

## Article

# Occurrence Prediction of Riffle Beetles (Coleoptera: Elmidae) in a Tropical Andean Basin of Ecuador Using Species Distribution Models

Gonzalo Sotomayor <sup>1,2,\*</sup> , Jorge Romero <sup>3</sup> , Daniela Ballari <sup>3</sup> , Raúl F. Vázquez <sup>2,4</sup> , Iván Ramírez-Morales <sup>5</sup> , Henrietta Hampel <sup>4</sup> , Xavier Galarza <sup>3</sup> , Bolívar Montesinos <sup>6</sup> , Marie Anne Eurie Forio <sup>1</sup> , and Peter L. M. Goethals <sup>1</sup>

- <sup>1</sup> Department of Animal Sciences and Aquatic Ecology, Faculty of Bioscience Engineering, Ghent University, Coupure Links 653, 9000 Ghent, Belgium; marie.forio@ugent.be (M.A.E.F.); peter.goethals@ugent.be (P.L.M.G.)
  - <sup>2</sup> Departamento de Ingeniería Civil, Facultad de Ingeniería, Universidad de Cuenca, Av. 12 de abril S/N, Cuenca, Azuay 010203, Ecuador; raulfvazquez@yahoo.co.uk
  - <sup>3</sup> Instituto de Estudios del Régimen Seccional del Ecuador (IERSE), Facultad de Ciencia y Tecnología, Universidad del Azuay, Cuenca 010204, Ecuador; jaromero@uazuay.edu.ec (J.R.); dballari@uazuay.edu.ec (D.B.); xavier.galarzag@hotmail.com (X.G.)
  - <sup>4</sup> Laboratorio de Ecología Acuática (LEA), Facultad de Ciencias Químicas, Universidad de Cuenca, Av. 12 de abril S/N, Cuenca 010203, Ecuador; hennihampel@gmail.com
  - <sup>5</sup> DINTA Research Group, Universidad Técnica de Machala, Machala 070213, Ecuador; iramirez@utmachala.edu.ec
  - <sup>6</sup> Ministerio del Ambiente, Agua y Transición Ecológica, Dirección Zonal 6, Cuenca 010104, Ecuador; bolivar.montesinos@ambiente.gob.ec
- \* Correspondence: gonzalo.sotomayor@ugent.be or gonpa83@yahoo.com



**Citation:** Sotomayor, G.; Romero, J.; Ballari, D.; Vázquez, R.F.; Ramírez-Morales, I.; Hampel, H.; Galarza, X.; Montesinos, B.; Forio, M.A.E.; Goethals, P.L.M. Occurrence Prediction of Riffle Beetles (Coleoptera: Elmidae) in a Tropical Andean Basin of Ecuador Using Species Distribution Models. *Biology* **2023**, *12*, 473. <https://doi.org/10.3390/biology12030473>

Academic Editors: Lijuan Ren, Xiaomin Xia and Jin Zeng

Received: 7 February 2023  
Revised: 16 March 2023  
Accepted: 16 March 2023  
Published: 20 March 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Simple Summary:** A machine learning algorithm, Random Forest, was used to establish species distribution models for five riffle beetle genera (Elmidae) in the Paute river basin (southern Ecuador), considering meteorology, land use, hydrology, and topography as environmental/explanatory variables. Alterations to riparian vegetation, canopy presence/absence, precipitation, elevation, and slope accounted for most of the Elmidae spatial variability. Clean and healthy streams were predicted to be the most likely places for Elmidae genera to occur. Additionally, specific ecological niches were predicted for each Elmidae genus. These findings can contribute significantly to conservation and restoration efforts in the study basin and could have implications for similar eco-hydrological systems.

**Abstract:** Genera and species of Elmidae (riffle beetles) are sensitive to water pollution; however, in tropical freshwater ecosystems, their requirements regarding environmental factors need to be investigated. Species distribution models (SDMs) were established for five elmidae genera in the Paute river basin (southern Ecuador) using the Random Forest (RF) algorithm considering environmental variables, i.e., meteorology, land use, hydrology, and topography. Each RF-based model was trained and optimised using cross-validation. Environmental variables that explained most of the Elmidae spatial variability were land use (i.e., riparian vegetation alteration and presence/absence of canopy), precipitation, and topography, mainly elevation and slope. The highest probability of occurrence for elmidae genera was predicted in streams located within well-preserved zones. Moreover, specific ecological niches were spatially predicted for each genus. *Macrelmis* was predicted in the lower and forested areas, with high precipitation levels, towards the Amazon basin. *Austrelmis* was predicted to be in the upper parts of the basin, i.e., páramo ecosystems, with an excellent level of conservation of their riparian ecosystems. *Austrolimnius* and *Heterelmis* were also predicted in the upper parts of the basin but in more widespread elevation ranges, in the *Heterelmis* case, and even in some areas with a medium level of anthropisation. *Neelmis* was predicted to be in the mid-region of the study basin in high altitudinal streams with a high degree of meandering. The main findings of this research are likely to contribute significantly to local conservation and restoration efforts being implemented in the study basin and could be extrapolated to similar eco-hydrological systems.

**Keywords:** Elmid genera; presence–absence records; species distribution models; random forest; streams

## 1. Introduction

Freshwater ecosystems have been severely altered by human activities and are significantly vulnerable to climate change [1,2]. Hence, there is an urgent need to understand the spatial and temporal patterns of aquatic organisms to maintain and restore aquatic biodiversity [3]. In this context, species distribution models (SDMs) relate taxa occurrence with the local environmental conditions and provide a spatial prediction of taxa habitat suitability on the entire study area and, optionally, across time [4]. Thus, SDMs can be particularly useful for biodiversity conservation policies, since they can identify suitable areas for preserving threatened taxa or priority areas for future sampling efforts [5]. SDMs have been used for identifying trends of spatial distribution and habitat limitations for some specific aquatic organisms, e.g., benthic macroinvertebrates [6–8], fishes [9,10], and algae [11]. Furthermore, SDMs could help manage water resources in a country such as Ecuador, which is facing a severe decline in the ecological integrity of its rivers and lakes [12,13]. However, to our knowledge, just one work has been carried out in the country using the SDMs framework, particularly targeting benthic macroinvertebrates taxa [14].

In this context, riffle beetles (i.e., Coleoptera: Elmidae) are cosmopolitan freshwater coleopterans that inhabit clean and well-oxygenated running waters in their larval and adult stages. Regarding ecosystem functionality, the Elmidae members are collector-gatherers and scrapers that feed mainly on algae and detritus. Further, elmid larvae and adults are an important part of the diet of fish. Elmidae taxa are sensitive to changes in the habitat structure and physicochemical conditions of aquatic ecosystems [15–17]. Therefore, elmids are considered excellent indicators of water quality integrity and perhaps also of climate change [18].

In the Paute river basin (PRB), which is one of the most important hydrological systems of Ecuador owing to its significant hydroelectric potential [19,20], elmids have been identified as the key taxa to establish adequate stream ecohydrological characterisation [21]. Notwithstanding the usefulness of elmids as bioindicators of freshwater ecosystems integrity in tropical zones [22], there is little research that focuses on the individual ecological requirements of the set of genera that the Elmidae family encapsulates. The problem with higher taxonomic resolution data is that they include several lower resolution taxa, which may have different environmental/ecological preferences. Thus, working on higher taxonomic resolution (e.g., Elmidae family) may mask the ecological sensitivities of taxa of lower resolution (e.g., genera of Elmidae) [23]. In this context, assessing the suitable habitats of different elmid genera is important to drive key study site conservation and restoration efforts. Thus, one way to cover the lack of knowledge about individual elmid ecohydrological preferences is through SDMs.

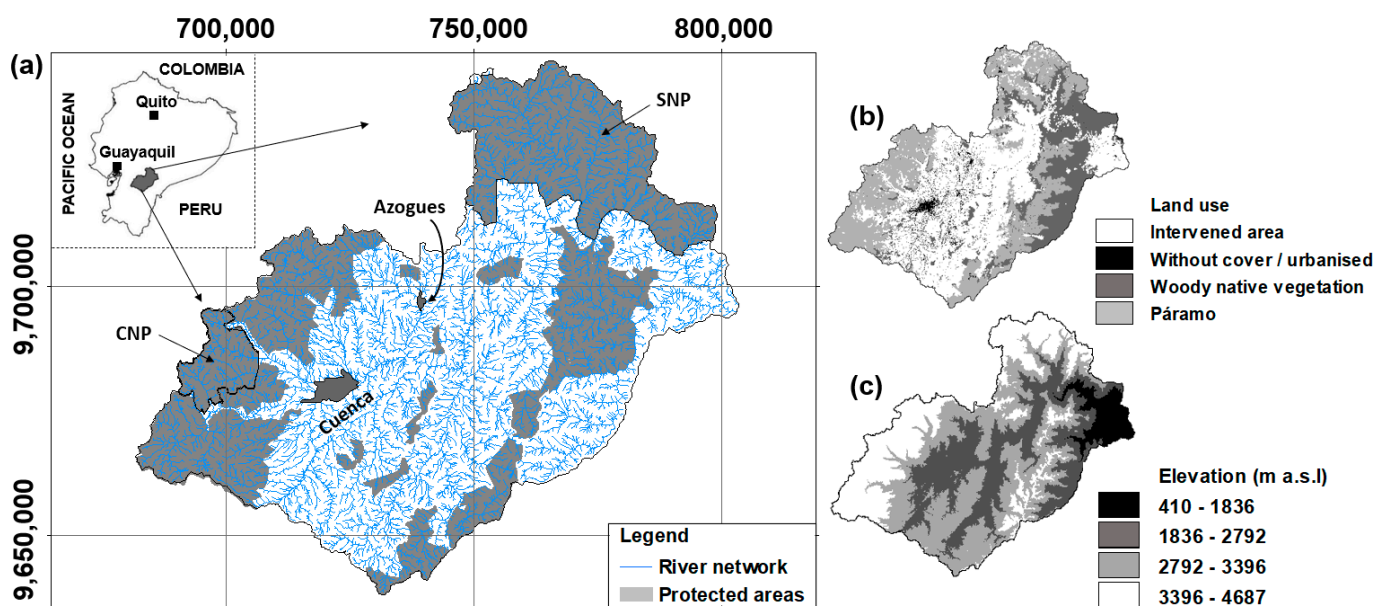
Further, while the species modelling framework is similar in the terrestrial, marine, and freshwater realms, each realm comprises specific challenges for combining the spatial scale, the environmental data, and the species records for building reliable models [24]. Thus, the choice of the modelling tool is an essential aspect of the development of SDMs. Worldwide, the Maximum Entropy Algorithm [25,26] is the most used tool for developing SDMs using the MaxEnt software [27]. Nevertheless, considering some common negative features of the SDMs, mainly dealing with the class imbalance nature of SDMs [28] and the availability of too few samples in large under-sampled areas [29], the use of the Random Forest (RF) algorithm is an attractive alternative [30–32]. Correspondingly, the current research echoes this latter trend by using the RF algorithm to model the occurrence probability of the elmid genera in the study basin. Within this frame of reference, the general goal of the current research was to develop and assess the SDMs models of riffle beetles (Coleoptera: Elmidae) in the Paute river basin. The main specific goals of the current research were (1) building

different SDMs for five genera of Elmidae recorded in the study basin, (2) identifying the most important environmental factors that explain the spatial distribution of the elmid genera, and (3) performing a congruence assessment of the different SDMs of the study elmid genera.

## 2. Materials and Methods

### 2.1. Study Area

The Paute river basin (PRB), in the south of Ecuador (Figure 1), has an area of 6442 km<sup>2</sup>, including the eastern lower portion towards the Amazon plateau. Its elevation ranges between 410 and 4687 m above sea level (a.s.l.), and slopes vary between 25% and 50%. The lower temperatures correspond to the western Andes range with a mean daily value of about 6 °C (at about 3500 m a.s.l.), while the warmest areas are situated in the Amazonian-influenced valleys and subtropical zones, with a mean daily value of 24 °C; nevertheless, a remarkable diurnal amplitude was observed. Due to the altitudinal gradient, mean annual rainfall oscillates in intensity and duration, with the lowest value of 660 mm at the basin's centre and the highest observed value exceeding 3400 mm near the basin outlet. On the other hand, meteorological stations located at higher elevations (above 3000 m a.s.l.) receive between 1000 and 1400 mm [33]. Two major cities, namely Cuenca and Azogues, are in the basin, with approximately 600,000 and 40,000 inhabitants, respectively. Important conservation zones are in the study basin, the most relevant (Figure 1) the Cajas National Park (CNP) and the Sangay National Park (SNP), both UNESCO World Heritage Sites. However, despite these conservation efforts, domestic wastewaters, agricultural runoff, animal husbandry, and industrial effluents are negative factors that are known to influence the surface water quality (WQ) of the study basin [21,34,35].

