Ecological Indicators 108 (2020) 105769

Contents lists available at ScienceDirect

Ecological Indicators

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

Biological water quality in tropical rivers during dry and rainy seasons: A model-based analysis

Rubén Jerves-Cobo[a,b,c,*], Marie Anne Eurie Forio[a], Koen Lock[a], Jana Van Butsel[a], Guillermina Pauta[d], Félipe Cisneros[c], Ingmar Nopens[b], Peter L.M. Goethals[a]

[a] Laboratory of Environmental Toxicology and Aquatic Ecology, Department of Animal Science and Aquatic Ecology, Ghent University, Coupure Links 653, 9000 Ghent, Belgium
[b] BIOMATH, Department of Data Analysis and Mathematical Modelling, Ghent University, Coupure Links 653, 9000 Ghent, Belgium
[c] PROMAS, Programa para el manejo del agua y del suelo, Universidad de Cuenca, Av. 12 de abril s/n y Agustín Cueva, 010103 Cuenca, Ecuador
[d] Facultad de Ingeniería, Universidad de Cuenca, Av. 12 de abril s/n y Agustín Cueva, 010103 Cuenca, Ecuador

ARTICLE INFO

Keywords:
Cuenca River basin
Ecological modelling
Generalized Linear Models
Pollution
Ecological assessment
Biological water quality
Andes

ABSTRACT

Recent studies on water quality in tropical rivers indicate substantial differences between seasons. However, investigations on the needs and added value of season-specific models are lacking. Thus, this paper aims to determine the accuracy and relevance of season-specific and season-overarching models to predict biological water quality. Additionally, we investigated the variation of prediction accuracy using sub-datasets from different parts of the Cuenca River basin. This study was accomplished in the rivers that pass through the urban and suburban areas of the city of Cuenca, which is located in the southern Andes of Ecuador. The Andean Biotic Index (ABI) was used as an indicator of biological water quality. Subsequently, models were developed to predict the ABI, with physicochemical and morphological variables as predictors, which were collected in 43 sites during both the dry and the rainy seasons. The predictions were obtained using three kinds of generalized linear models (GLMs): Gaussian, Gamma and Inverse Gaussian. The season-specific models were more accurate than the season-overarching models. Similarly, the predictions of the biological water quality in sites sampled in the urban area were more accurate than the forecasts performed in reference sites. The major variables predicting the ABI during the dry season were five-day biological oxygen demand (BOD_5), ammonium and orthophosphate, while dissolved oxygen (DO), oxygen saturation (OS), nitrate, total solids proved to be important during the rainy season. The results of this research emphasize the importance of developing season-specific models and the implementation of different key actions for river restoration during both the dry and rainy seasons. The accuracy and the replication of these models could be improved and checked with more data taken from new sampling events. The modelling approach developed in this study can be applied to similar basins in the tropics and reveals that environmental investments need to count on monitoring strategies and data and analyses of the biological water quality variation in dry and rainy seasons, within the context of sustainable development.

1. Introduction

Efforts to restore the good ecological status of water bodies throughout the world have been conducted in urban and suburban areas. These efforts have been developed using various measures such as the implementation of wastewater treatment plants and the application of temporary storage for sewer overflows. However, this approach only focused on the control of the point source releases without considering the impact of diffuse sources such as livestock and agricultural into the river (Erbe et al., 2002). Moreover, the impact on the ecology of running waters was not taken into account before the implementation of these measures in most cases (Mouton et al., 2009). For this reason, the identification of the main drivers that affect the ecological status of rivers and what impact produced by these drivers is paramount with regard to the implementation of river restoration measures (Holguin-Gonzalez et al., 2013a). Hence, restoration options must be evaluated using scenario analyses before their application to determine their effectiveness in improving river ecology (Palmer et al., 2005).

The assessment of ecological water quality on freshwater bodies has been frequently performed through the monitoring of benthic macroinvertebrates (Jacobsen, 1998; De Pauw et al., 2006; Gabriels et al.,

* Corresponding author.
E-mail address: rubenf.jervesc@ucuenca.edu.ec (R. Jerves-Cobo).

https://doi.org/10.1016/j.ecolind.2019.105769
Received 14 May 2019; Received in revised form 19 September 2019; Accepted 23 September 2019
1470-160X/ © 2019 Elsevier Ltd. All rights reserved.

2010). These bioindicators respond to both hydromorphological variation and physicochemical changes as a result of human and natural stressors in streams, rivers and their surrounding areas (Cairns and Pratt, 1993; Thorne and Williams, 1997). Furthermore, macroinvertebrates continuously respond to environmental variations over a long period, as opposed to physicochemical samples that reflect the water quality at a particular point in time (Džeroski et al., 2000; De Pauw et al., 2006). Thus, the Biological Monitoring Working Party (BMWP), an index that assesses the surface water quality in England, was developed based on the presence of taxa that are sensitive to organic pollution in water bodies (Armitage et al., 1983). This index has been adapted to tropical countries, such as Brazil (Junqueira and Campos, 1998), Thailand (Mustow, 2002) and Colombia (Roldán Pérez, 2003). The BMWP-Col – an adaptation to the BMWP – has been used in different regions of Colombia and Ecuador (Roldán Pérez, 2003; Álvarez, 2005; Damanik-Ambarita et al., 2016). The BMWP-Col was further updated and adapted to the high altitude regions (i.e. the Andes of Ecuador and Peru), and was designated as the Andean Biotic Index (ABI) (Ríos-Touma et al., 2014).

Predicting the ecological water quality in rivers has frequently been implemented through ecological modelling, using linear relationships between the biotic and abiotic information despite the non-linear behavior of the ecosystem (Džeroski et al., 2000; Forio et al., 2016; Forio et al., 2018). By selecting the most appropriate ones, these models have been recognized as powerful tools of predictive ecology (Holguin-Gonzalez et al., 2013b). They are instruments that can be used in the restoration and conservation of aquatic ecosystems (Mouton et al., 2009). However, biological models in tropical rivers often have been based on season-specific data collections such as in the Philippines (Forio et al., 2018), in Ecuador's coastal region (Damanik-Ambarita et al., 2016) and in Bolivia (Moya et al., 2011). Furthermore, biological studies in tropical countries indicate that there can be substantial differences between seasons, leading to dissimilarities in environmental conditions between the dry and rainy seasons (Arunachalam et al., 1991; Jacobsen and Encalada, 1998; Beauchard et al., 2003; Mereta et al., 2012). However, a lack of model-based comparisons in tropical countries needs to be addressed to answer questions such as "How well do models perform in one season compared to another season?", "What is the benefit of season-overarching models?" and "Is the use of a selective combination of season-specific models more effective than a season-overarching model?".

Ecological models have been developed by using different techniques such as artificial neural networks (ANNs), Bayesian belief networks (BBNs), classification and regression trees (CTs and RTs), genetic algorithms (GAs), generalized linear models (GLMs), and support-vector machines (SVMs) (Goethals, 2005). In particular, the GLMs often have been used to identify variables that could affect the ecology of water bodies (Heino et al., 2007; Wilson et al., 2008; Domisch et al., 2011). This technique is mainly used to predict outcomes with the use of continuous variables (McCullagh, 1984; Zuur et al., 2009). Similarly, it has been demonstrated that GLMs are very efficient in providing the main variables that contribute to obtaining accurate predictions using small datasets (Vayssières et al., 2000).

This research was carried out in the framework of the VLIR-UOS IUC Programme – University of Cuenca and the VLIR-UOS Ecuador Biodiversity Network Project. In this study, new developmental and methodological steps for environmental indication were constructed. For that, we aim to build and to evaluate season-specific and season-overarching models to predict the biological water quality based on the Andean Biotic Index (ABI) in the Cuenca River, a tropical river located in the southern Andes of Ecuador. Furthermore, we analyzed the predictability of sites throughout seasons and between the dry and the rainy season based on simulations. We also included an analysis whether the models can be effectively applied or extrapolated to the main rivers or to their tributaries, as well as be tested to urbanized or to natural sites. In order to link environmental variables with the biological index ABI, generalized linear models (GLMs) were constructed. The framework and objectives of the current study are presented in Fig. 1. The findings of this study can be applied to scenario analyses for season-specific river restoration based on the key disturbance variables obtained from the modeled indices. As indicated in the framework, the cause deduction (phase two in Fig. 1) is not addressed in this study because this requires additional analysis and simulations, either in the actual rivers or in river water quality models.

2. Materials and methods

2.1. Study area

The study area corresponds to the urban and suburban area of the Cuenca River basin, which is situated in the southern Province of Azuay in the Andes of Ecuador. The Cuenca River is an Andean mountain stream that is part of the Paute Upper Basin. The latter is one of the tributaries of the Santiago River, which is an affluent of the Amazon River. The Cuenca River is located downstream from the city of Cuenca (Fig. 2), and its four main tributaries are the Tarqui, Yanuncay, Tomebamba and Machangara Rivers that run from upstream and through the urban area.

The study area is 223 km^2, representing 13% of the Cuenca River basin, of which 16% is an urban area (i.e. the city of Cuenca), 73% is a mosaic between pastures and crops, 6% is forest and about 5% consists of lakes, moorland and bare land. The urban area had approximately 382,000 inhabitants in 2017 (INEC, 2010; SENPLADES, 2016). Of the four subcatchments, only the Machangara Basin is regulated all year by the presence of two hydropower dams: the Labrado and the Chanlud, which are located upstream from the city of Cuenca. Two natural reserves are also located upstream from the Cuenca River basin: Cajas National Park and the Machangara-Tomebamba protected forest. Both are water sources for the Tomebamba, Yanuncay and Machangara Rivers.

The mean altitude of the study area is 2655 m a.s.l., which varies from 2335 in the lowest part of the Cuenca River to 2770 m a.s.l. in the Yanuncay tributary. The highest altitudes are 2545, 2695, and 2622 m a.s.l. in the Machangara, Tomebamba and Tarqui Basins, respectively. The average annual rainfall in the study area is approximately 879 mm per year, while the average annual temperature is 16.3 °C (Aereopuerto-Mariscal-Lamar, 2012). There are two seasons during the year: the rainy season, which starts from the middle of February until the beginning of July and from the second half of September until the first two weeks of November, while the rest of the year constitutes the dry season (Fig. A1). The average flow of the Cuenca River measured before its discharge into the Paute River is around 28 $m^3 \cdot s^{-1}$ (Cordero Domínguez, 2013). The 95th percentile annual flow, which generally happens during August and September – dry season –, is around 6 $m^3 \cdot s^{-1}$, while the 5th percentile annual flow which is present during the rainy season, is about 71 $m^3 \cdot s^{-1}$ (ETAPA-EP, 2018).

2.2. Data collection

The study considered 43 sampling sites located in the city of Cuenca and nearby areas. Of these sites, 27 were sampled during the dry season of July 2015 (Fig. 2A), while 35 sites were sampled during the rainy season of March 2016 (Fig. 2B). Nineteen sites were sampled during both seasons. The samples were collected during the day, from 8:00 am to 5:00 pm. These sampling sites were selected based on two criteria: macro-locations and micro-locations. Macro-locations are river stretches sampled in a basin or sub-basin, while micro-locations are placed to represent critical points like outfalls (Strobl and Robillard, 2008). Application of the macro-locations was in order to understand the state of the rivers from upstream to downstream, for which the location of the sampled sites was determined according to similar contributing tributary subareas (Sharp, 1971) in the Cuenca River basin. In order to

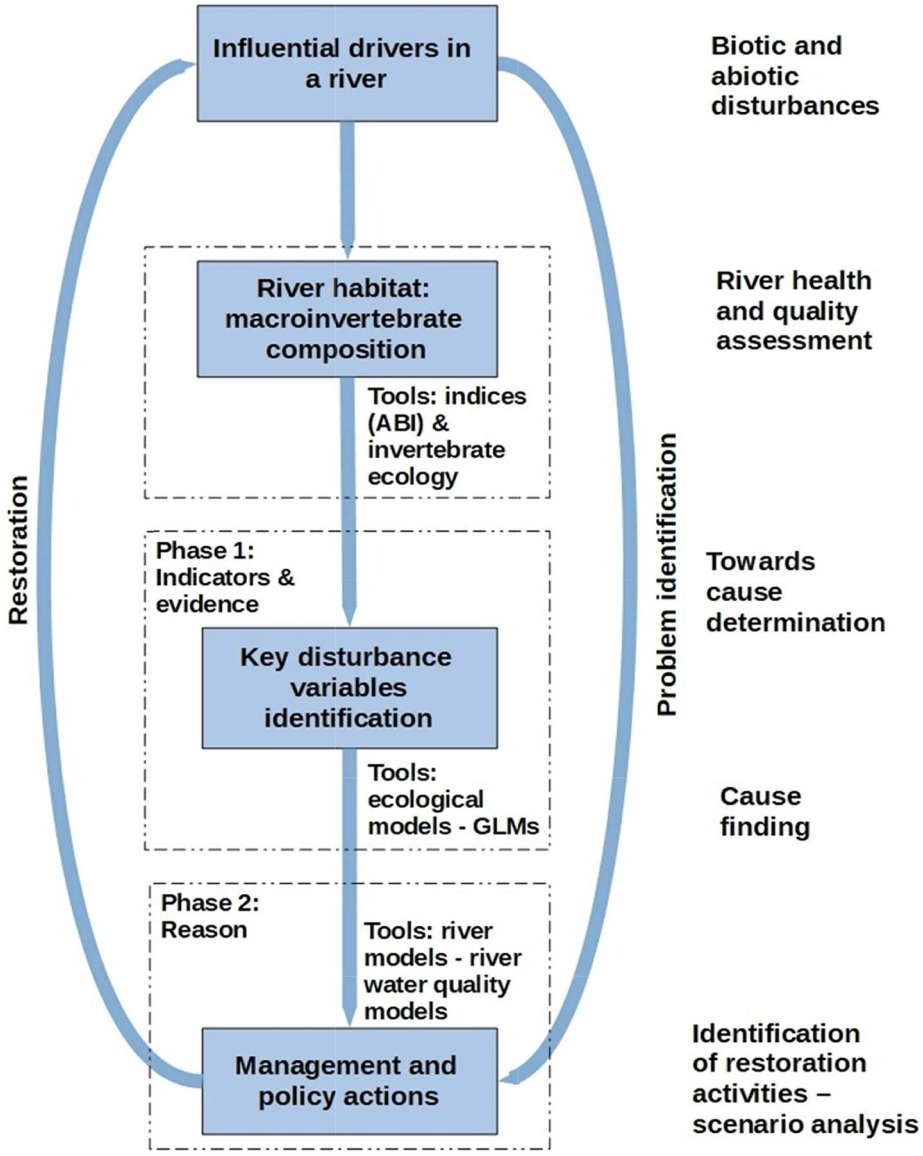

Fig. 1. The applied framework for the restoration of river water quality, of which Phase One was implemented in this study — . adapted from Forio et al. (2017)

assess local impacts and critical sites in the rivers, micro-locations were selected upstream and downstream from the outfalls (Ward, 1973; Strobl and Robillard, 2008). Thus, the first sampled sites were located prior to the start of the city's suburban area and situated upstream of the four main tributaries of the Cuenca River. The combined sewer overflow (CSOs) are in operation only during the rainy seasons; thus, additional sites were sampled providing insights into the impact of CSOs. The locations that were sampled in both seasons focus on the water quality variation between them.

At each sampling site, environmental (physical-chemical and hydromorphological) conditions were recorded and biological (macroinvertebrates) samples were collected. In particular, 28 physicochemical, hydraulic and microbiological variables that could influence the biological water quality were measured. Six of these variables were measured directly in the field with two YSI®6920-V2 (Yellow Spring, OH, USA) multi-parameter probes: water temperature, specific conductivity, dissolved oxygen (DO), turbidity, pH and chlorophyll-a. To obtain the flow in each sampling site, the river cross section was divided into various subsections, in which depth, width and mean stream velocity were measured. The latter variable was determined in 40% of the water depth in each subsection, using a water current meter (Rantz, 1982) Gurley 622A. The discharge was calculated as the sum of the subflows computed in each subsection. Table A1 presents the overview of the chemical data per season and over both seasons, which were measured in the laboratory of Sanitation at the University of Cuenca. Additionally, information was compiled on elevation, land use, river morphology, substrate characteristics, macrophytes, shading on the rivers at each site.

The samples of benthic macroinvertebrates were taken from the river and its tributaries with the kick-sampling procedure. This method is applied by shuffling the feet walking backwards against the current, while holding a standard net (a conical net with a frame size of 0.20 × 0.30 m and mesh size of 500 µm, attached to a stick) against the current for five minutes (Gabriels et al., 2010). The sampling was performed in a stretch approximately 10–20 m of length in different aquatic habitats, including bed substrates (stones, sand or mud), macrophytes (floating, submerged, and emerging) and other floating or submerged natural or artificial substrates. In addition to the hand net sampling, macroinvertebrates were also collected manually from stones, leaves and branches. Living animals were sorted in a white tray. Macroinvertebrates were sampled during the dry and rainy season to understand the seasonal effects in their assemblage's composition,

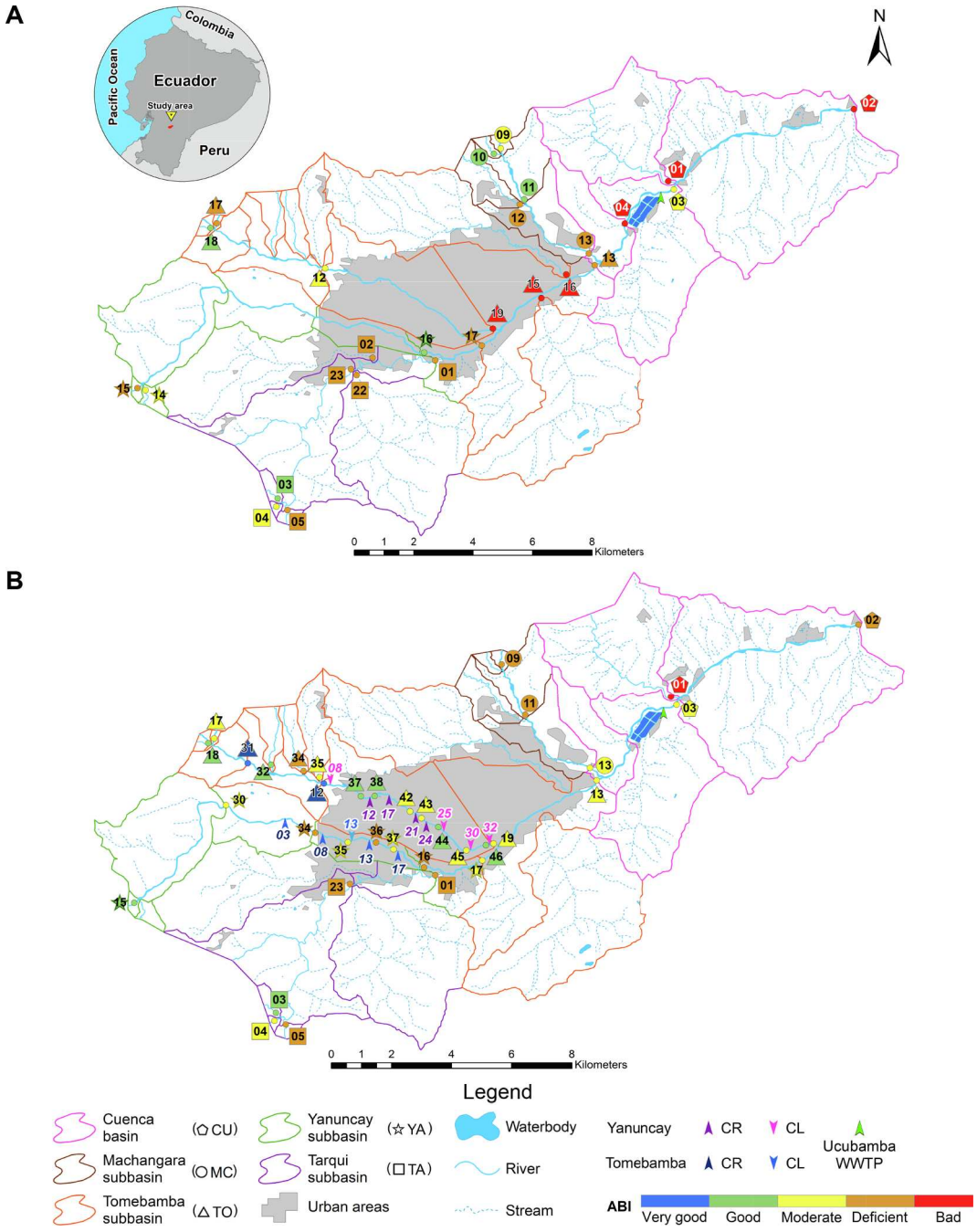

Fig. 2. Sampling sites location with their Andean Biotic Index (ABI) in both seasons: dry season, (B) rainy season. The figures (i.e. pentagon, square, triangle, star, and circle) indicate to what basin each site belongs.

produced either by the animals' life-cycle, by functional feeding group variety or by different environmental characteristics of the streams (Jacobsen and Encalada, 1998). During both seasons, 43 families of macroinvertebrates were found (Tables A2 and A3), identification was done by means of a stereomicroscope (Roldán Pérez, 1988; Álvarez, 2005; Encalada et al., 2011). For a detailed overview of the sampling campaigns and description of the locations, the authors refer to Jerves-Cobo et al. (2018a).

