Article

Regression Tree Analysis for Stream Biological Indicators Considering Spatial Autocorrelation

Mi-Young Kim [1] and Sang-Woo Lee [2],*

1 Graduate Program, Department of Environmental Science, Konkuk University, Gwangjin-Gu,
   Seoul 05029, Korea; kmy91321@konkuk.ac.kr
2 Department of Forestry and Landscape Architecture, Konkuk University, Gwangjin-Gu, Seoul 05029, Korea
* Correspondence: swl7311@konkuk.ac.kr; Tel.: +82-2-450-4120

Abstract: Multiple studies have been conducted to identify the complex and diverse relationships between stream ecosystems and land cover. However, these studies did not consider spatial dependency inherent from the systemic structure of streams. Therefore, the present study aimed to analyze the relationship between green/urban areas and topographical variables with biological indicators using regression tree analysis, which considered spatial autocorrelation at two different scales. The results of the principal components analysis suggested that the topographical variables exhibited the highest weights among all components, including biological indicators. Moran's I values verified spatial autocorrelation of biological indicators; additionally, trophic diatom index, benthic macroinvertebrate index, and fish assessment index values were greater than 0.7. The results of spatial autocorrelation analysis suggested that a significant spatial dependency existed between environmental and biological indicators. Regression tree analysis was conducted for each indicator to compensate for the occurrence of autocorrelation; subsequently, the slope in riparian areas was the first criterion of differentiation for biological condition datasets in all regression trees. These findings suggest that considering spatial autocorrelation for statistical analyses of stream ecosystems, riparian proximity, and topographical characteristics for land use planning around the streams is essential to maintain the healthy biological conditions of streams.

Keywords: regression tree analysis; biological indicators; spatial autocorrelation; land use; principal component analysis

1. Introduction

Streams and rivers are complex lotic systems witnessing continuous changes. Interactions among multiple factors, such as topography, physiochemistry, and hydraulics in watersheds directly and indirectly affect the stream condition [1,2]. Previous studies have examined the relationship between instream habitat quality and aquatic organisms [3–5]. As instream habitat quality is determined by watershed and riparian ecosystem conditions, landcover, topography, physical habitats, and organisms of these systems are critical factors that determine the stream water quality and biological condition [6,7]. In particular, the surrounding land use strongly influences the biological integrity of streams, and thus, multiple studies have been conducted to identify the relationship between watershed land use and biological condition [5,8–15]. Recently, watershed topography has significantly changed due to the increasing population and high concentration of industries in urban areas. Buildings, roads, and parking spaces in urbanized areas along with anthropogenic uses of watersheds for urbanization, cultivation, and other activities have altered surface water characteristics, hydraulic and hydrologic systems, and the amount of pollutant emissions and pollutant loads in watersheds. Consequently, these alterations have degraded physiochemical and ecological conditions of streams [16–18]. Additionally, previous studies reported that riparian forests adjacent to streams directly influence the flow mechanisms of sediments, organic matter, and nutrients from watersheds [19,20]. Riparian forests have

various functions, such as preventing bank erosion, reducing organic matter and nutrients, controlling water temperature, providing habitats for aquatic and terrestrial organisms, increasing ecological health and diversity, and mitigating the negative effects of land use on rivers [19–33]. Therefore, the relationships between land cover (particularly, green and urban areas) and stream conditions need to be studied in detail to develop programs for improving the stream health and implementing stream restoration and management strategies [34].

As reported previously, the extent of forested areas and the width of riparian areas strongly influence biological communities [35,36]. However, specifying buffer widths is challenging because the buffer scale typically used varies depending on topography, geology, hydrology, rainfall intensity, and vegetation type [37]. Additionally, the responses of stream ecosystems to anthropogenic modification (e.g., land use, subsurface modification, groundwater abstraction, stream channelization, and damming) depends on the modification types and scales [38,39]. Thus, buffer widths might need to be determined based on the specific response (e.g., water availability, productivity, water quality, the composition of species, microclimate regulation, habitat loss, flow regime) and scale (e.g., catchment scale, landscape scale, segment scale) of the stream ecosystem. Damanik-Ambarita et al. [38] conducted an extensive review about the various riparian and catchment scales used in previous studies as a part of a study in exploring the quantitative relationship between land use and the aquatic macroinvertebrate community in previous studies. According to their study, the scale of the riparian buffer used in previous studies is ranging from 30 to 5000 m width (or radius) [40–42]. Other researchers have proposed multiple spatial scales and extents to identify the relationship between land use and stream ecosystem [43,44]. For example, Burdon et al. [31] used 5 m buffer scale in testing the relationships between riparian integrity with ecological status in European streams. Furthermore, in another study, the buffer width differed according to the study purpose, for example, 30 m for water quality conservation, 150 m for protection from flood, and more than 500 m for riparian habitats and biological communities [45]. Yirigui et al. [37] reported that a buffer width greater than 500 m is necessary for nutrient control, bank stabilization, aquatic and terrestrial wildlife habitat, and microclimate change mitigation. The Korean Ministry of Environment (MOE) has adopted two buffer widths, 500 m and 1 km, to preserve riparian areas and conserve drinking water quality [46]. Kim et al. [47] studied the relationship between fish and land use using multiple spatial scales in a 5 km buffer zone and reported that the effects of forest areas were higher at a high spatial scale.

Studies conducted to identify the relationship between landscape variables and biological integrity of streams have used conventional correlation and regression analyses, which statistically assume that all variables are independent [48]. However, the impacts of landscape variables on the biological communities in streams are not completely independent and instead, exhibit spatial autocorrelation [49–51]. Spatial autocorrelation is a phenomenon in which a variable at one point correlates with the same variable around it in a two-dimensional space [48,52,53]. Lee et al. [54] indicated that spatial autocorrelation of landscape characteristics can occur because of the continuity and hierarchical structure of environmental conditions. Therefore, spatial autocorrelation is critical because the measured environmental variables are not random but have a common structural variance [55]. As conventional statistical methods assume randomness of the measured data, using a new approach to address spatial autocorrelation is necessary [56]. However, it is noteworthy that some studies argued that local conditions and sampling methods, such as watershed-based sampling [57], and multi-distance sampling can overrule the effects of spatial autocorrelation in some cases [58].

This study aims to analyze the relationships between green spaces, urban areas, and biological integrity of streams at two different scales in Nam-Han River Basin. The objectives were (i) to identify spatial autocorrelation of biological indicators represented by diatoms, macroinvertebrates, and fish, (ii) to analyze the relationship for each biological indicator with land cover (green and urban areas) and topographical variables using two

different riparian scales (5 km and 500 m), and (iii) to suggest scientific strategies for the restoration and management of green spaces and streams in the basin. The findings would thus, provide insights for improving water quality and the biological condition of streams in aquatic ecosystems.

2. Materials and Methods

2.1. Study Area and Sampling Site

The Nam-Han River Basin is located centrally in the Korean Peninsula and is a part of the major streams of the Han River, the largest water system in Korea (Figure 1). The Nam-Han River is a major river that forms the Han River System along with the Buk-Han River. It covers approximately one-third (12,577 km^2) of the total area of the Han River System (34,473 km^2), and its total length is 375 km [59]. The average annual rainfall in Korea is 1159 mm and that in the Nam-Han Basin is 1409 mm. Approximately 65% of the total annual precipitation occurs from June to September and the rivers are parched from March to May and October to November. The Nam-Han River Basin is located at a high elevation (more than 1000 m above sea level) [60]. Urbanized areas, including cultivation areas in the Nam-Han River Basin are concentrated downstream, and cover approximately 24% of the total area. Furthermore, forests cover approximately 70% of the total area of the basin [61].
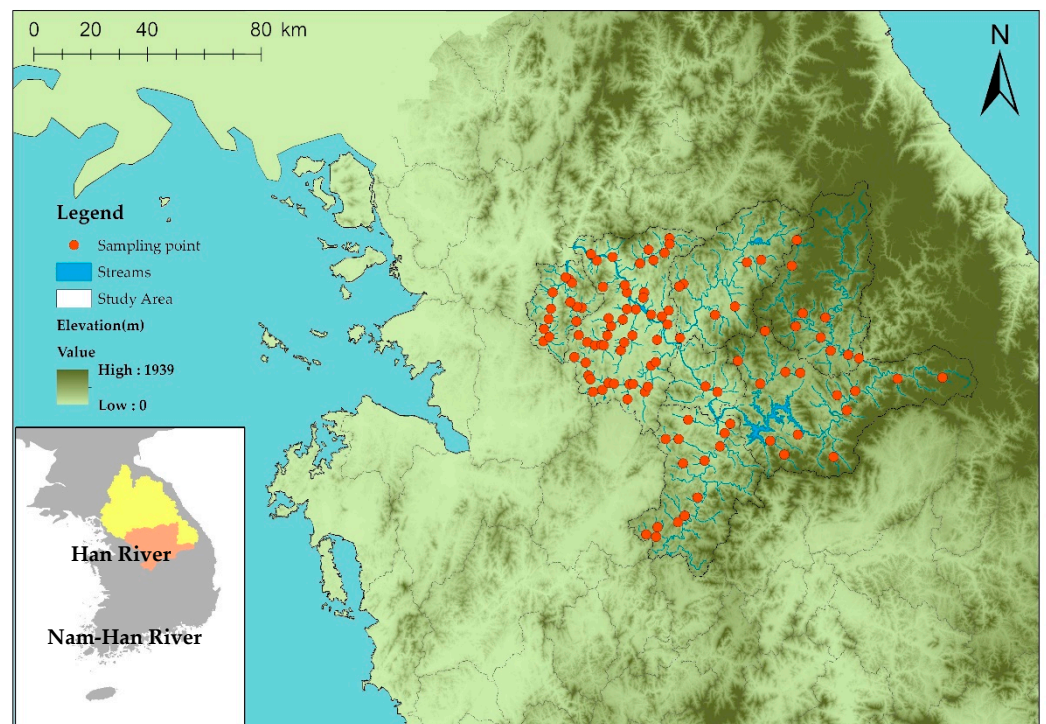

Figure 1. Sampling sites in the Nam-Han River Basin in Korea selected by the National Aquatic Ecological Monitoring Program.

