Science of the Total Environment 754 (2021) 142171

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

# Double constrained ordination for assessing biological trait responses to multiple stressors: A case study with benthic macroinvertebrate communities

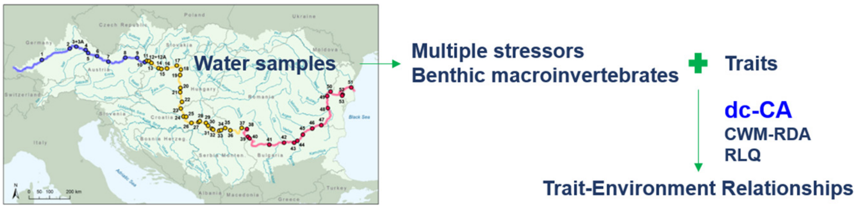
Feng-Jiao Peng [a,b,*], Cajo J.F. ter Braak [c], Andreu Rico [d], Paul J. Van den Brink [a,e]

[a] *Aquatic Ecology and Water Quality Management group, Wageningen University, P.O. Box 47, 6700 AA Wageningen, the Netherlands*
[b] *Human Biomonitoring Research Unit, Department of Population Health, Luxembourg Institute of Health, 1 A-B rue Thomas Edison, 1445 Strassen, Luxembourg*
[c] *Biometris, Wageningen University & Research, PO Box 16, 6700 AA Wageningen, the Netherlands*
[d] *IMDEA Water Institute, Science and Technology Campus of the University of Alcalá, Avenida Punto Com 2, 28805, Alcalá de Henares, Madrid, Spain*
[e] *Wageningen Environmental Research, Wageningen University & Research, P.O. Box 47, 6700 AA Wageningen, the Netherlands*

## HIGHLIGHTS

- CWM-RDA and RLQ have drawbacks to examine trait-environment relationships.
- We compare dc-CA with CWM-RDA and RLQ for analysis of trait-environment relationships.
- We illustrate the application and benefits of dc-CA to biological data analysis.
- Human impact on macroinvertebrates is of similar magnitude to the natural impact.
- dc-CA may be useful to assess mechanistic links between multiple stressors and ecosystem health.

## GRAPHICAL ABSTRACT



## ARTICLE INFO

*Article history:*
Received 2 June 2020
Received in revised form 1 September 2020
Accepted 1 September 2020
Available online 2 September 2020

Editor: Sergi Sabater

*Keywords:*
Biomonitoring
Benthic macroinvertebrates
Multiple stressors
RLQ
Double constrained correspondence analysis
Trait-environment relationships

## ABSTRACT

Benthic macroinvertebrate communities are used as indicators for anthropogenic stress in freshwater ecosystems. To better understand the relationship between anthropogenic stress and changes in macroinvertebrate community composition, it is important to understand how different stressors and species traits are associated, and how these associations influence variation in species occurrence and abundances. Here, we show the capacity of the multivariate technique of double constrained correspondence analysis (dc-CA) to analyse trait-environment relationships, and we compare it with the redundancy analysis method on community weighted mean values of traits (CWM-RDA), which is frequently used for this type of analysis. The analyses were based on available biomonitoring data for macroinvertebrate communities from the Danube River. Results from forward selection of traits and environmental variables using dc-CA analyses showed that aquatic stages, reproduction techniques, dispersal tactics, locomotion and substrate relations, altitude, longitudinal and transversal distribution, and substrate preferendum were significantly related to habitat characteristics, hydromorphological alterations and water quality measurements such as physico-chemical parameters, heavy metals, pesticides and pharmaceuticals. Environmental variables significantly associated with traits using the CWM-RDA method were generally consistent with those found in dc-CA analysis. However, the CWM-RDA does neither test nor explicitly select traits, while dc-CA tests and selects both traits and environmental variables. Moreover, the dc-CA analysis revealed that the set of environmental variables was much better in explaining the community data than the available trait set, a kind of information that can neither be obtained from CWM-RDA nor from RLQ

---

[*] Corresponding author at: Aquatic Ecology and Water Quality Management Group, Wageningen University, P.O. Box 47, 6700 AA Wageningen, the Netherlands.
*E-mail address:* fengjiaopeng@gmail.com (F.-J. Peng).

https://doi.org/10.1016/j.scitotenv.2020.142171

(Environment, Link and Trait data), which is a close cousin of dc-CA but not regression-based. Our results suggest that trait-based analysis based on dc-CA may be useful to assess mechanistic links between multiple anthropogenic stressors and ecosystem health, but more data sets should be analysed in the same manner.

## 1. Introduction

Freshwater ecosystems often receive a wide range of chemical inputs from urban, industrial and agricultural activities (Rico et al., 2019; Loos et al., 2017). Chemical mixtures together with other anthropogenic stressors such as habitat disturbance, hydrological alterations or organic matter loads may negatively affect benthic macroinvertebrate communities (Rico et al., 2016), which play an important role in energy flow transfer across trophic levels and contribute to organic matter decomposition and nutrient cycling (Wallace and Webster, 1996; Covich et al., 1999). On the other hand, benthic macroinvertebrates are good candidates to assess the impact of human pressure on freshwater ecosystems because of their short lifespan and their relatively high sensitivity to some chemical classes (e.g. pesticides, home-care products) (Resh and Rosenberg, 1993), and are regularly monitored to assess the ecological status of surface waters (EC, 2000).

Traits are phenotypic or ecological characteristics of an organism, usually measured at the individual level, but often applied to characterize the ecology of a given species, to reflect on their potential resilience and resistance to environmental disturbances (Townsend and Hildrew, 1994; Van den Brink et al., 2011). For instance, maximal size, reproductive method, or feeding strategies have the potential to reflect the adaptation of species to relatively disturbed environments (Gayraud et al., 2003). Trait-based approaches use species traits together with their taxonomic classification to assess ecosystem responses against environmental change (Baird and Van den Brink, 2007). Therefore, these approaches can improve the ability to understand the structure and dynamics of ecological communities and potentially predict their sensitivity and recovery potential to natural or human disturbances (Keddy, 1992; Díaz and Cabido, 1997; Dray et al., 2014). In fact, biological traits have been widely used in the field of ecological risk assessment (ERA) and biomonitoring of aquatic ecosystems to reveal impacts of various types of human disturbance on invertebrate communities, including organic pollution, heavy metal pollution, cargo-ship traffic, eutrophication, land use and hydrological alterations (e.g., Dolédec and Statzner, 2008; Statzner and Beche, 2010; Culp et al., 2011; De Castro-Català et al., 2015; Rico and Van den Brink, 2015; Kuzmanović et al., 2017; Lemm and Feld, 2017), especially in the temperate zone where biological attributes of invertebrates are well documented (Tachet et al., 2010). For example, Kuzmanović et al. (2017) found that pesticides, high metal pollution, nutrients and flow alterations significantly affected trait composition of benthic macroinvertebrate assemblages in polluted Iberian rivers.

There are several approaches to examine trait-environment relationships. Currently, the most widely used approach is perhaps redundancy analysis (RDA) of community weighted mean trait values (CWM-RDA), where the weights are the species abundances in the monitoring samples (Lavorel et al., 2008; Kleyer et al., 2012) with RLQ (Environment, Link and Trait data; Dray et al., 2014) as runner up. The CWM-RDA approach is a community-based analysis and thus can only test the significance of the effects of environmental variables but not the significance of traits (Peres-Neto et al., 2017). Moreover, it yields elevated Type I error rate when only the environment is important in structuring species distributions, i.e. when there is in fact no trait-environment association (Peres-Neto et al., 2017). As a result, Peres-Neto et al. (2017) proposed to replace the use of CWM-RDA by more robust statistical methods. Recent advancements on statistical methods show that double constrained correspondence analysis (dc-CA) is a direct and powerful method to predict composition of community assemblages from environmental and trait predictors, which has advantages over other methods like CWM-RDA and RLQ for measuring and testing the direct link between variations in species traits and environmental variables (Dolédec et al., 1996; Dray et al., 2014; Ter Braak et al., 2018a). Dc-CA is very similar in spirit as RLQ; it is in fact its regression-based cousin, whereas RLQ is covariance (coinertia) based. Both dc-CA and RLQ are based on an ordination of a rectangular response data table in which the scores of both rows (samples) and columns (species) are constrained by linear combinations of predictor variables (i.e., environmental variables and traits), which attempts to find a low-dimensional representation of the interaction effects of traits and environmental variables on invertebrate communities in a log-linear model (Ter Braak et al., 2018a). Both do so by finding linear combinations of traits and of environmental variables. RLQ does so by maximizing covariance, whereas dc-CA does so by maximizing the fourth corner correlation between these linear combinations using weighted-least squares, where the weights are the site and species totals (Ter Braak et al., 2018a). The power and efficiency of the fourth corner correlation for testing trait-environment interaction was shown in Ter Braak (2017). Compared to RLQ, dc-CA and CWM-RDA take into account correlations among environmental variables. Compared to CWM-RDA, dc-CA takes into account the correlation among traits as well (Ter Braak et al., 2017, 2018a). More importantly, in simulations dc-CA outperforms the RLQ and CWM-RDA in terms of statistics' sampling accuracy and statistical power (Peres-Neto et al., 2017; Ter Braak et al., 2018a). However, to the best of our knowledge, there are no studies that apply dc-CA on real biomonitoring data.

In the present study, we assessed trait-environment relations of benthic macroinvertebrate communities sampled from the Danube River by using dc-CA. We also performed a CWM-RDA analysis and (to a lesser extent) RLQ and compared their outcomes with those provided by dc-CA. As such, we aimed to illustrate the application and the benefits of dc-CA to biological data analysis, which is expected to expand its application to establish causal relationships between specific stressors and community responses in future evaluations.

## 2. Materials and methods

### *2.1. Background of dc-CA and comparison with CWM-RDA*

Double constrained ordination is an ordination of a rectangular response data table in which the scores of both rows (samples) and columns (species) are constrained by linear combinations of predictor variables (i.e., environmental variables and traits) (see Ter Braak et al., 2018a for algorithms). In the present study, the predictor variables for the rows are called environmental variables and the predictor variables for the columns are called traits. As in RLQ, there are thus three data tables: a rectangular $n \times m$ table Y containing the abundances (or presence/absence) of species ($m$) in samples ($n$), a rectangular $n \times q$ table E with the values of environmental variables ($q$) in samples ($n$), and a rectangular $m \times p$ table T describing traits ($p$) of species ($m$). A brief comparison of dc-CA with RLQ is provided in the Discussion section. In dc-CA, a fourth table is needed, which is the transpose of the first one, i.e. a rectangular $m \times n$ table $Y^T$ containing the abundances (or presence/absence) of species ($m$) in the samples ($n$) (Ter Braak et al., 2018a). The fourth table ($Y^T$) allows for species-level analysis, while the first table (Y) allows for community-level analysis. The fourth table allows a) for permutation tests with the response variables (columns of Y) as the statistical units so as to test for the significance of the trait effects on the response data Y and b) for determination of how much variation in the abundance values can be explained by the traits. Without any statistical testing and variation decomposition, the analysis can be performed in a single step (a singular value decomposition), but often it is also of interest how well the traits and the environmental variables each explain the community data using single constrained ordinations (steps 1 and 3 below) and how well the environmental variables explain the trait-structured variation and how well the traits explain the environmentally structured variation (Peres-Neto et al., 2017). The full analysis thus needs four steps. Two more steps are required a) to determine what the maximum strength of the trait-environment relation could have been for a given abundance matrix Y (step 5 below) and b) to clue the results of all steps together (step 6). The six steps for the dc-CA are given below:

1. The first step is a canonical correspondence analysis (CCA; Y ~ E) constraining the species composition by environmental variables. This step allows to determine how much variation in the abundance values can be explained by the environmental variables. It also gives species scores ($S^*$) to be used in the next step; $S^*$ is an $m \times q^*$ table of scores with $q^*$ the rank of the environmental variables and its entries are species niche centroids (Peres-Neto et al., 2017) of the orthonormalized environmental variables. $S^*$ represents the environmentally structured variation in the abundance table.
2. The second step is a weighed RDA ($S^*$ ~ T), which turns the single constrained ordination of step 1 into a double constrained ordination (Ter Braak et al., 2018a). This step is a species-level analysis, which gives a significance test of whether the traits modulate the species-environment relationship and a measure of how well the traits explain the environmentally structured variation, namely the percentage variance in the species niche centroids (in $S^*$) that is explained by the traits.
3. The third step is a CCA ($Y^T$ ~ T) constraining the transposed species composition by traits. This step allows to determine how much variation in the abundance values can be explained by the traits. It also gives sample scores ($R^*$) to be used in the next step; $R^*$ is an $n \times p^*$ table of scores with $p^*$ the rank of the trait data and its entries are community weighted means (Peres-Neto et al., 2017) of the orthonormalized traits. $R^*$ represents the trait-structured variation in the abundance table.
4. The fourth step is a weighted RDA ($R^*$ ~ E), which turns the single constrained ordination of step 3 into a double constrained ordination that is identical to the one obtained in step 2 (Ter Braak et al., 2018a). This step is a case-level (community-level) analysis, which gives a significance test of whether the environmental variables modulate the species-trait relationship and a measure of how well the environmental variables explain the trait-structured variation, namely the percentage variance in the community weighted means (in $R^*$) that is explained by the environmental variables.
5. The fifth step is a CA to find scores for samples and species that maximize the fourth-corner correlation without any constraints on rows and columns of Y.
6. The last step is used for calculating the dc-CA scores from steps 1–4, in which all resulting summary statistics and ordination scores are collected. The number of double constrained axes is equal to the minimum of the number of environmental variables and the number of traits.

Compared to dc-CA, CWM-RDA consists only of step 4 with $R^*$ being simply the community weighted means of the individual traits. CWM-RDA can therefore not establish whether the analysed traits are really related to the environmental variables, because even random traits may yield significant relationships (Peres-Neto et al., 2017). For this, one also needs to perform a species-level analysis (such as in step 2).

### *2.2. Danube River dataset*

The Danube River dataset was used in this study, which has been previously described in detail in Rico et al. (2016). In that study, variance partitioning was used in an RDA and CMW-RDA approach to establish taxonomic and trait-environment relationships, and to evaluate how much of the variation in trait means and species abundances can be explained by different anthropogenic stressors (Rico et al., 2016). It contains 55 sampling sites located in eight countries of Central and Eastern Europe, and 393 macroinvertebrate taxa (see Dataset S1 in Supporting Information A for taxa). Samples were collected between the 13th of August and the 25th of September of 2013 (Fig. 1). Benthic macroinvertebrates were sampled in the littoral zone (up to a maximum water depth of 1.5 m) using the Multi-Habitat Sampling (MHS) method, stored in formaldehyde (4%) in-situ and identified to the lowest taxonomic resolution. A more detailed description of sampling methods can be found in Liška et al. (2015).

The environmental data were obtained from Rico et al. (2016). We divided the full set of 285 environmental variables (Dataset S2) into nine groups (number of parameters in parentheses): habitat characteristics (13), hydromorphological alterations (10), physico-chemical parameters (19), heavy metals (8), pesticides (39), pharmaceuticals (134), industrial chemicals (46), home and personal care products (HPCPs; 4) and miscellaneous compounds (12). Here habitat characteristics and hydromorphological alterations were considered natural variables while the remaining seven groups were considered anthropogenic variables. The substrate PC1 and substrate PC2 in the habitat characteristics group were derived from PCA on seven substrate variables (Fig. S1), which respectively correspond to scores on the first and second principal components, as in Rico et al. (2016). Concentration values below the limit of detection (LOD) were converted to half of the LOD for data analysis. In the present study we examined the trait-environment relationships based on concentration data of individual chemicals, whereas Rico et al. (2016) used the Toxic Unit (TU) approach based on acute toxicity data ($EC_{50}$–48 h) for *Daphnia magna* to evaluate the toxic stress of heavy metals and organic contaminants at each sampling site. Accordingly, we were able to identify which evaluated environmental variables and traits were associated with the variation in species occurrence and abundances across the sampling sites.

Trait data of the benthic macroinvertebrate communities were also obtained from Rico et al. (2016). However, there were no trait data for ten macroinvertebrate taxa (i.e., *Hemimysis anomala*, *Katamysis warpachowsky*, *Limnomysis benedeni*, *Paramysis baukensis*, *Paramysis intermedia*, *Paramysis lacustris*, *Paramysis* sp., *Paramysis ullski*, *Schizorhamphus scabriusculus* and *Hydrachnidia* Gen. sp.). They were thus excluded from the present study. We divided the full trait dataset presented in Rico et al. (2016) into biological and ecological traits. Biological traits can provide a putative mechanistic link between chemical exposure and the sensitivity or recovery potential of benthic macroinvertebrate populations, and are less confounded by natural spatial gradients, contributing to reliably assess human impact as compared to ecological traits (Dolédec et al., 1999). The biological trait dataset (Datasets S3 and S4) was comprised of 11 traits divided into 63 categories, while the ecological trait dataset (Datasets S5 and S6) consisted of 11 traits classified into 55 categories. Biological traits describe life cycle features (maximal size, aquatic stages, life cycle duration, potential number of generations per year), resistance or resilience capacities (dispersal, substrate relation, resistance stages), general physiology (respiration), as well as reproduction and feeding behaviour (reproduction, feeding habits, food) (Usseglio-Polatera et al., 2000; Menezes et al., 2010). Ecological traits reflect habitat preferences, such as transversal distribution, longitudinal distribution, substrate, current velocity and trophic status preference (Usseglio-Polatera et al., 2000; Menezes et al., 2010). The original trait datasets (Datasets S3 and S5) were used for dc-CA, while average trait datasets (Datasets S4 and S6; see below for calculation) were used for CWM-RDA.

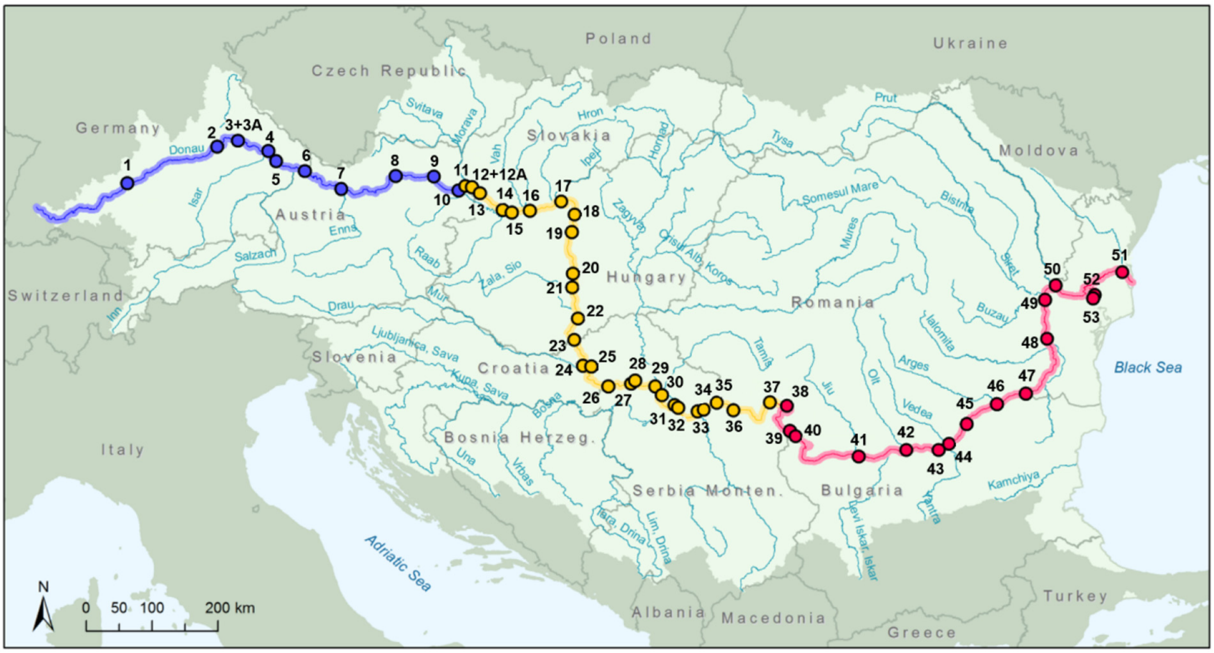


**Fig. 1.** Location of sampling sites in the Danube River. Blue, yellow and red circle symbols represent sampling sites in the upper, middle and lower stretches of the river, respectively. Sites 3A and 12A are located in the downstream of the dam of a power plant and downstream of the urban area of Bratislava, respectively. Taken with permission from Rico et al. (2016). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

### *2.3. Data analysis*

Because of the large number of environmental variables, we first determined which of the nine groups of variables were significantly associated with the species composition and then searched for trait-environment relations in each significant group separately. Because of the limited number of samples and the fact that human-impact trends are likely correlated with geographical upstream-downstream differences, we did not attempt to adjust the effects of one group for the effect of the others.

In detail, the procedure was as follows. We first applied CCA to each group with Monte Carlo (MC) permutation tests and *p*-values were adjusted by Holm's sequential Bonferroni method. However, the number of industrial chemicals ($n = 46$) and pharmaceuticals ($n = 134$) were large compared to the number of samples ($n = 55$), which would make the constrained ordination models effectively unconstrained and the permutation tests unreliable. We thus pre-screened environmental variables from these two groups for further statistical analyses; only variables that were individually significant after Holm correction were retained for further analysis (see Table S1 in Supporting Information A for adjusted *p*-values of individual variables). Subsequently, to search for trait-environment relationships, dc-CA analyses with forward selection of variables (see Text S1 in Supporting Information B for specific steps) were performed for each significant group of environmental variables in combination with biological traits and ecological traits, separately. During the forward selection procedure of dc-CA, *p*-values of environmental variables and traits were adjusted by false discovery rate for multiple testing at the 5% significance level. The relations of environmental variable-trait, environmental variable-species, environmental variable-site, trait-site, trait-species and species-site can be interpreted from corresponding biplots.

For comparison with dc-CA, CWM-RDA analyses were performed separately on aforementioned eight groups of environmental variables using average biological or ecological trait data (Datasets S4 and S6). CWM-RDA allowed for forward selection of environmental variables, but not of traits. Traits were selected post-hoc on the basis of their fit. This selection was solely to produce readable graphs.