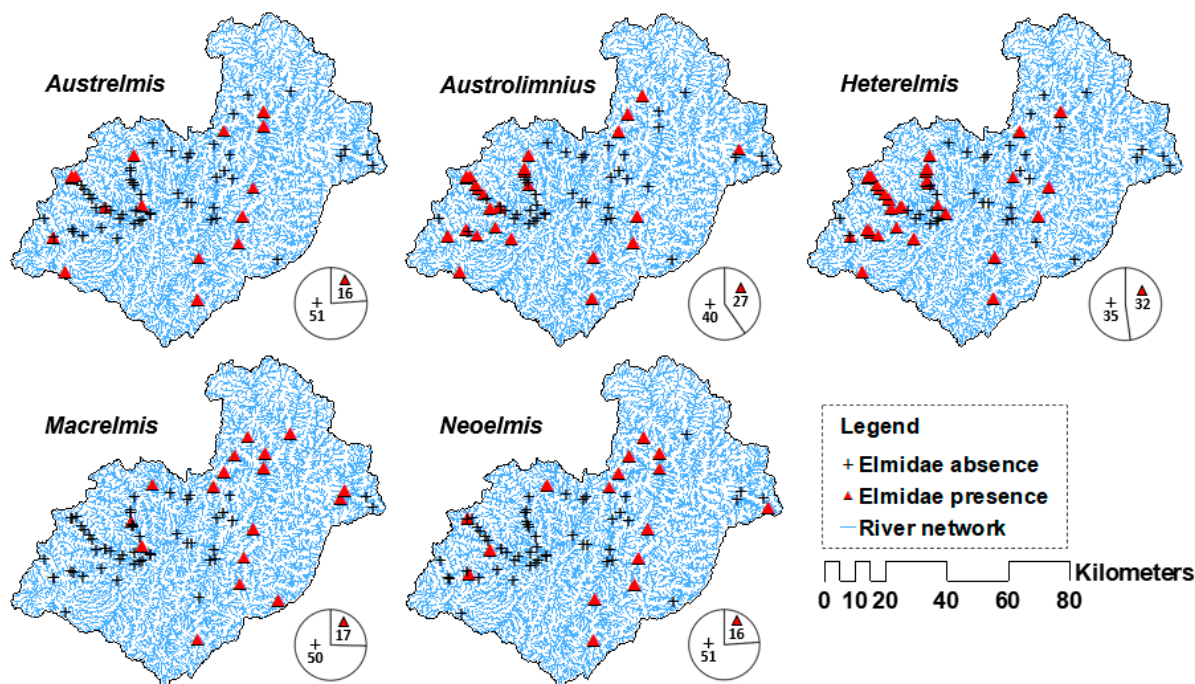


**Figure 1.** (a) Location of the Paute river basin in continental Ecuador and its two largest cities (Cuenca and Azogues); (b) main land cover classes distribution; and (c) elevation distribution in the basin. CNP = Cajas National Park; SNP = Sangay National Park. Coordinates system: WGS84 UTM 17S (m).

### 2.2. Sampling of Riffle Beetles

The benthic macroinvertebrate community was sampled at 67 sites located in the study basin throughout four years (2010–2012 and 2015) by the former Ecuadorian National Secretary of Water (SENAGUA), Santiago River Hydrographic Demarcation (DHS), and the Municipal Public Enterprise of Telecommunications, Drinking Water, Sewerage and Sanitation of Cuenca (ETAPA EP). Samples were collected using a D-frame kick net (25 cm aperture, 0.5 mm mesh) [36]. Sampling encompassed all existing microhabitats charac-

terised by different depths, substrates, and water velocities. Macroinvertebrate samples were preserved in 70% ethanol and sorted using a stereomicroscope. Using these samples, the presence–absence data records of elmids were obtained (Figure 2). The sampling sites were visited four times per year (on average). Some were sampled more frequently because they were located either at highly impacted sites or, on the contrary, at unaltered environmental (i.e., reference) locations.



**Figure 2.** Spatial distribution of the observed presence–absence of the five genera of Elmidae records in the Paute river basin. The pie charts indicate the number of sampling sites with presence–absence records (the total number of sampling sites is 67).

### Riffle Beetles and Their Presence–Absence Records

A total of 1672 elmid records were compiled and grouped into five genera (i.e.,  $n_{\text{gen}} = 5$ ) that belong to the subfamily Elminae [37], namely, *Austrelmis* ( $g_1$ , 8.4%), *Austrolimnius* ( $g_2$ , 26.6%), *Heterelmis* ( $g_3$ , 30.3%), *Macrelmis* ( $g_4$ , 20.0%), and *Neoelmis* ( $g_5$ , 14.7%). The research was limited to genera as most records of Elmidae from the study basin were predominantly larvae, which can only be identified at the genus level [38]. Herein, to use a record of an elmid genus, such as presence data, to perform the modelling process, the minimal sample size [39] was greater than two individuals per taxa. The latter was carried out to minimise the probability that an individual of a given taxa was recorded accidentally (i.e., fortuitous arrival through a strong current or a dead individual drifted downwards by the river current, etc.) in the sampling station of interest.

### 2.3. Environmental Variables

This study used twelve environmental variables ( $12en_v$ ) as the (independent) descriptive factors to explain the spatial variability of elmid genera. The twelve variables (Table 1) were selected from a previous set of 20 variables ( $en_v$ ) upon a Pearson’s correlation analysis that enabled excluding redundant  $en_v$  characterised by positive or negative correlation magnitudes above 0.75 [40]. This was done to achieve a parsimonious model and to minimise the risk of overfitting it. The correlation analysis was performed with the R package ‘ENMTools 1.0’ [41]. The eight excluded  $en_v$  were: solar radiation, roughness index, stream power index, flow accumulation, Strahler stream order, canopy height, evapotranspiration, and environmental temperature.

The unit of analysis for this research was the hydrographic network of the basin generated from a Digital Elevation Model (DEM), a LIDAR product of the SIGTIERRAS project (<http://www.sigtierras.gob.ec> accessed on 7 February 2022) of the Ecuadorian government [42]. Its original horizontal resolution is 3 m, whilst its vertical precision is  $\pm 1.5$  m. However, to reduce computational running times to reasonable levels, its horizontal resolution was resampled up to 12 m using the Bilinear algorithm available in the Resampling set of tools of ArcGIS 10.4.1 software [43,44]. The respective resampled product (DEM<sub>r</sub>) was used in the rest of the analysis.

The hydrographic network (Hy<sub>net</sub>) was obtained using the Hydrology toolbox of ArcGIS 10.4.1, which applies the method for extracting hydrographic networks from DEMs [45,46]. Thus, the following steps were applied: (1) pre-processing the DEM<sub>r</sub>; (2) determining the flow direction; (3) calculating the cumulative amount of the flow confluence; (4) determining the confluence threshold; and (5) generating the hydrographic network [47,48]. Hereafter, the spatial distribution of each one of the 12en<sub>v</sub> was incorporated into a raster layer that was previously cropped according to the Hy<sub>net</sub> mask, producing a continuum of environmental predictors along the stream network [49].

**Table 1.** Description of the twelve environmental variables (12en<sub>v</sub>) used in developing the species distribution models (SDMs).

Source	Variable		Used Tool in ArcGis/ Methodology	Unit	Abbreviation	Ecological Importance	Range
	Name	Type					
DEM of SIGTIERRAS project (Corral and Montiel Olea, 2020)	Elevation	Continuous	Spatial Analyst > Hydrology > Fill	m a.s.l	Elev	Temperature tends to be colder at higher elevations, e.g., in páramo ecosystems, influencing water dissolved oxygen values [50–52].	411–4212
	Slope	Continuous	Spatial Analyst > Surface > Slope	Degree	Slp	Water velocity and, consequently, oxygen content, are related to slope [21].	0–74.1
	Flow direction	Categorical	Spatial Analyst > Hydrology > Flow Direction	(-)	Fdir	Flow direction is related to substrate accumulation and streambed heterogeneity [53].	1–128
	Shreve stream order	Continuous	Spatial Analyst > Hydrology > Stream Order	(-)	Shreve	High-stream order values are indicators of bigger discharges [21,54].	1–5367
	Eastness	Continuous	Spatial Analyst > Map Algebra > Raster	(-)	East	These factors are related to the terrain declivity, stream course direction, and luminosity, which affect water temperature, oxygen [56], and algae growth. Algae are food sources for certain elmids [57].	–1–1
	Northness	Continuous	Calculator [55]		Ntns		
		Sinuosity	Continuous	Stream Gradient and Sinuosity > Shapefiles > Calculate Sinuosity [58]	(-)	Snty	The sinuosity is related to the accumulation of sediments and channel heterogeneity [59].
National Institute of Meteorology and Hydrology ( <a href="http://www.inamhi.gob.ec">http://www.inamhi.gob.ec</a> accessed on 7 February 2022)	Precipitation	Continuous	Spatial Analyst > Map Algebra > Raster Calculator	mm	PP	Precipitation is directly related to water availability and indirectly to water velocity and oxygen content [60].	586.5–3237.7
Geopedological map, scale 1:25,000; SIGTIERRAS project (Corral and Montiel Olea, 2020)	Lithology	Categorical	Conversion > To Raster > Polygon to Raster	(-)	Ltlgy	Elements in the water and sediments of rivers are present because of the natural weathering of the surrounding lithology [61]. These elements conditionate the elmids [62].	1–78
	Soil type	Categorical	Conversion > To Raster > Polygon to Raster	(-)	Soils	Water chemistry of rivers is affected by surrounding soil units [63].	1–10

Table 1. Cont.

Source	Variable		Used Tool in ArcGis/ Methodology	Unit	Abbreviation	Ecological Importance	Range
	Name	Type					
Land Use map, scale 1:100,000 (MAE, 2013)	Riparian alteration	Continuous	[64,65]	%	Rip-alt	The riparian zones regulate water temperature and allochthonous organic matter inputs and mitigate the effects of anthropogenic pressures [21,66].	0–99
Global Land Analysis and Discover ( <a href="https://glad.umd.edu/dataset/ge">https://glad.umd.edu/dataset/ge</a> accessed on 7 February 2022)	Canopy	Continuous	Data Management > Raster > Raster Processing > Resample	(-)	Cnpy	Canopy attenuates the sunlight, regulates the water temperature of streams and favours streambed heterogeneity [66,67].	0–100

Eastness (East) and northness (Ntns) provide continuous measures describing geographical orientation in combination with slope. For the Northness, +1 represents the north and –1 south directions. For Eastness, +1 represents the east and –1 the west directions [68]. Sinuosity (Snty) provides the degree of meandering of the stream channel. In general, Snty = 1 is linked to a straight channel, and Snty = 4.8 is the maximum degree of meandering in the Paute river basin hydrographic network. The Lithology (Ltlgy) variable implies 78 lithological groups for the Paute river basin. The first half of these 78 groups correspond to sands, sandstones, clasts, and schists; the second half corresponds to silts, clays, pyroclasts, and undifferentiated metamorphic rocks. The soil type (Soils) variable accounts for the ten soil units that exist in the study basin, i.e., Andisols (1), Inceptisols (2), Mollisols (3), Vertisols (4), Entisols (5), Alfisols (6), Oxisols (7), Histosols (8), Ultisols (9), and miscellaneous (10). Canopy (Cnpy) ranges from 1 to 100, with 1 representing riparian areas without the presence of forest and 100 riparian areas with high forest presence.

#### 2.4. Species Distribution Models (SDMs) Using Random Forest (RF) Algorithm

The Random Forest (RF) algorithm [69] is an ensemble of classification or regression trees and is widely used in research, including SDMs analyses [70]. It performs classification analysis by building many decision trees from bootstrap data set samples. The final model prediction is performed by averaging the predictions made by each tree in the forest. In this study, RF was implemented using the R Package ‘Biomod2’ [71] to model the spatial distribution of the presence–absence of elmid genera (i.e., occurrence probability) as a function of the 12en<sub>v</sub>.

Each elmid genus was separately modelled, i.e., five RF modelling processes were performed. The tuned parameters to estimate the different RF models were the number of trees ( $n_{\text{tree}}$ ) and the number of variables randomly selected at each node (mtry), given that the RF algorithm is prone to be sensitive to these parameters [72,73]. Herein, for parameterising the RF algorithm, the strategy of Strobl et al. [74,75] was implemented. It was based in a grid search through which all possible combinations of given discrete parameter regions were evaluated. Values of mtry = 5 and  $n_{\text{tree}}$  = 3000 were adopted in this study after a sensitivity analysis that showed more consistent results with these values.

The different RF runs were carried out using the K-fold cross-validation (CV) method, in which the data were divided into K disjoint sets (folds), and the K-th fold was used as an independent test (i.e., validation) set. The remaining (K – 1) folds were used to train the RF model and find its different parameters, after which model validation took place using the test set. This process was repeated n times. The error estimation was averaged over all n trials to get the total effectiveness of the model [76]. For the current research, K = 4 with n = 3 repetitions was used for each elmid genera, producing twelve runs (models) for each of the 5 genera. Trade-offs were involved when selecting K number of folds [77]. Using K = 4 (implying the use of 75% of the data for training and 25% for validation) has been reported as an excellent value to perform a realistic classification assessment [78].