Since 2007, the Korean MOE has been conducting a National Aquatic Ecological Monitoring Program (NAEMP) to monitor the water quality and ecological condition of streams and rivers across the country. NAEMP is significant for stream ecosystem research because it investigates and evaluates various aquatic ecosystem characteristics, such as structure, vegetation, water quality, and ecological health of rivers [62]. Additionally, the MOE has been monitoring aquatic properties, such as physiochemical water quality index, including total nitrogen (TN), total phosphorous (TP), biochemical oxygen demand (BOD), chemical oxygen demand (COD), dissolved oxygen (DO), chlorophyll a (Chl-a), and pH; biological indicators, including trophic diatom index (TDI), benthic macroinvertebrate index (BMI), and fish assessment index (FAI); and topographical characteristics, including

width, depth, and velocity of the five major rivers through 800 monitoring sites [63]. In this study, we sampled 107 sites mentioned in the tributary datasets of the Nam-Han Basin in 2018 (excluding the main stream because it has complex and diverse effects apart from effects of land cover) to investigate the relationship between stream biological conditions, green and urban areas, and topographic variables.

2.2. Biological Indicators of Streams

In this study, scores of three biological indexes developed by NAEMP were used as indicators of stream biological conditions in the Nam-Han River Basin. A score of 0 to 100 was assigned for each biological indicator, which were later classified into Class A (Excellent), Class B (Good), Class C (Fair), or Class D (Poor). The classification is relative to the biological state of streams having similar observations based on the national distribution and therefore, it is not considered to be an absolute outcome. Table 1 provides the equations for computing the biological indicators (TDI, BMI, and FAI) developed by NAEMP [64].

Table 1. Equations for computing the biological indicators developed by National Aquatic Ecological Monitoring Program (NAEMP).

Biological Indicators	Equations
TDI (Trophic Diatom Index)	TDI = 100 − {(WMS × 25) − 25} WMS: weighted mean sensitivity $WMS = \sum A_j \cdot S_j \cdot \frac{V_j}{\sum A_j \cdot V_j}$ Where, j = species A_j = abundance (proportion) of species j in the sample (%) S_j = pollution sensitivity ($1 \leq S \leq 5$) of species j V_j = indicator value ($1 \leq V \leq 3$)
BMI (Benthic Macroinvertebrates Index)	$\mathrm{BMI} = \left\{ 4 - \sum_{j=1}^{n} K_j H_j G_j \Big/ \sum_{j=1}^{n} H_j G_j \right\} \times 25$ where, j = number assigned to species n = number of species K_j = unit saprobic value of species j H_j = frequency of species j G_j = indicators weight value of species j
FAI (Fish Assessment Index)	FAI = sum of 8 metrics Metric 1 (M1): number of Korean native species Metric 2 (M2): number of rifle benthic species Metric 3 (M3): number of sensitive species Metric 4 (M4): percentage of tolerant species Metric 5 (M5): percentage of omnivores Metric 6 (M6): percentage of insectivores Metric 7 (M7): the amount of collection native species Metric 8 (M8): percentage of fish abnormalities

Algae are important indicators of water quality change or water pollution and are widely used to evaluate the health of rivers [65]. TDI estimates diatom conditions based on species abundance and sensitivity and evaluates the nutritional conditions of stream ecosystems [66]. It is computed by calculating the composition using weighted mean sensitivity (WMS) measurements and proportion of benthic diatom taxa (Table 1).

Benthic macroinvertebrates are important biotic components of the river ecosystem and a food source for fish. Since they are extremely sensitive to habitat disturbances, they serve as an important biological indicator [67]. They exhibit distinct differences in the cluster structure, such as the number of species and distribution of populations [68]. BMI represents the condition of macroinvertebrate communities in stream ecosystems and describes the state of the communities based on habitat and environmental changes. The

number assigned to each species, unit saprobic values, and frequencies are used as weight indicators for the species [69].

Fish as a biological indicator reflect the long-term environmental impacts within an aquatic ecosystem [70]. NAEMP developed the FAI using eight evaluation metric models (M1–M8) by analyzing the ecological characteristics of Korean fish communities. These metrics can be divided into four categories: composition of species, nutritional composition, fish abundance, and individual health [62,71].

2.3. Measurements and Selection of Scale

To measure the area of green and urban areas in the Nam-Han River Basin, we used land use land cover (LULC) data provided by the Korean MOE. The urbanized class was represented by "Urban," while the forest and grass classes were represented by "Green." The respective areas of each class were extracted using ArcMap software version 10.1. The areas were transformed into grids using the buffer features; additionally, the ratio of areas in each buffer was calculated as PLAND using the spatial pattern metric computing program Fragstats version 4.2 (The University of Massachusetts, Amherst, US) [72]. PLAND quantifies the area (in percentage) of each patch to the entire buffer and ranges between 0 and 100. It is a measure of landscaping composition that indicates the range of landscape with certain patches [73,74]. In addition, the DEM (digital elevation model) acquired from the Han River Flood Control Office and the National Geographic Information Institute was used to extract data on topographical variables in the Nam-Han River Basin for calculating the average elevation and slope for each buffer area.

In this study, two buffer scales were selected based on the sampling points to investigate the impact of green and urban areas and topographical variables on biological indicators. Multiple different scales have been used in previous studies [38,75]. Generally, the effectiveness of the riparian buffer improves with an increase in the buffer width, but various landscaping indicators have been reported to have different effects at different scales [76]. Furthermore, different maximum buffer sizes have been proposed based on previous research [77]. Accordingly, we selected (i) circle buffers with a 5 km radius from the sampling points and (ii) linear buffers of 500 m from the riparian areas to compare the large-scale effects of variables near the river (Figure 2).
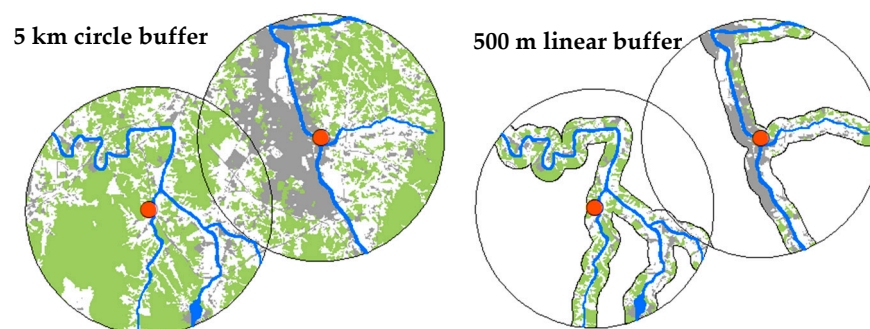

Figure 2. Circle buffer scale set to 5 km radius in the watershed and linear buffer set to 500 m in the riparian area near the sampling sites (green color; forest and grass areas, gray color; urbanization areas, blue color; streams, orange circle; sampling points).

2.4. Spatial Autocorrelation and Data Analysis

Spatial autocorrelation is commonly measured by methods that use a single exponent, such as Moran's *I* (MI, Equation (1); where *n* is the number of data, *X* is the pixel value, $\overline{X}$ is the mean of pixel values within the neighborhood range, and W_{ij} is the spatial weight matrix) and Geary's C, and methods that use variograms [78]. The single exponent method is useful for measuring the overall spatial autocorrelation of the study site, whereas variograms are useful for understanding the spatial distribution and structure of the autocorrelation [54,79]. In this study, we used MI to verify the spatial autocorrelation of the

biological indicators of the Nam-Han River watershed. The MI values for the entire region were calculated using GeoDa version. 1.18 designed by Luc Anselin [80]. The theoretical range of correlation coefficients for MI is between −1 and 1, where coefficient values closer to 1 or −1 indicate a higher spatial autocorrelation, while a value closer to 0 indicates low spatial autocorrelation [81]. Spatial autocorrelation in the observed data can be corrected by separate sampling at a certain effective distance or by performing statistical corrections [82,83]. However, sampling at effective distances is practically difficult. Therefore, in this study, we conducted an analysis that allowed us to consider spatial autocorrelation as a statistical correction.

$$I = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij} (X_j - \overline{X})}{S_0 \sum_{i=1}^{i=n} (X_i - \overline{X})^2} \tag{1}$$

Principal component analysis (PCA) was performed to identify the correlation between the variables. PCA is frequently used as a preliminary test prior to performing other statistical analyses, and not as a final analysis method. Furthermore, the relationship between the variables can be graphically visualized, facilitating the identification of correlation. Moreover, normalization can rescale the data to identify the weights of the principal components (PCs), thus allowing the reduction of the variable dimensions based on the accumulation of the explanatory ability for critical variables [84]. In this study, we conducted PCA to estimate the relationship between the biological indicators, topographical variables, and PLAND of green/urban areas in the Nam-Han River Basin using the "prcomp" and "PCA" packages in R statistical language. PCA biplots provide a graphical visualization, describing the interrelationships between multiple datasets using vectors. The angle and direction of vectors indicate correlations between the original dataset and PCs; if the vector is highly parallel to the PCs, the impact (i.e., correlation) on the particular PC is high. PCA analysis is often used because it provides excellent visualization of the variables from multiple perspectives in the form of two-dimensional graphs [85].

Thus, since MI values extracted from biological indicators have a strong spatial autocorrelation, we conducted regression tree analysis, which is a statistical correction method for dependent samples. It is useful to acquire a sufficient number of samples devoid of spatial autocorrelation to facilitate easy interpretation [86,87]. Compared to usual linear models, the regression tree provides better prediction because it is appropriate for analysis of complex ecological data that can include nonlinear relationships between various variables, missing values, and more complicated interactions [83,88]. Regression trees with having hierarchical structures explain variations of responding variables by splitting predictor variables at certain thresholds minimizing the variance of responding variables. Regression tree analysis has been widely used in ecological studies [83,89–93]. The tree of the model grows repeating binary splitting of the data, and each split that is defined based on a single explanatory variable forms two nodes [94,95]. Selecting the size of the tree is performed through cross-validation, and the selected tree has the smallest predicted mean square error [96]. Therefore, in addition to PCA, we used regression tree analysis to maintain the maximum number of observations and solve the spatial autocorrelation problem inherent in the dataset. Additionally, the regression tree can provide a foundation for step-by-step examination of the effects of variables and assist decision making, unlike previous studies that used regression or simply correlation analysis. We used the "rpart" package along with ANOVA (analysis of variance) in R statistical language and generated one tree for each of the three biological indicators to identify its specific relationship with the variables of the Nam-Han River Basin. According to Everaert et al. [95], different parameterizations (e.g., the number of cross-validations, pruning, and the minimum number of observations per leaf) in regression tree analysis can result in considerably different outcomes. However, parameters were not specified in regression tree analysis. We selected the models that minimize 'rel error' (i.e., error on the observations used to estimate the model) and 'xerror" (i.e., error on the observations from cross validation data) without overfit issue (i.e., simplest model) [97] (see Everaert

et al. [95], for a more detailed description on how regression trees can vary due to different parameter settings in regression tree analysis).