In above analyses, benthic macroinvertebrate abundance data were quarter-root transformed to down-weigh high abundance values and obtain approximately a normal distribution of the data (Van den Brink et al., 2000). CWM trait values were calculated from standardized traits and quarter-root-transformed abundance data. In addition, environmental variables (except substrate PC1 and PC2) were $\log_{10}$-transformed prior to analysis to reduce the influence of outliers, with log10 of half the smallest non-zero values being imputed for values equal to zero. The RDA in CWM-RDA used only centring of the CWM values, but no further standardization. The reason for this is that the CWM values are already on a common scale due to the standardization of the traits before the CWM-calculation. Moreover, the standard deviation of the CWM values of a particular trait is measure of the importance of the trait, so that further standardization (by division by the standard deviation) in the RDA would remove possibly important information (Ter Braak et al., 2018a). The statistical significance was set as 5% after *p*-value correction and was determined with Monte Carlo simulation tests (number of permutations = 5000). All multivariate analyses were performed with the Canoco 5.12 software (Ter Braak and Šmilauer, 2018), with an Canoco project file with description and analyses as Supporting Information C. Additionally, an R script for dc-CA is provided in Supporting Information D.

## 3. Results

### *3.1. Double constrained correspondence analysis (dc-CA)*

#### *3.1.1. dc-CA based on the biological trait dataset*

Concerning the effects of the groups of environmental variables on species abundance, eight of the nine groups of environmental variables showed significance as judged by overall tests using CCA ($p = 0.018$ for each, after Holm-correction for nine tests); only the HPCPs group was non-significant. The group of pesticide variables was the one that better explained the abundance variation, as judged by the adjusted $R^2$ (26%), closely followed by the pharmaceuticals (23%) and physico-chemical environmental variables (20%) (Table 1). The biological traits had an adjusted $R^2$ of 9.7% ($p = 0.0002$). When the variables selected by dc-CA (Table 1) were used to explain abundance by a CCA, the adjusted $R^2$

**Table 1**

Environment and biological traits: summary of single and double constrained correspondence analyses performed on each group of environmental variables. All biological traits jointly explain 22% of the abundance variation (52 df, adjusted $R^2 = 9.7\%$, $p = 0.0002$).

| Dataset | Single constrained-all environmental variables | | | | Single constrained-selected[a] variables | | | | | | | | Double constrained-selected variables | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Environmental variables | | | | Traits | | | | | | | | |
| | df[b] | $R^2$ | adjusted-$R^2$ | *p*-value | df | $R^2$ | adjusted-$R^2$ | *p*-value | df | $R^2$ | adjusted-$R^2$ | *p*-value | df_Env | df_Traits | $R^2$ | adjusted-$R^2$ | *p*-value |
| Habitat characteristics | 13 | 36% | 16% | 0.0018 | 6 | 23% | 14% | 0.0018 | 8 | 8.2% | 6.2% | 0.0018 | 6 | 8 | 4.0% | 3.8% | 0.0018 |
| Hydromorphological alterations | 10 | 24% | 7.3% | 0.0018 | 2 | 7.9% | 4.4% | 0.0018 | 8 | 7.1% | 5.2% | 0.0018 | 2 | 8 | 1.2% | 1.2% | 0.0018 |
| Physico-chemical parameters | 19 | 48% | 20% | 0.0018 | 5 | 22% | 14% | 0.0018 | 14 | 11% | 7.6% | 0.0018 | 5 | 14 | 4.8% | 4.4% | 0.0018 |
| Heavy metals | 8 | 25% | 11% | 0.0018 | 2 | 12% | 8.4% | 0.0018 | 9 | 8.5% | 6.3% | 0.0018 | 2 | 9 | 2.5% | 2.5% | 0.0018 |
| Pesticides | 39 | 79% | 26% | 0.0018 | 2 | 15% | 12% | 0.0018 | 9 | 8.7% | 6.5% | 0.0018 | 2 | 9 | 3.1% | 3.0% | 0.0018 |
| Pharmaceuticals | 27[c] | 62% | 23% | 0.0018 | 5 | 23% | 15% | 0.0018 | 10 | 9.5% | 7.0% | 0.0018 | 5 | 10 | 4.4% | 4.2% | 0.0018 |
| Industrial chemicals | 8[d] | 24% | 11% | 0.0018 | 3 | 14% | 8.7% | 0.0018 | 9 | 8.2% | 6.0% | 0.0018 | 3 | 9 | 2.1% | 2.0% | 0.0018 |
| Miscellaneous category | 12 | 29% | 8.5% | 0.0018 | 1 | 7.2% | 5.4% | 0.0018 | 7 | 7.6% | 5.9% | 0.0018 | 1 | 7 | 1.9% | 1.9% | 0.0018 |

[a] Selected by dc-CA using forward selection; no more variables are added, if the false discovery rate of each variable considered in a step exceeds 5%.

[b] df, degrees of freedom; $R^2$, explained variation; df_Env, degrees of freedom of the environmental variables; df_Traits, degrees of freedom of trait variables; *p*-value, significance level after Holm correction for 9 groups of variables (the unadjusted *p*-value is 0.0002).

[c] Selected from 134 variables by prescreening using significance of each individual variable at 5% level after Holm correction for multiple testing.

[d] Selected from 46 variables by prescreening using significance of each individual variable at 5% level after Holm correction for multiple testing.

dropped somewhat, for example from 26% to 12% for pesticides, and from 9.7% to 6.5% for the biological traits, selected in conjunction with pesticides. Constraining the analysis by both the selected traits and selected environmental variables reduced the explained variation further, e.g. to an adjusted $R^2$ of 3.0% for pesticides (Table 1). The physico-chemical environmental variables were most strongly associated with the biological traits (4.4%), closely followed by the pharmaceuticals (4.2%) and habitat characteristics (3.8%). The associations were significant, even after Holm-correction ($p = 0.018$), despite the low percentage of variance that they explained.

Biological trait modalities significantly modulating the species-environment relationships and environmental variables significantly modulating the species-biological trait relationships are summarized in Tables S1 and S2, respectively. Seven to fourteen out of 63 biological trait modalities were significantly associated with each group of environmental variables, where.

(a) habitat characteristics were represented by substrate PC1, area (area of cross sections), river-km (distance to river mouth), distance to shore (distance of the macroinvertebrate sampling point to the closest river shore), surface velocity (mean river flow velocity measured during the macroinvertebrate sampling) and mean sample depth (mean water depth measured during the macroinvertebrate sampling),
(b) hydromorphological alterations by affected banks (banks affected by artificial materials) and substrate alteration (near natural substrate mix or altered),
(c) physico-chemical parameters by $SO_4^{2-}$, dissolved oxygen, nitrates, potassium total and ammonium,
(d) heavy metals by As and Cu,
(e) pesticides by atrazine-2-hydroxy and picoxystrobin,
(f) pharmaceuticals by triazolam, hydrochlorothiazide, 2-ethylidene-1,5-dimethyl-3,3-diphenylpyrrolidine (EDDP), lamotrigine and nordiazepam,
(g) industrial chemicals by tris(2-butoxyethyl)phosphate (TBEP), perfluorooctanoic acid (PFOA) and dibutyltin-cation, and.
(h) miscellaneous chemicals by sucralose.

The results of the community-based and species-based dimensionality tests showed that i) for habitat characteristics the first four constrained axes were significant, ii) for hydromorphological alterations, heavy metals, pesticides and industrial chemicals the first two constrained axes were significant, iii) for physico-chemical parameters and pharmaceuticals the first three constrained axes were significant, and iv) for miscellaneous chemical group only the first axis was significant ($p \leq 0.0222$; Table S3).

The trait-environment relations are displayed in Fig. 2. In these diagrams, both traits and environmental variables are represented by arrows which point in the direction of higher values and which together form a biplot of their fourth corner correlation. A few trait modalities were commonly associated with several groups of environmental variables. For instance, locomotion and substrate relations of flier, laying clutches in terrestrial habitats, and aerial active dispersal were positively correlated with surface velocity, river-km, dissolved oxygen, nitrates, picoxystrobin, hydrochlorothiazide, EDDP and TBEP. In contrast, burrowing locomotion, egg and/or adult aquatic life stages, and asexual reproduction were inversely correlated with the above environmental variables, but were positively correlated with substrate PC1, $SO_4^{2-}$, potassium total, heavy metals (As and Cu), atrazine-2-hydroxy (a pesticide), pharmaceuticals (lamotrigine and nordiazepam), PFOA (an industrial chemical), and/or sucralose. Fig. S2 shows the relations between sampling sites and environmental variables. Sampling sites lying close together have similar environmental variables, while those with very different environmental variables lie far apart. For all groups of environmental variables, except for hydromorphological alterations, sampling sites in the upper (blue circles) are apart from those in the middle (yellow circles) and lower (red circles) part of the river. Higher surface velocity and river-km occurred in the upper part, while higher content of fine sediments (substrate PC1), area, distance to shore, mean sample depth occurred in the middle and lower parts of the river (Fig. S2a). Regarding the parameters representing hydromorphological alterations, most sampling sites overlapped (Fig. S2b). As to the remaining six groups representing water quality measurements (Fig. S2c-h), sampling sites in the upper (blue circles) and lower (red circles) part of the river are respectively placed on the right and left side of the diagrams, showing differences in water quality between these sampling sites. Fig. 3 shows the top 10 species' for which the most variation is explained by environmental variables and traits in Fig. 2. The species in the hydromorphological alteration group were different from those in the remaining seven groups. In these seven groups, most of the 10 species occurred more in the samples from the upper part of the river; six insects (i.e., *Cricotopus (cricotopus) sp.*, *cricotopus (cricotopus) triannulatus*, *Tanytarsus* sp. “Traun”, *Microtendipes* cf. *britteni*, *Potthastia gaedii-Gr.* and *Baetis fuscatus*) appeared in more than half of diagrams ($n = 4–7$) and were more abundant in the upper than the middle and lower parts of the river; the species

