

# Autecology of terrestrial diatoms under anthropic disturbance and across climate zones

Jasper Foets<sup>a,b,\*</sup>, Jadwiga Stanek-Tarkowska<sup>c</sup>, Adriaan J. Teuling<sup>a</sup>, Bart Van de Vijver<sup>d,e</sup>, Carlos E. Wetzel<sup>b</sup>, Laurent Pfister<sup>b,f</sup>

<sup>a</sup> Department of Environmental Sciences, Wageningen University and Research, Wageningen, the Netherlands

<sup>b</sup> Environmental Research and Innovation Department, Luxembourg Institute of Science and Technology, Belvaux, Luxembourg

<sup>c</sup> Department of Soil Science, Environmental Chemistry and Hydrology, University of Rzeszow, Poland

<sup>d</sup> Research Department, Meise Botanic Garden, Meise, Belgium

<sup>e</sup> Department of Biology, ECOBE, University of Antwerp, Wilrijk, Belgium

<sup>f</sup> Faculty of Science, Technology and Medicine, University of Luxembourg, Belval, Luxembourg

## ARTICLE INFO

### Keywords:

Diatoms  
Anthropogenic disturbance  
Indicator  
Autecological values  
Soil

## ABSTRACT

Like aquatic diatoms, terrestrial diatoms are sensitive to multiple environmental factors such as pH, anthropic disturbances, soil moisture and nitrogen. While most aquatic diatom species have well-known autecological values for several important variables, qualifying them as biological indicators, this potential has not yet been untapped for terrestrial diatoms. Here, we aim to establish autecological values for pH and soil moisture content using weighted averaging for common, widespread terrestrial diatoms and compare them with previous calculated indicator values and tolerance ranges. In order to provide robust autecological values, we combined the data of ecological studies carried out across a range of climate conditions. We found that our validation metrics improved considerably after removing samples collected in anthropic disturbed habitats, suggesting that anthropic disturbance is the principal factor defining taxon occurrences. Besides our results showing similarities with previous studies, they also showed a significant improvement on the existing indicator values. Moreover, we expanded the list of terrestrial diatoms to 249 taxa that may serve as environmental markers in different research fields.

## 1. Introduction

Diatoms are a group of microscopic, single-celled algae living in almost all moist and aquatic environments with sufficient light (Dixit et al., 1992). Most species show very specific preferences for a broad range of environmental factors such as pH, nutrients and salinity, making them one of the most commonly used bio-indicators for water quality assessment and paleolimnological analysis (Smol et al., 2001; Smol and Stoermer, 2010). In order to infer degradation levels of water bodies or reconstruct past environmental conditions, autecological indicator values are regularly established (Carayon et al., 2019; Van de Vijver et al., 2002). These values are generally derived from weighted averaging; a simple, reliable and extensively used technique for estimating taxa indicator values and assumes that taxa abundance follows a unimodal relationship with a given environmental variable. Indicator values are the basis of some widely used diatom indices such as the IPS

(Specific Pollution Sensitivity Index; Cemagref, 1982) and BDI (Biological diatom index; Lenoir and Coste, 1996). Also, they have been gathered in synthetic trait matrices to be used for other types of ecological diagnosis (Carayon et al., 2020; Taylor et al., 2007; Dam et al., 1994). Overall, the calculation of autecological values for diatoms has proven to be a very useful tool in many aspects of water quality research (Poikane et al., 2020).

While aquatic diatoms are commonly studied and ecologically well characterised, mainly due to their general use in water quality monitoring programmes, studies on limno-terrestrial diatom communities (i. e. here used as diatom assemblages that can be found on soils) are rather scarce. However, ecological studies showed that diatom communities on soils are also quite responsive to several environmental variables such as soil moisture and pH (Antonelli et al., 2017; Lund, 1945; Van de Vijver et al., 2002; Van de Vijver and Beyens, 1998; Van Kerckvoorde et al., 2000). As a result, Van de Vijver et al. (2002), who investigated 106 soil

\* Corresponding author at: 41 Rue du Brill, L-4422 Belvaux, Luxembourg.

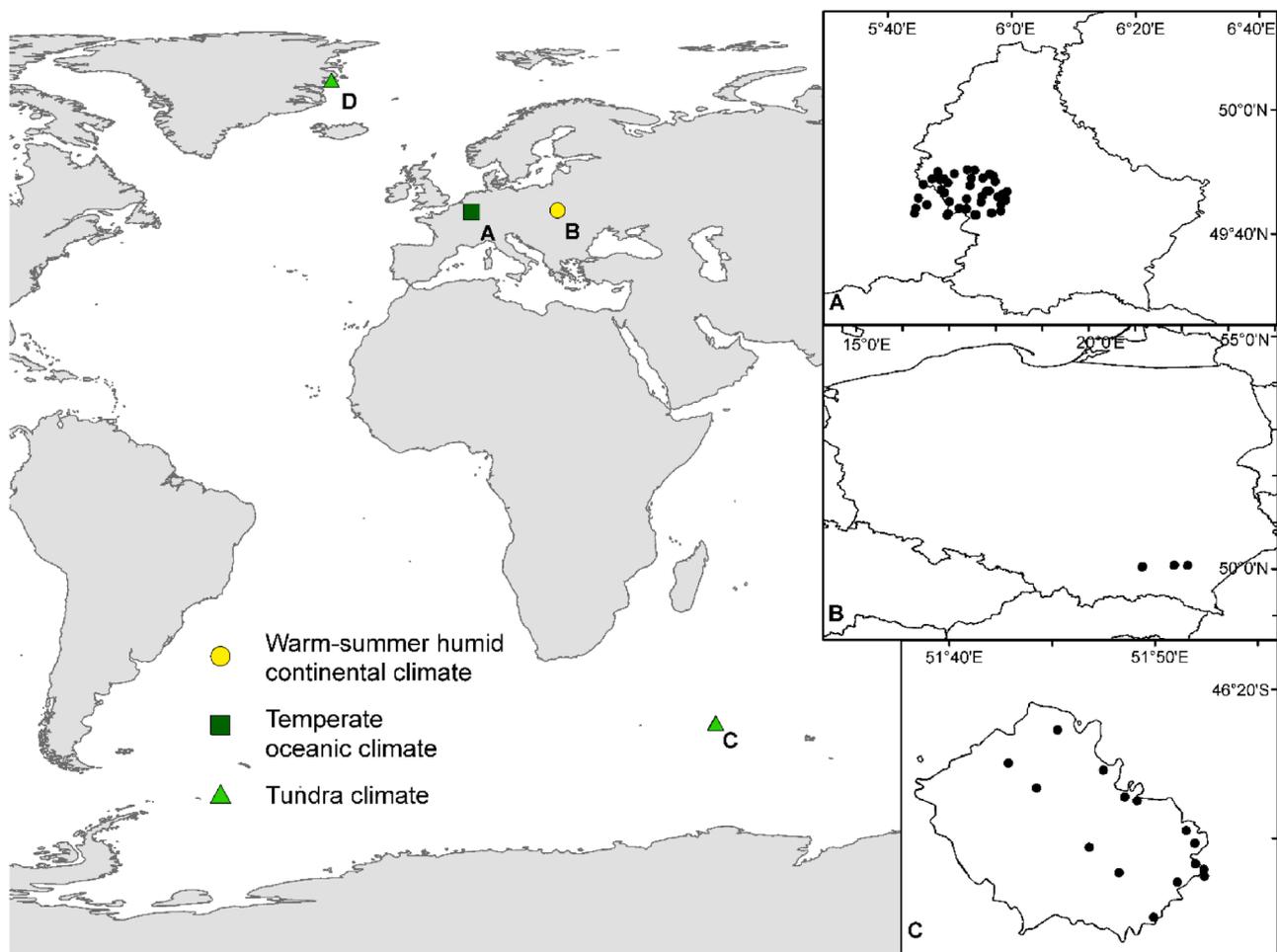
E-mail addresses: [jasper.foets@list.lu](mailto:jasper.foets@list.lu) (J. Foets), [carlos.wetzel@list.lu](mailto:carlos.wetzel@list.lu) (C.E. Wetzel).

<https://doi.org/10.1016/j.ecolind.2020.107248>

Received 7 September 2020; Received in revised form 30 October 2020; Accepted 2 December 2020

Available online 5 January 2021

1470-160X/© 2020 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).



**Fig. 1.** Locations of the sampling sites with indication of their climatic region according to the Köppen-Geiger classification (Beck et al., 2018). A, Luxembourg; B, Poland; C, Île de la Possession (Îles Crozet, sub-Antarctica); D, Zackenberg (Greenland). Source: Esri, Global Mapping International, U.S. Central Intelligence Agency (The World Factbook).

diatom samples on Île de la Possession (Crozet, sub-Antarctica), could establish optimum values of soil moisture for the most common taxa in their dataset occurring there. Also, Lund (1945), who sampled 66 different soils in the UK, was able to determine tolerance ranges to pH for 24 taxa. In addition to soil moisture and pH, disturbances caused by farming practices also play a key role in structuring terrestrial diatom assemblages (Antonelli et al., 2017; Foets et al., 2020a; Heger et al., 2012; Stanek-Tarkowska and Noga, 2012; Vacht et al., 2014). Foets et al. (2020a) found that disturbed areas were less diverse and that land uses with different disturbance levels could be differentiated solely based on the community composition. They also noticed that the species composition remains stable throughout the year, meaning that variation in soil moisture availability, irradiance and temperature does not play a significant role. Besides the direct influence of these disturbances on diatoms, they indirectly affect other variables as well such as organic matter, nitrogen and carbon content in the soil, which in turn also impact diatoms (Stanek-Tarkowska et al., 2018a). Both nitrogen and carbon are often found as significant explanatory variables for the diatom species composition (Antonelli et al., 2017; Vacht et al., 2014), while organic matter increases the moisture holding capacity of the soil and acts as a buffer against dryness (Stanek-Tarkowska et al., 2018a; Stanek-Tarkowska and Noga, 2012). Thus, terrestrial diatoms are sensitive to several environmental variables, but so far, most studies focused on the communities' structure and not at the species level preferences.

Pending more ecological knowledge on terrestrial diatoms, several studies explored their potential as environmental markers. Many of

them tried to explain the occurrence of diatoms on soils based on the ordinal classifications created by Dam et al. (1994) (see for instance Antonelli et al., 2017; Stanek-Tarkowska et al., 2013; Stanek-Tarkowska and Noga, 2012). Despite that these classifications are based on data acquired from aquatic samples in the Netherlands (and ecological values reported in the published literature), they are nevertheless frequently used in many European countries as a reference for autecological studies. However, for numerous terrestrial taxa, similar values are lacking, as they only occur sporadically in aquatic environments, and assigned values may not really reflect their behaviour in true terrestrial conditions (Barragán et al., 2018). There is a pressing need for assigning autecological values to terrestrial diatoms, reflecting their preferences in terrestrial environments and which would enable us to unlock their potential as environmental markers.

Despite the worldwide application of some diatom-based indices (e.g. IPS), there is nowadays a tendency to develop indices and indicator values specific to certain regions, since these values would better reflect environmental conditions in that area (Carayon et al., 2020, 2019; Lavoie et al., 2009). Unfortunately, data on terrestrial diatom ecology remains scarce and developing an index or defining indicator values at the local scale is at present still not possible. Therefore, we aim to provide robust autecological values for common diatom taxa living on soils based on data of several ecological studies carried out across a range of climate and geographical conditions. We will analyse several environmental variables, which previously showed to influence diatom communities living on soils. This shall enable us not only to use them later as environmental indicators, but also to expand the existing toolbox of

environmental markers applied in for example hydrology, soil and science.

## 2. Material and methods

### 2.1. Description of the data

In this study, the terrestrial diatom community and environmental data were gathered from the studies of Antonelli et al. (2017), Barragán et al. (2018), Foets et al. (2020a), Foets et al. (2020b), Stanek-Tarkowska et al. (2018a), Stanek-Tarkowska et al. (2015), Stanek-Tarkowska and Noga (2012), Van de Vijver et al. (2002) and Van Kerckvoorde et al. (2000). The data covers four distinct regions and three different climate types (Fig. 1).

The studies of Foets et al. (2020a) and Foets et al. (2020b) were conducted in the Attert River basin in Luxembourg. Terrestrial diatom samples were taken at the soil surface according to the method described by Coles et al. (2016) in 16 sites around the catchment every 3–5 weeks for a period of 14 months (October 2017 to November 2018) totalizing 206 samples that were included in the present study. Subsequently, the same soil samples were pooled and used for pH and nutrient analysis. Of the several environmental variables included in that study, only pH measured in 1:5 H<sub>2</sub>O, volumetric soil moisture content (VWC, expressed in percent) measured *in situ* (30 times around the sampling area), and the bioavailable fraction of total nitrogen (TN), Phosphorus (P) and Dissolved organic carbon (DOC) (both in mg L<sup>-1</sup> soil extraction) were used. The latter two were analysed with ICP-OES after a 0.01 M CaCl<sub>2</sub> extraction following (Houba et al., 2000). Furthermore, they incorporated five different types of land use (forest, undisturbed grassland, grassland disturbed by cattle grazing, grassland disturbed by agriculture and agricultural fields).

Similar to Foets et al. (2020a) and Foets et al. (2020b), the study of Antonelli et al. (2017) was also carried out in the Attert River basin from August 2014 to March 2015. Diatoms were sampled as in Foets et al. (2020a) and Foets et al. (2020b) at 34 locations during three sampling campaigns totalling 92 samples. Topsoil samples for physico-chemical analysis were collected simultaneously with a shovel. Contrary to Foets et al. (2020a), only three land use types were distinguished. However, based on photographs and additional information on the sampling areas, we were able to adjust the different land uses according to their classification. Besides, soil moisture was measured gravimetrically (%) rather than volumetrically, pH in H<sub>2</sub>O, KCl and CaCl and C, H and N (all expressed in percent) using a CHN analyser. Of those variables, pH<sub>H2O</sub>, N and C were included. While pH is merged with the results of Foets et al. (2020a) and Foets et al. (2020b), N and C are regarded as different variables.