At each sampling site, the Andean Biotic Index (ABI) was calculated to assess the water quality. The ABI was built based on adaptations of the Biological Monitoring Working Party score (BMWP) of South America (Encalada et al., 2011; Ríos-Touma et al., 2014). This index was used since it was shown to be more suitable for the Andes in Ecuador above 2400 m a.s.l. (Jerves-Cobo et al., 2018a). The sensitivity score ranges from one for very tolerant taxa to 10 for the most sensitive families. The ABI gives a five-water quality classification in function of the sum of sensitivity scores obtained in each location: bad ($\leq$15), deficient (16–35), moderate (36–60), good (61–99), and very good ($>$ 99) (Álvarez, 2005; Zúñiga et al., 2009). Fig. 2 shows the ABI score of each sampling site in both seasons by means of a color code.

2.3. Ecological model

An ecological model was developed to determine parameters such as BOD_5, DO nitrate, nitrite or bank material that could influence the performance of the aforementioned biological index in the study area. For this, the data was fitted by a generalized linear model (GLM), as this modelling technique can deal with the non-linear behavior of the

ecosystem and has frequently been used for ecological related studies (Guisan et al., 2006; Zuur et al., 2009; Forio et al., 2018). The ABI, which was the response variable, takes values that are positive and continuous in nature. Thus, Gaussian, Gamma and Inverse Gaussian distributions are considered as the most appropriate distributions for the response variable (McCullagh, 1984; Zuur et al., 2009; Hardin and Hilbe, 2013). Furthermore, these distributions were also chosen based on the best fitting of the ABI score (response variable) that could be adjusted to either symmetric or skewed distributions. A Gaussian (Normal) distribution is applicable to a wide range of phenomena and its shape is symmetric, while the skewedness of the Gamma distribution depends on their parameters α and β_i. The Inverse Gaussian is a distribution that is skewed to the right and is frequently used for studying the diffusion process (Forbes et al., 2011).

The GLMs with Gaussian, Gamma and Inverse Gaussian distributions are represented in Eqs. (1)–(3), respectively; where $E(Y_i)$, also expressed as μ_i, is the mean of the response (or dependent) variable Y, given the regressor X_i (or independent) variable, i.e. the mean is a function of the regressor. α and β_i are parameters, referred to as the intercept and the regression coefficient, respectively. The unexplained information is captured by the residuals, which are assumed to be normal, Gamma and Inverse Gaussian distributed, respectively. The variance σ^2 in Gaussian GLM is denoted as $var(Y) = \sigma^2$, while the variance in Gamma and Inverse Gaussian GLMs are represented as $var(Y) = \mu^2/\vartheta$ and $var(Y) = \mu^3/\vartheta$, respectively, in which ϑ^{-1}denotes the shape parameter (Zuur et al., 2009; Lovison et al., 2011; Hardin and Hilbe, 2013).

Gaussian GLM:

$$E(Y_i) = \mu_i = \eta_i = \alpha + \beta_1 \times X_{1i} + \beta_2 \times X_{2i} + \cdots\cdots + \beta_n \times X_{ni} \quad (1)$$

Gamma GLM:

$$E(Y_i) = \mu_i = \frac{1}{\eta_i} = \frac{1}{\alpha + \beta_1 \times X_{1i} + \beta_2 \times X_{2i} + \cdots\cdots + \beta_n \times X_{ni}} \quad (2)$$

Inverse Gaussian GLM:

$$E(Y_i) = \mu_i = \frac{1}{\sqrt{\eta_i}} = \frac{1}{\sqrt{\alpha + \beta_1 \times X_{1i} + \beta_2 \times X_{2i} + \cdots\cdots + \beta_n \times X_{ni}}} \quad (3)$$

2.3.1. Model development and validation

To build the model, five datasets as described in Fig. 3 were fitted with Gaussian, Gamma and Inverse Gaussian GLMs. The first dataset (complete dataset) included all the sites (i.e. 43 cases) that were sampled in both seasons in the study area. The second dataset (without-outliers dataset) contained the previous dataset, but without the influence of an (outlier) data point (i.e. one site). The influence observation (data point) was obtained using two techniques: Cook's distance and the hat elements. Cook's distance is a measure of the change in the regression coefficients, which reveals the most influential cases that affect the regression equation. Whereas, the hat elements identify points that do not necessarily affect the regression coefficients. However, this technique could be useful in the identification of isolated points (Stevens, 1984). The third dataset (main-rivers dataset) comprised the locations sampled in both seasons solely in the main four rivers and not in their tributaries (i.e. 29 cases). The fourth dataset (dry-season dataset) consisted of all the sites sampled during the dry season (i.e. 26 cases) and finally, the last dataset (rainy-season dataset) involved all locations taken during the rainy season (i.e. 35 cases). For the construction of the models, 12 variables were removed before analysis due to missing values, which are present in Table 1. A total of 18 variables were considered during model development of each dataset (Table 1). The bank material surveyed (Jerves-Cobo et al., 2018a) was classified as bedrock, boulder, cobble, pebble, gravel, sand, silt and clay. It entered the models as an ordinal categorical variable with values from one to seven. Prior to the model development, a correlation analysis was performed between the ABI and the response variables and between the response variables alone. The correlation coefficients obtained were lower than 0.8, which indicated that no strong correlations were detected and no redundant variables were included (Booth et al., 1994; Hering et al., 2006; Forio et al., 2018). The only variable discarded was temperature, which was highly correlated with DO (-0.9). Moreover, the correlation coefficient between ABI and oxygen saturation and between variables such as BOD_5 and DO, Orthophosphates and BOD_5, was in the range of $\pm$ 0.6. In this regard, Zuur et al. (2009) indicated a value of $\pm$ 0.6 is not large enough to exclude any variable. Thus, all 18 variables were included during the construction of the models. The statistical differences of ABI scores and the 18 aforementioned variables were performed between the data collected at different seasons (i.e. 19 sites). For that, two tests were used, the two tailed paired Student *t*-test when the explanatory variables or response variable (ABI) were normally distributed; otherwise the paired Wilcoxon test was applied (Demšar, 2006). Both tests were performed at the significance level of 5%. The normal distribution of the variables was analyzed by means of the Shapiro-Wilk test (Thode, 2002), while the homogeneity of their variance was checked using the Levene test (Anderson, 2006).

During the construction and validation of the GLMs, the three-fold cross-validation procedure was implemented (Goethals, 2005). For this, each dataset was randomly stratified according to the response variable (i.e. the ABI) and then partitioned into three equal sub-datasets resulting in three groups with a similar quantity of cases for each ABI class (Everaert et al., 2013; Forio et al., 2016). Subsequently, two sub-datasets were used for training the model while the third was used for testing (validation) (Fig. 3A and B). This procedure was repeated, so that all tree folds were used once for validation. In total, 15 training sets (i.e. based on two sub-datasets) were used for model development. Each training set was fitted with the three different types of GLMs indicated in Eqs. (1)–(3).

2.3.2. Variable selection

To identify which explanatory variables have an influence on the biological index ABI, a stepwise backward selection procedure was implemented. The least important input variables that had the highest *p-value* were consecutively eliminated. However, the previously removed variables were introduced again whenever they improve the model performance, that is, when the lowest Akaike information criterion (AIC) is attained (Gabriels et al., 2007; Jerves-Cobo et al., 2017). This is due to the fact that the AIC measures goodness-of-fit as well as the complexity of a model. When its value is low, GLMs tend to better fit the data, minimizing the number of parameters in the model, because a model is penalized for having too many parameters (Agresti and Kateri, 2011). The selection of the abiotic variables was accomplished with the help of the command stepAIC available in the MASS package in R (Venables and Ripley, 2002). The final model had the lowest AIC. However, based on the AIC value, one model was accepted when all or all minus one of its explanatory variables had *p-values* < 0.05. In the case of the model that had one explanatory variable with a *p-value* higher than 0.05, the model was accepted when the *p-value* of this explanatory variable was < 0.10. All statistical tests were performed at the 5% significance level.

2.3.3. Model selection and evaluation

From the models constructed in each dataset, the model that had the highest pseudo R^2 was preliminarily chosen. The pseudo R^2 was calculated based on the training set and their testing set (i.e. two-third sub-datasets used as a training set and the last sub-dataset was used as a testing set). The pseudo R^2 is a likelihood ratio index in a generalized regression model that is analogous with the R^2, which issued in multiple linear regression techniques. The value of pseudo R^2 gives a measurement of how well-observed outcomes are replicated by the model. Its value is calculated based on the Kullback-Leibler divergence (Cameron and Windmeijer, 1997). Similarly, the goodness-of-fit of each model

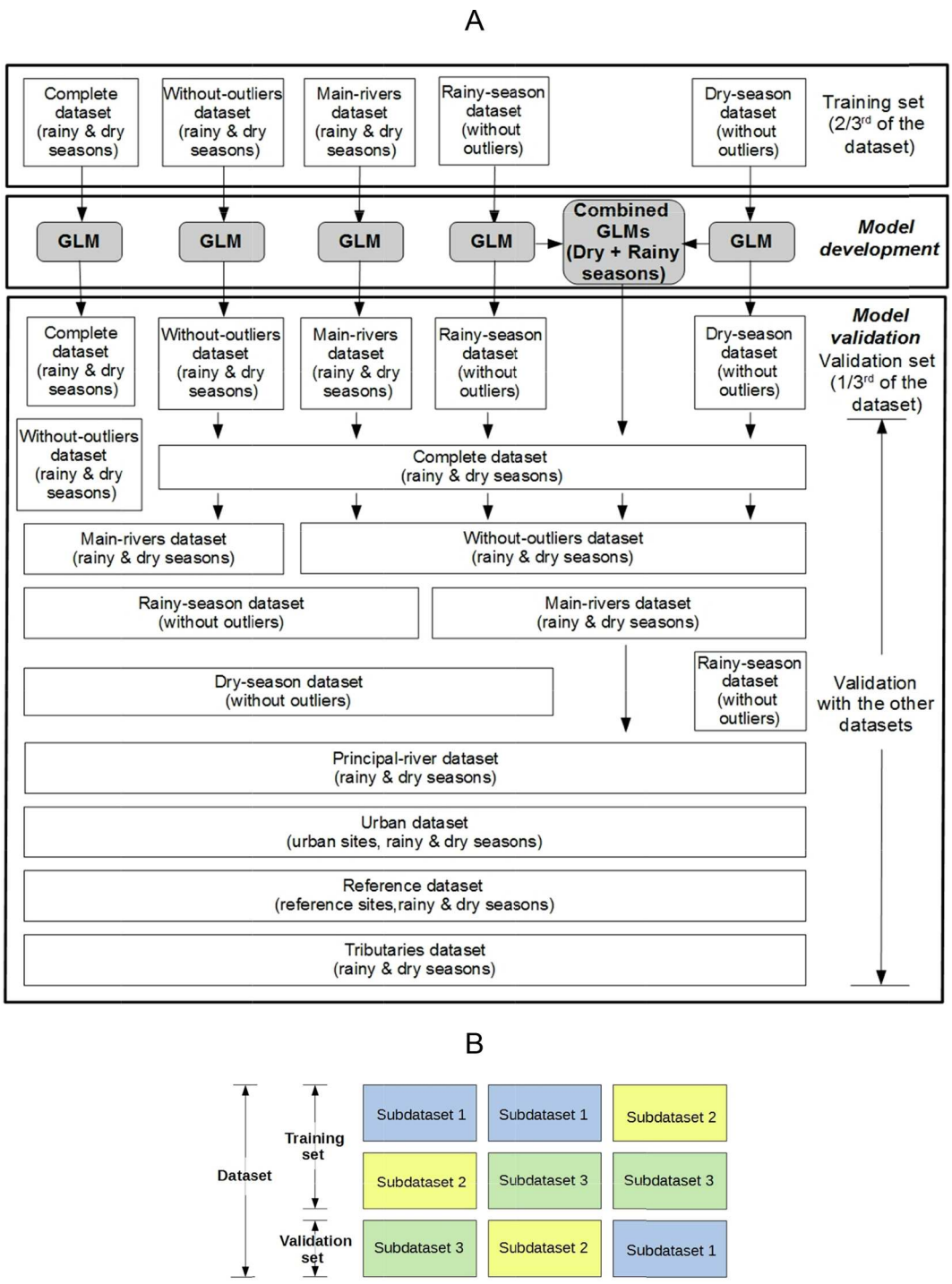

Fig. 3. Model development and validation scheme: (A) General schema of the models construction and (B) three-fold cross-validation procedure. The GLM represents generalized linear models.

Table 1
Variables included and removed for the construction of the ecological models.

Action	Variables
Included variables	Nitrate-N, nitrite-N, ammonium-N, five-day biological oxygen demand (BOD_5), dissolved oxygen (DO), oxygen saturation (OS), fecal coliforms –which was measured as most probable number (MPN) and analyzed on a logarithmic scale–, orthophosphate-P, chloride, total solids (TS), turbidity, flow velocity, water temperature, pH, conductivity, alkalinity, true color and bank material
Removed variables	Chemical oxygen demand (COD), chlorophyll-a, iron, nickel, tannins + lignins, copper, aluminum, silicon, chrome, manganese and fluorides

was also assessed with the residual plots (Zuur et al., 2009). Residual plots determine if the data meets the assumptions of the model such as homoscedasticity and normality. In case the assumptions were violated, the model was not withheld.

To choose the most appropriate model for each dataset, the predicted ABI values were transformed into a categorical variable based on the water quality classification as elaborated in Section 2.2. The categorical class obtained from the predicted ABI values was compared with the categorical class observed in the field. The accuracy of the predicted categorical classes was evaluated using the testing datasets with two measurements obtained from the confusion matrix: the correctly classified instances (CCI) and the Cohen's Kappa statistic (κ). This matrix identifies true positive (*TP*), false positive (*FP*), false negative (*FN*) and true negative (*TN*) cases. The CCI is calculated as the sum of the diagonal (i.e., TP + TN) divided by the sum of all values (i.e., TP + FP + TN + FN) (Kohavi and Provost, 1998). The value of CCI is expressed in percentage and it ranges from 0 to 100%, where a value of 100% means that the accuracy of the model is the highest (Fukuda et al., 2011). The Cohen's Kappa statistic (κ) (Eq. (4)) is a derived statistic that measures the proportion of possible cases of correct predictions (*TP* and *TN*) by a model after accounting for chance predictions (Cohen, 1960; Goethals et al., 2007; Jerves-Cobo et al., 2018b). According to Goethals et al. (2007), models are considered good when the Kappa statistic is higher than 0.4 and the CCI is at least 70% (Goethals, 2005). The models that comply with these criteria were chosen. However, in some datasets no model reached the aforementioned thresholds. Therefore, selection in these cases was done with the models achieving the highest Kappa statistic and CCI.

$$Kappa(\kappa) = \frac{(TP + TN) - \frac{(TP+FN)(TP+FP)+(FP+TN)(FN+TN)}{n}}{n - \frac{(TP+FN)(TP+FP)+(FP+TN)(FN+TN)}{n}} \quad (4)$$

Each chosen model had its own explanatory variables and its own coefficients. However, the model whose variables were most closely related to organic pollution was adopted. Because of this, the final model could be tested in a scenario analysis, under new improvements in the wastewater management in the city of Cuenca and its surrounding areas.

The final selected models from each dataset were validated with other datasets to check their accuracy (CCI) in different conditions. These datasets were featured with or without outliers in one or both seasons (Fig. 3A). In addition, these models were validated with subdatasets consisting of sites from the Tomebamba River (principal-river dataset), the urban sites (urban dataset), the reference sites (reference dataset) and the tributaries (tributaries dataset) (Fig. 3A). Finally, the models obtained during the dry and rainy seasons were combined and selectively validated according to the season in which the sites were sampled (Fig. 3A). The sites included in each dataset are presented in Table 2 and Fig. 2.

3. Results

3.1. Variable differences between seasons and influential points analysis

In sampling sites measured during dry and rainy seasons, the total number of taxa, the Shannon-Wiener index (SWD), water temperature, pH, DO, total solids and Nitrite-N were significantly different, according to the paired Student *t*-test and paired Wilcoxon test (Table A4). However, turbidity, oxygen saturation, alkalinity, chloride, ammonium and flow velocity did not show differences between seasons.

Site Ta02 of the total dataset was found to be a highly influential point based on both Cook's distance and hat elements tests when Gamma and Inverse Gaussian distributions were applied (Fig. A2A–D). According to the Cook distance, this highly influential point would affect the regression coefficients of the equations obtained from the aforementioned distributions (Stevens, 1984). However, no influential point was detected for Gaussian distribution (Fig. A2E and F). Despite this, the site Ta02 was excluded from the other datasets. The Ta02 site registered high concentrations of organic pollutants and nutrients that could sway its detection as a highly influential point. The detection of influential points was also applied to the remaining training datasets (i.e. without-outliers, main-rivers, dry-season and rainy-season datasets), in which no further sites were excluded (Figs. A3–A6). Furthermore, the Ta01 site was detected as an influential point during dry season in the main-rivers dataset, when the three GLMs (Gaussian, Gamma and Inverse Gaussian) were applied, but its influence was marginal (Fig. A4).