3. Results

3.1. Descriptive Statistics

Table 2 provides a summary of descriptive statistics for the biological indicators, including the mean values of topographical variables, such as slope and elevation, and green/urban areas in the riparian areas within the 5 km circle buffer and 500 m linear buffer. The minimum values of TDI, BMI, and FAI were 11.20, 22.60, and 21.90, whereas their maximum values were 95.20, 94.70, and 100.0, respectively, suggesting that the biological conditions of the Nam-Han River show variability based on different sites within the basin. The standard deviation of FAI was 18.57, which was slightly lower than that of TDI and BMI, indicating marginal changes in FAI at the sampling sites. Furthermore, the standard deviation of TDI (19.78) was the highest among the three indicators, thus, suggesting relatively substantial changes in the diatom status.

Table 2. Descriptive statistics of biological indicators, topographical features, and green/urban areas in the Nam-Han River Basin.

Classification	Variables	Min.	Max.	Mean	S.D.
Biological indicators	TDI	11.20	95.20	56.08	19.78
	BMI	22.60	94.70	65.52	19.19
	FAI	21.90	100.0	55.44	18.56
Topographical features	5 km Mean elevation (m)	48.85	764.73	212.00	139.73
	500 m Mean elevation (m)	38.90	589.63	158.91	109.00
	5 km Mean slope (%)	1.99	23.29	10.06	4.91
	500 m Mean slope (%)	1.64	21.54	7.85	4.87
Green spaces	5 km Forest and Grass area (%)	4.58	90.91	50.10	23.87
	500 m Forest and Grass area (%)	4.82	78.21	41.06	19.73
Urban areas	5 km Urban area (%)	0.98	91.74	14.56	19.48
	500 m Urban area (%)	2.33	66.06	13.85	12.52

The health score of the three biological indicators was defined by NAEMP as shown in Table 3 [64]. The mean values of TDI, BMI, and FAI indicated that the health score of the biological indicators in the Nam-Han River basin, were 56.08, 65.53, and 55.45, respectively (Table 2). Thus, according to the classification criteria defined by NAEMP, the condition of diatoms, macroinvertebrates, and fish in the Nam-Han River basin was "Fair," "Fair," and "Poor." FAI showed the lowest mean value, whereas the mean value of BMI was the highest.

Table 3. Classification of biological indicators according to NAEMP.

Biological Conditions	Class	TDI	BMI	FAI
Good	A	$60 \leq$ TDI ≤ 100	$80 \leq$ BMI ≤ 100	$87.5 \leq$ FAI ≤ 100
Fair	B	$45 \leq$ TDI < 60	$60 \leq$ BMI < 80	$56.2 \leq$ FAI < 87.5
Poor	C	$30 \leq$ TDI < 45	$45 \leq$ BMI < 60	$25 \leq$ FAI < 56.2
Very Poor	D	$0 \leq$ TDI < 30	$0 \leq$ BMI < 45	$0 \leq$ FAI < 25

Moreover, the mean value of urban areas (14.56) in the 5 km circle buffer was significantly lower than that of green areas (50.10), suggesting a higher proportion of forests and grasslands than urban areas in the Nam-Han River Basin. The mean values of urban and green areas for 500 m linear buffers were 13.85 and 41.07, respectively, although the mean value of urban areas did not differ significantly depending on buffer scale. The proportion of green areas was higher than that of riparian areas since the mean value of the 5 km buffer was greater than that of the 500 m buffer. The mean values of topographical variables for 5 km and 500 m buffer zones were 212.00 and 158.92, respectively, for elevation and 10.07

and 7.85, respectively, for slope, indicating higher elevation and steeper slope of the terrain in the wide range buffers than in the riparian area.

3.2. Principal Component Analysis

PCA was conducted to examine and describe the degree of variation of the variables by the newly developed PCs (Table 4). Additionally, PCA presents the variables that have the greatest influence on each PC. Positive and negative signs denote direction, and the influence of variables is dependent upon the absolute value of the coefficient (Table 5). For example, PC1 represents the first PC, which can be expressed as a linear combined equation of the following variables:

$$\text{PC1} = (\text{TDI} \times 0.70) + (\text{BMI} \times 0.83) + \ldots + (\text{Elev\_50 m} \times 0.85) + (\text{Slope\_500 m} \times 0.93). \tag{2}$$

Table 4. Proportion of variance and cumulative proportion for each principal component.

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
Proportion of Variance	0.50	0.15	0.09	0.07	0.06	0.03	0.02	0.02	0.02	0.00	0.00
Cumulative Proportion	0.50	0.65	0.75	0.82	0.88	0.92	0.95	0.97	0.99	0.99	1.00

Table 5. Rotated factor matrix extracted from principal component analysis. The underline values are the largest weighs on that PCs.

Variable	Principal Component				
	PC1	PC2	PC3	PC4	PC5
TDI	0.70	−0.02	0.11	0.48	−0.28
BMI	0.83	−0.11	0.14	0.16	−0.21
FAI	0.81	0.03	0.01	−0.07	−0.35
Urban_5 km	0.25	<u>0.82</u>	−0.21	0.17	0.18
Green_5 km	0.20	<u>−0.79</u>	0.39	−0.12	0.19
Elev_5 km	<u>0.88</u>	0.07	−0.15	−0.32	0.16
Slope_5 km	<u>0.92</u>	0.09	0.14	−0.02	0.10
Urban_500 m	−0.12	0.50	<u>0.78</u>	0.05	0.24
Green_500 m	0.52	-0.37	−0.28	<u>0.49</u>	<u>0.47</u>
Elev_500 m	<u>0.85</u>	0.05	−0.15	−0.39	0.13
Slope_500 m	<u>0.93</u>	0.10	0.14	−0.03	0.01

PC1, which represented topographical variables, such as elevation and slope, accounted for 50% of the total variation, whereas PC2, which represented the proportion of urban and green areas at a 5 km buffer, accounted for 15.87% of the total variation. Thus, the cumulative explanatory power of the dataset by PC1 was 65.87% (Table 4). If the weight of a PC is highly directed to the positive direction, it indicates good biological condition. Conversely, if weight is highly directed to the negative direction, it represents poor condition. In PC1, all topographical variables a high weight. However, in PC2, the proportion of urban and green areas within the 5 km buffer exhibited the highest weight in the opposite direction. Furthermore, the proportion of urban and green areas within the 500 m buffer represented the highest weight in PC3 (9% of the total variance) and PC4 (7% of the total variance), and 82% of the dataset were explained by PC1–PC4.

Figure 3 is a biplot of the PCA that presents an effective visualization of the PCs. The x-axis of the biplot denotes Dim1 (primary PC), and the y-axis denotes Dim2 (secondary PC). Variables in the same direction are seen to be corresponding variables. The graph briefly represents the weights of the variables that affect PC1 or PC2, and their corresponding spatial locations indicate similarities between them. Thus, the figures show that the biological indicators (TDI, BMI, and FAI) and topographical variables (elevation and slope) represent similar trends, and urban and green areas represent perpendicular (high and

low) trends. In particular, except for the topographical variables, the green area in the 500 m buffer was the closest related variable to biological indicators because the vectors were parallel to the PC axis as a closer distance indicates a stronger impact. Meanwhile, differences in the buffer scale were observed to be perpendicular to the vertical axis. Green and urban areas in the 5 km buffers had a stronger effect than 500 m buffer areas in PC2. Additionally, green areas exhibited a positive relationship with the three biological indicators, whereas urban areas were negatively associated in the 500 m buffer. However, the directionality of the 5 km and 500 m scales in urban areas differed as the urban areas did not have a significant impact on the biological indicators at a wider range. Therefore, the analysis focused on identifying the effect of the impact of the relationship on biological indicators at each scale and type of area.
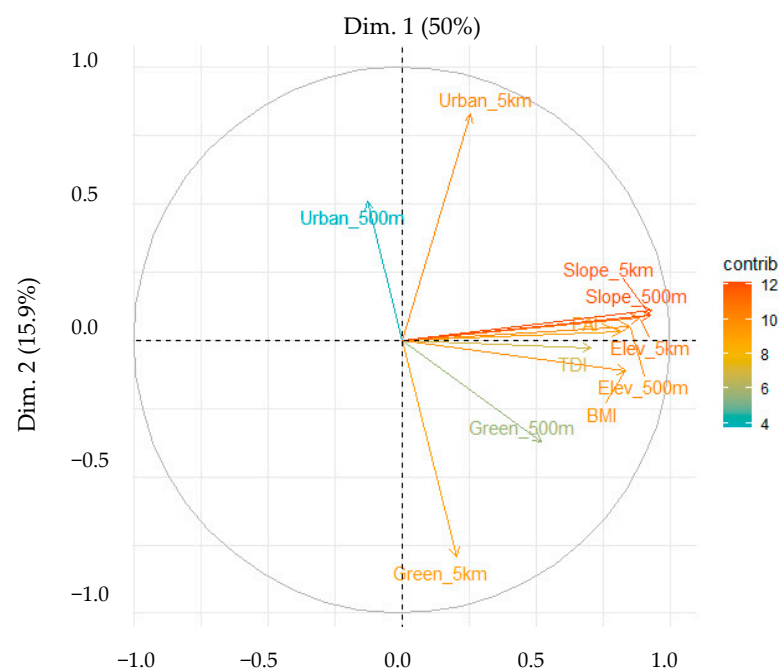

Figure 3. Principal component analysis biplot for the first two principal component scores.

3.3. Spatial Autocorrelation Verification of Biological Indicators

Analysis of the spatial autocorrelation of TDI, BMI, and FAI in the Nam-Han River Basin using GeoDa indicated high spatial autocorrelation of the biological indicators (Figure 4). In particular, the spatial autocorrelation coefficient (MI) for BMI was the highest (0.834), which was followed by that of FAI (0.785) and TDI (0.779). The topographical and land use variables and biological indicators showed MI values of 0.5 or higher, indicating a significant spatial dependency of the variables ($p < 0.01$) [98]. The high spatial autocorrelation of the biological indicators indicates that the spatial distribution of stream biological conditions depends on the hierarchical structure of environmental conditions [54]. Thus, the stream biological condition at any point is not entirely independent and instead, depends on the biological health around specific locations. These results suggest that conducting sampling and statistical analyses without considering spatial autocorrelation of stream biological indicators can cause potential errors.

