water

MDPI

Article

Determination of the Habitat Preferences of Dominant Epilithic Diatoms Using Statistical Models: A Case Study in the Han River, South Korea

Yuna Shin [1], Doyun Kim [2] and Tae-Young Heo [2,*]

[1] Department of Water Environment Research, National Institute of Environmental Research, Incheon 22689, Korea; marianshin@korea.kr
[2] Department of Information & Statistics, Chungbuk National University, Chungbuk 28644, Korea; doyun2516@gmail.com
* Correspondence: theo@cbnu.ac.kr; Tel.: +82-43-261-3741

Abstract: Diatoms have traditionally been used to assess water quality; however, current research suggests that physical factors, such as habitat and landscape, may be linked to the organization of diatom assemblages in streams. The aim of this study was to determine the environmental factors affecting the physiological and ecological changes of epilithic diatoms. To this end, the dominant diatom species were used, and a strategy based on epilithic diatom habitat characteristics was investigated for river restoration. The classification and regression tree (CART) and weighted averaging (WA) regression models were used to determine the habitat preference of epilithic diatoms and physicochemical environmental factors. The 35 environmental variables and six epilithic diatom species that were dominant at 58 sites along the Han River in South Korea were used. The species abundance and composition of adherent diatoms were affected by nutrient concentration and a variety of physicochemical environmental factors. These results suggest that when evaluating water quality, various factors that affect the abundance of epilithic diatoms should be considered. Research on the autecological characteristics and environmental preferences of indicator diatom species could help establish river restoration policies and quantitative evaluation criteria for biological assessments.

Keywords: classification and regression trees (CART); environmental variables; epilithic diatom; weighted averaging (WA)

Citation: Shin, Y.; Kim, D.; Heo, T.-Y. Determination of the Habitat Preferences of Dominant Epilithic Diatoms Using Statistical Models: A Case Study in the Han River, South Korea. *Water* **2022**, *14*, 956. https://doi.org/10.3390/w14060956

Academic Editor: Michele Mistri

Received: 14 February 2022
Accepted: 15 March 2022
Published: 18 March 2022

1. Introduction

Rivers and streams have complex ecosystem structures with a variety of environmental variables that make it impractical to assess all potentially important variables that influence various environmental conditions [1]. Diatoms are sensitive to nutrient and organic matter concentrations and integrate long-term water quality with environmental variables [2,3]. Diatoms have traditionally been used to assess water quality; however, current research suggests that physical factors, such as habitat and landscape, may be linked to the organization of diatom assemblages in streams [4–7].

When the water becomes nutrient-saturated owing to the domination of similar eutrophic species, small differences in diatom composition and structure are influenced more by physical factors rather than nutrients [8]. The degree to which species composition resembles that of the reference can be used as an indicator of biological circumstances, while deviations from the reference can more clearly explain the consequences of anthropogenic activity [9]. The impact of nutrients and numerous environmental elements should be considered when evaluating aquatic ecological health as a whole. It is also vital to conduct research on water quality evaluation and river restoration technologies by employing diatom species that reflect the features of the region.

The biological assessment of water quality is critical for an overall perception of ecosystem integrity, which includes aquatic ecosystem habitat conditions [10]. Bioindicators represent the impact of non-measurable environmental variables. Recently, epilithic diatom monitoring was used to assess biological water quality [11]. The interactions between diatoms and environmental variables in streams are complicated, frequently including multiple variables that function in a hierarchical manner [12]. Diatom autecology can benefit from explicitly incorporating interactions between variables. The current state of diatom autecology knowledge is insufficient and based on research that has not been explicitly targeted to assess common species' environmental requirements. As a result, environmental preference lists for common species are frequently imprecise and inconsistent, and autecological information on common species is often unavailable in many places.

Classification and regression trees (CART) are effective for analyzing complex ecological data. A CART has the following characteristics: (1) the ability to utilize different types of response variables; (2) the ability for interactive exploration, description, and prediction; (3) invariance to transformations of explanatory variables; (4) simple graphical interpretation of complex results involving interactions; (5) model selection by cross-validation; and (6) procedures for dealing with missing values [13]. Owing to its ability to handle both continuous and discrete variables, its inherent ability to model interactions among predictors, and its hierarchical structure, CART is appealing to many exploratory environmental and ecological studies. Regression trees (RT) and classification trees (CT) are useful for visualizing data structures and interactions [13,14]. A CART is well-suited for the relative abundance data of diatom species, as these frequently contain a large number of zero values [15].

CART techniques are increasingly being utilized to investigate species–environmental interactions in plants and animals [13,16,17]. Diatom autecology has been quantitatively evaluated using a weighted averaged regression model [18,19]. When a species is rare and has small ecological amplitude, and the distribution of the environmental variables among the sites is fairly homogeneous over the entire range of occurrence of the species, weighted averaging (WA) is just as effective as regression methods for estimating optima [20].

The WA regression model remains a simple and useful tool used to reveal structures in data tables by rearranging species and sites based on an exploratory variable. WA has been proposed as a biotic index for vascular plants, algae, and faunal communities in streams and rivers [21]. The WA regression model quantifies diatom autecology and, in many cases, expands our understanding of diatom species preferences. However, WA modeling assumes that the variable of interest is the only variable that influences species distribution. The significance of other environmental variables is implicitly accounted for in the calculation of the optimum WA. Conversely, WA cannot explicitly depict interactions between environmental variables. As a result, each environmental variable must be interpreted individually.

In Korea, the Trophic Diatom Index (TDI) [22] is used to assess water quality. However, as each epilithic diatom species reacts differently under different environmental conditions, it is necessary to establish a method that can be used as an indicator of water quality evaluation by analyzing the habitat preferences of epilithic diatoms. Therefore, it is necessary to develop biological indicators and evaluation methods that are appropriate for domestic conditions and reflect the effects of various environmental factors. The CART and WA were used in this study to determine the quantitative parameters of epilithic diatom habitat preference based on physicochemical environmental factors. Most previous studies have focused on the epilithic diatom community. However, as diatoms respond to complex environmental variables in a sensitive manner, habitat preferences for diatom organisms should be qualitatively indicated. In particular, the dominant species is a species that has been well adapted to the region's environmental conditions and defines the representative characteristics that characterize the group. Dominant species are those that appear at the highest density and frequency and are used for ecological river restoration as

reflected in river planning, design, construction, maintenance, and monitoring through the characterization of the river environment [23].

The goal of this study was to use CART and WA to determine the physicochemical environmental factors that influence the dominant diatom species as well as to determine the preferred habitat for each species. The results of this study can be used to evaluate and predict changes in the water environment.

2. Material and Methods

2.1. Study Area

The Han River is a major river in South Korea, formed by the confluence of the two major branches of the river, the Namhan River (South Han River) and the Bukhan River (North Han River). The river flows through Seoul before merging with the Imjin River just before it enters the Yellow Sea. The total length of the Han River (including its tributaries, the Namhan and Bukhan rivers) is approximately 494 km, with a catchment area of 25,953 km^2. Figure 1 depicts the 58 monitoring sites in South Korea used to collect epilithic diatom samples and assess 35 environmental variables.