3.2. Season-specific models

In this section and forward, we only present the models that were performing with the highest κ and CCI values during the validation process. The season-specific models indicated that different variables were associated with the ABI during each season. Thus, during the dry season, BOD_5 and orthophosphate were the best predictors of the ABI (Tables 3, 4 and A5). Gamma and Inverse Gaussian GLMs predicted the ABI better than the Guassian GLM. However, results suggest that Inverse Gaussian was the best performing GLM.

The models obtained to predict the ABI during the rainy season (Tables 3 and A6) displayed oxygen saturation (OS) as the only common explanatory variable. Whereas nitrate, nitrite, DO, total solids and bank material were found as predictors in two of the four models (Table A6). Ammonium was included in one of the four models. Gaussian and Inverse Gaussian GLMs had good accuracy for the ABI prediction in one of the three validation datasets (Table A6).

The accuracy of the models obtained during the dry season was higher than the model developed for the rainy season. This was based on the three-fold cross-validation results obtained with the evaluation metrics: pseudo R^2, CCI and κ (Table 5). The higher concentration of organic pollutants and nutrients could have positively influenced the prediction of the ABI during the dry season. Fig. 4 presents the comparison between the observed and the predicted ABI classes, which were obtained from the models, during both seasons. The residual plots of the models obtained in each season can be seen in Fig. A7.

3.3. Season-overarching models

This section summarizes the results obtained for the models trained with complete, without-outliers, and the main-rivers datasets. These models were developed to forecast the ABI class during any season. Additionally, we compared the accuracy of the season-overarching models with the combination of season-specific models, which were selectively applied according to the season in which the sites were sampled.

Orthophosphate, DO and oxygen saturation were the common predictors of the models developed to calculate the ABI (Table 4) trained with the complete, without-outliers and main-rivers datasets. The Gamma and Gaussian GLMs provided the most accurate predictions for the aforementioned datasets. The residual plots and scale location of the models developed from the three datasets are presented in Fig. A8.

In addition, ammonium, DO and OS were selected in two of the four models developed using the complete dataset (Table A7), while, BOD_5, DO and OS were the explanatory variables in two of the models obtained for the without-outliers dataset (Table A8). In the main-rivers dataset, orthophosphate was the common predictor variable for the ABI in the four constructed models, while DO and OS were predictor variables for the ABI in three of the four constructed models (Table A9). Figs. 5A, A9A and B show the deviation of observed values and the predicted values with the complete, the without-outliers and the main-rivers datasets, respectively. Consequently, the season-overarching models shown as predictors, a mixture of variables from both dry and rainy season models. Based on the three-fold cross-validation results

Table 2
Sites included in each dataset (i.e. complete, without-outliers, main-rivers, tributaries, urban and reference sites).

Sites	Complete dataset		Without-outliers		Main-rivers		Principal-river		Tributaries		Urban sites	Reference sites
	Dry season	Rainy season	Dry season	Rainy season	Dry season	Rainy season	Dry season	Rainy season	Dry season	Rainy season		
Tomebamba sub-basin												
CU01	X	X	X	X					X	X	X	
CU02	X	X	X	X	X	X	X	X				
CU03	X	X	X	X	X	X	X	X			X	
CU04	X		X		X		X				X	
TO12	X	X	X	X	X	X	X	X			X	
TO13	X	X	X	X	X	X	X	X			X	
TO15	X		X						X		X	
TO16	X		X						X		X	
TO17	X	X	X	X					X	X		X
TO18	X	X	X	X	X	X	X	X				X
TO19	X	X	X	X	X	X	X	X			X	
TO31		X		X		X		X			X	
TO32		X		X						X	X	
TO34		X		X						X	X	
TO35		X		X						X	X	
TO37		X		X		X		X			X	
TO38		X		X		X		X			X	
TO42		X		X		X		X			X	
TO43		X		X		X		X			X	
TO44		X		X		X		X			X	
TO45		X		X		X		X			X	
TO46		X		X		X		X			X	
Machangara Sub-basin												
MC09	X	X	X	X	X	X						X
MC10	X		X						X			X
MC11	X	X	X	X	X	X					X	
MC12	X		X						X		X	
MC13	X	X	X	X	X	X					X	
Yanuncay Sub-basin												
YA14	X		X						X			X
YA15	X	X	X	X	X	X						X
YA16	X	X	X	X	X	X					X	
YA17	X	X	X	X	X	X					X	
YA30		X		X		X					X	
YA34		X		X		X					X	
YA35		X		X		X					X	
YA36		X		X		X					X	
YA37		X		X		X					X	
Tarqui Sub-basin												
TA01	X	X	X	X	X	X					X	
TA02	X								X		X	
TA03	X	X	X	X					X	X		X
TA04	X	X	X	X					X	X		
TA05	X	X	X	X	X	X						
TA22	X		X						X		X	
TA23	X	X	X	X	X	X					X	

obtained with the evaluation metrics: pseudo R^2, CCI and κ (Table 5), the models obtained with the main-rivers dataset, showed a lower accuracy in comparison with the models obtained with the complete and without-outliers datasets.

Based on the CCI and κ, the models trained with the different datasets were validated with the other datasets (Fig. 3). The best performing models (CCI ~ 70% and $\kappa > 0.4$) were the selective combination of season-specific models validated with the principal-river dataset. Furthermore, the selective validation of the combination of season-specific models had good accuracy (CCI > 60% and $\kappa > 0.4$), when they were validated with the without-outliers and the dry-season datasets as well as the urban and the tributaries datasets. The CCI and the κ were similar when the season-specific models were validated with the complete and without-outliers datasets, obtaining a slightly higher accuracy for the without-outliers dataset. When the rainy season models were validated with the principal-river (Tomebamba) dataset, the CCI and κ was over 60% and 0.4, respectively. The models trained with the complete dataset showed marginally higher accuracy than the models developed from the without-outliers dataset. In general, the season-specific models performed better than the overarching models. Thus, the selective validation of the combination of the models developed from the specific-season datasets had points closer to the bisector, indicating that the predicted values are deviating to a lesser extent from the observed values (Fig. 5B). By contrast, the lowest accuracy was noted when the reference dataset was analyzed with the other datasets. A low accuracy also resulted when a season-specific model was validated with the dataset from a different season. Table 6 presents the validation results of models developed from the different datasets.

4. Discussion

4.1. Biological and environmental differences between seasons in tropical rivers

One of the strengths of our study design is the comparison between the combinations of season-specific models with season-overarching

Table 3
The best-performing season-specific models.

Explanatory variables	Regression parameters	Season-specific models					
		Dry-season Gamma model: mD1.3fcv2a3.gamma			Rainy-season Inverse Gaussian model: mR8.gaussian		
		Coefficient	Std. Error	p-Values	Coefficient	Std. Error	p-Values
	A	1.3E−02	3.1E−03	1.0E−03	−725.5	195.9	1.0E−03
Nitrate	B1				85.3	26.6	3.0E−03
Nitrite	B2				−1100.7	319.9	2.0E−03
Ammonium	B3	−4.2E−02	1.7E−02	2.5E−02			
BOD_5	B4	5.2E−03	1.3E−03	1.0E−03			
DO	B5				−42.8	19.7	3.9E−02
Oxygen Saturation	B6				10.9	2.8	1.0E−03
Orthophosphate	B8	1.4E−01	4.7E−02	9.0E−03			
Total solids	B10				−0.1	0.1	3.3E−02
Bank material	B18				10.6	4.1	1.6E−02
Training subset (2/3)							
AIC:		140.6			314.0		
Pseudo R^2:		0.7			0.6		
CCI:		72.2%			68.6%		
κ:		0.6			0.5		
Validation subset (1/3)							
Pseudo R^2:		0.4			0.3		
CCI:		62.5%			72.7%		
κ:		0.4			0.6		

Table 4
Best-performing season-overarching models.

Explanatory variables	Regression Parameters	Models for both season								
		Complete dataset Gamma model m03.3fcv2a3T.gamma			Without-outliers dataset Gamma model m1.3fcv2a3.gamma			Main-rivers dataset Gaussian model: m2um3fcv1a2.gaussian		
		Coefficient	Std. Error	p-Values	Coefficient	Std. Error	p-Values	Coefficient	Std. Error	p-Values
	A	3.0E−01	8.1E−02	< 0.01	2.2E−01	7.8E−02	0.01	−764.7	243.8	< 0.01
Ammonium	B3	−9.4E−03	2.3E−03	< 0.01						
BOD_5	B4				1.5E−03	7.2E−04	0.04			
DO	B5	1.8E−02	5.9E−03	0.01	2.0E−02	5.7E−03	< 0.01	−30.7	13.5	0.03
Oxygen saturation	B6	−4.1E−03	1.0E−03	< 0.01	−3.6E−03	9.7E−04	< 0.01	10.5	2.7	< 0.01
Orthophosphate	B8	6.7E−02	2.3E−02	0.01	3.3E−02	1.8E−02	0.07	−88.6	40.9	0.04
Training subset (2/3)										
AIC:			353.6			347.4			273.4	
Pseudo R^2:			0.5			0.6			0.5	
CCI:			48.8%			53.7%			53.3%	
κ			0.3			0.3			0.3	
Validation subset (1/3)										
Pseudo R^2:			0.5			0.6			0.4	
CCI (%):			61.9%			60.0%			57.1%	
κ			0.5			0.4			0.4	

Table 5
Three-fold cross-validation results of the models with different datasets. Mean and stand deviation of pseudo R^2, CCI and κ.

Dataset	Pseudo R^2 ± sd			CCI ± sd			κ ± sd		
Dry-season	0.35	±	0.30	42.3%	±	16.6%	0.21	±	0.20
Rainy-season	0.24	±	0.24	44.9%	±	10.4%	0.24	±	0.24
Complete	0.29	±	0.17	44.1%	±	10.1%	0.22	±	0.13
Without-outliers	0.33	±	0.24	44.6%	±	11.4%	0.22	±	0.16
Main-rivers	0.18	±	0.12	39.8%	±	12.1%	0.16	±	0.15

models in tropical places. With respect to the modelling of water quality in tropical rivers during the dry and rainy seasons, Maillard and Santos (2008) conducted a study in Brazil and found that some of the physicochemical variables used in the season-specific models had the opposite behavior between seasons. However, these researchers constructed physicochemical water quality models, in which they analyzed neither the results of a possible season-overarching model, nor the modelling of the biological status of the rivers. In another study, Jacobsen (1998) discovered that there was a variation of macroinvertebrate assemblage between the dry and rainy seasons at the high altitude tropical environment.

The BOD_5, orthophosphate, and ammonium mainly predicted the ABI in the Cuenca River basin during the dry season. The increase of BOD_5, orthophosphate and ammonium along the rivers during dry season in the study area was possibly due to three causes: (1) sewage discharges; (2) some sewage overflows whose operational level was incorrectly calibrated; and (3) diffuse pollution from runoff originated

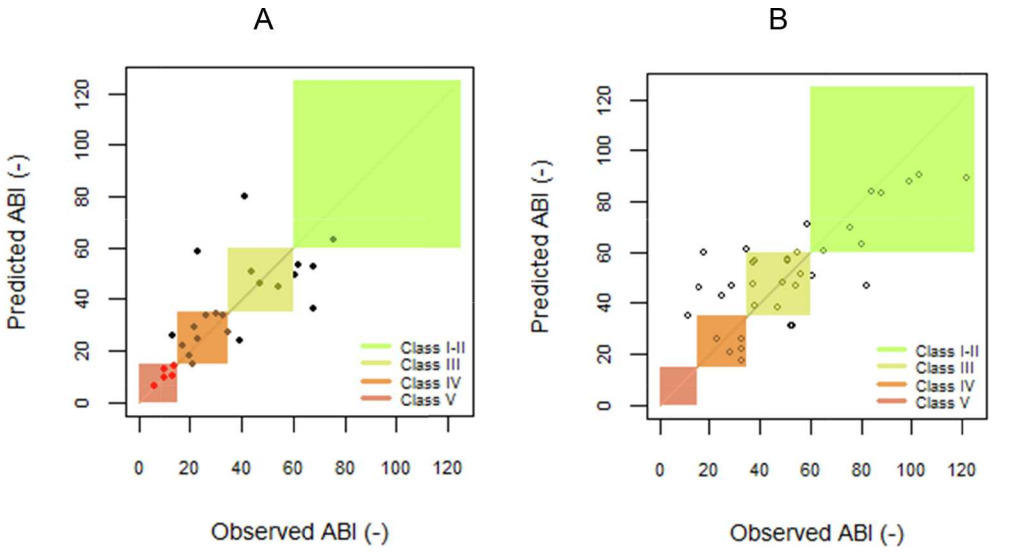

Fig. 4. Graph comparing observed data and simulated index obtained from the best models during the (A) dry season and (B) rainy season. The colors inside the squares indicate the biological water quality (ABI class). The dots inside the box are the sampling sites whose ABI classes were correctly predicted. The colors green, brown, orange and red represents good and very good (Class I-II), moderate (Class III), deficient (Class IV) and bad (Class V) biological water quality, respectively.

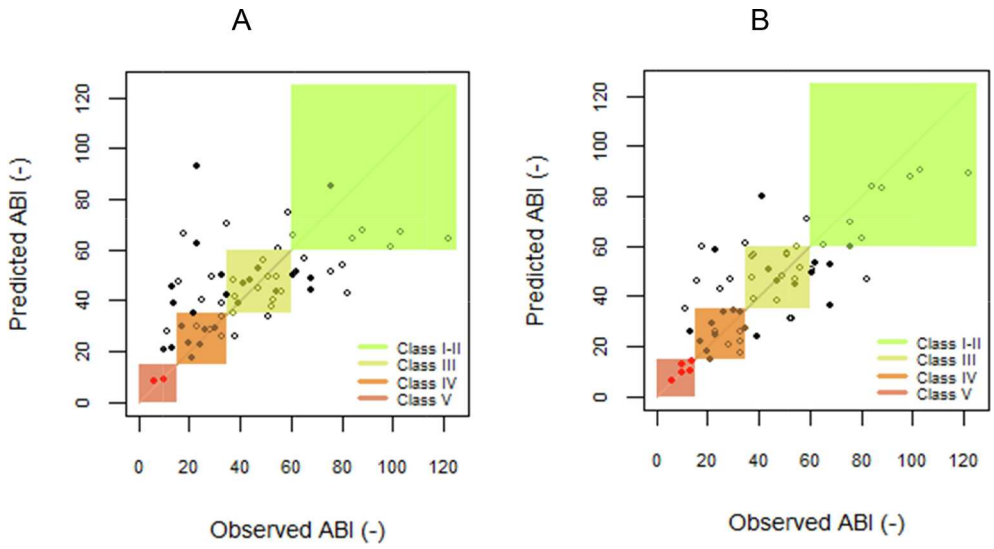

Fig. 5. Graph comparing simulated index and field data for models developed from: (A) the complete dataset, and (B) selective validation of the combination of the best models for the dry and the rainy seasons. The filled dots and the empty dots are the sampling sites taken during dry and rainy seasons respectively.

by irrigation of agriculture and grassland areas in the suburban areas of the city (Jerves-Cobo et al., 2018a). Thus, the concentration of these pollutants was higher during the dry season, affecting the presence of macroinvertebrates and consequently the biological index ABI. Macroinvertebrates per se are affected by oxygen saturation and not by pollutants such as orthophosphate, BOD_5 and ammonium (Jacobsen, 1998; Beyene et al., 2009). However, the depletion of oxygen concentration is influenced by these pollutants (Hynes, 1960), which were higher correlated to the oxygen concentration during the dry season than the rainy season. These results agreed with the findings of other studies carried out in the high altitude of the Andes in Ecuador (Jacobsen, 1998; Burneo and Gunkel, 2003; Jerves-Cobo et al., 2017). These authors indicated that the number of macroinvertebrate species was reduced by the increase of organic pollutants and phosphates. Similarly, the BOD_5 and nutrient concentrations were negatively associated with most taxa, both in tropical and higher latitude streams (Schleiter et al., 1999; Beyene et al., 2009; Friberg et al., 2010).

During the rainy season, the ABI was on average higher than that obtained during the dry season. The effect of organic pollution during the rainy season was diminished by a greater dilution of the wastewater (Jacobsen, 1998), resulting in the selection of variables such as oxygen saturation (OS), dissolved oxygen (DO), nitrite, nitrate and total solids as predictors of the ABI. During the rainy season, the percentage of OS, as well as the concentration of TS and nitrite, was higher. The DO concentration in most sites of the main rivers was close to the saturation level during both the dry and rainy seasons, with higher values during the dry season. Moreover, in both seasons the values of the DO were higher in the sites located upstream than the downstream sites, which was the opposite with the concentrations of BOD_5 and nutrients (Jerves-Cobo et al., 2018a). Similar findings concerning the DO in both seasons were found by Ríos-Touma et al. (2011) in the northern Andes of Ecuador. Additionally, the flow velocity and turbulence were also greater during the rainy season aiding the maintenance of relatively high oxygen saturation (Jacobsen, 1998; Jerves-Cobo et al., 2017). However, the oxygen saturation level is relatively low for 2500 m a.s.l. regions, in comparison with low altitude regions whose average values for the dry and rainy seasons, respectively, are of 8.1 $mg{\cdot}l^{-1}$ and 7.9 $mg.l^{-1}$. The average value of DO, obtained during the sampling

Table 6
Independent validation of the models developed from the different datasets (complete, without-outliers, main-rivers, dry-season and rainy-season datasets). Mean and standard deviation of CCI and κ were derived after the application of three-fold cross-validation. Values in bold represent the best-performing models: CCI > 60% and κ > 0.4.