Furthermore, the MI values estimated in this study do not only represent the spatial autocorrelation or dependency of the measurements across the variables but also clustering or dispersibility of the variables. The values represent a random distribution close to zero, strong clustering close to +1, and strong dispersion of uniform close to −1; therefore, the high MI values of the three biological indicators of the Nam-Han River Basin indicate that their conditions are clustered [81].
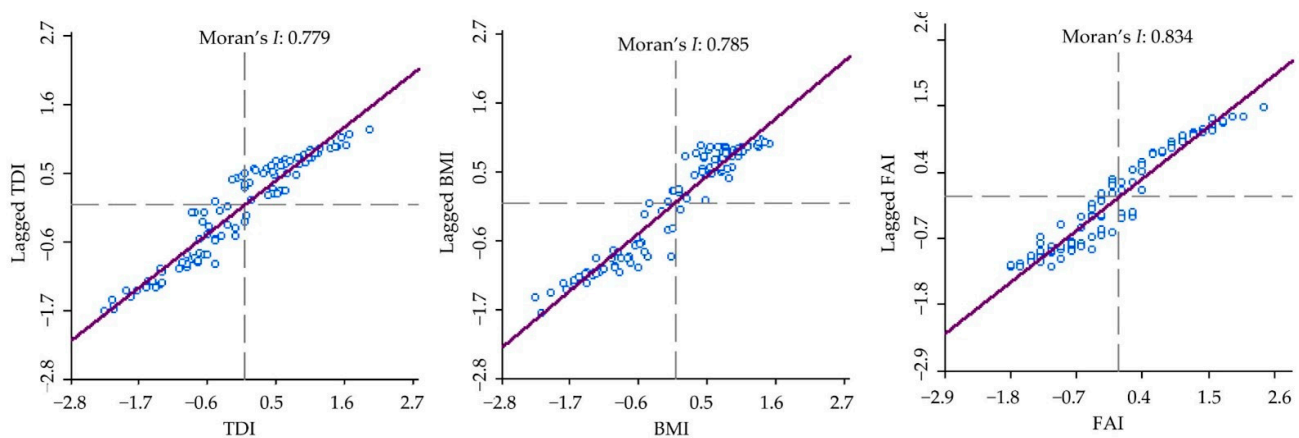
Int. J. Environ. Res. Public Health **2021**, *18*, 5150

10 of 19

Figure 4. Moran's *I* values of the three biological indicators.

3.4. Regression Tree Analysis

The slope of the 500 m buffer was the first variable to determine the differentiation for the biological indicators of the Nam-Han River Basin. This was consistent with the PCA results, which indicated that the slope of the 500 m buffer had the highest weight among the first PCs (0.93). However, the biological indicator results differed after Node1.

The TDI regression tree indicates that the slope within the 500 m linear buffer in the riparian areas is the most significant variable to describe TDI (Figure 5). The observations were divided into two groups according to the slope, wherein groups with a slope less than 7.6% denoted a poor TDI condition, while groups with slopes more than 7.6% indicated a good TDI condition. The elevation of the 5 km circle buffer was the second most important variable to describe the TDI score in the group with a slope less than 7.6% (left side of the tree). A high elevation increased the mean TDI value by 21% and was associated with a good status of diatoms. In areas with slopes of 7.6% or higher (right side of tree), the proportion of green areas in the 500 m linear buffer was the second most important variable to explain the TDI status. In areas with relatively high slopes, a high proportion of green areas near the riparian system indicated better biological conditions of diatoms. After the mean elevation of the 5 km buffer, the green area in the 500 m buffer and 5 km buffer represented the third important variable on the left and right sides of the tree, respectively. The R^2 value of the regression tree for TDI was 0.67, and the relative importance of the variables was 20% (500 m slope), 18% (5 km slope), 16% (5 km elevation), 15% (500 m elevation), 13% (500 m green area), 12% (5 km green area), 3% (500 m urban area), and 2% (5 km urban area). Other than the topographical variables of the stream, an increased proportion of the forest and grass area near the riparian area, represented by the 500 m buffer green rate, improved the condition of diatoms.

The first differentiation in the BMI regression tree was also the slope within 500 m of the riparian area (Figure 6). In the lower slope group (left of the tree), the second most important variable was the green area in the 5 km circle buffer, which was again classified as the urban area of the 500 m buffer at the divided node. However, the second most important variable in the high slope group (right side of the tree) was elevation. Low values of elevation in the 5 km buffer were affected by the BMI condition in the green area of the 500 m buffer. Furthermore, the wide range of green areas with lower slopes (i.e., relatively less fluid flow) had a significant impact on the condition of macroinvertebrates, suggesting that the green area near the stream was comparatively more important among the areas with high slopes (i.e., areas with more fluid flow). The R^2 value of the BMI regression tree was 0.78, and the relative importance of the variables was 20% (500 m slope), 15% (5 km elevation), 14% (500 m elevation), 14% (5 km green area), 13% (500 m green area), 3% (500 m urban area), and 2% (5 km urban area). Similar to the TDI results, the proportion of the green areas at wide ranges, represented by the proportion of green areas

within the 5 km buffer, influenced the improvement of macroinvertebrate conditions along with the stream topographical variables.
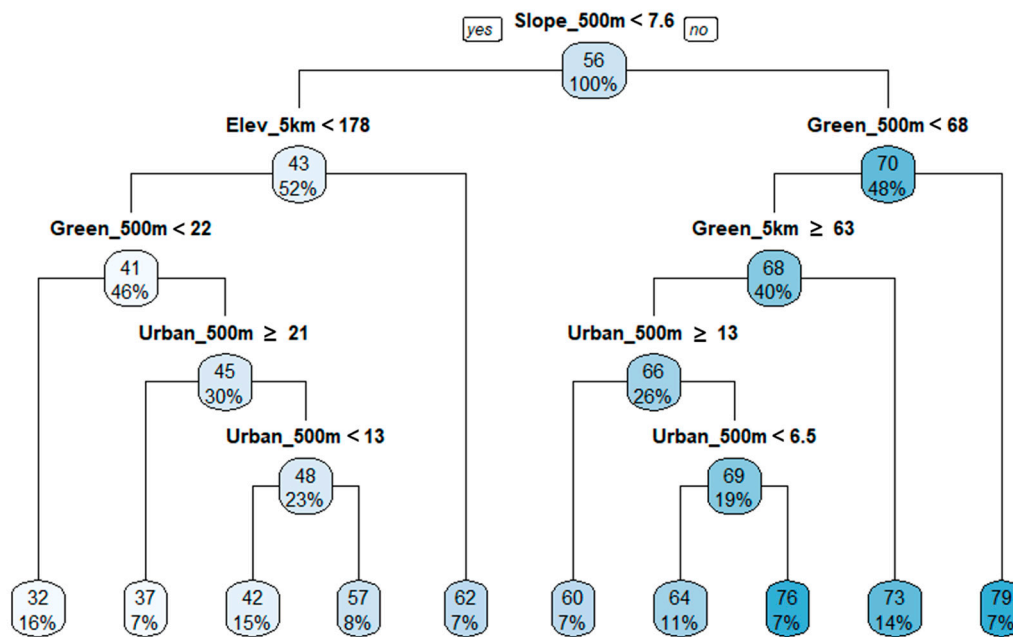

Figure 5. Regression tree for TDI with stream topographical variables and proportions of green and urban areas in buffer areas ($R^2 = 0.67$). The numbers located on top in color boxes are mean values of TDI.
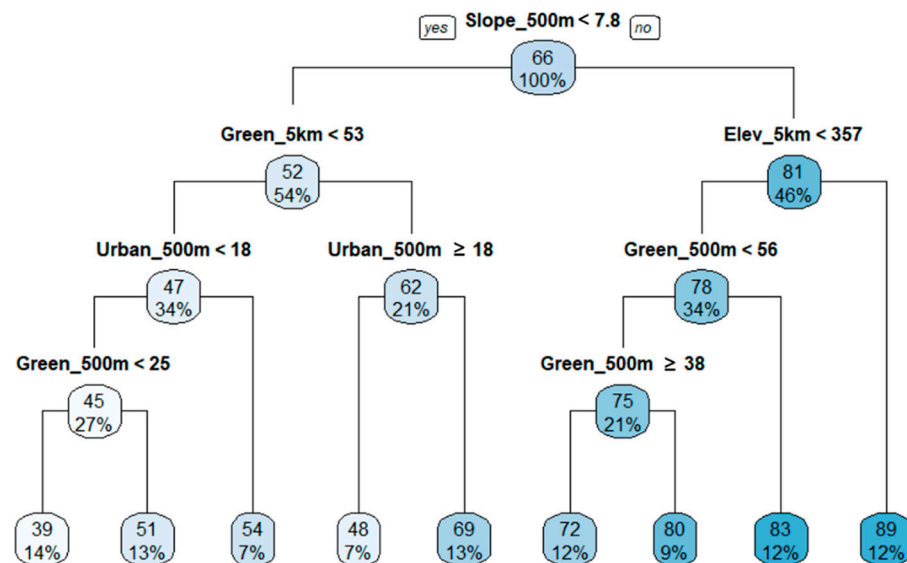

Figure 6. Regression tree for BMI with stream topographical variables and proportions of green and urban areas in buffer areas ($R^2 = 0.78$). The numbers located on top in color boxes are mean values of BMI.

In the FAI regression tree, fish condition scores were divided first by the 500 m slope (Figure 7). Observations with slope less than 9.3% formed a second node at the urbanization rate of the 5 km circle buffer. For FAI values classified in high urbanization areas (left of the tree), the third most important variable was the slope of the 500 m buffer, which was then split into the elevation and urbanization rates of the 500 m buffer at the next node. In groups with a high slope of 9.3% or more (right side of the tree), the second most important variable that determined fish condition was elevation of the 500 m buffer, which indicated the mean elevation of the riparian area from the sampling point. Unlike

Int. J. Environ. Res. Public Health **2021**, 18, 5150

12 of 19

the TDI and BMI regression trees, the FAI regression tree indicated that along with the topographical variables, the fish condition was more likely affected by the urban areas rather than green areas. In the low slope areas, the urban area within the 5 km buffer determined the biological status of fish. Additionally, in the high slope areas, the impact on the surrounding land use and fish condition was determined only by the elevation near the streams. This suggests that the biological conditions of fish in the Nam-Han River Basin were more likely affected by pollutants from urban areas around the stream than by the surrounding forests. The R^2 of the regression tree for FAI was 0.72, and the relative importance of the variables was 21% (500 m slope), 17% (5 km elevation), 17% (5 km slope), 13% (5 km urban rate), 10% (5 km green rate), 4% (500 m urban rate), and 1% (500 m green rate). Along with the stream topographical variables, the proportion of urban areas in a wide area, represented by the 5 km buffer urbanization rate, influenced the improvement of the fish condition.
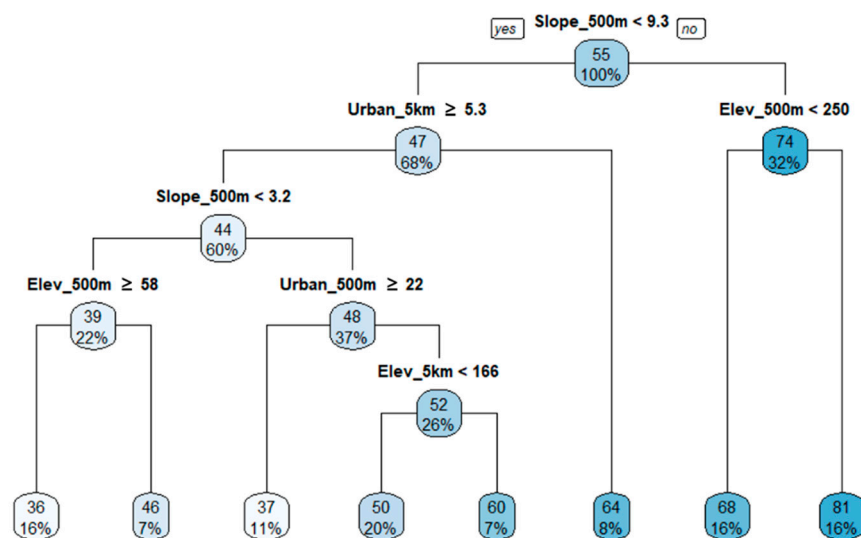

Figure 7. Regression tree for FAI with stream topographical variables and proportions of green and urban areas in buffer areas (R^2 = 0.72). The numbers located on top in color boxes are mean values of FAI.