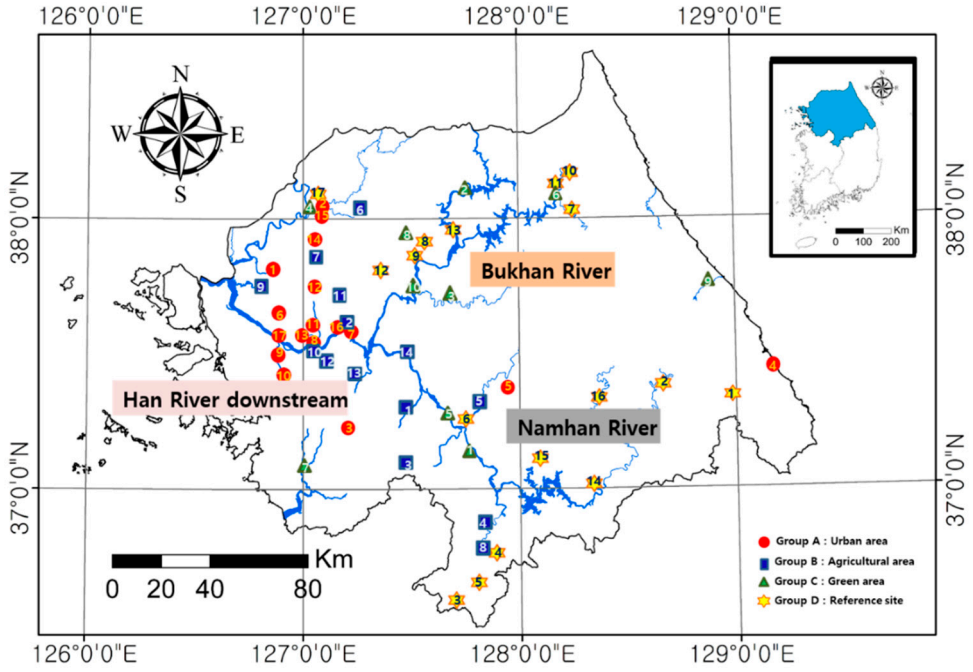


Figure 1. The Han River is a major river in South Korea, formed by the confluence of the two major branches of the river, the Namhan River and the Bukhan River. Fifty-eight monitoring sites (four groups) along the Han River in South Korea.

We chose 58 sites along the Han River in South Korea and classified them into four groups based on the land use of riverside areas within 500 m of the stream bank (A: urban area; B: agricultural area (mixed area of urban and agricultural use); C: green area (well-reserved area with minimal human activity); D: reference site), as shown in Table 1. The reference condition (group D) was composed of sites that experts considered to be a good riparian environment. The WFD (The European Water Framework Directive) defined reference condition sites as those with none or minimal human impact [24]; this definition was made using previous knowledge, expert judgments, and collected information [25]. Reference sites are usually selected from the sampling sites [26] and are mostly defined by their diatom community structure [11].

Table 1. The code and name of fifty-eight monitoring sites in the study area along the Han River in South Korea.

Code	Group A	Code	Group B	Code	Group C	Code	Group D
A-1	Munsancheon 1	B-1	Bockhacheon 2	C-1	Deogeunri	D-1	Buncheon
A-2	Hantangang 3	B-2	Wangsukcheon 4	C-2	Hwacheon	D-2	Jeongseon 2
A-3	Kyeongancheon 1	B-3	Cheongmicheon 1	C-3	Hongcheongang 2	D-3	Dalcheon 1
A-4	Sajik	B-4	Ssangcheon	C-4	Hantangang 3-1	D-4	Hyangmokdongcheon
A-5	Wonjuchun 1	B-5	Seomgang 4	C-5	Gangchun	D-5	Hwayangcheon
A-6	Changneungcheon 1	B-6	Yeongpyeongcheon 2	C-6	Naerincheon 2	D-6	Seomgang 4-1
A-7	Dosimcheon	B-7	Sincheon 1	C-7	Jinwicheon2	D-7	Naerincheon 1
A-8	Seongbukcheon	B-8	Dalcheon 3	C-8	Gapyeongcheon 1	D-8	Gapyeongcheon 2
A-9	Anyangcheon 4	B-9	Gongneungcheon 2	C-9	Gangneung	D-9	Gapyeongcheon 3
A-10	Anyangcheon 3	B-10	Yangiaecheon	C-10	Namiseom	D-10	Bukcheon
A-11	Jeongneungcheon	B-11	Wangsukcheon 1			D-11	Inbukcheon 2
A-12	Jungnangcheon 1	B-12	Tancheon 4			D-12	Jojongcheon 2
A-13	Cheonggyecheon 1	B-13	Kyeongancheon 4			D-13	Chuncheon
A-14	Sincheon 2	B-14	Gangsang			D-14	Maepocheon
A-15	Sincheon 3					D-15	Jecheoncheon 2
A-16	Wangsukcheon 3					D-16	Pyeongchanggang 2
A-17	Yeongdeungpo					D-17	Chatancheon

2.2. Data

Composition of Dominant Species

The physicochemical variables and epilithic diatom data for the 58 monitored sites (A, 17 sites; B, 14 sites; C, 10 sites; D, 17 sites) from 2008 to 2010 were obtained from the Korea Ministry of Environment and the Korea Meteorological Administration. The Korea Ministry of Environment and the Korea Meteorological Administration provided data on 337 samples from 58 monitoring sites in the Han River, as well as six epilithic diatoms (Table 2). In the aforementioned samples, 361 diatom species were identified. Among the 361 species identified, six dominant species (*Achnanthes minutissima*, *A. convergens*, *Nitzschia inconspicua*, *N. fonticola*, *N. amphibia*, and *N. palea*) were used.

Table 2. Seven descriptive statistics for six diatom species.

Code	Taxon	Minimum	Median	Mean	Proportion (%)	Maximum	No. of Dominance	No. of Sites
A.m.	*Achnanthes minutissima*	0	9658	128,302	29.37	3,707,386	65	230
A.c.	*Achnanthes convergens*	0	2236	96,885	22.67	8,275,400	59	178
N.i.	*Nitzschia inconspicua*	0	0	29,456	4.76	1,166,023	25	82
N.f.	*Nitzschia fonticola*	0	31	115,637	10.67	13,795,082	25	154
N.a.	*Nitzschia amphibia*	0	3291	40,023	14.81	1,562,954	24	195
N.p.	*Nitzschia palea*	0	4125	64,525	17.72	1,557,544	15	196

In this study, the top five species (*A. minutissima*, *A. convergens*, *N. amphibia*, *N. fonticola*, and *N. inconspicua*) and *N. palea*, which are widely used as indicators of water pollution as polluting species, were also chosen.

Data analysis was performed on 307 out of 337 samples and included six dominant species and 35 environmental variables. The analysis did not include 18 samples with no occurrence and 12 samples with outliers. The following variables listed in Table 3 were considered in this study: altitude, sunshine duration, silt–sand, gravel, bedrock, riffle, run, swamp, canopy, vegetation cover, herb, shrub, urban, forestry, agricultural land, river width, depth of water, water temperature, dissolved oxygen (DO), pH, conductivity, turbidity, chlorophyll a, biochemical oxygen demand (BOD), chemical oxygen demand (COD), suspended solids (SS), T-N, T-P, DTN, DTP, NH_3-N, NO_3-N, PO_4-P, fecal coliforms, and total coliforms.

Table 3. Four descriptive statistics for 35 environmental variables.

Code	Environmental Variable	Min	Median	Max	Mean
AT	Altitude (m)	1.0	70.000	744.0	105.908
Sun	Sunshine (h)	0.0	6.800	13.2	6.328
SSand	Silt–sand (%)	0.0	30.000	100.0	38.139
Gv	Gravel (%)	0.0	40.000	100.0	47.202
BR	Bedrock (%)	0.0	5.000	95.0	11.484
Rpp	Ripple (%)	0.0	20.000	100.0	20.766
run	Run (%)	0.0	70.000	100.0	64.139
SW	Swamp (%)	0.0	0.000	100.0	8.033
CP	Canopy (%)	0.0	0.000	80.0	8.917
VG	Vegetative cover (%)	0.0	60.000	100.0	51.122
Herb	Herb (%)	0.0	80.000	100.0	75.947
Sh	Shrub (%)	0.0	20.000	80.0	23.282
Ub	Urban (%)	0.0	20.000	100.0	39.487
Fr	Forest (%)	0.0	10.000	100.0	30.454
Agc	Agriculture (%)	0.0	0.000	100.0	17.727
Wd	River width (m)	2.0	50.000	1000.0	109.237
Depth	River depth of river	1.0	40.000	1000.0	56.220
Temp	Temperature (°C)	7.5	20.400	95.1	20.218
DO	DO	0.0	9.700	26.7	10.056
pH	pH	4.1	7.640	76.0	7.943
Cdv	Conductivity (μs/cm)	0.0	197.000	9840.0	341.271
Tbd	Turbidity	0.0	7.000	352.0	15.887
Chl_a	chlorophyll-a (μg/cm^3)	0.0	3.650	92.7	6.676
BOD	BOD (mg/L)	0.1	1.100	32.0	2.886
COD	COD (mg/L)	0.5	3.400	37.0	4.961
SS	SS (mg/L)	0.2	5.600	138.0	9.328
T_N	T-N (mg/L)	0.6	2.687	20.2	3.732
NH3_N	NH_3-N (mg/L)	0.0	0.066	15.5	0.736
NO3_N	NO_3-.N (mg/L)	0.0	1.919	8.8	2.211
T_P	T-P (mg/L)	0.0	0.041	4.4	0.164
FC	Fecal coliforms (/100 mL)	0.0	70.000	600,000.0	5313.629
TC	Total coliforms (/100 mL)	0.0	1800.000	2,500,000.0	43,120.801
DTN	DTN (mg/L)	0.6	2.525	18.6	3.476
DTP	DTP (mg/L)	0.0	0.029	3.9	0.121
PO4_P	PO_4-P (mg/L)	0.0	0.013	1.6	0.093

