, our results suggest that genus-level information can substitute species-level data to represent the main species-environment relationships of stream diatom communities. This has practical importance in the biomonitoring and ecological assessment context, for which useful surrogates or indicators need to detect the ecological status of the ecosystem of interest (Landeiro et al. 2012, Siqueira et al. 2012b). Moreover, we suggest that biovolume classes may be good surrogacy candidates to detect changes in the composition of diatom communities, but it may be premature to consider this approach to represent species-environment relationships.

This is because potential information may be lost due to the low number of environmental variables retained in the analysis in comparison with the species and genus-level approaches.

The environmental variables that best explained the variation in the composition of diatom communities clearly represent the ecological status of the studied streams. The importance of stream width in low-order Brazilian Pampa streams is probably related to habitat heterogeneity and differences in light incidence in the streambed (e.g., Pacheco et al. 2022). For example, despite the riparian zones of most of the sampled streams being dominated by natural grasslands, those streams less impacted by agriculture also have low-density arboreal riparian vegetation. In fact, the negative relationship of *Achnanthidium* with stream width reinforces the suggestion of a relation with light incidence, as this genus is characterised by organisms living in the bottom layer of the periphytic biofilm and thus adapted to the reduced light availability (Passy 2007, Schneck et al. 2022). Furthermore, nitrogen is related to the variation of diatom species among freshwater ecosystems (e.g., B-Berés et al. 2014, Huttunen et al. 2020). In our study, streams with high nitrogen concentrations are those within agricultural landscapes, mainly at the Santa Maria and Vacacaí-Mirim hydrographic basins,
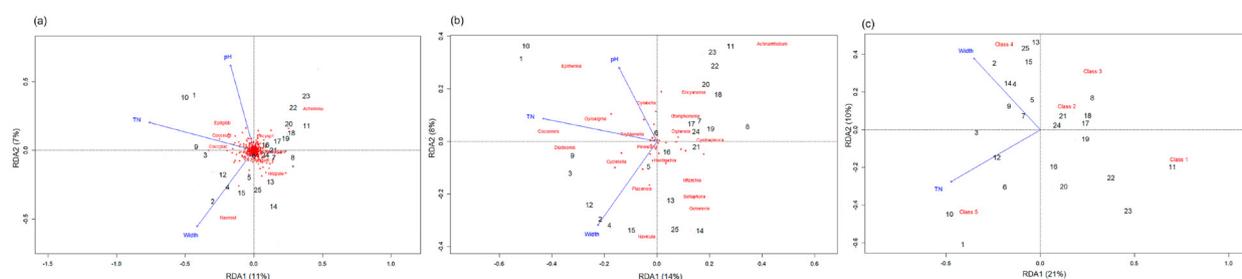
Figure 4. Redundancy Analysis (RDA) using diatom communities composition based on abundance data for species (a), genera (b) and biovolume classes (c). Species abbreviations: Achnminu: *Achnanthidium minutissimum*; Coccegl: *Cocconeis placentula* var. *euglypta*; Coccplac: *Cocconeis placentula*; Cymbtumi: *Cymbella tumida*; Encysp1: *Encyonema* sp1; Naviros: *Navicula* cf. *rostellata*; Nitzpale: *Nitzschia palea*. Numbers represent the streams sampled; TN = total nitrogen.

in which nutrients are more easily leached to streams (Carpenter et al. 1998). Although nitrogen is often a limiting factor for algal growth (Kumar & Bera 2020), the excessive input of nitrogen of anthropogenic origin (i.e. agricultural runoff in the context of our study) degrades water quality, leading to eutrophication (Schindler 2006) and changes in the composition of freshwater diatom communities (e.g. Tan et al. 2014, Huttunen et al. 2020), as we found here. For instance, *A. minutissimim* and the genus *Encyonema*, which showed a negative relationship with nitrogen-rich streams, are commonly classified as sensitive to multiple stressors from agricultural runoff (e.g., Waite et al. 2020). On the other hand, diatoms usually documented in polluted streams, such as *C. placentula* (e.g., Çelekli et al. 2019) and *Diadesmis* (e.g., Wood et al. 2019) were positively related with total nitrogen in our study. These results indicate that those species and genera more strongly related to nitrogen concentrations could be good indicators of agricultural impacts in Brazilian Pampa streams. Finally, regarding biovolume classes, the relation of large classes with agricultural streams can be explained by the limitation of nitrogen. This is because the increase in cell volume in algae communities can be drive by nitrogen availability (Peter & Sommer 2014). Thus, we suggested that larger biovolume classes may be indicative of streams influenced by nitrogen input, while smaller biovolume classes may be indicative of grassland streams in Brazilian Pampa.