Since the available response variables, i.e., presence–absence records of elmid genera, were imbalanced (Figure 2), RF was chosen for this study because it is known to work well

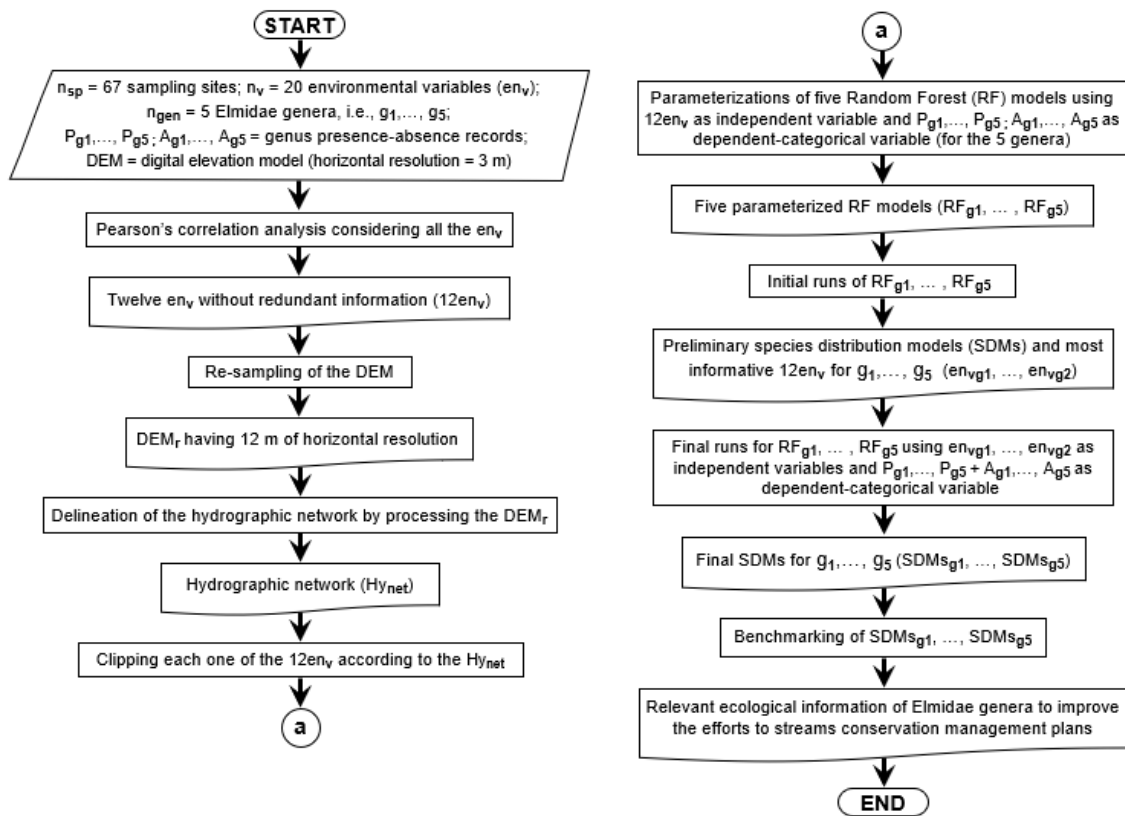
with imbalanced data sets in a classification framework [79–81]. For further details about RF, the reader is kindly referred, for instance, to [82,83].

For evaluating the RF outputs, the area under the receiver operating characteristic (ROC) curve was used, which was applied for the analysis of classification performances in the framework of binary classification of samples as positive (P) or negative (N). In this context, the ROC curve is defined [84] as a plot of  $x = 1 - Sp_p$  (specificity of the positive class, also known as False Positive Rate, FPR) versus  $y = Sn_p$  (sensitivity of the positive class, also known as True Positive Rate, TPR). Given the ROC curve for a classifier, the area under the curve (AUC) measures its overall diagnostic performance, with  $AUC = (Sp_p + Sn_p)/2$  [85]. Since the AUC is a portion of the area of a unit square, its value varies between 0 and 1, with 1 being its optimal value. For each elmid genus, there were 12 output models because of the cross-validation process; thereby, 60 AUC values (i.e.,  $n_{gen} \times 12$ ) were obtained in total. For each elmid genus, its 12 AUC values were aggregated into a single value using central tendency measures. Before this aggregating process, the normality of each set of 12 AUC values was checked [35,86] using the Shapiro–Wilk (S–W) test [87] considering a 95% confidence level. For a particular elmid genus, if the S–W test suggested normality, the mean AUC value was used for aggregating; otherwise, the median was assigned as the aggregated AUC value [88,89]. For the interpretation of the AUC values, it followed the proposal of Hosmer et al. [90], where an  $AUC = 0.5$  could be interpreted as “no discrimination”;  $0.5 < AUC < 0.7$  as “poor discrimination”;  $0.7 \leq AUC < 0.8$  as “acceptable discrimination”;  $0.8 \leq AUC < 0.9$  as “excellent discrimination”; and  $AUC \geq 0.9$  as “outstanding discrimination”.

#### Assessing Significant Environmental Variables

The ‘Biomod2’ R package [91] uses a random permutation procedure to estimate the importance ( $var_{imp}$ ) of each  $12en_v$ . The procedure (Figure 3) is independent of the modelling technique. It uses Pearson’s correlation between the standard prediction (i.e., fitted values) and the predictions obtained by focusing the simulations on a given environmental variable and randomly permutating its value for every simulation. If the correlation is high, i.e., showing little difference between the standard and a given prediction, the given variable is considered unimportant for the model. This is repeated several times for each given variable, and Pearson’s mean correlation coefficient over the runs is kept. Herein, the number of permutations to estimate the  $var_{imp}$  for every one of the  $12en_v$  was 5.

As a result, the R ‘Biomod2’ package produces a ranking of variables and their corresponding  $var_{imp}$  values. In this context, for each elmid genus, there were 12 output models because of the cross-validation process, thus 144  $var_{imp}$  values (i.e.,  $12en_v \times 12$  output models). In five separate analyses (i.e., one for each genus), for each  $12en_v$ , their 12  $var_{imp}$  values were aggregated. Thus, as in the case of AUC, before the aggregating process took place, the normality trend of each set of 12  $var_{imp}$  values was checked using the S–W test, considering a 95% confidence level. For each elmid genus, their aggregated  $var_{imp}$  values from each  $12en_v$  were expressed as percentages and ranked in descending order. However, there is not a statistically based  $var_{imp}$  threshold on distinguishing between important and non-important  $en_v$  to explain the spatial distribution of elmid. Thus, a variable segregation analysis was carried out for each genus. In this analysis, the important set of  $en_v$  ( $en_{v-imp}$ ) was identified by removing, one by one, the non-important  $en_v$  with respect to the (standard) RF-based model containing all the  $12en_v$ . On every occasion, after a given  $en_v$  was removed from a previous RF-based model, the complete modelling approach was repeated so that a cross-validation analysis was entirely performed, and the respective aggregated AUC value was obtained for the newer model. This variable segregation approach was carried out until the best RF-based model, formed by the set of  $en_{v-imp}$ , was identified by the highest aggregated AUC value. In this analysis, variables were removed by considering the ranked  $var_{imp}$  information so that the  $en_v$  associated with lower  $var_{imp}$  was removed first.



**Figure 3.** Flowchart of the methodology implemented in the current study.

Additionally, for each  $en_{v-imp}$  of elmid genera, response curves were created using the AUC data. Thus, to define the optimal range for the distribution of each genus per each  $en_{v-imp}$ , a cluster analysis through the k-means method was implemented using the AUC values. With this procedure, it is possible to distinguish the statistical cut-off AUC values (i.e., borders or thresholds) and, thereby, the optimal range of preference of each genus for each  $en_{v-imp}$ . K-means clustering is a non-hierarchical clustering method that assigns each object to the group with the closest centroid by calculating the centroid of each group [92]. This study applied this method using the Euclidean distance as the similarity measure between objects. In the k-means algorithm, the number of clusters is specified a priori, usually according to some hypothesis [93]; however, a more robust statistical procedure uses internal validation indices [94]. Using quantities and features inherent in the data, an internal index measures the appropriateness of clustering partitions without external information [95]. Herein, an internal validity index was applied, namely, the Silhouette Coefficient (SC). The Cluster Validity Analysis Platform (CVAP) was used for this purpose [94]. SC [96] is a dimensionless measure that evaluates the quality of compactness and separation of clusters; with an upper bound equal to 1, the optimum k value corresponds to its largest average. The inspected number of clusters k was from 2 to 5.

### 2.5. Prediction of Spatial Distribution

In the 'Biomod2' R package, the final generated models using the environmental space ( $en_{v-imp}$ ) were projected within the  $Hy_{net}$  to create the spatial predictions for each elmid genera [97]. These SDMs contain the occurrence probability values for each elmid genera. Correspondingly, values close to 0 indicate probable absences, and values close to 1 suggest probable presences. The twelve SDMs outputs for each genus of the Elmidae family were exported in ESRI raster format (GRID) to facilitate their processing/averaging using the Raster Calculator tool available in ArcGIS 10.4.1 [91]. As a result, one final SDM for each elmid genus was created, i.e.,  $SDM_{g1}, \dots, SDM_{g5}$ . Further, to improve the visualisation of



these SDMs, each one of them was reclassified considering three probability classes of the spatial occurrence of modelled taxa, i.e., low (C1, 0–0.33), medium (C2, 0.3–0.66) and high (C3, 0.66–1). This number of classes was chosen following a previous study (Sotomayor et al., 2020, 2021) that concluded that three classes are adequate for characterising the water quality in the Paute river basin.

2.6. Congruency of the Predicted Spatial Distribution of the C3 Probability of Occurrence of Elmid Genera

The congruency of the C3 class of occurrence probability of elmid genera was assessed by its visual comparison with the land use (LU)–land coverage (LC) distribution in the study basin [98]. The LU–LC data were reclassified to describe the spatial distribution of the anthropogenic impact (higher or lower) level, which was compared to the distribution of the C3 class of occurrence probability of elmid genera. This latter distribution was obtained by merging the respective spatial distribution of every one of the five study genera. The original LU–LC classes [98] were the following: (1) altered vegetation; (2) woody native vegetation; (3) without cover/urbanised, (4) páramo ecosystem; and (5) water. The higher anthropogenic impact class was defined upon the reclassification of LU–LC classes 1 and 3, whilst the lower anthropogenic impact class was defined upon classes 2, 4, and 5. Thus, these anthropogenic impact classes are not the result of any additional calculation of an index or a factor but just a simple reclassification of the original LU–LC information. ArcGIS 10.4.1 was used for all the respective Geographic Information Systems (GIS) analyses.