4. Discussion

4.1. Spatial Autocorrelation of Stream Biological Indicators

Previous studies investigated the biological indicators of stream condition without considering the spatial dependency inherent from the systemic structure of stream ecosystems [99–101]. Such inherent spatial dependency of variables violates the basic assumption of the conventional statistical analyses and may under/over-estimate the LULC in watershed and riparian areas based on the physiochemical properties and ecological condition of streams [48]. In this study, the three biological indicators (i.e., TDI, BMI, and FAI) in the Nam-Han River showed significant spatial autocorrelation.

Spatial dependence in stream biological communities can occur due to multiple reasons. By the definition, streams are a fluid ecosystem in which water continuously flows from headwater in mountainous areas to the estuary in lowlands which is the heart of river continuum concept [102]. Thus, polluted or ecologically impaired headwater can directly affect conditions downstream, despite the concentration of pollutants and the degree of impairments can be mediated or more deteriorated by other tributaries. Therefore, the physiochemical and biological conditions of streams are dependent on the condition of the headwater and the feeding streams resulting in spatial autocorrelations between upstream and downstream at various spatial scales [103–105]. Beside the nature of stream itself, the most common probable cause of spatial dependency in stream properties is topographic characteristics (e.g., elevation, and slope) in watersheds. Topographic features are not

randomly distributed and have structural patterns in space. Thus, the topographic characteristics of a certain location are dependent on surroundings [106,107]. In our study areas, we were able to observe significant spatial autocorrelations of elevation and slope. In literature, topographic characteristics have been shown to determine the hydrologic and hydraulics systems, the amount of pollutant loading, run-off process, and pollutant delivering mechanisms [108,109]. Topographic characteristics even affect the availability, types, and intensity of land uses in watersheds and riparian areas [110]. Thus, it is evident that stream systems in the watershed are strongly tied with topographic characteristics in various ways. We also observed strong correlations of elevation and slope with the amounts of forests and biological indicators including TDI, BMI, and FAI through PCA. Besides, slope and elevation appeared as the first and the second splitter in our estimated regression tree models for TDI (Figure 5), BMI (Figure 6), and FAI (Figure 7), which emphasize the importance in explaining the variances of TDI, BMI, and FAI across the study areas. Since the topographic characteristics are structured systems over space (i.e., spatially autocorrelated), and stream systems are strongly tied with topographic systems, the spatial dependency of physiochemical attributes and biological communities of streams are inevitably inherent from spatial autocorrelation of topographic properties. Numerous studies argued the problems using conventional statistical methods (e.g., Spearman's rank correlation and ordinary least square regression) assuming independency among observations [111,112]. Our findings suggest that spatial autocorrelation should be considered while assessing the relationship between environmental variables and stream quality.

4.2. Land Use and Topographical Effects on Biological Indicators of Stream

Diatoms, macroinvertebrates, and fish, as indicators of stream condition, exhibit different characteristics and complex mechanisms that interact with environmental and topographical variables [113,114]. According to the PCA and regression tree analyses results, topographic variables, including elevation and slope, had the strongest relationship with biological indicators, and the relationships between urban areas, green areas, and biological indicators at the 5 km scale were stronger at those at the 500 m scale. However, the variables that determined the second differentiation differed on the regression trees of the three biological indicators. The regression trees showed that slope at the 500 m scale had the greatest effect on the biological indicators. In the case of the TDI regression tree, the percentage of green areas at the 500 m scale was the most important factor after topographical characteristics. However, the BMI regression tree indicated that the green area at the 5 km scale had strong effects on the biological indicators at low slopes, while the green area at the 500 m scale had strong effects on the biological indicators at steep slopes. Moreover, the urban area at the 5 km scale had greater effects on FAI than at the 500 m scale.

The results of this study can be discussed in two aspects. First, the reason why the initial split was estimated as a topographical variable is that because of most of the tributaries in the Nam-Han River Basin are in forests. Furthermore, topographical factors, including elevation and slope are important variables for determining the biological indexes [17]. The forests are at high elevation and have a steep terrain; therefore, elevations and slopes are critical for determining the spatial characteristics of the stream ecosystem because they represent structural determinants of the amount and speed of water flow [115]. Subsequently, the amount and speed of stream water can significantly change the biological condition in the stream. In addition, elevation and slopes may restrict land use around streams. Urban and agricultural areas that are majorly built downstream at low elevation and slope, greatly influence the stream biological indicators [116]. Therefore, the highest elevation and slope variables affect the biological condition of the tributaries in the Nam-Han River Basin since forests comprise 70% of the total basin area. Second, the biological characteristics of each biological indicator are different. Diatoms are largely affected by the surrounding regions [117]. They are sensitive to light and water temperature, and that is greatly influenced by its conductivity. Urbanized rivers generally have a higher conductivity than

the rivers that flow through forested landscapes. This increases the intensity of light and increases the water temperature, making it a habitat disadvantage to the species [118]. Consistent with the results of this study, topographical variables and green areas in the riparian areas were critical for diatoms. Furthermore, macroinvertebrates and fish respond distinctly to changes in their physical habitats. This is more evident for fish, which directly represent stream conditions since they are the top predator in the food chain of the stream ecosystem [119,120]. According to studies by B. Schmalz et al. [121], macroinvertebrates have the strongest correlation with dominant substrates, and fish have a stronger correlation with flow variability and immediate land use. Additionally, another prior study of this ascribed greater predictive power to land use by comparison the covered a larger area, providing greater contrast among sub-catchments [122]. The results concluded that the importance of local habitat conditions is best revealed by comparing at the within sub-catchment scale. Thus, it is assumed that macroinvertebrates were most affected by proportion of green areas on a wide-area scale, which is the largest land use in this study. And except for the topographical variables, the proportion of urbanized area is acting as the biggest factor in the species status change of fish because the changes in land use are relatively immediate. Since the detailed life cycle, food chain and movement mechanisms of the three biological indicators are different, they have different environmental variables that are most affected by each of them, even in the same site [123,124].

5. Conclusions

This study examined the relationship between land cover (represented as green and urban areas) and biological indicators according to two buffer scales. Subsequently, different variables were the first differentiation criterion for the regression tree of each biological indicator, thus indicating that the effect of green and urban areas differed and depended on the condition of each biota in the stream. Additionally, the effect of the riparian proximity difference was compared by comparing the proportion of areas in the 5 km circle and 500 m linear buffers. Our findings indicate that these effects can vary significantly depending on the topographical variables of the watershed. Furthermore, as the topography of the study area was influenced by the characteristics of the forest terrain in the Nam-Han River Basin, considering the river topography during land use planning is essential in the basin.

This study provides several implications for planning and managing LULC in watershed and riparian areas. First, the study emphasizes that spatial autocorrelation originating from the spatial characteristics of river ecosystems along with their relationships with land use and topographical characteristics should be considered while analyzing the biological condition of the stream. Subsequently, this could provide additional information, such as the criteria of node splitting, and correlation of various effects on river ecosystems. Second, our results provide a basis for considering the impact of locations of green and urban areas on each biological index through a comparison on two different scales. Positive effects of green areas are commonly known, but sufficient literature on the effect of the spatial extent of different types of areas on each biological indicator at the same study site does not exist. Therefore, this study verified and compared the difference in the impact of the location of green and urban areas on each species. However, the present study was limited to green areas (including forests and grasslands) and urban areas (including cities, roads, and buildings) in the Nam-Han River Basin. Thus, future research should include other environmental characteristics of the basin. Third, the regression trees that were constructed for the biological indicators can assist in establishing and prioritizing stream restoration strategies by determining the significance of topography and land cover variables for the condition of each biological indicator in the basin. Different land cover types and topographical characteristics affect different biological indicators, which in turn indicate the biological condition of the basin by determining the unhealthy indicators. Thus, the results of this study can serve as a baseline for supplementing efficient management of the basin. However, arguably the most significant limitation of regression tree analysis is that it is a nonparametric approach which makes it hard to generalize the analysis results based

Int. J. Environ. Res. Public Health **2021**, *18*, 5150

15 of 19

on the observed data. The estimated regression tree model is a black-box model simply describing the relationship between predictors and responses [95] and is built based on the observed datasets without inference about the underlying probability distribution [125]. Another limit of the regression tree analysis is the complexity of the estimated regression tree when it has many nodes, and, in such cases, interpretation of the estimated tree can be challenging [95,126,127]. Therefore, additional studies should be conducted to investigate these limitations further and to develop reliable tools for moderating these issues.