Validation sets	Model					
	Complete dataset		Without-outliers dataset		Main-rivers dataset	
	CCI ± sd	κ ± sd	CCI ± sd	κ ± sd	CCI ± sd	κ ± sd
Complete dataset	54.0 ± 5.0	0.4 ± 0.1	52.0 ± 4.6	0.3 ± 0.1	51.2 ± 0.8	0.3 ± 0
Without-outliers	54.9 ± 5.1	0.4 ± 0.1	51.6 ± 5.1	0.3 ± 0.1	52.1 ± 0.8	0.3 ± 0
Main-rivers	54.6 ± 4.9	0.3 ± 0.1	51.7 ± 6	0.3 ± 0.1	50.6 ± 1.1	0.3 ± 0
Dry-season	50.0 ± 4.4	0.3 ± 0.1	48.1 ± 5	0.3 ± 0.1	50.0 ± 3.1	0.3 ± 0
Rainy-season	58.6 ± 6.8	0.4 ± 0.1	54.3 ± 7.7	0.3 ± 0.1	53.6 ± 1.4	0.3 ± 0
Principal-river (Tomebamba River)	58.3 ± 2.4	0.4 ± 0	58.3 ± 8.1	0.4 ± 0.1	58.3 ± 4.6	0.4 ± 0.1
Urban sites	55.7 ± 6.0	0.4 ± 0.1	52.8 ± 5.4	0.3 ± 0.1	52.8 ± 1.1	0.4 ± 0
Reference sites	37.5 ± 8.3	0.1 ± 0.1	39.6 ± 8	0.1 ± 0.1	35.4 ± 4.2	0 ± 0.1
Tributaries	52.8 ± 7.2	0.4 ± 0.1	52.8 ± 3.2	0.4 ± 0	52.8 ± 3.2	0.4 ± 0
Validation sets	Model				Selective validation of the	
	Dry-season dataset		Rainy-season dataset		Combination of dry season + rainy season models	
	CCI ± sd	κ ± sd	CCI ± sd	κ ± sd	CCI ± sd	κ ± sd
Complete dataset	45.6 ± 2	0.2 ± 0	47.6 ± 4.7	0.3 ± 0.1	59.3 ± 3.6	0.4 ± 0.1
Without-outliers	46.3 ± 2.1	0.2 ± 0	48.4 ± 4.7	0.3 ± 0.1	60.3 ± 3.6	0.5 ± 0.1
Main-rivers	46.6 ± 4.8	0.2 ± 0.1	48.3 ± 4.7	0.3 ± 0.1	58 ± 2.9	0.4 ± 0
Dry-season	60.6 ± 4.8	0.5 ± 0.1	37.5 ± 8.5	0.2 ± 0.1	60.6 ± 4.8	0.5 ± 0.1
Rainy-season	35.7 ± 5.5	0 ± 0.1	56.4 ± 6.3	0.4 ± 0.1	56.4 ± 6.3	0.4 ± 0.1
Principal-river (Tomebamba River)	36.9 ± 2.4	0.1 ± 0	60.7 ± 4.6	0.4 ± 0.1	69.1 ± 6.2	0.6 ± 0.1
Urban sites	50 ± 1.9	0.3 ± 0	50 ± 12.5	0.3 ± 0.1	65.3 ± 5	0.5 ± 0.1
Reference sites	20.8 ± 4.8	−0.1 ± 0.1	33.3 ± 17.4	0 ± 0.2	33.3 ± 6.8	0 ± 0.1
Tributaries	43.1 ± 2.8	0.2 ± 0	45.8 ± 7.9	0.3 ± 0.1	62.5 ± 9.5	0.5 ± 0.1

campaigns was 7.3 $mg.l^{-1}$ and 7.6 $mg.l^{-1}$ for the dry and rainy seasons, respectively. In this regard, Jacobsen (1998) pointed out that the oxygen concentration of saturated water decreases with increased altitude. Moreover, the author revealed that in the high Andes region with oxygen saturation between 80 and 90%, many Ephemeroptera, Plecoptera and Trichoptera (EPT) taxa almost disappeared completely. Hence, macroinvertebrates from the tropical highland Andes streams may be particularly sensitive to the diminishing oxygen levels by organic pollution. The depletion of DO in the rivers prevails during the aerobic conversion of organic pollutants (Hynes, 1960) by bacterial decomposition (Rauch et al., 1998). However, the higher water temperature measured during the rainy season could have had a more significant influence on the DO than the lower concentration of the organic pollutants. Thus, the dissolved oxygen saturation decreased on average around 6% due to the higher water temperature during the rainy season, while the DO diminished on average around 5% in main rivers, possibly due to the higher water temperature and the lower organic pollutant concentrations. It is important to note that the day and night cycles of the DO and OS were not analyzed in this study. According to Wilcock et al. (1978), when the water temperature increases at the same partial pressure, the solubility of oxygen decreases leading to a reduced DO concentration. Whereas oxygen saturation (OS) is related to water temperature and atmospheric pressure variation, diminishing with the increase of temperature and with a drop of atmospheric pressure (Mortimer, 1981). Oxygen saturation and the dissolved oxygen (DO) have also been established as key variables that explain the presence of macroinvertebrates in the Ecuadorian Andes (Jacobsen and Marín, 2008).

In the main rivers, during the rainy season, the concentration of ammonium was a little lower than that measured during the dry season, while nitrite registered a little higher value in its concentrations. In this regard, during the rainy season ammonium has been related to the wash-out from the sewer systems and discharged by means of CSOs (Holzer and Krebs, 1998). The concentration of nitrite, a pollutant measured in a very low range and expressed, as μg N/L, is apparently associated with the ammonium concentrations and higher flows resulting in higher stream velocities during the rainy season. In this regard, it is known that increased levels of nitrite can be found in fast-flowing aerobic small streams, a characteristic of Andean mountain rivers, in which ammonium oxidation via nitrification is responsible for the higher concentrations of this pollutant (Kelso et al., 1999; Philips et al., 2002). Moreover, runoff from the livestock areas that originated during rainy events could contribute to the diffuse fluxes of organic pollutants (Fernandez de Cordova and González, 2012; Jerves-Cobo et al., 2018a). The total solids (TS), which are the sum of dissolved solids plus suspended solids and settled solids, were similar in both seasons, with a concentration during dry seasons at approximately 75% of that registered during the rainy season. The increase of TS in rivers is attributable to multiple factors such as runoff, relief, lithology, rainfall pattern, vegetation, and basin size (Meybeck et al., 2003). In the study area, the TS may have been transported to the rivers from combined sewage outfalls (CSOs) from urban areas that discharge wastewater and runoff during rainfalls. Moreover, the runoff from rural areas, which is produced both by irrigation during the dry season and by rains during the rainy season, may have moved the total solids (TS) to the rivers. In this regard, it has been found that suspended solids influence the community structure of macroinvertebrates in streams (Gray and Ward, 1982; Doeg and Milledge, 1991).

4.2. Season-specific and season-overarching models

Our results demonstrate that GLMs can be used to select physico-chemical variables that best predict the biological water quality based on the ABI index. However, in most models, Ψ was lower than 0.4 and CCI was lower than 70%, indicating that only a few models were reliable. The accuracy of the models could also have been affected by the transformation of a continuous response variable ABI, into a categorical variable.

The selective combination of season-specific models that were applied according to the season in which the sites were sampled, displayed a higher accuracy than the season-overarching models. Thus, the season-specific models showed higher values in CCI and κ. at least 5% and 10%, respectively. This could be due to the fact that the main variables predicting the ABI were different between seasons. Similarly, the obtained models developed for the dry season were more precise and stable than the models developed for the rainy season. Thus, during the dry season, the ABI was mainly predicted by orthophosphate, ammonium, and BOD_5. In this regard, Jacobsen (1998) found that the biotic indices BMWP (Biological Monitoring Working Party) and ASPT (Average Score Per Taxon) were closely related to the dissolved oxygen and phosphate concentrations in the high Andes region during the dry season. Additionally, Holguin-Gonzalez et al. (2013b) established that the Trichoptera presence was mostly related to the BOD_5. Consequently, these sensitive taxa prefer water with low concentrations of BOD_5.Yet, the concentration of organic pollutants decreased during the rainy season. Thus, the ABI was primarily associated with variables such as OS, DO, nitrite and total solids during the rainy season. Similar outcomes were described by Jacobsen (1998), who found a weaker correlation between the biotic indices and organic pollutants during the rainy period. Furthermore, during the rainy season, the models revealed the influence of morphological variables such as bank material. In this regard, Burneo and Gunkel (2003) found in a study developed in the northern Andes of Ecuador that bank material, a morphological characteristic, affected invertebrate communities. By contrast, the presence of macrophytes and the effect of shading were not relevant; probably they were limited in the sampled sites (Jerves-Cobo et al., 2018a). The validation results obtained from the rainy dataset were quite stable as indicated by the standard deviation of. $^{c}\Psi$ with a variation between 0 and Ψ 10% of the mean κ. (Table 6). However, models developed for the rainy season showed a higher variation of explanatory variables in three-fold cross-validation than the models developed for the dry season (Tables A5 and A6).

As previously mentioned, the predictors of the season-specific models for ABI varied between seasons. Thus, during dry season, the ABI was mainly predicted by orthophosphate and the BOD_5, while variables such as OS, DO, nitrite, nitrate and total solids primarily influenced the prediction of the ABI during the rainy season. The application of the season-specific models developed for dry season showed a low accuracy when they were validated with the rainy-season dataset, and vice versa. In this regard, McCullagh (1984) revealed that a model developed with a particular dataset is not able to incorporate the inevitable changes given in another dataset that was collected under different conditions. In contrast, the season-overarching models had the following predictors: orthophosphate, BOD_5, DO and OS. This demonstrates that a mixture of predictors from both the dry and rainy season models influenced the season-overarching models. Variables such as nitrite, nitrate and total solids had less weight in the prediction of the ABI in the season-overarching models. Both the mixture and loss of predictors in the season-overarching models have likely caused their lower accuracy in comparison to the season-specific models. However, when the season-specific models were validated with the different season dataset, the accuracy of the season-overarching models was better.

4.3. Model development

The only correlated variable that was eliminated prior to model construction was temperature. Assuming that this research included multi-collinearity, in this regard, Shmueli (2010) pointed out that multi-collinearity is not damning because the prediction performance of the models was not affected. Consequently, either for supposed possible inclusion of the correlated variables or for their exclusion, the forecast capacity of the models was not altered.

During the construction of the models, we compared the results obtained from the datasets both with and without outliers. An outlier is a result from either an influence observation or a measurement error (Rousseeuw and Leroy, 2005). In this research, we classified the influence observations as outliers that were registered in the polluted tributaries of the main rivers. The models constructed in this study indicated that both CCI and κ. varied on average ± 2%. However, the models developed using these two datasets had few explanatory variables that were different (Tables A7 and A8). This would imply that outliers affected the selection of explanatory variables, but not the accuracy of the models. In this regard, Chatterjee and Hadi (1986) pointed out that outliers can affect the linear regression models, either concerning the lack of fit, the selection of explanatory variables, or both.

Gamma and Inverse Gaussian distributions demonstrated a better accuracy in the prediction of the ABI over all the datasets used in this research. This could be because most of the data was collected in urban and suburban areas, where the biological water quality varied principally between moderate and bad, particularly during the dry season. Therefore, the data were skewed, resulting in the suitability of the aforementioned GLMs (Folks and Chhikara, 1978; Attrill et al., 1996). However, the Gaussian GLM performed well during the rainy season. This was likely due to the Gaussian distribution of biological water quality classes that varied mainly between good and deficient; consequently, the data were less skewed.

The datasets used in this research were relatively small, ranging from 27 to 62 samples. According to Stockwell and Peterson (2002) and Vaughan and Ormerod (2005), the application of a large sample size – higher than 100 observations – allows models to obtain better accuracy in their prediction. However, Stockwell and Peterson (2002) concluded that the accuracy of samples was close to maximum when the size was higher than 50 observations while the accuracy diminishes to 90% when the sample size was 10 observations. Furthermore, Vayssières et al. (2000) indicated that the results obtained from small datasets were more efficient with GLMs, than with non-parametric techniques such as generalized additive models or classification trees. Similarly, biological indices from a small dataset have been modeled with GLMs, demonstrating a stable goodness-of-fit in their prediction (Holguin-Gonzalez et al., 2013a; Damanik-Ambarita et al., 2016). Moreover, to avoid possible overfitting in the linear models that could be produced with small datasets, we follow the recommendations given by Shmueli (2010). Thus, (1) for the construction of the models the cross-validation procedure was performed; and (2) the similarity of the accuracy of the models, measured as pseudo-R^2, obtained within the training and validation datasets were checked. Furthermore, the outliers that could produce overfitting were deleted before the model construction (Zuur et al., 2009).

The water quality of rivers is not static. Its variation is related to the frequency of discharges from anthropogenic sources, and the presence of substances transported by runoff from rainfall and irrigation.

Consequently, river water quality displays a dynamic variation over time. Similarly, with regard to macroinvertebrates, a taxon could be recorded absent in a site although it was present as it was not detected (Elith and Leathwick, 2009). Additionally, in habitats, it is possible that some taxa may not be present due to migration patterns, seasonal variation or unstable habitats (Jacobsen and Encalada, 1998). However, this water quality was assessed with discrete sampling campaigns with one sampling event per season. Consequently, the models were constructed with information from these sampling campaigns, which offered insight mainly into the impact generated from natural and anthropogenic sources.

An ecological model is an important tool that can be applied in river restoration. However, ecological models have two limitations: the models are constructed with information of a specific area, and the models are developed with datasets whose variables have a definite range of values. Consequently, the use of ecological models is restricted to the area for which they were generated (Forio et al., 2016; Jerves-Cobo et al., 2018b). Furthermore, the accuracy of these statistical models can be altered if their range of values for each variable is different than that which was obtained in their construction (McCullagh, 1984).

In this research, numerous hypothesis tests that included many environmental variables, model selection procedures and a biological index were performed. However, no stage corrections for multiplicity were developed. This implies that the presence of more false positives, which could be obtained with the applied individual statistical hypothesis test at 5%, is possible. Potential approaches that could be used to correct the p-values in a multiplicity analysis could be the false discovery rate (FDR) or the family-wise error rate (FWER) at 5%. However, these procedures are rarely applied in ecological studies (Forio, 2017). For this reason, this correction was not implemented in this research.

In this study, despite the limited number of samples, the use of the GLMs, the three cross-validation procedure and the standardized method applied during the collection of samples, all contributed to the reliability and accuracy of the models showing relevant ecological patterns. However, a future update to the ecological models constructed in this research is recommended. This update could be developed with future data collected from new sampling campaigns or with information from a monitoring program. Furthermore, this new information could help to check for the replication of the current results and to enhance the accuracy of the current models.

4.4. Main rivers and tributaries, and their impacts from urbanization

The expansion of sewage interceptors and the combined sewage network in the city of Cuenca and the Ucubamba Wastewater Treatment Plant (U-WWTP) (ETAPA-EP, 2007) have improved the water quality of the rivers (Fernandez de Cordova and González, 2012). However, degradation of the water quality is still occurring because of pollution point sources such as industrial discharges, combined sewage overflows (CSOs), surface water outfall (SWO), overflow before the U-WWTP and fluxes of organic pollution from an extensive livestock area (Jerves-Cobo et al., 2018a). For these reasons, we analyzed the accuracy of the developed models to assess their applicability in improving water quality status and their effectiveness in river management. In this way, the constructed models displayed the lowest accuracy when applied to the sub-dataset of the reference sites, wherein sites were classified as having only good and very good biological water quality. Possibly, this low accuracy was because the models were mainly constructed with information collected in disturbed sites, located in urban and suburban areas. In this regard, McCullagh (1984) indicated that a model provides good predictions according to the range of conditions used in its development. Furthermore, the complexity of the biological interactions is not incorporated into the models, which could also contribute to reduced accuracy. Conversely, the selective combination of the season-specific models that were applied according to the season in which the sites were sampled, appeared to be more suitable for their validation with the sub-datasets of the main-rivers, the principal-river and tributaries, obtaining the highest accuracy with the principal-river sub-dataset. In this regard, the combination of the season-specific models could be used to support the results of possible improvements for the main rivers (Tomebamba, Machangara, Yanuncay and Tarqui Rivers), the principal river (Tomebamba), or their tributaries. These improvements could be tested in a scenario analysis wherein possible actions can be simulated in order to effectively restore the rivers.

5. Conclusions

Three main conclusions could be drawn from this modelling study. First, the season-specific models were more reliable than the season-overarching model. Second, the prediction of the biological water quality was more accurate at sites sampled in the urban area than at the reference sites indicating the difficulty in predicting at more biodiverse sites. Finally, the main predictors of the Andean Biotic Index (ABI) varied between seasons. The five-day biological oxygen demand (BOD_5), orthophosphate and ammonium were the main explanatory variables during the dry season, while oxygen saturation (OS), dissolved oxygen (DO), nitrate and total solids were the major variables during the rainy season. Consequently, the development of season-specific models seems relevant as a basis to establish key actions for river restoration. The constructed models could be improved with information from new sampling campaigns. The modelling approach developed in this study can be applied and/or extrapolated to similar basins in tropical countries.

Acknowledgements

This research was executed in the context of the VLIR-UOS IUC Programme – University of Cuenca and the VLIR Ecuador Biodiversity Network project. Marie Anne Eurie Forio receives financial support from the special research fund of Ghent University. The authors would also like to extend their gratitude for the collaboration with the Universidad Técnica del Norte, Universidad del Azuay and Universidad Politécnica Salesiana. Similarly, appreciation is extended to the Water Supply and Wastewater Management Municipal Company ETAPA – EP, Electro Generator of the Austro Company – ELECAUSTRO S. A. and the Ecuadorian Environmental Ministry.

Appendix A

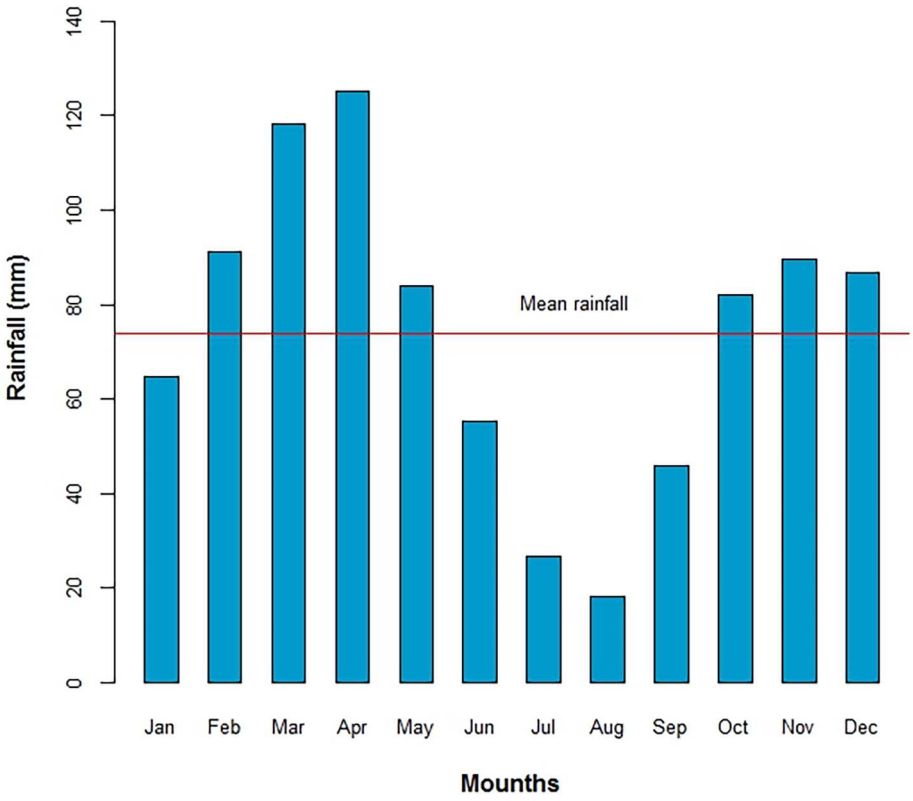

Fig. A1. Histogram of monthly average rainfall in the city of Cuenca (1990–2012).