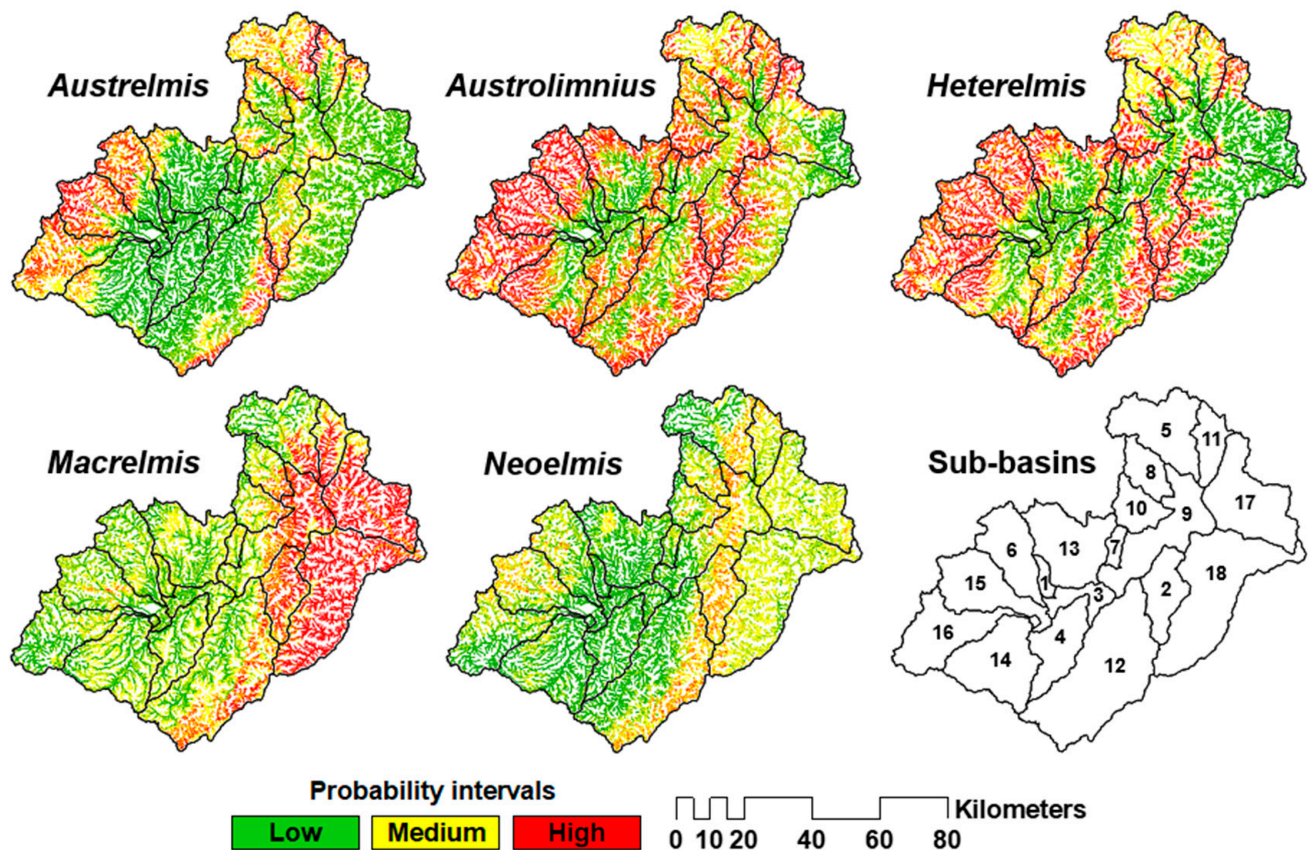
3. Results

3.1. Species Distribution Models (SDMs)

A significantly outperforming RF (i.e., aggregated AUC values) was observed when only the informative environmental variables (i.e.,  $en_{v-imp}$ ) were used in the modelling process. The statistical performance of the RF models, i.e., the aggregated AUC values linked to each SDM of elmid genera (Table 2), suggested that the RF model for *Austrolimnius*, *Austrelmis*, and *Macrelmis* had the best performance (i.e., excellent discrimination), followed by *Heterelmis* and *Neoelmis* (i.e., acceptable discrimination) [90]. The aggregated AUC values ranged from 0.76 to 0.89 (Table 2). The spatial extent for each genus of the Elmidae family upon their probability of occurrence ranges (i.e., C1, C2, and C3) indicated that *Austrolimnius* and *Heterelmis* are the taxa with the most widespread spatial probability of occurrence in the study site. *Macrelmis* would occur, in its great majority, in the lowest basin areas, toward the Amazon basin. On the contrary, *Austrelmis* is likely to occur in the higher elevations of the basin, especially in the protected zones such as the Cajas National Park (CNP) and the Sangay National Park (SNP). *Neoelmis* shows low and medium probabilities of occurrence in the studied basin (Table 2, Figure 4).

**Table 2.** Random Forest predictions for each genus of the Elmidae family. The aggregated area under curve (AUC) values are the result of considering either all the independent variables (“step 1”) or only the significant variables identified as truly important for explaining the spatial probability of occurrence of elmids (“step 2”). “\*” indicates that the median was chosen as the aggregated AUC value. SDM = species distribution model. Probability of occurrence class: low (C1), medium (C2), high (C3).

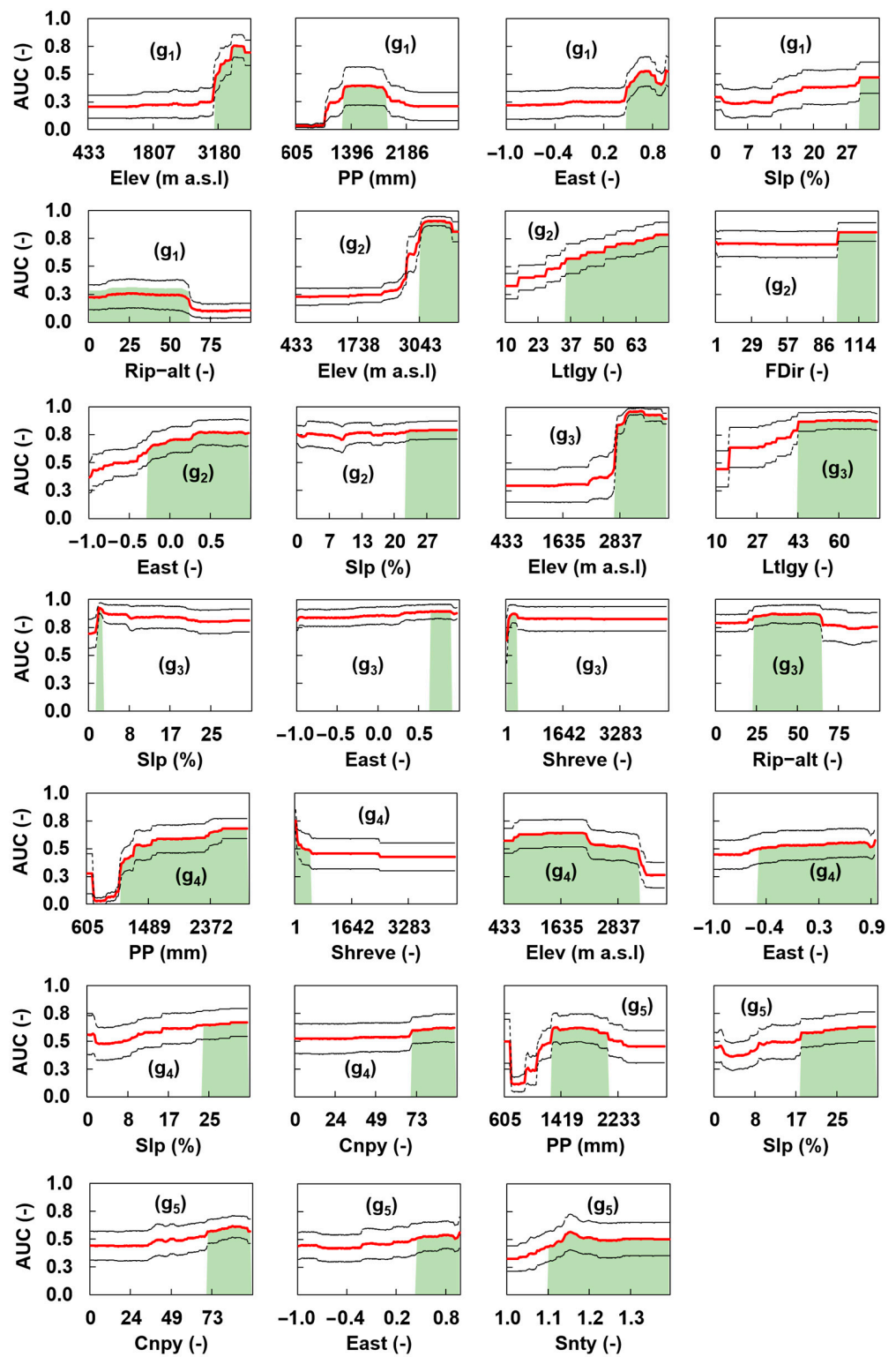
Genus	AUC		SDM of Probability of Occurrence			
	Mean/Median		Probability Range	Spatial Extent (%)		
	(Step 1)	(Step 2)		C1	C2	C3
<i>Austrelmis</i>	0.76	0.83	0.00–0.94	52.8	35.4	11.9
<i>Austrolimnius</i>	0.87	0.89 *	0.00–1.00	25.0	37.4	37.7
<i>Heterelmis</i>	0.76	0.79	0.01–0.99	33.3	34.9	31.7
<i>Macrelmis</i>	0.76	0.82	0.00–0.94	28.6	41.4	30.0
<i>Neoelmis</i>	0.70	0.76	0.00–0.87	48.1	48.9	2.9



**Figure 4.** Spatial distribution of the probability of occurrence for the five genera of Elmidae in the Paute river basin, considering three classes, i.e., low (C1), medium (C2), and high (C3). Sub-basins: 1 = Sidcay, 2 = Collay, 3 = Cuenca, 4 = Jadán, 5 = Juval, 6 = Machángara, 7 = Magdalena, 8 = Mazar, 9 = Paute, 10 = Pindilig, 11 = Púlpito, 12 = Santa Bárbara, 13 = Burgay, 14 = Tarqui, 15 = Tomebamba, 16 = Yanuncay, 17 = Paute bajo, and 18 = Negro.

### 3.2. Assessing Significant Environmental Variables

The  $en_v$  that showed the highest association with the spatial distribution of Elmidae were slope, eastness, elevation, precipitation, Shreve stream order, lithology, canopy, percentage of riparian alteration, flow direction, and sinuosity (Table 3). Northness and soil types were non-important variables. The response curves of the important variables ( $en_{v-imp}$ ) for each genus and their optimal probability range of preference are presented in Figure 5. The first  $en_{v-imp}$  for all genera of the Elmidae family was the most important for modelling the spatial distribution of the occurrence probability of a given elmid genus (Table 3). The curves of the first  $en_{v-imp}$  differed from the symmetric bell-shaped form. They exhibited clear peaks and depressions (Figure 5), indicating the variable ranges associated with higher (and lower) values of the probability of occurrence of elmid genera. Further, the  $en_{v-imp}$  that were lower in relevance (Table 3) had less discriminatory power in modelling the spatial distribution of the occurrence probability of a given elmid genus and, as such, exhibited fewer clear peaks and depressions (Figure 5).



**Figure 5.** AUC response curves (red line) of each genus of the Elmidae family, namely: *Austrelmis* (g<sub>1</sub>), *Austrolimnius* (g<sub>2</sub>), *Heterelmis* (g<sub>3</sub>), *Macrelmis* (g<sub>4</sub>), and *Neoelmis* (g<sub>5</sub>) as a function of the important environmental variables (en<sub>v-imp</sub>). Dashed lines define the AUC standard deviation band. The highlighted area under the AUC indicates the optimal range of preference for the different Elmidae genera. Cnpy = Canopy, Elev = Elevation; East = Eastness, Fdir = Flow direction; Ltlgy = Lithology, PP = Precipitation, Rip-alt = Riparian alteration, Shreve = Shreve stream order, Slp = Slope, Snty = Sinuosity.

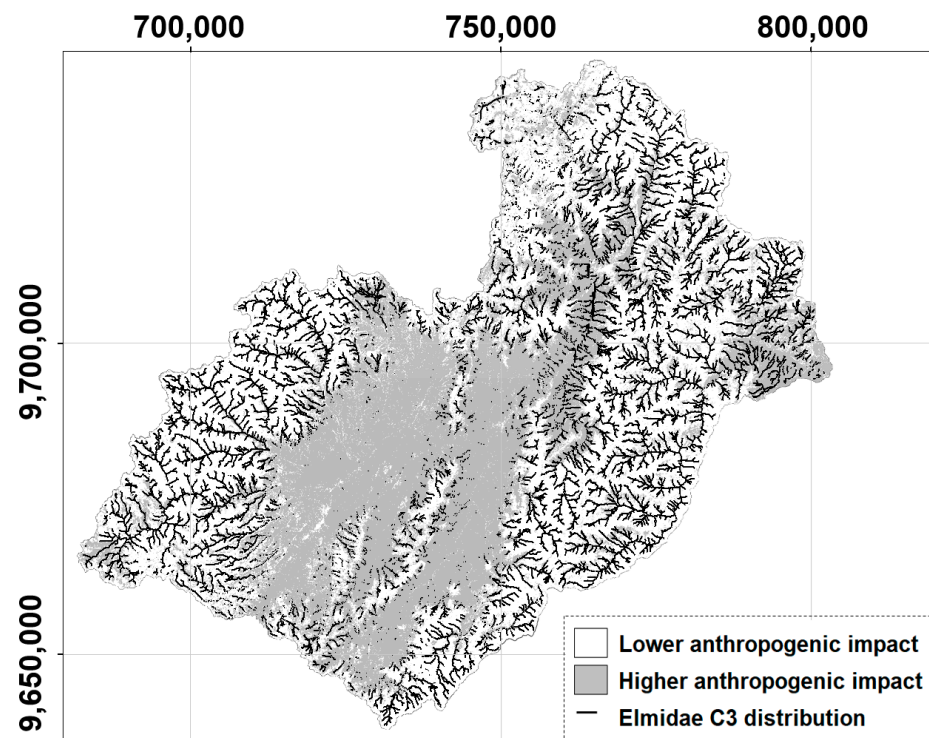
**Table 3.** Environmental variables that were identified as important to explain the spatial variability of each genus of the Elmidae family. The importance value ( $\text{var}_{\text{imp}}$ ) for each variable is expressed in percentage (“\*” indicates that the median central tendency measure was used to define the aggregated variable value). Cnpy = Canopy; Elev = Elevation; East = Eastness; Fdir = Flow direction; Ltlgy = Lithology; PP = Precipitation; Rip-alt = Riparian alteration; Shreve = Shreve stream order; Slp = Slope; Snty = Sinuosity.