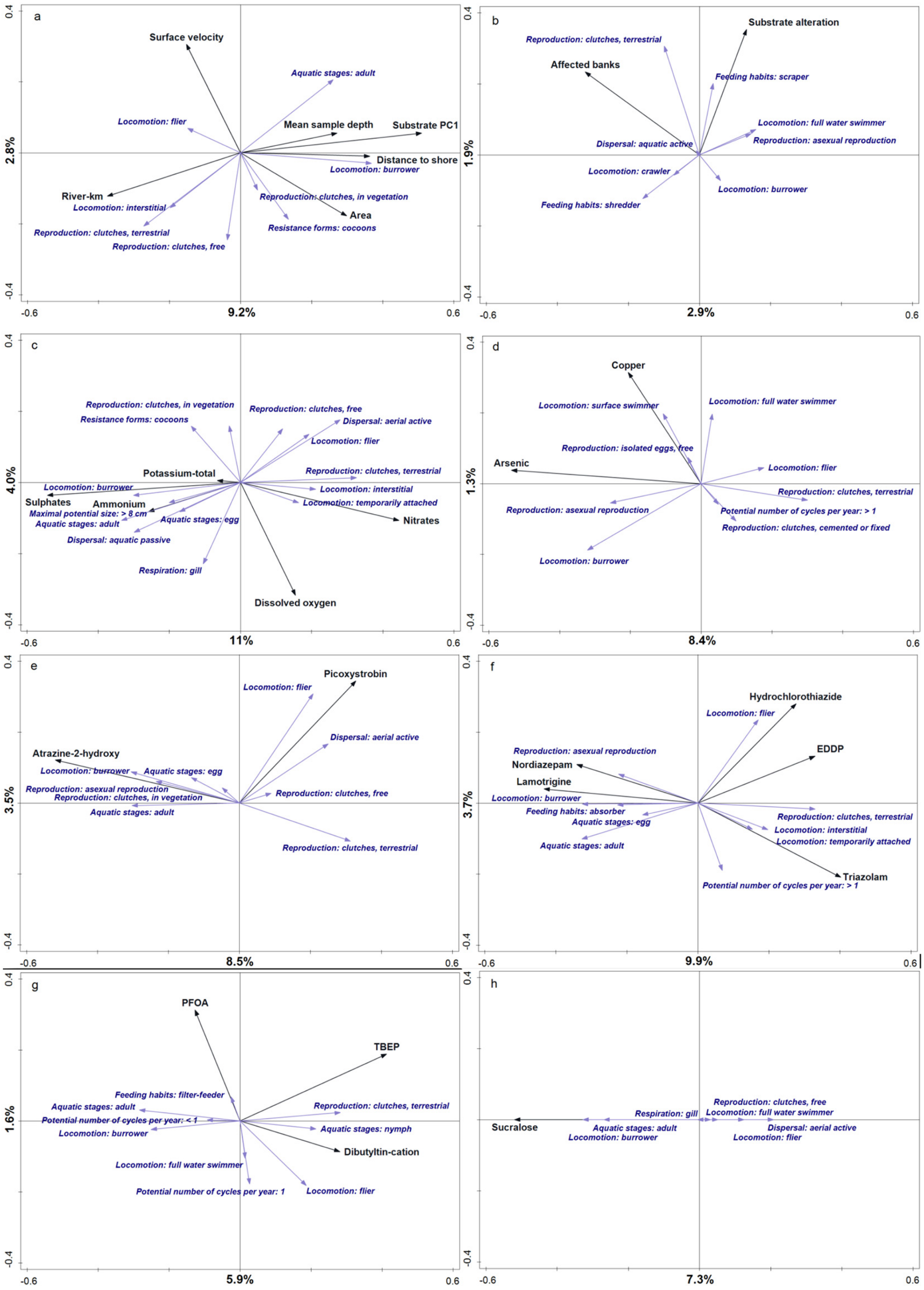


**Fig. 2.** Biplots of the associations (fourth-corner correlations) between the selected environmental variables (black arrows) and biological trait modalities (blue arrows) in the Danube River data set with 55 sampling sites and 383 invertebrate taxa, obtained from dc-CA on habitat characteristics (a), hydromorphological alterations (b), physico-chemical parameters (c), heavy metals (d), pesticides (e), pharmaceuticals (f), industrial chemicals (g) and miscellaneous chemicals (h), respectively. Values along the axes are their corresponding eigenvalues. EDDP, 2-ethylidene-1,5-dimethyl-3,3-diphenylpyrrolidine; PFOA, perfluorooctanoic acid; TBEP, Tris(2-butoxyethyl)phosphate. See text for interpretation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

more abundant in the middle and/or lower than the upper part were two snails (*Theodoxus danubialis* ssp. and *Lithoglyphus naticoides*), four insects (*Gomphus flavipes*, *Chironomus acutiventris*, *Polypedilum nubeculosum* and *Coenagrionidae* Gen. sp. juv.), one clam (*Dreissena polymorpha*) and one annelid worm (*Pristina aequiseta*) (Fig. 3a, c, d, h and Table S4).

### 3.1.2. dc-CA based on the ecological trait dataset

The environmental variables used for dc-CA on the ecological trait dataset were the same as the biological trait dataset (Table 2 versus 1). The ecological traits explained 8.2% (adjusted $R^2$) of the abundance variation ($p = 0.0002$), which is slightly lower than that explained by the biological traits (9.7%). Again, when the variables selected by dc-

CA (Table 2) were used to explain abundance by a CCA, the adjusted $R^2$ dropped, for example from 26% to 12% for pesticides, and from 8.2% to 6.7% for the ecological traits, selected in conjunction with pesticides. Constraining the analysis by both the selected traits and selected environmental variables reduced the explained variation further, e.g. to an adjusted $R^2$ of 3.1% for pesticides (Table 2). The habitat characteristics were most strongly associated with the ecological traits (4.4%), followed by the pharmaceuticals (4.1%), physico-chemical parameters (3.7%), pesticides (3.1%), heavy metals (2.4%), miscellaneous compounds (2.0%), industrial chemicals (1.6%) and hydromorphological alterations (1.2%). All associations were significant, even after Holm-correction ($p = 0.018$).

All the eight groups of environmental variables were significantly associated with ecological traits ($p_{(adj)} = 0.0018$; Table 2). For each group of environmental variables, three to eleven out of 55 ecological trait modalities significantly modulated the species-environment relationship ($p \leq 0.0460$; Table S1). The environmental variables significantly modulating the species-ecological trait relationships were similar to those for biological traits ($p \leq 0.0486$; Table S2). The percentage variation of the environmentally structured variation explained by traits (Table S1) was substantially lower than the reverse percentage, i.e. the percent variation of the trait structured variation explained by environmental variables (Table S2). The results of the community-based and species-based dimensionality tests showed that i) the first three constrained axes were significant concerning habitat characteristics, physico-chemical parameters, and pharmaceuticals ($p \leq 0.0036$; Table S3), ii) the first two constrained axes were significant concerning hydromorphological alterations and pesticides ($p \leq 0.0034$), and iii) only the first constrained axis was significant for heavy metals, industrial chemicals and miscellaneous chemicals ($p = 0.0002$).

The biplots relating the selected traits and environmental variables are shown in Fig. S3. Substrate preferendum for gravel, altitude of alpine level, and crenon, epirithron or metarithron as the longitudinal distribution type were positively related to several environmental variables that

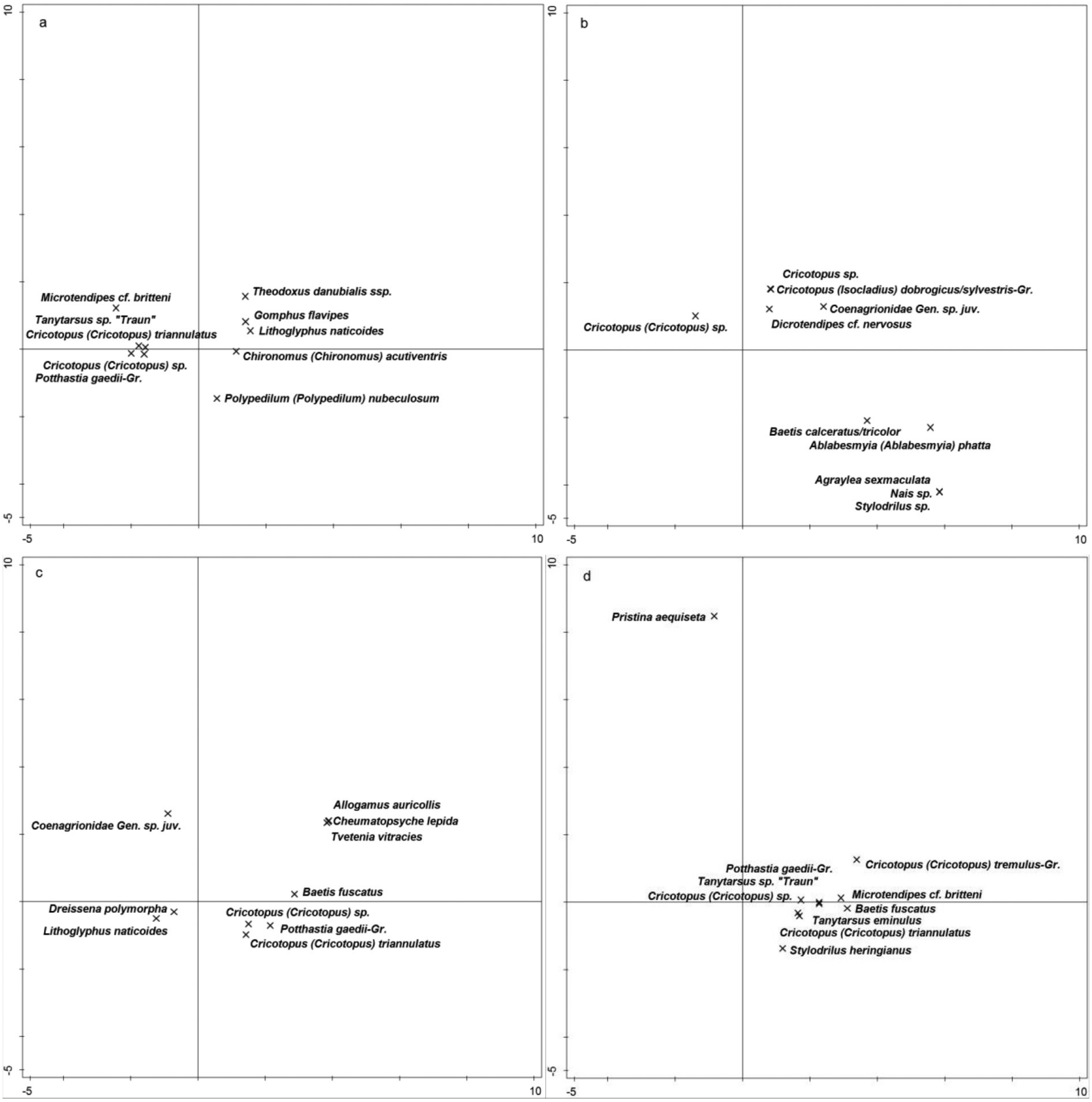


**Fig. 3.** Position of the species in the dc-CA biplots of Fig. 2. Shown are the 10 species, whose variation is best explained by the biological traits and environmental variables in Fig. 2. Positions are unconstrained (i.e. derived from species abundance, instead of being constrained positions, i.e. being a linear combination of the trait values) and were obtained from dc-CA on habitat characteristics (a), hydromorphological alterations (b), physico-chemical parameters (c), heavy metals (d), pesticides (e), pharmaceuticals (f), industrial chemicals (g) and miscellaneous chemicals (h), respectively. Species occurring only at one sampling site were excluded from the pesticide diagram (e). See text for interpretation.