2.3. Data Analysis

The CART and WA models were used to qualitatively infer environmental parameters affecting dominant diatom species and current environmental conditions that reflected the physiological and ecological aspects of each species in this study. CART analysis can help quantify the environmental conditions that influence the appearance of each species. As a result, research into the development of a water quality evaluation method and ecological river restoration technique based on the environmental sensitivity of the dominant species showing a high frequency of occurrence in a specific area, as opposed to the existing community-oriented evaluation method, is required. WA techniques allow for a quantitative assessment of diatom autecology and, in many cases, expand our understanding of diatom species preferences.

2.3.1. CART

CART can be used to visually aid interpretation, reveal data structures, and display interactions. CART is a binary recursive partitioning method that produces a tree-based model class. Owing to its ability to handle both continuous and discrete variables, its inherent ability to model interactions among predictors, and its hierarchical structure, the method is appealing to many exploratory environmental and ecological studies [13–15].

CART analysis was used to determine the habitat preferences of the dominant diatom species. CART analysis recursively divides observations in a matched data set into progressively smaller groups, using a categorical (for a classification tree (CT)) or continuous (for a regression tree (RT)) dependent (response) variable and one or more independent (explanatory) variables. Each partition consists of two binary splits. The splits of each explanatory variable were examined during each recursion, and the split that maximized

the homogeneity of the two resulting groups with respect to the dependent variable was chosen. The RT and CT analyses were performed using the R package rpart [27,28].

CT analysis was used to classify the presence or absence of the six dominant diatom species, and RT analysis was used to calculate the average dominant rate relative to the total epilithic diatoms. Cross-validation was utilized to obtain accurate estimations of prediction errors for CT and RT tree sizes as well as to determine when to stop splitting the data [29]. Cross-validation divides the original dataset into separate, mutually exclusive datasets, resulting in numerous splits [19].

2.3.2. Optimal Conditions of Dominant Species: Using WA Regression and Calibration

The WA regression model is used for the same purpose as regression analysis based on a Gaussian model and is utilized in various fields of ecology. The WA is used to estimate the optimal point of a species using environmental variables. The abundance of each algal species and the environmental conditions at each site are inextricably linked.

Environmental conditions may take precedence in a causal relationship because diverse environmental variables might alter the habitat conditions and abundance of algal species. However, the amount of generation may change the ambient conditions based on the biological properties of the algae. As a result, the analysis should be conducted using a model that reflects the interactive link between abundance and environment, such as the WA regression model.

The WA model calculates the optimal condition value (called optima) for each species as the weighted average of the occurrence of each species for one environmental variable, which results in the calculation of a new environmental variable value. As the newly calculated value of the environmental variable reflects the amount of species generation, regression and calibration can be performed using the existing environmental variable value and each other as the dependent and explanatory variables, respectively. The WA is used to estimate the optimal point of a species using environmental variables. If a species follows a unimodal or Gaussian form, depending on the environmental variables, the species will be concentrated at the peak of this form.

The WA of the k-th species to which the environmental variable x is weighted, which is referred to as the optimal environmental condition for each species, is calculated as follows (Equation (1)):

$$\overline{\mu} = \sum_{i=1}^{n} \frac{y_{ik} x_i}{y_{+k}}, \qquad y_{+k} = \sum_{i=1}^{n} y_{ik} \tag{1}$$

where y_{ik} is the abundance of the k-th species in the i-th region and x_i is the environmental variable of the i-th region.

The WA tolerance (T_i) value of each species in Equation (2) is obtained by weighting the abundance of each algal species in all regions and calculating the standard deviation of the WA.

$$T_i = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{\mu}_k)^2 y_{ik}}{\sum_{i=1}^{n} y_{ik}}} \tag{2}$$

Using the optimal point of the species calculated in Equation (1), the WA regression method can be used to estimate the environmental value of each site. If the species is closely related to the environmental variable, the optimal point of the species will naturally represent environmental variables. The WA is calculated using all species and is weighted for each site, according to Equation (3):

$$\widetilde{x}_i = \frac{\sum_{k=1}^{m} y_{ik} \overline{\mu}_k}{y_{i+}}, y_{i+} = \sum_{k=1}^{m} y_{ik} \tag{3}$$

3. Results

3.1. Habitat Preference in Epilithic Diatoms Using CART Analysis

3.1.1. CT

CT was used to determine the presence or absence (binary: 1 or 0) of six dominant epilithic diatom species and the existence probability of six species in Table 4. *A. minutissima* was more abundant than the other taxa, while *N. inconspicua* was less abundant.

Table 4. The existence probabilities of six diatom species.

Species	Probability
A. minutissima	0.749
A. convergens	0.580
N. inconspicua	0.267
N. fonticola	0.502
N. amphibia	0.635
N. palea	0.638

As is shown in Figure 2a, the first split (total 307 sites) was separated from the second (CP < 2.5, 159 sites) and seventh nodes (CP ≥ 2.5, 148 sites) in the CT change point for *A. minutissima*. The second node was split into the third (COD ≥ 3.25, 90 sites) and sixth (COD < 3.25, 69 sites) nodes and the third node was split into the fourth (Chl_a < 3.595, 58 sites) and fifth (Chl_a ≥ 3.595, 32 sites) nodes. The existence probabilities of *A. minutissima* in the fourth, fifth, and sixth nodes were 0.293 (58 sites), 0.75 (32 sites), and 0.826 (69 sites), respectively. The seventh node was separated from the eighth (DO ≥ 16.86, 8 sites) and ninth nodes (DO < 16.86, 140 sites), and the existence probabilities in the eighth and ninth nodes were 0.125 and 0.936, respectively.

In the presence of *A. minutissima*, the ninth node (CP ≥ 2.5 and DO < 16.86, 140 sites) had the highest value, while the fourth node (58 sites) had the lowest value. Based on these results, the probability of *A. minutissima* surviving in the CP ≥ 2.5 and DO < 16.86 conditions is high. Conversely, low levels of *A. minutissima* were found under the CP < 2.5, COD ≥ 3.25, and Chl_a < 3.595 conditions. There was a 0.749 chance of discovering *A. minutissima* in all 307 sites, with a 15.3% chance of misclassification.

As is shown in Figure 2b, the first split of *A. convergens* (307 sites) was separated from the second (DTN < 4.218, 59 sites) and third (DTN < 4.218, 248 sites) nodes. The chances of finding *A. convergens* in the second and third nodes were 0.203 (59 sites) and 0.669 (248 sites), respectively. Based on these results, a high presence of *A. convergens* is expected under the DTN < 4.218 condition. The third node was separated from the fourth (pH ≥ 7.175) and fifth (pH < 7.175) nodes. The fifth node was separated from the sixth (run ≥ 7.5) and seventh (run < 7.5) nodes. The probability of finding *A. convergens* in all 307 sites was 0.580, with a misclassification rate of 25.1%.