Author Contributions: M.-Y.K. designed the research, performed the data analysis, and wrote the manuscript, interpreted the results of the analysis and edited the manuscript; S.-W.L. performed the additional statistical analysis and finalized the manuscript. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Vannote, R.L.; Minshall, G.W.; Cummins, K.W.; Sedell, J.R.; Cushing, C.E. The River Continuum Concept. *Can. J. Fish. Aquat. Sci.* **1980**, *37*, 130–137. [CrossRef]
2. Yang, B.; Li, S. Design with Nature: Ian McHarg's Ecological Wisdom as Actionable and Practical Knowledge. *Landsc. Urban Plan.* **2016**, *155*, 21–32. [CrossRef]
3. Stone, M.L.; Whiles, M.R.; Webber, J.A.; Williard, K.W.; Reeve, J.D. Macroinvertebrate Communities in Agriculturally Impacted Southern Illinois Streams: Patterns with Riparian Vegetation, Water Quality, and In-stream Habitat Quality. *J. Environ. Qual.* **2005**, *34*, 907–917. [CrossRef]
4. Rankin, E.T. Habitat Indices in Water Resource Quality Assessments. In *Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making*; CRC Press: Boca Raton, FL, USA, 1995; Volume 1995, pp. 181–208.
5. Wang, L.; Lyons, J.; Kanehl, P.; Gatti, R. Influences of Watershed Land use on Habitat Quality and Biotic Integrity in Wisconsin Streams. *Fisheries* **1997**, *22*, 6–12. [CrossRef]
6. Lamberti, G.A. Grazing Experiments in Artificial Streams. *J. N. Am. Benthol. Soc.* **1993**, *12*, 337–342.
7. Leland, H.V.; Porter, S.D. Distribution of Benthic Algae in the Upper Illinois River Basin in Relation to Geology and Land Use. *Freshwat. Biol.* **2000**, *44*, 279–301. [CrossRef]
8. Lee, S.; Hwang, S. Investigation on the Relationship between Land use and Water Quality with Spatial Dimension, Reservoir Type and Shape Complexity. *J. Korean Inst. Landsc. Archit.* **2007**, *34*, 1–9.
9. Lee, S. Testing Non-Stationary Relationship between the Proportion of Green Areas in Watersheds and Water Quality using Geographically Weighted Regression Model. *J. Korean Inst. Landsc. Archit.* **2013**, *41*, 43. [CrossRef]
10. Meador, M.R.; Coles, J.F.; Zappia, H. *American Fisheries Society Symposium*; Fish Assemblage Responses to Urban Intensity Gradients in Contrasting Metropolitan Areas: Birmingham, AL, USA, 2005; pp. 409–423.
11. Mehaffey, M.H.; Nash, M.S.; Wade, T.G.; Ebert, D.W.; Jones, K.B.; Rager, A. Linking Land Cover and Water Quality in New York City's Water Supply Watersheds. *Environ. Monit. Assess.* **2005**, *107*, 29–44. [CrossRef]
12. Tong, S.T.; Chen, W. Modeling the Relationship between Land use and Surface Water Quality. *J. Environ. Manag.* **2002**, *66*, 377–393. [CrossRef]
13. Tu, J.; Xia, Z. Examining Spatially Varying Relationships between Land use and Water Quality using Geographically Weighted Regression I: Model Design and Evaluation. *Sci. Total Environ.* **2008**, *407*, 358–378. [CrossRef]
14. Weaver, L.A.; Garman, G.C. Urbanization of a Watershed and Historical Changes in a Stream Fish Assemblage. *Trans. Am. Fish. Soc.* **1994**, *123*, 162–172. [CrossRef]
15. Park, S.; Lee, H.; Lee, S.; Hwang, S.; Byeon, M.; Joo, G.; Jeong, K.; Kong, D.; Kim, M. Relationships between Land use and Multi-Dimensional Characteristics of Streams and Rivers at Two Different Scales. *Ann. Limnol. Int. J. Limnol.* **2011**, *47*, 107–116. [CrossRef]

16. Chelsea Nagy, R.; Graeme Lockaby, B.; Kalin, L.; Anderson, C. Effects of Urbanization on Stream Hydrology and Water Quality: The Florida Gulf Coast. *Hydrol. Process.* **2012**, *26*, 2019–2030. [CrossRef]

17. An, K.; Lee, S.; Hwang, S.; Park, S.; Hwang, S. Exploring the Non-Stationary Effects of Forests and Developed Land within Watersheds on Biological Indicators of Streams using Geographically-Weighted Regression. *Water* **2016**, *8*, 120. [CrossRef]

18. Hwang, S.; Hwang, S.; Park, S.; Lee, S. Examining the Relationships between Watershed Urban Land use and Stream Water Quality using Linear and Generalized Additive Models. *Water* **2016**, *8*, 155. [CrossRef]

19. Casotti, C.G.; Kiffer Jr, W.P.; Costa, L.C.; Rangel, J.V.; Casagrande, L.C.; Moretti, M.S. Assessing the Importance of Riparian Zones Conservation for Leaf Decomposition in Streams. *Nat. Conserv.* **2015**, *13*, 178–182. [CrossRef]

20. Chellaiah, D.; Yule, C.M. Effect of Riparian Management on Stream Morphometry and Water Quality in Oil Palm Plantations in Borneo. *Limnologica* **2018**, *69*, 72–80. [CrossRef]

21. Pusey, B.J.; Arthington, A.H. Importance of the Riparian Zone to the Conservation and Management of Freshwater Fish: A Review. *Mar. Freshw. Res.* **2003**, *54*, 1–16. [CrossRef]

22. Popov, V.H.; Cornish, P.S.; Sun, H. Vegetated Biofilters: The Relative Importance of Infiltration and Adsorption in Reducing Loads of Water-Soluble Herbicides in Agricultural Runoff. *Agric. Ecosyst. Environ.* **2006**, *114*, 351–359. [CrossRef]

23. Meek, C.S.; Richardson, D.M.; Mucina, L. A River Runs through it: Land-use and the Composition of Vegetation Along a Riparian Corridor in the Cape Floristic Region, South Africa. *Biol. Conserv.* **2010**, *143*, 156–164. [CrossRef]

24. Scott, M.L.; Nagler, P.L.; Glenn, E.P.; Valdes-Casillas, C.; Erker, J.A.; Reynolds, E.W.; Shafroth, P.B.; Gomez-Limon, E.; Jones, C.L. Assessing the Extent and Diversity of Riparian Ecosystems in Sonora, Mexico. *Biodivers. Conserv.* **2009**, *18*, 247–269. [CrossRef]

25. Broadmeadow, S.; Nisbet, T.R. The Effects of Riparian Forest Management on the Freshwater Environment: A Literature Review of Best Management Practice. *Hydrol. Earth Syst. Sci.* **2004**, *8*, 286–305. [CrossRef]

26. Li, M.; Huang, C.; Zhu, Z.; Shi, H.; Lu, H.; Peng, S. Assessing Rates of Forest Change and Fragmentation in Alabama, USA, using the Vegetation Change Tracker Model. *For. Ecol. Manag.* **2009**, *257*, 1480–1488. [CrossRef]

27. Taniwaki, R.H.; Cassiano, C.C.; Filoso, S.; de Barros, F.; Silvio, F.; de Camargo, P.B.; Martinelli, L.A. Impacts of Converting Low-Intensity Pastureland to High-Intensity Bioenergy Cropland on the Water Quality of Tropical Streams in Brazil. *Sci. Total Environ.* **2017**, *584*, 339–347. [CrossRef]

28. Popescu, C.; Oprina-Pavelescu, M.; Dinu, V.; Cazacu, C.; Burdon, F.J.; Forio, M.A.E.; Kupilas, B.; Friberg, N.; Goethals, P.; McKie, B.G. Riparian Vegetation Structure Influences Terrestrial Invertebrate Communities in an Agricultural Landscape. *Water* **2021**, *13*, 188. [CrossRef]

29. Forio, M.A.E.; De Troyer, N.; Lock, K.; Witing, F.; Baert, L.; Saeyer, N.D.; Rîșnoveanu, G.; Popescu, C.; Burdon, F.J.; Kupilas, B. Small Patches of Riparian Woody Vegetation Enhance Biodiversity of Invertebrates. *Water* **2020**, *12*, 3070. [CrossRef]

30. Mutinova, P.T.; Kahlert, M.; Kupilas, B.; McKie, B.G.; Friberg, N.; Burdon, F.J. Benthic Diatom Communities in Urban Streams and the Role of Riparian Buffers. *Water* **2020**, *12*, 2799. [CrossRef]

31. Burdon, F.J.; Ramberg, E.; Sargac, J.; Forio, M.A.E.; De Saeyer, N.; Mutinova, P.T.; Moe, T.F.; Pavelescu, M.O.; Dinu, V.; Cazacu, C. Assessing the Benefits of Forested Riparian Zones: A Qualitative Index of Riparian Integrity is Positively Associated with Ecological Status in European Streams. *Water* **2020**, *12*, 1178. [CrossRef]

32. Sargac, J.; Johnson, R.K.; Burdon, F.J.; Truchy, A.; Rîșnoveanu, G.; Goethals, P.; McKie, B.G. Forested Riparian Buffers Change the Taxonomic and Functional Composition of Stream Invertebrate Communities in Agricultural Catchments. *Water* **2021**, *13*, 1028. [CrossRef]

33. Kupilas, B.; Burdon, F.J.; Thaulow, J.; Håll, J.; Mutinova, P.T.; Forio, M.A.E.; Witing, F.; Rîșnoveanu, G.; Goethals, P.; McKie, B.G. Forested Riparian Zones Provide Important Habitat for Fish in Urban Streams. *Water* **2021**, *13*, 877. [CrossRef]

34. Lee, S.; Lee, J.; Park, S. A Structural Relationship of Topography, Developed Areas, and Riparian Vegetation on the Concentration of Total Nitrogen in Streams. *J. Korean Inst. Landsc. Archit.* **2020**, *48*, 25–34. [CrossRef]

35. King, R.S.; Baker, M.E.; Whigham, D.F.; Weller, D.E.; Jordan, T.E.; Kazyak, P.F.; Hurd, M.K. Spatial Considerations for Linking Watershed Land Cover to Ecological Indicators in Streams. *Ecol. Appl.* **2005**, *15*, 137–153. [CrossRef]

36. Grau, H.R.; Aide, T.M.; Zimmerman, J.K.; Thomlinson, J.R.; Helmer, E.; Zou, X. The Ecological Consequences of Socioeconomic and Land-use Changes in Postagriculture Puerto Rico. *Bioscience* **2003**, *53*, 1159–1168. [CrossRef]

37. Yirigui, Y.; Lee, S.; Nejadhashemi, A.P.; Herman, M.R.; Lee, J. Relationships between Riparian Forest Fragmentation and Biological Indicators of Streams. *Sustainability* **2019**, *11*, 2870. [CrossRef]

38. Damanik-Ambarita, M.N.; Everaert, G.; Goethals, P.L. Ecological Models to Infer the Quantitative Relationship between Land use and the Aquatic Macroinvertebrate Community. *Water* **2018**, *10*, 184. [CrossRef]

39. Ekka, A.; Pande, S.; Jiang, Y.; der Zaag, P.v. Anthropogenic Modifications and River Ecosystem Services: A Landscape Perspective. *Water* **2020**, *12*, 2706. [CrossRef]

40. Abouali, M.; Nejadhashemi, A.P.; Daneshvar, F.; Woznicki, S.A. Two-Phase Approach to Improve Stream Health Modeling. *Ecol. Inform.* **2016**, *34*, 13–21. [CrossRef]

41. Hrodey, P.J.; Sutton, T.M.; Frimpong, E.A.; Simon, T.P. Land-use Impacts on Watershed Health and Integrity in Indiana Warmwater Streams. *Am. Midl. Nat.* **2009**, *161*, 76–95. [CrossRef]