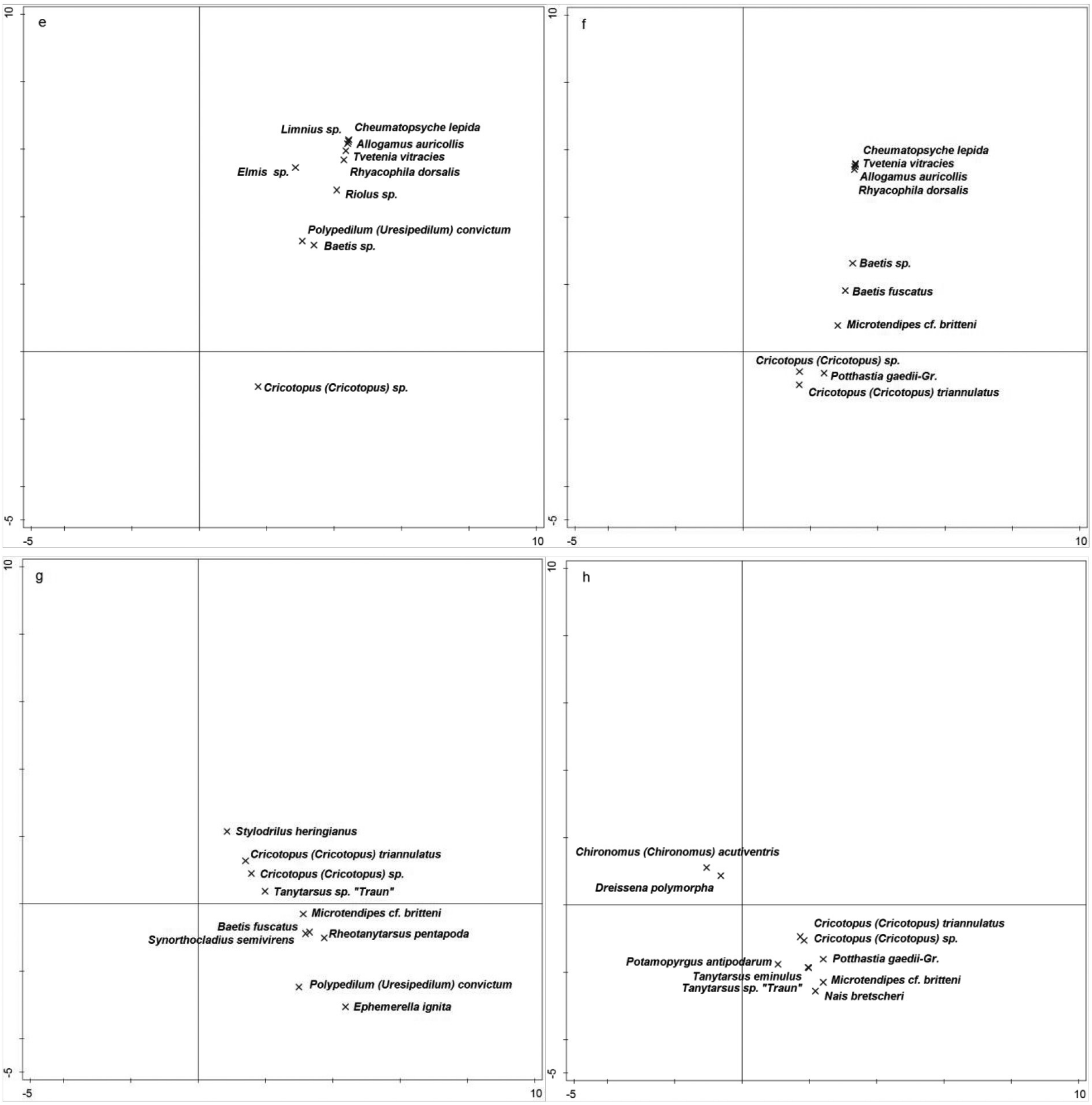


**Fig. 3** (*continued*).

had higher levels in the upper part of the river, including river-km, surface velocity, dissolved oxygen, nitrates, picoxystrobin, EDDP, hydrochlorothiazide, amisulpride, alpha-hydroxymidazolam, desvenlafaxine, 1H-benzotriazole, dibutyltin-cation and TBEP. On the contrary, substrate preferendum for mud and river channel as the transversal distribution type were positively related to several environmental variables that showed higher levels in the middle and lower parts of the river, including substrate PC1, area, distance to shore, $SO_4^{2-}$, potassium total, ammonium, As, Cu, atrazine-2-hydroxy, N-formyl-4-aminoantipyrine (FAA) and/or sucralose.

### 3.2. Redundancy analysis on community weighted means (CWM-RDA)

All the eight groups of environmental variables, except for hydromorphological alterations, significantly explained the variation in both biological and ecological trait means (Table S5). Hydromorphological alterations were only significantly associated with biological trait means. Each group of environmental variables were represented by one to three variables (Table S6), most of which also showed significant associations with traits in the dc-CA analyses (Table S2). For both datasets, habitat characteristics, physico-chemical parameters and pharmaceuticals explained the largest proportions of variance (adjusted estimates: 30–32% and 36–42% for biological and ecological traits, respectively), whereas hydromorphological alterations and industrial chemicals explained the least proportions of variance (adjusted estimates: 5.3–12% and 0–10% for biological and ecological traits, respectively). Additionally, heavy metals, pesticides, and miscellaneous chemicals explained comparable proportions of variance (adjusted estimates: 18–23% and 22–32% for biological and ecological traits, respectively; Table S5).

For the biological trait dataset, among the 10 trait modalities' for which the most variation is displayed in Fig. 4, i) laying free clutches or in terrestrial habitats, interstitial, life cycle duration ≤1, maximal potential size >0.25–0.5 cm, areal active/passive dispersal, and food preference for living microphytes were positively correlated with river-km, substrate alteration, EDDP, alpha-hydroxymidazolam and TBEP showing higher levels in the upper part, and ii) adult aquatic life stages,

**Table 2**
Environment and ecological traits: summary of single and double constrained correspondence analyses performed on each group of environmental variables. All ecological traits jointly explain 18.8% of the abundance variation (44 df, adjusted $R^2 = 8.2\%$, $p = 0.0002$).

| Datasets | Single constrained-all environmental variables | | | | Single constrained-selected[a] variables | | | | | | | | Double constrained-selected variables | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Environmental variables | | | | Traits | | | | | | | | |
| | df[b] | $R^2$ | adjusted-$R^2$ | *p*-value | df | $R^2$ | adjusted-$R^2$ | *p*-value | df | $R^2$ | adjusted-$R^2$ | *p*-value | df_Env | df_Traits | $R^2$ | adjusted-$R^2$ | *p*-value |
| Habitat characteristics | 13 | 36% | 16% | 0.0018 | 6 | 23% | 14% | 0.0018 | 11 | 9.4% | 6.7% | 0.0018 | 6 | 11 | 4.7% | 4.4% | 0.0018 |
| Hydromorphological alterations | 10 | 24% | 7.3% | 0.0018 | 2 | 7.6% | 4.1% | 0.0018 | 6 | 6.9% | 5.4% | 0.0018 | 2 | 6 | 1.2% | 1.2% | 0.0018 |
| Physico-chemical parameters | 19 | 48% | 20% | 0.0018 | 5 | 22% | 14% | 0.0018 | 8 | 8.1% | 6.1% | 0.0018 | 5 | 8 | 3.9% | 3.7% | 0.0018 |
| Heavy metals | 8 | 25% | 11% | 0.0018 | 2 | 12% | 8.4% | 0.0018 | 5 | 6.6% | 5.4% | 0.0018 | 2 | 5 | 2.4% | 2.4% | 0.0018 |
| Pesticides | 39 | 79% | 26% | 0.0018 | 2 | 15% | 12% | 0.0018 | 10 | 9.2% | 6.7% | 0.0018 | 2 | 10 | 3.2% | 3.1% | 0.0018 |
| Pharmaceuticals | 27[c] | 62% | 23% | 0.0018 | 6 | 24% | 15% | 0.0018 | 11 | 9.6% | 6.9% | 0.0018 | 6 | 11 | 4.5% | 4.1% | 0.0018 |
| Industrial chemicals | 8[d] | 24% | 11% | 0.0018 | 3 | 13% | 7.6% | 0.0018 | 3 | 4.9% | 4.2% | 0.0018 | 3 | 3 | 1.7% | 1.6% | 0.0018 |
| Miscellaneous category | 12 | 29% | 8.5% | 0.0018 | 1 | 7.2% | 5.4% | 0.0018 | 5 | 6.4% | 5.1% | 0.0018 | 1 | 5 | 2.1% | 2.0% | 0.0018 |

[a] Selected by dc-CA using forward selection; no more variables are added, if the false discovery rate of each variable considered in a step exceeds 5%.
[b] df, degrees of freedom; $R^2$, explained variation; df_Env, degrees of freedom of the environmental variables; df_Traits, degrees of freedom of trait variables; *p*-value, significance level after Holm correction for 9 groups of variables (the unadjusted *p*-value is 0.0002).
[c] Selected from 134 variables by prescreening using significance of each individual variable at 5% level after Holm correction for multiple testing.
[d] Selected from 46 variables by prescreening using significance of each individual variable at 5% level after Holm correction for multiple testing.

aquatic passive dispersal, life cycle duration >1 and burrower were positively associated with environmental variables with higher levels being found in the middle and lower parts of the river, including substrate PC1, $SO_4^{2-}$, chlorides, suspended solids, As, Cu, atrazine-2-hydroxy, FAA, PFOA and sucralose (Figs. 4 and S4).

For the ecological trait dataset, among the 10 trait modalities' for which the most variation is displayed in Fig. S5, i) altitude of alpine or piedmont level, epirithron or crenon as longitudinal distribution type, substrate preferendum for gravel and/or microphytes, trophic status preferendum for oligotrophic, pH preferendum for >4–5 and current velocity preferendum for fast (> 50 cm/s) were positively correlated with river-km, EDDP and TBEP showing higher levels in the upper part, and ii) altitude of lowlands, epipotamon or metarithron as longitudinal distribution type, substrate preferendum for mud, trophic status preferendum for eutrophic and pH preferendum for >6 were positively correlated with substrate PC1, $SO_4^{2-}$, As, atrazine-2-hydroxy, FAA, sucralose, cyclamate and/or genistein showing higher levels in the middle and lower parts of the river.

## 4. Discussion

Here we examined trait-environment relationships of the macroinvertebrate community in the Danube River by dc-CA and compared results from dc-CA with those from CWM-RDA. Overall, results from both dc-CA and CWM-RDA showed that there were significant relationships between biological/ecological traits and all the evaluated groups of environmental variables except for HPCPs, with the strongest association being found with physico-chemical parameters, habitat characteristics and pharmaceuticals and the weakest with hydromorphological alterations (Tables 1, 2 and S5). These results suggest that human impact on macroinvertebrate communities is of similar magnitude to the natural impact. An important benefit of dc-CA over CWM-RDA is that it contains a species-level analysis by which we ruled out that the relationships found are an artefact of strong environmental impacts.[1] Results from dc-CA also showed that the variation in macroinvertebrate communities could be better explained by the available environmental variables than by the available traits, which cannot be assessed using CWM-RDA. To obtain more explanatory power by traits, it might be worthwhile to establish in future research traits that can better represent macroinvertebrate responses to anthropogenic stress (Van den Berg et al., 2019).

All the evaluated groups of environmental variables except HPCPs significantly influenced the macroinvertebrate community composition in the Danube River. The lack of associations between HPCPs and species abundance data could be attributed to their small variations across sampling sites and/or their negligible effect on macroinvertebrates at the concentration levels detected in the present study. For instance, both benzophenone-3 and chlorophene were only detected at one sampling site in the lower part of the river (Dataset S2). Concerning the remaining environmental variables, several studies have found a significant correlation between macroinvertebrate composition and habitat characteristics, hydromorphological alterations, physico-chemical parameters, heavy metals, pesticides, pharmaceuticals and/or industrial organic chemicals in European rivers (e.g., Dolédec and Statzner, 2008; De Castro-Català et al., 2015; Kiesel et al., 2015; Kuzmanović et al., 2016; Doretto et al., 2018). For instance, Kiesel et al. (2015) reported that substrate composition, substrate stability and current velocity influenced occurrences and abundances of benthic invertebrates in streams. In four Mediterranean basins examining the relationships between emerging pollutants and invertebrate community, De Castro-Català et al. (2015) found that pharmaceuticals significantly explained the invertebrate community composition in the Llobregat River and fungicides primarily determined the structure of the invertebrate community in the Júcar River (Iberian Peninsula).