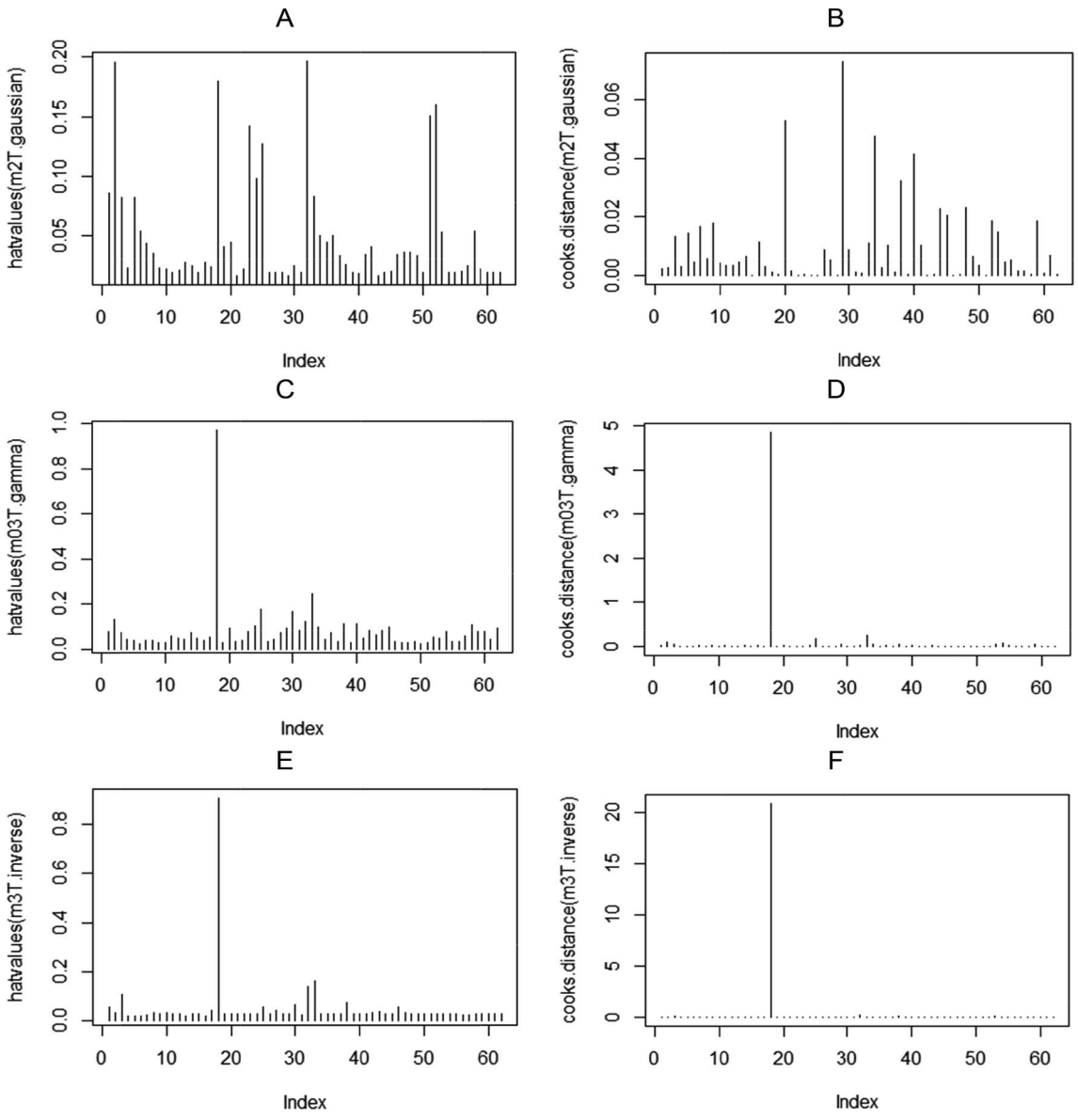

Fig. A2. Plots of influential points with the complete dataset, using Cook's distance and hat elements with different kind of GLM: (A and B) Gaussian, (C and D) Gamma, and (E – F) Inverse Gaussian.

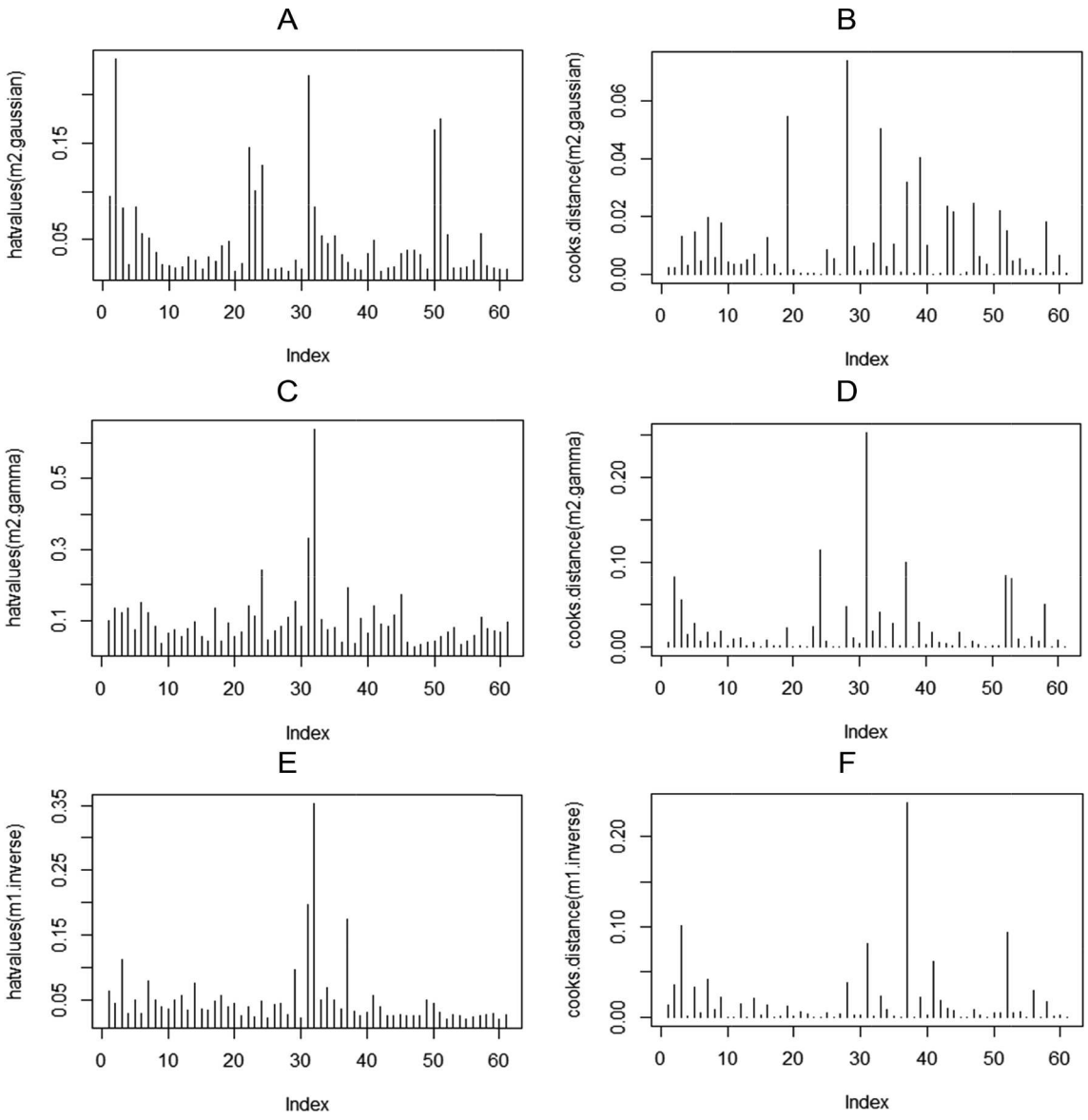

Fig. A3. Plots of influetial points with the without-outliers (Ta02) dataset, using Cook's distance and hat elements with different kind of GLM: (A – B) Gaussian, (C – D) Gamma, and (E – F) Inverse Gaussian.

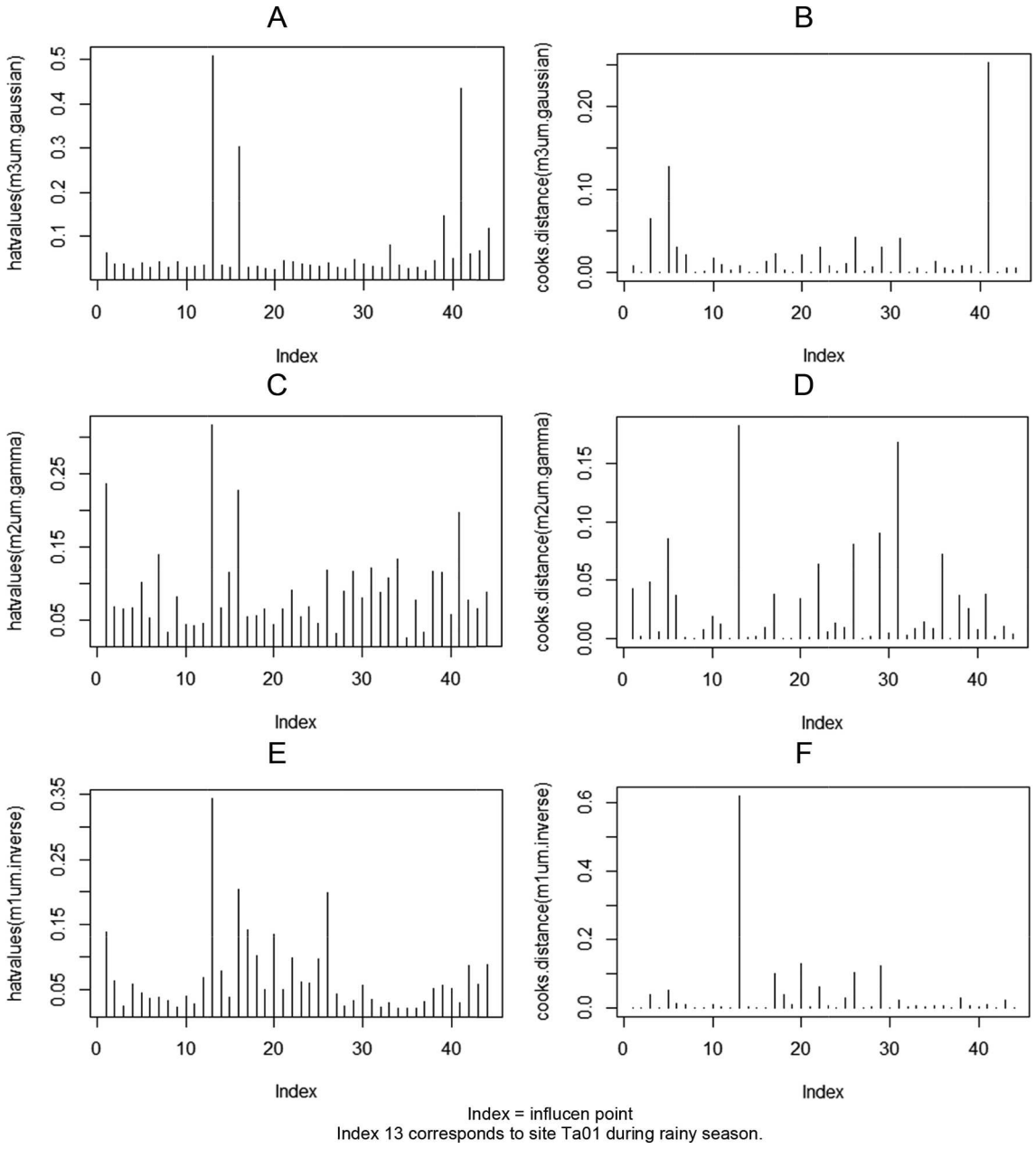

Fig. A4. Plots of influential points with the main-rivers dataset, using Cook's distance and hat elements with different kind of GLM: (A – B) Gaussian, (C – D) Gamma, and (E – F) Inverse Gaussian.

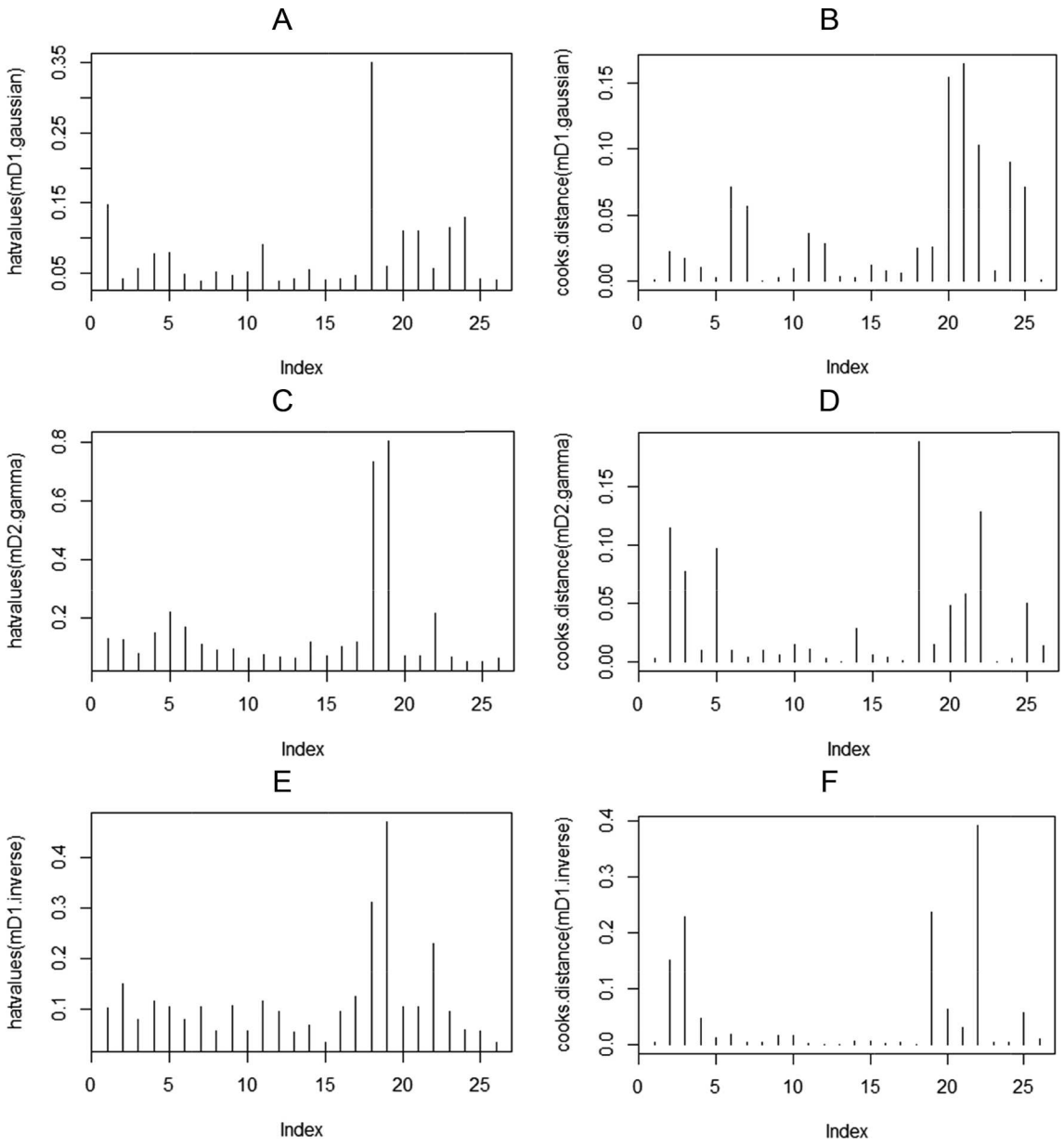

Fig. A5. Plots of influential points with the dry-season dataset, using Cook's distance and hat elements with different kind of GLM: (A – B) Gaussian, (C – D) Gamma, and (E – F) Inverse Gaussian.

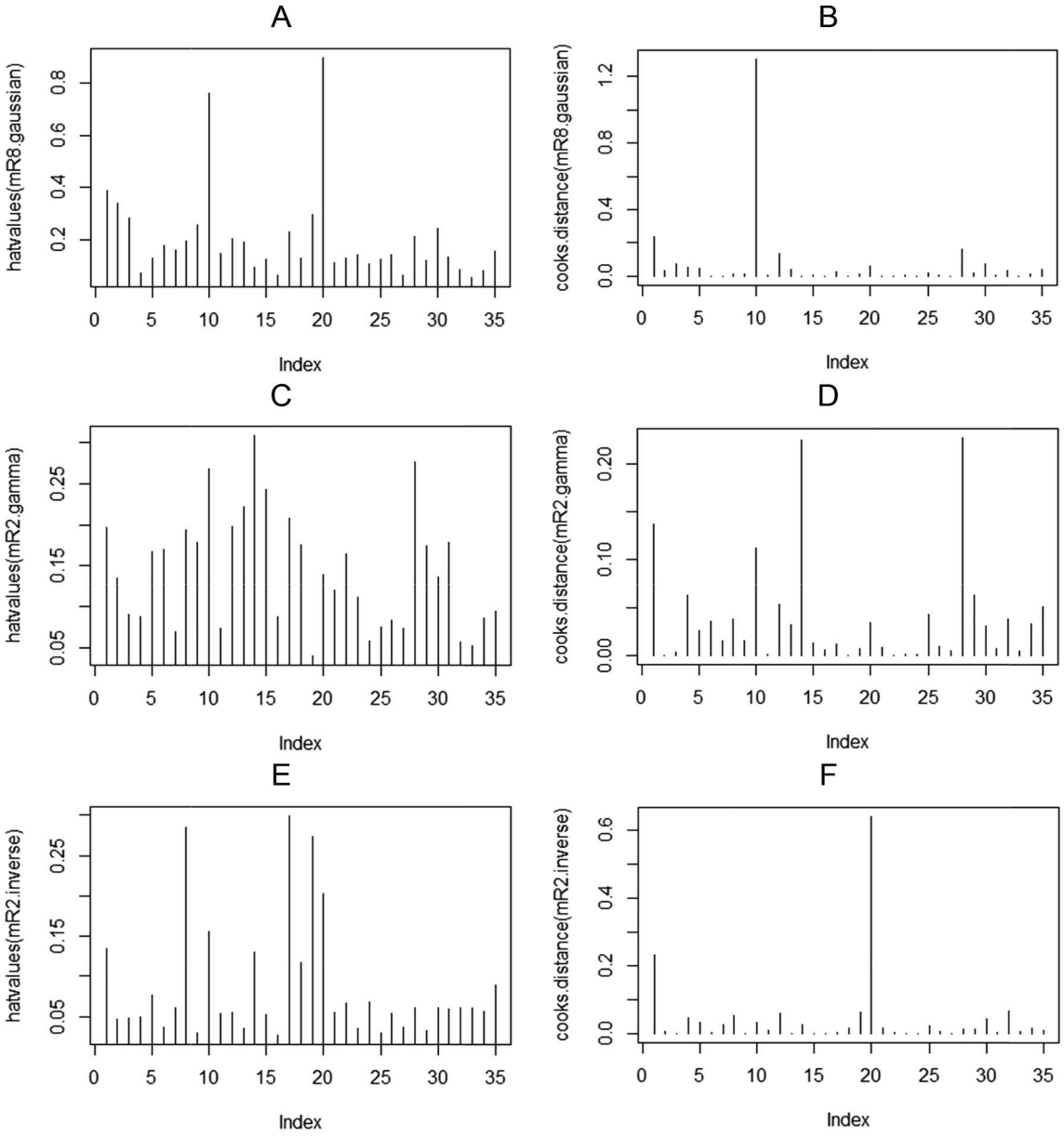

Fig. A6. Plots of influential points with the rainy-season dataset, using Cook's distance and hat elements with different kind of GLM: (A – B) Gaussian, (C – D) Gamma, and (E – F) Inverse Gaussian.