As is shown in Figure 2c, the first split (total 307 sites) of *N. inconspicua* was separated from the second (AT ≥ 90.5, 126 sites) and third (AT < 90.5, 181 sites) nodes. The fourth (CP < 7.5, 115 sites) and fifth (CP ≥ 7.5, 66 sites) nodes were separated from the third node. *Nitzschia inconspicua* had a 0.0635, 0.226, and 0.727 probability of being present in the first, fourth, and fifth nodes, respectively. Based on these results, the presence of *N. inconspicua* is expected to be high under the AT < 90.5 and CP ≥ 7.5 conditions and low under the AT 90.5 condition. *N. inconspicua* was found in 0.267 of the 307 sites, with a 16.9% misclassification rate.

As is shown in Figure 2d, the first split (307 sites) of *N. fonticola* was distinguished from the second (AT ≥ 134.5, 77 sites) and third nodes (AT < 134.5, 230 sites). The presence probability of *N. fonticola* in the second and third nodes was 0.234 and 0.591, respectively. Based on these results, the presence of *N. fonticola* is expected to be high under the AT < 134.5 condition and low under the AT ≥ 134.5 condition. The third node (PO4-P 0.185) was separated from the fourth and fifth nodes (PO4-P = 0.185). The existence probability of *N. fonticola* at all 307 sites was 0.502, with a 29.6% misclassification rate.

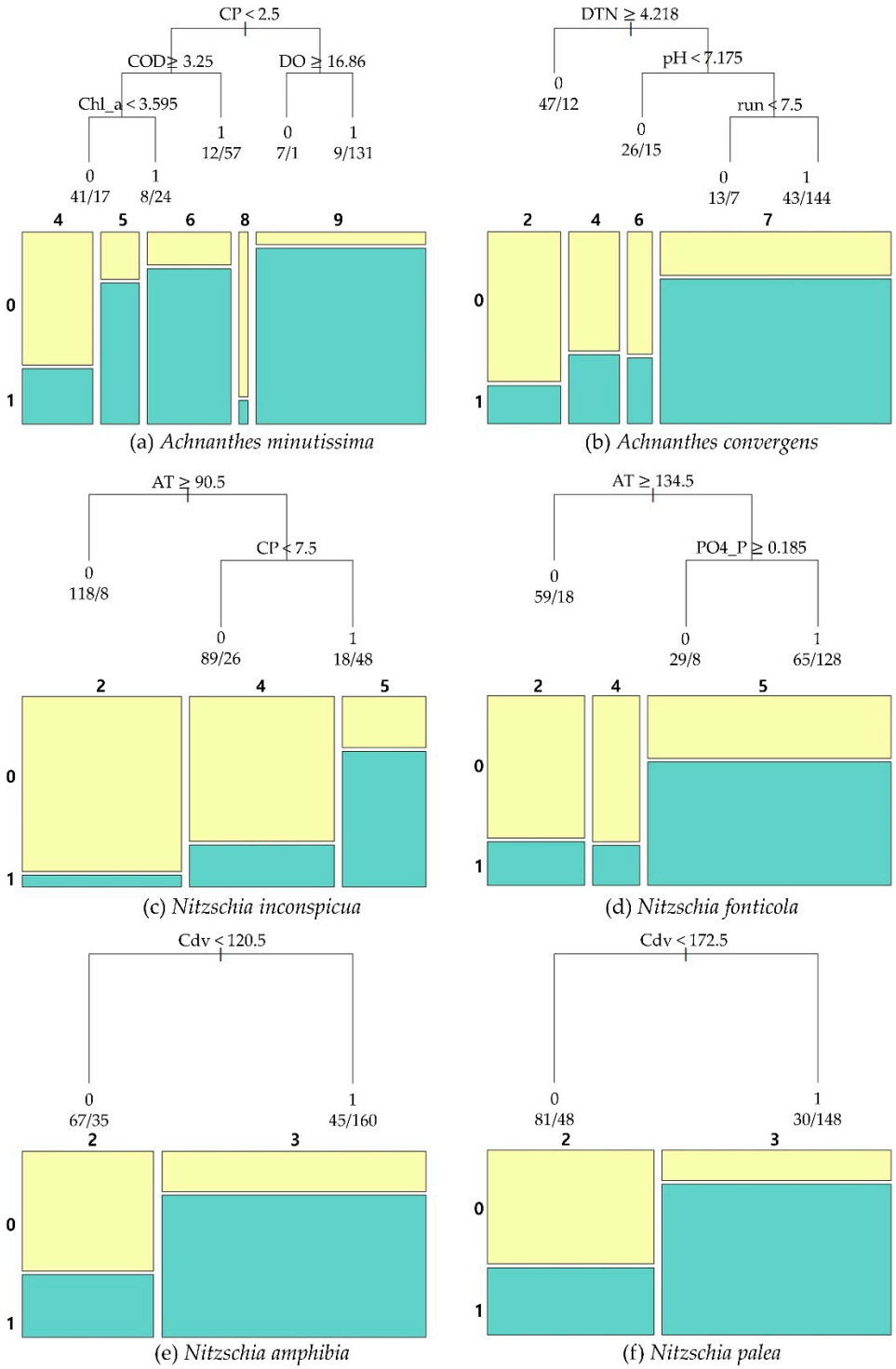


Figure 2. Classification tree results for six diatom species. (**a**) *Achnanthes minutissima*, (**b**) *Achnanthes convergens*, (**c**) *Nitzschia inconspicua*, (**d**) *Nitzschia fonticola*, (**e**) *Nitzschia amphibia*, (**f**) *Nitzschia palea*.

As is shown in Figure 2e, the first split (307 sites) of *N. amphibia* was separated from the second (Cdv < 120.5, 102 sites) and third nodes (Cdv ≥ 120.5, 205 sites). The probability of finding *N. amphibia* in the second and third nodes was 0.343 and 0.780, respectively. Based on these results, the presence of *N. amphibia* is expected to be high under the Cdv < 120.5 condition and low under the CDV ≥ 120.5 condition. The probability of the existence of *N. amphibia* at all 307 sites was 0.635, and the misclassification rate was 26.1%.

As is shown in Figure 2f, the first split (307 sites) of *N. palea* was separated from the second (Cdv < 172.5, 129 sites) and third nodes (Cdv ≥ 172.5, 178 sites). The chances of

finding *N. palea* in the second and third nodes were 0.372 and 0.831, respectively. Based on these results, the presence of *N. palea* is expected to be high under the Cdv ≥ 172.5 condition and low under the Cdv < 172.5 condition. The probability of finding *N. palea* in all 307 sites was 0.638, with a misclassification rate of 25.4%.

3.1.2. RT

The change points identified by RT consider hierarchical relationships between environmental variables and thus may provide more accurate information on where diatom species abundances shift along environmental gradients [15].

The relative abundances of *A. minutissima, A. convergens, N. inconspicua, N. fonticola, N. amphibia,* and *N. palea* were 27.0%, 20.4%, 6.2%, 24.4%, 8.4%, and 13.6%, respectively.

As is shown in Figure 3a, the first split of *A. minutissima* (307 sites) was separated from the third (Cdv ≥ 154 and DTP ≥ 0.0025, 176 sites), fourth (Cdv ≥ 54 and DTP < 0.0025, 15 sites), and fifth (Cdv < 154, 116 sites) nodes. *A. minutissima* had the lowest relative abundance of 0.0815 with 176 sites in the third node and the highest relative abundance of 0.696 with 116 sites in the fifth node. Based on these results, *A. minutissima* will have high relative abundance under the Cdv 154 condition. However, a box plot of the fifth node (Cdv < 154, 116 sites) showed a large gap between the 25th and 75th percentiles as well as a deviation from the relative abundance of *A. minutissima* with a high mean. Based on these results, a relative abundance of 0.667 in the fourth node (Cdv ≥ 154 and DTP < 0.0025) can be considered a good condition.

As is shown in Figure 3b, the first split of *A. convergens* (307 sites) was separated from the second (Wd < 152, 243 sites) and fifth (Wd ≥ 152, 64 sites) nodes. The second node was separated from the third (AT < 102, 143 sites) and fourth (AT ≥ 102, 100 sites) nodes. The relative abundance of *A. convergens* was 0.031 in the third node and 0.69 in the fifth node. Based on these results, *A. convergens* is expected to have high relative abundance under the Wd < 152 condition. The relative abundance of *A. convergens* at all 307 sites was 0.204.