Percentages of explained variation were generally low, and are logically lower in double constrained ordination compared to single constrained or unconstrained ordination, due to the extra constraints (Tables 1 and 2). These low percentages reflect the noisiness of macroinvertebrate community data. Nevertheless, the patterns found in the noisy data contain important information on the human and natural impacts and what shifts in trait values these impacts imply. It is worth noting that RLQ would have led to even lower fits. As an example we applied RLQ to similarly transformed abundance data and the selected physico-chemical parameters and the biological traits (Table S3 and Supporting Information D). The first two RLQ eigenvalues were 0.39 and 0.03, compared to 0.11 and 0.04 in dc-CA. However, the eigenvalues in RLQ are in terms of coinertia, which in this case is equal to the sum of pair-wise fitted squared fourth-corner correlations, whereas the eigenvalues in dc-CA are in terms of inertia of the abundance table analysed. To put the RLQ and dc-CA axes on equal footing we calculated the inertia fitted by the first RLQ axis. This inertia was only half that of the dc-CA first axis (Supporting Information D). The mathematical explanation is that dc-CA maximizes this fitted inertia, whereas RLQ is a compromise between the axes of three unconstrained analyses, namely the

---

[1] A blog worth reading on this is https://davidzeleny.net/blog/2019/01/20/five-years-with-community-weighted-mean/.

unconstrained analysis of the abundance, environment and trait tables (Dray et al., 2014). Whereas dc-CA allows comparison between double, single and unconstrained ordination (as all eigenvalues are in terms of inertia of the abundance table), RLQ compares the three-table analysis with unconstrained ordinations only (namely with the CA of the abundance table, the PCAs, or similar methods, of the environmental and trait tables). The focus in dc-CA is thus clearly on explaining the community assemblage by the traits and environmental variables, whereas in RLQ the focus is diffuse as it also aims to fit the trait and environment data tables. Because the eigenvalues in dc-CA are in terms of inertia of the abundance table analysed, the values in Tables 1, 2 and S3 can be validly compared among the groups of environmental variables, so that we are able to rank the groups in terms of explanatory power. It would not be straightforward to construct a similar table on the basis of RLQ.

The percentage variation explained by CWM-RDA is in our experience generally of the same order of magnitude as that of step 4 of the dc-CA algorithm (compare for example Tables S2 and S6 and Tables S5 and S7). Note, however, that these percentages are not strictly comparable as dc-CA in step 4 calculates CWMs with respect to orthonormalized traits instead of just standardized traits in CWM-RDA. In contrast with RLQ and CWM-RDA, dc-CA allows to determine how much of the environmentally structured variation can be explained by traits and we found on the basis this that the traits are currently the limiting factor in getting higher explanatory power.

The results of the dc-CA showed that macroinvertebrates (e.g., Oligochaeta: *Tubificidae* Gen. sp.) using asexual reproduction, egg and/or adult aquatic life stages and aquatic passive dispersal tactics, burrowing, and preferring to live in river channels characterized by muddy substrate were more abundant in the middle and lower parts

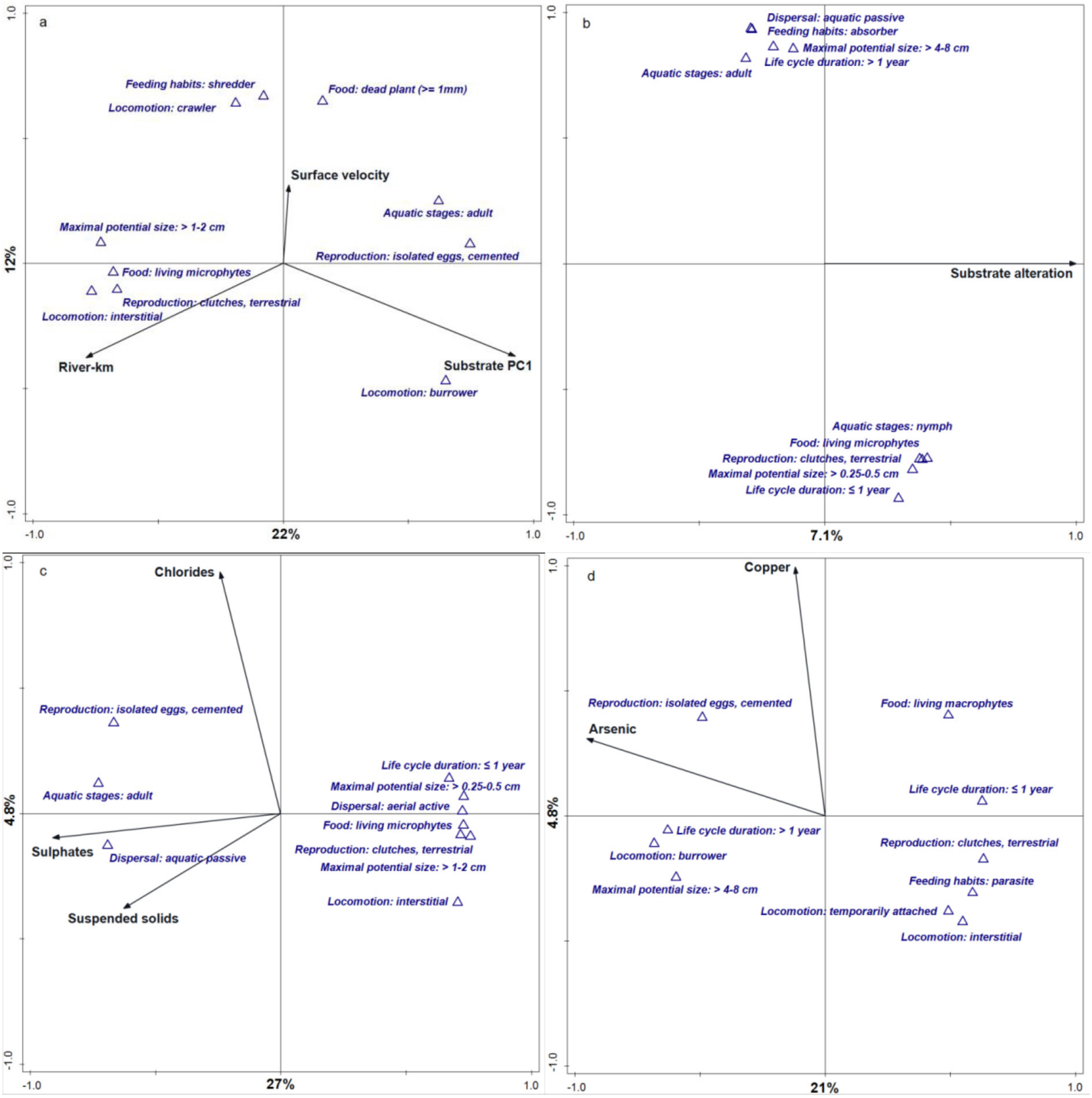


**Fig. 4.** Biplots of CWM-RDA relating the significant environmental variables (black arrows) and biological trait means (blue triangles) in the Danube River data set with 55 sampling sites and 383 invertebrate taxa, obtained from CWM-RDA on habitat characteristics (a), hydromorphological alterations (b), physico-chemical parameters (c), heavy metals (d), pesticides (e), pharmaceuticals (f), industrial chemicals (g) and miscellaneous chemicals (h), respectively. In all the eight graphs, only the 10 traits for which the most variation is explained in the diagram are shown. Values along the axes are their corresponding eigenvalues. FAA, N-formyl-4-aminoantipyrine; EDDP, 2-ethylidene-1,5-dimethyl-3,3-diphenylpyrrolidine; PFOA, perfluorooctanoic acid; TBEP, Tris(2-butoxyethyl)phosphate. See text for interpretation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