Genera	Environmental Variable and Its Weight (%)					
<i>Austrelmis</i>	Elev *	PP *	East *	Slp *	Rip-alt *	
	28.92	24.56	16.90	6.45	5.57	
<i>Austrolimnius</i>	Elev	Ltlgy *	East *	Fdir *	Slp *	
	51.70	35.39	4.10	2.11	1.86	
<i>Heterelmis</i>	Elev	Ltlgy	Slp *	East	Shreve *	Rip-alt *
	52.27	26.95	6.63	5.00	3.27	2.06
<i>Macrelmis</i>	PP	Shreve	Elev *	East	Slp *	Cnpy *
	53.56	19.56	7.17	6.23	5.76	2.83
<i>Neoelmis</i>	PP	Slp	Cnpy	East *	Snty	
	47.62	10.06	9.41	6.38	5.45	

Upon the AUC curves shown in Figure 5, the following environmental requirements for the Elmidae genera, reflected by the respective optimal probability of occurrence ranges, are distinguished. Higher values of the probability of occurrence for *Austrelmis* ( $g_1$ ) are in streams characterised by the environmental variable ranges: elevation [3111–3833] m a.s.l., precipitation [1279–1883] mm, eastness [0.49–0.99], slope [29–33]%, and riparian alteration [0–61]%. Higher probability values for *Austrolimnius* ( $g_2$ ) are in streams characterised by the ranges: elevation [3043–3833] m a.s.l., eastness [−2.7–0.990], flow direction [97.2–126.7], and slope [22.5–32.9]%. Higher probability values for *Heterelmis* ( $g_3$ ) are in streams characterised by the ranges: elevation [2734–3798] m a.s.l., lithology [43.3–75.3], slope [0.33–2.68]%, eastness [0.65–0.89], Shreve order [47.9–282.2], and riparian alteration [28.0–63.0]%. Higher probability values for *Macrelmis* ( $g_4$ ) are in streams characterised by the environmental ranges: precipitation [1093.4–2883.5] mm, Shreve order [1.0–422.9], elevation [433.0–3249.2] m a.s.l., eastness [−0.46–0.97], slope [23.8–32.9]%, and canopy [70.5–96.0]. Higher probability values for *Neoelmis* ( $g_5$ ), are in streams denoted by the ranges: precipitation [1279.4–2069.8] mm, slope [17.8–32.9]%, canopy [70.6–96.0], eastness [0.45–0.97], and sinuosity [1.1–1.4].

### 3.3. C3 Class of Occurrence Probability of Elmidae across the Paute River Basin

The spatial distribution of the C3 class of occurrence probability of Elmidae across the study basin and the respective distribution of the anthropogenic impact are shown in Figure 6. A lower anthropogenic level characterises 59.4% of the basin area, whilst the remaining 40.6% exhibits a higher anthropogenic level. The figure depicts that the C3 class of occurrence probability of Elmidae is, in average terms, not distributed in higher anthropogenic impacted zones, which is congruent with elmids being prone to be absent in (water quality) impacted zones.



**Figure 6.** Spatial distributions of the C3 class of occurrence probability of elmid genera and estimated anthropogenic impact in the Paute river basin, Ecuador.

## 4. Discussion

### 4.1. Model Selection

The use of the RF algorithm was successful in terms of the achieved modelling performance. Notwithstanding, most studies that chose the SDMs framework of analysis utilised the Maximum Entropy Algorithm [25,26] using the MaxEnt software [27] as their primary modelling tool since it has a user-friendly graphical interface (i.e., it is easy to use and enables a lucid visualisation of results) [99]. This trend of using MaxEnt has also been observed in Ecuador for spatially assessing organisms belonging to different biological communities [100–106]. However, some limitations of the MaxEnt modelling tool have been reported [107]. As it is the most relevant that this approach considers only presence (i.e., occurrence) data, this implies that the prevalence of the species (i.e., the proportion of occupied sites) cannot be precisely determined [108,109]. A second fundamental limitation of MaxEnt is that sample selection bias (whereby some areas in the landscape are sampled more intensively than others) has a much stronger effect on presence-only models than on presence–absence models. In this context, if the presence–absence survey data are available as in the current research, it is generally prudent to use a presence–absence modelling method [109], such as the RF algorithm, which has been tested as one of the most accurate tools for the construction of SDMs [29–32,110–112].

### 4.2. Model Performance

Some differences were observed in the performance of the RF algorithm (characterised by the AUC values) modelling the distribution of the five elmid genera (Table 2). Species distribution modelling quality could be assessed considering subjective AUC conditions/ranges. Alternatively, it could be adopted the AUC conditions/ranges suggested by Hosmer et al. [90] or the AUC condition (i.e., >0.7) used by Cha et al. [32] to regard the modelling quality as “excellent”. According to the first AUC criteria, the RF SDMs of *Austrelmis*, *Austrolimnius*, and *Macrelmis* could perform an overall “excellent discrimination”; the respective RF models of *Heterelmis* and *Neoelmis* performed an overall “acceptable

discrimination". In consonance with the second AUC criterion, given that the AUC values of all the RF SDMs of the study genera were higher or equal to 0.7, the overall performance of all the SDMs could be regarded as "excellent". The False Positive Rate (FPR) confirms the AUC findings, in the sense that the RF-based SDMs for *Austrelmis* (19%), *Austrolimnius* (16%) and *Macrelmis* (19%) performed better than the respective models for *Heterelmis* (27%) and *Neoelmis* (34%).

Model performance is likely to be compromised using genera instead of species taxonomic level since genera can contain several different congeners, which could have different ecological requirements. However, all study genera prefer clean water conditions [34,35,86]; as such, it is assumed that congeners within each genus also have similar preferences. Hence, it is assumed that, under the current conditions, the model performance is only being compromised in a marginal level.

#### 4.3. Basic Findings of the Developed SDMs

Elmidae family members are indicators of good stream ecosystem status [15,18,113,114]. The occurrence probabilities of elmids predicted in this study are notably lower in the human-impacted central region of the study catchment (Figures 4 and 6). Further, some substantial differences in ecological requirements were predicted by the SDMs of elmid genera in the current research (Figure 4). These differences are like previous studies [16,22,115,116], which found that genera in the Elmidae family differ regarding their ecological requirements. In this context, despite some relevant works that have been published on the taxonomy of the Ecuadorian Elmidae in the last few years [117–119], no work has been done about species distribution modelling of Elmidae. Further, to the best of our knowledge, just one SDM study using MaxEnt considered Elmidae in a southern Brazilian basin [60]. Contrasting the findings of this study (Figure 4), Braun et al. [60] found very similar spatial distributions of the occurrence probability of their study genera. These differences in the outcomes of both studies are likely to be the consequence of (i) the different characteristics (elevation range, climate, geology, etc.) of the study basins and (ii) the different modelling approaches that were used in either study (i.e., presence–absence or presence-only modelling).

#### 4.4. Important Predictors for Elmids Distribution

Despite the relatively similar identification of important variables for the models of the five study genera, each genus model has its own set of informative variables (Table 3, Figure 5). This finding emphasises the (predicted) dissimilarities of ecological requirements of some of the study genera.

González-Córdoba et al. [22] found in the Colombian Andean region that *Austrelmis* survive in a narrow temperature range of cold water. In contrast, *Austrolimnius* and *Heterelmis* tolerate a wide temperature range and survive cold and relatively warm water. Given the tight inverse relationship between elevation and temperature [120], both findings of González-Córdoba et al. [22] fit with the current research, where the higher probability of spatial occurrence of *Austrelmis* is between 3111 and 3833 m a.s.l. (Figures 4 and 5), whilst *Austrolimnius* and *Heterelmis* tend to occur at higher and lower elevation streams (ranges between 3043 and 3833 and 2734 and 3798 m a.s.l., respectively). This highlights the importance of elevation for the spatial distribution of the *Austrelmis*, *Austrolimnius* and *Heterelmis*. The elevation preference for *Austrelmis* is like what has been estimated for the Cañete River basin in the south of Perú [121].

For *Austrelmis*, *Austrolimnius*, and *Heterelmis*, the lithology and the percentage of riparian alteration were selected as informative (Figure 5). Consistent with this finding, in the Ocoa river basin, Colombia, the conservancy of the riparian ecosystems has been reported as a critical aspect for *Austrolimnius* and *Heterelmis* [122]. It is likely to be the consequence of the fact that elmid members frequently use riparian forests for their terrestrial pupation [123]. Further, *Austrolimnius* and *Heterelmis* apparently prefer rivers located in formations such as silts, clays, pyroclasts, and undifferentiated metamorphic rocks;

however, in the literature, there are no specific studies about the influence of this variable on Elmidae members. Overall, it is known that the macroinvertebrate communities are generally modified by local factors such as geology [124]; for example, the influence of the local geology was responsible for the high concentrations of salts in the Lincha river sub-basin (Perú), a factor that conditioned the presence or relative abundance of some taxa, including elmids [121].

The precipitation was another essential variable for the ecological requirements of *Macrelmis* and *Neoelmis* (Figure 5). In the case of *Macrelmis*, it is mainly predicted in the east of the study basin, where higher precipitations, in the range from 1090 to 2885 mm, are observed [125]. Because precipitation influences stream discharge, the latter is consistent with previous findings, as this genus is positively correlated with higher discharges [126].

Some studies found that *Macrelmis* is related to higher water temperatures [127–129]. Similarly, the respective SDM predicted a high probability of occurrence of *Macrelmis* towards the Amazon region where the temperature is higher. For *Macrelmis*, the canopy variable was important in the species distribution modelling process (Figure 5). Previous findings for *Macrelmis* suggest that they are closely related to areas with forested biomes [130], which is congruent with the findings of the current research as the high probability values for *Macrelmis* were predicted to occur in its great majority in the east of the basin (Figure 4), where high canopy values characterise sub-basins.

In the case of *Neoelmis*, the canopy variable was significant in its species distribution modelling (Figure 5), as members of this taxon were rarely predicted in the middle stripe of the study basin where forested areas are limited, and anthropogenic activities are important. González-Córdoba et al. [22] found that *Neoelmis* was present in a wide range of temperatures and even tolerates medium to high degrees of contamination. It is likely these differences are based on the hydrological systems and on the frameworks of research (i.e., the work of González-Córdoba et al. [22] is far from the concept of SDMs). Additionally, the possibility that different congeners vary in habitat preferences could explain the dissimilarity of the findings of *Neoelmis* between both studies. Just for *Neoelmis*, the sinuosity (i.e., Snty) was selected as an informative variable to explain its spatial distribution in the study basin, i.e., members of *Neoelmis* prefer streams with a high degree of meandering (Figure 5). However, although the importance of Snty to elmids has been described [131], for the specific case of *Neoelmis* no similar findings exist like the current research.

#### 4.5. Elmidae Genera's SDMs and Their Implications for the Surface Water Quality Management in the Study Basin

The high probability values of *Austrolimnius*, *Heterelmis*, and mainly *Austrelmis* are rarely predicted in most parts of the Burgay and Magdalena sub-basins (Figure 4), which have been described as polluted systems where the domestic and industrial wastewater discharge, extensive agriculture, cattle ranching, and the loss of native vegetation cover are the anthropogenic threats that cause severe surface water quality pollution and the subsequent loss of benthic macroinvertebrates taxa [21,34,35,132,133]. For the study basin, Sotomayor et al. [21,86] found that Elmidae is a keystone family in establishing adequate stream water quality assessments. Thus, despite the dissimilarity of genera in the Elmidae family regarding their ecological requirements, the overall trend of elmids members is that their occurrence probability in the study basin is higher in areas with good levels of conservation, e.g., protected areas (Figure 6). This is emphasised through the cross-check analysis of Figure 6, i.e., the high occurrence probability of Elmidae is less distributed in the areas with high anthropogenic impact. Likewise, in the high zones of the study basin, the anthropogenic activities are less than in their lower zones. That is, the biogeographical importance of the elevation for the potential distribution of elmids members is notorious. However, indirectly, anthropisation is a factor linked with the elevation, and both signals were detected in the current research.