As is shown in Figure 3c, the first split (307 sites) of *N. inconspicua* was separated from the second (Agc < 32.5, 235 sites) and fifth nodes (Agc ≥ 32.5, 72 sites). The second node was separated from the third (DO < 3.5, 223 sites) and fourth (DO ≥ 3.5, 12 sites) nodes. The relative abundance of *N. inconspicua* was 0.0192 in the third node (Agc < 32.5 and DO ≥ 3.5) and 0.508 in the fourth node. However, the relative abundance in the fourth node (12 sites) did not have a robust mean with one outlier. Therefore, *N. inconspicua* is expected to have high relative abundance in the fifth node under the Agc ≥ 32.5 condition. *N. inconspicua* had a relative abundance of 0.062 across all 307 sites.

As is shown in Figure 3d, the first split (307 sites) of *N. fonticola* was separated from the second (AT ≥ 36, 246 sites) and fifth nodes (AT < 36, 61 sites). The second node was separated from the third (CP < 25, 218 sites) and fourth (CP ≥ 25, 28 sites) nodes. The fifth node was then separated from the sixth (temp <16.75, 47 sites) and seventh (temp 16.75, 14 sites) nodes. The sample size of this species was small; however, the average abundance was lowest at the third node (0.0578) and highest at the seventh node (0.896). *N. fonticola* is expected to have high relative abundance in the seventh node under the AT < 36 and temp <16.75 conditions. Furthermore, the lowest relative abundance was found in the third node, with AT ≥ 36 and CP < 25. *N. fonticola* had a relative abundance of 0.244 across all 307 sites.

As is shown in Figure 3e, the first split (307 sites) of *N. amphibia* was separated from the second (T_ P < 0.0305, 132 sites) and third nodes (T_P ≥ 0.0305, 175 sites). *N. amphibia* had the lowest and highest relative abundance in the second (0.0197) and third (0.145) nodes, respectively. Based on these results, the presence of *N. amphibia* is expected to be high under the T_P ≥ 0.0305 condition. *N. amphibia* was found in 0.084 of 307 sites.

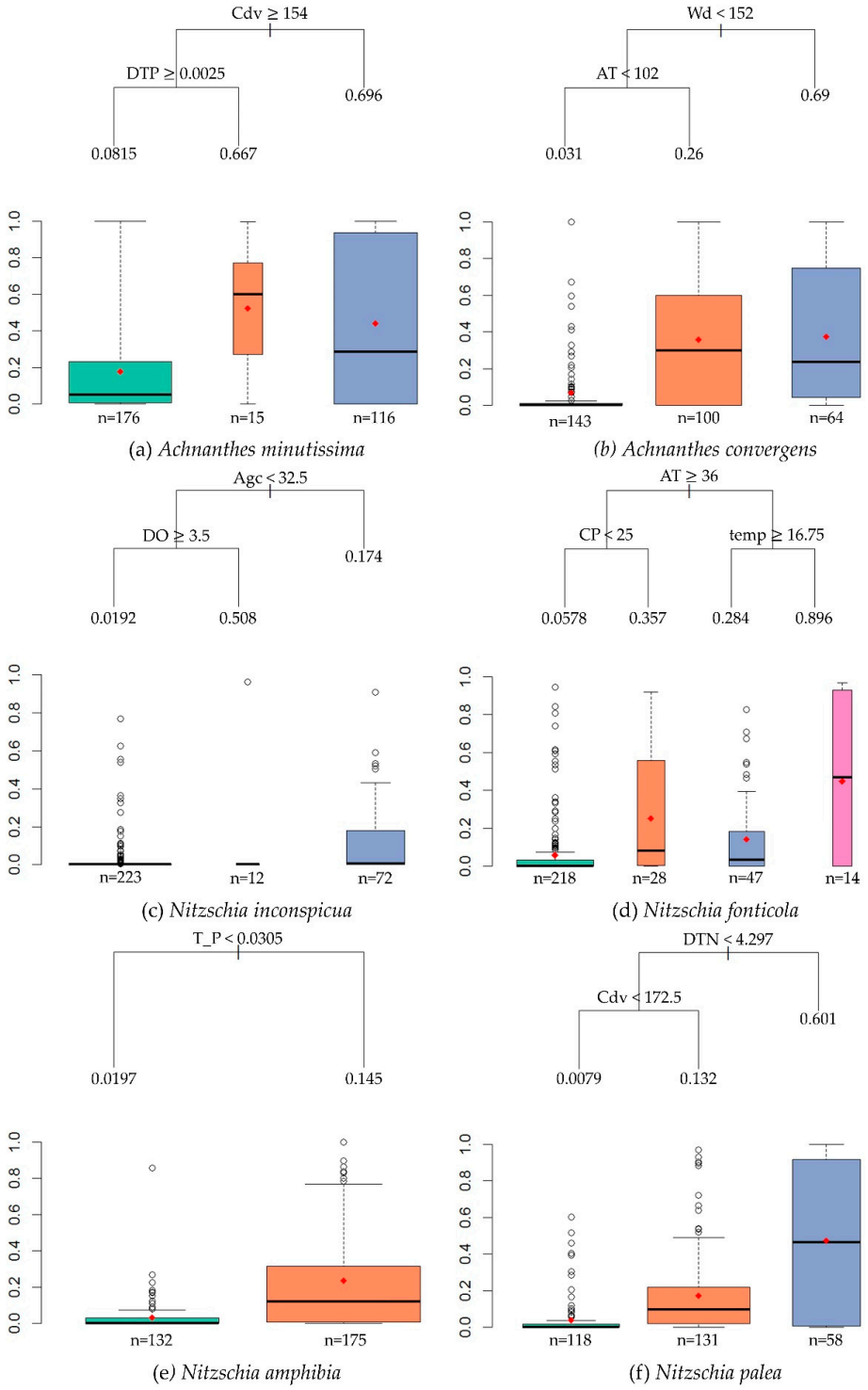


Figure 3. Regression tree results for six diatom species. (**a**) *Achnanthes minutissima*, (**b**) *Achnanthes convergens*, (**c**) *Nitzschia inconspicua*, (**d**) *Nitzschia fonticola*, (**e**) *Nitzschia amphibia*, (**f**) *Nitzschia palea*.

As is shown in Figure 3f, the first split (307 sites) of *N. palea* was separated from the second (DTN < 4.297, 249 sites) and fifth (DTN ≥ 4.297, 58 sites) nodes, and the third (Cdv < 172.5, 118 sites) and fourth (Cdv ≥ 172.5, 131 sites) nodes were separated from the second node. *N. palea* had the lowest relative abundance (0.0079) with 118 sites in the third node and a relative abundance of 0.601 with 58 sites in the fifth node. Based on these results, the relative abundance of *N. palea* is expected to be high under the DTN ≥ 4.297 condition

and low under the DTN < 4.297 and Cdv < 172.5 conditions. *N. palea* had a relative abundance of 0.136 across all 307 sites.

3.2. WA Model

WA regression using R (version 3.6) with the package RIOJA (version 0.8-7; Juggins, 2013) was used to estimate species optima and tolerance levels for environmental variables. [30]. The coefficient of determination (R^2) is a descriptive statistic that measures how much of the variance of the dependent variable y is explained by the explanatory factors given. As WA optima is a value weighted by the inferred environmental variable ($\widetilde{x}_i$), as in Equation (3), R^2 by regression analysis can be used as an index of the model validity. If R^2 for a specific environmental variable is high to a certain amount, it may be assumed that there is a relationship between a species' abundance and the environmental variable, and WA optima can be considered a trustworthy value [21]. The top 10 environmental variables are listed in Table 5 in the order of R^2.

Table 5. Ten environmental variables with a high R^2 value.