Barragán et al. (2018) sampled four locations in the Attert basin comprising four different anthropogenic pressures. At each location, they collected 10 soil samples for diatom analysis following the method described in Coles et al. (2016) and determined soil moisture volumetrically. Since soil moisture measurements were done as in Foets et al. (2020b), both variables were aggregated. Also, for this study, photographs and information on the locations were provided and enabled us to classify the different land uses.

In Stanek-Tarkowska and Noga (2012), two agricultural fields under different tillage systems, located in the Subcarpathian region (SE Poland), were sampled once. The samples were collected at a depth of 0–3 cm and put into petri dishes. In that study, some samples were also cultured, but these were excluded here. Besides, they also collected soil samples from the surface layer (0–10 cm) of which they measured pH in 1:2.5 KCl. These values were converted to pH 1:5 H<sub>2</sub>O following the Eq. (1) provided in Kabała et al. (2016) before aggregating the data.

$$pH_{H2O1.5} = -1.95 + 11.58 * \log_{10}(pH_{KCl1.2.5}) \quad (1)$$

The study of Stanek-Tarkowska et al. (2015) was also conducted in Poland. There, one site located in Pogórska Wola was sampled for nine

**Table 1**

Overview of the different datasets and variables included in this study. The number of cells per sample is the minimum number of diatom cells that was targeted in the corresponding study. However, sometimes this number could not be reached due to a low presence of diatoms in the samples. Converted or modified factors are indicated by '\*'. <sup>1</sup>Study only included the most abundant taxa, therefore it was only used for the calculation of the autecological values.

Dataset	Location	No. of samples	No. of diatom cells/samples	Variables used in this study
Van de Vijver et al. (2002)	Crozet, sub-Antarctica	77 (106 × 1)	500	pH, VWC
Van Kerckvoorde et al. (2000)	Zackenber, Greenland	30 (30 × 1)	500	pH, VWC, SOM
Foets et al. (2020a, b)	Attert basin, Luxembourg	206 (16 × 14)	200	pH, VWC, TN, P, DOC, land use
Antonelli et al. (2017)	Attert basin, Luxembourg	92 (34 × 3)	200	pH, C, N, land use*
Barragán et al. (2018)	Attert basin, Luxembourg	40 (4 × 10)	400	VWC, land use*
Stanek-Tarkowska and Noga (2012)	Dabrowa, Poland	2 (2 × 1)	400	pH*
Stanek-Tarkowska et al. (2015)	Pogórska Wola, Poland	9 (1 × 9)	400	pH*
Stanek-Tarkowska et al. (2018a) <sup>1</sup>	Rzeszów, Poland	60 (2 × 30)	400	pH, soil moisture

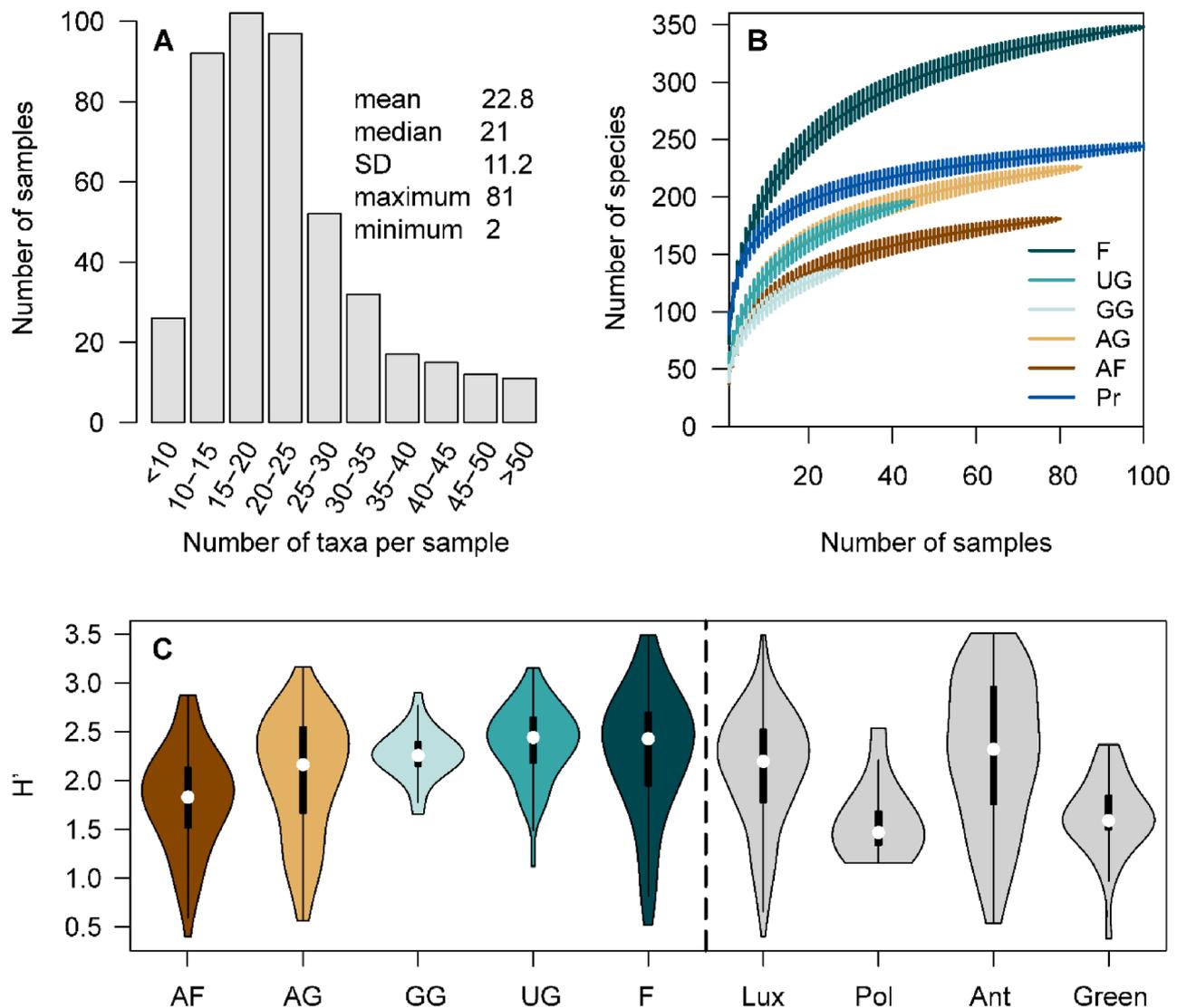
consecutive months in 2011. Soil samples for diatom analyses were collected from a 0–3 cm deep layer and placed in petri dishes (three replicates), whereas samples for physico-chemical analysis were taken from the topsoil layer (0–5 cm) at the same time. pH was measured in a 1:2.5 KCl solution and was converted following Eq. (1) before data merging.

Stanek-Tarkowska et al. (2018a) collected monthly samples from April to November for four consecutive years (2013–2016) at two agricultural fields near Rzeszów (Poland) (n = 60). There, two replicates for diatom analyses were taken from the 0–5 cm soil layer and placed in petri dishes, while undisturbed soil samples were taken with 100 m<sup>3</sup> cylinders for physico-chemical analysis. pH was determined as in Foets et al. (2020a) and Foets et al. (2020b), while gravimetric soil moisture content was converted to volumetric moisture content using soil bulk density. Both variables were used in our study without further adjustments.

The dataset of Van Kerckvoorde et al. (2000) included 30 diatom samples originating from 30 different sites in the Zackenberg area (NE Greenland) covering an area of approximately 4 km<sup>2</sup>. For those samples, the upper 3 cm of the soils was collected. In addition to diatom community data, soil moisture and pH were measured similar to Foets et al. (2020b), while SOM was derived from LOI (Loss-On-Ignition). All three variables were included without modification in the final dataset.

Finally, Van de Vijver et al. (2002) analysed 106 samples collected on the sub-Antarctic Île de la Possession (Îles Crozet). Of those, 29 originated from permanently wet areas or moistened rocks (i.e. moisture content > 75%) and were excluded from the final dataset, as they did not reflect the true terrestrial soil conditions as is the case in the other studies. In the sub-Antarctic study, several environmental variables were assessed, but only pH 1:5 H<sub>2</sub>O and VWC were incorporated, because the other variables were measured in a different way and could therefore not be converted.

After merging the community datasets, the entire dataset was made taxonomically consistent by updating species names to the most recent diatom taxonomy, putting synonyms together and assigning four-letter codes to each taxon according to the latest Omnidia version (March 2019; Lecointe et al., 1993). Since identification of terrestrial diatoms is



**Fig. 2.** Shannon-Wiener and species richness of the samples and comparison between land uses and regions. A, frequency bar plot of the taxon richness per sample. B, Sample-based species rarefaction curves per land use type. C, Shannon diversity per region and land use type. F, forest; UG, undisturbed grassland; GG, grazed grassland; AG, agricultural grassland; AF, agricultural field; Pr, pristine (i.e. Antarctica and Greenland). Lux, Luxembourg; Pol, Poland; Ant, sub-Antarctica; Green, Greenland. These analyses include 456 samples and rare species. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

often sparse in the literature, several publications were used, including Ettl and Gärtner (1995), Krammer (2000), Lange-Bertalot (2001), Lange-Bertalot (2011), Lange-Bertalot et al. (2017), Levkov et al. (2013) and Lund (1946), along with some essential studies on terrestrial algae (Brendemuhl, 1949; Hustedt, 1942; Petersen, 1915, 1928, 1935) and some recent studies on terrestrial diatoms (Reichardt, 2008, 2012; Wetzel et al., 2015). Available diatom pictures from the different studies were used to make the datasets taxonomic consistent as much as possible. Taxon names were kept in the broad sense (i.e. *sensu lato*) as our aim was to give general autecological values. This resulted in a final dataset including 516 soil diatom samples across 166 different sites, covering 710 taxa (including varieties, subspecies and forms), and the following variables: pH, VWC, TN, DOC, N, C and SOM. Table 1 provides an overview of all data.

## 2.2. Statistics

The Shannon-Wiener diversity and species richness were calculated for each sample. However, the communities given in Stanek-Tarkowska et al. (2018a) were excluded from further analysis, as the data only

contained the 14 most abundant taxa and not the entire communities. Furthermore, the significance of soil moisture, pH and type of land use in explaining the variation in those two variables was investigated. For this, a generalised mixed model with site as random variable to account for repeated measurements was set up. The Shannon-Wiener index between four different regions (i.e. Luxembourg, Poland, sub-Antarctica and Greenland) and five land use types (i.e. forest, undisturbed grassland, grazed grassland, agricultural grassland and agricultural field) was analysed with analysis of variance (ANOVA) and assessed the difference further with parametric *post hoc* tests. Species rarefaction curves for each region and land use type using the function *specaccum* incorporated in the *vegan* R-package.

Next, the community dataset was reduced by removing rare taxa. A taxon was considered 'rare' when it did not occur with a minimum relative abundance of 2.5% in at least five samples (i.e. 0.01% of the samples). The identification of most of those taxa was also uncertain (i.e. sp., cf. or aff.). Then, species autecological values were computed using weighted averaging regression (ter Braak and Looman, 1986). After the calculation of the ecological indicator values, the inverse algorithm (i.e. WA calibration) was used to build a model predicting value based on

floristic data. These two values were then compared and assessed after bootstrap cross-validations ( $n = 1000$ ) using root mean square error (RMSE; Wallach and Goffinet, 1989) and correlation coefficients as model validation metrics. The best results (i.e. lowest RMSE and highest  $r^2$ ) were obtained with data that was tolerance down-weighted and transformed according to inverse deshrinking models.

Additionally, a training-set containing half of the samples (i.e. all odd samples) for pH and soil moisture was created. Then the goodness-of-fit was assessed by passively fitting the passive samples (i.e. all even samples) into a constrained ordination of the training-set with pH or soil moisture as the sole constraint. The passive samples were positioned as supplementary samples within the ordination space by means of transition equations given in ter Braak and Smlauer (2002). This determines subsequently a score for each passive sample by taking the weighted average (canonical ordination analysis, CCA) of the species scores extracted from the ordination of the training-set samples. Hence, the even samples are positioned within the ordination without influencing the underlying ordination solely based on the training set (Birks et al., 2012). CCA as ordination method was chosen, since DCA on the reduced species data revealed a unimodal distribution for the first four axes (S.D. > 4). Next, the distribution of the squared residual distances between each sampling point and its fitted position on the first constrained axis for the training-set was calculated. Any passive sample that has a squared residual fit greater than the 90th percentile distance for the training-set samples is poorly fitted within the calibration function model (Birks et al., 1990). Afterwards, the two sample sets were switched and the analysis was done again to verify the previous outcome. All previous calculations were done with the *residLen* function from the *analogue* R-package (Simpson and Oksanen, 2020).

After estimating autecological values (optimum and tolerance) for each taxon, the optimum values of pH and soil moisture were compared with the updated indicator values of the same variables assigned by Dam et al. (1994). Differences between the two sets of optimum values were tested using ANOVA. In case of significant differences were found, the Tukey HSD test was used to reveal which categories deviated from the others. The normality and homoscedasticity of the model residuals were checked prior to statistical analysis. For all aforementioned statistical analyses were performed using the R statistical program (R v. 3.6.3.; <http://www.r-project.org/>) and additional functions from the R-packages *vegan* (version 2.5–6; Oksanen et al., 2019), *analogue* (version 0.17–4; Simpson and Oksanen, 2020) and *rioja* (version 0.9–21; Jugings, 2017).

### 3. Results

An overall mean species richness per sample of  $21.7 \pm 11.1$  with a maximum of 81 and a minimum of 2 (Fig. 2) was observed in the entire dataset. From the species rarefaction curves, we observe that communities sampled in undisturbed areas have generally a higher species richness per sample than disturbed sites. A closer look reveals that after approximately 80 samples counted, we should find between 300 and 350 different species in forested sites, which is twice the number found for agricultural fields. Also, the rarefaction curves for the samples collected in Greenland and sub-Antarctica reach a clear asymptote at approximately 250 taxa.