42. Dahm, V.; Hering, D. A Modeling Approach for Identifying Recolonisation Source Sites in River Restoration Planning. *Landscape Ecol.* **2016**, *31*, 2323–2342. [CrossRef]

43. de Paula, F.R.; Gerhard, P.; de Barros, S.F.; Wenger, S.J. Multi-Scale Assessment of Forest Cover in an Agricultural Landscape of Southeastern Brazil: Implications for Management and Conservation of Stream Habitat and Water Quality. *Ecol. Ind.* **2018**, *85*, 1181–1191. [CrossRef]
44. Bae, M.; Kwon, Y.; Hwang, S.; Chon, T.; Yang, H.; Kwak, I.; Park, J.; Ham, S.; Park, Y. Relationships between Three Major Stream Assemblages and their Environmental Factors in Multiple Spatial Scales. *Ann. Limnol. Int. J. Limnol.* **2011**, *47*, 91–105. [CrossRef]
45. Jontos, R. Vegetative Buffers for Water Quality Protection: An Introduction and Guidance Document. *Conn. Assoc. Wetl. Sci. White Pap. Veg. Buffers* **2004**, *1*, 22.
46. Lee, S.; Hwang, S.; Lee, S.; Hwang, H.; Sung, H. Landscape Ecological Approach to the Relationships of Land use Patterns in Watersheds to Water Quality Characteristics. *Landsc. Urban Plann.* **2009**, *92*, 80–89. [CrossRef]
47. Kim, J.; An, K.; Hwang, S.; Hwang, G.; Kim, D.; Kim, C.; Lee, S. Mediating Effect of Stream Geometry on the Relationship between Urban Land use and Biological Index. *Paddy Water Environ.* **2014**, *12*, 157–168. [CrossRef]
48. Cai, X.; Wang, D. Spatial Autocorrelation of Topographic Index in Catchments. *J. Hydrol.* **2006**, *328*, 581–591. [CrossRef]
49. Dormann, C.F. Effects of Incorporating Spatial Autocorrelation into the Analysis of Species Distribution Data. *Glob. Ecol. Biogeogr.* **2007**, *16*, 129–138. [CrossRef]
50. Teittinen, A.; Kallajoki, L.; Meier, S.; Stigzelius, T.; Soininen, J. The Roles of Elevation and Local Environmental Factors as Drivers of Diatom Diversity in Subarctic Streams. *Freshwat. Biol.* **2016**, *61*, 1509–1521. [CrossRef]
51. Soininen, J.; Paavola, R.; Muotka, T. Benthic Diatom Communities in Boreal Streams: Community Structure in Relation to Environmental and Spatial Gradients. *Ecography* **2004**, *27*, 330–342. [CrossRef]
52. Legendre, P. Spatial Autocorrelation: Trouble or New Paradigm? *Ecology* **1993**, *74*, 1659–1673. [CrossRef]
53. Schabenberger, O.; Gotway, C.A. *Statistical Methods for Spatial Data Analysis*; Chapman Hall/CRC: London, UK, 2005.
54. Lee, S.; Won, M.; Lee, H. Testing Spatial Autocorrelation of Burn Severity. *J. Korean Soc. For. Sci.* **2012**, *101*, 203–212.
55. Allen, T.F.; Starr, T.B. *Hierarchy: Perspectives for Ecological Complexity*; University of Chicago Press: Chicago, IL, USA, 2017.
56. Dorner, B.; Lertzman, K.; Fall, J. Landscape Pattern in Topographically Complex Landscapes: Issues and Techniques for Analysis. *Landsc. Ecol.* **2002**, *17*, 729–743. [CrossRef]
57. Detenbeck, N.E.; Cincotta, D.; Denver, J.M.; Greenlee, S.K.; Olsen, A.R.; Pitchford, A.M. Watershed-Based Survey Designs. *Environ. Monit. Assess.* **2005**, *103*, 59–81. [CrossRef]
58. Peterson, E.E.; Merton, A.A.; Theobald, D.M.; Urquhart, N.S. Patterns of Spatial Autocorrelation in Stream Water Chemistry. *Environ. Monit. Assess.* **2006**, *121*, 571–596. [CrossRef] [PubMed]
59. Yoo, C.; Kim, I.; Ryoo, S. Evaluation of Raingauge Density and Spatial Distribution: A Case Study for Nam Han River Basin. *J. Korea Water Resour. Assoc.* **2003**, *36*, 173–181. [CrossRef]
60. Lee, J.; Choi, J.; An, K. Influence of Landuse Pattern and Seasonal Precipitation on the Long-Term Physico-Chemical Water Quality in Namhan River Watershed. *J. Environ. Sci. Int.* **2012**, *21*, 1115–1129. [CrossRef]
61. Lee, J.; Lee, S.; An, K.; Hwang, S.; Kim, N. An Estimated Structural Equation Model to Assess the Effects of Land use on Water Quality and Benthic Macroinvertebrates in Streams of the Nam-Han River System, South Korea. *Int. J. Environ. Res. Public Health* **2020**, *17*, 2116. [CrossRef]
62. Lee, S.; Hwang, S.; Lee, J.; Jung, D.; Park, Y.; Kim, J. Overview and Application of the National Aquatic Ecological Monitoring Program (NAEMP) in Korea. *Ann. Limnol. Int. J. Limnol.* **2011**, *47*, 3–14. [CrossRef]
63. Won, D.; Byun, M.; Park, J.; Lee, S.; Hwang, S. Evaluation the current status of aquatic ecosystem health in five major rivers in Korea. *Korean Soc. Civ. Eng.* **2010**, *3*, 149–160.
64. National Institute of Environmental Research, (NIER). *The Survey and Evaluation of Aquatic Ecosystem Health in Korea*; NIER: Seong-Kyu Yoon, Korea, 2008.
65. Homes, N. Integrated Watershed Management: Ecohydrology and Phytotechnology. *Wetl. Ecol. Manag.* **2005**, *13*, 209–210.
66. 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]
67. Lee, Y.; Lee, S. Stream Classification Based on the Ecological Characteristics for Effective Stream Management-in the Case of Nakdong River. *J. Korean Soc. Environ. Restor. Technol.* **2012**, *15*, 103–114. [CrossRef]
68. Allan, J.D.; Castillo, M.M. *Stream Ecology: Structure and Function of Running Waters*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2007.
69. Jun, Y.; Won, D.; Lee, S.; Kong, D.; Hwang, S. A Multimetric Benthic Macroinvertebrate Index for the Assessment of Stream Biotic Integrity in Korea. *Int. J. Environ. Res. Public Health* **2012**, *9*, 3599–3628. [CrossRef] [PubMed]
70. Karr, J.R. *Assessing Biological Integrity in Running Waters: A Method and its Rationale*; Special publication/Illinois Natural History Survey: Champaign, IL, USA, 1986.
71. Ahn, C.H.; Joo, J.C.; Kwon, J.H.; Song, H.M. Evaluation on Functional Assessment for Fish Habitat of Underground Type Eco-Artificial Fish Reef using the Index of Biological Integrity (IBI) and Qualitative Habitat Evaluation Index (QHEI). *J. Korean Soc. Civil Eng.* **2011**, *31*, 565–575.
72. McGarigal, K. FRAGSTATS: Spatial Pattern Analysis Program for Categorical Maps. Computer Software Program Produced by the Authors at the University of Massachusetts. 2002. Available online: http://www.umass.edu/landeco/research/fragstats/fragstats.html (accessed on 12 March 2021).

73. Frate, L.; Carranza, M.L. Quantifying Landscape-Scale Patterns of Temperate Forests Over Time by Means of Neutral Simulation Models. *ISPRS Int. J. Geo Inf.* **2013**, *2*, 94–109. [CrossRef]

74. Tolessa, T.; Senbeta, F.; Kidane, M. Landscape Composition and Configuration in the Central Highlands of Ethiopia. *Ecol. Evol.* **2016**, *6*, 7409–7421. [CrossRef] [PubMed]

75. Yirigui, Y.; Lee, S.; Nejadhashemi, A.P. Multi-Scale Assessment of Relationships between Fragmentation of Riparian Forests and Biological Conditions in Streams. *Sustainability* **2019**, *11*, 5060. [CrossRef]

76. Poiani, K.A.; Richter, B.D.; Anderson, M.G.; Richter, H.E. Biodiversity Conservation at Multiple Scales: Functional Sites, Landscapes, and Networks. *Bioscience* **2000**, *50*, 133–146. [CrossRef]

77. Hawes, E.; Smith, M. Riparian Buffer Zones: Functions and Recommended Widths. *Eightmile River Wild Scen. Study Comm.* **2005**, *15*, 2005.

78. Fortin, M.; Drapeau, P.; Legendre, P. Spatial autocorrelation and sampling design in plant ecology. In *Progress in Theoretical Vegetation Science*; Springer: Berlin/Heidelberg, Germany, 1990; pp. 209–222.

79. Griffith, D.A. Spatial autocorrelation in spatial interaction. In *Complexity and Spatial Networks*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 221–237.

80. Anselin, L.; Syabri, I.; Kho, Y. GeoDa: An introduction to spatial data analysis. In *Handbook of Applied Spatial Analysis*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 73–89.

81. Anselin, L. Local Indicators of Spatial association—LISA. *Geogr. Anal.* **1995**, *27*, 93–115. [CrossRef]

82. Miralha, L.; Kim, D. Accounting for and Predicting the Influence of Spatial Autocorrelation in Water Quality Modeling. *ISPRS Int. J. Geo Inf.* **2018**, *7*, 64. [CrossRef]

83. 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]

84. Kurita, T. Principal Component Analysis (PCA). In *Computer Vision: A Reference Guide*; Springer: Tokyo, Japan, 2019; pp. 1–4.

85. Abdi, H.; Williams, L.J. Wiley interdisciplinary reviews: Computational statistics. *Princ. Compon. Anal.* **2010**, *2*, 433–459.

86. Cablk, M. *Assessment of Spatial Autocorrelation in Empirical Models in Ecology. Predicting Species Occurrences: Issues of Accuracy and Scale*; Island Press: Washington, DC, USA, 2001; pp. 429–440.