## 5. Conclusions

Using the Random Forest algorithm, the genera of the Elmidae family were predicted in a great majority with good statistical reliability in healthy streams of the Paute river basin located in areas with good conservation status, i.e., protected areas. Some clear differences in ecological and environmental requirements were registered for some of the modelled elmid genera in the basin. The high probability values of spatial occurrence for *Austrelmis* are linked chiefly to streams of high mountains, i.e., the páramo ecosystems. Additionally predicted in the upper parts of the basin were *Austrolimnius* and *Heterelmis*, but at more varied elevation ranges, and *Heterelmis* in some areas where human activity was moderate. Contrary, *Macrelmis* were the majority predicted in the east of the studied basin in forest areas with high canopy values towards the Amazon system (also with high precipitation levels). According to predictions, *Neoelmis* would inhabit high altitudinal streams with a high degree of meandering in the middle portion of the Paute river basin. There is relative consistency in the informative variables that explain the spatial distribution of elmid genera in the study basin. Thus, factors such as elevation, precipitation, the quantity of water, and land use are linked to the general ecological requirements of the elmid genera and thereby to their occurrence probability. However, for each modelled genus, a specific set of environmental factors was observed, which implies dissimilarities in the ecological and biogeographical factors that govern the spatial distribution of the studied genera. However, despite these dissimilarities between elmid genera, the overall finding is that Elmidae members are good indicators of healthy freshwater ecosystems. The study revealed that the use of robust machine learning methods such as the one applied here, in conjunction with appropriate spatial analysis and visualisation tools, could be a promising approach to derive plausible geographical distributions of species (and genera) in support of conservation and management purposes, and that could be applied in other locations where suitable spatially distributed data are available.

**Author Contributions:** G.S.: Conceptualisation, methodology, software, data curation, formal analysis, investigation, writing-original draft. J.R.: Conceptualisation, methodology, software, writing-review and editing. D.B.: Methodology, writing-original draft, visualisation, writing-review and editing. R.F.V.: Methodology, writing-original draft, visualisation, writing-review and editing. I.R.-M.: Methodology, writing-review and editing. H.H.: Methodology, writing-review and editing. X.G.: Methodology, writing-review and editing. B.M.: Methodology, writing-review and editing. M.A.E.F.: Methodology, writing-review and editing. P.L.M.G.: Methodology, writing-review and editing. All authors have read and agreed to the published version of the manuscript.

**Funding:** The APC was kindly paid by the VLIR-UOS Biodiversity Network Ecuador. Not specific grant number applicable.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Restrictions apply to the availability of data used in this research. The database belongs to the former Ecuadorian National Secretary of Water (SENAGUA), which was recently absorbed by the Ecuadorian Environmental, Water and Ecological Transition Ministry (MAATE). The SENAGUA issued the respective data use authorisation to the first author in the scope of his PhD research. This signed agreement explicitly forbids him sharing the information with a third party without their specific authorisation, which is not easy to obtain. The data might be available directly from the MAATE.

**Acknowledgments:** The authors would like to express their gratitude to the former Ecuadorian National Secretary of Water (SENAGUA)-Santiago River Hydrographic Demarcation (DHS) and the Municipal Public Enterprise of Telecommunications, Drinking Water, Sewerage and Sanitation of Cuenca (ETAPA EP) for making the raw data accessible to the current study. The preparation of the manuscript took place in the framework of the undergraduate research in Biology of the second author, which was supervised by the third author and co-supervised by the first author and was carried out at the Department of Biology, Ecology and Management, Faculty of Science and Technology of the University of Azuay, Ecuador. Participation of the fourth and sixth authors in the



preparation of the current manuscript took place in the context of the project “Uso de teledetección para el desarrollo de herramientas para la gestión de los recursos naturales del Parque Nacional Cajas (TELEANDES)”, financed by the Research Directorate of the University of Cuenca (DIUC).

**Conflicts of Interest:** The authors declare that they have not known competing for financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- Schmeller, D.S.; Loyau, A.; Bao, K.; Brack, W.; Chatzinotas, A.; De Vleeschouwer, F.; Friesen, J.; Gandois, L.; Hansson, S.V.; Haver, M.; et al. People, pollution and pathogens—Global change impacts in mountain freshwater ecosystems. *Sci. Total Environ.* **2018**, *622–623*, 756–763. [[CrossRef](#)]
- Albert, J.S.; Destouni, G.; Duke-Sylvester, S.M.; Magurran, A.E.; Oberdorff, T.; Reis, R.E.; Winemiller, K.O.; Ripple, W.J. Scientists’ warning to humanity on the freshwater biodiversity crisis. *Ambio* **2021**, *50*, 85–94. [[CrossRef](#)]
- Cañedo-Argüelles, M.; Hermoso, V.; Herrera-Grao, T.; Barquín, J.; Bonada, N. Freshwater conservation planning informed and validated by public participation: The Ebro catchment, Spain, as a case study. *Aquat. Conserv. Mar. Freshw. Ecosyst.* **2019**, *29*, 1253–1267. [[CrossRef](#)]
- Van Echelpoel, W.; Boets, P.; Landuyt, D.; Gobeyn, S.; Everaert, G.; Bennetsen, E.; Mouton, A.; Goethals, P.L. Species distribution models for sustainable ecosystem management. In *Developments in Environmental Modelling*; Elsevier B.V.: Amsterdam, The Netherlands, 2015; Volume 27.
- Zurell, D.; Franklin, J.; König, C.; Bouchet, P.J.; Dormann, C.F.; Elith, J.; Fandos, G.; Feng, X.; Guillera-Aroita, G.; Guisan, A. A standard protocol for reporting species distribution models. *Ecography* **2020**, *43*, 1261–1277. [[CrossRef](#)]
- Besacier Monbertrand, A.L.; Timoner, P.; Rahman, K.; Burlando, P.; Fatichi, S.; Gonseth, Y.; Moser, F.; Castella, E.; Lehmann, A. Assessing the vulnerability of aquatic macroinvertebrates to climate warming in a mountainous watershed: Supplementing presence-only data with species traits. *Water* **2019**, *11*, 636. [[CrossRef](#)]
- Mehler, K.; Burlakova, L.E.; Karatayev, A.Y.; Biesinger, Z.; Bruestle, E.; Valle-Levinson, A.; Castiglione, C.; Gorsky, D. Integrating remote sensing and species distribution modelling to predict benthic communities in a Great Lakes connecting channel. *River Res. Appl.* **2017**, *33*, 1336–1344. [[CrossRef](#)]
- Kusch, J. Interacting influences of climate factors and land cover types on the distribution of caddisflies (Trichoptera) in streams of a central European low mountain range. *Insect Conserv. Divers.* **2015**, *8*, 92–101. [[CrossRef](#)]
- Chucholl, C. Niche-based species distribution models and conservation planning for endangered freshwater crayfish in south-western Germany. *Aquat. Conserv. Mar. Freshw. Ecosyst.* **2017**, *27*, 698–705. [[CrossRef](#)]
- Azzurro, E.; Soto, S.; Garofalo, G.; Maynou, F. *Fistularia commersonii* in the Mediterranean Sea: Invasion history and distribution modeling based on presence-only records. *Biol. Invasions* **2013**, *15*, 977–990. [[CrossRef](#)]
- Rocha, J.C.; Peres, C.K.; Buzzo, J.L.; de Souza, V.; Krause, E.A.; Bispo, P.C.; Frei, F.; Costa, L.S.; Branco, C.C. Modeling the species richness and abundance of lotic macroalgae based on habitat characteristics by artificial neural networks: A potentially useful tool for stream biomonitoring programs. *J. Appl. Phycol.* **2017**, *29*, 2145–2153. [[CrossRef](#)]
- Deknock, A.; De Troyer, N.; Houbraken, M.; Dominguez-Granda, L.; Nolivos, I.; Van Echelpoel, W.; Forio, M.A.; Spanoghe, P.; Goethals, P. Distribution of agricultural pesticides in the freshwater environment of the Guayas river basin (Ecuador). *Sci. Total Environ.* **2019**, *646*, 996–1008. [[CrossRef](#)]
- Celi, J.E.; Villamarín, F. Freshwater ecosystems of mainland Ecuador: Diversity, issues and perspectives. *Acta Limnol. Bras.* **2020**, *32*, 1–8. [[CrossRef](#)]
- Gobeyn, S.; Volk, M.; Dominguez-Granda, L.; Goethals, P.L.M. Input variable selection with a simple genetic algorithm for conceptual species distribution models: A case study of river pollution in Ecuador. *Environ. Model. Softw.* **2017**, *92*, 269–316. [[CrossRef](#)]
- Brown, H. Biology Of Riffle Beetles. *Annu. Rev. Entomol.* **1987**, *32*, 253–273. [[CrossRef](#)]
- Miserendino, M.L.; Archangelsky, M. Aquatic coleoptera distribution and environmental relationships in a large Patagonian river. *Int. Rev. Hydrobiol.* **2006**, *91*, 423–437. [[CrossRef](#)]
- Dos Santos, D.A.; Molineri, C.; Reynaga, M.C.; Basualdo, C. Which index is the best to assess stream health? *Ecol. Indic.* **2011**, *11*, 582–589. [[CrossRef](#)]
- Elliott, J.M. The Ecology of Riffle Beetles (Coleoptera: Elmidae). *Freshw. Rev.* **2008**, *1*, 189–203. [[CrossRef](#)]
- Salazar, G.; Rudnick, H. Hydro power plants in Ecuador: A technical and economical analysis. In Proceedings of the 2008 IEEE Power and Energy Society General Meeting—Conversion and Delivery of Electrical Energy in the 21st Century, Pittsburgh, PA, USA, 20–24 July 2008; pp. 1–5. [[CrossRef](#)]
- Castillo, L.G.; Álvarez, M.A.; Carrillo, J.M. Numerical modeling of sedimentation and flushing at the Paute-Cardenillo Reservoir. In Proceedings of the ASCE-EWRI International Perspective on Water Resources and Environment, Quito, Ecuador, 8–10 January 2014; pp. 2–11.
- Sotomayor, G.; Henrietta, H.; Vázquez, R.F.; Goethals, P.L.M. Multivariate-statistics based selection of a benthic macroinvertebrate index for assessing water quality in the Paute river basin (Ecuador). *Ecol. Indic.* **2020**, *111*, 106037. [[CrossRef](#)]