Environmental Variable	R^2 of WA	Environmental Variable	R^2 of WA
T-N (mg/L)	0.3485	COD (mg/L)	0.2458
DTN (mg/L)	0.3477	NO_3-N (mg/L)	0.2330
Forest (%)	0.2797	BOD (mg/L)	0.2326
T-P (mg/L)	0.2677	PO_4-P (mg/L)	0.2293
DTP (mg/L)	0.2501	Altitude (m)	0.2245

As is shown in Table 6, the WA optima for T-N, DTN, T-P, DTP, COD, NO_3N, BOD, and PO_4P were found in *N. palea*, *N. amphibia*, *N. fonticola*, and *N. inconspicua*, while the WA optima for forest and altitude were found in *A. minutissima* and *A. convergens*. Non-pollution variables, such as forest and altitude, affected *A. minutissima* and *A. convergens*, while water quality variables affected *N. palea*, *N. amphibia*, *N. fonticola*, and *N. inconspicua*. As a result, each epilithic diatom species was resistant to pollution, and the optimal habitat environment for each species was significant.

Table 6. Optima and tolerance for six dominant species using weighted averaging (WA).

		Taxon (Code)					
		A.m.	***A.c.***	***N.i.***	***N.f.***	***N.a.***	***N.p.***
T-N (mg/L)	WA optima	2.5894	2.4983	3.0887	3.2953	4.9045	6.3064
	Tolerance	2.2177	1.6162	0.9323	1.9156	3.5212	4.5713
DTN (mg/L)	WA optima	2.3683	2.3215	2.9437	3.0774	4.6043	5.9082
	Tolerance	2.0619	1.5543	0.8918	1.7857	3.4263	4.2987
Forestry (%)	WA optima	49.5756	40.6140	19.1550	25.0513	14.4665	13.9408
	Tolerance	35.6040	35.7744	25.5176	31.3951	27.5419	24.1234
T-P (mg/L)	WA optima	0.0730	0.0621	0.0790	0.1081	0.2287	0.3240
	Tolerance	0.1810	0.1183	0.0974	0.1513	0.2598	0.3962
DTP (mg/L)	WA optima	0.0497	0.0435	0.0600	0.0682	0.1693	0.2376
	Tolerance	0.1289	0.0893	0.0817	0.1158	0.2090	0.3170
COD (mg/L)	WA optima	3.3335	3.3474	3.9642	5.2606	6.5497	7.8706
	Tolerance	3.3181	2.3368	3.2517	5.4492	4.5315	6.4098
NO_3-N (mg/L)	WA optima	1.7486	1.6983	2.4339	2.2705	2.5445	3.1202
	Tolerance	1.1345	0.9769	0.8867	1.1990	1.3341	1.8639
BOD (mg/L)	WA optima	1.5072	1.4302	1.5596	2.8786	4.4491	5.5318
	Tolerance	2.5537	2.4620	2.9810	4.8352	4.7232	6.3538
PO_4-P (mg/L)	WA optima	0.0387	0.0311	0.0438	0.0510	0.1397	0.2030
	Tolerance	0.1176	0.0792	0.0770	0.1099	0.1959	0.2935
Altitude (m)	WA optima	155.9967	147.7223	48.7005	62.5374	73.7021	64.4209
	Tolerance	143.1342	123.6208	25.1594	62.8436	67.0039	64.3272

4. Discussion

The complexity of hierarchical interactions between environmental variables and diatoms may contribute to the difficulty in linking indicator taxa to a specific set of environmental variables. Mismatches between diatoms and measured environmental variables at both spatial and temporal scales could also be a source of contention [7].

The various water pollution features have an impact on epilithic diatom assemblage changes, which could result in a cascade of impacts that determine diatom assemblage composition.

Epilithic diatom species are sensitive to a wide range of physicochemical conditions and have complex interactions with their environment. The most prevalent and dominant species, *A. convergens*, *A. minutissima*, *N. amphibian*, *N. fonticola*, and *N. inconspicua*, may be able to adapt to a variety of environmental changes. Typically, each species is found near specific environmental optima [31].

A. minutissima was discovered primarily in well-protected areas and reference sites. The best water quality was found in *A. minutissima* dominated areas, which were mostly located upstream [32,33]. *N. palea* [29] was the most frequently encountered taxon, was primarily influenced by the chemical properties of water and lived in heavily polluted areas [34]. *N. palea*, a nutrient-tolerant species, has been discovered in both intensively farmed agricultural areas and urban environments [35]. In the CART analysis of the six dominant diatom species, 10 variables (AT, CP, Agc, Wd, temp, DO, Cdv, Chl_a, COD, TP, DTP, and DTN) were used multiple times for decision making, as is shown in Table 5. The most commonly used variables in CT and RT were AT and Cdv (Table 7).

Table 7. The most common environmental variables used variables in the classification tree (CT) and regression tree (RT) analyses of six dominant species for decision making.

Code	Environmental Variable	Frequency	
		CT	RT
AT	Altitude (m)	2	2
CP	Canopy (%)	2	1
Agc	Agriculture (%)		1
Wd	River width (m)		1
Temp	Temperature (°C)		1
DO	DO (mg/L)	1	1
Cdv	Conductivity (μs/cm)	2	2
Chl_a	chlorophyll-a (μg/cm^3)	1	
COD	COD (mg/L)	1	
T_P	T-P (mg/L)		1
DTP	DTP (mg/L)		1
DTN	DTN (mg/L)	1	1

CT indicates whether an epilithic diatom is present, while RT indicates the relative likelihood of how many epilithic diatoms are likely to appear. Our findings distinguished CT from RT. Although certain environmental variables were not necessarily the most influential, based on the aforementioned analyses, we could determine the impact of environmental variables. Table 8 shows the change points of the CART and WA optima. CART denotes environmental variables with a changing point range, and the appearance ratio of the species has high values within the specified range, while WA denotes variables with a specific value, optima, which denotes the optimal condition for the species to appear.

The changing point range in CART for *A. minutissima* was not identical to the WA optima in the COD and DTP variables. WA optima were included in the range of changing points in CART for *A. convergens*. There was little difference between the range of changing points and WA optima for *N. inconspicua*, TP for *N. amphibia*, and DTN for *N. palea*. The values of the range of changing points in the CART and WA optima were not identical for *N. fonticola* (Table 6).

Table 8. Changing point for the classification and regression tree (CART) and the weighted average (WA) optima for six dominant species.

Code	CT Changing Point	RT Changing Point	WA Optima
A.m.	CP ≥ 2.50 COD < 3.25 DO < 16.86 Chl_a ≥ 3.595	Cdv < 154 DTP < 0.0025	COD = 3.33 DTP = 0.0497
A.c.	DTN < 4.218 pH ≥ 7.175 run ≥ 7.5	Wd ≥ 152 AT ≥ 102	AT = 147.722 DTN = 2.322
N.i.	AT < 90.5 CP ≥ 7.5	Agc ≥ 32.5 DO < 3.5	AT = 48.700
N.f.	AT < 134.5 PO4_P < 0.185	AT < 36 CP ≥ 25 temp < 16.75	AT = 62.537
N.a.	Cdv ≥ 120.5	T_P ≥ 0.0305	T_P = 0.229
N.p.	Cdv ≥ 172.5	DTN ≥ 4.297 Cdv ≥ 172.5	DTN = 5.908

CART and WA optima demonstrated that various physicochemical environmental factors, such as AT, Cdv, and nutrient concentration, influenced the occurrence and composition of epilithic diatoms. Changes in specific environmental variables may affect the abundance of species that prefer the same conditions.