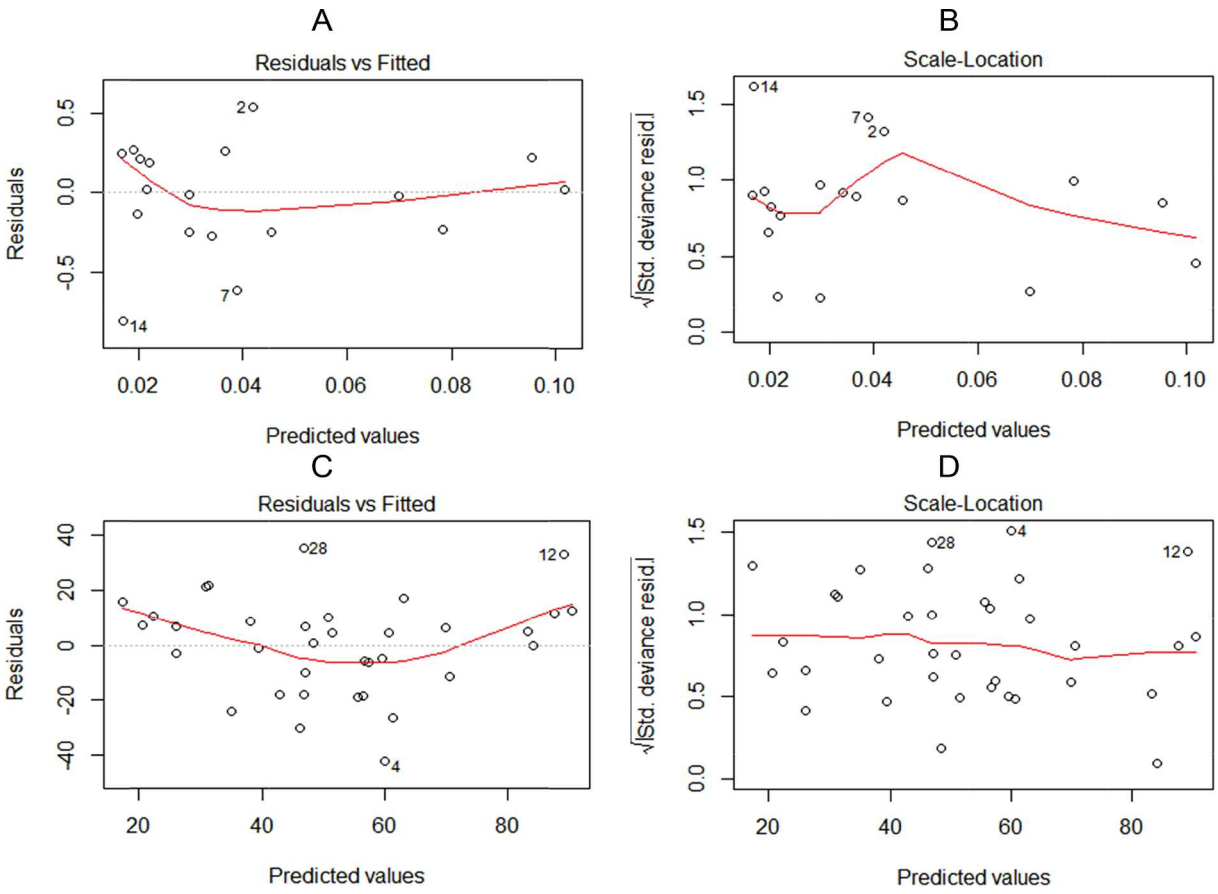

Fig. A7. Plots of Residual vs. Fitted of the best models: (A) Dry season, (B) Rainy season, and plots of Scale-Location of the best models: (C) Dry season, (D) Rainy season.

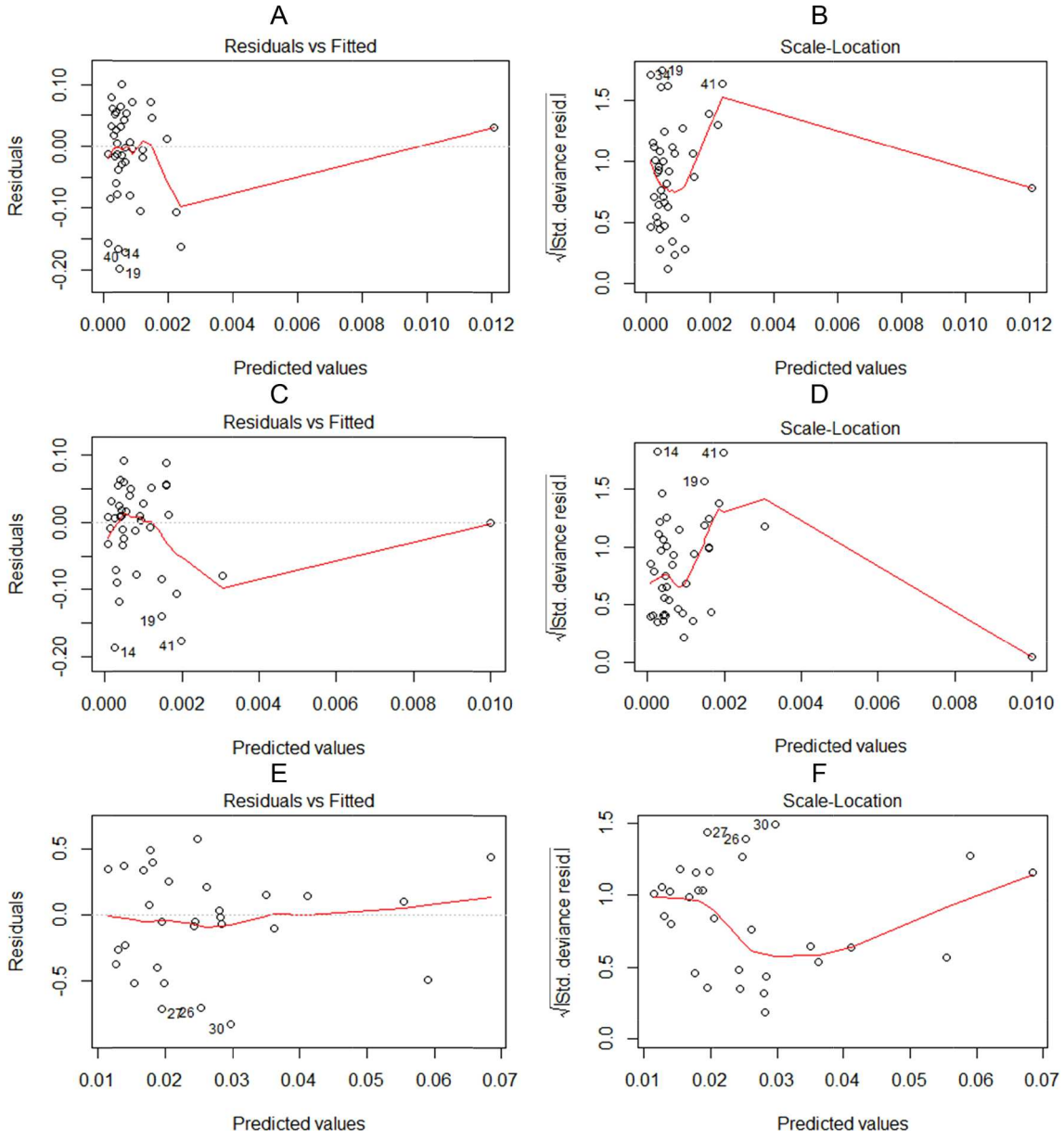

Fig. A8. Plots of Residual vs. Fitted of the best models: (A) with the complete dataset, (B) with the without-outliers dataset, (C) with the main-rivers dataset, and plots of Scale-Location of the best models: (D) with the complete dataset, (E) the without-outliers dataset, and (F) with the main-rivers dataset.

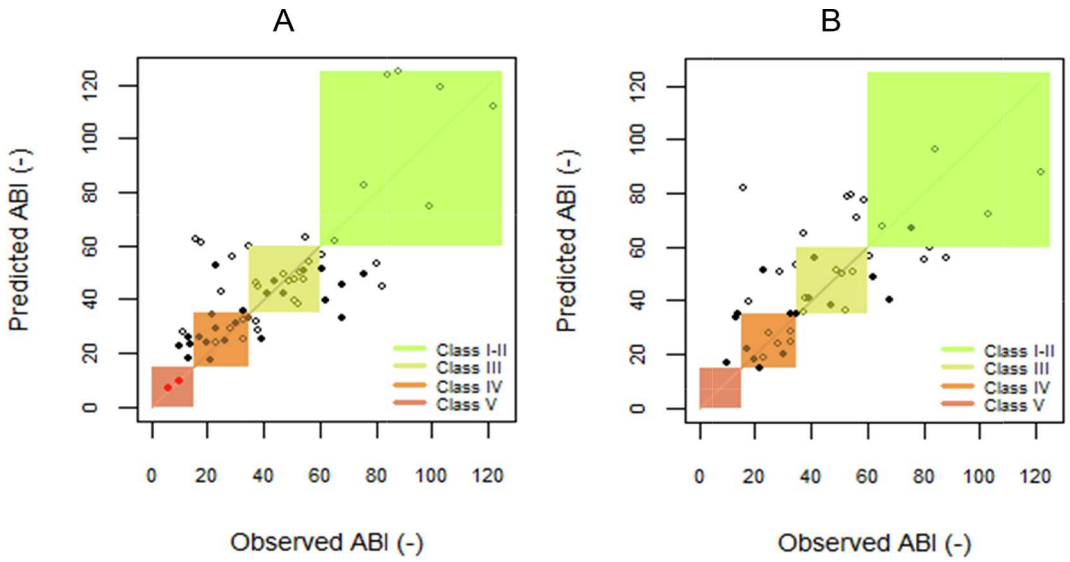

Fig. A9. Graph comparing simulated index and field data for models developed from: (A) the without-outlier dataset, and (B) the main-rivers dataset. The filled dots and the empty dots are the sampling sites taken during dry and rainy seasons respectively.

Table A1
Summary of the physical, chemical and microbiological data collected in the study area based on 43 samples in 2015 and 2016.

Parameter	Units	Mean Value		Standard Deviation	Min Value	Max Value	Median Value	Mean Dry Season	Mean Rainy Season
Temperature	°C	13.9	±	2.0	10.4	19.8	13.9	13.2	14.4
Specific conductivity	$\mu S \cdot cm^{-1}$	135.7	±	194.4	29.0	1396.5	90.9	177.6	103.4
pH		7.5	±	0.4	6.6	8.8	7.6	7.7	7.5
Turbidity	NTU	27.3	±	33.4	0.8	187.0	11.8	30.5	24.8
Chlorophyll-a	$\mu g \cdot L^{-1}$	9.7	±	7.3	2.4	29.3	6.9	9.7	–
Dissolved oxygen (DO)	$mg \cdot L^{-1}$	7.5	±	1.3	0.7	8.5	7.7	**7.3**	**7.6**
Dissolved oxygen saturation (DO_Sat)	%	97.0	±	15.1	9.6	104.4	100.7	92.3	100.6
Biochemical oxygen demand 5 day (BOD_5)	$mg \cdot L^{-1}$	11.6	±	49.9	0.8	384.0	2.4	**22.2**	**3.4**
Chemical oxygen demand (COD)	$mg \cdot L^{-1}$	98.2	±	195.8	7.9	1036.8	53.8	98.2	–
True color (color)	HU	58.9	±	55.0	12.0	293.0	40.5	57.2	60.3
Alcalinity	$mg \cdot L^{-1}$ $CaCO_3$	47.0	±	39.6	16.2	209.4	35.9	57.8	38.7
Phenophftaleina	$mg \cdot L^{-1}$ $CaCO_3$	0.8	±	3.2	BDL	17.4	0.0	0.6	0.9
Total Hardness	$mg \cdot L^{-1}$ $CaCO_3$	55.7	±	55.6	16.6	421.2	45.2	60.8	51.8
Ca + +	$mg \cdot L^{-1}$	16.9	±	16.2	3.8	119.8	13.8	17.5	16.5
Mg + +	$mg \cdot L^{-1}$	1.5	±	2.2	BDL	9.4	0.5	0.0	2.6
Chloride	$mg \cdot L^{-1}$	10.1	±	13.6	3.2	95.3	6.1	13.3	7.7
Ortophosphate	$mg \cdot L^{-1}$	0.2	±	0.4	BDL	2.2	0.0	**0.3**	**0.1**
Total phosphorus	$mg \cdot L^{-1}$	1.2	±	1.3	BDL	5.4	0.6	0.9	1.4
Ammonium-N	$mg \cdot L^{-1}$	1.0	±	4.0	BDL	26.4	0.1	**2.1**	**0.2**
Nitrate-N	$mg \cdot L^{-1}$	0.3	±	0.2	BDL	1.7	0.3	0.3	0.3
Nitrite-N [1]	$\mu g \cdot L^{-1}$	28	±	54	BDL	365	12	38	21
Total Solids	$mg \cdot L^{-1}$	155.1	±	144.9	29.0	998.0	105.5	176.4	138.7
Total coliforms	MPN.100 mL^{-1}	4.1E + 05	±	3.5E + 01	1.4E + 03	4.3E + 10	1.9E + 05	8.4E + 05	2.3E + 05
	CFU.100 mL^{-1}	7.5E + 04	±	1.7E + 01	3.9E + 02	8.4E + 09	5.6E + 04	1.2E + 05	5.4E + 04
Fecal coliforms	MPN.100 mL^{-1}	1.3E + 05	±	3.1E + 01	4.9E + 02	9.2E + 09	8.6E + 04	2.0E + 05	9.4E + 04
	CFU.100 mL^{-1}	3.1E + 04	±	1.7E + 01	1.1E + 02	4.9E + 08	2.2E + 04	3.6E + 04	2.9E + 04
Mean stream width	m	13.8	±	8.9	0.9	30.5	13.2	12.6	14.6
Mean depth	m	0.6	±	0.4	0.1	1.6	0.6	0.5	0.6
Flow velocity	$m \cdot s^{-1}$	1.1	±	0.5	0.2	2.0	1.1	1.0	1.2
Flow	$m3 \cdot s^{-1}$	13.4	±	16.7	0.0	86.7	9.9	14.6	12.2

[1]Nitrite-N is expressed as $\mu g \cdot L^{-1}$. For its determination the following APHA 4500-NO_2 colorimetric method with a detection value of 2 $\mu g \cdot L^{-1}$ was used.
Descriptive statistics of physicochemical and microbiological variables are given as mean values ± standard deviations, minimums and maximums.
NTU = Nephelometric turbidity units; HU = Hazen units; MPN = Most probable number; CFU = Colony-forming unit.
BDL = Below Detection Limit.
Bolded numbers mean the concentration difference between main organic pollutants.

Table A2
Taxa present in sampling sites in the Tomebamba River during dry and rainy season.

Taxa \ Site ->	Sensitivity Score: ABI Encalda 2011	Tomebamba River (from upstream to downstream)																														Sampling sites with the same taxa
		Dry Season											Rainy Season																				
		TO18	TO17	TO12	TO19	TO15	TO16	TO13	CU04	CU03	CU01	CU02	TO18R	TO17R	TO31R	TO32R	TO34R	TO35R	TO12R	TO37R	TO38R	TO42R	TO43R	TO44R	TO45R	TO46R	TO19R	TO13R	CU03R	CU01R	CU02R	
Water quality ->		2	4	3	5	5	5	4	5	4	5	5	2	3	2	2	4	3	1	2	2	3	3	2	4	2	3	3	3	5	4	
Tubificidae	1	p	p	p	p	p	p	p	p	p	p	p	-	p	p	p	p	p	p	p	p	p	p	p	-	p	p	p	p	p	p	28
Chironomidae	2	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	30
Muscidae	2	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	p	3
Hydrophilidae	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Lymnaeidae	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	1
Physidae	3	-	-	p	-	-	-	p	-	p	p	p	-	-	p	p	p	-	p	p	p	p	p	p	p	p	p	p	p	-	p	20
Psychodidae	3	-	-	p	p	p	p	-	-	p	p	p	-	p	-	p	p	-	p	-	p	p	-	p	p	-	p	p	p	p	-	19
Sphaeriidae	3	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	2
Acari	4	p	-	p	-	-	-	-	-	-	-	-	-	p	p	p	-	-	p	p	-	p	-	-	-	p	p	-	-	-	-	10
Baetidae	4	p	p	p	p	-	-	p	p	p	p	p	p	p	p	p	-	p	p	p	p	p	p	p	p	p	p	p	p	-	p	26
Ceratopogonidae	4	p	-	p	-	-	-	-	-	p	-	-	-	-	p	-	-	-	p	p	p	-	-	-	-	-	p	-	-	-	-	8
Empididae	4	-	-	p	-	-	p	-	-	-	-	-	p	-	-	p	p	p	-	-	p	p	-	-	-	p	-	-	-	-	-	9
Pyralidae	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
Tabanidae	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
Elmidae	5	p	-	p	-	-	-	-	-	-	-	-	p	-	p	p	p	p	p	p	p	p	p	p	p	p	p	-	p	-	p	18
Hydropsychidae	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
Limoniidae	5	p	p	p	-	-	-	p	-	-	-	-	p	p	p	p	p	p	p	p	p	p	-	p	-	-	-	-	-	-	p	16
Psephenidae	5	p	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	3
Scirtidae	5	-	-	-	-	-	-	-	-	p	-	-	p	-	p	p	p	p	p	p	p	-	-	-	p	-	-	-	-	-	-	10
Simuliidae	5	p	p	p	-	-	-	-	p	-	-	-	p	p	p	p	-	p	p	p	p	p	p	p	p	p	p	p	p	p	p	22
Tipulidae	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	p	-	p	-	-	-	p	-	-	-	-	-	-	-	4
Aeshnidae	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
Coenagrionidae	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
Hyallelidae	6	p	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	p	p	p	-	p	p	-	p	-	p	p	-	p	11
Hydroptilidae	6	-	p	-	-	-	-	-	-	-	-	-	p	-	p	p	-	-	p	p	p	-	-	p	-	p	p	p	-	-	-	11
Limnephilidae	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	1
Glossosomatidae	7	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Leptohyphidae	7	p	-	p	-	-	-	-	-	p	-	-	p	p	p	-	-	-	p	p	-	p	-	p	-	p	-	-	-	-	-	11
Hydrobioscidae	8	p	-	-	-	-	-	-	-	-	-	-	p	-	p	p	-	-	p	p	p	p	p	p	-	p	p	p	p	-	-	14
Leptoceridae	8	-	-	-	-	-	-	-	-	-	-	-	-	-	p	p	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	3
Polycentropodidae	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
Blepharoceridae	10	p	-	-	-	-	-	-	-	p	-	-	-	p	-	-	-	-	p	-	p	-	p	p	p	-	p	-	-	-	-	9
Calamoceratidae	10	p	-	-	-	-	-	-	-	-	-	-	-	p	p	p	-	-	p	p	p	-	p	p	-	p	-	-	-	-	-	10
Gripopterygidae	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	1
Helicopsychidae	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
Leptophlebiidae	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
Perlidae	10	-	-	-	-	-	-	-	-	-	-	-	p	-	p	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	3
Dugesiidae	-	p	-	p	-	-	-	p	-	-	-	-	-	-	p	-	p	-	p	-	p	p	-	-	p	p	p	p	-	-	p	13
Glossiphoniidae	-	-	-	-	-	-	p	p	p	p	p	p	-	p	-	p	-	p	p	p	p	-	p	-	p	-	p	p	-	p	p	17
Lumbricidae	-	p	p	p	-	-	-	p	p	-	-	-	-	p	p	-	p	-	p	p	p	p	p	p	p	p	p	p	-	-	p	19
Stenopsychidae	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	1
Total taxa per site →		16	7	14	4	3	5	9	7	10	6	6	11	12	21	16	12	26	10	19	20	15	11	17	10	16	14	12	10	4	12	

p = Taxa present

Biological water quality classification → 1 Very good 2 Good 3 Moderate 4 Deficient 5 Bad

Table A3

Taxa present in sampling sites in the Machangara, Yanuncay and Tarqui Rivers during dry and rainy season (from upstream to downstream).