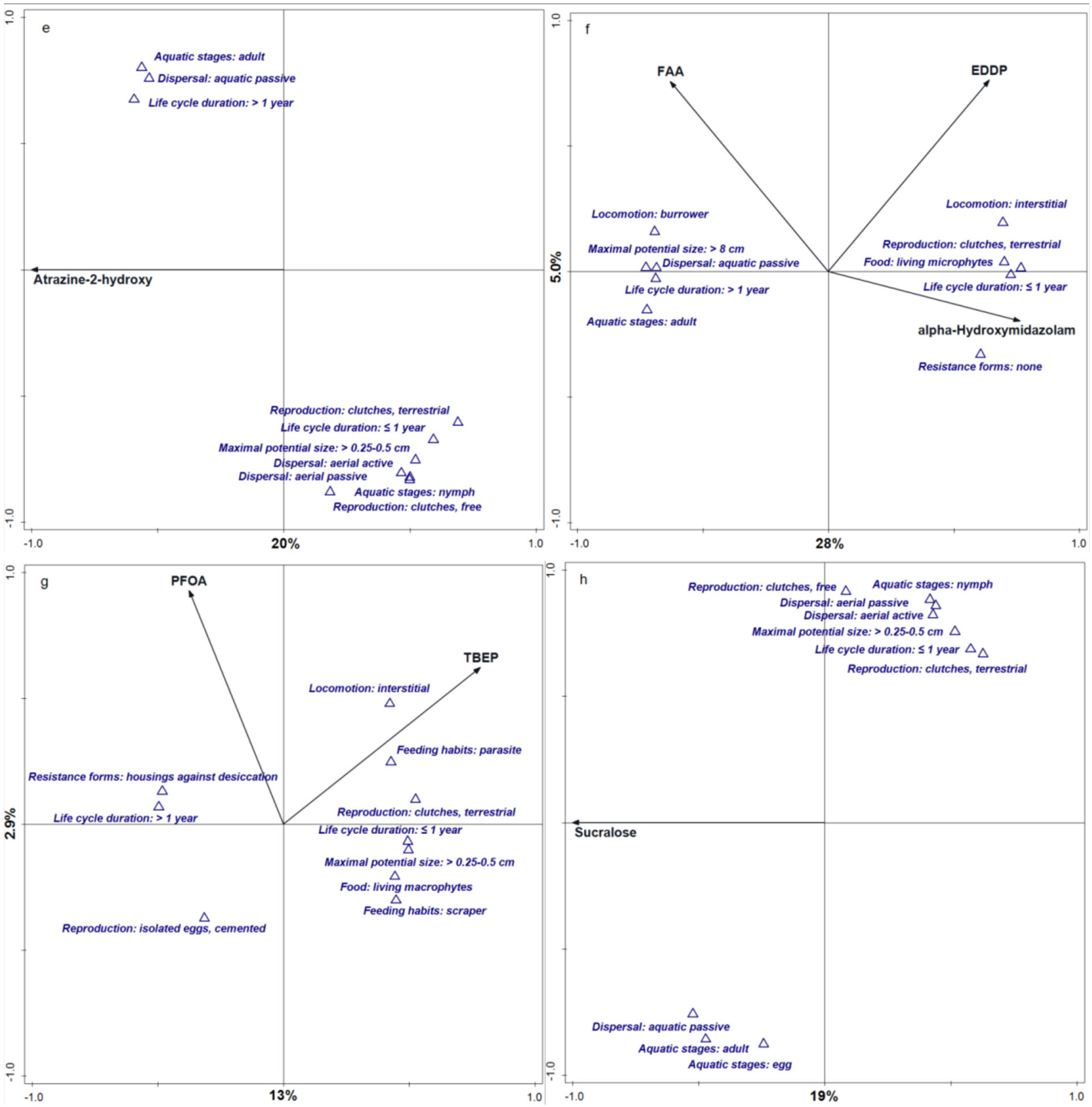


**Fig. 4** (*continued*).

of the Danube River where higher content of fine sediments and higher concentrations of $SO_4^{2-}$, potassium, heavy metals (e.g., As and Cu), pesticides (e.g., atrazine-2-hydroxy), pharmaceuticals (e.g., FAA, lamotrigine and nordiazepam), industrial chemicals (e.g., PFOA) and sucralose occurred compared to the upper part (Figs. 2, S2 and S3). This may suggest that different trait combinations are specific to different habitats in running waters (e.g., Dolédec et al., 1999; Statzner and Beche, 2010; Mondy et al., 2016; Berger et al., 2018). Similar results have been reported for macroinvertebrates from 422 monitoring sites in Germany, where adult aquatic life stages and substrate preferendum for mud were found to be consistently associated with taxa tolerant to the evaluated 20 water quality stressors including basic water quality parameters (e.g., ammonium, conductivity and sulphate) and organic micropollutants (e.g., N,N-diethyl-m-toluamide and caffeine) (Berger et al., 2018). Also, it has been reported that biological traits describing reproductive techniques had high potential to discriminate the level of multiple human impacts on invertebrate communities in 190 large river reaches in Europe (Gayraud et al., 2003). In contrast, macroinvertebrates (e.g., Insecta: *Cricotopus (cricotopus) sp.) flying* or temporarily attaching to substrate, using aerial active dispersal tactics, laying clutches in terrestrial habitats, and preferring to live in habitats characterized by gravelly substrate, altitude of alpine level and crenon/epirithron/metarithron were more abundant in the upper part showing larger distance to the river mouth, higher surface velocity and higher levels of dissolved oxygen, nitrates, pesticides (e.g., picoxystrobin), pharmaceuticals (e.g., EDDP) and industrial chemicals (e.g., TBEP) as compared to the middle and lower parts of the river (Figs. 2, 3, S2 and S3). This combination of attributes are likely to reflect the adaptation and/or recovery of organisms to stress caused by pesticides, pharmaceuticals and/or industrial chemicals. Indeed, it has been reported that locomotion and substrate relations and dispersal reflect resistance and resilience capacities of European benthic invertebrate genera (Dolédec et al., 2017). From the upper to the middle and lower parts of the river, there was a substrate preferendum gradient from gravel to mud (Figs. S2 and S3), which perfectly corresponded to the changes in substrate composition along the river (Fig. S1), suggesting an effect of habitat utilization on macroinvertebrates. A similar distribution pattern of substrate preferendum has been reported for benthic macroinvertebrate taxa identified in French freshwaters (Usseglio-Polatera et al., 2000). Moreover, in the upper part of the river the longitudinal distribution was dominated by crenon, epirithron, and/or metarithron and the altitude by alpine level (Figs. S2 and S3). Taken together, these findings clearly indicate the crucial roles of substrate, longitudinal and altitude gradient in the composition of instream invertebrate communities (Grubaugh et al., 1996; Kiesel et al., 2015; Van Looy et al., 2017; Doretto et al., 2018). Overall, our results suggest that both biological and ecological traits reflected adaptations of macroinvertebrates to dominant environmental characteristics and/or stresses of habitats. Unlike the present study, a previous study found that ecological traits of macroinvertebrate communities poorly indicated human impact on the French Rhône River (Dolédec et al., 1999). This might be attributed to the larger-scale environmental conditions in the Danube River relative to Rhône River, which would result in larger anthropogenic and/or natural environmental gradients, such as pollutants, flow velocity, substrate and geographic location. Therefore, at large scales the ecological traits are likely to modulate the species-environment relations better.

Environmental variables had a stronger influence on species abundance compared to traits. Unlike the present study, a recent study reported that environmental variables (i.e., Cd and phosphorus) and the trait (reproduction: asexual reproduction) explained a comparable amount of the variance (53% and 55%) in the macroinvertebrate assemblages collected from six rivers in Guangzhou City (South China) (Peng et al., 2020). This difference may be related to the difference in the examined environmental variables (e.g., measured in the water versus sediment samples), which would affect their relations with the traits and/or species composition. Additionally, the larger number of taxa and sampling sites in the current study than that study could have lowered the trait-environment relationship. Overall, the results of dc-CA suggest that life cycle attributes (e.g., aquatic stages), dispersal tactics, locomotion and substrate relation, and reproduction techniques can serve as multiple probes for different human impact and/or habitat characteristics.

Environmental variables significantly associated with biological or ecological traits in the CWM-RDA analyses were generally consistent with those found by dc-CA analyses (Tables S2 and S6). However, in the CWM-RDA analyses the number of significant environmental variables was somewhat lower than those identified by the dc-CA analyses, except for the analyses on the miscellaneous chemicals group for ecological traits. For instance, concerning physico-chemical parameters only $SO_4^{2-}$ and water temperature were significantly associated with the variance in ecological trait means (Table S6), whereas the dimensionality tests showed that there were three significant gradients in the dc-CA on ecological traits (Table S3). In fact, there are a few drawbacks on using CWM-RDA to examine trait-environment relationships. For instance, it does not have a principled way to select traits and it neglects any existing correlations among traits and only does a sample-level analysis (Peres-Neto et al., 2017). In contrast, double constrained ordination accounts for such correlations, allows for trait selection, and does both sample-level and species-level analysis (Ter Braak et al., 2018a, 2018b). These differences are likely to be responsible for the different trait-environment relationships detected by the CWM-RDA and dc-CA. For instance, we found significant correlations between several ecological traits and dams/substrate alteration in the dc-CA but not in the CWM-RDA (Figs. S3b and S5). Additionally, in the dc-CA we can rank traits and environmental variables according to their relative importance for driving macroinvertebrate community assembly (Tables S1 and S2). However, there is a weakness in dc-CA compared to CWM-RDA. If the underlying correspondence analysis detects outliers, e.g. sites with species that are unique to these sites, then dc-CA can be heavily influenced by these outliers. This influence becomes reality when the constraining environmental variables and traits are able to distinguish such sites and species from the other sites and species. An example is in the Danube data set the most upstream site, which contains so many species that are unique to this site, that it becomes an outlier on the second axis of correspondence analysis. But this site had high picoxystrobin level and there is a combination of trait modalities that largely identifies the species that are unique to this site. Therefore, this site becomes an outlier in dc-CA as well (Fig. S2e). We kept the site and picoxystrobin in the analysis, as picoxystrobin (and the site) also contributes importantly to the first axis, which represents a dominant gradient in pesticides.

We now compare results of the CWM-RDA in the present study with those in Rico et al. (2016), where chemical concentrations were replaced by TUs for CWM-RDA analysis to explore relations between chemicals and trait means. Our results indicate that all groups of chemicals, except for HPCPs, correlated significantly with both biological and ecological trait means (Table S5), whereas in Rico et al. (2016) only pharmaceutical and miscellaneous contaminant group showed significant correlations with biological trait means or with biological in combination with ecological trait means. Additionally, usage of $TU_{\text{pharmaceuticals}}$ implicitly assumes that all pharmaceuticals are related in the same type of way to traits, whereas the results of dc-CA suggest that pharmaceuticals differ in the way they correlate with traits as the arrows for different pharmaceuticals in Figs. 2f and S3f point in different directions. Therefore, replacing chemical concentrations by TUs has drawbacks for such data analysis.

Wagner (2004) shows how to apply geostatistical methods in the context of CA and CCA leading to multi-scale ordination by which one can determine whether species-environment relations are dependent on spatial scale. If strong, such scale dependence makes the overall 'scale-ignorant' effect less meaningful, analogously to the case in regression analysis of interpreting main effects in the presence of interaction. The author also developed a diagnostic tool to determine whether there is spatial autocorrelation in the residuals. With spatial correlation, statistical tests become too liberal. Braga et al. (2018) solve this issue by proposing a statistical significance test of the fourth-corner correlation that adjusts for spatial and phylogenetic correlation. The test is based on correlation-preserving randomization. Whether or not to adjust is not only a statistical question but also a matter of which question one wants to answer precisely. The practical statistical consequence of adjustment for spatial and phylogenetic correlation is that the statistical test implicitly uses effectively fewer samples and species compared to the unadjusted analysis so that the resulting test becomes more conservative and has less power. The ideas from these papers could also be applied in the context of dc-CA.

Pavoine et al. (2011) propose an integrated, RLQ-based method to relate phylogeny and traits to space and environment with five RLQ-based significance tests; a global test for the joint relationship (phylogeny & traits ↔ space & environment) and four tests on the individual components, namely traits ↔ environment, phylogeny ↔ environment, traits ↔ space and phylogeny ↔ space. In their approach, the data on phylogeny and space are each converted to rectangular data tables (if they are not so already) using principal coordinate analysis. For the joint analysis, these matrices must be brought to a comparable scale for which the user must select a scaling method. The above mentioned analyses and statistical tests can also be performed using dc-CA. In our view, this would be advantageous for two main reasons: (1) dc-CA is scale-invariant, so that the user does not need to select a scaling method and (2) dc-CA allows variation partitioning (Peres-Neto et al., 2006; Rico et al., 2016) whereas RLQ does not, because dc-CA is regression-based while RLQ is covariance-based. Variation partitioning allows one to determine which part of the variation in the abundance data is uniquely attributable to the environmental variables, which part uniquely attributable to space and which part is shared, because of the spatial structure in the environmental variables. With dc-CA, these parts concern the trait-structured variation. For example, analogously to standard variation partitioning (Peres-Neto et al., 2006), the part that is uniquely attributable to the environment is obtained by subtracting the explained variance of the dc-CA analysis using environment and traits-with-phylogeny from the explained variance of the analysis using environment-with-space and, again, traits-with-phylogeny. Such a subtraction makes mathematical sense in dc-CA as its explained variance is in terms of inertia of the abundance table, whereas in RLQ such analyses can be performed, and are so in Pavoine et al. (2011), but a similar comparison of explained variances cannot be made (and Pavoine et al. (2011) did not do so), as RLQ does not use the same scale to measure variance in the different analyses. Using partial dc-CA (in Canoco, obtained by specifying the environmental variables as explanatory data and the space variables as covariate data), the statistical significance of the uniquely environmental part can be tested. In this paper, we chose not to try to separate the spatial component from the environmental one for the reason that there is hardly a statistical way to separate these components in a river data set in which most variables are highly correlated with distance from the source. Our point of view is that the large-scale spatial variation can largely be explained by changes in the environment and that our sampling points are generally far enough apart to rule out small-scale spatial correlation. We acknowledge that the relations that we discovered require validation in later studies.

In summary, in this large-scale study we found that aquatic life stages, dispersal tactics, reproduction techniques, locomotion and substrate relations, transversal and longitudinal distribution, altitude, and substrate preferendum were significantly associated with all the evaluated nine groups of environmental variables except HPCPs. These environmental variables can thus function as filters selecting those traits that allow species to adapt to or recover from stress (Dolédec and Statzner, 2008; Kuzmanović et al., 2017). These results suggest that trait composition of macroinvertebrate assemblages can reflect the strategies that assemblages used to cope with environmental stressors. These results also demonstrate that double constrained ordination is a promising, powerful approach to monitor specific effects of multiple stressors on macroinvertebrate communities in rivers and streams, and helps to improve our diagnostic ability on stressor-specific alterations.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2020.142171.