The Shannon diversity values are significantly lower in areas that are highly disturbed by agriculture (AF) ( $F = 13.74$ ,  $df = 443$ ,  $R^2 = 0.028$ ,  $P < 0.01$ ), while pastures, undisturbed grassland and forest did not present a significantly different diversity ( $P > 0.05$ ). Furthermore, we notice that the communities collected in Greenland and Poland have a lower diversity than the ones taken in Luxembourg and sub-Antarctica ( $F = 11.22$ ,  $df = 452$ ,  $R^2 = 0.063$ ,  $P < 0.05$ ). It must be noted however, that all samples from the Polish sampling sites used here are coming from anthropic disturbed areas and are therefore comparable to samples from disturbed fields (AF, AG) elsewhere. However, the sites were not assigned to a land use type, since verification material (i.e. pictures and

**Table 2**

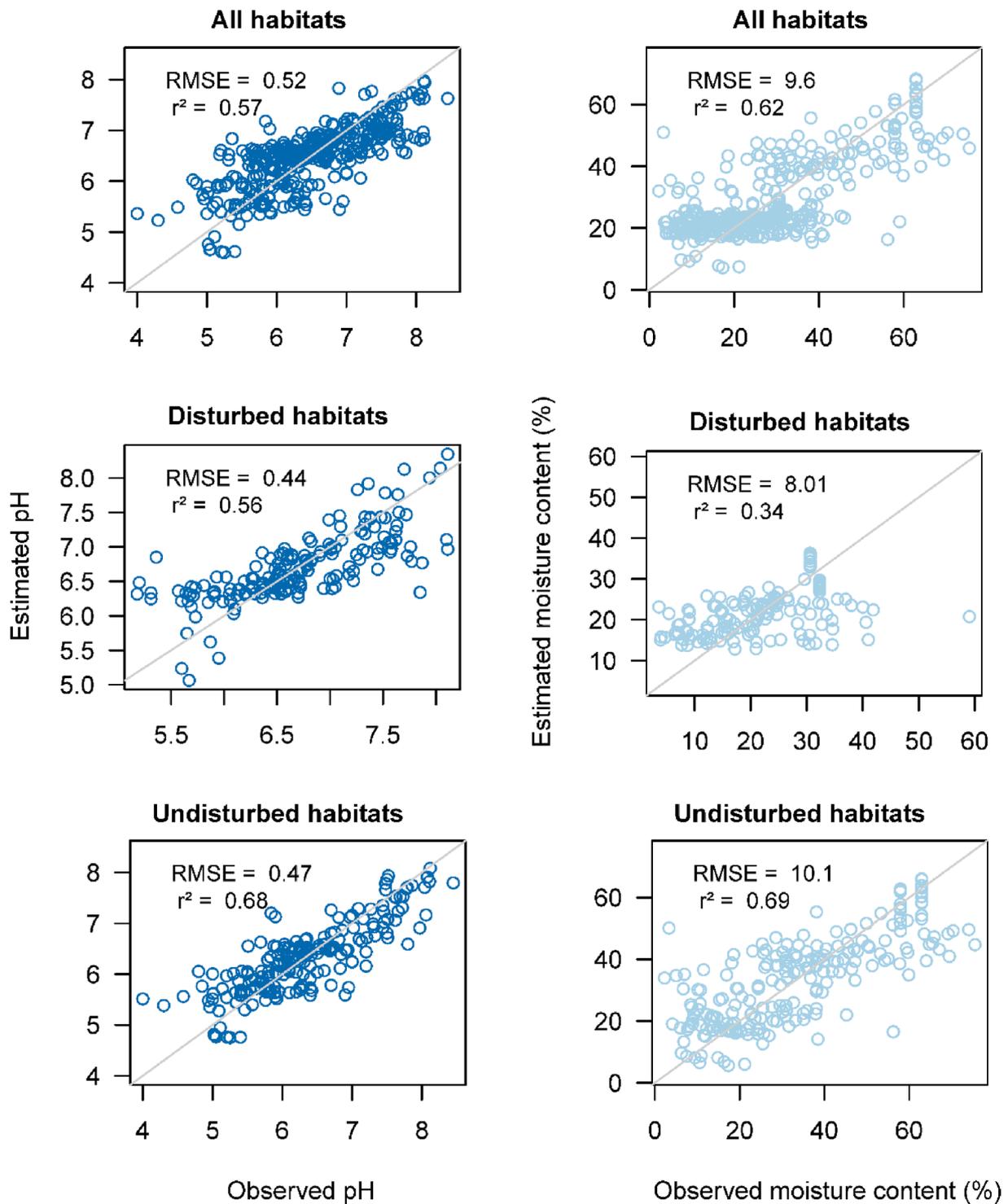
Validation of the weighted average of different environmental variables. Results of the validation ( $r^2$  and RMSE) are given for the data that was tolerance down-weighted and inverse deshrinked. The reference between brackets gives the study where the variable was analysed. Also, the number of samples included in the analysis and their range are provided. SOM, Soil Organic Matter; VWC, Volumetric Moisture Content; RMSE, Root Mean Square Error.

Variable (Study)	n	Range (units)	$r^2$	RMSE
TN (Foets et al., 2020a, 2020b)	206	0.25–53.70 (mg L <sup>-1</sup> soil extraction)	0.16	9.32
DOC (Foets et al., 2020a, 2020b)	206	1.70–57.07 (mg L <sup>-1</sup> soil extraction)	0.18	9.20
P (Foets et al., 2020a, 2020b)	206	0–0.88 (mg L <sup>-1</sup> soil extraction)	0.31	0.16
C/N (Antonelli et al., 2017)	89	4.61–29.83	0.07	4.63
SOM (Van Kerckvoorde et al., 2000)	30	9.93–79.85 (%)	0.29	17.34
pH (see Table 1)	388	4–8.45	0.57	0.52
VWC (see Table 1)	411	2.27–75.5 (%)	0.62	9.61

additional notes on the sampling areas) was not available. In addition, we ran a generalised mixed model with 'site' as random variable to check whether differences in species richness and diversity could be explained by land use, soil moisture and/or pH. The first model revealed that the type of land use ( $P < 0.01$ ) and soil pH ( $P < 0.05$ ), which was positively correlated, explained 84.5% of species richness, whereas the second one showed that they both explained 62.7% of the species diversity ( $P_{landuse} < 0.05$ ,  $P_{pH} < 0.01$ ). Soil moisture was not significant in both models ( $p > 0.05$ ). Although samples are coming from contrasted environments, the results indicate that the species richness and diversity are similar between the samples and are likely driven by the same environmental factors.

After using the cut-off criteria of a 2.5% relative abundance in five samples, 248 of the 710 taxa were kept in the statistical analyses, covering 48 genera of which *Pinnularia* (45 taxa), *Nitzschia* (18), *Navicula* (13), *Mayamaea* (13), *Humidophila* (13) and *Eunotia* (12) were the most species rich genera. The most abundant taxa in the soil dataset are *Hantzschia amphioxys* (Ehrenberg) Grunow (12.0% of total relative abundance), *Mayamaea atomus* (Kützing) Lange-Bertalot (5.9%), *M. permissis* (Hustedt) Bruder & Medlin (5.6%), *Pinnularia obscura* Krasske (5.1%) and *Nitzschia pusilla* (Kützing) Grunow (4.6%). Of all the taxa, only *Luticola mutica* s.L. (Kützing) D.G. Mann, species complex *Pinnularia borealis* Ehrenberg (see Pinseel et al., 2019), *P. obscura*, *M. atomus*, and *Encyonema silesiacum* (Bleisch) D.G.Mann occurred in all four regions, while *Caloneis lancettula* (Schulz) Lange-Bertalot & Witkowski, *Mayamaea excelsa* (Krasske) Lange-Bertalot, *Stauroneis parathermicola* Lange-Bertalot, *Craticula molestiformis* (Hustedt) Lange-Bertalot and *Humidophila irata* (Krasske) R.L.Lowe et al. were some of the taxa only observed in the samples from Luxembourg. *Achnanidium pyrenaicum* Hustedt was only noted in Polish samples, whereas *Stauroneis thermicola* (J.B.Petersen) J.W.G.Lund and *Sellaphora nana* (Hustedt) Lange-Bertalot et al. occurred only in samples from both Luxembourg and Poland. Typical taxa exclusively found in sub-Antarctica include for instance *Psammothidium aretasii* (Manguin) Lange-Bertalot, *Frankophila maillardii* (Le Cohu) Lange-Bertalot, *Humidophila comperei* (Le Cohu & Van de Vijver) R.L. Lowe et al., *Pinnularia subantarctica* (Manguin) Van de Vijver & Le Cohu, *Achnanidium sieminskiae* Witkowski, Kulikovskiy & Riaux-Gobin and, *Humidophila ingeae* (Van de Vijver) R.L. Lowe et al., whereas *Caloneis aerophila* Bock, *Eunotia fallax* Cleve and *Pinnularia lagerstedtii* (Cleve) Cleve-Euler were only present in samples from Greenland.

We calculated weighted averages for TN, P, DOC, C/N, SOM, pH and VWC and validated the calculated with the observed values. Bar the latter two, all other variables are coming from one single study (Table 1) and therefore those analyses are restricted to smaller sample sizes. The outcome of the validations show generally (very) low correlation coefficients and relative high RMSE values for TN, C/N, P, DOC and SOM

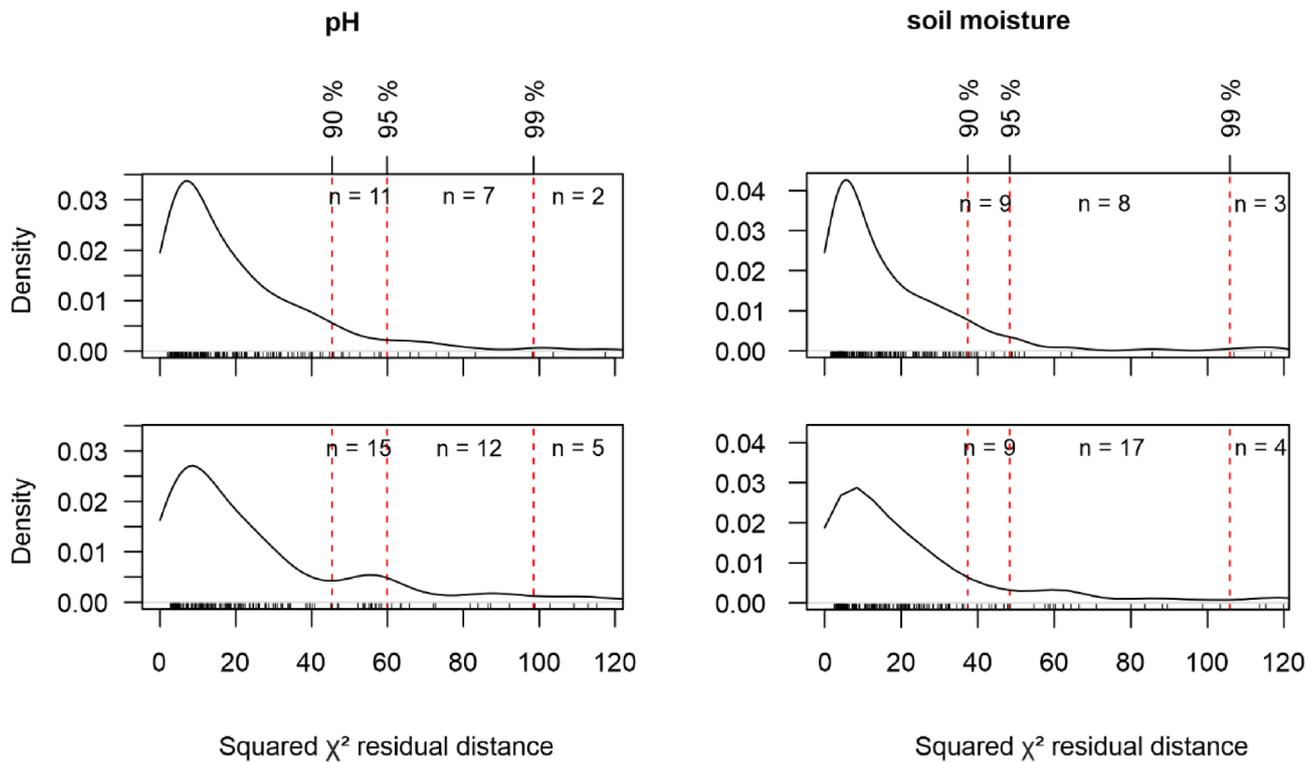


**Fig. 3.** Relationships between observed and estimated values using weighted averaging. A separate analysis has been done for the disturbed habitats (AG, GG, AF) and undisturbed (UG, F, Pr). Here, only taxa were included with a minimum abundance of 2.5% in at least five samples. Regarding pH, 202 and 173 data points were included in the validation of the undisturbed and disturbed habitats respectively, while there were 209 and 142 for soil moisture.

(Table 2), while pH and soil moisture reach an overall  $r^2$  of respectively 0.57 and 0.62 (Fig. 3). Furthermore, we divided the samples based on the type of land use (anthropic disturbed vs. undisturbed) and ran the same analyses again. For both pH and soil moisture content, we now observed increased correlation coefficients. pH reached 0.68 and had a RMSE of 0.47 compared to 0.56 and 0.44 when only disturbed habitats were considered, whereas for soil moisture a substantial difference of  $r^2$  was seen between disturbed (0.34) and undisturbed areas (0.69). The

RMSE also increased with 0.5 compared to the overall value.