Site -> / Taxa	Sensitivity Score :ABI Encalda 2011	Machangara River								Yanuncay River												Tarqui River												Sampling sites with the same taxa in the three rivers
		Dry Season					Rainy Season			Dry Season				Rainy Season								Dry Season							Rainy Season						
		MC09	MC10	MC11	MC12	MC13	MC09R	MC11R	MC13R	YA15	YA14	YA16	YA17	YA15R	YA30R	YA34R	YA35R	YA36R	YA37R	YA16R	YA17R	TA05	TA04	TA03	TA23	TA22	TA02	TA01	TA05R	TA04R	TA03R	TA23R	TA01R	
Water quality ->		3	2	2	4	4	4	4	3	4	3	2	4	2	3	4	3	4	3	4	3	4	3	2	4	4	4	4	4	3	2	4	4	
Tubificidae	1	p	-	p	p	p	p	p	p	-	-	p	p	p	-	p	-	p	p	p	p	p	p	p	p	p	p	p	p	-	p	p	p	26
Chironomidae	2	p	p	p	p	p	-	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	-	p	p	p	p	-	p	p	p	p	29
Muscidae	2	-	p	p	p	-	-	-	p	-	-	-	-	-	p	-	-	-	p	-	p	-	-	-	-	p	-	-	-	-	-	-	-	8
Dytiscidae	3	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	2
Hydrophilidae	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	p	-	2
Lymnaeidae	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	2
Physidae	3	-	p	p	p	-	-	p	-	-	-	p	-	p	p	p	p	p	p	p	p	p	p	-	p	p	-	-	p	p	-	p	p	21
Psychodidae	3	-	p	-	p	p	p	-	p	-	-	-	p	-	-	-	p	p	p	p	p	-	p	-	-	-	p	-	-	-	-	p	p	15
Sphaeriidae	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	p	-	-	p	p	p	-	-	-	-	-	p	p	-	p	p	9
Acari	4	-	-	-	-	-	-	p	-	-	p	p	-	p	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	5
Baetidae	4	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	-	-	p	-	-	28
Ceratopogonidae	4	-	-	p	-	-	-	-	p	p	p	p	p	p	-	-	-	p	p	-	-	-	-	p	-	-	p	-	-	-	p	-	-	12
Dolichopodidae	4	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Empididae	4	-	p	p	-	-	-	-	-	-	-	p	-	p	-	-	-	-	p	-	-	-	p	-	-	-	-	p	p	p	-	-	-	9
Pyralidae	4	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Tabanidae	4	-	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Elmidae	5	p	p	p	-	-	-	p	p	p	p	p	p	p	p	-	-	-	-	-	-	-	-	p	-	-	p	-	-	p	p	p	-	16
Hydropsychidae	5	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Limoniidae	5	-	p	p	-	p	-	p	p	-	-	p	-	p	p	-	-	-	p	-	p	-	p	p	-	-	p	-	-	p	p	-	p	15
Scirtidae	5	-	p	-	-	-	-	-	-	-	p	-	-	p	p	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	p	-	-	6
Simuliidae	5	p	p	-	p	-	p	p	p	p	p	p	-	p	-	-	p	p	p	p	p	p	p	p	-	-	-	-	p	p	p	p	-	22
Tipulidae	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	p	-	-	-	-	2
Aeshnidae	6	-	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	p	-	-	-	-	3
Coenagrionidae	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	1
Hyallelidae	6	-	p	-	p	-	-	-	-	p	-	p	p	-	p	p	p	-	p	p	p	p	p	-	p	-	-	p	p	p	p	p	p	20
Hydroptilidae	6	p	p	-	-	-	-	p	p	-	-	-	-	-	-	-	-	-	p	-	p	-	-	-	p	-	-	-	-	p	p	-	-	9
Limnephilidae	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
Leptohyphidae	7	-	-	-	-	-	-	-	-	p	p	p	-	p	p	-	p	-	-	-	p	-	p	p	-	-	-	-	-	-	p	-	-	10
Hydrobioscidae	8	p	p	p	-	p	-	-	p	-	p	p	-	-	-	-	-	-	-	-	-	-	p	p	-	-	-	-	-	p	p	-	-	11
Leptoceridae	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
Polycentropodidae	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	p	-	-	2
Blepharoceridae	10	p	-	p	-	-	-	-	-	-	-	p	p	p	p	-	p	-	p	-	-	-	-	-	-	-	-	-	-	-	p	-	-	9
Calamoceratidae	10	-	-	-	-	-	-	-	-	-	-	-	-	p	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	3
Gripopterygidae	10	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Helicopsychidae	10	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Leptophlebiidae	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	2
Perlidae	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	1
Dugesiidae	-	-	-	-	p	-	-	p	-	-	-	-	-	-	-	-	p	p	-	p	p	p	p	-	p	p	p	p	p	-	p	p	p	16
Glossiphoniidae	-	-	-	-	p	-	-	-	-	-	p	-	-	-	-	-	-	-	-	p	p	-	-	-	p	-	-	p	p	p	-	p	p	10
Lumbricidae	-	p	p	p	-	p	-	-	p	p	-	p	p	p	p	p	p	p	p	p	p	p	p	-	-	-	p	-	-	p	p	-	p	22
Total taxa per point →		9	16	14	10	7	5	10	13	8	10	15	9	17	12	6	13	10	14	11	16	10	15	11	8	8	9	7	10	12	19	10	10	

p = Taxa present

Biological water quality classification → 1 Very good | 2 Good | 3 Moderate | 4 Deficient | 5 Bad

Table A4
Results of the variables difference between seasons.

Parameter	Normality tests		Paired student test			Paired wilcoxon test		
	Levene test	Shapiro test	Two sided	Greater during		Two sided	Greater during	
				rainy season	dry season		rainy season	dry season
Biological parameters								
Andean Biotic Index (ABI)	0.59	0.99	0.10	**0.05**	0.95			
Total number of taxa	0.28	0.80	**0.04**	**0.02**	0.98			
Ephemeroptera – Plecoptera – Trichoptera (EPT) taxa	0.06	**0.02**				0.33	0.17	0.85
Number of sensitive taxa NST	0.82	0.31	0.08	**0.04**	0.96			
Shannon-Wiener index (SWD)	0.86	0.45	**0.01**	< **0.01**	1.00			
Mean Tolerance Sore (MTS)	0.72	0.74	0.62	0.31	0.69			
Physicochemical parameters								
Temperature	1.00	0.39	< **0.01**	< **0.01**	1.00			
Conductivity	0.67	< **0.01**				0.36	0.18	0.83
pH	0.33	0.47	< **0.01**	1.00	< **0.01**			
Turbidity	0.26	0.63	0.20	0.10	0.90			
Dissolved oxygen (DO)	0.94	0.29	< **0.01**	1.00	< **0.01**			
Oxygen saturation (OS)	0.96	0.85	0.20	0.10	0.90			
Five-day biological oxygen demand (BOD_5)	0.69	0.01				0.37	0.19	0.82
True color (color)	0.19	0.05	0.11	0.06	0.94			
Alkalinity	0.75	0.14	0.39	0.81	0.20			
Chloride	0.87	0.01				0.23	0.12	0.89
Orthophosphate	0.41	< **0.01**				0.51	0.25	0.76
Ammonium-N	0.81	**0.05**				0.79	0.62	0.40
Nitrate-N	0.40	< **0.01**				0.12	0.06	0.94
Nitrite-N	0.10	**0.04**				< **0.01**	< **0.01**	1.00
Total solids	0.24	0.30	**0.03**	**0.02**	0.99			
Log Fecal Coliforms (MPN/100 mL)	0.79	0.54	0.31	0.15	0.85			
Flow velocity	0.65	0.94	0.70	0.35	0.65			

Bolded numbers have a p-value < 0.05.

Table A5
Best model outcomes: dry season.

Explanatory variables	Regression Parameters	Models obtained from the dry-season dataset (without outliers)							
		Complete data Inverse Gaussian model: mD1.inverse		Fold-1 Inverse Gaussian model: mD9.3fcv1a2.inverse		Fold-2 Inverse Gaussian model: mD6.3fcv1a3.gamma		Fold-3 Gamma model: mD1.3fcv2a3.gamma	
		Coefficient	p-Values	Coefficient	p-Values	Coefficient	p-Values	Coefficient	p-Values
	A	1.9E−04	0.15	−4.4E−04	0.09	1.6E−02	< 0.01	1.3E−02	< 0.01
Nitrate	B1								
Nitrite	B2								
Ammonium	B3					−2.6E−02	0.04	−4.2E−02	0.03
BOD_5	B4	1.7E−04	< 0.01	1.9E−04	0.06	3.4E−03	0.04	5.2E−03	< 0.01
DO	B5								
Oxygen saturation	B6								
Log Fecal Coliforms	B7								
Orthophosphate	B8	4.9E−03	< 0.01			1.2E−01	0.02	1.4E−01	< 0.01
Chloride	B9			1.2E−04	0.02				
Total solids	B10								
Turbidity	B11								
Velocity	B12								
pH	B13								
Conductivity	B14								
Alkalinity	B15								
True color	B16								
Bank material	B17								
AIC:		198.2		126.8		134.4		140.6	
Training subset (2/3)									
Pseudo R^2:		0.7		0.7		0.5		0.7	
CCI:		57.7%		58.8%		52.9%		72.2%	
κ:		0.4		0.4		0.4		0.6	
Validation subset (1/3)									
Pseudo R^2:				9.4E−04		**0.8**		0.4	
CCI:				44.4%		**55.6%**		**62.5%**	
κ:				0.25		**0.4**		**0.4**	

Table A6
Best model outcomes: rainy season.

Explanatory variables	Regression parameters	Models obtained from the rainy-season dataset (without outliers)							
		Complete data Gaussian model mR8.gaussian		Fold-1 Gaussian model mr29.3fcv1a2. gaussian		Fold-2 Gaussian model mR07.3fcv1a3. gaussian		Fold-3 Inverse Gaussian model mR2.3fcv2a3. inverse	
		Coefficient	p-Values	Coefficient	p-Values	Coefficient	p-Values	Coefficient	p-Values
	A	−725.5	< 0.01	−641.7	0.01	−865.4	0.01	8.7E−03	< 0.01
Nitrate	B1	85.3	< 0.01			68.9	0.02		
Nitrite	B2	−1100.7	< 0.01			−6.4E−01	0.07		
Ammonium	B3			−32.8	0.08				
BOD_5	B4								
DO	B5	−42.8	0.04			−52.4	0.04		
Oxygen saturation	B6	10.9	< 0.01	7.1	0.03	13.1	0.01	−8.3E−05	< 0.01
Log Fecal Coliforms	B7								
Orthophosphate	B8								
Chloride	B9								
Total solids	B10	−0.1	0.03			−1.2E−01	0.04		
Turbidity	B11			−0.5	0.03				
Velocity	B12			−27.8	0.04				
pH	B13								
Conductivity	B14								
Alkalinity	B15								
True color	B16								
Bank material	B17	10.6	0.02	12.6	0.03				
AIC:		314.0		218.2		202.8		206.7	
Training subset (2/3)									
Pseudo R^2:		0.6		0.6		0.6		0.3	
CCI:		66.7%		62.5%		56.5%		47.8%	
κ:		0.5		0.4		0.3		0.2	
Validation subset (1/3)									
Pseudo R^2:		0.3		0.1		0.4		0.6	
CCI:		**72.7%**		63.6%		50.0%		50.0%	
κ:		**0.6**		0.5		0.3		0.3	

Table A7
Best model outcomes with the complete dataset.

Explanatory variables	Regression Parameters	Models obtained from both season with the complete dataset							
		Complete data Gamma model: m03T.gamma		Fold-1 Gamma model m2.3fcv1a2T.gamma		Fold-2 Inverse Gaussian model m5T.3fcv1a3T.inverse		Fold-3 Gamma model m03.3fcv2a3T.gamma	
		Coefficient	p-Values	Coefficient	p-Values	Coefficient	p-Values	Coefficient	p-Values
	A	2.0E−01	0.08	−1.29E−02	0.09	2.01E−04	< 0.01	3.0E−01	< 0.01
Nitrate	B1								
Nitrite	B2								
Ammonium	B3	−7.4E−03	< 0.01					−9.4E−03	< 0.01
BOD_5	B4					−2.3E−05	< 0.01		
DO	B5	1.2E−02	< 0.01					1.8E−02	0.01
Oxygen saturation	B6	−2.7E−03	< 0.01					−4.1E−03	< 0.01
Log Fecal Coliforms	B7								
Orthophosphate	B8	7.5E−02	< 0.01			8.4E−03	< 0.01	6.7E−02	0.01
Chloride	B9			3.4E−03	< 0.01				
Total solids	B10								
Turbidity	B11								
Velocity	B12			1.1E−02	0.02				
pH	B13			−1.29E−02	0.09	2.01E−04	< 0.01		
Conductivity	B14								
Alkalinity	B15								
True color	B16								
Bank material	B17								
AIC:		527.0		356.1		353.9		353.6	
Training subset (2/3)									
Pseudo R^2:		0.5		0.5		0.5		0.5	
CCI:		53%		61.9%		43.9%		48.8%	
κ:		0.3		0.5		0.2		0.3	
Validation subset (1/3)									
Pseudo R^2:				0.1		0.3		**0.5**	
CCI:				40.0%		47.6%		**61.9%**	
κ:				0.2		0.3		**0.5**	

Table A8
Best model outcomes without-outliers dataset.

Explanatory variables	Regression parameters	Models obtained from both season without-outliers dataset							
		Complete data Gamma model m2.gamma		Fold-1 Gamma model m3.3fcv1a2.gamma		Fold-2 Inverse Gaussian model m2.3fcv1a3.inverse		Fold-3 Gamma model m1.3fcv2a3.gamma	
		Coefficient	p-Values	Coefficient	p-Values	Coefficient	p-Values	Coefficient	p-Values
	A	1.4E−01	0.02	−1.2E−02	0.11	1.5E−04	0.06	2.2E−01	0.01
Nitrate	B1								
Nitrite	B2								
Ammonium	B3	−1.3E−02	0.03						
BOD_5	B4	2.1E−03	< 0.01					1.5E−03	0.04
DO	B5	1.4E−02	< 0.01					2.0E−02	< 0.01
Oxygen saturation	B6	−2.3E−03	< 0.01					−3.6E−03	< 0.01
Log Fecal Coliforms	B7								
Orthophosphate	B8	6.1E−02	< 0.01			8.1E−03	< 0.01	3.3E−02	0.07
Chloride	B9			3.4E−03	< 0.01				
Total solids	B10								
Turbidity	B11								
Velocity	B12			1.1E−02	0.03				
pH	B13								
Conductivity	B14								
Alkalinity	B15								
True color	B16								
Bank material	B17								
AIC:		513.7		347.6		343.9		347.4	
Training subset (2/3)									
Pseudo R^2:		0.6		0.6		0.5		0.6	
CCI:		57.4%		63.4%		45.0%		53.7%	
κ:		0.4		0.5		0.2		0.3	
Validation subset (1/3)									
Pseudo R^2:				0.1		0.3		**0.6**	
CCI:				45.0%		52.4%		**60.0%**	
κ:				0.2		0.3		**0.4**	

Table A9
Best model outcomes with the main-rivers dataset.

Explanatory variables	Regression parameters	Models obtained from both season with the main-rivers dataset							
		Complete data Gamma model: m6um.gamma		Fold-1 Gaussian model: m2um3fcv1a2.gaussian		Fold-2 Gamma model: m13um3fcv1a3.gamma		Fold-3 Gaussian model m13um3fcv.gaussian23	
		Coefficient	p-Values	Coefficient	p-Values	Coefficient	p-Values	Coefficient	p-Values
	A	2.6E−01	0.02	−764.7	< 0.01	2.4E−01	0.01	117.5	< 0.01
Nitrate	B1							−84.6	0.05
Nitrite	B2								
Ammonium	B3								
BOD_5	B4								
DO	B5	1.2E−02	0.02	−30.7	0.03	1.3E−02	0.06		
Oxygen saturation	B6	−3.4E−03	0.02	10.5	< 0.01	−3.2E−03	0.01		
Log Fecal Coliforms	B7								
Orthophosphate	B8	9.5E−02	< 0.01	−88.6	0.04	1.2E+01	< 0.01	264.1	0.06
Chloride	B9								
Total solids	B10								
Turbidity	B11							−9.2E−01	< 0.01
Velocity	B12								
pH	B13								
Conductivity	B14							−4.7E−01	0.03
Alkalinity	B15								
True color	B16								
Bank material	B17								
AIC:		378.5		273.4		247.0		254.3	
Training subset (2/3)									
Pseudo R^2:		0.5		0.5		0.5		0.5	
CCI:		52.3%		53.3%		55.2%		62.1%	
κ:		0.3		0.3		0.4		0.5	
Validation subset (1/3)									
Pseudo R^2:				0.4		0.1		0.1	
CCI:				**57.1%**		40.0%		40.0%	
κ:				**0.4**		0.1		0.2	

References

Aereopuerto-Mariscal-Lamar (2012) Información meteorológica aereopuerto Mariscal Lamar Cuenca. In. Dirección de aviación civil del Ecuador, Quito – Ecuador.