## CRediT authorship contribution statement

**Feng-Jiao Peng:** Formal analysis, Writing - original draft. **Cajo J.F. ter Braak:** Conceptualization, Methodology, Formal analysis, Writing - original draft. **Andreu Rico:** Investigation, Writing - review & editing. **Paul J. Van den Brink:** Conceptualization, Funding acquisition, Supervision, Writing - review & editing.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

A. Rico is supported by a postdoctoral grant provided by the Spanish Ministry of Science, Innovation and University (IJCI-2017-33465). The data used in this study were collected during the Joint Danube Survey 3, which was organized by the International Commission for the Protection of the Danube River (ICPDR).

## References

Baird, D.J., Van den Brink, P.J., 2007. Using biological traits to predict species sensitivity to toxic substances. Ecotoxicol. Environ. Saf. 67, 296–301.

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? Sci. Total Environ. 619, 144–154.

Braga, J., ter Braak, C.J.F., Thuiller, W., Dray, S., 2018. Integrating spatial and phylogenetic information in the fourth-corner analysis to test trait-environment relationships. Ecology 99, 2667–2674.

Covich, A.P., Palmer, M.A., Crowl, T.A., 1999. The role of benthic invertebrate species in freshwater ecosystems: zoobenthic species influence energy flows and nutrient cycling. BioScience 49, 119–127.

Culp, J.M., Hose, G.C., Armanini, D.G., Dunbar, M.J., Orlofske, J.M., Poff, N.L., Pollard, A.I., Yates, A.G., 2011. Incorporating traits in aquatic biomonitoring to enhance causal diagnosis and prediction. Integr. Environ. Assess. Manag. 7, 187–197.

De Castro-Català, N., Muñoz, I., Armendáriz, L., Campos, B., Barceló, D., López-Doval, J., et al., 2015. Invertebrate community responses to emerging water pollutants in Iberian river basins. Sci. Total Environ. 503, 142–150.

Díaz, S., Cabido, M., 1997. Plant functional types and ecosystem function in relation to global change. J. Veg. Sci. 8, 463–474.

Dolédec, S., Statzner, B., 2008. Invertebrate traits for the biomonitoring of large European rivers: an assessment of specific types of human impact. Freshw. Biol. 53, 617–634.

Dolédec, S., Chessel, D., ter Braak, C.J.F., Champely, S., 1996. Matching species traits to environmental variables: a new three-table ordination method. Environ. Ecol. Stat. 3, 143–166.

Dolédec, S., Statzner, B., Bournard, M., 1999. Species traits for future biomonitoring across ecoregions: patterns along a human-impacted river. Freshw. Biol. 42, 737–758.

Dolédec, S., Tilbian, J., Bonada, N., 2017. Temporal variability in taxonomic and trait compositions of invertebrate assemblages in two climatic regions with contrasting flow regimes. Sci. Total Environ. 599, 1912–1921.

Doretto, A., Piano, E., Bona, F., Fenoglio, S., 2018. How to assess the impact of fine sediments on the macroinvertebrate communities of alpine streams? A selection of the best metrics. Ecol. Indic. 84, 60–69.

Dray, S., Choler, P., Doledec, S., Peres-Neto, P.R., Thuiller, W., Pavoine, S., et al., 2014. Combining the fourth-corner and the RLQ methods for assessing trait responses to environmental variation. Ecology 95, 14–21.

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 ofwater policy. Off. J. Eur. Communities 2000.

Gayraud, S., Statzner, B., Bady, P., Haybachp, A., Schöll, F., Usseglio-Polatera, P., et al., 2003. Invertebrate traits for the biomonitoring of large European rivers: an initial assessment of alternative metrics. Freshw. Biol. 48, 2045–2064.

Grubaugh, J.W., Wallace, J.B., Houston, E.S., 1996. Longitudinal changes of macroinvertebrate communities along an Appalachian stream continuum. Can. J. Fish. Aquat. Sci. 53, 896–909.

Keddy, P.A., 1992. Assembly and response rules: two goals for predictive community ecology. J. Veg. Sci. 3, 157–164.

Kiesel, J., Schröder, M., Hering, D., Schmalz, B., Hörmann, G., Jähnig, S.C., et al., 2015. A new model linking macroinvertebrate assemblages to habitat composition in rivers: development, sensitivity and univariate application. Fundamental and Applied Limnology/Archiv für Hydrobiologie 186, 117–133.

Kleyer, M., Dray, S., Bello, F., Lepš, J., Pakeman, R.J., Strauss, B., et al., 2012. Assessing species and community functional responses to environmental gradients: which multivariate methods? J. Veg. Sci. 23, 805–821.

Kuzmanović, M., López-Doval, J.C., De Castro-Català, N., Guasch, H., Petrović, M., Muñoz, I., et al., 2016. Ecotoxicological risk assessment of chemical pollution in four Iberian river basins and its relationship with the aquatic macroinvertebrate community status. Sci. Total Environ. 540, 324–333.

Kuzmanović, M., Dolédec, S., de Castro-Catala, N., Ginebreda, A., Sabater, S., Muñoz, I., et al., 2017. Environmental stressors as a driver of the trait composition of benthic macroinvertebrate assemblages in polluted Iberian rivers. Environ. Res. 156, 485–493.

Lavorel, S., Grigulis, K., McIntyre, S., Williams, N.S., Garden, D., Dorrough, J., et al., 2008. Assessing functional diversity in the field–methodology matters. Funct. Ecol. 22, 134–147.

Lemm, J.U., Feld, C.K., 2017. Identification and interaction of multiple stressors in central European lowland rivers. Sci. Total Environ. 603, 148–154.

Liška, I., Wagner, F., Sengl, M., Deutsch, K., Slobodník, J., 2015. Joint Danube survey 3-a comprehensive analysis of Danube water quality. Final Scientific Report, International Commission for the Protection of the Danube River, Vienna, Austria, p. 335.

Loos, R., Tavazzi, S., Mariani, G., Suurkuusk, G., Paracchini, B., Umlauf, G., 2017. Analysis of emerging organic contaminants in water, fish and suspended particulate matter (SPM) in the Joint Danube Survey using solid-phase extraction followed by UHPLC-MS-MS and GC-MS analysis. Sci. Total Environ. 607, 1201–1212.

Menezes, S., Baird, D.J., Soares, A.M., 2010. Beyond taxonomy: a review of macroinvertebrate trait-based community descriptors as tools for freshwater biomonitoring. J. Appl. Ecol. 47, 711–719.

Mondy, C.P., Muñoz, I., Dolédec, S., 2016. Life-history strategies constrain invertebrate community tolerance to multiple stressors: a case study in the Ebro basin. Sci. Total Environ. 572, 196–206.

Pavoine, S., Vela, E., Gachet, S., de Bélair, G., Bonsall, M.B., 2011. Linking patterns in phylogeny, traits, abiotic variables and space: a novel approach to linking environmental filtering and plant community assembly. J. Ecol. 99, 165–175.

Peng, F.J., Pan, C.G., Zhang, N.S., ter Braak, C.J.F., Salvito, D., Selck, H., Ying, G.G., Van den Brink, P.J., 2020. Benthic invertebrate and microbial biodiversity in sub-tropical urban rivers: correlations with environmental variables and emerging chemicals. Sci. Total Environ. 709, 136281.

Peres-Neto, P.R., Legendre, P., Dray, S., Borcard, D., 2006. Variation partitioning of species data matrices: estimation and comparison of fractions. Ecology 87, 2614–2625.

Peres-Neto, P.R., Dray, S., ter Braak, C.J., 2017. Linking trait variation to the environment: critical issues with community-weighted mean correlation resolved by the fourth-corner approach. Ecography 40, 806–816.

Resh, V.H., Rosenberg, D.M., 1993. Freshwater Biomonitoring and Benthic Macroinvertebrates.

Rico, A., Van den Brink, P.J., 2015. Evaluating aquatic invertebrate vulnerability to insecticides based on intrinsic sensitivity, biological traits, and toxic mode of action. Environ. Toxicol. Chem. 34, 1907–1917.

Rico, A., Van den Brink, P.J., Leitner, P., Graf, W., Focks, A., 2016. Relative influence of chemical and non-chemical stressors on invertebrate communities: a case study in the Danube River. Sci. Total Environ. 571, 1370–1382.

Rico, A., Arenas-Sánchez, A., Alonso-Alonso, C., López-Heras, I., Nozal, L., Rivas, D., Vighi, M., 2019. Identification of contaminants of concern in the upper Tagus river basin (central Spain). Part 1: screening, quantitative analysis and comparison of sampling methods. Sci. Total Environ. 666, 1058–1070.

Statzner, B., Beche, L.A., 2010. Can biological invertebrate traits resolve effects of multiple stressors on running water ecosystems? Freshw. Biol. 55, 80–119.

Tachet, H., Richoux, P., Bournaud, M., Usseglio-Polatera, P., 2010. Invertébrés d'eau Douce: Systématique, Biologie, Écologie. CNRS éditions, Paris.

Ter Braak, C.J., 2017. Fourth-corner correlation is a score test statistic in a log-linear trait-environment model that is useful in permutation testing. Environ. Ecol. Stat. 24, 219–242.

Ter Braak, C.J., Šmilauer, P., 2018. Canoco Reference Manual and user's Guide: Software for Ordination (Version 5.1). Microcomputer Power, Ithaca, USA.

Ter Braak, C.J., Peres-Neto, P., Dray, S., 2017. A critical issue in model-based inference for studying trait-based community assembly and a solution. PeerJ 5, 2885.

Ter Braak, C.J., Šmilauer, P., Dray, S., 2018a. Algorithms and biplots for double constrained correspondence analysis. Environ. Ecol. Stat. 25, 171–197.

Ter Braak, C.J., Peres-Neto, P.R., Dray, S., 2018b. Simple parametric tests for trait-environment association. J. Veg. Sci. https://doi.org/10.1111/jvs.12666.

Townsend, C.R., Hildrew, A.G., 1994. Species traits in relation to a habitat templet for river systems. Freshw. Biol. 31, 265–275.

Usseglio-Polatera, P., Bournaud, M., Richoux, P., Tachet, H., 2000. Biological and ecological traits of benthic freshwater macroinvertebrates: relationships and definition of groups with similar traits. Freshw. Biol. 43, 175–205.

Van den Berg, S.J.P., Baveco, H., Butler, E., De Laender, F., Focks, A., Franco, A., Rendal, C., Van den Brink, P.J., 2019. Modelling the sensitivity of aquatic macroinvertebrates to chemicals using traits. Environ. Sci. Technol. 53, 6025–6034.

Van den Brink, P.J., Hattink, J., Bransen, F., Van Donk, E., Brock, T.C., 2000. Impact of the fungicide carbendazim in freshwater microcosms. II. Zooplankton, primary producers and final conclusions. Aquat. Toxicol. 48, 251–264.

Van den Brink, P.J., Rubach, M.N., Culp, J.M., Pascoe, T., Maund, S.J., Baird, D.J., 2011. Traits-based ecological risk assessment (TERA): realizing the potential of ecoinformatics approaches in ecotoxicology. Integr. Environ. Assess. Manag. 7, 169–171.

Van Looy, K., Piffady, J., Floury, M., 2017. At what scale and extent environmental gradients and climatic changes influence stream invertebrate communities? Sci. Total Environ. 580, 34–42.

Wagner, H.H., 2004. Direct multi-scale ordination with canonical correspondence analysisS. Ecology 85, 342–351.

Wallace, J.B., Webster, J.R., 1996. The role of macroinvertebrates in stream ecosystem function. Annu. Rev. Entomol. 41, 115–139.