22. González-Córdoba, M.; Del Carmen Zúñiga, M.; Manzo, V. The Elmidae family (Insecta: Coleoptera: Byrrhoidea) in Colombia: Taxonomic richness and distribution. *Rev. la Acad. Colomb. Ciencias Exactas Fis. y Nat.* **2020**, *44*, 522–553. [CrossRef]
23. Legendre, P.; Legendre, L. *Numerical Ecology*; Elsevier Science B.V.: Amsterdam, The Netherlands, 2012.
24. Domisch, S.; Jähnig, S.C.; Simaika, J.P.; Kuemmerlen, M.; Stoll, S. Application of species distribution models in stream ecosystems: The challenges of spatial and temporal scale, environmental predictors and species occurrence data. *Fundam. Appl. Limnol.* **2015**, *186*, 45–61. [CrossRef]
25. Jaynes, E.T. Information Theory and Statistical Mechanics. II. *Phys. Rev.* **1957**, *108*, 171–190. [CrossRef]
26. Jaynes, E.T. Information theory and statistical mechanics. *Phys. Rev.* **1957**, *106*, 620–630. [CrossRef]
27. Phillips, S.J.; Dudík, M.; Schapire, R.E. A Maximum Entropy Approach to Species Distribution Modeling. In Proceedings of the Twenty-First International Conference on Machine Learning, Banff, AB, Canada, 4–8 July 2004. [CrossRef]
28. Johnson, R.A.; Chawla, N.V.; Hellmann, J.J. Species distribution modeling and prediction: A class imbalance problem. In Proceedings of the 2012 Conference on Intelligent Data Understanding, CIDU 2012, Boulder, CO, USA, 24–26 October 2012; pp. 9–16. [CrossRef]
29. Mi, C.; Huettmann, F.; Guo, Y.; Han, X.; Wen, L. Why choose Random Forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. *PeerJ* **2017**, *5*, e2849. [CrossRef] [PubMed]
30. Fukuda, S.; De Baets, B.; Waegeman, W.; Verwaeren, J.; Mouton, A.M. Habitat prediction and knowledge extraction for spawning European grayling (*Thymallus thymallus* L.) using a broad range of species distribution models. *Environ. Model. Softw.* **2013**, *47*, 1–6. [CrossRef]
31. Piri Sahragard, H.; Ajourlo, M.; Karami, P. Modeling habitat suitability of range plant species using random forest method in arid mountainous rangelands. *J. Mt. Sci.* **2018**, *15*, 2159–2171. [CrossRef]
32. Cha, Y.; Shin, J.; Go, B.; Lee, D.S.; Kim, Y.; Kim, T.; Park, Y.S. An interpretable machine learning method for supporting ecosystem management: Application to species distribution models of freshwater macroinvertebrates. *J. Environ. Manag.* **2021**, *291*, 112719. [CrossRef]
33. Celleri, R.; Willems, P.; Buytaert, W.; Feyen, J. Space–time rainfall variability in the Paute Basin, Ecuadorian Andes. *Hydrol. Process. Int. J.* **2007**, *21*, 3316–3327. [CrossRef]
34. Sotomayor, G. Evaluación de la Calidad de las Aguas Superficiales Mediante Técnicas de Estadística Multivariante: Un Estudio de Caso en la Cuenca del Río Paute, al Sur de Ecuador. Master’s Thesis, Universidad Nacional de La Plata, La Plata, Argentina, 2016.
35. Sotomayor, G.; Hampel, H.; Vázquez, R.F. Water quality assessment with emphasis in parameter optimisation using pattern recognition methods and genetic algorithm. *Water Res.* **2018**, *130*, 353–362. [CrossRef]
36. Jacobsen, D.; Schultz, R.; Encalada, A. Structure and diversity of stream invertebrate assemblages: The influence of temperature with altitude and latitude. *Freshw. Biol.* **1997**, *38*, 247–261. [CrossRef]
37. Segura, M.O.; Da Silva Dos Passos, M.I.; Fonseca-Gessner, A.A.; Froehlich, C.G. Elmidae curtis, 1830 (coleoptera, polyphaga, byrrhoidea) of the neotropical region. *Zootaxa* **2013**, *3731*, 1–57. [CrossRef]
38. Curiel, J.; Morrone, J.J. Association of larvae and adults of Mexican species of Macrelmis (Coleoptera: Elmidae): A preliminary analysis using DNA sequences. *Zootaxa* **2012**, *3361*, 56–62. [CrossRef]
39. Cao, Y.; DeWalt, R.E.; Robinson, J.L.; Tweddale, T.; Hinz, L.; Pessino, M. Using Maxent to model the historic distributions of stonefly species in Illinois streams: The effects of regularization and threshold selections. *Ecol. Modell.* **2013**, *259*, 30–39. [CrossRef]
40. Shen, C.; Shi, Y.; Fan, K.; He, J.S.; Adams, J.M.; Ge, Y.; Chu, H. Soil pH dominates elevational diversity pattern for bacteria in high elevation alkaline soils on the Tibetan Plateau. *FEMS Microbiol. Ecol.* **2019**, *95*, fiz003. [CrossRef] [PubMed]
41. Warren, D.L.; Matzke, N.J.; Cardillo, M.; Baumgartner, J.B.; Beaumont, L.J.; Turelli, M.; Glor, R.E.; Huron, N.A.; Simões, M.; Iglesias, T.L.; et al. ENMTools 1.0: An R package for comparative ecological biogeography. *Ecography* **2021**, *44*, 504–511. [CrossRef]
42. Corral, L.R.; Montiel Olea, C.E. What Drives Take-up in Land Regularization: Ecuador’s Rural Land Regularization and Administration Program, Sigtierras. *J. Econ. Race Policy* **2020**, *3*, 60–75. [CrossRef]
43. McCoy, J.; Johnston, K.; Kopp, S.; Borup, B.; Willison, J.; Payne, B. Using ArcGIS™ Spatial Analyst, GIS by ESRI. Redlands, California: Environmental Systems Research Institute Inc. 2002. Available online: [http://downloads.esri.com/support/documentation/ao\\_/776Using\\_Spatial\\_Analyst.pdf](http://downloads.esri.com/support/documentation/ao_/776Using_Spatial_Analyst.pdf) (accessed on 7 February 2022).
44. Arif, F.; Akbar, M. Resampling air borne sensed data using bilinear interpolation algorithm. In Proceedings of the IEEE International Conference on Mechatronics, 2005, ICM ’05, Taipei, Taiwan, 10–12 July 2005; Volume 2005, pp. 62–65.
45. Bajjali, W. *ArcGIS for Environmental and Water Issues*; Springer Textbooks in Earth Sciences, Geography and Environment; Springer: Berlin/Heidelberg, Germany, 2018. [CrossRef]
46. Tarboton, D.G.; Bras, R.L.; Rodriguez-Iturbe, I. On the extraction of channel networks from digital elevation data. *Hydrol. Process.* **1991**, *5*, 81–100. [CrossRef]
47. Vázquez, R.F.; Feyen, J. Assessment of the effects of DEM gridding on the predictions of basin runoff using MIKE SHE and a modelling resolution of 600 m. *J. Hydrol.* **2007**, *334*, 73–87. [CrossRef]
48. Li, Y.; Lei, N.; Xiong, Y. Research on Watershed Extraction Method Based on GIS. *IOP Conf. Ser. Earth Environ. Sci.* **2019**, *300*, 022168. [CrossRef]
49. Kuemmerlen, M.; Schmalz, B.; Guse, B.; Cai, Q.; Fohrer, N.; Jähnig, S.C. Integrating catchment properties in small scale species distribution models of stream macroinvertebrates. *Ecol. Modell.* **2014**, *277*, 77–86. [CrossRef]

50. Jacobsen, D. Altitudinal changes in diversity of macroinvertebrates from small streams in the Ecuadorian Andes. *Arch. Hydrobiol.* **2003**, *158*, 145–167. [[CrossRef](#)]
51. Jacobsen, D. Contrasting patterns in local and zonal family richness of stream invertebrates along an Andean altitudinal gradient. *Freshw. Biol.* **2004**, *49*, 1293–1305. [[CrossRef](#)]
52. Braun, B.M.; Salvarrey, A.V.B.; Kotzian, C.B.; Spies, M.R.; Pires, M.M. Diversity and distribution of riffle beetle assemblages (Coleoptera, Elmidae) in montane rivers of Southern Brazil. *Biota Neotrop.* **2014**, *14*, 2. [[CrossRef](#)]
53. Harrison, E.T.; Norris, R.; Wilkinson, S.N. The impact of fine sediment accumulation on benthic macroinvertebrates: Implications for river management. In Proceedings of the 5th Australian Stream Management Conference: Australian Rivers: Making a Difference, Albury, NSW, Australia, 21–25 May 2007; pp. 139–144.
54. Miserendino, M.L. Macroinvertebrate assemblages in Andean Patagonian rivers and streams: Environmental relationships. *Hydrobiologia* **2001**, *444*, 147–158. [[CrossRef](#)]
55. Roberts, D.W. Ordination on the basis of fuzzy set theory. *Vegetatio* **1986**, *66*, 123–131. [[CrossRef](#)]
56. Eichenberg, D.; Pietsch, K.; Meister, C.; Ding, W.; Yu, M.; Wirth, C. The effect of microclimate on wood decay is indirectly altered by tree species diversity in a litterbag study. *J. Plant Ecol.* **2017**, *10*, 170–178. [[CrossRef](#)]
57. Vannucchi, P.E.; López-Rodríguez, M.J.; Tierno de Figueroa, J.M.; Gaino, E. Structure and dynamics of a benthic trophic web in a Mediterranean seasonal stream. *J. Limnol.* **2013**, *72*, 606–615. [[CrossRef](#)]
58. Dilts, T.; Yang, J. *Stream Gradient and Sinuosity Toolbox for ArcGIS 10.1*; University of Nevada: Las Vegas, NV, USA, 2015.
59. Ferreira, W.R.; Ligeiro, R.; Macedo, D.R.; Hughes, R.M.; Kaufmann, P.R.; Oliveira, L.G.; Callisto, M. Importance of environmental factors for the richness and distribution of benthic macroinvertebrates in tropical headwater streams. *Freshw. Sci.* **2014**, *33*, 860–871. [[CrossRef](#)]
60. Braun, B.M.; Kotzian, C.B.; Gonçalves, A.S.; Pires, M.M. Potential distribution of riffle beetles (Coleoptera: Elmidae) in southern Brazil. *Austral Entomol.* **2018**, *58*, 646–665. [[CrossRef](#)]
61. Smith, J.V. *Colloquium on Geology, Mineralogy, and Human Welfare*; National Academies Press: Washington, DC, USA, 1999; Volume 96.
62. Wolmarans, C.T.; Kemp, M.; de Kock, K.N.; Wepener, V. The possible association between selected sediment characteristics and the occurrence of benthic macroinvertebrates in a minimally affected river in South Africa. *Chem. Ecol.* **2017**, *33*, 18–33. [[CrossRef](#)]
63. Battle, J.; Golladay, S.W. Water quality and macroinvertebrate assemblages in three types of seasonally inundated limesink wetlands in southwest Georgia. *J. Freshw. Ecol.* **2001**, *16*, 189–207. [[CrossRef](#)]
64. Endries, M. *Aquatic Species Mapping in North Carolina Using Maxent*; US Fish and Wildlife Service, Ecological Services Field Office: Lakewood, CO, USA, 2011.
65. Tchoukanski, I. Create Non-Overlapping Buffers with Attributes. 2021. Available online: [https://www.ian-ko.com/ET\\_SolutionCenter/gw\\_create\\_non\\_overlapping\\_buffers.htm](https://www.ian-ko.com/ET_SolutionCenter/gw_create_non_overlapping_buffers.htm) (accessed on 16 May 2022).
66. Forio, M.A.; Burdon, F.J.; De Troyer, N.; Lock, K.; Witing, F.; Baert, L.; De Saeyer, N.; Rîşnoveanu, G.; Popescu, C.; Kupilas, B. A Bayesian Belief Network learning tool integrates multi-scale effects of riparian buffers on stream invertebrates. *Sci. Total Environ.* **2022**, *810*, 152146. [[CrossRef](#)] [[PubMed](#)]
67. Beschta, R.L. Riparian shade and stream temperature: An alternative perspective. *Rangelands* **1997**, *19*, 25–28.
68. Amatulli, G.; McInerney, D.; Sethi, T.; Strobl, P.; Domisch, S. Geomorpho0m, empirical evaluation and accuracy assessment of global high-resolution geomorphometric layers. *Sci. Data* **2020**, *7*, 162. [[CrossRef](#)] [[PubMed](#)]
69. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
70. Valavi, R.; Elith, J.; Lahoz-Monfort, J.J.; Guillera-Arroita, G. Modelling species presence-only data with random forests. *Ecography* **2021**, *44*, 1731–1742. [[CrossRef](#)]
71. Thuiller, W.; Georges, D.; Gueguen, M.; Engler, R.; Breiner, F. *biomod2: Ensemble Platform for Species Distribution Modeling. R Package Version 3.5.1*; R Team: Vienna, Austria, 2021.
72. Fox, E.W.; Hill, R.A.; Leibowitz, S.G.; Olsen, A.R.; Thornbrugh, D.J.; Weber, M.H. Assessing the accuracy and stability of variable selection methods for random forest modeling in ecology. *Environ. Monit. Assess.* **2017**, *189*, 316. [[CrossRef](#)]
73. Probst, P.; Wright, M.N.; Boulesteix, A.L. Hyperparameters and tuning strategies for random forest. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2019**, *9*, e1301. [[CrossRef](#)]
74. Strobl, C.; Boulesteix, A.L.; Zeileis, A.; Hothorn, T. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinform.* **2007**, *8*, 25. [[CrossRef](#)]
75. Strobl, C.; Boulesteix, A.L.; Kneib, T.; Augustin, T.; Zeileis, A. Conditional variable importance for random forests. *BMC Bioinform.* **2008**, *9*, 307. [[CrossRef](#)]
76. Xiong, Z.; Cui, Y.; Liu, Z.; Zhao, Y.; Hu, M.; Hu, J. Evaluating explorative prediction power of machine learning algorithms for materials discovery using k-fold forward cross-validation. *Comput. Mater. Sci.* **2020**, *171*, 109203. [[CrossRef](#)]
77. Grimm, K.J.; Mazza, G.L.; Davoudzadeh, P. Model Selection in Finite Mixture Models: A k-Fold Cross-Validation Approach. *Struct. Equ. Model.* **2017**, *24*, 246–256. [[CrossRef](#)]
78. Pal, K.; Patel, B.V. Data Classification with k-fold Cross Validation and Holdout Accuracy Estimation Methods with 5 Different Machine Learning Techniques. In Proceedings of the 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 11–13 March 2020; pp. 83–87. [[CrossRef](#)]