In addition to evaluating the regression between the observed and estimated environmental variables, we also calculated 'goodness-of-fit' statistics from a CCA ordination for pH and soil moisture. Therefore, we subdivided the samples to create two equally large datasets: a training set and a passive sample set. For pH, both sets contained 194 samples, whereas for soil moisture the sets contained respectively 206 and 205 samples. The distribution of the squared residuals for all four sets of



**Fig. 4.** Density plots of the squared residual fit to pH and soil moisture. The results are given for the training data set (odd samples; upper panels) and the passive samples (even samples; lower panels) derived from passively overlaying the odd samples on to a canonical correspondence analysis ordination (CCA) of the training-set samples. The pH training set includes 194 samples, while the set for soil moisture yields 206 samples. The labelled dashed lines are for the 90th, 95th, and 99th percentiles of the distributions of the two sets of squared residual lengths. Samples lying beyond the 99th percentile are extremely poorly fitted to pH and soil moisture, respectively, those between the 95th and 99th percentiles are very poorly fitted, and those samples between the 90th and 95th percentiles are poorly fitted.

**Table 3**

Previous and calculated tolerance ranges. Previous ranges are based on studies of Van de Vijver et al. (2002) for soil moisture and Lund (1945) concerning pH. Lund determined pH colorimetrically, while we determined it electrometrically in a 1:5 water solution. Therefore, values may deviate a bit. \*, Species only present in the samples of Van de Vijver et al. (2002).

Taxon	Van de Vijver et al. (2002)	This study	Taxon	Lund (1945)	This study
PBOR	0–75	12.9–43.5	PBOR	4.4–6.0	5.4–6.6
POBS	18–76	11.7–40	NCHH	6.9–7.4	7.1–7.9
ABRY	21–80	13.5–54	SEAT	6.2–7.2	6.0–7.3
HAMP	0–50	10.3–30.7	TDEB	7–7.2	6.0–7.8
MAAT	5–46	17.2–36.6	HMON	6.9–8.2	7.0–8.1
LMUT	19–43	12.3–40.4	LMUT	6.0–8.3	5.6–6.9
HUCO	13–62	8.7–33.4	HUCO	3.9–8.3	5.7–7, 5.18–6.28
			(incl. DBIC)		
NPAL	17–69	13.5–30.3	NPAL	5.9–7.4	6.1–7.6
HABU	31–71	7.5–31.3	PMIC	4.6–6.6	5.1–6.8
DCOP*	0–43	2.2–34.4			
HUCZ*	20–58	23.9–51.1			
PKOL*	5–40	10.2–41.2			
PAUE*	20–79	18.8–56.8			

samples is shown in Fig. 4. In the training sets, we observe that for both pH and VWC 10% of the samples were poorly fitted (i.e. outside 90th percentile), whereas those values increased to 16.5 (pH) and 14.6% (VWC) for the passive samples. Similar percentages were found when we switched the sample sets. So, the analysis of the squared residual distances for the passive samples to pH and soil moisture indicates that 82–85% of the samples are well fitted within the unimodal response-model framework and that we can therefore be confident of the estimated autecological values.

The obtained autecological values were then compared with previous studies of Van de Vijver et al. (2002) and Lund (1945) for respectively volumetric soil moisture content and pH (Table 3). Prior to comparing the values, the diatom identifications made by Lund (1946) were checked if his identifications were the same as ours. Lund (1945) unfortunately determined pH using a colorimeter, whereas in our study, an electrometric method was used. Although the outcomes of the two methods should not deviate much more than 0.2 units from each other according to Haines et al. (1983), comparisons should be interpreted with care. Nevertheless, it seems that all ranges for pH overlap well with the values of Lund (1945). However, *Tryblionella debilis* Arnott has a very small tolerance range (7–7.2) specified by Lund (1945), whereas we calculated a rather broad pH tolerance ranging from 6.0 to 7.8. Also, *Humidophila contenta* s.l. (Grunow) R.L.Lowe et al. (including *Humidophila biceps* (Grunow) P.Furey, K.Manoylov & R.L.Lowe) occurs on very acidic soil (pH = 3.9) according to Lund (1945), while this study indicates that both taxa are normally present on soils with a pH between respectively 5.7 and 7 and 5.18 and 6.28. Furthermore, we see some larger differences between the ranges for soil moisture content. The ranges assigned by Van de Vijver et al. (2002) tend to go more towards higher moisture concentrations and are generally broader, compared to ours. This is not only the case for common, widespread taxa, but also for taxa such as *Humidophila crozetikerguelensis* (Le Cohu & Van de Vijver) R. L.Lowe et al., *H. comperei* and *Planothidium aueri* (Krasske) Lange-Bertalot which were only found in the samples of Van de Vijver et al. (2002). However, this did not result in non-overlapping ranges, except for the very common *Hantzschia abundans* Lange-Bertalot. For that species, we found a range between 7.5 and 31%, while the tolerance calculated by Van de Vijver et al. (2002) ranged between 31 and 71%. Bar this exception, the tolerance ranges for both soil moisture and pH overlap well with previous research.

Optima and tolerance values for common on soils occurring diatom

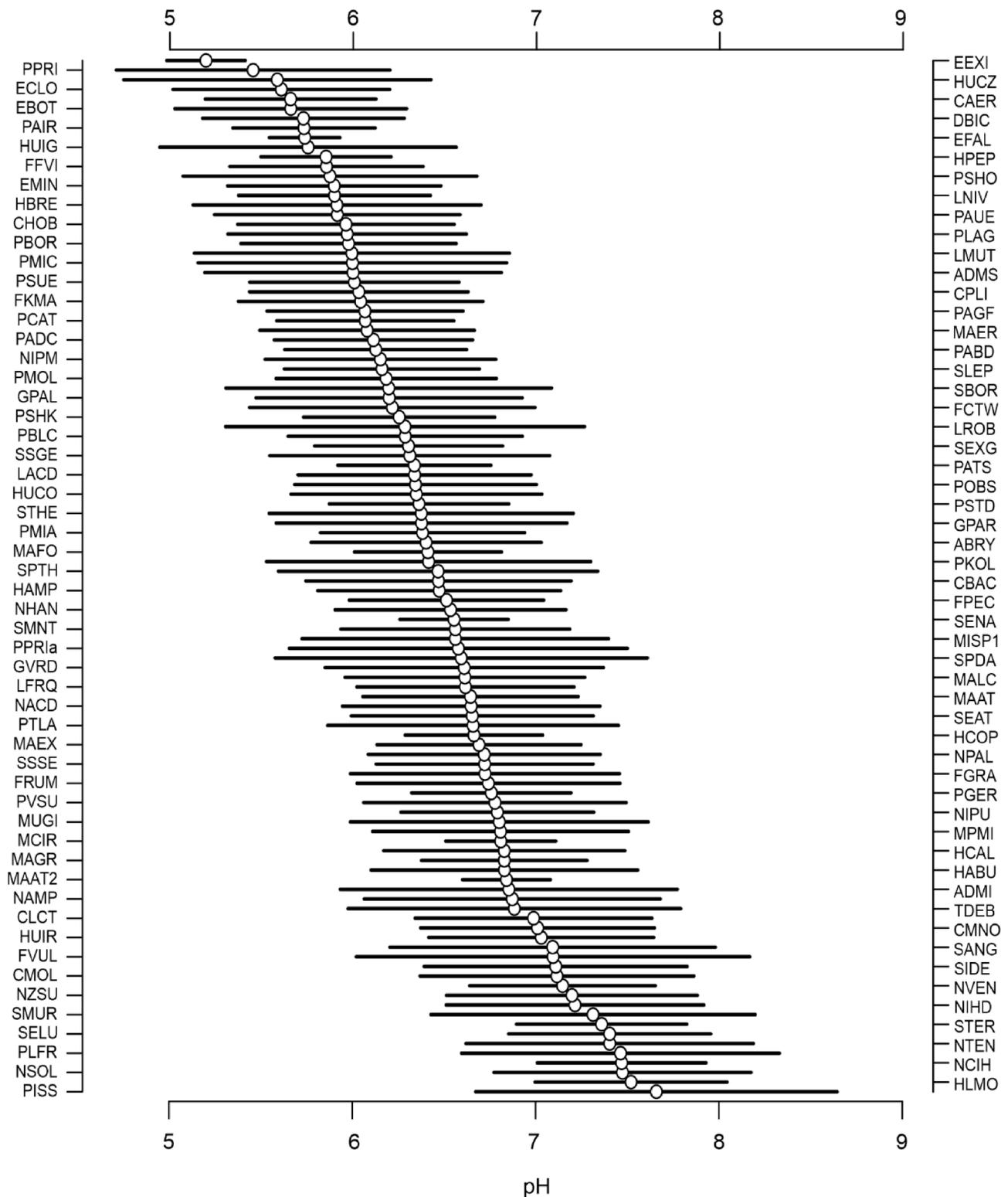


Fig. 5. Optimum and tolerance values of pH for the dominant soil diatom taxa (10% in 10 samples). Diatom taxa are abbreviated with a four-letter code following Omnidia. The diatom taxa with their respective Omnidia codes are given in Table 4.

taxa with respect to pH and soil moisture are shown in Figs. 5 and 6. There, we observe that some taxa such as *Eunotia exigua* (Brébisson ex Kützing) Rabenhorst, *S. nana* and *Meridion circulare* (Greville) C.A. Agardh are very sensitive to pH. For example, *E. exigua* only occurs in very acidic environments, while *S. nana* is rather restricted to soils with a neutral pH. On the contrary, several species including *H. ingae*, *Luticola robusta* Van de Vijver, Ledeganck & Beyens, *Frustulia vulgaris*

(Thwaites) De Toni, *H. crozetikerguelensis* and *Stauroneis pseudoagrestis* Lange-Bertalot & Werum are rather tolerant to pH and could be present on a wide range of soils. We observe similar patterns for VWC. Certain species such as *Adlafia minuscula* (Grunow) Lange-Bertalot var. *minuscula*, *Adlafia minuscula* var. *muralis* (Grunow) Lange-Bertalot, *E. exigua* and *Chamaepinnularia obsoleta* (Hustedt) C.E.Wetzel & Ector only require a low moisture content to grow and reproduce (i.e. low optimum

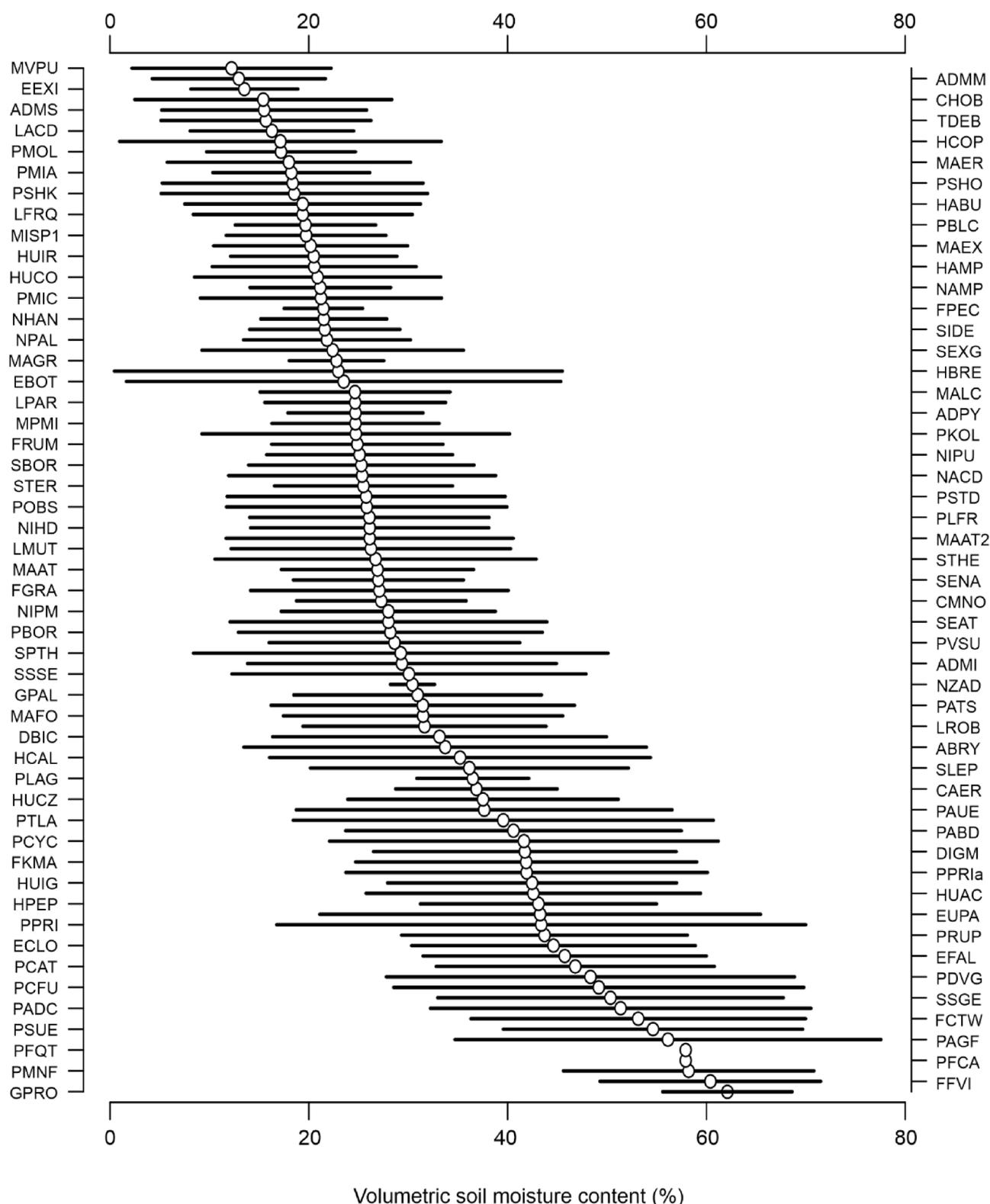
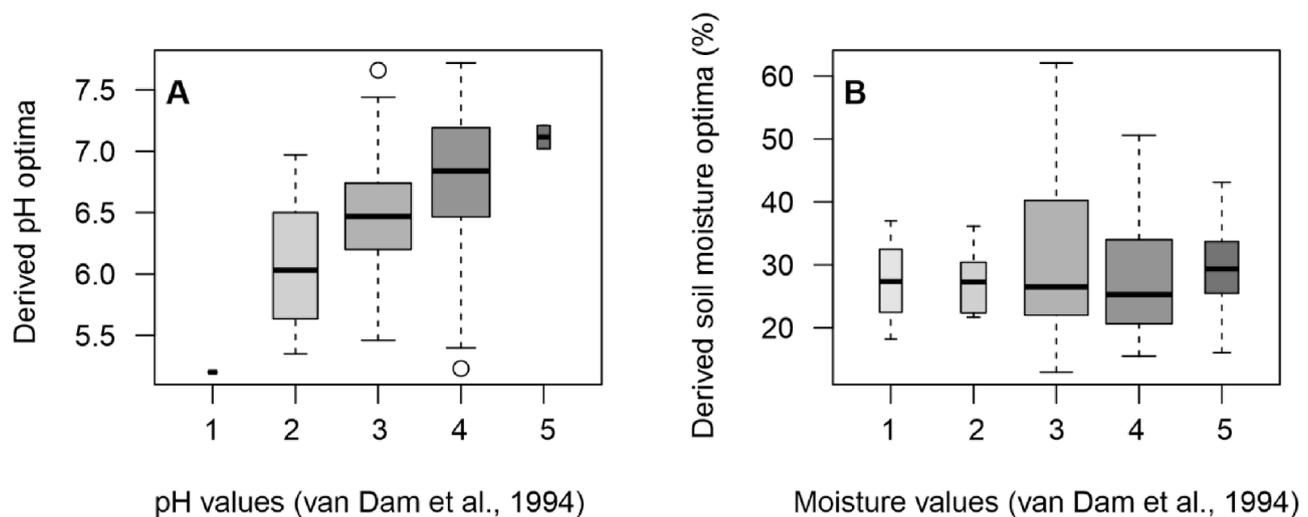