Agresti, A., Kateri, M., 2011. Categorical Data Analysis. Springer.

Álvarez, L.F. (2005) Metodología para la utilización de los macroinvertebrados acuáticos como indicadores de la calidad del agua, Bogota – Colombia.

Anderson, M.J., 2006. Distance-based tests for homogeneity of multivariate dispersions. Biometrics 62, 245–253.

Armitage, P., Moss, D., Wright, J., Furse, M., 1983. The performance of a new biological water quality score system based on macroinvertebrates over a wide range of unpolluted running-water sites. Water Res. 17, 333–347.

Arunachalam, M., Nair, K.M., Vijverberg, J., Kortmulder, K., Suriyanarayanan, H., 1991. Substrate selection and seasonal variation in densities of invertebrates in stream pools of a tropical river. Hydrobiologia 213, 141–148.

Attrill, M.J., Rundle, S.D., Thomas, R.M., 1996. The influence of drought-induced low freshwater flow on an upper-estuarine macroinvertebrate community. Water Res. 30, 261–268.

Beauchard, O., Gagneur, J., Brosse, S., 2003. Macroinvertebrate richness patterns in North African streams. J. Biogeogr. 30, 1821–1833.

Beyene, A., Legesse, W., Triest, L., Kloos, H., 2009. Urban impact on ecological integrity of nearby rivers in developing countries: the Borkena River in highland Ethiopia. Environ. Monit. Assess. 153, 461.

Booth, G.D., Niccolucci, M.J. & Schuster, E.G. (1994) Identifying proxy sets in multiple linear regression: an aid to better coefficient interpretation. Research paper INT (USA).

Burneo, P.C., Gunkel, G., 2003. Ecology of a high Andean stream, Rio Itambi, Otavalo, Ecuador. Limnol. Ecol. Manage. Inland Waters 33, 29–43.

Cairns, J., Pratt, J.R., 1993. A history of biological monitoring using benthic macroinvertebrates. Freshw. Biomonitor. Benthic Macroinvertebrates 10, 27.

Cameron, A.C., Windmeijer, F.A., 1997. An R-squared measure of goodness of fit for some common nonlinear regression models. J. Econom. 77, 329–342.

Chatterjee, S., Hadi, A.S., 1986. Influential observations, high leverage points, and outliers in linear regression. Stat. Sci. 379–393.

Cohen, J., 1960. A Coefficient of agreement for nominal scales. Educ. Psychol. Measur. 20, 37–46.

Cordero Domínguez, I.R., 2013. Evaluación de la gestión territorial de la cuenca del río Paute, estrategias y líneas de acción para superarlas. Master Thesis. Universidad de Cuenca, Cuenca – Ecuador.

Damanik-Ambarita, M.N., Everaert, G., Forio, M.A.E., Nguyen, T.H.T., Lock, K., Musonge, P.L.S., Suhareva, N., Dominguez-Granda, L., Bennetsen, E., Boets, P., 2016. Generalized linear models to identify key hydromorphological and chemical variables determining the occurrence of macroinvertebrates in the guayas river basin (Ecuador). Water 8, 297.

De Pauw, N., Gabriels, W., Goethals, P.L., 2006. River monitoring and assessment methods based on macroinvertebrates. In: Biological Monitoring of Rivers. John Wiley and Son Ltd, Chichester, UK, pp. 113–134.

Demšar, J., 2006. Statistical comparisons of classifiers over multiple data sets. J. Mach. Learn. Res. 7, 1–30.

Doeg, T., Milledge, G., 1991. Effect of experimentally increasing concentration of suspended sediment on macroinvertebrate drift. Mar. Freshw. Res. 42, 519–526.

Domisch, S., Jaehnig, S.C., Haase, P., 2011. Climate-change winners and losers: stream macroinvertebrates of a submontane region in Central Europe. Freshw. Biol. 56, 2009–2020.

Džeroski, S., Demšar, D., Grbović, J., 2000. Predicting chemical parameters of river water quality from bioindicator data. Appl. Intell. 13, 7–17.

Elith, J., Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. Annu. Rev. Ecol. Evol. Syst. 40.

Encalada, A.C., Sant, M.R., Prat i Fornells, N., 2011. Protocolo Simplificado y guía de Evaluación de la Calidad Ecológica de Ríos andinos (CERA-S). Proyecto FUCARA, Quito – Ecuador.

Erbe, V., Frehmann, T., Geiger, W., Krebs, P., Londong, J., Rosenwinkel, K.-H., Seggelke, K., 2002. Integrated modelling as an analytical and optimisation tool for urban watershed management. Water Sci. Technol. 46, 141–150.

ETAPA-EP, 2007. Evolución de la calidad del agua en los ríos que atraviesan la ciudad de Cuenca. ETAPA-EP Direccion de Manejo Ambiental.

ETAPA-EP, 2018. Cuenca Basin gauging station network – records 2014-2017. In.

Everaert, G., Pauwels, I.S., Boets, P., Verduin, E., de la Haye, M.A.A., Blom, C., Goethals, P.L.M., 2013. Model-based evaluation of ecological bank design and management in the scope of the European Water Framework Directive. Ecol. Eng. 53, 144–152.

Fernandez de Cordova, J., González, H., 2012. Evolución de la calidad del agua de los tramos bajos de los ríos de la ciudad de Cuenca. ETAPA-EP, Cuenca – Ecuador.

Folks, J.L., Chhikara, R.S., 1978. The inverse Gaussian distribution and its statistical application-a review. J. Roy. Stat. Soc. Ser. B Methodol. 263–289.

Forbes, C., Evans, M., Hastings, N., Peacock, B., 2011. Statistical Distributions. John Wiley & Sons.

Forio, M.A.E., 2017. Statistical Analysis of Stream Invertebrate Traits in Relation to River Conditions. Ghent University.

Forio, M.A.E., Lock, K., Radam, E.D., Bande, M., Asio, V., Goethals, P., 2017. Assessment and analysis of ecological quality, macroinvertebrate communities and diversity in rivers of a multifunctional tropical island. Ecol. Ind. 77, 228–238.

Forio, M.A.E., Goethals, P.L., Lock, K., Asio, V., Bande, M., Thas, O., 2018. Model-based analysis of the relationship between macroinvertebrate traits and environmental river conditions. Environ. Modell. Softw. 106, 57–67.

Forio, M.A.E., Van Echelpoel, W., Dominguez-Granda, L., Mereta, S.T., Ambelu, A., Hoang, T.H., Boets, P., Goethals, P.L.M., 2016. Analysing the effects of water quality on the occurrence of freshwater macroinvertebrate taxa among tropical river basins from different continents. AI Commun. 29, 665–685.

Friberg, N., Skriver, J., Larsen, S.E., Pedersen, M.L., Buffagni, A., 2010. Stream macroinvertebrate occurrence along gradients in organic pollution and eutrophication. Freshw. Biol. 55, 1405–1419.

Fukuda, S., De Baets, B., Mouton, A.M., Waegeman, W., Nakajima, J., Mukai, T., Hiramatsu, K., Onikura, N., 2011. Effect of model formulation on the optimization of a genetic Takagi-Sugeno fuzzy system for fish habitat suitability evaluation. Ecol. Model. 222, 1401–1413.

Gabriels, W., Lock, K., De Pauw, N., Goethals, P.L., 2010. Multimetric macroinvertebrate index flanders (MMIF) for biological assessment of rivers and lakes in Flanders (Belgium). Limnol. Ecol. Manage. Inland Waters 40, 199–207.

Gabriels, W., Goethals, P.L., Dedecker, A.P., Lek, S., De Pauw, N., 2007. Analysis of macrobenthic communities in Flanders, Belgium, using a stepwise input variable selection procedure with artificial neural networks. Aquat. Ecol. 41, 427–441.

Goethals, P., 2005. Data Driven Development of Predictive Ecological Models for Benthic Macroinvertebrates in Rivers. Ghent University, Ghent – Belgium.

Goethals, P.L., Dedecker, A.P., Gabriels, W., Lek, S., De Pauw, N., 2007. Applications of artificial neural networks predicting macroinvertebrates in freshwaters. Aquat. Ecol. 41, 491–508.

Gray, L.J., Ward, J.V., 1982. Effects of sediment releases from a reservoir on stream macroinvertebrates. Hydrobiologia 96, 177–184.

Guisan, A., Lehmann, A., Ferrier, S., Austin, M., Overton, J., Aspinall, R., Hastie, T., 2006. Making better biogeographical predictions of species' distributions. J. Appl. Ecol. 43, 386–392.

Hardin, J.W., Hilbe, J.M., 2013. Generalized Estimating Equations, Second Edition ed. CRC Press Taylor & Francis Group, Boca Raton, FL.

Heino, J., Mykrä, H., Hämäläinen, H., Aroviita, J., Muotka, T., 2007. Responses of taxonomic distinctness and species diversity indices to anthropogenic impacts and natural environmental gradients in stream macroinvertebrates. Freshw. Biol. 52, 1846–1861.

Hering, D., Feld, C.K., Moog, O., Ofenböck, T., 2006. Cook book for the development of a Multimetric Index for biological condition of aquatic ecosystems: experiences from the European AQEM and STAR projects and related initiatives. Hydrobiologia 566, 311–324.

Holguin-Gonzalez, J.E., Everaert, G., Boets, P., Galvis, A., Goethals, P.L.M., 2013a. Development and application of an integrated ecological modelling framework to analyze the impact of wastewater discharges on the ecological water quality of rivers. Environ. Modell. Softw. 48, 27–36.

Holguin-Gonzalez, J.E., Boets, P., Alvarado, A., Cisneros, F., Carrasco, M.C., Wyseure, G., Nopens, I., Goethals, P.L.M., 2013b. Integrating hydraulic, physicochemical and ecological models to assess the effectiveness of water quality management strategies for the River Cuenca in Ecuador. Ecol. Model. 254, 1–14.

Holzer, P., Krebs, P., 1998. Modelling the total ammonia impact of CSO and WWTP effluent on the receiving water. Water Sci. Technol. 38, 31–39.

Hynes, H., 1960. BN (1960)'.The Biology of Polluted Waters'. Liverpool UP.

INEC (2010) Proyección de la Población Ecuatoriana, por años calendario, según cantones 2010-2020. In. Instituto Nacional de Estadísticas y Censos del Ecuador, Quito – Ecuador.

Jacobsen, D., 1998. The effect of organic pollution on the macroinvertebrate fauna of Ecuadorian highland streams. Arch. Hydrobiol. 143, 179–195.

Jacobsen, D., Encalada, A., 1998. The macroinvertebrate fauna of Ecuadorian highland streams in the wet and dry season. Arch. Hydrobiol. 142, 53–70.

Jacobsen, D., Marín, R., 2008. Bolivian Altiplano streams with low richness of macroinvertebrates and large diel fluctuations in temperature and dissolved oxygen. Aquat. Ecol. 42, 643–656.

Jerves-Cobo, R., Lock, K., Van Butsel, J., Pauta, G., Cisneros, F., Nopens, I., Goethals, P.L.M., 2018a. Biological impact assessment of sewage outfalls in the urbanized area of the Cuenca River basin (Ecuador) in two different seasons. Limnologica 71, 8–28.

Jerves-Cobo, R., Everaert, G., Iñiguez-Vela, X., Córdova-Vela, G., Díaz-Granda, C., Cisneros, F., Nopens, I., Goethals, P.L., 2017. A methodology to model environmental preferences of EPT taxa in the machangara river basin (Ecuador). Water 9, 195.

Jerves-Cobo, R., Córdova-Vela, G., Iñiguez-Vela, X., Díaz-Granda, C., Van Echelpoel, W., Cisneros, F., Nopens, I., Goethals, P., 2018b. Model-based analysis of the potential of macroinvertebrates as indicators for microbial pathogens in rivers. Water 10, 375.

Junqueira, V., Campos, S., 1998. Adaptation of the "BMWP" method for water quality evaluation to Rio das Velhas watershed (Minas Gerais, Brazil). Acta Limnol. Bras. 10, 125–135.

Kelso, B.H., Smith, R.V., Laughlin, R.J., 1999. Effects of carbon substrates on nitrite accumulation in freshwater sediments. Appl. Environ. Microbiol. 65, 61–66.

Kohavi, R., Provost, F., 1998. Glossary of terms. Mach. Learn. 30, 271–274.

Lovison, G., Sciandra, M., Tomasello, A., Calvo, S., 2011. Modeling Posidonia oceanica growth data: from linear to generalized linear mixed models. Environmetrics 22, 370–382.

Maillard, P., Santos, N.A.P., 2008. A spatial-statistical approach for modeling the effect of non-point source pollution on different water quality parameters in the Velhas river watershed–Brazil. J. Environ. Manage. 86, 158–170.

McCullagh, P., 1984. Generalized linear models. Eur. J. Oper. Res. 16, 285–292.

Mereta, S.T., Boets, P., Bayih, A.A., Malu, A., Ephrem, Z., Sisay, A., Endale, H., Yitbarek, M., Jemal, A., De Meester, L., 2012. Analysis of environmental factors determining the abundance and diversity of macroinvertebrate taxa in natural wetlands of Southwest Ethiopia. Ecol. Inf. 7, 52–61.

Meybeck, M., Laroche, L., Dürr, H., Syvitski, J., 2003. Global variability of daily total

suspended solids and their fluxes in rivers. Global Planet. Change 39, 65–93.
Mortimer, C.H., 1981. The oxygen content of air-saturated fresh waters over ranges of temperature and atmospheric pressure of limnological interest: with 6 figures and 1 table in the text and on 1 folder, and 4 appendices. Int. Vereinig. Theoret. Angew. Limnol. Mitteilungen 22, 1–23.
Mouton, A.M., Van Der Most, H., Jeuken, A., Goethals, P.L., De Pauw, N., 2009. Evaluation of river basin restoration options by the application of the water framework directive explorer in the Zwalm River basin (Flanders, Belgium). River Res. Appl. 25, 82–97.
Moya, N., Hughes, R.M., Domínguez, E., Gibon, F.-M., Goitia, E., Oberdorff, T., 2011. Macroinvertebrate-based multimetric predictive models for evaluating the human impact on biotic condition of Bolivian streams. Ecol. Ind. 11, 840–847.
Mustow, S., 2002. Biological monitoring of rivers in Thailand: use and adaptation of the BMWP score. Hydrobiologia 479, 191–229.
Palmer, M.A., Bernhardt, E., Allan, J., Lake, P., Alexander, G., Brooks, S., Carr, J., Clayton, S., Dahm, C., Shah, J.F., 2005. Standards for ecologically successful river restoration. J. Appl. Ecol. 42, 208–217.
Philips, S., Laanbroek, H.J., Verstraete, W., 2002. Origin, causes and effects of increased nitrite concentrations in aquatic environments. Rev. Environ. Sci. Biotechnol. 1, 115–141.
Rantz, S.E., 1982. Measurement and Computation of Streamflow: volume 2, computation of discharge. In. USGPO.
Rauch, W., Henze, M., Koncsos, L., Reichert, P., Shanahan, P., Somlyódy, L., Vanrolleghem, P., 1998. River water quality modelling: I. State of the art. Water Sci. Technol. 38, 237–244.
Ríos-Touma, B., Acosta, R., Prat, N., 2014. The Andean Biotic Index (ABI): revised tolerance to pollution values for macroinvertebrate families and index performance evaluation. Rev. Biol. Tropical 62, 249–273.
Ríos-Touma, B., Encalada, A.C., Prat Fornells, N., 2011. Macroinvertebrate assemblages of an andean high-altitude tropical stream: the importance of season and flow. Int. Rev. Hydrobiol. 96, 667–685.
Roldán Pérez, G.A., 1988. Guía para el estudio de los macroinvertebrados acuáticos del Departamento de Antioquia. Fondo para la Protección del Medio Ambiente "José Celestino Mutis", Bogotá – Colombia.
Roldán Pérez, G.A., 2003. Bioindicación de la Calidad del Agua en Colombia: Uso del Método BMWP/Col. Imprenta Universidad de Antioquia, Medellín, Colombia.
Rousseeuw, P.J., Leroy, A.M., 2005. Robust Regression and Outlier Detection. John Wiley & Sons.
Schleiter, I.M., Borchardt, D., Wagner, R., Dapper, T., Schmidt, K.-D., Schmidt, H.-H., Werner, H., 1999. Modelling water quality, bioindication and population dynamics in lotic ecosystems using neural networks. Ecol. Model. 120, 271–286.
SENPLADES (2016) Proyecciones referenciales de población a nivel de distritos de planificación: 2010–2020. In. Secretaría Nacional de Planificación y Desarrollo Ecuador, Subsecretaría de Información, Dirección de Normas y Metodología, Quito – Ecuador.
Sharp, W., 1971. A topologically optimum water-sampling plan for rivers and streams. Water Resour. Res. 7, 1641–1646.
Shmueli, G., 2010. To explain or to predict? Stat. Sci. 25, 289–310.
Stevens, J.P., 1984. Outliers and influential data points in regression analysis. Psychol. Bull. 95, 334.
Stockwell, D.R., Peterson, A.T., 2002. Effects of sample size on accuracy of species distribution models. Ecol. Model. 148, 1–13.
Strobl, R.O., Robillard, P.D., 2008. Network design for water quality monitoring of surface freshwaters: a review. J. Environ. Manage. 87, 639–648.
Thode, H.C., 2002. Testing for Normality. CRC Press.
Thorne, R., Williams, P., 1997. The response of benthic macroinvertebrates to pollution in developing countries: a multimetric system of bioassessment. Freshw. Biol. 37, 671–686.
Vaughan, I.P., Ormerod, S.J., 2005. The continuing challenges of testing species distribution models. J. Appl. Ecol. 42, 720–730.
Vayssières, M.P., Plant, R.E., Allen-Diaz, B.H., 2000. Classification trees: an alternative non-parametric approach for predicting species distributions. J. Veg. Sci. 11, 679–694.
Venables, W.N., Ripley, B.D., 2002. Tree-based methods. In: Modern Applied Statistics with S. Springer, New York, pp. 251–269.
Ward, R.C. (1973) Data acquisition systems in water quality management. In.
Wilcock, R.J., Battino, R., Danforth, W.F., Wilhelm, E., 1978. Solubilities of gases in liquids II. The solubilities of He, Ne, Ar, Kr, O2, N2, CO, CO2, CH4, CF4, and SF6 in n-octane 1-octanol, n-decane, and 1-decanol. J. Chem. Thermodyn. 10, 817–822.
Wilson, A., Watts, R., Stevens, M., 2008. Effects of different management regimes on aquatic macroinvertebrate diversity in Australian rice fields. Ecol. Res. 23, 565–572.
Zúñiga, M.D.C., Cardona, W., Cantera, J., Carvajal, Y. & Castro, L. (2009) Bioindicadores de calidad de agua y caudal ambiental. Caudal ambiental: Conceptos Experiencias y Desafíos. Universidad del Valle: Cali, Colombia, 1, pp. 303–310.
Zuur, A., Ieno, E.N., Walker, N., Saveliev, A.A., Smith, G.M., 2009. Mixed Effects Models and Extensions in Ecology With. R. Springer Science & Business Media.