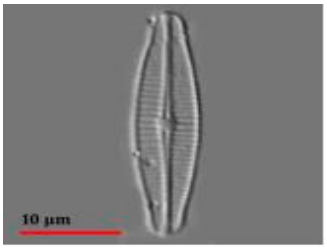
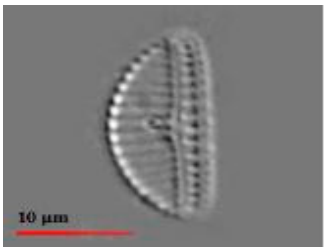
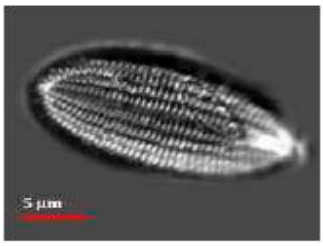
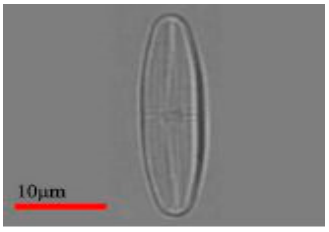
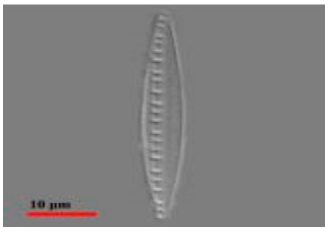
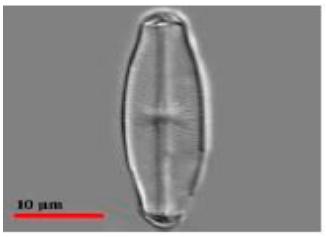
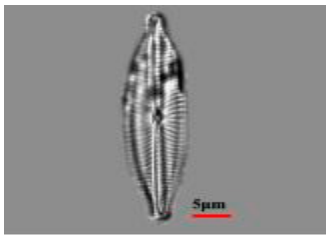
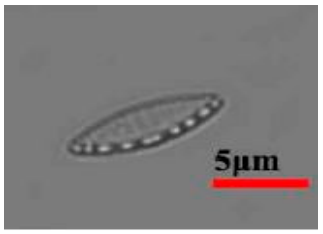
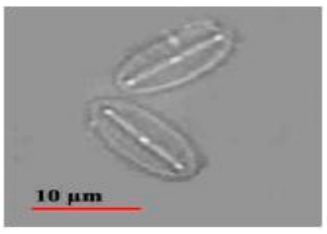
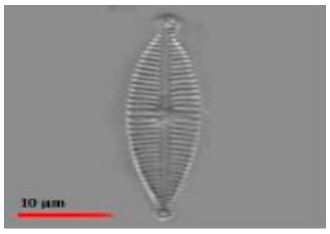
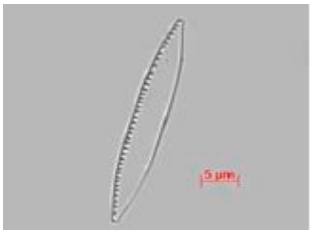
The CART and WA optima results for the six dominant species (*A. minutissima*, *A. convergens*, *N. amphibian*, *N. fonticola*, *N. inconspicua*, and *N. palea*) revealed that each species was influenced by a variety of complex environmental variables. Furthermore, the range of the changing point and WA optima did not always correspond with each other owing to the upper–lower rank interaction. The CART and WA approaches yielded distinct yet complementary information on the complex relationships between common stream diatoms and environmental variables [15].

A. minutissima and *A. convergens* were the dominant species in the reference condition in this study and were highly prevalent under the following conditions: conductivity <154 and DTP < 0.003, and altitudes >102 and width >152, respectively. As changing physical factors, such as altitude and width, which are the habitat conditions of *A. convergens*, to suit the habitat conditions of one type of diatom is difficult, the habitat environment of *A. minutissima*, which is primarily affected by chemical factors, should be considered. For example, if the conductivity and DTP in a specific river are changed to conductivity <154 and DTP < 0.003, the habitat conditions of *A. minutissima*, it is projected that the chance of maintaining good aquatic health would rise in tandem with the rate of the appearance of *A. minutissima*. However, more research and verification of the properties of organisms living in the same habitat are required before they can be used for river restoration.

The autecological properties of diatom species may be useful for objectively determining a reference condition. The environmental preferences of indicator diatoms are considered when developing river restoration policies and quantitative evaluation criteria for biological assessment.

Based on the findings of our study, *A. minutissima* and *A. convergens*, which had high dominant frequencies in groups C and D, respectively, belonged to the "excellent" and "good" classes in the TDI grade, while *N. inconspicua*, which had high dominant frequencies in groups A and B, also belonged to the "excellent" and "good" classes in the TDI grade, appeared to be a member of the "fair" class. The main species in Groups A and B, *N. palea*, belonged to the "poor" class and it was determined that the concentration of nutrients and the water conversion process were strongly associated (Table 9).

Table 9. Determination of biological grade using epilithic diatoms in ecological river restoration technical guide, Korea Ministry of Environment [23,36].

Class by TDI	Indicator Species		
Excellent			
	Achnanthes minutissima	*Cymbella minuta*	*Cocconeis placentula var. lineata*
Good			
	Achnanthes convergens	*Synedra ulna*	*Nizschia dissipata*
Fair			
	Navicula pupula	*Navicula viridula*	*Nitzschia inconspicua*
Poor			
	Navicula saprophila	*Gomphonema parvulum*	*Nitzschia palea*

According to the findings of this study, the abundance and composition of epilithic diatom species can be influenced not only by nutrient concentrations but also by a variety of physicochemical environmental factors. Furthermore, it was determined that even the frequency of occurrence of species living in the same habitat may change owing to changes in specific environmental factors.

5. Conclusions

Physical and chemical factors influence the composition of diatom assemblages. Diatom species respond differently to physicochemical variations and complex interactions among environmental variables. The autecological characteristics of diatom species may be useful for conducting an objective search for reference conditions. The environmental preferences of indicator diatoms should be considered when developing river restoration policies and quantitative evaluation criteria for biological assessment.

In this study, environmental factors influencing structural changes in epilithic diatom species were quantified. The composition and number of epilithic diatom species appeared to be influenced by a variety of complex environmental factors depending on the habitat. CART and WA analyses were performed on dominant species to quantitatively derive environmental factors affecting each species' relative abundance and to present environmental conditions reflecting each species' physiological and ecological characteristics.

The presence (0.75) of *A. minutissima* was determined as high under the CP $\geq$ 2.5% and DO < 16.9 mg/L conditions. The presence of *A. convergens* (0.58) was determined by the DTN < 4.218 mg/L condition. The presence (0.27) of *N. inconspicua* was determined by altitude < 90.5 m and CP $\geq$ 7.5%. The presence (0.50) of *N. fonticola* was determined by altitude < 134.5 m. The presence (0.64) of *N. amphibia* was determined by conductivity <120.5 μS/cm. The presence (0.64) of *N. palea* was determined by conductivity $\geq$172.5 μS/cm as well. CART analysis may help identify the hierarchical interactions among environmental variables in predicting the relative abundance of epilithic diatoms.

Research on the autecological characteristics and environmental preferences of indicator diatom species could aid in making objective decisions for the establishment of river restoration policies and quantitative evaluation criteria for biological assessments. In the future, it is expected that a clear set of integrated water quality evaluation criteria reflecting optimal diatom habitat conditions will be established. Basic data for the establishment of standards and the development of ecological river restoration technologies are expected to be available.

Author Contributions: Data curation and software, T.-Y.H.; conceptualization, Y.S.; methodology, T.-Y.H.; data analysis, Y.S. and T.-Y.H.; investigation, T.-Y.H. and Y.S.; writing—original draft, Y.S.; visualization, D.K. writing—review & editing, D.K. and T.-Y.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by a grant from the National Institute of Environment Research (NIER), funded by the Ministry of Environment of the Republic of Korea (NIER-2019-01-01-038). This work was partially supported by the research fund of the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education (2018R1D1A1B07047712, 2019R1 | 1A3A01057696).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Barbour, M.T.; Gerritsen, J.; Snyder, B.D.; Stribling, J.B. *Revision to Rapid Bioassessment Protocols for Use in Streams and Rivers: Periphyton, Benthic Macroinvertebrates, and Fish*; EPA 841-D-97-002; United States Environmental Protection Agency: Washington, DC, USA, 1999.
2. Hering, D.; Johnson, R.K.; Kramm, S.; Schmutz, S.; Szoszkiewicz, K.; Verdonschot, P.F.M. Assessment of European streams with diatoms, macrophytes, macroinvertebrates and fish: A comparative metric-based analysis of organism response to stress. *Freshw. Biol.* **2006**, *51*, 1757–1785. [CrossRef]
3. Reavie, E.D.; Jicha, T.M.; Angradi, T.R.; Bolgrien, D.W.; Hill, B.H. Algal assemblages for large river monitoring: Comparison among biovolume, absolute and relative abundance metrics. *Ecol. Indic.* **2010**, *10*, 167–177. [CrossRef]