Fig. 6. Optimum and tolerance values of volumetric soil moisture content for the dominant soil diatom taxa (10% in 10 samples). Diatom taxa are abbreviated with a four-letter code following Omnidia. The diatom taxa with their respective Omnidia codes are given in Table 4.

values), whereas *Pinnularia isselana* Krammer, *Halamphora montana* (Krasske) Levkov, *Nitzschia solgensis* Cleve-Euler and *Navicula cincta* var. *heuffleri* Grunow need a soil that has a VWC of at least 40 to 50%. Besides, for *Humidophila brekkaensis* (Petersen) R.L.Lowe et al., *Mayamaea alci-monica* (E.Reichardt) C.E. Wetzel, Barragán & Ector, *Pinnularia perir-rorata* Krammer and *Pinnularia angliciformis* Van de Vijver & Beyens it

seems that other environmental factors are more important than soil moisture, since they have large tolerance values. On the other hand, taxa such as *Fragilaria pectinalis* Lyngbye, *Nitzschia adamata* Hustedt, *Maya-maea agrestis* (Hustedt) Lange-Bertalot and *P. lagerstedtii* are very sensi-tive to soil moisture. Interestingly, *H. amphioxys*, the most generally known terrestrial diatom, has also a rather limited moisture range and is



**Fig. 7.** Derived autecological values for pH (A) and soil moisture (B) in comparison with indicator values assigned by [Dam et al. \(1994\)](#) (updated version). The width of the boxes indicates the number of species with smaller boxes yielding less species. **A:** 1, pH < 5.5; 2, pH < 7; 3, pH = 7; 4, mainly occurring on pH > 7; 5, exclusively occurring on pH > 7. **B:** 1, rarely occurring outside water bodies; 2, sometimes occurring on wet places; 3, regularly occurring on wet and moist places; 4, mainly occurring on terrestrial places; 5, almost exclusively occurring on terrestrial places.

generally not present on soils anymore when the moisture content reaches 35 to 40%. Overall, these autecological values will improve our knowledge of the ecology of terrestrial diatoms and enable us to better use these organisms as environmental markers.

Finally, we analysed if our inferred optimum values correspond to the categories for pH and moisture assigned by [Dam et al. \(1994\)](#) ([Fig. 7](#)). The latter indicates how likely it is that a species will occur in a terrestrial environment and at which prevailing moisture conditions. Concerning pH, we see that our optimum values correspond well with the different [Dam et al. \(1994\)](#) categories ( $F = 9.87$ ,  $df = 115$ ,  $R^2 = 0.23$ ,  $P < 0.001$ ). However, our autecological values indicate that it would be better to take a pH of 6.5 instead of 7 to differentiate the categories 2, 3 and 4 from each other when evaluating soils. Contrary to pH, there was no significant difference of our optimum values for VWC between the different categories ( $F = 0.35$ ,  $df = 81$ ,  $R^2 = -0.03$ ,  $P = 0.84$ ). This is interesting, since we expected that taxa categorised as aquatic require a higher threshold for soil moisture than terrestrial taxa to be present on terrestrial habitats. Now it seems that, independent of their assigned category, diatoms generally need a VWC of around 20% to grow and reproduce on soils.

#### 4. Discussion

For this research, we combined eight different datasets on terrestrial diatoms coming from four distinct regions, encompassing 516 soil samples from 166 different sites and covering 710 taxa. A first analysis of this data revealed significant relationships of species diversity and richness with pH and anthropic disturbance, while soil moisture was not important. These observations are in line with previous research. [Hoffmann \(1989\)](#) and [Lund \(1945\)](#) noted that diatoms generally prefer neutral to alkaline soils. However, both pointed out that pH is often correlated with other variables (e.g.  $\text{CaCO}_3$ ) and that it will change the nutrient availability in soils, which could also affect diatoms. Regarding different land uses, [Barragán et al. \(2018\)](#) and [Foets et al. \(2020a\)](#) observed that less disturbed areas were more species rich and diversified. The difference between the two extremes (agricultural field and forest) observed here was 150 taxa, doubling the number of species occurring on agricultural fields. Even though soil moisture is crucial for diatoms to grow and reproduce, recent studies indicated that soil moisture does not influence the community composition as such ([Foets et al., 2020a](#); [Zhang et al., 2020](#)). However, it does play a key role in the absolute diatom abundances (i.e. primary production) ([Foets et al.,](#)

[2020b](#)). Overall, these analyses confirm the main patterns in the variability of diatom communities and reveal a huge difference in species richness between undisturbed and disturbed areas.

In a next step of our analysis, we checked the sensitivity of diatom taxa for pH, VWC, TN, P, DOC and SOM. Despite that these variables all have been documented in having an effect on the community composition ([Antonelli et al., 2017](#); [Lund, 1945](#); [Vacht et al., 2014](#); [Van de Vijver et al., 2002](#); [Van Kerckvoorde et al., 2000](#)), we only found that the optimum values assigned for pH and soil moisture gave decent results with  $r^2$ -values of respectively 0.57 and 0.62. A side note here is that considerably less data has been included in the calculation for TN, P, DOC, SOM and C/N. Thus, adding more samples to the dataset may improve the model metrics for those variables. However, research by [Gremmen et al. \(2007\)](#) and [Van de Vijver et al. \(2002\)](#) on terrestrial diatoms revealed high  $r^2$ -values of 0.85 and 0.68 for respectively altitude and soil moisture, including less data points than we did. However, these studies only used data coming from anthropogenic undisturbed environments situated on sub-Antarctic islands. This latter factor is also very important, as our validation metrics improved considerably after removing the samples collected in anthropogenic disturbed habitats, whilst reaching values similar to [Van de Vijver et al. \(2002\)](#). This observation indicates that anthropogenic disturbance is perhaps the principal factor defining taxon occurrences. In order to improve the validation metrics for nitrogen, carbon and organic matter, it will be important that, apart from adding data, a significant part of the data should come from anthropogenic undisturbed sites. Only then, we will know if we could get acceptable autecological values for those variables and eventually apply them in future research and management.

After validating the calculated autecological values for pH and soil moisture, we compared them with previous research of [Lund \(1945\)](#) and [Van de Vijver et al. \(2002\)](#). Although pH was measured differently, the ranges seem to overlap well and there is no real indication that the ones established by [Lund \(1945\)](#) are constantly more acidic or alkali than ours, suggesting that the outcome of the two methods does indeed not seem to diverge much from each other. We noticed that the tolerance ranges for soil moisture calculated by [Van de Vijver et al. \(2002\)](#) tend to go more towards higher values and are generally broader. This is probably due to the selection and removal of the data, since the same observation is done for taxa exclusively occurring in sub-Antarctic samples. While [Van de Vijver et al. \(2002\)](#) included diatoms found on soils ranging between 0 and 100% saturation, we only selected sites with a maximum saturation of 75.5%. In addition, our samples generally had

Table 4

Optimum and tolerance values of pH and volumetric soil moisture content (VWC) for the most abundant soil diatoms. The diatom taxa, which occurred in at least 5 samples with an abundance of minimum 2.5%. The four-letter code is retrieved from OMNIDIA.

Diatom taxa	Code	pH optima	pH tolerance	VWC optima (%)	VWC tolerance (%)
<i>Adlafia bryophila</i>	ABRY	6.40	0.63	33.71	20.26
<i>Adlafia bryophiloides</i>	ABYD	NA	NA	40.66	13.77
<i>Achnanthydium eutrophilum</i>	ADEU	6.44	0.60	14.83	10.29
<i>Adlafia linearis</i>	ADFL	NA	NA	38.95	12.74
<i>Adlafia frenotii</i>	ADFR	NA	NA	14.86	13.67
<i>Achnanthydium kranzii</i>	ADKR	6.15	0.45	35.33	28.75
<i>Achnanthydium minutissimum</i>	ADMI	6.85	0.92	29.36	15.55
<i>Adlafia minuscula</i> var. <i>muralis</i>	ADMM	5.40	0.96	12.95	8.72
<i>Adlafia minuscula</i>	ADMS	6.00	0.81	15.51	10.32
<i>Achnanthydium pyrenaicum</i>	ADPY	NA	NA	24.69	6.82
<i>Achnanthydium hoffmannii</i>	AHOF	6.90	0.55	NA	NA
<i>Achnanthydium modestiforme</i>	AMDF	5.68	0.59	53.47	15.14
<i>Achnanthydium sieminskae</i>	ASIE	5.89	0.71	45.74	19.13
<i>Achnanthydium strictum</i>	ASTU	7.47	0.64	NA	NA
<i>Caloneis aerophila</i>	CAER	5.66	0.47	36.86	8.14
<i>Caloneis leptosoma</i>	CALE	6.39	0.82	NA	NA
<i>Chamaepinnularia aliena</i>	CALN	NA	NA	59.42	11.01
<i>Caloneis</i> sp.3	CALS3	6.79	1.00	NA	NA
<i>Chamaepinnularia australomediocris</i>	CAUM	NA	NA	46.80	21.72
<i>Caloneis</i> cf. <i>bacillum</i>	CBAC	6.47	0.72	34.31	13.54
<i>Caloneis fontinalis</i>	CFON	7.21	1.04	NA	NA
<i>Chamaepinnularia evanida</i>	CHEV	5.97	0.60	50.58	22.09
<i>Chamaepinnularia muscicola</i>	CHMC	6.60	0.95	44.19	19.18
<i>Chamaepinnularia obsoleta</i>	CHOB	5.97	0.59	15.41	12.94
<i>Chamaepinnularia parsura</i>	CHPP	NA	NA	53.38	20.24
<i>Cavinula intractata</i>	CITT	6.03	0.52	36.93	22.42
<i>Caloneis lancetula</i>	CLCT	6.99	0.65	NA	NA
<i>Craticula molestiformis</i>	CMLF	7.05	0.40	NA	NA
<i>Craticula minusculoides</i>	CMNO	7.01	0.64	27.29	8.55
<i>Caloneis molaris</i>	CMOL	7.12	0.75	NA	NA
<i>Cocconeis costata</i> var. <i>costata</i>	COCC	NA	NA	12.57	18.57
<i>Cocconeis</i> sp.	COCS	NA	NA	28.24	34.35
<i>Cocconeis pediculus</i>	CPED	7.19	0.62	NA	NA
<i>Cocconeis placentula</i> var. <i>lineata</i>	CPLI	6.04	0.60	NA	NA
<i>Diatomella balfouriana</i>	DBAL	6.15	0.60	34.67	18.81
<i>Humidophila biceps</i>	DBIC	5.73	0.55	33.15	16.82
<i>Denticula</i> sp.	DENS	NA	NA	30.62	0.00
<i>Distrionella germainii</i>	DIGM	6.20	0.69	41.73	15.24
<i>Diploneis subovalis</i>	DSBO	5.86	0.57	42.12	12.32
<i>Eunotia botuliformis</i>	EBOT	5.66	0.63	23.50	21.87
<i>Eonotia clotii</i>	ECLO	5.61	0.59	44.61	14.29
<i>Eunotia exigua</i>	EEXI	5.20	0.22	13.51	5.40
<i>Eunotia fallax</i>	EFAL	5.74	0.19	45.74	14.27
<i>Eunotia incisa</i>	EINC	NA	NA	59.51	14.65
<i>Eunotia minor</i>	EMIN	5.90	0.58	49.76	22.75
<i>Eunotia muscicola</i>	EMUS	NA	NA	46.75	16.54
<i>Eunotia palatina</i>	EPLT	5.23	0.45	12.19	5.79
<i>Eunotia praerupta</i>	EPRA	5.79	0.39	41.45	12.58
<i>Encyonema silesiacum</i>	ESLE	6.13	0.74	NA	NA
<i>Eunotia tenella</i>	ETEN	5.47	0.45	31.94	31.13
<i>Eunotia bidens</i>	EUBI	5.67	0.17	32.61	6.54
<i>Eunotia paludosa</i>	EUPA	NA	NA	43.27	22.18
<i>Frustulia cirisiae</i>	FCIR	6.60	0.38	24.54	11.17
<i>Fragilaria capucina</i> var. <i>capucina</i> morphotype 2	FCTW	6.22	0.78	53.13	16.85
<i>Fragilariforma virescens</i>	FFVI	5.86	0.53	60.39	11.12
<i>Fragilaria gracilis</i>	FGRA	6.72	0.73	27.09	12.98
<i>Frankophila maillardii</i>	FKMA	6.05	0.67	41.86	17.17
<i>Fragilaria nevadensis</i>	FNEV	NA	NA	62.93	0.00
<i>Fragilaria pectinalis</i>	FPPE	6.51	0.53	21.46	3.97
<i>Fistulifera pelliculosa</i>	FPEL	6.45	0.88	NA	NA
<i>Fragilariopsis kerguelensis</i>	FPKE	6.40	0.69	NA	NA
<i>Fragilaria rumpens</i>	FRUM	6.74	0.72	24.88	8.63
<i>Fragilaria vaucheriae</i>	FVAU	6.89	0.70	NA	NA
<i>Frustulia vulgaris</i>	FVUL	7.09	1.07	40.93	19.84
<i>Gomphonema affine</i>	GAFF	5.96	0.52	44.82	13.61
<i>Gomphonema angustatum</i>	GANG	6.49	0.65	23.47	7.99
<i>Gomphonema micropus</i>	GMIC	6.62	0.93	27.26	17.52
<i>Gomphonema minutum</i> f. <i>minutum</i>	GMIN	6.71	0.78	26.83	9.84
<i>Geissleria paludosa</i>	GPAL	6.20	0.73	30.94	12.47
<i>Gomphonema parvulum</i> var. <i>parvulum</i>	GPAR	6.38	0.79	NA	NA
<i>Gomphonema productum</i>	GPRO	5.90	0.61	62.10	6.52
<i>Gomphonema varioreduncum</i>	GVRD	6.61	0.76	NA	NA
<i>Hantzschia abundans</i>	HABU	6.83	0.73	19.38	11.87