In our study, the spatial component was important only to explain the variation in the composition of species. The relevance of space for diatom communities may be associated with the differences in their dispersal capacity due to the auto-ecological characteristics of species within communities. For example, planktonic and motile species can benefit from dispersal through mass effects due to water flow in relation to attached diatoms (e.g., Jamoneau et al. 2018). Moreover, by grouping species into genera or biovolume classes, we simplified the data matrix in a way that most groups occur in all sites, reducing the pure spatial effect that may hinder the responses of organisms to major environmental factors of interest for biomonitoring and assessment programmes. As highlighted by Siqueira et al. (2012b), better indicator groups are those for which the metacommunity structure is determined mostly by environmental factors and not by spatial ones.

We found that the use of abundance data resulted in higher percentages of explained variance than presence-absence data, for both species and genera datasets. This effect of numerical resolution in analyses of communities is commonly observed (e.g., Mueller et al. 2013, Szabó et al. 2018, Sgarbi et al. 2020) and related to the larger amount of ecological information provided by common species compared to rare species (Blanchet et al. 2016) Abundance data improve the detection of more subtle differences among communities, whereas incidence data are better in detecting large differences in species composition (Melo 2005, Blanchet et al. 2016), as may be the case in studies performed at larger spatial or environmental extents than ours (Wilson 2012). Thus, the choice between using incidence or abundance data should be a compromise between the effort needed to obtain abundance information, the spatial extent of the sampling, and the detail of ecological information needed for a specific purpose.

We recognise that the main caveat of our study is that our analyses were based on a single sampling in 25 streams. In this context, we cannot predict whether the surrogates would represent the variation in diatom species composition and the ecological responses of species at different seasons. Moreover, the

snap-shot sampling may partly explain the high percentage of unexplained variation in the composition of diatom communities. This high percentage of unexplained variation is common in metacommunity studies with diatoms (e.g., Soininen et al. 2016, Szabó et al. 2018), and may be related to the stochasticity of community organization or to processes involving unmeasured environmental variables (e.g., Szabó et al. 2018), such as silica content, heavy metals and pesticides or natural disturbance regimes resulting in uneven spate-producing rainfall. In this sense, to improve the knowledge of the validity of using surrogates for biomonitoring purposes in streams from the Pampa region, we suggest that a necessary further step is to test the surrogacy approach over time and, most importantly, over a larger spatial extent.

In conclusion, there is a high congruence between diatom communities at the levels of species and genus and between species and biovolume classes when abundance data are used. We also found that all datasets showed similar responses in relation to environmental factors and that the surrogates were not related to spatial factors, a desired characteristic for bioindicator candidates. We suggest that genus level is a good surrogate candidate to provide information about the species-environment relationships, whereas biovolume classes may be useful to detect changes in diatom communities. However, it is important to highlight that the use of such surrogates may affect our ability to detect species losses or more subtle changes in the composition of communities due to anthropogenic impacts, and should thus be used for specific purposes. Moreover, the use of biovolume classes may prevent the detection of disturbances based on the environmental tolerances of species. Finally, our results contribute to a deeper knowledge of surrogacy in freshwater ecosystems from the Brazilian Pampa biome and its importance to simplify aquatic biomonitoring and ecological assessments when using diatoms as bioindicators, especially in relation to taxonomic difficulties at the species level. Thus, the choice of surrogates to the detriment of species-level information may be an alternative to be evaluated for reducing the time and effort required for biomonitoring and ecological assessment studies.