4. Hlúbiková, D.; Novais, M.H.; Dohet, A.; Hoffmann, L.; Ector, L. Effect of riparian vegetation on diatom assemblages in headwater streams under different land uses. *Sci. Total Environ.* **2014**, *475*, 234–247. [CrossRef] [PubMed]
5. Jyrkänkallio-Mikkola, J.; Siljander, M.; Heikinheimo, V.; Pellikka, P.; Soininen, J. Tropical stream diatom communities—The importance of headwater streams for regional diversity. *Ecol. Indic.* **2018**, *95*, 183–193. [CrossRef]
6. Newall, P.; Walsh, C.J. Response of epilithic diatom assemblages to urbanization influences. *Hydrobiologia* **2005**, *532*, 53–67. [CrossRef]
7. Pan, Y.; Hill, B.H.; Husby, P.; Hall, R.K.; Kaufmann, P.R. Relationships between environmental variables and benthic diatom assemblages in California Central Valley streams (USA). *Hydrobiologia* **2006**, *561*, 119–130. [CrossRef]
8. US Geological Survey. Biological-Community composition in small streams and its relations to habitat, nutrients, and land use in agriculturally dominated landscapes in Indiana and Ohio, 2004, and implications for assessing nutrient conditions in Midwest Streams. In *National Water-Quality Assessment Program: Scientific Investigations Report 2009–5055*; United States Geological Survey: Reston, VA, USA, 2009; p. 21. [CrossRef]
9. Stevenson, R.J.; Pan, Y.; van Dam, H. Assessing environmental conditions in reivers and streams with diatoms. In *The Diatoms: Applications for the Environmental and Earth Sciences*, 2nd ed.; Smol, J.P., Stoermer, E.F., Eds.; Cambridge University Press: Cambridge, UK, 2010; pp. 57–85.
10. Korea Ministry of Environment. *A Study on the Development and Survey of Comprehensive Evaluation Method for Water Environment (II)-Final Report*; Korea Ministry of Environment: Sejong, Korea, 2005; p. 614.
11. Grenier, M.; Lavoie, I.; Rousseau, A.N.; Campeau, S. Defining ecological thresholds to determine class boundaries in a bioassessment tool: The case of the Eastern Canadian Diatom Index (IDEC). *Ecol. Indic.* **2010**, *10*, 980–989. [CrossRef]
12. Leland, H.V. Distribution of phytobenthos in the Yakima River basin, Washington, in relation to geology, land use, and other environmental factors. *Can. J. Fish. Aquat. Sci.* **1995**, *52*, 1108–1129. [CrossRef]
13. De'ath, G.; Fabricius, K.E. Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology* **2000**, *81*, 3178–3192. [CrossRef]
14. Clark, L.A.; Pregibon, D. Tree-based models. In *Statistical Models*; Chambers, S.J.M., Hastie, T.J., Eds.; Chapman & Hall: London, UK, 1993.
15. Weilhoefer, C.L.; Pan, Y. Using change-point analysis and weighted averaging approaches to explore the relationships between common benthic diatoms and in-stream environmental variables in mid-atlantic highlands streams, USA. *Hydrobiologia* **2008**, *614*, 259–274. [CrossRef]
16. Hershey, A.E.; Beaty, S.; Fortino, K.; Keyse, M.; Mou, P.P.; O'Brien, W.J.; Ulseth, A.J.; Gettel, G.A.; Lienesch, P.W.; Luecke, C.; et al. Effect of landscape factors on fish distribution in arctic Alaskan lakes. *Freshw. Biol.* **2006**, *51*, 39–55. [CrossRef]
17. Iverson, L.R.; Prasad, A.M. Predicting abundance of 80 tree species following climate change in the eastern United States. *Ecol. Monogr.* **1998**, *68*, 465–485. [CrossRef]
18. Hall, R.I.; Smol, J.P. A weighted-averaging regression and calibration model for inferring total phosphorus concentration from diatoms in British Columbia (Canada) lakes. *Freshw. Biol.* **1992**, *27*, 417–434. [CrossRef]
19. Pan, Y.; Stevenson, R.J.; Hill, B.H.; Herlihy, A.T.; Collins, G.B. Using diatoms as indicators of ecological conditions in lotic systems: A regional assessment. *J. N. Am. Benthol. Soc.* **1996**, *15*, 481–495. [CrossRef]
20. ter Braak, C.J.F.; Juggins, S. Weighted averaging partial least squares regression (WA-PLS): An improved method for reconstructing environmental variables from species assemblages. *Hydrobiologia* **1993**, *269–270*, 485–502. [CrossRef]
21. Bere, T.; Tundisi, J.G. Weighted average regression and calibration of conductivity and pH of benthic assemblages in stream influenced by urban pollution-São Carlos/SP, Brazil. *Acta Limnol. Bras.* **2009**, *21*, 317–325.
22. Kelly, M.G.; Whitton, B.A. The trophic diatom index: A new index for monitoring eutrophication in rivers. *J. Appl. Phycol.* **1995**, *7*, 433–444. [CrossRef]
23. Korea Ministry of Environment. *Ecological River Restoration Technical Guide*; Korea Ministry of Environment: Sejong, Korea, 2011; p. 320.
24. European Commission. Directive 2000/60/EC of the European parliament of the Council of 23rd October 2000 establishing a framework for community action in the field of water policy. *Off. J. Eur. Commun.* **2000**, *327*, 1–72.
25. USEPA. *Best Practices for Identifying Reference Condition in Mid-Atlantic Streams*; US Environmental Information: Washington, DC, USA, 2006; 8p. Available online: https://nepis.epa.gov/Exe/ZyPDF.cgi/P1000D3D.PDF?Dockey=P1000D3D.PDF (accessed on 14 March 2022).
26. Bailey, R.C.; Norris, R.H.; Reynoldson, T.B. *Bioassessment of Freshwater Ecosystems: Using the Reference Condition Approach*; Kluwer Academic Publishers: New York, NY, USA, 2004; 170p.
27. Atkinson, E.J.; Therneau, T.M. *An Introduction to Recursive Partitioning Using the RPART Routines*; Mayo Foundation: Rochester, MN, USA, 2000; 33p. Available online: https://www.stat.auckland.ac.nz/~{}yee/784/files/minitech.pdf (accessed on 14 March 2022).
28. Therneau, T.M.; Atkinson, E.J. *An Introduction to Recursive Partitioning Using the RPART Routines*; Mayo Foundation: Rochester, MN, USA, 1997; 52p. Available online: https://www.mayo.edu/research/documents/biostat-61pdf/doc-10026699 (accessed on 14 March 2022).
29. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.G. *Classification and Regression Trees*; Routledge: New York, NY, USA, 1984.
30. Juggins, S. Quantitative reconstructions in palaeolimnology: New paradigm or sick science? *Quat. Sci. Rev.* **2013**, *64*, 20–32. [CrossRef]

31. Green, R.H. A multivariate statistical approach to the Hutchinsonian niche: Bivalve molluscs of central Canada. *Ecology* **1971**, *52*, 543–556. [CrossRef]
32. Kwandrans, J.; Eloranta, P.; Kawecka, B.; Wojtan, K. Use of benthic diatom communities to evaluate water quality in rivers of southern Poland. *J. Appl. Phycol.* **1998**, *10*, 193–201. [CrossRef]
33. Medley, C.N.; Clements, W.H. Responses of diatom communities to heavy metals in streams: The influence of longitudinal variation. *Ecol. Appl.* **1998**, *8*, 631–644. [CrossRef]
34. Lange-Bertalot, H. Pollution tolerance of diatoms as a criterion for water quality estimation. *Nova Hedwig.* **1979**, *64*, 285–305.
35. Fore, L.S.; Grafe, C. Using diatoms to assess the biological condition of large rivers in Idaho (U.S.A.). *Freshw. Biol.* **2002**, *47*, 2015–2037. [CrossRef]
36. Korea Ministry of Environment. *Aquatic Ecosystem Health Survey and Evaluation*; Korea Ministry of Environment: Sejong, Korea, 2010; p. 554.