(continued on next page)

Table 4 (continued)

Diatom taxa	Code	pH optima	pH tolerance	VWC optima (%)	VWC tolerance (%)
<i>Hantzschia amphioxys</i>	HAMP	6.47	0.66	20.54	10.28
<i>Humidophila brekkaensis</i>	HBRE	5.92	0.79	22.96	22.55
<i>Hantzschia calcifuga</i>	HCAL	6.83	0.66	35.22	19.18
<i>Humidophila comperei</i>	HCOP	6.66	0.38	17.14	16.21
<i>Hantzschia hyperborea</i>	HHYP	6.15	0.34	46.65	12.17
<i>Halamphora montana</i>	HLMO	7.52	0.52	24.53	9.13
<i>Humidophila nienta</i>	HNIE	6.25	0.61	40.93	14.65
<i>Humidophila perpusilla</i>	HPEP	5.86	0.36	43.10	11.89
<i>Humidophila arcuata</i>	HUAC	5.89	0.42	42.59	16.84
<i>Humidophila contenta</i> s.L.	HUCO	6.35	0.69	20.88	12.41
<i>Humidophila crozetikerguelensis</i>	HUCZ	5.59	0.84	37.52	13.62
<i>Humidophila ingeae</i>	HUIG	5.76	0.81	42.45	14.55
<i>Humidophila irata</i>	HUIR	7.03	0.61	20.49	8.38
<i>Humidophila latestriata</i>	HULT	NA	NA	39.13	8.17
<i>Humidophila subantarctica</i>	HUSA	6.23	0.49	28.71	21.73
<i>Humidophila vidalii</i>	HUVI	NA	NA	40.98	11.46
<i>Luticola acidoclinata</i>	LACD	6.34	0.64	16.29	8.24
<i>Luticola frequentissima</i>	LFRQ	6.62	0.59	19.38	11.04
<i>Lecohuia geniculata</i>	LGEN	6.03	0.70	41.64	15.19
<i>Lemnicola hungarica</i>	LHUN	NA	NA	26.73	7.11
<i>Luticola intermedia</i>	LINT	6.47	0.57	21.72	2.98
<i>Luticola muticopsis</i>	LMTP	NA	NA	20.74	14.04
<i>Luticola mutica</i> s.L.	LMUT	6.00	0.86	26.24	14.08
<i>Luticola nivalis</i>	LNIV	5.90	0.52	NA	NA
<i>Luticola paramutica</i>	LPAR	6.00	0.56	24.66	9.10
<i>Luticola robusta</i>	LROB	6.29	0.98	31.64	12.25
<i>Luticola ventricifusa</i>	LVCF	7.44	0.53	NA	NA
<i>Luticola ventricosa</i>	LVEN	7.44	0.65	16.25	5.06
<i>Mayamaea atomus</i>	MAAT	6.65	0.59	26.91	9.68
<i>Mayamaea</i> cf. <i>atomus</i> 2	MAAT2	6.84	0.24	26.12	14.45
<i>Microcostatus aerophilus</i>	MAER	6.08	0.59	18.00	12.25
<i>Mayamaea excelsa</i>	MAEX	6.69	0.56	20.18	9.77
<i>Mayamaea fossalis</i>	MAFO	6.41	0.40	31.49	14.07
<i>Mayamaea agrestis</i>	MAGR	6.83	0.45	22.78	4.78
<i>Mayamaea alcimonica</i>	MALC	6.61	0.65	24.64	9.58
<i>Mayamaea asellus</i>	MASE	6.57	0.84	NA	NA
<i>Mayamaea terrestris</i>	MAYT	7.22	0.64	NA	NA
<i>Meridion circulare</i> var. <i>circulare</i>	MCIR	6.81	0.30	18.20	9.08
<i>Meridion constrictum</i>	MCON	6.21	0.58	NA	NA
<i>Melosira</i> sp.2	MELS2	NA	NA	34.47	20.41
<i>Mayamaea fossalis</i> var. <i>obsidialis</i>	MFOB	6.46	0.96	17.22	12.03
<i>Mayamaea fossiloides</i>	MFSL	7.18	0.59	31.19	8.46
<i>Melosira guillaumini</i>	MGUI	NA	NA	24.99	32.65
<i>Microcostatus</i> sp.1	MISP1	6.56	0.84	19.73	8.04
<i>Microcostatus krasskei</i>	MKRA	5.35	0.78	17.35	17.25
<i>Mayamaea permissis</i>	MPMI	6.81	0.70	24.69	8.43
<i>Mayamaea recondita</i>	MRCD	6.40	0.65	28.02	20.40
<i>Muelleria gibbula</i>	MUGI	6.80	0.81	16.04	9.91
<i>Muelleria terrestris</i>	MUTE	7.04	1.34	19.25	8.91
<i>Muelleria varipunctata</i>	MVPU	6.72	0.30	12.22	10.03
<i>Nitzschia acidoclinata</i>	NACD	6.65	0.70	25.36	13.45
<i>Neidium alpinum</i>	NALP	5.57	0.49	NA	NA
<i>Nitzschia amphibia</i>	NAMP	6.87	0.81	21.15	7.10
<i>Navicula bicephala</i>	NBIC	6.34	0.47	33.59	19.53
<i>Navicula cincta</i> var. <i>heutleri</i>	NCIH	7.47	0.46	21.52	9.60
<i>Navicula cryptocephala</i>	NCRY	6.22	0.49	NA	NA
<i>Navicula cryptotenella</i>	NCTE	7.42	0.60	NA	NA
<i>Nitzschia dissipata</i> var. <i>media</i>	NDME	6.50	0.63	16.84	9.15
<i>Navicula exilis</i>	NEXI	6.63	0.84	NA	NA
<i>Navicula fritschii</i>	NFRH	7.01	0.62	25.56	15.44
<i>Navicula frugalis</i>	NFRU	7.28	0.76	NA	NA
<i>Navicula gregaria</i>	NGRE	6.78	0.83	30.94	19.28
<i>Nitzschia hantzschiana</i>	NHAN	6.54	0.63	21.50	6.35
<i>Nitzschia frustulum</i>	NIFR	7.26	0.74	NA	NA
<i>Nitzschia harderi</i>	NIHD	7.22	0.70	26.12	12.00
<i>Nitzschia perminuta</i>	NIPM	6.15	0.63	28.00	10.78
<i>Nitzschia pusilla</i>	NIPU	6.79	0.53	25.10	9.36
<i>Nitzschia solita</i>	NISO	NA	NA	30.62	0.00
<i>Navicula lanceolata</i>	NLAN	7.49	0.68	NA	NA
<i>Navicula moenofranconica</i>	NMFC	7.72	0.54	NA	NA
<i>Nitzschia obtusangula</i>	NOTA	7.83	0.41	NA	NA
<i>Nitzschia palea</i> var. <i>debilis</i>	NPAD	6.50	0.67	22.19	12.32
<i>Nitzschia paleacea</i>	NPAE	6.28	0.58	21.69	7.79
<i>Nitzschia palea</i> var. <i>palea</i>	NPAL	6.72	0.63	21.83	8.40
<i>Nitzschia</i> cf. <i>palea</i>	NPALc	6.54	0.38	12.02	7.45
<i>Nitzschia palea</i> var. <i>tenuirostris</i>	NPAT	6.68	0.53	22.29	6.65

(continued on next page)

Table 4 (continued)

Diatom taxa	Code	pH optima	pH tolerance	VWC optima (%)	VWC tolerance (%)
<i>Nitzschia solgensis</i>	NSOL	7.47	0.70	24.10	11.11
<i>Navicula tenelloides</i>	NTEN	7.40	0.79	NA	NA
<i>Navicula trivialis</i>	NTRV	7.57	0.51	NA	NA
<i>Navicula vaucheriae</i>	NVAU	7.20	0.47	24.27	8.17
<i>Navicula veneta</i>	NVEN	7.15	0.51	25.47	6.99
<i>Nitzschia adamata</i>	NZAD	NA	NA	30.43	2.24
<i>Nitzschia supralitorea</i>	NZSU	7.20	0.69	20.79	5.53
<i>Odontidium mesodon</i>	OMES	6.41	0.17	16.86	6.45
<i>Orthoseira roeseana</i>	OROE	5.23	1.19	29.36	11.30
<i>Psammothidium abundans</i>	PABD	6.13	0.50	40.58	16.90
<i>Pinnularia acidicola</i>	PADC	6.12	0.54	51.36	19.16
<i>Pinnularia angliciformis</i>	PAGF	6.07	0.54	56.13	21.44
<i>Pinnularia appendiculata</i> var. <i>irrorata</i>	PAIR	5.74	0.39	NA	NA
<i>Psammothidium aretasii</i>	PATS	6.34	0.42	31.46	15.28
<i>Planothidium aueri</i>	PAUE	5.92	0.67	37.64	18.92
<i>Pinnularia brebissonii</i> var. <i>bicuneata</i>	PBBB	6.03	0.97	17.30	6.52
<i>Pinnularia bullacostae</i>	PBLC	6.29	0.64	19.67	7.09
<i>Pinnularia borealis</i>	PBOR	5.98	0.59	28.20	15.30
<i>Pinnularia bottmica</i>	PBOT	5.99	0.84	31.80	18.66
<i>Pinnularia brebissonii</i>	PBRE	6.09	1.00	NA	NA
<i>Pinnularia carteri</i>	PCAT	6.07	0.48	46.80	14.00
<i>Psammothidium confusum</i>	PCFU	5.88	0.61	49.16	20.64
<i>Pinnularia cuneoestrata</i>	PCNR	6.44	0.52	20.65	13.37
<i>Pinnularia crozetii</i>	PCRO	NA	NA	36.47	16.54
<i>Planothidium cyclophorum</i>	PCYC	6.38	0.66	41.64	19.57
<i>Parlibellus delognei</i>	PDEL	7.02	0.53	31.20	28.42
<i>Pinnularia domestica</i>	PDMS	6.41	0.90	NA	NA
<i>Pinnularia divergentissima</i>	PDVG	6.06	0.63	48.33	20.55
<i>Placoneis elginensis</i>	PELG	7.72	1.14	NA	NA
<i>Pinnularia frauenbergiana</i> var. <i>caloneiopsis</i>	PFCA	7.02	1.13	57.91	0.00
<i>Pinnularia frequentis</i>	PFQT	NA	NA	57.91	0.00
<i>Psammothidium germainii</i>	PGER	6.76	0.44	25.13	17.20
<i>Psammothidium incognitum</i>	PICO	5.95	0.55	35.42	14.70
<i>Psammothidium investians</i>	PINV	NA	NA	52.18	18.02
<i>Pinnularia isselana</i>	PISS	7.66	0.99	NA	NA
<i>Pinnularia kolbei</i>	PKOL	6.42	0.89	24.73	15.50
<i>Pinnularia lagerstedtii</i>	PLAG	5.97	0.65	36.50	5.64
<i>Planothidium frequentissimum</i>	PLFR	7.46	0.87	26.09	12.06
<i>Placoneis hambergii</i>	PLHA	6.44	0.79	NA	NA
<i>Placoneis ignorata</i>	PLIG	7.02	0.65	NA	NA
<i>Psammothidium manguinii</i>	PMAN	6.17	0.54	28.56	15.72
<i>Pinnularia microstauron</i> var. <i>angusta</i>	PMIA	6.38	0.56	18.24	7.91
<i>Pinnularia microstauron</i> var. <i>microstauron</i>	PMIC	6.00	0.84	21.22	12.15
<i>Pinnularia microstauron</i> var. <i>nonfasciata</i>	PMNF	5.48	0.39	58.21	12.61
<i>Pinnularia molaris</i>	PMOL	6.19	0.60	17.20	7.50
<i>Pinnularia obscuriformis</i>	POBF	6.46	0.27	39.95	26.16
<i>Pinnularia obscura</i>	POBS	6.34	0.66	25.82	14.12
<i>Pinnularia petersenii</i>	PPET	6.67	0.44	27.66	20.60
<i>Pinnularia perirrorata</i>	PPRI	5.46	0.75	43.38	26.61
<i>Pinnularia</i> aff. <i>perirrorata</i>	PPRIa	6.58	0.92	41.91	18.21
<i>Planothidium quadripunctatum</i>	PQDP	NA	NA	42.44	20.38
<i>Pinnularia rabenhorstii</i>	PRAB	NA	NA	34.56	9.69
<i>Pinnularia rhombarea</i> var. <i>serrata</i>	PRHS	6.35	0.82	37.41	20.04
<i>Pinnularia rabenhorstii</i> var. <i>subantarctica</i>	PRSA	NA	NA	13.82	15.38
<i>Pinnularia rupestris</i>	PRUP	5.46	0.15	43.71	14.39
<i>Psammothidium subatomoides</i>	PSAT	6.59	0.54	27.98	13.35
<i>Pinnularia subcapitata</i> var. <i>subcapitata</i>	PSCA	6.26	0.51	NA	NA
<i>Pinnularia subcapitata</i> f. <i>typica</i>	PSCTx	6.04	0.45	20.53	12.51
<i>Pseudostaurosira elliptica</i>	PSDE	6.48	0.32	36.98	8.92
<i>Pinnularia schimanskii</i>	PSHK	6.26	0.52	18.53	13.41
<i>Pinnularia schoenfelderi</i>	PSHO	5.88	0.80	18.37	13.14
<i>Pinnularia silvatica</i>	PSIL	NA	NA	57.25	17.59
<i>Pinnularia sinistra</i>	PSIN	6.00	0.55	59.92	2.57
<i>Pseudostaurosira naveana</i>	PSNA	6.51	0.51	31.84	17.94
<i>Pinnularia subrostrata</i>	PSRO	5.54	0.19	34.37	6.15
<i>Pinnularia subrupestris</i>	PSRU	6.97	0.61	40.18	19.58
<i>Psammothidium stauroneioides</i>	PSTD	6.36	0.49	25.77	14.01
<i>Pinnularia stricta</i>	PSTI	6.29	0.60	15.11	10.66
<i>Pinnularia subantarctica</i> var. <i>elongata</i>	PSUE	6.01	0.57	54.62	15.08
<i>Planothidium lanceolatum</i>	PTLA	6.66	0.79	39.56	21.15
<i>Pinnularia viridis</i> var. <i>sudetica</i>	PVSU	6.78	0.72	28.61	12.63
<i>Rhoicosphenia abbreviata</i>	RABB	7.35	0.59	NA	NA
<i>Rhopalodia rupestris</i>	RRUP	NA	NA	37.02	15.72
<i>Reimeria sinuata</i>	RSIN	6.20	0.66	NA	NA
<i>Surirella angusta</i> var. <i>constricta</i>	SACO	6.75	0.59	38.06	11.08
<i>Surirella angusta</i>	SANG	7.09	0.89	NA	NA