REFERENCES

ALGARTE VM, RODRIGUES L, LANDEIRO VL, SIQUEIRA T & BINI LM. 2014. Variance partitioning of deconstructed periphyton communities: does the use of biological traits matter? Hydrobiologia 722(1): 279-790.

ALLEN S, GRIMSHA WM, PARKINSON JA & QUARMBY C. 1974. Chemical analysis of ecological materials. London: Blackwell Scientific Publishers, 565 p.

ALVARES CA, STAPE JL, SENTELHAS PC, GONÇALVES JDM & SPAVOREK G. 2013. Köppen's climate classification map for Brazil. Meteorol Z 22(6): 711-726.

BAEZA S ET AL. 2022. Two decades of land cover mapping in the Río de la Plata grassland regions: The MapBiomas Pampa initiative. Remote Sens Appl: Soc Environ 28: 100835.

BAUMGARTEN MGZ, ROCHA JMB & NIENCHESKI LFH. 1996. Manual de análises em oceanografia química. Rio Grande: Editora da FURG, 132 p.

B-BERÉS V, TÖRÖK P, KÓKAI E, KRASZNAI T, TÓTHMÉRÉSZ B & BÁCSI I. 2014. Ecological diatom guilds are useful but not

sensitive enough as indicators of extremely changing water regimes. Hydrobiologia 738: 191-204.

BEVILACQUA S, TERLIZZI A, MISTRI M & MUNARI C. 2015. New frameworks for species surrogacy in monitoring highly variable coastal ecosystems: Applying the BestAgg approach to Mediterranean coastal lagoons. Ecol Indic 52: 207-218.

BLANCHET FG, LEGENDRE P & BORCARD D. 2008. Forward selection of explanatory variables. Ecography 89(9): 2623-2632.

BLANCHET FG, LEGENDRE P & HE F. 2016. A new cost-effective approach to survey ecological communities. Oikos 125: 975-987.

BORCARD D & LEGENDRE P. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. Ecol Modell 153(1-2): 51-68.

CARNEIRO FM, BINI LM & RODRIGUES LC. 2010. Influence of taxonomic and numerical resolution on the analysis of temporal changes in phytoplankton communities. Ecol Indic 10(2): 249-255.

CARPENTER SR, CARACO NF, CORRELL DL, HOWARTH RW, SHARPLEY AN & SMITH VH. 1998. Nonpoint pollution of surface waters with phosphorus and nitrogen. Ecol Appl 8(3): 559-568.

ÇELEKLI A, KAYHAN S, LEKESIZ O, TOUDJANI AA & ÇETIN T. 2019. Limno-ecological assessment of Aras River surface waters in Turkay: applications of diatom indices. Environ Sci Pollut Res 26: 8028-8038.

COSTA APT & SCHNECK F. 2022. Diatoms as indicators in running waters: trends of studies on biological assessment and monitoring. Environ Monit Assess 194(10): 695.

DE OLIVEIRA SS, ORTEGA JCG, RIBAS LGS, LOPES VG & BINI LM. 2020. Higher taxa are sufficient to represent biodiversity patterns. Ecol Indic 111: 105994.

DRAY S, BLANCHET G, BORCARD D, GUENARD G, JOMBART T, LAROCQUE G, LEGENDRE P, MADI N & WAGNER HH. 2016. adespatial: Multivariate Multiscale Spatial Analysis. R Package version 0.0-7 [available on internet at https://CRAN.R-project.org/package=adespatial].

FOX J & WEISBERG S. 2011. An R Companion to Applied Regression. Sage, Thousand Oaks.

GALLEGO I ET AL. 2012. Taxonomic or ecological approaches? Searching for phytoplankton surrogates in the determination of richness and assemblage composition in ponds. Ecol Indic 18: 575-585.

GODOY BS, FARIA APJ, JUEN L, SARA L & OLIVEIRA LG. 2019. Taxonomic sufficiency and effects of environmental and spatial drives on aquatic insect community. Ecol Indic 107: 105624.