87. Collins, B.M.; Kelly, M.; Van Wagtendonk, J.W.; Stephens, S.L. Spatial Patterns of Large Natural Fires in Sierra Nevada Wilderness Areas. *Landsc. Ecol.* **2007**, *22*, 545–557. [CrossRef]

88. Ripley, B.D. *Pattern Recognition and Neural Networks*; Cambridge University Press: Cambridge, UK, 2007.

89. Goethals, P.L.; Dedecker, A.P.; Gabriels, W.; Lek, S.; De Pauw, N. Applications of Artificial Neural Networks Predicting Macroinvertebrates in Freshwaters. *Aquat. Ecol.* **2007**, *41*, 491–508. [CrossRef]

90. Boets, P.; Lock, K.; Messiaen, M.; Goethals, P.L. Combining Data-Driven Methods and Lab Studies to Analyse the Ecology of Dikerogammarus Villosus. *Ecol. Inform.* **2010**, *5*, 133–139. [CrossRef]

91. Dominguez-Granda, L.; Lock, K.; Goethals, P.L. Using Multi-Target Clustering Trees as a Tool to Predict Biological Water Quality Indices Based on Benthic Macroinvertebrates and Environmental Parameters in the Chaguana Watershed (Ecuador). *Ecol. Inform.* **2011**, *6*, 303–308. [CrossRef]

92. Everaert, G.; Boets, P.; Lock, K.; Džeroski, S.; Goethals, P.L. Using Classification Trees to Analyze the Impact of Exotic Species on the Ecological Assessment of Polder Lakes in Flanders, Belgium. *Ecol. Model.* **2011**, *222*, 2202–2212. [CrossRef]

93. Everaert, G.; Pauwels, I.S.; Boets, P.; Verduin, E.; de la Haye Michelle, A.A.; Blom, C.; Goethals, P.L. Model-Based Evaluation of Ecological Bank Design and Management in the Scope of the European Water Framework Directive. *Ecol. Eng.* **2013**, *53*, 144–152. [CrossRef]

94. De'Ath, G. Multivariate Regression Trees: A New Technique for Modeling Species–environment Relationships. *Ecology* **2002**, *83*, 1105–1117.

95. Everaert, G.; Pauwels, I.; Bennetsen, E.; Goethals, P.L. Development and Selection of Decision Trees for Water Management: Impact of Data Preprocessing, Algorithms and Settings. *Communications* **2016**, *29*, 711–723. [CrossRef]

96. Breiman, L.; Friedman, J.; Stone, C.J.; Olshen, R.A. *Classification and Regression Trees*; CRC Press: Boca Raton, FL, USA, 1984.

97. Therneau, T.M.; Atkinson, E.J. An Introduction to Recursive Partitioning Using the RPART Routines; Technical report Mayo Foundation. 1997. Available online: https://r.789695.n4.nabble.com/attachment/3209029/0/zed.pdf (accessed on 14 March 2021).

98. Wood, S.N. *Generalized Additive Models: An Introduction with R*; CRC Press: Boca Raton, FL, USA, 2017.

99. Ayllón, D.; Almodóvar, A.; Nicola, G.G.; Parra, I.; Elvira, B. A New Biological Indicator to Assess the Ecological Status of Mediterranean Trout Type Streams. *Ecol. Ind.* **2012**, *20*, 295–303. [CrossRef]

100. Hill, B.H.; Stevenson, R.J.; Pan, Y.; Herlihy, A.T.; Kaufmann, P.R.; Johnson, C.B. Comparison of Correlations between Environmental Characteristics and Stream Diatom Assemblages Characterized at Genus and Species Levels. *J. N. Am. Benthol. Soc.* **2001**, *20*, 299–310. [CrossRef]

101. Wang, B.; Liu, D.; Liu, S.; Zhang, Y.; Lu, D.; Wang, L. Impacts of Urbanization on Stream Habitats and Macroinvertebrate Communities in the Tributaries of Qiangtang River, China. *Hydrobiologia* **2012**, *680*, 39–51. [CrossRef]

102. Ward, J.V. The Four-Dimensional Nature of Lotic Ecosystems. *J. N. Am. Benthol. Soc.* **1989**, *8*, 2–8. [CrossRef]

103. Gomi, T.; Sidle, R.C.; Richardson, J.S. Understanding Processes and Downstream Linkages of Headwater Systems: Headwaters Differ from Downstream Reaches by their Close Coupling to Hillslope Processes, More Temporal and Spatial Variation, and their Need for Different Means of Protection from Land Use. *Bioscience* **2002**, *52*, 905–916.

104. Dodds, W.K.; Oakes, R.M. Headwater Influences on Downstream Water Quality. *Environ. Manag.* **2008**, *41*, 367–377. [CrossRef] [PubMed]

Int. J. Environ. Res. Public Health **2021**, *18*, 5150

19 of 19

105. Le Pichon, C.; Tales, É.; Belliard, J.; Torgersen, C.E. Spatially Intensive Sampling by Electrofishing for Assessing Longitudinal Discontinuities in Fish Distribution in a Headwater Stream. *Fish. Res.* **2017**, *185*, 90–101. [CrossRef]
106. Beven, K.J.; Kirkby, M.J. A Physically Based, Variable Contributing Area Model of Basin Hydrology/Un Modèle À Base Physique De Zone D'Appel Variable De L'Hydrologie Du Bassin Versant. *Hydrol. Sci. J.* **1979**, *24*, 43–69. [CrossRef]
107. O'loughlin, E.M. Prediction of Surface Saturation Zones in Natural Catchments by Topographic Analysis. *Water Resour. Res.* **1986**, *22*, 794–804. [CrossRef]
108. Goh, K.; Härdter, R.; Fairhurst, T. *Fertilizing for Maximum Return. Oil Palm: Management for Large and Sustainable Yields*; International Plant Nutrition Institute: São Paulo, Brazil, 2003; pp. 279–306.
109. Gharabaghi, B.; Rudra, R.; Whiteley, H.R.; Dickinson, W.T. Development of a Management Tool for Vegetative Filter Strips. *J. Water Manag. Model.* **2002**. [CrossRef]
110. Yu, S.; Xu, Z.; Wu, W.; Zuo, D. Effect of Land use Types on Stream Water Quality Under Seasonal Variation and Topographic Characteristics in the Wei River Basin, China. *Ecol. Ind.* **2016**, *60*, 202–212. [CrossRef]
111. Karp, D.S.; Chaplin-Kramer, R.; Meehan, T.D.; Martin, E.A.; DeClerck, F.; Grab, H.; Gratton, C.; Hunt, L.; Larsen, A.E.; Martínez-Salinas, A. Crop Pests and Predators Exhibit Inconsistent Responses to Surrounding Landscape Composition. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 7863–7870. [CrossRef]
112. Henebry, G.M. Spatial Model Error Analysis using Autocorrelation Indices. *Ecol. Model.* **1995**, *82*, 75–91. [CrossRef]
113. Villeneuve, B.; Souchon, Y.; Usseglio-Polatera, P.; Ferréol, M.; Valette, L. Can we Predict Biological Condition of Stream Ecosystems? A Multi-Stressors Approach Linking Three Biological Indices to Physico-Chemistry, Hydromorphology and Land Use. *Ecol. Ind.* **2015**, *48*, 88–98. [CrossRef]
114. Zuellig, R.E.; Carlisle, D.M.; Meador, M.R.; Potapova, M. Variance Partitioning of Stream Diatom, Fish, and Invertebrate Indicators of Biological Condition. *Freshw. Sci.* **2012**, *31*, 182–190. [CrossRef]
115. Lee, S.; Seo, J.; Lee, Y.; Kim, S.; Chun, K. Analysis of Research Trends on Mountain Streams in the Republic of Korea: Comparison to International Research Trends. *Korean J. Environ. Ecol.* **2019**, *33*, 216–227. [CrossRef]
116. Aguiar, F.C.; Martins, M.J.; Silva, P.C.; Fernandes, M.R. Riverscapes Downstream of Hydropower Dams: Effects of Altered Flows and Historical Land-use Change. *Landsc. Urban Plan.* **2016**, *153*, 83–98. [CrossRef]
117. Min, H.; Kim, N.; Kim, M.; Lee, S.; Hwang, K.; Hwang, S. Relation of Stream Shape Complexity to Land use, Water Quality and Benthic Diatoms in the Seom River Watershed. *Korean J. Ecol. Environ.* **2012**, *45*, 110–122.
118. Teittinen, A.; Taka, M.; Ruth, O.; Soininen, J. Variation in Stream Diatom Communities in Relation to Water Quality and Catchment Variables in a Boreal, Urbanized Region. *Sci. Total Environ.* **2015**, *530*, 279–289. [CrossRef]
119. Beyene, A.; Addis, T.; Kifle, D.; Legesse, W.; Kloos, H.; Triest, L. Comparative Study of Diatoms and Macroinvertebrates as Indicators of Severe Water Pollution: Case Study of the Kebena and Akaki Rivers in Addis Ababa, Ethiopia. *Ecol. Ind.* **2009**, *9*, 381–392. [CrossRef]
120. Karr, J.R. Assessment of Biotic Integrity using Fish Communities. *Fisheries* **1981**, *6*, 21–27. [CrossRef]
121. Schmalz, B.; Kuemmerlen, M.; Kiesel, J.; Cai, Q.; Jähnig, S.C.; Fohrer, N. Impacts of Land use Changes on Hydrological Components and Macroinvertebrate Distributions in the Poyang Lake Area. *Ecohydrology* **2015**, *8*, 1119–1136. [CrossRef]
122. Nerbonne, B.A.; Vondracek, B. Effects of Local Land use on Physical Habitat, Benthic Macroinvertebrates, and Fish in the Whitewater River, Minnesota, USA. *Environ. Manag.* **2001**, *28*, 87–99. [CrossRef] [PubMed]
123. Miserendino, M.L.; Casaux, R.; Archangelsky, M.; Di Prinzio, C.Y.; Brand, C.; Kutschker, A.M. Assessing Land-use Effects on Water Quality, in-Stream Habitat, Riparian Ecosystems and Biodiversity in Patagonian Northwest Streams. *Sci. Total Environ.* **2011**, *409*, 612–624. [CrossRef] [PubMed]
124. Clapcott, J.E.; Collier, K.J.; Death, R.G.; Goodwin, E.O.; Harding, J.S.; Kelly, D.; Leathwick, J.R.; Young, R.G. Quantifying Relationships between Land-use Gradients and Structural and Functional Indicators of Stream Ecological Integrity. *Freshwat. Biol.* **2012**, *57*, 74–90. [CrossRef]
125. Prasad, A.M.; Iverson, L.R.; Liaw, A. Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. *Ecosystems* **2006**, *9*, 181–199. [CrossRef]
126. Poor, C.J.; Ullman, J.L. Using Regression Tree Analysis to Improve Predictions of Low-Flow Nitrate and Chloride in Willamette River Basin Watersheds. *Environ. Manag.* **2010**, *46*, 771–780. [CrossRef]
127. Breslow, L.A.; Aha, D.W. Simplifying Decision Trees: A Survey. *Knowl. Eng. Rev.* **1997**, *12*, 1–40. [CrossRef]