(continued on next page)

Table 4 (continued)

Diatom taxa	Code	pH optima	pH tolerance	VWC optima (%)	VWC tolerance (%)
<i>Stauroneis borrichii</i>	SBOR	6.20	0.89	25.28	11.36
<i>Stauroneis</i> aff. <i>Borrichii</i>	SBORa	7.39	0.70	17.16	8.59
<i>Sellaphora atomoides</i>	SEAT	6.65	0.66	28.01	15.95
<i>Sellaphora elorantana</i>	SELO	6.39	0.95	43.55	18.95
<i>Sellaphora lundii</i>	SELU	7.40	0.55	NA	NA
<i>Sellaphora nana</i>	SENA	6.56	0.30	26.99	8.58
<i>Stauriforma exiguiformis</i>	SEXG	6.31	0.51	22.40	13.17
<i>Sellaphora hustedii</i>	SHUS	NA	NA	57.91	0.00
<i>Simonsenia delognei</i>	SIDE	7.11	0.72	21.60	7.58
<i>Stauroneis lecohui</i>	SLEC	NA	NA	41.60	18.27
<i>Staurisirella leptostauron</i>	SLEP	6.16	0.53	36.15	16.02
<i>Stauroneis minutula</i>	SMNT	6.56	0.62	NA	NA
<i>Stauroneis muriella</i>	SMUR	7.31	0.89	NA	NA
<i>Surirella ovata</i>	SOVA	7.07	0.46	20.94	6.48
<i>Stauroneis</i> cf. <i>pseudoagrestis</i>	SPAGx	7.10	0.61	26.80	12.13
<i>Stauroneis pseudoagrestis</i>	SPDA	6.59	1.02	NA	NA
<i>Stauroneis parathermicola</i>	SPTH	6.47	0.87	29.24	20.87
<i>Sellaphora saugeresii</i>	SSGE	6.31	0.76	50.34	17.40
<i>Sellaphora subseminulum</i>	SSSE	6.72	0.59	30.07	17.83
<i>Surirella terricola</i>	STER	7.36	0.46	25.50	8.97
<i>Stauroneis thermicola</i>	STHE	6.38	0.83	26.72	16.17
<i>Stauroneis kriegeri</i>	STKR	6.62	0.88	NA	NA
<i>Surirella minuta</i>	SUMI	7.19	0.73	NA	NA
<i>Tryblionella debilis</i>	TDEB	6.88	0.91	15.68	10.59

a lower volumetric moisture content averaging 28%, whereas Van de Vijver et al. (2002) had an average content of around 45%. Considering those differences, the ranges of moisture (and pH) overlap very well and show that data selection plays an important role in calculating indicator values.

Apart from comparing the tolerance ranges of pH and soil moisture, we also checked whether the updated Dam et al. (1994) classification is in accordance with those optimum values. We found that diatom taxa occurring in both aquatic and terrestrial environments have similar preferences in both environments, meaning that the categories for pH are still useful in terrestrial settings. However, if used for soils, the threshold value between categories 2, 3 and 4 should be set around 6.5 instead of 7 to have a better interpretation. Contrary to pH, the categories for soil moisture did not work for terrestrial environments and results indicate that taxa, either assigned as rather aquatic or terrestrial by Dam et al. (1994), can (frequently) occur on soils if the moisture concentration is minimum 20%. A similar observation was reported by Stokes (1940), who found that algae function and grow best when 40 to 60% of moisture-holding capacity of the soil is reached. Besides, this outcome also affects the use of diatoms as hydrological tracers, since in those studies Dam et al. (1994) categories 4 and 5 were used to classify diatoms as being 'terrestrial' (Pfister et al., 2017, 2009). Knowing that diatoms from other categories also regularly occur on soils, we should revise the classification, which eventually would lead to the inclusion of more diatoms that can be used for tracing hydrological connectivity. Although the classification of Dam et al. (1994) has proven to be very useful, it does not always provide the correct ecological answers for diatoms present on soils.

As mentioned before, adding data, independent of the location, to the existing dataset might still improve the validation and optimum values for the investigated variables provided that the community data is taxonomically consistent. This is, however, not easy since taxonomy in diatom research changes constantly and verification is rather time-consuming and not always possible. In this study, we could not verify everything and due to this inconsistency, it is possible that (much) 'noise' ended up in the results. However, since terrestrial diatoms are far less studied than strictly aquatic diatoms, taxonomical changes occur in a slower pace. Very often, old diatom publications such as Ettl and Gärtner (1995), which is a compilation of terrestrial diatom studies in Europe and Lund (1945) and Lund (1946), among other references

mentioned earlier are still consulted for identification, making taxonomical harmonization easier between the different datasets. Noise will also result from differences in soil moisture between the soil moisture observation and the diatom sample, since soil moisture can be highly variable even at small distances (Teuling and Troch, 2005). Also, we did not account for (pseudo-)cryptic species (i.e. genetically different species, but morphologically (almost) undistinguishable) such as *P. borealis* and *H. amphioxys* (Maltsev and Kulikovskiy, 2017; Pinseel et al., 2020, 2019; Souffreau et al., 2013). Both are known to have developed different adaptations and/or tolerances to certain climatic conditions (Souffreau, 2011; Souffreau et al., 2010). As they are widespread, common diatoms, we can imagine that many more cosmopolitan species such as *H. contenta*, *H. abundans*, *P. obscura* are also (pseudo-)cryptic. A way to solve this is to switch to molecular techniques (i.e. DNA metabarcoding and High-Throughput Sequencing) for community analysis. These techniques are seen as a fast, efficient and low-cost solution to the rather time-consuming microscopic diatom identifications (Rivera et al., 2020; Vasselon et al., 2017). It will also enable us to increase sampling frequency and to combine and link different soil organisms with each other, paving the way for monitoring programmes and other related applications (Fløjgaard et al., 2019; Orgiazzi et al., 2015). Moreover, it will be also a solution for the cryptic diversity and thus it will still be possible to combine datasets from different climatic regions.

Another point that we should be aware of when collecting data for intercomparing purposes is to align and standardise the soil sampling and analysis of environmental variables. Due to this, we were not able to fully utilise our combined dataset. Of the different variables included here, we believe more focus should be on soil organic matter, since an encouraging  $r^2$  of 0.29 was obtained based on a limited number of only 30 samples. Besides, it is strongly related to anthropic disturbances and the moisture-holding capacity of the soil (Hudson, 1994; Stanek-Tarkowska et al., 2018a), both of which diatoms are responsive to and likely sensitive as well. Furthermore, in regards of their potential usage as a measure for soil quality, SOM is often seen as an important indicator (Bünemann et al., 2018) and it will therefore be interesting to explore this variable more in relation to diatoms.

## 5. Conclusions

In this study, we defined autecological values for pH and soil moisture content for the most common, widespread soil diatoms and compared them with previous research of Lund (1945), Dam et al. (1994) and Van de Vijver et al. (2002). Besides our results showing similarities with those studies, they also indicated a significant improvement on the existing indicator values. Moreover, we expanded the list of terrestrial diatoms to 249 taxa to be used as environmental markers in different research fields. We believe that future studies should focus on molecular techniques as it will be possible to speed up the identification process and discriminate better between cryptic species. Additionally, soil sample collection and nutrient analysis should be standardised in order to enable a better pooling of data concerning diatom ecology in the future.

## 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.

## Acknowledgements

The authors thank Mrs A. Mertens and Dr H. van Dam for providing updated indicator values. They are also grateful to Dr G. de Deyn for critically reviewing a first version of the manuscript. Two reviewers are thanked for their detailed and constructive comments that improved the manuscript. Dr A. Van Kerckvoorde is thanked for providing his dataset. The work was supported by the Luxembourg National Research Fund (FNR) (PRIDE15/10623093/HYDRO-CSI).

## Appendix A

## Appendix B. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ecolind.2020.107248>.

## References

- Antonelli, M., Wetzel, C.E., Ector, L., Teuling, A.J., Pfister, L., 2017. On the potential for terrestrial diatom communities and diatom indices to identify anthropic disturbance in soils. *Ecol. Ind.* 75, 73–81. <https://doi.org/10.1016/j.ecolind.2016.12.003>.
- Barragán, C., Wetzel, C.E., Ector, L., 2018. A standard method for the routine sampling of terrestrial diatom communities for soil quality assessment. *J. Appl. Phycol.* 30 (2), 1095–1113.
- Beck, H.E., Zimmermann, N.E., McVicar, T.R., Vergopolan, N., Berg, A., Wood, E.F., 2018. Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Sci. Data* 5, 1–12. <https://doi.org/10.1038/sdata.2018.214>.
- Birks, H.J.B., Line, J.M., Juggins, S., Stevenson, A.C., Ter Braak, C.J.F., 1990. Diatoms and pH reconstruction. *Philos. Trans. - R Soc. London, B* 327, 263–278. <https://doi.org/10.1098/rstb.1990.0062>.
- Birks, H.J.B., Lotter, A.F., Juggins, S., Smol, J.P., 2012. In: *Tracking Environmental Change using Lake Sediments: Data Handling and Numerical Techniques*. Springer. <https://doi.org/10.1007/978-94-007-2745-8>.
- Brendemuhl, I., 1949. Über die Verbreitung der Erddiatomeen. *Archiv Mikrobiol* 14 (3), 407–449.
- Bünemann, E.K., Bongiorno, G., Bai, Z., Creamer, R.E., De Deyn, G., de Goede, R., Fleksens, L., Geissen, V., Kuypers, T.W., Mäder, P., Pulleman, M., Sukkel, W., van Groenigen, J.W., Brussaard, L., 2018. Soil quality – A critical review. *Soil Biol. Biochem.* 120, 105–125. <https://doi.org/10.1016/j.soilbio.2018.01.030>.
- Carayon, D., Eulin-Garrigue, A., Vigouroux, R., Delmas, F., 2020. A new multimetric index for the evaluation of water ecological quality of French Guiana streams based on benthic diatoms. *Ecol. Ind.* 113, 106248.
- Carayon, D., Tison-Rosebery, J., Delmas, F., 2019. Defining a new autoecological trait matrix for French stream benthic diatoms. *Ecol. Ind.* 103, 650–658. <https://doi.org/10.1016/j.ecolind.2019.03.055>.
- Cemagref, 1982. Etude des méthodes biologiques d'appréciation quantitative de la qualité des eaux.