HORTAL J, BELLO F, DINIZ-FILHO JAF, LEWINSOHN TM, LOBO JM & LADLE RJ. 2015. Seven shortfalls the beset large-scale knowledge of biodiversity. Annu Rev Ecol Evol Syst 46(1): 523-549.

HUTTUNEN KL, MUOTKA T, KARJALAINEN SM, LAAMANEN T & AROVIITA J. 2020. Excess of nitrogen reduces temporal variability of streams diatom assemblages. Sci Total Environ 713: 136630.

JACKSON DA. 1993. Stopping rules in a principal component analysis: a comparison of heuristical and statistical approaches. Ecology 74(8): 2204-2214.

JACKSON DA. 1995. PROTEST: a PROcrustean randomization TEST of community environment concordance. Ecoscience 2(3): 297-303.

JAMONEAU A, PASSY SI, SOININEN J, LEBOUCHER T & TISON-ROSEBERY J. 2018. Beta diversity of diatom species and ecological guilds: Response to environmental and spatial mechanisms along the stream watercourse. Freshw Biol 63(1): 62-73.

KUMAR A & BERA S. 2020. Revisiting nitrogen utilization in algae: a review on the process of regulation and assimilation. Bioresour Technol Rep 12: 100584.

LANDEIRO VL, BINI LM, COSTA FRC, FRANKLIN E, NOGUEIRA A, SOUZA JLP, MORAES J & MAGNUSSON WE. 2012. How far can go in simplifying biomonitoring assessment? An integrated analysis of taxonomic surrogacy, taxonomic sufficiency and numerical resolution in a megadiverse region. Ecol Indic 23: 366-373.

LEGENDRE P & LEGENDRE L. 1998. Numerical Ecology. Amsterdam: Elsevier, 853 p.

LEGENDRE P & GALLAGHER ED. 2001. Ecologically meaningful transformations for ordination of species data. Oecologia 129(2): 271-280.

MARGULES CR & PRESSEY RL. 2000. Systematic conservation planning. Nature 405(6783): 243-253.

MARRONE F, FONTANETO D & NASSELI-FLORES L. 2023. Cryptic diversity, niche displacement and our poor understanding of taxonomy and ecology of aquatic microorganisms. Hydrobiologia 850(6): 1221-1236.

MELO AS. 2005. Effects of taxonomic and numeric resolution on the ability to detect ecological patterns at a local scale using stream macroinvertebrates. Arch Hydrobiol 164: 309-323.

METZELTIN D, LANGE-BERTALOT H & GARCÍA-RODRÍGUEZ F. 2005. Diatoms of Uruguay - taxonomy, biogeography and diversity. Königstein: A.R.G. Gantner Verlag, 736 p.

METZELTIN D & GARCÍA-RODRÍGUEZ F. 2012. Las diatomeas uruguayas. Montevideo: DIRAC, 208 p.

MUELLER M, PANDER J & GEIST J. 2013. Taxonomic sufficiency in freshwater ecosystems: effects of taxonomic resolution, functional traits, and data transformation. Aquat Sci 32(3): 762-778.

OKSANEN J, BLANCHET FG, KINDT R, LEGENDRE P, MINCHIN PR, O'HARA RB, SIMPSOM GL, SOLYMOS P, STEVENS MHH & WAGNER H. 2017. vegan: Community Ecology Package. R Package version 2.4-3 [available on internet at http://CRAN.R-project.org/package=vegan].

OLIVEIRA TE, FREITAS DS, GIANEZINI M, RUVIARO CF, ZAGO D, MÉRCIO TZ, DIAS EA, LAMPERT VN & BARCELLOS JOJ. 2017. Agricultural land use change in the Brazilian Pampa biome: The reduction of natural grasslands. Land Use Policy 63: 394-400.

OVERBECK GE, MULLER SC, FIDELIS A, PFADENHAUER J, PILLAR VD, BLANCO CC, BOLDRINI II, BOTH R & FORNECK ED. 2007. Brazil's neglected biome: the southern Brazilian Campos. Plant Ecol Evol Syst 9(2): 101-116.