79. Khalilia, M.; Chakraborty, S.; Popescu, M. Predicting disease risks from highly imbalanced data using random forest. *BMC Med. Inform. Decis. Mak.* **2011**, *11*, 51. [[CrossRef](#)] [[PubMed](#)]
80. Brown, I.; Mues, C. An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Syst. Appl.* **2012**, *39*, 3446–3453. [[CrossRef](#)]
81. Larras, F.; Coulaud, R.; Gautreau, E.; Billoir, E.; Rosebery, J.; Usseglio-Polatera, P. Assessing anthropogenic pressures on streams: A random forest approach based on benthic diatom communities. *Sci. Total Environ.* **2017**, *586*, 1101–1112. [[CrossRef](#)]
82. Boulesteix, A.L.; Janitza, S.; Kruppa, J.; König, I.R. Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2012**, *2*, 493–507. [[CrossRef](#)]
83. Rebala, G.; Ravi, A.; Churiwala, S. Random forests. In *An Introduction to Machine Learning*; Springer Nature: Cham, Switzerland, 2019; pp. 77–94. [[CrossRef](#)]
84. Ballabio, D.; Grisoni, F.; Todeschini, R. Multivariate comparison of classification performance measures. *Chemom. Intell. Lab. Syst.* **2018**, *174*, 33–44. [[CrossRef](#)]
85. Sokolova, M.; Japkowicz, N.; Szpakowicz, S. Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. In *AI 2006: Advances in Artificial Intelligence*; Carbonell, J.G., Siekmann, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2006; pp. 1015–1021. [[CrossRef](#)]
86. Sotomayor, G.; Hampel, H.; Vázquez, R.F.; Forio, M.A.E.; Goethals, P.L.M. Implications of macroinvertebrate taxonomic resolution for freshwater assessments using functional traits: The Paute River Basin (Ecuador) case. *Divers. Distrib.* **2021**, *28*, 1735–1747. [[CrossRef](#)]
87. Shapiro, S.S.; Wilk, M.B. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika Trust* **1965**, *52*, 591–611. [[CrossRef](#)]
88. Anderson, T.; Finn, J. *The New Statistical Analysis of Data*; Springer: New York, NY, USA, 1996.
89. Helsel, D.R.; Hirsch, R.M.; Ryberg, K.R.; Archfield, S.A.; Gilroy, E. Statistical Methods in Water Resources. In *Book 4, Hydrologic Analysis and Interpretation*; U.S. Geological Survey: Washington, DC, USA, 2020; p. 458.
90. Hosmer, D.W., Jr.; Lemeshow, S.; Sturdivant, R.X. *Applied Logistic Regression*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2013; Volume 38.
91. Thuiller, W.; Lafourcade, B.; Engler, R.; Araújo, M.B. BIOMOD—A platform for ensemble forecasting of species distributions. *Ecography* **2009**, *32*, 369–373. [[CrossRef](#)]
92. Jain, A.K. Data clustering: 50 years beyond K-means. *Pattern Recognit. Lett.* **2010**, *31*, 651–666. [[CrossRef](#)]
93. Hammer, Ø. *PAST: Paleontological Statistics Version 4.03—Reference Manual*; Natural History Museum University of Oslo: Oslo, Norway, 2020; pp. 1–283.
94. Wang, K.; Wang, B.; Peng, L. CVAP: Validation for Cluster Analyses. *Data Sci. J.* **2009**, *8*, 88–93. [[CrossRef](#)]
95. Thalamuthu, A.; Mukhopadhyay, I.; Zheng, X.; Tseng, G. Evaluation and comparison of gene clustering methods in microarray analysis. *Bioinformatics* **2006**, *22*, 2405–2412. [[CrossRef](#)] [[PubMed](#)]
96. Kaufman, L.; Rousseeuw, P.J. *Finding Groups in Data—An Introduction to Cluster Analysis*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 1990.
97. Guisan, A.; Thuiller, W.; Zimmermann, N.E. *Habitat Suitability and Distribution Models: With Applications in R*; Cambridge University Press: Cambridge, UK, 2017. [[CrossRef](#)]
98. Ministerio del Ambiente del Ecuador (MAE). *Sistema de Clasificación de Ecosistemas del Ecuador Continental*; Subsecretaría de Patrimonio Natural—Proyecto Mapa de Vegetación: Quito, Ecuador, 2013.
99. Lissovsky, A.A.; Dudov, S.V. Species-Distribution Modeling: Advantages and Limitations of Its Application. 2. *MaxEnt. Biol. Bull. Rev.* **2021**, *11*, 265–275. [[CrossRef](#)]
100. Cárdenas, R.E.; Buestán, J.; Dangles, O. Diversity and distribution models of horse flies (diptera: Tabanidae) from Ecuador. *Ann. La Soc. Entomol. Fr.* **2009**, *45*, 511–528. [[CrossRef](#)]
101. Escobar, L.E.; Romero-Alvarez, D.; Leon, R.; Lepe-Lopez, M.A.; Craft, M.E.; Borbor-Cordova, M.J.; Svenning, J.C. Declining Prevalence of Disease Vectors Under Climate Change. *Sci. Rep.* **2016**, *6*, 39150. [[CrossRef](#)]
102. Cuesta, F.; Peralvo, M.; Merino-Viteri, A.; Bustamante, M.; Baquero, F.; Freile, J.F.; Muriel, P.; Torres-Carvajal, O. Priority areas for biodiversity conservation in mainland Ecuador. *Neotrop. Biodivers.* **2017**, *3*, 93–106. [[CrossRef](#)]
103. Moya, W.; Jacome, G.; Yoo, C.K. Past, current, and future trends of red spiny lobster based on PCA with MaxEnt model in Galapagos Islands, Ecuador. *Ecol. Evol.* **2017**, *7*, 4881–4890. [[CrossRef](#)] [[PubMed](#)]
104. Yañez-Arenas, C.; Díaz-Gamboa, L.; Patrón-Rivero, C.; López-Reyes, K.; Chiappa-Carrara, X. Estimating geographic patterns of ophidism risk in Ecuador. *Neotrop. Biodivers.* **2018**, *4*, 55–61. [[CrossRef](#)]
105. Jácome, G.; Vilela, P.; Yoo, C.K. Present and future incidence of dengue fever in Ecuador nationwide and coast region scale using species distribution modeling for climate variability's effect. *Ecol. Modell.* **2019**, *400*, 60–72. [[CrossRef](#)]
106. Kübler, D. Effect of Topography on the Distribution of tree Species and Radial Diameter Growth of Potential Crop Trees in a Tropical Mountain Forest in Southern Ecuador. Ph.D. Thesis, Technische Universität München, Munich, Germany, 2020.
107. Yackulic, C.B.; Chandler, R.; Zipkin, E.F.; Royle, J.A.; Nichols, J.D.; Campbell Grant, E.H.; Veran, S. Presence-only modelling using MAXENT: When can we trust the inferences? *Methods Ecol. Evol.* **2013**, *4*, 236–243. [[CrossRef](#)]
108. Ward, G.; Hastie, T.; Barry, S.; Elith, J.; Leathwick, J.R. Presence-Only Data and the EM Algorithm. *Biometrics* **2009**, *65*, 554–563. [[CrossRef](#)]

109. Elith, J.; Phillips, S.J.; Hastie, T.; Dudík, M.; Chee, Y.E.; Yates, C.J. A statistical explanation of MaxEnt for ecologists. *Divers. Distrib.* **2011**, *17*, 43–57. [[CrossRef](#)]
110. Marmion, M.; Parviainen, M.; Luoto, M.; Heikkinen, R.K.; Thuiller, W. Evaluation of consensus methods in predictive species distribution modelling. *Divers. Distrib.* **2009**, *15*, 59–69. [[CrossRef](#)]
111. Drew, C.A.; Wiersma, Y.F.; Huettmann, F. Predictive species and habitat modeling in landscape ecology: Concepts and applications. In *Predictive Species and Habitat Modeling in Landscape Ecology: Concepts and Applications*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 1–313. [[CrossRef](#)]
112. De Luis, M.; Álvarez-Jiménez, J.; Rejos, F.J.; Bartolomé, C. Using species distribution models to locate the potential cradles of the allopolyploid *Gypsophila bermejoi* G. López (Caryophyllaceae). *PLoS ONE* **2020**, *15*, e0232736. [[CrossRef](#)] [[PubMed](#)]
113. García-Criado, F.; Fernández-Alaéz, M. Hydraenidae and Elmidae assemblages (Coleoptera) from a Spanish river basin: Good indicators of coal mining pollution? *Arch. Hydrobiol.* **2001**, *150*, 641–660. [[CrossRef](#)]
114. Von Ellenrieder, N. Composition and structure of aquatic insect assemblages of Yungas mountain cloud forest streams in NW Argentina. *Rev. La Soc. Entomol. Argent.* **2007**, *66*, 57–76.
115. Albanesi, S.A.; Cristobal, L.; Manzo, V.; Nieto, C. Dataset of the baetidae (Ephemeroptera) and elmidae (coleoptera) families from the yungas of Argentina. *Rev. La Soc. Entomol. Argent.* **2020**, *79*, 17–23. [[CrossRef](#)]
116. García-Ríos, R.F.; Moi, D.A.; Peláez, O.E. Efectos del gradiente altitudinal sobre las comunidades de macroinvertebrados bentónicos en dos períodos hidrológicos en un río altoandino neotropical. *Ecol. Austral* **2020**, *30*, 033–044. [[CrossRef](#)]
117. Monte, C.; Mascagni, A. Review of the Elmidae of Ecuador with the description of ten new species (Coleoptera: Elmidae). *Zootaxa* **2012**, *38*, 1–38. [[CrossRef](#)]
118. Linský, M.; Čiamporová-Zaťovičová, Z.; Čiampor, F. Four new species of hexanchorus sharp from Ecuador (Coleoptera, elmidae) with dna barcoding and notes on the distribution of the genus. *Zookeys* **2019**, *2019*, 85–109. [[CrossRef](#)]
119. Čiampor, F.; Kodada, J.; Božánová, J.; Čiamporová-Zaťovičová, Z. Disersus otongachi a new species of Larainae riffle beetles from Ecuador (Coleoptera: Elmidae). *Zootaxa* **2021**, *4963*, 193–199. [[CrossRef](#)]
120. Kattel, D.B.; Yao, T.; Yang, W.; Gao, Y.; Tian, L. Comparison of temperature lapse rates from the northern to the southern slopes of the Himalayas. *Int. J. Climatol.* **2015**, *35*, 4431–4443. [[CrossRef](#)]
121. Acosta, R. Estudio de la Cuenca Altoandina del Río Cañete (Perú): Distribución Altitudinal de la Comunidad de Macroinvertebrados Bentónicos y Caracterización Hidroquímica de sus Cabeceras Cársticas. Doctoral Thesis, Universitat de Barcelona, Barcelona, Spain, 2009.
122. Aguilera Giraldo, I.A.; Vásquez-Ramos, J.M. Distribución espacial y temporal de Elmidae (Insecta: Coleoptera) y su relación con los parámetros fisicoquímicos en el río Ocoa, Meta, Colombia. *Rev. La Acad. Colomb. Cienc. Exactas Físicas Nat.* **2019**, *43*, 108. [[CrossRef](#)]
123. Burk, R.A.; Kennedy, J.H. Invertebrate communities of groundwater-dependent refugia with varying hydrology and riparian cover during a suprasedasonal drought. *J. Freshw. Ecol.* **2013**, *28*, 251–270. [[CrossRef](#)]
124. Pacheco, G.S.M.; Pellegrini, T.G.; Lopes Ferreira, R. Cave lithology influencing EPT (Ephemeroptera, Plecoptera, Trichoptera) assemblages and habitat structure in south-eastern Brazil. *Mar. Freshw. Res.* **2021**, *72*, 1546–1552. [[CrossRef](#)]
125. Mora, D.E.; Willems, P. Decadal oscillations in rainfall and air temperature in the Paute River Basin-Southern Andes of Ecuador. *Theor. Appl. Climatol.* **2012**, *108*, 267–282. [[CrossRef](#)]
126. Ríos-Touma, B.; Encalada, A.C.; Prat Fornells, N. Macroinvertebrate assemblages of an Andean high-altitude tropical stream: The importance of season and flow. *Int. Rev. Hydrobiol.* **2011**, *96*, 667–685. [[CrossRef](#)]
127. Spangler, P.J.; Santiago-Fragoso, S. *The Aquatic Beetle Subfamily Larainae (Coleoptera: Elmidae) in Mexico, Central America, and the West Indies*; Smithsonian Contributions to Zoology; Smithsonian: Washington, DC, USA, 1992; pp. 1–74. [[CrossRef](#)]
128. Spangler, P.J. Two new species of the aquatic beetle genus *Macrelmis* Motschulsky from Venezuela (Coleoptera: Elmidae: Elminae). *Insecta Mundi* **1997**, *11*, 1–8.
129. Fernandes, A.S. *Taxonomia de Elmidae (Insecta, Coleoptera) do Município de Presidente Figueiredo, Amazonas, Brasil*; Instituto Nacional de Pesquisas da Amazônia: Manaus, Brazil, 2010; p. 140.
130. Braun, B.M.; Pires, M.M.; Stenert, C.; Maltchik, L.; Kotzian, C.B. Effects of riparian vegetation width and substrate type on riffle beetle community structure. *Entomol. Sci.* **2018**, *21*, 66–75. [[CrossRef](#)]
131. Brown, A.G.; Rhodes, E.J.; Davis, S.; Zhang, Y.; Pears, B.; Whitehouse, N.J.; Bradley, C.; Bennett, J.; Schwenninger, J.L.; Firth, A.; et al. Late Quaternary evolution of a lowland anastomosing river system: Geological-topographic inheritance, non-uniformity and implications for biodiversity and management. *Quat. Sci. Rev.* **2021**, *260*, 106929. [[CrossRef](#)]
132. Da Ros, G. *La Contaminación de Aguas en Ecuador: Una Aproximación Económica*; Instituto de Investigaciones Económicas, Pontificia Universidad Católica del Ecuador: Quito, Ecuador, 1995.
133. Pauta Calle, G.; Chang Gómez, J. Índices de calidad del agua de fuentes superficiales y aspectos toxicológicos, evaluación del Río Burgay. *Maskana* **2014**, *5*, 165–176.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.