- Coles, A.E., Wetzel, C.E., Martínez-Carreras, N., Ector, L., McDonnell, J.J., Frenress, J., Klaus, J., Hoffmann, L., Pfister, L., 2016. Diatoms as a tracer of hydrological connectivity: are they supply limited? *Ecology* 9, 631–645. <https://doi.org/10.1002/eco.1662>.
- Dixit, S.S., Smol, J.P., Kingston, J.C., Charles, D.F., 1992. ES&T diatoms: powerful indicators of environmental change. *Environ. Sci. Technol.* 26 (1), 22–33.
- Ettl, H., Gärtner, G., 1995. In: *Syllabus der Boden-, Luft- und Flechtenalgen*, first ed. Springer Spektrum. <https://doi.org/10.1007/978-3-642-39462-1>.
- Fløjgaard, C., Frøsløv, T.G., Brunbjerg, A.K., Bruun, H.H., Moeslund, J., Hansen, A.J., Ejrnæs, R., 2019. Predicting provenance of forensic soil samples: Linking soil to ecological habitats by metabarcoding and supervised classification. *PLoS One* 14. <https://doi.org/10.1371/journal.pone.0202844>.
- Foets, J., Wetzel, C.E., Teuling, A.J., Laurent, P., 2020a. Temporal and spatial variability of terrestrial diatoms at the catchment scale: controls on communities. *PeerJ* 1–19. <https://doi.org/10.7717/peerj.8296>.
- Foets, J., Wetzel, C.E., Teuling, A.J., Pfister, L., 2020b. Temporal and spatial variability of terrestrial diatoms at the catchment scale: controls on productivity and comparison with other soil algae. *PeerJ* 8, 1–21. <https://doi.org/10.7717/peerj.9198>.
- Gremmen, N.J.M., van de Vijver, B., Frenot, Y., Lebouvier, M., 2007. Distribution of moss-inhabiting diatoms along an altitudinal gradient at sub-Antarctic Îles Kerguelen. *Antarct. Sci.* 19 (1), 17–24.
- Haines, T.A., Akielaszek, J.J., Norton, S.A., Davis, R.B., 1983. Errors in pH measurement with colorimetric indicators in low alkalinity waters. *Hydrobiologia* 107 (1), 57–61.
- Heger, T.J., Straub, F., Mitchell, E.A.D., 2012. Impact of farming practices on soil diatoms and testate amoebae: a pilot study in the DOK-trial at Therwil, Switzerland. *Eur. J. Soil Biol.* 49, 31–36. <https://doi.org/10.1016/j.ejsobi.2011.08.007>.
- Hoffmann, L., 1989. Algae of terrestrial habitats. *Bot. Rev.* 55 (2), 77–105.
- Houba, V.J.G., Temminghoff, E.J.M., Gaikhorst, G.A., van Vark, W., 2000. Soil analysis procedures using 0.01 M calcium chloride as extraction reagent. *Commun. Soil Sci. Plant Anal.* 31 (9–10), 1299–1396.
- Hudson, B.D., 1994. Soil organic matter and available water capacity. *J. Soil Water Conserv.* 49, 189–194.
- Hustedt, F., 1942. Aerophile Diatomeen in der nordwestdeutschen Flora. *Ber. Dtsch. Bot. Ges.* 60, 55–73. <https://doi.org/10.1111/j.1438-8677.1942.tb00439.x>.
- Juggins, S., 2017. rioja: Analysis of Quaternary Science Data. R package version (0.9–21).
- Kabala, C., Muszyfaga, E., Galka, B., Łabuńska, D., Mańczyńska, P., 2016. Conversion of soil pH 1:2.5 KCl and 1:2.5 H<sub>2</sub>O to 1:5 H<sub>2</sub>O: conclusions for soil management, environmental monitoring, and international soil databases. *Polish J. Environ. Stud.* 25, 647–653. <https://doi.org/10.15244/pjoes/61549>.
- Krammer, K., 2000. Diatoms of Europe: Diatoms of the European inland WATERS and Comparable Habitats, The Genus Pinnularia. ARG Gantner Verlag KG, Ruggell.
- H. Lange-Bertalot 2001. Diatoms of Europe: Diatoms of the European inland waters and comparable habitats, Navicula sensu stricto; 10 genera separated from Navicula sensu lato Frustulia. Gantner Verlag KG, Ruggell. vol. 2.
- Lange-Bertalot, H., Bak, M., Witowski, A., Tagliaventi, N., 2011. Eunotia and some related genera. In: Lange-Bertalot, H. (Ed.), *Diatoms of Europe*. A.R.G. Gantner Verlag K.G., Ruggell.
- Lange-Bertalot, H.; Hofmann, G.; Werum, M. & Cantonati, M. (2017): Freshwater Benthic Diatoms of Central Europe: Over 800 Common Species Used in Ecological Assessments. English edition with updated taxonomy and added species. – 942 pp., Koeltz Botanical Books, Schmittens-Oberreifenberg.
- Lavoie, I., Hamilton, P.B., Wang, Y.-K., Dillon, P.J., Campeau, S., 2009. A comparison of stream bioassessment in Quebec (Canada) using six European and North American diatom-based indices. *Nov Hedwigia, Beih.* 135, 37–56.
- Lecointe, C., Coste, M., Prygiel, J., 1993. “Omnia”: software for taxonomy, calculation of diatom indices and inventories management. *Hydrobiologia* 269, 509–513. <https://doi.org/10.1007/BF00028048>.
- Lenoir, A., Coste, M., 1996. Development of a practical diatom index of overall water quality applicable to the French Water Board network. In: Rott, E. (Ed.), *Proceedings of International Symposium Use of Algae for Monitoring Rivers II*. Institut für Botanik, Universität Innsbruck, pp. 29–45.
- Levkov, Z., Metzeltin, D., Pavlov, A., 2013. Lenticular and Lenticularis. In: Lange-Bertalot, H. (Ed.), *Diatoms of Europe*, Vol. 7. Koeltz Scientific Books, Königstein.
- Lund, J.W.G., 1945. Observations on Soil Algae. I. The ecology, size and taxonomy of British soil diatoms (part 1). *New Phytol.* 44 (2), 196–219.
- Lund, J.W.G., 1946. Observations on soil algae. I. The ecology, size and taxonomy of British soil diatoms (part 2). *New Phytol.* 45 (1), 56–110.
- Maltsev, Y.I., Kulikovskiy, M.S., 2017. Morphological and genetic variability of *Hantzschia amphioxys* (Bacillariophyceae) in terrestrial and aquatic habitats. *Bot. Zhurnal* 102, 1–12.
- Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGinn, D., Minchin, P. R., O'Hara, R.B., Simpson, G.L., Solyomos, P., Stevens, M., Szoecs, E., Wagner, H., 2019. *Vegan: community ecology package*. R package version 2.5-6.
- Orgiazzi, A., Dunbar, M.B., Panagos, P., de Groot, G.A., Lemanceau, P., 2015. Soil biodiversity and DNA barcodes: opportunities and challenges. *Soil Biol. Biochem.* 80, 244–250. <https://doi.org/10.1016/j.soilbio.2014.10.014>.
- Petersen, J.B., 1915. Studier over danske aerofile alger, Kongel Danske Vidensk Selsk Skr. Naturvidensk Math Afd 12, 271–380.
- Petersen, J.B., 1928. The aerial Algae of Iceland. In: *The Botany of Iceland*. pp 325–447.
- Petersen, J.B., 1935. *Studies on the Biology and Taxonomy of Soil Algae*. Copenhagen, C. A. Reitzels Forlag.
- Pfister, L., McDonnell, J.J., Wrede, S., Hlubíková, Daša, Matgen, P., Fenicia, F., Ector, L., Hoffmann, L., 2009. The rivers are alive: on the potential for diatoms as a tracer of water source and hydrological connectivity. *Hydrol. Process.* 23 (19), 2841–2845.

- Pfister, L., Wetzel, C.E., Klaus, J., Martínez-Carreras, N., Antonelli, M., Teuling, A.J., McDonnell, J.J., 2017. Terrestrial diatoms as tracers in catchment hydrology: a review. *Wiley Interdiscip. Rev. Water*.
- Pinseel, E., Janssens, S.B., Verleyen, E., Vanormelingen, P., Kohler, T.J., Biersma, E.M., Sabbe, K., Van de Vijver, B., Vyverman, W., 2020. Global radiation in a rare biosphere soil diatom. *Nat. Commun.* 11 (1).
- Pinseel, E., Kulichová, J., Scharfen, V., Urbánková, P., Van de Vijver, B., Vyverman, W., 2019. Extensive cryptic diversity in the terrestrial diatom *Pinnularia borealis* (Bacillariophyceae). *Protist* 170 (2), 121–140.
- Poikane, S., Salas Herrero, F., Kelly, M.G., Borja, A., Birk, S., van de Bund, W., 2020. European aquatic ecological assessment methods: a critical review of their sensitivity to key pressures. *Sci. Total Environ.* 740, 140075.
- Reichardt, E., 2008. Bemerkenswerte Diatomeenfunde aus Bayern VI. *Hantzschia hyperborea* (Grun.) Lange-B. *Ber. Bayer Bot. Ges.* 78, 17–22.
- Reichardt, E., 2012. Neue Diatomeen (Bacillariophyceae) aus dem Gebiet der Stadt Treuchtlingen. *Ber. Bayer Bot. Ges.* 82, 19–32.
- Rivera, S.F., Vasselon, V., Bouchez, A., Rimet, F., 2020. Diatom metabarcoding applied to large scale monitoring networks: optimization of bioinformatics strategies using Mothur software. *Ecol. Ind.* 109, 1–13. <https://doi.org/10.1016/j.ecolind.2019.105775>.
- Simpson, G.L., Oksanen, J., 2020. Analogue Methods in Palaeoecology: Using the analogue Package. R package version 0.17-4.
- Smol, J.P., Birks, H.J.B., Last, W.M., 2001. Sedimentary pigments. In: Smol, J.P., Birks, H.J.B., Last, W.M. (Eds.), *Tracking Environmental Change Using Lake Sediments. Volume 3: Terrestrial, Algal, and Siliceous Indicators*. Kluwer Academic Publishers, Dordrecht, pp. 295–327. <https://doi.org/10.1007/0-306-47668-1>.
- Smol, J.P., Stoermer, E.F., 2010. *The Diatoms: Applications for the Environmental and Earth Sciences*. Cambridge University Press.
- Souffreau, C., 2011. *Biogeography, Species Diversity and Stress Tolerance of Aquatic and Terrestrial Diatoms*. Ghent university.
- Souffreau, C., Vanormelingen, P., Van de Vijver, B., Isheva, T., Verleyen, E., Sabbe, K., Vyverman, W., 2013. Molecular Evidence for Distinct Antarctic Lineages in the Cosmopolitan Terrestrial Diatoms *Pinnularia borealis* and *Hantzschia amphioxys*. *Protist* 164 (1), 101–115.
- Souffreau, C., Vanormelingen, P., Verleyen, E., Sabbe, K., Vyverman, W., 2010. Tolerance of benthic diatoms from temperate aquatic and terrestrial habitats to experimental desiccation and temperature stress. *Phycologia* 49 (4), 309–324.
- Stanek-Tarkowska, J., Czyz, E.A., Dexter, A.R., Sławiński, C., 2018a. Effects of reduced and traditional tillage on soil properties and diversity of diatoms under winter wheat. *Int. Agrophys.* 32, 403–409. <https://doi.org/10.1515/intag-2017-0016>.
- Stanek-Tarkowska, J., Noga, T., 2012. Diversity of diatoms (Bacillariophyceae) in the soil under traditional tillage and reduced tillage. *Inżynieria Ekol.* 287–296.
- Stanek-Tarkowska, J., Noga, T., Kochman-Kędzióra, N., Peszek, Ł., Pajączek, A., Kozak, E., 2015. The diversity of diatom assemblages developed on fallow soil in Pogórska Wola near Tarnów (southern Poland). *Acta Agrobot.* 68, 33–42. <https://doi.org/10.5586/aa.2015.011>.
- Stanek-Tarkowska, J., Noga, T., Pajączek, A., Peszek, U., 2013. The occurrence of *Sellaphora nana* (Hust.) Lange-Bert. Cavacini, Tagliaventi & Alfinito, *Stauroneis borichii* (J.B. Petersen) J.W.G. Lund, *S. parathermicola* (J.B. Petersen) J.W.G. Lund on agricultural soils. *Arch. Hydrobiol. Suppl. Algal. Stud.* 142, 109–119. <https://doi.org/10.1127/1864-1318/2013/0114>.
- Stokes, J.L., 1940. The influence of environmental factors upon the development of algae and other micro-organisms in soil. *Soil Sci.* 49 (3), 171–184.
- Taylor, J.C., Harding, W.R., Archibald, C.G.M., 2007. An illustrated guide to some common diatom species from South Africa. Report to the Water Research Commission.
- Braak, C.J.F., Looman, C.W.N., 1986. Weighted averaging, logistic regression and the Gaussian response model. *Vegetatio* 65 (1), 3–11.
- ter Braak, C.J.F., Šmilauer, P., 2002. *CANOCO Reference Manual and CanoDraw for Windows User's Guide: Software for Canonical Community Ordination (version 4.5)*. (Microcomputer Power), Ithaca NY, USA.
- Teuling, A.J., Troch, P.A., 2005. Improved understanding of soil moisture variability dynamics. *Geophys. Res. Lett.* 32, 1–4. <https://doi.org/10.1029/2004GL021935>.
- Vacht, P., Puusepp, L., Koff, T., Reitalu, T., 2014. Variability of riparian soil diatom communities and their potential as indicators of anthropogenic disturbances. *Est. J. Ecol.* 63, 168–184. <https://doi.org/10.3176/eco.2014.3.04>.
- Dam, H., Mertens, A., Sinkeldam, J., 1994. A coded checklist and ecological indicator values of freshwater diatoms from the Netherlands. *Netherl. J. Aquat. Ecol.* 28 (1), 117–133.
- Van De Vijver, B., Beyens, L., 1998. A preliminary study on the soil diatom assemblages from Île de la Possession (Crozet, Subantarctica). *Eur. J. Soil Biol.* 34, 133–141.
- Van de Vijver, B., Ledeganck, P., Beyens, L., 2002. Soil diatom communities from Île de la Possession (Crozet, sub-Antarctica). *Polar Biol.* 25, 721–729. <https://doi.org/10.1007/s00300-002-0392-9>.
- van Kerckvoorde, A., Trappeniers, K., Nijs, I., Beyens, L., 2000. Terrestrial soil diatom assemblages from different vegetation types in Zackenberg (Northeast Greenland). *Polar Biol.* 23 (6), 392–400.
- Vasselon, V., Rimet, F., Tapolczai, K., Bouchez, A., 2017. Assessing ecological status with diatoms DNA metabarcoding: Scaling-up on a WFD monitoring network (Mayotte island, France). *Ecol. Ind.* 82, 1–12. <https://doi.org/10.1016/j.ecolind.2017.06.024>.
- Wallach, D., Goffinet, B., 1989. Mean squared error of prediction as a criterion for evaluating and comparing system models. *Ecol. Model.* 44 (3-4), 299–306.
- Wetzel, C.E., Ector, L., Van de Vijver, B., Compere, P., Mann, D.G., 2015. Morphology, typification and critical analysis of some ecologically important small naviculoid species (Bacillariophyta). *Fottea* 15 (2), 203–234.
- Zhang, Y., Ouyang, S., Nie, L., Chen, X.u., 2020. Soil diatom communities and their relation to environmental factors in three types of soil from four cities in central-west China. *Eur. J. Soil Biol.* 98, 103175.