PACHECO JP, CALVO C, AZNAREZ C, BARRIOS M, MEERHOFF M, JEPPESEN E, PEDERSEN AB. 2022. Periphyton biomass and life-form responses to a gradient of discharge in contrasting light and nutrient scenarios in experimental lowland streams. Sci Total Environ 806: 150505.

PANDEY LK, SHARMA YC, PARK J, CHOI S, LEE H, LYU J & HAN T. 2018a. Evaluating features of periphytic diatom communities as biomonitoring tools in fresh, brackish and marine waters. Aquat Toxicol 194: 67-77.

PANDEY LK, LAVOIE I, MORIN S, PARK J, LYU S, LEE H & HAN T. 2018b. River water quality assessment based on a multi-descriptor approach including chemistry, diatom assemblage structure, and non-taxonomical diatom metrics. Ecol Indic 84: 140-151.

PASSY SI. 2007. Diatom ecological guilds display distinct and predictable behavior along nutrient and disturbance gradients in running waters. Aquat Bot 86(2): 171-178.

PERES-NETO PR & JACKSON DA. 2001. How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. Oecologia 129: 169-178.

PERES-NETO PR, LEGENDRE P, DRAY S & BORCARD D. 2006. Variation partitioning of species data matrices: estimation and comparison of fractions. Ecology 87(10): 2614-2625.

PETER KH & SOMMER U. 2014. Interactive effect of warming, nitrogen and phosphorus limitation on phytoplankton cell size. Ecol Evol 5(5): 1011-1024.

PETSCH DK. 2016. Causes and consequences of biotic homogenization in freshwater ecosystems. Int Rev Hydrobiol 101(3-4): 113-122.

R CORE TEAM. 2020. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. http://www.R-project.org/.

RIATO L, HILL RA, HERLIHY AT, PECK DV, KAUFMANN PR, STODDARD JL & PAULSEN SG. 2022. Genus-level, trait-based multimetric diatom indices for assessing the ecological condition rivers and streams across the conterminous United States. Ecol Indic 141: 109131.

RIBEIRO L, BROTAS V, HERNÁNDEZ-FARIÑAS T, JESUS B & BARRILÉ L. 2020. Assessing alternative microscopy-based approaches to species abundance description of intertital diatom communities. Front Mar Sci 7: 36.

RIMET F & BOUCHEZ A. 2012a. Biomonitoring river diatoms: implications of taxonomic resolution. Ecol Indic 15(1): 92-99.

RIMET F & BOUCHEZ A. 2012b. Life form, cell-sizes and ecological guilds of diatoms in European rivers. Knowl Manag Aquat Ecosyst 406: 1.

SCHINDLER DW. 2006. Recent advances in the understanding and management of eutrophication. Limnol Oceanogr 51(1part2): 356-363.

SCHNECK F, BINI LM, MELO AS, PETSCH DK, SAITO VS, WENGRAT S & SIQUEIRA T. 2022. Catchment scale deforestation increases the uniqueness of subtropical stream communities. Oecologia 199: 671-683.

SGARBI LF ET AL. 2020. Sampling effort and information quality provided by rare and common species in estimating assemblage structure. Ecol Indic 110: 105937.

SIQUEIRA T, BINI LM, ROQUE FO, COUCEIRO FRM, TRIVINHO-STRIXINO S & COTTENIE K. 2012a. Common and rare species respond to similar niche processes in macroinvertebrate metacommunities. Ecography 35(2): 183-192.

SIQUEIRA T, BINI LM, ROQUE FO & COTTENIE K. 2012b. A metacommunity framework for enhancing the effectiveness of biological monitoring strategies. PloS One 7: e43626.

SIQUEIRA T, LACERDA CGLT & SAITO VS. 2015. How does landscape modification induce biological homogenization in tropical stream communities? Biotropica 47(4): 509-516.

SOININEN J, JAMONEAU A, ROSEBERY J & PASSY SI. 2016. Global patterns and drivers of species and trait composition in diatoms. Global Ecol Biogeogr 25: 940-950.

SORIANO A. 1991. Rio de Plata Grasslands. In: Coupland RT (Ed), Ecosystems of the World: Natural Grasslands, New York: Elsevier, p. 367-408.

STEVENSON RJ, PAN Y & VAN DAM H. 2010. Assessing environmental conditions in rivers and streams with diatoms. In: Smol J & Stoermer E (Eds), The Diatoms: Applications for the Environmental and Earth Sciences, Cambridge: Cambridge University Press, p. 57-85.

SZABÓ B, LENGYEL E, PADISÁK J & STENGER-KOVÁCS C. 2018. Benthic diatom metacommunity across small freshwater lakes: driving mechanisms, ß-diversity and ecological uniqueness. Hydrobiologia 828: 183-198.

TAN X, XIA X, ZHAO Q & ZHANG Q. 2014. Temporal variations of benthic diatom community and its main influencing factors in a subtropical river, China. Environ Sci Pollut Res 21: 434-444.

TRINDADE CRT, LANDEIRO VL & SCHNECK F. 2018. Macrophyte functional groups elucidate the relative role of environmental and spatial factors on species richness and assemblage structure. Hydrobiologia 823: 217-230.

VALDERRAMA JC. 1981. The simultaneous analysis of total nitrogen and total phosphorus in natural waters. Mar Chem 10(2): 109-112.

VIEIRA MC, ORTEGA JCG, VIEIRA LCG, VELHO LFM & BINI LM. 2022. Evidence that dams promote biotic differentiation of zooplankton communities in two Brazilian reservoirs. Hydrobiologia 849(3): 697-709.

VIRTANEN L & SOININEN J. 2012. The roles of environment and space in shaping stream diatom communities. Eur J Phycol 47(2): 160-168.

WAITE IR, PAN I & EDWARDS PM. 2020. Assessment of multi-stressors on compositional turnover of diatom, invertebrate and fish assemblages along an urban gradient in Pacific Northwest streams (USA). Ecol Indic 112: 106047.

WATSON JEM, JONES KR, FULLER RA, DI MARCO M, SEGAN DB, BUTCHART SHM, ALLAN JR, MCDONALD-MADDEN E & VENTER O. 2016. Persistent disparities between recent rates of habitat conversion and protection and implications for future global conservation targets. Conserv Lett 9(6): 413-421.

WICKHAM H. 2016. Ggplot2: Elegant graphics for data analysis. Houston: Springer.

WILSON JB. 2012. Species presence/absence sometimes represents a plant community as well as species abundances do, or better. J Veg Sci 23: 1013-1023.

WOOD RJ, MITROVIC SM, LIM RP, WARNE MSJ, DUNLOP J & KEFFORD BJ. 2019. Benthic diatoms as indicators of herbicide toxicity in rivers - A new SPEcies at Risk (SPEARherbicides) index. Ecol Indic 99: 203-213.

WUNSAM S, CATTANEO A & BOURASSA N. 2022. Comparing diatom species, genera and size in biomonitoring: a case study from streams in the Laurentians (Québec, Canada). Freshw Biol 47(2): 325-340.

SUPPLEMENTARY MATERIAL

Table SI.

ANA PAULA TAVARES COSTA
https://orcid.org/0000-0003-1938-1299

FABIANA SCHNECK
https://orcid.org/0000-0002-0018-444X

Universidade Federal do Rio Grande - FURG, Programa de Pós-Graduação em Biologia de Ambientes Aquáticos Continentais, Instituto de Ciências Biológicas, Avenida Itália, Km 8, 96203-900 Rio Grande, RS, Brazil

Correspondence to: **Ana Paula Tavares Costa**
E-mail: anapaulatavarescosta@gmail.com

Author contributions

Conceived and designed the study: APTC and FS. Collected the data: APTC. Analyzed and interpreted the data: APTC and FS. Drafted and revised the manuscript: APTC and FS.